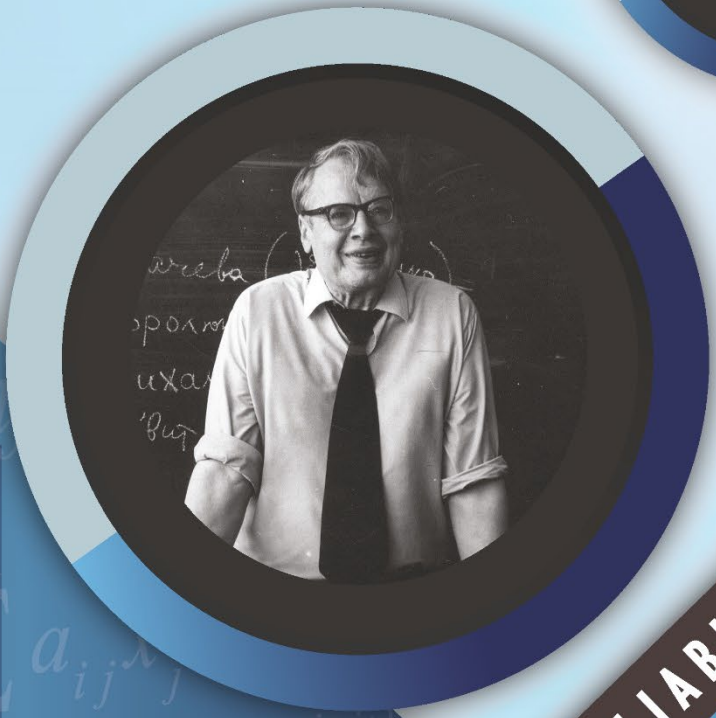
RTA

ISSN 1932-2321

JOURNAL IS REGISTERED IN THE LIBRARY OF THE U.S. CONGRESS

RELIABILITY: THEORY & APPLICATIONS

INTERNATIONAL GROUP ON RELIABILITY



RELIABILITY
RISK ANALYSIS
MAINTENANCE
SAFETY

GNEDENKO FORUM PUBLICATIONS

#4

(80) VOL.19

DECEMBER

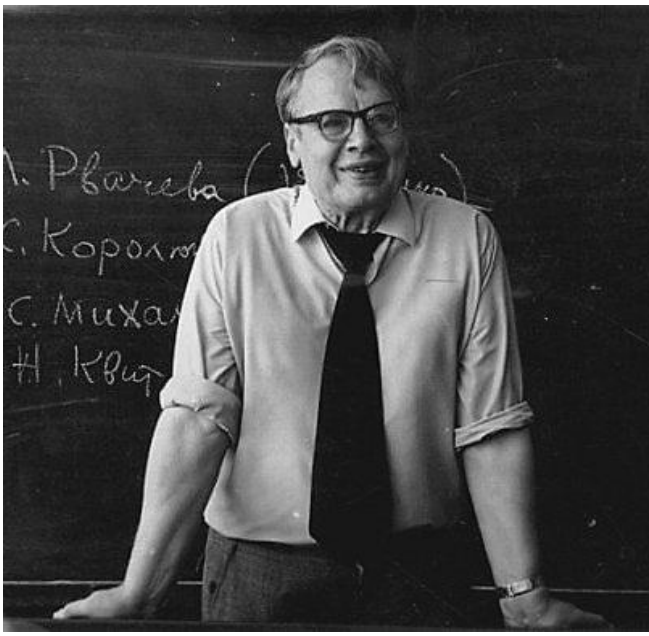
2024

SAN DIEGO

ISSN 1932-2321

http://www.gnedenko.net/Journal/index.htm





RELIABILITY:

THEORY & APPLICATIONS

Vol.19 No.4 (80),
December 2024

San Diego
2024

Editorial Board

e-Journal ***Reliability: Theory & Applications*** publishes papers, reviews, memoirs, and bibliographical materials on Reliability, Quality Control, Safety, Survivability and Maintenance.

Theoretical papers must contain new problems, finger practical applications and should not be overloaded with clumsy formal solutions.

Priority is given to descriptions of case studies.
General requirements for presented papers.

1. Papers must be presented in English in MS Word or LaTeX format.
2. The total volume of the paper (with illustrations) can be up to 15 pages.
3. A presented paper must be spell-checked.
4. For those whose language is not English, we kindly recommend using professional linguistic proofs before sending a paper to the journal.

The manuscripts complying with the scope of journal and accepted by the Editor are registered and sent for external review. The reviewed articles are emailed back to the authors for revision and improvement.

The decision to accept or reject a manuscript is made by the Editor considering the referees' opinion and considering scientific importance and novelty of the presented materials. Manuscripts are published in the author's edition. The Editorial Board are not responsible for possible typos in the original text. The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Publication in this e-Journal is equal to publication in other International scientific journals.

Papers directed by Members of the Editorial Boards are accepted without referring.
The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Send your papers to Alexander Bochkov, e-mail: a.bochkov@gmail.com

Table of Contents

or heterogeneous resource availability on cluster nodes. Again, all task containers are treated similarly, irrespective of their specific resource-oriented nature. This paper proposes the weighted resource optimization algorithm for calculating the weighted score of each node. Score depends on CPU and memory weight for a given task and the availability of that resource on the node. The task container is placed on the node with the highest score. This approach improves CPU and memory load balancing in a Docker cluster and also improves the completion time of the task container as compared to the spread strategy.

Muthukrishnan. R, Karthika Ramakrishnan

Aim: To provide a new Liu regression procedure for predictive modeling in cases of multicollinearity and with/without outliers. Methods: Regression analysis is employed in many statistical research domains for both estimation and prediction. Liu and Robust Estimators were developed in a classical linear regression model to address the issues of multicollinearity and outliers, respectively. In order to jointly handle the issues of multicollinearity and outliers, this research paper explores a new Robust Liu regression estimator based on the MM estimator, which is then demonstrated using real and simulated data sets. The performances of various regression estimators such as Least Square, Ridge, Liu and the Robust Liu are compared based on the Mean Square Error criterion. Findings: According to the computed error measure, the study concludes that the Robust Liu regression estimator provides more reliable results than the other mentioned regression procedures in situations where datasets have both multicollinearity and outliers.

Farzaliyev Y.Z., Farhadzadeh E.M.

This article deals with economic aspects, i.e. identification of reserves of thermal efficiency of obsolete equipment in the example of power units of thermal power plants, which have a useful life exceeding 50%. As a result of operation of such equipment, useful heat required for power generation is lost. The developed new approach allows to detect in time those reserves, which are not possible with the use of energy characteristics due to wear and tear of the equipment and in the end these reserves will remain latent. With the help of the new approach when comparing it with the intuitive one, by which the technical staff wastes more time, it is shown that by taking into account the actual technical condition, reliability and efficiency of equipment operation it is possible to achieve the desired result. The results showed themselves brilliantly when distributing the load between power units of a thermal power station. The exploitation data for solving the problem are technical and economic indicators that characterize the wear and tear of the equipment

A.R. Gokul, M. Pachamuthu

Robust missing observations have emerged as a crucial study area in statistical research. Response Surface Methodology (RSM), a recognized and extensively utilized area in experimental design, has determined that the absence of observations in an experiment can introduce complexity and interfere with the estimation of parameters. Previous literature reviews reveal that most studies on missing Central Composite Design (CCD) data were conducted using optimality and minimax loss criteria. Our study explores the spherical region of interest in the missing observation of CCD, represented through Variance Dispersion Graph (VDG) and Fraction of Design Space (FDS) graphs. Practitioners primarily focus on the region of interest rather than employing various alpha values. We investigate the predictive capabilities of each factorial, axial, and center missing design point against different radii(r) and fractions of the design space region, and we also measure relative G- and D- efficiency. We scrutinize various factors (k) from two to seven, including five center runs. Our research explores the region of interest in operating the experiment under robust conditions through visual aids of VDG and FDS graphs.

algorithm. Additionally, a comparison is provided. A numerical example for comparison analysis with previous publications is provided, utilising appropriate graphical layout and tables, to elucidate both approaches.

Survival analysis is one of the key techniques utilized in the domains of reliability engineering, statistics, and medical domains. It focuses on the period between the initialization of an experiment and a subsequent incident. Censoring is one of the key aspects of survival analysis, and the techniques created in this domain are designed to manage various censoring schemes with ease, ensuring accurate and insightful time-to-event data analysis. The statistical efficiency of parameter estimates is improved by accurately incorporating censoring information by making use of the available data. This paper reviews the concepts, model descriptions, and applications of conventional and hybrid censoring schemes. The introduction of new censoring schemes from conventional censoring schemes has evolved by rectifying the drawbacks of the previous schemes, which are explained in detail in this study. The evolution of hybrid censoring schemes through the combination of various conventional censoring schemes, the data structures, concepts, methodology, and existing literature works of hybrid censoring schemes are reviewed in this work.

The aim of the present study is to investigate reliability, availability, maintainability, and dependability (RAMD) of crystallization system of a sugar production plant. Previous studies attentive on the reliability and availability analysis of sugar plants specially its subsystems like evaporation units. This study is focus on the RAMD analysis of the crystallization system of sugar plant having four subsystems with different number of components. Failure and repair rates of all subsystems are taken as exponentially distributed. The transition diagram and Chapman- Kolmogorov differential equations for each subsystem are derived by using Markov birth-death process. For all four subsystems, reliability, availability, mean time between failure (MTBF), mean time to repair (MTTR), and dependability ratio are computed using simple probabilistic concepts. The effect of change in failure rates of subsystem in system performance is also observed. It is shown that the crystallization subsystem found to be more sensitive among four subsystems from reliability point of view. This study can be helpful to system designer for further modeling/designing of reliable systems and enhancement in system's performance through planning efficient maintenance strategies.

Wave loads are critical factor for the design and safe operation of offshore structures. The accurate determination of these loads is essential to ensure the structural reliability and operational efficiency of such platforms at sea. This study develops analytical expressions for calculating wave loadings that affect the support of various Condeep-type offshore structures. In this regard, wave load calculations for the Draugen Monopile Condeep platform, previously constructed in Norway, were analyzed in the context of a case study. The results of this assessment provide useful information regarding the characteristics of wave loads and their relevance to the overall structural analysis. Furthermore, the investigation also covers recommendations for design and safety improvements that consider the calculated wave loads and the assessment of the structural reliability. Study is expected to contribute to the knowledge base surrounding offshore engineering practices and improve resilience and functionality against dynamic wave forces.

optimal conditions, the proposed estimators are not only unbiased but also more efficient than traditional methods, including the sample mean and the double sampling ratio and product type estimators developed by Naik and Gupta [11] and Singh et al. [17]. To further substantiate the theoretical results, we conducted a numerical study, which demonstrates the practical effectiveness of the proposed estimators in improving estimation accuracy.

Upasana Rana, Tanuj Kumar

Enhancing inventory control for perishable goods is challenging since their shelf life is short and their demand is constantly changing. The current research examines into a more advanced inventory model for perishable goods, where demand is affected by both time and the reliability the product. The model occupies a pentagonal-fuzzy environment to assist with inherent uncertainties with these kinds of systems. This gives a more accurate picture of how demand fluctuates over time. Using analytical optimization techniques, the model targets to minimize total inventory costs, consisting of ordering cost, holding cost, and deterioration cost while maintaining high service levels. The total cost function is defuzzified using the Graded Mean Integration Representation (GMIR) method. The study's results, which were verified by numerical evaluations, demonstrate that the model is better at cost reduction and boosting dependability than other models using the MATLAB software. This research contributes a robust framework for handling perishable inventory with uncertain situations, which has a major impact on optimizing the supply chain.

Ashish Negi, Ompal Singh

Inventory control is vital in supply chain management, especially for perishable goods. The paper depicts a probabilistic inventory model for robust products where deterioration and demand change over time and depend on reliability. This paper also talks about the conventional back-order reliability inventory model in a fuzzy, cloudy environment. This is because products deteriorate and demand fluctuates all the time. This study shows a novel approach to modeling inventory that deals with these problems. It does this by including uniform distribution deterioration, demand that depends on both time and product reliability, and cloudy-fuzzy numbers to show uncertainty. Although we start with the crisp model and fuzzifying it to obtain a decision under the cloudy fuzzy demand rate (which is an extension of dense fuzzy) demand rate, before putting it to use in practice. For ranking the fuzzy numbers, a new defuzzification method was used. Subsequently, extensive analysis is done to compare the crisp, general fuzzy solutions to the cloudy fuzzy solutions. The numerical examples and graphical are examined to demonstrate that the novel approach is useful in the model itself. The suggested model aims to maintain high service reliability while minimizing the total cost of inventory. Numerical analyses indicate that the model is effective, exhibiting that it can lower costs and improve reliability compared to older models using MATLAB software. This study builds a strong framework for managing inventory in supply lines for perishable goods, which opens up opportunities for more progress in this area.

Krishan Kumar Yadav, Ajay Singh Yadav, Shikha Bansal

The main objective of this study is to demonstrate how a company's inventory management can be significantly impacted by its ability to provide reliable, high-quality products and to balance stock availability in order to maintain customer satisfaction. Such measures can ultimately lead to an increase in a company's market share, efficiency, and profitability. In order to analyze the impact of reliability and time-based demand rate on inventory management system, an economic order quantity (EOQ) model with two-warehouse is established. Complete backlog allows for the consequences of constant degradation and shortages. The holding and degradation costs are considered while analyzing

the effect of carbon emissions. This study's primary goal is to optimize overall cost while maintaining item reliability and total cycle time. Analytical optimization is used to yield an algorithm for the inventory model that determines the optimal output. A numerical example-based sensitivity analysis using MATLAB Software version R2021b is also presented to illustrate the effect of carbon emission and validation of the model.

Acceptance sampling is a statistical quality control technique used in manufacturing to determine whether to accept or reject a batch of products based on the number of defects obtain in a sample. Among the various sampling plans, the double sampling plan more effective because it often delivers more reliable results in selecting quality lots than other plans. In most of the real-life situation, it is not easy found the product as strictly defective or non-defective. In some situation, quality of the product can be classified several types which are expressed as good, almost good, bad, not so bad and so on. This is causes fuzzy logic comes into play. Fuzzy set theory is most powerful mathematical tool, it can deal incomplete and imprecise information. In this paper Double Sampling Plans (DSPs) are derived when non conformities are said imprecise and these imprecisions are model through ZIP distribution. It analyzes, the effectiveness of these sampling plans by comparing vital metrics such as Average Outgoing Quality (AOQ) and Average Total Inspection (ATI) using both fuzzy and crisp environments. These findings are appraised as both numerically and graphically, showing that whether the process quality is either extremely good or very bad, the AOQ curve will be lower, the plan's able to effectively control product quality.

This paper addresses the problem of designing an acceptance sampling plan for a truncated life test where the lifetime of the product follows a generalized Rayleigh distribution. The study identifies the minimum sample sizes needed to ensure the specified mean life for various acceptance numbers, confidence levels, and ratios of the fixed experiment time to the specified mean life. The operating characteristic values of the sampling plans, along with the producer's risk, are discussed. Additionally, tables are provided to facilitate the application of these sampling plans, and a numerical example is included to illustrate the use of these tables.

The article explores the important role those modern sensors play in ensuring the safety and efficient operation of electrical equipment. Advances in sensor technologies make it possible to effectively monitor the condition of equipment, identify potential problems and prevent accidents. The article examines the principles of operation of sensors, their diversity and application in various areas of the electric power industry. Particular attention is paid to predictive maintenance technologies, which allow optimizing resources and increasing the reliability of electrical equipment. Ultimately, the use of sensors to monitor the condition and safety of electrical equipment leads to reduced risk of accidents and increased efficiency of electrical power systems.

Diagnostics of electrical equipment at thermal power plants plays a key role in ensuring reliable operation of power systems. This article examines methods and technologies for diagnosing electrical equipment at thermal power plants

financial market price, business process, fuel prices and etc., They are particularly relevant in the context of price movements because they provide a framework for understanding and predicting the future state of a system based on its current state. In a Markov chain process, there are a set of states and we progress from one state to another based on a fixed probability. In these decades many articles are showed that modeling a market as a random walk was applicable and that a market may be viewed as having the Markov property. The objective of this paper is to construct the Markov chain model for daily fruit price movement in Salem District, Tamil Nadu. Two models are highlighted, where the price movement is considered as being in a state of gain, loss and no change and large gain, or small gain or loss, or large loss and no change. Ten different types of fruits are considered which are cultivated Salem areas and above two models are used to analyze the price movement of each fruit. These models were used to obtain transitional probabilities, steady state probabilities and mean recurrence times. Our results indicate that the pattern of price movement of Banana is similar to price movements of other fruits, in both models. The investor is encouraged to invest in the fruit market at any time in away which leads to a greater chance of getting more gain than loss.

Vyshnavi. M, Muthukumar. M

This work explores using fuzzified techniques to enhance the performance of Hidden Markov Models (HMMs) in handling uncertainties and imprecise inputs. We construct and evaluate three types of fuzzy HMMs: the Trapezoidal fuzzy HMM, the Sigmoidal fuzzy HMM, and the Gaussian fuzzy HMM. As part of our process, parameter estimations are calculated and models are chosen based on AIC, BIC, AICc, and HQIC criteria. Each state's mean, variance, and stationary distribution are calculated and examined to evaluate the predictability and stability of the models. We use the Viterbi technique to identify the most likely state sequences for the next five years. According to the results, the Gaussian Fuzzy HMM offers superior predicted accuracy and durability when compared to the other models. This paper emphasizes the advantages of using fuzzy membership functions in HMMs and provides the foundation for future research in different areas, such as agricultural data prediction.

Krishan Pal, Ajay Singh Yadav, Seema Agarwal

A multi-warehouse shortage model has been developed where demand is assumed to be deterministic. In reality, machines run for long periods during production and random failures may occur as the system transitions from a controlled to an uncontrolled state. During this time the production system produces defective products. Demand is assumed to be deterministic. Retailers offer a quantity discount per unit on the selling price of an item and in return receive a quantity- based discount on the purchase price of the item. A retailer has limited storage capacity and therefore requires additional space with unlimited storage capacity. This additional space is called a rented warehouse and its storage cost is higher than accompany- owned warehouse. The objective of this model is to study a multiple inventory model of defective items under quantity- based discounts, where defective items can be sorted and sold in a single batch with decision variables set to the optimal order quantity and optimal inventory and shipment quantity to increase overall profits to maximize the value for the retailer. A solution procedure for determining the optimal solution is presented and a numerical example is given to illustrate this study. A sensitivity analysis is also performed to examine the effect of changing parameter values on the optimal solution.

Sasirekha D, Senthilkumar P

Renewable energy provides more environmentally friendly sources of energy, which reduces the demand for fossil fuels and is therefore necessary to reach zero emissions of carbon. But the need for systems that are capable of capturing and storing this energy is expanding as the world gets a growing amount of electricity from these forms of renewable energy.

In present-day society, renewable energy storage is widely used, and governments are concentrating on developing suitable storage technologies together with a plan for upcoming energy storage reduction. Energy storage technologies have been proposed as potential solutions for this issue due to their ability to store energy and lower energy consumption. Aspects of technology, economy, society, and environment are the four main criteria used in this study to examine different energy storage techniques. The most effective strategy was identified in this paper. In this study, we use the ELECTRE-III approach to suggest the optimal storage technology under the linguistic neutrosophic fuzzy set. Finally, a numerical example of this area of study is provided. A comparison and sensitivity analysis are shown for the effectiveness of the proposed method.

The study of robust missing observations has gained prominence in statistical research. In particular, the Response Surface Methodology (RSM), a widely applied approach in experimental design, faces challenges when dealing with missing data. This paper investigates two design variants: the three- level second-order Box-Behnken design (BBD) with one missing observation and the Small Box- Behnken Design (SBBD), which involves fewer experimental runs than the standard BBD. We evaluate prediction performance using a fraction of design space (FDS) plot, revealing the distribution of scaled prediction variance (SPV) values across the design space. Additionally, we assess the efficiency of design model parameters using information-based criteria (A, D, and G relative efficiency). Our analysis spans k factors, ranging from k = 3 to 9. The findings guide practitioners in selecting optimal design points for efficient parameter estimation and accurate prediction within the context of missing observations. This comparative study sheds light on the trade-offs between BBD and SBBD, providing valuable insights for experimental design practitioners.

With the introduction of renewable energy sources, in particular wind power and photovoltaic installations, in the autonomous generation systems, the problem of reliability of the equipment used and the entire energy complex becomes one of the main ones. It is necessary to develop and improve methods for analyzing and calculating reliability, which will make it possible at the design stage to take into account the probabilistic characteristics of renewable energy resources, reliability indicators and operating experience of the equipment used. The article discusses the scheme of an autonomous energy complex based on renewable energy sources. A graph of the dependence of failure rate on recovery time is presented. This paper discusses various methods for assessing the reliability of autonomous generation systems based on renewable energy sources: analytical methods, state space method (Markov process theory), Monte Carlo method, fault tree method and state enumeration method. The advantages and disadvantages of these methods are considered.

In this manuscript, we have introduced a new model of the Kumaraswamy distribution known as SMP Kumaraswamy (SMPK) distribution using SMP technique. The SMPK distribution has the desirable feature of allowing greater flexibility than some of its well-known extensions. A comprehensive account of statistical properties along with the estimation of parameters using classical estimation method is presented. Furthermore, a simulation study is carried out to assess the behavior of estimators based on their biases and mean square errors. Finally, we consider two real-life data sets; we observe that the proposed model outperforms other competing models using goodness of fit measures.

The process of recognizing an underwater object and detecting potentially hazardous underwater object is very important in underwater operations. To facilitate the work of the side scan sonar operator, this paper proposes to increase the reliability of recognizing hydroacoustic images of potentially hazardous underwater objects in automatic mode. Based on the analysis of sonar images received from the side scan sonar, an image of an object is formed, which is then recognized (classified) as belonging to a certain class of objects. Five classes of recognized objects are defined. A convolutional neural network used to determine whether an underwater potentially dangerous object belongs to one of the classes is described. Filters for initial sonar images for acceleration of neural network operation are defined. Algorithms and software for forming an image of the object and making a decision on its belonging to one or another class are developed. It is shown that the use of convolutional neural network allows to determine the correct class of the object with an accuracy of 91%.

Based on the limitations of the Inverse Lomax distribution and exponential distribution as outlined in the literature, a new extension of the exponential distribution is introduced in this paper. Some statistical properties of the ILOEED such as mean, variance, skewness, quantile function, moment, moment generating function, as well as kurtosis were demonstrated. The shapes of the hazard function of the proposed distribution suggest that it can be used to fit a dataset with increasing and bath-tube shapes. A simulation study for three different cases was also presented. The result of the simulation for three different cases (I, II, and III) indicated that ILOEED's estimates are consistent. Lastly, an application to Industry datasets was demonstrated based on the ILOEED. Having minimum values of the Goodness-of-fit criteria and Goodness-of-fit statistics, the ILOEED can be recommended to fit these three datasets, in preference to other distributions considered in this paper.

The study addresses the challenges of estimating the population mean in two-stage cluster sampling, where there is an equal chance of random non-response at the first-stage unit. The researchers propose some regression-type imputation schemes and regression-type estimators that incorporate measurement error parameters for both the study and supplementary variables. The properties of the proposed estimators were derived and numerically compared using a simulated sample population. The proposed estimators outperformed the existing estimators consider in the study. The researchers conclude that their proposed methodology can be practically applied, using the actual responses of the respondents and including the measurement error parameters to estimate the finite population mean.

The concept of Queuing system is most commonly used in our everyday life. It is essential to characterize the practical queuing characteristics in order to improve the performance of the queuing model. This study investigates M/M(a,b)/1/MWV queuing model with heterogeneous encouraged arrival occurring in the regular busy period. The considered model follows General bulk service rule and if the system is not in use, or when it is vacant, the server goes on vacation, thus there occurs multiple working vacations which are exponentially distributed. In this study, a model of multiple working vacation queues in which with heterogeneous encouraged arrivals following Poisson process is examined. With the mentioned conditions, the explicit formulations for the steady state probabilities and the

performance measures of the proposed model are derived. Also, some particular cases have been developed and compared with existing models. Finally, the numerical impact of various parameters on performance attributes are also analysed.

A. Shivaji, B. Harika, D. Rajaiah, L.P. Rajkumar

Based on the Triangular and Skew Symmetric (TSS) splitting method, a novel iterative approach is proposed to solve a class of fuzzy regularized linear system of equations with fuzzy coefficient stochastic rate matrix. The non-homogeneous fully fuzzy linear system is same as the non-homogeneous linear system which is derived from the homogeneous linear system with stochastic rate matrix and steady state vector. An iterative procedure is developed for finding a unique non-trivial solution. Numerical results shown that the proposed method is effective and efficient when compared with the existing classical methods.

Shakuntla Singla, Diksha Mangla, Shilpa Rani, Umar Muhammad Modibbo

The availability of uninterrupted performance time has become essential for any industry seeking to maximize profits while incurring minimal maintenance costs. However, the system's components become weary as a result of the constant burden, resulting in decreased system efficiency and automatic full failure in the end. Complete failure is not always manageable; it might result in a significant loss of profit or productivity. In this regard, preventative maintenance is critical to ensuring that the industry runs smoothly, even with lower efficiency. Preventive maintenance is required in any sector to satisfy the demands of maximum profit and low cost for good output. This study examines the reliability of a complexly organized system of three units, A, B and C in order to determine its sensitivity to the effects of deteriorated rate and preventative maintenance rate over time. The three units are further made up of subunits which are in series or parallel configuration. The mathematical design work is based on the Markov process and the Laplace transformation. Different system parameters such as mean time to system failure, Available performance time, reliability, and profit, are analysed with respect to time and various rates. Further, A sensitivity analysis is used to explore how the rate of deterioration and preventative maintenance affects the system over time. Various malfunction and repair rates effect the system parameters in increasing or decreasing manner and sensitive analysis evaluated the impact of one unit on another or whole system. Here is a numerical example generated with the help of an appropriate model; the results are visually represented which concluded that with the passage of time reliability and other system parameters of system decreased under the influence of different rates. Utilizing the service cost, Profit is analysed which help to estimate the overall gain by the presented system. Also, by sensitive analysis it is concluded that out of three units A, B and C, Unit C has more effect as compared to B and C which is shown graphically. The purposed study can elaborate the profit after examined the reliability indices which become a key point for different industries like as diary plant, fertilizer plant etc. to have good outcomes with less maintenance cost.

Mushtaq A. Lone, S. A. Mir, Kaisar Ahmad, Aafaq A. Rather, Danish Qayoom, S. Ramki

This article addresses the challenges of determining the optimal allocation of sample sizes in stratified sampling design to minimize the cost function. Researchers employed the iterative procedure of Rosen's Gradient projection method and obtained optimal allocation of non-linear programming problem through manual calculation, which are often susceptible to human errors, such as rounding or arithmetic mistakes especially for complex nonlinear programming problems. R software performs calculations with high precision and consistency. In this paper, we demonstrate how to solve the non-linear programming problem by using iterative based procedure of Rosen's Gradient projection method through R software.

simulations and sensitivity analyses, offering managers a valuable instrument to enhance operational efficiency in scenarios including time-varying variables. This research improves the field by offering a strong solution framework for inventory management in complex scenarios with dual storage considerations, paving the way for more reliable and effective production strategies.

Control charts are essential in production processes to maintain quality of the products. Inspite of numerous control charts existing for process location under normal model, there is a need for developing control charts when situations demand production process under other distributions. In this paper, a class of control charts based on various midranges is proposed for monitoring location parameter of an ongoing process when process variables follow exponential distribution. The midranges are defined and their distributions are obtained. The performance of some members of the proposed class are evaluated in terms of their power, average run length (ARL), median run length (MRL) and standard deviation of run length (SDRL). Also, optimality and effectiveness of members of the class are discussed along with their illustration through an example.

In our research paper, we introduce an innovative statistical distribution known as the New Transmuted Rayleigh Distribution. This distribution serves as a versatile expansion of the traditional Rayleigh distribution and has been developed using a novel transmutation technique. We provide an in-depth analysis of several statistical properties of this new distribution. The resulting model has the ability to represent complex shapes, making it suitable for a wide range of applications. Our manuscript thoroughly examines the fundamental characteristics of the new model, outlining the methodology for estimating its unknown parameters through maximum likelihood estimation. Additionally, we demonstrate the practical significance of the model by applying it to an empirical dataset and conclusively establishing its superiority over some existing prominent models.

The strength of load-bearing and enclosing structures largely depends on the parameters of their materials. Complex shear testing of concrete is a non-destructive method used to determine the parameters and quality of the mixtures used with high accuracy. This concrete testing method has become widespread due to its versatility and convenience. The material strength is tested by directly impacting the concrete of the structure and causing its partial shearing. During the test, the force needed to tear off a fragment of the structure using a leafed anchor embedded in the bore hole is determined. This method can provide more accurate data on the concrete strength to make a decision on the need for further operation of the building. The concrete test to be shear tested must be located at a sufficient distance from pre-stressed rods. In addition, the test area should not be subjected to heavy operational loads.

A variant of an expert-statistical approach to solving the problem of forecasting parametric deviations of critical systems condition is proposed. Issues of development of specialized software (case-based reasonong approach) with the necessary problem orientation (forecasting degradation of technical condition) and allowing to improve the quality of forecast are

discussed. An approach to case describing using an ontological model of degradation processes with allowing to take into account both external influences, and internal processes characteristic of specific types of elements, is proposed.

K.M. Sakthivel, Vidhya G

In the current era, it is quite challenging to find symmetric data, as the form of most real-world data is asymmetric, meaning it tends to slant towards one side or another. These types of data emerge from various fields, including finance, economics, medicine, and reliability. Traditional statistical models often fail to handle such type of data as most of the statistical procedures are developed under normality assumptions. Therefore, the usual way of modeling these data results in incorrect predictions or leads to wrong decisions. There is no familiar methodology available in the research for modeling asymmetric data. Hence, there is a need to address this research gap as an emerging area of research in statistical modeling. In this paper, we propose a new systematic approach called the Model Selection Algorithm for modeling asymmetric data. In this algorithm, we incorporate various statistical tools and provide a guideline for a step-by-step procedure. Further, we have applied maximum likelihood estimation for parameter estimation, and model selection criteria such as Cramer Von Mises, Anderson Darling, and Kolmogorov Smirnov tests. We used real-time data to demonstrate the effectiveness of the algorithm.

Kavithanjali S, Sheik Abdullah A

Acceptance sampling plan by attributes is a statistical measure used in quality control in various production process. It is mainly determined for identifying whether the lot or the batch of the product is accepted or rejected based on the number of defective items in the sample. Appropriate sampling plan provides defect-free lot. There are several sampling plans are available for determine the sample size. Among the sampling plan, double sampling plan is more effective because it is always giving best result in lot selection compared with other sampling plan. In most of the practical situation, it is very hard to found the product as strictly defective or non-defective. In some situation, quality of the product can be classified several types which are expressed as good, almost good, bad not so bad and so on. This causes ambiguity deficiency in proportion value of lot or process. In mathematical tools, fuzzy set or fuzzy logic is one of the powerful modeling, which has incomplete and imprecise information. The fuzzy set theory is adopted to cope the vagueness in these linguistic expressions for the accepting sampling. In this article double sampling plans, are determined when non-conformities are fuzzy number and being modeled based on Zero-Inflated Poisson (ZIP) distribution. The Operating Characteristic (OC) function and Average Sample Number (ASN) function are evaluated both numerically and graphically in fuzzy and crisp environments.

G.V. Mamedova

Modern electromechanical devices for continuous process control are used in a variety of industrial applications. A control system or electromechanical device used to control the speed and torque of AC motors by varying the frequency and supply voltage converts alternating current of one frequency to alternating current of another frequency. The power section and the control device are the main elements of the control system. The main elements of a control system or electromechanical device are the power part (electrical energy converter) and the control device (controller). Modern frequency converters have a modular architecture, which expands the capabilities of the device, and also, in most cases, allows the installation of additional interface modules for input-output channels. The control device (microcontroller) is controlled by software and is controlled by the main parameters (speed or torque).

HOW TO PROPERLY APPLY SYSTEMS OF ARTIFICIAL INTELLIGENCE

H. Schäbe[1] ,I.B. Shubinsky[2]

[1] Dr.rer.nat., TÜV Rheinland InterTraffic, Cologne, Germany, dr.hendrik.schaebe@gmail.com
[2] DSc, prof., NIIAS, Moscow, Russia, Igor-shubinsky@yandex.ru

Abstract

The authors present their views on the essence of systems with artificial intelligence and point out the limitations to the use of those systems. Based on these considerations, an approach for the correct and effective use of artificial intelligence is proposed. A system with artificial intelligence (SAI) is in fact a very flexible statistical model with many parameters, which cannot be interpreted. Therefore, the use of an SAI is like a brute force attack using a very flexible statistical model to a problem. The sample which is used to train the SAI becomes much more important than the method itself. SAI can be used for safety applications, but the result of an SAI must be verified and that a proof of safety must be maintained. Mostly, this proof must be based on statistical arguments. A best approach for a use of a SAI is if it supports the developer for specific and well specified problem.

Keywords: artificial intelligence, effective application

I. Introduction

Systems using Artificial Intelligence (SAI) are widely used, one could even say that they have become fashionable see e.g. [1-5]. Apart from the obvious and recognised applications, it sometimes seems as if almost all problems could be solved using SAI.
In this paper, the authors present their views on the essence of SAI and point out the limitations to the use of SAI. Based on these considerations, an approach for the correct and effective use of SAI is proposed.

Chapter two presents our ideas about the nature of SAIs. The third chapter discusses the application of SAIs to security systems. The last chapter summarises and discusses the possibilities of effective application of SAIs.

II. The essence of artificial intelligence systems

Many types of SAIs are known, e.g.

- neural networks,
- deep learning,
- machine learning,
- support vector machines.

Vapnik's works [6,7] describe the mathematical core of artificial intelligence systems. In principle, AI systems are very flexible models of mathematical statistics having a huge number of parameters. Moreover, direct interpretation of these parameters is difficult, if not impossible.

This will mean that the data set that is used to train the SAI - in essence it must be a representative sample - has enormous value. In fact, the data almost entirely determines the behaviour of the SAI.

In this way, some problems of mathematical statistics affect the behaviour of the SAI.

The **first problem** is the representativeness of the sample that is used to train the SAI. The sample should cover all cases and situations that are relevant to the problem to be solved by the SAI. The crux of the issue is what is essential to the problem and hence should be reflected in the sample. Since there is no model or other abstract representation - this is the reason for using an SAI - it is difficult to decide which part of the sample is essential and which is not. In the end, it is often necessary to select the sample directly, make it as large as possible and accompany the sampling process with a rough analysis of what factors should be taken into account.

The **second problem** is the verification of the SAI. This means that for statistical verification, a second sample must be available that is completely diverse from the first sample used for training the SAI, but it must also be representative. This can be difficult if there are only single items for special cases that need to be included in both samples, for training and for verification.

The third problem is the difference between model fit to the data on one side versus model predictability on the other. This problem has been known for a long time, cf. [7]. It consists in the fact that a very flexible model with many parameters can approximate the data well, but it cannot predict new, additional cases. There are a lot of models describing well the weather in the past but that have poor predictability of future weather, despite the large number of parameters. SAIs are complex and comprehensive models with a large number of parameters and a researcher using an SAI can fall into this trap - the applied SAI describes the data well but gives a poor prediction for future cases.

The **fourth problem** is false (misleading) correlation. This problem has also been known for a long time [7]. If we study many factors that are not correlated and perform a correlation test with a statistical confidence of 90%, for example, a false correlation may occur with a probability of 10%. Examination of 10 uncorrelated values will thus result in an average of one false correlation. In NRIs, possibly hundreds of correlations are defined as intermediate parameters, e.g. in neural networks or deep learning systems. Thus, false correlations can occur in these SAI parameters and remain undetected.

The **fifth problem** is how to construct a model as a law of nature. There is a notion especially in the social sciences that a law of nature can be derived simply and directly from data using a mathematical formula that approximates the data well. However, it is additionally necessary to have an idea of what this formula expresses and what theoretical imaginations underlie the law. As an example, we can refer to the fact that Einstein's theory of gravity did not arise simply as a generalisation of experimental data, but on the basis of axioms that generalise human knowledge about gravity. Experiments then confirmed Einstein's theory. This process is absent when SAI is applied to process data with the hope of extracting a new law of nature.

In terms of mathematical statistics, SAIs are very complex and flexible models. They need representative samples, which in part require a huge amount of manual work. In addition, at least two different, representative samples are needed.

In this connection, one important concept should be discussed - the concept of **sufficient statistics** [9]. The essence of this concept is that all information contained in a sample of data that fit a parametric model, is contained in these sufficient statistics. Usually these are the parameters of that model. As an example, consider a sample taken from a population of normally distributed numbers. Then all information is contained in the sample mean and standard deviation. Of course, one could apply SAI to such a sample, but then the advantage of use of that parametric model would be lost.

It follows that the existence of a particular model contains additional information that can be used in data processing. This additional information is what makes this processing more efficient. Applying SAI directly does not lead to these results.

To summarise, SAIs are complex and flexible models that describe data well. But they are not able to replace the abstraction process leading to simpler models if an SAI is applied directly.

Of course, SAI can also be applied to the task "Find me a suitable parametric model for the data set" instead of the task "Describe me the data set". However, one should then check the result of the SAI with information about what processes led to the data under study and check the plausibility of the result suggested by SAI.

III Use of SAI for safety functions

An important question is whether the SAI can be applied to safety functions. This task has already been considered in [1]. The answer is in the affirmative. But the application of SAI to safety functions requires certain caveats.

The **first question** is what safety integrity level should be applied and which safety requirements should be derived. Without exception, all functional safety standards prescribe a risk analysis phase in the system life cycle. In this phase, the hazards, the possible accidents caused by them and the corresponding risks are analysed. In this phase, the system is treated as a black box, without yet setting specific requirements for the system development itself. What is derived in this phase are the requirements for the safety functions and methods for reducing the risk to an acceptable level, including the permissible hazardous failure rates of the safety functions of the system, and hence the levels of safety completeness. Since the risk analysis is not based on the internal structure of the future system, it is also acceptable to the SAI. Details can be found in [1].

The development of SAI is partly the same as for other electronic safety systems: hardware and software are developed according to the requirements of standards and the safety integrity level.

But, such a system is not yet operable, it needs to be trained with a large and representative sample. Thus, in addition to the classical elements of hardware and software, a third element is included, and that is the training sample.

The second question is about this third element. It also complicates the proof of safety, because one has to take this third element into account.

If we follow the ideas of [2], there are two approaches.

The **first approach** is based on the statistical ideas of the proven in use approach. In this case, the SAI is tested using a second, independent, representative sample. This sample should be of sufficient size to allow statistical proof for the required level of safety integrity. Especially for the third and fourth safety integrity levels, sample sizes that are needed that are practically difficult to obtain. Note that a second sample is required for all SAIs for system validation reasons. However, more stringent requirements apply for safety proofs, since this sample is used for statistical proofs.

The second approach is based on the possibility of explaining the behaviour of SAI. Here we can selectively cite the work of [10]. This standard provides a comparison with the ISO 26262 standard [11]. Further, safety management for artificial intelligence is introduced. The sampling issues for training and verification of SAI are thereby covered. It describes the lifecycle, methods of data verification and validation, conducting relevant safety analyses, the process of training of the SAI, the approach to failures, and so on. This standard is just one example of the development of standards for SAI in various fields.

Approximately in the same manner the argumentation in [12] is carried out. Here it is proposed to prove safety by a combination of formal methods, statistical methods and with the help of explicable SAI. And this paper is only an example of a number of articles.

One important aspect of safety systems is the tolerable hazardous rate. For conventional safety systems without the use of artificial intelligence, the hardware must have a hazardous failure rate less than an acceptable value corresponding to the safety level. Software has only systematic failures, which can be neglected if the software is designed following the requirements of the applicable safety integrity level. The same is true for systematic hardware failures. And they can be neglected if the development followed the requirements of the standard.

For SAI, there is an additional category of dangerous failures caused by erroneous decisions of artificial intelligence algorithms. This problem can be solved in such a way that a part of the permissible intensity of failures refers to erroneous and thus dangerous reactions of artificial intelligence algorithms. Consequently, the SAI for the safety function has the same total dangerous failure rate as a system without the use of artificial intelligence. This is achieved at the cost that the hardware for SAI must be better in order to transfer some of its tolerable hazardous failure rate budget to artificial intelligence.

IV. Conclusions

The question now is how to use SAI effectively. In fact, the use of an SAI is like a brute force attack using a very flexible statistical model to a problem. It must be remembered that the result of an SAI must be verified and that a proof of safety must be maintained when SAIs are used in systems with safety functions.

- **SAI supports the developer**

In this approach, the SAI collects information, compiles a list of references, etc. In this way, the SAI supports the developer, but the developer takes full responsibility for the results of his work.

- **Selection of the mathematical model**

Artificial intelligence helps in the selection of a mathematical model. In this case, the artificial intelligence suggests a model, and the developer checks the applicability of the model and shows with the help of statistical methods that the model is suitable. And in this case, the full responsibility lies with the developer.

- **Data analysis**

The SAI is used to analyse data. Trends, relationships, etc. can be identified. However, it is recommended to look for patterns that led to the data. In this case, the responsibility remains with the developer

These three cases are only selected examples of the application of SAI. It is recommended that the SAI be used as an assistant who carries out routine work, the results of which are checked by the developer himself.

In this way it is possible to avoid voluminous proofs of correctness of the SAI or even proofs of safety if the SAI would be used in systems with safety functions.

In any case, the use of SAI should not lead to the neglect of natural intelligence.

References

[1] Schäbe, H. and J. Braband, J., (2020). On safety assessment of artificial intelligence, *Dependability*, 19: 25-34.

[2] Schäbe, H. and Braband, J. (2022) The application of artificial intelligence in railway technology for safety-relevant applications - opportunities and problems , Signal and Data Communication 114 / no.5: 14-21.

[3] Shubinsky, I.B., Rozenberg E.N. and Schäbe, H. (2024) Methods for ensuring and proving functional safety of automatic train operation systems, RT&A, 19: 360-375.

[4] Posthoff, C. Computers and Artifical Intelligence, Past – Present – Future (In German: Computer und Künstliche Intelligenz Vergangenheit - Gegenwart - Zukunft), Springer Vieweg, 2022

[5] Berghoff, C., Biggio, B., Brummel, E. Danos, V., Doms, T., Ehrich, H., Gantevoort, T. Hammer, B., Iden, J., Jacob, S. Khlaaf, H. Komrowski, L., Kröwing, R., Metzen, J. H., Neu, M., Petsch, F., Poretschkin, M. Samek, W., Schäbe, H., von Twickel, A., Vechev M., and Wiegand, T., Towards Auditable AI Systems Current status and future directions, TÜV Verband, BSI, Fraunhofer, 2021

[6] Vapnik, V.N., Chervonenkis, A.Y., Pattern Recognition Theory. Statistical problems of learning (In Russian). Moscow, Nauka, 1974

[7] Vapnik, V.N., The Nature of Statistical Learning Theory, Springer, 2010.

[8] A.N. Kolmogorov, The Theory of Probabiity and Mathematical Statistics, Moscow, Nauka, On the question of applicability of prediction formulae, derived by statistical methods (in Russian), 161...167

[9] L. Schmetterer Introduction to Mathematical Statistics (In German: Einführung in die

mathematische Statistik,) Springer, Wien, New York 1966, I

[10] Road Vehicles - Safety and artificial intelligence, ISO DPAS 8800, 2024

[11] ISO 26262 *Road vehicles - Functional safety, 2018*

[12] G. Henzal, T. Strobel, J. Großmann, B.-H. Schlingloff, M. Leuschel, S. Sadeghipour, J. Firnkörn, (2021) KI-LOK - A joint test procedure project for AI based components used in railway operations, Signal and Data Communication (113) 10/2021 p. 6-15

APPLICATION OF THE DPSIR FRAMEWORK FOR SIBERIAN COMMUNITIES

Olga Derendiaeva, Valery Akimov
•
All-Russian Institute for Civil Defense and Emergency Situations of the Emergencies Ministry of Russia (Federal Center for Science and High Technologies)
derendiaeva.o@gmail.com

Abstract

The development of Siberia is a priority for the Russian government as it has great economic potential. However, the benefits for local populations are unclear, as economic expansion affects traditional livelihoods and social development. The challenges faced by the local population are, to some extent, relevant for all traditional communities in the world. While a huge amount of research is devoted to ongoing socio-economic processes in developing countries, about the transformations in Russian Siberia. In this paper DPSIR approach used to identify driving forces, pressures, states, impact and responses within Siberian communities. New indicators were proposed for a policy analysis.

Keywords: DPSIR Framework, sustainable communities, Siberian environment, sustainability measures

I. Introduction

Although there are several approaches for conducting these studies, a commonly used method is the DPSIR framework developed by the European Environment Agency in 1999 [13]. The DPSIR (The Driving forces - Pressure - State - Impact - Response) is a conceptual framework for describing the interactions between society and the environment. The flexibility is one of the main advantages of the DPSIR. It allows building the framework addressing the problem or management concern [71]. This advantage is embodied through a step-by-step approach to build the DPSIR. The initial step towards the creation of the DPSIR is identifying main concepts and indicators for specific problem.

II. Methods

DPSIR method highlights cause-and-effect relationships and systematises information to solve environmental problems. It was an extension of Pressure-State-Response (PSR) model developed by Organisation for Economic Cooperation and Development [43]. This technique appeared as a response to the need to analyse a Social–Ecological System (SES). According to [11] SES is a type of complex adaptive system. These systems comprise many interdependent parts that interact in ways that give rise to many patterns that cannot be predicted. The SES concept developed at works [10, 45]. According to [5] DPSIR or its modifications help to formalise relationships between society's adverse effects on the environment, and responses to such effects. Moreover, DPSIR shows these relationships in its dynamic. The ability to see the dynamics of the process is important for

adaptive management. It can help to monitor and evaluate the current situation to progressively implement a project. The DPSIR framework was applied for many projects and communities around the world and has been repeatedly improved and modified.

However, there are still vast geographical and conceptual areas that have hardly been studied in terms of adaptive management of SES. Recently, some increased in the use of the DPSIR method can be observed. Figure 1 shows a distribution of published papers by six countries during last 10 years (2012-2022). Three countries are northern states since part of their territory is located beyond the Arctic Circle; another group represented by three southern countries.

Despite varieties of approaches to the application of the DPSIR framework many studies have been carried out in relation to the southern communities, or at countries belong to Global South according to the UN classification [69]. There may be several reasons. A significant part of the world's population living at southern communities. Many countries at South and Southeast Asia are part of the so-called the Pacific Ring of Fire [40] is a region around much of the rim of the Pacific Ocean where many volcanic eruptions and earthquakes occur. This increases the number of natural disasters happened in this region and induced the vulnerability of communities. All these require additional measures to reduce risk and maintain resilience. Another reason may lie in the additional risks caused by the developing nature of the economies of these countries, which are in an active process of industrialization [41]. It can create an additional burden on ecosystems and risk for local communities. Given the two circumstances described above, international organisations, including the UN pay specific attention to the countries, united by the South-South development strategy.

Meanwhile the northern regions of our planet are experiencing no less pressure due to climate change. For example, over the past 30 years the Arctic has warmed more than three times faster than the world average [34]. The fragility of local ecosystems and their susceptibility to temperature changes; increased vulnerability of infrastructure, most of which is built on permafrost, the layer of which is constantly decreasing; life and cultural characteristics of the local population, adapted to harsh weather conditions that can be destroyed due to climate change – all these define the vulnerability of northern communities in the Arctic and Siberian regions [68]. Facts above describe the severity of the problem for communities in the northern latitudes makes a comprehensive assessment of the natural, social and economic development of this region extremely relevant. It is necessary to consider the actual lack of relevant research in such places like Siberia. Such study would fill gaps in scientific knowledge for these areas and enrich existing methodology of decision supporting tools with new concepts and methods from the northern perspective.

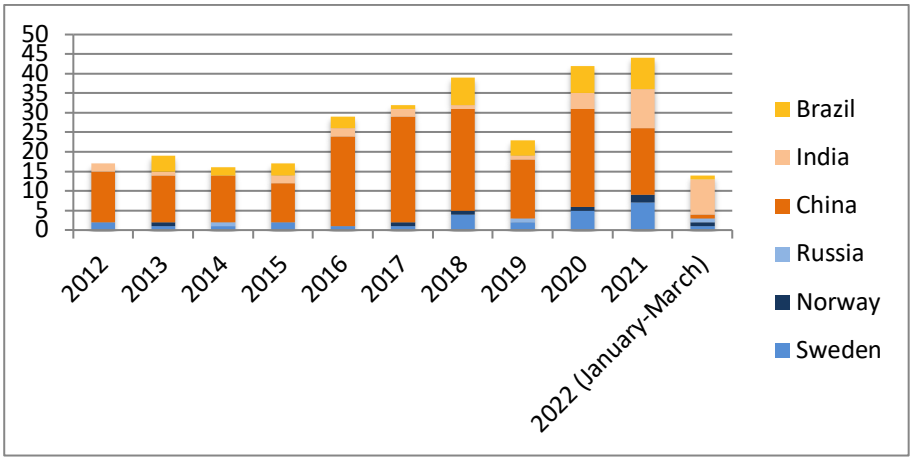


Figure 1: *The search results at Google Scholar, Russian Scientific Electronic Library and China National Knowledge Infrastructure (number of published papers regarding application of DPSIR framework by selected countries by years)*

In this paper DPSIR framework will be applied to Siberian region. Driving forces will be identified within current socio-economic and environmental situation; Pressures and States will be studied based on environmental reports for Siberian region during the last three decades; scientific

literature review will be done for identify the impact of adverse impacts; Russian legislation system will be analysed to study responses proved by the government. Further, based on the DPSIR analysis new indicators will be proposed for complex policy related study of communities in Siberia.

III. Results

Driving forces

Driving forces in the DPSIR model are fundamental human needs in the society, which cause different activities and lead to environmental problems. In the context of community vulnerability to climate change Driving forces can be described as human activities which aggravating climate change by inducing variability of ecosystems. Changes in climate lead to further variability of ecosystems, which creates a "vicious circle", with human activity as a prime Driving force. Further main Driving forces in Siberia are defined.

Siberia is the main Russian region for mining production [6, 28, 29]. Kemerovo oblast called the forge of Russia. That is why industrial raw material extraction and its processing (machine tool building, locomotive manufacturing, steel production, chemical production etc.) are the main economic drivers in this region and determine Siberian Federal District as a national centre of heavy industry [7]. Economic growth leads to intensification of raw materials production and the development of heavy industry in Siberia and as a result affects local ecosystems and the ability of communities for sustainable development and growth. The main centres of mining in the Siberian Federal District are located unevenly. It is important to note that some of the industrial centres are located on the territory of permafrost. This region has traditionally been a centre for the extraction of coal and other minerals. A new promising direction is the extraction of hydrocarbon raw materials, which is just beginning in this region. Significant prospects for increasing the Russian raw material production are associated with further study of the oil and gas potential of Siberia [62]. In addition to heavy industry, agriculture is well developed in the Siberian Federal District. It is another important driver of the economy. The share of the Siberian Federal District in the total volume of agricultural production of all agricultural producers in Russia (agricultural organisations, individual entrepreneurs) in the first quarter of 2021 amounted to 10.7% [59].

Despite the development of the heavy industry and agriculture, all regions of the Siberian Federal District have a negative balance of migration within the country for a long time. The Siberian Federal District occupies 25.5% of Russia's territory and accounts only for 11.7% of the country's population [59]. Only the city of Novosibirsk has a positive migration trend. In the regions, the outflow of the population is partly compensated by incoming migrants from Central Asia. Despite the positive impact of these migrants on the demographic situation of the region, they cannot compensate for the outflow of the local population. Migrants also aggravate ethnic problems because of different cultural values. Due to the low level of professional qualification, they can little contribute to modernisation of heavy industry. Such negative tendencies decrease ability for collective response and its adaptive capacity. Many experts [61] consider the impact of migration on the demographic situation, economy and culture to be negative. And in the opinion of other scientists [58, 38], even the local population is replacing by migrants. At the same time, some experts [63, 74] point to the low level of education and professional qualifications of newcomers, their cultural priorities and poor knowledge of the Russian language. These migration processes reduce the ability of communities to cooperate and increase their vulnerability to climate risks.

Another ethno-demographic problem is the presence of indigenous populations, living in traditional habitats and have a traditional lifestyle. They are among the most socially vulnerable groups of the population. According to [51] their places of residence are characterised by a low level of socio-economic development, lack of social and transport infrastructure. The problems of education of migrants and indigenous peoples directly affect human capital and the ability of peoples to adapt to climate change and conduct effective climate change policy. One of the main

tasks is the preservation and development of the national system of education of the indigenous peoples of Siberia, integration of migrants into multinational communities of Siberian Federal District. According to [67] the main problem of education is the lack of qualified personnel, as well as ignoring the cultural and historical characteristics of different nationalities, which reduces the effectiveness of educational programs. At the same time, part of the researcher emphasises the reluctance of the indigenous peoples themselves to receive education due to traditional lifestyle, which doesn't require high education [18, 39], while some other scientists see the problem in the preparation of educational programs [67, 57].

Complex infrastructure problems aggravate the vulnerability of local communities to climate change. In [19, 75, 24] described a low level of gasification of the Siberian regions, problems of energy efficiency, the absence of central heating in the regions of the Far North, which makes these territories dependent on the so-called Northern delivery of energy resources. However, it is necessary to note the additional effect of climate change that is associated with the possibility of using the Northern Sea Route more actively in the future for the transportation of hydrocarbons [62].

Pressures

These Driving forces create Pressures on ecosystem processes. Pressures in the DPSIR model show the degree of exploitation of resources. In general, the Siberian Federal District suffers from a complex degradation of ecosystems, including air, water, soil, flora and fauna [73, 60, 55, 21]. However, the pressure on some areas of the ecosystem is especially high, which, for example, makes the region a national leader in air pollution. These problems will be discussed below.

The high proportion of raw materials extraction and heavy industries cause the Siberian Federal District became a leader in terms of air pollution among other Russian regions. In particular, the Krasnoyarsk Krai, Novosibirsk city, Omsk city and Kemerovo Oblast took the last lines of the national environmental rating [21]. Another acute problem is the disposal of production waste. More than half of all Russian waste was generated in the Siberian Federal District [55]. The largest amount of waste (86.8%) is generated during the extraction of raw materials, mainly represented by products of mineral processing [21]. Many experts [48, 20] notify the lack of a systematic approach to the separation, collection and disposal of waste, including these containing toxic components, increasing the environmental pollution.

In addition to the factors mentioned above agriculture and forests clear-cutting contributes to the deterioration of the ecosystems of the Siberian Federal District. The problem of soil pollution is extremely acute in the region, both due to the development of the raw materials industry and due to the use of fertilisers in agriculture [56]. According to [8] the resource potential of Siberia forests has been significantly undermined by depleted forest management in the last century. Also, many authors [12, 2] point out that the problem of deforestation is connected not only with clear cutting, but also with fires, where the human factor and "consuming" attitude to the environment play an important role.

State

Adverse human activities create Pressures, which lead to the variability of the State of the environment and increase the Exposure of climate factors to local communities. [4] describes these factors as external to the community system, while Sensitivity is an internal factor of community vulnerability. According to [52, 2] two categories of changes in the State are observed in the territory of the Siberian Federal District: man-made changes and climate changes. Some authors [72, 49] claim that man-made changes in the State of ecosystems prevail in the region. While others [8, 65] believe that even though man-made transformations of ecosystem's State remain leading positions, at the beginning of the 21st century, the priority has shifted to the synergy of climate and man-made changes. Man-made changes are associated with a decrease in the forest fund, polluted territories,

degradation of agricultural land, as well as mining areas. Since the beginning of the century, the Siberian Federal District has been the leader in terms of the deforestation resulting both from clear cutting and from wildfires [35, 49]. The high industrialisation of the region leads to changes in the state of soils, including in areas of Arctic soils [65]. The area of the most valuable soils for agricultural use in the Krasnoyarsk Krai and Altai Krai is decreasing. Up to 50% of all chernosems of the Siberian Federal District is concentrated in these two regions [56].

According to [65, 25] the man-made changes described above are only intensifying due to increasing changes in climate. Climate changes are observed at the whole territory of Siberian Federal District, such as an increase in temperature, changes in precipitation, a decrease in permafrost cover in the North of the region, where much of the heavy industry infrastructure is located. The scenario made by [25] of the most intensive increase in greenhouse gas emissions by the middle of the 21st century in comparison to the end of the 20th century will lead to an increase in summer air temperature by 1.0-1.5°C, and in the South by 2-3°C. Precipitation in summer in the North and in the centre parts of Siberia may increase by 10-20%, and in the South decrease by 5-10%. In winter, in the northern and central regions, air temperature may rise by 5-8°C, in the southern part - by 3-5°C. The increase in precipitation in the winter season in most of the district is expected to be in the range of 20-40%, and on the coast of the Arctic ocean - 50-80%

Impact

The consequences of these Pressures and changes in the State lead to an Impact on local ecosystem and community wellbeing. Main Impacts on Siberian communities can be divided into three categories: economics Impact, environmental Impact and social Impact. Currently, there is no unequivocal opinion in Russian science regarding the impact of climate change on the communities of the Siberian Federal District. Some researchers [30, 50] see climate change as an opportunity, mainly due to increase in the period of use of the Northern Sea Route, availability of raw materials, development of tourism due to milder temperatures, reduction of central heating costs. While others [47, 36, 33] highlight its disruptive impact on community functioning, and believe that the benefits of new opportunities will not offset the damage from climate warming, especially for communities whose economic activity: takes place in permafrost regions; associated with the agricultural sector and fisheries; depends on "Northern delivery"; as well as for the indigenous peoples of the North, who are engaged in traditional economic activities. The distribution of hazardous natural phenomena across the territory of Russia largely depends on the relief and the degree of proximity to large water areas [46]. The largest number of hazardous phenomena in recent years has been observed in the Siberian Federal District [14]. The geographical position, flatness of the territory of Western Siberia creates prerequisites for the penetration of air masses both from the South and from the North. From the West, the territory of Siberia is bounded by the Ural Mountains, and its eastern part in the cold half of the year is subject to the influence of the western spur of the Siberian anticyclone. All this leads to instability in the atmosphere. At the same time, in the South of Western Siberia, according to the results published in 2017 [54] the risk of heavy and prolonged precipitation has significantly increased in recent years. All this leads to great destruction from natural disasters, for example, flooding of residential buildings.

In addition to the uneven distribution of natural disasters depending on the landscape, it has seasonal features. According to [33], at the beginning of the 21st century, in the spring-summer period, almost the entire territory of Western Siberia is covered by warming. In the cold season, this process is disrupted. In autumn, the temperature begins to decrease, and in winter this process is clearly expressed in the South of the region (including the Tomsk Oblast), while in its northern part the trend is stable positive. In summer, there is a possibility of dry periods, as there is an increase in negative extremes of precipitation. This is also facilitated by the fact that in the South of the region there is an increase in the likelihood of high temperatures (above 30°C) and a decrease in wind speed

[33]. This is especially notable in the South of Western Siberia and in the Tomsk Oblast and Altay Krai (the region engaged in agriculture, which certainly has a negative impact on the economic and social development of local communities). In addition, there is a threat to the local population, flora and fauna due to the increased number of forest fires. While many communities in the South of the region depend on agro-industrial production, the population of the northern communities is often engaged in fishing [66]. The climate change described above affects the region's aquatic systems, fisheries and aquaculture, and endangers the communities that depend on them and their livelihoods.

In the light of the above changes, particular concerns are caused by disruptive tendencies in the regions of the Far North. In the regions of the Far North, climate change has contributed to the intensification of new natural hazards, which can be called geocryological [3]. These include processes and phenomena that are associated with the melting of ice and have an adverse impact on the environment. One of the problems that have become aggravated in recent years is the destructive impact of permafrost melting on infrastructure. Many industrial and residential buildings, oil rigs, pumping stations and pipelines, roads, bridges, runways in the northern regions are built on permafrost and are designed for operation in a certain range of changes in external conditions. Melting of permafrost can affect many logistics processes, for example, "North delivery".

All these changes become a serious economic challenge for the region. The development of non-environmentally friendly, extractive industries in the region, as well as high climate risks, may reduce the potential for attracting investment. The development of the ESG approach in the world practice also reduces the attractiveness of many companies in the Siberian Federal District for potential investment. At the same time, according to [9, 53] in Siberia, the development eco-products market based on the agricultural production of the Altai Mountains is planned. This new market can potentially create new "green" development opportunities for this region. This potential is based on the ecosystem richness of this region. The nature of the Altai Mountains has diverse natural potential. Its flora and fauna make it possible to classify Altai as one of the world's centres of biodiversity concentration on the planet.

Resposes

Resolutions, National Strategies, Federal and regional laws, decrees of the Government of the Russian Federation - all these are the main sources of information about the Responses in the field of reducing the vulnerability of communities to climate change. However, as [42] mentions in the legislation of the Russian Federation the prevention of climate change is not established as a goal of legal regulation. Climate change combating policy is reflected in the following fundamental documents: [17, 25]. In addition, there are several Federal laws aimed at protecting forests, waters and lands from pollution. However, the Strategy for Social and Economic development of Siberia until 2020 [27] did not reflect the problems of vulnerability to climate change. All measures to combat climate change can be divided into three categories: prevention, mitigation and adaptation; and two areas: environmental and socio-economic. Table 1 provides a list of examples of various measures and legislative acts in specific categories.

Table 1: *Examples of legislations and other measures to reduce the vulnerability of the population to climate change in Russia*

	Environmental	Socio-economic
Prevention	Fedortsova and Sidorenko, 2017; (PJSC Gazprom Neft is implementing the Corporate Biodiversity Conservation Program based on the List of flora and fauna species that are indicators of the sustainable state of marine ecosystems	Federal Law on Fire Safety of December 21, 1994 (as amended on June 11, 2021) On Fire Safety. Federal Law No. 82 of April 30, 1999 (as amended on July 13, 2020) On

	in the Russian Arctic zone.) Federal Law on Fire Safety of December 21, 1994 (as amended on June 11, 2021) On Fire Safety. List of requirements and mandatory measures to improve, protect land and protect soil from winds, water erosion, 2019. (Federal service for veterinary and phytosanitary supervision Order July 8, 2019, no. 662)	Guarantees of the Rights of Indigenous Peoples of the Russian Federation.
Mitigation	Federal Law of June 11, 2021, No. 175 On the development of agriculture. Federal Law No. 254 dated July 31, 2020, On the Peculiarities of Regulating Certain Relations for the Purpose of Modernising and Expanding the Main Infrastructure and on Amending Certain Legislative Acts of the Russian Federation. Federal Law on Production and Consumption Wastes of June 24, 1998 (as amended by Resolution of the Constitutional Court of the Russian Federation of July 19, 2019, N 30-P)	Guidelines for organising the preparation and support of a flood-hazardous period on the territory of a constituent entity of the Russian Federation (approved by the Russian Ministry of Emergency Situations on November 10, 2021). Oltyan, et al., 2021 Implementation of the Sendai Framework for Disaster Reduction in the Russian Federation. Kharitonov, et al, 2021 Problems and prospects of socio-economic development of rural areas: a regional aspect
Adaptation	The National Action Plan for the first stage of adaptation to climate change for the period up to 2022 was created, which was created for various sectors of the economy. Order of the Ministry of Economic Development of the Russian Federation of May 13, 2021, N 267 "On approval of methodological recommendations and indicators on adaptation to climate change».	Order of the Federal Agency for Ethnic Affairs of Russia dated November 17, 2020, N 142 On approval of the Guidelines for public authorities of the constituent entities of the Russian Federation "On the social and cultural adaptation and integration of foreign citizens in the Russian Federation". The National Action Plan for the first stage of adaptation to climate change for the period up to 2022 was created, which was created for various sectors of the economy.

Further main features of prevention, mitigation and adaptation measures held in Russian are considered. In general, measures to prevent the impact of climate change on the communities are mainly based on the protection of the local population from natural and man-made disasters, as well as the protection of the rights of vulnerable population; local authorities are developing guidance to prevent soil destruction and the reduction of the forest fund. Despite the fact, that guarantees of the rights of indigenous peoples are described in the Federal law there are need for additional support of these communities. According to [64] there are some measures of support on reindeer husbandry, as this is the main area of traditional business of the indigenous peoples of the

North of Russia. These measures include: the support for reindeer breeding as a separate area of subsidies, including the purchase of breeding stock. Grants for the development of a family farm, purchase snowmobiles etc.

Measures to mitigate climate change impact include increasing the preparedness of the population for natural disasters, implemented as part of international initiatives such as: "My city is getting ready!" [1]. In addition, special attention is paid to rural communities. All regions of the country are divided into four types with the first type, where the rural population prevail over urban population, and the fourth type, where the urban population prevails. It should contribute to targeted assistance within the framework of the implementation of the Strategy for the Sustainable Development of Rural Territories of the Russian Federation until 2030 [44]. The waste management reform aims to harmonise the sorting and disposal of various types of waste. Also, it is planned to increase opportunities for waste processing [22]. [23] considers measures to create infrastructure in specially protected natural areas and reserves, as well as on the territory of Lake Baikal.

Comprehensive adaptation measures have only just begun to be implemented in Russia. Their development is associated with the publication of The National Action Plan for the first stage of adaptation to climate change for the period up to 2022. It is supposed that this plan evolves in 2022 and beyond to create comprehensive climate change adaptation measures.

IV. Discussion

DPSIR indicators

Based on main DPSIR concepts for Siberian region, indicators for the application of the DPSIR framework are proposed at Table 2.

Table 2: *DPSIR indicators for Siberian region*

DPSIR component	№ of Indicator	Indicator
Driving forces	D1	Gross regional product and its % of GDP
	D2	Population
	D3	Financing of geological exploration (billion rubles)
	D4	The share of agricultural production from the all-Russian (%)
	D5	The share of extraction of natural resources from the all-Russian (%)
	D6	Number of indigenous people
	D7	Share of communities employed in agriculture (%)
	D8	The level of gasification of the population (%)
	D9	Urbanisation rate (%)
	D10	Number of industrial centres in permafrost areas
	D11	Migration rate
Pressures	P1	Greenhouse gas emission by per sector
	P2	Non-greenhouse gas emission by per sector
	P3	Solid waste generation rate m3/ person
	P4	Spread of fires (hectar)
	P5	Applied chemicals (kg active ingredient / hectar)
State	S1	Temperature variability
	S2	Changes in the forest fund (m2)
	S3	Precipitation variability
	S4	Permafrost variability
	S5	Degradation of agricultural land (%)
	I1	Grain yield

Impact	I2	Human Development Index
	I3	Number of households depend on North Delivery
	I4	Number of households affected by disasters
	I5	Tourism development
	I6	Fishery development
	I7	Development of reindeer breeding
	I8	Infrastructure development
	I9	Investments
Responses	R1	Number of normative acts and regulations
	R2	Share of recycling and neutralisation of waste
	R3	Human capital development
	R4	Dynamics of modernisation
	R5	Dynamics of gasification
	R6	Dynamics of deforestation

The DPSIR can be a useful tool for creating comprehensive plans for the development of the region. The creation of DPSIR for community-based management in climate risk in Siberia needs a set of special indicators. The study needs to start with the defining of the study area, which is a difficult task due to the vagueness of the term Siberia. To solve this problem, it is proposed to proceed from the modern administrative-territorial boundaries of the regions of Russia, due to climate change policy remain state-led in Russia. Also, it is proposed to consider ethnic and climatic factors. After studying the materials, it becomes clear that the territory of the Siberian Federal District, including the Tyumen Oblast, which has a trans-boundary position, has boundaries that correspond to the historical definition of Siberia. Ethnic processes in the territory of this region led to the formation of a special sub-ethnos of Siberians, which can also serve to determine the study area within one sub-ethnic group. As for the climatic features, the longitude location of the region makes it possible to study the impact of climate change on the activities of communities in the Arctic, sub-Arctic and Continental climatic zones, as well as strategies for interzonal interaction. Based on DPSIR concepts a set of indicators was determined for the quantitative assessment of the components of the DPSIR. All this work is a preparatory stage for the creation of the DPSIR for community-based management in climate risk. Further research could be carried out in the field of studying the main indicators and identifying the relationships between the main elements of the DPSIR.

References

[1] Akimov, V., Bedilo, M., Derendiaeva, O., Ivanova, E., & Oltyan, I. (2022). Forecast of natural emergency situations with modern methods. *Reliability: Theory & Applications,* 17(SI 4 (70)): 71-77.

[2] Andreeva, D., Korobova, Yu. (2018). Problems of forest protection from fire. *Scientific and practical electronic journal Alley of Science,* 10 (26): 734-738.

[3] Anisimov, O., Belolutskaya, M. (2002). Assessing the Impact of Climate Change and Permafrost Degradation on Infrastructure in the Northern Regions of Russia. *Meteorology and Hydrology,* 6: 15-22.

[4] Babu, M., & Ray, P. (2023). Sensitivity analysis, optimal design, cost and energy efficiency study of a hybrid forecast model using HOMER pro. *Journal of Engineering Research,* 11(2): 100033.

[5] Barnard, S., Elliott, M., (2015). The 10-tenets of adaptive management and sustainability — applying an holistic framework for understanding and managing the socio-ecological system. *Environmental Science and Policy*, 51: 181–191.

[6] Basareva, V. (2016). Problems of the Siberian Federal District in the context of the

objectives of restoring economic growth. Scientific and technical statements of SPbSPU. *Economic sciences*, 6(256): 73-85.

[7] Basic catalog of high-tech industrial products and services for the needs of the Arctic zone of the Russian Federation. *Minpromtorg.* Available at: https://minpromtorg.gov.ru/common/upload/docs/sfo.pdf (Accessed: 08/07/2024).

[8] Belonovskaya, E., Tishkov, A., Vaisfel'd, M., Glazov, P., Krenke, Jr., Morozova, O., Pokrovskaya, I., Tsarevskaya, N., Tertitskii, G. (2016). "Greening" of the Arctic and current trends of its biota. *Izvestial Russian Academy of Science. Geography,* 3: 28—39.

[9] Belova, I., Karslyants, E. (2014). Market of organic products: global trends and development prospects in Russia. *Vestnik RUDN University. Series: Economy,* 2: 40-49.

[10] Berkes, F. 1989. Common property resources. Ecology and community-based sustainable development. *Belhaven Press,* London.

[11] Biggs, Reinette, et al. (2022). Social-ecological change: insights from the Southern African Program on Ecosystem Change and Society. *Ecosystems and People,* 18, 1: 447-468.

[12] Bruykhanov, A. (2009). Environmental assessment of the state of forests in Siberia: an alarming result. *WWF Russia.* 2 (21): 21-31.

[13] Classification of ecosystem services (EEA). *UNCEEA.* Available at: https://seea.un.org/sites/seea.un.org/files/5_74.pdf (Accessed: 08/07/2024).

[14] Climate Report. *Meteorf.* Available at: https://yandex.ru/search/?text=климатический+отчет+2018+&lr=213 (Accessed: 08/07/2024).

[15] Commissioner for the rights of indigenous peoples in the Krasnoyarsk Krai. Available at: https://www.ombudsmankk.ru/media/Doklad_KMNS_2020.pdf (Accessed: 08/07/2024).

[16] COP 26 Glasgow Climate Pact. *COP 26.* Available at: https://unfccc.int/sites/default/files/resource/cop26_auv_2f_cover_decision.pdf (Accessed: 08/07/2024).

[17] Decree of the President of the Russian Federation of November 4, 2020 No. 666. Kremlin.ru. Available at: http://www.kremlin.ru/acts/bank/45990 (Accessed:08/07/2024).

[18] Dicansky, N., Popkov, Y., Radchenko, V., Sviridov, I. Education for the indigenous peoples of Siberia: the social role of the Novosibirsk State University, Nonparel, Novosibirsk, 2005.

[19] Difficulties in the way of the Eastern gas vector. *Energy Bulletin March 2016*. Available at: https://ac.gov.ru/files/publication/a/8435.pdf (Accessed: 08/07/2024).

[20] Drugov, Yu. Analysis of contaminated soil and hazardous waste, Binom. Knowledge Laboratory, Moscow, 2019.

[21] Ecology. *Kommersant.* Available at: https://www.kommersant.ru/region/novosibirsk/files/novosibirsk/PDF/INSTRUMENTY_ekologia_04_2020.pdf (Accessed: 08/07/2024).

[22] Federal law on production and consumption waste. N 89-FZ. June 24, 1998. As amended by Resolution of the Constitutional Court of the Russian Federation of July 19, 2019 N 30-P. Available at: http://www.consultant.ru/document/Cons_doc_LAW_19109/ (Accessed: 08/07/2024).

[23] Federal Law No. 254-FZ dated July 31, 2020 On the Peculiarities of Regulating Certain Relations for the Purpose of Modernising and Expanding the Main Infrastructure and on Amending Certain Legislative Acts of the Russian Federation (last edition). Konsultant.ru Available at: http://www.consultant.ru/document/cons_doc_LAW_358735/ (Accessed: 24/05/2022).

[24] Fuel and Energy Complex of Russia. *Industry directory of enterprises in Russia.* Available at: https://www.cdu.ru/upload/medialibrary/21c/ectaufnfsqdmk32k9ej77ueqtyds6tzk/TEC_Russia_02_2021_Full.pdf (Accessed: 08/07/2024).

[25] Global climate change and the Siberian Federal District. On the way to adaptation. *Federal Service for Hydrometeorology and Environmental Monitoring.* Available at:

https://meteoinfo.ru/images/media/books-docs/klim-riski-2017.pdf (Accessed: 08/07/2024)

[26] Government of the Russian Federation Order N 3187-p December 25, 2019. Government.ru. Available at: http://government.ru/docs/all/125386/. (Accessed: 08/07/2024).

[27] Government of the Russian Federation order dated July 5, 2010, No. 1120-r. Government.ru. Available at: http://government.ru/docs/all/73524/. (Accessed: 08/07/2024).

[28] Innovative development of the Russian Fderation in 2019 Siberian Federal District. *FGBNU Research Institute RINKCE.* Available at: https://www.miiris.ru/digest/analitika_SibFO.pdf (Accessed: 08/07/2024).

[29] Information Bulletin on the state of the mineral resource base of solid minerals. *FSBI VIMS.* Available at: https://vims-geo.ru/documents/203/Sibir_03072018.pdf (Accessed: 08/07/2024).

[30] Isaev, A. Ecological climatology. Scientific world, Moscow, 2001.

[31] Khajuria, A., Ravindranath, N. (2012). Climate Change Vulnerability Assessment: Approaches DPSIR Framework and Vulnerability Index. *Earth science and climate change,* 3 (1): 1-6.

[32] Kharlamova, N., Sukhova, M., J Chlachula, J. (2019). Present climate development in Southern Siberia: a 55-year weather observations record. *IOP Conference Series: Earth and Environmental Science*, 395: 012027.

[33] Kharyutkina, E., Loginov, S., Usova, E., Martynova, Yu., Pustovalov, K. 2019. Change trends in Western Siberia's climate extreme in the end of the 20th - beginning of the 21-th centuries. *Fundamental and applied climatology,* 2: 45-65.

[34] Koenigk, T., Key, J., & Vihma, T. (2020). Climate change in the Arctic. *Physics and chemistry of the Arctic atmosphere,* 673-705.

[35] Kolesnikova, A. (2015). Forest use on the territory of Siberia and the Far East: state and dynamics. *Bulletin of the ZabGU*, 6 (121): 127-143.

[36] Koroleva, T., Konstantinov, A., Shunkin, E. (2015). Threats and socio-economic impacts of climate change on the forestry sector. *Proceedings of the St. Petersburg Research Institute of Forestry,* 3: 55-71.

[37] Kronik, J., Verner., D. (2010). Indigenous Peoples and Climate Change in Latin America and the Caribbean. *The World Bank.* Available at: https://documents1.worldbank.org/curated/en/654311468010837927/pdf/555400PUB0Indi1EPI1958810601PUBLIC1.pdf (Accessed: 08/07/2024).

[38] Ledenyova, V., Kononov, L. State and municipal regulation of the processes of adaptation and integration of migrants in modern Russia. RUDN University, Moscow, 2021.

[39] Malysheva, E., Nabok, I. (2015). Education of indigenous peoples of the Arctic: problems and development prospects. *Society. Environment. Development (Terra Humana),* 1 (34): 139-144.

[40] Masum, M., & Akbar, M. A. (2019). The Pacific ring of fire is working as a home country of geothermal resources in the world. *In IOP Conference Series: Earth and Environmental Science*, 249, 1: 012020.

[41] Myers, G. (2021). Urbanisation in the global south. *Urban ecology in the global south*, 27-49.

[42] Nikonov, I. (2022). Using a macroalgal cultivation system in wastewater bioremediation: a case study of Bolungarvík, Iceland (Doctoral dissertation).

[43] OECD core set of indicators for environmental performance reviews. *OECD.* Available at: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD(93)179&docLanguage=En (Accessed: 08/07/2024).

[44] Order of the Government of the Russian Federation of February 2, 2015 No. 151-r on the approval of the Strategy for the Sustainable Development of Rural Territories of the Russian Federation for the period up to 2030. Government of the Russian Federation. Available at: http://static.government.ru/media/files/Fw1kbNXVJxQ.pdf (Accessed: 08/07/2024).

[45] Ostrom, E. Governing the commons, Cambridge University Press, Cambridge, 1990.

[46] Panfutova, Yu. Dangerous meteorological phenomena in the flat territory of the Russian

Federation and the risks they create, Main Geophysical Observatory named after A.I. Voeikova, St. Petersburg, 2008.

[47] Pavlenko, V., Sergeev, A. (2006). Climate warming in Western Siberia and possible environmental and economic consequences. *Interexpo Geo-Siberia,* 6: 176-181

[48] Pavlenkov, M., Voronin, P. (2018). Problems of development of the sphere of solid municipal waste of the municipality. *Bulletin of the Kemerovo State University. Series: Political, sociological and economic sciences,* 3: 130–139.

[49] Pimenov, A., Danilina, D., Konovalova, M., Stepanov, N. (2020). On the conservation of intra-specific diversity of the Siberian cedar in the mountains of the Altai-Sayan region. *Scientific basis for sustainable forest management,* 90-93.

[50] Porfiriev, B., Voronina, S., Semikashev, V. (2017). Climate change impact on economic growth and specific sectors' development of the Russian Arctic. *The Arctic: ecology and economy,* 4 (28): 4—17.

[51] Report on the problems of realising the constitutional rights and freedoms of indigenous peoples in the territory of the Krasnoyarsk Krai in 2020. Ombudsmankk.ru. Available at: https://yandex.ru/search/?text=Доклад+о+проблемах+реализации+конституционных+прав+и+свобод+коренных+народов+на+территории+Красноярского+края+в+2020+году.&search_source=dzen_desktop_safe&lr=213 (Accessed: 15/07/2024).

[52] Rozhnova, V. Study and conservation of biodiversity in South Siberia and Central Asia in transboundary territories, Association of scientific publications KMK, Moscow, 2018.

[53] Ruschitskaya, O., Noskova, K., Fetisova, A., Zhelvis, S. (2019). The market of environmentally friendly products in Russia. Problems and prospects of development. *Moscow Economic Journal,* 4, 33.

[54] Sharapova, A., Kuzhevskaya, I., Kashtanova, K., Polyakov, D. 2017. Characteristics of extreme precipitation in Western Siberia. *Geographic Bulletin. Meteorology*, 3 (42): 88-98.

[55] Shilkina, S. (2020). World trends in waste management and analysis of the situation in Russia. *Waste and Resources Online Journal*, 1 (7): 1-17.

[56] Shpedt, A., Aksenova, Yu., Zhulanova, V., Rassypnov, V., Erunova, M., Butyrin, M. (2019). Assessment of Siberian agrochernosems based on modern approaches. *Zemledelie,* 4: 8-14.

[57] Sinitsa, A. (2019). Improving the level and quality of education of the indigenous peoples of the North: problems and prospects. *The level of life of the population of the regions of Russia,* 3 (213): 70-81.

[58] Soboleva, S., Smirnova, N., Chudaeva, O.V. (2014). Migration of the population in the border regions of Siberia. *EKO,* 8 (482),18-31.

[59] Socio-economic situation of the Siberian Federal District in q1 2021. *Federal Service of State Statistics*. Available at: https://rosstat.gov.ru/storage/mediabank/tgniSwUP/Bul_Sib-fo.pdf (Accessed: 08/07/2024).

[60] Soyan, Sh. (2019). Ecological problems of the Siberian regions. *Natural resources, environment and society,* 3 (3): 34-38.

[61] Starkov, P., Sbrodova, N. (2017). Problems and consequences of internal migration of labor resources from the Ural, Siberian and Far Eastern federal districts. *Human Progress,* 3 (7), 6.

[62] State balance of mineral reserves. Rosgeolfond. Available at: file:///Users/derendaeva94/Downloads/5e7047359b706ba031120c98201d3e8d.pdf (Accessed: 08/07/2024).

[63] Suslova, T. (2016). Adaptation of migrants: understanding the problem in the context of a constructivism approach. *Psychology*, 9-3 (51): 100-103.

[64] The Ministry of Agriculture will finalise the measures of state support for reindeer breeding. Interfax. Available at: https://www.interfax.ru/russia/768738 (Accessed: 08/07/2024).

[65] Tishkov, A., Belonovskaya, E., Vaisfel'd, M., Glazov, P., Krenke, A., Tertitskii, G. (2018). "Greening" of the tundra as a driver of modern dynamics of the Arctic biota. *Arktika: ecology and*

economics, 2 (30): 31—44.

[66] Turaev, V. (2017). Traditional fishing as a conflict of interests of aboriginal communities, business and authorities. *Russia and the Asia-Pacific Region,* 2 (96): 177-194.

[67] Ulturgasheva, N., Turamuratova, I., Alekseeva, A. (2017). The problem of preservation and development of the national-regional system of ethno-cultural education of the indigenous and small peoples of Siberia. *Uchenye zapiski (AGAKI),* 3 (13): 130-134.

[68] Ulturgasheva, O., & Stelmaszyk, M. (2024). Embracing uncertainty: porous and actionable responses to climate change at the borders of Indigenous and scientific expertise (s) in Siberia. *Journal of the Royal Anthropological Institute,* 31 (1): 1-19.

[69] UN countries classification. Un.org. Available at : https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf (Accessed: 08/07/2024).

[70] UNFCCC report of the thematic in-session workshop: enhancing the participation of local communities, in addition to indigenous peoples, in the local communities and indigenous peoples platform 25 november 2019. *UNFCCC.int.* Available at: https://unfccc.int/sites/default/files/resource/LCIPP%20workshop%20report_2019.pdf (Accessed: 08/07/2024).

[71] Using the DPSIR Framework to Develop a Conceptual Model: Technical Support Document. *US EPA*. Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NHEERL&dirEntryId=311236 (Accessed: 08/07/2024).

[72] Vladimirov, I. (2018). Anthropogenic disturbance and dynamics of geosystems of Baikal Siberia. *Nature of Inner Asia*, 1 (6): 19-30.

[73] Zabelina, I., Klevakina, E. (2016). Economic development and negative impact on the environment in the regions of cross-border interaction. *ECO*, 8 (506): 67-82.

[74] Zavyalov, A. (2019). Topical issues of state regulation of social adaptation of migrants. *State Administration. Electronic Bulletin.* 72: 143-164.

[75] Zemnukhova, E. (2018). Directions of gasification of Siberian regions as a key direction for improving the environmental situation in the region. *Interexpo Geo-Siberia,* 2: 216-221.

RT&A, No 4(80)
Volume 19, December, 2024

DETERMINATION OF STABILITY AND RELIABILITY OF SHORTEST PATHS IN A GRAPH THROUGH LISTS OF LABELS IN DIJKSTRA'S ALGORITHM

G.Sh. Tsitsiashvili

•

[1]Insitute for Applied Mathematics, Far Eastern Branch of Russian Academy Scinces, Russia
guram@iam.dvo.ru

Abstract

In this paper, the problem of determining all shortest paths is solved in a weighted graph. For a weighted graph, the path length is defined as the sum of the lengths of its edges. This problem is solved by generalizing the well-known Dijkstra algorithm by introducing a list of labels. In the list of labels at each vertex of the graph, the first label determines the length of the shortest path. The second label is defined by a set of vertices, from which directed edges exit to the vertex in question. To reduce the required memory and determine the reliability of the shortest paths, the number of edges of the shortest paths entering the vertices of the graph is introduced and recursively calculated. The stability of shortest paths is calculated recursively, as the number of edges, paths entering the vertices of the graph and deviating from the minimum length by a given amount. These results extend to unweighed and planar graphs.

Keywords: weighted graph, shortest path, vertex neighbours, path length, reliability, stability

1. Introduction

In this paper, we consider the problem of enumerating all the edges of the shortest paths from the starting vertex to other vertices of a directed graph (digraph) or from the face of a planar graph to the outer face. This problem is solved for weighted and planar graphs at the request of specialists in the field of transport, cartography and biotechnology, where it is currently relevant [1], [2].

For a weighted graph, the path length is defined as the sum of the lengths of its edges. For a planar graph, the path length is the number of boundaries between the faces that the path from the face to the outer face intersects. The problem of determining the minimum length from the starting vertex to other vertices of the graph in a weighted digraph is solved on the basis of the Dijkstra algorithm [3] or on the basis of the Bellman algorithm [4], [5]. A similar problem, but in an unweighed graph, is solved based on the wave algorithm [6]. And the problem of determining the length of the shortest path from each face to the outer face in a planar graph is solved using the concept of a set of neighbours - surrounding a face of a planar graph with other faces.

In this paper, the algorithms under consideration are supplemented by the concept of a list of labels. The concept of a label list is taken from mathematical programming (see, for example, [7], [8]). The list of labels consists of the first label, which determines the length of the shortest path to the vertex from the starting vertex or the length of the shortest path from the face to the outer face. For weighted graphs, the second label is a list of vertices from which the edges of the shortest paths are directed to the vertex in question. Thus, using the second labels at the vertex of a weighted or unweighed digraph, it is possible to list all the edges of the shortest paths to this vertex of the graph. For a planar graph, the second label is a list of faces that are adjacent to the face in question in the shortest paths to the outer face. Thus, using the lists of labels, it is possible

to determine all the shortest paths from the starting vertex to the other vertices of the weighted digraph. Similarly, for a planar graph, using the lists of labels at the edges of the graph, you can list all the shortest paths to the outer face.

2. Inclusion of new labels in Dijkstra's algorithm

The Dijkstra algorithm solves the problem of determining the length of the shortest paths from the selected vertex of a weighted (equipped with a positive edge length) graph G. However, there is no algorithm for enumerating all edges included in the shortest paths. At this point of the paper, a modification of Dijkstra's algorithm is being built, which allows us to list all the edges of the shortest paths using lists of additional labels. Dijkstra's algorithm is constructed for a directed weighted graph (without multiple edges) G with a set of vertices V, consisting of n vertices, a set of edges E and a set of edge weights $\{s(e) > 0,\ e \in E\}$. In the graph G there is a starting vertex $\mathbf{0}$ from which all shortest paths to the other vertices of this graph are searched.

The first and second vertex labels of the graph G are determined by a recurrent procedure. First, set $R(\mathbf{0}) = 0,\ R(v) = \infty,\ v \in V,\ v \neq \mathbf{0}$; $S(v) = \emptyset,\ v \in V,\ G_0 = G,\ V_0 = V,\ E_0 = E$. Next, for $0 \leq j \leq n$, select the vertex k_j from the condition $R(k_j) = \min\limits_{v \in V_j} R(v)$, redefine the labels $R(v),\ S(v),\ v \in V_j,\ v \neq k_j$ and construct the graph G_{j+1} from the following conditions.

($\mathbf{A}$). If $R(k_j) + s(k_j, v) < R(v)$, then $R(v) := R(k_j) + s(k_j, v),\ S(v) := k_j$.

($\mathbf{B}$). If $R(k_j) + s(k_j, v) = R(v)$, then $R(v) := R(v),\ S(v) := S(v) \cup k_j$.

($\mathbf{C}$). If $R(k_j) + s(k_j, v) > R(v)$, then $R(v) := R(v),\ S(v) := S(v)$.

After such a redefinition of the labels, the vertex k_j is assumed to be visited and the graph G_{j+1} is constructed by removing the vertex k_j from the set of vertices V_j of the graph G_j and from the set of edges E_j incident k_j in E_j. As a result, a graph G_{j+1} is constructed with a set of vertices V_{j+1} and a set of edges E_{j+1}.

In the proposed modification of Dijkstra's algorithm, a list of the first and second labels is mapped to each vertex $v \in V$. At the end of the algorithm, the first label is $R(v)$ is equal to the length of the shortest path from the vertex $\mathbf{0}$ to the vertex v. The second label $S(v)$ is the set of neighbours u of the vertex v in the graph G, through which all the last edges (u, v) of the shortest paths from $\mathbf{0}$ to the vertex v pass. In the list of vertex labels $v \in V$, the first label is separated from the second label by the icon ; .

For the first label, this recurrent procedure coincides with Dijkstra's algorithm [3], [6]. And for the second label, using matinduction, it is easy to prove that for each j all the last edges (u, v) of the shortest paths from $\mathbf{0}$ to the vertex v form the set of vertices $S(k_j)$. Thus, the presented modification of Dijkstra's algorithm allows using the second vertex labels to determine all the shortest routes in the graph G from the starting vertex $\mathbf{0}$ to the remaining vertices of the graph.

Example 1. Figure 1 shows an example of a weighted undirected graph (a special case of a digraph with the same weights of multi directional edges) with vertex numbers in circles. Table 1 shows the lists of labels for the vertices of this graph.

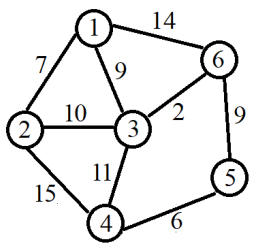


Figure 1. An example of a weighted graph.

Vertices	1	2	3	4	5	6
Label Lists	0	7;1	9; 1	20; 3	20; 6	11;3

Table 1. Lists of labels at the vertices of a weighted graph.

Remark 1. It should be noted that the presence of several vertices in the second label may be a slight change in the weights of the edges eliminated. It follows that in the case of the general situation in the (non-degenerate case), the second label at the vertices of a weighted graph can consist of only one vertex. This procedure reduces the computational complexity and the amount of memory required to implement this algorithm. However, the presence of several vertices in the second label provides the decision maker with additional degrees of freedom and makes these decisions more reliable. With small changes in the edge weights, in some cases it is even possible, on the contrary, to increase the number of vertices in the second label to increase reliability. This circumstance can be used, for example, in transport systems.

3. Edges of shortest paths in an unweighed graph

Let's now consider an unweighed graph G with a set of vertices V and a set of edges E. For such a graph, let's put $s(e) = 1,\ e \in E$. Therefore, we can proceed to the modified Dijkstra algorithm described in the previous section. However, it is more convenient to use the wave algorithm. Let $V_0 = \{a\}$ and denote V_1 the set of all neighbors of the vertex a. Let $V^1 = V \setminus V_1$, define V_2 a collection of vertices of the set V^1, which include edges from the vertices of the set V_1. Next, let's put $V^2 = V \setminus (V_1 \cup V_2)$. Continuing by induction , we get $V^k = V \setminus (V_1 \cup V_2 \cup ... \cup V_k)$,then V_{k+1} is the collection of all vertices of the set V^k, which include edges from the vertices of the set V_k. It is possible to assume the length of the shortest path from the vertex **0** to any other vertex v of the graph (the first vertex label v) to be equal to k, if $v \in V_k$. Then the second vertex label $v \in V_k$ becomes the set of vertices u of the set V_{k+1} such that the edges $(v, u) \in E$.

Figure 2 shows an example of an unweighed graph. Here, the starting vertex is marked in green, the vertices of the set V_1 are orange, and the vertices of the set V_2. are blue in the table. 2 lists of graph vertex labels are given.

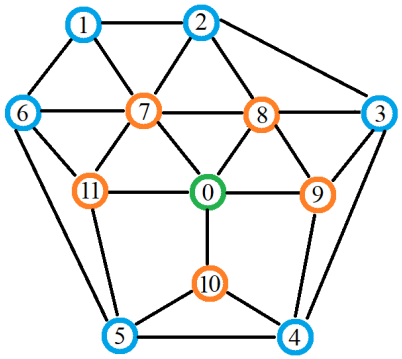


Figure 2. An example of an unweighted graph.

Vertices	0	7	8	9	10	11	1	2	3	4	5	6
Label Lists	0	1; 0	1; 0	1; 0	1; 0	1; 0	2; 7	2;7,8	2; 8,9	2; 9,10	2; 10,11	2; 7,11

Table 2. Lists of labels at the vertices of an unweighed graph.

Remark 2. For an unweighted graph, it is also sufficient to write out a set of sets $V_k,\ k = 1, \ldots,$ for which it is not difficult to restore the second vertex labels using the graph G. The presence of several vertices in the second label plays an important role in increasing the reliability of connections between proteins in the protein network represented by an unweighed graph.

4. Shortest paths in a planar graph

Consider the planar graph Γ[9, Chapter 1] and denote $U_0 = \{u\}$ the set of its faces with the exception of the outer face. We will look for the shortest paths from the faces of the set U_0 to

the outer face. Here, the shortest path is understood to be such a polyline that intersects the minimum number of boundaries between the faces. Each face $u \in U_0$ is mapped to a subset of faces $L(u) \subseteq U_0$, touching the boundaries with the face u. Let's call $L(u)$ the set of neighbours of the face u. The face adjacent to the outer face is called the boundary, and not adjacent to the outer faces - internal. Let's denote $V_0 \subseteq U_0$ the set of boundary faces and $U_1 = U_0 \setminus V_0$ - the set of inner faces.

Define recurrent algorithm

$$U_{K+1} = \{u \in U_k : \ l(u) \subseteq U_k\}, \ V_k = U_k \setminus U_{k+1}. \tag{1}$$

We continue this recursion until the step n, when for the first time $U_{n+1} = \emptyset$ or $U_{n+1} = U_n$. It is not difficult to prove that for any $k \leq n, \ u_k \in V_k$, there exists a face $v_{k-1} \in V_{k-1}$ such that $v_{k-1} \in S(v_k)$. Moreover, for any $j > 1$ intersection of $L(v_k) \cap V_{k-j} = \emptyset$. Thus, the minimum number of boundaries that a poly line must cross from the face $u_k \in V_k$ to the outer face is $k+1$. Therefore, $k+1$ can be determined by the first label of the face u_k. In turn, the second label can be taken as $L(u_k) \cap V_{k-1}$. Moreover, the directed edges coming out of the vertex u_k are represented as $(u_k, u_{k-1}), \ u_{k-1} \in L(u_k) \cap V_{k-1}$. It should be noted that, unlike the first two paragraphs for a planar graph, the second label is defined differently.

Example 2. Figure 3 shows an example of a planar graph in which the faces of the set V_0 are colored blue, the faces of the set V_1 are orange, and the faces of the set V_2 are green. Table 3 lists the labels of these faces.

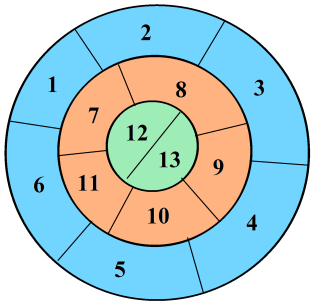


Figure 3. An example of a planar graph with faces highlighted on it.

Vertices	12	13	7	8	9	10	11	1,2,3,4,5,6
Label Lists	3; 7,8,11	3; 8,9,10,11	2;1,2,6	2; 2,3	2; 3,4	2; 4,5	2;5,6	1

Table 3. Lists of labels at the vertices of a planar graph.

Remark 3. In a planar graph, the presence of several vertices in the second label is quite common. Moreover, in cartographic applications it is not necessary to write out the second labels, it is enough to specify the sets $V_k, \ k = 0, \dots, n$, and represent these sets as shaded faces.

5. Reliability and stability of shortest paths

However, building label lists increases the amount of memory required to implement this algorithm. Therefore, we can replace the second label with the number of neighbours u of the vertex v in the graph G, through which all the last edges (u, v) of the shortest paths from $\mathbf{0}$ to the vertex v pass, denoting this number $N(v)$. To do this, you can only slightly change the algorithm $(\mathbf{A})$, $(\mathbf{B})$, $(\mathbf{C})$, replacing it with the following algorithm.

First we put $N(\mathbf{v}) = 0; \ v \in V, \ R(\mathbf{0}) = 0, \ R(v) = \infty, \ v \in V, \ v \neq \mathbf{0}; \ G_0 = G, \ V_0 = V, \ E_0 = E$. Next, for $0 \leq j \leq n$, select the vertex k_j from the condition $R(k_j) = \min\limits_{v \in V_j} R(v)$, redefine the labels $R(v), \ N(v), \ v \in V_j, \ v \neq k_j$ and construct the graph G_{j+1} from the following conditions.

$(\mathbf{A}.\mathbf{1})$. If $R(k_j) + s(k_j, v) < R(v)$, then $R(v) := R(k_j) + s(k_j, v), \ N(v) := 1$.

(**B**.**1**). If $R(k_j)+s(k_j,v)=R(v)$, then $R(v):=R(v),\ N(v):=N(v)+1$.
(**C**.**1**). If $R(k_j)+s(k_j,v)>R(v)$, then $R(v):=R(v),\ N(v):=N(v)$.

After such a redefinition of the labels, the vertex k_j is assumed to be visited and the graph G_{j+1} is constructed by removing the vertex k_j from the set of vertices V_j of the graph G_j and from the set of edges E_j incident k_j E_j. As a result, a graph G_{j+1} is constructed with a set of vertices V_{j+1} and a set of edges E_{j+1}.

Obviously, the set of values $\{N(v),\ v\in V\}$ can be considered as a characteristic of the reliability of the shortest paths of the graph G. After all, it defines each vertex of the graph in terms of the number of edges of the shortest paths entering it. On the other hand, this set is quite simple to calculate and practically does not require additional memory.

Let us now turn to the question of the stability of shortest paths with variations in edge weights. This question is conveniently related to the computational complexity of the algorithm for determining the reliability characteristics of $N(v)$. For this purpose, the fluctuation value $\varepsilon>0$ is set and the following modification of the previously introduced algorithm is constructed.

First we put $N(v)=0;\ v\in V,\ R(\mathbf{0})=0,\ R(v)=\infty,\ v\in V,\ v\neq\mathbf{0};\ S(v)=\emptyset,\ v\in V;\ G_0=G,\ V_0=V,\ E_0=E$. Then for $0\le j\le n$ we select the vertex k_j from the condition $R(k_j)=\min_{v\in V_j}R(v)$, redefine redefine labels $R(v),\ N(v),\ v\in V_j,\ v\neq k_j$ and construct a graph G_{j+1} from the conditions.

(**A**.**2**). If $R(k_j)+s(k_j,v)<R(v)-\varepsilon$, then $R(v):=R(k_j)+s(k_j,v),\ N(v):=1$.
(**B**.**2**). If $|R(k_j)+s(k_j,v)-R(v)|\le\varepsilon$, then $R(v):=R(v),\ N(v):=N(v)+1$.
(**C**.**2**). If $R(k_j)+s(k_j,v)>R(v)+\varepsilon$, then $R(v):=R(v),\ N(v):=N(v)$.

After such a redefinition of the labels, the vertex k_j is assumed to be visited and the graph G_{j+1} is constructed by removing the vertex k_j from the set of vertices V_j of the graph G_j and from the set of edges E_j incident k_j in E_j. As a result, a graph G_{j+1} is constructed with a set of vertices V_{j+1} and a set of edges E_{j+1}. This modification of Dijkstra's algorithm allows us to calculate how much edge weight perturbations affect the final result of determining the value of $N(v),\ v\in V$.

Let's now give a table of values of $N(v),\ v\in V$ for examples 1, 2.

Vertices	0	7	8	9	10	11	1	2	3	4	5	6
Lists of $N(v)$	0	1	1	1	1	1	1	2	2	2	2	2

Table 4. Values $N(v)$ at the vertices of an unweighed graph for example 1.

Vertices	12	13	7	8	9	10	11	1,2,3,4,5,6
Values $N(v)$	3	4	3	2	2	2	2	1

Table 5. Values $N(v)$ for faces of planar graph in example 2.

6. Conslusion

In this paper, the emphasis is not on determining the lengths of the shortest paths in the graph, but on determining the shortest paths themselves. As a matter of fact, the paper does not give an algorithm for enumerating all shortest paths. The lists of labels introduced in the work only allow you to determine the shortest paths from some vertex to the starting vertex or from some face to the outer face. However, the proposed technique of labelling lists, especially the second label, makes it possible to follow well-known algorithms for determining the lengths of shortest paths: Dijkstra algorithm, wave algorithm, etc. This allows, by introducing a second label, to recursively determine its change when moving to a new vertex/face of the graph. A feature of the proposed algorithms is the introduction of a second list label into them, which at the same time is an empty or non-empty set of graph vertices. In this paper, the emphasis is not on evaluating the computational complexity of the algorithm and on estimating the amount of memory required for implementation. Therefore, this algorithm can be used for relatively small weighted graphs. It should be noted that the requirements for the uniqueness of vertices in the second label of the graph, on the one hand, reduces the computational complexity of the presented algorithms and

the required amount of memory. On the other hand, the presence of several vertices in the second label makes it possible to increase the reliability of the decision about the shortest paths in the graph. Moreover, in various applications (transport systems, cartography, biotechnology) these conditions have different effects on the convenience of the above algorithms.

The research was carried out within the state assignment for IAM FEB RAS (N 075-00459-24-00).

References

[1] Tsitsiashvili G. Sh., Bocharnikov V. N., Krasnopeev S. M. (2024). Zoning of areas of the region in proximity to the external border. *Mathematical game theory and its applications*, 16 (1), p. 78-91. (In Russian).
[2] Tsitsiashvili G. Sh., Bulgakov V. P. (2022). New Applied Problems in the Theory of Acyclic Digraphs. *Mathematics*, 10(1), 45.
[3] Dijtstra E. W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik*, 1 (1), p. 269-271.
[4] Bellman R. (1958). On a Routing Problem. *Quarterly of Applied Mathematics*, 16 (1), p 87-90.
[5] Ford L. R., Jr., Fulkerson D. R. Flows in Networks. Princeton University Press, 1962.
[6] Cormen Th. H., Leiserson Ch. E., Rivest R. L. Introduction to Algorithms. MIT Press and McGraw-Hill, 1990.
[7] https://ru.wikipedia.org/wiki/
[8] Goloveshkin A. V., Mikhalkovich S. S. (2022). Stable algorithmic binding to an arbitrary section of the program code. *Software systems: theory and applications*, 13 (1), p. 3-33. (In Russian).
[9] Prasolov V. V. Elements of combinatorial and differential topology. Moscow: MCNMO, 2004. (In Russian).

THE WEIGHTED SABUR DISTRIBUTION WITH APPLICATIONS OF LIFE TIME DATA

Suvarna Ranade [1], Aafaq A. Rather[2,*]

•

[1,2]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[1]Dr. Vishwanath Karad MIT World Peace University, Pune
[1]suvarnamacs@gmail.com, [2,*]Corresponding Author: aafaq7741@gmail.com

Abstract

In this paper, we propose a weighted version of Sabur distribution. The Stability of distribution are studied with structural properties, moments generating functions, likelihood ratio test, entropy measures, order statistics and Fisher's information matrix. The new model provides flexibility to analyse complex real data. Application of model on real data sets shows that the weighted Sabur distribution is quite effective. In this paper we utilize Monte Carlo simulation to evaluate the effectiveness of estimators. We used our weighted Sabur distribution on two real data set, Anderson-Darling and Cramer-von Mises class of quadratic EDF statistics utilize to test whether a given sample of data is drawn from a weighted Sabur distribution.

Key Words: Weighted distribution, Sabur distribution, Entropy, Order statistics.

1.Introduction

The concept of weighted distribution was first utilized by Fisher 1934 [8] in study of effect on form of distribution of recorded observations because of methods of ascertainment. The same concept was demonstrated and formulated by Rao 1965 [18] on modelling statistical data. The weighted distribution reduces to length biased distribution when the weight function considers only the length of units. The concept of length biased sampling was introduced by Cox [7] and Zelen [21]. Many newly introduced distributions along with their weighted versions exist in literature whose statistical behaviour is extensively studied during decades.

In recent years, researchers have made significant advancements in the study of the Lindley distribution and have proposed various one and two-parameter distributions to model complex datasets effectively. A notable contribution was made by Ghitney et al. [12], who conducted an extensive study on the Lindley distribution. They demonstrated that the Lindley distribution outperforms the exponential distribution when applied to modelling waiting times before bank customer service. Additionally, they highlighted that the contours of the hazard rate function for the Lindley distribution show an increasing trend, while the mean residual life function is a decreasing function of the random variable. Many authors modify the Lindley distribution by introducing new parameters and evaluating performance of these extended distribution with various dataset.

In this paper, we introduce a new distribution with three parameter, namely as weighted Sabur distribution with the hope that it provides more flexibility in various applications of Reliability, Survival Analysis, Biology etc.

2. Weighted Sabur Distribution

2.1 Density and Cumulative Density functions

The probability density function (pdf) of the Sabur distribution with two parameters α and β is defined as

$$f(x,\alpha,\beta) = \frac{\beta^2}{\alpha\beta+\beta^2+1}(\alpha+\beta+\frac{\beta}{2}x^2)e^{-\beta x} \text{ x } > 0,\ \alpha,\beta\ > 0 \quad (1)$$

Suppose X is a non-negative random variable with pdf f(x). Let w(x) be the non-negative weight function, then the pdf of the weighted random variable X_w is given by

$$f_w(x) = \frac{w(x)f(x)}{E(w(x))}, x > 0$$

Where w(x) is a non-negative weight function and

$$E(w(x)) = \int w(x)f(x)\,dx$$

.In this paper, we will consider the weight function was w(x) = xc, and using the definition of weighted distribution, the pdf of the weighted Sabur distribution is given as

$$f_w(x) = \frac{x^c f(x)}{E(x^c)}, c\ > 0 \quad (2)$$

Expected value is defined as

$$E(x^c) = \int_0^\infty x^c f(x)\ dx$$

$$E(x^c) = \frac{\beta^2}{\alpha\beta+\beta^2+1}\left[\frac{\alpha+\beta}{\beta^{c+1}}\Gamma c+1+\frac{\Gamma c+3}{2\beta^{c+2}}\right] \quad (3)$$

Substituting equation (1) and (3) in equation (2) we obtain the density function of weighted Sabur distribution as follows

$$f_w(x,\alpha,\beta) = \frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)} \quad (4)$$

and the cumulative density function (cdf) of weighted Sabur distribution is obtained by

$$F_w(x) = \int_0^x f_w(x)dx$$

$$F_w(x)\ = \int_0^x \frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\,dx \quad (5)$$

After simplification, the cdf of the weighted Sabur distribution is given by

$$F_w(x) = \frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)} \quad (6)$$

Fig.1 and Fig. 2 visually illustrates the pdf and cdf of Weighted Sabur Distribution.

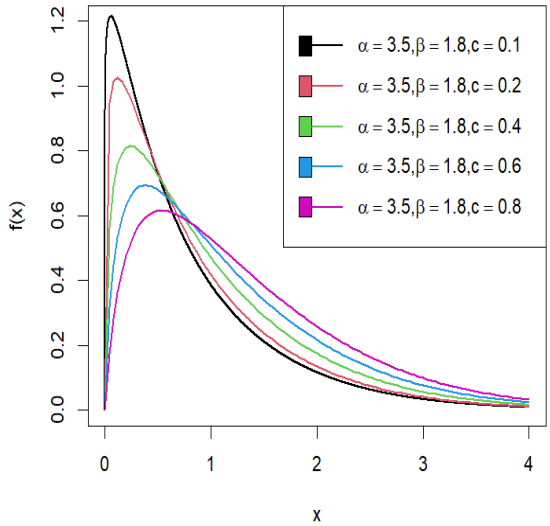


Fig. 1: pdf plot of Weighted Sabur distribution

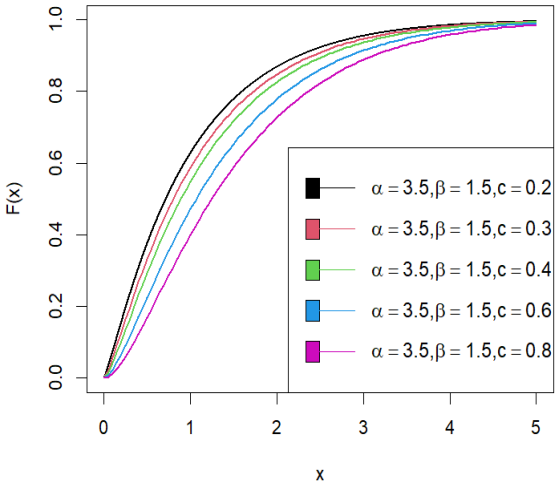


Fig. 2: pdf plot of Weighted Sabur distribution

2.2 Survival, Hazard and Reversed Hazard Functions

In this section we discuss about the survival function, hazard and reverse hazard functions of the weighted Sabur distributions. The survival function or the reliability function of weighted Sabur distribution is given by

$$S(x) = 1 - F_w(x)$$

$$S(x) = 1 - \left(\frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right) \quad (7)$$

The hazard function is also known as the hazard rate function, instantaneous failure rate or force of mortality and is given for the weighted Sabur distribution as

$$h(x) = \frac{f_w(x)}{s(x)}$$

$$h(x) = \frac{\frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}}{1-\left(\frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right)} \quad (8)$$

$$h(x) = \frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{(2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3))-(2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x))} \quad (9)$$

The reverse hazard function of the weighted Sabur distribution is given by

$$h_r(x) = \frac{f_w(x)}{F_w(x)}$$

$$h_r(x) = \frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)} \quad (10)$$

Fig. 3 and Fig. 4 depicts the graphical survival function and Hazard function plot of Weighted Sabur distribution.

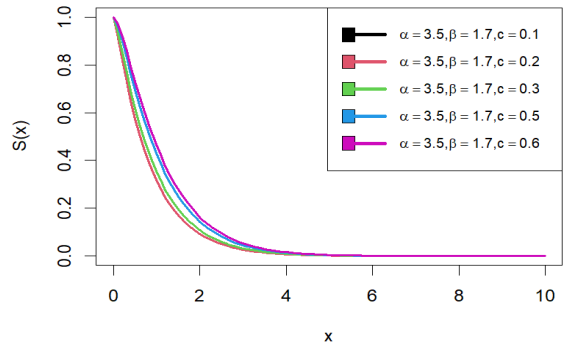


Fig. 3: Plots of Survival function

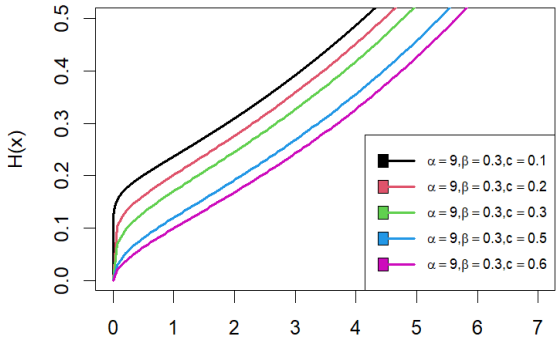


Fig. 4: Plots of hazard rate function

3. Structural properties

In this section we investigate various structural properties of the weighted Sabur distribution. Let X denote the random variable of weighted Sabur distribution with parameters α, β and c, then its r th order moment about origin is given by

$$E(x^r) = \mu_r' = \int_0^\infty x^r f_w(x)\, dx$$

$$E(x^r) = \int_0^\infty x^r \frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\, dx \tag{11}$$

After simplifying the expression, we get

$$E(x^r) = \frac{[2\beta(\alpha+\beta)(\Gamma r+c+1)+(\Gamma r+c+3)]}{\beta^r[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]} \tag{12}$$

Putting r = 1, we get the expected value of weighted Sabur distribution as follows

$$E(x) = \frac{[2\beta(\alpha+\beta)(\Gamma c+2)+(\Gamma c+4)]}{\beta[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]} \tag{13}$$

Put r = 2, we obtained second moment as

$$E(x^2) = \frac{[2\beta(\alpha+\beta)(\Gamma c+3)+(\Gamma c+5)]}{\beta^2[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]} \tag{14}$$

The variance of Weighted Sabur distribution is calculated as

$$V(x) = E(x^2) - [E(x)]^2$$

$$V(x) = \frac{2[\beta(\alpha+\beta)(\Gamma c+3)+(\Gamma c+5)]}{\beta^2[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]} - \left[\frac{[2\beta(\alpha+\beta)(\Gamma c+2)+(\Gamma c+4)]}{\beta[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]}\right]^2 \tag{15}$$

3.1 Harmonic mean

The harmonic mean of the weighted Sabur distribution of random variable x can be written as

$$H = E\left(\frac{1}{x}\right) = \int_0^\infty \frac{1}{x} f_w(x) dx$$

$$H = \int_0^\infty \frac{1}{x} \frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}dx \tag{16}$$

After simplification we get

$$\mathrm{H} = \frac{\beta[2\ \beta(\alpha+\beta)\Gamma c+\Gamma c+2]}{2\ \beta(\alpha+\beta)\Gamma c+1+\Gamma c+3} \tag{17}$$

3.2 Moment generating function and characteristic function

Let X have a weighted Sabur distribution, then the Moment generating function of X is obtained as

$$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} f_w(x)dx$$

Using Tayler's series, we obtain

$$M_X(t) = E(e^{tx}) = \int_0^\infty \left(1 + tx + \frac{(tx)^2}{2!} + ..\right) f_w(x)dx \tag{18}$$

$$M_X(t) = \int_0^\infty \sum_{i=0}^\infty \frac{t^i}{i!} x^i\, f_w(x)dx \tag{19}$$

$$M_X(t) = \sum_{j=0}^\infty \frac{t^j}{j!} E(x^j)dx$$

$$M_X(t) = \sum_{j=0}^\infty \frac{t^j}{j!} \frac{2[\beta(\alpha+\beta)(\Gamma j+c+1)+(\Gamma \mathrm{j}+\mathrm{c}+3)]}{\beta^j[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]} \tag{20}$$

Similarly, the characteristic function of weighted Sabur distribution of random variable X can obtain as

$$\Phi_X(t) = M_X(it) = \sum_{j=0}^\infty \frac{(it)^j}{j!} \frac{2[\beta(\alpha+\beta)(\Gamma j+c+1)+(\Gamma \mathrm{j}+\mathrm{c}+3)]}{\beta^j[2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)]} \tag{21}$$

4. Likelihood Ratio Test

Let X_1, X_2, X_3….be a random sample from the weighted Sabur distribution, we use the hypothesis

$$H_0 : f(x) = f(x{:}\,\alpha,\beta)\ \ against\ \ H_1{:}\, f(x) = f_w(x{:}\,\alpha,\beta,c)$$

In order to test whether the random sample of size n comes from the Sabur distribution or weighted Sabur distribution, we will use following statistics

$$\Delta = \frac{L_1}{L_0} = \prod_{i=1}^n \frac{f_w(x;\alpha,\beta,c)}{f(x;\alpha,\beta)} \tag{22}$$

$$\Delta = \prod_{i=1}^n {x_i}^c \frac{2\beta^c(\alpha\beta+\beta^2+1)}{2\beta(\alpha+\beta)\Gamma c+1\ +\Gamma c+3} \tag{23}$$

$$\Delta = A^n \prod_{i=1}^n {x_i}^c \quad \text{where}$$

$$A = \frac{2\beta^c(\alpha\beta+\beta^2+1)}{2\beta(\alpha+\beta)\Gamma c+1\ +\Gamma c+3} \tag{24}$$

We reject the null hypothesis, if

$$\Delta = A^n \prod_{i=1}^n {x_i}^c > \mathrm{k}$$
$$\Delta^* = \prod_{i=1}^n {x_i}^c > \mathrm{k}\, A^n$$

For large sample size n, 2log Δ is distributed as chi square distribution with one degree of freedom and also p-value is obtained from the chi-square distribution. Thus, we reject the null hypothesis, when the probability value is given by

$$P(\Delta^* > a^*)$$

Where a^* is less than a specified level of significance and $\prod_{i=1}^n {x_i}^c$ is the observed value of the statistics Δ^*.

5. Entropy Measures

The concept of entropy is important in different areas such as probability and statistics, physics, communication theory and economics. Entropy measures quantify the diversity, uncertainty or randomness of a system. Entropy of a random variable X is measure of variation of the uncertainty.

5.1 Renyi Entropy

It was proposed by Renyi (1957). The Renyi entropy of order ξ for a random variable X is given by

$$e(\xi) = \frac{1}{1-\xi}\log\left(\int_0^\infty f^{\xi}(x)\,dx\right) \text{ where } \xi > 0 \text{ and } \xi \neq 1$$

$$e(\xi) = \frac{1}{1-\xi}\log\left(\int_0^\infty \left(\frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right)^{\xi} dx\right) \tag{25}$$

After simplifying the equation, we get

$$e(\xi) = \frac{1}{1-\xi}\log\left(\left(\frac{2\beta^{c+2}}{2\beta(\alpha+\beta)\Gamma c+1+\Gamma c+3}\right)^{\xi}\sum_{i=0}^{\infty}\binom{\xi}{i}(\alpha+\beta)^{\xi-i}\left(\frac{\beta}{2}\right)^{i}\frac{\Gamma(c\xi+2i+1)}{\beta\xi^{c\xi+2i+1}}\right) \tag{26}$$

5.2 Tsallis Entropy

A generalization of Boltzmann-Gibbs(B-G) statistical mechanics initiated by Tsallis has focussed a great deal to attention. This generalization of B-G statistics was proposed firstly by introducing the mathematical expression of Tsallis entropy (Tsallis, 1988) for a continuous random variable. Tsallis entropy of order λ of the weighted Sabur distribution is given by

$$S_\lambda = \frac{1}{\lambda-1}\left(1-\int_0^\infty f^{\lambda}(x)\,dx\right) \tag{27}$$

$$S_\lambda = \frac{1}{\lambda-1}\left(1-\int_0^\infty \left(\frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right)^{\lambda} dx\right) \tag{28}$$

After simplifying the expression, we get

$$S_\lambda = \frac{1}{\lambda-1}\left[\left(1-\left(\frac{2\beta^{c+2}}{2\beta(\alpha+\beta)\Gamma c+1+\Gamma c+3}\right)^{\lambda}\right)\sum_{i=0}^{\infty}\binom{\lambda}{i}(\alpha+\beta)^{\lambda-i}\left(\frac{\beta}{2}\right)^{i}\frac{\Gamma(c\lambda+2i+1)}{\beta\lambda^{c\lambda+2i+1}}\right] \tag{29}$$

6. Order Statistics

Let $X_{(1)}$, $X_{(2)}$, $X_{(3)}$…...$X_{(n)}$ be the order statistics of a random sample $X_1, X_2, X_3 \ldots X_n$ drawn from the continuous population with pdf $f_x(x)$ and cdf $F_x(x)$ then the pdf of r th order statistic X (r) is given by

$$f_{x(r)}(x) = \frac{n!}{(r-1)!(n-r)!}f_x(x)[F_x(x)]^{r-1}[1-F_x(x)]^{n-r} \tag{30}$$

Substituting equation (4) and (5) in equation (6), the pdf of order statistics X(r) of the weighted Sabur distribution is given by

$$f_{x(r)}(x) = \frac{n!}{(r-1)!\,(n-r)!}\left(\frac{2\beta^{c+2}x^c\left(\alpha+\beta+\frac{\beta}{2}x^2\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+\ (\Gamma c+3)}\right)$$
$$\times\left(\frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)+\ (\Gamma c+3)}\right)^{r-1}$$
$$\times\left(1-\ \left(\frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)\ +\ (\Gamma c+3)}\right)\right)^{n-r} \tag{31}$$

Therefore, the pdf of the higher order statistics X(n) can be obtained as

$$f_{x(n)}(x) = n\left(\frac{2\beta^{c+2}x^{c}\left(\alpha+\beta+\frac{\beta}{2}x^{2}\right)e^{-\beta x}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right)\times\left(\frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right)^{n-1} \quad (32)$$

And the pdf of the first order statistics $X_{(1)}$ can be obtained as

$$f_{x(1)}(x) = n\left(\frac{x^{c}\left(\alpha+\beta+\frac{\beta}{2}x^{2}\right)e^{-\beta x}}{\left(\frac{\alpha+\beta}{\beta^{2}}\right)\Gamma(c+1)+\frac{1}{2\beta}\Gamma(c+3)}\right)\times\left(1-\left(\frac{2\beta(\alpha+\beta)\gamma(c+1,\beta x)+\gamma(c+3,\beta x)}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)}\right)\right)^{n-1} \quad (33)$$

7. Income Distribution Curve

The Bonferroni and the Lorenz curves are not only used in economics in order to study the income and poverty, but it is also being used in other fields like reliability, medicine and demography. The Bonferroni and Lorenz curves are given by

$$B(p) = \frac{1}{p\mu_1'}\int_0^q x\,f(x)dx \quad \text{and}$$

$$L(p) = PB(p) = \frac{1}{\mu_1'}\int_0^q x\,f(x)dx$$

Here, we define the first raw moments as

$$\mu'_1 = \frac{[2\beta(\alpha+\beta)\Gamma c+2+\Gamma c+4]}{\beta[2\beta(\alpha+\beta)\Gamma c+1+\Gamma c+3]} \quad (34)$$

And q = $F^{-1}(p)$, Then we have

$$B(p) = \frac{2\beta(\alpha+\beta)\gamma(c+2,Bq)+\gamma(c+4,Bq)}{p\left(2\beta(\alpha+\beta)\Gamma(c+2)+\Gamma(c+4)\right)} \quad (35)$$

$$L(p) = \frac{2\beta(\alpha+\beta)\gamma(c+2,Bq)+\gamma(c+4,Bq)}{(2\beta(\alpha+\beta)\Gamma(c+2)+\Gamma(c+4))} \quad (36)$$

8. Estimation

We will discuss the maximum likelihood estimators (MLEs) of the parameters of the weighted Sabur distribution. Consider $X_1, X_2, X_3 \ldots, X_n$ be the random sample of size n from the weighted Sabur distribution, then the likelihood function is given by

$$L(x;\ \alpha,\beta,c=1) = \prod_{i=1}^{n} {x_i}^{c}\frac{2\beta^{c+2}\left(\alpha+\beta+\frac{\beta}{2}{x_i}^{2}\right)e^{-\beta x_i}}{2\beta(\alpha+\beta)(\Gamma c+1)+(\Gamma c+3)} \quad (37)$$

$$L(x;\ \alpha,\beta,c) = \frac{2^{n}\beta^{n(c+2)}}{(2\beta(\alpha+\beta)\Gamma c+1+\Gamma c+3)^{n}}\prod_{i=1}^{n}{x_i}^{c}\left(\alpha+\beta+\frac{\beta}{2}{x_i}^{2}\right)e^{-\beta x_i} \quad (38)$$

The loglikelihood function is obtained as

$$Log\ L = nlog2 + n(c+2)log\beta - nlog(2\beta(\alpha+\beta)\Gamma c+1+\Gamma c+3) + clog\sum x_i + \sum log(\alpha+\beta\frac{\beta}{2}\sum {x_i}^{2}) - \beta\sum x_i \quad (39)$$

The MLEs of α, β, c can be obtained by differentiating Log L with respect to α, β, c and must satisfy the normal equation.

$$\frac{\partial log\ L}{\partial\beta} = \frac{n(c+2)}{\beta} - \frac{n}{\beta} - \frac{2n}{\beta} + \frac{1+\frac{1}{2}\sum {x_i}^{2}}{\alpha+\beta+\frac{\beta}{2}\sum {x_i}^{2}} - \sum x_i = 0 \quad (40)$$

$$\frac{\partial log\ L}{\partial\alpha} = \left[-\frac{n}{\alpha} + \frac{1}{\alpha+\beta+\frac{\beta}{2}\sum {x_i}^{2}}\right] = 0 \quad (41)$$

$$\frac{\partial log\ L}{\partial c} = nlog\beta - nlog\Psi(2\beta(\alpha+\beta)\Gamma c+1+\Gamma c+3) + log\sum x_i = 0 \quad (42)$$

Where $\Psi(.)$ is the digamma function. Because of the complicated form of the above likelihood equations, algebraically it is very difficult to solve the system of nonlinear equations. Therefore, we use R and Wolfram Mathematica for estimating the required parameters. To obtain confidence interval we use the asymptotic normality results. We have that, if $\hat{\lambda} = (\hat{\alpha}, \hat{\beta}, \hat{c})$ denotes the MLE of $\lambda = (\alpha, \beta, c)$ we can state the results as follows

$$\left(\hat{\lambda} - \lambda\right) \rightarrow N_3\left(0, I^{-1}(\lambda)\right)$$

Where I(λ) is Fisher's Information matrix given by

$$I(\lambda) = -\frac{1}{n}\begin{pmatrix} E\left(\frac{\partial^2 \log l}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 \log l}{\partial \alpha \partial \beta}\right) & E\left(\frac{\partial^2 \log l}{\partial \alpha \partial c}\right) \\ E\left(\frac{\partial^2 \log l}{\partial \beta \partial \alpha}\right) & E\left(\frac{\partial^2 \log l}{\partial \beta^2}\right) & E\left(\frac{\partial^2 \log l}{\partial \beta \partial c}\right) \\ E\left(\frac{\partial^2 \log l}{\partial c \partial \alpha}\right) & E\left(\frac{\partial^2 \log l}{\partial c \partial \beta}\right) & \left(\frac{\partial^2 \log l}{\partial c^2}\right) \end{pmatrix} \quad (43)$$

Here we define

$$\frac{\partial^2 \log L}{\partial \beta^2} = -\frac{(cn-n)}{\beta^2} - \frac{\left(1+\frac{1}{2}\sum {x_i}^2\right)^2}{(\alpha+\beta+\frac{\beta}{2}\sum {x_i}^2)^2} \quad (44)$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -\frac{n}{\alpha^2} - \frac{1}{(\alpha+\beta+\frac{\beta}{2}\sum {x_i}^2)^2} \quad (45)$$

$$\frac{\partial^2 \log L}{\partial c^2} = -n\Psi'\left((2\beta(\alpha+\beta)\Gamma c + 1 + \Gamma c + 3)\right) \quad (46)$$

$$\frac{\partial^2 \log L}{\partial c \partial \beta} = \frac{n}{\beta}$$

$$\frac{\partial^2 \log L}{\partial c \partial \alpha} = -n\Psi'\left((2\beta(\alpha+\beta)\Gamma c + 1 + \Gamma c + 3)\right) \quad (47)$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \alpha} = -\frac{\left(1+\frac{\sum {x_i}^2}{2}\right)^2}{(\alpha+\beta+\frac{\beta}{2}\sum {x_i}^2)^2} \quad (48)$$

where $\Psi'(.)$ is the first order derivative of digamma function. Since λ being unknown, we estimate $I^{-1}(\lambda)$ by $I^{-1}\left(\hat{\lambda}\right)$ and this can be used to obtain asymptotic confidence intervals for α, β, c.

9. Simulation

Simulations offer a comprehensive and flexible approach to comprehending the behaviour of maximum likelihood estimators across diverse sample sizes. This understanding serves as a valuable guide for enhanced decision-making, risk mitigation, and the enhancement of reliability and efficiency in statistical analysis within various domains, including finance, healthcare, and engineering. By utilizing simulations, we gain the ability to anticipate the behaviour of maximum likelihood estimators across a broad spectrum of sample sizes, even those challenging to attain in practical scenarios. This predictive capability aids in grasping how the bias, variance, and efficiency of the estimator evolve with fluctuations in sample size. Simulations play a crucial role in identifying the optimal sample size for the application of maximum likelihood estimators. Our investigation has delved into the performance of ML estimators across different sample sizes, namely n=25, 50, 75, 100, 200, and 300.

The inverse cumulative distribution function (cdf) technique was utilized for data simulation using the R-software, and this process was iterated 700 times to compute bias, variance, and mean squared error (MSE). Analysis of Table 1 reveals a consistent trend across various parameter values and sample sizes of the Weighted Sabur distribution, indicating a decrease in variance, bias, and MSE as the sample size increases. The diminishing bias suggests that Maximum Likelihood (ML) estimation approaches the true parameter values with an expanding sample size. Simultaneously, the declining

variance indicates that the estimators exhibit increased precision and stability with larger sample sizes, displaying reduced variability across repeated simulations.

Table 1: *Estimation of Bias, Variance and MSE for different sample sizes*

n	$\beta=1$			$\alpha=1.5$			C=2		
	Bias	Variance	MSE	Bias	Var.	MSE	Bias	Var.	MSE
20	1.288	0.54226	2.20300	-1.4594	0.01567	2.1456	4.01555	11.02994	27.15457
30	1.452	0.36401	2.47277	-1.4999	0	2.2470	5.12954	7.216834	33.52903
50	1.520	0.21295	2.52622	-1.4999	0	2.2470	5.48211	5.147884	35.20144
75	1.400	0.08699	2.04928	-1.4999	0	2.2470	4.84093	0.868977	24.30359
100	1.626	0.19394	2.83889	-1.4999	0	2.2470	5.64059	4.11263	35.92895
200	1.458	0.11823	2.24448	-1.499	0	2.2470	5.08705	2.425165	28.30325
300	1.436	0.02933	2.09309	-1.499	0	2.2470	5.18147	0.397811	27.24549
	$\beta=1$			$\alpha=2$			C=1.2		
	Bias	Variance	MSE	Bias	Var.	MSE	Bias	Var.	MSE
20	1.775	1.155957	3.91724	-1.999	0	3.9960	3.91724	7.448396	22.79317
30	1.041	0.081519	1.16448	-1.9999	0	3.9603	3.67425	6.102347	19.60251
50	1.216	0.132173	1.61203	-1.9999	0	3.9960	2.96087	1.464504	10.23129
75	1.271	0.16421	1.78079	-1.999	0	3.9960	2.65069	1.099146	8.12533
100	1.229	0.102218	1.61407	-1.999	0	3.9960	2.68728	0.816670	8.038174
200	1.079	0.073882	1.23887	-1.9999	0	3.9960	2.24187	0.459616	5.485583
300	1.046	0.018731	1.11406	-1.9999	0	3.9960	2.30047	0.149156	5.441353
	$\beta=1$			$\alpha=1.5$			C=0.9		
	Bias	Variance	MSE	Bias	Var.	MSE	Bias	Var.	MSE
20	1.279	0.360584	1.99845	-1.999	0	3.9960	2.6694	3.71722	10.84287
30	1.161	0.211514	1.56024	-1.9999	0	3.9603	2.1697	1.086474	5.794367
50	1.199	0.202216	1.64052	-1.9999	0	3.9960	2.3040	1.318587	6.627044
80	1.139	0.170998	1.47002	-1.999	0	3.9960	2.0790	1.098943	5.421277
100	1.014	0.157921	1.18775	-1.999	0	3.9960	1.6903	0.799668	3.656934
200	0.976	0.094923	1.04934	-1.9999	0	3.9960	1.6897	0.579610	3.434744
300	0.918	0.025004	0.8694	-1.9999	0	3.9960	1.5269	0.128867	2.460371
	β= 1.5			α= 0.5			C=0.8		
	Bias	Variance	MSE	Bias	Var.	MSE	Bias	Variance	MSE
20	1.631	0.73142	3.3928	-0.499	0	0.2490	2.1509	2.752424	7.379041
30	1.413	0.61655	2.6138	-0.499	0	0.2490	1.7945	1.282839	4.503185
50	1.173	0.19589	1.5725	-0.499	0	0.2490	1.5009	0.681053	2.933925
75	1.516	0.26232	2.5634	-0.499	0	0.2490	1.9730	0.482302	4.375146
100	1.354	0.14775	1.9834	-0.499	0	0.2490	1.5451	0.351610	2.738881
200	1.210	0.05982	1.5258	-0.499	0	0.2490	1.4653	0.141682	2.288914
300	1.177	0.03884	1.4216	-0.499	0	0.2490	1.4216	0.101695	2.122741

9. Application

In this section we consider survival period data of 45 patients treated with chemotherapy only were made by Bekker et al.4 and Fulment et al.10. The data set are:

0.047, 0.115, 0.121, 0.132, 0.164, 0.197, 0.203, 0.260, 0.282, 0.296, 0.334, 0.395, 0.458, 0.466, 0.501, 0.507, 0.529, 0.534, 0.540, 0.641, 0.644, 0.696, 0.841, 0.863, 1.099, 1.219, 1.271, 1.326, 1.447, 1.485, 1.553, 1.581, 1.589, 2.178, 2.343, 2.416, 2.444, 2.825, 2.830, 3.578, 3.658, 3.743, 3.978, 4.003, 4.033.
The second data set represents the failure time of 50 items [13]
0.12, 0.43, 0.92, 1.14, 1.24, 1.61, 1.93, 2.38, 4.51, 5.09, 6.79, 7.64, 8.45, 11.9, 11.94, 13.01, 13.25, 14.32, 17.47, 18.1, 18.66, 19.23, 24.39, 25.01, 26.41, 26.8, 27.75, 29.69, 29.84, 31.65, 32.64, 35, 40.7, 42.34, 43.05, 43.4, 44.36, 45.4, 48.14, 49.1, 49.44, 51.17, 58.62, 60.29, 72.13, 72.22, 72.25, 72.29, 85.2, 89.52.

In order to compare the weighted Sabur distribution with the Erlang Truncated Exponential distribution, Exponential distribution, Power Lindley distribution, we consider the criteria like Bayesian information criterion (BIC), Akaike Information criterion (AIC), Akaike Information Criterion Corrected (AICC) and -2logL. The distribution having lower values of BIC, AIC, AICC and -2log L can be consider better. Along with this we calculate goodness of (GoF) metrics statistic Schwarz Information (SIC), Hannan-Quinn Information (HQIC) criteria, we also assess the Anderson–Darling (A*), Cramer-Von Mises (W*), Kolmogorov–Smirnov (K–S) statistic and associated P-value (PV). Table 2 and Table 3 represents parameter estimation of data set 1 and set 2 with GoF metrics. Figure 5 and Figure 6 shows the diagrammatic representation of density curve of data set 1 and data set 2. Figure 7 and Figure 8 shows the QQ plot of weighted Sabur distribution of data set 1 and data set 2.

$$AIC = 2k - 2logL, \qquad BIC = klogn - 2logL, AICC = AIC + \frac{2k(k+1)}{(n-k-1)}$$

Table 2 : *Parameter estimation and goodness of fit test statistics for survival data set 1*

Distributions and estimations	Weighted Sabur distribution	Erlang Truncated Exp distribution	Power Lindley Distribution	Exponential Distribution
MLE	$\hat{\alpha}$=0.00100 $\hat{\beta}$=2.09088 C=0.66942	$\hat{\beta}$= 1.140789 $\hat{\theta}$= 1.059767	$\hat{\beta}$= 0.9465414 $\hat{\theta}$=1.135077	$\hat{\theta}$= 0.7454655
SE	$\hat{\alpha}$=NaN $\hat{\beta}$=0.278 C=0.276	$\hat{\beta}$=70.36128 $\hat{\theta}$= 116.3132	$\hat{\beta}$= 0.107625 $\hat{\theta}$=0.146510	$\hat{\theta}$=0.111127
-2log L	96.2613	116.437	116.805	116.437
AIC	107.681	120.4372	120.8056	118.4372
BIC	107.681	124.0506	124.4189	120.2439
AICC	108.26666	120.72291	118.72291	120.89862
SIC	107.681	124.7229	124.4189	124.0506
HQIC	104.282	121.7842	122.1526	121.7842

A*	2.507	0.44535	0.56555	0.44535
W*	0.255	0.05897	0.08453	0.05897
K-S	0.255	0.09083 (0.819)	0.11044 (0.603)	0.16968 (0.1332)

Table 3: *Parameter estimation and goodness of fit test statistics for failure data set 2*

Distributions and estimations	Weighted Sabur distribution	Erlang Truncated Exp distribution	Power Lindley Distribution	Exponential Distribution
MLE	$\hat{\alpha}$=6.997702e+04 $\hat{\beta}$=.00332 C=.001	$\hat{\beta}$= 0.1437169 $\hat{\theta}$= 0.2621020	$\hat{\beta}$= 0.72807201 $\hat{\theta}$=0.07252873	$\hat{\theta}$= 0.033142
SE	$\hat{\alpha}$=1.186329e+04 $\hat{\beta}$=.0074 C=.0176	$\hat{\beta}$= 1.9452922 $\hat{\theta}$= 4.0561881	$\hat{\theta}$=0.16807702 $\hat{\theta}$=0.04478477	$\hat{\theta}$=0.004683
-2log L	440.7604	440.7133	442.1001	440.7133
AIC	446.7604	444.7133	446.1001	442.7133
BIC	452.4965	448.5374	449.9242	444.6253
AICC	447.28214	444.96861	446.35542	446.35541
SIC	452.4965	448.5374	449.9242	448.5374
HQIC	448.9448	446.1695	447.5563	446.1695
A*	0.8861	0.8852	0.9367	169
W*	0.11426	0.13171	0.13959	8.571
K-S	0.11426 (0.495)	0.11434 (0.4947)	0.13959 (0.4631)	0.6958 (8.882e-16)

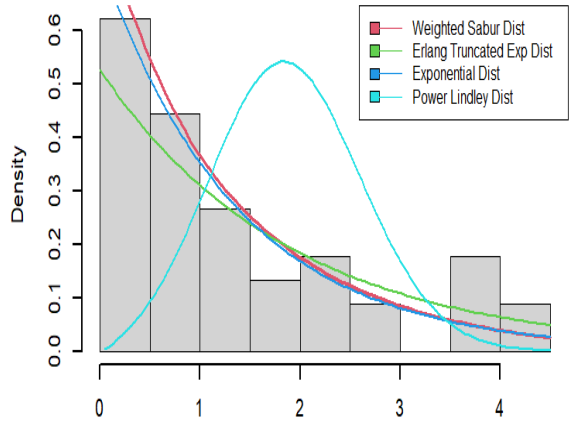


Fig. 5: Fitting density curves of data set 1

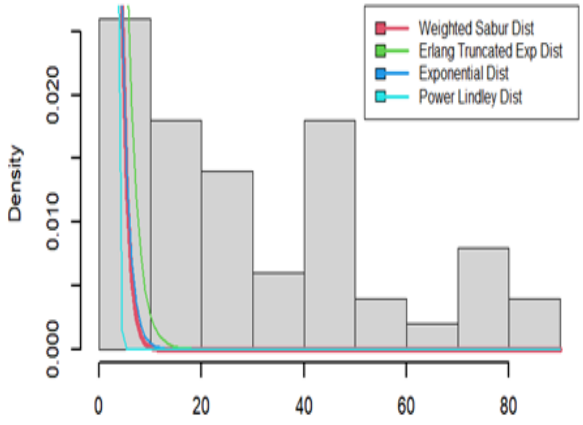


Fig. 6: Fitting density curves of data set 2

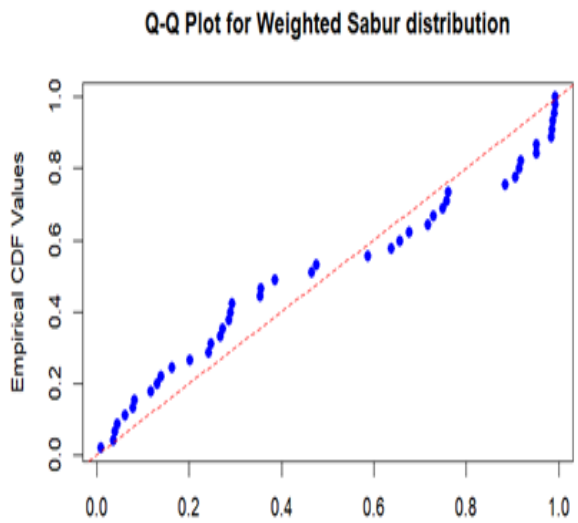


Fig. 7: Custom cdf values for data set 1

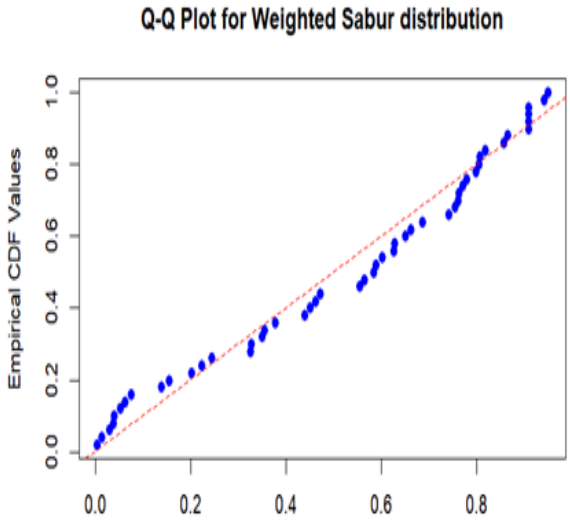


Fig.8 : Custom cdf values for data set 2

10. Conclusion

This paper introduces the weighted Sabur distribution with three parameters, a novel extension of the Sabur distribution, and explores its comprehensive statistical properties. The model parameters are estimated using maximum likelihood estimation, incorporating a weighted approach to enhance precision. The analysis encompasses various mathematical aspects and reliability measures, including the hazard rate function, to evaluate the distribution's performance as a lifetime model. Additionally, we benchmark the weighted Sabur distribution against other established distributions such as the exponential, power Lindley, and Erlang truncated exponential, using two sets of real-world data for validation. This comparative analysis confirms the potential of the weighted Sabur distribution as a robust and versatile model for lifetime data analysis.

References

[1] Ahmad Aijaz, Rather A. A., Gemeay A. M., Nagy M., Sapkota L. P., Mansi A. H. (2024), Novel Sin-G Class of Distributions with an Illustration of Lomax Distribution: Properties and Data Analysis, *AIP Advances*, Vol. 14, No. 035132, pp 1-17.

[2] Aijaz, A., Ahmad, A, Tripathi, R. (2020), Transmuted inverse Lindley distribution: Statistical properties Applications, *Science, technology and development,* 9(7), 1-10.

[3] Aijaz, A., Ahmad, Afaq Ahmad, Aafaq A. Rather, The Sabur Distribution: Properties and application related to engineering data, *RT&A*, No 4 (76), pp 882-892.

[4] Aijaz. A, Jallal. M, Qurat Ul Ain.S, Tripathi . R (2020). The Hamza distribution with Statistical properties and Applications. *Asian Journal of Probability and Statistics*, 8(1):28-42.

[5] Bekker, A., Roux, J. J. J. & Mostert, P. J. A generalization of the compound Rayleigh distribution: Using a Bayesian method on cancer survival times. *Commun. Stat. Theory Methods* 29(7), 1419–1433 (2000).

[6] Bhaumik, D. K., Kapur, K. and Gibbons, R. D. (2009). Testing parameters of a gamma distribution for small samples, *Technimetrics*, 51(3), 326-334.

[7] Cox D. R. (1969). Some sampling problems in technology, In New Development in Survey Sampling, Johnson, N. L. and Smith, H., Jr .(eds.) New York, *Wiley- Inte sciences*, 506-527

[8] Fisher, R.A. (1934). The effects of methods of ascertainment upon the estimation of frequencies, *Annals of Eugenics*, 6, 13- 25.

[9] Flaih A., Elsalloukh, H., Mendi, E., Milanova, M., 2012. The exponentiated inverted Weibull distribution. *Appl. Math. Inf. Sci* 6 (2), 167–171.

[10] Fuller E.J., Friema S., Quinn, J., Quinn, G., and Carter, W. (1994): Fracture mechanics approach to the design of glass aircraft windows: A case study, *SPIE* Proc 2286, 419-430.

[11] Fulment A. K., Gadde, S. R. & Peter, J. K. Te odd log-logistic generalized exponential distribution: Application on survival times of chemotherapy patients' data. F1000research 11, 1444 (2023). 30. Aarset, M. V. How to identify a bathtub hazard rate. *IEEE* Trans. R

[12] Ghitany, M.E., Atieh B., Nadarajah S., (2008). Lindley distribution and its applications, Mathematics Computing and Simulation, 78, 493-506

[13] Gross, A. J., Clark, V. A., (1975). Survival Distributions: Reliability Applications in the Biometrical Sciences, *John Wiley*, New York

[14] Lai, C.D.; Murthy, D.; Xie, M. Weibull Distributions and Their Applications. In Springer Handbook of Engineering Statistics; Pham, H., Ed.; *Springe*r: London, UK, 2006.

[15] Lindley, D, V. (1958). Fiducial distributions and Bayes' theorem, *Journal of Royal Statistical Society*, Series B, 20, 102-107.

[16] Merovci, F., (2013) Transmuted Lindley distribution, *International Journal of Open problems in Computer Science and Mathematics*, 6, 63-72.

[17] Nadarajah, S., BAKOUCH, H. S., TAHMASBI, R., (2011). A Generalized Lindley distribution, *Sankhya Series* B, 73, 331-359.

[18] Rao, C. R. (1965), ON Discrete Distributions Arising out of Methods of Ascertainment. *Sankhya*; The Ind. J. Statist, Series A, 27, 311-324.

[19] Rather, A. A. & Ozel G. (2021). A new length-biased power Lindley distribution with properties and its applications, Journal of Statistics and Management Systems, DOI: 10.1080/09720510.2021.1920665

[20] Rather, A. A. and Subramanian, C. (2020), A New Exponentiated Distribution with *Engineering Science Applications, J. Stat. Appl*. Pro. 9, No. 1, pp 127-137.

[21] Zelen, M. (1974). Problems in cell kinetic and the early detection of disease, in Reliability and Biometry, F. Proschan & R. J. Sering, eds, *SIAM*, Philadelphia, 701-706

STATISTICAL AND DEEP-LEARNING BASED DISASTER IDENTIFICATION MODELLING USING UNMANNED AERIAL VEHICLE SYSTEMS FOR EMERGENCY RESPONSE

Mustafa Kamal[1]*, Mohammad Faisal Khan[2], Shahnawaz Khan[3]

•

[1]* Department of Basic Sciences, College of Science and Theoretical Studies, Saudi Electronic University, Dammam Branch 32256, Saudi Arabia
Emails: m.kamal@seu.edu.sa

[2]Department of Basic Sciences, College of science and theoretical studies, Saudi Electronic University, Riyadh, 11673, Saudi Arabia
Email: f.khan@seu.edu.sa

[3]Faculty of Engineering, Design and Information and Communications Technology, Bahrain Polytechnic, Isa Town, Bahrain
Email: shahnawaz.khan@polytechnic.bh

Abstract

Unmanned aerial vehicle systems offer a significant impact for the prediction of disaster identification and management by integrating both statistical and neural network techniques. Existing disaster response systems primarily rely on manual reporting or satellite imagery which are prone to delays and inefficiencies. The present study presents a statistical modelling using structural equation model integrated with deep learning-based model to enhance prediction accuracy. The model takes input variables such as unmanned aerial vehicle altitude, speed, area coverage, temperature, and population density to predict a disaster index. The structural equation model analysis revealed that all the input variables unmanned aerial vehicle altitude, speed, area coverage, temperature, and population density have a significant impact on disaster index. The proposed multi-layer perceptron model achieves an overall r2 score of 0.86, demonstrating its effectiveness in differentiating disaster severity. The study concludes that integrating unmanned aerial vehicle systems with statistical and deep learning techniques for disaster index is a feasible and impactful solution to mitigate human and economic losses during extreme events.

Keywords: Unmanned aerial vehicle data, Disaster Index, Multi-class regression approach, Metrological parameters, Multi-Layer Perceptron model

I. Introduction

Natural and man-made disasters are persistent global challenges, threatening lives, infrastructure, and economies. Floods, fires, and traffic accidents are among the most frequent and

devastating disasters, causing immense destruction each year [1-3]. The unpredictability and intensity of such events have underscored the need for faster, more efficient disaster identification and response systems. Traditionally, manual reporting and satellite imagery have been used to assess disaster damage and manage emergency responses [4,5]. However, these methods often suffer from delays, inaccuracies, and inefficiencies due to their dependence on human intervention or the long revisit times associated with satellite imagery. This delay in identifying and assessing the severity of disasters can lead to a slow response, resulting in more casualties and greater economic loss [6]. Therefore, there is an urgent need to develop advanced systems that can quickly and accurately identify and assess the severity of disasters in real time.

In recent years, Unmanned Aerial Vehicle (UAV) systems have emerged as a revolutionary tool for disaster management [7]. UAVs offer several advantages, including rapid deployment, high maneuverability, and the ability to capture detailed images of disaster-stricken areas from various angles. UAVs can operate in challenging environments where traditional systems struggle, such as during adverse weather conditions or in remote areas [8]. These aerial systems provide real-time data that can be analyzed to determine the extent of damage, allowing emergency services to respond more effectively [9]. Ample research focused on analyzing the factor affecting severity of disaster using statistical models [10-11]. But these statistical models came with certain limitations such as low accuracy and robustness [12]. Nowadays with the advancement of AI techniques, the deep learning techniques have got special attention for prediction problem. Existing research has used deep learning techniques are widely used for other domain applications such as Environment, Banking, and Tourism [13-15]. While UAV technology is advancing, there is still a significant research gap in integrating UAV data with advanced deep learning models to improve disaster identification, particularly in assessing the severity of disasters.

The deep learning techniques contain several techniques such as Multi-Layer Perceptron (MLP), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Long-Short Term Memory (LSTM). One of the main challenges in disaster identification is the ability to classify the severity of a disaster. Most of the disaster severity is classified into low, moderate, and extreme based on the available data. The present study is a regression problem statement. Most of the current systems are limited to binary classifications, such as flooded versus non-flooded areas, without further granularity which are mostly analyzed by the help of CNN technique [16, 17]. While these approaches can be useful for simple disaster identification, they lack the depth needed for effective resource allocation and emergency response. A more nuanced understanding of the disaster's severity would enable emergency services to prioritize high-risk areas and allocate resources more efficiently, ultimately saving more lives and reducing economic losses.

Therefore, the present study uses MLP model for the prediction of disaster severity as the problem is a regression problem. The proposed model integrates both statistical methods i.e. structural equation model and deep learning techniques, specifically MLP, to analyze disaster severity. By incorporating variables such as UAV altitude, speed, area coverage, temperature, and population density, the model aims to predict a Disaster Index, which classifies the severity of the disaster into three categories: low, moderate, and extreme. The ability to accurately classify disaster severity will greatly enhance decision-making processes, enabling authorities to prioritize emergency responses based on real-time data. This makes UAV-based deep learning systems highly efficient for disaster identification, especially in regions prone to frequent disasters such as floods, fires, and traffic accidents. Table 1 shows some more recent literature review of various research on the impact of various UAVs parameters for the prediction of risk assessment using statistical modeling and artificial intelligence techniques.

The novelty of this research lies in its multi-class regression approach for disaster severity, in contrast to the binary classification models used in most previous studies [32, 33]. By focusing on the severity levels (low, moderate, and extreme), the model offers a more comprehensive analysis,

which is crucial for emergency response and resource allocation. Additionally, the use of both statistical and deep learning methods enhances the model's robustness and generalizability across different disaster scenarios. The findings of this study will have significant implications for regions prone to frequent natural and man-made disasters, particularly in the context of improving emergency response times and resource allocation.

Table 1: *Literature review of the various studies conducted for the assessment of risk.*

Author(s)	Country	Parameters	Conclusion
[18]	Saudi Arabia	UAV altitude, area coverage	Demonstrated the effectiveness of UAV in flood risk assessment in urban areas.
[19]	Kuwait	UAV data, field survey, and satellite image	Proposed a model strategic plan to diminish flood vulnerability.
[20]	China	Multiple source satellite datasets	Found that UAVs provide significant advantages in flood monitoring by noisy learning method.
[21]	India	Topography, forest, soil, and geologic factor	Developed a deep learning model yielding a high accuracy for landslide identification.
[22]	Italy	Technical features of UAV	Established a correlation between UAV flight parameters and disaster mapping accuracy.
[23-26]	Multiple sampling locations	Environmental factors, sun glint, vegetation	Highlighted the impact of environmental parameters on UAV image quality.
[27]	Dubai	Image data from UAVs	Achieved high accuracy (85.4%) in vegetation cover accuracy using UAV imagery with DL techniques.
[28]	Indus River, Pakistan	UAV-based aerial imagery	CNN model achieved an accuracy of 91% for flood detection.
[29]	China–Russia Border	Humidity, temperature, precipitation, and wind speed	Heilongjiang province is best suited place to travel during summer.
[30]	USA	Population density, tropical cyclones, annual variation of mortality, and topography	Topography and population density has a direct correlation for Flood-induced mortality.
[31]	Not mentioned	Coverage area, height, velocity	Particle swarm optimization (PSO) was used to obtain the severity of the disaster.

II. Methods and Material

I. Study Area

Jeddah city, situated within three primary sub-basins (northern, middle, and southern), serves as a focal point for this case study due to its susceptibility to flash floods. The northern sub-basin comprises several wadis, including Wadi Daghbaj, Wadi Brayman, Wadi Muraygh, Wadi Quraa, Wadi Ghaia, and Wadi Um Hablain. The middle sub-basin encompasses Wadi Mraikh and Wadi Bani Malik, while the southern sub-basin includes Wadi Qaws, Wadi Methweb, Asheer, Wadi Al

Khomra, and Wadi Ghulail. As of 2022, Jeddah city had an estimated population of 4.78 million residents. Bordered to the west by the Red Sea and to the east by mountain ranges with a maximum elevation of 675 meters, the drainage area delineated by a 30-meter Digital Elevation Model (DEM) covers approximately 1,821 km² [34]. The city's residential zones, located on the coastal plain, are vulnerable to the impacts of flash floods originating from the adjacent mountains. The topography of the Jeddah watershed reveals two distinct geomorphological units: the coastal plain and the mountainous regions that surround the city. Despite Jeddah's arid climate, it is recurrently affected by flash floods, with significant events recorded multiple times. Notably, on November 25, 2009, flash floods severely impacted urban areas, leading to extensive damage to infrastructure, buildings, vehicles, and roads, resulting in approximately 113 fatalities. Another destructive event occurred in 2011, further highlighting the flood risk in the region. The watershed has various drainage channels that traverse neighborhoods such as Al-Harazat, King Abdul Aziz University, Al-Haramin Highway, Al-Mesaid, Queza, and Al-Sawaid, all of which experienced substantial effects from the 2009 flash flood incident. Figure 1 shows the conceptual framework of the present study.

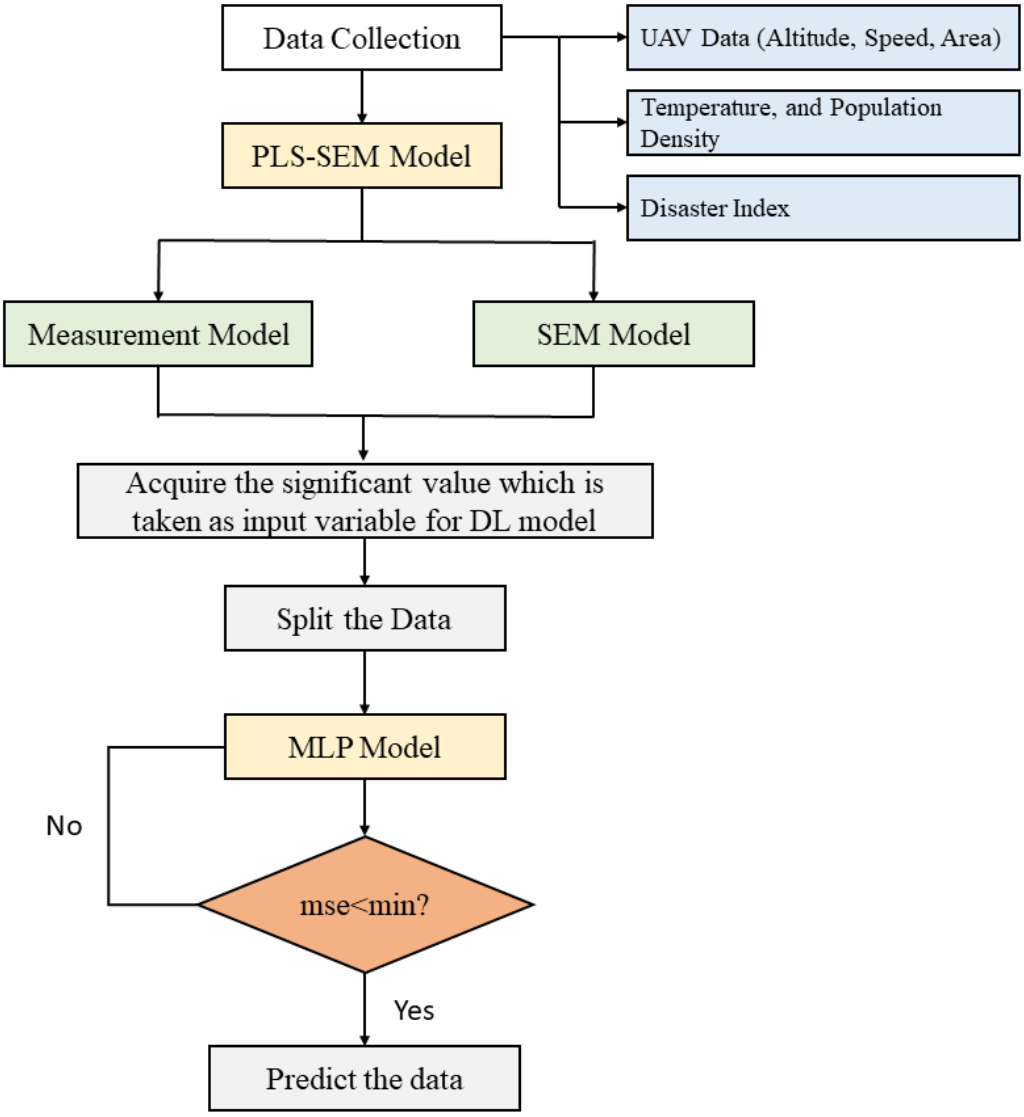


Figure 1: *Conceptual framework of present study.*

II. Data Collection

The UAV data, including altitude, speed, temperature, humidity, area coverage, and population data, were taken into account from the different sources such as: The National Centre for Meteorology (NCM), Saudi Arabia and Shuttle Radar Topography Mission (SRTM) [35-37]. The only feature that has been taken based on the severity is disaster index. This feature has been divided into three subcategories such as low, moderate, and high. Table 2 presents the standard deviation and mean of the data used for modelling and prediction. UAVs equipped with sensors gather altitude, speed, and area coverage data, providing quantitative information on UAV flight dynamics and coverage area.

Table 2: *Statistical parameters of the data used in the study.*

Variable	Mean	Standard Deviation
UAV Altitude (m)	130	25
UAV Speed (m/s)	10.5	2.5
Area Coverage (km²)	1.75	0.4
Temperature (°C)	28	3
Humidity (%)	45	8
Population Density (people/km²)	2600	800
Disaster Index (0-1)	0.68	0.1

III. SEM Modelling

SEM is a statistical technique that enables the analysis of complex relationships between the constructs [38]. In the present study, SEM is used to model the relationships between multiple factors influencing variables, such as UAV altitude, speed, area coverage, temperature, and population density on disaster index. SEM allows for the simultaneous examination of direct and indirect effects among these variables, providing deeper insights into their collective impact on the disaster index. In SEM analysis, two models are used to predict disaster index i.e. measurement model and structural equation model. Measurement model also known as inner model is used to check the reliability and validity of the constructs whereas the structural equation model which is known as outer model is used to analyze the significant impact of independent constructs on disaster index.

IV. Hypothesis Testing

The study utilizes five set of hypotheses for the prediction of disaster index as shown in Figure 2. The hypothesis uses UAV altitude, speed, coverage area, temperature, and population density for the prediction of disaster index.

- **H1:** UAV altitude (m) has a significant effect on Disaster Index.
- **H2:** UAV speed (m/s) has a significant effect on Disaster Index.
- **H3:** Area coverage (km²) has a significant effect on Disaster Index.
- **H4:** Temperature (^{0}C) has a significant effect on Disaster Index.
- **H5:** Population density (people/km²) has a significant effect on disaster index.

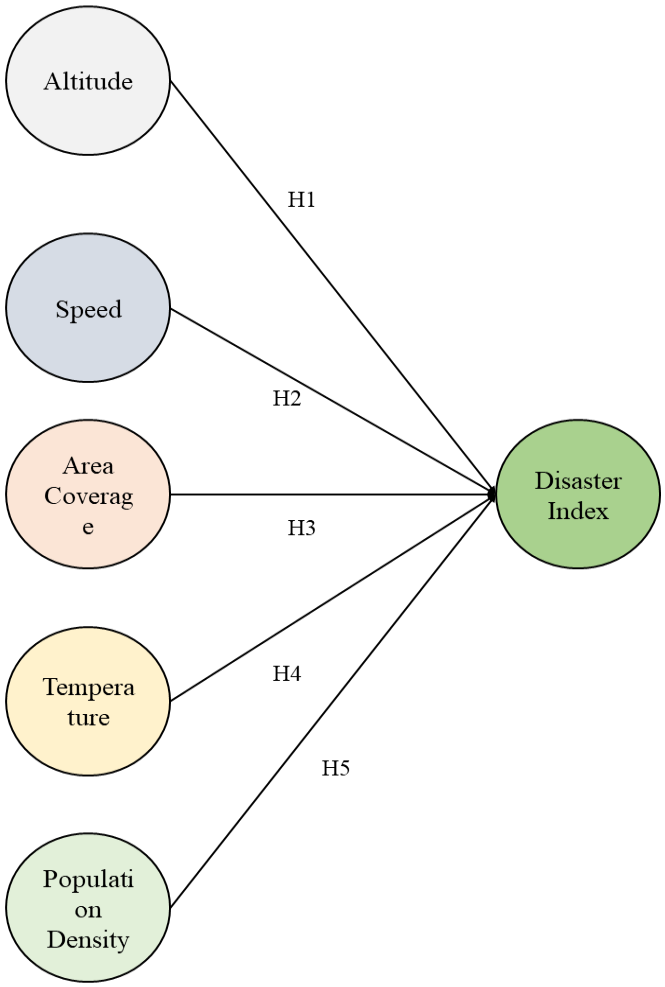


Figure 2: *Various hypothesis for the present study.*

V. Data Pre-processing

Data pre-processing is a crucial step to prepare the UAV images and auxiliary datasets for input into the deep learning model. The regression data was splitted in the ratio of 75:25 (training: testing). The pre-processing involves resizing images to ensure uniform dimensions, typically $H \times W$ (height and width), and applying geometric transformations to correct for distortions [3]. Normalization is performed to scale pixel values between 0 and 1, ensuring that the model training is not biased by large numerical values. For input variables like temperature, humidity, and population density, standardization is applied:

$$X' = (X - \mu)/\sigma \tag{1}$$

where X is the original value, μ is the mean, and σ is the standard deviation, ensuring all features are on a comparable scale.

VI. Multi-Layer Perceptron (MLP)

The MLP model is a type of feed-forward neural network that maps input variables to output

variables by learning complex non-linear relationships [39]. An MLP consists of an input layer, one or more hidden layers, and an output layer, where each layer contains multiple neurons. Each neuron in a layer performs a weighted sum of its inputs, applies an activation function, and passes the result to the next layer. Figure 3 shows the architecture of MLP model. For each neuron j in a hidden or output layer, the neuron computes:

$$z_j = \sum_{i=1}^{n} w_{ij} x_i + b_j \quad (2)$$

where x_i is the input from the previous layer, w_{ij} represents the weight connecting input i to neuron j, and b_j is the bias term. The result z_j is then passed through an activation function f to introduce non-linearity $a_j = f(z_j)$. Common activation functions include the **ReLU** function $f(z) = max\ (0, z)$ for hidden layers and a linear function for regression tasks in the output layer.

The goal of training an MLP is to minimize the difference between the predicted outputs and the actual values (e.g., Disaster Index). This is achieved by optimizing the loss function, often Mean Squared Error (MSE) in regression:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (3)$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the number of samples.

The training process involves backpropagation and gradient descent to adjust the weights and biases to minimize the loss function. By iterating through multiple epochs of training, the MLP model learns to approximate the mapping from inputs to outputs, enabling it to make accurate predictions on new data.

The algorithm used in the study for the prediction of disaster index is illustrated below:

1. Input: X= {x_i |i= 1, 2 ,…, N} ⊳ Input samples including UAV altitude, speed, area coverage, temperature, humidity, and population density
2. Ci ← Classifier input features
3. Y ← y_{train}, y_{test} ⊳ True labels for Disaster Index for training and test sets
4. X ← x_{train}, x_{test} ⊳ Split data into training and testing
5. S ← No. of samples ⊳ Total number of samples
6. C ← Model complexity (hidden layers, neurons) ⊳ Define MLP architecture
7. Reg ← MLP regressor ⊳ MLP model for regression
8. L ← Loss function (Mean Squared Error)
9. Train ← Train model with x_{train} ⊳ Fit MLP model to training data
10. Evaluate ← Evaluate model with x_{test}, y_{test}
11. Prediction ← MLP(x_{test}) ⊳ Predict Disaster Index for test samples
12. Calculate Metrics ← MSE, MAE, R^2
13. Plot ← Observed vs. Predicted plot
14. While L(x) is not minimized:
15. If Loss > threshold:

Adjust parameters ← Fine-tune MLP architecture or learning rate
Else
Stop Training
Return Final Model

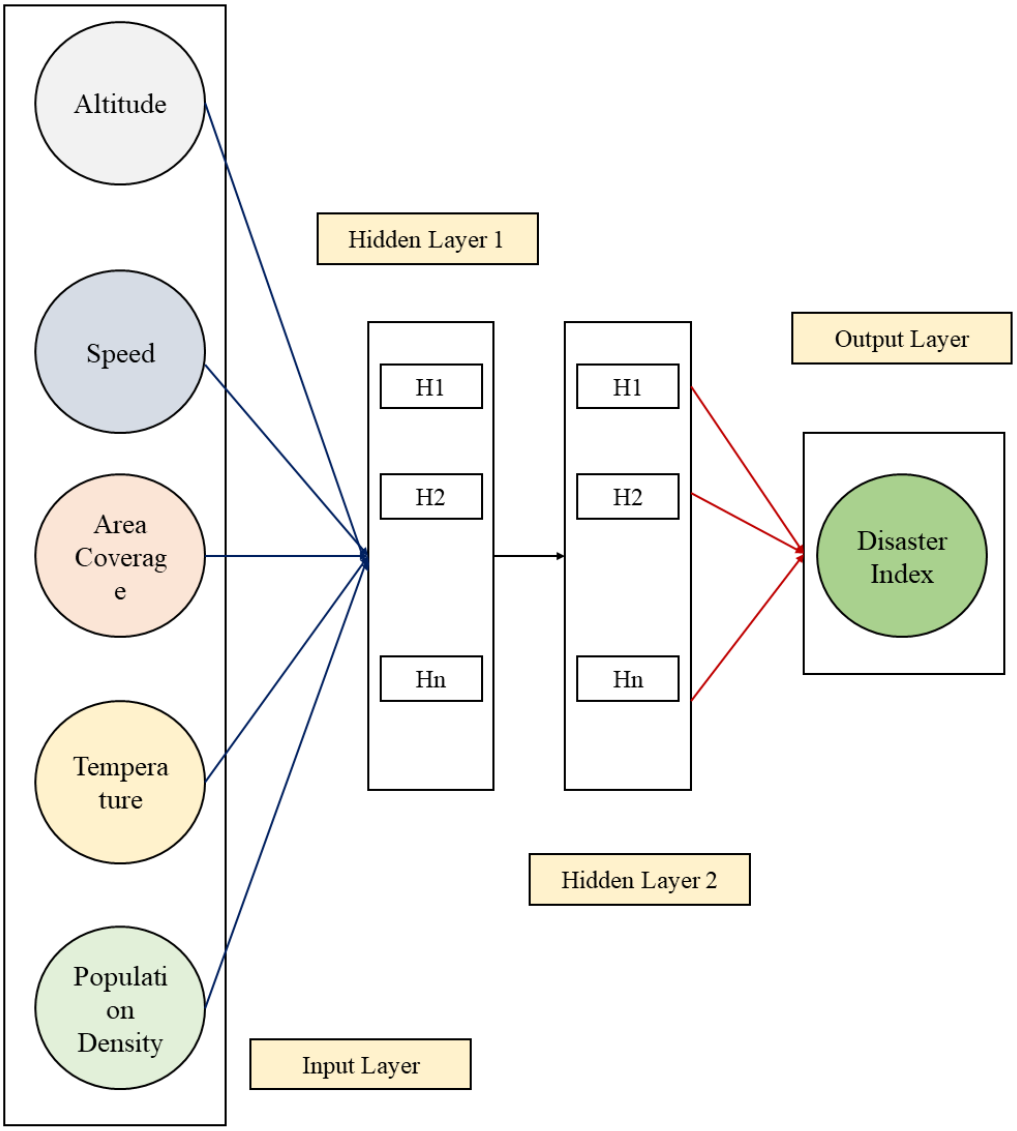


Figure 3: *Architecture of MLP model.*

III. Results and Discussion

The present research organized all indicators into six constructs: UAV altitude, UAV speed, area coverage, temperature, population density, and Disaster Index. To assess the reliability and validity of these variables, a measurement model was employed.

Table 3: *Reliability and Validity test of the hypothesis.*

Latent Variable	Cronbach's Alpha	Composite Reliability (CR)	Average Variance Extracted (AVE)
UAV Altitude	1	1	1
UAV Speed	1	1	1
Area Coverage	1	1	1
Temperature	1	1	1
Population Density	1	1	1

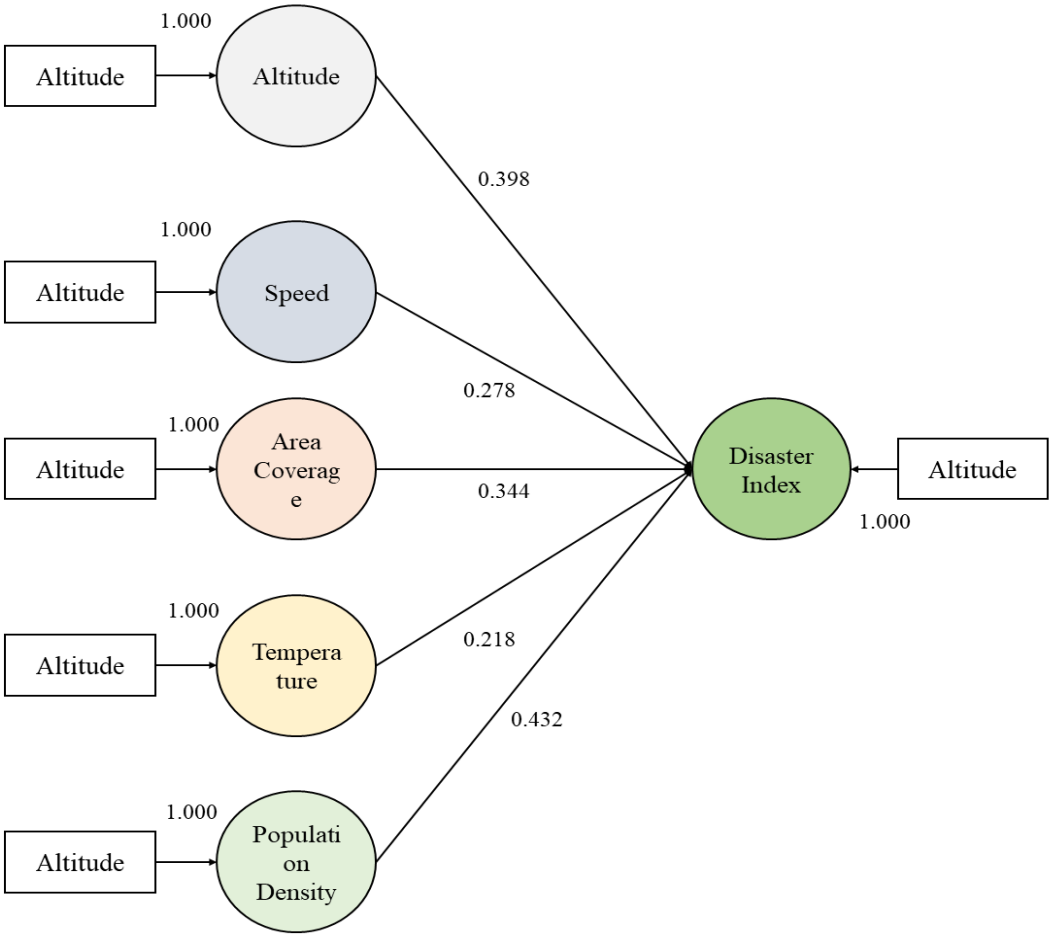


Figure 4: *The SEM model.*

Confirmatory factor analysis was executed using SmartPLS 4.0 software to evaluate the measurement properties of the items concerning their designated factors. Indicators with a standard loading below 0.05 were excluded from the measurement model. The results derived from the measurement model are presented in Table 3. It is evident that all indicators achieved a significance level of 0.001 and a standard loading exceeding 0.5. The reliability of the latent variables was confirmed as the Cronbach's alpha, composite reliability (CR), and average variance extracted (AVE) values all exceeded the thresholds of 0.7, 0.6, and 0.5, respectively. The values of all the constructs were 1 only because there was only one indicator to represent the construct. As demonstrated in Table 3, all six constructs (UAV altitude, UAV speed, area coverage, temperature, population density, and Disaster Index) met these criteria, indicating conformity with the requirements for convergent validity.

The SmartPLS 4.0 bootstrapping method was employed to evaluate and analyze the interrelationships among the constructs. The SEM model was constructed using a set of six constructs and six indicators. Additionally, the structural model was utilized to examine the path relationships among these variables. The findings from the structural path model are detailed in Table 4. The path coefficient serves as a statistical indicator that estimates both the strength and direction of the relationship between two latent variables. For a significant relationship to be established between the latent variables, the path coefficient must exceed 0.2 at a 95% confidence interval. Figure 4 illustrates that all path coefficients achieved a significance level of 0.001. Mean, standard deviation, t-value, and p-values were assessed to determine the significant relationships among the variables. It was found that Population Density (t: 6.1), Altitude (t: 5.24) significantly

influenced Disaster Index. Furthermore, Speed (t: 4.73), Area Coverage (t: 4.91), and Temperature (t: 3.45) exhibited significant effects on Disaster Index.

Table 4: *SEM model statistics for the prediction of disaster index.*

Hypothesis	Mean	Standard Deviation	t-value	p-value	Significant Effect?
H1: Altitude → Disaster Index	0.65	0.12	5.24	< 0.001	Yes
H2: Speed → Disaster Index	0.58	0.15	4.73	< 0.001	Yes
H3: Area Coverage → Disaster Index	0.63	0.13	4.91	< 0.001	Yes
H4: Temperature → Disaster Index	0.6	0.14	3.45	0.002	Yes
H5: Population Density → Disaster Index	0.67	0.11	6.1	< 0.001	Yes

Table 5: *Performance matrix of MLP model.*

Metric	Training Set	Testing Set
Mean Squared Error (MSE)	0.0052	0.0068
Mean Absolute Error (MAE)	0.058	0.062
R2 Score	0.89	0.86

Table 5 shows the performance metrics for the MLP model on both training and testing demonstrate its predictive accuracy in assessing disaster index. The Mean Squared Error (MSE) values of 0.0052 for the training set and 0.0068 for the testing set indicate that the model achieves low average squared errors, showing it effectively minimizes large prediction deviations. The Mean Absolute Error (MAE) values of 0.058 for training and 0.062 for testing further confirm the model's accuracy by providing insight into the average absolute difference between predicted and actual values. Additionally, the R^2 Score values of 0.89 for training and 0.86 for testing imply that the model explains a high proportion of the variance in disaster severity, with only minor overfitting or under-fitting present. Together, these metrics suggest that the MLP model generalizes well to unseen data and is reliable for predicting disaster severity based on UAV parameters and environmental conditions.

Figure 5 illustrates the relationships between various parameters related to UAV-based disaster identification, including altitude, UAV speed, area coverage, temperature, humidity, population density, and the disaster index. Each cell shows the correlation coefficient between two variables, with values closer to 1 indicating a strong positive correlation and values closer to -1 indicating a strong negative correlation. For example, the disaster index has a strong positive correlation with altitude (0.97), temperature (0.95), and population density (0.96). This suggests that as altitude, temperature, or population density increase, the disaster index also tends to increase. Additionally, UAV speed and area coverage have a strong correlation (0.94), indicating that faster UAV speeds tend to be associated with larger areas covered. These insights can help in fine-tuning UAV parameters for better disaster identification and emergency response effectiveness.

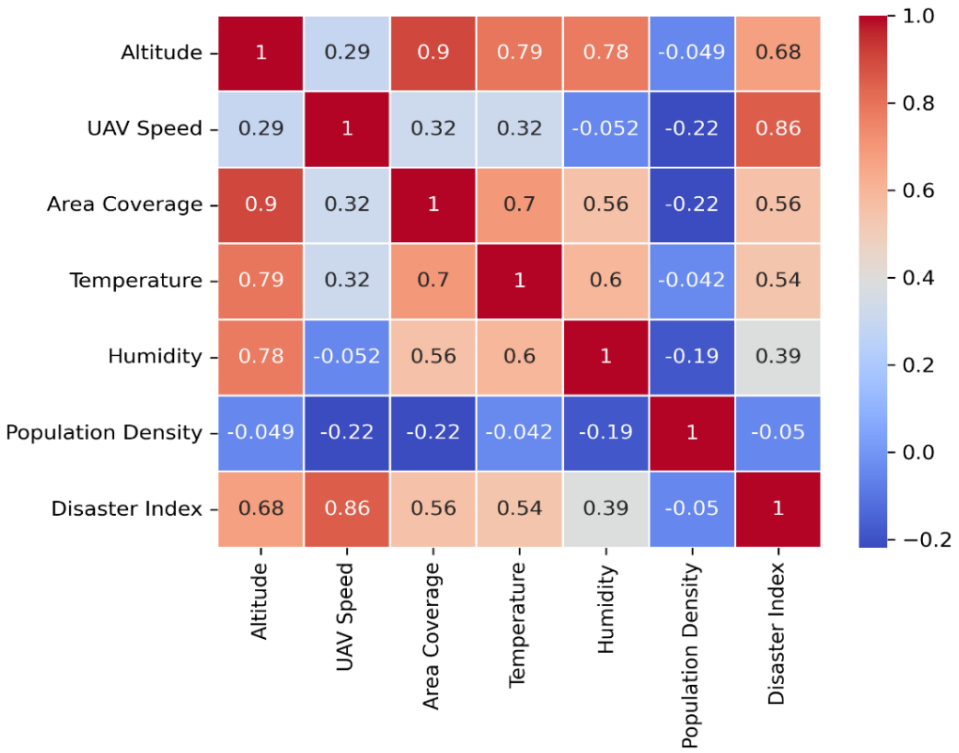


Figure 5: *Correlation matrix between the variables*

Table 6 presents a sensitivity analysis of various variables contributing to the disaster index specific to flood severity in Jeddah City, Saudi Arabia. This disaster index measures the potential impact of flooding through multiple influencing factors. Among the variables listed, altitude exhibits the highest sensitivity at a value of 23.5, indicating that changes in altitude significantly affect flood severity; areas at lower elevations are more susceptible to flooding risks due to greater water accumulation. The speed of water flow during floods shows even greater sensitivity, with a value of 32.8, suggesting that faster water flow can lead to more severe flooding and greater damage to infrastructure. Area coverage, with a sensitivity value of 12.3, indicates a moderate impact on flood severity, where larger flooded areas could imply extensive effects on communities and ecosystems; however, its lower sensitivity compared to altitude and speed suggests that changes in area coverage have a less pronounced effect. Temperature and humidity follow, with sensitivity values of 10.5 and 4.9, respectively. While temperature influences rainfall intensity and snowmelt, making it significant in the flood context, humidity seems to have a minor effect on the disaster index. Lastly, population density, with a sensitivity of 16.0, reflects the impact of how population distribution can modify disaster consequences—higher population density typically results in greater impacts, as more individuals are exposed to potential flood hazards.

Table 6: *Sensitivity analysis of the disaster index.*

Variables	Sensitivity
Altitude	23.5
Speed	32.8
Area Coverage	12.3
Temperature	10.5
Humidity	4.9
Population Density	16.0

IV. Limitations and Scope of Future Work

The limitations and scope of future work are as follows:

- One limitation of this study is the reliance on historical data for model training, which may introduce biases based on past disaster events. Such biases can limit the model's ability to generalize and accurately predict disaster scenarios that differ significantly from those previously encountered. Additionally, the statistical assumptions inherent in SEM may not always hold true in the complex and dynamic environment of disaster management.
- Another limitation is the potential for sensor inaccuracies in the UAV systems utilized in data collection. Factors such as environmental conditions and technical malfunctions can lead to variability in the input variables, thereby affecting the reliability and precision of the predictions made by the models. This variability necessitates careful calibration and validation of the UAV systems to ensure consistency in the data used for analysis.
- Furthermore, the study focuses primarily on specific input variables, such as altitude, speed, and population density, which may not encompass all critical factors influencing disaster severity. Variables like socioeconomic factors, infrastructure resilience, and local governance mechanisms are also significant in real-world disaster scenarios but were not included in the current modeling approach. Ignoring these additional variables could limit the comprehensiveness of the disaster index and its applicability in diverse contexts.
- Lastly, the proposed deep learning model, while demonstrating a high $r2$ score, may exhibit overfitting if not appropriately regularized. Overfitting can lead to a model that performs well on training data but poorly in real-world applications, where new, unseen data may differ from the training set. Continuous monitoring and updating of the model based on incoming data will be necessary to maintain its accuracy and reliability over time.

V. Conclusion

The present study demonstrates the potential of integrating statistical mode with deep learning techniques, specifically MLP for the prediction of efficient disaster severity. The disaster severity was classified into three levels: low, moderate, and extreme. The structural equation model demonstrates that Altitude, Speed, Area Coverage, Temperature, and Humidity has a significant impact on the disaster index with a p-values less than 0.001. All of the significant variables are taken as the input variables for the MLP model. The MLP model achieves an r2 score of 0.86 for the prediction of disaster severity. The study highlights several limitations, including challenges related to data acquisition and the generalizability of the model to diverse environments. The current model focuses primarily on data, and incorporating other data sources such as ground-based sensors or weather information could improve its robustness. Additionally, the sensitivity analysis shows that UAV speed plays an important predictor of disaster index. Future research should address these limitations by expanding the model's capabilities to handle multimodal data and adapt to changing conditions in real-time. In conclusion, this study provides a strong foundation for the use of UAV based deep learning systems in disaster management, particularly for emergency response and resource allocation. The high regression accuracy and real-time decision-making potential of this model make it a valuable tool for disaster identification.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at Saudi Electronic University for funding this research work through the project number (8398).

Acknowledgments: The authors thank to the Deanship of Scientific Research at Saudi Electronic University for funding this research work through the project number (8398).

References

[1] Adelekan, I. O. (2020). Urban dynamics, everyday hazards and disaster risks in Ibadan, Nigeria. Environment and Urbanization, 32(1), 213-232.
[2] Sharma, S. C. (1994). Disaster management. KHANNA PUBLISHING HOUSE.
[3] Tiwari, S. K., Kumaraswamidhas, L. A., Patel, R., Garg, N., & Vallisree, S. (2024). Traffic noise measurement, mapping, and modeling using soft computing techniques for mid-sized smart Indian city. Measurement: Sensors, 33, 101203.
[4] Puttinaovarat, S., & Horkaew, P. (2020). Internetworking flood disaster mitigation system based on remote sensing and mobile GIS. Geomatics, Natural Hazards and Risk, 11(1), 1886-1911.
[5] Mazzoglio, P., Ajmar, A., Schumann, G. J., Balbo, S., Boccardo, P., Perez, F., & Borgogno-Mondino, E. (2021). Satellite-based approaches in the detection and monitoring of selected hydrometeorological disasters. The Increasing Risk of Floods and Tornadoes in Southern Africa, 19-37.
[6] Dell'Acqua, F., & Gamba, P. (2012). Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. Proceedings of the IEEE, 100(10), 2876-2890.
[7] Bushnaq, O. M., Mishra, D., Natalizio, E., & Akyildiz, I. F. (2022). Unmanned aerial vehicles (UAVs) for disaster management. In Nanotechnology-Based Smart Remote Sensing Networks for Disaster Prevention (pp. 159-188). Elsevier.
[8] Nikhil, N., Shreyas, S. M., Vyshnavi, G., & Yadav, S. (2020, August). Unmanned aerial vehicles (UAV) in disaster management applications. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 140-148). IEEE.
[9] Alawad, W., Halima, N. B., & Aziz, L. (2023). An unmanned aerial vehicle (UAV) system for disaster and crisis management in smart cities. Electronics, 12(4), 1051.
[10] Kakooei, M., & Baleghi, Y. (2017). Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment. International journal of remote sensing, 38(8-10), 2511-2534.
[11] Hayajneh, A. M., Zaidi, S. A. R., McLernon, D. C., Di Renzo, M., & Ghogho, M. (2018). Performance analysis of UAV enabled disaster recovery networks: A stochastic geometric framework based on cluster processes. IEEE Access, 6, 26215-26230.
[12] Tiwari, S. K., Kumaraswamidhas, L. A., Gautam, C., & Garg, N. (2022). An auto-encoder based LSTM model for prediction of ambient noise levels. Applied Acoustics, 195, 108849.
[13] Tiwari, S. K., Kumaraswamidhas, L. A., Prince, Kamal, M., & Rehman, M. U. (2023). A hybrid deep leaning model for prediction and parametric sensitivity analysis of noise annoyance. Environmental Science and Pollution Research, 30(17), 49666-49684.
[14] Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. Annals of tourism research, 75, 410-423.
[15] Hassani, H., Huang, X., Silva, E., & Ghodsi, M. (2020). Deep learning and implementations in banking. Annals of Data Science, 7, 433-446.
[16] Munawar, H. S., Ullah, F., Qayyum, S., Khan, S. I., & Mojtahedi, M. (2021). UAVs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection. Sustainability, 13(14), 7547.
[17] Lohumi, K., & Roy, S. (2018, December). Automatic detection of flood severity level from flood videos using deep learning models. In 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) (pp. 1-7). IEEE.

[18] Hussain Shah, S. M., Yassin, M. A., Abba, S. I., Lawal, D. U., Hussein Al-Qadami, E. H., Teo, F. Y., ... & Aljundi, I. H. (2023). Flood Risk and Vulnerability from a Changing Climate Perspective: An Overview Focusing on Flash Floods and Associated Hazards in Jeddah. Water, 15(20), 3641.

[19] Hassan, A., Albanai, J. A., Al-Ali, J., Fayad, M., A. Atalla, M., Abdelkarim, A., & Badawy, H. (2024). Assessment of flash flood risks in the desert cities: a case study on Sabah Al-Ahmad, Kuwait. Journal of Water and Climate Change, jwc2024191.

[20] Zhang, L., & Xia, J. (2021). Flood detection using multiple Chinese satellite datasets during 2020 China summer floods. Remote Sensing, 14(1), 51.

[21] Pradhan, A. M. S., & Kim, Y. T. (2020). Rainfall-induced shallow landslide susceptibility mapping at two adjacent catchments using advanced machine learning algorithms. ISPRS International Journal of Geo-Information, 9(10), 569.

[22] Boccardo, P., Chiabrando, F., Dutto, F., Giulio Tonolo, F., & Lingua, A. (2015). UAV deployment exercise for mapping purposes: Evaluation of emergency response applications. Sensors, 15(7), 15717-15737.

[23] Zeng, C., Richardson, M., & King, D. J. (2017). The impacts of environmental variables on water reflectance measured using a lightweight unmanned aerial vehicle (UAV)-based spectrometer system. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 217-230.

[24] El Hoummaidi, L., Larabi, A., & Alam, K. (2021). Using unmanned aerial systems and deep learning for agriculture mapping in Dubai. Heliyon, 7(10).

[25] Khan, S., Alourani, A., Mishra, B., Ali, A., & Kamal, M. (2022). Developing a credit card fraud detection model using machine learning approaches. *International Journal of Advanced Computer Science and Applications*, *13*(3).

[26] Wasiq, M., Kamal, M., & Ali, N. (2023). Factors influencing green innovation adoption and its impact on the sustainability performance of small-and medium-sized enterprises in Saudi Arabia. *Sustainability*, *15*(3), 2447.

[27] Zeraibi, A., Jahanger, A., Usman, M., Balsalobre-Lorente, D., Adebayo, T. S., & Kamal, M. (2024). The role of fiscal decentralization and technological innovations in curbing sulfur dioxide emissions: formulating SDGs policies for China. *Environment, Development and Sustainability*, *26*(8), 19659-19684.

[28] Munawar, H. S., Ullah, F., Qayyum, S., Khan, S. I., & Mojtahedi, M. (2021). UAVs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection. Sustainability, 13(14), 7547.

[29] Zhou, Y., Wang, J., Grigorieva, E., Egidarev, E., & Zhang, W. (2020). Comprehensive Spatio-Temporal Analysis of Travel Climate Comfort Degree and Rainstorm-Flood Disaster Risk in the China–Russia Border Region. Sustainability, 12(8), 3254.

[30] Hu, P., Zhang, Q., Shi, P., Chen, B., & Fang, J. (2018). Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. Science of the Total Environment, 643, 171-182.

[31] Munawar, H. S., Hammad, A. W., & Waller, S. T. (2022). Disaster region coverage using drones: Maximum area coverage and minimum resource utilisation. Drones, 6(4), 96.

[32] Pereira, J., Monteiro, J., Silva, J., Estima, J., & Martins, B. (2020). Assessing flood severity from crowdsourced social media photos with deep neural networks. Multimedia Tools and Applications, 79(35), 26197-26223.

[33] Lohumi, K., & Roy, S. (2018, December). Automatic detection of flood severity level from flood videos using deep learning models. In 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) (pp. 1-7). IEEE.

[34] Youssef, A. M., Sefry, S. A., Pradhan, B., & Alfadail, E. A. (2016). Analysis on causes of flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS. Geomatics, Natural Hazards and Risk, 7(3), 1018-1042.

[35] https://www.my.gov.sa/wps/portal/snp/agencies/agencyDetails/AC040/!ut/p/z0/04_Sj9CPykssy0xPLMnMz0vMAfIjo8zivQIsTAwdDQz9LQwNzQwCnS0tXPwMvYwNDAz0g1Pz9L30o_ArAppiVOTr7JuuH1WQWJKhm5mXlq8f4ehsYGKgX5DtHg4A39XO3w!!/

[36] https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1

[37] https://www.stats.gov.sa/en

[38] Tiwari, S. K., Kumaraswamidhas, L. A., & Garg, N. (2023). Assessment of noise pollution and associated subjective health complaints in Jharia Coalfield, India: A structural equation model analysis. Noise Mapping, 10(1), 20220172.

[39] Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. Acta numerica, 8, 143-195.

A CRITICAL REVIEW OF RAM METHODOLOGY: ANALYSIS AND PERFORMANCE EVALUATION IN INDUSTRIAL COMPLEXITIES

Pardeep kumar [1*], Dinesh Kumar [2], Rupesh Chalisgaonkar [3], Vipin Kumar Sharma[4], Santosh Kumar Rai [5]

•

[1,2]Department of Mechanical Engineering, MMEC, Maharishi Markandeshwar (Deemed to be University), Ambala, India
[3]Department of Mechanical Engineering, Medi-Caps University, Indore, India
[4]Department of Mechanical Engineering, IIMT University, Meerut, India
[5]Department of Mechanical Engineering, SRM Institute of Science & Technology, Delhi NCR Campus, Ghazibad, India

[1]pardeepkumar456@gmail.com
[2]dinesh_kumar@mmumullana.org
[3]rupesh_chalisgaonkar2000@yahoo.com
[4]vipin2871985@gmail.com
[5]08rai.santosh@gmail.com

Abstract

This paper investigates the reliability, availability and maintainability (RAM) characteristics of a in different systems of the process industries. Critical mechanical subsystems with respect to failure frequency, reliability and maintainability are identified for taking necessary measures for enhancing availability of the respective industries. As complexity of the systems increasing across the various sectors so performance evaluation becomes necessary for the smooth functioning of all the systems of respective industry. The study explores the evolution of RAM approaches over time, highlighting their significance in ensuring the efficient operation of intricate systems. It provides an overview of the historical development and current state of RAM practices in the complex system of the industries. A comprehensive review of academic literature from the past two decades, including books, journals, and scholarly articles, is conducted to expand the analysis, mainly focus on the evaluating RAM methodology in diverse industrial contexts, different complex system and other process industries.

Keywords: Reliability, Availability, Maintainability, Performance Evaluation.

I. Introduction

Today, industries are increasingly focusing on the principles of Reliability, Availability, and Maintainability (RAM). Growing competition, tighter production schedules, budget constraints, and the constant demand for cheaper, higher-quality products have heightened the importance of utilizing RAM engineering tools. As industrial systems become more complex, reliability simulation has emerged as the most effective and preferred method for addressing modern real-world challenges—challenges that are often difficult or even impossible to solve using traditional analytical methods. This simulation-based approach is now widely adopted across various

industries, with RAM analysis becoming the standard tool.

Reliability of a system is associated with the prediction and occurrence of failures (or more correctly, the lack of failures). The term MTBF (Mean Time between Failures) is generally used to measure the reliability of repairable systems and similarly MTTF (Mean Time to Failure) term is commonly used for non-repairable systems. Reliability is more accurately expressed as it is the probability of a machine that can perform its intended function successfully (i.e. without failure) over a given duration of time under specified conditions. The performance of the plant depends upon the reliability of different components and equipment used. Reliability is crucial factor because it affects such as product quality, production capacity and profitability.

Availability is a key parameter used to evaluate the performance of repairable systems, as it depends on both reliability and maintainability. It is defined as the probability that a system will be operational when needed. In other words, availability is the likelihood that a repairable system is functioning and not experiencing a failure or undergoing repairs at a specific time. This metric ranges from 0 to 1 and accounts for both failure and repair rates.

According to British Standards 4778, availability is the ability of a product, considering its reliability and maintainability, to perform its intended function over a specified time period and under defined environmental conditions. Put simply, it reflects the probability that a system is operational at a given moment—the period during which the equipment is actively functioning. Availability can be calculated by measuring the downtime of a machine or piece of equipment. However, for repairable systems, availability is typically analyzed by considering the failure and repair rates of its sub-systems. It's important to note that a system may be available but not necessarily reliable.

Maintainability, while related to both reliability and availability, has its own distinct meaning and significance. It is typically defined as the probability that a machine or its components can be repaired and restored to working condition within a specified time frame. Maintainability and availability are closely linked—availability decreases as repair time increases for a failed system. In other words, maintainability refers to how quickly a system can be restored to its original state after a failure. Also known as serviceability, maintainability involves various methods aimed at detecting problems early and minimizing system downtime. Ultimately, the combination of reliability and maintainability forms the basis for the concept of availability.

This paper provides a critical review of the literature on RAM (Reliability, Availability, and Maintainability) methodologies, focusing on improving system availability while reducing operational costs. Additionally, the paper explores how various optimization techniques, based on performance analysis and maintenance requirements can further enhance system performance. Modelling tools such as Monte Carlo simulations, Petri Nets, Weibull analysis, Pareto analysis, and Markov modeling are highlighted as valuable in selecting effective maintenance and replacement strategies.

II. Literature Review

The literature review comprehensively study of literature has led to a detailed investigation into the aspects of reliability, availability and maintainability. The literature review primarily covers the period from 2000 to 2023, providing a contemporary perspective on the field. Among the various tools and methodologies for modeling system performance, RAMS stands out as a comprehensive approach that addresses both the engineering and management aspects of a plant. RAMS is particularly valuable as it assesses system performance across the design and production phases, offering insights that span the entire lifecycle of the system.

Bansal et al. [1] purposed a study approach based on neural network for optimized the parameters and performance of the casting process. With the help of this technique, availability, reliability and profitability of the system has been analyzed which help to predict the critical

component of the system so that an effective maintenance scheduled can be planned.

Kumar et al. [2] analyzed the availability of the system by dividing the thermal power plant into six distinct systems and found that availability is significantly low, with boiler tube failure identified as a major contributing factor. Then, with the help of implementing an effective maintenance schedule to reduce the failure rate which help to decrease the plant shutdown periods and enhance overall system availability.

Daya and Lazakis [3] investigated the specific challenges faced by operators related to environmental conditions, operational demands, and technological issues impacting availability, maintenance and reliability of the system. A framework was developed using outputs from DFTA minimal cut set analysis and RPN from FMECA, which served as inputs for a Bayesian Belief Network (BBN) to analyze the availability of a ship's marine diesel generator (DG) system. With the help of BBN influence diagram, a decision support system (DSS)has been developed for maintenance strategy selection. The overall outcome of the DSS indicated a relatively high level of system unavailability.

Singla et al. [4] examined the most critical subsystem by taking variable failure and repair rates from all different subsystems. It focuses on the sensitivity analysis of a complex repairable threshing machine, which consists of 21 subsystems arranged in a series configuration. The machine operates at full capacity when the threshing drum and feeding hopper are fully functional, while the concave subsystem and blower operate at reduced power.

Kumar et al. [5] examined the operational availability of the Veneer layup system at a plywood production facility using Petri Nets for modeling. The study examined how different subsystem failure and repair rates affect system availability, highlighting crucial subsystems that need special attention. The results help maintenance engineers develop efficient planning techniques.

Kumar et al. [6] analyzed the factors which are significant causes of poor availability in thermal power plant. The authors developed the mathematical equations which have been used to find out the availability and it was found that the value of availability is very low and boiler tube failure is one of the most critical factors for this low availability of system.

Berrouane et al. [7, 8] proposed the maintenance decisions support system for the Coal Ash Handling System of a subcritical Thermal Power Plant. By using state probabilities and normalization conditions, a model was developed and then performability matrices were created based on repair and failure rates for all the subsystems. On further analysis, it was observed that ESP as the most critical and ash silo as the least critical subsystem, which help to make a Decision support system.

Kumar et al. [9] analyzed the poor availability of a thermal power plant, which was divided into six systems, including a waste gas heating system. Boiler tube failure, especially in the economizer zone due to erosion from high-velocity flue gas, was identified as a key factor. Reducing erosion in economizer tubes through CFD analysis can minimize failures, shorten shutdown periods, and improve system availability.

Parkash & Tewari [10] reviewed the effectiveness of Reliability, Availability, Maintainability, and Safety (RAMS) approaches in different mechanical systems of process industries. It examines a wide range of research, including articles, conference papers, and books, that address RAMS applications in both industry and research. The review covers various tools, techniques, and methods useful for both qualitative and quantitative analysis, providing insights into the past and present state of RAMS practices.

Parkash & Tewari [11] focused on performance modeling and proposed a Decision Support System (DSS) in an assembly line for prioritizing maintenance system using a probabilistic approach. The system divided into various subsystems: Shot Peening, Painting Machine, Assembly Platform, and Riveting Machine. A Markovian approach was used to model performance, with

steady-state probabilities calculated through a transition diagram and differential equations and decision matrices were created with different failure and repair rates.

Kumar et al. [12, 13] developed a maintenance decision support system for repairable equipment of a thermal power plant so that the availability and performance have been improved.

Lu et al. [14] suggested a method of quality improvement and preventative maintenance for a system. To give the financial gains, a machine reliability model that considers the effects of component deterioration states on quality is built based on the proportional hazard model.

Pan et al. [15] assessed the reliability of the Reactor Protection System (RPS) using the proposed method, which is then compared to models based on two other contrasting methods. This comparison confirms the effectiveness and accuracy of the proposed approach.

Kumar et al. [16] presented a decision support system for providing maintenance orders, critical for efficient operations. It also evaluates the performability of a coal handling system in a thermal power plant using stochastic Petri nets (SPN) and GRIF software. By varying failure and repair rates, performability metrics were used to predict subsystem maintenance priorities. The study analyzes how these variations affect system performance and availability.

Patel and Joshi [17] proposed a model for complex manufacturing systems with deadlock, followed by Petri Net analysis to generate the reachability tree. The aim is to determine the minimum number of pieces required to maintain a desired throughput level while ensuring sufficient time to meet batch size constraints.

Kołowrocki et al. [18] presented a simulation method based on Monte Carlo for evaluating the reliability of a multistate system under a variable operation process. By linking the system reliability model and operation process model, a general reliability model is proposed to assess the system's performance under changing conditions and determine its reliability characteristics.

Kumar [19] examined the impact of failure and repair data on the availability and reduced capacity of a liquid milk processing plant. Using the Petri Nets approach, the system is modeled by dividing it into four subsystems in series and findings are useful not only for plant owners but also for professionals in other fields who wish to analyze and predict equipment behavior.

Kajal et al. [20] suggested a decision support system based on their study of the butter oil unit of a dairy plant. The butter oil unit's mathematical model was designed by utilizing the Markov Birth-Death process. The unit's differential equations were derived via a probabilistic approach that utilized a transition diagram.

Sachdeva et al. [21] evaluated an enhanced maintenance planning approach to reduce operating and maintenance costs by modeling and examining the performance of a feeding system using Petri Nets (PN) and a Markovian method in the paper industry.

Gupta et al. [22] Utilized a Markovian method to develop a Decision Support System (DSS) after assessing the performance of the feedwater system in a thermal power plant. The maintenance department uses the DSS to determine optimal maintenance intervals.

Sania et al. [23] presented a model to assess system reliability, including availability, profit, and mean time to failure (MTTF), considering three stages of deterioration (minor, medium, major). Markov models were used, based on state transitions and differential equations, to evaluate availability, busy period, profit, and MTTF.

Aven [24] reviewed key concepts of risk analysis and management by addressing common myths, such as equating risk with expected value, uncertainty, or probabilities. It concludes that risk involves both consequences and uncertainties, extending beyond expected values and probabilities.

Shubinsky et al. [25] proposed a method of technical systems for calculating reliability and functional safety, which allows deriving exact formulas for stationary parameters directly from the system state graph. This method applied to both Markov and semi-Markov models and also

included the examples of calculating safety, availability factors, and time parameters for a two-channel safety-related device.

Ramirez-Marquez et al. [26] examined a complex system called a multi-state series-parallel system (MSPS). This system consists of binary components with limited capacity, which is capable of delivering various degrees of performance for the overall system. They claim that the varying degrees of demand that need to be fulfilled throughout the system's operation make the system multi-state, and that the new solution technique provides several unique advantages.

Ebrahimi N.B. [27] outlined the aspects of Grid and Grid modeling, a distributed software-hardware environment with a new computation and job flow management structure. A general model scheme was developed to analyze user-resource interaction issues and formulated specific mathematical tasks and discussed methods for solving them.

Tang J. [28] developed a novel method for assessing mechanical system reliability was proposed, utilizing graph theory and Boolean functions. Graph theory was first applied to develop a reliability formula that accounts for the connections between subsystems and components. Then, the failure interactions between two specific system components were analyzed using Boolean functions.

Savsar, M. et al. [29] developed an appropriate mathematical model incorporating a versatile machine, a pallet handling system, and a loading/unloading robot. They made the assumption that the times of operation, loading/unloading, and material handling were random. Afterwards, the operation of a fully reliable and an unreliable flexible manufacturing cell (FMC) were analysed and compared.

Kumar et al. [30] suggested a maintenance plan for the proper working of a beer plant so that the availability can be improved. With the help of Markov, they found the performance and then improved the performance by reducing the failure rate with a well plan maintenance policy.

III. Research Gap

After thoroughly reviewing the literature, some concise observations regarding the use of RAM methodology in different complex system of the process industries:

- Most of Most research has focused on reliability in process industries like thermal power plants, ply-board processing unit, paper industries, sugar, fertilizer, dairy plants, and oil refineries, rather than in product-based industries.
- Many researchers have studied methods to improve plant availability through effective maintenance, but few have examined the link between availability and break-even analysis. It's important to note that higher plant availability directly reduces the time needed to reach the break-even point.
- The literature review reveals that several methods are used for both quantitative and qualitative plant analysis, including Reliability Block Diagrams (RBD), Fault Tree Analysis, Markov models, and Petri Nets. Each method has its pros and cons. For example, the Markov chain is a powerful tool for stochastic reliability and availability analysis, but it faces the challenge of state explosion, even in small systems. Petri Nets, which consist of Places, Transitions, Arcs, and Tokens, have gained attention due to their simplicity and balanced modeling and decision-making capabilities, though they tend to increase system complexity.
- Most studies have used steady-state availability to evaluate the performance of industrial systems through Markov modeling. However, only a few have tackled these mathematical models under more realistic industrial conditions.

IV. Conclusion and Future Scope

A critical review of the available literature highlights the effectiveness of various RAMS tools and techniques applied in different plants to reduce the cost of plant non-availability. Plants are typically divided into multiple systems or subsystems for effective maintenance planning, ensuring prolonged system availability. A decision support framework using statistical analysis has been developed, utilizing tools such as Reliability Hazard Analysis, FMEA, Reliability Block Diagrams, Fault Tree Analysis, Reliability Growth Analysis, Root Cause Analysis, Finite Element Analysis, Markov Analysis, and Petri Nets. Each of these techniques has distinct advantages and disadvantages, which are discussed in the paper.

Conventional modeling and optimization techniques may be insufficient for addressing issues in plants with continuous wear and tear, like those in product industries. Therefore, advanced methodologies such as Fuzzy Logic, Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Artificial Neural Networks (ANN), and MATLAB software are being implemented to tackle these challenges more effectively.

References

[1] Bansal, S., & Tyagi, S. L. (2024). Availability and profit optimization of continuous casting system of the steel industry using artificial neural network technique. *Reliability: Theory & Applications*, *19*(3 (79)); 47-58.

[2] Kumar, P., Sharma, V.K. and Kumar, D. (2024). Availability analysis for identification of critical factor of a thermal power plant. *Reliability: Theory & Applications*, *19*(3 (79)), pp.717-724.

[3] Daya, A. A., & Lazakis, I. (2023). Developing an advanced reliability analysis framework for marine systems operations and maintenance. *Ocean Engineering*, *272*, 113766.

[4] Singla, S., Modibbo, U. M., Mijinyawa, M., Malik, S., Verma, S., & Khurana, P. (2022). Mathematical model for analysing availability of threshing combine machine under reduced capacity. *Yugoslav Journal of Operations Research*, *32*(4), 425-437.

[5] Kumar, N., Tewari, P. C., & Sachdeva, A. (2020). Petri Nets modelling and analysis of the Veneer Layup system of plywood manufacturing plant. *International Journal for Engineering Modelling*, *33*(1-2 Regular Issue); 95-107.

[6] Kumar, P., Ali, A., & Kumar, S. (2020). Coal-fired thermal power plant performance optimization using Markov and CFD analysis. *SN Applied Sciences*, *2*; 1-10.

[7] Malik, S., & Tewari, P. C. (2023). Performability and maintenance decisions for coal ash handling system of a subcritical thermal power plant. *International Journal of System Assurance Engineering and Management*, *14*(1), 45-54.

[8] Malik, S., Verma, S., Gupta, A., Sharma, G., & Singla, S. (2022). Performability evaluation, validation and optimization for the steam generation system of a coal-fired thermal power plant. *MethodsX*, *9*, 101852.

[9] Kumar, P., Ali, A., & Kumar, S. (2020). Coal-fired thermal power plant performance optimization using Markov and CFD analysis. *SN Applied Sciences*, *2*, 1-10.

[10] Parkash, S., & Tewari, P. C. (2021). Critical Review of Rams Tools and Techinques For The Analysis of Multi Component Complex Systems. *Reliability: Theory & Applications*, *16*(3 (63)), 220-227.

[11] Parkash, S., & Tewari, P. C. (2022). Performance modeling and dss for assembly line system of leaf spring manufacturing plant. *Reliability: Theory & Applications*, *17*(2 (68)), 403-

412.

[12] Kumar, P., Tewari, P. C., & Khanduja, D. (2016). Maintenance Priorities for a Repairable System of a Thermal Power Plant Subject to Availability Constraint. *International Journal of Performability Engineering*, *12*(6).

[13] Kumar, P., Khanduja, D., & Tewari, P. C. (2017). Maintenance strategy for a system of a thermal power plant. *Int. J. Oper. Quant. Manag.*, *23*(1); 23-40.

[14] Lu, B., Zhou, X., & Li, Y. (2016). Joint modelling of preventive maintenance and quality improvement for deteriorating single-machine manufacturing systems. *Computers & Industrial Engineering*. *91*; 188-196.

[15] Pan, X., Chen, H., Shen, A., Zhao, D., & Su, X. (2024). A Reliability Assessment Method for Complex Systems Based on Non-Homogeneous Markov Processes. *Sensors*, *24*(11), 3446.

[16] Kumar, S., Tewari, P. C., & Sachdeva, A. K. (2024). Decision support system for maintenance order priority of multistate coal handling system with hot redundancy. *International Journal of Intelligent Enterprise*, *11*(2), 103-119.

[17] Patel, A. M., & Joshi, A. Y. (2013). Modeling and analysis of a manufacturing system with deadlocks to generate the reachability tree using petri net system. *Procedia Engineering*, *6*;, 775-784.

[18] Kołowrocki, K., Kuligowska, E., & Soszyńska-Budny, J. (2015). Reliability of complex system under operation process influence: Monte Carlo simulation Approach. *Journal of Polish Safety and Reliability Association*, *6*.

[19] Kumar, N., Tewari, P. C., & Sachdeva, A. (2020). Performance assessment of milk processing system using Petri nets. *SN Applied Sciences*, *2*, 1-12.

[20] Kajal, S., Tewari, P. C., & Kausik, D. (2010). Decision Support System for Butter Oil System of a Dairy Plant at NDRI. *IUP Journal of Science & Technology*, *6*(1).

[21] Sachdeva, A., Kumar, D., & Kumar, P. (2008). Reliability analysis of pulping system using Petri nets. *International Journal of Quality & Reliability Management*, *25*(8); 860-877.

[22] Gupta, P., Lal, A. K., Sharma, R. K., & Singh, J. (2007). Analysis of reliability and availability of serial processes of plastic-pipe manufacturing plant: A case study. *International Journal of Quality & Reliability Management*, *24*(4); 404-419.

[23] Sania, B., Gatawab, R. I., & Yusufc, I. (2017). Reliability assessment of deteriorating system. *Reliability: Theory & Applications*, *12*(3 (46)), 20-29.

[24] Aven, T. (2009). Risk analysis and management. Basic concepts and principles. *Reliability: Theory & Applications*, *4*(1 (12)), 57-73.

[25] Shubinsky Igor, B., & Zamyshlyaev Alexey, M. (2012). Topological semi-markov method for calculation of stationary parameters of reliability and functional safety of technical systems. *Reliability: Theory & Applications*, *7*(2).

[26] Ramirez-Marquez, J. E., & Coit, D. W. (2004). A heuristic for solving the redundancy allocation problem for multi-state series-parallel systems. *Reliability Engineering & System Safety*, *83*(3); 341-349.

[27] Gaeta, M., Konovalov, M., & Shorgin, S. (2006). Development of mathematical models and methods of task distribution in distributed computing system. *Reliability: Theory & Applications*, *1*(4 (4)), 16-21.

[28] Tang, J. (2001). Mechanical system reliability analysis using a combination of graph theory and Boolean function. *Reliability Engineering & System Safety*, *72*(1); 21-30.

[29] Savsar, M. (2000). Reliability analysis of a flexible manufacturing cell. *Reliability Engineering & System Safety*, *67*(2); 147-152.

[30] Kumar, R., Kumar, P., & Joshi, A. (2014). Maintenance strategy for beer manufacturing plant. *International Journal of Mechanical Engineering and Robotics Research*, *3*(3); 95.

MANAGEMENT OF REGIONAL RESILIENCE THROUGH GOVERNANCE OF INFRASTRUCTURE OPERATIONAL RISK

Sviatoslav Timashev[1,2], Tatyana Kovalchuk[1]

•

[1]Science & Engineering Center «Reliability and Safety of Large Systems and Machines» Ural Branch Russian Academy of Sciences, Russia, 620049, Yekaterinburg, St. Studencheskaya, 54-A
[2]Ural Federal University, Russia, 620002, Yekaterinburg, St. Mira, 19
timashevs@gmail.com

Abstract

In this paper the notion of urban infrastructure resilience, expressed verbally and strictly in conditional probability terms, is formulated. It is then used to formulate several most important features of a smart city. This multidisciplinary and multifaceted approach is used to explain the concept of quantitative resilience in urban design, operation, managing urban risk and mitigating of the consequences of a natural or industrial disaster. The super urgent problem is formulated on how to connect the physical and spatial (core) resiliencies with the functional, organizational, economic and social resiliencies.

Keywords: urban resiliencies, interdependent critical infrastructures, regional risk, potentially dangerous objects, systems of systems, average life expectancy, life quality index, carbon foot print.

I. Introduction

Currently, the concept of resilience emerged as a central theme of industrial and urban development (there are more than 120 definitions of resilience, most of them are qualitative). It is capable of serving as the basis and tool for solving the most urgent issues of modern civilization, including strategic investments by leading development institutions and humanitarian communities around the world. Despite the importance of critical infrastructures and systems and expected growth of future climatic hazards, relatively few studies have addressed these issues and no methodology for the analysis of such an impact has ever reached a general consensus. As of now, it seems (to our knowledge) that there is no quantitative definition of resilience and strategic preparedness to which a majority would subscribe.

Cities are conglomerates where the majority of humanity lives. In the last century and decades urban ecosystems have evolved from simple clusters of buildings to extremely complex systems of systems (vast networks of services) strongly interdependent on one another. In order to become a resilient city, it is necessary that different stakeholders find a way to compromise through thoughtful governance, and carry out operational procedures of complex or territorial systems in groups.

The quantitative and qualitative analysis of resilience as related to urban infrastructures takes

its roots from the notion and concept of industrial resilience [1–4]. In this paper the urban infrastructure resilience is defined both verbally and strictly in conditional probabilistic terms, as all the parameters which describe resilience quantitatively, are random. The conditionality of the resilience probabilities is due to the probabilistic and uncertain/fuzzy nature of the impact, and of the financial, social and other restrictions on the critical infrastructure, for which the resilience is assessed.

Resilience is considered as a probability vector, which components include the physical, environmental, and spatial resilience, as well as functional, organizational, economic and social resilience. Each component of the resilience vector can be considered as a partial resilience as related to different aspects of the considered type of resilience. The physical and spatial resilience of a system of critical infrastructures is defined through its reliability and operational risk.

It is possible to implement a multidisciplinary and multifaceted approach to urban critical and strategic infrastructures of different nature, using the above novel concept of quantitative resilience in design, operation and mitigation of the consequences of an urban disaster.

The paper describes verbally and quantitatively how to manage urban risk by exercising governance which balances technological innovation, economic competitiveness, environmental protection and social flourishing using such regional criteria as resilience, life expectance, life quality index, carbon foot print and regional entropy.

The overarching idea of this research is that *human life time is the only true and universal measure of things*. Probably, the first who formulated this was Henry David Thoreau, who in his novel «Walden», 1852, wrote: «The cost of a thing is the amount of what I will call life which is required to be exchanged for it, immediately or in the long run».

Before addressing these issues, it is necessary to introduce several definitions and notions, which follow. Main sources of regional risk are: nature; systems of interdependent critical infrastructures (SICIs); people (population of the region, and vandals). Critical infrastructure is any large localized or, more often, distributed across a specific territory multicomponent geotechnical man-machine-environment system, which consists of many potentially dangerous objects (PDOs) and groups of people, who operate and/or live near these objects. Practically, the CIs are PDOs, which *cannot function without each other* when producing some sellable product or/and service, *and hence, are interdependent*. The ICI must operate effectively, meet the standards of well-being and safety for the population, and provide environmental sustainability of the region.

II. Short history of the resilience concept

In Russian language the word for resilience (*zhivuchest*) has as its root the Russian word *zhivoy* (*alive*) and means capacity of an object (*animate* or *inanimate*) to continue, *without interruption*, its functioning (staying «*alive*»), while being damaged by extreme loads, forces and/or influences. The Russian Dictionaries define *zhivuchest* as human longevity, or as capability for robust performance of a ship under the influence of wind, waves, fire, and enemy artillery. It, obviously, contains in itself the notion of *endurance*. The English word resilience originated from *resilire*, the Latin word meaning skip *backward* or *rebound*. In the Western world the concept of *resilience* as a common notion was adopted early 17-th century and, by the end of that century, the concept evolved to mean the *ability to react after (not during) a shock*.

It seems that Admiral O.S. Makarov of Russian Navy was the first who introduced, way back in 1894, the notion of infrastructure *resilience* in reference to the ability of battle ships (as elements of the war infrastructure) *to continue their effective performance* under artillery fire and subsequent damage [5]. Seven years later, in 1901 the Charpy impact V-notch high strain-rate test, that determines the amount of energy absorbed by a material during fracture, was standardized, and its *resilience* was defined as the capacity of the test material to absorb an impact. So the modern

concept of resilience comes from the metal industry, in which it is defined as the capacity of steel to withstand an impact, to maintain its shape, and to recover fast after receiving an impact. The notion of resilience was then gradually used in defining the behavior of different mechanical systems (beams, trusses, shell-like structures, bridges, as well as industrial and military vehicles and machines, etc.).

Fast forward, during the 20-th century, the term *resilience* penetrated many branches of science and is now widely used in physics, psychology, anthropology, economy, and ecology. During this period resilience is linked to such concepts as elasticity, longevity, durability and persistence, ability to adapt and rebound.

The notion of resilience spread, starting from 1936, to social psychology, when describing human elasticity, good mood, resourcefulness and spirit. G. Vickers stated [5] (1965) that a system is resilient if it's able to survive despite being beaten by shocks, hazards, or distress from internal or external turbulences. In the seventies, the idea of resilience becomes present in the ecology field, where resilience is defined as a way to measure the persistence of a system, its capacity to absorb turbulences, and its ability to keep the healthy relationships between the population and the different layers of the system [6]. At the end of the last century, the idea of resilience was utilized in economics, where it was understood as the capacity to stay on course, keep a juridical entity or business going and maintaining stability in a volatile environment. In the early years of the 21-st century, the resilience concept has been embraced by management, organizational, FEMA and HSA (when dealing with incidents, catastrophes and terrorist attacks), safety engineering and computing. The concept of resilience is also becoming more and more relevant in climate change and sociology [7].

III. Urban resilience

Currently, cities contain in their bowels 50+ % of the world's population. With rural population decreasing, the global population growth will continue in urban areas. Cities attract people because of the quality of work, services, communications, security and social relations, etc. Although some of the living conditions in cities can be worse than in rural areas, but, in general, the urban quality of life is overwhelmingly better that rural existence.

Cities can be classified as: 1) large, medium and small; 2) metropolitan, urban, semi-urban and rural; 3) consolidated and developmental. They can belong to the four existing global environments – rich, modest, poor and miserable worlds. Any city can be classified by these three characteristic factors and the four world types in which it is located.

Cities grow at an ever increasing speed, and are currently subjected to a multitude of pressures, out of which the following four are of greatest concern: 1) evolution of technology and globalization at a neck breaking pace; 2) fast socioeconomic changes; 3) obvious climate change; 4) the growing needs of demanding citizens, due to a combination of the above elements.

All of the above creates new threats and the necessity to protect citizens from these threats. They could be effectively mitigated by the novel concept of *urban resilience*. According to the Rockefeller Foundation [8]: «*Urban Resilience is the capacity of individuals, communities, institutions, businesses, and systems within a city to survive, adapt, and grow, no matter what kinds of chronic stresses and acute shocks they experience, and even transform when conditions require it.*» The European Commission defines resilience (in general) as «*the ability of … a community, a country or a region to withstand, adapt, and quickly recover from stresses and shocks such as drought, violence, conflict or natural disaster*». Resilient systems refer to those institutions, cities or states that have the ability to reconstruct and to recover, using the right tools, assets and skills to deal with impacts and resist, absorb and adapt. The challenge here actually is how to manage these risks effectively, and how to transform a generic resilience concept into a tailored resilience of a specific infrastructure or a

system.

The world community adapted this idea in economics, social issues, and *urban contexts* when overcoming disaster and climate changes. The scale of urban risk is increasing due to the increase of the number of people living in cities. Risk is also increasingly unpredictable, due to the complexity of city's systems of systems (SoS) and the uncertainty associated with natural disasters, industrial catastrophes and climate change.

Consequences of climate change are increasingly dangerous for citizens and urban network services. Related events can cause tremendous economic impacts and losses for cities, due to their population density and asset concentration. Global economic losses from natural disasters have averaged almost $200 billion (1998) per year over the past decade, up from just $50 billion per year in the 1980s, according to the World Bank and Munich Re, the world's largest reinsurer. Total costs sum up to $3.8 trillion from 1980 to 2012. Three-quarters of this total was due to extreme weather-related events. According to some analysis [10], the economic impacts of these types of emergencies can be reduced by up to 10 %, using appropriate diagnostics, monitoring, maintenance and decision making tools. Environmental risks (extreme weather and climate change) outnumber economic (unemployment, underemployment or fiscal crises) risks, according to the latest report of the World Economic Forum (WEF).

The network of urban services is becoming more and more complex due to a high degree of interconnection between them and the evolution of the technology and requirements/demands of the citizens. Confusion may arise when dealing with the enormous databases (Big Data) that are created by the new information technology. All this implies that failure of one specific service endangers the ability of many other services to provide service, unleashing domino effects that spread far from the initial event. Another threat lies in that contemporary cities experience a decrease of city budgets and outsource the services to minimize costs. This leads to pronounced deficit of coordination and communication between the networks of urban services.

The total number of cities around the world interested in urban resilience is hard to estimate, but currently there are 1520 emerging cities associated to the UNISDR program *Making Cities Resilient* (mostly from Central and South America, Balkan Europe, India and Far East), a hundred cities in the Rockefeller Foundation *100 Resilient Cities Program*, located in North America and Europe, including such metropolitan cities like New York, London, Paris, Rome, and Barcelona [8]. This means that modern world wide urban society centers also on metropolitan regions, and the security, resilience and viability of metropolises also relies on their *critical infrastructures*, core public services and economic bases.

I. Brief verbal description of urban resilience

The definition of urban resilience was given above. Resilience is the driver and, simultaneously, a precious quality of *sustainable* urban development. Considering a city as a system of systems (SoS), resilience recognizes all of them as dynamic and complex systems that have to continuously adapt to various challenges of *stochastic, probabilistic, uncertain, or vague character* in an integrated and holistic manner. Each part of these systems has an inherent reliance on all the other parts.

In general, factors that influence city resilience include: the range and severity of hazards; the risk to life, limb and property; the vulnerability and exposure of human, social, and environmental systems, and the degree of (strategic) preparedness of the physical and the governance systems to any Natural or urban and industrial shocks and stressors and their consequences during an incident, accident or catastrophe. The resilience concept adopts a multiple hazards approach, considering resilience against all types of plausible hazards, and refers not only to reducing risks and damages from disasters (i.e. loss of lives, limbs and assets), but also the ability to quickly recover back to the pre-shock stable state. Using resilience concept permits correct placing of

strategic investments by leading development institutions and humanitarian communities around the world.

II. The essence and components of urban resilience

The essence and components of urban resilience consists of working to: 1) prevent any potential threat; 2) withstand any impact caused; 3) react to the crises derived from the impact; 4) recover the city's functionalities; 5) learn from the experience.

Urban resilience is based on humanitarian, vertical-operational and cross-cutting approaches. The *humanitarian approach* focuses on resilience of *cities*, regions and countries, taking into account how they are able to deal with industrial and natural disasters, climate and social changes, and how these impacts may affect the welfare of citizens in terms of life expectancy, life quality index (LQI), etc., and how the sustainability of the environment is affected. *The cross-cutting approach* to resilience implementation holistically tackles functioning of a city in a comprehensively manner, observing the city strategically as an ecosystem of interdependent urban services. The v*ertical-operational approach* studies the complex anatomy of a city, and its different elements (society – individuals, groups of people, government, etc.); services and infrastructures that contribute to the welfare of citizens, and the interactions between the society and its services and infrastructures.

Urban resilience consists of four phases: 1) *Operating* – the ability to withstand, respond or recover after an impact; 2) *Planning* – simulate incident and disaster events and decide accordingly; 3) *Improving* – learning from the experience in order to correct and project current strategies to ensure city continuity in the event of an impact; 4) *Preventing* – foresee potential hazards before they happen by identifying risks and diagnosing the environment.

The four main components of urban resilience are: industrial disaster and climate resilience, economic resilience, social resilience and urban resilience. All this is achieved when the city becomes smart. There are two competing concepts of Smart Cities. The first concept is used to equip the infrastructures and the services for the optimal management of the city as the path to reach a Resilient City. In the second concept the optimal management of a city goes through already resilient network of services and infrastructures equipped with smart technology to create a Smart City. In the latter case the concept of a smart city is formulated and builds up around optimizing implementation of following five key ideas [9, 10]:

- The win-win exchanging/sharing of goods and services between citizens and communities, using the common heritage or private property;
- The minimum environmental consumption and energy efficiency (minimal environmental/carbon footprint of the city), by recomposing the mix of energy consumption and the self-production of renewables;
- The free and fluid communication among social stakeholders (citizens, communities, companies, and institutions) using new digital technologies;
- City wide integration of new information and communication technologies, robotics and intelligent systems that maximize delivering needed information just in time;
- The network operation, which is the basis of resilience, to: 1) achieve maximum security of supply of goods and services with the right energy and environmental consumption; 2) make good use of the available infrastructure, and 3) provide the necessary social communication that will enable the city to adapt and recover functionalities in case of an impact.

Implementation of these ideas may include changes in the design and management of: 1) infrastructures, with emphasis on the redundancies and interconnections; 2) interdependent services, focusing on the ways they could support each other in case of an incident; 3) behavior of citizens in critical situations (the fundamental strategic element for improving urban resilience).

Ecosystem Resilience is the capacity of an ecosystem to *resist damage* and *quickly recover* after experiencing such stochastic disturbances as extremely high or low ambient temperatures, fires, flooding, wind and sand storms, tsunami, tornados, insect population explosions, and harmful human activities – deforestation, fracking for oil extraction, spraying of pesticides, and introduction of exotic plants or animal species, to name a few.

Disturbances of sufficient magnitude or duration can profoundly affect and may force an ecosystem to reach a tipping point (threshold) beyond which the system starts operating under entirely different set of regimes. Human activities that adversely affect ecosystem resilience (reduction of biodiversity, over usage of natural resources, pollution, and anthropogenic land-use and climate change) cause regime shifts in ecosystems, most often to less desirable and degraded conditions.

Interdisciplinary discourse on resilience now includes consideration of the interactions of humans and ecosystems via *socio-ecological* systems, and the need for shifting from the maximum sustainable yield paradigm to environmental resource management which aims to harmonize and build ecological and social resilience.

Climate resilience can be generally defined as the capacity of a socio-ecological system to: absorb and adapt stresses imposed upon it by climate change, reorganize and evolve into more desirable configurations that improve the sustainability of the system, leaving it better prepared for future climate change impacts. The key focus of climate resilience efforts is to address the vulnerability that communities, states, and countries currently have due to the environmental consequences of climate change. Currently, climate resilience efforts encompass social, economic, technological, and political strategies that are being implemented at all levels of society, with technological strategies being the core issue which attracts all other strategies as its satellites. From local community action *via cities* to global treaties, addressing climate resilience is becoming a priority. Although a significant amount of the theory has yet to be translated into practice, there is a robust and ever-growing movement, fueled by local (urban) and national bodies alike, towards building and improving climate resilience.

Urban resilience and sustainability. There are different system traits directly related to urban resilience: sustainability, vulnerability, flexibility and the like. As such, the concept of resilience is of great relevance for both urban city managers and politicians and they should be encouraged to manage cities through a resilience-based approach.

Urban resilience and vulnerability. Cities have always faced multiple risks over time. All cities that survived over the centuries and millennia of their existence have demonstrated their resilience in the face of resource shortages, natural hazards, and conflicts of various nature. The 21st century tectonic global shifts – climate change, disease pandemics, economic fluctuations, and terrorism – pose new challenges to the places we live. The *urban vulnerability* of a city is understood as *weakness against any harm* (like, exposure to floods, earthquakes, release of toxic chemicals and tsunami).

Dynamic urban resilience. Consequences of welfare impacts also depend on *micro-economic resilience,* which depends on the *distribution of losses***;** on *household vulnerability,* such as pre-disaster income and ability to smooth shocks over time with savings, borrowing, and insurance; and on the *social protection system,* or the mechanisms for sharing risks across the population. The (economic) welfare disaster risk in a city can be reduced by: (1) decreasing exposure or vulnerability of people and assets (reducing asset losses); (2) increasing macroeconomic resilience (reducing aggregate consumption losses for a given level of asset losses), or (3) increasing microeconomic resilience (reducing welfare losses for a given level of aggregate consumption losses).

IV. Quantitative description of resilience

The above *verbal* description of the resilience concept lacks tools that would allow solving numerous problems related to assessing and controlling (managing) resilience as a quantity. Below such an apparatus is described. A generalized quantitative definition of resilience and preparedness is given by taking into account that most of the multiple parameters on which resilience/preparedness is dependent, are random variables (RV), random functions or random fields (RF).

Therefore, resilience is also a RV or a RF, and also is an explicit function of time. Hence, it is possible to quantitatively define resilience (as a rough first approximation) as follows [11]:

$$Rsl(t) = P\begin{pmatrix} N_t < N_*(0),\ E < E_*; \\ \Delta RDP \le \Delta RDP_*; \\ \Delta t_r < \Delta t_*,\ C \le C_*,\ 0 < \tau < t \end{pmatrix} \quad (1)$$

where $P\left(N_t < N_*(0)\right)$ is the probability that the number of injuries/lethalities during the incident or catastrophe (and after, while mitigating its consequences) will not exceed a specific number during the time t; $P\left(E < E_*\right)$ is the probability that the volume/monetary value of the environmental damage during mitigating the catastrophe will not exceed a specific value during the time t; $P\left(\Delta RDP \le \Delta RDP_*\right)$ is the probability that the decrease of the regional domestic product will not be larger than a specific value during the time t; $P\left(\Delta t_r < \Delta t_*\right)$ is the probability that the acceptable recovery envelope time will not exceed a specific time; $P\left(C \le C_*\right)$ is the probability that the cost of recovery of the region will not exceed the forecasted value.

To this definition some quantitative measures should be added that refer the incident mitigation to the human factor (applied to the whole population involved in the crises): amount of suffering – total hours of being out of the comfort zone; number of mild, medium and serious illnesses; numbers of injuries, limb losses and lethalities (classified by age, gender, profession).

Now the strategic preparedness would be defined as a complex characteristic of a city, which resilience parameters [see formula (1)] are not less than some benchmark values. The latter could be obtained through solving corresponding optimization problems or real life statistics.

I. The architecture of the urban resilience system

The architecture of the urban resilience system (URS) mimics the long time existing on the market different monitoring and maintenance optimization systems designed to optimize performance of critical industrial infrastructures [2–4]. The difference is in that the urban infrastructure, in its entirety, is a very specific complex system of interdependent systems (SoIS) and is widely spread and some of its part *continuously moves* over the whole territory of a municipality. The URS is designed to provide, in the first place, raw and processed data about how this SoIS functions, and to some extent, but much less, about how it degrades in time [9, 10].

A typical URS consist of following elements (which can be purchased at affordable prices that have lowered more than 80 % in the last 10 years, and continue to drop): *sensors, geolocation subsystem*, *information subsystem*, *security subsystem*, and the *situation room SR*, which serves as the ultimate place where the decision makers *formulate, simulate* and *calibrate* their actions in response to different incidents, emergency situations, and catastrophes.

Sensors for an URS measure, in real-time, following parameters of the urban SoIS: traffic over different routes; occupation of public spaces; track public transport vehicles position in time; measure the levels of utilization of urban services (to optimize operations); health parameters (to protect the population); understand and track deficits and remaining works on the sheets. Sensors can be embedded into not moving and mobile objects, smart things, which also are a part of an overall URS architecture.

The *geolocation subsystem* consists of: GPS units attached to all moving components of the CI, information panels in public spaces; information (Data Bases) concentrated in social networks; capabilities to exchange information through transportation systems.

The *information subsystem* is comprised of: transversal overview in real time; simulation scenarios and decision outcomes; exchange of information between all the stakeholders. It is tethered to the *Internet of things* (IoT). The latter is a fast emerging industry and a wide meaning concept. Basically IoT refers to a remotely controlled «thing» connected to the Internet. In general, the word «*thing*» in an urban environment means *consumer things* [wearables, different home and garden appliances (heaters, air conditioners, water sprinklers, etc.)]. *Enterprise Internet of Things* (EIoT) connects things within the Business-to-Business environment (i.e., diagnostics, control, monitoring and maintenance system for a plant or a smart city district). A viable IoT solution contains *six distinct layers:* 1) *Devices* (heart of the IoT, the hardware able to gather the relevant to UR data); 2) *Enablers* (hardware that offers intelligence to *dumb* devices; 3) *Connectivity* (means that enable sending the data collected from the device to the server); 4) *Middleware/General Purpose Platforms* **(**transform the *machine language* into the *application language*, store the data, manage the devices (read data, send control commands to the device, authorize, authenticate, audit and account), as well as any alarms, events and notifications; 5) *Intelligent Services Suppliers* companies that offer services for the hosting applications (*Public Cloud*, *Private Cloud* or *Space* for your own server, take care of *Big Data Analysis*, creates patterns for predictive maintenance, data compression systems, etc.); 6) *System Integrator* companies have a key role in this industry as they design solutions, install and connect things, adapt software, unify systems, and interconnect software from different vendors (new and pre-existing software) In other words, they pull all the above together to make it work. *Telecom Operators* own the Internet connectivity, allowing to connect the things to the server where the application will be installed. The main parts of this URS architecture are shown in Fig. 1 [9].

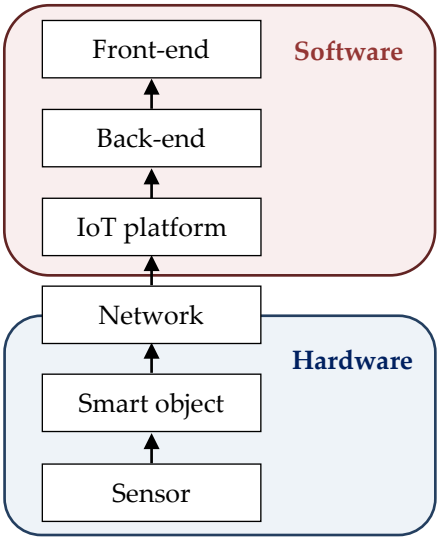


Figure 1: *The main components of the urban resilience subsystem*

V. Critical infrastructure models

The *software of the urban resilience subsystem* consists of a set of models that can be divided into three groups: 1) conceptual (verbal) models, which are results of corresponding identification processes combined with discretization procedures and accuracy assessments; they are used as input data for simulation; 2) calibrated models that are considered as a valid representation of the *natural system.* They have to account for the second type of uncertainty, which arises from the improper extension of calibrating conditions to prediction, i.e. not taking into account changes in the natural system or the scale of the problem); 3) mathematical (quantitative) analog, analytical, and numerical models. Approximate models are often used as decision support tools, as they are able to evaluate the relevant advantages and disadvantages of each alternative and rank the options. In practice, this is usually all that is needed for decision-making.

One of the most relevant problems in risk analysis of complex systems is the construction of adequate models of critical infrastructures (CIs). These models should lead to simple and effective quantitative methods of risk analysis and management of urban systems of CIs. Complex systems are characterized by that: 1) The interaction of its subsystems (elements) is hard to present explicitly; 2) The input data which describes functioning of the various elements of CIs is heterogeneous; 3) The traditional models, based on series-parallel connection of CI elements not always give an adequate description of CIs; 4) Implementation of the logic-probability models and graph theory models demands great efforts and presence of substantial prior information on the subject of research, which is not always available. Most effective in many cases are descriptions of CI as transportation, supply and Bayesian networks.

Consider an important and complex interdisciplinary integration problem – managing safety of a critical infrastructure (CI) of a large municipal area or a region, embedded into the context and realities of the modern society of risk. When analyzing risk it is necessary to answer three mutually dependent questions: Is the considered system of CIs safe enough? What is the size of risk? At what level of expenditures and efforts it is worthwhile to spend to save lives? There are two major approaches to this daunting task.

The first one could be characterized as the *from top to bottom – FTTB* or *from top-down* models. This approach largely ignores the small and medium scale events and zeroes in on state or regional scale consequences [e,g., the Leontieff Input-Output (I-O) model]. The second approach could be called the *from-bottom-up – FBU* models. It builds the risk model as a *quilt*, using as its *bricks* corresponding results of solutions of problems that relate to elements of critical infrastructures, and, finally, systems of CIs. It also is comprehensive, repeatable, rigorous, sophisticated, and based on real life statistics, permits true multidisciplinary block-module approach, when the output of the i-th problem is the input for the (i+1)-th problem. It also provides a natural way of assessing the domino-type disaster scenarios. This approach (*as related to industrial and natural disasters*) is institutionalized by a series of Federal laws and EMERCOM regulations and ruling documents. According to these rulings *every entity* that is considered a potentially dangerous object (PDO) is obliged to provide a *declaration* and a *passport of its safety*, and a *risk map*, which depicts the individual risk in the territory of the site and its surroundings. These documents contain: quantitative description of the operational risk of this entity; description of the needed mitigating means (machines, transportation, materials, workforce, and financial means) for the *worst case scenario* and the *average scenario*. The risk (failure) analysis is conducted using a set of state approved recommended practices (RP). These RP's are based on solutions of relevant problems of fracture mechanics, blast, fire, spill, filtration, water and air pollution, and descriptions of their consequences in typical scenario settings; provide some guidelines as to how to assess the number of fatalities and the monetary value of lost life or limb; prescribe how to assess the damage inflicted by a catastrophe and to present the collective risk specific for the PDO in consideration.

The *FBU* approach to command and control operations in emergency situations does not as yet fully address the problem of evaluation of the emergency resource management for regional resilience and strategic preparedness. Currently, there is no coherent, collaborative system in place to evaluate different risk scenarios, and to enhance communication to benefit from the unique and specific experience of each PDO as an organization.

VI. Selecting risk mitigating control (management) means

The last problem that crowns the full solution of urban risk management is designing and implementing risk mitigation control means. There are two approaches to solve this problem. The direct problem is posed as follows: With *given means* for improving CIs safety S_{giv} choose such a set of measures that *maximizes reduction* of incident probability $Q_i = P / A$. The inverse problem is formulated in following terms: With *minimal expenditures* EX choose such a set of measures, implementation of which lowers the incident probability $Q_i = P / A$ down to an *acceptable* (preassigned) level $P_{acc}(A)$. The above methodology was successfully implemented in creating risk maps for large municipalities and its satellites, power grids of several regions, pipeline systems, and other types of PDOs [12].

VII. Intrinsic specifics of critical infrastructures

CIs are large distributed renewable geotechnical man – machine – environment systems with: (1) non-economical responsibility (human factor, environment); (2) functional, structural and time-wise redundancy; (3) geometrical, physical, statistical and economical non-linearity.

Modern CIs have following indispensible components: risk-based diagnostic subsystems; monitoring and/or control subsystem(s), et al; risk-based integrity maintenance subsystems; assets safety and security/defense subsystems, and other. The total risk of operating CIs is carried by its full group of scenarios (100 %). All these specifics should be consistently accounted for during the design, operation and risk assessment of urban PDOs and CIs.

Urban ICI networks can be considered as *conduits* and at the same time as *intermediaries* between the natural environment and the resource demands of the urban society [13]. ICI is also the principal source of technological hazards of the city. A point failure anywhere in the ICI can rapidly propagate through the city with broad impacts on the citizens and the environment. Hence, it stands for reason that *management of urban risk may be boiled down to management of risk for the whole urban system of ICIs*.

The problem of urban resilience management consists of following two parts: Assessing the full possible damage and all of its components; Designing means and methods for reduction the potential consequences of an initial failure in the system of ICIs. This problem can be solved *only through interdisciplinary approach*, and by convoluting the heterogeneous parameters, which define the operation of the CI, into few integral parameters, which should be simple to understand and use.

The main conceptual problem of assessing, monitoring, and managing resilience/risk of ICIs is defined by following three factors: the dimension of the problem is huge (could be tens of thousands of interdependent parameters); the problem is multi-disciplinary, and the parameters involved when solving the problem are from different sciences and branches of engineering, and currently are, as a rule, hard, if not impossible to convolute; the ICI risk cannot be adequately described without explicitly accounting for the human factor. Hence, before attempting to solve the problem in consideration, it is necessary to introduce some *unified measures of safety/risk*, which account for the human factor in socially meaningful terms.

VIII. Proposed quantitative unified criteria for resilience/risk management

Authors proposes [14] following five generalized criteria: ICI Resilience; Regional Average Life Expectancy (RALE); Regional Life Quality Index (RLQI); ICI Carbon Footprint: ICI Entropy. Important comment: Public safety and security is an important objective, but diminishing of risk requires additional expenditures. The share of resources that is being devoted by society for achieving safety must be continuously evaluated, having in mind other needs of society, such as clean air and water, healthy food, housing, health care, social security benefits, pensions, education, etc., which also improve the longevity and quality of life.

Therefore, the central problem of regional (ICIs) risk management becomes optimization of the distribution of the always limited resources to improve the overall safety of systems of ICIs, and via this, the urban safety. This paper describes principles and methodology which lead to achieving this goal.

I. Probabilistic definition of urban resilience and strategic preparedness

It is best to first visualize the Resilience Factor, see Fig. 2, where *RDP* is the regional (urban) domestic product; $N(L,l)$ is the number of casualties (L), injuries (l); $(E+A)$ is the environmental and property losses; $U(\tau_0)$ is the vector of full losses. The corresponding problems are solved using appropriate probabilistic methods.

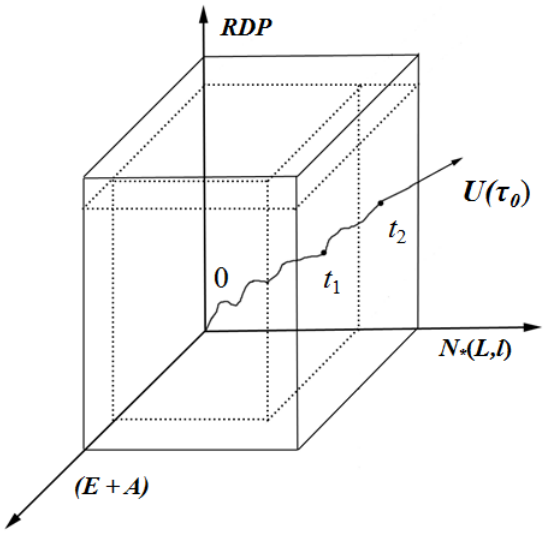


Figure 2: *Full and partial resilience factors*

Actually, Resilience Factor RF is an n-dimensional vector. In order to visualize each component of this vector it is recommended to deploy the vector, portraying each component of the RF vector as a two-dimensional function of time (Fig. 3). In this figure it can be seen that after the disaster the CI output O is decreased and it takes some time to restore O to predisaster values. The same pattern is observed for the RDP of the damaged CI. The losses S of limb and health of the citizens can be compensated, but it takes more time than in the previous case. Finally, the loss of life is permanent, as it will last forever. In reality, there are more components of the resilient factor RF.

Further analysis involves considering different scenarios of development of the restoration phase of the damaged CI, each of which has its own probability of developing (Fig. 4). Assessment of these probabilities is a very important, but separate part of the analysis and is derived via computer simulation.

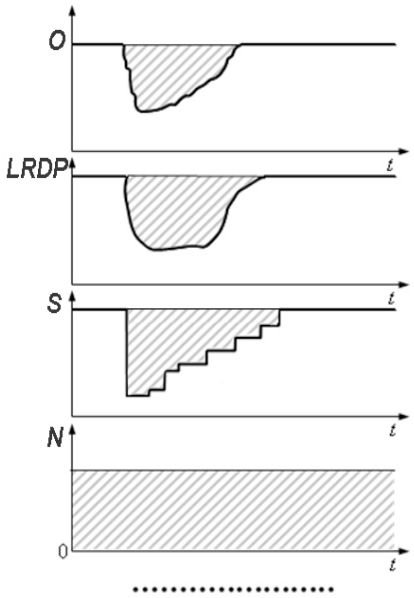


Figure 3: *n-dimensional collapse of infrastructure operating quality due to a disaster or a catastrophe: O – Output; LRDP – Lost growth RDP; N – non recoverable losses; S – recoverable losses; the dashed area relates to the n-dimensional volume/area of the quality collapse*

The FAQs when solving this problem are: What are the quantities of supplies required to meet the contingent demand of sheltered citizens? How far in advance must these supplies be ordered and pre-staged? What *level of uncertainty* in disaster forecasts exists at the critical decision points? What are the consequences of delaying such response decisions as related to supply? These questions have only probabilistic answers.

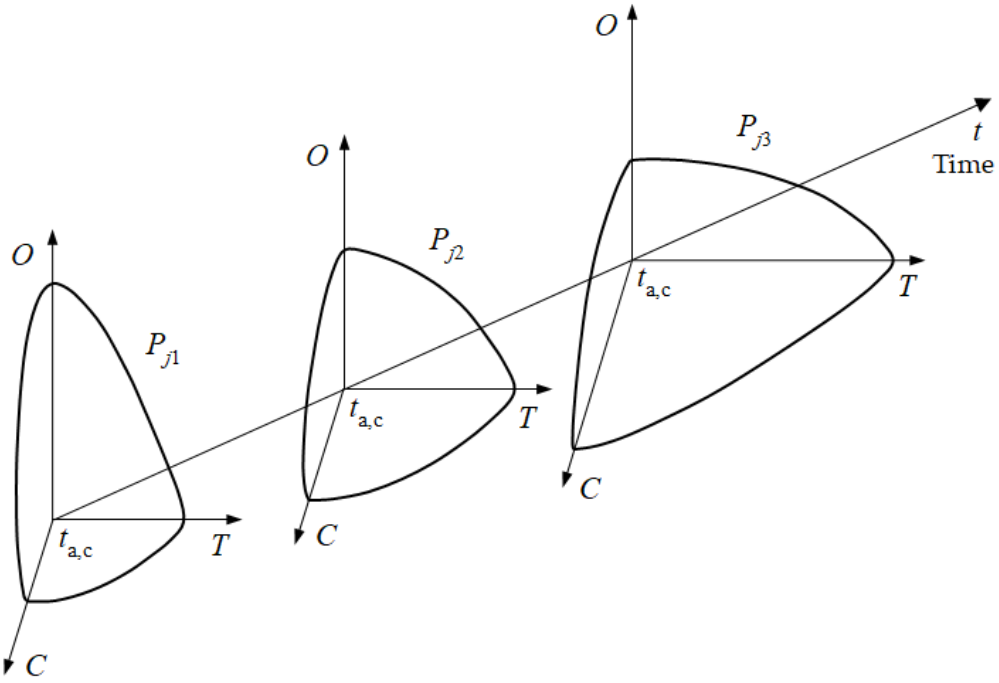


Figure 4: *The change of the partial infrastructure resilience $Res_j(t)$ in time for different probabilities of an incident/disaster/catastrophe and size of losses.*
$t_{a,c}$ – time of the disaster; C – cost of recovery; T – duration of recovery; O – volume of lost production/services

IX. Urban average life expectancy (UALE) as generalized criteria for optimizing city resilience

One may ask, why UALE? There are many reasons why UALE could be effectively used in urban risk analysis and management. The most valuable asset of any society is its people. The most valuable trait of a human being is her/his life. The most valuable parameter of a human life is its longevity in good health. UALE at birth is a non additive (non linear) parameter which permits combining parameters of complex safety of elements, structures and ICI systems with economic parameters of the operation and social aspects of sustainable development of the region. UALE provides seamless tying up of separate specific problems of safety/reliability of CIs and their elements with the generalized problem of regional risk management. UALE is a convenient characteristic for assessing the quality of life, because it continues to make sense with the size of the society in consideration shrinking. It is possible to calculate ALE for the nation/country as a whole, as well as for a separate region, industry, ICI, PDO and even for an individual. UALE has a biological «ceiling» (currently, just below 125 years) and some properties of a fractal, is a solution of a system of differential equations and has the form of a logistic curve which is a function of time. It depends on the current value of UALE and on how optimal the society distributes year by year the regional DP on accumulation of wealth, consumption, and on safety of the system of the ICIs, its employees, and the population adjacent to the same ICIs from the possible influence of incidents of different nature (Nature, technological, premeditated). In the last component it is necessary to single out those means (shares of UDP and of the PDOs budget), which could be (are) allotted to mitigate disasters and catastrophes of the ICIs' components and, accordingly, define, what would be the decrease/increase of the number of fatalities/injuries in the region in consideration due to natural/technogenic incidents, and assess how quantitatively this will influence the UALE.

The necessity of balancing the benefit from increasing safety (i.e., in the context of the problem in consideration, increasing UALE), and the cost of decreasing risk is an imperative of the XXI century and professional obligation of decision makers who are responsible for the safety of people. The ability of any society to prevent premature death/injury of its people is finite and restricted by its capability to create societal wealth. Hence, the *central problem of management of any risk (including technological risk) becomes optimization of the distribution (by volume and place of application) of the always limited resources to mitigate risk using the UALE criterion.*

X. The stakeholders of urban resilience

The stakeholders of urban resilience are four major different groups of organizations: 1) multilateral bodies (key to the urban resilience market, as they provide financing for UR improvement projects in emerging and developed cities, management and its benefits to citizens, economic operators, and decision-makers, support of policies for improving resilience; 2) research centers, which play a fundamental and vital role in this regard; 3) businesses that catalyze UR development; 4) governments of all caliber cities who are the end users of the UR product.

The hotspot in implementing the resilience concept in every day management of cities is in creating and using specially tailored software. Several such packages already exist and are being offered in the market [9, 15, 16]. For instance the HAZUR® software package [9] has been used to optimize management of several large and small cities in Catalonia (Spain). Besides improving public image, it helps creating an effective resilience subsystem of the city (installation of sensors for diagnostics and monitoring), for studying, analyzing and managing the city resilience, and *optimizing* the compatibility, coordination, operation of different services (agents) based on gathered data, allows to run simulations to illustrate how the city will react in the case of an impact

[17, 18].

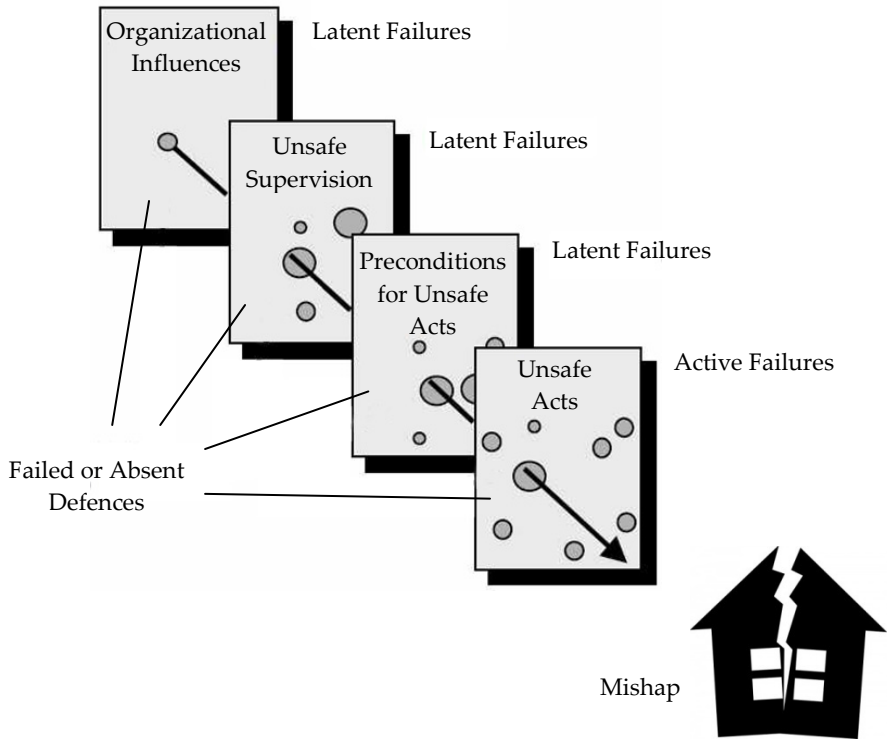


Figure 5: *The «Swiss Cheese» model of human error causation [19] adapted for the HFACS taxonomy by D. Wiegmann and S. Shappell [20]*

XI. Conclusion

1) Initial results of an interdisciplinary project on developing a methodology of urban risk management via risk governance of ICIs systems are presented.

2) The proposed methodology may serve as a useful tool for managing risk of PDOs, critical infrastructures and their systems according to the RALE criterion.

3) Results of the research may be useful to the municipal level decision makers (DMs), who make decisions related to optimal distribution of their budgets, taking into account sustainable growth of entities under their jurisdiction. They will also be able to monitor how their decisions influence the quality of life/level of happiness of their constituents as related to the decisions they make in the disaster and ordinary times.

4) In order to implement the resilience methodology to create a smart sustainable city it is necessary to build up for it an urban resilience subsystem URS, its architecture outlined in this paper, and create in its frame work a Resilience Office that is the core of dealing with urban crises and systemic stress. This URS would identify the weakest spots in the urban System of Systems and react faster and more efficiently during and after an impact or crisis.

References

[1] Timashev S.A. and Yablonskikh I.L. (1996). Expert System for Assessing Main Pipeline Reliability and Residual Lifetime. *Proc. of the 7-th Specialty Conf. ASCE «Probabilistic Mechanics and Structural Reliability»*, Worcester, Massachusetts, USA, 322–329.

[2] Timashev S.A., Hoperskiy G.G. and Chepurskiy V.N. (1997). Reliability and residual life monitoring system for oil pumping station equipment. *Pipeline transport of oil*, 8:5–9, 9:8–13.

[3] Galagan P.I., Semihatov N.A., Poluyan L.V. and Timashev S.A. (1998). Stationary system

of vibration protection and vibrodiagnostics of oil pumping units. *Diagnosis and control*, 6.

[4] Timashev S.A. (2000). Optimal machinery integrity and maintenance control. *COMADEM Magazine*.

[5] Makarov S.O. (1894). The analysis of elements that make up the combat power of ships. *Marine Collection*, Petrograd, 6:1–106.

[6] Claudel P. (1936). The American Elasticity, Wiley & Sons.

[7] Crawford Stanley Holling. (1965). Resilience and stability of ecological systems, NY.

[8] ARUP. City Resilience Framework. City Resilience Index. (2014). The Rockefeller Foundation.

[9] HAZUR Software. (2017).

[10] Opticity LLC. (2016).

[11] Timashev S.A. (2016). Infrastructure resilience: definition, calculation, application. *WEEF, IGIP/ICL Conf.*, Pisa, Italy.

[12] Guryev E.S., Poluyan L.V., and Timashev S.A. (2014). Construction of dynamic risk maps for large metropolitan areas. *J. of Risk Analysis and Crisis Response*, Paris, France : Atlantis Press, 4:272–276.

[13] Timashev S.A. (2014). Average life expectancy as a criterion for regional risk management. *J. of Risk Analysis and Crisis Response*, Paris, France : Atlantis Press, 4-1:10–19.

[14] Timashev S.A. (2013). Unified quantitative criteria for managing regional risk. *Proc. of the 11th Intern. Conf. on Structural Safety & Reliability, ICOSSAR,* Columbia University, NY.

[15] Constructing a Resilience Index for the Enhanced Critical Infrastructure Protection Program. (2010). Argonne National Lab, ANL/DIS-19-9.

[16] National Institute of Building Sciences. The Brashier Group LLC. Two Open Source Solutions for Advancing Resilience. (2012). Washington, USA.

[17] Timashev S.A. (2017). Urban infrastructures resilience – key element for creating smart sustainable cities. *Proc. of the Intern. Conf. «Critical Infrastructures and Sustainable Cities»*, Yekaterinburg, Russia, 474–487.

[18] Timashev S.A. (2016). Infranetics – the new convergent science of the 21st century. *Proc. of the II Intern. Conf «Problems of safety of civil engineering critical infrastructures»*, Yekaterinburg, Russia, 234–237.

[19] Reason J. (1990). Human Error, Cambridge, Cambridge University Press.

[20] Wiegmann D. and Shappell S. (2003). A human error approach to aviation accident analysis, London, Routledge.

RT&A, No 4(80)
Volume 19, December, 2024

MARSHALL-OLKIN EXPONENTIATED NADARAJAH HAGHIGHI DISTRIBUTION AND ITS APPLICATIONS

Nicy Sebastian[1*], Jeena Joseph[2], Muhsina C. S.[3] and Sandra I. S.[4]

•

[1,2,3,4] Department of Statistics, St Thomas College, Thrissur
affiliated to the University of Calicut, Kerala, India- 680 001
[1*]nicycms@gmail.com,[2]sony.jeena@gmail.com,
[3]muhsinacs2000@gmail.com , [4]sandrais490@gmail.com
[*]Corresponding Author

Abstract

In this paper, we introduce a new generalization of exponentiated Nadarajah Haghighi distribution, namely Marshall-Olkin exponentiated Nadarajah Haghighi (MOENH) distribution and study its properties. The stress-strength parameter estimation is also taken into account. Characterizations of the new distribution are obtained. The unknown parameters of the distribution are estimated using the maximum likelihood method. It is established how important this distribution is to the research of the minification process. Simulation studies are done, and sample path properties are explored. A real data set is fitted to the new distribution to demonstrate the model's adaptability and effectiveness.

Keywords: Marshall-Olkin family, Exponentiated Nadarajah Haghighi distribution, Stress-Strength reliability, Maximum likelihood, Minification process.

1. Introduction

Marshall and Olkin introduced a new family of distributions in 1997 with one additional parameter α, known as the Marshall-Olkin Family of Distributions. The distribution function of the family is given by,

$$G(x) = \frac{F(x)}{\alpha + (1-\alpha)F(x)}, -\infty < x < \infty, \alpha > 0 \tag{1}$$

Here, $F(x)$ represents the distribution function of a random variable X. The Marshall-Olkin family of distributions is a class of continuous probability distributions that are used to model lifetime data. It has the advantage of being flexible and able to fit a wide range of data sets. Cordeiro and Lemonte [1] discussed some mathematical properties of the Marshall-Olkin extended Weibull distribution and estimation of the model parameters by the maximum likelihood method. Ristić and Kundu [14] introduced a third shape parameter to the two-parameter generalized exponential distribution, adopted from the Marshall-Olkin method so that the hazard function of the proposed model can have all the four major shapes, namely increasing, decreasing, bathtub or inverted bathtub types. Recent developments in the Marshall-Olkin family of distributions can be given in George and Thobias [4], Gillariose and Tomy [5], etc.

The exponential distribution is a versatile and widely used probability distribution that has many important applications in various fields, particularly in modeling time-related events. The generalized exponential distribution, sometimes referred to as the exponentiated exponential (EE) distribution, failure rate function that might be either increasing, decreasing, or constant,

see Gupta, et al. [8]. According to Gupta and Kundu [9], the failure rate function of the EE distribution has similar behaviour to that of the gamma distribution and in many cases, it can be utilized as an alternate distribution to the gamma and Weibull distributions. Nadarajah and Haghighi [13] proposed a new generalization of the exponential distribution as an alternative to the gamma, Weibull, and EE distributions. Lemonte [11] defined a new three-parameter exponential-type distribution family, exponentiated Nadarajah Haghighi, that can be used to model survival data and reliability issues. A three-parameter distribution called Exponentiated Nadarajah and Haghighi (ENH) with cdf is given in (2)

$$F(x) = \begin{cases} (1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta} & ; x > 0, \gamma, \beta, \lambda > 0 \\ 0 & ; \text{Otherwise,} \end{cases} \tag{2}$$

where the parameters γ, β control the shape of the distribution and parameter λ is the scale parameter. The major goal of the present study is to create a new model of the four-parameter distribution, in the expectation that, in some instances, the new distribution will "fit better" than the exponential, ENH, Nadarajah-Haghighi (NH), and Marshall-Olkin-Nadarajah-Haghighi Distribution (MONH) distributions.

The rest of the paper is organized as follows: In Section 2, we propose a new generalization of ENH distribution, namely MOENH distribution. Various structural properties of the MOENH distribution, such as moments, quantile function, order statistics and stress-strength reliability are studied in Section 3. Characterizations of MOENH distribution are obtained in Section 4. In Section 5, we study the estimation of parameters of the MOENH distribution using the method of Maximum Likelihood. In Section 6, we discuss the MOENH Minification process and the corresponding sample paths. In Section 7, we have fitted the model to real-life data to show the flexibility of the new distribution. Concluding remarks are presented in Section 8.

2. MOENH distribution

In this section, we discuss the MOENH distribution introduced by using (1) and (2) we have the distribution function of MOENH distribution as follows:

$$G(x) = \begin{cases} \frac{\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}}{\alpha+(1-\alpha)\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}} & ; x > 0, \alpha, \gamma, \beta, \lambda > 0 \\ 0 & ; \text{ Otherwise.} \end{cases} \tag{3}$$

where the parameters β, γ are shapes of the distribution, and α, λ are the scale parameter. If $\beta = 1$, the $MOENH$ distribution reduces to $MONH$ distribution. We have the NH distribution when $\alpha = 1$, $\beta = 1$. For $\alpha = 1$, $\beta = 1$, $\gamma = 1$, we obtain the exponential distribution. The probability density function of MOENH distribution is,

$$g(x) = \begin{cases} \frac{\alpha\beta\gamma\lambda(1+\lambda x)^{\gamma-1}\left(e^{1-(1+\lambda x)^{\gamma}}\right)}{\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{1-\beta}\left(\alpha+(1-\alpha)\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right)^{2}} & ; x > 0, \alpha, \gamma, \beta, \lambda > 0 \\ 0 & ; \text{ Otherwise.} \end{cases} \tag{4}$$

The survival function and failure rate function of MOENH distribution are respectively as,

$$S(x) = 1 - \frac{\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}}{\alpha + (1-\alpha)\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}} = \frac{\alpha - \alpha(1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta}}{\alpha + (1-\alpha)(1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta}}, \tag{5}$$

$$h(x) = \frac{\alpha\beta\gamma\lambda(1+\lambda x)^{\gamma-1}e^{1-(1+\lambda x)^{\gamma}}}{\left[\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)\left(\alpha + (1-\alpha)\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right)\right]} \times \frac{1}{\left[\left(\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{-\beta}\left(\alpha + (1-\alpha)\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right)\right) - 1\right]}.$$

It can demonstrate that the distribution exhibits increasing, decreasing, bathtub-shaped, inverse bathtub-shaped, and constant hazard functions. In Figure 1, we can see the plots of pdf and hazard function of MOENH distribution for different values of the parameters.

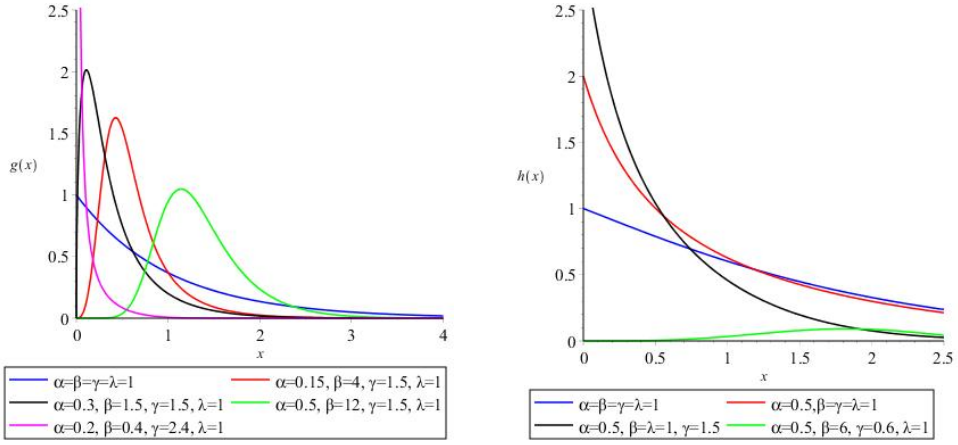


Figure 1: *Plots of pdf(left) and hazard function (right) for different values of parameters.*

3. Statistical properties

In this section, some statistical properties of MOENH distribution are discussed.

3.1. Moment Generating Function

The moment generating function of MOENH distribution is given by

$$M_X(t) = \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{h+k+j}\frac{(t)^h}{h!}\binom{h}{k}\binom{k/r}{m}\frac{(n+1)\beta}{\lambda^h\alpha j!} \times\left(\frac{\alpha-1}{\alpha}\right)^n\frac{\Gamma(\beta n+\beta)\Gamma(m+1)}{\Gamma(\beta n+\beta-j)(j+1)^{m+1}}.$$

3.2. Quantile Function

The quantile function has a number of applications. It can be used to obtain median, skewness, and kurtosis and can also be used to generate random variables. The quantile function of MOENH distribution is obtained as,

$$X=\frac{\left(1-\ln\left(1-e^{\frac{1}{\beta}\ln\left(\frac{p\alpha}{1-p+p\alpha}\right)}\right)\right)^{\frac{1}{\gamma}}-1}{\lambda}, 0<p<1, \alpha,\gamma,\beta,\lambda>0. \tag{6}$$

3.3. Order Statistic

Order statistic makes their appearance in many areas of statistical theory. Let $X_1, X_2, \ldots, X_n$ be a random sample from the MOENH family of distributions, and let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the corresponding order statistic. The pdf of i^{th} order statistic, say $X_{i:n}$, can be written as

$$\begin{aligned} g_{i:m}(x) &= \frac{m!g(x)}{(i-1)!\,(m-i)!}\left[G(x)\right]^{i-1}\left[1-G(x)\right]^{m-i} \\ &= \frac{m!g(x)}{(i-1)!\,(m-i)!}\sum_{j=0}^{m-i}(-1)^j\binom{m-i}{j}\left[G(x)\right]^{j+i-1} \end{aligned} \tag{7}$$

by using binomial expansion $[1-G(x)]^{m-i} = \sum_{j=0}^{m-i}(-1)^j \binom{m-i}{j} \sum_{j=0}^{m-i}(-1)^j \binom{m-i}{j} [G(x)]^j$ The pdf of the 1^{st} and n^{th} ordered statistic will be,

$$g_{1:m}(x) = \frac{\sum_{j=0}^{m-1}(-1)^j \binom{m-1}{j} m!}{(m-1)!} \times \frac{\alpha\beta\gamma\lambda(1+\lambda x)^{\gamma-1}e^{1-(1+\lambda x)^\gamma}}{\left(1-e^{1-(1+\lambda x)^\gamma}\right)^{1-\beta}\left(\alpha+(1-\alpha)\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right)^2} \times \frac{\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta}{\alpha+(1-\alpha)\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta},$$

$$g_{n:m}(x) = \frac{\sum_{j=0}^{m-n}(-1)^j \binom{m-n}{j} m!}{(m-n)!(n-1)!} \times \frac{\alpha\beta\gamma\lambda(1+\lambda x)^{\gamma-1}e^{1-(1+\lambda x)^\gamma}}{\left(1-e^{1-(1+\lambda x)^\gamma}\right)^{1-\beta}\left(\alpha+(1-\alpha)\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right)^2} \times \frac{\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta}{\alpha+(1-\alpha)\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta}.$$

3.4. Stress-strength reliability

In order to estimate the stress-strength parameter, considering two random variables X and Y with $MOENH(\alpha_1, \beta, \gamma, \lambda)$ and $MOENH(\alpha_2, \beta, \gamma, \lambda)$ distributions, respectively, with the same baseline parameters β, γ, λ. We assume that X and Y are independent random variables. Then the stress-strength parameter is obtained in the form

$$\begin{aligned} R &= P(Y<X) = \int_0^\infty \left[\int_0^x g_Y(y)\right] g_X(x)dx \\ &= \int_0^\infty G_Y(x)g_X(x)dx = -\alpha_1\left[\frac{\alpha_2\left(ln\left(\alpha_1\right) - ln(\alpha_2)+1\right)+\alpha_1}{(\alpha_1-\alpha_2)^2}\right]. \end{aligned} \tag{8}$$

4. Characterization

This section deals with the characterization of the MOENH distribution based on the ratio of two truncated moments. To present the characterization of the distribution, consider the theorem presented in Glänzel [7].

Theorem 1. Let (Ω, G, P) be a given probability space and let $H = [a, b]$ be an interval for some $a < b$ ($a = -\infty, b = \infty$ might be allowed). Let $X : \Omega \to H$ be a continuous random variable with the distribution function G and let q_1 and q_2 be two real functions defined on H that

$$E[q_2(x)|X \geq x] = E[q_1(x)|X \geq x]\xi(x), x \in H$$

are defined with some real function ξ. Assume that q_1 and $q_2 \in C^1(H)$, $\xi \in C^2(H)$ and G is a twice continuously differentiable and strictly monotone function on the set H. Finally, assume that the equation $\xi q_1 = q_2$ has no real solution in the interior of H. Then G is uniquely determined by

the function q_1, q_2 and ξ, particularly

$$G(x) = \int_a^x C \left[\frac{\xi^{'}(u)}{\xi(u)q_1(u) - q_2(u)} \right] exp\left(-S(u)\right)du,$$

where the function is a solution of the differential equation $S^{'} = \frac{\xi^{'} q_1}{\xi q_1 - q_2}$ and C is the normalization constant, such that $\int_H dG = 1$ and $\xi(x) = \frac{E[q_2(x)|X \geq x]}{E[q_1(x)|X \geq x]}$.

Proposition 1. Let $X : \Omega \to (0, \infty)$ be a continuous random variable and

$$q_1(x) = (\alpha + (1-\alpha)(1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta})^2(1 - e^{1-(1+\lambda x)^{\gamma}})^{1-\beta}$$

and

$$q_2 = q_1(x)e^{1-(1+\lambda x)^{\gamma}}, x > 0.$$

Then the random variable X has pdf (4) if and only if the function ξ defined in the Theorem 1 is of the form

$$\xi(x) = \frac{e^{1-(1+\lambda x)^{\gamma}}}{2}. \tag{9}$$

Proof. Suppose the random variable X has pdf (4), then

$$(1 - G(x))E\left[q_1(X) \mid X \geq x\right] = -Ce^{1-(1+\lambda x)^{\gamma}}$$
$$(1 - G(x))E\left[q_2(X) \mid X \geq x\right] = -\frac{C}{2}\mathrm{e}^{2(1-(1+\lambda x)^{\gamma})}$$

where, $C = \alpha\beta$. Further,

$$\xi(x)q_1(x) - q_2(x) = -\frac{1}{2}q_1(x)\mathrm{e}^{1-(1+\lambda \mathrm{x})^{\gamma}} \neq 0, x > 0.$$

Conversely, if ξ is of the above form, then

$$S^{'}(x) \quad = \quad \frac{\xi^{'}(x)q_1(x)}{\xi(x)q_1(x) - q_2(x)} = \gamma\lambda(1+\lambda x)^{\gamma-1}$$

and hence,

$$S(x) \quad = \quad (1+\lambda x)^{\gamma}.$$

Now, in view of Theorem 1, X has density (4). ■

Corollary 1. Let $X : \Omega \to (0, \infty)$ be a continuous random variable and $q_1(x)$ be as in Proposition 1. The pdf of X in (4) if and only if there exist functions q_2 and ξ defined in Theorem 1 satisfying the differential equation

$$\frac{\xi^{'}(x)q_1(x)}{\xi(x)q_1(x) - q_2(x)} \quad = \quad \gamma\lambda(1+\lambda x)^{\gamma-1}, x > 0.$$

The general solution of the differential equation given in Corollary 1 is,

$$\xi(x) = \mathrm{e}^{(1+\lambda \mathrm{x})^{\gamma}}\left[-\int \gamma\lambda(1+\lambda \mathrm{x})^{\gamma-1}\mathrm{e}^{-(1+\lambda \mathrm{x})^{\gamma}}\left(\mathrm{q}_1(\mathrm{x})\right)^{-1}\mathrm{q}_2(\mathrm{x})\mathrm{dx} + \mathrm{D}\right],$$

where D is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 1 with $D = 0$. However, it should be noted that there are other triplets (q_1, q_2, ξ) satisfying the conditions of Theorem 1.

Proposition 2. The MOENH density function is log-convex for fixed α say, $\alpha = 1$ and if $\gamma < 1$ and $\beta < 1$, and it is log-concave if $\gamma > 1$ and $\beta > 1$.

Proof. Let $z = (1+\lambda x)^{\gamma}$, which implies that $z > 1$ for $x > 0$. We have $x = \left(z^{1/\gamma} - 1\right)/\lambda$. The MOENH density is now rewritten as a function of $z, \xi(z)$ say, we obtain

$$\xi(z) = f\left(\left(z^{1/\gamma} - 1\right)/\lambda\right) = \gamma\lambda\beta \frac{z^{(\gamma-1)/\gamma} e^{(1-z)}}{[1 - e^{(1-z)}]^{1-\beta}}, \quad z > 1.$$

The result follows by noting that the second derivative of $\log[\xi(z)]$ is

$$\frac{d^2 \log[\xi(z)]}{dz^2} = -\left[\frac{(\gamma-1)}{\gamma z^2} + \frac{(\beta-1)e^{1-z}}{[1-e^{1-z}]^2}\right].$$

■

Proposition 3. For any $\lambda > 0, \alpha = 1$, the MOENH distribution has an increasing failure rate function if $\gamma > 1$ and $\beta > 1$, and it has a decreasing failure rate function if $\gamma < 1$ and $\beta < 1$. The failure rate function is constant if $\gamma = \beta = 1$.

Proof. Using the log-convexity of the density function, the conclusion is valid. ■

5. Estimation of parameters

There are several methods in the literature for estimating unknown parameters. In this section, maximum likelihood method of estimation is used for estimating the parameters of MOENH distribution. Let us consider $x_1, x_2, \ldots, x_n$ be the random variables having MOENH distribution. Then the likelihood function is given by,

$$L(x_i, \alpha, \beta, \gamma, \lambda) = \prod_{i=1}^{n} \frac{\alpha\beta\gamma\lambda(1+\lambda x)^{\gamma-1} e^{1-(1+\lambda x)^{\gamma}}}{\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{1-\beta} \left(\alpha + (1-\alpha)\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right)^2}$$

The log-likelihood function is given by,

$$\begin{aligned} l =& n\ln(\alpha\beta\gamma\lambda) + \sum_{i=1}^{n} \ln(1+\lambda x)^{\gamma-1} + \sum_{i=1}^{n} 1 - (1+\lambda x)^{\gamma} \\ & - 2\sum_{i=1}^{n} \ln\left(\alpha + (1-\alpha)\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right) - \sum_{i=1}^{n} (1-\beta)\ln\left(1 - e^{1-(1+\lambda x)^{\gamma}}\right) \end{aligned} \tag{10}$$

Partial derivatives of (10) with respect to the unknown parameters $\alpha, \beta, \gamma, \lambda$

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - 2\sum_{i=1}^{n} \frac{1 - (1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta}}{(\alpha + (1-\alpha)(1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta})^2} \tag{11}$$

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} - 2\sum_{i=1}^{n} \frac{\beta(1-\alpha)(1 - e^{1-(1+\lambda x)^{\gamma}})^{(\beta-1)}}{\alpha + (1-\alpha)(1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta}} \tag{12}$$

$$\begin{aligned} \frac{\partial l}{\partial \gamma} &= \frac{n}{\gamma} + \sum_{i=1}^{n} \ln(1+\lambda x) - \sum_{i=1}^{n} (1+\lambda x)^{\gamma} \ln(1+\lambda x) - 2(1-\alpha)\beta \\ &\times \sum_{i=1}^{n} \frac{e^{1-(1+\lambda x)^{\gamma}} (1+\lambda x)^{\gamma} \ln(1+\lambda x)(1 - e^{1-(1+\lambda x)^{\gamma}})^{(\beta-1)}}{\alpha - (\alpha-1)(1 - e^{1-(1+\lambda x)^{\gamma}})^{\beta}} \\ &- \sum_{i=1}^{n} \frac{(1-\beta)\ln(1+\lambda x) e^{1-(1+\lambda x)^{\gamma}}}{e^{1-(1+\lambda x)^{\gamma}}} \end{aligned} \tag{13}$$

$$\begin{aligned}\frac{\partial l}{\partial \lambda} &= \frac{n}{\lambda}+\sum_{i=1}^{n}\frac{(\gamma-1)x}{1+\lambda x}-\sum_{i=1}^{n}\gamma x(1+\lambda x)^{\gamma-1}\\ &-2\sum_{i=1}^{n}\frac{x\gamma\beta(1-\alpha)e^{1-(1+\lambda x)^{\gamma}}(1+\lambda x)^{\gamma-1}(1-e^{1-(1+\lambda x)^{\gamma}\beta-1}}{\alpha-(\alpha-1)(1-e^{1-(1+\lambda x)^{\gamma}})^{\beta}}\\ &-\sum_{i=1}^{n}\frac{\gamma x(1-\beta)(1+\lambda x)^{\gamma-1}e^{1-(1+\lambda x)^{\gamma}}}{1-e^{1-(1+\lambda x)^{\gamma}}}\end{aligned} \tag{14}$$

To obtain the maximum likelihood estimates of the unknown parameters, we equate,

$$\frac{\partial l}{\partial \alpha}=0;\frac{\partial l}{\partial \beta}=0;\frac{\partial l}{\partial \gamma}=0;\frac{\partial l}{\partial \lambda}=0 \tag{15}$$

The ML estimators are found through the solution of the nonlinear system. Hence, using R or MATLAB, a numerical approximation of the software's solution to this system of equations is possible.

6. Autoregressive Time Series Modelling

The autoregressive model is a stochastic process used in statistical modeling in which future values are forecasted based on a weighted sum of past values. The idea behind an autoregressive process is that the values of the past have an impact on the values of the present. A first-order autoregressive time series model with exponential stationary marginal distribution was developed by Gaver and Lewis [3]. In recent years, many authors Jayakumar and Babu [10] and Gillariose and Tomy [6] have developed various autoregressive models with minification structures. In this section, we develop various Autoregressive models of order 1 with Marshall-Olkin Exponentiated Nadarajah Haghighi as marginals, namely MIN AR (1) Model I and Model II and MAX - MIN AR (1) Model I and Model II, and explore some properties.

6.1. MIN AR(1) Model - I with MOENH Marginal Distribution

Consider an AR(1) structure,

$$X_n=\begin{cases}\epsilon_n; & \text{with probability} \quad \delta\\ \min(X_{n-1},\epsilon_n); & \text{with probability} \quad 1-\delta\end{cases} \tag{16}$$

where $\{\epsilon_n\}$ is a sequence of iid random variables independent of $\{X_n\}$ and $\delta\in(0,1)$. Then the process is Stationary Markovian with MOENH Distribution.

Theorem 2. In an AR(1) process with structure (16), $\{X_n\}$ is Stationary Markovian with MOENH distribution with parameters δ, γ, β, λ iff $\{\epsilon_n\}$ is distributed as ENH(γ, β, λ)

Proof. Let $\epsilon_n\sim ENH(\gamma,\beta,\lambda)$ From(16)

$$\bar{G}_{X_n}(x)=\delta\bar{G}_{\epsilon_n}(x)+(1-\delta)\bar{G}_{X_{n-1}}(x)\bar{G}_{\epsilon_n}(x).$$

Under stationary equilibrium,

$$\bar{G}_X(x)=\frac{\delta\bar{G}_\epsilon(x)}{1-(1-\delta)\bar{G}_\epsilon(x)},$$

and hence

$$\bar{G}_\epsilon(x)=\frac{\bar{G}_X(x)}{\delta+(1-\delta)\bar{G}_X(x)}.$$

If $\epsilon_n \sim ENH(\gamma, \beta, \lambda)$

$$\bar{G}_\epsilon(x) = 1 - (1 - e^{1-(1+\lambda x)^\gamma})^\beta.$$

Thus

$$\begin{aligned}\bar{G}_X(x) &= \frac{\delta\{1 - (1 - e^{1-(1+\lambda x)^\gamma})^\beta\}}{1 - (1-\delta)\{1 - (1 - e^{1-(1+\lambda x)^\gamma})^\beta\}} \\ &= \frac{\delta - \delta(1 - e^{1-(1+\lambda x)^\gamma})^\beta}{\delta + (1-\delta)(1 - e^{1-(1+\lambda x)^\gamma})^\beta},\end{aligned}$$

which is the survival function of MOENH(δ). Conversely, if

$$\bar{G}_X(x) = \frac{\delta - \delta(1 - e^{1-(1+\lambda x)^\gamma})^\beta}{\delta + (1-\delta)(1 - e^{1-(1+\lambda x)^\gamma})^\beta},$$

then $\bar{G}_{\epsilon_n}(x)$ is distributed as ENH(γ, β, λ) and the process is stationary. In order to establish stationarity, assume that $X_{n-1} \sim MOENH(\delta)$ and $\epsilon_n \sim ENH(\gamma, \beta, \lambda)$ then,

$$\bar{G}_X(x) = \frac{\delta - \delta(1 - e^{1-(1+\lambda x)^\gamma})^\beta}{\delta + (1-\delta)(1 - e^{1-(1+\lambda x)^\gamma})^\beta}.$$

This means that X_n is distributed as MOENH(δ). ■

Remark 1. If X_0 has an arbitrary distribution G_{X_0}, the minification process is asymptotically stationary with MOENH($\delta, \gamma, \beta, \lambda$).
Since

$$\begin{aligned}\bar{G}_{X_n}(x) &= \delta\bar{G}_\epsilon(x)\sum_{i=0}^{n-1}(1-\delta)^i\bar{G}_\epsilon^i(x) + (1-\delta)^n\bar{G}_{X_0}(x)\bar{G}_\epsilon^n(x) \\ &= \frac{\delta\bar{G}_\epsilon(x)}{1-(1-\delta)\bar{G}_\epsilon(x)} \\ &= \frac{\delta - \delta(1 - e^{1-(1+\lambda x)^\gamma})^\beta}{\delta + (1-\delta)(1 - e^{1-(1+\lambda x)^\gamma})^\beta},\end{aligned}$$

the survival function of $MOENH(\delta, \gamma, \beta, \lambda)$.

6.2. MIN AR(1) Model - II with MOENH Distribution

Here we discuss a more general structure which allows probabilistic selection of process values, innovations and combinations of both. Consider the AR(1) structure given by

$$X_n = \begin{cases} X_{n-1}; & \text{with probability } \delta_1 \\ \epsilon_n; & \text{with probability } \delta_2 \\ \min(X_{n-1}, \epsilon_n); & \text{with probability } 1 - \delta_1 - \delta_2, \end{cases} \tag{17}$$

where $\delta_1, \delta_2 > 0$, $\delta_1 + \delta_2 < 1$ and $\{\epsilon_n\}$ is a sequence of iid random variables independent of $\{X_n\}$. Then the process is stationary with Marshall-Olkin Exponentiated Nadaraja Haghighi distribution.

Theorem 3. In an AR(1) process with structure (17), $\{x_n\}$ is stationary Markovian with MOENH distribution with parameters τ, γ, β, and λ iff $\{\epsilon_n\}$ is distributed as ENH with parameters γ, β and λ, where $\tau = \frac{\delta_2}{1-\delta_1}$.

Proof. Let $\epsilon_n \sim ENH(\gamma, \beta, \lambda)$. From (17)

$$\bar{G}_{X_n}(x) = \delta_1\bar{G}_{X_{n-1}}(x) + \delta_2\bar{G}_{\epsilon_n}(x) + (1 - \delta_1 - \delta_2)\,\bar{G}_{X_{n-1}}(x)\bar{G}_{\epsilon_n}(x)$$

Under stationary equilibrium we have,

$$\begin{aligned}\bar{G}_X(x) &= \frac{\tau\bar{G}_\epsilon(x)}{1-(1-\tau)\bar{G}_\epsilon(x)} \\ &= \frac{\tau\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}{1-(1-\tau)\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]} \\ &= \frac{\frac{\delta_2}{1-\delta_1}\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}{1-\left(1-\frac{\delta_2}{1-\delta_1}\right)\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]},\end{aligned}$$

where $\tau = \frac{\delta_2}{1-\delta_1}$, which is in the Marshall-Olkin form. Now let us assume that $\{X_n\} \sim$ MOENH$(\tau, \gamma, \beta, \lambda)$. From (17) under stationarity,

$$\bar{G}_\epsilon(x) = \frac{(1-\delta_1)\,\bar{G}_X(x)}{\delta_2 + (1-\delta_1-\delta_2)\,\bar{G}_X(x)}.$$

Now by using X_n as MOENH $(\tau, \gamma, \beta, \lambda)$, we have

$$\begin{aligned}\bar{G}_\epsilon(x) &= \frac{(1-\delta_1)\left[\frac{\frac{\delta_2}{1-\delta_1}\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}{1-\left(1-\frac{\delta_2}{1-\delta_1}\right)\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}\right]}{\delta_2 + (1-\delta_1-\delta_2)\left[\frac{\frac{\delta_2}{1-\delta_1}\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}{1-\left(1-\frac{\delta_2}{1-\delta_1}\right)\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}\right]} \\ &= 1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta.\end{aligned}$$

Which is the Survival function of Exponentiated Nadaraja Haghighi distribution with parameters γ, β and λ. ■

6.3. MAX-MIN AR(1) Model - I with MOENH Distribution

Consider the AR(1) structure given by,

$$X_n = \begin{cases}\max(X_{n-1}, \epsilon_n); & \text{with probability } \delta_1 \\ \min(X_{n-1}, \epsilon_n); & \text{with probability } \delta_2 \\ X_{n-1}; & \text{with probability } 1-\delta_1-\delta_2,\end{cases} \tag{18}$$

where $0 < \delta_1, \quad \delta_2 > 1, \quad \delta_2 < \delta_1, \quad \delta_1 + \delta_2 < 1$ and $\{\epsilon_n\}$ is a sequence of iid random variables independent of $\{X_n\}$. Then the process is Stationary Markovian with Marshall-Olkin Exponentiated Nadaraja Haghighi distribution.

Theorem 4. In an $AR(1)$ MAX-MIN process with structure (18), $\{X_n\}$ is a stationary Markovian AR(1) MAX-MIN process with MOENH stationary distribution with parameters τ, γ, β and λ iff $\{\epsilon_n\}$ is distributed as ENH with parameters γ, β and λ, where $\tau = \frac{\delta_1}{\delta_2}$.

Proof. Let $\epsilon_n \sim$ ENH(γ, β, λ). From (18) we have,

$$\bar{G}_{X_n}(x) = \delta_1\left[1-\left(1-\bar{G}_{X_{n-1}}(x)\right)\left(1-\bar{G}_{\epsilon_n}(x)\right)\right] + \delta_2\bar{G}_{X_{n-1}}(x)\bar{G}_{\epsilon_n}(x) + (1-\delta_1-\delta_2)\,\bar{G}_{X_{n-1}}(x).$$

Under stationary equilibrium,

$$\begin{aligned}\bar{G}_{X_n}(x) &= \frac{\tau \bar{G}_{\epsilon(x)}}{1-(1-\tau)\bar{G}_{\epsilon}(x)} \\ &= \frac{\tau[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}]}{1-(1-\tau)[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}]} \\ &= \frac{\frac{\delta_1}{\delta_2}[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}]}{1-(1-\frac{\delta_1}{\delta_2})[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}]}.\end{aligned}$$

Where $\tau = \frac{\delta_1}{\delta_2}$ and $\bar{G}_{X_n}(x)$ is in the form of Marshall-Olkin distribution. Now Let $X_n \sim$ MOENH$(\tau, \gamma, \beta, \lambda)$. Then from (18), under stationarity,

$$\bar{G}_{\epsilon}(x) = \frac{\delta_2 \bar{G}_{X_n}(x)}{\delta_1 + (\delta_2 - \delta_1)\bar{G}_{X_n}(x)}.$$

Thus, after simplification it can be written as

$$\begin{aligned}\bar{G}_{\epsilon}(x) &= \frac{\delta_2 \left[\frac{\frac{\delta_1}{\delta_2}\left[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right]}{1-\left(1-\frac{\delta_1}{\delta_2}\right)\left[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right]}\right]}{\delta_1 + (\delta_2-\delta_1)\left[\frac{\frac{\delta_1}{\delta_2}\left[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right]}{1-\left(1-\frac{\delta_1}{\delta_2}\right)\left[1-\left(1-e^{1-(1+\lambda x)^{\gamma}}\right)^{\beta}\right]}\right]} \\ &= 1-(1-e^{1-(1+\lambda x)^{\gamma}})^{\beta},\end{aligned}$$

which is the survival function of Exponentiated Nadaraja Haghighi distribution with parameters γ, β and λ. ■

6.4. MAX-MIN AR(1) Model - II with MOENH Distribution

Consider the more general MAX-MIN process that includes minimum, maximum innovations and the process values, The AR(1) structure is given by

$$X_n = \begin{cases} \max(X_{n-1}, \epsilon_n); & \text{with probability } \delta_1 \\ \min(X_{n-1}, \epsilon_n); & \text{with probability } \delta_2 \\ \epsilon_n; & \text{with probability } \delta_3 \\ X_{n-1}; & \text{with probability } 1-\delta_1-\delta_2-\delta_3, \end{cases} \tag{19}$$

where $0 < \delta_1, \quad \delta_2, \quad \delta_3 < 1, \quad \delta_1+\delta_2+\delta_3 < 1$ and $\{\epsilon_n\}$ is a sequence of iid random variables independent of $\{X_n\}$. Then the process is stationary Markovian with Marshall-Olkin Exponentiated Nadaraja Haghighi distribution.

Theorem 5. AR(1) MAX-MIN process $\{X_n\}$ with structure (19) is a stationary Markovian AR(1) MAX-MIN process with MOENH distribution $(\tau, \gamma, \beta, \lambda)$ if $\{\epsilon_n\}$ is distributed as ENH with parameters γ, β and λ where $\tau = \frac{\delta_1+\delta_3}{\delta_2+\delta_3}$.

Proof. Let $\epsilon_n \sim$ ENH(γ, β, λ). From (19) we have,

$$\begin{aligned}\bar{G}_{Xn}(x) &= \delta_1[1-(1-\bar{G}_{X_{n-1}}(x))(1-\bar{G}_{\epsilon_n}(x))] + \delta_2 \bar{G}_{X_{n-1}}(x)\bar{G}_{\epsilon_n}(X) + \delta_3 \bar{G}_{\epsilon_n}(X) \\ &\quad + (1-\delta_1-\delta_2-\delta_3)\bar{G}_{X_{n-1}}(x)\end{aligned}$$

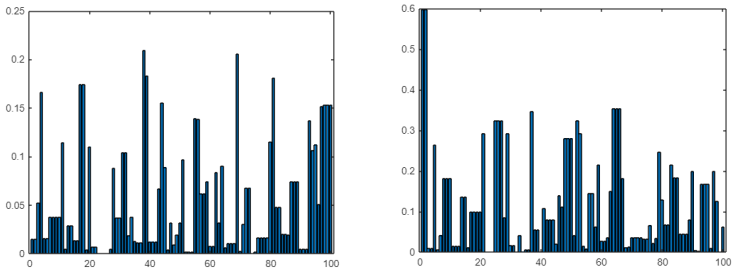


Figure 2: *Sample path for AR(1) Minification Model-I for* $p = 0.6, 0.7$, β=0.5, γ=1.2 *and* λ=1.

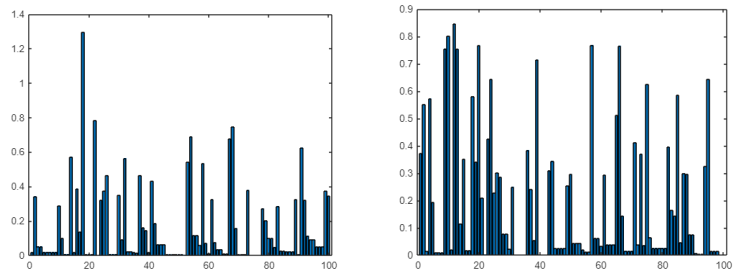


Figure 3: *Sample path of AR(1) Minification Model-II for different sets of* $(p_1, p_2) = (0.3, 0.4), (0.2, 0.5), \beta = 0.5, \gamma = 1.5,$ *and* $\lambda = 1$.

Under stationary equilibrium it gives,

$$\begin{aligned}\bar{G}_{X_n}(x) &= \frac{\tau\bar{G}_\epsilon(x)}{1-(1-\tau)\bar{G}_\epsilon(x)} \\ &= \frac{\frac{\delta_1+\delta_3}{\delta_2+\delta_3}[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta]}{1-(1-\frac{\delta_1+\delta_3}{\delta_2+\delta_3})[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta]},\end{aligned}$$

where $\tau = \frac{\delta_1+\delta_3}{\delta_2+\delta_3}$, which is in the form of Marshall-Olkin distribution. Now let $X_n \sim \text{MOENH}(\tau, \gamma, \beta, \lambda)$. Then from (19), we have

$$\bar{G}_\epsilon(x) = \frac{(\delta_2+\delta_3)\,\bar{G}_X(x)}{(\delta_1+\delta_3)+(\delta_2-\delta_1)\,\bar{G}_X(x)}$$

By simplifying, we get

$$\begin{aligned}\bar{G}_\epsilon(x) &= \frac{(\delta_2+\delta_3)\left[\frac{\frac{\delta_1+\delta_3}{\delta_2+\delta_3}\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}{1-\left(1-\frac{\delta_1+\delta_3}{\delta_2+\delta_3}\right)\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}\right]}{(\delta_1+\delta_3)+(\delta_2-\delta_1)\left[\frac{\frac{\delta_1+\delta_3}{\delta_2+\delta_3}\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}{1-\left(1-\frac{\delta_1+\delta_3}{\delta_2+\delta_3}\right)\left[1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta\right]}\right]} \\ &= 1-\left(1-e^{1-(1+\lambda x)^\gamma}\right)^\beta.\end{aligned}$$

Which is the Survival function of Exponentiated Nadaraja Haghighi distribution with parameters γ, β and λ. Figure 2-5 displays the sample path features of the four AR(1) models created in this section and how these measures change with different parameter settings. ■

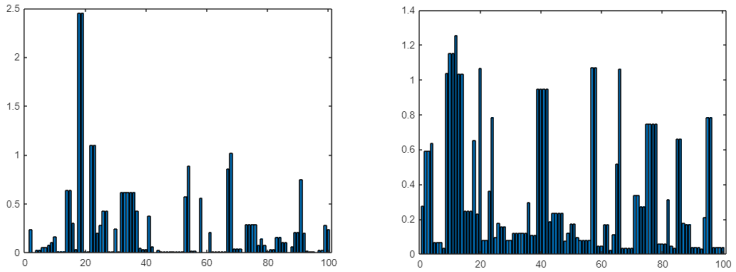


Figure 4: *Sample path for AR(1) MAX-MIN Model -I for various combinations of* $(p_1, p_2) = (0.2, 0.1), (0.3, 0.2), \beta = 0.25, \gamma = 1.2,$ *and* $\lambda = 1$.

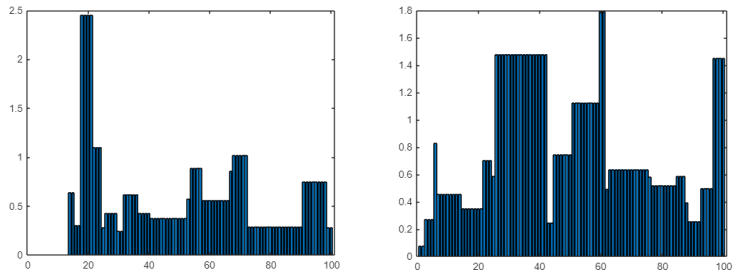


Figure 5: *Sample path for AR(1) MAX-MIN Model -II for* $(p_1, p_2, p_3) = (0.4, 0.2, 0.2), (0.5, 0.2, 0.1), \beta = 0.25, \gamma = 1.2,$ *and* $\lambda = 1$.

7. Numerical illustration

7.1. Simulation study

A simulation study is conducted to evaluate the effectiveness of the MLEs for estimating the parameters of MOENH distribution. For this, we take into account, $\alpha = 0.5, \beta = 1.2, \gamma = 0.8,$ and $\lambda = 0.09$. For different sample sizes of n = 1000, n = 2000, and n = 4000, we simulate data from the MOENH model and determine the MLEs by maximizing the likelihood function. We carry out the procedure, 10000 times, and the results show that bias and root-mean-square error (RMSE) decreases as sample size increases. The results are in the Table 1.

Table 1: *Simulation*

n	Parameters	Estimate	Bias	RMSE
1000	α	0.6352	0.0135	0.1687
	β	1.2098	0.0009	0.0246
	γ	0.9561	0.0156	0.1306
	λ	0.1352	0.0045	0.0512
2000	α	0.5895	0.0089	0.1295
	β	1.2074	0.0007	0.0168
	γ	0.9000	0.0100	0.0995
	λ	0.1193	0.0029	0.0373
4000	α	0.5509	0.0051	0.0881
	β	1.2027	0.0002	0.0113
	γ	0.8511	0.0051	0.0652
	λ	0.1058	0.0015	0.0242

7.2. Data illustration

We take into account the following DARWin data set for numerical illustration. DARWin is a project, within the Swedish industry organization Swedish Energy, which collects and annually presents outage data from most of the Swedish electricity system operators. The annual reports from Swedish Energy are open-accessible and can be downloaded from the Swedish Energy website. The unplanned events from 2012 divided into voltage level (12 k V) and failure causes are given by Ekstedt et al. [2]. The parameters are estimated using the maximum likelihood method. Akaike information criteria (AIC), Bayesian information criteria (BIC), Kolmogrov-Smirnov (K-S), and $p - value$ are the goodness-of-fit metrics that we take into consideration.

Table 2: *Parameter estimates and goodness of fit statistics for models fitted to the data*

Model	MLE	-Log L	AIC	BIC	KS	p-VALUE
MOENH	$\alpha = 2.000$ $\beta = 0.2777$ $\gamma = 1.2189$ $\lambda = 0.00015$	73.6221	155.24	156.03	0.10819	0.9994
ENH	$\beta = 0.2777$ $\gamma = 1.2189$ $\lambda = 0.00013$	78.6904	163.38	163.9725	0.2306	0.6449
NH	$\gamma = 1.2189$ $\lambda = 0.0003$	79.124	162.249	132.6435	0.2370	0.6122
Exp	$\lambda = 0.0004$	78.8909	159.78	159.982	0.2240	0.6782

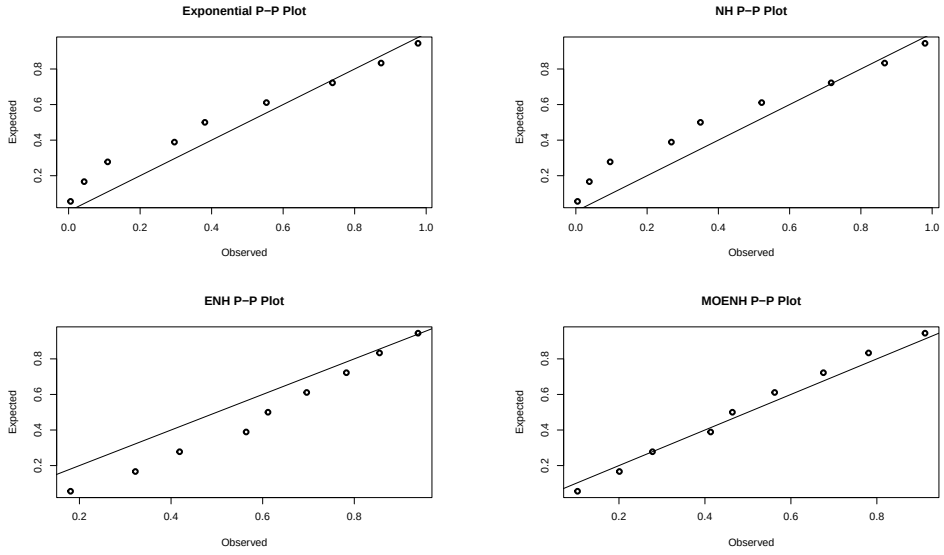


Figure 6: *pp-plots*

Table 2 lists the parameter estimates and goodness of fit statistics for DARWin voltage level (12 k V) failure data. The MOENH model is more suitable for this data since the values of $-\log L$, AIC, BIC, K-S and p-value for the MOENH distribution are lower than those of the other competing models. Figure 6 represents the pp-plot for the fitted models.

7.3. Testing of Hypothesis

In this section, we present the likelihood ratio test procedure for testing the significance of the parameters of the MOENH model. We consider LR statistics to check if the fitted MOENH

distribution for a given data set is statistically superior to the fitted exponential, NH, ENH distributions. In any case, the hypothesis test of the type $H_0 : \theta = \theta_0$ *vs* $H_1 : \theta \neq \theta_0$ using the generalized likelihood ratio test. The test statistic is,

$$-2\ln\lambda(x) = 2[\ln L(\hat{\Theta}; x) - \ln L(\hat{\Theta}^*; x)] \tag{20}$$

where $\hat{\Theta}$ is the maximum likelihood estimator with no restriction, and $\hat{\Theta}^*$ is the maximum likelihood estimator with restriction. The test statistic follows a Chi-square distribution with degrees of freedom $(df = df_{alt} - df_{null})$. So here we consider the following likelihood ratio tests.

1. $H_{01} : \alpha = \beta = \gamma = 1$, the sample is from $Exp(\lambda)$
   $H_{11} : \alpha \neq \beta \neq \gamma \neq 1$, the sample is from $MOENH(\alpha, \beta, \gamma, \lambda)$
2. $H_{02} : \alpha = \beta = 1$, the sample is from NH
   $H_{12} : \alpha \neq \beta \neq 1$, the sample is from $MOENH(\alpha, \beta, \gamma, \lambda)$
3. $H_{03} : \alpha = 1$, the sample is from ENH
   $H_{13} : \alpha \neq 1$, the sample is from $MOENH(\alpha, \beta, \gamma, \lambda)$

Table 3: *Likelihood ratio test*

Model	Hypothesis	Test statistic	p-value
Exp vs $MOENH$	$H_{01} : \alpha, \beta, \gamma = 1$ $H_{11} : H_{01}$ is false	10.5406	0.0145
NH vs $MOENH$	$H_{02} : \alpha, \beta = 1$ $H_{12} : H_{02}$ is false	11.0046	0.0041
ENH vs $MOENH$	$H_{03} : \alpha = 1$ $H_{13} : H_{03}$ is false	10.1365	0.0014

The test statistic $-2\ln\lambda(x)$ given in (20) is asymptotically distributed as χ^2 with three degrees of freedom for test 1, 2 degrees of freedom for test 2, and 1 degree of freedom for test 3. The computed values of the test statistic in the case of the DARWin data set are listed in Table 3. From Table 3, we can see that $p - value$ is less than the significant level of 0.05.LR tests reject the three sub-models in favour of the MOENH distribution. Since the critical values at the significance level 0.05 and degree of freedom three, two, and one for the two-tailed tests are 9.348, 7.378, and 5.024 respectively the null hypothesis is rejected in all cases, which shows the appropriateness of the MOENH distribution to the DARWin data.

8. Conclusion

In this paper, as a generalization of the exponential distribution, the MOENH distribution is introduced. Statistical properties, characterization properties, and autoregressive time series models of MOENH distribution are obtained. It is shown that the new model is a competitor to the exponential distribution for modeling certain types of data sets. Also, the generation of random variables from the new model is simple. The new model may attract the attention of researchers as a viable competitor to the exponential distribution.

References

[1] Cordeiro, G. M. and Lemonte, A. J. (2013). On the Marshall-Olkin extended Weibull distribution. *Statistical papers*, 54: 333–353.

[2] Ekstedt, N., Wallnerström, C. J., Babu, S., Hilber, P., Westerlund, P., Jürgensen, J. H. and Lindquist, T. M. (2014). Reliability Data: A Review of Importance, Use, and Availability. *In: NORDAC 2014 (Eleventh Nordic Conference on Electricity Distribution System Management and Development, Stockholm, 8 - 9 September 2014).*

[3] Gaver, D. P. and Lewis, P. A. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, 12(3): 727–745.

[4] George, R. and Thobias, S. (2019). Kumaraswamy Marshall-Olkin exponential distribution. *Communications in Statistics-Theory and methods*, 48(8): 1920–1937.

[5] Gillariose, J. and Tomy, L. (2020). The Marshall-Olkin extended power Lomax distribution with applications. *Pakistan Journal of Statistics and Operation Research*, 16(2): 331–341.

[6] Gillariose, J. and Tomy, L. (2023). On an extension of the two-parameter Lindley distribution. Reliability: Theory & Applications, 18(72): 385–402.

[7] Glänzel, W. (1987). A Characterization Theorem Based on Truncated Moments and its Application to Some Distribution Families. In: Bauer, P., Konecny, F., Wertz, W. (eds) Mathematical Statistics and Probability Theory. Springer, Dordrecht: 75–84.

[8] Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and methods*, 27(4): 887–904.

[9] Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 43(1): 117–130.

[10] Jayakumar, K. and Babu, M. G. (2015). Some Generalizations of Weibull Distribution and Related Processes. *Journal of Statistical Theory and Applications*, 14(4): 425–434.

[11] Lemonte, A. J. (2013). A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics & Data Analysis*, 62: 149–170.

[12] Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions, with application to the exponential and Weibull families. *Biometrika*, 84(3): 641–652.

[13] Nadarajah, S. and Haghighi, F. (2011). An extension of the exponential distribution. *Statistics*, 45(6): 543–558.

[14] Ristić, M. M. and Kundu, D. (2015). Marshall-Olkin generalized exponential distribution. *Metron*, 73: 317–333.

QUADRASOPHIC FUZZY MATRIX AND ITS APPLICATION

G. Aruna[1], J. Jesintha Rosline[2]

•

[1,2] PG and Research Department of Mathematics, Auxilium College (Autonomous) ,
Affiliated to Thiruvalluvar University, Vellore Dist., Tamil Nadu, India
anu9117@gmail.com[1] , jesi.simple@gmail.com[2]

Abstract

Quadrasophic Fuzzy Set is one of the generalizations of Fuzzy set theory. In this artifact, a definition of the Quadrasophic Fuzzy Algebra and its characteristics are provided. The definition of a Quadrasophic Fuzzy Matrix is explored with the aid of Quadrasophic Fuzzy Algebra. The binary operators of Fuzzy Matrices are used to describe various kinds and specific operations on Quadrasophic Fuzzy Matrices. The theorems and results of Quadrasophic Fuzzy Matrix are demonstrated with pertinent examples and proofs. Additionally, the illustration of the identification of paddy illnesses is analyzed with the tool of Quadrasophic Fuzzy Matrix in the decision-making process.

Keywords: Quadrasophic Fuzzy Set, Quadrasophic Fuzzy Algebra, Quadrasophic Fuzzy Matrix, Operations, Plant illness identification

1. Introduction

The ability to recognize both the advantages and disadvantages of a problem is crucial in problem-solving. In recent years, we have considered other elements while making judgements, like the environment and peer pressure. The bipolar fuzzy set addresses both the problems in positive and negative aspects. However, the environment's impact on the problem is crucial to manage the unclear scenario. The current fuzzy set extension is insufficient to account for the environment's influence or the rate at which the issue is becoming more prevalent. With the unique characteristics of positive reluctant membership and negative reluctant membership functions, QFS facilitates solving such issues.

One of the generalizations of fuzzy set theory is the Quadrasophic Fuzzy Set (QFS) [1]. Many industries, including medicine, business, decision-making, agriculture, etc., rely heavily on QFS. Any sort of work has both a positive and a negative aspect.These aspects might change due to influential variables, producing different kinds of outcomes. QFS considers the impact of influential factors as the reluctant positive side, i.e., the partial positive side, and the reluctant negative side [1]. Let's take an example where someone is driving from City A to City B. Positive membership indicates that the traveller has reached City B; negative membership indicates that there has been a delay in reaching the destination because of traffic or road construction; these variations in positive and negative membership values are therefore regarded as positive and negative reluctant memberships. The QFS quantifies the value of positive membership in the interval [0,1], negative membership in the interval [-1,0], negative reluctant ranges in [-0.5,0], and postive reluctant ranges in [0,0.5] [2].

In numerous distinct domains, matrices are significant. The traditional matrix is insufficient to cover unpredictable circumstances. The ambiguous situations make it possible to create

the idea of a fuzzy matrix to depict the actual situation. Thomason discussed the convergence of power in a fuzzy matrix in 1977 [3]. Fuzzy matrices were first conceptualized by Hashimoto. Several authors established and validated numerous results, including determinant, trace, Nil potency, adjoint matrices, adjacent matrices, Idem potency, features of transitive and closure, and power series convergence in fuzzy matrices as well as in their different expansions [4], [5], [6].

Cho conducted research on the regularity characteristics and rankings of fuzzy matrices in 1999 [7]. Pal applied numerous matrix concepts to the fuzzy sets. Pal investigated and established Interval-valued Intuitionistic fuzzy matrices as well as the Intuitionistic fuzzy determinant [5], [8]. The notion of a matrix was applied to triangular fuzzy numbers by A.K. Shyamal et al. [9]. They also developed triangular fuzzy matrices and addressed their specific types of properties. Pal (2016) developed four binary fuzzy operators, expanded the idea of fuzzy matrices with fuzzy rows and columns, and applied them to image processing [10]. In 2016, Mondal and Pal investigated the eigen vectors and eigen values of bipolar fuzzy matrices [11]. In 2019, Pal and Sanjib defined and introduced the concept of bipolar fuzzy matrices, their power convergence, and their applications[12]. Dogra and Pal introduced the picture fuzzy matrix with the determinant and adjoint of its unique restricted square picture fuzzy matrices [13]. Fuzzy matrices were created and used in a variety of fields to handle uncertain situations. Numerous authors, including Sanchez, Pal, Muthuraji, Nagoor Gani, and Meenakshi, have examined and applied various results on fuzzy intuitionistic matrix, fuzzy interval valued matrix, triangular fuzzy matrix, bipolar fuzzy matrix, picture fuzzy matrix, neutrosophic fuzzy number, and fuzzy soft matrix in the field of decision making [14],[15], [16], [17], [18]. Inspired by these inventions, we define the Quadrasophic Fuzzy Matrix (QFM) in this artifact. Also, we explore the application of Quadrasophic Fuzzy Matrix in the detection of plant illnesses.

In this artifact, Section 2 provides an overview of QFS's fundamental definitions and functions. In Section 3, we defined the operations of QFS, and Section 4 presented the introduction to Quadrasophic Fuzzy Algebra. The Quadrasophic Fuzzy Matrix is defined in Section 5 and includes relevant examples along with its operations and propositions. In Section 6, the identification of plant diseases serves as an example of the use of the Quadrasophic Fuzzy Matrix in decision-making. The conclusion and the scope of future investigation potential are presented in Section 7.

2. Preliminaries

Quadrasophic fuzzy set [1]: The Quadrasophic fuzzy set q on the set U is defined as

$$q = \{(x, \eta_q(x) , \lambda_{\eta_q}(x) , \lambda_{\mu_q}(x) , \mu_q(x)) | x \in U\}$$

where the degree of positive membership grade is $\mu_q(x) : U \to [0, 1]$, the degree of negative membership is $\eta_q(x) : U \to [-1, 0]$ the degree of restricted positive membership is $\lambda_{\mu_q}(x) : U \to [0, 0.5]$ the degree of restricted negative membership is $\lambda_{\eta_q}(x) : U \to [-0.5, 0]$ And the condition follows: $-1 \leq \mu_q(x) + \eta_q(x) \leq 1$, $-0.5 \leq \lambda_q \leq 0.5$ and $0 \leq \mu_q^2 + \eta_q^2 + \lambda_q^2 \leq 3$ for all $x \in U$, such that $\lambda_q =$ Length of $(\lambda_{\mu_q}, \lambda_{\eta_q})$.

Intersection operation of QFS [2]: Let q_a, $q_b \in$ QFS, then for all $x \in X$ the intersection of q_a and q_b is represented as:

$$q_a \cap q_b = (\eta_{q_a}(x) \vee \eta_{q_b}(x) , \lambda_{\eta_{q_a}} (x) \vee \lambda_{\eta_{q_b}} (x) , \lambda_{\mu_{q_a}} (x) \wedge \lambda_{\mu_{q_b}} (x) , \mu_{q_a}(x) \wedge \mu_{q_b}(x))$$

Union operation of QFS [2]: Let q_a, $q_b \in$ QFS, then for all $x \in X$ the union of q_a and q_b is represented as:

$$q_a \cup q_b = (\eta_{q_a}(x) \wedge \eta_{q_b}(x) , \lambda_{\eta_{q_a}} (x) \wedge \lambda_{\eta_{q_b}} (x) , \lambda_{\mu_{q_a}} (x) \vee \lambda_{\mu_{q_b}} (x) , \mu_{q_a}(x) \vee \mu_{q_b}(x))$$

Complement[2]: Let $q \in$ QFS, then the complement of q is described as $q^c = (-1-\eta_q, -0.5-\lambda_{\eta_q}, 0.5-\lambda_{\mu_q}, 1-\mu_q)$.

Score- valued function [2]: The score- valued function $score(q)$ of QFS is represented as $score(q) = \frac{\mu_q(x)+\lambda_{\mu_q}(x)+\eta_q(x)+\lambda_{\eta_q}(x)}{3}$, for all $score(q) \in [-1,1]$.

The $\ominus$ operator: [6] For $u, v \in [0,1]$, the $\ominus$ operator can be defined as :

$$u \ominus v = \begin{cases} u, & \text{if } u > v \\ 0, & \text{if u} \leq \text{v.} \end{cases}$$

3. Certain Operations on Quadrasophic Fuzzy Sets

Operations on Quadrasophic Fuzzy Sets: Let $q_1 = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1})$ and $q_2 = (-\eta_{q_2}, -\lambda_{\eta_{q_2}}, \lambda_{\mu_{q_2}}, \mu_{q_2})$ be two Quadrasophic Fuzzy sets in the set Q_F.

Disjunction: $q_1 + q_2 = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (-\eta_{q_2}, -\lambda_{\eta_{q_2}}, \lambda_{\mu_{q_2}}, \mu_{q_2})$

$$= \{-max(\eta_{q_1}, \eta_{q_2}), -max(\lambda_{\eta_{q_1}}, \lambda_{\eta_{q_2}}), max(\lambda_{\mu_{q_1}}, \lambda_{\mu_{q_2}}), max(\mu_{q_1}, \mu_{q_2})\}$$
$$= \{-(\eta_{q_1} \vee \eta_{q_2}), -(\lambda_{\eta_{q_1}} \vee \lambda_{\eta_{q_2}}), (\lambda_{\mu_{q_1}} \vee \lambda_{\mu_{q_2}}), (\mu_{q_1} \vee \mu_{q_2})\}$$

Parallel conjunction: $q_1 \cdot q_2 = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (-\eta_{q_2}, -\lambda_{\eta_{q_2}}, \lambda_{\mu_{q_2}}, \mu_{q_2})$

$$= \{-min(\eta_{q_1}, \eta_{q_2}), -min(\lambda_{\eta_{q_1}}, \lambda_{\eta_{q_2}}), min(\lambda_{\mu_{q_1}}, \lambda_{\mu_{q_2}}), min(\mu_{q_1}, \mu_{q_2})\}$$
$$= \{-(\eta_{q_1} \wedge \eta_{q_2}), -(\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}), (\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}}), (\mu_{q_1} \wedge \mu_{q_2})\}$$

Serial conjunction: $q_1 \otimes q_2 = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (-\eta_{q_2}, -\lambda_{\eta_{q_2}}, \lambda_{\mu_{q_2}}, \mu_{q_2})$

$$= \{-((\eta_{q_1} \wedge \mu_{q_2}) \vee (\mu_{q_1} \wedge \eta_{q_2}))\},$$
$$\{-((\lambda_{\eta_{q_1}} \wedge \lambda_{\mu_{q_2}}) \vee (\lambda_{\mu_{q_1}} \wedge \lambda_{\eta_{q_2}}))\},$$
$$\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}) \vee (\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}})\},$$
$$\{-((\eta_{q_1} \wedge \eta_{q_2}) \vee (\mu_{q_1} \wedge \mu_{q_2}))\}$$

Negation of q_1**:** $-q1 = -(-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1})$

$$= (-\mu_{q_1}, -\lambda_{\mu_{q_1}}, \lambda_{\eta_{q_1}}, \eta_{q_1})$$

Complement of q_1**:** $\rightharpoonup q1 = \rightharpoonup (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1})$

$$= (-1+\eta_{q_1}, -0.5+\lambda_{\eta_{q_1}}, 0.5-\lambda_{\mu_{q_1}}, 1-\mu_{q_1})$$

QFS $\ominus$ operator:

$$q_1 \ominus q_2 = \begin{cases} -\eta_{q_1}, & \text{for } -(\eta_{q_1} > \eta_{q_2}) \\ -\lambda_{\eta_{q_1}}, & \text{for } -(\lambda_{\eta_{q_1}} > \lambda_{\eta_{q_2}}) \\ \lambda_{\mu_{q_1}}, & \text{for } (\lambda_{\mu_{q_1}} > \lambda_{\mu_{q_2}}) \\ \mu_{q_1}, & \text{for } (\mu_{q_1} > \mu_{q_2}) \\ 0, & \text{otherwise.} \end{cases}$$

Zero element of QFS: The zero element of QFS is represented by 0_q and is described as $0_q = (0, 0, 0, 0)$.
Unit element of QFS: The unit element of QFS is represented by i_q and is described as $i_q = (-1, -0.5, 0.5, 1)$.

4. Quadrasophic Fuzzy Algebra

A Quadrasophic Fuzzy Algebra is a mathematical structure (QF, +,.) with two binary operations (+ and .) defined on QFS that fulfill the following principle.

Theorem 1. Consider $q_1 = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1})$ $q_2 = (-\eta_{q_2}, -\lambda_{\eta_{q_2}}, \lambda_{\mu_{q_2}}, \mu_{q_2})$ $q_3 = (-\eta_{q_3}, -\lambda_{\eta_{q_3}}, \lambda_{\mu_{q_3}}, \mu_{q_3})$ belongs to QFS. Then the results as follows:
i. Law of Idempotent: $q_1 + q_1 = q_1, q_1.q_1 = q_1$
ii. Law of Commutativity: $q_1 + q_2 = q_2 + q_1, q_1.q_2 = q_2.q_1$
iii. Law of Associativity: $q_1 + (q_2 + q_3) = (q_1 + q_2) + q_3, q_1.(q_2.q_3) = (q_1.q_2).q_3$
iv. Law of Absorption: $q_1 + (q_1.q_2) = q_1, q_1.(q_1 + q_2) = q_1$
v. Law of Distributivity: $q_1.(q_2 + q_3) = (q_1.q_2) + (q_1.q_3), q_1 + (q_2.q_3) = (q_1 + q_2).(q_1 + q_3)$
vi. Universal Law: $q_1 + 0_q = q_1, q_1 + i_q = i_q, q_1.0_q = 0_q, q_1.i_q = q_1$

Proof. *i.* $q_1 + q_1 = (-\{\eta_{q_1} \vee \eta_{q_1},\}, -\{\lambda_{\eta_{q_1}} \vee \lambda_{\eta_{q_1}}\}, \{\lambda_{\mu_{q_1}} \vee \lambda_{\mu_{q_1}}\}, \{\mu_{q_1} \vee \mu_{q_1}\})$
$= (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1})$
$= q_1$
Similarly, $q_1 \cdot q_1 = (-\{\eta_{q_1} \wedge \eta_{q_1},\}, -\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_1}}\}, \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_1}}\}, \{\mu_{q_1} \wedge \mu_{q_1}\})$
$= (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1})$
$= q_1$

ii. $q_1 + q_2 = (-\{\eta_{q_1} \vee \eta_{q_2},\}, -\{\lambda_{\eta_{q_1}} \vee \lambda_{\eta_{q_2}}\}, \{\lambda_{\mu_{q_1}} \vee \lambda_{\mu_{q_2}}\}, \{\mu_{q_1} \vee \mu_{q_2}\})$
$= (-\eta_{q_1}, -\lambda_{\eta_{q_2}}, \lambda_{\mu_{q_1}}, \mu_{q_2})$
$= (-\{\eta_{q_2} \vee \eta_{q_1},\}, -\{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_1}}\}, \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_1}}\}, \{\mu_{q_2} \vee \mu_{q_1}\})$
$= (-\eta_{q_2}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_2}}, \mu_{q_1})$
$= q_2 + q_1$
Similarly, $q_1 \cdot q_2 = q_2 \cdot q_1$.

$$\begin{aligned} \textit{iii. } q_1 + (q_2 + q_3) =& (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (-\{\eta_{q_2} \vee \eta_{q_3},\}, -\{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_3}}\}, \\ & \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}, \{\mu_{q_2} \vee \mu_{q_3}\}) \\ & = -max(\eta_{q_1}, \{\eta_{q_2} \vee \eta_{q_3}\}), -max(\lambda_{\eta_{q_1}}, \{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_3}}\}), \\ & max(\lambda_{\mu_{q_1}}, \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}), max(\mu_{q_1}, \{\mu_{q_2} \vee \mu_{q_3}\}) \\ & = -(\eta_{q_1} \vee \{\eta_{q_2} \vee \eta_{q_3}\}), -(\lambda_{\eta_{q_1}} \vee \{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_3}}\}), \\ & (\lambda_{\mu_{q_1}} \vee \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}), (\mu_{q_1} \vee \{\mu_{q_2} \vee \mu_{q_3}\}) \end{aligned}$$

$$\begin{aligned} (q_1 + q_2) + q_3 =& (-\{\eta_{q_1} \vee \eta_{q_2},\}, -\{\lambda_{\eta_{q_1}} \vee \lambda_{\eta_{q_2}}\}, \{\lambda_{\mu_{q_1}} \vee \lambda_{\mu_{q_2}}\}, \{\mu_{q_1} \vee \mu_{q_2}\}) \\ & + (-\eta_{q_3}, -\lambda_{\eta_{q_3}}, \lambda_{\mu_{q_3}}, \mu_{q_3}) \\ & = -max(\{\eta_{q_1} \vee \eta_{q_2}\}, \eta_{q_3}), -max(\{\lambda_{\eta_{q_1}} \vee \lambda_{\eta_{q_2}}\}, \lambda_{\eta_{q_3}}), \\ & max(\{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}, \lambda_{\mu_{q_1}}), max(\mu_{q_1}, \{\mu_{q_2} \vee \mu_{q_3}\}) \\ & = -(\eta_{q_1} \vee \{\eta_{q_2} \vee \eta_{q_3}\}), -(\lambda_{\eta_{q_1}} \vee \{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_3}}\}), \\ & (\lambda_{\mu_{q_1}} \vee \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}), (\mu_{q_1} \vee \{\mu_{q_2} \vee \mu_{q_3}\}) \end{aligned}$$

Hence, $q_1 + (q_2 + q_3) = (q_1 + q_2) + q_3$
In similar way, $q_1 \cdot (q_2 \cdot q_3) = (q_1 \cdot q_2) \cdot q_3$

$$\begin{aligned} iv.\ q_1 + (q_1 \cdot q_2) =& (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (-\{\eta_{q_1} \wedge \eta_{q_2},\}, -\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}\}, \\ & \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}}\}, \{\mu_{q_1} \wedge \mu_{q_2}\}) \\ & = -(\eta_{q_1} \vee \{\eta_{q_1} \wedge \eta_{q_2}\}), -(\lambda_{\eta_{q_1}} \vee \{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}\}), \\ & (\lambda_{\mu_{q_1}} \vee \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}}\}), (\mu_{q_1} \vee \{\mu_{q_1} \wedge \mu_{q_2}\}) \\ & = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) = q_1 \end{aligned}$$

Similarly, $q_1.(q_1 + q_2) = q_1$

$$\begin{aligned} v.\ q_1 \cdot (q_2 + q_3) =& (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) \cdot (-\{\eta_{q_2} \vee \eta_{q_3},\}, -\{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_3}}\}, \\ & \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}, \{\mu_{q_2} \vee \mu_{q_3}\}) \\ & = -(\eta_{q_1} \wedge \{\eta_{q_2} \vee \eta_{q_3}\}), -(\lambda_{\eta_{q_1}} \wedge \{\lambda_{\eta_{q_2}} \vee \lambda_{\eta_{q_3}}\}), \\ & (\lambda_{\mu_{q_1}} \wedge \{\lambda_{\mu_{q_2}} \vee \lambda_{\mu_{q_3}}\}), (\mu_{q_1} \wedge \{\mu_{q_2} \vee \mu_{q_3}\}) \\ & = -(\{\eta_{q_1} \wedge \eta_{q_2}\} \vee \{\eta_{q_1} \wedge \eta_{q_3}\}), (\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}\} \vee \{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_3}}\}), \\ & (\{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}}\} \vee \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_3}}\}), (\{\mu_{q_1} \wedge \mu_{q_2}\} \vee \{\mu_{q_1} \wedge \mu_{q_3}\}) \\ (q_1 \cdot q_2) + (q_1 \cdot q_3) =& (-\{\eta_{q_1} \wedge \eta_{q_2}\}, -\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}\}, \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}}\}, \{\mu_{q_1} \wedge \mu_{q_2}\}) \\ & + (-\{\eta_{q_1} \wedge \eta_{q_3}\}, -\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_3}}\}, \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_3}}\}, \{\mu_{q_1} \wedge \mu_{q_3}\}) \\ & = -(\{\eta_{q_1} \wedge \eta_{q_2}\} \vee \{\eta_{q_1} \wedge \eta_{q_3}\}), -(\{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_2}}\} \vee \{\lambda_{\eta_{q_1}} \wedge \lambda_{\eta_{q_3}}\}), \\ & (\{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_2}}\} \vee \{\lambda_{\mu_{q_1}} \wedge \lambda_{\mu_{q_3}}\}), (\{\mu_{q_1} \wedge \mu_{q_2}\} \vee \{\mu_{q_1} \wedge \mu_{q_3}\}) \end{aligned}$$

Thus, $q_1 \cdot (q_2 + q_3) = (q_1 \cdot q_2) + (q_1 \cdot q_3)$
Similarly, $q_1 + (q_2 \cdot q_3) = (q_1 + q_2) \cdot (q_1 + q_3)$

$$\begin{aligned} vi.\ q_1 + 0_q =& (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (0,0,0,0) \\ & = (-\{\eta_{q_1} \vee 0\}, -\{\lambda_{\eta_{q_1}} \vee 0\}, \{\lambda_{\mu_{q_1}} \vee 0\}, \{\mu_{q_1} \vee 0\}) \\ & = (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) \\ & = q_1 \\ q_1 + i_q =& (-\eta_{q_1}, -\lambda_{\eta_{q_1}}, \lambda_{\mu_{q_1}}, \mu_{q_1}) + (-1, -0.5, 0.5, 1) \\ & = (-\{\eta_{q_1} \vee -1\}, -\{\lambda_{\eta_{q_1}} \vee -0.5\}, \{\lambda_{\mu_{q_1}} \vee 0.5\}, \{\mu_{q_1} \vee 1\}) \\ & = (-1, -0.5, 0.5, 1) \\ & = i_q \end{aligned}$$

Similarly, we can prove $q_1 \cdot 0_q = 0_q$ and $q_1.i_q = q_1$. ■ Thus, $(QFS, +, \cdot)$ is Quadrasophic Fuzzy Algebra (QFA).

Example 2: Let us consider three Quadrasophic Fuzzy Sets q_1=$(-0.4, -0.1, 0.3, 0.5)$, $q_2 =$

$(-0.5,-0.3,0.4,0.7)$ and $q_3 = (-0.7,-0.4,0.2,0.6)$

$$\begin{aligned}
i. q_1 + (q_1 \cdot q_2) &= q_1 + (-0.4,-0.1,0.3,0.5) = (-0.4,-0.1,0.3,0.5) \\
&= q_1 \\
ii. q_1 \cdot (q_1 + q_2) &= q_1.(-0.5,-0.3,0.4,0.7) \\
&= (-0.4,-0.1,0.3,0.5) \\
&= q_1. \\
iii.\ q_1 + (q_2 \cdot q_3) &= q_1 + (-0.5,-0.3,0.2,0.6) \\
&= (-0.5,-0.3,0.3,0.6) \\
iv. (q_1 + q_2) \cdot (q_1 + q_3) &= (-0.5,-0.3,0.4,0.7) \cdot (-0.7,-0.4,0.3,0.6) \\
&= (-0.5,-0.3,0.3,0.6)
\end{aligned}$$

Hence, $q_1 + (q_2 \cdot q_3) = (q_1 + q_2) \cdot (q_1 + q_3)$

5. Quadrasophic Fuzzy Matrix

Let $Q_{r\times s}$ be a Quadrasophic Fuzzy Matrix, where

$$Q_{r\times s} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1s} \\ q_{21} & q_{22} & \cdots & q_{2s} \\ \vdots & & \ddots & \vdots \\ q_{r1} & q_{r2} & \cdots & q_{rs} \end{bmatrix}$$

Here, $q_{ij} = (-\eta_{q_{ij}}, -\lambda_{\eta_{q_{ij}}}, \lambda_{\mu_{q_{ij}}}, \mu_{q_{ij}})$, $(\eta_{q_{ij}}, \mu_{q_{ij}}) \in [0,1]$ represents the positive and negative grade, and $(\lambda_{\eta_{q_{ij}}}, \lambda_{\mu_{q_{ij}}}) \in [0,0.5]$ represents the restricted positive and restricted negative grade of $q_i j$. A Quadrasophic Fuzzy Matrix (QFM) is the matrix defined over the Quadrasophic Fuzzy Algebra.

Zero Quadrasophic Matrix: The zero Quadrasophic Fuzzy Matrix 0_{q_s} of order $s \times s$ is the matrix where all the elements are $0_q = (0,0,0,0)$.

Identity Quadrasophic Matrix: The identity Quadrasophic Matrix I_{q_s} of order $s \times s$ is the matrix where all the diagonal entries are $i_q = (-1,-0.5,0.5,1)$ and all the other entries are $0_q = (0,0,0,0)$.

Remark 1. M_{rs} indicates the set of all QFM's of order $r \times s$, while M_s indicates the set of square QFM's of order $s \times s$.

5.1. Operations on Quadrasophic Fuzzy Matrix

Binary opertors of QFM: Let $Q_1 = (q_{1ij})$ and $Q_2 = (q_{2ij})$ be two QFMs of order $r \times s$. Then the following operations as follows:

$$\begin{aligned}
Q_1 + Q_2 &= (q_{1ij} + q_{2ij})_{r\times s} \\
&= \{-max(\eta_{q_{1ij}}, \eta_{q_{2ij}}), -max(\lambda_{\eta_{q_{1ij}}}, \lambda_{\eta_{q_{2ij}}}), \\
&\quad max(\lambda_{\mu_{q_{1ij}}}, \lambda_{\mu_{q_{2ij}}}), max(\mu_{q_{1ij}}, \mu_{q_{2ij}})\} \\
Q_1 \cdot Q_2 &= (q_{1ij} \cdot q_{2ij})_{r\times s} \\
&= \{-min(\eta_{q_{1ij}}, \eta_{q_{2ij}}), -min(\lambda_{\eta_{q_{1ij}}}, \lambda_{\eta_{q_{2ij}}}), \\
&\quad min(\lambda_{\mu_{q_{1ij}}}, \lambda_{\mu_{q_{2ij}}}), min(\mu_{q_{1ij}}, \mu_{q_{2ij}})\}
\end{aligned}$$

$\odot$ and $\otimes$ operators of QFM: Let $Q_1 = (q_{1ij})_{r\times s}$ and $Q_2 = (q_{2ij})_{s\times t}$ be two QFMs. Then the following operations as follows:

$$\begin{aligned} Q_1 \odot Q_2 &= \left(\sum_{l=1}^{s} q_{1il}\cdot q_{2lj}\right)_{r\times t} \\ &= \left(-max_{l=1}^{s}(min\{\eta_{q_{1il}},\eta_{q_{2lj}}\}), (-max_{l=1}^{s}(min\{\lambda_{\eta_{q_{1il}}},\lambda_{\eta_{q_{2lj}}}\}),\right. \\ &\left. max_{l=1}^{s}(min\{\lambda_{\mu_{q_{1il}}},\lambda_{\mu_{q_{2lj}}}\}), max_{l=1}^{s}(min\{\mu_{q_{1il}},\mu_{q_{2lj}}\})\right) \end{aligned}$$

$$\begin{aligned} Q_1 \otimes Q_2 &= \left(\prod_{l=1}^{s} q_{1il}+q_{2lj}\right)_{r\times t} \\ &= \left(-min_{l=1}^{s}(max\{\eta_{q_{1il}},\eta_{q_{2lj}}\}), (-min_{l=1}^{s}(max\{\lambda_{\eta_{q_{1il}}},\lambda_{\eta_{q_{2lj}}}\}),\right. \\ &\left. min_{l=1}^{s}(max\{\lambda_{\mu_{q_{1il}}},\lambda_{\mu_{q_{2lj}}}\}), min_{l=1}^{s}(max\{\mu_{q_{1il}},\mu_{q_{2lj}}\})\right) \end{aligned}$$

Theorem 2. Let Q_1, Q_2, and Q_3 be three Quadrasophic Fuzzy matrices, and then we have the following results:
i. $Q_1 + Q_2 = Q_2 + Q_1$, $Q_1 \cdot Q_2 = Q_2 \cdot Q_1$
ii. $Q_1 + (Q_2 + Q_3) = (Q_1 + Q_2) + Q_3$, $Q_1 \cdot (Q_2 \cdot Q_3) = (Q_1 \cdot Q_2) \cdot Q_3$
iii. $Q_1 \cdot (Q_2 + Q_3) = (Q_1.Q_2) + (Q_1 \cdot Q_3)$, $Q_1 + (Q_2 \cdot Q_3) = (Q_1 + Q_2) \cdot (Q_1 + Q_3)$
iv. $Q_1 + 0_q = 0_q + Q_1 = Q_1$, $Q_1 \cdot 0_q = 0_q \cdot Q_1 = 0_q$, where 0_q is the zero QFM.

Proof. To prove the results, let us consider the following example:

$$Q_1 = \begin{bmatrix} (-0.3,-0.1,0.3,0.5) & (-0.4,-0.4,0.5,0.7) \\ (-0.2,-0.1,0.4,0.7) & (-0.7,-0.3,0.4,0.8) \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} (-0.4,-0.3,0.4,0.7) & (-0.6,-0.2,0.5,0.8) \\ (-0.6,-0.2,0.3,0.9) & (-0.7,-0.4,0.3,0.5) \end{bmatrix}$$

$$\textit{(i.)}\ Q_1 + Q_2 = \begin{bmatrix} \{-max(0.3,0.4), -max(0.1,0.3), & \{-max(0.4,0.6), -max(0.4,0.2), \\ max(0.3,0.4), max(0.5,0.7)\} & max(0.5,0.5), max(0.7,0.8)\} \\ \{-max(0.2,0.6), -max(0.1,0.2), & \{-max(0.7,0.7), -max(0.3,0.4), \\ max(0.4,0.3), max(0.7,0.9)\} & max(0.4,0.3), max(0.8,0.5)\} \end{bmatrix}$$

$$Q_1 + Q_2 = \begin{bmatrix} (-0.4,-0.3,0.4,0.7) & (-0.6,-0.4,0.5,0.8) \\ (-0.6,-0.2,0.4,0.9) & (-0.7,-0.4,0.4,0.8) \end{bmatrix}$$

$$Q_1 + Q_2 = \begin{bmatrix} \{-max(0.4,0.3), -max(0.3,0.1), & \{-max(0.6,0.4), -max(0.2,0.4), \\ max(0.4,0.3), max(0.7,0.5)\} & max(0.5,0.5), max(0.8,0.7)\} \\ \{-max(0.2,0.6), -max(0.1,0.2), & \{-max(0.7,0.7), -max(0.4,0.3), \\ max(0.4,0.3), max(0.7,0.9)\} & max(0.3,0.4), max(0.5,0.8)\} \end{bmatrix}$$

$$Q_2 + Q_1 = \begin{bmatrix} (-0.4,-0.3,0.4,0.7) & (-0.6,-0.4,0.5,0.8) \\ (-0.6,-0.2,0.4,0.9) & (-0.7,-0.4,0.4,0.8) \end{bmatrix}$$

Hence, $Q_1 + Q_2 = Q_2 + Q_1$

$$Q_1 \cdot Q_2 = \begin{bmatrix} \{-min(0.3,0.4), -min(0.1,0.3), & \{-min(0.4,0.6), -min(0.4,0.2), \\ min(0.3,0.4), min(0.5,0.7)\} & min(0.5,0.5), min(0.7,0.8)\} \\ \{-min(0.2,0.6), -min(0.1,0.2), & \{-min(0.7,0.7), -min(0.3,0.4), \\ min(0.4,0.3), min(0.7,0.9)\} & min(0.4,0.3), min(0.8,0.5)\} \end{bmatrix}$$

$$Q_1 \cdot Q_2 = \begin{bmatrix} (-0.3,-0.1,0.3,0.5) & (-0.4,-0.3,0.5,0.7) \\ (-0.2,-0.1,0.3,0.7) & (-0.7,-0.3,0.3,0.5)] \end{bmatrix}$$

$$Q_2 \cdot Q_1 = \begin{bmatrix} \{-min(0.4,0.3), -min(0.3,0.1), & \{-min(0.6,0.4), -min(0.4,0.3), \\ min(0.4,0.3), min(0.7,0.5)\} & min(0.3,0.4), min(0.5,0.8)\} \\ \{-min(0.6,0.2), -min(0.2,0.1), & \{-min(0.7,0.7), -min(0.4,0.3), \\ min(0.3,0.4), min(0.9,0.7)\} & min(0.3,0.4), min(0.5,0.8)\} \end{bmatrix}$$

$$Q_2 \cdot Q_1 = \begin{bmatrix} (-0.3,-0.1,0.3,0.5) & (-0.4,-0.3,0.5,0.7) \\ (-0.2,-0.1,0.3,0.7) & (-0.7,-0.3,0.3,0.5)] \end{bmatrix}$$

$\implies Q_1 \cdot Q_2 = Q_2 \cdot Q_1$,
Thus, (i.) is proved.
(ii.) The proof is obvious.
(iii.)

$$Q_1 \cdot (Q_2 + Q_3) = \begin{bmatrix} (-0.3,-0.1,0.3,0.5) & (-0.4,-0.3,0.5,0.7) \\ (-0.2,-0.1,0.3,0.7) & (-0.7,-0.3,0.4,0.8) \end{bmatrix}$$

$$(Q_1 \cdot Q_2) + (Q_1 \cdot Q_3) = \begin{bmatrix} (-0.3,-0.1, & (-0.4,-0.3, \\ 0.3,0.5) & 0.5,0.7) \\ (-0.2,-0.1, & (-0.7,-0.3, \\ 0.3,0.7) & 0.3,0.5) \end{bmatrix} + \begin{bmatrix} (-0.3,-0.1, & (-0.4,-0.3, \\ 0.3,0.5) & 0.2,0.7) \\ (-0.2,-0.1, & (-0.7,-0.3, \\ 0.2,0.5) & 0.4,0.8) \end{bmatrix}$$

$$(Q_1 \cdot Q_2) + (Q_1 \cdot Q_3) = \begin{bmatrix} (-0.3,-0.1,0.3,0.5) & (-0.4,-0.3,0.5,0.7) \\ (-0.2,-0.1,0.3,0.7) & (-0.7,-0.3,0.4,0.8) \end{bmatrix}$$

Hence, $Q_1 \cdot (Q_2 + Q_3) = (Q_1 \cdot Q_2) + (Q_1 \cdot Q_3)$.

$$Consider, Q_1 + (Q_2 \cdot Q_3) = \begin{bmatrix} (-0.4,-0.3,0.3,0.6) & (-0.6,-0.4,0.5,0.7) \\ (-0.5,-0.2,0.4,0.7) & (-0.7,-0.4,0.4,0.8) \end{bmatrix}$$

$$(Q_1 + Q_2) \cdot (Q_1 + Q_3) = \begin{bmatrix} (-0.4,-0.3, & (-0.6,-0.4, \\ 0.4,0.7) & 0.5,0.8) \\ (-0.6,-0.2, & (-0.7,-0.4, \\ 0.4,0.9) & 0.4,0.8) \end{bmatrix} + \begin{bmatrix} (-0.7,-0.4, & (-0.6,-0.4, \\ 0.3,0.6) & 0.5,0.7) \\ (-0.5,-0.3, & (-0.8,-0.4, \\ 0.4,0.7) & 0.4,0.9) \end{bmatrix}$$

$$(Q_1 + Q_2) \cdot (Q_1 + Q_3) = \begin{bmatrix} (-0.4,-0.3,0.3,0.6) & (-0.6,-0.4,0.5,0.7) \\ (-0.5,-0.2,0.4,0.7) & (-0.7,-0.4,0.4,0.8) \end{bmatrix}$$

Hence, $Q_1 + (Q_2 \cdot Q_3) = (Q_1 + Q_2) \cdot (Q_1 + Q_3)$ Thus, (iii.) is proved.

$$(iv). Q_1 + 0_q = \begin{bmatrix} \{-max(0.3,0), -max(0.1,0), & \{-max(0.4,0), -max(0.4,0), \\ max(0.3,0), max(0.5,0)\} & max(0.5,0), max(0.7,0)\} \\ \{-max(0.2,0), -max(0.1,0), & \{-max(0.7,0), -max(0.3,0), \\ max(0.4,0), max(0.7,0)\} & max(0.4,0), max(0.8,0)\} \end{bmatrix}$$

$$= \begin{bmatrix} (-0.3,0.1,0.3,0.5) & (-0.4,-0.4,0.5,0.7) \\ (-0.2,-0.1,0.4,0.7) & (-0.7,-0.3,0.4,0.8) \end{bmatrix} = Q_1$$

$$0_q + Q_1 = \begin{bmatrix} (-0.3, 0.1, 0.3, 0.5) & (-0.4, -0.4, 0.5, 0.7) \\ (-0.2, -0.1, 0.4, 0.7) & (-0.7, -0.3, 0.4, 0.8) \end{bmatrix} = Q_1$$

$$Consider,\ Q_1 \cdot 0_q = \begin{bmatrix} \{-min(0.3,0), -min(0.1,0), & \{-min(0.4,0), -min(0.4,0), \\ min(0.3,0), min(0.5,0)\} & min(0.5,0), min(0.7,0)\} \\ \{-min(0.2,0), -min(0.1,0), & \{-min(0.7,0), -min(0.3,0), \\ min(0.4,0), min(0.7,0)\} & min(0.4,0), min(0.8,0)\} \end{bmatrix}$$

$$= \begin{bmatrix} (0,0,0,0) & (0,0,0,0) \\ (0,0,0,0) & (0,0,0,0) \end{bmatrix} = 0_q$$

Hence, $Q_1 \cdot 0_q = 0_q \cdot Q_1 = 0_q$
Thus, (i.), (ii.) , (iii.) and (iv.) are easily verified when Q_1, Q_2, Q_3 are of same order. ■

Theorem 3. Let Q_1, Q_2, and Q_3 be three Quadrasophic Fuzzy matrices, and then we have the following results:
i. $Q_1 \odot (Q_2 \odot Q_3) = (Q_1 \odot Q_2) \odot Q_3$
ii.$Q_1 \otimes (Q_2 \otimes Q_3) = (Q_1 \otimes Q_2) \otimes Q_3$
iii.$Q_1 \odot i_q = i_q \odot Q_1 = i_q$
iv.$Q_1 \otimes i_q = i_q \otimes Q_1 = i_q$
v.$Q_1 \odot Q_2 \neq Q_2 \odot Q_1$
vi.$Q_1 \otimes Q_2 \neq Q_2 \otimes Q_1$
vii.$Q_1 \odot (Q_2 + Q_3) \neq (Q_1 \odot Q_2) + (Q_1 \otimes Q_3)$
viii.$Q_1 \otimes (Q_2 \cdot Q_3) \neq (Q_1 \otimes Q_2) \cdot (Q_1 \otimes Q_3)$

Proof. By using the definitions (i.),(ii.), (iii.) and (iv.) are easily verified.
(v.) Let $Q_1 = (q_{1ij})_{r\times s}$ and $Q_2 = (q_{2ij})_{s\times t}$ be two QFMs. Then $Q_1 \odot Q_2 = (q_{12ij})_{r\times t}$ exist. But $Q_2 \odot Q_1$ does not exist. Suppose Q_1 and Q_2 are square matrices. Then by using previous example we can easily verify the result.

$$Q_1 \odot Q_2 = \begin{bmatrix} (-0.4, -0.2, 0.3, 0.7) & (-0.4, -0.4, 0.3, 0.5) \\ (-0.6, -0.2, 0.4, 0.7) & (-0.7, -0.3, 0.4, 0.7) \end{bmatrix}$$

$$Q_2 \odot Q_1 = \begin{bmatrix} (-0.3, -0.1, 0.4, 0.7) & (-0.6, -0.3, 0.4, 0.8) \\ (-0.3, -0.1, 0.3, 0.5) & (-0.7, -0.3, 0.3, 0.7) \end{bmatrix}$$

Hence, $Q_1 \odot Q_2 \neq Q_2 \odot Q_1$
In similar way, (vi.) $Q_1 \otimes Q_2 \neq Q_2 \otimes Q_1$ inequality holds.
(vii.) By computing,

$$Q_1 \odot (Q_2 + Q_3) = \begin{bmatrix} (-0.4, -0.2, 0.3, 0.7) & (-0.4, -0.4, 0.4, 0.7) \\ (-0.6, -0.2, 0.4, 0.7) & (-0.7, -0.3, 0.4, 0.7) \end{bmatrix}$$

$$(Q_1 \odot Q_2) + (Q_1 \odot Q_3) = \begin{bmatrix} (-0.4, -0.2, & (-0.4, -0.4, \\ 0.3, 0.7) & 0.3, 0.5) \\ (-0.6, -0.2, & (-0.7, -0.3, \\ 0.4, 0.7) & 0.4, 0.7) \end{bmatrix} + \begin{bmatrix} (-0.4, -0.3, & (-0.4, -0.4, \\ 0.3, 0.5) & 0.4, 0.7) \\ (-0.5, -0.3, & (-0.7, -0.3, \\ 0.3, 0.6) & 0.4, 0.8) \end{bmatrix}$$

$$= \begin{bmatrix} (-0.4, -0.3, 0.3, 0.7) & (-0.4, -0.4, 0.4, 0.7) \\ (-0.5, -0.3, 0.4, 0.8) & (-0.7, -0.3, 0.4, 0.8) \end{bmatrix}$$

Hence, $Q_1 \odot (Q_2 + Q_3) \neq (Q_1 \odot Q_2) + (Q_1 \odot Q_3)$
Similarly, (viii.) $Q_1 \otimes (Q_2 \cdot Q_3) \neq (Q_1 \otimes Q_2) \cdot (Q_1 \otimes Q_3)$. holds. ■

6. Identification of Plant diseases by employing a Quadrasophic Fuzzy Matrix

The implementation of the Fuzzy Matrix framework uses several techniques to diagnose a wide range of uncertain problems. A Fuzzy Matrix is a crucial tool for decision-making analysis. In India, the primary factor influencing employment and economic standing is agriculture. Approximately 60% of India's rural community works in agriculture [19]. Moreover, the backbone of the Indian economy is agriculture [19]. The damage that natural catastrophes, plant diseases, and climate change have caused to agriculture. To increase the pace of agricultural production, we might focus on plant diseases among these elements. Insects, bacteria, viruses, fungi, and protozoa can all have an impact on plants. In Tamil Nadu, the most important crops farmed are rice, oilseeds, pulses, sugarcane, cotton, and millets. Tamil Nadu is the third-largest producer of paddy in India among these crops. The main staple food in Tamil Nadu is paddy.

In this section, a new ranking mechanism for the Quadrasophic Fuzzy Matrix is proposed and applied to the detection of agricultural plant disease.

Let us consider four paddy plants $\rho = \{p_1, p_2, p_3, p_4\}$ that are afflicted with bacterial infections, with indications represented by $\delta = \{i_1, i_2, i_3, i_4\}$ are presented in Table 1. Let $\kappa = \{l_1, l_2, l_3, l_4\}$ indicates the collection of paddy illnesses.

Table 1: *Notations of paddy illness and indicators*

Notation	Description
i_1	Spindle shaped spots with brown margin and grey centre.
i_2	yellowing of leaves and wilting
i_3	Irregular greyish brown water -soaked lesions on flag leaf sheath
i_4	Oval or cylindrical dark brown spots with a yellow halo
l_1	Rice blast
l_2	Bacterial leaf blight of rice
l_3	Sheath rot of rice
l_4	Rice brown spot

Procedure:
Step 1: Construct the Quadrasophic Fuzzy Matrix Q_1, illustrating the association between indicators and illnesses of the paddy, and Q_2, which illustrates the association between illnesses and paddy.

$$Q_1 = \begin{bmatrix} (-0.7,-0.3,0.2,0.6) & (-0.6,-0.4,0.3,0.8) & (-0.5,-0.2,0.3,0.8) & (-0.6,-0.2,0.4,0.9) \\ (-0.5,-0.2,0.3,0.7) & (-0.7,-0.3,0.2,0.5) & (-0.7,-0.3,0.2,0.8) & (-0.9,-0.3,0.4,0.6) \\ (-0.6,-0.3,0.4,0.8) & (-0.8,-0.4,0.3,0.7) & (-0.3,-0.2,0.3,0.7) & (-0.7,-0.3,0.4,0.9) \\ (-0.9,-0.4,0.3,0.7) & (-0.3,-0.2,0.4,0.8) & (-0.7,-0.3,0.4,0.9) & (-0.7,-0.3,0.3,0.9) \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} (-0.6,-0.3,0.3,0.7) & (-0.4,-0.2,0.3,0.6) & (-0.3,-0.1,0.2,0.5) & (-0.6,-0.2,0.3,0.8) \\ (-0.8,-0.4,0.2,0.4) & (-0.9,-0.4,0.3,0.7) & (-0.3,-0.1,0.4,0.7) & (-0.5,-0.3,0.4,0.9) \\ (-0.9,-0.3,0.2,0.5) & (-0.7,-0.2,0.2,0.6) & (-0.3,-0.2,0.4,0.7) & (-0.6,-0.3,0.3,0.8) \\ (-0.3,-0.1,0.2,0.4) & (-0.2,-0.1,0.4,0.8) & (-0.3,-0.2,0.4,0.9) & (-0.4,-0.2,0.5,0.9) \end{bmatrix}$$

Step 2: Generate the complement matrices Q_1^c and Q_2^c.

$$Q_1^c = \begin{bmatrix} (-0.3,-0.2,0.3,0.4) & (-0.4,-0.1,0.2,0.2) & (-0.5,-0.3,0.2,0.5) & (-0.4,-0.3,0.1,0.1) \\ (-0.5,-0.3,0.2,0.3) & (-0.3,-0.2,0.3,0.5) & (-0.3,-0.2,0.3,0.2) & (-0.1,-0.2,0.1,0.4) \\ (-0.4,-0.2,0.1,0.2) & (-0.2,-0.1,0.2,0.3) & (-0.7,-0.3,0.2,0.3) & (-0.3,-0.2,0.1,0.1) \\ (-0.1,-0.1,0.2,0.3) & (-0.7,-0.3,0.1,0.2) & (-0.3,-0.2,0.1,0.1) & (-0.3,-0.2,0.2,0.4) \end{bmatrix}$$

$$Q_2^c = \begin{bmatrix} (-0.4,-0.2,0.2,0.3) & (-0.6,-0.3,0.2,0.4) & (-0.7,-0.4,0.3,0.5) & (-0.4,-0.3,0.2,0.2) \\ (-0.2,-0.1,0.3,0.6) & (-0.1,-0.1,0.2,0.3) & (-0.7,-0.4,0.1,0.3) & (-0.5,-0.2,0.1,0.1) \\ (-0.1,-0.2,0.3,0.5) & (-0.3,-0.3,0.3,0.4) & (-0.7,-0.3,0.1,0.3) & (-0.4,-0.2,0.2,0.2) \\ (-0.7,-0.4,0.3,0.6) & (-0.8,-0.4,0.1,0.2) & (-0.7,-0.3,0.1,0.1) & (-0.6,-0.3,0.0,0.1) \end{bmatrix}$$

Step 3: Construct $Q_3 = Q_2 \otimes Q_1$, which describes the relationship between Q_1 and Q_2.

$$Q_3 = \begin{bmatrix} (-0.5,-0.2,0.3,0.7) & (-0.6,-0.2,0.3,0.6) & (-0.3,-0.2,0.3,0.7) & (-0.6,-0.3,0.3,0.6) \\ (-0.6,-0.3,0.2,0.6) & (-0.5,-0.3,0.3,0.7) & (-0.3,-0.2,0.3,0.5) & (-0.7,-0.3,0.4,0.7) \\ (-0.6,-0.2,0.2,0.6) & (-0.6,-0.3,0.2,0.6) & (-0.3,-0.2,0.2,0.5) & (-0.7,-0.3,0.3,0.6) \\ (-0.5,-0.2,0.2,0.6) & (-0.4,-0.2,0.3,0.8) & (-0.3,-0.2,0.3,0.5) & (-0.6,-0.2,0.4,0.8) \end{bmatrix}$$

Step 4: Construct $Q_4 = Q_2^c \otimes Q_1^c$, which provides the relation between Q_1^c and Q_2^c.

$$Q_4 = \begin{bmatrix} (-0.4,-0.2,0.2,0.5) & (-0.4,-0.2,0.2,0.2) & (-0.4,-0.3,0.2,0.2) & (-0.4,-0.3,0.2,0.3) \\ (-0.3,-0.2,0.1,0.3) & (-0.3,-0.1,0.1,0.2) & (-0.3,-0.2,0.1,0.1) & (-0.1,-0.2,0.1,0.3) \\ (-0.3,-0.2,0.1,0.3) & (-0.3,-0.2,0.2,0.2) & (-0.3,-0.2,0.2,0.2) & (-0.3,-0.2,0.1,0.3) \\ (-0.6,-0.3,0.1,0.2) & (-0.7,-0.3,0.1,0.2) & (-0.6,-0.3,0.1,0.1) & (-0.6,-0.3,0.1,0.2) \end{bmatrix}$$

Step 5: By using the definition of QFS $\ominus$ operator, compute $Q_3 \ominus Q_4$.

$$Q_5 = (Q_3 \ominus Q_4) = \begin{bmatrix} (-0.5,0, & (-0.6,0, & (0,0, & (-0.6,0, \\ 0.3,0.7) & 0.3,0.6) & 0,0.7) & 0.3,0.6) \\ (-0.6,-0.3, & (-0.5,-0.3, & (0,0, & (-0.7,-0.3, \\ 0.2,0.6) & 0.3,0.7) & 0.3,0.5) & 0.4,0.7) \\ (-0.6,0, & (-0.6,-0.3, & (0,0, & (-0.7,-0.3, \\ 0.2,0.6) & 0,0.6) & 0,0.5) & 0.3,0.6) \\ (0,0, & (0,0, & (0,0, & (0,0, \\ 0.2,0.6) & 0.3,0.8) & 0.3,0.5) & 0.4,0.8) \end{bmatrix}$$

Step 6: By using the score function, compute the scores of Q_5.

$$Score(Q_5) = \begin{bmatrix} 0.166 & 0.1 & 0.233 & 0.1 \\ -0.03 & 0.066 & 0.266 & 0.033 \\ 0 & -0.1 & 0.166 & -0.033 \\ 0.266 & 0.366 & 0.266 & 0.4 \end{bmatrix}$$

The score values indicated the phase of plant disease. The impact of plant illnesses is identified by the lowest value.
1. i_2 and i_4 illnesses may have an impact on p_1.The impact of illnesses on p_1 can be mitigated by applying appropriate manure or fertilizer, as every value falls within the restricted positive range.
2. p_2 is influenced by i_1's illness. Due to its negative restricted range, the condition of p_2 can be effectively managed to prevent the plant from rotting.
3. p_3 is affected by illnesses i_2 and i_4. Because of its negatively restricted range, p_3 can be most effectively regulated in order to prevent the plant from rotting. However, yield has not increased.
4. The illnesses i_1 and i_3 afflict p_4, with each value falling within a restricted positive range, allowing for the application of appropriate manure or fertilizer to prevent all the illnesses.
Hence, Quadrasophic Fuzzy Matrix is a powerful tool for estimating uncertainty in agriculture, aiding in the identification of illness stages in targeted plant and improving production levels.

7. Conclusion

Certain operations of the Quadrasophic Fuzzy Set are defined in this artifact. The properties of Quadrasophic Fuzzy Algebra are defined. The development of the Quadrasophic Fuzzy Matrix and its properties are facilitated by the Quadrasophic Fuzzy Algebra. Propositions and theorems of the Quadrasophic Fuzzy Matrix are explained through relevant cases. Further, a new ranking

technique utilizing the Quadrasophic Fuzzy Matrix tool is proposed and implemented to identify paddy ailments in the farming industry. The conceptual framework of the Quadrasophic Fuzzy Matrix can be used in decision-making, business statistics, medicine, and operations research. The forthcoming investigation will explore the features of the Quadrasophic Fuzzy Matrix and their uses in various disciplines.

References

[1] Aruna, G., J. Jesintha Rosline., and A. Anthoni Amali (2023). Quadrasophic Fuzzy Sets and its Applications in OCD - Decision Making Analysis. *Tuijin Jishu/Journal of Propulsion Technology* ISSN:1001-4055, Vol. 44 No. 4: 3241-3258.

[2] Aruna, G., and J. Jesintha Rosline.(2023). PROPERTIES OF QUADRASOPHIC FUZZY SET AND ITS APPLICATIONS. *Reliability: Theory & Applications* 18, no. 4 (76) : 208-220.

[3] Thomason, Michael G.(1977). Convergence of powers of a fuzzy matrix. *Journal of mathematical analysis and applications* 57, no. 2: 476-480.

[4] Mondal, Sanjib, and Madhumangal Pal.(2016). Rank of interval-valued fuzzy matrices.*Afrika Matematika* 27 : 97-114.

[5] Pal, M. (2001). Intuitionistic fuzzy determinant. *V. U. J. Phys Sci*7: 87-93.

[6] Shyamal, Amiya K., and Madhumangal Pal. (2004). Two new operators on fuzzy matrices. *Journal of Applied Mathematics and Computing* 15: 91-107.

[7] Cho, Han Hyuk.(1999). Regular fuzzy matrices and fuzzy equations. *Fuzzy sets and systems* 105, no. 3 : 445-451.

[8] Pal, Madhumangal. (2015). Interval-valued fuzzy matrices with interval-valued fuzzy rows and columns. *Fuzzy Information and Engineering* 7, no. 3 : 335-368.

[9] Shyamal, Amiya K., and Madhumangal Pal.(2007). Triangular fuzzy matrices. *Iranian Journal of Fuzzy Systems* 4(1): 75-87.

[10] Pal, Madhumangal.(2016).Fuzzy matrices with fuzzy rows and columns. *Journal of Intelligent & Fuzzy Systems* 30, no. 1 : 561-573.

[11] Mondal, Sanjib, and Madhumangal Pal.(2016). Similarity relations, eigenvalues and eigenvectors of bipolar fuzzy matrix. *Journal of Intelligent & Fuzzy Systems* 30, no. 4 : 2297-2307.

[12] Pal, Madhumangal, and Sanjib Mondal.(2019). Bipolar fuzzy matrices. *Soft computing* 23, no. 20 : 9885-9897.

[13] Dogra, Shovan, and Madhumangal Pal.(2020) Picture fuzzy matrix and its application.*Soft Computing* 24, no. 13: 9413-9428.

[14] GANI, A. NAGOOR, M. AFFROSE BEGUM, and P. MURUGANANTHAM.(2023). AN APPLICATION OF TRIANGULAR FUZZY NUMBER MATRICES WITH TRIPLET OPERATOR IN MEDICAL DIAGNOSIS. .*Advances and Applications in Mathematical Sciences* 22,no. 8: 1715-1720.

[15] Khan, Susanta Kumar, and Madhumangal Pal.(2014). Interval-valued intuitionistic fuzzy matrices. *arXiv preprint* arXiv:1404.6949. https://doi.org/10.48550/arXiv.1404.6949

[16] Meenakshi, A. R., and M. Kaliraja.(2011). An application of interval valued fuzzy matrices in medical diagnosis. *International Journal of Mathematical Analysis* 5, no. 36 : 1791-1802.

[17] Muthuraji, T., and P. Punitha Elizabeth.(2023). Application of bipolar fuzzy matrices in Flood Damage.*In 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE.:pp. 1-5.

[18] Sanchez, Elie.(1979). Inverses of fuzzy relations. Application to possibility distributions and medical diagnosis.*Fuzzy sets and systems* 2, no. 1: 75-86.

[19] https://en.wikipedia.org/wiki/Agriculture_in_India

BAYESIAN APPROACH FOR HEAVY-TAILED MODEL FITTING IN TWO LOMAX POPULATIONS

Vijay Kumar Lingutla[1], Nagamani Nadiminti[2]

•

[1,2]Department of Mathematics, School of Advanced Sciences, VIT-AP University, Inavolu, Beside AP Secretariat, Amaravati AP-522237, India
[1]vijay.lingutla@gmail.com, [2]maniinadiminti999@gmail.com

Abstract

Heavy-tailed data are commonly encountered in various real-world applications, particularly in finance, insurance, and reliability engineering. This study focuses on the Lomax distribution, a powerful tool for modeling heavy-tailed phenomena. We investigate the estimation of parameters in two Lomax populations characterized by a common shape parameter and distinct scale parameters. Our analysis employs both Maximum Likelihood Estimation (MLE) and Bayesian estimation techniques, recognizing the absence of closed-form solutions for the estimators. We utilize the Newton-Raphson method for numerical evaluation of the MLE and implement Lindley's approximation for Bayesian estimators with different priors, under symmetric loss function. Additionally, we estimate posterior densities using Gibbs sampling and bootstrapping methods to manage uncertainty. A Monte Carlo simulation study is conducted to assess the performance of the proposed estimators, providing insights into their behavior under various scenarios. This paper also discusses the application of these methodologies through a real-life example, demonstrating the practical utility of the proposed estimation techniques for analyzing heavy-tailed data.

Keywords: Lomax Distribution, Bayes estimation, Lindley's Approximation, Gibbs Sampling, Bootstrapping.

1. Introduction

In many real-world applications, data often exhibit heavy tails, meaning extreme values occur more frequently than predicted by normal distributions, impacting risk assessment and decision-making in fields such as finance, insurance, and reliability engineering. To accurately model these datasets, the Lomax distribution, also known as the Pareto Type II distribution proposed by Lomax [1], is particularly effective. Its suitability for heavy-tailed data is highlighted by Bryson [2], who emphasized its superiority over traditional distributions like the Exponential, Gamma, or Weibull. The Lomax distribution's flexibility and capacity to model various tail behaviors have led to its widespread use, as demonstrated by Hassan et al. [3] and Aljohani [4] in optimal step stress accelerated life testing and Ijaz [5] in characterizing electronic device lifespans. Moreover, Chakraborty et al. [6] proposed Generalized Lomax Models (GLM) to capture the non-linearities and heavy-tailed nature of complex network degree distributions, further underscoring the Lomax distribution's versatility in addressing real-world challenges.

Building on the importance of the Lomax distribution, researchers have extensively explored the estimation of its parameters using various methodologies. Okasha [7] utilized Bayesian and E-Bayesian methods for estimating the shape parameter, reliability, and hazard functions based on type-II censored data. Fitrilia et al. [8] employed Bayesian and E-Bayesian methods under the balanced square error loss function for estimating the shape parameter with right-censored data. Ellah [9] applied Maximum Likelihood Estimation (MLE) and Bayesian methods,

considering symmetric and asymmetric loss functions, to estimate both parameters, reliability, and hazard functions from recorded values. Hasanain et al. [10] implemented MLE and Bayesian estimation with three distinct loss functions for parameter estimation. Al-Bossly [11] employed MLE, Bayesian, and E-Bayesian methods for estimating the shape parameter while considering six different loss functions. Additionally, Kumari et al. [12] utilized MLE and Bayesian estimation under entropy and precautionary loss functions. These studies collectively contribute to enhancing our understanding of parameter estimation in the context of the Lomax distribution, showcasing a variety of approaches and methodologies.

Despite significant advancements in parameter estimation for single-population models, a notable gap persists in extending these methodologies to more complex scenarios involving two or more Lomax populations. Estimating a common parameter across two or more populations is a widely employed statistical method with diverse applications, aiding in comparative analyses and supporting risk assessment by identifying similarities or differences in variable distributions. The pioneering investigation into estimating the common mean of two normal populations was conducted by Graybill and Deal [13], who introduced a combined estimator that surpasses individual sample means concerning variance, subject to certain constraints on sample sizes. For further insights into estimating the common mean of two or more normal populations, one can refer to Moore and Krishnamoorthy [14], Tripathy and Kumar [15], and the relevant citations therein, which provide valuable perspectives from both classical and decision-theoretic standpoints.

In addition to normally distributed populations, researchers have extensively explored estimating common parameters for non-normally distributed populations. For instance, Ghosh and Razmpour [16] considered two exponential distributions and examined a common location parameter using UMVUE (Uniformly Minimum Variance Unbiased Estimator), Maximum Likelihood Estimation (MLE), and modified MLE approaches. Similarly, Jin and Pal [17] introduced enhanced estimators that surpassed MLE for estimating common location parameters of exponential distributions, utilizing convex loss functions. Azhad et al. [18] delved into several heterogeneous exponential distributions, estimating common location parameters through UMVUE, MLE, and modified MLE approaches. Additionally, Nagamani and Tripathy [19] investigated the estimation of common scale parameters for two Gamma populations, employing both MLE and Bayesian estimation methods, including simulation studies to assess the performance of the proposed methods. In a different context, Nagamani et al. [20] addressed two inverse Gaussian populations and estimated the common dispersion parameter, conducting simulation studies to evaluate their results. These studies collectively contribute to advancing parameter estimation methodologies across diverse distributional settings, offering valuable insights for analyzing various types of data.

In our study, we focus on two Lomax populations characterized by a common shape parameter but distinct scale parameters. To estimate the parameters, we employ both Maximum Likelihood and Bayesian estimation techniques, as closed-form estimators do not exist in our scenario. The numerical evaluation of these estimators is facilitated by the Newton-Raphson technique for Maximum Likelihood Estimation. For Bayesian estimations, we utilize Lindley's approximation with different priors, under symmetric loss function. Additionally, we estimate the posterior densities of Bayesian estimators using Gibbs and bootstrapping sampling methods. Gibbs sampling, a technique for generating samples from a joint distribution, is particularly valuable in Bayesian statistics for handling complex posterior distributions. Conversely, bootstrapping, a resampling method, aids in estimating the sampling distribution of a statistic and can be adapted to the Bayesian context for uncertainty estimation. Both Gibbs sampling and bootstrapping play crucial roles in Bayesian data analysis, offering essential tools for managing complex models and estimating uncertainties. To assess the behavior of various estimates, we conduct a Monte Carlo simulation study employing a well-constructed algorithm.

The writing is organized as follows: In Section 2, we derive the Maximum Likelihood Estimates (MLE) and asymptotic confidence intervals for the scale parameters δ_1, δ_2, and the common shape parameter λ. Section 3 discusses Bayesian estimators for the parameters under symmetric and

asymmetric loss functions, deriving the Bayes estimators using vague priors, Jeffreys priors, and conjugate priors. It is worth noting that none of these estimators has a closed-form expression. To approximate Bayes estimators, we utilize an approximation for the ratio of integrals suggested by Lindley. Sections 4 and 5 cover the generation of posterior densities using Gibbs and bootstrapping algorithms, respectively. Section 6 presents the numerical results and discusses the rigorous simulation analysis comparing all the offered estimators. In Section 7, we provide a real-life example to illustrate the estimation methods for estimating parameters. Finally, in Section 8, the study concludes with some remarks.

2. Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a widely employed method for parameter estimation and inference within statistics. The principal aim of MLE is to identify the parameters that maximize the probability or likelihood of the sample data. This section is dedicated to acquiring the Maximum Likelihood Estimates (MLEs) for the model parameters.

Let us assume that two independent random samples are drawn from two Lomax populations $X_1 = (x_{11}, x_{12}, \ldots, x_{1m})$ and $X_2 = (x_{21}, x_{22}, \ldots, x_{2n})$, of sizes m and n, respectively. These samples share a common shape parameter λ but may have different scale parameters, denoted as δ_1 and δ_2. The two populations are represented as $L(\lambda, \delta_1)$ and $L(\lambda, \delta_2)$, respectively. The corresponding probability density functions are given as:

$$f(x_{1i}, \lambda, \delta_1) = \frac{\lambda}{\delta_1}\left(1 + \frac{x_{1i}}{\delta_1}\right)^{-(\lambda+1)}, \qquad x_{1i} > 0, \quad \lambda > 0, \delta_1 > 0, \tag{1}$$

$$f(x_{2j}, \lambda, \delta_2) = \frac{\lambda}{\delta_2}\left(1 + \frac{x_{2j}}{\delta_2}\right)^{-(\lambda+1)}, \qquad x_{2j} > 0, \quad \lambda > 0, \delta_2 > 0. \tag{2}$$

From equations (1) and (2), the joint likelihood function is obtained as:

$$l(\lambda, \delta_1, \delta_2 | X_1, X_2) = \frac{\lambda^{m+n}}{\delta_1^m \delta_2^n} \prod_{i=1}^{m}\left(1 + \frac{x_{1i}}{\delta_1}\right)^{-(\lambda+1)} \prod_{j=1}^{n}\left(1 + \frac{x_{2j}}{\delta_2}\right)^{-(\lambda+1)}$$

Taking the logarithm of the likelihood function, we obtain the log-likelihood function:

$$L = (m+n)\log\lambda - m\log\delta_1 - n\log\delta_2 - (\lambda+1)\left[\sum_{i=1}^{m}\log\left(1 + \frac{x_{1i}}{\delta_1}\right) + \sum_{j=1}^{n}\log\left(1 + \frac{x_{2j}}{\delta_2}\right)\right]. \tag{3}$$

To find the Maximum Likelihood (ML) estimates of the parameters λ, δ_1, and δ_2, we differentiate the log-likelihood function with respect to each parameter and set the derivatives to zero. This yields a system of three non-linear equations:

$$\begin{aligned}
\frac{\partial L}{\partial \lambda} &= \frac{m+n}{\lambda} - T_1 - T_2 = 0, \\
\frac{\partial L}{\partial \delta_1} &= -\frac{m}{\delta_1} - (\lambda+1)T_1' = 0, \\
\frac{\partial L}{\partial \delta_2} &= -\frac{n}{\delta_2} - (\lambda+1)T_2' = 0.
\end{aligned}$$

Here, T_1 and T_2 represent the summation terms involving the samples from populations X_1 and X_2, and T_1' and T_2' are their first derivatives with respect to δ_1 and δ_2, respectively. Full expressions for T_1, T_2, T_1', and T_2' are provided in Appendix A.

As the system of non-linear equations cannot be solved analytically, we employ the Newton-Raphson method to obtain numerical solutions. The MLE results after solving these equations are presented in Section 6.

Following this, we calculate the Fisher information matrix for the model parameters λ, δ_1, and δ_2:

$$I(\lambda,\delta_1,\delta_2)=\begin{bmatrix} \frac{-(m+n)}{\lambda^2} & -T_1' & -T_2' \\ -T_1' & \frac{m}{\delta_1^2}-(\lambda+1)T_1'' & 0 \\ -T_2' & 0 & \frac{n}{\delta_2^2}-(\lambda+1)T_2'' \end{bmatrix}.$$

Here, T_1'' and T_2'' denote the second derivatives of T_1 with respect to δ_1 and T_2 with respect to δ_2, respectively. Detailed expressions for T_1'', T_2'', and d are also provided in Appendix A.

Using the information matrix, we construct 95% asymptotic confidence intervals for the model parameters as follows:

$$\hat{\lambda}_{ML}\pm 1.96\sqrt{\frac{1}{d}\left(\frac{m}{\delta_1^2}-(\lambda+1)T_1''\right)\left(\frac{n}{\delta_2^2}-(\lambda+1)T_2''\right)},$$

$$\hat{\delta}_{1ML}\pm 1.96\sqrt{\frac{1}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{n}{\delta_2^2}-(\lambda+1)T_2''\right)+(T_2')^2\right]},$$

$$\hat{\delta}_{2ML}\pm 1.96\sqrt{\frac{1}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{m}{\delta_1^2}-(\lambda+1)T_1''\right)+(T_1')^2\right]}.$$

Numerical results for these confidence intervals, obtained using fixed sample sizes, are presented in Section 6.

3. Bayesian Study

In recent decades, the Bayesian perspective has gained significant attention for statistical inference, offering a powerful and valid alternative to classical statistical methods. This section considers the Bayesian estimation of the model's parameters. The Bayes estimator is particularly useful when there is prior knowledge about the distribution of parameters. Let $\rho_1(\lambda)$ be the prior density function of the shape parameter λ, and let $\rho_2(\delta_1)$ and $\rho_3(\delta_2)$ be the prior densities for the scale parameters δ_1 and δ_2, respectively.

The likelihood function of $(\lambda,\delta_1,\delta_2)$ for the given data (X_1,X_2) is obtained as:

$$l(\lambda,\delta_1,\delta_2|X_1,X_2)=\frac{\lambda^{m+n}}{\delta_1^m\delta_2^n}\prod_{i=1}^{m}\left(1+\frac{x_{1i}}{\delta_1}\right)^{-(\lambda+1)}\prod_{j=1}^{n}\left(1+\frac{x_{2j}}{\delta_2}\right)^{-(\lambda+1)}.$$

We can obtain the joint density function of $(\lambda,\delta_1,\delta_2,X_1,X_2)$ by combining the likelihood and the priors, as follows:

$$f(\lambda,\delta_1,\delta_2,X_1,X_2)=\frac{\lambda^{m+n}}{\delta_1^m\delta_2^n}\prod_{i=1}^{m}\left(1+\frac{x_{1i}}{\delta_1}\right)^{-(\lambda+1)}\prod_{j=1}^{n}\left(1+\frac{x_{2j}}{\delta_2}\right)^{-(\lambda+1)}\rho_1(\lambda)\rho_2(\delta_1)\rho_3(\delta_2).$$

The posterior joint density function of $(\lambda,\delta_1,\delta_2)$ for (X_1,X_2) is:

$$f(\lambda,\delta_1,\delta_2|X_1,X_2)=\frac{f(\lambda,\delta_1,\delta_2,x_1,x_2)}{\int_0^\infty\int_0^\infty\int_0^\infty f(\lambda,\delta_1,\delta_2,x_1,x_2)d\lambda d\delta_1 d\delta_2}.$$

The posterior expectation of $g(\lambda,\delta_1,\delta_2)$ is given by:

$$E[g(\lambda,\delta_1,\delta_2)|X_1,X_2] = \frac{\int_0^\infty\int_0^\infty\int_0^\infty g(\lambda,\delta_1,\delta_2)f(\lambda,\delta_1,\delta_2,x_1,x_2)d\lambda d\delta_1 d\delta_2}{\int_0^\infty\int_0^\infty\int_0^\infty f(\lambda,\delta_1,\delta_2,x_1,x_2)d\lambda d\delta_1 d\delta_2}. \quad (4)$$

It is challenging to calculate the ratio of integrals in equation (4) using analytical methods. However, certain approximations can be used to obtain a numerical value. To calculate the ratio, we employ the method proposed by Lindley [21], which is explained in detail below. Moreover, by using different priors and loss functions for the parameters, Bayes estimators can be derived.

3.1. Lindley's Approximation

In Bayesian analysis, we frequently encounter the problem of the ratio of integrals. Lindley [21] proposed an asymptotic solution for the ratio of two integrals. We use this method to evaluate the expression in equation (4). Lindley's method allows us to approximate expressions such as:

$$I = \frac{\int \mu(\theta) v(\theta) \exp L(\theta) d\theta}{\int v(\theta) \exp L(\theta) d\theta} = E\left[\mu(\theta)|x\right], \tag{5}$$

where $L(\theta)$ is the log-likelihood function of the data $X = (x_1, x_2, ..., x_n)$, $\theta = (\theta_1, \theta_2, ..., \theta_m)$, $\mu(\theta)$ is any function of θ, and $v(\theta)$ is the prior function of θ. Lindley's approximation to equation (5) is given by:

$$E\left[\mu(\theta)|x\right] = \left[\mu + \frac{1}{2}\sum_i\sum_j(\mu_{ij} + 2\mu_i\rho_j)\sigma_{ij} + \frac{1}{2}\sum_i\sum_j\sum_k\sum_r L_{ijk}\sigma_{ij}\sigma_{kr}\mu_r\right]_{\hat{\theta}_{ML}} + O\left(\frac{1}{n^2}\right), \tag{6}$$

where μ or $\mu(\theta)$ is any function of θ, μ_i is the partial derivative of μ with respect to θ_i, μ_{ij} represents the second partial derivative of the function μ with respect to the parameters θ_i and θ_j, L_{ijk} represents the third partial derivative of the function L with respect to the parameters θ_i, θ_j, and θ_k, σ_{ij} represents the $(i,j)^{\text{th}}$ element of the matrix $[-L_{ij}]^{-1}$, and $\hat{\theta}$ is the MLE of θ. All terms are evaluated at the MLEs of θ. Further, $\rho(\theta) = \log[v(\theta)]$ and ρ_j is the partial derivative of ρ with respect to θ_j.

In the subsequent section, Lindley's approximation method is employed to derive Bayes estimators for the parameters λ, δ_1, and δ_2 under symmetric loss function. The primary role of a loss function is to assess the efficacy of a model by assigning penalties based on the extent of deviation between predictions and true values. Using equation (6), we can obtain Bayes estimators for the parameters $(\lambda, \delta_1, \delta_2)$ under symmetric (Squared Error) loss function.

3.2. Symmetric loss function

In this section, we obtain the Bayes estimators under the Symmetric(SE) Loss function, after ignoring the terms of order $\frac{1}{(m+n)^2}$ and smaller, the expression in (6), reduces to

$$\begin{aligned} E[\mu(\theta)|(x_1,x_2)] &= \mu + \mu_1 a_1 + \mu_2 a_2 + \mu_3 a_3 + a_4 + a_5 + \frac{1}{2}[A(\mu_1\sigma_{11} + \mu_2\sigma_{12} + \mu_3\sigma_{13}) \\ &\quad + B(\mu_1\sigma_{21} + \mu_2\sigma_{22} + \mu_3\sigma_{23}) + D(\mu_1\sigma_{31} + \mu_2\sigma_{32} + \mu_3\sigma_{33})]. \end{aligned} \tag{7}$$

In our notation, we have $\theta = (\theta_1, \theta_2, \theta_3) = (\lambda, \delta_1, \delta_2)$. For more details, refer to Tripathy, & Nagamani[22]. We obtain the Bayesian estimators of the parameters using various priors under symmetric loss function in the subsequent sections.

3.2.1 Vague Prior

We use Vague prior for the parameters λ, δ_1, and δ_2 to estimate the Bayesian estimators. The prior densities shape λ and scale δ_1 and δ_2 parameters is considered as.

$$\rho_1(\lambda) = 1, \;\; \rho_2(\delta_1) = \frac{1}{\delta_1^2}, \;\; \rho_3(\delta_2) = \frac{1}{\delta_2^2}$$

We can derive the joint prior density for the parameters λ, δ_1, and δ_2 by combining their individual prior densities. This can be expressed as follows:

$$\begin{aligned} v_V(\lambda, \delta_1, \delta_2) &= \frac{1}{\delta_1^2 \delta_2^2} \\ \rho(\theta) = log v(\theta) &= -2log\delta_1 - 2log\delta_2. \end{aligned}$$

From $\rho(\theta)$ we get ρ_1, ρ_2, ρ_3, a_1, a_2, a_3 and the details of the notations are provided in Appendix [B]

Let $\mu(\theta) = \lambda$; then $\mu_1 = 1$, $\mu_2 = \mu_3 = 0$, $\mu_{ij} = 0$, i, j = 1, 2, 3, $a_4 = 0$, $a_5 = 0$. These values, when substituted into (7), give the Bayes estimator for λ.

$$E(\lambda|(x_{1i}, x_{2j})) = \hat{\lambda}_{ML} - \left[\frac{2}{\hat{\delta}_{1ML}}\sigma_{12} + \frac{2}{\hat{\delta}_{2ML}}\sigma_{13}\right] + \frac{1}{2}\left[A\sigma_{11} + B\sigma_{21} + D\sigma_{31}\right] \quad (8)$$

Consider $\mu(\theta) = \delta_1$; then $\mu_2 = 1$, $\mu_1 = \mu_3 = 0$, $\mu_{ij} = 0$, i, j = 1, 2, 3, and $a_4 = 0$, $a_5 = 0$. We can obtain the Bayesian estimator for δ_1 by substituting these values into (7) as follows:

$$E(\delta_1|(x_{1i}, x_{2j})) = \hat{\delta}_{1ML} - \left[\frac{2}{\hat{\delta}_{1ML}}\sigma_{22} + \frac{2}{\hat{\delta}_{2ML}}\sigma_{23}\right] + \frac{1}{2}\left[A\sigma_{12} + B\sigma_{22} + D\sigma_{32}\right] \quad (9)$$

Again, consider $\mu(\theta) = \delta_2$; then $\mu_3 = 1$, $\mu_1 = \mu_2 = 0$, $\mu_{ij} = 0$, i, j = 1, 2, 3, and $a_4 = 0$, $a_5 = 0$. We can obtain the Bayesian estimator for δ_2 by substituting these values into (7) as follows:

$$E(\delta_2|(x_{1i}, x_{2j})) = \hat{\delta}_{2ML} - \left[\frac{2}{\hat{\delta}_{1ML}}\sigma_{32} + \frac{2}{\hat{\delta}_{2ML}}\sigma_{33}\right] + \frac{1}{2}\left[A\sigma_{13} + B\sigma_{23} + D\sigma_{33}\right] \quad (10)$$

All the terms A, B, D, and σ_{ij}'s are provided in Appendix [B], and these notations remain consistent throughout the subsequent derivations.

3.2.2 Jeffrey's Prior

Here, we utilize Jeffrey's prior to formulate Bayes estimators for the parameters λ, δ_1, and δ_2. Jeffrey initially developed this prior using the Fisher information matrix, denoted as $I(\lambda, \delta_1, \delta_2)$, and it is expressed as:

$$v_J(\lambda, \delta_1, \delta_2) \propto \sqrt{\det(I(\lambda, \delta_1, \delta_2))},$$

$$v_J = \sqrt{\left(-\frac{d_1 d_2 (m+n)}{\lambda^2} - d_1 (T_2')^2 - d_2 (T_1')^2\right)},$$

$$\rho(\theta) = \frac{1}{2}\log\left[-\frac{d_1 d_2 (m+n)}{\lambda^2} - d_1 (T_2')^2 - d_2 (T_1')^2\right],$$

From $\rho(\theta)$, we obtain ρ_1, ρ_2, and ρ_3. The notations d_1, d_2, ρ_1, ρ_2, and ρ_3 are given in Appendix [C].

Let $\mu(\theta) = \lambda$; then $\mu_1 = 1$, $\mu_2 = \mu_3 = 0$, $\mu_{ij} = 0$ for $i, j = 1, 2, 3$, $a_4 = 0$, $a_5 = 0$. These values, when substituted into (7), give the Bayes estimator for λ:

$$E(\lambda|(x_{1i}, x_{2j})) = \hat{\lambda}_{ML} + [p_1\sigma_{11} + p_2\sigma_{12} + p_3\sigma_{13}] + \frac{1}{2}\left[A\sigma_{11} + B\sigma_{21} + D\sigma_{31}\right] \quad (11)$$

Next, consider $\mu(\theta) = \delta_1$; then $\mu_2 = 1$, $\mu_1 = \mu_3 = 0$, $\mu_{ij} = 0$ for $i, j = 1, 2, 3$, and $a_4 = 0$, $a_5 = 0$. Substituting these values into (7) gives the Bayes estimator for δ_1:

$$E(\delta_1|(x_{1i}, x_{2j})) = \hat{\delta}_{1ML} + [p_1\sigma_{21} + p_2\sigma_{22} + p_3\sigma_{23}] + \frac{1}{2}\left[A\sigma_{12} + B\sigma_{22} + D\sigma_{32}\right] \quad (12)$$

Finally, for $\mu(\theta) = \delta_2$, with $\mu_3 = 1$, $\mu_1 = \mu_2 = 0$, $\mu_{ij} = 0$ for $i, j = 1, 2, 3$, and $a_4 = 0$, $a_5 = 0$, the Bayes estimator for δ_2 is obtained as follows:

$$E(\delta_2|(x_{1i}, x_{2j})) = \hat{\delta}_{2ML} + [p_1\sigma_{31} + p_2\sigma_{32} + p_3\sigma_{33}] + \frac{1}{2}\left[A\sigma_{13} + B\sigma_{23} + D\sigma_{33}\right] \quad (13)$$

3.2.3 Conjugate Prior

In this context, for the parameters λ, δ_1, and δ_2 we estimate Byes estimators using conjugate priors. For the shape parameter, we employ a gamma prior, and for the scale parameters, we employ inverse gamma priors, with their respective probability density functions given below:

$$\rho_1(\lambda) = \frac{b_1^{c_1}}{\Gamma(c_1)}\lambda^{c_1-1}e^{-b_1\lambda},\ \ \rho_2(\delta_1) = \frac{b_2^{c_2}}{\Gamma(c_2)}\delta_1^{-(c_2+1)}e^{\frac{-b_2}{\delta_1}},\ \ \rho_3(\delta_2) = \frac{b_3^{c_3}}{\Gamma(c_3)}\delta_2^{-(c_3+1)}e^{\frac{-b_3}{\delta_2}}$$

We can derive the joint prior density for the parameters λ, δ_1, and δ_2 by combining their individual prior densities. This can be expressed as follows:

$$v_C(\lambda,\delta_1,\delta_2) = \frac{b_2^{c_2}}{\Gamma(c_2)}\frac{b_2^{c_2}}{\Gamma(c_2)}\frac{b_3^{c_3}}{\Gamma(c_3)}\lambda^{c_1-1}\delta_1^{-(c_2+1)}\delta_2^{-(c_3+1)}e^{-b_1\lambda-\frac{b_2}{\delta_1}-\frac{b_3}{\delta_2}}$$

$$\begin{aligned}\rho(\theta) = logv(\theta) &= c_1logb_1 + c_2logb_2 + c_3logb_3 - log\Gamma(c_1) - log\Gamma(c_2) - log\Gamma(c_3)\\ &+(c_1-1)log\lambda - (c_2-1)log\delta_1 - (c_3-1)log\delta_2 - b_1\lambda - \frac{b_2}{\delta_1} - \frac{b_3}{\delta_2}\end{aligned}$$

From $\rho(\theta)$ we get ρ_1, ρ_2, and ρ_3. The detailed notations for ρ_1, ρ_2, and ρ_3, a_1, a_2, a_3 are given in [D]

Let $\mu(\theta) = \lambda$; then $\mu_1 = 1$, $\mu_2 = \mu_3 = 0$, $\mu_{ij} = 0$, i, j = 1, 2, 3, $a_4 = 0$, $a_5 = 0$. These values, then substituted into (7), give the Bayes estimator for λ.

$$\begin{aligned}E(\lambda|(x_{1i},x_{2j})) &= \hat{\lambda}_{ML} + \left[\left(\frac{c_1-1}{\hat{\lambda_{ML}}} - b_1\right)\sigma_{11} + \left(\frac{b_2}{\hat{\delta}^2_{1ML}} - \frac{c_2+1}{\hat{\delta}_{1ML}}\right)\sigma_{12} + \left(\frac{b_3}{\hat{\delta}^2_{2ML}} - \frac{c_3+1}{\hat{\delta}_{2ML}}\right)\sigma_{13}\right]\\ &+\frac{1}{2}\left[A\sigma_{11} + B\sigma_{21} + D\sigma_{31}\right] \qquad (14)\end{aligned}$$

Consider $\mu(\theta) = \delta_1$; then $\mu_2 = 1$, $\mu_1 = \mu_3 = 0$, $\mu_{ij} = 0$, i, j = 1, 2, 3, and $a_4 = 0$, $a_5 = 0$. These values, then substituted into (7), give the Bayes estimator for δ_1.

$$\begin{aligned}E(\delta_1|(x_{1i},x_{2j})) &= \hat{\delta_1}_{ML} + \left[\left(\frac{c_1-1}{\hat{\lambda_{ML}}} - b_1\right)\sigma_{21} + \left(\frac{b_2}{\hat{\delta}^2_{1ML}} - \frac{c_2+1}{\hat{\delta}_{1ML}}\right)\sigma_{22} + \left(\frac{b_3}{\hat{\delta}^2_{2ML}} - \frac{c_3+1}{\hat{\delta}_{2ML}}\right)\sigma_{23}\right]\\ &+\frac{1}{2}\left[A\sigma_{12} + B\sigma_{22} + D\sigma_{32}\right] \qquad (15)\end{aligned}$$

Again, consider $\mu(\theta) = \delta_2$; then $\mu_3 = 1$, $\mu_1 = \mu_2 = 0$, $\mu_{ij} = 0$, i, j = 1, 2, 3, and $a_4 = 0$, $a_5 = 0$. These values, then substituted into (7), give the Bayes estimator for δ_2.

$$\begin{aligned}E(\delta_2|(x_{1i},x_{2j})) &= \hat{\delta_{2ML}} + \left[\left(\frac{c_1-1}{\hat{\lambda_{ML}}} - b_1\right)\sigma_{31} + \left(\frac{b_2}{\hat{\delta}^2_{1ML}} - \frac{c_2+1}{\hat{\delta}_{1ML}}\right)\sigma_{32} + \left(\frac{b_3}{\hat{\delta}^2_{2ML}} - \frac{c_3+1}{\hat{\delta}_{2ML}}\right)\sigma_{33}\right]\\ &+\frac{1}{2}\left[A\sigma_{13} + B\sigma_{23} + D\sigma_{33}\right] \qquad (16)\end{aligned}$$

4. Gibbs Sampling

Gibbs sampling is a method for generating samples from a joint probability distribution by iteratively sampling from the conditional distributions of each variable while keeping others fixed. This technique is particularly useful when direct sampling from the joint distribution is challenging. Gibbs sampling is prevalent in Bayesian statistics, probabilistic modeling, and fields requiring sampling from complex multivariate distributions.

Working rule for Gibbs sampling:

1. Start with initial values for the parameters.
2. Define the Prior, $P(\lambda, \delta_1, \delta_2)$, and the Likelihood, $P(\lambda, \delta_1, \delta_2 | X_1, X_2)$.
3. Generate the Joint Posterior Density function using the defined Prior and Likelihood.
4. Randomly draw parameters from conditional densities as follows:
   - Draw λ from $P(\lambda | \delta_1, \delta_2, X_1, X_2)$ using current values of δ_1, δ_2, X_1, and X_2.
   - Draw δ_1 from $P(\delta_1 | \lambda, \delta_2, X_1, X_2)$ using current values of λ, δ_2, X_1, and X_2.
   - Draw δ_2 from $P(\delta_2 | \lambda, \delta_1, X_1, X_2)$ using current values of λ, δ_1, X_1, and X_2.
5. Iterate through the above steps for N times to obtain N draws of the parameters.
6. After obtaining N draws, calculate Highest Posterior Density (HPD) intervals for each parameter, representing the credible range of values.
7. To assess convergence and stability, calculate the average for each parameter and compare it with the initial values of $(\lambda, \delta_1, \delta_2)$.

5. Bootstrapping

Bootstrapping is a method that involves repeated resampling of a single dataset, allowing the creation of multiple simulated samples used to compute standard errors, confidence intervals, and conduct hypothesis tests. In Bayesian statistics, bootstrapping can extend to resampling from posterior samples obtained through methods like Markov Chain Monte Carlo (MCMC). This approach provides a means to estimate uncertainty in Bayesian inference by generating simulated datasets through resampling with replacement. These datasets are then utilized to compute summary statistics or parameters of interest, such as standard errors or confidence intervals, enhancing the robustness of uncertainty assessment in Bayesian estimates.

Steps in Bootstrapping:

1. Set the initial values for the parameters $(\lambda, \delta_1, \delta_2)$.
2. Define the data (X_1, X_2) from Lomax distributions with parameters $(\lambda, \delta_1, \delta_2)$.
3. Choose the number of bootstrap samples $N = 10000$.
4. Draw N bootstrap samples D_k (for $k = 1, 2, \ldots, N$) by randomly sampling with replacement from the observed dataset (X_1, X_2).
5. For each bootstrap sample D_k, conduct Bayesian parameter estimation to obtain posterior samples of the parameters using the Joint Posterior Density function with Prior $P(\lambda, \delta_1, \delta_2)$ and Likelihood $P(\lambda, \delta_1, \delta_2 | X_1, X_2)$.
6. Calculate Highest Posterior Density (HPD) intervals for each parameter based on the obtained posterior samples, representing the credible range of values for the parameters.
7. To assess convergence and stability, calculate the average for each parameter and compare it with the initial values of $(\lambda, \delta_1, \delta_2)$.

6. Simulation Study

This study focuses on estimating parameters of two Lomax distributions, assuming a common shape parameter (λ) and potentially two distinct scale parameters (δ_1 and δ_2). Maximum Likelihood Estimators (MLEs) for the scale parameters and the shape parameter were computed in Section 2, utilizing computational techniques to compare these estimators numerically through simulations. In Section 3, we delve into the development of Bayes estimators. While these estimators lack a precise analytical form, we derive parameter approximations based on Lindley's method. Various priors, including the Vague Prior, Jeffreys Prior, and Conjugate Prior, are employed to calculate these estimators and assess their performance under symmetric (as described in Sections 8 to 16) loss function.

In Sections 4 and 5, we further include Bayesian estimators estimated using Gibbs and Bootstrapping algorithms to compare these results with the parameters obtained using MLE and Lindley's approximation. Additionally, 95% Highest Posterior Density (HPD) credible intervals for the parameters are estimated using Gibbs and Bootstrapping, facilitating comparisons with 95% asymptotic confidence intervals. Trace plots and density plots are also generated to evaluate the performance and convergence of the MCMC chains.

The performance of these estimators is evaluated using bias and mean square error (MSE) metrics. To quantitatively compare these estimators, we generate $10,000$ random samples from two Lomax populations across various sample sizes and parameter combinations. Specific hyperparameters ($c_1 = c_2 = c_3 = 1.5$ and $b_1 = b_2 = b_3 = 0.5$) are employed to calculate biases and MSE for all estimators, with results presented in tabular form in Tables 1 to 2.

Table 1 presents the bias and MSE of MLE and Bayes estimators under a symmetric loss function. The first column denotes various sample sizes, while the second column represents the parameters λ, δ_1, and δ_2. Columns 3 and 4 display the bias and MSE of the MLE estimates, respectively. The subsequent columns represent the bias and MSE of Bayes estimators using Jeffreys prior in columns 5 and 6, Conjugate prior in columns 7 and 8, Gibbs in columns 9 and 10, and Bootstrapping in columns 11 and 12. Table 2 displays 95% asymptotic and HPD intervals using Gibbs and Bootstrapping for the parameters. The first two columns represent sample sizes and parameters λ, δ_1, and δ_2. The third column indicates asymptotic confidence intervals estimated using the information matrix. The fourth and fifth columns denote the HPD intervals estimated using Gibbs and Bootstrapping algorithms.

The observations derived from our simulation study provide valuable insights into the performance of different estimators under varying conditions:

- As sample sizes increase, both bias and mean square error for each estimator decrease.
- Estimators of the shape parameter obtained using MLE, Lindley's method, Gibbs sampling, and Bootstrapping converge to constant values with increasing sample sizes, indicating consistency. The same trend is observed for scale parameters.
- For small sample sizes, Gibbs estimates outperform MLE and Bootstrapping in terms of bias and mean square error for parameters λ, δ_1, and δ_2.
- Bayes estimation with informative priors yields minimum error compared to non-informative priors like Vague and Jeffreys under symmetric loss functions.
- Lindley's approximation, employing a conjugate prior in Bayes estimation, demonstrates superior performance compared to ML estimators, Gibbs sampling, and Bootstrapping under symmetric loss function.
- Highest Posterior Density (HPD) confidence intervals obtained through Gibbs sampling tend to be more precise than Bootstrapping and traditional asymptotic confidence intervals.

Table 1: *For various sample sizes, we compare biases and mean square errors of multiple estimators under squared error loss for* $\theta = (\lambda, \delta_1, \delta_2)$.

		MLE		**Bayes (J)**		**Bayes(C)**		**Gibbs**		**Bootstrap**	
m,n	$\theta \downarrow$	**Bias**	**Mse**	**Bias**	**Mse**	**Bias**	**Mse**	**Bias**	**Mse**	**Bias**	**Mse**
	1.5	-0.14	0.264	0.022	0.316	0.076	0.244	-0.041	0.148	0.26	0.316
	1	0.209	0.26	0.209	0.26	0.041	0.116	-0.139	0.110	0.036	0.232
	2	0.469	1.087	-0.196	0.221	-0.173	0.151	0.096	0.552	-0.228	0.501
	2	-0.196	0.45	0.019	0.531	0.056	0.404	-0.287	0.294	0.006	0.672
(10,10)	1	0.215	0.247	0.215	0.247	0.069	0.122	-0.084	0.107	0.173	0.298
	2	0.456	1.058	-0.198	0.207	-0.122	0.147	0.001	0.484	-0.44	0.73
	2.5	-0.266	0.719	0	0.833	0.02	0.638	0.077	0.488	0.125	0.862
	1	0.232	0.269	0.232	0.269	0.095	0.141	0.023	0.112	0.024	0.15
	2	0.503	1.289	-0.194	0.213	-0.091	0.161	0.022	0.426	0.3	0.854
	1.5	-0.076	0.099	-0.01	0.104	0.011	0.093	-0.131	0.074	-0.029	0.102
	1	0.081	0.068	0.081	0.068	0.024	0.051	-0.173	0.073	0.146	0.147
	2	0.182	0.257	-0.022	0.139	0.004	0.132	-0.159	0.203	-0.497	0.418
	2	-0.078	0.18	0.012	0.193	0.023	0.172	-0.441	0.269	-0.043	0.263
(20,30)	1	0.08	0.071	0.08	0.071	0.029	0.054	-0.008	0.059	0.091	0.169
	2	0.163	0.257	-0.037	0.146	0.004	0.142	-0.274	0.225	-0.177	0.296
	2.5	-0.123	0.295	-0.011	0.311	-0.008	0.277	-0.362	0.267	0.124	0.329
	1	0.083	0.07	0.083	0.07	0.035	0.054	-0.15	0.059	0.197	0.125
	2	0.193	0.263	-0.013	0.142	0.038	0.144	-0.259	0.202	0.203	0.313
	1.5	-0.03	0.06	0.009	0.063	0.02	0.059	0.11	0.057	0.255	0.098
	1	0.039	0.028	0.039	0.028	0.016	0.024	0.055	0.035	0.18	0.103
	2	0.078	0.126	-0.017	0.098	0.005	0.097	-0.23	0.151	0.309	0.391
	2	-0.051	0.098	0.001	0.101	0.007	0.095	-0.18	0.093	0.311	0.298
(50,40)	1	0.037	0.025	0.037	0.025	0.017	0.022	0.001	0.029	0.143	0.089
	2	0.097	0.136	-0.001	0.101	0.029	0.105	-0.207	0.142	0.133	0.239
	2.5	-0.042	0.165	0.023	0.174	0.024	0.163	0.214	0.194	0.171	0.227
	1	0.036	0.026	0.036	0.026	0.018	0.023	0.074	0.037	0	0.036
	2	0.091	0.124	-0.006	0.094	0.029	0.098	0.144	0.153	0.071	0.164
	1.5	-0.016	0.042	0.014	0.044	0.022	0.042	-0.022	0.031	-0.092	0.05
	1	0.03	0.019	0.03	0.019	0.012	0.017	0	0.025	0	0.04
	2	0.047	0.08	-0.023	0.067	-0.008	0.067	-0.13	0.104	-0.222	0.156
	2	-0.038	0.083	0.001	0.085	0.005	0.081	0.055	0.061	-0.18	0.121
(60,60)	1	0.034	0.021	0.034	0.021	0.018	0.019	0.001	0.022	-0.099	0.043
	2	0.068	0.088	-0.004	0.071	0.017	0.072	0.073	0.099	-0.022	0.16
	2.5	-0.053	0.121	-0.005	0.123	-0.004	0.117	-0.253	0.132	0.106	0.186
	1	0.031	0.021	0.031	0.021	0.017	0.019	-0.044	0.021	0.045	0.033
	2	0.078	0.096	0.005	0.077	0.03	0.079	-0.116	0.085	0.419	0.43

Table 2: 95% *Asymptotic, HPD intervals using Gibbs and Bootstrap confidence interval for the parameters* $\theta = (\lambda, \delta_1, \delta_2)$ *at various sample sizes*

(m,n)	$\theta \downarrow$	**Asymptotic**	**Gibbs**	**Bootstrap**
	$\delta_1 = 1$	[0.372,1.677]	[0.299,1.433]	[0.385,1.631]
	$\delta_2 = 2$	[1.42,3.521]	[0.804,3.267]	[0.676,2.804]
(10,10)	$\lambda = \{1.5$	[0.562,2.974]	[0.773,2.218]	[1.204,3.642]
	2	[0.447,2.131]	[0.874,2.615]	[1.227,3.527]
	2.5$\}$	[0.837,4.003]	[1.32,3.96]	[1.128,3.484]
	$\delta_1 = 1$	[0.456,1.292]	[0.621,2.021]	[0.831,2.577]
	$\delta_2 = 2$	[1.745,2.361]	[0.62,2.616]	[0.898,2.938]
(20,10)	$\lambda = \{1.5$	[0.705,2.535]	[0.802,1.961]	[1.261,3.168]
	2	[1.281,4.627]	[0.984,2.443]	[1.719,4.321]
	2.5$\}$	[0.978,3.431]	[1.374,3.36]	[1.426,3.486]
	$\delta_1 = 1$	[0.595,1.559]	[0.622,1.832]	[0.435,1.282]
	$\delta_2 = 2$	[1.887,3.025]	[0.871,2.619]	[0.791,2.378]
(20,20)	$\lambda = \{1.5$	[0.736,2.05]	[0.79,1.709]	[1.042,2.35]
	2	[1.225,3.476]	[1.405,3.151]	[1.282,2.796]
	2.5$\}$	[1.178,3.253]	[1.374,3.36]	[1.846,4.175]
	$\delta 1 = 1$	[0.604,1.499]	[0.437,1.238]	[0.515,1.443]
	$\delta 2 = 2$	[1.583,2.309]	[0.987,2.474]	[1.149,2.754]
(20,30)	$\lambda = \{1.5$	[0.516,1.258]	[0.941,1.871]	[0.908,1.788]
	2	[1.189,2.979]	[1.015,2.081]	[1.415,2.912]
	2.5$\}$	[1.46,3.654]	[1.456,2.878]	[1.883,4.013]
	$\delta_1 = 1$	[0.6571.271]	[0.653,1.419]	[0.717,1.535]
	$\delta_2 = 2$	[1.489,2.239]	[1.277,2.777]	[1.383,2.827]
(40,40)	$\lambda = \{1.5$	[1.28,2.614]	[0.959,1.657]	[1.171,2.088]
	2	[0.973,1.918]	[1.411,2.837]	[1.746,3.038]
	2.5$\}$	[1.347,2.706]	[1.813,3.181]	[1.57,2.807]
	$\delta_1 = 1$	[0.721,1.309]	[0.75,1.442]	[0.697,1.361]
	$\delta_2 = 2$	[1.721,2.393]	[1.178,2.379]	[1.307,2.595]
(50,40)	$\lambda = \{1.5$	[0.828,1.597]	[1.225,2.046]	[1.221,2.098]
	2	[1.634,3.224]	[1.359,2.328]	[1.651,2.856]
	2.5$\}$	[1.402,2.699]	[1.989,3.475]	[2.010,3.465]
	$\delta 1 = 1$	[0.738,1.253]	[0.707,1.316]	[0.701,1.291]
	$\delta 2 = 2$	[1.819,2.334]	[1.357,2.417]	[1.471,2.59]
(60,60)	$\lambda = \{1.5$	[1.146,2.017]	[1.151,1.819]	[1.075,1.664]
	2	[1.235,2.158]	[1.598,2.539]	[1.298,2.084]
	2.5$\}$	[2.028,3.617]	[1.731,2.758]	[1.784,2.867]

Table 3: *Maximum Likelihood estimators and Bayes estimators of the combined model are given below.*

θ	**MLE**	**Mse**	**Bayes(S)**	**Mse**	**Gibbs**	**Mse**	**Bootstrap**	**Mse**
λ=0.5	0.623	0.015	0.595	0.009	0.316	0.034	0.320	0.032
δ_1=10	9.044	0.912	9.742	0.066	10.115	2.390	10.390	2.473
δ_2=11	10.773	0.051	11.753	0.567	11.460	3.315	12.098	5.891

Table 4: 95% *Asymptotic, HPD intervals using Gibbs and Bootstrap confidence interval for the parameters* $\theta = (\lambda, \delta_1, \delta_2)$ *of the model.*

θ	**Asymptotic**	**Gibbs**	**Bootstrap**
λ	[0.6181, 0.6268]	[0.2784, 0.3561]	[0.2764, 0.3519]
δ_1	[5.2672, 11.8221]	[7.0932, 13.1585]	[7.3985, 13.5131]
δ_2	[6.4066, 15.1402]	[8.212, 15.0033]	[8.3435, 15.1913]

7. Empirical Example

In evaluating our model's accuracy, we've collected data on annual deaths from Meningitis and Nutritional Deficiencies across 158 countries. Our aim is to assess its performance by comparing it with observed patterns in the dataset. To facilitate this comparison, we'll create histograms to visually represent the data and calculate the joint probability to understand simultaneous occurrences of both types of deaths. By employing a joint density function, we can explore various scenarios and better understand the probability associated with different combinations of Meningitis and Nutritional Deficiencies deaths. It's crucial for our model to align well with observed data patterns and predict a range of related health outcomes. The data for annual deaths is as follows:

Deaths due to Meningitis: 1563, 13, 292, 2520, 453, 46, 31, 62, 2323, 49, 45, 1975, 22, 134, 123, 2008, 39, 5258, 1239, 129, 2791, 89, 18, 902, 4623, 113, 6465, 377, 70, 284, 35, 2259, 17, 94, 38, 6147, 31, 110, 217, 118, 764, 52, 218, 55, 693, 73, 11283, 13, 16, 226, 86, 212, 21, 216, 3487, 30, 242, 3260, 267, 997, 109, 50, 34736, 4715, 577, 427, 10, 28, 185, 21, 408, 83, 136, 4396, 11, 12, 38, 162, 34, 162, 432, 43, 16, 2084, 2369, 291, 6260, 217, 522, 20, 24, 400, 2729, 1246, 100, 469, 96, 13, 44, 7772, 44914, 1235, 223, 19, 4493, 14, 17987, 28, 36, 599, 65, 143, 2056, 135, 36, 60, 1143, 995, 180, 25, 1563, 27, 1630, 20, 17, 4672, 2221, 77, 1968, 155, 297, 662, 31, 27, 214, 105, 125, 3765, 1037, 31, 625, 11, 85, 351, 52, 3941, 399, 54, 265, 1146, 14, 175, 254, 747, 14, 479, 2065, 1450.

Deaths due to Nutritional Deficiencies: 1244, 5, 114, 3015, 1330, 164, 12, 29, 4402, 45, 371, 820, 7, 894, 185, 8221, 11, 4048, 2048, 532, 965, 354, 12, 1247, 2454, 576, 16863, 1332, 66, 275, 23, 756, 3, 65, 134, 6355, 81, 206, 464, 734, 816, 293, 126, 31 1051, 92, 8989, 19, 12, 4734, 68, 116, 10, 782, 1973, 10, 1847, 1741, 100, 953, 218, 34, 26868, 20348, 230, 104, 10, 52, 773, 65, 1832, 18, 121, 4614, 13, 1, 9, 201, 15, 194, 323, 18, 4, 5285, 2062, 239, 14865, 279, 7558, 2, 15, 215, 3530, 1386, 189, 1300, 210, 15, 145, 2449, 5496, 6445, 443, 101, 26438, 38, 14631, 12, 126, 205, 304, 1300, 3611, 125, 165, 75, 456, 1142, 66, 18, 425, 20, 1180, 10, 26, 7626, 2101, 318, 2180, 467, 163, 795, 127, 95, 92, 126, 25, 6887, 992, 65, 237, 26, 29, 1043, 11, 3937, 139, 7, 159, 6090, 134, 41, 552, 954, 10, 1010, 1899, 2884.

Following the provided data, we estimate the parameters using Maximum Likelihood Estimation (MLE) and Bayes estimation techniques, as outlined in the preceding sections. Substituting the values, the results obtained are as follows.

In Table 3, we have estimated the parameters (λ, δ_1, and δ_2) of the combined model for deaths due to Meningitis and Nutritional Deficiencies. The parameter values and Mean Squared Error (MSE) of Maximum Likelihood Estimators (MLE), Bayes estimators using symmetric loss function, as well as Bayes estimators using Gibbs and Bootstrapping methods have been computed. From the results, it is evident that Bayes estimators, when estimated using the symmetric loss function, yield the minimum error. Additionally, asymptotic and Highest Posterior Density (HPD) intervals

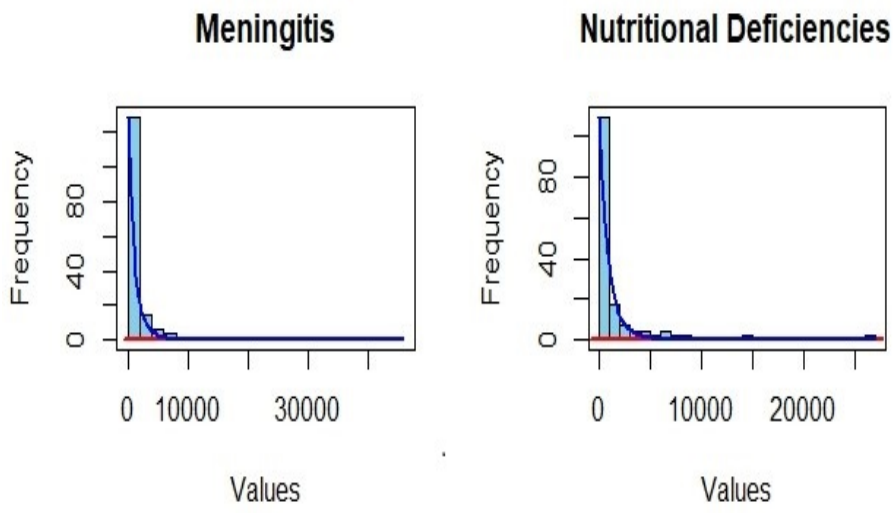


Figure 1: *Illustration of Deaths due to Meningitis and Nutritional Deficiencies.*

have been calculated using Gibbs and Bootstrapping, and the results are presented in Table 4.

Fig. 1 illustrates the number of deaths attributed to Meningitis and Nutritional Deficiencies. Each bar represents the frequency of deaths, while the curve depicts the density of the Lomax distribution. Based on the graph, we observe that deaths due to Meningitis approximately follow a Lomax distribution with parameters: shape=8, scale=0.2. Similarly, deaths due to Nutritional Deficiencies also approximately follow a Lomax distribution with parameters: shape=8, scale=0.

In summary, our analysis begins with trace plots and density plots, which serve as valuable tools for understanding the behavior of the Markov chain over iterations. Trace plots, illustrated in Figures 2 and 3, show the Markov chain's values against iteration number. A stable and random pattern across iterations is observed, indicating convergence and the adequate representation of the posterior distribution by the chain. Meanwhile, density plots, depicted in Figures 4 and 5, provide estimates of simulated marginal posterior distributions, resembling smoothed histograms. Importantly, the unimodal density plot indicates that the posterior distribution is well-behaved, lacking multimodality or skewness. Overall, our analysis underscores the robustness and reliability of our Bayesian inference process.

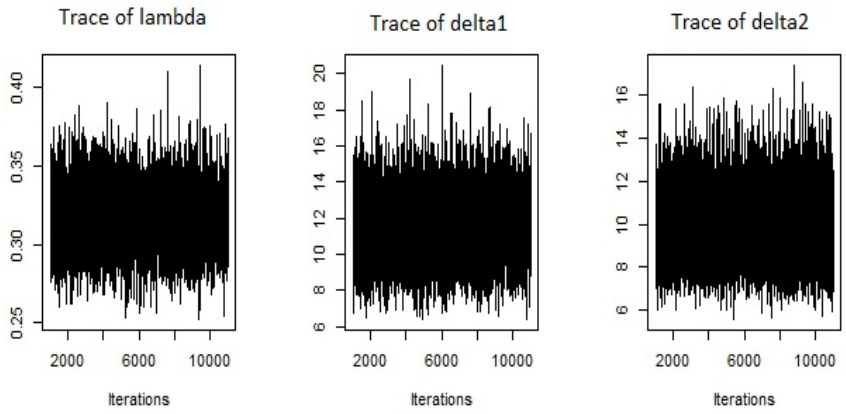


Figure 2: *Illustration of Trace of the parameters estimated using Gibbs*

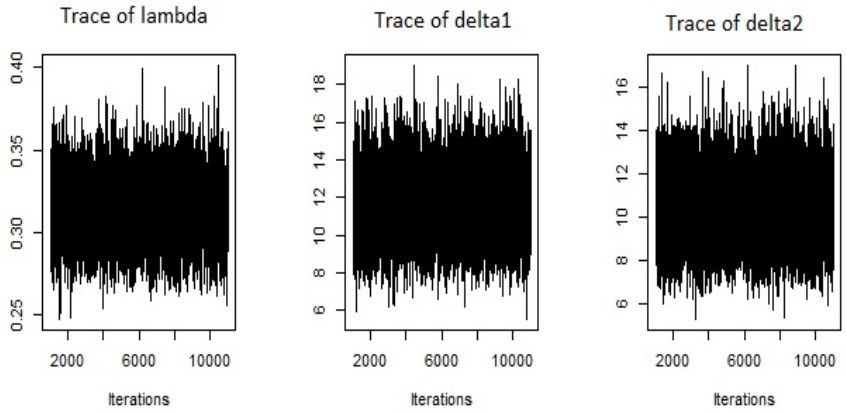


Figure 3: *Illustration of Trace of the parameters estimated using Bootstrapping*

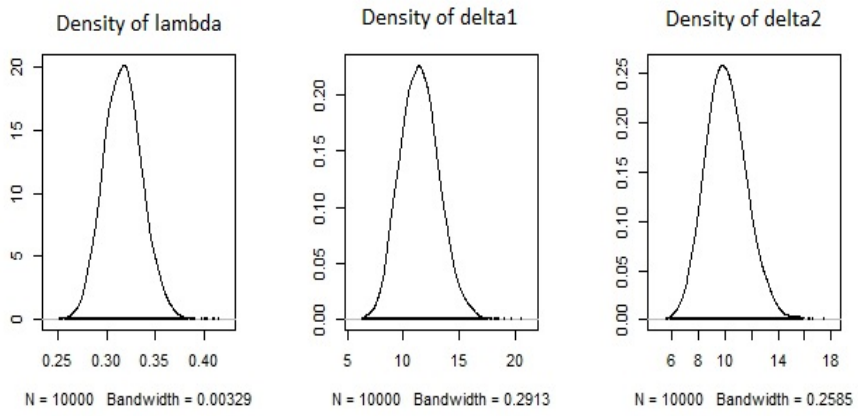


Figure 4: *Illustration of Density of the parameters estimated using Gibbs*

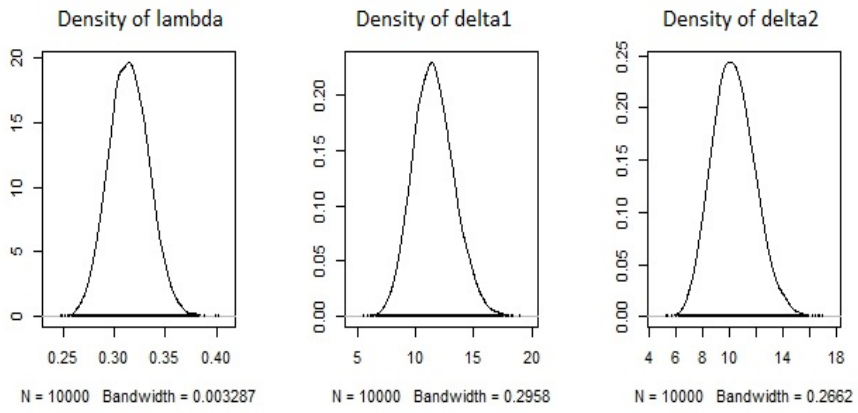


Figure 5: *Illustration of Density of the parameters estimated using Bootstrapping*

8. Conclusion

The focus of our study was to estimate the common shape parameter λ for two Lomax populations, where the scale parameters δ_1 and δ_2 are unknown and potentially different. It is important to note that this problem has not been explored in the existing literature. Similar to the case of a single population, it is not possible to obtain exact expressions for Maximum Likelihood Estimates (MLE) and Bayes estimates for our model. Hence, we employed a numerical approach to derive approximate MLEs for the associated parameters. Using these MLEs, we also obtained 95% asymptotic confidence intervals for the parameters.

Furthermore, we developed approximate Bayes estimators under various priors (vague, Jeffreys, and conjugate), incorporating symmetric loss function. A comprehensive assessment of all proposed estimators was conducted, evaluating their performance in terms of biases and risk values. Our numerical investigation highlighted the superiority of Bayes estimators under a conjugate prior, particularly when utilizing the symmetric loss function. These estimators demonstrated better performance compared to all other alternatives, specifically with respect to mean squared error.

It is imperative to emphasize that our conclusions regarding the suitability of these estimators are exclusively drawn from the outcomes of our numerical simulations. To elucidate the method of estimation, we presented a real-life example. We hope that our study will inspire researchers to explore alternative estimators for the common shape parameter, potentially offering competitive performance against our proposed estimators.

Acknowledgment

I would like to extend my heartfelt gratitude to the individuals and the esteemed VIT-AP Institution whose unwavering support and invaluable contributions have been instrumental in the successful completion of this research endeavor.

Conflict of Interest

The authors declare no conflicts of interest.

A. MLE

The detailed explanation of the notations used in Section 2 are given below.

$$
\begin{aligned}
T_1 &= \sum_{i=1}^{m}\left(1+\frac{x_{1i}}{\delta_1}\right),\ T_2=\sum_{j=1}^{n}\left(1+\frac{x_{2j}}{\delta_2}\right)\\
T_1' &= \sum_{i=1}^{m}\frac{x_{1i}}{\delta_1^2+\delta_1 x_i},\ T_2'=\sum_{j=1}^{n}\frac{x_{2j}}{\delta_2^2+\delta_2 x_{2j}}\\
T_1'' &= \sum_{i=1}^{m}\frac{x_{1i}\left(2\delta_1+x_{1i}\right)}{\left(\delta_1^2+\delta_1 x_{1i}\right)^2},\ T_2''=\sum_{j=1}^{n}\frac{x_{2j}\left(2\delta_2+x_{2j}\right)}{\left(\delta_2^2+\delta_2 x_{2j}\right)^2}\\
T_1''' &= \sum_{i=1}^{m}\frac{2x_{1i}\left(3\delta_1^2+3\delta_1 x_{1i}+x_{1i}^2\right)}{\left(\delta_1^2+\delta_1 x_{1i}\right)^3},\ T_2'''=\sum_{j=1}^{n}\frac{2x_{2j}\left(3\delta_2^2+3\delta_2 x_{2j}+x_{2j}^2\right)}{\left(\delta_2^2+\delta_2 x_{2j}\right)^3}\\
d &= -\frac{m+n}{\lambda^2}\left(\frac{m}{\delta_1^2}-(\lambda+1)T_1''\right)\left(\frac{n}{\delta_2^2}-(\lambda+1)T_2''\right)-(T_1')^2\left(\frac{n}{\delta_2^2}-(\lambda+1)T_2''\right)\\
&\quad -(T_2')^2\left(\frac{m}{\delta_1^2}-(\lambda+1)T_1''\right)
\end{aligned}
$$

B. Bayes

All the notations used in section 3.2.1 are given below.

$\rho_1 = 0, \rho_2 = -\frac{2}{\delta_1}, \rho_3 = -\frac{2}{\delta_2},$

$a_1 = -\frac{2}{\delta_1}\sigma_{12} - \frac{2}{\delta_2}\sigma_{13}, \quad a_2 = -\frac{2}{\delta_1}\sigma_{22} - \frac{2}{\delta_2}\sigma_{23}, \quad a_3 = -\frac{2}{\delta_1}\sigma_{32} - \frac{2}{\delta_2}\sigma_{33}.$

$$
\begin{aligned}
A &= \frac{2T_2'T_1''}{d}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right) - \frac{T_2''}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right) + (T_1')^2\right] \\
&\quad -\frac{2(m+n)}{d\lambda^3}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right)\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right), \\
B &= \frac{2T_1'T_1''}{d}\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right) - \frac{1}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right) + (T_2')^2\right]\left(\frac{2m}{\delta_1^3} - (\lambda+1)T_1'''\right), \\
D &= \frac{2T_2'T_2''}{d}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right) - \frac{1}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right) + (T_1')^2\right]\left(\frac{2n}{\delta_2^3} - (\lambda+1)T_2'''\right), \\
d &= -\frac{m+n}{\lambda^2}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right)\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right) - T_1'^2\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right) - T_2'^2\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right), \\
\sigma_{11} &= \frac{1}{d}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right)\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right), \\
\sigma_{12} &= -\frac{T_1'}{d}\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right) = \sigma_{21}, \\
\sigma_{13} &= -\frac{T_2'}{d}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right) = \sigma_{31}, \\
\sigma_{22} &= -\frac{1}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{n}{\delta_2^2} - (\lambda+1)T_2''\right) + (T_2')^2\right], \\
\sigma_{23} &= -\frac{T_1'T_2'}{d} = \sigma_{32}, \\
\sigma_{33} &= -\frac{1}{d}\left[\frac{m+n}{\lambda^2}\left(\frac{m}{\delta_1^2} - (\lambda+1)T_1''\right) + (T_1')^2\right].
\end{aligned}
$$

C. Jeffrey

All the notations used in section 3.2.2 are given below.

$$
\begin{aligned}
\rho_1 &= \frac{1}{2\left[\frac{d_1d_2(m+n)}{\lambda^2} + d_1(T_2')^2 + d_2(T_1')^2\right]}\left[\frac{(m+n)}{\lambda}\left(\frac{2d_1d_2}{\lambda} - d_1T_2'' - d_2T_1''\right) - (T_1')^2T_2'' - (T_2')^2T_1''\right] \\
\rho_2 &= \frac{1}{2\left[\frac{d_1d_2(m+n)}{\lambda^2} + d_1(T_2')^2 + d_2(T_1')^2\right]}\left[\frac{(m+n)d_1'd_2}{\lambda^2} + 2T_1'T_1''d_2 + (T_2')^2d_1\right] \\
\rho_3 &= \frac{1}{2\left[\frac{d_1d_2(m+n)}{\lambda^2} + d_1(T_2')^2 + d_2(T_1')^2\right]}\left[\frac{(m+n)d_1d_2'}{\lambda^2} + 2T_2'T_2''d_1 + (T_1')^2d_2\right]
\end{aligned}
$$

$a_1 = \rho_1\sigma_{11} + \rho_2\sigma_{12} + \rho_3\sigma_{13},$

$$a_2 = \rho_1\sigma_{21} + \rho_2\sigma_{22} + \rho_3\sigma_{23},$$

$$a_3 = \rho_1\sigma_{31} + \rho_2\sigma_{32} + \rho_3\sigma_{33},$$

$$d_1 = \frac{m}{\delta_1^2} - (\lambda+1)T_1'', \quad d_2 = \frac{n}{\delta_2^2} - (\lambda+1)T_2''.$$

$$d_1' = \frac{-2m}{\delta_1^3} - (\lambda_1+1)T_1''', \quad d_2' = \frac{-2n}{\delta_2^3} - (\lambda+1)T_2'''$$

D. Conjugate

All the notations used in section 3.2.3 are given below.

$$\rho_1 = \frac{c_1-1}{\lambda} - b_1, \quad \rho_2 = \frac{b_2}{\delta_1^2} - \frac{c_2+1}{\delta_1}, \quad \rho_3 = \frac{b_3}{\delta_2^2} - \frac{c_3+1}{\delta_2}$$

$$a_1 = \frac{c_1-1}{\lambda} - b_1\sigma_{11} + \frac{b_2}{\delta_1^2} - \frac{c_2+1}{\delta_1}\sigma_{12} + \frac{b_3}{\delta_2^2} - \frac{c_3+1}{\delta_2}\sigma_{13},$$

$$a_2 = \frac{c_1-1}{\lambda} - b_1\sigma_{21} + \frac{b_2}{\delta_1^2} - \frac{c_2+1}{\delta_1}\sigma_{22} + \frac{b_3}{\delta_2^2} - \frac{c_3+1}{\delta_2}\sigma_{23},$$

$$a_3 = \frac{c_1-1}{\lambda} - b_1\sigma_{31} + \frac{b_2}{\delta_1^2} - \frac{c_2+1}{\delta_1}\sigma_{32} + \frac{b_3}{\delta_2^2} - \frac{c_3+1}{\delta_2}\sigma_{33},$$

References

[1] Lomax, K. S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, **49**(268), 847–852.

[2] Bryson, M. C. (1974). Heavy-tailed distributions: Properties and tests. *Technometrics*, **16**, 61–68.

[3] Hassan, A. M., and Al-Ghamdi, A. S. (2009). Optimum step stress accelerated life testing for Lomax distribution. *Journal of Applied Sciences Research*, **5**, 2153–2164.

[4] Aljohani, H. M. (2024). Estimation for the P(X > Y) of Lomax distribution under accelerated life tests. *Heliyon*, **10**(3).

[5] Ijaz, M. (2021). Bayesian estimation of the shape parameter of Lomax distribution under uniform and Jeffery prior with engineering applications. *Gazi University Journal of Science*, **34**(2), 562–577.

[6] Chakraborty, T., Chattopadhyay, S., Das, S., Kumar, U., and Senthilnath, J. (2022). Searching for heavy-tailed probability distributions for modeling real-world complex networks. *IEEE Access*, **10**, 115092–115107.

[7] Okasha, H. M. (2014). E-Bayesian estimation for the Lomax distribution based on type-II censored data. *Journal of Egyptian Mathematical Society*, **22**, 489–495.

[8] Fitrilia, A., Fithriani, I., and Nurrohmah, S. (2018). Parameter estimation for the Lomax distribution using the E-Bayesian method. *Journal of Physics: Conference Series*, **1108**.

[9] Ellah, H. (2007). Comparison of estimates using record statistics from Lomax model: Bayesian and non-Bayesian approaches. *Journal of Statistical Research of Iran JSRI*, **3**, 139–158.

[10] Hasanain, W. S., Al-Ghezi, N. A. O., and Soady, A. M. (2022). Bayes estimation of Lomax parameters under different loss functions using Lindley's approximation. *Italian Journal of Pure and Applied Mathematics*, **48**, 630–640.

[11] Al-Bossly, A. (2021). E-Bayesian and Bayesian estimation for the Lomax distribution under weighted composite LINEX loss function. *Computational Intelligence and Neuroscience*, **2021**.

[12] Kumari, P., Kumar, V., and Aditi. (2022). Bayesian analysis for two-parameter Lomax distribution under different loss functions. *Communications in Mathematics and Applications*, **13**, 163–170.

[13] Graybill, F. A., and Deal, R. B. (1959). Combining unbiased estimators. *Biometrics*, **15**(4), 543–550.

[14] Moore, B. C., and Krishnamoorthy, K. (1997). Combining independent normal sample means by weighting with their standard errors. *Journal of Statistical Computation and Simulation*, **58**(2), 145–153.

[15] Tripathy, M. R., and Kumar, S. (2014). Equivariant estimation of common mean of several normal populations. *Journal of Statistical Computation and Simulation*, **85**(18), 3679–3699.

[16] Ghosh, M., and Razmpour, A. (1984). Estimation of the common location parameter of several exponentials. *The Indian Journal of Statistics*, **46**, 383–394.

[17] Jin, C., and Pal, N. (1992). On common location of several exponentials under a class of convex loss functions. *Calcutta Statistical Association Bulletin*, **42**, 167–168.

[18] Azhad, Q. J., Arshad, M. A., and Misra, A. (2020). Estimation of common location parameter of several heterogeneous exponential populations based on generalized order statistics. *Journal of Applied Statistics*, **48**(10), 1798–1815.

[19] Nagamani, N., and Tripathy, M. R. (2017). Estimating common scale parameter of two gamma populations: A simulation study. *American Journal of Mathematical and Management Sciences*, **36**, 346–362.

[20] Nagamani, N., Tripathy, M. R., and Kumar, S. (2020). Estimating common scale parameter of two logistic populations: A Bayesian study. *American Journal of Mathematical and Management Sciences*, **40**, 44–67.

[21] Lindley, D. V. (1980). Approximate Bayes method. *Trabajos de Estad?stica*, **31**, 223–237.

[22] Tripathy, M. R., and Nagamani, N. (2017). Estimating common shape parameter of two gamma populations: A simulation study. *Journal of Statistics and Management Systems*, **20**(3), 369–398.

RT&A, No 4(80)
Volume 19, December, 2024

DOUBLE SAMPLING INSPECTION PLAN UNDER ZERO-ONE FAILURE SCHEME FOR GENERALIZED INVERTED EXPONENTIAL DISTRIBUTION

S. Singh[1], A. Kaushik[2]

•

[1,2]Department of Statistics, Banaras Hindu University, Varanasi, India.
arundevkauhsik@gmail.com
Corresponding Author Email: [1]statsshubham@bhu.ac.in

Abstract

This article presents a double acceptance sampling plan for products whose lifetimes follow a generalized inverted exponential distribution. The plan uses a zero-one failure scheme, where a lot is accepted if there are no failures observed in the first sample, and it is rejected if more than one failure occurs. In cases where there is only one failure from the first sample, a second sample is drawn and tested for the same duration as the first sample. To ensure that the true median lifetime is longer than the specified lifetime at a given consumer's confidence level, the minimum sample sizes of the first and second samples are determined. The operating characteristics of the plan are analyzed for various ratios of the true median lifetime to the specified lifetime. Finally, an example is given to explain the results. The example shows how the double acceptance sampling plan can be used to determine the sample size and acceptance criteria for a product with a specified lifetime and a given consumer's confidence level. The results of the example demonstrate the effectiveness of the plan in ensuring that the true median lifetime of the product is longer than the specified lifetime at the desired level of confidence.

Keywords: Generalized inverted exponential distribution$(GIED)$, Double acceptance sampling plan, Operating Characteristics $(O.C.)$, Producer's risk, Consumer's risk

1. Introduction

Statistical quality control involves the study of the quality of manufactured items, which is often characterized by their lifetimes. However, items produced under identical conditions may have variable lifetimes due to chance causes. To avoid issues with consumer acceptance, producers prioritize the quality of their products. However, it may not be feasible to inspect the entire lifetime of all products due to destruction during inspection and time/cost limitations. As such, a sampling inspection plan is recommended. However, both producers and consumers face risks when making acceptance/rejection decisions based on samples. If a sample from a lot is accepted, and several products have a mean or median lifetime less than the specified lifetime (a "bad" lot), this is known as consumer's risk. On the other hand, the probability that a "good" lot of products is rejected is known as the producer's risk. Increasing the sample size can minimize these risks, but it can also lead to higher costs and longer inspection times for producers. To minimize risks while considering time and cost constraints, an efficient acceptance sampling scheme should be used. Therefore, implementing a sampling inspection plan is recommended in this process, as both the producer and consumer are exposed to risks when deciding whether to accept or reject products. If a sample from a lot is accepted, and several products have a mean or median lifetime less than the specified lifetime (known as a "bad" lot), this represents

consumer's risk. The probability of rejecting a "good" lot of products is known as the producer's risk. Increasing the sample size can be an effective way to minimize the risks faced by both the producer and consumer, but this approach may incur higher costs and longer inspection times for the producer. To minimize these risks while adhering to time and cost constraints, an efficient acceptance sampling scheme should be utilized.

A statistical method for product control known as acceptance sampling or sampling inspection as an alternative to 100% complete inspection was developed by [1]. An acceptance sampling plan for exponential distribution was the first time considered by [2]. The problem of acceptance sampling plans when the life test is truncated at a pre-assigned time under the gamma distribution by taking mean lifetime as the average lifetime of the product studied in detail by [3]. The acceptance sampling plans at various confidence levels for various values of the ratio of the prefixed experimental time to the specified mean lifetime under inverse Rayleigh distribution investigated by [4]. To use median lifetime as the average lifetime of the product for acceptance sampling plans under the generalized Birnbaum-Saunders distribution proposed by [7]. Truncated life tests for log-logistic distribution at various confidence levels for different values of the ratio of the pre-fixed experimental time to the specified average lifetime considered by [5]. Time truncated acceptance sampling plans for generalized exponential distribution by taking the median lifetime of the product as the preferred average lifetime of the product developed by [9]. An acceptance sampling plans based on truncated life tests in the Pareto distribution of the second kind provided by [6]. A double acceptance sampling plans based on truncated life tests for Marshall Olkin extended exponential distribution considered by [10], he also calculate the consumer's risk for the various choices of prefixed producer's risk. An acceptance sampling plan based on truncated life tests for exponentiated Frechet distribution developed by [11]. A comparison study of the three sampling plans namely single acceptance sampling plan, double acceptance sampling plan and group acceptance sampling plan provided by [14]. A new attribute sampling inspection plan and its applications has been discussed by [23]. An acceptance sampling plan based on truncated life tests for generalized inverse weibull distribution by assumming mean lifetime as the preferred average lifetime of the product developed [16]. The time truncated double acceptance sampling inspection plan for Maxwell distribution considered by [12]. A double acceptance sampling inspection plan for time truncated life tests based on the transmuted generalized inverse Weibull distribution provided by [17]. A double acceptance sampling inspection plan for time truncated life tests based on the transmuted new Weibull-Pareto distribution by taking median lifetime as the average lifetime of the product considered by [15]. An acceptance sampling plan based on the time truncated life tests for Sushila distribution by assumming mean lifetime as the average life of the product and also illustrate an example by taking real data application considered [18]. A single and double acceptance sampling plans for truncated life tests based on transmuted Rayleigh distribution by taking different values of the parameters developed by [21]. A double and group acceptance sampling plan for truncated life test based on inverse log-logistic distribution by taking mean lifetime as the average lifetime of the product compared by [22]. An improved acceptance sampling plans based on truncated life tests for Garima distribution developed by [19].

GIED is a very useful lifetime distribution given by [8] and it is also found to be very useful in the reliability testing. This distribution has non- constant hazard rate function, which is also unimodal and right positively skewed distribution.

The objective of this paper is to propose a double acceptance sampling plan for life tests assuming that the product's lifetime follows a GIED. The zero-one failure scheme is primarily considered, as it is generally more effective than the ordinary sampling scheme. Under this scheme, a lot is accepted if no failures are observed from the first sample, and it is rejected if there are more than one failures. If there is exactly one failure, a second sample is selected and tested for the same duration as the first sample. The zero-one failure scheme is particularly useful in situations where the cost of the product is high, and it is not feasible to bear the high cost of inspection. The minimum sample sizes for the first and second samples are determined at a specified consumer's confidence level. The operating characteristics are analyzed as a function of

the ratio of the time to the specified median lifetime, and the minimum of such ratios is obtained to minimize the producer's risk at the specified level.

The model (GIED) is introduced in Section 2. The methodology used in double sampling plan is given in Section 3 and its operating characteristics(OC) function, Producer's risk and Consumer's risk are analyzed in Section 4, Section 4.1 and Section 4.2 respectievely. An example to explain the double sampling plan based on zero-one failure scheme are given in Section 5.

2. The Model

GIED is the generalization of the one parameter IED which was proposed by [8]. The various statistical properties and reliability estimation has also been studied in detail by [8].

A two parameter GIED has the following PDF and CDF as follows:

$$\begin{aligned} f(x) &= \frac{\alpha\lambda}{x^2}\exp\left(\frac{-\lambda}{x}\right)\left(1-\exp\left(\frac{-\lambda}{x}\right)\right)^{\alpha-1}; \qquad x>0, \alpha>0, \lambda>0 \\ &\text{and,} \\ F(x) &= 1-(1-exp\,(-\lambda/x))^{\alpha}. \end{aligned} \tag{1}$$

Where, α is the shape parameter and λ is the scale parameter.
Let t_p be the p^{th} percentile of the GIED then it is given by

$$p = 1-\left(1-exp\left(-\frac{\lambda}{t_p}\right)\right)^{\alpha} \tag{2}$$

This implies that

$$t_p = \frac{-\lambda}{\ln(1-(1-p)^{1/\alpha})}$$

and median of the distribution is given by

$$m_d = \frac{-\lambda}{\ln(1-(0.5)^{1/\alpha})}. \tag{3}$$

For eliminating λ, we can write

$$\begin{aligned} \frac{t_p}{m_d} &= \frac{ln(1-(0.5)^{1/\alpha})}{ln(1-(1-p)^{\frac{1}{\alpha}})}, \\ p &= 1-(1-\ exp(A))^{\alpha}. \end{aligned} \tag{4}$$

Where,

$$A = \frac{ln(1-(0.5)^{1/\alpha})}{\frac{t_p}{m_d}}. \tag{5}$$

3. Methodology

Let us consider the median lifetime of a product as m_d, which is used to determine the quality of a submitted lot. A lot is considered to be of acceptable quality (i.e., good) if the null hypothesis $H_0 : m_d \geq m_d^0$ is supported by the data against the alternative hypothesis $H_1 : m_d < m_d^0$, where m_d^0 is a specified lifetime. The consumer's risk is used as the significance level for the test, denoted by $1-P^{\star}$, where $P^{\star}$ is the consumer's confidence level. To implement the double sampling plan, the following method is used: the first sample is drawn from the lot, and if no failures are observed in this sample, the lot is accepted. If more than one failure occurs, the lot is rejected. When only one failure occurs in the first sample, a second sample is drawn and tested for the

same duration as the first sample. The minimum sample sizes of the first and second samples are determined to ensure that the true median lifetime is longer than the specified lifetime at the specified consumer's confidence level. The method is designed to minimize the producer's risk while keeping the time and cost of inspection within acceptable limits. Assume that the lifetime of a products can be represented by its median lifetime, denoted by m_d. We say that the submitted lot has a good quality (lot of acceptable quality) and the lot will be accepted if the data supports the following null hypothesis, $H_0 : m_d \geq m_d^0$ (m_d^0 is a specified lifetime) against the alternative hypothesis, $H_1 : m_d < m_d^0$. As the significance level for the test, the consumer's risk is used through $1 - P^\star$, where $P^\star$ is the consumer's confidence level. The method of double sampling plan is given as follows:

- To begin the inspection process, a random sample of size n_1 is drawn from the lot and put to the test. If the number of failures observed before a pre-fixed experiment time t is equal to or less than c_1, the lot is accepted. However, if the number of failures observed is greater than c_2, the experiment is terminated before t, and the lot is rejected. Here, it is important to note that c_1 must be less than c_2.
- A double sampling plan is used to make a decision about whether to accept or reject a lot based on two samples that have been inspected. If the lot has not been rejected or accepted, which means that the number of failures by time t is between $c_1 + 1$ and c_2, a second sample of size n_2 is drawn and tested during time t. The lot is deemed acceptable if no more than c_2 failures are observed from both samples. On the other hand, if there are more than c_2 failures, the lot is rejected. A sampling plan in which a decision about the acceptance or rejection of a lot is based on two samples that have been inspected is known as a double sampling plan.

Double sampling is utilized when a definite verdict regarding the approval or disapproval of a batch cannot be made based on a single sample. Typically, in a double sampling plan, the approval or disapproval of a batch is determined by evaluating two samples. If the initial sample is satisfactory, the batch is approved without requiring a second sample. Conversely, if the initial sample is unsatisfactory, the batch is rejected without necessitating a second sample. However, if the initial sample is inconclusive and there are doubts about its outcome, a second sample is taken and the approval or disapproval decision is made based on the evidence obtained from both samples.

Assuming that N lots of equal size are received from either the supplier or the final assembly line and examined individually, the acceptance probability can be calculated using the binomial distribution since the lot size is sufficiently large. The double acceptance sampling plan, which is based on a zero-one failure scheme, determines whether the lot is accepted or rejected.

$$P_a = \sum_{i=0}^{c_1} \binom{n_1}{i} p^i (1-p)^{n_1-i} + \sum_{i=c_1+1}^{c_2} \binom{n_1}{i} p^i (1-p)^{n_1-i} \left[\sum_{j=0}^{c_2-i} \binom{n_2}{j} p^j (1-p)^{n_2-j} \right]. \tag{6}$$

Where, p is given in (4)

In the context of the zero-one failure scheme for double acceptance sampling, we are specifically concerned with the case where $c_1 = 0$ and $c_2 = 1$. In this acceptance sampling plan, the consumer prefers a smaller number of accepted lots. The acceptance probability of the lot in the zero-one failure scheme can be expressed as follows:

$$P_a = (1-p)^{n_1} + n_1 p (1-p)^{n_1-1} (1-p)^{n_2}. \tag{7}$$

Therefore, the minimum required sample sizes n_1 and n_2 ensuring $m_d \geq m_d^0$ at the consumer confidence level P^* can be found as follows:

$$(1-p)^{n_1} + n_1 p (1-p)^{n_1-1} (1-p)^{n_2} \leq 1 - P^*. \tag{8}$$

There will be infinite many solutions satisfying Eq(8). Ultimately our goal is to minimize the average sample number (ASN) under the constraint that $n_2 \leq n_1$.

The ASN for double sampling plan based on zero-one failure scheme is given by

$$ASN= n_1P_1 + (n_1 + n_2)(1 - P_1).$$

Where, P_1 is probability of acceptance or rejection of the lot based on the first sample is given by

$P_1 = 1 - \sum_{i=c_1+1}^{c_2} \binom{n_1}{i} p^i (1-p)^{n_1-i}$.

Table 1: *Minimum sample size required for the double acceptance sampling plan with shape parameter* $\alpha = 1$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.33	(5, 3)	(7, 5)	(8, 8)	(12, 9)
0.942	0.479	(3, 2)	(4, 4)	(5, 5)	(8, 4)
1.257	0.576	(3, 1)	(3, 3)	(4, 3)	(6, 3)
1.571	0.643	(2, 2)	(3, 2)	(4, 2)	(5, 3)
2.356	0.745	(2, 1)	(2, 2)	(3, 2)	(4, 2)
3.142	0.802	(2, 1)	(2, 2)	(3, 1)	(3, 3)

Table 2: *Minimum sample size required for the double acceptance sampling plan with shape parameter* $\alpha = 2$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.2630	(6, 5)	(9, 6)	(11, 8)	(16, 10)
0.942	0.4694	(3, 3)	(5, 2)	(5, 5)	(8, 4)
1.257	0.6112	(2, 2)	(3, 3)	(4, 2)	(5, 5)
1.571	0.7059	(2, 1)	(3, 1)	(3, 2)	(4, 3)
2.356	0.8350	(2, 1)	(2, 1)	(2, 2)	(3, 2)
3.142	0.8954	(1, 1)	(2, 1)	(2, 1)	(3, 1)

Table 3: *Minimum sample size required for the double acceptance sampling plan with shape parameter* $\alpha = 4$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.1976	(8, 7)	(12, 10)	(15, 11)	(22, 14)
0.942	0.4583	(3, 3)	(5, 3)	(6, 3)	(8, 5)
1.257	0.6515	(2, 2)	(3, 2)	(4, 2)	(5, 3)
1.571	0.7738	(2, 1)	(2, 2)	(3, 1)	(4, 2)
2.356	0.9139	(1, 1)	(2, 1)	(2, 1)	(2, 2)
3.142	0.9615	(1, 1)	(1, 1)	(2, 1)	(2, 1)

Table 4: *Minimum sample size required for the double acceptance sampling plan with shape parameter* $\alpha = 6$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.1638	(10, 8)	(15, 11)	(18, 15)	(27, 18)
0.942	0.4513	(3, 3)	(5, 3)	(6, 3)	(8, 6)
1.257	0.6769	(2, 1)	(3, 2)	(3, 3)	(5, 2)
1.571	0.8134	(2, 1)	(2, 1)	(2, 2)	(3, 2)
2.356	0.9487	(1, 1)	(1, 1)	(2, 1)	(2, 1)
3.142	0.9832	(1, 1)	(1, 1)	(1, 1)	(2, 1)

4. OC function

The operating characteristic function of the sampling plan provides the probability of accepting the lot of incoming lot quality p. For the above acceptance sampling plan this probability is given by

$$\begin{aligned} OC(p) &= Prob(\text{accepting the lot of the incoming lot quality}) \\ &= (1-p)^{n_1} + n_1 p(1-p)^{n_1-1}(1-p)^{n_2}. \end{aligned} \tag{9}$$

Table 5: *OC function of the incoming lot quality with shape parameter* $\alpha = 1$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.33	0.2350	0.0888	0.0471	0.0095
0.942	0.479	0.2473	0.0936	0.0452	0.0084
1.257	0.576	0.2079	0.0999	0.0457	0.0094
1.571	0.643	0.1860	0.0768	0.0312	0.0082
2.356	0.745	0.1619	0.0897	0.0260	0.0074
3.142	0.802	0.1021	0.0517	0.0264	0.0085

Table 6: *O.C. function of the incoming lot quality with shape parameter* $\alpha = 2$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.2630	0.2348	0.0972	0.0467	0.0096
0.942	0.4694	0.2086	0.0944	0.0499	0.0098
1.257	0.6112	0.2230	0.0750	0.0446	0.0095
1.571	0.7059	0.2087	0.0793	0.0413	0.0093
2.356	0.8350	0.0727	0.0727	0.0347	0.0063
3.142	0.8953	0.1983	0.0306	0.0306	0.0042

Table 7: *O.C. function of the incoming lot quality with shape parameter* $\alpha = 4$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.1976	0.2443	0.0945	0.0489	0.0098
0.942	0.4583	0.2231	0.0780	0.0457	0.0098
1.257	0.6515	0.1766	0.0711	0.0281	0.0072
1.571	0.7738	0.1304	0.0691	0.0385	0.0045
2.356	0.9139	0.1648	0.0210	0.0210	0.0086
3.142	0.9615	0.0755	0.0755	0.0043	0.0043

Table 8: *O.C. function of the incoming lot quality with shape parameter* $\alpha = 6$

t_p/m_d	p	P^*			
		0.75	0.90	0.95	0.99
0.628	0.1638	0.2455	0.0965	0.0496	0.0097
0.942	0.4513	0.2326	0.0835	0.0495	0.0097
1.257	0.6769	0.2458	0.0559	0.0409	0.0074
1.571	0.8134	0.0914	0.0914	0.0454	0.0094
2.356	0.9487	0.0999	0.0999	0.0076	0.0076
3.142	0.9832	0.0333	0.0333	0.0333	0.00083

4.1. Producer's risk

The probability of rejecting a lot of acceptance quality level (AQL) p_1 is called producer's risk. It is given for the double sampling inspection plan based on zero-one failure scheme as follows:

$$\begin{aligned} PR(p) &= Prob(\text{ rejecting a lot}) \\ &= 1 - Prob(\text{accepting a lot of the quality } p_1) \\ &= 1 - (1-p_1)^{n_1} - n_1 p_1 (1-p_1)^{n_1-1}(1-p_1)^{n_2}. \end{aligned} \tag{10}$$

Table 9: *Producer's risk of the ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 1$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.1704	0.3020	0.4048	0.5717
0.10	0.20	0.942	0.0742	0.1526	0.2158	0.3185
0.10	0.25	1.257	0.0523	0.0939	0.1313	0.2103
0.05	0.15	1.571	0.0118	0.0204	0.0324	0.0516
0.05	0.20	2.356	0.0073	0.0118	0.0204	0.0307
0.05	0.25	3.142	0.0073	0.0118	0.0140	0.0266

Table 10: *Producer's risk of the ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 2$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.2594	0.4067	0.5211	0.6998
0.10	0.20	0.942	0.0939	0.1438	0.2158	0.3185
0.10	0.25	1.257	0.0442	0.0939	0.1077	0.2158
0.05	0.15	1.571	0.0073	0.0140	0.0204	0.0385
0.05	0.20	2.356	0.0073	0.0073	0.0118	0.0204
0.05	0.25	3.142	0.0025	0.0073	0.0073	0.0140

Table 11: *Producer's risk of the ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 4$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.3865	0.5863	0.6864	0.8465
0.10	0.20	0.942	0.0939	0.1704	0.2103	0.3436
0.10	0.25	1.257	0.0442	0.0742	0.1077	0.1704
0.05	0.15	1.571	0.0073	0.0118	0.0140	0.03074
0.05	0.20	2.356	0.0025	0.0073	0.0073	0.0118
0.05	0.25	3.142	0.0025	0.0025	0.0073	0.0073

Table 12: *Producer's risk of the ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 6$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.4845	0.6864	0.7881	0.9157
0.10	0.20	0.942	0.0939	0.1704	0.2103	0.3662
0.10	0.25	1.257	0.0280	0.0742	0.0939	0.1438
0.05	0.15	1.571	0.0073	0.0073	0.0118	0.0204
0.05	0.20	2.356	0.0025	0.0025	0.0073	0.0073
0.05	0.25	3.142	0.0025	0.0025	0.0025	0.0073

4.2. Consumer's risk

The probability of accepting the lot of the rejecting quality level (RQL) p_2 is called consumer's risk. The consumer's risk for the double acceptance sampling plan based on zero-one failure scheme is given by

$$c(p) = (1-p_2)^{n_1} + n_1 p_2 (1-p_2)^{n_1-1}(1-p_2)^{n_2}. \tag{11}$$

Where, p be the percentage defective in the lot that is lot quality(p), p_1 be the acceptance quality level of lot quality p, p_2 be the rejecting quality level of lot quality p.
Also, P_1 be the probability of accepting or rejecting the lot on the basis of the first sample and similar argument can be given for P_2.

Table 13: *Consumer's risk of ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 1$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.6610	0.4696	0.3773	0.2120
0.10	0.20	0.942	0.2103	0.5774	0.4619	0.3052
0.10	0.25	1.257	0.7207	0.5998	0.4944	0.3281
0.05	0.15	1.571	0.9067	0.8490	0.7838	0.6841
0.05	0.20	2.356	0.8960	0.8448	0.7578	0.6717
0.05	0.25	3.142	0.8906	0.7734	0.7207	0.5999

Table 14: *Consumer's risk of ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 2$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.5543	0.37034	0.2559	0.1155
0.10	0.20	0.942	0.7086	0.5898	0.4619	0.3052
0.10	0.25	1.257	0.7734	0.5999	0.5537	0.3312
0.05	0.15	1.571	0.9393	0.8905	0.8490	0.7483
0.05	0.20	2.356	0.896	0.896	0.8448	0.7578
0.05	0.25	3.142	0.9375	0.8437	0.8437	0.7383

Table 15: *Consumer's risk of ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 4$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.3958	0.2015	0.1260	0.0392
0.10	0.20	0.942	0.7086	0.5374	0.4635	0.2777
0.10	0.25	1.257	0.7734	0.6592	0.5537	0.4042
0.05	0.15	1.571	0.9393	0.9067	0.8905	0.7882
0.05	0.20	2.356	0.96	0.896	0.896	0.8448
0.05	0.25	3.142	0.9375	0.9375	0.8437	0.8437

Table 16: *Consumer's risk of ($n_1, n_2,\ t_p/m_d$) with shape parameter $\alpha = 6$*

AQL(p_1)	RQL(p_2)	t_p/m_d	p^*			
			0.75	0.90	0.95	0.99
0.10	0.15	0.628	0.2915	0.1260	0.0685	0.0156
0.10	0.20	0.942	0.7086	0.5374	0.4635	0.2557
0.10	0.25	1.257	0.8437	0.6592	0.5999	0.4598
0.05	0.15	1.571	0.9393	0.9393	0.9067	0.8490
0.05	0.20	2.356	0.9600	0.9600	0.8960	0.8960
0.05	0.25	3.142	0.9375	0.9375	0.9375	0.8437

5. Description of tables with examples

The choices of $t_p/m_d = 0.628, 0.942, 1.257, 1.571, 2.356, 3.142$ are consistent with [3], [5], [7].
Table 1 indicates that with a constant termination ratio (t_p/m_d), fixed proportion defective (p), and a consistent value of the shape parameter, there exists a clear relationship between the consumer's confidence level and the sample size; as the confidence level increases, the sample size also increases. This identical pattern holds true from Table 2 to Table 4, considering diverse values of the shape parameter α. Table 5 demonstrates that under a constant termination ratio (t_p/m_d), fixed proportion defective (p), and a given shape parameter value, an increase in the consumer's confidence level leads to a decrease in the Operating Characteristic (O.C.) function value. This consistent trend is also evident from Table 6 to Table 8 for varying values of the shape parameter α. Table 9 shows that under a constant termination ratio (t_p/m_d), shape parameter, and given AQL (p_1) and RQL (p_2) values, an increase in the consumer's confidence level leads to a reduction in the Producer's risk. This trend is consistent across Table 10 to Table 12 when considering different values of the shape parameter α.
Table 13 demonstrates how, with a constant termination ratio (t_p/m_d) as well as fixed AQL (p_1) and RQL (p_2) values, an increase in Consumer's confidence level leads to a decrease in Consumer's risk. This trend is consistent across Table 14 to Table 16 for different values of the shape parameter α.
Example:- The first table determines the minimum sample size required from a lot in order to make a decision about accepting or rejecting it. For instance, in analyzing the reliability of laptops using a double acceptance sampling plan under the zero-one failure scheme, the failure times in days (denoted by T) are used to represent the lifetimes of the laptops from the start of their operation. It is assumed that T follows a generalized inverted exponential distribution.

Assuming a specified median life of 1000 days and a testing time of 942 days, the manufacturer would like to determine if the median life of the laptops is at least 1000 days with a confidence level of 0.99. The experiment will end after 942 days using the zero-one failure scheme of the double sampling plan, which refers to a ratio of 0.942. From the information provided in Table 1, the minimum sample sizes required are $n_1 = 8$ and $n_2 = 4$. To start, eight items are monitored for 942 days, and the lot is accepted if there are no failures during the experiment. If more than one failure occurs, the lot is rejected. Next, a second sample of four items is selected from the lot and tested for 942 days, during which only one failure occurred.

The laptop has a specified median life of 1000 days, and the manufacturer wants to determine whether the median life is at least 1000 days with 99% confidence. The experiment will use a double sampling plan with a zero-one failure scheme, ending at 942 days. The ratio of testing time to specified median life is 0.942. According to Table 1, the minimum required sample sizes are $n_1 = 8$ and $n_2 = 4$. To conduct the experiment, the first step is to monitor eight items for 942 days. If no failures occur during this time, the lot is accepted. If there are one or more failures, the lot is rejected. If the lot is accepted, a second sample of size four is selected from the lot and tested for 942 days. If only one failure occurs during this time, the manufacturer can conclude with 99% confidence that the median life of the laptop is at least 1000 days.

6. Conclusion

This article presents a novel double acceptance sampling plan for the zero-one failure scheme based on the median life of products with a fixed tenure life. The plan is developed for the GIED and assumes that the life test is truncated with the end of the product's lifetime. The goal of the plan is to ensure that the product has a better life than the specified one, and it considers the percentile ratio of lifetimes to determine the design parameters of the double sampling plan, namely, $(n_1, n_2,\ t_p/m_d)$. The study shows that the zero-one failure scheme is superior to the ordinary sampling plan in terms of the required sample size for inspection. This work can be extended for various lifetime models under different sampling plan.

References

[1] Montgomery, D. C. (2007). Introduction to statistical quality control. John Wiley Sons.
[2] Epstein, B. (1954). Truncated life tests in the exponential case. The Annals of Mathematical Statistics, 555-564.
[3] Gupta, S. S., Gupta, S. S. (1961). Gamma distribution in acceptance sampling based on life tests. Journal of the American Statistical Association, 56(296), 942-970.
[4] Rosaiah, K., Kantam, R. R. L. (2005). Acceptance sampling based on the inverse Rayleigh distribution.
[5] Kantam, R. R. L., Rosaiah, K., Rao, G. S. (2001). Acceptance sampling based on life tests: log-logistic model. Journal of applied statistics, 28(1), 121-128.
[6] Baklizi, A. (2003). Acceptance sampling based on truncated life tests in the Pareto distribution of the second kind. Advances and Applications in Statistics, 3(1), 33-48.
[7] Balakrishnan, N., Leiva, V., L'opez, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum-Saunders distribution. Communications in Statistics-Simulation and Computation, 36(3), 643-656.
[8] Abouammoh, A. M., Alshingiti, A. M. (2009). Reliability estimation of generalized inverted exponential distribution. Journal of statistical computation and simulation, 79(11), 1301-1315.
[9] Aslam, M., Kundu, D., Ahmad, M. (2010). Time truncated acceptance sampling plans for generalized exponential distribution. Journal of Applied Statistics, 37(4), 555-566.
[10] Rao, G. S. (2011). Double acceptance sampling plans based on truncated life tests for the Marshall-Olkin extended exponential distribution. Austrian journal of Statistics, 40(3), 169-176.
[11] Al-Nasser, A. D., Al-Omari, A. I. (2013). Acceptance sampling plan based on truncated life tests for exponentiated Frechet distribution. Journal of Statistics and Management Systems, 16(1), 13-24.
[12] Gui, W. (2014). Double acceptance sampling plan for time truncated life tests based on Maxwell distribution. American Journal of Mathematical and Management Sciences, 33(2), 98-109.
[13] Oguntunde, P. E., Adejumo, A., Balogun, O. S. (2014). Statistical properties of the exponentiated generalized inverted exponential distribution. Applied Mathematics, 4(2), 47-55.
[14] Singh, S., Tripathi, Y. M., Jun, C. H. (2015). Sampling plans based on truncated life test for a generalized inverted exponential distribution. Industrial Engineering and Management Systems, 14(2), 183-195.
[15] Al-Omari, A. I., Zamanzade, E. (2017). Double acceptance sampling plan for time truncated life tests based on transmuted generalized inverse Weibull distribution. Journal of Statistics Applications and Probability, 6(1), 1-6.
[16] Al-Omari, A. I. (2016). Time truncated acceptance sampling plans for Generalized Inverse Weibull Distribution. Journal of Statistics and Management Systems, 19(1), 1-19.
[17] Al-Omari, A. I., Zamanzade, E. (2017). Double acceptance sampling plan for time truncated life tests based on transmuted generalized inverse Weibull distribution. Journal of Statistics Applications and Probability, 6(1), 1-6.

[18] Al-Omari, A. I. (2018). Acceptance sampling plans based on truncated life tests for Sushila distribution. Journal of Mathematical and Fundamental Sciences, 50(1), 72-83.
[19] Al-Omari, A. (2018). Improved acceptance sampling plans based on truncated life tests for Garima distribution. International Journal of System Assurance Engineering and Management, 9(6), 1287-1293.
[20] Ahmadi Nadi, A., Sadeghpour Gildeh, B. (2019). A group multiple dependent state sampling plan using truncated life test for the Weibull distribution. Quality Engineering, 31(4), 553-563.
[21] Saha, M., Tripathi, H., Dey, S. (2021). Single and double acceptance sampling plans for truncated life tests based on transmuted Rayleigh distribution. Journal of Industrial and Production Engineering, 38(5), 356-368.
[22] Tripathi, H., Dey, S., Saha, M. (2021). Double and group acceptance sampling plan for truncated life test based on inverse log-logistic distribution. Journal of Applied Statistics, 48(7), 1227-1242.
[23] Tripathi, H., Al-Omari, A. I., Saha, M., Mali, A. (2022). Time truncated life tests for new attribute sampling inspection plan and its applications. Journal of Industrial and Production Engineering, 39(4), 293-305.

DISCRETE-TIME QUEUEING ANALYSIS OF POWER-SAVING MECHANISMS IN LTE DRX SYSTEMS WITH DIFFERENTIATED VACATION AND DISASTER

A Mohammed Shapique[1], A Vaithiyanathan[2]

•

[1,2]IFET College of Engineering, Villupuram, India
shapique@gmail.com, vaithi05@gmail.com

Abstract

This paper investigates the power-saving mechanisms of Discontinuous Reception (DRX), a technique used in wireless communication networks to reduce energy consumption. By employing a discrete-time Geo/Geo/1 queueing model with differentiated vacations and system disasters, we aim to more accurately capture the intermittent nature of data arrivals, often overlooked in continuous-time models. Our research addresses the existing gap in the literature by providing a more realistic representation of DRX behaviour. Understanding the performance and characteristics of DRX is crucial for optimizing energy efficiency and improving the overall performance of wireless networks. This paper contributes to this understanding by deriving steady-state probabilities, calculating key performance metrics, and visualizing the system behaviour through graphical analysis.

Keywords: Geo/Geo/1, Differentiated vacation, Disaster, Repair, Generating function, Steady-state probabilities

1. Introduction

Queueing theory is a mathematical framework used to analyse systems where customers or jobs wait in line for service. It is essential in healthcare, manufacturing, communication, Internet of Things (IoT) and computer systems. For example, in a communication network, queueing theory helps predict data packet delays and optimise network performance. Network engineers can design more efficient and reliable systems by understanding the dynamics of queues. The proliferation of wireless devices and the increasing demand for seamless connectivity have created a critical need for energy-efficient communication systems. Power-Saving Mechanisms (PSMs) have emerged as a pivotal solution to address this challenge. Recently, queueing theory with vacations has emerged as a powerful tool for analysing PSM in wireless sensor networks [14], Internet of things [10] and wireless communication systems [12, 7]. By modelling the system as a queue where the server (e.g., a processor or network node) takes periodic vacations (e.g., entering a low-power state), queueing theory can provide valuable insights into the trade-offs between energy consumption and system performance. The paper aims to analyse the functions of Discontinuous Reception (DRX), a PSM applied in communication systems using the Geo/Geo/1 queueing model with differentiated vacation and system disaster.

Wireless devices are usually battery-powered, and their battery life significantly influences user experience [1]. Despite recent improvements in battery capacity, many users still complain about the limited battery life of their devices. Consequently, substantial research is focused on mitigating this issue. While one approach involves improving battery efficiency, increasing attention is given to incorporating energy-saving techniques within communication protocols. To

address energy efficiency concerns, the 3rd Generation Partnership Project (3GPP) introduced Discontinuous Reception (DRX) [21] as a mechanism to reduce energy consumption in wireless devices, referred to as User Equipment (UE) in 3GPP standards. The DRX mechanism is designed to extend the battery life of UEs by alternating between active and sleep modes while maintaining essential network connectivity.

The primary goal of DRX is to minimize power consumption when no data transmission occurs, primarily by allowing the UE's radio receiver to enter sleep mode for extended periods. In LTE DRX, there are three key states: Inactivity (Idle), Active, and Sleep periods, which are further divided into Short Sleep and Long Sleep periods. The Inactive and active periods represent power-active modes, while the Short and Long Sleep periods are power-saving modes. The functions of various modes of DRX are as follows:

- Inactivity Period: During this period, the DRX Inactivity Timer is active, and the UE monitors the Physical Downlink Control Channel (PDCCH) for potential downlink transmissions. If a transmission is detected, the UE switches to Active mode. The Short Sleep period begins if the Inactivity Timer expires without any transmission.
- Short Sleep Period: The DRX Short Cycle Timer is activated in this phase. The UE periodically wakes up to monitor the PDCCH. If a downlink transmission is detected, the UE transitions back to Active mode and restarts the Inactivity Timer. If no transmission is detected, the UE returns to Short Sleep until the Short Cycle Timer expires.
- Long Sleep Period: Once the Short Cycle Timer expires, the UE enters a deeper power-saving state called the Long Sleep period. Similar to the Short Sleep phase, the UE periodically wakes up to monitor the PDCCH but remains in Long Sleep until a transmission is detected or other network triggers occur.

This paper focuses on downlink transmissions, where the UE transitions between sleep and active states based on the status of the DRX Inactivity Timer. Several analytical studies have investigated the performance of DRX using vacation queuing models. Yang and Lin [19] proposed a Markov chain model to evaluate the DRX performance in UMTS systems, assuming downlink packet arrivals follow a Poisson process. Yang et al. [20] applied a semi-Markov process to analyze the PSM of DRX. Turck et al. [4] analyzed the PSM of WiMAX and DRX using a discrete-time queuing model. Recently, numerous authors have examined the PSM of DRX using the Differentiated Vacation (DV) policy introduced by Ibe and Isijola [5]. Unlike the classical multiple vacation queuing model, where the server embarks on another vacation of the same duration if no customers are present at the end of a vacation, the DV model is characterized by varying durations for subsequent vacations. The DV model operates with three states: busy, type-I vacation (T_{V_1}), and type-II vacation (T_{V_2}). After a busy period, the server enters T_{V_1} if no customers are waiting. If the system remains empty at the end of the T_{V_1} period, the server transitions to a T_{V_2}.

The DV policy has been analysed in various contexts. Ibe and Isijola [5] employed the DV policy to study human behaviour. Vijayashree and Janani [18] explored time-dependent probabilities in an M/M/1/DV queuing model. Sampath and Liu [13] introduced the concept of customer impatience into the M/M/1/DV model, presenting both transient and steady-state analyses. Kumar et al. [8] discussed a congestion model incorporating the DV policy and customer impatience. More recently, Shapique et al. [11] applied the M/M/1/DV queuing model to examine the PSM in WiMAX.

Disaster in a queueing system can be considered server failures [2], which can severely disrupt the functioning of the systems. Integrating these events into queueing models allows us to evaluate system robustness, analyse performance under adverse conditions, and formulate effective risk management strategies. Network outages, power failures, and electromagnetic interference can critically affect essential infrastructure and emergency response mechanisms in communication systems. Many authors analysed queueing systems with disaster [6, 9, 15]. Recently, continuous-time queueing systems with DV policy and system Disaster captured the

interest of several authors. Dimitriou [3] applied the M/G/1 queueing model to analyse the PSM of DRX with fault-tolerant. Sudhesh et al. [16] applied an M/M/1/DV queueing system with system disaster to analyse the PSM of DRX.

The existing literature shows that many studies have focused on continuous-time models to analyse the PSMs of DRX, providing valuable insights into performance under various conditions. However, the discrete nature of data transmission in wireless communication systems requires a more detailed and granular analysis. This paper addresses this gap by examining DRX functions within a discrete-time framework, explicitly using a Geo/Geo/1 queueing model with differentiated vacations. By adopting this discrete-time approach, the analysis more accurately captures the intermittent nature of data arrivals and the discrete-time operations intrinsic to DRX, leading to a more precise evaluation of system behaviour.

The structure of the paper is as follows: Section 2 presents the model description. Section 3 discusses the steady-state analysis of various states. Section 4 presents various performance indices of the proposed model. Section 5 deals with a numerical illustration of the model. Finally, Section 6 presents the conclusion of the work.

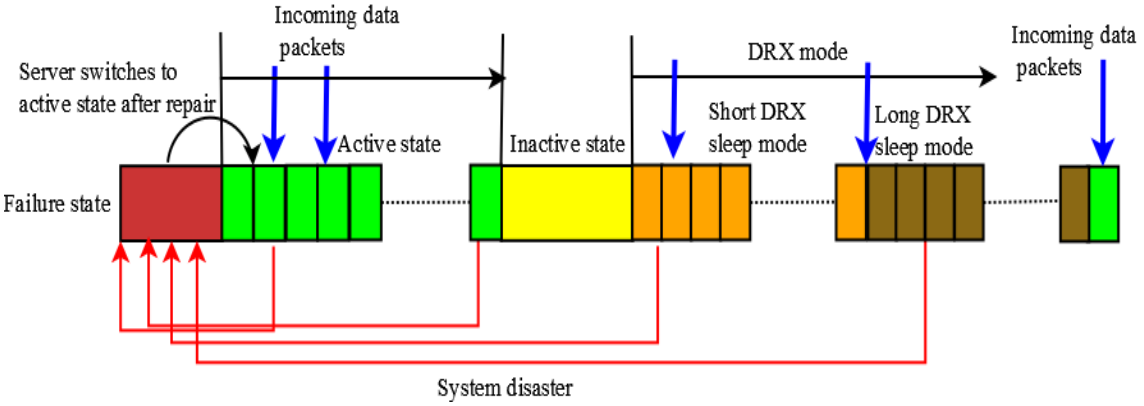


Figure 1: *Pictorial representation of the investigated model*

2. Model description

This section presents the model description of a Geo/Geo/1 queueing system with a DV policy, where the system is subject to disaster and repair mechanisms. Data packets join the buffer with parameter λ, and the server provides service to these packets with rate μ. After serving all data packets, the server enters an idle state. If no new packets arrive within a specified idle timer Ψ, the server transitions to a T_{V_1} state, activating a timer κ_1. If no packets are waiting at the end of the T_{V_1} timer κ_1, the server transitions to a T_{V_2} state with rate ϑ, activating a timer κ_2. Additionally, the system is subject to disasters that can occur at any state with rate ζ. When a disaster occurs, all packets in the system are lost, and the server enters a failure state. After repairs, which occur with rate ω, the server returns to the idle state. The system operates under a first-come, first-served service discipline. The detailed assumptions of the model are given below.

1. The data packets arrivals occur at the slot $t = n^-$, $n = 0, 1, 2$ and inter-arrival time τ follows a geometric distribution whose probability mass function is given by

$$P(\tau = n) = \lambda\bar{\lambda}^{n-1}, n = Z^+; \lambda \in (0, 1)$$

2. The probability mass function of service distribution is given by

$$P(S = n) = \mu\bar{\mu}^{n-1}, n = Z^+; \mu \in (0, 1)$$

3. The probability mass function of T_{V_1} and T_{V_2} is given by

$$P\left(V_i = n\right) = \begin{cases} \kappa_1 \cdot \bar{\kappa_1}^{n-1}, & n \in \mathbb{Z}^+, \kappa_1 \in (0,1) \text{ if } i = 1 \\ \kappa_2 \cdot \bar{\kappa_2}^{n-1}, & n \in \mathbb{Z}^+, \kappa_2 \in (0,1) \text{ if } i = 2. \end{cases}$$

Let N_n represent the number of data packets in the system at time $t = n^+$ and M_n represent the state of the system at time $t = n^+$. We define

$$M_n = \begin{cases} A, & \text{if the server is in busy state at time } t = n^+, \\ B, & \text{if the server is in type-I vacation state at time } t = n^+, \\ C, & \text{if the server is in type-II vacation state at time } t = n^+, \\ F, & \text{if the server is in failure state.} \end{cases}$$

It is observed that $\left\{N_n, M_n\right\}$ is a Markov chain with state space

$$\varpi = \left\{(n, A), n = 0, 1, 2, ..\right\} \cup \left\{(n, B), n = 0, 1, 2, ..\right\} \cup \left\{(n, C), n = 0, 1, 2, ..\right\} \cup F.$$

The one-step transition probabilities of $\left\{N_n, M_n\right\}$ is represented by

$$S_{(n_1,m_1),(n_2,m_2)} = P\left[X_{n+1} = (n_2, m_2) / X_n = (n_1, m_1)\right], n_1, n_2 \in N_n ; m_1, m_2 \in M_n.$$

If $n_1 = n$ then $n_2 = n - 1$ or n or $n + 1$
State transitions and their corresponding transition parameters are presented in Table 1.

Table 1: *State Transitions and its corresponding transition parameter*

$S_{(n_1,m_1),(n_2,m_2)}$	**Transition parameter**
$S_{(n,A),(n+1,A)}$	$\begin{cases} \lambda\bar{\Psi}\bar{\zeta}, n = 0 \\ \lambda\bar{\mu}\bar{\zeta}, n = 1, 2, 3, ... \end{cases}$
$S_{(n,A),(n-1,A)}$	$\bar{\lambda}\mu\bar{\zeta}, n = 1, 2, 3, ...$
$S_{(n,B),(n+1,B)}$	$\begin{cases} \lambda\bar{\vartheta}\bar{\zeta}, n = 0 \\ \lambda\bar{\kappa}_1\bar{\zeta}, n = 1, 2, 3, ... \end{cases}$
$S_{(n,C),(n+1,C)}$	$\begin{cases} \lambda\bar{\zeta}, n = 0 \\ \lambda\bar{\kappa}_2\bar{\zeta}, n = 1, 2, 3, ... \end{cases}$
$S_{(n,A),(n,A)}$	$\begin{cases} \alpha_1, n = 0 \\ \alpha_5, n = 1, 2, 3, ... \end{cases}$
$S_{(n,B),(n,B)}$	$\begin{cases} \alpha_2, n = 0 \\ \alpha_3, n = 1, 2, 3, ... \end{cases}$
$S_{(n,C),(n,C)}$	$\begin{cases} \bar{\lambda}, n = 0 \\ \alpha_4, n = 1, 2, 3, ... \end{cases}$
$S_{(0,A),(0,B)}$	$\Psi\bar{\lambda}\bar{\zeta}$
$S_{(0,B),(0,C)}$	$\vartheta\bar{\lambda}\bar{\zeta}$

$S_{(n_1,m_1),(n_2,m_2)}$	**Transition parameters**
$S_{(n,B),(n,A)}$	$\bar{\lambda}\kappa_1\bar{\zeta}, n = 1,2,3,...$
$S_{(n,C),(n,A)}$	$\bar{\lambda}\kappa_2\bar{\zeta}, n = 1,2,3,...$
$S_{(n,A),(F)}$	$\begin{cases}\beta_1, n = 0\\ \beta_5, n = 1,2,3,...\end{cases}$
$S_{(n,B),(F)}$	$\begin{cases}\beta_2, n = 0\\ \beta_3, n = 1,2,3,...\end{cases}$
$S_{(n,C),(F)}$	$\begin{cases}\zeta, n = 0\\ \beta_4, n = 1,2,3,...\end{cases}$
$S_{(F),(0,A)}$	ω
$S_{(F,F)}$	$\bar{\omega}$

where

$$\beta_1 = \lambda\bar{\Psi}\zeta + \bar{\lambda}\Psi\zeta + \lambda\Psi\zeta + \bar{\lambda}\bar{\Psi}\zeta, \beta_2 = \lambda\bar{\vartheta}\zeta + \bar{\lambda}\vartheta\zeta + \lambda\vartheta\zeta + \bar{\lambda}\bar{\vartheta}\zeta,$$
$$\beta_3 = \lambda\bar{\kappa}_1\zeta + \bar{\lambda}\kappa_1\zeta + \lambda\kappa_1\zeta + \bar{\lambda}\bar{\kappa}_1\zeta, \beta_4 = \lambda\bar{\kappa}_2\zeta + \bar{\lambda}\kappa_2\zeta + \lambda\kappa_2\zeta + \bar{\lambda}\bar{\kappa}_2\zeta,$$
$$\beta_5 = \lambda\bar{\mu}\zeta + \bar{\lambda}\mu\zeta + \lambda\mu\zeta + \bar{\lambda}\bar{\mu}\zeta, \alpha_1 = \left(\bar{\lambda}\bar{\Psi} + \lambda\Psi\right)\bar{\zeta},$$
$$\alpha_2 = \left(\bar{\lambda}\bar{\vartheta} + \lambda\vartheta\right)\bar{\zeta}, \alpha_3 = \left(\bar{\lambda}\bar{\kappa}_1 + \lambda\kappa_1\right)\bar{\zeta} \text{ and } \alpha_4 = \left(\bar{\lambda}\bar{\kappa}_2 + \lambda\kappa_2\right)\bar{\zeta}.$$

The Kolmogorov equations for the investigated model are as follows:

$$\pi_F = (1-\omega)\,\pi_F + \beta_1\pi_{0,A} + \beta_5\sum_{n=1}^{\infty}\pi_{n,A} + \beta_2\pi_{0,B} + \beta_3\sum_{n=1}^{\infty}\pi_{n,B} + \zeta\pi_{0,C} + \beta_4\sum_{n=1}^{\infty}\pi_{n,C}, \quad (1)$$
$$\pi_{0,A} = \alpha_1\pi_{0,A} + \omega\pi_F + \bar{\lambda}\mu\bar{\zeta}\pi_{1,A}, \quad (2)$$
$$\pi_{1,A} = \lambda\bar{\Psi}\bar{\zeta}\pi_{0,A} + \alpha_5\pi_{1,A} + \bar{\lambda}\mu\bar{\zeta}\pi_{2,A} + \bar{\lambda}\kappa_1\bar{\zeta}\pi_{1,B} + \bar{\lambda}\kappa_2\bar{\zeta}\pi_{1,C}, \quad (3)$$
$$\pi_{n,A} = \alpha_5\pi_{n,A} + \lambda\bar{\mu}\bar{\zeta}\pi_{n-1,A} + \bar{\lambda}\mu\bar{\zeta}\pi_{n+1,A} + \bar{\lambda}\kappa_1\bar{\zeta}\pi_{n,B} + \bar{\lambda}\kappa_2\bar{\zeta}\pi_{n,B}, n = 2,3,4,..., \quad (4)$$
$$\pi_{0,B} = \alpha_2\pi_{0,B} + \Psi\bar{\lambda}\bar{\zeta}\pi_{0,A}, \quad (5)$$
$$\pi_{1,B} = \alpha_3\pi_{1,B} + \lambda\bar{\vartheta}\bar{\zeta}\pi_{0,B}, \quad (6)$$
$$\pi_{n,B} = \alpha_3\pi_{n,B} + \lambda\bar{\kappa}_1\bar{\zeta}\pi_{n-1,B}, n = 2,3,4,..., \quad (7)$$
$$\pi_{0,C} = \bar{\lambda}\bar{\zeta}\pi_{0,C} + \vartheta\bar{\lambda}\bar{\zeta}\pi_{0,B}, \quad (8)$$
$$\pi_{1,C} = \alpha_4\pi_{1,C} + \lambda\bar{\zeta}\pi_{0,C}, \quad (9)$$
$$\pi_{n,C} = \alpha_4\pi_{n,C} + \lambda\bar{\kappa}_2\bar{\zeta}\pi_{n-1,C}, n = 2,3,4,.... \quad (10)$$

3. Stationary analysis

The steady state probabilities of $\pi_{n,B}$, $\pi_{n,C}$, $\pi_{n,A}$ and π_F are presented in this section.

3.1. Evaluation of vacation state probabilities $\pi_{n,B}$ and $\pi_{n,C}$

Using Equations (5) − (7), we get

$$\pi_{n,B} = \begin{cases}\dfrac{\Psi\bar{\lambda}\bar{\zeta}}{1-\alpha_2}\pi_{0,A}, & \text{if } n = 0,\\ \dfrac{\lambda\bar{\lambda}\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)(1-\alpha_3)}\bar{\zeta}_1^{n-1}\pi_{0,A}, & \text{if } n = 1,2,3,\ldots\end{cases} \quad (11)$$

Similarly, using Equations (8) − (10), we obtain

$$\pi_{n,C} = \begin{cases} \dfrac{\vartheta\Psi\bar{\lambda}^2\bar{\zeta}^2}{(1-\bar{\lambda}\bar{\zeta})(1-\alpha_2)}\pi_{0,A}, & \text{if } n=0, \\ \dfrac{\lambda\bar{\lambda}^2\vartheta\Psi\bar{\zeta}^3}{(1-\bar{\lambda}\bar{\zeta})(1-\alpha_2)(1-\alpha_4)}\bar{\zeta}_2^{n-1}\pi_{0,A}, & \text{if } n=1,2,3,\ldots \end{cases} \tag{12}$$

where $\phi_1 = \dfrac{\lambda\bar{\kappa}_1\bar{\zeta}}{1-\alpha_3}$ and $\phi_2 = \dfrac{\lambda\bar{\kappa}_2\bar{\zeta}}{1-\alpha_4}$. Thus we have expressed the vacation state probabilities $\pi_{n,B}$ and $\pi_{n,C}$ in-terms of $\pi_{0,A}$. An explicit expression for $\pi_{0,A}$ is presented in Section 3.2.

3.2. Evaluation of $\pi_{n,A}$

The busy state probability $\pi_{n,A}$ is presented in this section. We define a generating function as follows:

$$H_z(n) = \sum_{n=1}^{\infty}\pi_{n,A}z^n.$$

Multiplying suitable powers of z on Equations (3) and (4), we get

$$\left\{\lambda\bar{\mu}\bar{\zeta}z^2 - (1-\alpha_5)z + \bar{\lambda}\mu\bar{\zeta}\right\}H_z(n) = \bar{\lambda}\mu\bar{\zeta}\pi_{1,A}z - \lambda\bar{\Psi}\bar{\zeta}\pi_{0,A}z^2 - \bar{\lambda}\kappa_1\bar{\zeta}z\sum_{n=1}^{\infty}\pi_{n,B}z^n - \bar{\lambda}\kappa_1\bar{\zeta}z\sum_{n=1}^{\infty}\pi_{n,C}z^n,$$

where $\alpha_5 = \left(\bar{\lambda}\bar{\mu} + \lambda\mu\right)\bar{\zeta}$.
The above equation can be expressed as

$$\lambda\bar{\mu}\bar{\zeta}(z-\alpha)(z-\beta)H_z(n) = \bar{\lambda}\mu\bar{\zeta}\pi_{1,A}z - \lambda\bar{\Psi}\bar{\zeta}\pi_{0,A}z^2 - \bar{\lambda}\kappa_1\bar{\zeta}z\sum_{n=1}^{\infty}\pi_{n,B}z^n\bar{\lambda}\kappa_2\bar{\zeta}z\sum_{n=1}^{\infty}\pi_{n,C}z^n, \tag{13}$$

where

$$\alpha = \frac{(1-\alpha_5) + \sqrt{(1-\alpha_5)^2 - 4\lambda\bar{\lambda}\mu\bar{\mu}\bar{\zeta}^2}}{2\lambda\bar{\mu}\bar{\zeta}},$$

and

$$\beta = \frac{(1-\alpha_5) - \sqrt{(1-\alpha_5)^2 - 4\lambda\bar{\lambda}\mu\bar{\mu}\bar{\zeta}^2}}{2\lambda\bar{\mu}\bar{\zeta}}.$$

It is observed that $\beta < 1$ for any $0 < \lambda < 1$, $0 < \mu < 1$ and $0 < \zeta < 1$. Setting $z = \beta$ in Equation (13), we obtain

$$\bar{\lambda}\mu\bar{\zeta}\pi_{1,A} = \lambda\bar{\Psi}\bar{\zeta}\pi_{0,A}\beta + \bar{\lambda}\kappa_1\bar{\zeta}\sum_{n=1}^{\infty}\pi_{n,B}\beta^n + \bar{\lambda}\kappa_2\bar{\zeta}\sum_{n=1}^{\infty}\pi_{n,C}\beta^n. \tag{14}$$

Applying Equation (14) in Equation (13), we get

$$H_z(n) = \frac{\bar{\zeta}}{\bar{\mu}\alpha}\sum_{n=0}^{\infty}\left[\frac{1}{\alpha^n} + \frac{\lambda^2\kappa_1\Psi\bar{\vartheta}\bar{\zeta}}{(1-\alpha_2)(1-\alpha_3)(1-\beta\phi_1)}\sum_{k=0}^{n}\frac{\bar{\zeta}_1^{n-k}}{\alpha^k} + \frac{\lambda^3\kappa_2\Psi\vartheta\bar{\zeta}^2}{(1-\alpha_2)(1-\alpha_4)} \times \frac{1}{(1-\bar{\lambda}\bar{\zeta})(1-\beta\phi_2)}\sum_{k=0}^{n}\frac{\bar{\zeta}_2^{n-k}}{\alpha^k}\right]z^{n+1}\pi_{0,A}. \tag{15}$$

Comparing the coefficient of z^n on both sides of Equation (15), we obtain

$$\pi_{n,A} = \frac{1}{\bar{\mu}\alpha^n}\left[\bar{\Psi} + \frac{\bar{\lambda}^2\kappa_1\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)(1-\alpha_3)(1-\beta\phi_1)(\alpha\phi_1-1)}\left\{(\alpha\phi_1)^n - 1\right\} + \frac{\bar{\lambda}^3\kappa_2\Psi\vartheta\bar{\zeta}^3}{(1-\alpha_2)(1-\alpha_4)\left(1-\bar{\lambda}\bar{\zeta}\right)(1-\beta\phi_2)(\alpha\phi_2-1)}\left\{(\alpha\phi_2)^n - 1\right\}\right]\pi_{0,A}. \tag{16}$$

The result (16) presents the explicit expression for the busy state probability in terms of $\pi_{0,A}$

3.3. Evaluation of Failure state probability π_F and $\pi_{0,A}$

The failure state probability π_F is obtained by using the results (11), (12) and (16) in Equation (1).

$$\pi_F = \frac{1}{\omega}\left[\beta_1 + \frac{\beta_2\Psi\bar{\lambda}\bar{\zeta}}{1-\alpha_2} + \frac{\beta_5}{\bar{\mu}\,(\alpha-1)}\left\{1 + \frac{\bar{\lambda}^2\kappa_1\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)\,(1-\alpha_3)\,(1-\beta\phi_1)\,(1-\phi_1)}\right.\right.$$
$$\left.\left. + \frac{\bar{\lambda}^3\kappa_2\Psi\vartheta\bar{\zeta}^3}{(1-\alpha_2)\,(1-\alpha_4)\,\left(1-\bar{\lambda}\bar{\zeta}\right)\,(1-\beta\phi_2)\,(1-\phi_2)}\right\} + \frac{\zeta\vartheta\Psi\bar{\lambda}^2\bar{\zeta}^2}{\left(1-\bar{\lambda}\bar{\zeta}\right)\,(1-\alpha_2)}\right.$$
$$\left. + \frac{\beta_3\lambda\bar{\lambda}\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)\,(1-\alpha_3)\,(1-\phi_1)} + \frac{\beta_4\lambda\bar{\lambda}^2\vartheta\Psi\bar{\zeta}^3}{\left(1-\bar{\lambda}\bar{\zeta}\right)\,(1-\alpha_2)\,(1-\alpha_4)\,(1-\phi_2)}\right]\pi_{0,A} \tag{17}$$

An explicit expression for the idle state probability $\pi_{0,A}$ can be as follows: The normalisation condition for the investigated model is given by

$$\sum_{n=0}^{\infty}\pi_{n,A} + \sum_{n=0}^{\infty}\pi_{n,B} + \sum_{n=0}^{\infty}\pi_{n,C} + \pi_F = 1$$

Substituting the results (11), (12), (16) and (17) in the normalisation condition, we obtain

$$\pi_{0,A} = \frac{1}{D}, \tag{18}$$

where

$$D = \left[\left(1+\frac{\beta_1}{\omega}\right) + \left(1+\frac{\beta_2}{\omega}\right)\frac{\Psi\bar{\lambda}\bar{\zeta}}{1-\alpha_2} + \left(1+\frac{\beta_3}{\omega}\right)\frac{\lambda\bar{\lambda}\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)\,(1-\alpha_3)\,(1-\phi_1)}\right.$$
$$+ \left(1+\frac{\zeta}{\omega}\right)\frac{\vartheta\Psi\bar{\lambda}^2\bar{\zeta}^2}{\left(1-\bar{\lambda}\bar{\zeta}\right)\,(1-\alpha_2)} + \left(1+\frac{\beta_4}{\omega}\right)\frac{\lambda\bar{\lambda}^2\vartheta\Psi\bar{\zeta}^3}{\left(1-\bar{\lambda}\bar{\zeta}\right)\,(1-\alpha_2)\,(1-\alpha_4)\,(1-\phi_2)}$$
$$+ \left(1+\frac{\beta_5}{\omega}\right)\frac{1}{\bar{\mu}\,(\alpha-1)}\left\{\bar{\Psi} + \frac{\bar{\lambda}^2\kappa_1\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)\,(1-\alpha_3)\,(1-\beta\phi_1)\,(1-\phi_1)}\right.$$
$$\left.\left. + \frac{\bar{\lambda}^3\kappa_2\Psi\vartheta\bar{\zeta}^3}{(1-\alpha_2)\,(1-\alpha_4)\,\left(1-\bar{\lambda}\bar{\zeta}\right)\,(1-\beta\phi_2)\,(1-\phi_2)}\right\}\right].$$

Remark It is noted that if $\zeta = 0$ and $\Psi = 0$ then the result (16) reduces to

$$\pi_{n,A} = \frac{1}{1-\mu}\left\{\frac{\lambda}{\mu}\left(\frac{1-\mu}{1-\lambda}\right)\right\}^n \pi_{0,A}$$

where

$$\pi_{0,A} = 1 - \frac{\lambda}{\mu}$$

The above expression coincides with Equation (24) of Taha [17]

4. Performance measures

This section presents the performance measures of the investigated system

4.1. Expected system size

Let $E(N_{DP})$ denote the expected number of data packets in the system, then

$$\begin{aligned}E\left(N_{DP}\right) &= \sum_{n=1}^{\infty} n\left(\pi_{n,A}+\pi_{n,B}+\pi_{n,C}\right)\\ &= \frac{1}{\bar{\mu}}\Bigg[\frac{\Psi\alpha}{(\alpha-1)^2}+\frac{\bar{\lambda}^2\kappa_1\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)(1-\alpha_3)(1-\beta\phi_1)(\alpha\phi_1-1)}\left\{\frac{\phi_1}{(1-\phi_1)^2}-\frac{\alpha}{(\alpha-1)^2}\right\}\\ &+\frac{\bar{\lambda}^3\kappa_2\Psi\vartheta\bar{\zeta}^3}{(1-\alpha_2)(1-\alpha_4)\left(1-\bar{\lambda}\bar{\zeta}\right)(1-\beta\phi_2)(\alpha\phi_2-1)}\left\{\frac{\phi_2}{(1-\phi_2)^2}-\frac{\alpha}{(\alpha-1)^2}\right\}\\ &+\frac{\lambda\bar{\lambda}\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)(1-\alpha_3)\left(1-\bar{\zeta}_1\right)^2}+\frac{\lambda\bar{\lambda}^2\vartheta\Psi\bar{\zeta}^3}{\left(1-\bar{\lambda}\bar{\zeta}\right)(1-\alpha_2)(1-\alpha_4)\left(1-\bar{\zeta}_2\right)^2}\Bigg]\pi_{0,A}.\end{aligned}$$

4.2. Probability that the server is in vacation state

Let $\pi_{\bullet,B}$ and $\pi_{\bullet,C}$ denote the probability that the server is in short and long vacation respectively. Then

$$\pi_{\bullet,i}=\sum_{n=0}^{\infty}\pi_{n,i}=\begin{cases}\dfrac{\Psi\bar{\lambda}\bar{\zeta}}{1-\alpha_2}\left[1+\dfrac{\lambda\bar{\vartheta}\bar{\zeta}}{(1-\alpha_3)(1-\zeta_1)}\right]\pi_{0,A}, & \text{if } i=B\\[2ex] \dfrac{\vartheta\Psi\bar{\lambda}^2\bar{\zeta}^2}{\left(1-\bar{\lambda}\bar{\zeta}\right)(1-\alpha_2)}\left[1+\dfrac{\lambda^2\bar{\zeta}}{(1-\alpha_4)(1-\zeta_2)}\right]\pi_{0,A}, & \text{if } i=C\end{cases}$$

4.3. Probability that the server is in busy

Let $\pi_{\bullet,A}$ denote the probability that the server is busy, then

$$\begin{aligned}\pi_{\bullet,A} &= \sum_{n=1}^{\infty}\pi_{n,A}\\ &= \frac{1}{\bar{\mu}(\alpha-1)}\Bigg[\Psi+\frac{\bar{\lambda}^2\kappa_1\Psi\bar{\vartheta}\bar{\zeta}^2}{(1-\alpha_2)(1-\alpha_3)(1-\beta\phi_1)(1-\phi_1)}+\frac{\bar{\lambda}^3\kappa_2\Psi\vartheta\bar{\zeta}^3}{(1-\alpha_2)(1-\alpha_4)\left(1-\bar{\lambda}\bar{\zeta}\right)}\\ &\times\frac{1}{(1-\beta\phi_2)(1-\phi_2)}\Bigg]\pi_{0,A}\end{aligned}$$

The expression for $\pi_{0,A}$ is presented in the result (18).

5. Numerical illustration

This section presents the numerical illustration for the investigated model. The values of the parameter as choose as follows: $\lambda=0.3$, $\mu=0.7$, $\psi=0.2$, $\vartheta=0.2$, $\zeta=0.01$, $\omega=0.4$, $\kappa_1=0.2$ and $\kappa_2=0.4$. Figure 2 presents the behaviour of $\pi_{n,B}$ for various arrival rates λ. The graph shows that as λ increases, the probability $\pi_{n,B}$ decreases, indicating that higher arrival rates result in lower probabilities of having more packets in the system during the T_{V_1} period.

Figure 3 displays the probability curves of $\pi_{n,B}$ for different values of κ_1. It is evident that as κ_1 increases, the probability $\pi_{n,B}$ for $n=0$ rises slightly, indicating a higher likelihood of an empty system when the server switches more frequently from the vacation to the busy state. This suggests that with faster switching, the system is cleared more efficiently, reducing the chance of data packets accumulating during the server™s vacation period. On the other hand, for $n=1$ to $n=10$, the values of $\pi_{n,B}$ consistently decrease as κ_1 increases. This trend indicates that when the server switches back to the busy state more frequently, the system spends less time in the

vacation state with non-zero packets. Consequently, the probability of data packets in the system during the vacation state decreases with increasing κ_1.

In Figure 4, we can see the behaviour of the T_{V_2} state for different arrival rates λ. As the arrival rate λ increases, the probability $\pi_{n,C}$ generally decreases when $n = 0$, indicating that the chance of having an empty system is lower at higher arrival rates. As the value of n increases, the values of $\pi_{n,C}$ consistently decrease across all λ values. This suggests that as the arrival rate increases, the likelihood of having a larger number of packets in the vacation state becomes smaller.

In Figure 5, the probability curves of $\pi_{n,C}$ are shown for different values of κ_2. As κ_2 increases, the probabilities $\pi_{n,C}$ decrease for higher values of n, indicating that a higher switching rate reduces the likelihood of a large number of packets building up in the system. This is because as κ_2 increases, the server switches from vacation to busy mode more quickly, processing data packets more frequently and reducing congestion during the vacation state. For smaller values of n, such as $n = 0, 1, 2$, the differences between the probabilities for different κ_2 values are less noticeable. However, as n increases, particularly beyond $n = 4$, the impact of a higher κ_2 becomes more apparent, with a more pronounced decline in $\pi_{n,C}$.

In Figure 6, we can see how the busy state $\pi_{n,A}$ behaves for different values of λ. When $n = 0$, the probability $\pi_{0,A}$ decreases as λ increases, indicating that a higher arrival rate makes it less likely for the system to be idle. As n increases, the probabilities $\pi_{n,A}$ generally decrease for larger n, meaning it becomes less likely for many packets to accumulate in the system. However, for intermediate values of n, like $n = 1$ and $n = 2$, the probabilities slightly increase with higher λ, suggesting that more packets are expected in the system when the arrival rate is higher. Overall, the behaviour of $\pi_{n,A}$ across different λ values indicates that higher arrival rates increase the likelihood of having more data packets in the system during the busy state.

Figure 7 shows the probability curves of the busy state $\pi_{n,A}$ for various values of μ. Observing the probabilities, we see that for $n = 0$, $\pi_{0,A}$ increases as μ increases. This trend suggests that a higher service rate leads to a higher probability of the system being idle, likely because packets are processed more quickly, reducing the chance of any being present in the busy state. For values of n greater than 0, the probability $\pi_{n,A}$ generally decreases as μ increases. This indicates that higher service rates effectively reduce the likelihood of the system having many packets, as the server can handle incoming traffic more efficiently.

Figure 8 presents the probability of having n packets in the busy state $\pi_{n,A}$ for various values of κ_1. As the rate κ_1 increases, the probability $\pi_{n,A}$ of having n data packets in the busy state shows a clear pattern. For $n = 0$ and $n = 1$, $\pi_{n,A}$ increases, meaning the system is more likely to have fewer packets as the server switches faster. For instance, $\pi_{0,A}$ rises as κ_1 increases from 0.1 to 0.4. However, for larger n, $\pi_{n,A}$ decreases with higher κ_1, indicating that faster switching reduces the likelihood of accumulating more packets in the busy state. This reflects the server™s improved efficiency in managing data packets.

Figure 9 presents the behaviour $\pi_{n,A}$ for various values of κ_2. It is observed that $\pi_{n,A}$ decreases as n increases, indicating that the probability of having a higher number of packets in the system diminishes as n increases. Additionally, as κ_2 increases, the probabilities for smaller values of n rise, signifying that when the server switches more quickly from the vacation state to the busy state, the system is more likely to have fewer packets. Conversely, for larger values of n, such as $n = 10$, the probability decreases more rapidly for higher κ_2. This trend highlights the role of κ_2 in maintaining system stability, as higher switching rates lead to lower probabilities of larger packet queues, suggesting that the system is more stable with fewer packets when the server can quickly switch from vacation to the busy state.

Figure 10 displays the mean system size $E(N_s)$, which represents the average number of data packets in the system, for different values of the arrival rate λ and the disaster rate ζ. From the figure, we observe that the mean system size increases as λ increases, regardless of the disaster rate ζ. This indicates that higher arrival rates of packets lead to larger system sizes, which is expected as more packets are entering the system. Additionally, for each value of λ, increasing ζ results in a smaller system size, reflecting the effect of more frequent disasters clearing the system.

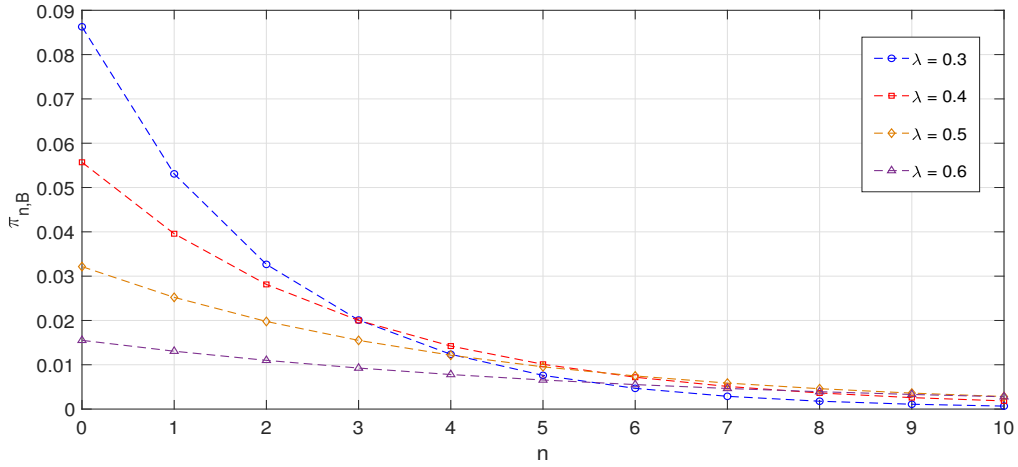


Figure 2: *Probabilities of the type-I vacation $\pi_{n,B}$ for different values of λ.*

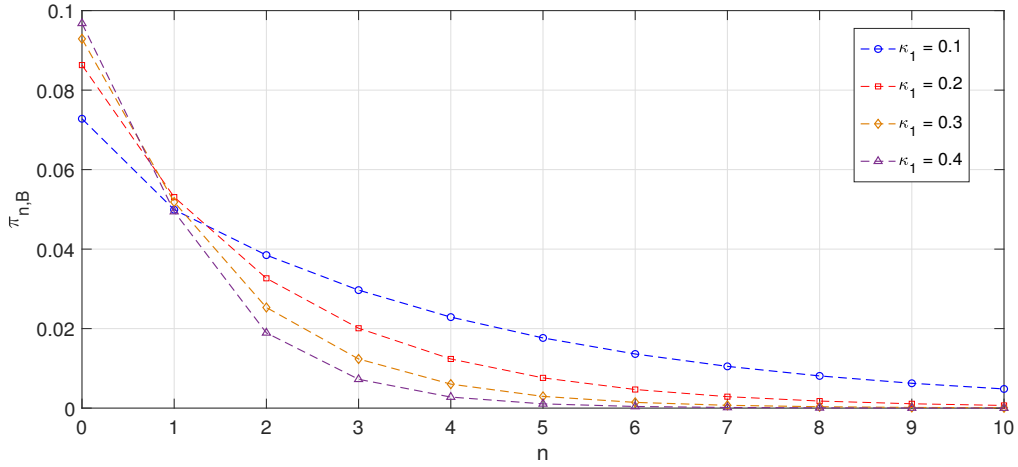


Figure 3: *Probabilities of the type-I vacation $\pi_{n,B}$ for different values of κ_1.*

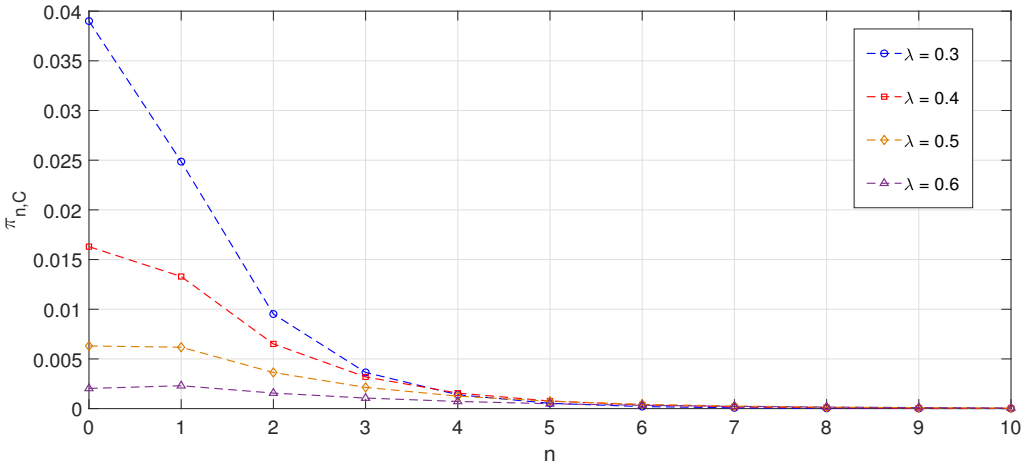


Figure 4: *Probabilities of the type-II vacation $\pi_{n,C}$ for different values of λ.*

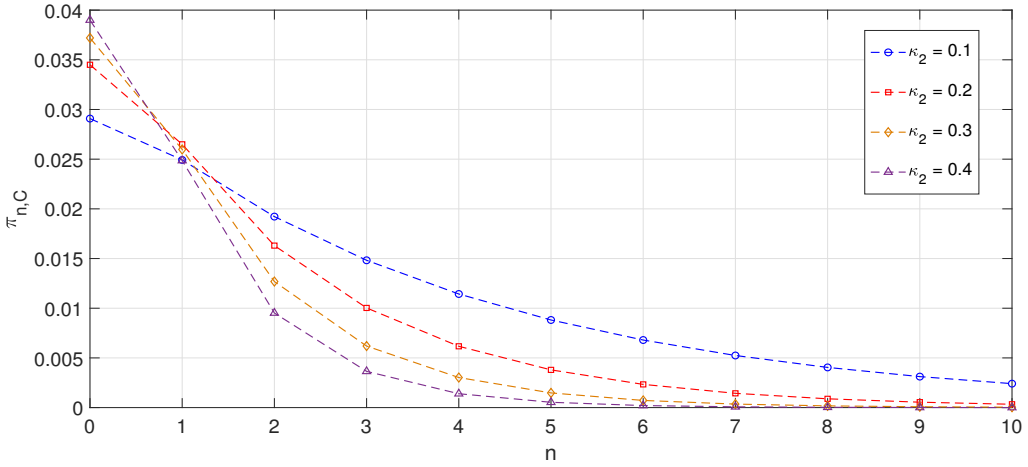


Figure 5: *Probabilities of the type-II vacation $\pi_{n,C}$ for different values of κ_2.*

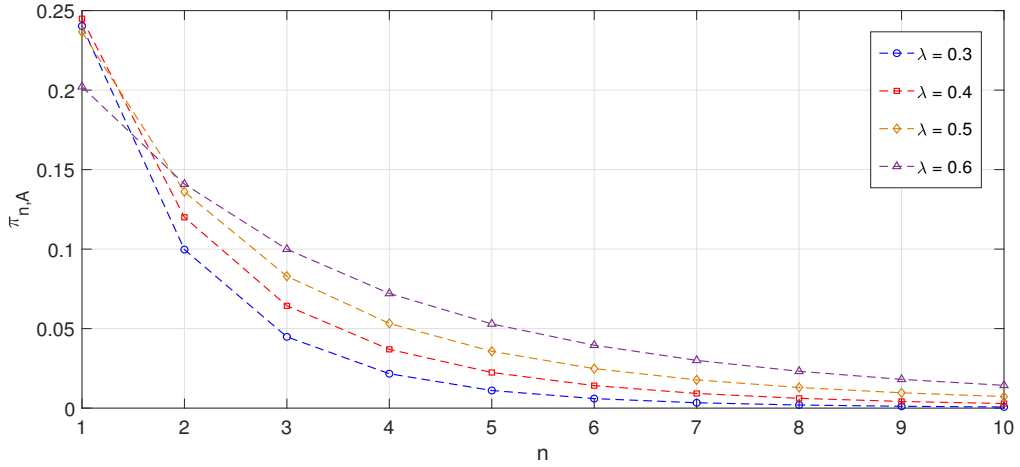


Figure 6: *Probabilities of the busy state $\pi_{n,A}$ for different values of λ.*

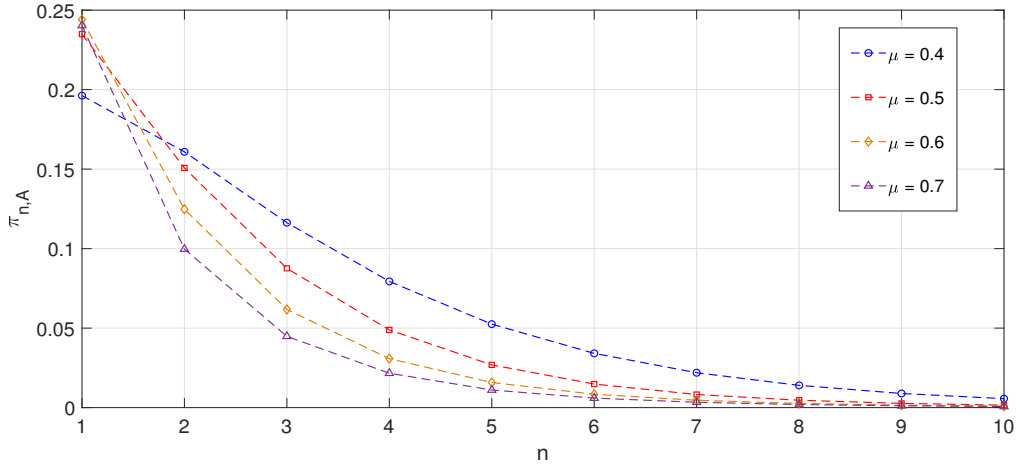


Figure 7: *Probabilities of the busy state $\pi_{n,A}$ for different values of μ.*

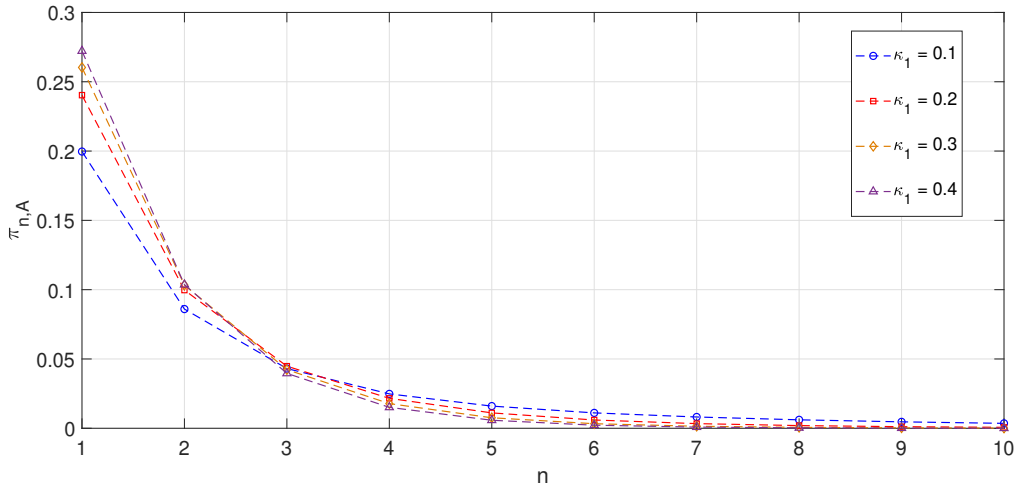


Figure 8: *Probabilities of the busy state $\pi_{n,A}$ for different values of κ_1.*

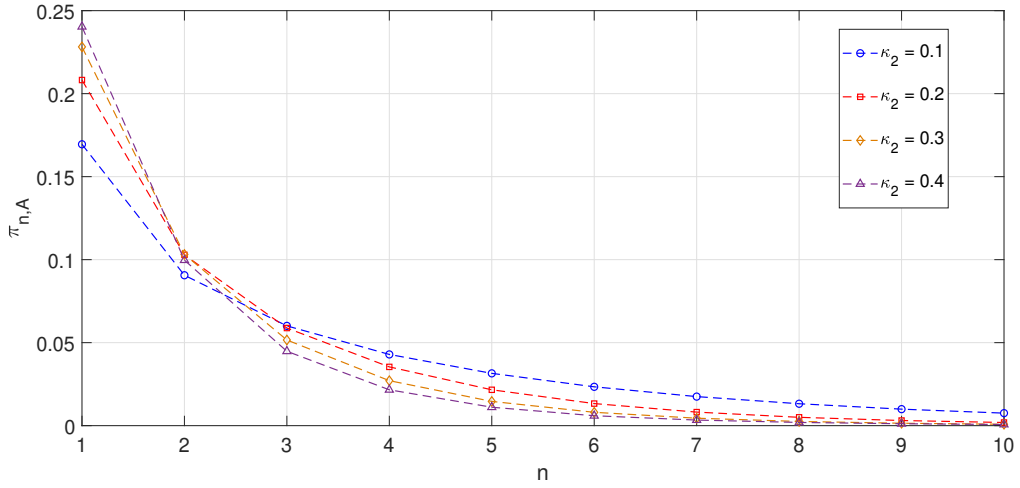


Figure 9: *Probabilities of the busy state $\pi_{n,A}$ for different values of κ_2.*

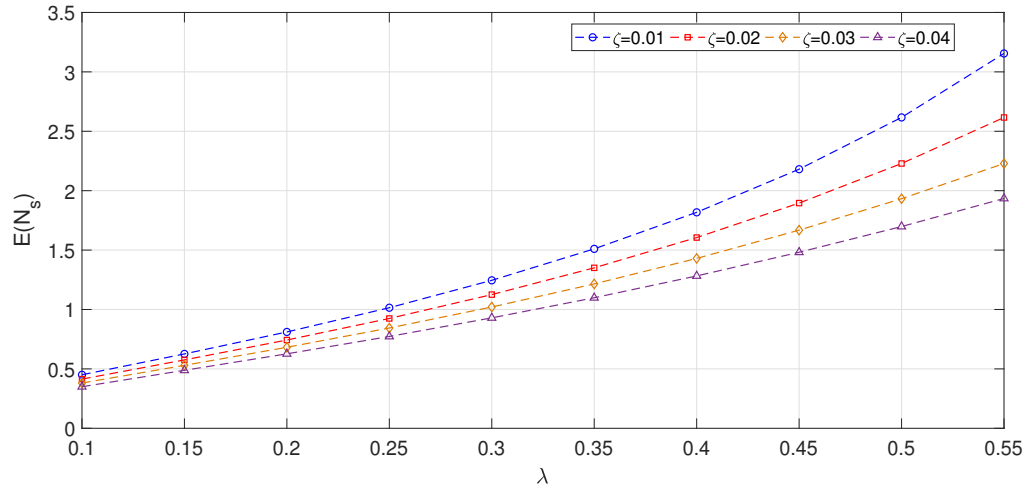


Figure 10: *Expected system size for different values of ζ.*

6. Conclusion

This paper analyzed the DRX mechanism, a critical power-saving feature in wireless communication systems, using a discrete-time Geo/Geo/1 queueing model with DV policy and disaster. By incorporating the DV policy, we captured the varying duration of sleep periods, providing a more granular and realistic evaluation of DRX performance compared to traditional continuous-time models. Our findings demonstrate that the DRX mechanism, when modelled through differentiated vacations, can effectively balance energy efficiency and network performance. The current study focused on a single server setup. One can extend this work by investigating multi-server configurations.

References

[1] Ayman E (2014) Extending the battery life of smartphones and tablets: a practical approach to optimizing the lte network. IEEE Veh Technol Mag 9(2):38“49.
[2] Di Crescenzo, A., Giorno, V., Nobile, A. G., & Ricciardi, L. M. (2003). On the M/M/1 queue with catastrophes and its continuous approximation. Queueing Systems, 43, 329-347.
[3] Dimitriou, I. (2016). Queueing analysis of the DRX power saving mechanism in fault-tolerant 3GPP LTE wireless networks. Annals of Operations Research, 239(2), 521-552.
[4] De Turck, K., De Vuyst, S., Fiems, D.,Wittevrongel, S., & Bruneel, H. (2012). Performance analysis of sleep mode mechanisms in the presence of bidirectional traffic. Computer Networks, 56, 2494“2505.
[5] Ibe, O. C., & Isijola, O. A. (2014). M/M/1 multiple vacation queueing systems with differentiated vacations. Modelling and Simulation in Engineering, 2014(1), 158247.
[6] Jain, G., & Sigman, K. (1996). A Pollaczek“Khintchine formula for M/G/1 queues with disasters. Journal of Applied Probability, 33(4), 1191-1200.
[7] Kalita, P., & Selvamuthu, D. (2023). Stochastic modelling of sleeping strategy in 5G base station for energy efficiency. Telecommunication Systems, 83(2), 115-133.
[8] Kumar, A., Kaswan, S., Devanda, M., & Shekhar, C. (2023). Transient analysis of queueing-based congestion with differentiated vacations and customer™s impatience attributes. Arabian Journal for Science and Engineering, 48(11), 15655-15665.
[9] Lee, D. H., & Yang, W. S. (2013). The N-policy of a discrete time Geo/G/1 queue with disasters and its application to wireless sensor networks. Applied Mathematical Modelling, 37(23), 9722-9731.
[10] Mittal, N., Jain, V., & Dharmaraja, S. (2023). Power efficient stochastic modeling for Narrowband Internet of Things devices in 5G networks. Transactions on Emerging Telecommunications Technologies, 34(9), e4825.
[11] Mohammed Shapique, A., Sudhesh, R., & Dharmaraja, S. (2024). Transient Analysis of a Modified Differentiated Vacation Queueing System for Energy-Saving in WiMAX. Methodology and Computing in Applied Probability, 26(3), 23.
[12] Raj, R., & Dharmaraja, S. (2023). Stochastic modelling of multi-layer HAP-LEO systems in 6G for energy saving: An analytical approach. Computer Communications, 210, 22-34.
[13] Suranga Sampath, M. I. G., & Liu, J. (2020). Impact of customers™ impatience on an M/M/1 queueing system subject to differentiated vacations with a waiting server. Quality Technology & Quantitative Management, 17(2), 125-148.
[14] Sudhesh, R., & Mohammed Shapique, A. (2022). Transient analysis of power management in wireless sensor network with start-up times and threshold policy. Telecommunication Systems, 80(1), 1-16.
[15] Sudhesh, R., & Vaithiyanathan, A. (2021). Stationary analysis of infinite queueing system with two-stage network server. RAIRO-Operations Research, 55, S2349-S2357.
[16] Sudhesh, R., Mohammed shapique, A., & Dharmaraja, S. (2022). Analysis of a multiple dual-stage vacation queueing system with disaster and repairable server. Methodology and Computing in Applied Probability, 24(4), 2485-2508.

[17] El-Taha, M. (2023). A Review of Birth?Death and Other Markovian Discrete?Time Queues. Advances in Operations Research, 2023(1), 6620393.
[18] Vijayashree, K. V., & Janani, B. (2018). Transient analysis of an M/M/1 queueing system subject to differentiated vacations. Quality Technology & Quantitative Management, 15(6), 730-748.
[19] Yang, S.-R., & Lin, Y.-B. (2005). Modeling UMTS discontinuous reception mechanism. IEEE Transactions on Wireless Communications, 4(1), 312“319.
[20] Yang, S.-R., Yan, S., & Hung, H. (2007).Modeling UMTS power saving with bursty packet data traffic. IEEE Transactions on Mobile Computing, 6(12), 1398“1409.
[21] 3GPP TS 36.321, Evolved Universal Terrestrial Radio Access (EUTRA); Medium Access Control (MAC) protocol specification (version 9.3.0 Release 9), 2010.

SIMULATIONS AND BAYESIAN ESTIMATION OF TRUNCATED EXPONENTIAL LOG-TOPP-LEONE GENERALIZED FAMILY WITH APPLICATION TO SURVIVAL TIME DATA

Usman Abubakar[1], Abdulhameed A. Osi[2], Ahmed Shuaibu[3], liyasu A. salisu[4]

•

[1,2,3,4]Department of Statistics, Faculty of Computing and Mathematical Sciences
Aliko Dangote University of Science and Technology Wudil, Kano, Nigeria
[1]usmanabubakar@jigpoly.edu.ng; [2]abuammarosi@gmail.com
[3]talk2ahmedomare@gmail.com; [4]iliyasuabubakarsalis@gmail.com

Abstract

Due to the requirements for the flexible statistical model to fit the lifetime data, we extended the truncated exponential topp-leone family due to its bounded interval, and introduced a truncated exponential log topp-leone generalized family of distributions. we examine some properties including survival function, hazard rate function, residual lifetime, reverse residual lifetime, moment, moment generating function, Shannon entropy, quantile, and parameter estimation using maximum likelihood, maximum product spacing, and bayesian estimation. Two simulation studies were conducted to investigate the properties (i.e. mean, variance, skewness, and kurtosis), and behavior of the maximum likelihood estimate using mean, bias, and RMSE. Finally, we apply the data on the survival times of breast cancer patients and suggest that the family of the proposed distribution outperforms other standard distributions based on information criteria and goodness of fit.

Keywords: Truncated Exponential Log Topp-Leone-G; Statistical Properties; Estimation; Simulation; Application.

I. Introduction

Lifetime data modeling has become a significant area in many fields, including actuarial science, economics, life science, engineering, business, and industries, among other [11] scientific fields, as data sets get more diverse and complicated [21]. This attracts much literature, which requires an appropriate model for accurate data realization [21]. To overcome such an issue, [22] developed a new distribution of empirical data with J-shaped histograms. It is a bounded support continuous distribution that can be used to simulate a distribution's lifetime. There has been little discussion prior to its discovery by [15], who examined a few of its properties, including moments and central moments. The flexibility of its hazard rate function makes Topp-leone a suitable distribution for modeling lifetime data [14].

In order to extend the existing truncated exponential topp-leone distribution by [6] with a bounded interval, there is need to modify the existing work of [12] on generalized topp-leone distribution using the concept of log transformation on log topp-leone [23] to become the log topp-leone generalized distribution, and incorporate it with truncated exponential as in [6] to provide a distribution with an unbounded interval.

Many researchers have introduced a generalization of distribution in relation to topp-leone including Topp-Leone Generalized Family of Distribution by [12], A New Topp-Leone Generalized Family of Distribution by [10], Topp-Leone Exponentiated-Generalized by [20], Sin Topp-Leone Generalized Distribution by [**?**], Exponentiated Topp-Leone Exponentiated-Generalized Distibution by [18], Frechet Topp-Leone Generalized Distribution by [18], Transmuted Topp-Leone Generalized by [24], New Power Topp-Leone Generated Distribution by [9], Poisson Topp-Leone Generator of Distribution by [13], truncated exponential topp-leone exponential by [6], truncated exponential topp-leone rayleigh by [5], among others.

In this paper, we conducted a Monte Carlo simulation to examine the behavior and consistency of the maximum likelihood estimate on the family of the Truncated exponential log-topp-leone-G introduced in [1] and also in [2]. The bayesian estimation would also be introduced in this paper, where the prior, conditional, and posterior distributions are to be discussed, while some properties were derived in including residual, reverse residual, Shannon entropy, and other methods of estimation parameters (see [1]), while other properties were already derived in [2].

II. Methods

1. Truncated Exponential Log-topp-leone Generalized Family of Distributions

The proposed Truncated exponential Log Topp-leone Generalized Family of distributions is drive from the cdf of truncated exponential distribution in [3] and the cdf of Log Topp-Leone-G distribution (which is an extension to the work of [23]) by integrating the pdf of truncated exponential distribution in equation with limit from 0 to the cdf of log topp-leone generalized family, and is drive as follows.

$$F_{TELTL-G}(y,\beta) = \int_0^{(1-e^{-2H(y,\psi)})^\theta} \frac{\beta e^{-\beta y}}{1-e^{-\beta}} dy = \frac{1-e^{-\beta(1-e^{-2H(y,\psi)})^\theta}}{1-e^{-\beta}} \tag{1}$$

Therefore

$$F(y,\beta,\theta) = \frac{1-e^{-\beta(1-e^{-2H(y,\psi)})^\theta}}{1-e^{-\beta}} \tag{2}$$

And we can find the probability density function by differentiating the Cdf using quotient rule. Let The cumulative distribution and probability density function of the truncated exponential Log Topp-leone-G family are given by

$$f_{TELTL-G}(y,\beta,\theta,\psi) = \frac{2\beta\theta h(y,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-1}e^{-\beta(1-e^{-2H(y,\psi)})^\theta}}{1-e^{-\beta}} \qquad y,\ \theta,\beta > 0 \tag{3}$$

2. Mathematical properties of TELTL-G

In this part, we discussed some mathematical properties of TELTL-G as follows:

2.1 Survival function and Hazard rate function of TELTL-G

The Survival function S(Y) of a TELTL-G as one of the important tools for measuring the failure time of a system is given by

$$S(y) = \frac{e^{-\beta(1-e^{-2H(y,\psi)})^\theta} - e^{-\beta}}{1-e^{-\beta}} \tag{4}$$

The hazard rate function H(Y) of a TELTL-G is given by

$$H(y) = \frac{2\beta\theta g(t,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-1}e^{-\beta(1-e^{-2H(y,\psi)})^\theta}}{e^{-\beta(1-e^{-2H(y,\psi)})^\theta} - e^{-\beta}} \tag{5}$$

2.2 Residual and Reverse residual of TELTL-G

one of the tools with application in acturial science, biometry, risk management, survival analysis is residual and reverse residual life. The residual and reverse residual life denoted by $r_t(y)$ and $\bar{r}_t(y)$ in the following equation.

$$r_t(y) = \frac{S(y+t)}{S(t)} = \frac{e^{-\beta(1-e^{-2H(y+t,\psi)})^{\theta}} - e^{-\beta}}{e^{-\beta(1-e^{-2H(t,\psi)})^{\theta}} - e^{-\beta}} \tag{6}$$

$$\bar{r}_t(y) = \frac{S(y-t)}{S(t)} = \frac{e^{-\beta(1-e^{-2H(y-t,\psi)})^{\theta}} - e^{-\beta}}{e^{-\beta(1-e^{-2H(t,\psi)})^{\theta}} - e^{-\beta}} \tag{7}$$

2.3 Moment and moment generating function

Moments is a crucial part of any statistical study [10]. They may be used to characterize key distributional features and forms, such as dispersion and spread as determined by mean and variance and peakiness of the distribution as determined by kurtosis. They can also be used to look at the symmetry of the distribution's shape as determined by skewness. Using the pdf from (3), the rth moment of a TELTL-Gdistribution is given by,

$$E(y^r) = \mu^r \int_{-\infty}^{\infty} y^r f(y,\beta,\theta,\psi)dy \tag{8}$$

$$\Rightarrow \mu^r = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{e=0}^{\infty}\int_0^{\infty} \kappa y^r h(y,\psi)(H(y,\psi))^{j+p}dy \tag{9}$$

$$\Rightarrow \mu^r = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{e=0}^{\infty}\kappa\Delta \tag{10}$$

Where

$$\kappa = \binom{\theta-1}{i}\binom{k\theta}{m}\frac{(-1)^{i+j+k+p}2^{i+P+1}(i+1)^j\beta^{k+1}\theta m^p t^e}{j!k!p!e!(1-e^{-\beta})} \tag{11}$$

And

$$\Delta = \int_0^{\infty} y^r h(y,\psi)(H(y,\psi))^{j+p}dy \tag{12}$$

The MGF of the random variable that follows TELTL-G having pdf in equation (3) is drive as.

$$E(e^{ty}) = M_y(t) = \int_{-\infty}^{\infty} e^{ty} f(y,\beta,\theta,\psi)dy \tag{13}$$

$$M_y(t) = \int_0^{\infty} \tau y^{e+1} h(y,\psi)(H(y,\psi))^{j+p}dy \tag{14}$$

$$\Rightarrow M_y(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{e=0}^{\infty}\tau\varrho \tag{15}$$

where

$$\tau = \binom{\theta-1}{i}\binom{k\theta}{m}\frac{(-1)^{i+j+k+p}2^{i+P+1}(i+1)^j\beta^{k+1}\theta m^p t^e}{j!k!p!e!(1-e^{-\beta})} \tag{16}$$

And

$$\varrho = \int_0^{\infty} y^{e+1} h(y,\psi)(H(y,\psi))^{j+p}dy \tag{17}$$

2.4 Quantile function and Shannon Entropy

The Quantile of TELTL-G is derived as,

$$\Rightarrow y_u = -\frac{1}{2}\ln\left\{1-\left[\frac{-log(1-u(1-e^{-\beta}))}{\beta}\right]^{\frac{1}{\beta}}\right\} \tag{18}$$

Shannon entropy is a concept from information theory, introduced by [19]. It measures the amount of uncertainty or randomness in a probability distribution. The Shannon entropy is calculated as

$$\gamma_y = E(-log(f(y))) \tag{19}$$

Implies that

$$log(f(y)) = log\left[\frac{2\beta\theta h(y,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-1}e^{-\beta(1-e^{-2H(y,\psi)})^{\theta}}}{1-e^{-\beta}}\right] \tag{20}$$

$$= log\frac{2\beta\theta}{1-e^{-\beta}} + log\left(\frac{h(y,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-1}e^{-\beta(1-e^{-2H(y,\psi)})^{\theta}}}{1-e^{-\beta}}\right) \tag{21}$$

$$= log\frac{2\beta\theta}{1-e^{-\beta}} + loge^{-2H(y,\psi)} + log\left(h(y,\psi)(1-e^{-2H(y,\psi)})^{\theta-1}\right) - \beta(1-e^{-2H(y,\psi)})^{\theta} \tag{22}$$

$$= log\frac{2\beta\theta}{1-e^{-\beta}} - 2H(y,\psi) + log(h(y,\psi)) + log(1-e^{-2H(y,\psi)})^{\theta-1} - \beta(1-e^{-2H(y,\psi)})^{\theta} \tag{23}$$

$$E(-log(f(y))) = E\left[-log\tfrac{2\beta\theta}{1-e^{-\beta}}\right] + 2E\left[H(y,\psi)\right] - E\left[log(h(y,\psi))\right] -$$
$$E\left[log(1-e^{-2H(y,\psi)})^{\theta-1}\right] + \beta E\left[(1-e^{-2H(y,\psi)})^{\theta}\right] \tag{24}$$

2.5 Order statistics

Let $y_1, y_2..., y_n$ be a random sample from the TELTL-G distribution and let y(1), . . . , y(n) be the corresponding order statistics. The pdf of nth order statistic can be written as

$$f_{(i,n)}(y) = \frac{n!(-1)^j}{(i-1)(n-i-j)!j!}\sum_{j=0}^{n-i}\binom{n-i}{j}f(y)[F(y)]^{i+j-1} \tag{25}$$

$$f_{(i,n)}(y) = \sum_{j=0}^{n-i}\binom{n-i}{j}\frac{n!(-1)^j\eta(1-e^{-\beta(1-e^{-2H(y,\psi)})^{\theta}})^{j+i-1}}{(i-1)(n-i-j)!j!(1-e^{-\beta})^{j+i}} \tag{26}$$

where $\eta = 2\beta\theta h(y,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-1}e^{-\beta(1-e^{-2H(y,\psi)})^{\theta}}$

3. Estimation of Parameters

3.1 Maximum likelihood estimation (MLE)

Let $y_1, y_2, ..., y_n$ be a random sample from the TELTL-G family of distribution with pdf in equation (3) with $\bar{\omega} = (\beta, \theta, \psi)$, the TELTL-G's n sample log-likelihood is drive as:

$$l = logl(y/\bar{\omega}) = log\prod_{i=1}^{n} f(y/\bar{\omega}) \tag{27}$$

$$l(\bar{\omega}) = \prod_{i=1}^{n} \frac{2\beta\theta h(y,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-1}e^{-\beta(1-e^{-2H(y,\psi)})^{\theta}}}{1-e^{-\beta}} \tag{28}$$

$$l(\bar{\omega}) = nlog(2) + nlog(\beta) + nlog(\theta) + \sum_{i=1}^{n} logh(y,\psi) - 2\sum_{i-1}^{n} H(y,\psi) +$$

$$(n\theta - n)log(1-e^{-2H(y,\psi)}) - \beta\sum_{i=1}^{n}(1-e^{-2H(y,\psi)})^{\theta} + nlog(1) - nlog(1-e^{-\beta}) \tag{29}$$

By differentiating the log likelihood with respect to β, θ, and ψ, we have:

$$\frac{dl}{d\beta} = \frac{n}{\beta} - \frac{ne^{-\beta}}{(1-e^{-\beta})} - \sum_{i=1}^{n}(1-e^{-2H(y,\psi)})^{\theta} \tag{30}$$

$$\frac{dl}{d\theta} = \frac{n}{\theta} + nlog(1-e^{-2H(y,\psi)}) - \theta\beta\sum_{i=1}^{n}\left\{1-e^{-2H(y,\psi)}\right\}^{\theta-1} \tag{31}$$

$$\frac{dl}{d\psi} = \sum_{i=1}^{n}\frac{dh(y,\psi)/d\psi}{h(y,\psi)} + \frac{2(n\theta-n)h(y,\psi)e^{-2H(y,\psi)}}{(1-e^{-2H(y,\psi)})} - 2\sum_{i=1}^{n} h(y,\psi) -$$

$$2\theta\beta\sum_{i=1}^{n}\left\{1-e^{-2H(y,\psi)}\right\}^{\theta-1} h(y,\psi)e^{-2H(y,\psi)} \tag{32}$$

Where $\frac{dH(y,\psi)}{d\psi} = h(y,\psi)$

The 3x3 observed information matrix $J(\omega)$ will be obtain for the interval estimation of β,θ, and ψ and test of hypothesis for the parameters.

$$J(\omega) = \begin{pmatrix} J_{\beta\beta}(\omega) & J_{\beta\theta}(\omega) & J_{\beta\psi}(\omega) \\ J_{\theta\beta}(\omega) & J_{\theta\theta}(\omega) & J_{\beta\psi}(\omega) \\ J_{\psi\beta}(\omega) & J_{\psi\theta}(\omega) & J_{\psi\psi}(\omega) \end{pmatrix} \tag{33}$$

Where the element are

$$J_{\beta\beta} = \frac{-n}{\beta^2} - \frac{e^{-\beta}}{(1-e^{-\beta})^2} \tag{34}$$

$$J_{\theta\theta} = \frac{-n}{\theta^2} - \beta\sum(1-e^{-2H(y,\psi)})^{\theta-1} - \beta\theta(\theta-1)\sum(1-e^{-2H(y,\psi)})^{\theta-1} \tag{35}$$

$$J_{\psi\psi} = \frac{h(y,\psi)h''(y,\psi)-(h'(y,\psi))^2}{(h(y,\psi))^2} + 2(n\theta-n)\frac{d}{d\psi}\left(\frac{h(y,\psi)e^{-2H(y,\psi)}}{1-e^{-2H(y,\psi)}} - 2\sum h'(y,\psi)\right) -$$

$$2\theta\beta\frac{d}{d\psi}\left(\sum(1-e^{-2H(y,\psi)})^{\theta-1}h(y,\psi)e^{-2H(y,\psi)}\right) \tag{36}$$

$$J_{\beta\theta} = -\sum(1-e^{-2H(y,\psi)})^{\theta-1}log(1-e^{-2H(y,\psi)}) \tag{37}$$

$$J_{\beta\psi} = -2\theta h(y,\psi)e^{-2H(y,\psi)}\sum(1-e^{-2H(y,\psi)})^{\theta-1} \tag{38}$$

$$J_{\theta\psi} = \frac{-2nh(y,\psi)e^{-2H(y,\psi)}}{1-e^{-2H(y,\psi)}} - 2\beta\theta(\theta-1)\sum h(y,\psi)e^{-2H(y,\psi)}(1-e^{-2H(y,\psi)})^{\theta-2} \tag{39}$$

3.2 Maximum Product Spacing

The method of maximum product spacing (MPS) is a technique used in statistics and data analysis to estimate the parameters of a distribution. It is a alternative method to maximum likelihood estimation (MLE) and method of moments. In MPS, the goal is to find the parameter values that maximize the product of the spacings between the order statistics of the data. Order statistics are the values in the data set arranged in increasing order. The MPS estimate can be obtained

by maximizing the geometric mean of the spacings between the order statistics, rather than the product of the spacings.

$$GM = \left(\prod_{j=1}^{n+1} K_j\right)^{\frac{1}{n+1}} \qquad j = 1, 2, 3...., n+1 \tag{40}$$

Where the jth difference K_j is define as

$$K_j = \int_{y(j-1)}^{y(j)} f(y, \beta, \theta, \psi) dy \tag{41}$$

Where $f(y(0), \beta, \theta, \psi) = 0$ and $f(y(n+1), \beta, \theta, \psi) = 1$. Therefore, the mps of $\beta, \theta, and \psi$ are the result of maximizing the GM of difference.

$$logGM = \frac{1}{n+1} \sum_{j=1}^{n+1} log\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right] \tag{42}$$

$$\frac{dlogGM}{d\beta} = \frac{1}{n+1} \sum_{j=1}^{n+1} \frac{d\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right]/d\beta}{\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right]} \tag{43}$$

$$\frac{dlogGM}{d\theta} = \frac{1}{n+1} \sum_{j=1}^{n+1} \frac{d\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right]/d\theta}{\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right]} \tag{44}$$

$$\frac{dlogGM}{d\psi} = \frac{1}{n+1} \sum_{j=1}^{n+1} \frac{d\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right]/d\psi}{\left[H(y(j); \beta, \theta, \psi) - H(y(j-1); \beta, \theta, \psi)\right]} \tag{45}$$

3.3 Bayesian Estimation of TELTL-G

Supposed that $\beta \sim \Gamma(a, b)$, $\theta \sim \Gamma(c, d)$, and $\psi \sim \Gamma(m, l)$ respectively, where a,b,c,d,l,m are positive constant. The Gamma prior of β, θ, ψ take the forms

$$\tau_1(\beta) = \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta} \qquad \beta, a, b > 0 \tag{46}$$

$$\tau_2(\theta) = \frac{d^c}{\Gamma(c)} \theta^{c-1} e^{-d\theta} \qquad \theta, c, d > 0 \tag{47}$$

$$\tau_3(\psi) = \frac{l^m}{\Gamma(m)} \psi^{m-1} e^{-l\psi} \qquad \psi, m, l > 0 \tag{48}$$

The joint density function were given based on conditional distribution of $\beta, \theta, and \psi$ given by

$$f(\beta/\theta, \psi, y) \propto \beta^{a+n-1} \prod_{i=1}^{n} (2h(y, \psi)(1 - e^{-\beta})^{-1}(1 - e^{-2H(y,\psi)})^{\theta}) e^{-b\beta - 2\sum H(y,\psi) - \beta\sum(1-e^{-2H(y,\psi)})^{\theta}} \tag{49}$$

$$f(\theta/\beta, \psi, y) \propto \theta^{c+n-1} \prod_{i=1}^{n} (2h(y, \psi)(1 - e^{-\beta})^{-1}((1 - e^{-2H(y,\psi)})^{\theta}) e^{-d\theta - 2\sum H(y,\psi) - \beta\sum(1-e^{-2H(y,\psi)})^{\theta}} \tag{50}$$

$$f(\psi/\beta, \theta, y) \propto \psi^{m-1} \prod_{i=1}^{n} (2h(y, \psi)(1 - e^{-\beta})^{-1}(1 - e^{-2H(y,\psi)})^{\theta}) e^{-l\psi - 2\sum H(y,\psi) - \beta\sum(1-e^{-2H(y,\psi)})^{\theta}} \tag{51}$$

And the posterior distribution is given by

$$\tau * (\beta, \theta, \psi|y) = \propto \tau(\beta, \theta, \psi) \prod_{i=1}^{n} f(y; \beta, \theta, \psi) \tag{52}$$

$$f(\omega/y) \propto \beta^{a+n-1} \theta^{c+n-1} \psi^{m-1} \prod_{i=1}^{n} \left(2h(y, \psi)(1 - e^{-\beta})^{-1}(1 - e^{-2H(y,\psi)})^{\theta}\right) \times$$
$$e^{-b\beta - d\theta - l\psi - 2\sum H(y,\psi) - \beta\sum(1-e^{-2H(y,\psi)})^{\theta}} \tag{53}$$

4. Sub-model of TELTL-G

The TELTL-G family's unique sub-models, the Truncated exponential log topp-leone Pareto Distribution (TELTL-PD) and the Truncated exponential log topp-leone lomax Distribution (TELTL-LD), are addressed in this section.

4.1 Truncated exponential log topp-leone pareto distribution

Let $H(y;\psi)$ be cdf of the pareto random variable given by $H(y;\psi) = 1 - (\frac{\omega}{y})^{\alpha}, \quad y, \alpha, \omega > 0$, and $h(y;\psi) = \frac{\alpha\omega^{\alpha}}{y^{\alpha+1}}$. Then, the cdf, pdf, and quantile of the TELTL-P distribution is given as,

$$F(y,\beta,\theta,\alpha,\omega) = \frac{1 - e^{-\beta(1-e^{-2(1-(\frac{\omega}{y})^{\alpha})})^{\theta}}}{1 - e^{-\beta}} \tag{54}$$

$$f(y,\beta,\theta,\alpha,\omega) = \frac{2\alpha\beta\theta\omega^{\alpha}y^{-\alpha-1}(1 - e^{-2(1-(\frac{\omega}{y})^{\alpha})})^{\theta-1}e^{-2(1-(\frac{\omega}{y})^{\alpha})}e^{-\beta(1-e^{-2(1-(\frac{\omega}{y})^{\alpha})})^{\theta}}}{1 - e^{-\beta}} \quad y,\theta,\beta,\alpha,\omega > 0 \tag{55}$$

$$y_u = \left(\frac{1}{\omega^{\alpha}}\left(1 + \frac{1}{2}log\left(1 - \left(\frac{-log(1 - u(1 - e^{-\beta}))}{\beta}\right)\right)\right)^{\frac{1}{\theta}}\right)^{\frac{-1}{\alpha}} \tag{56}$$

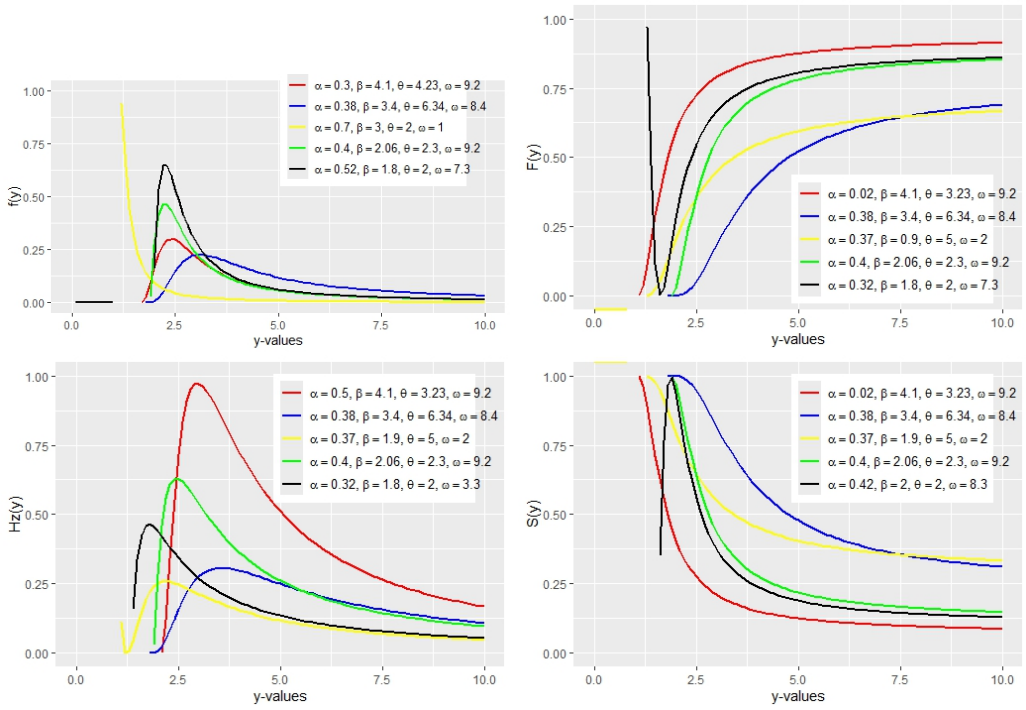


Figure 1: *CDF, PDF, survival and Hazard plot of TELTL-Pareto for values of parameters*

4.2 Truncated exponential log top-leone lomax Distribution

Let $H(y;\delta,\sigma)$ be cdf of the lomax random variable given by $H(y;\delta,\sigma) = 1 - (\frac{\delta}{\delta+y})^{\sigma}, \quad y, \delta, \sigma > 0$, and $h(y;\delta,\sigma) = \frac{\sigma\delta^{\sigma}}{(\delta+y)^{\sigma}}$. Then, the cdf, pdf, and quantile of the TELTL-L distribution is given as,

$$F(y,\beta,\theta,\delta,\sigma) = \frac{1 - e^{-\beta(1-e^{-2(1-(\frac{\delta}{\delta+y})^{\sigma}))})^{\theta}}}{1 - e^{-\beta}} \tag{57}$$

$$f(y,\beta,\theta,\delta,\sigma) = \frac{2\sigma\beta\theta\delta^{\sigma}\delta(1-e^{-2(1-(\frac{\delta}{\delta+y})^{\sigma}))})^{\theta-1}e^{-2(1-(\frac{\delta}{\delta+y})^{\sigma}))}e^{-\beta(1-e^{-2(1-(\frac{\delta}{\delta+y})^{\sigma}))})^{\theta}}}{1-e^{-\beta}} \quad y,\theta,\beta,\delta,\sigma > 0 \tag{58}$$

$$y_u = \delta - \left(\frac{1}{\delta^{\sigma}}\left(1+\frac{1}{2}log\left(1-\left(\frac{-log(1-u(1-e^{-\beta}))}{\beta}\right)\right)\right)^{\frac{1}{\theta}}\right)^{\frac{-1}{\sigma}} \tag{59}$$

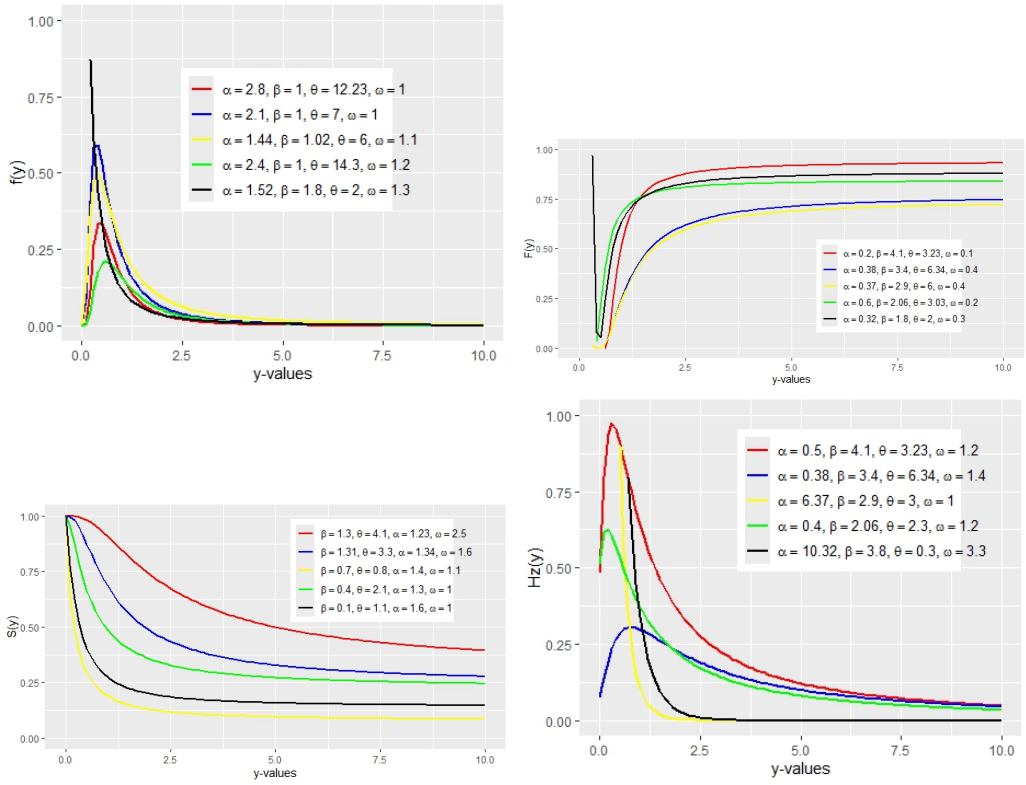


Figure 2: *CDF, PDF, survival and Hazard rate plot of TELTL-Lomax for some values of parameters*

III. Results

1. Simulation study

The simulations on TELTL-Pr and TELTL-Lmx were conducted to examine the behavior of the distribution. The consistency of the maximum likelihood of TELTL-Pr and TELTL-Lmx for set values (i.e. $\alpha = 1.9, \beta = 2.2, \theta = 3.3, \omega = 4.1$), while for TELTL-Lmx with four parameters ($\beta = 3.6, \theta = 1.2, \delta = 2.5, \sigma == 4.1$) were investigated using a finite sample of n=20, 50, 150, 250, 500, and 1000 were created. The random numbers for the TELTL-Pr were generated using the quantile function. For 1000 repeated samples. The Means, Bias, and RMSE were then calculated. Table 2 and 3 presents the outcomes of the simulation. It is concluded that the family member TELTL-G yields consistent results when predicting parameters for the mode based on the results of the monte-carlo simulation. Likewise, some properties including the mean, variance, skewness, and kurtosis were obtained from the simulation for a specific values of parameters as in Table 1. The bias and root mean square error (RMSE) are given by the following equation.

$$Bias(\hat{\theta}) = \frac{\sum_{i=1}^{N}\hat{\theta}_i}{N} - \theta \tag{60}$$

$$RMSE(\hat{\theta}) = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2}{N}} \tag{61}$$

Table 1: *Properties from Simulation result*

Distribution	$\alpha, \beta, \theta, \omega$	Mean	Variance	Skeweness	Kurtosis
TPr	(1, 2, 2, 3)	3.057956	2.479738e-03	1.0395213	3.329671
TLmx	(1, 2, 1, 3)	-5.57991e-02	2.270049e-03	-1.03586576	3.369155
TETL	(1, 2, 2)	-0.643513	0.07888371	0.7992454	2.510671
TGPr	(1, 2, 2)	0.5286839	0.04769423	-0.1457797	1.92661
NHTLmx	(1, 2, 2)	0.2232752	0.01735384	0.9802952	4.097517
RP	(1, 2, 2)	1.804450	4.221214e+01	-0.56238231	2.394527

Table 2: *Simulation result for first set of parameters of TELTL-Pr*

Sample	Properties	$\alpha = 1.9$	$\beta = 2.2$	$\theta = 3.3$	$\omega = 4.1$
n=20	Means	2.3348	4.6704	2.3998	1.3606
	Bias	0.4348	2.4704	-0.9002	-2.7394
	RMSE	0.6240	2.5323	0.9787	2.7587
n=50	Means	2.2570	4.5514	2.5102	1.2743
	Bias	0.3570	2.3514	-0.7898	-2.8257
	RMSE	0.4916	2.3813	0.8390	2.8346
n=150	Means	2.1654	4.5406	2.5857	1.2525
	Bias	0.2654	2.3406	-0.7143	-2.8475
	RMSE	0.3382	2.3540	0.7318	2.8510
n=250	Means	2.1622	4.5524	2.5897	1.2454
	Bias	0.2622	2.3524	-0.7103	-2.8546
	RMSE	0.3227	2.3672	0.7205	2.8580
n=500	Means	2.1495	4.6217	2.5829	1.2478
	Bias	0.2495	2.4217	-0.7171	-2.8522
	RMSE	0.3171	2.4407	0.7241	2.8558
n=1000	Means	0.7741	2.5453	3.0949	4.6305
	Bias	-0.0259	0.0453	0.0949	0.1305
	RMSE	0.0713	0.2629	0.2457	0.3592

2. Application to survival times of breast cancer patients data

The dataset was collected from 1929 to 1938, which represents the survival times of breast cancer patients. The data was used by [16]. The observations are given as 0.3, 0.3, 4.0, 5.0, 5.6, 6.2, 6.3, 6.6, 6.8, 7.4, 7.5, 8.4, 8.4, 10.3, 11.0, 11.8, 12.2, 12.3, 13.5, 14.4, 14.4, 14.8, 15.5, 15.7, 16.2, 16.3, 16.5, 16.8, 17.2, 17.3, 17.5, 17.9, 19.8, 20.4, 20.9, 21.0, 21.0, 21.1, 23.0, 23.4, 23.6, 24.0, 24.0, 27.9, 28.2, 29.1, 30.0, 31.0, 31.0, 32.0, 35.0, 35.0, 37.0, 37.0, 37.0, 38.0, 38.0, 38.0, 39.0, 39.0, 40.0, 40.0, 40.0, 41.0, 41.0, 41.0, 42.0, 43.0, 43.0, 43.0, 44.0, 45.0, 45.0, 46.0, 46.0, 47.0, 48.0, 49.0, 51.0, 51.0, 51.0, 52.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 60.0, 60.0, 61.0, 62.0, 65.0, 65.0, 67.0, 67.0, 68.0, 69.0, 78.0, 80.0, 83.0, 88.0,

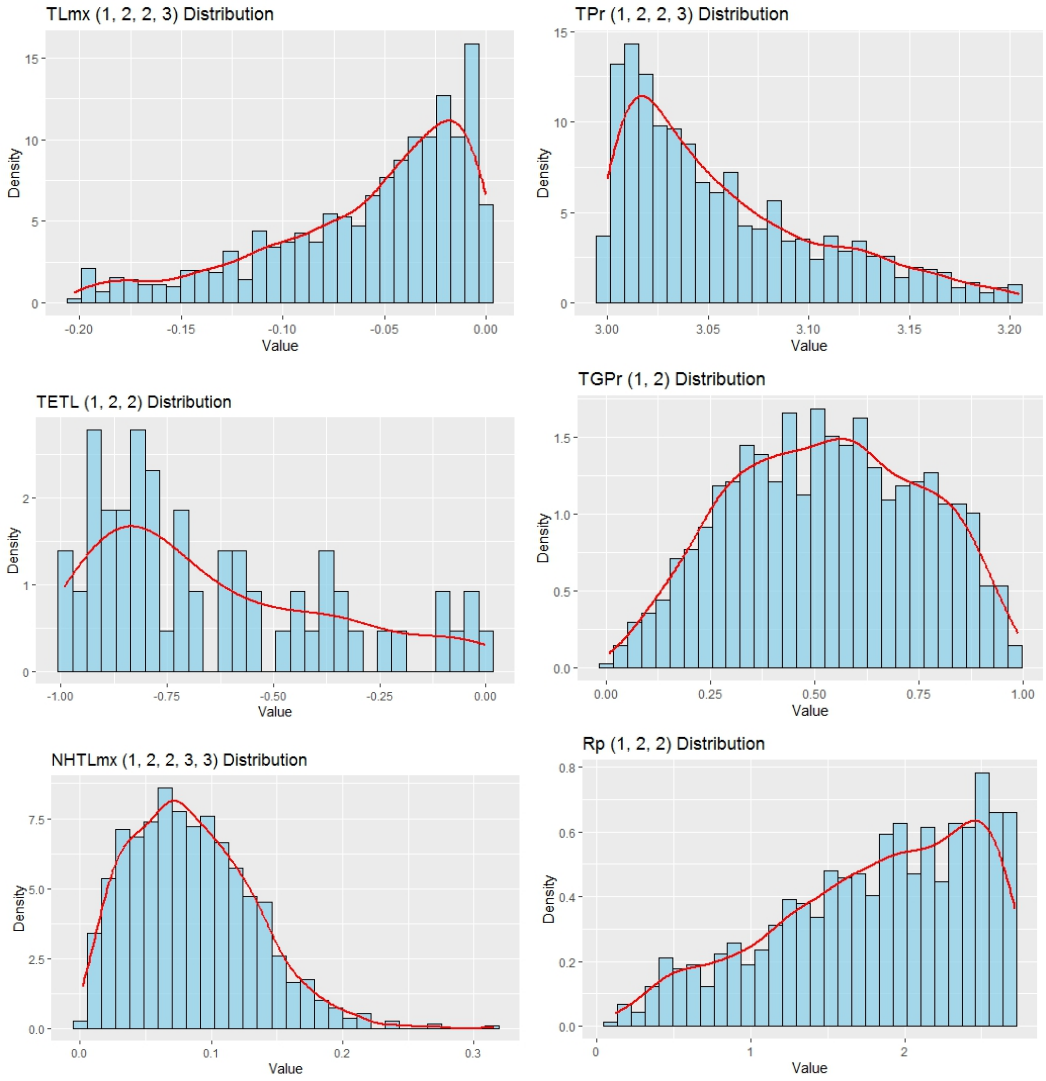


Figure 3: *Skewness and kurtosis for proposed and existing models*

Table 3: *Simulation result for second set of parameters of TELTL-Lmx*

Sample	Properties	$\beta = 3.6$	$\theta = 1.2$	$\delta = 2.5$	$\sigma == 4.1$
n=20	Means	5.8301	4.1297	2.2794	1.9739
	Bias	2.2301	2.9297	-0.2206	-2.1261
	RMSE	2.2301	2.9297	-0.2206	-2.1261
n=50	Means	2.2301	2.9297	-0.2206	-2.1261
	Bias	1.9216	2.6165	-0.2111	-2.2393
	RMSE	2.2505	2.9445	0.2468	2.2915
n=150	Means	5.1789	3.4406	2.2681	1.8987
	Bias	1.5789	2.2406	-0.2319	-2.2013
	RMSE	1.7777	2.4579	0.2542	2.2292
n=250	Means	5.0407	3.2331	2.2515	1.9199
	Bias	1.4407	2.0331	-0.2485	-2.1801
	RMSE	1.6257	2.2035	0.2679	2.2058
n=500	Means	4.9989	3.0595	2.2294	1.8953
	Bias	1.3989	1.8595	-0.2706	-2.2047
	RMSE	1.5425	1.9954	0.2842	2.2245
n=1000	Means	5.0224	2.9238	2.2215	1.8293
	Bias	1.4224	1.7238	-0.2785	-2.2707
	RMSE	1.5313	1.8206	0.2889	2.2893

89.0, 90.0, 93.0, 96.0, 103.0, 105.0, 109.0, 109.0, 111.0, 115.0, 117.0, 125.0,126.0, 127.0, 129.0, 129.0, 139.0, 154.0.

Calculations were made to compare the fitted models using the goodness-of-fit metrics, which include the kolmogrov smirnov, crimer von-mises, and Andason darling. The information log-likelihood and the information criteria were also examine using Akaike information criteria (AIC), Bayesian information criteria (BIC), and Akaike information corrected criteria (AICc). The model with smaller value of information criteria is the model with best fit [8].

Table 4: *Competitors Distributions*

Distributions	Author(s)	Year	Citation
TETLE	Al-noor and Hilal	2021	[**?**]
TGPr	Al-quraishy et al	2022	[**?**]
NHTLmx	Reyad et al	2019	[**?**]
Rp	Al-kadim and Muhammad	2018	[**?**]

Table 5: *MLE of the parameter(s) using Survival time data*

Distributions	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\gamma}$
TPr	0.92342	1.1666	7.3870	0.3609	-
TLmx	0.06907	1.643	41.81	19.84	-
TETL	-7.8510	0.01562	0.251287	-	-
TGPr	1.7716	0.5148	-	-	-
NHTLmx	20.002471	16.485334	0.123569	0.02324	0.3241
Rp	672.80819	-80.6895	0.0044	-	-

Table 6: *Information Criteria for the fitted models*

Distributions	$\hat{LL}$	AIC	AICc	BIC
TPr	-184.1018	374.2036	374.4087	392.5910
TLmx	-194.5029	397.0059	397.3507	408.1891
TETLE	-578.8903	1163.781	1163.986	1172.168
TGPr	-152.1718	411.2006	412.8517	418.1152
NHTLmx	-491.124	924.418	923.181	931.201
RP	-989.0158	1972.032	1971.826	1979.644

Table 7: *Goodness of fit test for the fitted models*

Distributions	KS	A	W	P-value
TPr	1.0	0.3242	0.34343	<0.00
TLmx	0.96138	5.6062	0.97944	<0.00
TETL	0.06105	0.38433	0.05142	0.7579
TGPr	0.10742	5.6358	0.10444	0.2132
NHTLmx	0.28944	1.5209	0.26993	<0.00
RP	0.11552	0.41323	0.058538	0.0914

IV. Discussion

The bayesian estimate derived in Section 2.7 gives the prior, conditional, and posterior distributions for the parameters of TELTL-G, which is an alternative estimation method of parameters. The first simulation conducted was to observe the behavior of the distribution, and it shows that the accuracy of the estimate is better as the sample size increases for the TELTL-Pr and for the TELTL-LMx; the accuracy is decreasing as the sample size is increasing. Likewise, Table 2 shows the properties for the family of TELTL-G, which includes the mean, variance, skewness, and kurtosis. As illustrated in the table, the TELTL-P is more skewed positively than competitors distributions with a value of 1.0395213, but the kurtosis of NHTLmx (4.097517) is greater than that of the proposed distribution. The TELTL-Lmx is negatively skewed with a value of -1.0358657 and a kurtosis of 3.369155, and these show that the proposed distributions have more skewness and kurtosis than the existing TETL-E distribution with a skewness of 0.799245 and a kurtosis of 2.520671. The plots illustrated in Figure 3 show the shape of the proposed and competitors distributions, while Figures 1 and 2 are the pdf, cdf, survival, and hazard rate functions for the different values of parameters of the TELTL-Pr and TELTL-Lmx distributions, respectively.

Conclusion

In this paper, we introduce the simulation and bayesian estimation of a truncated exponential log topp-leone generalized family of distributions with additional properties and estimation methods, since the properties and other characteristics were examined in our previous study, see [1]. The family of the proposed distribution demonstrates outstanding performance, and shows how it is good and flexible in term of fit compared to other standard distributions. Finally, we suggested that the TELTL-G is an alternative distributions in modeling heavy tail or skewed data set.

Acknowledgment

A profound gratitude to the head of Statistics department, Aliko Dangote university of science and technology Wudil, Abdulhameed Ado Osi. He has been the inspiration and backbone for carrying out this research.

Ethical Consideration

No any Ethical clearance required before the commencement of these research.

Data availability statement

All data used in this work can be found within the manuscript.

References

[1] Abubakar, U., Osi, A. A., and Shuaibu, A. (2024a). Truncated exponential log-topp-leone rayleigh distributions: Properties with application to bladder cancer data. *UMYU Scientifica*, 3(3):173“180.

[2] Abubakar, U., Osi, A. A., Shuaibu, A., Abubakar, A., Salisu, I. A., and Muhammad, Y. I. (2024b). Truncated exponential log-topp-leone generalized family of distributions: properties and application to real data sets. *International Journal of Research and technopreneurial innovation*, 1(1):147“162.

[3] Akahira, M. (2017). Statistical estimation for truncated exponential families. *Springer.*

[4] Al-Babtain, A. A., Elbatal, I., Chesneau, C., and Elgarhy, M. (2020). Sine topp-leone-g family of distributions: Theory and applications. *Open Physics*, 18(1):574“593.

[5] Al-Kadim, K. A. and Mohammed, B. D. (2018). Rayleigh pareto distribution. *Journal of University of Babylon for Pure and Applied Sciences*, 26(1):84“95.

[6] Al-Noor, N. H. and Hilal, O. A. (2021). Truncated exponential topp leone exponential distribution: properties and applications. In Journal of Physics: *Conference Series*, volume 1879, page 032039. IOP Publishing.

[7] Al Quraishy, K. A. and Manalmousa, S. E. (2022). Reliability estimation of the topp-leone g-pareto distribution: Properties and application. *Journal of Algebraic Statistics*, 13(3):4425“4439.

[8] Aryal, G. R., Ortega, E. M., Hamedani, G., and Yousof, H. M. (2017). The topp-leone generated weibull distribution: regression model, characterizations and applications. *International Journal of Statistics and Probability*, 6(1):126“141.

[9] Bantan, R. A., Jamal, F., Chesneau, C., and Elgarhy, M. (2019). A new power topp“leone generated family of distributions with applications. *Entropy*, 21(12):1177.

[10] Hassan, A. S., Elgarhy, M., and Ragab, R. (2020). Statistical properties and estimation of inverted topp-leone distribution. *J. Stat. Appl. Probab*, 9(2):319“331.

[11] Lawless, J. F. (2011). *Statistical models and methods for lifetime data. John Wiley & Sons*.

[12] Mahdavi, A. (2017). Generalized topp-leone family of distributions. *Journal of Biostatistics and Epidemiology*, 3(2):65“75.

[13] Merovci, F., Yousof, H. M., and Hamedani, G. (2020). The poisson topp leone generator of distributions for lifetime data: theory, characterizations and applications. *Pakistan Journal of Statistics and Operation Research.*

[14] Muhammad, M., Liu, L., Abba, B., Muhammad, I., Bouchane, M., Zhang, H., and Musa, S. (2023). A new extension of the topp“leone-family of models with applications to real data. *Annals of Data Science*, 10(1):225“250.

[15] Nadarajah, S. and Kotz, S. (2003). Moments of some j-shaped distributions. *Journal of Applied Statistics*, 30(3):311“317.

[16] Nanga, S., Sayibu, S. B., Angbing, I. D., Alhassan, M., Benson, A.-M., Abubakari, A. G., and Nasiru, S. (2024). Secant kumaraswamy family of distributions: Properties, regression model, and applications. *Computational and Mathematical Methods*, 2024(1):8925329.

[17] Reyad, H., Selim, M. A., and Othman, S. (2019a). The nadarajah haghighi topp leone-g family of distributions with mathematical properties and applications. *Pakistan Journal of Statistics and Operation Research,* pages 849“866.

[18] Reyad, H. M., Alizadeh, M., Jamal, F., Othman, S., and Hamedani, G. G. (2019b). The exponentiated generalized topp leone g family of distributions: properties and applications. *Pakistan Journal of Statistics and Operation Research*, pages 1“24.

[19] Shannon, C. (1948). Claude shannon. *Information Theory,* 3:224.

[20] Sule, I., Sani, I. D., Audu, I., and Jibril, H. M. (2020). On the topp leone exponentiated-g family of distributions: properties and applications.

[21] Suleiman, A. A., Daud, H., Ishaq, A. I., Othman, M., Sokkalingam, R., Usman, A., and Osi, A. A. (2023). The odd beta prime inverted kumaraswamy distribution with application to covid-19 mortality rate in italy. *Engineering Proceedings*, 56(1):218.

[22] Topp, C. W. and Leone, F. C. (1955). A family of j-shaped frequency functions. *Journal of the American Statistical Association,* 50(269):209“219.

[23] Usman, A., Ishaq, A. I., Suleiman, A. A., Othman, M., Daud, H., and Aliyu, Y. (2023). Univariate and bivariate log-topp-leone distribution using censored and uncensored datasets. *In Computer Sciences & Mathematics Forum*, volume 7, page 32. MDPI.

[24] Yousof, H. M., Alizadeh, M., Jahanshahi, S., Ramires, T. G., Ghosh, I., and Hamedani, G. (2017). The transmuted topp-leone g family of distributions: theory, characterizations and applications. *Journal of Data Science,* 15(4):723“740

SINE GENERALIZED ODD LOG-LOGISTIC FAMILY OF DISTRIBUTIONS: PROPERTIES AND APPLICATION TO REAL DATA

Abdulhameed A. Osi*, Usman Abubakar and Lawan A. Isma'il

•

Department of Statistics, Aliko Dangote University of Science and Technology, Wudil, Nigeria.
abuammarosi@gmail.com, aaosi@kust.edu,ng

Abstract

In this research, we introduce and analyze a new family of distributions called the sine generalized odd log-logistic-G family. This is driven by the reality that no single distribution can effectively model all types of data across different fields. Consequently, there is a need to develop distributions that possess desirable properties and are flexible enough to accommodate data with diverse characteristics. We explore its statistical properties, including the survival function, hazard function, moments, moment-generating function, and order statistics. A special case of the family of distributions is also presented. The maximum likelihood estimation method is used to obtain estimators of the family of distributions and the performance of the maximum likelihood estimators is evaluated in terms of bias and root mean squared errors through two simulation studies. Additionally, we demonstrate the practicality of this family using two real data sets, where it consistently provides better fits compared to other competitive distributions.

Keywords: Odd Log-Logistic family of distribution, Sine transformation, Maximum likelihood estimation, Monte Carlo Simulation, Breast tumor data

1. Introduction

The generalized odd family has gained attention in the literature because it can capture various distribution shapes, such as unimodality, bimodality, and heavy-tailed behavior [1]. Recent studies have extensively examined the generalized odd log-logistic family of distributions, providing enhanced flexibility for modeling lifetime data. This distribution extends the odd log-logistic family, which is commonly used in modeling lifetime and survival data [1]. A new class of continuous distributions with additional shape parameters was introduced by [2] and [7], leading to improved fits for various datasets. These families include special cases such as the proportional reversed hazard rate and odd log-logistic classes. The odd log-logistic Gompertz distribution was proposed by [3] and can model decreasing, increasing, and bathtub-shaped failure rates. Building on this concept, [4] introduced the odd log-logistic transmuted-G family, which combines the odd log-logistic distribution with the transmuted distribution.

A new family of continuous distributions, called the "exponentiated odd log-logistic family," was proposed by [5]. A new three-parameter lifetime distribution, called the odd log-logistic generalized Lindley (OLLGL) distribution, was introduced by [14]. All of these articles discuss the mathematical properties, parameter estimation methods, and real-world applications of their proposed distributions. These new families have demonstrated superior performance in modeling survival data and reliability problems compared to existing distributions, highlighting their importance in statistical analysis. The Odd Lomax log-logistic distribution (OLLLD) was introduced by [12] as a generalized parental distribution, enhancing flexibility to capture the

characteristics of real-world data sets, with parameters estimated using maximum likelihood estimation. The Odd Log-Logistic Generalized Exponential Distribution was developed by [13], enhancing distribution flexibility for survival data analysis, with a focus on statistical properties and parameter estimation methods. The odd log-logistic generalized Lindley distribution, developed by [14], is a three-parameter lifetime model that includes structural properties, estimation methods, a simulation study, and empirical illustrations using real data sets.

Recent advancements in trigonometric distributions have introduced several new families with broad applications in fields such as physics, engineering, and medicine. The Sine Type II Topp-Leone-G family, proposed by [16], offers flexibility and efficient data fitting. The sine-G family expanded by [18] to include the alpha-sine-G family, which adds a parameter for enhanced statistical inference. The exponentiated sine-G family was introduced by [19] for lifetime studies, emphasizing properties such as density, reliability, and entropy. The Sin-G family was extended by jamal2021beyond with a transformed version that enhances modeling. The sine Kumaraswamy-G family was developed by [21], exploring its statistical applications. Meanwhile, the sine-G family, an extension of the exponentiated Burr XII distribution for heavy-tailed datasets in insurance, was proposed by [15]. .

2. Methods

This paper presents a comprehensive study of the theoretical properties and applications of the sine generalized odd log-logistic distribution. By introducing the trigonometric function in the sine generalized odd log-logistic distribution, researchers and practitioners can now model a wide range of real-world phenomena with greater flexibility, making it a versatile tool for working with lifetime and survival data. The cdf of generalized log-logistic families of distributions by [9] is given as;

$$P(x) = \frac{G(x)^{\alpha}}{G(x)^{\alpha} + \bar{G}(x)^{\alpha}} \tag{1}$$

Where $\bar{G}(x) = 1 - G(x)$. And the cdf anf pdf of generalized odd log-logistic family by [11] is given by;

$$F(x) = \frac{G(x)^{\alpha\theta}}{G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha}} \tag{2}$$

$$f(x) = \frac{\alpha\theta g(x) G(x)^{\alpha\theta-1}(1 - G(x)^{\theta})^{\alpha-1}}{(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})^2} \tag{3}$$

2.1. Sine Generalized Odd Log-logistic Family

The cumulative distribution function and Probability density function of SGOLL-G is derived as;

$$F(x) = Sin\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})}\right] \tag{4}$$

The corresponding pdf of SGOLL-G is,

$$f(x) = \frac{\pi\alpha\theta g(x) G(x)^{\alpha\theta-1}(1 - G(x)^{\theta})^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta} + (1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})^2} \tag{5}$$

2.2. Mathematical Properties of SGOLL-G

In this part, we derive some properties of SGOLL-G Distribution like Survival function, hazard rate function, and quantile function.

$$S(x) = 1 - Sin\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})}\right] \tag{6}$$

$$H(x) = \frac{\pi\alpha\theta g(x)G(x)^{\alpha\theta-1}(1 - G(x)^{\theta})^{\alpha-1}Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})^2\left[1 - Sin\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]\right]} \tag{7}$$

$$x_u = G^{-1}\left[\frac{\left(\frac{2arcsinU}{\pi(1-\frac{2}{\pi}arcsinU)}\right)^{\frac{1}{\alpha}}}{1 + \left(\frac{2arcsinU}{\pi(1-\frac{2}{\pi}arcsinU)}\right)^{\frac{1}{\alpha}}}\right]^{\frac{1}{\theta}} \tag{8}$$

Where$G^{-1}(.)$ is the inverse of the baseline cdf. hence U is a uniform random variable on (0,1), then has SDOLL-G distribution.

However, other properties are entropy, moment, moment generating function, and order statistics were derived below;

2.2.1 Renyi's Entropy of SGOLL-G

The definition of Renyi's Entropy is given as,

$$RE = I_x = \frac{1}{1-k}log\int_{-\infty}^{\infty} f(x)^k dx \tag{9}$$

Substituting the PDF of SGLLO-G we have,

$$= \frac{1}{1-k}log\int_{-\infty}^{\infty}\left(\frac{\pi\alpha\theta g(x)G(x)^{\alpha\theta-1}(1 - G(x)^{\theta})^{\alpha-1}Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})^2}\right)^k dx \tag{10}$$

$$= \frac{1}{1-k}\left[log(\pi\alpha\theta)^k - klog2 + log\int_{-\infty}^{\infty}\left(\frac{g(x)G(x)^{\alpha\theta-1}(1 - G(x)^{\theta})^{\alpha-1}Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})^2}\right)^k dx\right] \tag{11}$$

$$= \frac{1}{1-k}\left[log(\pi\alpha\theta)^k - klog2 + log\beta\right] \tag{12}$$

$$= \frac{1}{1-k}\left[klog(\pi\alpha\theta) - klog2 + log\beta\right] \tag{13}$$

Where

$$\beta = \int_{-\infty}^{\infty}\left(\frac{g(x)G(x)^{\alpha\theta-1}(1 - G(x)^{\theta})^{\alpha-1}Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{(G(x)^{\alpha\theta} + (1 - G(x)^{\theta})^{\alpha})^2}\right)^k dx \tag{14}$$

2.2.2 Moment of SGOLL-G

The definition of Moment is given as'

$$\mu^r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x) dx \tag{15}$$

$$\mu^r = \frac{\pi\alpha\theta}{2} \int_0^{\infty} \frac{x^r g(x) G(x)^{\alpha\theta-1} (1-G(x)^{\theta})^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})^2} dx \tag{16}$$

$$\mu^r = \eta\tau \tag{17}$$

Where,

$$\tau = \int_0^{\infty} \frac{x^r g(x) G(x)^{\alpha\theta-1} (1-G(x)^{\theta})^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})^2} \tag{18}$$

and,

$$\eta = \frac{\pi\alpha\theta}{2} \tag{19}$$

2.2.3 Moment Generating function of SGOLL-G

The definition of Moment Generating function is given,

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \tag{20}$$

Substituting the PDF of SGOLL-G we have,

$$M_x(t) = \frac{\pi\alpha\theta}{2} \int_0^{\infty} \frac{e^{tx} g(x) G(x)^{\alpha\theta-1} (1-G(x)^{\theta})^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})^2} dx \tag{21}$$

Using taylor series expansion,

$$e^{tx} = \sum \frac{t^j x^j}{j!} \tag{22}$$

implies that,

$$M_x(t) = \frac{\pi\alpha\theta t^j}{2j!} \sum_{j=0}^{\infty} \int_0^{\infty} \frac{x^j g(x) G(x)^{\alpha\theta-1} (1-G(x)^{\theta})^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})^2} dx \tag{23}$$

$$\Rightarrow M_x(t) = \sum_{j=0}^{\infty} \varphi\vartheta \tag{24}$$

Where,

$$\vartheta = \int_0^{\infty} \frac{x^j g(x) G(x)^{\alpha\theta-1} (1-G(x)^{\theta})^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})}\right]}{2(G(x)^{\alpha\theta}+(1-G(x)^{\theta})^{\alpha})^2} dx \tag{25}$$

and,

$$\varphi = \frac{\pi\alpha\theta t^j}{2j!} \tag{26}$$

2.2.4 Order Statistics of SGOLL-G

The definition of Order Statistics is,

$$f_{(i,n)}(x) = \frac{n!}{(i-1)(n-i)} f(x)[F(x)]^{i-1}[1-F(x)]^{n-i} \tag{27}$$

Now using power series expansion;

$$[1-F(x)]^{n-1} = \sum_{j=0}^{n-i}(-1)^j \binom{n-i}{j}[F(x)]^j \tag{28}$$

$$f_{(i,n)}(x) = \frac{n!}{(i-1)(n-i)} \sum_{j=0}^{n-i}\binom{n-i}{j}(-1)^j f(x)[F(x)]^{i+j-1} \tag{29}$$

$$f_{(i,n)}(x) = \frac{n!(-1)^j}{(i-1)(n-i-j)} \sum_{j=0}^{n-i}\binom{n-i}{j} f(x)[F(x)]^{i+j-1} \tag{30}$$

$$\text{Now, } f_{(i,n)}(x) = \sum_{j=0}^{n-i}\binom{n-i}{j}\left[\frac{n!(-1)^j \pi\alpha\theta g(x)G(x)^{\alpha\theta-1}(1-G(x)^\theta)^{\alpha-1} Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]}{2(i-1)!(n-i-j)(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)^2}\right] \times \left[Sin\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]\right]^{i+j-1} \tag{31}$$

2.2.5 Maximum Likelihood Estimation

Let $x_1, x_2, ..., x_n$ be a random sample from the SGOLL-G family of distribution with pdf in equation (5) with $\bar{\kappa} = (\alpha, \lambda, \theta)^T$, the SGOLL-G's n sample log-likelihood is driven as:

$LL(\kappa) = nlog\pi + nlog\theta + \sum logg(x) + \sum logG(x)^{\alpha\theta-1} - \sum log(1-G(x)^\theta) - nlog2 - 2\alpha\theta \sum logG(x) +.$

$$\sum logCos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right] \tag{32}$$

$$\frac{dLL(\kappa)}{d\alpha} = \frac{n}{\alpha} - \theta\sum logG(x) + \sum \frac{\pi\theta G(x)^{\alpha\theta}(1-G(x)^\theta)^\alpha(1-log(1-G(x))tan\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)^2} \tag{33}$$

$$\frac{dLL(\kappa)}{d\theta} = \frac{n}{\theta} - \alpha\sum logG(x) + \sum \frac{G(x)^\theta logG(x)}{(1-G(x)^\theta)} + \sum \frac{\frac{d}{d\theta}Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]}{Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]} \tag{34}$$

$$\frac{dLL(\kappa)}{d\psi} = \sum\frac{g'(x)}{g(x)} - \sum\frac{g(x)}{G(x)} - \alpha\theta\sum\frac{g(x)}{G(x)} + \sum\frac{\theta g(x)G(X)^{\theta-1}}{(1-G(x)^\theta)} + \sum\frac{\frac{d}{d\psi}Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]}{Cos\left[\frac{\pi G(x)^{\alpha\theta}}{2(G(x)^{\alpha\theta}+(1-G(x)^\theta)^\alpha)}\right]} \tag{35}$$

2.3. Special member of the SGOLL-G family

In this section family member sine generalized odd log-logistic weibul were introduced as a special member of SGOLL-G, which were used for the simulation and application to real data. Considering Weibull distribution as a baseline ditribution we derive the Sine Generalized Odd Log-Logistic Weibul (SGOLL-W)Distribution, with CDF given by;

$$F(x,\alpha,\theta,\lambda,\omega) = Sin\left[\frac{\pi(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}}{2((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})}\right] \tag{36}$$

The corresponding pdf of SGOLL-W is;

$$f(x) = \frac{\pi\alpha\theta\omega x^{\omega-1}e^{-(\frac{x}{\lambda})^{\omega}}(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta-1}(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha-1}Cos\left[\frac{\pi(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}}{2((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})}\right]}{2\lambda^{\omega}((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})^{2}} \tag{37}$$

$$S(x) = 1 - Sin\left[\frac{\pi(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}}{2((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})}\right] \tag{38}$$

$$H(x) = \frac{\pi\alpha\theta\omega x^{\omega-1}e^{-(\frac{x}{\lambda})^{\omega}}(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta-1}(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha-1}Cos\left[\frac{\pi(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}}{2((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})}\right]}{2\lambda^{\omega}((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})^{2}\left[1-Sin\left[\frac{\pi(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}}{2((1-e^{-(\frac{x}{\lambda})^{\omega}})^{\alpha\theta}+(1-(1-e^{-(\frac{x}{\lambda})^{\omega}})^{\theta})^{\alpha})}\right]\right]} \tag{39}$$

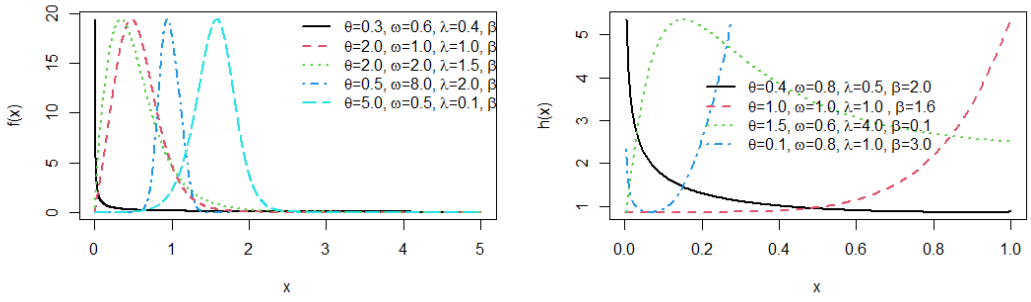


Figure 1: *The pdf and the hazard plot for some values of parameters of SGOLL-W*

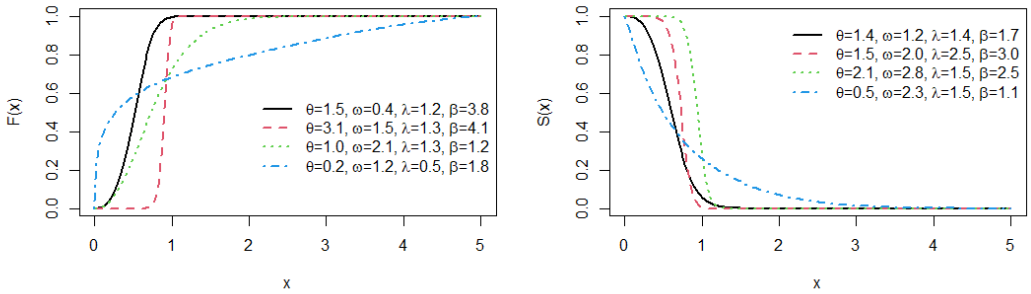


Figure 2: *The CDF and Survival plot for some values of parameters of SGOLL-W*

Figures 1 and 2 present PDF and HR, CDF and hazard plots of the SGOLLW distribution for the specified parameter values. The diagrams in Figure 1 demonstrate that the HR feature of the SGOLLW distribution can exhibits increasing, decreasing, bathtub shapes. While the PDF can have revered-j shape or right-skewed. This can be considered an advantage of the SGOLLW distribution because of its ability to model phenomena with increasing and decreasing shapes or bathtub failure rates, making SGOLLW more versatile for analyzing lifetime data.

3. Results

3.1. simulations Study

The behavior of the maximum likelihood of SGOLL-W for certain parameter values in the first trial (i.e. $\alpha = 0.8$, $\theta = 2.5$, $\omega = 3.4$, $\lambda = 4.5$) and second trial (i.e. $\alpha = 1.5$, $\theta = 1.3$, $\omega = 3.8$,$\lambda = 2.8$) was investigated using a finite sample. The random numbers for the SGOLL-W were generated using the quantile function. For 1000 repeats, samples with sizes of n=20, 50, 100, 250, 500, and 1000 were created. The Means, Bias, and RMSE were then calculated.Table 1 presents the outcomes of the simulation. We conclude that our model yields consistent results when predicting parameters for the mode based on the results of the Monte Carlo simulation.

Table 1: *Result of simulation for different values of parameters*

		$\alpha = 0.8$	$\theta = 2.5$	$\omega = 3.4$	$\lambda = 4.5$	$\alpha = 1.5$	$\theta = 1.3$	$\omega = 3.8$	$\lambda = 2.7$
n=20	Means	0.8415	2.6521	3.2461	4.9523	1.6197	1.3773	3.9759	2.7684
	Bias	0.0415	0.1521	0.2461	0.4523	0.1197	0.0773	0.1759	0.0684
	RMSE	0.2592	1.1345	0.7826	1.2290	0.4204	0.4576	0.4977	0.5142
n=50	Means	0.7953	2.6518	3.2039	4.8429	1.5253	1.3359	3.9630	2.7728
	Bias	-0.0047	0.1518	0.2039	0.3429	0.0253	0.0359	0.1630	0.0728
	RMSE	0.1759	0.9456	0.6222	1.0789	0.2406	0.3157	0.4162	0.3885
n=100	Means	0.7811	2.6235	3.1533	4.7364	1.4928	1.3362	3.9529	2.7533
	Bias	-0.0189	0.1235	0.1533	0.2364	-0.0072	0.0362	0.1529	0.0533
	RMSE	0.1247	0.7014	0.4316	0.8227	0.1652	0.2272	0.3190	0.2669
n=250	Means	0.7756	2.5626	3.1093	4.6981	1.4784	1.3170	3.9331	2.7568
	Bias	-0.0244	0.0626	0.1093	0.1981	-0.0216	0.0170	0.1331	0.0568
	RMSE	0.0969	0.4986	0.3363	0.6252	0.1120	0.1599	0.2783	0.2053
n=500	Means	0.7733	2.5596	3.1050	4.6583	1.4758	1.3119	3.9164	2.7477
	Bias	-0.0267	0.0596	0.1050	0.1583	-0.0242	0.0119	0.1164	0.0477
	RMSE	0.0881	0.3698	0.3023	0.4949	0.0904	0.1168	0.2648	0.1758
n=1000	Means	0.7741	2.5453	3.0949	4.6305	1.4752	1.3056	3.9074	2.7463
	Bias	-0.0259	0.0453	0.0949	0.1305	-0.0248	0.0056	0.1074	0.0463
	RMSE	0.0713	0.2629	0.2457	0.3592	0.0735	0.0837	0.2368	0.1347

3.2. Application

The sub-model were compared with some standard models using the waiting time and bladder tumors Data sets, we used methods of goodness of fit and information criteria, which comprises the log-likelihood function evaluated at the MLEs, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the corrected Akaike information criterion (CAIC), Hannan Quinn information criterion (HQIC), Anderson-Darling (A), Cramer-von Mises (W), and Kolmogorov-Smirnov (Ks). The competitors model were generalized odd log-logistic weibul, odd log-logistic weibul, topp-leone odd log-logistic weibul. Therefore, the better the fit is the one with smaller values of information criteria. aryal2017topp.

$$AIC = -2L + 2k \tag{40}$$

$$BIC = -2L + klog(n) \tag{41}$$

$$CAIC = -2L + \frac{2kn}{n-k-1} \tag{42}$$

$$HQIC = -2L_{max} + 2klog(log(n)) \tag{43}$$

3.2.1 Application to waiting time data

In this research, we have considered real data that contains 100 observations on minutes waiting time before a client receives the desired service in a bank: the data was used by zeineldin2021generalized and the data is given as:

0.8, 0.8, 1.3, 1.5, 1.8, 1.9, 1.9, 2.1, 2.6, 2.7, 2.9, 3.1, 3.2, 3.3, 3.5, 3.6, 4, 4.1, 4.2, 4.2, 4.3, 4.3, 4.4, 4.4, 4.6, 4.7, 4.7, 4.8, 4.9, 4.9, 5.0, 5.3, 5.5, 5.7, 5.7, 6.1, 6.2, 6.2, 6.2, 6.3, 6.7, 6.9, 7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8, 8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6, 9.7, 9.8, 10.7, 10.9, 11.0, 11.0, 11.1, 11.2, 11.2, 11.5, 11.9, 12.4, 12.5, 2.9, 13.0, 13.1, 13.3, 13.6, 13.7, 13.9, 14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0, 19.9, 20.6, 21.3, 21.4, 21.9, 23, 27, 31.6, 33.1, 38.5

Table 2: *Information criteria measure for the fitted models using waiting time data*

Distribution	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\omega}$	$\hat{\lambda}$	$\hat{LL}$	AIC	CAIC	BIC	HQIC
SGOLLW	2.13	26.68	2.46	0.16	317.00	642.00	642.43	652.42	646.22
GOLLW	66.66	59.78	4.41	0.01	319.27	646.54	646.96	656.96	650.76
OLLW	7.59	0.45	0.21	—	318.04	642.09	642.34	649.91	645.25
TLOLLW	5.88	28.65	O.71	0.09	325.88	659.78	660.20	670.20	663.99

Table 3: *Goodness of fit measure for the fitted models using waiting time data*

Distribution	KS	A	W	P-value
SGOLLW	0.042081	0.15976	0.021594	0.9944
GOLLW	0.0656	0.3507	0.0505	0.7825
OLLW	0.0456	0.2555	0.0365	0.9853
TLOLLW	0.3532	11.9933	2.1270	<0.000

Tables 2 displays the maximum likelihood estimates (MLEs) along with their standard errors and information criteria for four fitted models. These results pertain to the SGOLLW distribution and three competing models. All parameter estimates are significant across the fitted models, with the SGOLLW distribution demonstrating the lowest values in all information criteria. Therefore, the SGOLLW model emerges as a strong alternative to the other fitted models. Table 3 presents the gooness of fit statistics values specifically the Cram©r-von Mises (W*), Anderson"Darling (A*), and Kolmogorov"Smirnov (KS) along with its corresponding P values, for competing models fitted to the waiting time data sets. the SGOLLW distribution achieves the highest P value and the lowest distances for the Kolmogorov"Smirnov (KS), W*, and A* values. Figure 3 displays the empirical PDF and CDF plots for the SGOLLW distribution based on waiting time data. These findings suggest that the SGOLLW model offers a better fit compared to other distributions.

3.2.2 Application to breast tumors data

The second set of real data was obtained at Benha University Hospital in Egypt from June to October 2014. This data represents the ages of 155 patients with breast tumors in the early detection unit for breast cancer. The data are 46, 32, 50, 46, 44, 42, 69, 31, 25, 29, 40, 42, 24, 17, 35, 48, 49, 50, 60, 26, 36, 56, 65, 48, 66, 44, 45, 30, 28, 40, 40, 50, 41, 39, 36, 63, 40, 42, 45, 31, 48, 36, 18, 24, 35, 30, 40, 48, 50, 60, 52, 47, 50, 49, 38, 30, 52, 52, 12, 48, 50, 45, 50, 50, 50, 53, 55, 38, 40, 42, 42, 32, 40, 50, 58, 48, 32, 45, 42, 36, 30, 28, 38, 54, 90, 80, 60, 45, 40, 50, 50, 40, 50, 50, 50, 60, 39, 34, 28,

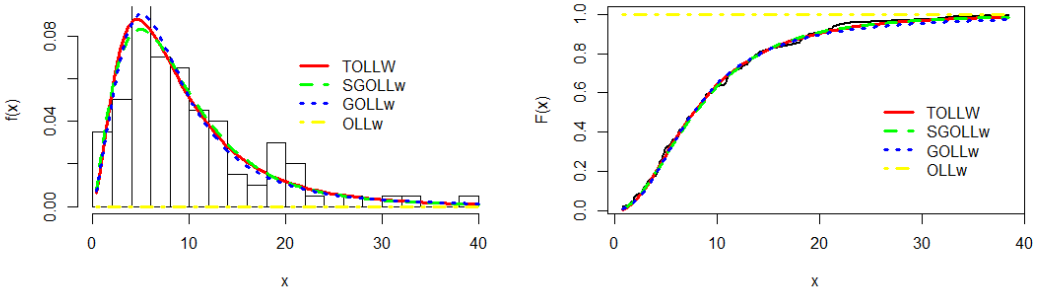


Figure 3: *Estimated pdf and cdf of the SGOLL-W and other distributions for waiting time data*

18, 60, 50, 20, 40, 50, 38, 38, 42, 50, 40, 36, 38, 38, 50, 50, 31, 59, 40, 42, 38, 40, 38, 50, 50, 50, 40, 65, 38, 40, 38, 58, 35, 60, 90, 48, 58, 45, 35, 38, 32, 35, 38, 34, 43, 40, 35, 54, 60, 33, 35, 36, 43, 40, 45, 56´.

Table 4: *Table 4: Information criteria measure for the fitted models using breast tumors data*

Distribution	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\omega}$	$\hat{\lambda}$	$\hat{LL}$	AIC	CAIC	BIC	HQIC
SGOLLW	18.53	90.55	4.15	0.04	599.76	1207.532	1207.79	1219.71	1212.47
GOLLW	55.48	90.37	4.57	0.02	601.62	1211.24	1211.50	1223.41	1216.18
OLLW	0.0509	6.2462	0.3948	—	958.16	1922.31	1922.47	1931.44	1926.02
TLOLLW	0.14	50.14	O.14	1.23	602.59	1213.19	1213.46	1225.37	1218.14

Table 5: *Goodness of fit measure for the fitted models using breast tumors data*

Distribution	KS	A	W	P-value
SGOLLW	0.0801	0.8190	0.1431	0.2874
GOLLW	0.0818	1.0636	0.1758	0.2509
OLLW	0.7211	1.3358	0.2117	<0.001
TLOLLW	0.0917	1.1823	0.1902	0.1474

Table 4 presents the maximum likelihood estimates (MLEs), their standard errors, and information criteria for four fitted models, including the SGOLLW distribution and three competing models. All parameter estimates across the fitted models are significant, with the SGOLLW distribution yielding the lowest values for all information criteria. As a result, the SGOLLW model is identified as a strong alternative to the other fitted models. Table 5 provides the goodness-of-fit statistics, specifically the Cram©r-von Mises (W*), Anderson-Darling (A*), and Kolmogorov-Smirnov (KS) values, along with their corresponding P values for the competing models fitted to the waiting time data sets. The SGOLLW distribution stands out with the highest P value and the lowest distances in the Kolmogorov-Smirnov (KS), W*, and A* statistics. Figure 4 illustrates the empirical PDF and CDF plots for the SGOLLW distribution based on the breast tumors data. These results indicate that the SGOLLW model offers a superior fit compared to the other distributions.

4. Discussion

In this paper, we introduce the Sine generalized odd log-logistic family of distributions. We derive several properties of this distribution, including the quantile, entropy, hazard rate function, survival function, moments, moment generating function, and order statistics. Additionally, we estimate the parameters of the SGOLL-G distribution using maximum likelihood estimation. We examine the generalized Sine odd log-logistic Weibull distribution using two data sets. Through

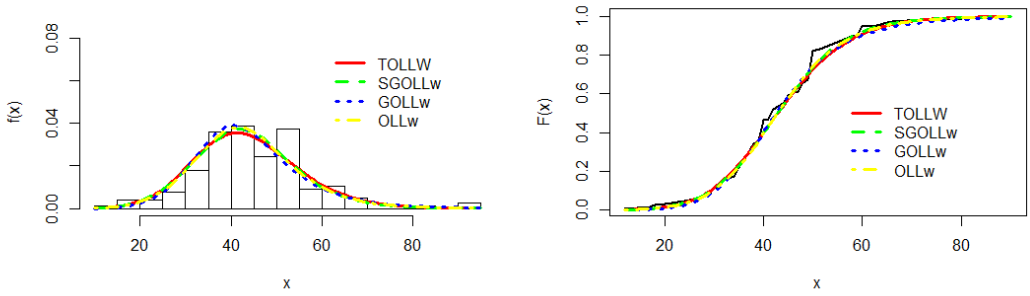


Figure 4: *Estimated pdf and cdf of the SGOLL-W and other distributions for breast tumors data*

analysis and simulation, we demonstrate the performance of the proposed distribution compared to other underlying models. The goodness-of-fit measures and information criteria, such as AIC, CAIC, and BIC, indicate that the SGOLLW distribution outperforms its competitors.

References

[1] Cordeiro, G. M., Alizadeh, M., Silva, R. B., and Ramires, T. G. (2017). Experimental observations of bedload tracer movement: Effects of mixed particle sizes and bedforms. *Water Resources Research*, 59:215–232.

[2] Singh, A. Wu, Z., Wilcock, P. and Foufoula-Georgiou, E. (2023). A new wider family of continuous models: the extended Cordeiro and De Castro family. *Hacettepe Journal of Mathematics and Statistics*, 47(4):937–961.

[3] Alizadeh, M., Tahmasebi, S., Kazemi, M. R., Nejad, H. S. and Hamedani, G. H. (2019). The odd log-logistic gompertz lifetime distribution: properties and applications. *Studia Scientiarum Mathematicarum Hungarica*, 56(1):55–80.

[4] Alizadeh, M., Rasekhi, M., Yousof, H. M., Hamedani, G. and Ataei, A. (2022). The Odd Log-Logistic Transmuted-G family of distributions: properties, characterization, applications and different methods of estimation. *Statistics, Optimization & Information Computing*, 10(3):904–924.

[5] Alizadeh, M., Tahmasebi, S. and Haghbin, H. (2020). The exponentiated odd log-logistic family of distributions: Properties and applications. *Journal of Statistical Modelling: Theory and Applications*, 1(1):29–52.

[6] Ranjbar, V., Eftekharian, A., Kharazmi, O. and Alizadeh, M. (2023). Odd log-logistic generalised Lindley distribution with properties and applications. *Statistics in Transition new series*, 24(4):71–92.

[7] Haghbin, H., Ozel, G., Alizadeh, M. and Hamedani, G. (2017). A new generalized odd log-logistic family of distributions. *Communications in Statistics-Theory and Methods*, 46(20):9897–9920.

[8] Aryal, G. R., Ortega, E. M., Hamedani, G. and Yousof, H. M. (2017). The Topp-Leone generated Weibull distribution: regression model, characterizations and applications. *International Journal of Statistics and Probability*, 6(1):126–141.

[9] Gleaton, J. U. and Lynch, J. D. (2010). Extended generalized log-logistic families of lifetime distributions with an application. *J. Probab. Stat. Sci*, 8(1):1–17.

[10] ZeinEldin, R. A., Chesneau, C., Jamal, F., Elgarhy, M., Almarashi, A. M. and Al-Marzouki, S. (2021). Generalized Truncated Fŕechet Generated Family Distributions and Their Applications. *Computer Modeling in Engineering & Sciences*, 126(2):791–819.

[11] Cordeiro, G. M., Alizadeh, M., Ozel, G. Hosseini, B., Ortega, E. M. and Altun, E. (2017). The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of statistical computation and simulation*, 87(5):908–932.

[12] Kailembo, B. B., Gadde, S. R. and Kirigiti, P. J. (2024). Application of Odd Lomax log-logistic distribution to cancer data. *Heliyon*, 10(6):108–132.

[13] Fulment, A. K., Gadde, S. R. and Peter, J. K. (2022). The odd log-logistic generalized exponential distribution: Application on survival times of chemotherapy patients data. *F1000research*, 11:1–13.

[14] Ranjbar, V., Eftekharian, A., Kharazmi, O. and Alizadeh, M. (2023). Odd log-logistic generalised Lindley distribution with properties and applications. *Statistics in Transition new series*, 24(4):71–92.

[15] Abonongo, J. (2024). Actuarial measures, aggregate loss models, and insurance applications of the sine exponentiated Burr XII distribution. *Research in Mathematics*, 11(1).

[16] Isa, A. M., Doguwa, S. I., Alhaji, B. B. and Dikko, H. G. (2023). Sine Type II Topp-Leone G Family of Probability Distribution: Mathematical Properties and Application. *Arid Zone Journal of Basic and Applied Research*, 24(4):124–138.

[17] Sapkota, L. P., Kumar, P. and Kumar, V. (2023). A New Class of Sin-G Family of Distributions with Applications to Medical Data. *Reliability: Theory & Applications*, 18(74):734–750.

[18] Benchiha, S., Sapkota, Laxmi P., Al Mutairi, A., Kumar, V., Khashab, R. H., Gemeay, A. M., Elgarhy, M. and Nassr, S. G. (2023). A new sine family of generalized distributions: Statistical inference with applications. *Mathematical and Computational Applications*, 28(4):83.

[19] Muhammad, M., Alshanbari, H, M., Alanzi, R., A., Liu, L., Sami, W., Chesneau, C. and Jamal, F. (2021). A new generator of probability models: the exponentiated sine-G family for lifetime studies. *Entropy*, 23(11):1394.

[20] Jamal, F., Chesneau, C., Bouali, D. L. and Ul Hassan, M. (2021). Beyond the Sin-G family: The transformed Sin-G family. *PLoS One*, 16(5).

[21] Chesneau, C. and Jamal, F. (2020). The sine Kumaraswamy-G family of distributions. *Journal of Mathematical Extension*, 15.

DOCKER CONTAINER PLACEMENT IN DOCKER SWARM CLUSTER BY USING WEIGHTED RESOURCE OPTIMIZATION APPROACH

Jalpa M Ramavat[1], Dr Kajal S Patel[2]

•

Research Scholar, Gujarat Technological University[1]
jalpa.ramavat.2012@gmail.com
Associate Professor, VGEC, Chandkheda[2]
kajalpatel@vgecg.ac.in

Abstract

The use of Docker containers and their orchestration tools is rapidly improving as Web application deployment shifts from a server- or VM-based approach to a container-based approach. Docker Swarm is a flexible and simple container orchestration tool. it is widely used by application developers for the deployment of their applications in a containerized environment. Docker Swarm uses the default spread strategy for placing new containers on cluster nodes. This strategy distributes containers evenly on all nodes of the cluster, but it will not consider the current resource utilization of nodes or heterogeneous resource availability on cluster nodes. Again, all task containers are treated similarly, irrespective of their specific resource-oriented nature. This paper proposes the weighted resource optimization algorithm for calculating the weighted score of each node. Score depends on CPU and memory weight for a given task and the availability of that resource on the node. The task container is placed on the node with the highest score. This approach improves CPU and memory load balancing in a Docker cluster and also improves the completion time of the task container as compared to the spread strategy.

Keywords: Cloud, VM, Containers, Docker Swarm, Orchestration

I. Introduction

Virtual machines (VMs) are widely utilized in organizations, even for running minor applications, leading to system inefficiencies. There's a growing demand for lightweight application deployment to enhance efficiency. Docker containers have emerged as a solution, offering lightweight virtualization [2].

As the demand for containers grows, efficient scheduling and orchestration mechanisms are paramount for distributing container executions across numerous nodes in cloud clusters. Consequently, various container scheduling tools have emerged, including Docker Swarm from Docker, Mesos from Apache, and Kubernetes from Google [10].

Docker Swarm is a container orchestration tool developed by Docker [11]. The Docker Swarm cluster has two types of nodes: masters and workers. Any node can leave the cluster at any time. Again, a new node can be added to the cluster by using tokens generated [12]. A Docker Swarm cluster has components like tasks, services, Raft Consensus Groups, internal distributed state stores, managers, and worker nodes [13]. A task is a combination of a container and a command to run it, and a service is represented by a single task or replicas of a task [13]. Manager nodes handle API requests, orchestration, allocation of IP addresses, scheduling, and control over worker nodes [11].

Worker nodes receive task containers and execute them [11]. Raft Consensus Groups are for handling fault tolerance by selecting a new leader if the current one becomes unavailable [14]. Docker Swarm integrates seamlessly with the Docker environment, simplifying container deployment and management. It allows decentralized design, declarative model, scaling, error correction, networking, load balancing, security, and version updates with restore points [15]. External load balancing and graphical interface for management were added after May 2019 [1].

Docker Swarm is commonly implemented without prior information about the workload or the specific resource requirements of containers. Consequently, it exclusively utilizes a single scheduling strategy referred to as "Spread" [10].

This policy aims to ensure that the workload is evenly distributed. It tries maintain equal number of running containers on each node within the cluster. If sufficient resources are available on a node having least number of running containers than new container will be placed on that node irrespective of resource utilization of that node by currently running containers. So, it tries to spread container equally among all Docker cluster nodes.

The spread strategy will not consider current resource utilization of node before placing containers. There might be load imbalance when containers are specific resource oriented i.e. CPU bound, Memory bound etc.

Over the past years, container scheduling has garnered considerable research attention. The author [7] introduces an Availability-Based Prioritization (ABP) scheduler to enhance service availability and scheduling efficiency within Docker Swarm. It begins by detailing a Service Availability Model. This model calculates service availability based on task distribution across nodes and considers mean time between failures (MTBF) and mean time to repair (MTTR). This model defines availability as the valid probability of nodes hosting service replicas, with replica requirements determined by node validity. The Modules Design section outlines the functionality of various components within Docker Swarm, with the Spread strategy serving as the default scheduling mechanism. The heart of the paper lies in the Scheduling Strategy of the ABP Scheduler. The periodic decisions are made to scale services based on task requirements and current distributions. These decisions prioritize tasks and nodes based on replica gaps, image layer coincidence, and dominant shares, ensuring efficient resource utilization. The Scale-Out and Scale-In decisions sections elaborate on how the ABP scheduler handles service scaling. Overall, the ABP scheduler presents a promising approach to dynamically adjusting service replicas to meet availability targets while optimizing resource allocation within Docker Swarm.

Mao et al. [4] analyzed two cloud platforms, Docker and Kubernetes for finding how well resource management is done on both of these domains. Author created a container monitoring system using Prometheus and Grafana which continuously monitor how much resource usage is done by each job on the worker nodes. Results show that by altering the default configurations on Docker and Kubernetes, the completion times were lowered by up to 79.4% and 69.4%.

The paper introduces ECSched [8], a container scheduling solution for heterogeneous clusters. This approach overcomes limitations of traditional queue-based schedulers by leveraging a graph-based approach and modeling scheduling as a Minimum Cost Flow Problem (MCFP). While ECSched demonstrates superior performance in terms of container completion time and resource utilization compared to baseline schedulers. But it introduces increased complexity in algorithm runtime, particularly noticeable when processing large numbers of concurrent container requests. This heightened computational demand could potentially pose challenges in highly dynamic or resource-constrained environments which is limiting the scalability of ECSched in certain scenarios. For container deployment of application tasks Wu and Xia [5] proposed a model for deploying containers at lowest possible deployment cost. They also proposed an improved PSO, known as the CD-PSO algorithm, to offer the best solution for application task loading

The adCFS [9] (Adaptive CPU Fair Sharing) policy aims to dynamically adjust CPU resource

allocation in containerized workflow systems based on workload characteristics, such as task runtime, CPU usage, and number of tasks. It utilizes a CPU State Predictor (CSP) to forecast CPU usage states and a Container's CPU Weight Scaler to redistribute CPU resources among containers accordingly. Two variations of the policy, soft (L1) and force (L2), are implemented based on CPU contention levels (cautious and severe states). L2 is enforcing strict CPU allocation based on estimated weights. This adaptive approach aims to improve fairness and efficiency in CPU sharing for scientific workflow processing. Experimental findings demonstrate a 12% improvement in container response time when compared to the default CFS policy.

Guanqaun Wu [6] proposed an improved ant colony algorithm to optimize Docker swarm cluster resource allocation. To reduce the start time of search they initialized pheromone by using scheduling algorithm minimum task first completion. Then they used balance factor to guide local and global pheromone updates for next iteration. For improving global search ability of algorithm, they use volatilization coefficient adjustment mechanism. The results show improvement of overall performance of cluster.

II. Methods

I. Docker Swarm Default strategy.

Docker swarm is a popular orchestration tool for containerized cloud environment. It has default scheduling strategy known as spread strategy. As its name suggest this strategy spread the container across all nodes in cluster equally. It tries to balance number of running containers on each node. However, it will not consider resource utilization of a node. This may lead to resource cluster node load unbalance and increasing execution time of service task container.

II. Proposed strategy.

To consider the resource availability and current resource utilization, this paper proposes a strategy which can be merge with existing docker swarm spread strategy to improve overall performance of the Service.

The proposed strategy will not only consider the resource of node but also focus on task type. Different Services require different resources. Services can be categorized depending on the intensity of resource they used. User has to specify the weight of resources i.e. CPU, Memory. If Service is CPU intensive than weight of CPU is higher and if Service is memory oriented then the weight of memory is higher. Weight can be given between 0 to 1.

Let N be the number of nodes in Docker cluster.

Let WCPU and WMEM be the weight assigned to CPU and Memory resources.

Let UCi is CPU usage of ith node and UMi is the memory usage of ith node in percentage.

The score Si for a node i is calculated as follows:

$Si = WCPU \times (100 - UCi) + WMEM \times (100 - UMi)$

The node selection process involves computing the score S for each node and selecting the node with the highest score:

So, Selected node is having score:

$Smax = max(S1, S2, \ldots, SN)$ Where N is total number of Nodes.

III. Proposed Algorithm (Weighted Resource Optimization):

Step 1. Start the Docker Swarm Cluster
Step 2: Input WCPU and WMEM
Step 3: For each service in the list do step 3 to 7
Step 4: For each node i in cluster
Step 5: Find score using

$$Si = WCPU \times (100 - UCi) + WMEM \times (100 - UMi)$$

Step 6: Selected node is having score Smax=max (S1, S2, …, SN)
Step 7: Create service for image and place its container on selected node

III. Results and Discussion

I. Experiment Setup

Three virtual machines were created using VirtualBox version 6.1.36. Manager virtual machine was allocated 2 CPU cores, 4 GB of RAM, and 40 GB of storage. Worker1 virtual machine was allocated 2 CPU cores, 3 GB of RAM, and 24 GB of storage. Worker2 virtual machine was allocated 2 CPU cores, 3 GB of RAM, and 26 GB of storage. Ubuntu 20.04 was installed on each virtual machine following the standard installation process.

Docker Engine version 24.0.2 was installed on each Ubuntu 20.04 virtual machine using the official Docker installation script provided by Docker. The Docker Python API (version 6.0.1) was utilized to programmatically interact with the Docker Swarm cluster for dynamic container placement. Python scripts were developed to leverage the Docker API for tasks such as creating Docker services, removing services, calculating Completion time etc.

Swarmprom[3] is a comprehensive toolkit designed for monitoring Docker Swarm environments. It includes essential monitoring components such as Prometheus, Grafana, cAdvisor, Node Exporter, Alert Manager, and Unsee, providing a complete solution for monitoring and managing Docker Swarm clusters.

Despite deploying all Swarmprom services, the focus of the experiment was primarily on utilizing Prometheus for metric collection and Grafana for visualization.

While Prometheus and Grafana were utilized for the experiment, it's noted that the Node Exporter, cAdvisor, and Docker Exporter containers were deployed on the worker nodes. These components contribute to comprehensive monitoring by collecting system and container metrics from the worker nodes.

II. Implementation and Results

Initially the status of three nodes is as shown in Figure 1 with containers for swarmprom are running in all the three nodes. It is shown that Manager node is having total 8 running containers. Worker1 and Worker2 nodes are having 3 running containers.

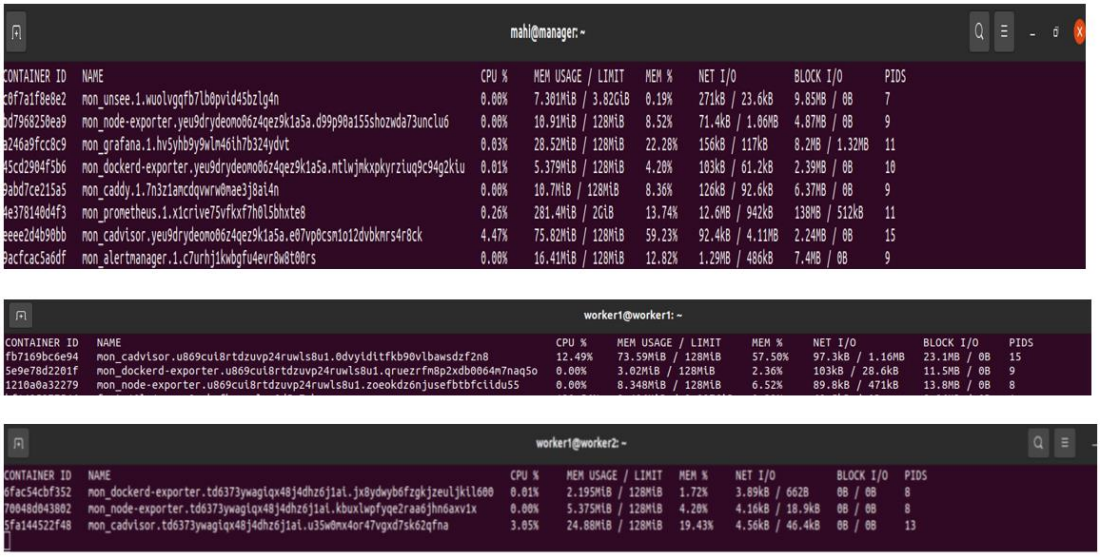


Figure 1 *Initial Docker Swarm Cluster Node -Manager, Worker1 and Worker2 with Swarmprom stack*

Initially, created images for the service of finding factorial of 500k ,300k and 100k. As factorial service requires more calculations, they are CPU intensive. Same way the images for creating and manipulating 4k by 4k and 5k by 5k arrays are also created and pushed in Docker hub. As storing 4K by 4K and 5K by 5K matrix required more memory, these are categorized under Memory intensive tasks.

After that to see the behavior of default spread strategy, five Nginx service containers are created and placed on Worker1 node by using docker placement constraint. The placement of these service containers will be as shown in Figure 2.

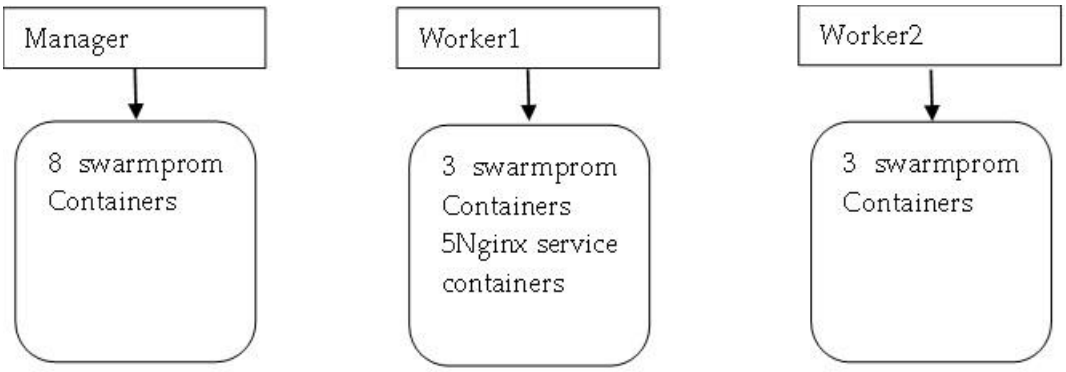


Figure 2 *Initial stat of Nodes after Swarmprom and Nginx container placement*

Now all services for images fact_500k, mem_5k, fact_300k, mem_4k and fact_100k are created and placed by default placement strategy. As spread distributes equal number of containers to all nodes the placement was seen in Figure 3. As manager is having 8 containers of swarmprom, worker1 is having 3 containers of swarmprom plus 5 containers of Nginx and worker2 is having least number of 3 containers of swarmprom stack so new created 5 service containers are placed on worker2 by Spread Strategy.

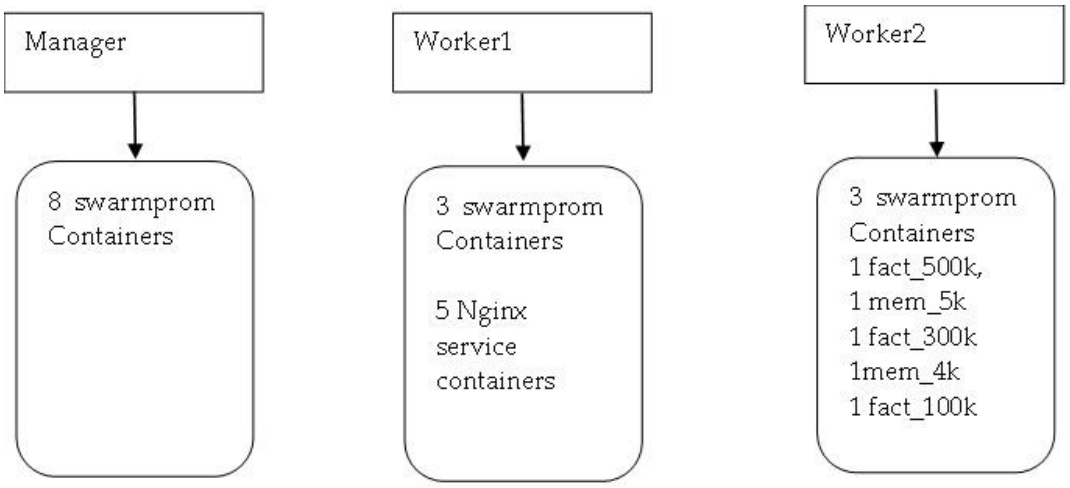


Figure 3 *Factorial and matrix manipulation task container placement (By Default Spread Strategy)*

Table 1 *CPU Utilization of Docker Swarm Cluster Nodes (Spread Strategy)*

Time(in Seconds)		0	90	180	270	360	450	540	630	720	810	900
CPU Usage (in %)	Manager	12.1	8.94	8.961	13.98	10.09	12.94	9.03	11.95	25	25.8	24.89
	Worker1	14.7	15.09	15.78	19.44	16.08	14.96	14.7	14.5	16.24	15.23	14.96
	Worker2	93.43	99.9	95.23	98.76	97	94.75	96.83	99.66	97.43	99.9	98

Table 2 *Memory Utilization of Docker Swarm Cluster Nodes (Spread Strategy)*

Time(in Seconds)		0	30	60	90	120	150	180	210	240	270	300
Memory Usage (in %)	Manager	61.61	62.47	62.76	61.89	57.64	57.67	57.77	58.84	59.30	59.10	58.33
	Worker1	33.58	33.65	33.65	33.61	33.58	33.58	33.65	33.92	33.85	33.82	33.68
	Worker2	68.48	59.41	60.99	60.86	59.44	67.97	51.01	55.33	64.57	58.46	57.01

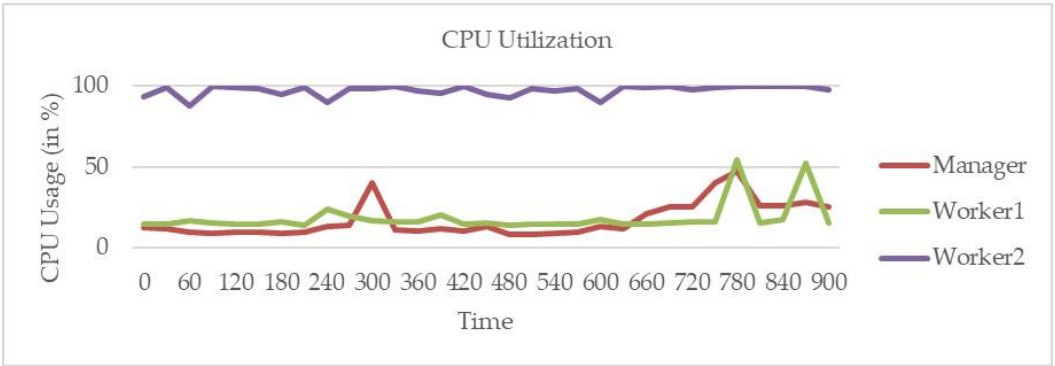


Figure 4 *Docker Swarm Cluster CPU Utilization after container placement by Default Spread Strategy*

Figure 5 *Docker Swarm Cluster Memory Utilization after container placement by Default Spread Strategy*

The CPU and Memory utilization of all cluster nodes are taken from Prometheus and Graphana as shown in Table 1 and Table 2. Now the readings are plotted as displayed in Figure 4 and 5. It is clear that the CPU load distribution is totally not even because number of containers in Worker1 and Manager node are more, all new created task containers are placed on Worker2 node by Spread Strategy without considering the resource utilization of all cluster nodes. Similarly, the memory usage is also not evenly distributed.

To avoid this unbalanced load distribution, the same containers are placed by using proposed strategy. First the results are taking using WCPU =0.7 and WMEM = 0.3. The task containers placement is shown in Figure 6.

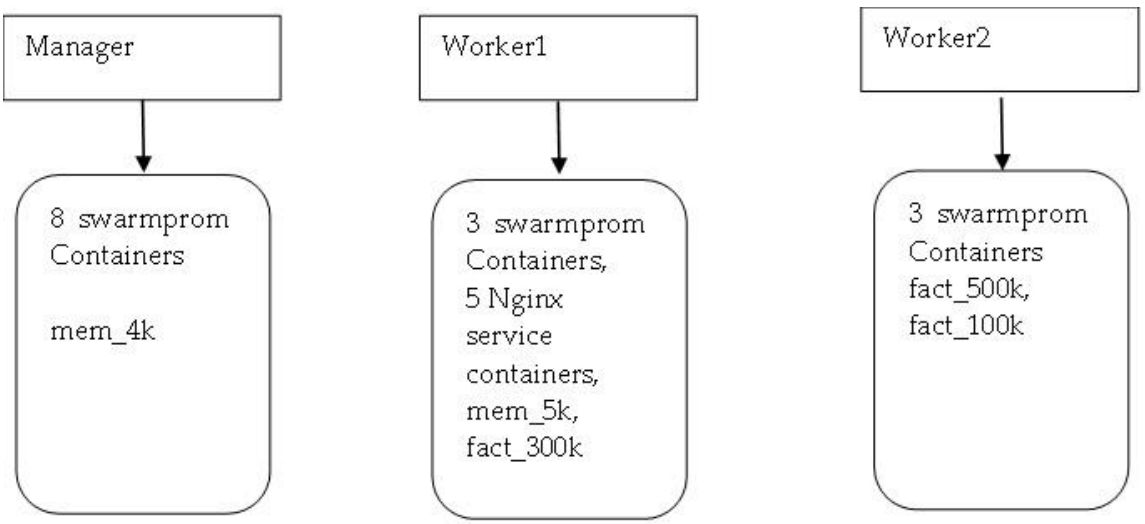


Figure 6 *Factorial and matrix manipulation task container placement (By Proposed Strategy: WCPU >WMEM)*

The CPU utilization of all three cluster nodes is as shown in Table 3 and plotted in Figure 7. The Memory utilization of all three cluster nodes is as shown in Table 4 and plotted in Figure 8.

Table 3 *CPU Utilization of Docker Swarm Cluster Nodes (Proposed Strategy: WCPU > WMEM)*

Time(in Seconds)		0	90	180	270	360	450	540	630	720	810	900
CPU Usage (in %)	Manager	37.53	50.40	42.46	44.45	31.57	34.56	32.44	33.23	33.00	49.81	34.26
	Worker1	76.50	71.77	77.53	76.11	73.94	75.57	84.43	77.37	83.28	77.36	79.77
	Worker2	60.43	65.83	72.80	68.23	48.31	68.67	69.23	57.33	68.33	69.17	65.30

Table 4 *Memory Utilization of Docker Swarm Cluster Nodes (Proposed Strategy: WCPU > WMEM)*

Time (in Seconds)		0	30	60	90	120	150	180	210	240	270	300
Memory Usage (in %)	Manager	53.86	53.37	53.50	59.25	53.25	56.24	53.32	53.73	53.20	68.48	62.81
	Worker1	51.18	65.75	62.88	64.30	62.91	61.77	62.41	61.46	64.67	61.83	63.65
	Worker2	32.97	33.34	33.61	33.88	33.88	34.36	34.29	34.52	34.12	34.39	34.19

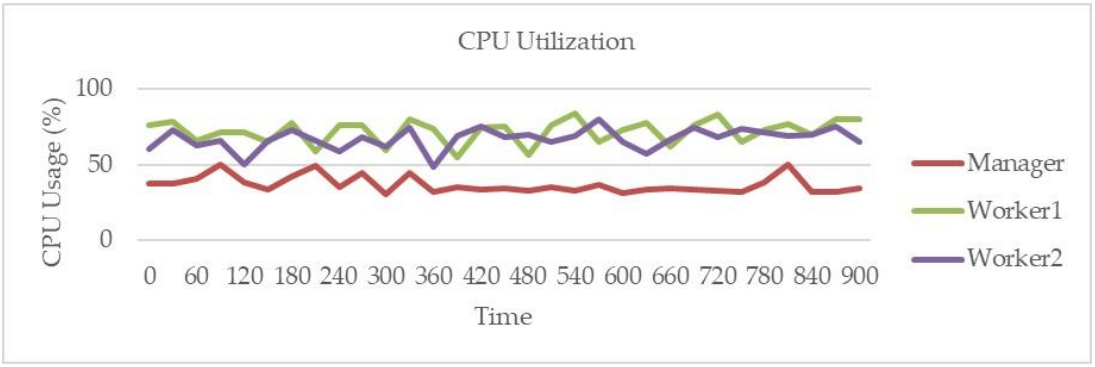


Figure 7 *Docker Swarm Cluster CPU Utilization after container placement by Proposed Strategy (WCPU > WMEM)*

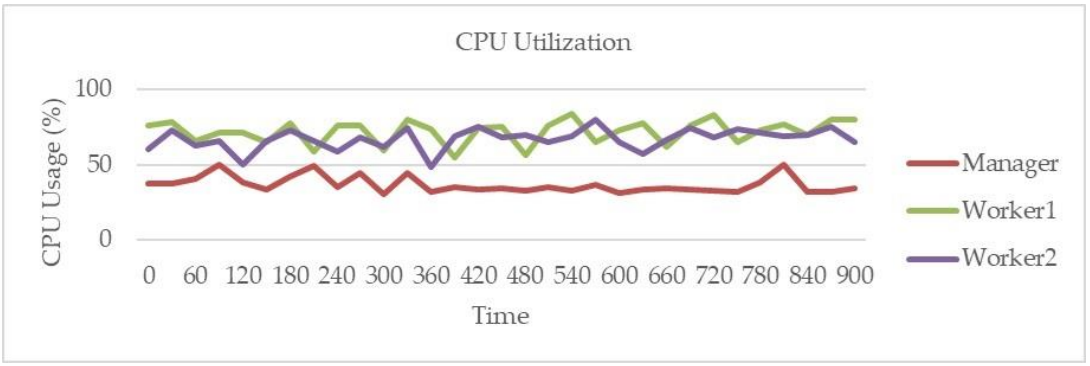


Figure 8 *Docker Swarm Cluster Memory Utilization after container placement by Proposed Strategy (WCPU > WMEM)*

As shown in above Figure 7 the CPU utilization of node is much balanced than as in spread technology shown in Figure 4. As shown in Figure 8 the memory Utilization of worker2 node is less as both are CPU oriented tasks are running on it.

Now the results are taken by using Proposed approach with W_{CPU}=0.3 and W_{MEM}-0.7. The new service containers are distributed as shown in Figure 9.

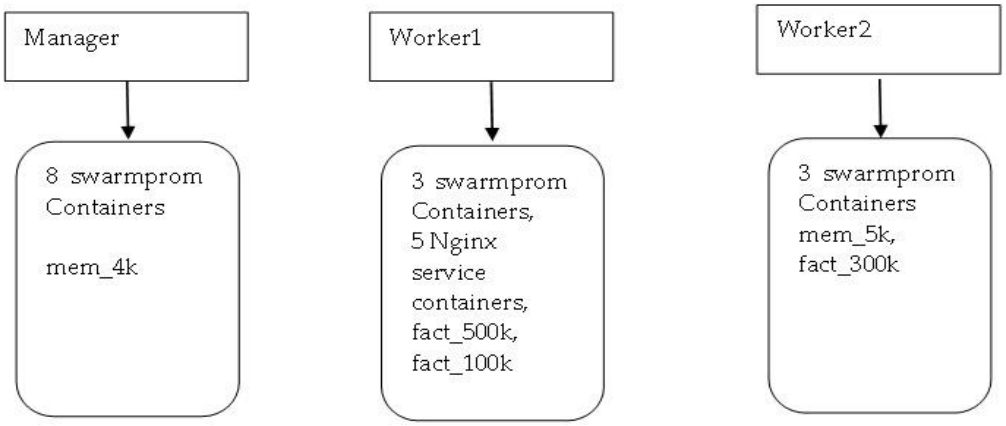


Figure 9 *Factorial and matrix manipulation task container placement (By Proposed Strategy: WCPU < WMEM)*

Table 5 *CPU Utilization of Docker Swarm Cluster Nodes (Proposed Strategy: WCPU < WMEM)*

Time (in Seconds)		0	90	180	270	360	450	540	630	720	810	900
CPU Usage (in %)	Manager	47.53	28.79	32.96	26.42	33.19	31.66	42.73	37.88	36.66	36.68	37.20
	Worker1	78.53	20.50	66.80	59.57	70.05	57.35	60.23	63.43	51.16	69.87	63.73
	Worker2	69.96	55.43	65.57	72.07	65.23	47.63	60.40	56.23	72.72	68.48	60.98

Table 6 *Memory Utilization of Docker Swarm Cluster Nodes (Proposed Strategy: WCPU < WMEM)*

Time(in Seconds)		0	30	60	90	120	150	180	210	240	270	300
Memory Usage (in %)	Manager	38.60	50.46	53.67	38.94	42.95	51.67	44.27	43.70	44.03	53.66	43.39
	Worker1	66.49	56.91	36.61	44.17	68.71	53.64	65.00	62.10	71.44	41.03	40.93
	Worker2	57.26	57.52	67.26	57.26	57.31	57.95	51.39	57.03	60.57	57.57	54.16

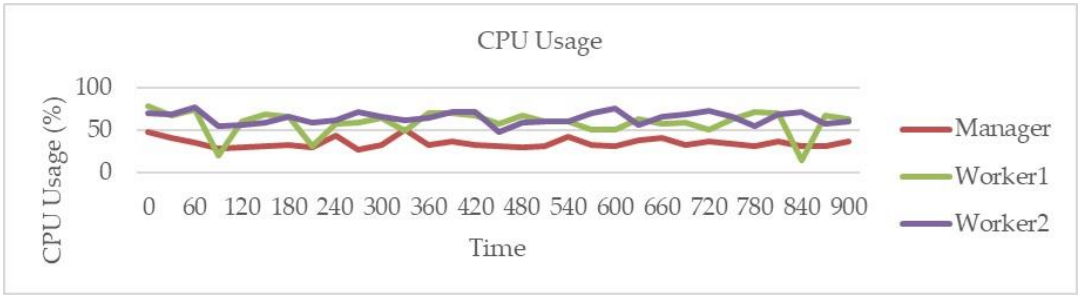


Figure 10 *Docker Swarm Cluster CPU Utilization after container placement by Proposed Strategy (WCPU < WMEM)*

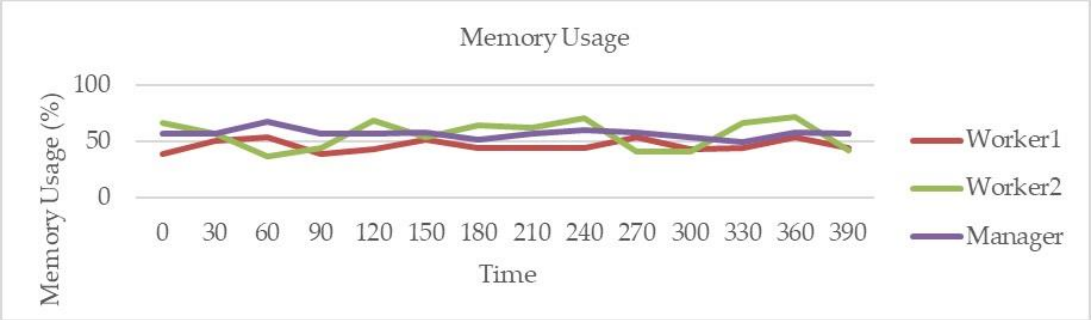


Figure 11 *Docker Swarm Cluster Memory Utilization after container placement by Proposed Strategy (WCPU < WMEM)*

Table 5 and Table 6 shows the CPU and Memory utilization of nodes by placing container using proposed algorithm by taking memory-oriented weight. As shown in Figure 11the memory load distribution is done more evenly. Also, CPU utilization is balanced as shown in Figure 10.

Now the completion time of containers of all services are calculated as shown in Table 7 and plotted in Figure 12. These readings are taken for 10 containers of each service and as factorial of 500k takes more time so 7 containers of it are taken. It can be seen from Figure 12 that the noticeable completion time reduction is achieved by using proposed approach.

Table 7 *Completion Time of task containers for Spread, Proposed (WCPU < WMEM, WCPU > WMEM)*

Services		ontainer1	ontainer2	ontainer3	ontainer4	ontainer5	ontainer6	ontainer7	ontainer8	ontainer9	ontainer10	Avg
Proposed $W_{CPU} > W_{MEM}$	fact_500k	124.65	166.55	139.70	135.67	116.23	115.80	116.90				130.79
	fact_100k	10.49	11.58	19.10	21.44	16.20	10.58	13.76	18.81	13.47	15.64	15.11
	fact_300k	49.83	51.26	45.23	46.85	43.98	44.74	47.44	44.90	44.54	46.49	46.53
	Mem_4k	7.95	8.68	8.80	7.86	8.02	7.76	9.58	7.97	8.15	8.23	8.23
	Mem_5k	14.13	16.86	14.93	17.42	18.03	11.20	10.37	10.32	12.28	11.97	13.75
Proposed $W_{CPU} < W_{MEM}$	fact_500k	121.12	126.29	125.52	123.28	119.64	123.60	123.73	-	-	-	123.28
	fact_100k	10.68	11.52	12.68	10.73	12.57	11.62	28.95	28.87	25.30	25.30	17.82
	fact_300k	43.33	43.36	40.63	39.58	42.76	41.54	41.21	43.07	43.44	38.75	41.77
	Mem_4k	8.10	7.81	8.40	8.46	8.28	8.15	7.94	8.56	8.11	8.04	8.18
	Mem_5k	12.43	10.15	9.91	9.76	11.17	12.84	10.81	10.27	11.42	9.27	10.80
Spread	fact_500k	234.44	230.93	251.08	246.21	227.47	277.72	233.15	-	-	-	243.00
	fact_100k	19.54	19.04	25.75	16.45	17.37	17.20	16.98	19.41	19.30	17.47	18.85
	fact_300k	90.72	94.23	82.33	88.72	81.10	86.17	83.51	91.76	82.94	83.31	86.48
	Mem_4k	16.31	24.48	22.00	22.57	16.31	20.22	19.02	19.29	17.52	16.57	19.43
	Mem_5k	23.88	17.84	22.80	23.95	19.56	16.61	20.12	21.68	15.89	15.77	19.81

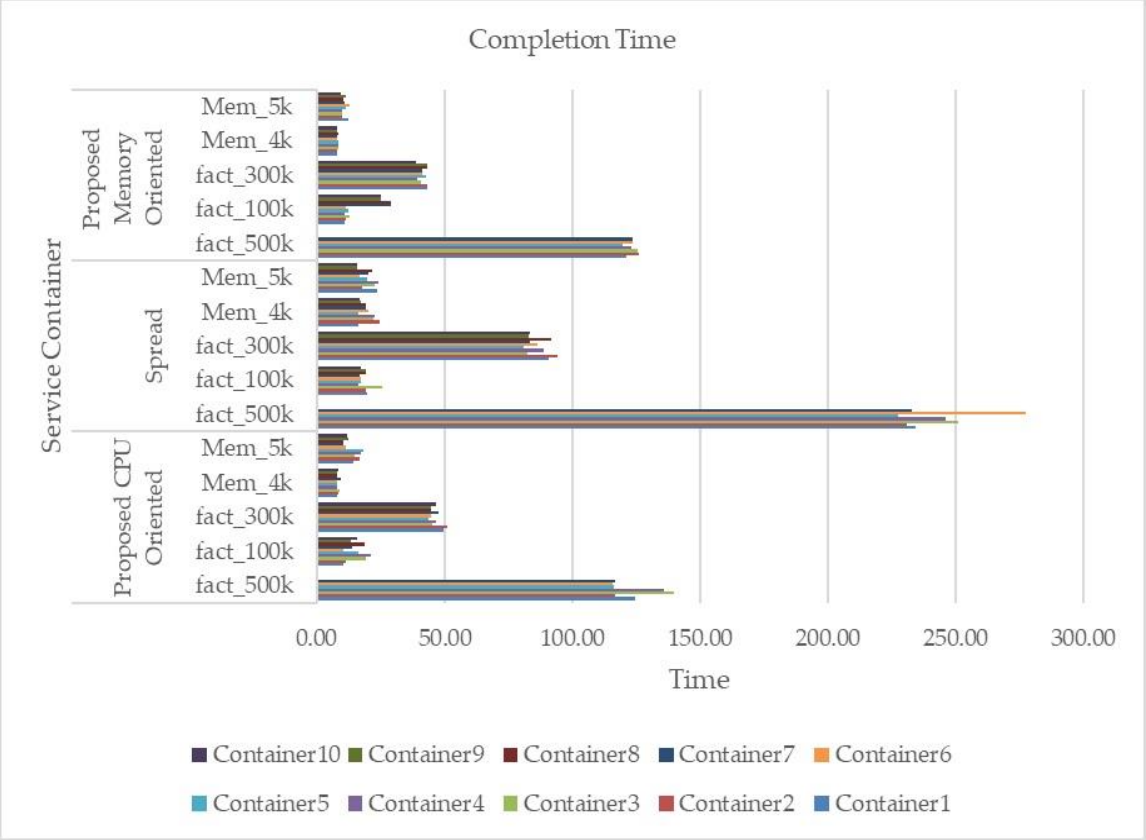


Figure 12 *Completion Time of task containers for Spread, Proposed (WCPU < WMEM, WCPU > WMEM)*

For visualizing the difference in completion time, the average completion time for the same services is calculated as displayed in Table 8 and plotted in Figure 13. It is shown that Proposed Weight resource Optimization approach is having almost same average completion time in both case (WCPU>WMEM and WCPU < WMEM), Whereas default Spread strategy is having higher average completion time for all services.

Table 8 *Average Completion Time of task containers for Spread, Proposed (WCPU < WMEM, WCPU > WMEM)*

Algo	fact_500k	fact_100k	fact_300k	Mem_4k	Mem_5k
Proposed W_{CPU} >W_{MEM}	123.6431	15.40614	46.52694	8.22835	13.75083
Proposed W_{CPU} <W_{MEM}	123.3108	17.8223	41.76793	8.184008	10.8016
Spread	243.0003	18.84889	86.47988	19.43128	19.80899

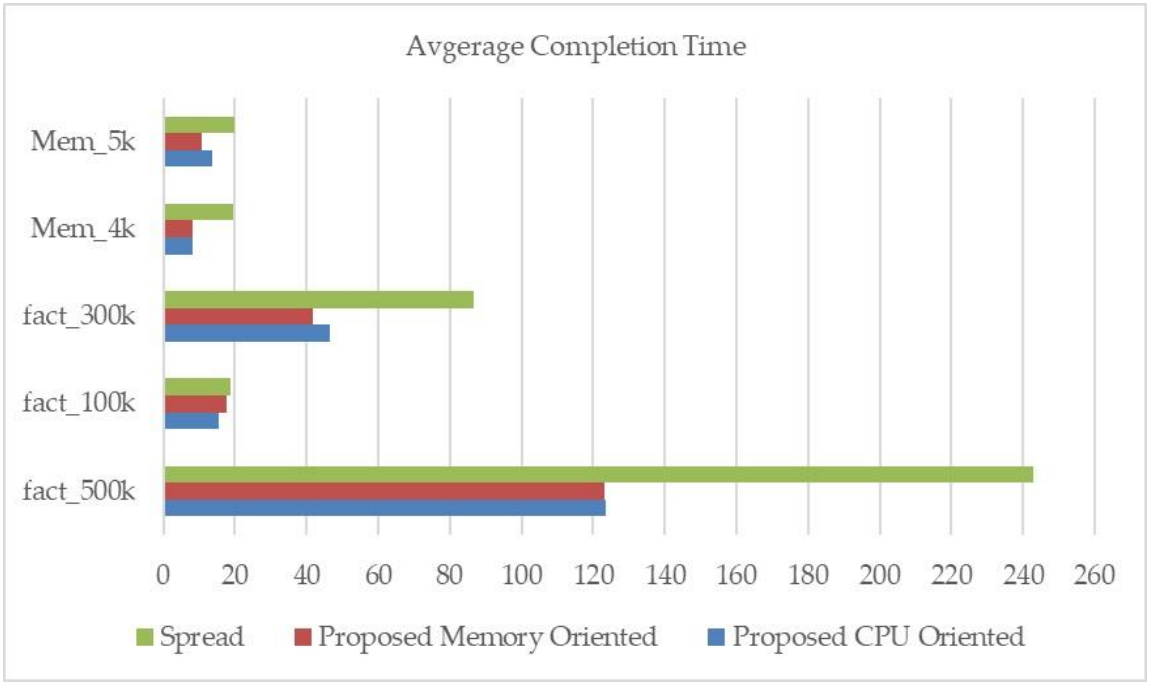


Figure 13 *Average Completion Time of task containers for Spread , Proposed (WCPU < WMEM , WCPU > WMEM)*

IV. Conclusion

This paper proposes a weighted resource optimization approach for finding a weighted score by using the available resources of nodes and the weight assigned to each resource. Resources currently considered are memory and CPU. Docker Swarm Cluster uses the default spread strategy for placing new containers on cluster nodes. This strategy tries to manage an equal number of containers on all nodes. but the spread strategy places the container without considering the resource requirements of task containers and the currently available resources of nodes. The proposed approach considers both conditions for calculating the score and places the container on the node with the highest score. The paper shows the improvement in load balancing among cluster nodes, and the completion time of task containers is also decreased. So overall performance is increased by the proposed approach.

References

[1] M. Moravcik and M. Kontsek (2020). Overview of Docker container orchestration tools. *18th* Slovenia,475-480.
[2] Potdar AM, Narayan DG, Kengond S, Mulla MM. Performance evaluation of docker container and virtual machine. Procedia Computer Science. 2020 Jan 1;171:1419-28.
[3] Swarmprom, 2021[Online]. Available https://github.com/stefanprodan/swarmprom
[4] Mao Y, Fu Y, Gu S, Vhaduri S, Cheng L, Liu Q. Resource management schemes for cloud-native platforms with computing containers of docker and kubernetes. arXiv preprint arXiv:2010.10350. 2020 Oct 20.
[5] Wu L, Xia H. Particle swarm optimization algorithm for container deployment. InJournal of Physics: Conference Series 2020 May 1 (Vol. 1544, No. 1, p. 012020). IOP Publishing.
[6] Wu G, Chen R, Zen D, Chen X. Using Ant Colony Algorithm on Scheduling Strategy Based on Docker Cloud Platform.

[7] Wu Y, Chen H. ABP scheduler: Speeding up service spread in docker swarm. In2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC) 2017 Dec 12 (pp. 691-698). IEEE.

[8] Hu Y, Zhou H, de Laat C, Zhao Z. Ecsched: Efficient container scheduling on heterogeneous clusters. InEuropean Conference on Parallel Processing 2018 Aug 1 (pp. 365-377). Cham: Springer International Publishing.

[9] Alzahrani EJ, Tari Z, Lee YC, Alsadie D, Zomaya AY. adCFS: Adaptive completely fair scheduling policy for containerised workflows systems. In2017 IEEE 16th International Symposium on Network Computing and Applications (NCA) 2017 Oct 30 (pp. 1-8). IEEE.

[10] Ghammam A, Ferreira T, Aljedaani W, Kessentini M, Husain A. Dynamic software containers workload balancing via many-objective search. IEEE Transactions on Services Computing. 2023 Feb 17;16(4):2575-91.

[11] Moravcik M, Kontsek M. Overview of Docker container orchestration tools. In2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA) 2020 Nov 12 (pp. 475-480). IEEE.

[12] Lyu T. *Evaluation of Containerized Simulation Software in Docker Swarm and Kubernetes* (Master's thesis).

[13] Pan Y, Chen I, Brasileiro F, Jayaputera G, Sinnott R. A performance comparison of cloud-based container orchestration tools. In2019 IEEE International Conference on Big Knowledge (ICBK) 2019 Nov 10 (pp. 191-198). IEEE.

[14] Raft consensus in swarm mode [Internet]. Docker Documentation. 2024 [cited 2024 Aug 21]. Available from: https://docs.docker.com/engine/swarm/raft/

[15] Swarm mode overview [Internet]. Docker Documentation. 2024 [cited 2024 Aug 21]. Available from: https://docs.docker.com/engine/swarm/

A NEW ROBUST LIU REGRESSION ESTIMATOR FOR HIGH-DIMENSIONAL DATA

Muthukrishnan. R, Karthika Ramakrishnan

•

Department of Statistics, Bharathiar University, Coimbatore, 641046, Tamil Nadu, India
muthukrishnan70@buc.edu.in, karthikaramakrishnan45@gmail.com

Abstract

Aim: *To provide a new Liu regression procedure for predictive modeling in cases of multicollinearity and with/without outliers.* ***Methods:*** *Regression analysis is employed in many statistical research domains for both estimation and prediction. Liu and Robust Estimators were developed in a classical linear regression model to address the issues of multicollinearity and outliers, respectively. In order to jointly handle the issues of multicollinearity and outliers, this research paper explores a new Robust Liu regression estimator based on the MM estimator, which is then demonstrated using real and simulated data sets. The performances of various regression estimators such as Least Square, Ridge, Liu and the Robust Liu are compared based on the Mean Square Error criterion.* ***Findings:*** *According to the computed error measure, the study concludes that the Robust Liu regression estimator provides more reliable results than the other mentioned regression procedures in situations where datasets have both multicollinearity and outliers.*

Keywords: Regression, Multicollinearity, Outliers, Liu, Robust Liu

I. Introduction

The Least Squares estimator is frequently utilized to predict the parameters of a regression model, provided that all the assumptions of the model are fully satisfied. Multicollinearity and outliers are the issues that could skew the outcomes of this approach. When there is a significant correlation between the independent variables, the situation is called multicollinearity, defined by Farrar and Glauber [7]. It will increase the error values and thus making the estimator not good. An outlier is a unique observation in the data. It causes the estimator to become inefficient and modifies the regression coefficients' sign. The existence of outliers, according to Chatterjee and Hadi [4], may leads to influence the parameter estimation and inaccurate predictions for traditional approaches. Ridge and Liu regression procedures were developed to overcome multicollinearity. When the data deviates from key assumptions, robust regression offers an alternative to the classical regression model. This research described a regression technique with a better estimate when multicollinearity and outliers are present in the dataset.

The rest of the paper is organized as follows. Various regression estimators like Least Squares, Ridge, Liu and Robust Liu are explained briefly in section 2. A numerical study is carried

out based on real and simulated datasets to compare the Mean Square Error of different regression estimators in section 3 and section 4 will give the conclusion.

II. Regression Procedures

Regression analysis serves as a means to glean insights from data by identifying relationships between the response and predictor variables, as outlined by Draper and Smith [6]. These methodologies within machine learning manifest in various forms, selected based on the nature of the dataset. It stands as the primary approach for addressing machine learning challenges through data modeling. This study encompasses Least Squares, Ridge, Liu and Robust Liu methods, comparing mean square error measures of different real and simulated datasets having the presence of both outliers and multicollinearity. Outliers in the actual data are detected and removed using Cook's distance, with the analysis conducted utilizing R software.

Least Squares Estimator

The Least Squares (LS) is a standard approach in regression analysis for estimating the parameters of a linear model. This method is employed to forecast the dependent variable (y) using several predictor variables (X). It stands as the widely utilized and optimal linear unbiased estimator, when all the suppositions of the classical regression model are satisfied. The standard model of LS with k independent variables is represented as follows.

$$y = X\beta + \varepsilon \tag{1}$$

here, y is an (m×1) vector of response variables, X is an (m×k) matrix of predictors, β is a (k×1) vector of unknown regression parameters, and ϵ is an (m×1) vector of residuals assumed to be independently and identically distributed as normal with a mean of zero and a fixed variance σ^2. The LS estimator for the unknown parameter is given by

$$\widehat{\beta_{LS}} = (X'X)^{-1}(X'y) \tag{2}$$

The performance of the LS estimator $\widehat{\beta_{LS}}$ becomes statistically insignificant when multicollinearity exists among the explanatory variables.

Ridge Estimator

Ridge regression estimator was provided by Hoerl and Kennard [9] to deal with the problem of multicollinearity. It gives a biased estimator and will depend on the ridge constant k which is used for minimizing the bias. The complexity parameter k needs to be selected appropriately in order to optimize the prediction accuracy. Hoerl et al. [10] find out a formula for the calculation of an optimal ridge constant k such that

$$k = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \widehat{\beta_i}^2} \tag{3}$$

where p is number of independent variables, $\hat{\sigma}^2$ is the estimated variance and $\widehat{\beta_i}$ is an LS regression parameter of canonical form. Ridge regression is depend on this constant k and will give a biased estimator as given below.

$$\widehat{\beta_{Ridge}} = (X'X + kI)^{-1}(X'y) \tag{4}$$

Liu Estimator

Liu Estimator is a class of biased estimators used to deal with datasets having multicollinearity. It was introduced by Liu [12]. These estimators are depending upon a biasing parameter d called the Liu parameter which lies between 0 and 1. The estimator of Liu regression is given by

$$\widehat{\beta_{Liu}} = (X'X + I_p)^{-1}(X'y + d\,\widehat{\beta_{LS}}) \tag{5}$$

where $0 \leq d \leq 1$, I_p is the identity matrix of order $p \times p$ and $\widehat{\beta_{LS}}$ is the LS estimator. The biasing parameter d of Liu is computed by the formula,

$$\hat{d} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^{p} \frac{1}{\lambda_i (\lambda_i + 1)}}{\sum_{i=1}^{p} \frac{\widehat{\beta_i}^2}{(\lambda_i + 1)^2}} \right] \tag{6}$$

where, $\hat{\sigma}^2$ and $\widehat{\beta_i}^2$ are the mean square error and the regression estimates computed via LS respectively. $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigen values of the matrix $X'X$. $\widehat{\beta_{Liu}}$ is named as the Liu estimator by Akdeniz and Kaciranlar [2]. The d value with minimum mean square error gives an efficient estimator as compared to other values . The R package liureg was developed by Muhammad Imdadullah et al. [15] provides the tools for the computation of the Liu estimator and the biasing parameter.

MM Estimator

The MM estimator is a robust regression technique introduced by Yohai [23], used to estimate parameters in the presence of outliers. It is a modification of the M-estimator, designed to provide robustness and high efficiency. The construction of this estimator starts with an initial robust estimator as S estimator obtained using a robust method such as minimizing the scale of residuals. The MM estimator is defined as follows.

$$\widehat{\beta_{MM}} = arg \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{s}\right) \tag{7}$$

where $r_i(\beta)$ are the residuals, s is a scale estimate based on the initial robust estimator, and ϱ is a loss function like Tukey's biweight.

Robust Liu Estimator

The objective of robust regression is to overcome some of the limitations of traditional regression analysis. The estimation and reference methods in robust regression should be straight forward to implement. Under a normal distribution without outliers, this robust method should yield results similar to those of LS. In this section, a new Robust Liu (RLiu) regression estimator was described to deal with the datasets having both multicollinearity and outliers by incorporating the properties of both Liu and MM regression procedures. The estimator of RLiu regression is given by

$$\widehat{\beta_{RLiu}} = \left(X'X + I_p\right)^{-1}(X'y + d_{MM}\widehat{\beta_{MM}}) \tag{8}$$

where $0 \leq d_{MM} \leq 1$, I_p is the identity matrix of order $p \times p$, $\widehat{\beta_{MM}}$ is the MM estimator. The biasing parameter $d_{MM}$of RLiu is computed by the formula,

$$\widehat{d_{MM}} = 1 - \widehat{\sigma_{MM}}^2 \left[\frac{\sum_{i=1}^{p} \frac{1}{\lambda_i (\lambda_i+1)}}{\sum_{i=1}^{p} \frac{\widehat{\beta_i}^2}{(\lambda_i+1)^2}} \right] \quad (9)$$

where, $\widehat{\sigma_{MM}}^2$ and $\widehat{\beta_i}^2$ are the mean square error and the regression estimates computed via MM respectively. $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigen values of the matrix $X'X$.

III. Experimental Results

Table 1: *Computed MSE under various regression methods (Real Data)*

	Methods			
Datasets	LS	Ridge	Liu	RLiu
Case 1:Prostate Cancer	0.46	0.46	0.16	0.14
	(1.5)	(0.33)	(0.13)	0.12
Case 2:Hald	3.68	3.32	1.19	0.14
	(2.57)	(2.46)	(1.02)	(0.89)

(.)Without outlier

Table 2: *Computed MSE under various regression methods (Simulation Data)*

		Methods			
n	Contamination	LS	Ridge	Liu	RLiu
50	0%	11.15	10.59	9.24	8.44
	5%	15.20	15.08	10.46	4.88
	10%	17.40	17.17	14.31	13.42
	15%	19.17	19.07	12.19	11.99
100	0%	12.31	12.13	4.53	4.41
	5%	12.78	12.68	5.96	5.42
	10%	12.62	12.44	4.99	4.82
	15%	18.88	18.42	7.06	5.73
200	0%	15.43	15.41	3.17	3.04
	5%	17.01	16.67	3.14	2.99
	10%	14.76	14.29	3.17	2.87
	15%	15.68	15.40	3.02	2.98

This section presents numerical analyses conducted on both real and simulated datasets. The first real dataset explained in case 1 exhibits moderate multicollinearity with outliers. The second dataset presented in case 2 displays high multicollinearity along with the presence of outliers. Outliers in the actual datasets were detected and eliminated using Cook's distance method introduced by Cook [5], and the analyses were performed using R software. A statistical technique called the Variance Inflation Factor (VIF) by Frisch [8] can detect and measure the amount of multicollinearity in a multiple regression model. The VIF assesses how much the regressors collectively impact the variance of each term within the model. The computed MSE measures of different regression estimators are summarized and presented in tables.

Case1. Prostate Cancer Dataset: The data come from a study that looked at how males undergoing radial prostatectomy correlated their level of prostate-specific antigen with several

clinical measures. This data set has 97 observations. There are seven independent variables namely lweight (log of prostate weight), age, lbph (log of benign prostatic hyperplasia amount), svi (seminal vesicle invasion), lcp (log of capsular penetration), gleason (Gleason score), lpsa (log of prostate specific antigen) and one dependent variable lcavol (log of cancer volume). Seven outliers are found in this dataset. Since the VIFs of the independent variables are in between 1 and 5, there is an indication of moderate multicollinearity.

Case 2. Hald Data: Woods et al [22] was introduced the Hald or Portland Cement Data. This data frame contains 13 observations with four independent variables. They are tricalcium aluminate (X1), tricalcium silicate (X2), tetracalcium aluminoferrite (X3) and β-dicalcium silicate (X4). The response variable Y is the evolved heat after 180 days in a cement mix. Since the VIFs of this Hald data set was greater than 10, the explanatory variables are highly correlated. As a result, the dataset has high multicollinearity. Also this data set has one outlier. The computed Mean Square Error (MSE) based on with and without outliers of different estimators in Cases 1 and 2 is given in Table 1.

Simulation studies were carried out to examine the efficiency of different regression estimators. In the study, the data was generated from a multivariate normal distribution with mean $\mu = [0]_{px1}$and the variance $\sum = [\sigma_{ij}]$ for the level of correlation, $\rho = 0.90$ and number of variables $p = 5$. Different levels of contamination (0%, 5%, 10%, and 15%) were studied for sample size $n = 50, 100, 200$. The performance of various regression procedures were compared using the MSE criterion and the results obtained for different number of observations with various levels of contamination are shown in Table 2.

The results obtained from Table 1 and Table 2 show that the error values for different estimators are slightly different from each other. Also the RLiu estimator has the smallest MSE of all others. Hence Robust Liu (RLiu) estimator is more efficient than the other estimators in the case of datasets having indication of multicollinearity and has outliers.

IV. Conclusion

Statistical learning techniques are crucial in various research fields, with regression analysis being a prominent method. Traditional linear regression often falls short when data deviates from its assumptions, necessitating alternative approaches. This paper explores several regression methods, including Least Squares, Ridge, Liu, and Robust Liu, and assesses their performance across different real and simulated datasets. The study addresses issues of multicollinearity and outliers by calculating Mean Square Error (MSE). The findings indicate that the Robust Liu regression method provides better estimates for datasets having both multicollinearity and/or outliers. This approach can be particularly advantageous for researchers employing machine learning techniques that need to account for these factors.

References

[1]Aitken, A. C. (1935). On least Squares and linear combinations of observations. *Proceedings of the Royal Statistical Society*, Edinburgh, 55: 42-48.

[2]Akdeniz, F. and Kacıranlar, S. (1995). on the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. *Communications in Statistics, Theory and Methods*, 24: 1789–1797.

[3]Arslan, O. and Billor, N. (2000). Robust Liu estimator for regression based on an M-estimator. *Journal of applied statistics*, 27: 39-47.

[4] Chatterjee, S. and Hadi, A. S. Sensitivity Analysis in Linear regression. John Wiley & Sons, New York, 2009.

[5]Cook, R. D. (2000). Detection of influential observation in linear regression. *Technometrics*, 42: 65-68.

[6] Draper, N. R. and Smith, H. Applied Regression Analysis. John Wiley & Sons, New York, 1998.

[7]Farrar, D. E. and Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, 49: 92-107.

[8]Frisch, R. Statistical confluence analysis by means of complete regression systems. University of Oslo, 1934.

[9]Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, 12: 55–67.

[10]Hoerl, A. E., Kennard, R. W. and Baldwin K. F. (1975). Ridge regression: Some simulation. *Communications in Statistics*, 4.

[11]Huber, P. H. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35: 7-101.

[12]Liu, K. J. (1993). A new class of biased estimate in linear regression. *Communications in Statistics*, 22: 393-402.

[13]Maronna, R. A. (2011). Robust ridge regression for high-dimensional data. *Technometrics*, 53: 44–53.

[14]Mendenhall, W. and Sincich, A. (2014). Second Course in Statistics: Regression Analysis, 7th ed (Harlow: Pearson), 105–123.

[15]Muhammad Imdadullah, Muhammad Aslam and Saima Altaf. (2017). liureg: A Comprehensive R Package for the Liu Estimation of Linear Regression Model with Collinear Regressors. *The R Journal 9*.

[16]Muthukrishnan, R. and Kalaivani, S. (2022). Data depth approach in fitting linear regression models, *Materials Today: Proceedings*, 57: 2212-2215.

[17]Muthukrishnan, R. and Karthika Ramakrishnan (2024). Effect of Classical and Robust Regression Estimators in the context of High dimensional data with Multicollinearity and Outliers. *Reliability: Theory & Applications*, 19: 335-341.

[18]Muthukrishnan, R. & Maryam Jamila, S. (2020). Predictive Modeling Using Support Vector Regression, *International Journal of Scientific and Technology Research*, 9: 4863-4865.

[19]Rousseeuw, P. J.(1983). Multivariate estimation with high breakdown point, The Fourth Pannonian Symposium on Mathematical Statistics and Probability, Bad Tatzmannsdorf, Austria.

[20]Rousseeuw, P. J and Leroy, A. M. Robust Regression and Outlier Detection , John Wiley & Sons, 1987.

[21]Susanti, Y., Pratiwi, H., Sri Sulistijowati, H. and Liana, T. (2014). M estimation, S estimation and MM estimation in Robust Regression. *International Journal of Pure and Applied Mathematics*, 91: 349-360.

[22]Woods, H., Steinour, H. H. and Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. *Industrial and Engineering Chemistry*, 24: 1207–1214.

[23]Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, 642-656.

DETECTION AND UTILIZATION OF THERMAL RESERVES IN OPERATION OF OBSOLETE POWER UNITS OF THERMAL POWER STATIONS

Farzaliyev Y.Z., Farhadzadeh E.M.

•

Azerbaijan State University of Economics
yusif.farzaliyev@unec.edu.az

Abstract

This article deals with economic aspects, i.e. identification of reserves of thermal efficiency of obsolete equipment in the example of power units of thermal power plants, which have a useful life exceeding 50%. As a result of operation of such equipment, useful heat required for power generation is lost. The developed new approach allows to detect in time those reserves, which are not possible with the use of energy characteristics due to wear and tear of the equipment and in the end these reserves will remain latent. With the help of the new approach when comparing it with the intuitive one, by which the technical staff wastes more time, it is shown that by taking into account the actual technical condition, reliability and efficiency of equipment operation it is possible to achieve the desired result. The results showed themselves brilliantly when distributing the load between power units of a thermal power station. The exploitation data for solving the problem are technical and economic indicators that characterize the wear and tear of the equipment

Keywords: reserves of thermal efficiency, obsolete equipment, new approach, actual technical condition, reliability, efficiency, load distribution between power units, technical and economic indicators

I. Introduction

Traditionally, thermal efficiency reserves are developed by increasing the reliability of "weak links" and improving the quality of control of power unit modes [1]. Technical economic indicators (TEI) analysis is carried out by comparing actual and nominal indicators. Nominal indicators are understood as indicators that reflect the actually achievable "economy of equipment operation under actual loads and external conditions, the condition and level of operation of the equipment, meeting the requirements of the current rules for the operation of power plants and networks".

Timely delivery of these calculations to operating personnel allows making decisions on control actions on equipment.

The high technical level of operational analysis of the reliability and efficiency of power units allows us to hope for achieving this goal [2]. However, unfortunately, the quality of maintenance and wear restoration of power unit installations does not meet the requirements [3].

The technical literature has repeatedly noted the increasing influence of the "human factor" [4]. And this influence is manifested in an increase in the share of fault of personnel, especially boiler-turbine shop personnel, in accidents and damage to power plant equipment, and a decrease in the

reliability and efficiency of their operation. The recommended approach to additional analysis of the TEI provides operational personnel with information about the "weak links", quality of management, maintainability and repair of the power unit and its installations relative to other similar power units of power plants. Ranking power units based on the reliability and efficiency of their operation allows us to take into account the operating experience of other power units, compare the quality of our work with the work of the operating personnel of other power units, and focus not only on some calculated TEI values, but also on the actual successes achieved by the operating personnel of other power units. This introduces into the process of organizing maintenance and repair elements of competition and material incentives for improving the TEI of the power unit [5].

The problem of load distribution between similar power units of thermal power plants is well known. Appropriate algorithms and calculation programs have been developed. Practical implementation requires, first of all, reliable energy characteristics, which in the conditions of increasing aging of the main and auxiliary equipment of the electrical energy supply in itself represents a serious problem. In this regard, heuristic approaches are often used, when, based on work experience, the workload of electronic equipment is assigned. Under these conditions, methodological support to the management of thermal power plants in the form of recommendations on the appropriate distribution of load between electrical units, depending on the reliability and efficiency of their operation, becomes important [6].

These recommendations can be obtained from estimates of integral indicators (B) of the reliability and efficiency of operation of energy blocks, calculated from the actual values of the technical and economic indicators of electric power plants [7]. Note that the desire to simultaneously increase the reliability and efficiency of operation in some cases is surprising, because additional costs are required to ensure operational reliability [8]. And that's true. However, in the formulation under consideration we are talking only about operating costs, which, with greater reliability, are naturally lower [9].

II. Methods

For the method of calculation of load distribution between power units - the initial data are [10]:

n_Σ - total number of single-type EBs

n_b - number of EBs in operation

$P_{min,per}$ - minimum permissible load of the EB

P_{rat} - rated power of EB

B - integral indicator of reliability and economic efficiency of EB operation

$P_{ave}=P_\Sigma/n_b$ - average load per EB, where P_Σ - TPS load

- The calculation of the load sharing between n_b of EBs is carried out in the following sequence:

$$b_i = \frac{B_i}{B_\Sigma} \quad (1)$$

where i=1,n_b; $B_\Sigma = \sum_{i=1}^{n_б^+} B_i^+ = \left|\sum_{j=1}^{n_б^-} B_j^-\right|$; B+ and B- are, respectively, positive (+) and negative (-) values of Bi, $n_б^+$ and - $n_б^-$ respectively the number of EBs with B+ and B- in operation;

- The minimum (b_{min}) and maximum (b_{max}) values of realizations of the integral indicator b_i are determined according to the formulas:

$$b_{min}= \min (b_1, b_2,\ldots\ldots b_{nб}) \quad (2)$$

$$b_{max}= \max (b_1, b_2,\ldots\ldots b_{nб}) \quad (3)$$

It's obvious that $b_{min}<0$, $b_{max}>04$;

- The intervals of possible decrease (ΔP^-) and increase ($\Delta P+$) of the average load of EB are determined by the formulas:

$$\Delta P^- = P_{cp} - P_{min,per} \quad (4)$$

$$\Delta P^+ = P_{rat} - P_{ave} \quad (5)$$

- If $\Delta P^- \leq \Delta P^+$,then the calculation of load distribution between n_b EBs, taking into account their reliability and economic efficiency, is carried out according to the formula:

$$P_i = P_{ave} + \Delta P^- \cdot b_i \quad (6)$$

- If, however, $\Delta P- > \Delta P+$, then by the formula:

$$P_i = P_{ave} + \Delta P^+ \cdot b_i \quad (7)$$

Main directions for increasing the efficiency of EBs
_____ month 20__ year

The results of the calculations allowed us to establish and recommend:

1.Technical and economic indicators of EBs not meeting the requirements for _____ month

№ EB	Technical and economic indicators			
	Identification	Relative dev.	Actual val.	Recommen. val.
2	Maximum electrical load	**-1,439**	220,00	270,00
	Share of Electricity for own needs	**-1,333**	5,50	4,10
	Average load	**-1,330**	160,00	179,29
	Installed capacity utilization factor	**-1,310**	11,10	44,84
	Specific consumption of fuel equivalent	**-1,186**	344,00	329,61
3	Specific consumption of fuel equivalent	**-1,072**	342,80	329,61
	Installed capacity utilization factor	**-0,996**	19,20	44,84
	Share of Electricity for own needs	**-0,857**	5,00	4,10
	Maximum electrical load	**-0,571**	250,00	270,00
6	Average load	**-0,296**	175,00	179,29
7	Average load	**-0,365**	174,00	179,29

2. The coefficients of significance of the technical condition (TC) of EBs are equal to

№ EB	1	2	3	4	5	6	7	8
The coefficients of sign. of TC	0,486	**-1,318**	**-0,634**	0,482		0,126	0,255	0,604

3. The group of "bad" includes 2 and 3 EBs. It is recommended to reduce their load in inverse proportion to the relative values of TC significance coefficients.

4. The least efficient of the operating EBs should be considered as 2 EBs. This EB is recommended to be shut down for scheduled repair, and preliminary-to be placed in reserve or to reduce the load as much as possible.

5. The group of "good" includes 8, 1, 4, 7, and 6 EBs. It is allowed to increase their performance in proportion to the relative values of TC significance coefficients.

6. The most efficient is 8 EB. It is advisable to operate it with the maximum permissible capacity.

Figure1: Fragment of the monthly result of the analysis of technical and economic indicators and recommendations on the main directions for improving the efficiency of EB operation.

1. The results of calculations of relative values of bi coefficients according to the formula (1) characterizing TC EB are given in **Table 1.**

Table 1: Realization of relative values of coefficients of significance of TC of EBs

N (i)	Serial number of power units							
	1	2	3	4	5	6	7	8
b_i	0.249	-0.675	-0.325	0.242	-	0.064	0.13	0.309

2. The results of calculations of the load distribution between EB TPS for a number of values of P_{ave} are given in **Table 2.**

Table 2: Results of calculations of load distribution between the TPP EBs for a number of P_{ave} values

Loads	Conditional numbers of EBs							
P_{ave}, MVt	1	2	3	4	5	6	7	8
110	115	96,5	103,5	114,9	-	111,3	112,6	116,2
130	140	103	117	139,9	-	132,6	135,2	142,4
150	164,9	109,5	130,5	164,8	-	153,9	157,8	168,6
170	189,9	116	144	189,7	-	175,2	180,4	194,8
190	214,9	122,5	157,5	214,7	-	196,4	203	220,9
210	232,4	149,2	180,8	232,2	-	215,8	221,7	237,9
230	247,4	182,7	207,3	247,3	-	234,5	239,1	251,7
250	262,4	216,2	233,8	262,3	-	253,2	256,5	265,5

Experience of calculations of load distribution between EBs shows that application of formulas (6) and (7), despite their error-free nature, does not sufficiently utilize the adjustment intervals of EBs (ΔP^+ and ΔP^-). A substantially greater effect is obtained if, instead of formulas (6) and (7), formulas (8) and (9) are used, which are of the form:

$$P_i = P_{ave} - \Delta P^- \frac{b_i}{b_{min}} = P_{ave} - (P_{ave} - P_{min,per}) \frac{b_i}{b_{min}} \quad (8)$$

$$P_i = P_{ave} + \Delta P^+ \frac{b_i}{b_{max}} = P_{ave} + (P_{rat} - P_{ave}) \frac{b_i}{b_{min}} \quad (9)$$

where i=1,n_b

Thus formula (8) is used if $\frac{\Delta P^-}{b_{min}} \leq \frac{\Delta P^+}{b_{max}}$, if, on the other hand $\frac{\Delta P^-}{b_{min}} > \frac{\Delta P^+}{b_{max}}$ then the formula is used (9). The results of calculations of load distribution between EBs of TPS according to formulas (8) and (9) are given in **Table 3.**

Table 3: Recommended power plant load distribution between EBs for a number of P_{ave} values.

Loads	Conditional numbers of EBs							
P_{ave}, MVt	1	2	3	4	5	6	7	8
110	117,4	90	100,4	117,3	-	111,9	113,9	119,2
130	144,8	90	110,7	144,6	-	133,8	137,7	148,3
150	172,1	90	121,1	171,9	-	155,7	161,6	177,5
170	199,5	90	131,5	199,2	-	177,6	185,5	206,7
190	226,9	90	141,9	226,5	-	199,5	209,3	235,8
210	254,3	90	152,2	253,9	-	221,5	233,2	265
230	281,6	90	162,6	281,2	-	243,4	257	294,2
250	290,2	140,9	197,5	289,9	-	260,4	271,1	300

Let's determine the interval of change in the EB load in the first and second methods of calculating load distribution. Suppose that $\Delta P^- < \Delta P^+$. When calculating by the first method:

- нижнее граничное значение нагрузки $(\underline{P})$ в соответствии с формулой (6) будет

равно:

$$\underline{P^{(1)}} = P_{ave} + \Delta P^{-} \cdot b_{min}$$

- верхнее граничное значение нагрузки $(\overline{P})$ в соответствии с формулой (7) будет равно:

$$\overline{P^{(1)}} = P_{ave} + \Delta P^{-} \cdot b_{max}$$

- the value of the load change interval is calculated by the formula:

$$\Delta_1 = \overline{P^{(1)}} - \underline{P^{(1)}} = \Delta P^{-}(b_{max} - b_{min} \quad (10)$$

When working in the second (2) method, the value of the load change interval (Δ_2) is calculated by the formula:

$$\Delta_2 = \overline{P^{(2)}} - \underline{P^{(2)}} = \Delta P^{-} \left\lceil \frac{b_{max} - b_{min}}{b_{min}} \right\rceil \quad (11)$$

Let's determine the degree of change of EB load interval from the ratio Δ_2 and Δ_1

$$\frac{\Delta_2}{\Delta_1} = b_{min}^{-1} \quad (12)$$

Thus, the load variation interval increases by a factor of $|b_{max}^{-1}| = \frac{1}{0.675} = 1.48$

If $\Delta P^{-} > \Delta P^{+}$, then similar calculations allow us to establish that,

$$\frac{\Delta_2}{\Delta_1} = b_{max}^{-1} = \frac{1}{0.309} = 3.23 \quad (13)$$

The significant excess of (Δ_2) over Δ_1 indicates the undoubted economic advantages of the second method.

III. Results

Evaluating the effectiveness of an intuitive approach to load distribution.

Analysis of the relationship between the average load of power units (P_{ave}) with the specific consumption of equivalent fuel (S_f) and with the integral indicator of the technical condition of power units (In), i.e. $P_{ave}=f1(S_f)$ and $P_{ave}=f2(In)$ allows us to judge the features of the existing approach to load distribution between power units of thermal power plants. We will evaluate this relationship by comparing the values of $P_{ave,i}$, ranked in descending order of significance, with S_f and with In [11].

The results of ranking P_{ave}, S_f and In_i power units by month of the year are given in **Table** 4.

Table 4: Information on the relationship of $P_{ave,i}$ with S_f and with In_i

Month of year	Indicators	Power unit number in the order of TEI deterioration							
1	P_{ave}	1	4	8	3	disabled	6	7	2
	In	8	4	1	7	disabled	6	3	2
	S_f	1	8	6	7	disabled	4	3	2
2	P_{ave}	1	7	8	4	disabled	2	6	3
	In	1	8	7	2	disabled	6	4	3
	S_f	1	8	7	4	disabled	2	6	3
3	P_{ave}	7	1	8	4	disabled	2	6	3
	In	1	7	7	6	disabled	2	3	4
	S_f	1	8	7	4	disabled	2	3	6
4	P_{ave}	7	disabled	4	1	disabled	8	3	6
	In	4	disabled	7	7	disabled	8	3	6
	S_f	1	disabled	1	8	disabled	4	3	6
5	P_{ave}	8	disabled	1	4	disabled	disabled	3	7
	In	8	disabled	3	1	disabled	disabled	4	7
	S_f	8	disabled	1	4	disabled	disabled	3	7
6	P_{ave}	1	7	3	disabled	2	disabled	8	5
	In	1	7	3	disabled	8	disabled	2	5
	S_f	7	1	3	disabled	8	disabled	2	5
7	P_{ave}	4	5	1	2	7	disabled	3	disabled
	In	7	4	2	1	3	disabled	5	disabled
	S_f	7	1	2	4	5	disabled	3	disabled
8	P_{ave}	1	5	2	6	3	7	4	disabled
	In	1	7	5	2	3	4	6	disabled
	S_f	7	3	2	1	4	5	6	disabled
9	P_{ave}	1	5	4	6	3	2	7	8
	In	4	2	7	1	3	6	5	8
	S_f	7	1	6	2	4	5	3	8
10	P_{ave}	4	8	7	3	5	disabled	2	1
	In	7	8	4	3	1	disabled	2	5
	S_f	7	3	2	1	4	disabled	6	5
11	P_{ave}	7	3	4	2	1	disabled	8	5
	In	8	7	1	2	8	disabled	4	5
	S_f	3	1	2	3	7	disabled	4	5

Analysis of the data in this table shows:

1. On an intuitive level, based on experience in operating power units:
   - The highest load P_{ave}^{max} at power units with the best technical condition is set only in 36.4% of cases;
   - The lowest load P_{ave}^{min} is set at power units with the worst technical condition of 77.8%;
   - In general, the coincidence of load distributions on the power unit with the integral characteristic of its technical condition is observed in 34.2%.
2. If the value of specific consumption of fuel equivalent is taken as the basis for load distribution, then:
   - The highest load of the power unit coincides with the lowest value of specific fuel consumption only in 27.3% of cases;
   - The lowest load of the power unit coincides with the highest value of specific fuel

consumption in 77.8% of cases;

- In the general case, the coincidence of load distributions between power units with the distribution of specific fuel consumption of a power unit is 30.1%.

Thus, operational data show that at load distribution between power units the degree of intuitive consideration of their reliability and efficiency and the degree of consideration of only the value of specific consumption of fuel equivalent is practically not different and amounts to about 30%. In this case, naturally, the question arises about interchangeability of the integral indicator and specific consumption of fuel equivalent arises [12]. According to the data of **Table** 4 it is easy to establish that coincidence of serial numbers of power units ranked by P_{ave}, In and S_f indicators is observed only in 16.4% of cases, of which in 58% of cases coincidence takes place at the worst power unit. And without taking into account these power units the coincidence takes place only in 6.9% of cases. These calculations show that both indicators (In and S_f) cannot be interchangeable and, first of all, because they are independent [13].

The above-mentioned indicates large reserves of thermal efficiency allowing to reduce the specific consumption of conditional fuel by eliminating the shortcomings of maintenance.

IV. Discussion

The results of power units ranking by average monthly values of TEI show:

- When distributing the load between power units, the degree of intuitive consideration of their reliability (technical condition) and efficiency (specific fuel equivalent consumption) and the degree of consideration of only the efficiency of operation are approximately the same and equal to 34.2%, in other words, the technical condition is practically not taken into account;
- The minimum load of the most "bad" power unit (power unit with the maximum value of the integral index) is observed in 77.8% of cases, and the maximum load of the most "good" power unit (power unit with the minimum value of the integral index) - in 36% of cases;
- The economic effect of load distribution using the recommended method (with seven or eight power units operating) is approximately 0.2÷0.5% of the total fuel costs.

References

[1] Farkhadzade, E.M., Muradaliev, A.Z., Rafieva, T.K. Development of methods for planning the restoration of the technical condition of power unit devices // - Moscow: Electric Stations. – 2007. No. 4, - p. 17-21.

[2] Dyakov, A.F., Isamukhamedov, Ya.Sh. Current state of the Russian electric power industry and factors reducing the reliability of power supply // Methodological issues in studying the reliability of large energy systems. Issue 63. Problems of reliability of energy systems in market conditions, - Baku: - 2013, - p. 7-15.

[3] STO-34.01-35-001-2020. Guidelines for the technical examination of substation equipment, power lines. "Rosseti", 2020, 20p.

[4] Akimova, G.P. Methodological approach to determining the influence of the human factor on the performance of information systems / G.P. Akimova, A.V. Soloviev, E.V. Pashkina. - Moscow: Proceedings of ISA RAS, - v.27. -2007. - With. 102-117.

[5] Farkhadzade E.M., Farzaliev Yu.Z., Muradaliev A.Z. Method and algorithm for ranking boiler installations of block power plants according to the criterion of reliability and efficiency of operation // - Moscow: Teploenergetika, - 2015. No. 10, - p. 22-29.

[6] Farkhadzade E.M., Muradaliev A.Z., Farzaliev Yu.Z. Ranking of power units of power

plants according to the reliability and efficiency of their operation // - Baku: Problems of Energy, - 2014. No. 2, - p. 8-16.

[7] Potrebich, A.A. Integrated system for solving technological problems of IASU PES // - Electric Power Plants, - Moscow: - 2001. No. 8, - p. 28-31.

[8] Makoklyuev, B.I., Polizharov, A.S. Information systems for solving technological problems at power facilities // - Moscow: Energetik, - 2007. No. 8, - p. 35-38.

[9] Yagovkin, G.N., Yagovkin, N.G., Panyukova, S.A. Reliability management in the implementation of the maintenance and repair strategy in power supply systems // - Moscow: Industrial Energy, - 2010. No. 9, - p. 12-15.

[10] Farhadzadeh, E.M., Muradaliyev, A.Z., Farzaliyev, Y.Z. To the question on distribution of loading between power units TPS // Journal of «Reliability: Theory&applications», – USA: RT&A (Vol.10 No.2(37), - 2015. June; - pp. 42-49.

[11] Farkhadzadeh, E.M. Automated analysis of a fleet of switches / E.M. Farkhadzade, A.Z. Muradaliev, T.K. Rafieva, S.A. Abdullaeva // Energetik, - Moscow: - 2014. No. 5, p. 11-15.

[12] Duel, M.A. Automation of determination of energy characteristics of power equipment / M.A. Duel, G.I. Kanyuk, T.N. Fursova, E.N. Bliznichenko // energy saving, energy, energy audit - Kyiv: - 2013. No. 2(108), - pp. 13-19.

[13] RD 34.09.454 Standard algorithm for calculating the technical and economic indicators of condensing power units with a capacity of 300, 500, 800 and 1200 MW // - Moscow: part 1, VTI Soyuz Techenergo. Date of update - 2016. - p. 37.

A GRAPHICAL STUDY ON THE MISSING DATA OF CENTRAL COMPOSITE DESIGN WITHIN A SPHERICAL REGION

A.R. Gokul[1] and M. Pachamuthu[2*]

•

[1]Research Scholar, Department of Statistics, Periyar University, Salem-636011, India
gokulmalar6@gmail.com
[2*]Assistant Professor, Department of Statistics, Periyar University, Salem-636011, India
pachamuthu@periyaruniversity.ac.in

Abstract

Robust missing observations have emerged as a crucial study area in statistical research. Response Surface Methodology (RSM), a recognized and extensively utilized area in experimental design, has determined that the absence of observations in an experiment can introduce complexity and interfere with the estimation of parameters. Previous literature reviews reveal that most studies on missing Central Composite Design (CCD) data were conducted using optimality and minimax loss criteria. Our study explores the spherical region of interest in the missing observation of CCD, represented through Variance Dispersion Graph (VDG) and Fraction of Design Space (FDS) graphs. Practitioners primarily focus on the region of interest rather than employing various alpha values. We investigate the predictive capabilities of each factorial, axial, and center missing design point against different radii(r) and fractions of the design space region, and we also measure relative G- and D- efficiency. We scrutinize various factors (k) from two to seven, including five center runs. Our research explores the region of interest in operating the experiment under robust conditions through visual aids of VDG and FDS graphs.

Keywords: Central Composite Design, Fraction of Design Space, Scaled Prediction Variance, Optimality, Variance Dispersion Graph

I. Introduction

Response Surface Methodology is a potent combination of statistical and mathematical techniques used for model building. It's specifically designed to evaluate the effects of several independent variables and determine their optimal values to achieve the most favorable results. This methodology is particularly beneficial when the goal is to optimize a product or process. It allows for a comprehensive understanding of the relationships between different variables and the response, facilitating efficient and effective optimization. The empirical model is constructed using data gathered directly from the system or process under study. *RSM* involves building empirical models using multiple linear regression and statistical techniques [12] In the literature review, various authors have conducted studies on the missing data of second-order composite designs. The details of these studies will be briefly discussed sequentially. Draper [8] examined the studies on sturdy methods for handling missing observations in response surface design and acknowledged the pioneer who formulated the parameter estimation equation. Indeed, Akhtar et

al. [2] introduced a minimax loss criterion to assess missing observations. This approach has become the most used method in response surface designs. It's a significant contribution to the field as it provides a robust and effective way to handle missing data, thereby improving the accuracy and reliability of the designs. Alrweili et al. [4] discussed the minimax loss criterion to develop more resilient models for missing data. They achieve this by integrating the most recent *CCD*s from *GSA* and *AEK*, both novel designs. Hayat et al. [10] delved into designs derived from regular and irregular structure subsets. They evaluate how these designs handle missing points using the minimax loss criterion. Furthermore, they scrutinize their alphabetic optimality and predictive capabilities through FDS plots depicting the variance in response difference. Alanazi et al. [3] introduced closed-form expressions that account for two missing observations. These expressions are based on α, the axial value utilized in *CCDs* that handle up to 10 factors. Hemavathi et al.[11] examined the ability of sequential third-order rotatable design to manage missing data without significant information loss. Additionally, the study quantifies the loss of information from one or two absent experimental runs at varying distances from the design's center. Park et al. [14] compared *CCD*, *SCD*, and *MinResV* designs. They focus on spherical regions with k = 3 to 7 factors, using the optimality criteria and the variance dispersion graph as benchmarks. Interestingly, their findings reveal that none of these designs consistently outperforms others. Li et al. [13] evaluated various *CCD*, *SCD*, and *MinResV* designs. These designs are applied to spherical and cuboidal regions with different axial values. To analyze the prediction variance properties of these designs, they use *FDS* plots and box plots. Onwuamaeze [15] employed graphical techniques like *VDG* and *FDS* plots. They use these methods to assess the prediction variance performance of *CCD*, *SCD*, and *MinResV* designs within the hypercube region. Ahmad et al. [1] compared Augmented Pairs (*AP*) designs and Subset designs based on standard optimality criteria and graphical criteria in spherical and cuboidal regions of experimentation, which provide more insight into the prediction performance of the designs. The article [17] employed *VDG* and *FDS* plots to depict the scaled prediction variance attributes of the second-order design and *G-* and *I-* optimality designs within a cuboidal area of interest. The study by *G. G.* Vining et al. [18]) involved a graphical method that plots the maximum and average mean squared prediction error across spheres of different radii within the design space.

This study thoroughly investigates the impact of missing data on the central composite design within a spherical region of interest. Utilizing *VDG* and *FDS* determines the robustness of the design in various regions based on missing factorial, axial, and center design points. The paper is segmented into various parts. The methodology is detailed in Section 2. Section 3 presents the findings and discussions, including the interpretation of *VDG* and *FDS* plots using spherical and rotatable alpha values and Section 4 concludes the study.

II. Methodology

I. Description of Second-Order Central Composite Design

In many cases where we apply RSM, the relationship between the predictor variables and the response might be unclear. A first-order model, while applicable in some cases, might not be able to accurately capture the curvature of the response function due to its linear nature. This is where higher-degree polynomial models, such as second-order models, come into play. These models can capture more complex relationships and better evaluate curvature in optimization experiments. They provide a more nuanced understanding of the data, allowing for more accurate predictions and more effective optimization.

For k quantitative factors denoted by $x_1, x_2, \ldots, x_k$, a second-order model is

$$y = \beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1}^{k}\beta_{ii} x_i^2 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\beta_{ij} x_i x_j + \varepsilon \quad (1)$$

In this context, $\beta_0, \beta_i, \beta_{ii}$, and β_{ij}represent the intercept, linear, quadratic, and bilinear coefficients, respectively. ϵ_i is a random error term with a mean of zero, a variance of σ^2, and is independent for each pair of runs. The total number of parameters that need to be estimated, denoted as p, is calculated as$p = k + k + \binom{k}{2} + 1$. To ensure enough degrees of freedom to estimate these model coefficients, the number of runs, represented as n, must equal or exceed p.

CCD is extensively utilized to estimate second-order response surfaces. Since its introduction by [7], the *CCD* has been the subject of numerous studies and has seen widespread use in various fields. The flexibility in utilizing the *CCD* lies in choosing alpha (α) , which represents the axial distance and n_c, denoting the number of centre runs. The selection of these two parameters can be crucial. The operational region and the area of interest largely influence the determination of α. Here, we use spherical and rotatable α values, where the spherical value of α is set to the square root of the number of factors, represented as $\alpha_S = k^{1/2}$. Indeed, in αs, all the design points are situated on a common geometric sphere, and it is nearly rotatable, and Box et al. [6] suggested a value, $\alpha_R = F^{1/4}$, called rotatable α, where F is the number of factorial runs in a design. Indeed, a design is rotatable if the variance remains the same for all points that are an equal distance from the design's center.

II. Scaled Prediction Variance and Relative G- and D- Efficiency

Borkoski [5] have developed an analytical form for calculating scaled prediction variance values of *CCD* and Box-Behnken design (*BBD*). Spherical Prediction Variance (*SPV*) enables accurate prediction of response variables at different points of interest within the experimental area. The prediction variance at a point *x* is given by

$$v(x) = \frac{n \cdot var[\hat{y}_{(x)}]}{\sigma^2} = n \cdot X^{(m)'}(X'X)^{-1}X^{(m)} \quad (2)$$

The vector $X^{(h)}$ represents the array of coordinates of a point in the design space that has been magnified to align with the model form, where n is the quantity of experimental runs design, and σ^2 is the observation error.

G- optimality is a standard that seeks to reduce the highest possible variance in any forecasted value across the entire experimental domain. This efficiency can be understood as the proportion between the determinant of the information matrix for a specific design and the determinant of the information matrix for the best optimal design.

$$G_{\text{eff}} = \frac{p}{n \cdot MAX_{X \in R} v(x)} \quad (3)$$

In this context, p represents the estimated model's parameters, and n denotes the number of observations in the corresponding design. The term $MAX_{X \in R} v_{(x)}$signifies the maximum value of the variance of the predicted response. As a result, the relative *G*-efficiency, denoted as RE_G, is determined by the ratio of the $G_{eff}$for the reduced design to the G_{eff} for the complete design.

$$RE_G = \frac{G_{\text{eff (reduced)}}}{G_{\text{eff}}} = \frac{n \cdot MAX_{X \in R} v(x)}{n_r \cdot MAX_{X \in R} v(x)_{\text{reduced}}} \quad (4)$$

In this scenario, n represents the size of the complete design, while n_r denotes the size of the reduced design. Based on the definition of RE_G, a design that yields a higher value of RE_G would be more desirable. This is because a higher RE_G value indicates a more efficient design relative to the complete design. By utilizing equations (3) and (4), we are able to compute the relative G-efficiency value. These values are then presented in tables 1 and 2.

D efficiency is defined as maximizing the determinant of the information matrix minimizing the determinant of the inverse of the information matrix. Thus, relative D- efficiency is given as

$$RE_D = \left(\frac{|X'X|_{\text{reduced}}}{|X'X|} \right)^{\frac{1}{p}} \quad (5)$$

Where, p is the number of parameters of the model to be estimated, $|X'X|_{\text{reduced}}$ is the determinant of the information matrix of reduced design and $|X'X|$ is the determinant of the complete design matrix. A value approaching one will represent a minor loss, whereas a value below one will represent a more significant loss in model estimation. Through the application of equation (5), we are able to determine the relative D-efficiency value. These computed values are then listed in tables 1 and 2.

III. Variance Dispersion Graph and Fraction of Design Space

Giovannetti-Jensen and Myers [9] presented the concept of variance dispersion graphs to evaluate the comprehensive predictive capability of a *RSM* within a region of interest. A variance dispersion graph allows one to visualize the uniformity of the scaled variance of a predicted value in multidimensional space. It consists of three curves: the maximum, the minimum and the average scaled variance of a predicted value on a hypersphere. However, these *VDG* plots do not consider the fraction of the total design space between concentric spheres of radius *r* at different distances from the center of the design space. Variance Dispersion Graph handle the *SPV* on a sphere with radius *r*, but they overlook the volume related to this information. To gain insight into the complete picture of the prediction performance of a design, one should consider the volume. An *FDS* plot [16] is generated by taking a substantial number of samples, represented as n, from the entire design area and calculating the associated *SPV* values. The underlying concept is that the design quality improves if a greater portion of the design space is near the minimum *SPV* value.

Furthermore, a flatter line indicates a more stable design. The *FDS* plot effectively encapsulates the range and distribution of *SPV* values in the design space, facilitating the comparison of designs through a single curve. Furthermore, it [19] equips the investigator with a unique graph for contrasting designs or examining the characteristics of a particular design.

III. Result and Discussion

I. Relative *D*- and *G*- efficiency of *CCD k* = 2 to 7 using spherical and rotatable α

When a data point is missing in a Central Composite Design (CCD) with factor k = 5 and a factorial (1,1,-1,1,-1) run, the relative G efficiency is poorest. G-efficiency measures the precision of parameter estimates in a design. Specifically, relative G efficiency compares the efficiency of a design with missing observations to the efficiency of the same design without any missing data. When a data point is missing, the relative G efficiency decreases, indicating reduced precision in estimating model parameters.

Additionally, relative D efficiency approaches 1 in this scenario, indicating maximum determinant value. D-efficiency is related to the determinant of the information matrix (also known as the Fisher information matrix). Relative D efficiency compares the determinant of the information matrix for a design with missing observations to the determinant for the complete design. A value closer to 1 indicates better efficiency in terms of information content.

Axial missing points outperform factorial points in terms of both relative G and D efficiency. Interestingly, missing a center run has minimal impact compared to the no-missing scenario, except for relative G efficiency when k = 2. Overall, relative G and D efficiency provide valuable insights for comparing missing and complete designs

Table 1: *Relative D and G efficiency of CCD of factors k = 2 to 7 using spherical α*

Factors	Number of runs	Types of Missing runs	Alpha value (α_S)	Relative G efficiency (RE_G)	Relative D efficiency (RE_D)
K=2	13	None	1.41421	1.0000	1.0000
	12	Factorial (-1,1)		0.4062	0.8492
		Axial (0,1.41421)		0.4062	0.8492
		Centre (0,0)		0.3008	0.9635
K=3	19	None	1.73205	1.0000	1.0000
	18	Factorial (-1, -1,1)		0.3581	0.8975
		Axial (0,1.73205,0)		0.4292	0.9080
		Centre (0,0,0,)		1.0555	0.9779
K = 4	29	None	2	1.0000	1.0000
	28	Factorial (-1,1,1,1)		0.4316	0.9433
		Axial (0,-2,0,0)		0.4316	0.9433
		Centre (0,0,0,0)		1.0358	0.9852
K = 5	31	None	2.23607	1.0000	1.0000
	30	Factorial (1,1,-1,1,-1)		0.1491	0.9120
		Axial (0,0,0,2.23607,0)		0.5177	0.9537
		Centre (0,0,0,0,0)		1.0334	0.9894
K = 6	49	None	2.4495	1.0000	1.0000
	48	Factorial (-1,-1,-1,1,1,-1)		0.3800	0.9653
		Axial (0,0,0,2.44949,0,0)		0.4723	0.9699
		Centre (0,0,0,0,0,0)		1.0209	0.9921
K = 7	83	None	2.64575	1.0000	1.0000
	82	Factorial (1,-1,-1,-1,1,1,-1)		0.7106	0.9845
		Axial (0,0,0,0,2.64575,0,0)		0.4747	0.9792
		Centre (0,0,0,0,0,0,0)		1.0122	0.9938

Table 2 : *Relative D and G efficiency of CCD of factors k =3,4,5 and 7 using rotatable* α

Factors	Number of runs	Types of Missing runs	Alpha value (α_S)	Relative G efficiency (RE_G)	Relative D efficiency (RE_D)
K = 3	19	None	1.68179	1.0000	1.0000
	18	Factorial (-1,-1,1)		0.3486	0.8951
		Axial (0,1.68179,0)		0.4178	0.9108
		Centre (0,0,0,)		1.0555	0.9780
K = 5	31	None	2	1.0000	1.0000
	30	Factorial (1,1,-1,1,-1)		0.1251	0.9044
		Axial (0,0,0,2,0)		0.4722	0.9574
		Centre (0,0,0,0,0)		1.0331	0.9901
K = 6	49	None	2.37841	1.0000	1.0000
	48	Factorial (-1,-1,-1,1,1,-1)		0.3770	0.9651
		Axial (0,0,0,2.37841,0,0)		0.4888	0.9704
		Centre (0,0,0,0,0,0)		1.0209	0.9921
K = 7	83	None	2.82843	1.0000	1.0000
	82	Factorial (1,-1,-1,-1,1,1,-1)		0.5790	0.9457
		Axial(0,0,0,0, 2.82843,0,0)		0.5340	0.9239
		Centre (0,0,0,0,0,0,0)		1.0121	0.9790

II. Interpretation of VDG of factors k = 2 to 7 of spherical $\sqrt{k}$ alpha value

Figure 1 depicts all the factors of *SPV* distribution against radius; when k = 2, the trajectory of the maximum, average, and minimum *SPV* for both factorial and axial design points with missing observations is identical. This trajectory remains consistent for each design point, beginning at a radius of 0 and extending to $\sqrt{2}$, resulting in a *maxSPV* value of 24.13. In this case, both the original and missing center design points exhibit the same maximum, minimum, and average prediction variance across all radii.

When k = 3, excluding *Fmax* and *Amax*, all other design points for maximum, average, and minimum prediction variance display a certain level of consistency in their *SPV* values below 5 up to a radius $r \leq 1.20$. Past a radius of about 1.20, the curve steadily rises until the radius hits $\sqrt{3}$. *Fmax* can achieve an *SPV* as elevated as 39.40, while *Amax* can attain an *SPV* as high as 33.40. For factor k = 4, a center run is missing at the radius origin, exhibiting an *SPV* value of 7.03. As the radius extends to r = 1, this *SPV* value decreases to 5.87. Beyond r = 1, the *SPV* value escalates with the radius, reaching a *maxSPV* of 16.26. A factorial and an axial design point are absent, with a maxSPV of 37.98 at a radius $\sqrt{4}$. The pattern followed by factor k = 4 resembles factor k = 2.

For factor k = 5, the lack of a factorial observation leads to a significant *SPV* value of 190.2 at the radius $\sqrt{5}$. This suggests that insufficient data under these conditions leads to subpar predictive performance in the experiment. The axial design point's maximum prediction variance is 52.39 at the radius $\sqrt{5}$. The *None-max, None-avg, Cmax, Cavg*, and *Aavg* of all these prediction variance design points are almost identical for all radii with an *SPV* of ~ 26.15 or less, except for *Favg SPV*, which slightly deviates from another curve from a radius of 1.30. Despite misinformation, *minSPVs* follow the same path for all design points.

When k = 6, the absence of a factorial and an axial observation result in the maximum *SPV* staying below 10 from a radius of $0 \leq r \leq 1.20$. However, as the radius increases beyond 1.20, the *SPV* value rises, reaching a maximum *SPV* of 73.90 for factorial and 63.20 for axial at radius $\sqrt{6}$.

Specifically, we can observe that the *SPV* design points *Favg*, *Fmin*, *Aavg*, and *Amin* follow the same trajectory from a radius of $0 \leq r \leq 1.40$, with an *SPV* of less than 8.65. Beyond a radius of 1.40, the *minSPVs* diverge from the path, reaching a *minSPV* of 29.40. A similar pattern is observed for *None-max*, *avg*, *min* and *Cmax*, *avg*, *min*, but they start from different *SPV* values at radius 0 and travel closely up to a radius of approximately 1.40. Beyond a radius of 1.40, the *minSPVs* diverge up to a radius of $\sqrt{6}$.

For factor $k = 7$, the lack of an axial design point leads to a predictive performance that is inferior to the factorial. The axial design point has an *SPV* of 71.85, while the factorial design point has an *SPV* of 53.75 at the radius $\sqrt{7}$. Excluding the center design point *SPV*, all other points start with an *SPV* of 16.20 at a radius of 0 and maintain the same path up to a radius slightly less than 1.0. Beyond a radius of 1.0, the prediction variances of each design point diverge from each other. It can be observed that the average and minimum *SPV* of various design points falls within an *SPV* interval of 35.31 to 33.50 at radius $\sqrt{7}$.

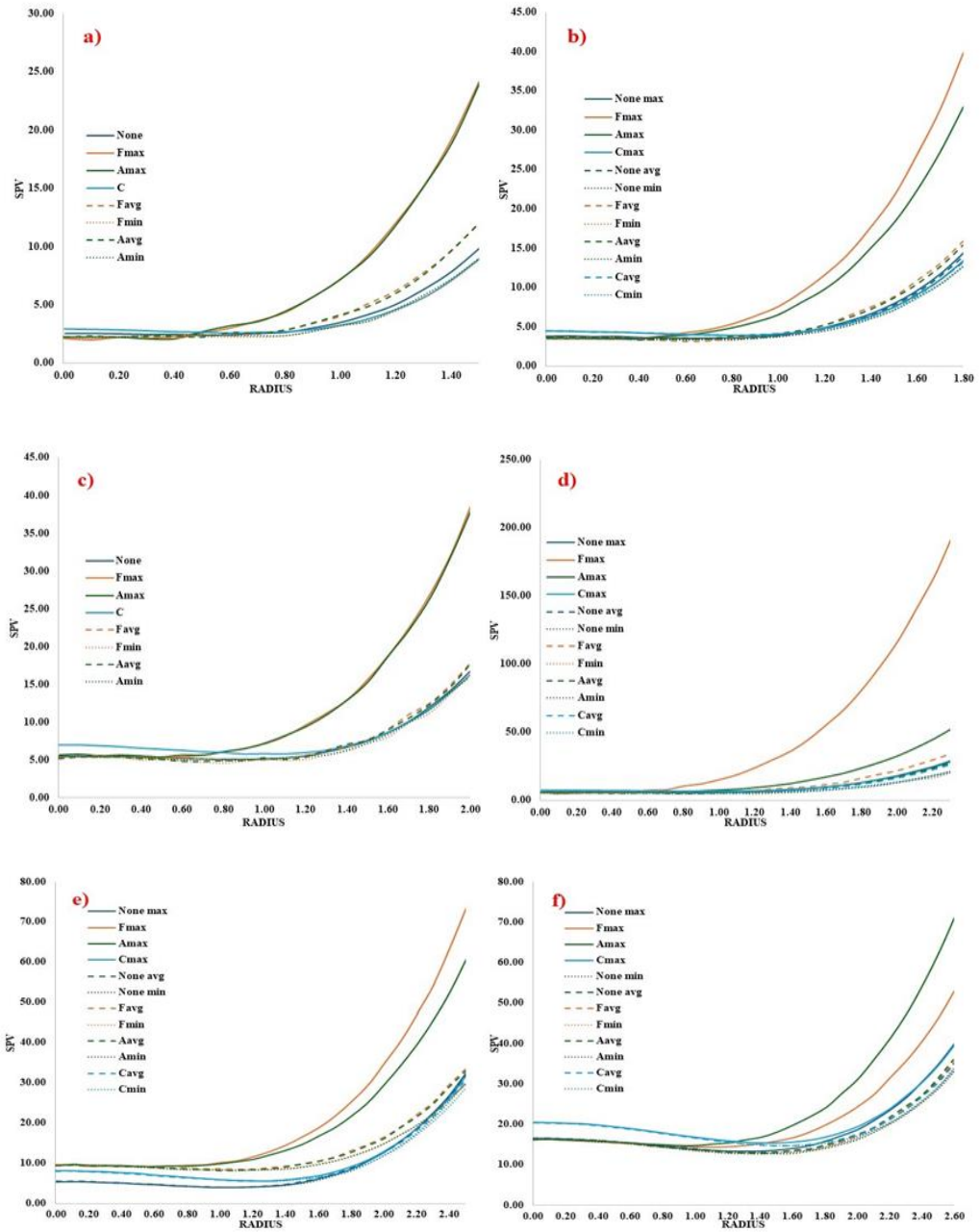


Figure 1: *(a) VDG for CCD K = 2. (b) VDG for CCD K = 3. (c) VDG for CCD K = 4. (d) VDG for CCD K = 5. (e) VDG for CCD K = 6. (F) VDG for CCD K = 7*

III. Interpretation of VDG of factors K = 3, 5, 6 and 7 of rotatable alpha value

Figure 2 depicts all the factors of *SPV* distribution against radius; For the factor k = 3, the absence of a center run results in the lowest *maxSPV* of 11.18. Excluding *Fmax* and *Amax*, which have SPVs of 34.11 and 28.76, respectively, all other average and minimum SPVs and design points with no missing data maintain an *SPV* value of less than 3.63 up to a radius of less than 1.00. Beyond a radius of 1.00, these design points begin to diverge slightly from each other.

For factor k = 5, when a factorial observation is absent up to a radius of 0.7, the SPV stays below 8.43. However, once the radius surpasses 0.7, the SPV undergoes a substantial increase, reaching a peak of 199.90 at $\sqrt{5}$. The lack of an axial observation results in a *maxSPV* of 8.41 at a radius of 1.00, and as the radius grows, the *SPV* steadily increases, attaining a *maxSPV* of 55.23 at $\sqrt{5}$. Apart from the two scenarios mentioned earlier, all other situations, such as no missing design points, a missing Centre run, average, and *minSPVs*, lie within the range of 33.71 to 24.81 at a radius of 2.23.

For factor k = 6, the lack of a factorial and an axial design point in an experimental setup results in nearly identical *SPV* values up to a radius of less than 1.20. Beyond this point, the *SPV* value slowly diverges, reaching 74.79 for factorial and 65.00 for axial at a radius of $\sqrt{6}$. For the factor k = 7, the absence of an axial observation leads to a lower prediction performance than when a factorial observation is missing. However, for radii less than 1.60, both design points maintain a nearly identical SPV value of 17.27. Factors k = 6 and 7 display the same conditions for no missing observation, average, and *minSPVs*, following the same path across all radii. The absence of a center run does not exhibit a consistent increasing trend across all radii; instead, it varies.

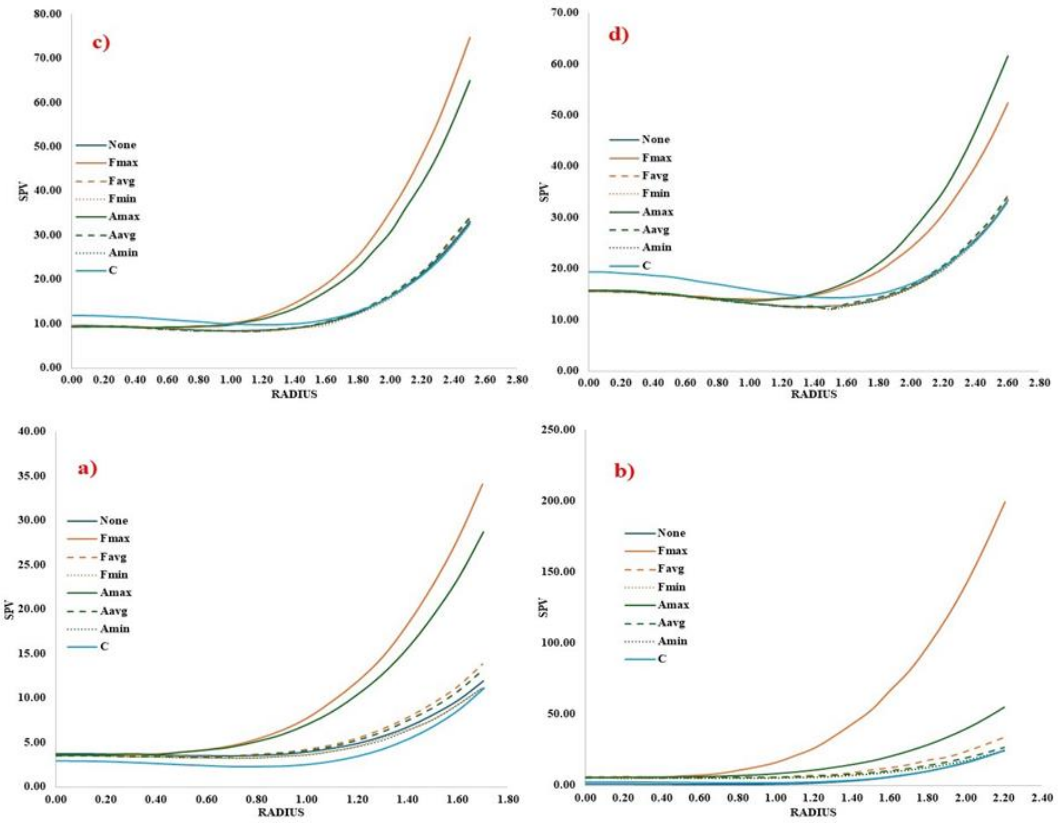


Figure 2: *(a) VDG for CCD K = 3. (b) VDG for CCD K = 5. (c) VDG for CCD K = 6. (d) VDG for CCD K = 7.*

IV. Interpretation of FDS of factors k = 2 to 7 of spherical $\sqrt{k}$ alpha value:

Figure 3 of *FDS* plot depicts that the impact of missing and non-missing observations is not significantly different across most factors. For factor k = 2, a missing center run must be distinct from all other factorial, axial, and non-missing design points. In contrast, the various observations for all other factors follow similar trajectories, staying close to each other within 75% to 85% of the design space region. When considering missing factorial and axial observations, it's notable that only for factor k = 7 does the axial observation have the highest *SPV*. The factorial observation has the highest *SPV* for all other factors, with factor k = 5 showing an exceptional *SPV* value of 191.86. A *G*-efficiency of 100% is achieved by all factors at various design points when the *FDS* region is approximately 80%. Beyond 85% of the *FDS* region, all factors experience a significant increase in the maximum *SPV*.

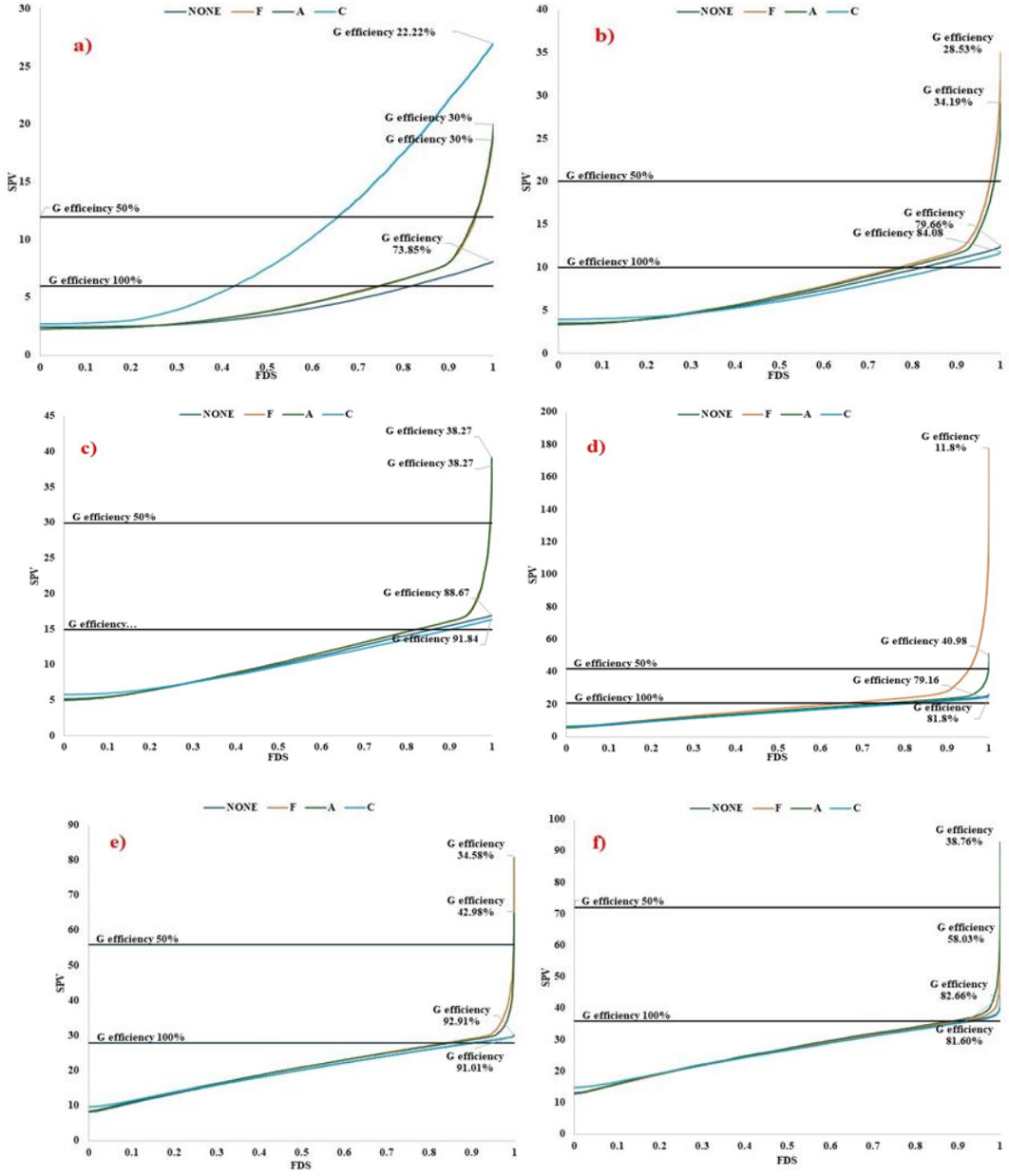


Figure 3: *(a) FDS for CCD K = 2. (b) FDS for CCD K = 3. (c) FDS for CCD K = 4. (d) FDS for CCD K = 5. (e) FDS for CCD K = 6. (F) FDS for CCD K = 7*

V. Interpretation of FDS of factors K = 3, 5, 6 and 7 of rotatable alpha value

As the factor increases (such as 5, 6, and 7), the coverage of the *FDS* region over 100% of the *G*-efficiency also increases. Specifically, for factor k = 5, both missing and non-missing observations reach the 100% *G*-efficiency line at 55% of the *FDS* region. For factor k = 6, they reach this line at 85% of the *FDS* region; for factor k = 7, they reach it at 95% of the *FDS* region

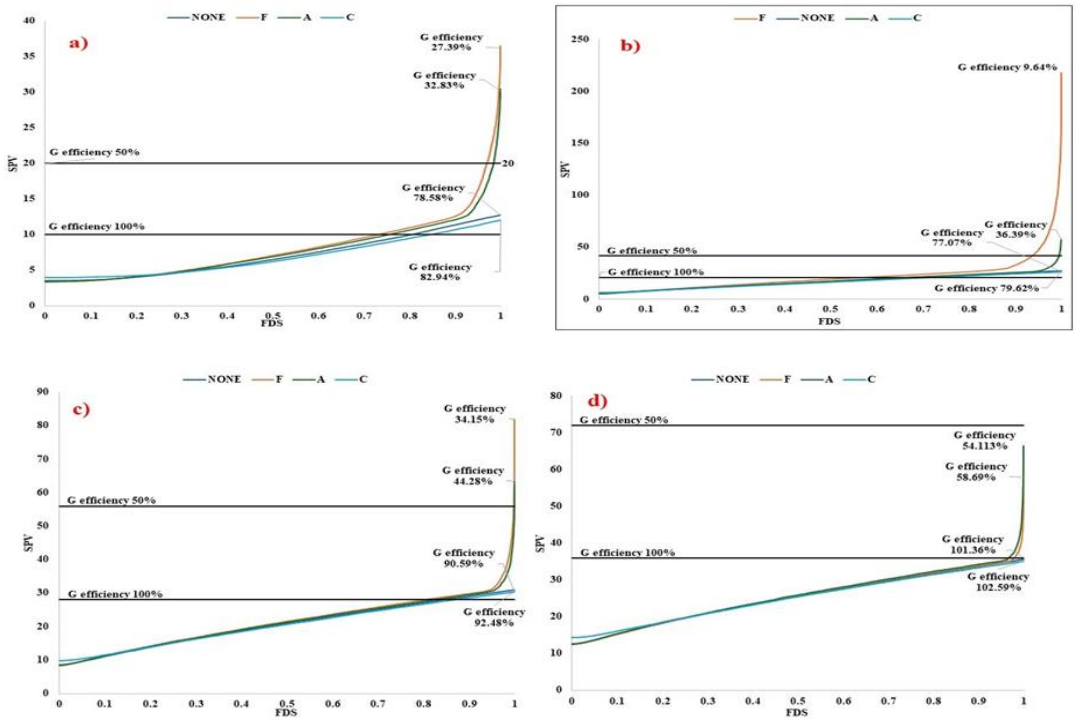


Figure 4: *(a) FDS for CCD K = 3. (b) FDS for CCD K = 5. (c) FDS for CCD K = 6. (d) FDS for CCD K = 7.*

IV. Conclusion

In an experimental setup, the absence of a single observation could potentially influence the results. There are diverse design points in a central composite design context, each with unique characteristics. In the context of a spherical region of interest or near a rotatable alpha for each factor, it's shown that missing a factorial observation influences the *SPV* value, leading to a maximum variance. Through *VDG*, we observe that the *SPV* value begins to rise approximately when the radius $r \geq 1$. On the other hand, in the *FDS* region, factorial data results occupying more than 85% of the fraction of the design space exhibit a significant SPV value. The average and minimum variances across all factors and various design points tend to remain close to each other. The results indicate that the absence of a center run, when compared to scenarios with no missing design points, does not influence the experiment's outcome within any spherical region of interest. The study highlights the benefits of employing graphical representations when dealing with missing observations in the region of interest. This approach can be expanded to include various second-order composite, computer-generated, and optimal designs to examine robust scenarios, making it a valuable tool in experimental design.

References

[1] Ahmad, T. Gilmour, S. G. and Muhammad Arshad, H. (2020). Comparisons of augmented pairs designs and subset designs. *Communications in Statistics: Simulation and Computation*, *49*(7), 1898–1921. https://doi.org/10.1080/03610918.2018.1508703

[1] Ahmad, T. Gilmour, S. G. and Muhammad Arshad, H. (2020). Comparisons of augmented pairs designs and subset designs. *Communications in Statistics: Simulation and Computation*, *49*(7), 1898–1921. https://doi.org/10.1080/03610918.2018.1508703

[2] Akhtar, M. and Prescott, P. (1986). Response surface designs robust to missing observations. *Communications in Statistics - Simulation and Computation*, *15*(2), 345–363. https://doi.org/10.1080/03610918608812512

[3] Alanazi, K. Georgiou, S. D. and Stylianou, S. (2023). On the robustness of central composite designs when missing two observations. *Quality and Reliability Engineering International*, *39*(4), 1143–1171. https://doi.org/10.1002/qre.3281

[4] Alrweili, H. Georgiou, S. and Stylianou, S. (2019). Robustness of response surface designs to missing data. *Quality and Reliability Engineering International*, *35*(5), 1288–1296. https://doi.org/10.1002/qre.2524

[5] Borkowski, J. J. (1995). Spherical prediction-variance properties of central composite and Box—Behnken designs. *Technometrics*, *37*(4), 399–410. https://doi.org/10.1080/00401706.1995.10484373

[6] Box, G. E. P. and Hunter, J. S. (1957). Multi-Factor Experimental Designs for Exploring Response Surfaces. *The Annals of Mathematical Statistics*, *28*(1), 195–241. https://doi.org/10.1214/aoms/1177707047

[7] Box, G. E. P. and Wilson, K. B. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *13*(1), 1–38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x

[8] Draper, N. R. (1961). Missing Values in Response Surface Designs. *Technometrics*, *3*(3), 389. https://doi.org/10.2307/1266729

[9] Giovannitti-Jensen, A. and Myers, R. H. (1989). *American Society for Quality Graphical Assessment of the Prediction Capability of Response Surface Designs* (Vol. 31).

[10] Hayat, H. Akbar, A. Ahmad, T. Bhatti, S. H. and Ullah, M. I. (2023). Robustness to missing observation and optimalities of response surface designs with regular and complex structure. *Communications in Statistics - Simulation and Computation*, *52*(11), 5213–5230. https://doi.org/10.1080/03610918.2021.1977954

[11] Hemavathi, M. Varghese, E. Shekhar, S. Athulya, C. K. Ebeneezar, S. Gills, R. and Jaggi, S. (2022). Robustness of sequential third-order response surface design to missing observations. *Journal of Taibah University for Science*, *16*(1), 270–279. https://doi.org/10.1080/16583655.2022.2046398

[12] Khuri, A. I. and Mukhopadhyay, S. (2010, March). Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*. https://doi.org/10.1002/wics.73

[13] Li, J. Li, L. Borror, C. M. Anderson-Cook, C. Montgomery, D. C. (2009). Graphical Summaries to Compare Prediction Variance Performance for Variations of the Central Composite Design for 6 to 10 Factors. *Quality Technology & Quantitative Management*, *6*(4), 433–449. https://doi.org/10.1080/16843703.2009.11673209

[14] On comparing the prediction variances of some central composite designs in spherical regions a review. (n.d.).

[15] Onwuamaeze, C. U. (2021). Optimal prediction variance properties of some central composite designs in the hypercube. *Communications in Statistics - Theory and Methods*, *50*(8), 1911–1924. https://doi.org/10.1080/03610926.2019.1656746

[16] Ozol-Godfrey, A. Anderson-Cook, C. M. and Montgomery, D. C. (2005). Fraction of Design Space Plots for Examining Model Robustness. *Journal of Quality Technology*, *37*(3), 223–235. https://doi.org/10.1080/00224065.2005.11980323

[17] Park, Y. J. Richardson, D. E. Montgomery, D. C. Ozol-Godfrey, A. Borror, C. M., & Anderson-Cook, C. M. (2005). Prediction variance properties of second-order designs for cuboidal regions. *Journal of Quality Technology*, *37*(4), 253–266. https://doi.org/10.1080/00224065.2005.11980329

[18] Vining, G. G. and Myers, R. H. (1991). A graphical approach for evaluating response surface designs in terms of the mean squared error of prediction. *Technometrics*, *33*(3), 315–326.

https://doi.org/10.1080/00401706.1991.10484837

[19] Zahran, A. Anderson-Cook, C. M. and Myers, R. H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, *35*(4), 377–386. https://doi.org/10.1080/00224065.2003.11980235

ANALYSIS OF A THREE-NODE SERIES QUEUE WITH ENCOURAGED ARRIVAL

Ismailkhan E [1*], R.Jeyachandhiran [2], P.Thangaraja[3], R.Karuppaiya[4]

•

[1*, 3] Department of Mathematics, Panimalar Engineering College, Chennai-600123
[2]Research Scholar, Vellore institute of technology, Vellore-632014
[4]Department of Mathematics, St. George's Arts and Science College, Chennai-600030
*ismailkhanmubarak@gmail.com, jayachandhiran.r@vit.ac.in, thangarajap1991@gmail.com
rajaanju.40@gmail.com

Abstract

This article deals with the three node series queues with encouraged arrival. We increase the expected number of subscribers by using encouraged arrival in this study. Performance metrics is developed by analytic method. After developing the governing-equations and utilizing the Burke's theorem, we resolve the steady-state probabilities and performance metrics of the three-node series queuing system. The study of learning series queues has received substantial interest in a variety of sectors, including manufacturing lines, computer systems, tollgates, telecommunications, and others. Researchers are becoming interested in the series queuing model because of its real-world application. A series queue is a line that runs through a chain of service stations, with subscribers always going along a single track from station to station studied a finite series queue and the view of approximate decomposition.

Keywords: Series queues, Burke's theorem, Encouraged arrival, Governing equations.

I. INTRODUCTION

This article deals with the three node series queues with encouraged arrival. A series queue is a line that runs through a chain of service stations, with subscribers always going along a single track from station to station studied a finite series queue and the view of approximate decomposition. A queue with the Poisson input and exponential service, and disclosed that the exit also Poisson with similar constraint as in input considered [4]. A double-node series queue with general arrivals and analyzed the performance metrics of a limited series queueing system analyzed [14]. An approximation method have predicted [3]. An admission control policy by learning a two-phase with no-space series queueing system analyzed in [5]. A limited series queue by utilizing the idea of critical paths discussed in [11]. Markovian modeling, they have attained the expected, variance of subscribers queue length and expected subscribers delay in the series queue analysed in [13]. They also governed the generating function for attaining the subscriber's queue-size distribution. A double-stage series retrial queueing model with a batch arrival process and attained results studied [10]. Using a fuzzy simulation results with the optimization of a series queue is analyzed in [1]. The server assigning problem without buffers series queueing system considered in [12]. In view of input flow Markovian and service is generalized distribution with multiple stages with heterogeneous subscribers in [9]. They derived performance metrics and also solved the condition of ergodicity. Utilizing matrix-analytic technique has analyzed a series queueing system with heterogeneous servers in [7]. They estimated obvious outcomes for performance indices has taken into attention a series queue with double heterogeneous service providers and also obtained a closed form solution in [6].The performance metrics of multi-service provider series queues in which input is Markovian and service is phase-type discussed in [2]. An M/M/1/N encouraged arrival quality control queueing system discussed in [15]. The research is systematized as in the following sections: Introduction in section 1.

Model description is in section 2. Governing equations are in section 4. Burke's theorem in section 5. Section 6 contains Performance measures. Numerical examples are solved in section 7. Results are in section 8. Conclusions are concluded in section 9.

II. MODEL DESCRIPTION

A series queueing model with 3 service nodes is considered with a Poisson input λ the subscribers from external arrive to the node T1. After concluding the service at the nodeT1, the subscribers will go to the node T2 from the node T2 for getting service, in front of the node T3 they join queue. The subscribers leave the system only after the achievement of service at node T3. The queue assumed to be unlimited in capacity and at nodes T1, T2 and T3 service-time follows exponential distributions with parameters μa, μb and μc. It is distinguished that one subscriber can access service from each node at once.

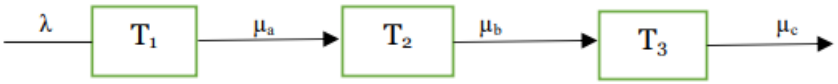


Figure 1: *Three - Phase Series Queue*

To estimate the steady-state probability P, K, L, M nodes of K subscribers at T1 and M subscribers respectively at nodes T2 and T3. Where K, L, M ≥ 0. Utilizing the state-transition figures the governing equations can be written (fig.2 and fig.3).

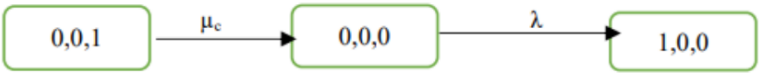


Figure 2.1: *represents that from state(0,0,1) only the (0,0,0)th state can be attained.*

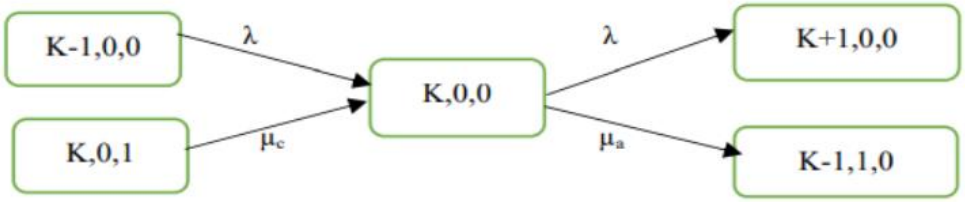


Figure 2.2: *represent that only the (K,0,0)th state can be reached only from the states (K-1,0,0)th or (K,0,1)th and arrival to the (K,0,0)th state becomes the (K+1,0,0)th state and the departure (K,0,0)th state becomes (K-1,1,0)th state*

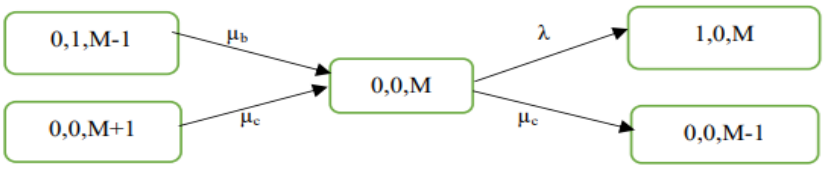


Figure 2.3: *Presents the (0,0,M)th state can be reached only from the states (0,1,M-1)th or (0,0,M+1)th state and arrival to the (0,0,M)th state becomes the (1,0,M)th state and the departure of (0,0,M)th state becomes (0,0,M-1)th state*

Figure 2.4: *Presents the (0,L,0)th state can be reached only from the states (1,L-1,0)th or (0,L,1)th and arrival to the (0,L,0)th state becomes the (1,L,0)th state and the departure of (0,L,0)th state becomes (0,L-1,1)th state*

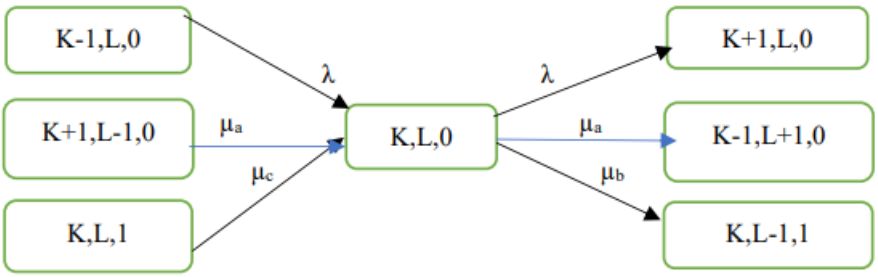


Figure 2.5: *Presents the (K,L,0)th state can be reached only from the states (K-1,L,0)th or (K+1,L-1,0)th or (K,L,1) and arrival to the (K,L,0)th state becomes the (K+1,L,0)th state and the departure of (K,L,0)th state becomes (K-1,L+1,0)th state or (K,L-1,1)th state*

Figure 2: *State Transition Diagram*

Figure 2 represents that from state (0,0,1) only the (0,0,0)th state can be attained.

The only state (1,0,0) into which the middle state (0,0,0) can traverse and the mid state cannot traverse to any other state expect (1,0,0)th state .

Similarly from (M-1,0,0) and (M,0,1) states, only the (M,0,0)th state can be reached.

The subscribers can move either to state (M+1,0,0)th or to (M-1,1,0) state from the (M,0,0)th state.

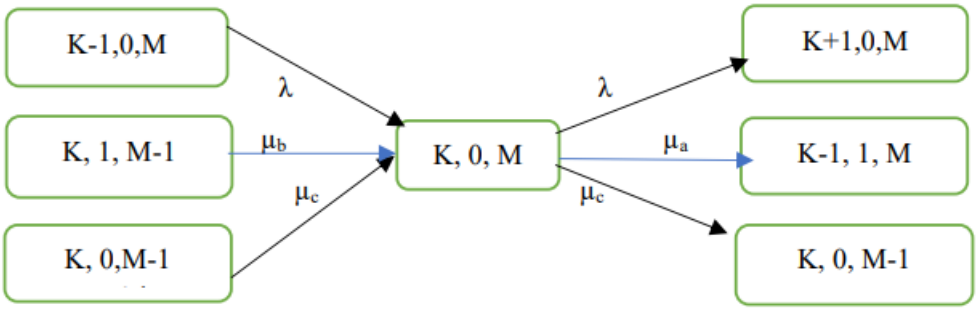


Figure 3.1: *Presents (K,0,M)th state can be reached only from the states (K-1,0,M)th or (K,1,M-1)th or (K,0,M-1) and arrival to the (K,0,M)th state becomes the (K+1,0,M)th state and the departure of (K,0,M)th state becomes (K-1,1,M)th state or (K,0,M-1)th state*

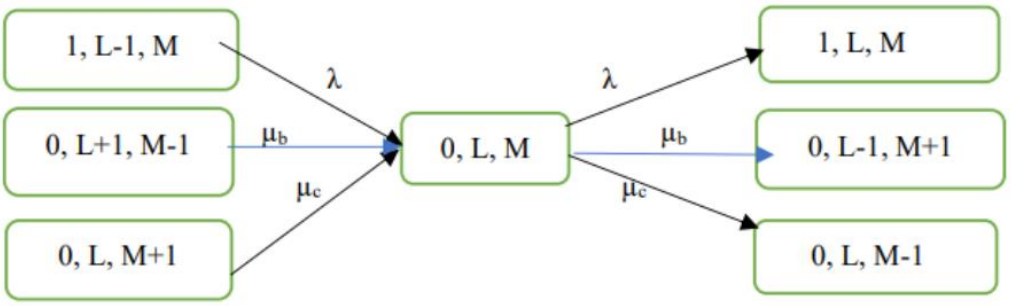


Figure 3.2:*Presents the (0,L,M)th state can be reached only from the states (1,L-1,M)th or (0,L+1,M-1)th or (0,L,M+1) and arrival to the (0,L,M)th state becomes the (1,L,M)th state and the departure of (0,L,M)th state becomes (0,L-1,M+1)th state or (0,L,M-1)th state .*

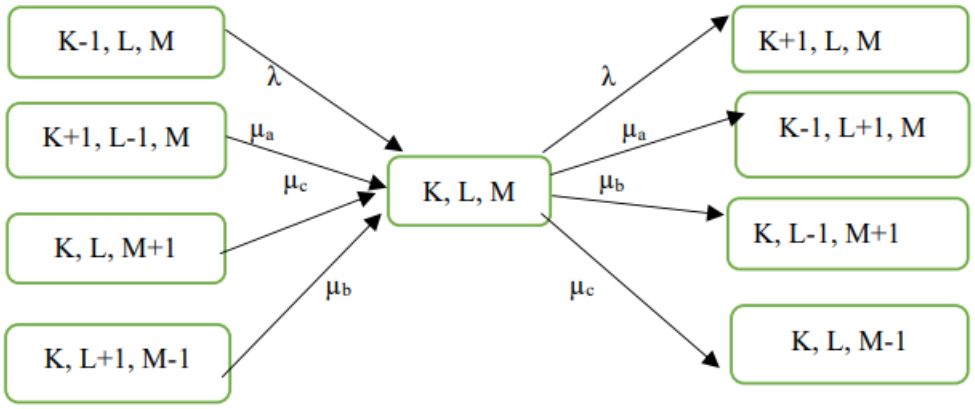


Figure 3.3: *Presents the (K,L,M)th state can be reached only from the states (K-1,L,M)th or (K+1,L-1,M)th or (K,L,M+1)th or (K,L+1,M-1)th and arrival to the (K,L,M)th state becomes the (K+1,L,M)th state and the departure of (K,L,M)th state becomes (K-1,L+1,M)th state or (K,L-1,M+1)th state or (K,L,M-1)th state.*

III. NOTATIONS

The notations are used to formulate the series queue with 3-nodes T1, T2, T3. Where $P_{K,L,M}$ denotes the steady-state probabilities that there are K number of subscribers at node T1, L-number of subscribers in T2 and M-subscribers in T3. States of the model are given by (K, L, M).

P_K − Probability that there are K subscribers at the node T_1

P_L − Probability that there are L subscribers at the node T_2

P_M − Probability that M subscribers at the node T_3

$P_{0,0,0}$ − Probability that no subscribers at any node T_1, T_2 and T_3.

$P_{K,0,0}$ − The probability that K subscribers at node T_1 and no subscribers at nodes T_2 and T_3.

$P_{0,L,0}$ − Probability that no subscribers at node T_1 and T_3, but L subscribers are at node T_2.

$P_{0,0,M}$ − Probability that no subscribers at node T_1 and T_2, but M subscribers are at node T_3.

$P_{K,L,0}$ − Probability that K subscribers at node T_1 and L subscribers at node T_2 no subscribers at node T_3.

$P_{K,0,M}$ - Probability K subscribers are at node T_1 and no subscribers at node T_2 and also M subscribers at node T_3.

$P_{0,L,M}$ − Probability that no subscribers at node T_1, but L subscribers at node T_2 and M subscribers at node T_3.

$P_{K,L,M}$ − Probability that k subscribers at node T_1, L subscribers at node T_2 and M subscribers at T_3.

IV. GOVERNING EQUATIONS

By utilizing the transition figure in figure 2.1, 2.2, 2.3, 2.4, 2.5 and figure 3.1, 3.2, 3.3 we developed the steady state equations as below;

$$\lambda(1+\gamma)P_{0,0,0} = \mu_c P_{0,0,1} \quad (1)$$

$$(\lambda(1+\gamma)+\mu_c)P_{0,0,M} = \lambda(1+\gamma)P_{0,1,M-1} + \mu_a P_{0,0,L+1} \quad (2)$$

$$(\lambda(1+\gamma)+\mu_a)P_{K,0,0} = \lambda(1+\gamma)P_{K-1,0,0} + \mu_c P_{K,0,1} \quad (3)$$

$$(\lambda(1+\gamma)+\mu_b)P_{0,L,0} = \mu_a P_{1,L-1,0} + \mu_c P_{0,L,1} \quad (4)$$

$$(\lambda(1+\gamma)+\mu_a+\mu_b)P_{K,L,0} = \lambda(1+\gamma)P_{K-1,L,0} + \mu_a P_{K+1,L-l,0} + \mu_c P_{K,L,1} \quad (5)$$

$$(\lambda(1+\gamma)+\mu_a+\mu_c)P_{k,0,m} = \lambda(1+\gamma)P_{K-1,0,M} + \mu_b P_{K,1,M-1} + \mu_c P_{K,0,M+1} \quad (6)$$

$$(\lambda(1+\gamma)+\mu_b+\mu_c)P_{0,l,m} = \mu_a P_{1,L-1,M} + \mu_b P_{0,L+1,M-1} + \mu_c P_{0,L,M+1} \quad (7)$$

$$(\lambda(1+\gamma)+\mu_a+\mu_b+\mu_c)P_{K,L,M} = \lambda(1+\gamma)P_{K-1,L,M} + \mu_a P_{K+1,L-1,M} + \mu_c P_{K,L,M+1} + \mu_b P_{K,L+1,M-1} \quad (8)$$

Moreover, the addition of entire probabilities must equal to one. i.e.,

$$\sum_K \sum_L \sum_M P_{K,L,M} = 1 \quad (9)$$

Burke's theorem is used in this instance to solve the governing equations mentioned above. The following is the theorem's statement:

V. BURKE'S THEOREM

Poisson arrival queue with single waiting queue with no departures, and exponentially distributed service times, the equilibrium distribution of the number of service accomplishments in a random time duration is equal as the arrival distribution, for every number of service providers.

To illustrate the outcome, Burke developed the supposition that the span of the expected inter-arrival $\frac{1}{\lambda(1+\gamma)}$, there are c service providers with exponentially distributed mean rate $\frac{1}{\mu}$ and $c\mu > \lambda(1+\gamma)$. By utilizing the conventions in its place of verifying the equilibrium distribution of the number of subscribers completing service between an arbitrary duration of span t follow poisson with rate $\lambda(1+\gamma)t$, Burke verified an equivalent output, that the time durations among successive service accomplishments are independent, exponentially distributed and it is identical as the inter-arrival times (visit [4]).

The process of entering (arriving) at T2, which is identical to the process of leaving T1

is also Poisson with parameter $\lambda(1+\gamma)$ from the theorem. In case of the service providers T_2 and T_3 same is detained. Hence, the queueing models are M/M/1 queueing system. so far M/M/1 queueing model is taken into account, now we govern the probability distribution for the three phase series queue by utilizing the Burke's theorem.

P_K − Probability that k number of subscribers at the node T_1

$$= \left[\frac{\lambda(1+\gamma)}{\mu_a}\right]^K \left[1 - \frac{\lambda(1+\gamma)}{\mu_a}\right] \tag{10}$$

P_L − Probability that l number of subscribers at the node T_2

$$= \left[\frac{\lambda(1+\gamma)}{\mu_b}\right]^L \left[1 - \frac{\lambda(1+\gamma)}{\mu_b}\right] \tag{11}$$

P_M − Probability that m number of subscribers at the node T_3

$$= \left[\frac{\lambda(1+\gamma)}{\mu_c}\right]^M \left[1 - \frac{\lambda(1+\gamma)}{\mu_c}\right] \tag{12}$$

At each node T_1, T_2, T_3 the number of subscribers are random variables which is independent, hence the probability of K amount of subscribers at node T_1 and L number of subscribers at node T_2 and m subscribers at node T_3 are jointly assumed by,

$$P_{K,L,M} = \left[\frac{\lambda(1+\gamma)}{\mu_a}\right]^K \left[1 - \frac{\lambda(1+\gamma)}{\mu_a}\right] \left[\frac{\lambda(1+\gamma)}{\mu_b}\right]^L \left[1 - \frac{\lambda(1+\gamma)}{\mu_b}\right] \left[\frac{\lambda(1+\gamma)}{\mu_c}\right]^M \left[1 - \frac{\lambda(1+\gamma)}{\mu_c}\right] \tag{13}$$

where $K, L, M \geq 0$

VI. PERFORMANCE MEASURES

For analyzing the series queueing model performance, different performance measures are developed utilizing steady state probability distributions as-

(i) The expected number of subscribers in the system

$$L=\sum_{K} KP_K+\sum_{l} LP_L+\sum_{m} MP_M$$
$$=\sum_{K=0}^{\infty} K\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]^K\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]+\sum_{L=0}^{\infty} L\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]^L\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]$$
$$+\sum_{M=0}^{\infty} M\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]^M\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$
$$=\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}-\lambda(1+\gamma)}+\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}-\lambda(1+\gamma)}+\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}-\lambda(1+\gamma)}$$

(ii) The expected waiting time of the subscribers in the system

$$\mathrm{W}=\frac{1}{\lambda(1+\gamma)}L$$
$$=\frac{1}{\mu_{\mathrm{a}}-\lambda(1+\gamma)}+\frac{1}{\mu_{\mathrm{b}}-\lambda(1+\gamma)}+\frac{1}{\mu_{\mathrm{c}}-\lambda(1+\gamma)}$$

(iii) The probability that the three service nodes are free

$$P_{0,0,0}=\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$
$$=\left[\frac{\mu_{\mathrm{a}}-\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\mu_{\mathrm{b}}-\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\mu_{\mathrm{c}}-\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$

(iv) The probability that K subscribers at node T_1 and no subscribers at T_2 and T_3

$$P_{K,0,0}=\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]^K\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$
$$=\lambda(1+\gamma)^K\left[\frac{\mu_1-\lambda(1+\gamma)}{\mu_{\mathrm{a}}^{K+1}}\right]\left[\frac{\mu_{\mathrm{b}}-\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\mu_{\mathrm{c}}-\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$

(v) The probability that no subscribers at nodes T_1 and T_3 and L number of subscribers at node T_2

$$P_{0,L,0}=\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]^L\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$
$$=\lambda(1+\gamma)^L\left[\frac{\mu_{\mathrm{a}}-\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\mu_{\mathrm{b}}-\lambda(1+\gamma)}{\mu_{\mathrm{b}}^{L+1}}\right]\left[\frac{\mu_{\mathrm{c}}-\lambda(1+\gamma)}{\mu_3}\right]$$

(vi) The probability that zero subscribers at nodes T_1 and T_2 and M subscribers at node T_3

$$P_{0,0,M}=\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]^0\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]^M\left[1-\frac{\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$
$$=\lambda(1+\gamma)^M\left[\frac{\mu_{\mathrm{a}}-\lambda(1+\gamma)}{\mu_{\mathrm{a}}}\right]\left[\frac{\mu_{\mathrm{b}}-\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\mu_{\mathrm{c}}-\lambda(1+\gamma)}{\mu_{\mathrm{c}}^{M+1}}\right]$$

(vii) The Probability that K, M subscribers at nodes T_1 and T_2 respectively and no subscribers at node T_3

$$P_{K,0,M}=\lambda(1+\gamma)^{K+M}\left[\frac{\mu_{\mathrm{a}}-\lambda(1+\gamma)}{\mu_{\mathrm{a}}^{K+1}}\right]\left[\frac{\mu_{\mathrm{b}}-\lambda(1+\gamma)}{\mu_{\mathrm{b}}}\right]\left[\frac{\mu_{\mathrm{c}}-\lambda(1+\gamma)}{\mu_{\mathrm{c}}^{M+1}}\right]$$

(viii) The Probability that K, L, zero subscribers at nodes T, T_2T_3 respectively

$$P_{K,L,0}=\lambda(1+\gamma)^{K+l}\left[\frac{\mu_{\mathrm{a}}-\lambda(1+\gamma)}{\mu_{\mathrm{a}}^{K+1}}\right]\left[\frac{\mu_{\mathrm{b}}-\lambda(1+\gamma)}{\mu_{\mathrm{b}}^{L+1}}\right]\left[\frac{\mu_{\mathrm{c}}-\lambda(1+\gamma)}{\mu_{\mathrm{c}}}\right]$$

(ix) The Probability that 0, L, M subscribers at nodes T, T_2T_3 respectively

$$P_{0,L,M} = \lambda(1+\gamma)^{L+M}\left[\frac{\mu_a - \lambda(1+\gamma)}{\mu_a}\right]\left[\frac{\mu_b - \lambda(1+\gamma)}{\mu_b^{L+1}}\right]\left[\frac{\mu_c - \lambda(1+\gamma)}{\mu_c^{M+1}}\right]$$

(x) The Probability that K, L, M subscribers at nodes $T_1, T_2 T_3$ respectively

$$P_{K,L,M} = \lambda(1+\gamma)^{K+L+M}\left[\frac{\mu_a - \lambda(1+\gamma)}{\mu_a^{K+1}}\right]\left[\frac{\mu_b - \lambda(1+\gamma)}{\mu_b^{L+1}}\right]\left[\frac{\mu_c - \lambda(1+\gamma)}{\mu_c^{M+1}}\right]$$

(xi) At the nodes T_1, T_2 and T_3 the probability that the subscribers h,l is where h>K, and i>L and j>M is given by

$$\begin{aligned} P_{h>K,i>L,j>M} &= \sum_{h=K+1}^{\infty} P_h \cdot \sum_{i=L+1}^{\infty} P_i \cdot \sum_{j=M+1}^{\infty} P_j \\ &= \left[\frac{\lambda(1+\gamma)}{\mu_a}\right]^{K+1} \cdot \left[\frac{\lambda(1+\gamma)}{\mu_b}\right]^{L+1} \cdot \left[\frac{\lambda(1+\gamma)}{\mu_c}\right]^{M+1} \\ &= \frac{\lambda(1+\gamma)^{k+l+m+3}}{\mu_a^{K+1} \cdot \mu_b^{L+1} \cdot \mu_c^{M+1}} \end{aligned}$$

(xii) The expected number of subscribers in the queue

$$\begin{aligned} L_Q &= \sum_{K=1}^{\infty} (K-1)P_K + \sum_{L=1}^{\infty} (i-1)P_i + \sum_{M=1}^{\infty} (M-1)P_M \\ &= \frac{\lambda(1+\gamma)}{\mu_a} \cdot \frac{\lambda(1+\gamma)}{\mu_a - \lambda(1+\gamma)} + \frac{\lambda(1+\gamma)}{\mu_b} \cdot \frac{\lambda(1+\gamma)}{\mu_b - \lambda(1+\gamma)} + \frac{\lambda(1+\gamma)}{\mu_c} \cdot \frac{\lambda(1+\gamma)}{\mu_c - \lambda(1+\gamma)} \\ &= (\lambda(1+\gamma))^2 \left[\frac{1}{\mu_a(\mu_b - \lambda(1+\gamma))} + \frac{1}{\mu_b(\mu_b - \lambda(1+\gamma))} + \frac{1}{\mu_c(\mu_c - \lambda(1+\gamma))}\right] \end{aligned}$$

(xiii) The expected waiting time of a subscribers in the queue

$$W_Q = \frac{L_q}{\lambda(1+\gamma)}$$

$$W_Q = \lambda(1+\gamma)\left[\frac{1}{\mu_a(\mu_a - \lambda(1+\gamma))} + \frac{1}{\mu_b(\mu_b - \lambda(1+\gamma))} + \frac{1}{\mu_c(\mu_c - \lambda(1+\gamma))}\right]$$

VII. NUMERICAL EXAMPLES

In this section, to formulate the series queue with 3-nodes T1, T2, T3. The steady-state probabilities that there are K number of subscribers at node T1, L-number of subscribers in T2 and M-subscribers in T3. States of the model are given by (K, L, M).The performance of the queuing system is examined numerically in relation to the parameters $\lambda(1+\gamma)$. Where γ =10% and 20%. Since Little's law values for the chosen parameter values are displayed in the table as a distinct column

Table1: *Represents the value of L and W when γ=0.0*

Λ	L	W	L =λ(1 + γ)W
1	0.869	0.869	0.869
1.1	0.9849	0.8954	0.9849
1.2	1.1080	0.9233	1.1080
1.3	1.2391	0.9531	1.2391
1.4	1.3790	0.9850	1.3790
1.5	1.5286	1.0190	1.5286

Table2: *Represents the value of L and W when γ=0.1*

Λ	L	W	L=λ(1 + γ)W
1	0.9849	0.8954	0.9849
1.1	1.1207	0.9262	1.1207
1.2	1.2663	0.9593	1.2663
1.3	1.4228	0.9950	1.4228
1.4	1.5914	1.0334	1.5914
1.5	1.7736	1.0745	1.7736

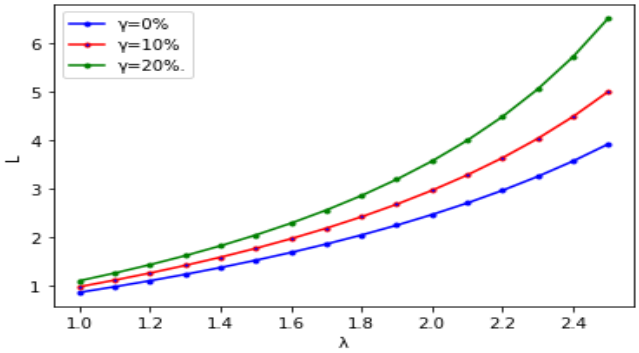


Figure 4: *plot against L and λ*

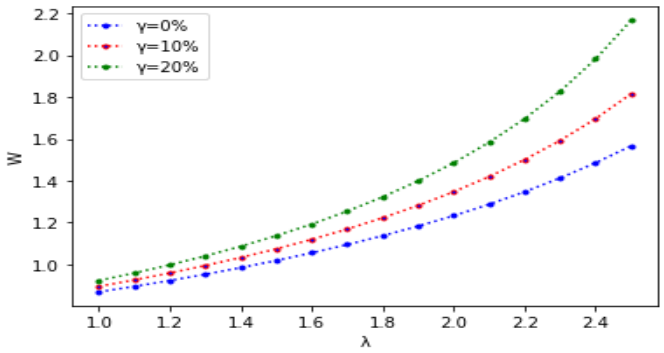


Figure5: *plot against W and λ*

VIII. RESULTS AND DISCUSSIONS

In table 1, the values of L and W increases when we increase the arrival rate λ with γ =0. Table 2, the values of L and W increases when we increase the arrival rate λ with γ =0.1. Table 3, the system of L and W increases when we increase the arrival rate λ with γ =0.2. Figure 4, When γ =0.2 (20%) we got maximum value of L than in table 1 and table 2. Figure 5, the results of W for different values of γ=0 to 0.2.

IX. CONCLUSIONS

In terms of operations management, planning, outlining, and execution, this study is quite helpful. Developing services for customers, and other areas. We increased the expected number of subscribers in the system. We developed the performance metrics by applying Burke's theorem. We enlarged the system size by using encouraged arrival discounts. Increasing the service subscribers will surely improve the company's revenue. We can further extend this article by increasing the number of servers and adding optimal cost to study the behavior of the series queues.

References

[1] Azadeh, A., Mohammad Ebrahima, R., and Eivazy, H. (2010). Parameter optimization of series queue systems with finite intermediate buffers via fuzzy simulation, *Performance Evaluation*, 67: 353-360.

[2] Baumann, H. and Sandmann, W. (2017). Multiserver series queue with Markovian arrival process, phase-type service times and finite buffers, *European Journal of Operational Research*, 256: 187-195.

[3] Bouhchouch, A., Frein, Y. and Dallery, Y. (1996). Performance evaluation of closed series queueing networks with finite buffers, *Performance Evaluation*, 26: 115-132.

[4] Burke, P. J. (1956). The output of a queueing system, *Operations Research*, 4: 699-704.

[5] Chang, K. H. and Chen, W. F. (2003). Admission control policies for two-stage series queues with no waiting spaces, *Computers and Operations Research*, 30: 589-601.

[6] Do,T.V. (2015). A closed-form solution for a toll booth series queue with two heterogeneous servers and exponential service times, *European Journal of Operationa Research*, 247: 672- 675.

[7] He, Q. M. and Chao, X. (2014). A tollbooth series queue with heterogeneous servers, *European Journal of Operational Research*, 236: 177- 189.

[8] Hunt, G. C. (1956). Sequential arrays of waiting lines, *Operations Research*, 4: 674-683.

[9] Kim, C., Dudin, A., Dudina, O., and Dudin, S. (2014). Series queueing system with infinite and finite intermediate buffers and generalized phase-type service time distribution, *European Journal of Operational Research*, 235: 170-179.

[10] Kim, C. S., Park, S. H., Dudin, A.,Klimenok , V. and Tsarenkov, G. (2010). Investigaton of the BMAP/G/1 → /P H/1/M series queue with retrials and losses, *Applied Mathematical Modelling*, 34: 2926-2940.

[11] Li, Y., Cai, X., Tu, F. and Shao, X. (2004). Optimization of series queue systems with finite buffers, *Computers and Operations Research*, 31: 963-984.

[12] Mohammad, H., Yarmand, Douglas, G. and Down (2013). Server allocation for zero buffer series queues, *European Journal of Operational Research*, 230: 596-603.

[13] Song, X. and Mehmet Ali, M. (2009). A performance metrics of discrete-time series queues with Markovian sources, *Performance Evaluation*, 66: 524-543.

[14] Tsiaris, G. D. (1992). Closed form expected queueing times of a series system with finite intermediate buffer: A recursive approach, *Applied Mathematical Letter*, 5: 33-35.

[15] Khan I.E. and Paramasivam R. (2022). Reduction in waiting time in an M/M/1/N encouraged arrival queue with feedback, balking and maintaining of reneged customers. *Symmetry*, 14(8):1743.

ANALYSIS OF TWO VACATION POLICIES UNDER RETRIAL ATTEMPTS, MARKOVIAN ENCOURAGED ARRIVAL QUEUING MODEL

Rajeswaran K[1],Rajendran P[2,*] Sanjay K[3],Shivali S[4] Ismailkhan E[5*]

•

[1,3,4]BTech student, vellore institute of technology, Vellore, 632014, Tamilnadu, India
[2*]Professor, vellore institute of technology, Vellore, 632014, Tamilnadu, India
[5*]Assistant Professor, Panimalar Engineering College,Chennai, 600123, Tamilnadu, India

prajendran@vit.ac.in, ismailkhanmubarak@gmail.com

Abstract

In this study,Markovian queuing models, which follow encouraged arrival rates and exponential service rates, are used in a variety of systems, including manufacturing, production, telecommunications, computers, and transportation. Every one has a hectic schedule and little free time in the modern world. Because the customer's arrival is unpredictable, they cannot complete their task in the allotted time because they cannot predict it. The encouraged arrival, idle server state, busy server state, vacation state, and breakdown and repair state conditions for a single-server Markovian queuing system were all taken into consideration. Vacation time grows acceleratory, and vacation policies abound. This Markovian-encouraged arrival queuing model takes into account customer impatience and retrial efforts to ensure service completion. We calculate the combined probability of these states and compare first-come, first-served with bulk service. The different performance measures have also been explained.

Keywords: Two vacation, encouraged arrival, reneging, retrial queue, Markov model

1. Introduction

We consider the single server to be encouraged for working vacations, breakdowns, and repairs where the server may be down at any time. A Poisson distribution governs customer arrival rates, whereas an exponential distribution governs customer service rates. Arriving customers will join the orbit group if they discover that the server is too busy serving another customer. Kalyanaraman and Sundaramoorthy [1] demonstrated that a Markovian has dependent arrival and breakdown. A single vacation on an unstable bulk server is detailed in Haridass and Arumuganathan [2]. Batch arrivals on an infinite server are investigated in Daw and Pender [3]. Examples of batch queues with inadequate identification are explored in Bar-Lev et al[4].
A generalized bulk queue with the Poisson model is addressed in Neuts [5]. Parveen and Begum [6] Investigated a general bulk queuing approach with a dual working vacation. In Li and Zhao [7], a retrial model with a consistent retrial rate, breakdowns, and dissatisfied customers were studied. A dual server queuing model for bulk arrival and service was examined in Kumar and Shinde [8]. Working vacations on bulk arrival queues combined with reneging and interruptions were studied in Vijaya Laxmi and Rajesh [9]. Dual service and two vacations are investigated in the bulk arrival Markovian system by Srivastava et al [10]. They investigated batch arrival and retrial queues using a dual vacation policy and the Markovian queuing model Singh and

Srivastava [11]. Som and Seth [12] Explored an $M/M/1/N$ system along with encouraged arrivals. Reduced wait times in an M/M/1/N encouraged customer arrivals, as seen in Khan and Paramasivam [13]. Reneging customers were observed in Som [14] when the M/M/c/N model was used in conjunction with encouraged arrivals.

2. Model elaboration

The following presumptions have been taken into account:

- The customers follow the First-Come-First-Serve rule
- Customers arrive according to an encouraged arrival with mean $\frac{1}{\lambda*(1+\omega)}$. ω representing the offered value and accelerated distribution with mean $\frac{1}{\mu}$.
- Customers will enter the retrial queue with probability 's'
- After a few tries, the customer notices the accelerated distribution with an average of 1- φ and attempts the request from the retrial space.
- When a customer attempts to receive service after a certain amount of time has passed, with probability δ_0
- The server will fail time following the accelerated distribution under the fail rate ν_1 and from the service to the customer under the repair rate ν_2.

We have given that the random variable $J(t)$ describes the absolute customers of the system at period 't', and $R(t) = 0, 1, 2, 3 \ldots$ consider the states, a server is free, a server is busy, the server is on vacation and the server is in breakdown & repair state at period 't'.

3. Analysis of system-size distribution

From first state $0, j \geq 0$ when the server is idle state:

$$\lambda * (1+\omega)\, \acute{P}_{0,0} = \mu \acute{P}_{1,0} \tag{1}$$

$$(\lambda * (1+\omega) + l\psi\delta_0)\, \acute{P}_{0,\,l} = l\mu \acute{P}_{1,\,l} \tag{2}$$

$$(\lambda * (1+\omega) + j\varphi\delta_0)\, \acute{P}_{0,j} = j\mu \acute{P}_{1,j} \tag{3}$$

From second-state $1, j \geq 0$ when the -.server is busy state:

$$(\nu_1 + \Delta_0 + s\lambda * (1+\omega) + \mu)\, \acute{P}_{1,0} = (1-s)\, \varphi \acute{P}_{1,1} + \nu_2 \acute{P}_{3,0} + \lambda * (1+\omega) \acute{P}_{2,0} + \varphi\delta_0 \acute{P}_{0,1}$$

$$+\lambda * (1+\omega) \acute{P}_{0,0}\, (1 - \Delta_0 + s\lambda * (1+\omega) + l\mu + \nu_1)\, P_{1,\,l} \tag{4}$$

$$= \nu_2 \acute{P}_{3,\,l} + \mu \acute{P}_{2,\,l} + \lambda * (1+\omega)\, \acute{P}_{0,l} + s\lambda * (1+\omega)\, \acute{P}_{1,\,l-1}$$

$$+(1-s)(l+1)\varphi P_{1,\,l+1} + (l+1)\varphi\delta_0 \acute{P}_{0,\,l+1}\, (1 - \Delta_0 + \nu_1 + j\mu + (1-s)j\varphi + s\lambda * (1+\omega))\, \acute{P}_{1,j} \tag{5}$$

$$= \mu \acute{P}_{2,j} + \nu_2 \acute{P}_{3,j} + \lambda * (1+\omega)\, \acute{P}_{0,j} + s\lambda * (1+\omega)\, \acute{P}_{1,-1}$$

$$+s\lambda * (1+\omega)\, (j+1)\, \varphi \acute{P}_{1,j+1} + (j+1)\, \varphi\delta_0 \acute{P}_{0,j+1} \tag{6}$$

From third state 2, when the server is on vacation:

$$2\lambda * (1+\omega)\, \acute{P}_{2,0} = \Delta_0 \acute{P}_{0,1} \tag{7}$$

$$(\mu + \lambda * (1+\omega))\, \acute{P}_{2,\,l} = (1-\Delta_0)\, P_{1,\,l} + \lambda * (1+\omega)\, \acute{P}_{1,\,l-1} \tag{8}$$

$$\mu \acute{P}_{2,j} = (1-\Delta_0)\, \acute{P}_{1,j} + \lambda * (1+\omega)\, \acute{P}_{1,j-1} \tag{9}$$

For fourth state 3, when the- server breakdown and is repaired state:

$$\nu_2 \acute{P}_{3,0} = \nu_1 \acute{P}_{1,0} \tag{10}$$

$$\nu_2 \acute{P}_{3,\,l} = \nu_1 \acute{P}_{1,\,l} \tag{11}$$

$$\nu_2 \acute{P}_{3,j} = \nu_1 \acute{P}_{1,j} \tag{12}$$

The following conclusions may be drawn from equations (1), (2), and (3).

$$\left.\begin{array}{c} \acute{P}_{1,0} = \left(\frac{\lambda*(1+\omega)}{\mu}\right) \acute{P}_{0,0} \\ \acute{P}_{1,m} = \left(\frac{\lambda*(1+\omega)+l\varphi\delta_0}{l\mu}\right) \acute{P}_{0,\,l} \\ \acute{P}_{1,j} = \left(\frac{\lambda*(1+\omega)+j\varphi\delta_0}{j\mu}\right) \acute{P}_{0,j} \end{array}\right\} \tag{13}$$

Using (4), now with j = 0 and (10), we obtain

$$(\nu_1 + \Delta_0 + s^{\smile} * (1+\omega) + \mu)\, \acute{P}_{1,0} = (1-s)\, \varphi \acute{P}_{1,1} + \nu_1 \acute{P}_{1,0} + \lambda * (1+\omega) \acute{P}_{2,0} + \varphi\delta_0 \acute{P}_{0,1}$$

$$+\lambda * (1+\omega) \acute{P}_{0,0}\, (\Delta_0 + s^{\smile} * (1+\omega) + \mu)\, \acute{P}_{1,0} = (1-s)\, \varphi \acute{P}_{1,1}$$

$$+\lambda * (1+\omega) \acute{P}_{2,0} + \varphi\delta_0 \acute{P}_{0,1} + \lambda * (1+\omega) \acute{P}_{0,0}$$

If we use (1), we obtain

$$\left(\frac{\Delta_0 \lambda * (1+\omega) + s(\lambda * (1+\omega))^2}{\mu}\right) \acute{P}_{0,0} = (1-s)\varphi \acute{P}_{1,1} + \lambda * (1+\omega) \acute{P}_{2,0} + \varphi\delta_0 \acute{P}_{0,1}$$

When k=1, use (13)

$$\left.\begin{array}{c} \acute{P}_{1,1} = \left(\frac{\lambda*(1+\omega)+\varphi\delta_0}{\mu}\right) \acute{P}_{0,1} \\ \acute{P}_{1,1} = a_1 \acute{P}_{0,1} \\ \left(\frac{\Delta_0\lambda*(1+\omega)+s(\lambda*(1+\omega))^2}{\mu}\right) \acute{P}_{0,0} = \end{array}\right. \tag{14}$$

$$(1-s)\varphi\ a_1 \acute{P}_{0,1} + \lambda * (1+\omega) \acute{P}_{2,0} + \varphi\delta_0 \acute{P}_{0,1}$$

If we use (7)

$$\left(\frac{\Delta_0 \lambda * (1+\omega) + s(\lambda * (1+\omega))^2}{\mu}\right) \acute{P}_{0,0} = \left[(1-s+a_1+\delta_0)\, \varphi + \frac{\Delta_0}{2}\right] \acute{P}_{0,1}$$

$$\acute{P}_{0,1} = \frac{2\left(\Delta_0 \lambda * (1+\omega) + s(\lambda * (1+\omega)\Delta)^2\right)}{\mu\left[(1-s+a_1+\delta_0)\, \varphi + \frac{\Delta_0}{2}\right]} \acute{P}_{0,0} \acute{P}_{0,1} = \frac{b_2}{c_2} \acute{P}_{0,0} \tag{15}$$

Applying j=1 in (6)

$$(1-\Delta_0+\nu_1+\mu+(1-s)\varphi+s^{\check{}}*(1+\omega))\acute{P}_{1,1}=\mu\acute{P}_{2,1}+\nu_2\acute{P}_{3,1}+\lambda*(1+\omega)\acute{P}_{0,1}$$

$$+s^{\check{}}*(1+\omega)\acute{P}_{1,0}+2s^{\check{}}*(1+\omega)\varphi\acute{P}_{1,2}+2\varphi\delta_0\acute{P}_{0,2}$$

To do this, use (12) and assign j=1.

$$(1-\Delta_0+\nu_1+(1-s)\,\varphi+s\lambda\,(1+\omega)+\mu)\,\acute{P}_{1,1}=\mu\acute{P}_{2,1}+v_1\acute{P}_{1,1}+\lambda\,(1+\omega)\,\acute{P}_{0,1}+s\lambda\,(1+\omega)\,\acute{P}_{1,0}$$

$$+2s_1\lambda\,(1+\omega)\,\acute{P}_{1,2}+2\varphi\delta_0\acute{P}_{0,2}\,[1-\Delta_0+(1-s)\,\varphi+\mu+s\lambda\,(1+\omega)]\,\acute{P}_{1,1}$$

$$=\mu\acute{P}_{2,1}+\lambda\,(1+\omega)\,\acute{P}_{0,1}+s_{1,0}\acute{P}_{1,0}+2s_1\lambda\,(1+\omega)\,\acute{P}_{1,2}+2\varphi\delta_0\acute{P}_{0,2}$$

If we use (9) for $j=1$ and (14), we get

$$\begin{gathered}\mu\acute{P}_{2,1}=(1-\Delta_0)\,\acute{P}_{1,1}+\lambda\,(1+\omega)\,\acute{P}_{1,0}\\((1-s)\,\varphi+\mu+s^{\check{}}\,(1+\omega))\,a_1\acute{P}_{0,1}\\=\lambda\,(1+\omega)\,\acute{P}_{1,0}+\lambda\,(1+\omega)\,\acute{P}_{0,1}+s^{\check{}}\,(1+\omega)\,\acute{P}_{1,0}\\+2s^{\check{}}\,(1+\omega)\,\delta_0\acute{P}_{1,2}+2\varphi\delta_0\acute{P}_{0,2}\\((1-s)\varphi\;a_1+\mu a_1+sa_1-\lambda\,(1+\omega))\,\acute{P}_{0,1}\\=\lambda\,(1+\omega)\,\acute{P}_{1,0}+s^{\check{}}\,(1+\omega)\,\acute{P}_{1,0}\\+2s^{\check{}}\delta_0\acute{P}_{1,2}+2\varphi\delta_0\acute{P}_{0,2}\end{gathered}\tag{16}$$

Now, we use (13) put $j=2$, and we obtain

$$\acute{P}_{1,2}=\left(\frac{\lambda\,(1+\omega)+2\varphi\delta_0}{2\mu}\right)\acute{P}_{0,2}=\frac{b_1}{c_1}\acute{P}_{0,2}$$

Additionally, using (1) and (15), we can solve (16) to obtain

$$\acute{P}_{0,2}=\frac{b_3}{c_3}\acute{P}_{0,0}\tag{17}$$

$$b_3=\frac{\begin{gathered}s^{\check{}}\,(1+\omega)\,a_1\;b_2+\varphi\,(1-s)\,a_1\;b_2\\+\mu a_1\;b_2-\lambda\,(1+\omega)\,b_2\end{gathered}}{c_2}$$

$$\frac{-(\lambda\,(1+\omega))^2(1+s)}{\mu}$$

$$c_3=2s^{\check{}}\,(1+\omega)\,\delta_0\;b_1/c_1+2\delta_0\varphi$$

When you enter j=2 in equations 6, 9, and 12, we get

$$\begin{gathered}(1-\Delta_0+\nu_1+2\mu+2\,(1-s)\,\varphi+s^{\check{}}\,(1+\omega))\,\acute{P}_{1,2}\\=\mu\acute{P}_{2,2}+\nu_2\acute{P}_{3,2}+\lambda\,(1+\omega)\,\acute{P}_{0,2}\\+s\lambda\,(1+\omega)\,\acute{P}_{1,1}+3s^{\check{}}\,(1+\omega)\,\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\ldots\end{gathered}\tag{18}$$

$$\mu\acute{P}_{2,2}=(1-\Delta_0)\,\acute{P}_{1,2}+\lambda\,(1+\omega)\,\acute{P}_{1,1}\tag{19}$$

$$\nu_2\acute{P}_{3,2}=\nu_1\acute{P}_{1,2}\tag{20}$$

If we use (19) & (20) in (18), we have

$$\begin{gathered}(2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\acute{P}_{1,2}\\=\lambda(1+\omega)(1+s)\acute{P}_{1,1}+\lambda(1+\omega)\acute{P}_{0,2}\\+3s^{\backsim}(1+\omega)\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\end{gathered}\tag{21}$$

$$\begin{gathered}(2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}\acute{P}_{0,2}\\=\lambda(1+\omega)(1+s)\acute{P}_{1,1}+\lambda(1+\omega)\acute{P}_{0,2}\\+3s^{\backsim}(1+\omega)\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\end{gathered}$$

$$\begin{gathered}\left((2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}-\lambda(1+\omega)\right)\acute{P}_{0,2}\\=\lambda(1+\omega)(1+s)\acute{P}_{1,1}+3s^{\backsim}(1+\omega)\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\end{gathered}$$

$$\begin{gathered}\left((2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}-\lambda(1+\omega)\right)\frac{b_3}{c_3}\acute{P}_{0,0}\\=\lambda(1+\omega)(1+s)a_1\acute{P}_{0,1}+3s^{\backsim}(1+\omega)\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\end{gathered}$$

$$\begin{gathered}\left((2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}-\lambda(1+\omega)\right)\frac{b_3}{c_3}\acute{P}_{0,0}\\=\lambda(1+\omega)(1+s)a_1\frac{b_2}{c_2}\acute{P}_{0,0}+3s^{\backsim}(1+\omega)\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\end{gathered}$$

$$\begin{gathered}\left(\begin{gathered}\left\{(2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}-\lambda(1+\omega)\right\}\frac{b_3}{c_3}\\-\lambda(1+\omega)(1+s)a_1\frac{b_2}{c_2}\end{gathered}\right)\acute{P}_{0,0}\\=3s^{\backsim}(1+\omega)\varphi\acute{P}_{1,3}+3\varphi\delta_0\acute{P}_{0,3}\end{gathered}$$

If, we Put $j=3$ in (13), and we have

$$\acute{P}_{1,3}=\left(\frac{\lambda(1+\omega)+j'\delta_0}{j^{-}}\right)\acute{P}_{0,3}=\frac{b_4}{c_4}\acute{P}_{0,3}$$

$$\begin{gathered}(\left\{(2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}-\lambda(1+\omega)\right\}\frac{b_3}{c_3}\\-\lambda(1+\omega)(1+s)a_1\frac{b_2}{c_2}\\=\left(3s'\frac{b_4}{c_4}+3\varphi\delta_0\right)\acute{P}_{0,3}\end{gathered}$$

$$\acute{P}_{0,3}=\frac{\left(\left\{(2*\mu+2(1-s)\varphi+s^{\backsim}(1+\omega))\frac{b_1}{c_1}-\lambda(1+\omega)\right\}\frac{b_3}{c_3}-\lambda(1+\omega)(1+s)a_1\frac{b_2}{c_2}\right)}{\left(\frac{3\lambda(1+\omega)\varphi\,b_4}{c_4}+3\varphi\delta_0\right)}\acute{P}_{0,0}$$

$$\acute{P}_{0,3}=\frac{b_5}{c_5}\acute{P}_{0,0}$$

In general, we get

$$\acute{P}_{0,n}=\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\tag{22}$$

If we use (13), we get

$$\acute{P}_{1,j}=\left(\frac{\lambda(1+\omega)+j\psi\delta_0}{j\mu}\right)(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots)\tag{23}$$

If, we use (9)

$$\acute{P}_{2,j} = (1-\Delta_0)\left(\frac{\lambda(1+\omega)+j\Sigma\delta_0}{j*\mu^2}\right)$$
$$\left(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\right)+\frac{\lambda(1+\omega)\acute{P}_{1,j-1}}{\mu} \tag{24}$$

Similarly, if, we use (12), we get

$$\acute{P}_{3,j} = \left(\frac{\nu_1}{\nu_2}\right)\left(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\right) \tag{25}$$

Now, if we use equations (22), (23), (24), and (25), we obtain

$$\acute{P}_{r,n} = \acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots \text{for} s=0\left(\frac{\lambda(1+\omega)+j'\delta_0}{j}\right)\left(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\right),$$
$$\text{for} s=1\,(1-\Delta_0)\left(\frac{\lambda(1+\omega)+j'\delta_0}{j*\mu^2}\right)\left(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\right)+\frac{\lambda(1+\omega)\acute{P}_{1,j-1}}{\mu},$$
$$\text{for} s=2,\left(\frac{\nu_1}{\nu_2}\right)\left(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\right),\text{for} s=3.$$

To calculate the value of $\acute{P}_{0,0}$, the normalization function, we have

$$\sum_{s=0}^{3}\sum_{j=0}^{\infty}P_{i,s}=1$$
$$\left(\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}\right)\left(1+\frac{\lambda(1+\omega)+j\varphi\delta_0}{j*\mu^2}+\frac{\nu_1}{\nu_2}\right)$$
$$\frac{+\lambda(1+\omega)}{\mu}\acute{P}_{1,j-1}=1$$
$$\acute{P}_{0,0}\left(\frac{b_2}{c_2}+\frac{b_3}{c_3}+\frac{b_5}{c_5}\right)\left(1+\frac{\lambda(1+\omega)+j\varphi\delta_0}{j*\mu^2}+\frac{\nu_1}{\nu_2}\right)$$
$$\frac{+\lambda(1+\omega)}{\mu}P_{1,j-1}=1$$
$$\acute{P}_{0,0}=\frac{1-\frac{\lambda(1+\omega)}{\mu}\acute{P}_{1,j-1}}{\left(\frac{b_2}{c_2}+\frac{b_3}{c_3}+\frac{b_5}{c_5}\right)\left(1+\frac{\lambda(1+\omega)+j'\delta_0}{j*\mu^2}+\frac{\nu_1}{\nu_2}\right)}$$

4. Validation of the model

(i) When the server is free:

$$\acute{P}_0 = \sum_{j=0}^{\infty}\acute{P}_{0,j}$$
$$=\frac{1-\frac{\lambda(1+\omega)}{\mu}\acute{P}_{1,j-1}}{\left(1+\frac{\lambda(1+\omega)+j'\delta_0}{j*\mu^2}+\frac{v_1}{v_2}\right)\left(\frac{b_2}{c_2}+\frac{b_3}{c_3}+\frac{b_5}{c_5}\right)}$$
$$\left[\acute{P}_{0,1}+\acute{P}_{0,2}+\acute{P}_{0,3}+\cdots\right]$$

(ii) When the server is busy:

$$\acute{P}_1 = \sum_{j=0}^{\infty}\acute{P}_{1,j}$$

$$= \frac{1 - \frac{\lambda(1+\omega)}{\mu} \acute{P}_{1,j-1}}{\left(1 + \frac{\lambda(1+\omega)+j'\delta_0}{j*\mu^2} + \frac{v_1}{v_2}\right)\left(\frac{b_2}{c_2} + \frac{b_3}{c_3} + \frac{b_5}{c_5}\right)}$$

$$\left(\frac{\lambda\,(1+\omega) + j'\delta_0}{j*\mu}\right) \times \left[\acute{P}_{0,1} + \acute{P}_{0,2} + \acute{P}_{0,3} + \cdots\right]$$

(iii) When the server is on vacation:

$$\acute{P}_2 \quad = \sum_{j=0}^{\infty} P_{2,j}$$

$$= \frac{1 - \frac{\lambda(1+\omega)}{\mu} \acute{P}_{1,j-1}}{\left(1 + \frac{\lambda(1+\omega)+j'\delta_0}{j*\mu^2} + \frac{v_1}{v_2}\right)\left(\frac{b_2}{c_2} + \frac{b_3}{c_3} + \frac{b_5}{c_5}\right)}$$

$$\times \left[(1 - \Delta_0)\left(\frac{\lambda\,(1+\omega) + j'\delta_0}{j*\mu^2}\right)\right]$$

$$\left(\acute{P}_{0,1} + \acute{P}_{0,2} + \acute{P}_{0,3} + \cdots\right) + \frac{\lambda\,(1+\omega)\,\acute{P}_{1,j-1}}{\mu}$$

(iv) When the server is in a breakdown and repaired state:

$$\acute{P}_3 = \sum_{j=0}^{\infty} P_{3,j}$$

$$= \frac{1 - \frac{\lambda(1+\omega)}{\mu} \acute{P}_{1,j-1}}{\left(1 + \frac{\lambda(1+\omega)+j'\delta_0}{j*\mu^2} + \frac{v_1}{v_2}\right)\left(\frac{b_2}{c_2} + \frac{b_3}{c_3} + \frac{b_5}{c_5}\right)} \left(\frac{v_1}{v_2}\right)$$

$$\times \left(\acute{P}_{0,1} + \acute{P}_{0,2} + \acute{P}_{0,3} + \cdots\right)$$

5. Conclusion

The Markovian Encouraged Arrival Queuing Model has been developed with the inclusion of customer retry efforts, balking, and reneging behavior. The four system states idle state, busy state, vacation state, breakdown state, and repair state have all been taken into consideration utilizing the concept of encouraged arrival. We have examined and verified the possibilities of the various conditions. Neural networks, communication systems, post offices, and supermarkets can all benefit from using this model to reduce the reneging and balking behavior of their customers.

References

[1] Kalyanaraman, R., & Sundaramoorthy, A. (2019) A Markovian Working Vacation Queue with Server State Dependent Arrival Rate and with Partial Breakdown. International Journal of Recent Technology and Engineering. 7(6s2), 664-668.

[2] Haridass M., & Arumuganathan R. (2008) Analysis of a Bulk Queue with Unreliable Server and Single Vacation. International Journal Open Problems Computational Mathematics. 1(2), 37-55.

[3] Daw, A., & Pender, J. (2019) On the Distributions of Infinite Server Queues with Batch Arrivals. Queueing Systems, 91, 367-401.

[4] Bar-Lev, S.K., Parlar, M., Perry, D., Stadje, W., & Van Der Duyn Schouten, F.A. (2007) Applications of Bulk Queues to Group Testing Models with Incomplete Identification. European journal of operational research 183, 226-237.

[5] Neuts, M.F. (1967) A General Class of Bulk Queue with Poisson Input. Annals of Mathematical Statistics 38(3), 759-770.

[6] Parveen, M.J, & Begum, M.I.A. (2013) General Bulk Service Queueing System with Multiple Working Vacation. International Journal of Mathematics Trends and Technology 4(9), 163-173.

[7] Li, H., & Zhao, Y.Q. (2005) A Retrial Queue with a Constant Retrial Rate, Server Break Downs and Impatient Customers. Stochastic Models 21(2-3), 531-550.

[8] Kumar, J., & Shinde, V. (2018) Performance Evaluation Bulk Arrival and Bulk Service with Multi Server Using Queue Model. International Journal of Research in Advent Technology. 6, 3069-3076.

[9] Vijaya Laxmi, P., & Rajesh P,(2018) Variant Working Vacations on Batch Arrival Queue with Reneging and Server Breakdowns. East African Scholars Multidisciplinary Bulletin. 1(1), 7-20.

[10] Srivastava, R.K., Singh, S., & Singh, (2020) A Bulk Arrival Markovian Queueing System with Two Types of Services and Multiple Vacations. International Journal of Mathematics and Computer Research. 8(8), 2130-2136.

[11] Singh, S., & R.K. Srivastava, (2021) Markovian Queueing System for Bulk Arrival and Retrial Attempts with Multiple Vacation Policy. International Journal of Mathematics Trends and Technology ·.

[12] Som, B.K., & Seth, S., (2017) An M/M/1/N queuing system with encouraged arrivals.Global journal and pure applied mathematics vol 13,no7.

[13] Khan, I. E., & Paramasivam, R. (2022). Reduction in Waiting Time in an M/M/1/N Encouraged Arrival Queue with Feedback, Balking and Maintaining of Reneged Customers. Symmetry,14(8), 1743.

[14] Som, B.K., (2018) M/M/c/N queuing systems with encouraged arrivals, reneging, retention, and Feedback customers. Yugoslav journal of operation research, 28(3), 333-344

RT&A, No 4(80)
Volume 19, December, 2024

EVALUATION OF REPETITIVE DEFERRED SAMPLING PLAN FOR TRUNCATED LIFE TESTS BASED ON PERCENTILES USING KUMARASWAMY EXPONENTIATED RAYLEIGH DISTRIBUTION

Neena Krishna P. K. Jayalakshmi. S

•

Department of Statistics, BJCSI College,Mulayara Thiruvananthapuram, Kerala.
Department of Statistics, Bharathiar University,Coimbatore.
neenakrishnapk94@gmail.com, statjayalakshmi16@gmail.com

Abstract

This paper focuses on the designing of the Repetitive Deferred sampling plan for truncated life test for percentiles using Kumaraswamy Exponentiated Rayleigh distribution. A truncated life test may be conducted to evaluate the smallest sample size to insure certain percentile life time of products. The main objective of the proposed sampling plan is to minimize the sample size because the analogous inspection time and inspection cost will be reduced. The operating characteristic function values are calculated according to various quality levels and the minimum ratios of the true average life to the specified average life at the specified producer's risk are derived. Certain real life examples are provided.

Keywords: Kumaraswamy Exponentiated Rayleigh Distribution, Repetitive Deferred Sampling Plan (RDS), Percentiles, life test, Producer's risk.

1. Introduction

A most commonly used technique in quality control is the Acceptance Sampling Plan. The acceptance sampling plans relates with the acceptance or rejection of a large-sized lot of products on the basis of quality of the products in a sample taken from the lot. Reliability sampling plans are the inspection techniques which are embraced for taking decisions on the disposition of the lot of an item based on assessment of the quality using the lifetimes of an items as quality variables. A life test is the process of estimating the life time of the product through experiments. A reliability sampling plan is also termed as the life test sampling plan for making decision about the disposition of lots based on the information obtained from a life test.

Many authors studied the designing of acceptance sampling plans based on the life test. Truncated life tests for the exponential distribution was first introduced by Epstein [4]. Further, truncated life tests were considered by many authors using various distributions. Gupta and Groll [5] developed the reliability acceptance sampling under the gamma distribution. Acceptance Sampling for the truncated life test based on the half logistic distribution was developed by Kantam and Rosaiah [8]. Baklizi and EI Masri [1] further designed reliability acceptance plan assuming the life time distribtion follows Birnbaum-Saunders distribution. Wu and Tsai [15] introduced the acceptance sampling truncated life test plan assuring mean lifetime under Birnbaum-Saunders distribution, which is outlined as an algorithm to obtain the plans. Balakrishnan et al. [2] developed the reliability acceptance sampling for generalized Birnbaum-Saunders distribution.

Percentiles bring more information about a life distribution than the mean life. When the life distribution is symmetric, the 50^{th} percentile or the median is equivalent to the mean life. Hence, developing acceptance sampling plans based on percentiles of a life distribution can be served as a generalization of the developing acceptance sampling plans based on the mean life of items. Lio .et.al [9] studied acceptance sampling for generalized Birnbaum-Saunders distribution using percentiles. Rosaiah et al. [13] developed an acceptance sampling procedure for the inverse Rayleigh distribution percentile under a truncated life test. Pradeepa Veera Kumari and Ponneswari [12] proposed the designing of acceptance sampling plan for life tests based on Percentiles of Exponentiated Rayleigh Distribution. Jayalakshmi and Neena Krishna [6] studied the designing of Special Type Double Sampling Plan for life tests based on percentiles using Exponentiated Frechet Distribution. Jayalakshmi and Vijilamery [7] developed Special Double Sampling Plan for truncated life tests based on percentile using Gompertz Frechet Distribution.

Lord Rayleigh [10] derived the Rayleigh distribution on the resultant of a large number of vibrations of the same pitch and of arbitrary phase. Kundu et al. [3] gave different methods and estimations of Generalized Rayleigh distribution. Nasr Ibrahim Rashwan [11] developed the Kumaraswamy Exponentiated Rayleigh Distribution. Shankar and Mohapatra [14] was first introduced the Repetitive Deferred Sampling Plan. This paper presents the idea of Repetitive Deferred Sampling Plans using Kumaraswamy Exponentiated Rayleigh Distribution, when the life test is truncated at a pre specified time. The main objective of the proposed sampling plan is to minimize the sample size because the analogous inspection time and inspection cost will be reduced.

2. Kumaraswamy Exponentiated Rayleigh Distribution

Nasr Ibrahim Rashwan(2016) defined the Cumulative Distribution Function(CDF) and Probability Density Function(PDF) of Kumaraswamy Exponentiated Rayleigh distribution.
The CDF of Kumaraswamy Exponentiated Rayleigh Distribution is given by

$$F(t,\lambda,\theta,a,b) = 1 - \left[1 - \left(1 - e^{-(\lambda t)^2}\right)\right)^{\theta a^b} \tag{1}$$

And the corresponding PDF is given by

$$f(t,\lambda,\theta,a,b) = 2ab\theta\lambda^2 t e^{-(\lambda t)^2}\left[1 - e^{-(\lambda t)^2}\right]^{\theta a-1}\left[1 - e^{-(\lambda t)^2}\right]^{b-1}, t,\lambda,\theta,a,b > 0 \tag{2}$$

Where λ is the scale parameter and θ, a, b are the shape parameters.
The hazard function of the distribution is given by

$$H(t) = \frac{f(t)}{1 - F(t)} \tag{3}$$

The percentile or the q^{th} quantile of any distribution is given by,

$$p_r\left(T \leq_{tq}\right) = q \tag{4}$$

$$t_q = \frac{1}{\lambda}\left[-\ln\left[1 - \left(1 - (1-q)^{1/b}\right)^{1/\theta a}\right]\right]^{1/2} \tag{5}$$

t_q and q are directly proportional. Let

$$\eta_q = \left[-\ln\left[1 - \left(1 - (1-q)^{1/b}\right)^{1/\theta a}\right]\right]^{1/2} \tag{6}$$

By changing the scale parameter $\lambda = \frac{\eta_q}{t_q}$, then the cumulative distribution function can be written in the form

$$F(t) = 1 - \left[1 - \left(1 - e^{-\left(\eta_q t/tq\right)^2}\right)^{\theta a}\right]^b \tag{7}$$

Let, $\delta_q = t/tq$

$$F(t) = 1 - \left[1 - \left(1 - e^{-\left(\eta_q \delta q\right)^2}\right)^{\theta a}\right]^b \tag{8}$$

Taking partial derivative with respect to δ, we have

$$\frac{\partial F(t)}{\partial \delta} = b\left[1 - \left(1 - e^{-\left(\eta_q \delta_q\right)^2}\right)^{\theta a}\right]^{b-1} \theta a \left(1 - e^{-\left(\eta_q \delta_q\right)^2}\right)^2 \theta a - 1_e^{-\left(\eta_q \delta_q\right)^2}{}_{2\eta_q \delta_q} \tag{9}$$

Since $\frac{\partial F}{\partial \delta_q} > 0, F\left(t, \delta_q\right)$ is a non-decreasing function of δ_q.

3. Formation of Repetitive Deferred Sampling Plan Using Kumaraswamy Exponentiated Rayleigh Distribution

The Repetitive Deferred Sampling Plan is represented as $(n, c_1, c_2, i, t/t_q^0)$ Here, the sample size be denoted as n, $c_j, j = 1, 2$ represents the acceptance number and i denotes the preceeding or succeeding lots can be taken. In life testing study, the test that will terminate at a pre-determined time t. The probability of rejecting a bad lot is P* and the maximum number of allowable defectives are c_1 and c_2. The RDS plan for percentiles is considered to obtain the minimum sample size "n" for the specified acceptance numbers c_1 and c_2 such that the consumer's risk (probability of accepting the bad lot) does not exceed $1 - P^*$.

A bad lot means that the true 100q^{th} percentile t_q^0 is below the specified percentile. Hence, the probability P^* is a minimum confidence level in the sense that lot of true average life below the specified life is rejected by the proposed sampling plan.

3.1. Operating Procedure

Shankar and Mohapatra (1991) developed Repetitive Deferred Sampling Plan indexed through producer and consumer quality levels. The operating characteristic function of the Repetitive Deferred Sampling plan under the truncated life test can be given as follows:

- Select a random sample of n units and count the number of defectives, then put on the test for pre-assigned experimental time t_0.
- Accept the lot if $d \leq c_1$,
- Reject the lot if $d > c_2$.
- If $c_1 < d < c_2$, if , accept the current lot provided
  - i Immediately preceding i lots are accepted in the case of deferred sampling or
  - ii Succeeding i lots are accepted in the case of deferred sampling plan.

3.2. Operating Characteristic Function

Let us represent the Repetitive Deferred Sampling Plan under Kumaraswamy Exponentiated Rayleigh Distribution as $n, c_1, c_2, i, \frac{t}{t_q^0}, j = 1, 2$. Here, the sample size be denoted as $n, c_j, j = 1, 2$ represents the acceptance number and i denotes the preceeding or succeeding lots.

For the proposed Repetitive Deferred Sampling Plan, the probability of acceptance of lot is given by

$$L(p) = \frac{p_a \left(1 - p_C\right)^i + p_c p_a^i}{\left(1 - p_C\right)^i} \tag{10}$$

$$\text{where, } p_a = p\left(d \leq_{c1}\right) = \sum_{x=0}^{c1} \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}$$

$$p_c = p\left(_{c1} < d <_{c2}\right) = \sum_{x=0}^{c2} \left(\begin{array}{c} n \\ x \end{array} \right)^x (1-p)^{n-x} - \sum_{x=0}^{c1} \left(\begin{array}{c} n \\ x \end{array} \right)^x (1-p)^{n-x}$$

Where, p is the failure probability before the time t, given a specified $100q$ th percentile life time ${t_q}^0$, is obtained from the equation $p = F(t,\delta) = 1 - \left[1 - \left(1 - e^{-\left(\eta_q \delta q\right)^2}\right)^{\theta a}\right]^b$

$$\text{Where } F(t:\delta) \leq F\left(t, \delta_q\right) \Leftrightarrow t_q \geq t_q^0.$$

3.3. Minimum Sample Size

In Repetitive Deferred Sampling Plan, to determine the minimum sample size 'n' for the known $P^*, c_1, c_2, i, \frac{t}{t_q^0}$ should be satisfy the following condition,

$$L(p) \leq 1 - P^* \tag{11}$$

Here, $L(p)$ is taken from the equation (10) and where $P^* = 0.99, 0.95, 0.90$ and 0.75 is the probability of rejecting the bad lot. Thus the smallest sample size ' n ' can be simulated using the search procedure for various values of P^*, c_1, c_2, i and $\frac{t}{t_q^0}$ are calculated. Since $\frac{\partial F\left(t,\delta_q\right)}{\partial \delta_q} > 0, F\left(t, \delta_q\right)$ is a non-decreasing function of δ_q with respect to the time t. So $F\left(t, \delta_q\right) \leq F\left(t, \delta_q^0\right) \Leftrightarrow \boldsymbol{t}_q \geq \boldsymbol{t}_q^0$ or equivalently $F\left(t, \delta_q\right) \leq F\left(t, \delta_q^0\right) \Leftrightarrow \delta_q < \delta_q^0$.

4. Numerical Example

Suppose a quality engineer examines the lifetime of an Air Conditioner and the engineer works with the Repetitive Deferred Sampling Plan for the lifetime of the product. Assume that the lifetime of the product follows Kumaraswamy Exponentiated Rayleigh distribution with $\theta = 2, a = 1, b = 0.5, \alpha = 0.05, \beta = 0.10$. The engineer is interested to adopt the sampling plan to insure 10^{th} percentile that the lifetime is at least 650 hours with the confidence level P^* =0.99. The experimenter wants to breaks the experiment at t = 1300 hours. From table 2.3.1,one can obtain the required sample size corresponding to the values of $P^* = 0.99$, $\frac{t}{t_q^0} = 2.00$ and $c_1 = 0, c_2 = 2, i = 1$ is $n = 6$. Thus we have to put the test upto 6 units. The corresponding operating characteristic value $L(p)$ with a confidence level 0.99 for the Repetitive Deferred Sampling Plan $(n = 6, c_1 = 0, c_2 = 2, i = 1, \frac{t}{t_q^0} = 2.00)$ under Kumaraswamy Exponentiated Rayleigh Distribution from Table 2 is given as,

$\frac{t_q}{t_q^0}$	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75
$L(P)$	0.00922	0.59769	0.88879	0.97103	0.99164	0.99730	0.99877	0.99902

This shows that the true 10^{th} percentile is equal to the required 10^{th} percentile ($\frac{t_q}{t_q^0}$ =1.00),the producer's risk is approximately 0.99078 (1-0.00922). The producer's risk is almost equal to 0.05 or less when the actual 10^{th} percentile is greater than or approximately equal to 2.00 times the

specified 10^{th} percentile. Table 3 gives the values of $d_{0.10}$ for $c_1 = 0$ and $c_2 = 2, i = 1$ and $\frac{t_q}{t_q^0}$ to guarantee that the producer's risk is less than or equals 0.05. In this illustration, the corresponding value of $d_{0.10}$ is 1.93421 for $c_1 = 0, c_2 = 2; t/t_{0.10} = 2.00$ and $\alpha = 0.05$. This means that the product can have a 10^{th} percentile life of 1.93421 times the required 10^{th} percentile lifetime under the above Repetitive Deferred Sampling Plan the product is accepted with probability of at least 0.95. Figure

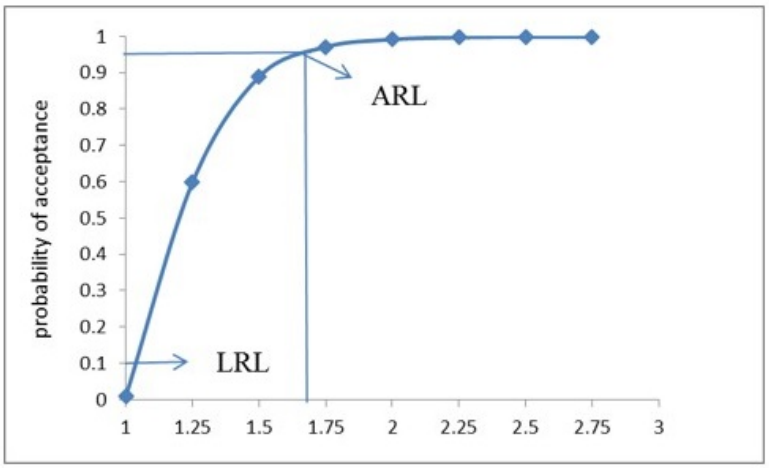


Figure 1: *Operating Characteristic Curve of RDS Plan for 10th Percentile under Kumaraswamy Exponentiated Rayleigh Distribution*

1 represents the operating characteristic curve of RDS plan under Kumaraswamy Exponentiated Rayleigh Distribution for the 10^{th} percentile $(n = 6, c_1 = 0, c_2 = 2, i = 1, t/t_{0.10} = 2.00)$. From figure 1, it is clear that the Repetitive Deferred Sampling Plan attains ARL when the actual 10^{th} life time percentile is 2.00 times greater than required 10^{th} percentile and attains LRL when the actual 10^{th} life time percentile is 1.00 times greater than required 10^{th} percentile.

Table 1: *Minimum Sample Size values 'n'for the* 10^{th} *percentile of Repetitive Deferred Sampling Plan under Kumaraswamy Exponentiated Rayleigh Distribution when* $\theta = 2, a = 1, b = 0.5$

P*	i	c_1	c_2	$\frac{t}{t_q^0}$							
				0.80	0.85	0.90	1.00	1.50	2.00	2.50	3.00
0.99	1	0	1	97	79	65	48	13	6	4	3
		0	2	97	79	65	46	13	6	4	3
		1	2	136	113	97	67	19	9	6	4
		2	3	179	146	121	88	24	7	5	4
		3	4	209	169	140	101	29	14	9	7
	2	0	1	96	78	64	46	13	6	4	3
		0	2	94	76	63	44	13	6	4	3
		1	2	135	112	95	65	19	9	6	3
		2	3	176	142	117	85	24	7	5	4
		3	4	207	167	138	98	29	14	9	7
0.95	1	0	1	64	52	43	30	9	4	3	2
		0	2	69	55	46	32	11	5	3	3
		1	2	99	81	67	47	14	7	5	3
		2	3	131	107	88	62	19	10	6	5
		3	4	161	131	108	77	23	12	8	6
	2	0	1	64	52	43	30	9	4	3	2
		0	2	62	50	42	29	9	4	3	3
		1	2	97	79	65	46	14	7	4	3
		2	3	130	105	87	61	19	9	6	5
		3	4	160	129	107	76	23	12	8	6
0.90	1	0	1	52	42	35	25	7	4	2	2
		0	2	58	47	39	28	8	4	3	3
		1	2	84	68	56	40	12	6	4	3
		2	3	113	92	76	54	16	8	6	4
		3	4	141	114	94	67	21	11	7	6
	2	0	1	48	39	32	23	7	3	2	2
		0	2	50	41	34	24	7	4	3	3
		1	2	80	65	54	38	12	6	4	3
		2	3	110	89	74	52	16	8	5	4
		3	4	138	112	92	66	20	10	7	5
0.75	1	0	1	36	29	24	17	5	3	2	2
		0	2	44	36	30	21	7	4	3	3
		1	2	63	51	42	30	9	5	3	3
		2	3	88	71	59	42	13	7	5	4
		3	4	112	91	75	54	17	9	6	5
	2	0	1	32	26	21	15	5	3	2	2
		0	2	37	30	25	18	6	3	3	3
		1	2	58	47	39	28	9	5	3	3
		2	3	83	68	56	41	12	7	5	4
		3	4	108	88	72	52	16	9	6	5

Table 2: *Operating Characteristic value for the 10^{th} percentile of Repetitive Deferred Sampling Plan under the assumption of Kumaraswamy Exponentiated Rayleigh Distribution for $\theta = 2, a = 1, b = 0.5$, $c_1 = 0, c_2 = 2, i = 1$*

P*	n	$\frac{t}{t_q^0}$	$\frac{t_q}{t_q^0}$							
			1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75
0.99	97	0.8	0.00962	0.76963	0.95633	0.99083	0.99771	0.99934	0.99994	0.99998
	79	0.85	0.00941	0.75872	0.95334	0.99011	0.99752	0.99928	0.99991	0.99995
	65	0.9	0.00943	0.75034	0.95083	0.98948	0.99735	0.99923	0.99959	0.99987
	46	1	0.00913	0.72892	0.94438	0.98787	0.99691	0.99909	0.99912	0.99937
	13	1.5	0.00963	0.65500	0.91598	0.97982	0.99451	0.99831	0.99987	0.99992
	6	2	0.00922	0.59769	0.88879	0.97103	0.99164	0.99730	0.99877	0.99902
	4	2.5	0.00395	0.43613	0.80623	0.94470	0.98324	0.99439	0.99564	0.99875
	3	3	0.00173	0.32259	0.72661	0.91639	0.97384	0.99101	0.99354	0.99547
0.99	69	0.8	0.04819	0.52216	0.91156	0.98477	0.99681	0.99920	0.99968	0.99997
	55	0.85	0.05197	0.53085	0.91277	0.98482	0.99679	0.99919	0.99954	0.99978
	46	0.9	0.04871	0.51023	0.90502	0.98321	0.99642	0.99909	0.99925	0.99945
	32	1	0.05142	0.50819	0.90139	0.98213	0.99613	0.99900	0.99921	0.99938
	11	1.5	0.02296	0.29891	0.77796	0.95137	0.98849	0.99686	0.99885	0.99955
	5	2	0.02571	0.29724	0.76262	0.94358	0.98571	0.99588	0.99774	0.99798
	3	2.5	0.02802	0.32259	0.78254	0.94798	0.98638	0.99592	0.99755	0.99787
	3	3	0.00173	0.04774	0.32259	0.72661	0.91639	0.97384	0.98458	0.98652
0.90	58	0.8	0.09622	0.66684	0.94771	0.99114	0.99814	0.99953	0.99978	0.99991
	47	0.85	0.09692	0.66223	0.94588	0.99072	0.99803	0.99950	0.99965	0.99987
	39	0.9	0.09440	0.65072	0.94247	0.99001	0.99787	0.99945	0.99955	0.99975
	28	1	0.08826	0.62454	0.93441	0.98831	0.99746	0.99934	0.99945	0.99965
	8	1.5	0.09724	0.60083	0.91870	0.98378	0.99618	0.99895	0.99912	0.99954
	4	2	0.08272	0.54909	0.89646	0.97763	0.99439	0.99837	0.99901	0.99921
	3	2.5	0.02802	0.32259	0.78254	0.94798	0.98638	0.99592	0.99874	0.99892
	3	3	0.00173	0.04774	0.32259	0.72661	0.91639	0.97384	0.98456	0.98974
0.75	44	0.8	0.24151	0.83647	0.97783	0.99626	0.99921	0.99980	0.99989	0.99992
	36	0.85	0.23688	0.82978	0.97643	0.99598	0.99914	0.99978	0.99989	0.99991
	30	0.9	0.22993	0.82123	0.97469	0.99564	0.99906	0.99976	0.99989	0.99990
	21	1	0.23734	0.81955	0.97360	0.99534	0.99898	0.99973	0.99985	0.99989
	7	1.5	0.16445	0.71826	0.94902	0.98997	0.99763	0.99935	0.99954	0.99987
	4	2	0.08272	0.54909	0.89646	0.97763	0.99439	0.99837	0.99885	0.99921
	3	2.5	0.02802	0.32259	0.78254	0.94798	0.98638	0.99592	0.99745	0.99865
	3	3	0.00173	0.04774	0.32259	0.72661	0.91639	0.97384	0.98254	0.98578

Table 3: *The ratio $d_{0.10}$ for accepting the lot with the producer's risk of 0.05 when $\theta = 2, a = 1, b = 0.5$*

P*	i	c_1	c_2	$\frac{t}{t_q^0}$							
				0.80	0.85	0.90	1.00	1.50	2.00	2.50	3.00
0.99	1	0	1	2.14911	2.16170	2.16758	2.20685	2.35443	2.49349	2.73476	2.96066
		0	2	1.76254	1.77179	1.78171	1.80171	1.87442	1.93421	2.06504	2.15984
		1	2	1.84268	1.85210	1.86341	1.88323	2.00143	2.10385	2.28006	2.30653
		2	3	1.69322	1.70460	1.71002	1.73359	1.82091	1.97192	1.66703	1.74091
		3	4	1.60473	1.60901	1.61803	1.63569	1.71425	1.78578	1.87400	1.98824
	2	0	1	2.18777	2.20348	2.21590	2.24439	2.41161	2.55693	2.81052	3.05413
		0	2	1.79778	1.80595	1.81777	1.83256	1.93704	2.00521	2.14501	2.25000
		1	2	1.85840	1.87156	1.88244	1.90351	2.02437	2.13583	2.31224	2.42538
		2	3	1.70666	1.71258	1.72297	1.73652	1.83405	1.98919	2.18927	2.36513
		3	4	1.61016	1.62090	1.62673	1.63886	2.29936	2.49900	2.88681	3.00250
0.95	1	0	1	1.93524	1.94877	1.96188	1.97880	2.11514	2.18779	2.46423	2.51298
		0	2	1.60939	1.60587	1.61867	1.62820	1.78207	1.80969	1.97759	2.15984
		1	2	1.69354	1.70295	1.71634	1.72783	1.82410	1.92950	2.11791	2.38377
		2	3	1.57579	1.58263	1.58879	1.60337	1.69220	1.81089	1.83545	2.00250
		3	4	1.50111	1.50625	1.51297	1.52802	1.59270	1.68852	1.77204	1.82271
	2	0	1	1.97824	1.99251	2.00767	2.02425	2.16669	2.24788	2.54396	2.60143
		0	2	1.60919	1.61264	1.62608	1.63119	1.72467	1.74510	1.87613	2.25001
		1	2	1.70172	1.71237	1.71755	1.73616	1.84588	1.95395	1.99628	2.02538
		2	3	1.57862	1.58634	1.59478	1.60693	1.70478	1.75715	1.85193	2.02947
		3	4	1.50172	1.50762	1.51632	1.52946	2.13344	2.29875	2.78604	2.83418
0.90	1	0	1	1.83432	1.84119	1.85660	1.88370	1.95998	2.18779	2.29473	2.51298
		0	2	1.53381	1.53780	1.54805	1.56558	1.60653	1.65345	1.79759	2.15984
		1	2	1.61879	1.62752	1.63153	1.65387	1.73680	1.82329	1.92689	1.98377
		2	3	1.51205	1.51879	1.52699	1.54081	1.60054	1.66208	1.83545	1.94091
		3	4	1.44662	1.45026	1.45627	1.46985	1.54535	1.62897	1.65824	1.82271
	2	0	1	1.83703	1.84768	1.85509	1.88492	2.01392	2.03596	2.16784	2.60143
		0	2	1.51745	1.52552	1.53482	1.54489	1.58856	1.71510	1.87613	2.25001
		1	2	1.61695	1.62202	1.63318	.64842	1.75733	1.84732	1.95628	2.02538
		2	3	1.51228	1.51561	1.52606	1.53498	1.61286	1.68053	1.68927	1.76513
		3	4	1.44564	1.45019	1.45457	1.47070	2.03666	2.17876	2.26193	2.61080
0.75	1	0	1	1.66288	1.66714	1.67397	1.69320	1.76599	1.97404	2.09473	2.51298
		0	2	1.42197	1.43068	1.43966	1.44238	1.53317	1.65345	1.79759	2.15984
		1	2	1.50037	1.50308	1.50922	1.52493	1.57901	1.69845	1.72046	1.98377
		2	3	1.41260	1.41566	1.42327	1.43368	1.49288	1.57192	1.66703	1.74091
		3	4	1.35853	1.36143	1.36425	1.38090	1.43839	1.49635	1.51606	1.58734
	2	0	1	1.64897	1.65667	1.65257	1.67466	1.81167	2.03596	2.16784	2.60143
		0	2	1.39634	1.40072	1.40874	1.42238	1.50587	1.54877	1.87613	2.25001
		1	2	1.48019	1.48605	1.49501	1.51223	1.60207	1.72367	1.78685	2.02538
		2	3	1.39977	1.40836	1.40980	1.43346	1.46418	1.58919	1.68927	1.76513
		3	4	1.35150	1.35537	1.35564	1.37274	1.88807	1.91080	1.93254	1.96080

5. Conclusion

This study establishes the designing of Repetitive Deferred sampling plan for the truncated life test when the life time of the product follows Kumaraswamy Exponentiated Rayleigh Distribution. The work designed and developed with the aim that sampling plan in this paper may helpful for the engineers and statistical plan developers in the field of statistical quality control especially in attribute reliability sampling plan. The results from the sampling plan developed in this article can easily adaptable in practical situation with small sample sizes and fewer experimental times and hence the plan yields a better result for reliability sampling plan. The tables are generated which is useful for both producer and consumer. This plan can highly recommendable since the products are randomly sampled and also the procedure for sampling involves less experiment time and with minimum sample size.

References

[1] Baklizi, A, El Masri, A.E.K.(2004). Acceptance sampling plans based on truncated life tests in the Birnbaum Saunders model. *Risk Analysis*, 24:453-457.

[2] Balakrishnan, N., Leiva, V., Lopez, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum-Saunders distribution. *Communications in statistics - Simulation and Computation*, 36:643-656.

[3] Debasis Kundu, Mohammed, Raqab, Z. (2005). Generalized Rayleigh distribution: Different methods and estimations. *Computational Statistics and Data Analysis*.

[4] Epstein, B. (1954). Truncated life tests in the exponential case. *Annals of Mathematical Statistics*, 25:555-564.

[5] Gupta, S.S., Groll, P.A. (1962). Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association*, 56:942-970.

[6] Jayalakshmi, S., Neena Krishna, P.K. (2021). Designing of Special Type Double Sampling Plan for Life Tests Based on Percentiles Using Exponentiated Frechet Distribution. *Reliability Theory & Applications*, 16(1):117-123.

[7] Jayalakshmi, S., Vijilamery, S. (2022). Designing of Special Type Double Sampling Plan for Truncate Life Test using Gompertz Frechet Distribution. *International Journal of Mathematics Trends and Technology*, 6870-76.

[8] Kantam, R. R. L., Rosaiah, K. (1998). Half Logistic distribution in acceptance sampling based on life tests. *IAPQR Transactions*, 23(2):117-125.

[9] Lio. Y. L., Tsai, T. R., Wu, S.J. (2009). Acceptance sampling plans from truncated life tests based on the Birnbaum-Saunders distribution for percentiles. *Communications in Statistics-Simulation and Computation*, 39(1):119-136.

[10] Lord Rayleigh. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary Phase. *The London, Edinburg and Dublin philosophical magazine and Journal of Science*, 10:73-78.

[11] Nasr Ibrahim Rashwan. (2016). A Note on Kumaraswamy Exponentiated Rayleigh distribution, *Journal of Statistical Theory and Applications*, 15(3):286-295.

[12] Pradeepaveerakumari. K, Ponneeswari. P. (2016). Designing of Acceptance Sampling Plan for life tests based on Percentiles of Exponentiated Rayleigh Distribution. *International Journal of Current Engineering and Technology*, 6(4):1148-1153.

[13] Rao, G. S., Kantam, R. R. L., Rosaiah, K., Reddy, J. (2012). Acceptance sampling plans for percentiles based on the Inverse Rayleigh Distribution. *Electronic Journal of Applied Statistical Analysis*, 5(2):164-177.

[14] Shankar. G., Mahopatra. (1991). GERT Analysis of Repetitive Deferred Sampling Plans, *IAPQR Transactions*, 16(2)2:17-25.

[15] Tsai. T. R., Wu, S. J., (2006). Acceptance sampling based on truncated life tests for generalized Rayleigh distribution. *Journal of Applied Statistics*, 33:595-600.

RT&A, No 4(80)
Volume 19, December, 2024

SOLVING GENERALIZED FUZZY LEAST COST PATH PROBLEM OF SUPPLY CHAIN NETWORK

Pratibha[1,*], Rajesh Dangwal[2]

•

[1,2]Department of Mathematics, H.N.B. Garhwal University, BGR Campus,
Pauri Garhwal-246001, Uttarakhand, India
[1,*]pratibhakd2706@gmail.com, [2]dangwalrajeshhnbgu@gmail.com
[*]Corresponding author

Abstract

Optimal route selection for delivering product is the major concern for organizations related to supply chain management. The choice of route is crucial as it has a big impact on an organization's finances. In this research, an optimum solution with inaccurate and hazy parameters to a fuzzy least cost route issue is presented. Costs can be represented by time, distance or other criteria that could represent edge weights and these are defined by the user. In this paper we are using term cost as activity time. More specifically, the cost value is taken as Generalized hexagonal fuzzy numbers. The paper discusses optimal route selection problem to reduce distance-driven costs. By using ranking method optimal cost value obtained in form of crisp numbers. Also, for the validation of our result and obtained optimal cost in form of fuzzy number, we use fuzzy dynamic programming. We obtain an improved result using our ranking algorithm. Additionally, a comparison is provided. A numerical example for comparison analysis with previous publications is provided, utilising appropriate graphical layout and tables, to elucidate both approaches.

Keywords: Dynamic programming, Generalized hexagonal fuzzy numbers, Supply chain network, Fuzzy least cost path, Ranking Function.

1. Introduction

Managing supply chains with an integrated approach is essential for effectively planning and supervising the movement of products from suppliers to end users. A supply chain is an intricate web of interrelated business networks that includes a number of different organizations, such as distributors, suppliers, manufacturers, warehouses, and shipping companies. The goal is to guarantee that goods flow smoothly through the manufacturing, shipping, and distribution phases, arriving at consumers' locations on schedule and within budget. Modern organizations depend on their supply chains to succeed because they have a direct impact on the profitability, customer satisfaction, and operational efficiency of the organization. For example, a well-managed supply chain can boost product availability, cut down on delays, and lower production costs, all of which can raise sales. However, delays, lost money, and strained client relations can result from supply chain disruptions. Supply chain management has a big impact on corporate success, which is why experts and researchers are interested in learning more about it and improving it. Creating plans that streamline supply chain operations from inventory control and demand forecasting to customer service and transportation scheduling is the main goal.

Cost-effectiveness plays a vital role in today's business environment through process optimization, waste reduction, and better resource utilisation, effective supply chain networks contribute to cost savings. Improved customer satisfaction and loyalty are the result of responsive and dependable supply chain networks, which guarantee prompt delivery, product availability, and high-quality services. Due to shifting market conditions, uncertainity in lead times, and changing demand, supply chain networks are by their very nature unpredictable. Supply chain managers may create more durable and dependable solutions by modelling and optimising distribution networks, inventory flows, and transportation routes under erratic circumstances with FLCPP.

The goal of the least-cost path problem is to find the least expensive way or route between a given starting point and destination. Generally, least-cost path problems are represented using graphs. A set of vertices and edges makes up a mathematical structure called a graph. Vertices are points, while edges are the connections between pairs of vertices. These edges allow movement between vertices. Graphs are either directed or undirected depending on whether movement is allowed in both directions or only in one. If just one direction is permitted for movement along a graph's edges, then that graph is said to be directed. On the other hand, if movement is allowed along the edges in both directions, the graph is categorized as undirected.

A graph's edges are typically given weights in the least-cost path issue, which indicate how much it costs to traverse each edge. The objective is to determine the path that minimizes the overall cost between the beginning and destination vertices. Numerous industries, including logistics, transportation, and communication networks, face these issues. The least-cost path problem is typically solved using algorithms such as Dijkstra's and Bellman-Ford, which identify the shortest path between two places while minimizing the overall cost by allocating weights to the graph's edges.

The traditional least-cost path issue is extended by the fuzzy least-cost path problem (FLCPP), which aims to determine the best cost-effective path between two nodes in a network while taking uncertainty and imprecision into account. Fuzzy set theory is used in FLCPP to model uncertainty, in contrast to the traditional least-cost path problem, which assigns precise values to the costs of traversing arcs or nodes. Fuzzy numbers are employed in this framework to represent the costs, which reflect real-world unpredictability like variable trip charges, irregular travel schedules, or imprecise user preferences. A more flexible and reliable method of modeling complex scenarios is provided by fuzzy numbers, which are inaccurate but realistic representations of these uncertainties. Taking into account the inherent uncertainties in the system, FLCPP seeks to identify the fuzzy path with the maximum degree of satisfaction or the lowest predicted cost. In contrast, standard approaches miss the subtleties of uncertain settings since they depend on stable values. With FLCPP, pathways must be evaluated not only on direct costs but also on decreasing overall uncertainty while satisfying the decision-maker's preferences, which makes the decision-making process more dynamic. The ultimate resolution is to offer the optimal equilibrium between economical viability and minimizing ambiguity. Because unforeseen conditions frequently develop in these domains where flexible decision-making is essential for optimization FLCPP finds applications in a variety of fields, including supply chain management, transportation, and telecommunications.

In supply chain operations, efficient inventory management is essential for reducing holding costs and adjusting stock levels to changing demand. Fuzzy least cost path problem (FLCPP): This sophisticated method of supply chain optimization takes into account a variety of uncertainties, including inaccurate demand estimates, erratic lead times from suppliers, and variable costs associated with retaining inventories. Businesses can use FLCPP to optimize distribution routes and make better decisions about inventory replenishment, which will ultimately save costs and increase efficiency. One significant area of supply chain spending that FLCPP can help with is transportation. Businesses can determine the most cost-effective transportation routes with the use of FLCPP, which accounts for uncontrollable factors like weather, traffic, and fluctuations in fuel prices. As a result, transportation planning becomes more economical and effective, which is essential to sustaining profitability. In general, FLCPP gives companies a competitive edge and boosts productivity in a volatile market by helping them better handle uncertainty in supply chain

operations. Businesses can successfully handle difficult supply chain problems by implementing fuzzy logic-based optimization techniques like FLCPP, which promotes long-term growth, better resource allocation, and higher financial gains. Businesses can overcome uncertainty and optimize their operations in a global market that is highly competitive by using this strategic strategy.

Supply chain networks are complex networks of linked companies, resources, activities, and technology that make it easier to produce, distribute, and deliver goods and services from suppliers to end users. Suppliers who offer the necessary parts, raw materials, or services to enable manufacturing are the first in the process. These inputs are subsequently used by manufacturers to create the finished goods. Distributors handle the handling of the products' storage, shipping, and delivery to retailers after manufacturing is finished. Retailers offer the products to final customers through a variety of tactics include direct marketing, internet platforms, and physical storefronts. Eventually, people or businesses use or consume the products or services. From the first stage of production to the last stage of consumption, every link in this chain is essential to the smooth flow of products. Businesses are able to meet customer needs while optimizing costs, time, and resources thanks to this networked system.

In conclusion, supply chain networks are critical to modern businesses' ability to satisfy customer needs, keep costs under control, and effectively allocate resources. By streamlining processes and boosting responsiveness, efficient supply chain management gives businesses a competitive edge in today's linked and dynamic global marketplaces. Businesses can attain operational excellence, which guarantees timely delivery of goods and services while cutting costs, by optimizing the manufacturing, distribution, and delivery procedures. An effective supply chain also fosters business sustainability and growth, generating long-term value for all parties involved. In the end, a company's capacity to succeed and adapt in a more competitive business environment ultimately depends on its supply chains being well-managed.

Motivation of this Study

The path with the lowest cost for individuals going between a source and a destination is referred to as the least cost path. In a directed acyclic network, exact values are typically used as weights on the edges. But a number of circumstances including shifting traffic patterns, unfavorable weather patterns, and natural disasters make it challenging to determine these weights precisely. Under such circumstances, traditional techniques might not be sufficient to capture the uncertainties prevalent in real-world scenarios. This emphasizes the need for alternative methods that can account for imprecision and fluctuation in the weights, like fuzzy set theory. Decision-makers can improve route optimization by more skillfully assessing the most cost-effective routes under dynamic conditions by including these uncertainties.

Consider a scenario in which a transportation company is tasked with determining the least-cost route for delivering goods across various locations. The total cost for any given route is influenced by uncertain factors such as fluctuating fuel prices, varying toll fees, and unpredictable road conditions. Traditional least-cost path algorithms typically rely on precise, fixed values for these factors, which may not adequately reflect real-world conditions. For instance, fuel prices can change frequently and vary significantly from one state to another. Within the state of Himachal Pradesh, for example, the petrol prices differ across various cities: Una has a price of 93.20 rupees, Hamirpur 93.68 rupees, Solan 94.20 rupees, Nahan 94.47 rupees, Chamba 95.00 rupees, and Kullu 95.20 rupees. Given this variability, the fuel cost can be effectively represented using a hexagonal fuzzy number, specifically $\tilde{F} = (93, 93.50, 94, 94.50, 95, 95.50)$. This approach allows for a more flexible representation of fuel costs, accounting for the inherent uncertainties present in the transportation network.

Objective

The objective of this paper is to apply an existing ranking methodology to solve the fuzzy least-cost route selection problem in supply chain management, where uncertainties like time and cost are critical factors. By using generalized hexagonal fuzzy numbers to represent these

uncertain parameters, the study aims to better capture the imprecision inherent in real-world transportation scenarios. The goal is to convert fuzzy values into crisp numbers, enabling more accurate and informed decision-making for optimal route selection. The method is validated through fuzzy dynamic programming, ensuring that the results are robust and applicable across various uncertain conditions. A comparative analysis with previous methods is provided to highlight the effectiveness of the approach. Additionally, a numerical example demonstrates how this model can be practically applied, offering insights for organizations seeking to minimize costs and manage uncertainty in their supply chain operations.

Advantages of our result

1. A comparison is provided with existing approaches, highlighting the effectiveness of the proposed method.
2. The use of a ranking function allows decision-makers the flexibility to incorporate multiple criteria, enabling more informed decision-making.
3. Fuzzy ranking handles uncertainties and imprecise data better than deterministic approaches, enhancing the robustness of the solution against variations and errors in the input parameters.
4. To validate the results, a previously established dynamic programming technique has been applied.
5. The results of fuzzy ranking are often easier to interpret, providing insights into the relative importance of different factors, thus supporting transparent decision-making.
6. The methodology used in this paper is straightforward and can be easily applied in practice.
7. The outcome is represented in the form of generalized hexagonal fuzzy numbers, which allow for comparison through various defuzzification methods to obtain crisp values.

Contributions

- Introduced the use of generalized hexagonal fuzzy numbers to represent costs in route selection, enhancing flexibility and accuracy in cost estimation under uncertainty.
- Applied fuzzy dynamic programming to validate the results, ensuring the reliability and applicability of the optimal solutions in dynamic and uncertain environments.
- Conducted a comprehensive comparative analysis with previous publications, providing insights into the effectiveness of the proposed method through graphical representations and tables.
- Utilized numerical examples to illustrate the practical implications of the proposed approach

The format of this document is as follows: We have provided some fundamental definitions and introductions in Section 2. The generalised hexagonal fuzzy number ranking algorithm is defined in Section 3. The least-cost path problem and the ranking method's solution are described in Section 4 with the conventional fuzzy approach method for addressing dynamic programming. Furthermore, a numerical case was resolved using each approach. Section 5 provide the outcome. The comparison is presented in Sections 6 respectively. Section 7 provides the conclusion and future research scope.

Literature Review

Initial studies (1960s-1980s), Initial mathematical models for inventory control and transportation were the focus of supply chain optimisation research. To maximise choices about production and distribution, linear programming techniques were frequently applied. In 1970 Bellman et al. [1] intoduced fuzzy dynamic programming. In which the concept of fuzzy set theory was used in decision making problems.

Expanding the Scope (1990s): During this decade, there was a change in the focus of supply chain optimisation to include more facilities, decision points, and tiers. More complex models, such as mixed-integer programming and heuristic methods for handling larger-scale issues, were introduced at this time. Later, the applications of fuzzy dynamic programming have given by Kacprzyk. With the help of this Hussien et al. [2] solved multiple criteria resources allocation problem. For multistage problem Baldwin et al. [3] have worked on fuzzy dynamic programming.

The 2000s saw the integration of information technology: As the internet and information technology grew in popularity, supply chain optimisation studies began to use real-time data and sophisticated analytics. During this time, supply networks were encouraged to be transparent, cooperative, and responsive.

Within the framework of supply chain optimisation, sustainability and risk management have received increasing attention in the 2010s. Environmental factors, such reducing carbon footprints and implementing green logistics, have been included by researchers into optimisation models. The mitigation of risks related to interruptions, such as natural catastrophes and geopolitical events, was also emphasised.

Industry 4.0 and digitalization (2015-present): The introduction of Industry 4.0 technology and the digitization of supply networks define the contemporary age. These days, big data analytics, blockchain, artificial intelligence, and the Internet of Things (IoT) are all included in optimisation research. These technologies give supply chain operations more flexibility, more visibility, and real-time decision-making capabilities. Supply chain optimisation research has developed across these phases in response to novel problems and chances brought about by modifications in company practices, technological breakthroughs, and evolving consumer expectations. A article on addressing a fuzzy optimum subdivision problem utilising the fuzzy least cost route problem by employing the generalised trapezoidal number has been proposed by Nagalakshmi et al. [4].

Chang and Zadeh established the idea of fuzzy mapping and control. A multitude of techniques have been presented for ranking fuzzy numbers ([5], [6], [7]). Additionally, a crucial part of decision-making is the ranking of fuzzy numbers. To defuzzify the fuzzy numbers, several efforts have been undertaken ([8], [9], [10], [11]). According to Chen et al.[12], for regular fuzzy numbers, there is no requirement for the membership function. The idea of generalised fuzzy numbers was first out by him. In order to rank fuzzy numbers, Yager [13] employed the centroids technique. Using the modes, rank, divergence, and spread that Rajarajeshwari and Sudha [14] presented, we may order generalised hexagonal fuzzy numbers.

Many studies have taken into account the multi-objective aspect of FLCPP by include many competing elements in the optimisation model, including cost, travel time, and dependability. Hybrid techniques, which integrate many approaches to enhance the quality and effectiveness of solutions, have been suggested in numerous research. To incorporate the benefits of both techniques, a hybrid algorithm may, for instance, combine a metaheuristic optimisation algorithm with a fuzzy logic-based decision-making module. Within supply chain networks, where uncertainty and imprecision are significant variables, there are several real-world applications of the fuzzy least cost path problem (FLCPP).

Supply chain logistics experts may utilise FLCPP to optimise transportation routes by accounting for irregular factors including weather fluctuations, traffic congestion, and bad road conditions. Atkinson et al. [15] provided a multi-criteria analysis-based least-cost-path technique for the route of an all-weather road. Because fuzzy logic enables the model to adapt to changing conditions and provide dependable route recommendations, it lowers transportation costs and

delivery delays.

Numerous research projects have been carried out on the fuzzy shortest path problem (FSPP). Within two locations, there are several ways and routes. When making decisions on the least amount of time and money to travel, the most reliable route, the least amount of traffic, etc., this information is crucial. Many real-world issues, such as scheduling [16, 17] and telecommunication communications [18, 19], were modelled by decision makers using SPP. The shortest path must be found by decision makers, who are viewed as a graph, in order to solve this optimisation challenge. For instance, the primary goal of the fire stations is to get the fire van to the house in their service area as quickly as possible. The lowest damaged path for the fastest evacuation is a highly valuable tool for disaster management.

The fuzzy shortest route problem was first proposed by Dubois and Prade in [20]. A fuzzy shortest route technique, based on the classic Ford Moore Bellman algorithm, was presented by them. In their study, they used the concept of a path's criticality.When Chanas and Kamburowski in [21] suggested utilising fuzzy preference relationship to find the shortest path and included it in their SPP approach, they introduced the concept of fuzzy preference relationship. Each edge was represented by an integer number between a given higher integer number and 1 by the authors of [22].

Taking into consideration the dynamic programming approach, they created an algorithm that selects a path depending on the membership grade ascertained by an expert. Linet al. [23] considered an edge to be the most significant (vital) edge in the graph if its removal increased the shortest path maximum. They have produced a fuzzy membership function for the shortest path by applying a linear programming approach. An algorithmic method has been suggested by them to identify the single most important edge in a fuzzy network.

Takahashi et al. introduced the fuzzy arc length SPP in [24], expanding upon Okada [25]'s originally proposed technique. A large fuzzy graph's estimated shortest path may be found using a genetic method, according to the authors. Interval and triangle fuzzy numbers were used by Nayeem and Pal [26] to define a graph's edge weights. Both kinds of numbers can be handled by their suggested approach, which can also solve the fuzzy shortest route issue.

An algorithmic method for resolving the fuzzy shortest route issue was presented by Hernandes et al. [27]. The weight of each edge has been represented by the authors using a triangular fuzzy number. Mahdavi and colleagues [28] introduced a dynamic programming method for solving the fuzzy shortest path chain issue. For their algorithm, they proposed a ranking system, which can assist in preventing the set of shortest pathways from being created because there may be a large number of shortest paths overall from a large fuzzy graph. Even for a specialist, selecting the precise shortest path might be quite challenging. For the purpose of solving the shortest path issue in an unpredictable environment, Deng et al. have given an expanded traditional Dijkstra's approach in [29]. By employing trapezoidal fuzzy numbers, the authors have expressed the fuzzy network's edge weights.They used the graded mean integration technique of fuzzy numbers in their approach to calculate the fuzzy route's length and contrast the fuzzy path distances between two different fuzzy pathways.

Hassanzadeh et al. suggested a method in [30] using fuzzy edge weights on a fuzzy graph, to determine the shortest path. The authors used an addition procedure based on α cut to determine the path's length. Using a least squares method for their addition operation, they suggest building an approximation fuzzy membership function for the proper addition process. A genetic strategy is also proposed to solve the fuzzy SPP and deal with the challenge of adding fuzzy numbers for a large-scale fuzzy network.

FLCPP may be used to address inventory routing problems while replenishing inventory levels across several locations while taking erratic demand patterns and supply disruptions into consideration. Routing strategies that balance inventory costs and service levels can be produced via an optimisation model that accounts for fuzzy demand estimates and fuzzy trip lengths.Fuzzy least cost path algorithms, which identify reliable supply chain routes with low impact from uncertainty and disturbances, can be used to reduce supply chain risk. Using probabilistic models and fuzzy risk assessments, decision-makers may proactively predict and avoid risks related to

supplier failures, natural disasters, and delays in transportation.

Because demand patterns and market circumstances are unpredictable, FLCPP's analysis of possible configurations can assist in making judgements regarding the placement of facilities and network architectures. By accounting for factors like fuzzy location preferences, transportation costs, and projected market demand, businesses may strategically place their facilities and build supply chain networks that are robust and flexible to changing customer wants. These real-world applications demonstrate FLCPP's versatility in addressing a range of supply chain network optimisation problems, including inventory management, transportation, and network design in addition to risk reduction. Businesses may make educated decisions that improve the adaptability, efficiency, and resilience of their supply chains in the face of uncertainty by utilising fuzzy logic and optimisation techniques.

2. Defination & Preliminaries

2.1. Fuzzy Set

A set $\tilde{I}$ defined as $\tilde{I} = \{(u, \mu_{\tilde{I}}(u)) : x \in I, \mu_{\tilde{I}}(u) \in [0,1]\}$, where $\mu_{\tilde{I}}(u)$ is membership function of $\tilde{I}$, is called a fuzzy set.

2.2. Hexagonal Fuzzy Number

Let $\tilde{I_h} = (i_1, i_2, i_3, i_4, i_5, i_6)$, be a fuzzy set defined on $\Re = (-\infty, \infty)$, which is called a hexagonal fuzzy number if the membership function of $\tilde{I_h}$ is given by

$$\mu_{\tilde{I_h}}(t) = \begin{cases} \frac{1}{2}(\frac{t-i_1}{i_2-i_1}), & \text{if} \quad i_1 \le t \le i_2 \\ \left(\frac{1}{2} + \frac{t-i_2}{2(i_3-i_2)}\right), & \text{if} \quad i_2 \le t \le i_3 \\ 1, & \text{if} \quad i_3 \le t \le i_4 \\ \left(1 - \frac{i_4-t}{2(i_5-i_4)}\right), & \text{if} \quad i_4 \le t \le i_5 \\ \left(\frac{i_6-t}{2(i_6-i_5)}\right), & \text{if} \quad i_5 \le t \le i_6 \\ 0, & \textit{otherwise} \end{cases}$$

2.3. Generalized Hexagonal Fuzzy Number

Let $\tilde{I_{gh}} = (i_1, i_2, i_3, i_4, i_5, i_6; w)$ be a fuzzy set defined on $\Re = (-\infty, \infty)$ and w is its maximum degree of membership function. The membership function of $\tilde{I_{gh}}$ is given by

$$\mu_{\tilde{I_{gh}}}(t) = \begin{cases} 0, & \text{if} \quad t \le i_1 \\ \frac{w}{2}(\frac{t-i_1}{i_2-i_1}), & \text{if} \quad i_1 \le t \le i_2 \\ \frac{w}{2} + \frac{w}{2}(\frac{t-i_2}{i_3-i_2}), & \text{if} \quad i_2 \le t \le i_3 \\ w, & \text{if} \quad i_3 \le t \le i_4 \\ 1 - \frac{w}{2}\frac{i_4-t}{i_5-i_4}, & \text{if} \quad i_4 \le t \le i_5 \\ \frac{w}{2}(\frac{i_6-t}{i_6-i_5}), & \text{if} \quad i_5 \le t \le i_6 \\ 0, & \text{if} \quad t \ge i_6 \end{cases}$$

2.4. Arithmetic operation on GHFN (Generalized Hexagonal Fuzzy Number):

Let $\tilde{I}_{gh_1} = (i_1, i_2, i_3, i_4, i_5, i_6; w_1)$ and $\tilde{I}_{gh_2} = (j_1, j_2, j_3, j_4, j_5, j_6; w_2)$ be two GHFNs then

1. **Equality of two GHFNs:**

   $\tilde{I}_{gh_1} = \tilde{I}_{gh_2}$ iff $i_1 = j_1, i_2 = j_2, i_3 = j_3, i_4 = j_5, i_5 = j_5, i_6 = j_6$ and $w_1 = w_2$

2. **Addition of two GHFNs:**

   $\tilde{I}_{gh_1} + \tilde{I}_{gh_2}$=$(i_1 + j_1, i_2 + j_2, i_3 + j_3, i_4 + j_4, i_5 + j_5, i_6 + j_6; w)$ where $w = min(w_1, w_2)$

3. **Subtraction of two GHFNs:**

   $\tilde{I}_{gh_1} - \tilde{I}_{gh_2}$=$(i_1 - j_1, i_2 - j_2, i_3 - j_3, i_4 - j_4, i_5 - j_5, i_6 - j_6; w)$ where $w = min(w_1, w_2)$

3. Ranking of Generalized Hexagonal Fuzzy Numbers [31]

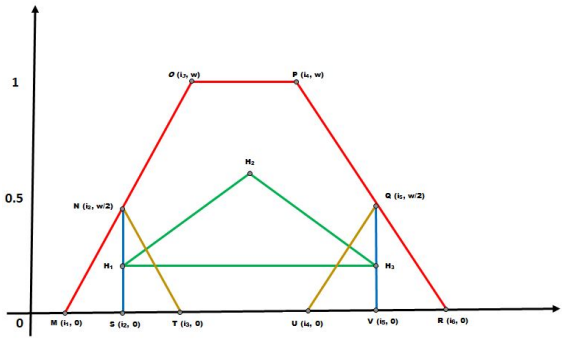


Figure 1: *Generalized hexagonal fuzzy numbers. Dividing hexagonal into three plain figure and taking centeriod say H_1, H_2, H_3 of these three figures will here define the ranking function.*

In Figure 1, the hexagon's balance point is thought to represent the centroid of a fuzzy number. The hexagonal may be divided into three planar forms. The triangle MNT, hexagon $OPQVTN$, and triangle VQR are the three figures, in that order. The generalised hexagonal fuzzy number ranking is defined by using the circumcenter of the centroids of these three planar figures as the reference point. Let H_1, H_2 and H_3 be the centroid of the corresponding three planar figures.

The centroid of the three plane figures is

$$H_1 = (\frac{i_1 + i_2 + i_3}{3}, \frac{w}{6}), \qquad H_2 = (\frac{i_2 + 2i_3 + 2i_4 + i_5}{6}, \frac{w}{2}), \qquad H_3 = (\frac{i_4 + i_5 + i_6}{3}, \frac{w}{6})$$

The line H_1, H_3 has the equation y=$dfracw6$, and H_2 is not on the line H_1, H_3. As a result, H_1, H_2, and H_3 form a non-collinear triangle.
The generalised hexagonal fuzzy number $\tilde{I_{gh}} = (i_1, i_2, i_3, i_4, i_5, i_6; w)$, and its ranking function $R(\tilde{f_{gh}})$ map the set of all fuzzy numbers to $\Re = (-\infty, \infty)$.

$$R(\tilde{f_{gh}}) = (x_0)(y_0) = (\frac{2i_1 + 3i_2 + 4i_3 + 4i_4 + 3i_5 + 2i_6}{18} \times \frac{5w}{18}) \qquad ....(1)$$

4. Least cost path Problem [4]

The least-cost path problem is a critical aspect of transportation and logistics that focuses on allocating resources efficiently along various routes to achieve the lowest possible cost between a starting point and a destination. The objective is to identify the most economical path from the initial node (Node 1) to the target node (Node n). In this context, each edge connecting the nodes is weighted to reflect the associated costs, which may include factors such as travel time, fuel expenses, and toll fees. Each city within the network is represented as a node, creating a framework for analyzing potential routes.

Determining the optimal route to connect these locations is essential for effective decision-making, as it directly impacts operational efficiency and cost management. By concentrating on minimizing costs, this approach not only enhances the overall performance of transportation systems but also aids organizations in making informed choices regarding resource allocation. Additionally, addressing the least-cost path problem contributes to the development of more sustainable transportation solutions, as it encourages the utilization of routes that optimize both time and financial resources. Ultimately, a thorough understanding of this problem is vital for improving supply chain management and enhancing service delivery in the logistics sector.

In addressing the problem at hand, we recognize that it unfolds in distinct phases, with each level requiring a decision to be made among several available options. At step ①, the decision-maker must select one of the three potential pathways: (1,2), (1,3), or (1,4). Each of these options presents different implications for the overall route and associated costs. The ultimate goal is to identify the optimal policy, which comprises a series of interconnected pathways or routes that effectively link the starting point ① to the destination ⓝ. This optimal policy will not only minimize costs but also accommodate the inherent uncertainties in the decision-making process. By systematically evaluating each pathway at every phase, the decision-maker can navigate through the complexities of the problem, ensuring that the selected route aligns with the overarching objectives of efficiency and cost-effectiveness. The following numerical example illustrates this decision-making process, providing clarity on how these pathways interact and contribute to the overall solution.

The problem can be approached through either forward or backward recursive equations, each providing a unique pathway to arrive at the solution. In forward recursion, the process begins at the initial state and progresses sequentially toward the final state. This method allows for a step-by-step evaluation of each decision point as the solution unfolds, ultimately leading to the desired outcome. Conversely, backward recursion starts at the end of the process and works its way back to the beginning. This approach is particularly useful when the final outcomes are known, allowing for the determination of optimal decisions by tracing back through the various pathways. By evaluating the potential consequences of each decision in reverse order, the decision-maker can identify the best route to achieve the final objectives. Both methods have their advantages, depending on the specific context of the problem. Forward recursion may be more intuitive for problems where the sequence of actions is clear and linear, while backward recursion can be advantageous in complex scenarios where understanding the end conditions is crucial. Ultimately, the choice between forward and backward recursion will depend on the nature of the problem and the preferences of the decision-maker.

4.1. Fuzzy Least Cost Path Problem

In a fuzzy environment, the problem at hand is addressed by considering activity time as a generalized hexagonal fuzzy number. This approach allows for a more accurate representation of uncertainties inherent in real-world scenarios, facilitating better decision-making throughout the optimization process.

4.2. Procedure of solution using Ranking method

To address the least-cost path problem, the following steps should be undertaken:

Step 1: We analyze the acyclic network, utilizing hexagonal fuzzy numbers to calculate the edge weights for improved decision-making.

Step 2: Find the value of (x_0) & (y_0).

Step 3: Defuzzify the edge weights by using ranking method mentioned in section 3.

Step 4: At each stage of the network, systematically calculate the cost value associated with each pathway.

Step 5: Find the total cost value for each path(p_i).

Step 6: Now, to find least, cost consider the minimum cost value obtained in step 5. i.e. suppose $\exists$ n path then-

$$\text{Least cost value} = \text{Min}[costvalue(p_1), costvalue(p_2), costvalue(p_3), ...costvalue(p_n)]$$

The cost value of end node is assumme as 0 which is in the last stage.

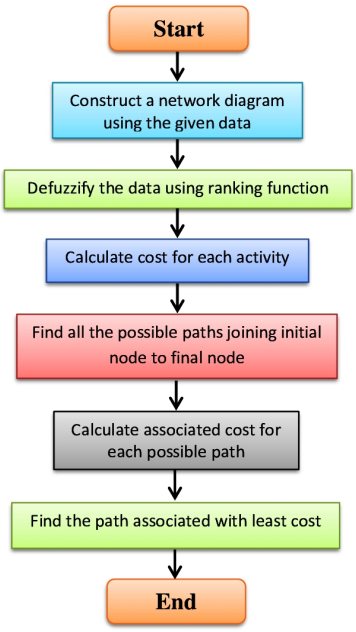


Figure 2: *Flow chart of the method.*

Validation

A potent mathematical method for resolving complicated decision-making issues is dynamic programming (DP), which divides larger problems into smaller, linked subproblems. It is an optimization technique that breaks down a more complex issue into a number of easier-to-manage issues, each of which adds to the final answer. In contrast to conventional mathematical formulations, decision-making process optimization (DP) offers a methodical approach to identify the best possible collection of decisions, rather than imposing a rigid mathematical model. This method provides a flexible framework for handling many different kinds of problems, with individual equations and solutions tailored to the unique circumstances of each one.

Dynamic programming is built around the concept of solving multistage problems, where a complex problem is broken down into several smaller stages, each linked in a specific order. At each stage, a decision needs to be made, and that decision immediately affects what happens next. The goal in these problems is to find a sequence of decisions that leads to the best possible outcome for the entire process. As you make decisions at each stage, the sequence gradually forms, guiding you toward the optimal solution step by step [32].

Dynamic programming (DP) was introduced to solve problems where decision-making factors change over time, which is common in real-world scenarios. It plays a crucial role in optimization, where the goal is to find either the maximum or minimum of a given objective. DP works by breaking down complex problems into smaller, interrelated sub-problems, known as phases, making them easier to solve. Fuzzy set theory, introduced by Zadeh [33], provides an effective way to handle uncertainty and vagueness in such problems. It has become an essential tool for dealing with real-world challenges where data may be imprecise or incomplete. Bellman et al.[1] linked fuzzy set theory to decision-making because the decision-making process often involves uncertainty and ambiguity. By integrating fuzzy set theory with dynamic programming, fuzzy dynamic programming (FDP) emerged as a method for solving optimization problems where parameters are not clearly defined.

In this research, we apply fuzzy dynamic programming to find the least-cost path, where the activity times (or costs) are represented as fuzzy numbers. This approach allows us to handle the uncertainties inherent in real-world logistics and supply chain problems, providing a more flexible and accurate solution to complex decision-making challenges.

Fuzzy Forward and Fuzzy Backward recursive equations

The dynamic programming model include

1. Let s_p be any state, then then it describes a specific city at stage p.
2. The path starting from one stage to the following stage known as decision alternatives.
3. Decision variable from state $p-1$ to p represent by d thus the state changes from s_{p-1} to s_p.
4. From decision $d(s_{p-1}, s_p)$ return will be represented by $f_p(s_p)$.
5. From s_1 to s_p the minimum cost will be denoted by ${f_p}^*(s_p)$.

The Fuzzy Forward recursive equations is

$$\tilde{f}_1(s_1) = \tilde{d}(s_1, s_2)$$
$$\tilde{f}_p(s_p) = Min\{\tilde{d}(s_p, s_{p+1}) + \tilde{f}_{p-1}(s_{p-1})\} \text{ where } p = 2, 3, ..., n$$

The Fuzzy Backward recursive equations is

$$\tilde{f}_n(s_n) = \tilde{d}(s_n, s_{n+1})$$
$$\tilde{f}_p(s_p) = Min\{\tilde{d}(s_p, s_{p+1}) + \tilde{f}_{p+1}(s_{p+1})\} \text{ where } p = n-1, ..., 1$$

Numerical Example

Consider a logistics company responsible for delivering goods from a central warehouse (①) to a retail store (⑧), with several distribution centers in between, namely ②, ③, ④, ⑤, and ⑥. The travel times between these nodes are uncertain, influenced by factors such as traffic conditions and operational delays. To account for this uncertainty, the travel times are represented by Generalized Hexagonal Fuzzy Numbers (GHFNs).

Table 1 presents the fuzzy activity times associated with each activity, reflecting the variability in delivery durations. Figure 3 illustrates the graphical representation of these fuzzy activity times,

providing a visual understanding of the distribution and uncertainty in the delivery process. The objective of this analysis is to determine the optimal route time between nodes ① and ⑧, ensuring efficient delivery while accommodating the inherent uncertainties in travel times. Our aim is to analyze different routes and identify the one that minimizes overall delivery time in its logistics operations.

Table 1: *Cost value in terms of Hexagonal fuzzy numbers*

Cost variable	Fuzzy activity duration
C_1: 1-2	(20,25,30,35,40,45)
C_2: 1-3	(11,12,13,14,15,16)
C_3: 1-4	(30,32,35,36,40,42)
C_4: 2-5	(11,15,17,18,20,25)
C_5: 3-6	(11,11,12,12,13,16)
C_6: 4-5	(60,62,65,66,68,70)
C_7: 4-6	(41,43,45,47,49,50)
C_8: 4-7	(2,4,6,8,10,12)
C_9: 5-8	(80,82,87,89,90,95)
C_{10}: 6-8	(52,53,55,55,56,59)
C_{11}: 7-8	(15,17,19,25,28,30)

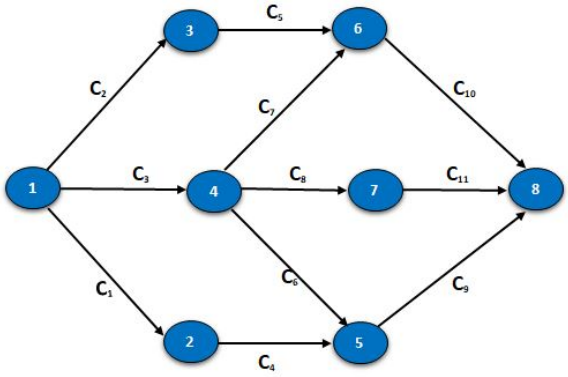


Figure 3: *Acyclic network in terms of hexagonal fuzzy numbers having total 8 events and 11 activities.*

The ranking approach for generalized hexagonal fuzzy numbers, which is discussed in Section 3, is used to transform hexagonal fuzzy costs into crisp values. These expenses are calculated as the cost connecting the two nodes.

SOLUTION:

(i) Using ranking method mentioned in Section 3:

Step 1: Considering the cost variable C_1 between nodes (1, 2), the edge weight associated with this connection is represented as a hexagonal fuzzy number, specifically $(20, 25, 30, 35, 40, 45)$. This representation reflects the uncertainty in the cost estimates for this route.

Step 2: Find the value of (x_0) & (y_0).

(i) For the value of x_0:

$$x_0 = \frac{2i_1 + 3i_2 + 4i_3 + 4i_4 + 3i_5 + 2i_6}{18} = \frac{2(20) + 3(25) + 4(30) + 4(35) + 3(40) + 2(45)}{18} = 32.5$$

(ii) For the value of y_0:

$$y_0 = \frac{5w}{18} = \frac{5(1)}{18} = 0.277777778$$

Step 3: Using the ranking technique outlined in Section 3 to defuzzify the edge weights. This method makes it easier to transform fuzzy values into precise numbers, which improves the

ability to compare edge weights and makes route selection decision-making easier.

$$R(\tilde{f_{gh}}) = (x_0)(y_0) = (32.5)(0.277777778) = 9.03$$

Step 4: To calculate the time required for each activity:

Table 2: *Cost value associated with each activity.*

Activity	Fuzzy activity duration	x_0	y_0	$R(\tilde{f_{gh}})$
① → ②	(20,25,30,35,40,45)	32.5	0.277777778	9.03
① → ③	(11,12,13,14,15,16)	13.5	0.277777778	3.75
① → ④	(30,32,35,36,40,42)	35.78	0.277777778	9.94
② → ⑤	(11,15,17,18,20,25)	17.61	0.277777778	4.89
③ → ⑥	(11,11,12,12,13,16)	12.33	0.277777778	3.43
④ → ⑤	(60,62,65,66,68,70)	65.22	0.277777778	18.12
④ → ⑥	(41,43,45,47,49,50)	45.89	0.277777778	12.75
④ → ⑦	(2,4,6,8,10,12)	7.00	0.277777778	1.94
⑤ → ⑧	(80,82,87,89,90,95)	87.22	0.277777778	24.23
⑥ → ⑧	(52,53,55,55,56,59)	54.94	0.277777778	15.26
⑦ → ⑧	(15,17,19,25,28,30)	22.28	0.277777778	6.19

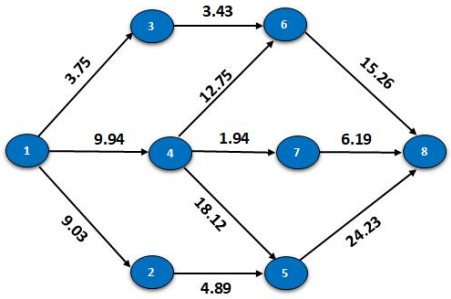


Figure 4: *Acyclic network in crisp form.*

Step 5: To determine the total duration for each path from the starting node ① to the ending node ⑧, we sum the time values associated with the edges along each route. This calculation includes the fuzzy durations for each activity, providing a clear estimate of the total time needed to travel between the two nodes.

Table 3: *Cost associated with each path.*

Path	Associated Cost Value
P_1 : ①→③→⑥→⑧	22.44
P_2 : ①→④→⑥→⑧	37.95
P_3 : ①→④→⑦→⑧	18.07
P_4 : ①→④→⑤→⑧	52.29
P_5 : ①→②→⑤→⑧	38.15

5. Result

In this study, we utilized the ranking function method to determine the least cost path, with Generalized Hexagonal Fuzzy Numbers (GHFNs) representing the uncertain activity times. By using GHFNs, we were able to effectively model various levels of uncertainty and imprecision inherent in the transportation process. This approach is especially useful in real-world scenarios where activity times, such as travel durations or fuel costs, are not fixed but subject to variability and ambiguity. The use of GHFNs allows for a detailed and flexible representation of these uncertainties, ensuring that fluctuations in activity times are accounted for. This method provides a more accurate and comprehensive understanding of the potential costs involved, enhancing decision-making in least-cost path analysis.

To determine the least fuzzy activity time of the path connecting ① to ⑧, we employ the forward procedure outlined in the proposed approach. The fuzzy activity time of the optimal path is computed as:

$$\text{Least cost value= Min } [22.44, 37.95, 18.07, 52.29, 38.15] = 18.07$$

Least cost value route = P_3 : ①→ ④ → ⑦ → ⑧ This value represents the fuzzy duration of the optimal path, taking into account uncertainty and imprecision in the activity times. The path linking ① to ④, ⑦, and ⑧ is determined to be the optimal one. By further applying the fuzzy backward recursive equations, we validate this optimal path, ensuring consistency with the forward approach.

To translate the fuzzy activity time into a more actionable result, we use the ranking function described in Section 3 of the Generalized Hexagonal Fuzzy Method (GHFM). This ranking function converts the fuzzy values into a crisp form, allowing for a precise evaluation of the optimal path. The least crisp activity time for the path from ① to ⑧ is calculated as:

$$\text{Crisp Activity Time} = 18.07$$

The conversion from fuzzy to crisp values provides a concrete measure of the total duration, which aids decision-makers in assessing and comparing potential routes. Furthermore, this result is visually demonstrated in Figure 5, where the least-cost path is depicted with a red dotted line connecting node ① to the destination node ⑧. The optimal path is highlighted, confirming that the path $C_3 \rightarrow C_8 \rightarrow C_{11}$ has the least cost in fuzzy form as $(47, 53, 60, 69, 78, 84)$ and in crisp form as 18.07. The graphical representation supports the findings of the ranking function and enhances the clarity of the optimal route selection.

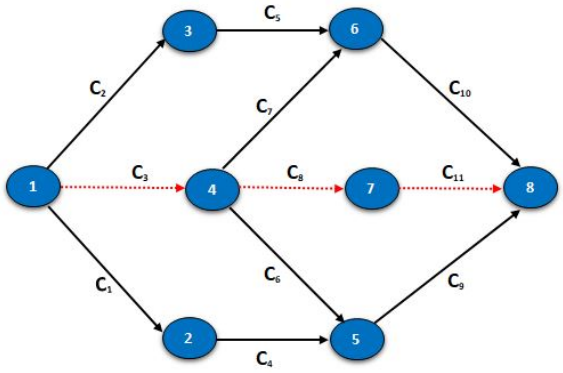


Figure 5: *Least cost value path.*

Validation

To solve the problem of identifying the optimal route for transportation, we break the process down into distinct stages, state variables, and decision variables. We begin by calculating the distances between the cities, starting with the first city, ① $(= x_1)$. From this initial point, we evaluate the distances to the next city, x_2, followed by x_3, and finally x_4. Each decision variable represents the choice of the route taken between consecutive cities. The state variables consist of the current city being analyzed, which helps track progress through the network of cities. By systematically calculating the distances at each stage, we can assess potential routes and determine the one that minimizes overall transportation costs. This approach not only allows for the identification of the best route but also incorporates any uncertainties in the distance data, ultimately facilitating a more effective decision-making process in route selection.

Fuzzy forword recursive equation:

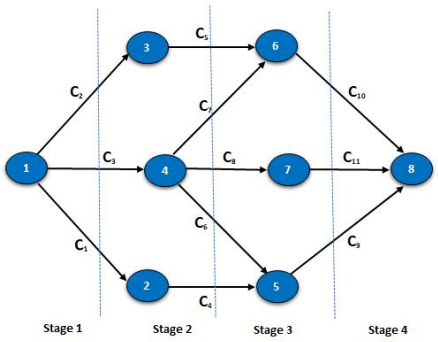


Figure 6: *Stages in network diagram*

Stage 2: In Stage 1, we start by finding the minimum distance from city ① to the cities in Stage 2: $s_2 =$ ②, $s_2 =$ ③, and $s_2 =$ ④. This step helps us understand the potential routes we can take from our starting point. As we move forward, we continue to assess the distances, working our way from city ① all the way to city ⑧. To make this process clearer, table 4 displays the distances necessary to get from Stage 1 to Stage 2. By looking closely at these values, we can pinpoint the most efficient routes and fine-tune our calculations to find the optimal path for our transportation needs.

$$\tilde{f}_2(s_2) = \tilde{d}(s_2, s_1)$$

Table 4: *Stage 2*

s_2	$s_1 = ①$	$\tilde{f}_2(s_2)$	s_1^*
③	$(11, 12, 13, 14, 15, 16)$	$(11, 12, 13, 14, 15, 16)$	①
②	$(20, 25, 30, 35, 40, 45)$	$(20, 25, 30, 35, 40, 45)$	①
④	$(30, 32, 35, 36, 40, 42)$	$(30, 32, 35, 36, 40, 42)$	①

The lowest value among $\tilde{d}(s_2, s_1)$ represents the ideal value of $\tilde{f}_2(s_2)$. In this step, for state $s_2 = ④$, we find the minimal value to be $(30, 32, 35, 36, 40, 42)$. This result reflects the most favorable outcomes for the distances associated with reaching state ④, providing a clear indication of the optimal route's effectiveness at this stage in our analysis. By identifying this minimal value, we can make more informed decisions as we proceed to the next stages of our route optimization process.

Stage 3: This Stage will follow a similar process to that of Stage 2. We will obtain the next optimal value by adding the minimal value from this Stage to the minimum value identified in Stage 2. By maintaining this consistent approach, we ensure that each stage builds upon the previous calculations, allowing us to systematically refine our route selection. This cumulative method enhances our ability to identify the most efficient path while considering the uncertainties and varying parameters at each step.

$$\tilde{f}_3(s_3) = Min\{\tilde{d}(s_3, s_2) + \tilde{f}_2(s_2)\}$$

Table 5: *Stage 3*

s_3	$s_2 = ③$	$s_2 = ②$	$s_2 = ④$	$\tilde{f}_3(s_3)$	s_2^*
⑥	$(22, 23, 25, 26, 28, 32)$	-	$(71, 75, 80, 83, 89, 92)$	$(22, 23, 25, 26, 28, 32)$	③
⑤	-	$(31, 40, 47, 53, 60, 70)$	$(90, 94, 100, 102, 108, 112)$	$(31, 40, 47, 53, 60, 70)$	②
⑦	–	–	$(32, 36, 41, 44, 50, 54)$	$(32, 36, 41, 44, 50, 54)$	④

Consequently, we achieve $\tilde{f}_3(s_3)$ at its ideal value. For state $s_3 = ⑦$, we find the associated values in Stage 3 to be $(32, 36, 41, 44, 50, 54)$. This set of values reflects the minimal distances for reaching state ⑦, indicating the optimal routes available at this stage.

Stage 4: Following the same procedure as before, we will add the minimum value from this stage to the minimum value obtained in stage 3 to determine the next best value. This approach ensures that each stage builds on the previous results, helping us refine our route selection as we move forward in the optimization process.

$$\tilde{f}_4(s_4) = Min\{\tilde{d}(s_4, s_3) + \tilde{f}_3(s_3)\}$$

Table 6: *Stage 4*

s_4	$s_3 = ⑥$	$s_3 = ⑤$	$s_3 = ⑦$	$\tilde{f}_4(s_4)$	s_3^*
⑧	$(74, 76, 80, 81, 84, 91)$	$(111, 122, 134, 142, 150, 165)$	$(47, 53, 60, 69, 78, 84)$	$(47, 53, 60, 69, 78, 84)$	⑦

Thus, we obtain the optimal value of $\tilde{f}4(s4)$ in stage 4, which is $(47, 53, 60, 69, 78, 84)$, associated with state $s_4 = ⑧$. This set of values represents the minimal distances for reaching state ⑧ and reflects the best routes available at this stage.
Notably, the results from dynamic programming match those from the ranking function, confirming the reliability of our method. This consistency shows how well our approach handles uncertainties and improves transportation logistics.

6. Comparison with existing methods

Table 7 presents a comprehensive comparison between the solutions obtained using the ranking function employed in this study and those derived from various existing methods found in the literature. The comparison highlights key performance indicators, including the least-cost path, the associated costs, the capacity to manage uncertainty in activity durations, and the overall computational efficiency in determining the least-cost path. This comparative analysis demonstrates that the proposed method consistently yields more cost-efficient solutions, establishing it as a competitive alternative to the other approaches considered.

1. Deng et al. [34] utilized the graded mean integration method. Following their approach, we apply graded mean integration to a hexagonal fuzzy number $\tilde{I}_h = (i_1, i_2, i_3, i_4, i_5, i_6; w)$. The graded mean integration in this case is expressed as follows:

$$GM = \frac{w(i_1 + 2i_2 + 3i_3 + 3i_4 + 2i_5 + i_6)}{12}$$

2. Nagalakshmi et al. [4] proposed an approach to solve the Generalized Trapezoidal Fuzzy Least-Cost Route problem using the centroid method. In this paper, we extend their method by applying the centroid technique for Hexagonal Fuzzy Numbers, employing the corresponding formula to achieve optimal route selection.

$$C(\tilde{I}_h) = \frac{w(i_1 + i_2 + i_3 + i_4 + i_5 + i_6)}{6}$$

3. Ramkumar et al. [35] adopt a statistical approach to the hexagonal fuzzy shortest travelling path problem, utilizing measures of central tendency. This method closely aligns with the approach employed by [4], indicating a similarity in their techniques for addressing fuzzy shortest path problems.

4. Rajkumar et al. [36] determined the shortest path in an acyclic network using hexagonal, heptagonal, and octagonal fuzzy numbers. The results were compared by defuzzifying the fuzzy numbers through the magnitude measure, providing an effective approach for handling complex fuzzy representations in network optimization problems.

$$GM = \frac{w(-i_1 + 6i_2 + i_3 + i_4 + 6i_5 - i_6)}{12}$$

The methodologies reviewed in this paper present a comprehensive framework for addressing fuzzy shortest path and least-cost route problems utilizing various types of fuzzy numbers. Specifically, the extension of the centroid method to hexagonal fuzzy numbers facilitates an effective mechanism for optimal route determination, enhancing the existing literature on fuzzy optimization. The application of graded mean integration, the centroid technique, and magnitude measures exemplifies the versatility of these approaches in managing uncertainty inherent in network optimization scenarios.

Furthermore, the close alignment between the techniques proposed by Deng et al. [34], Nagalakshmi et al. [4], Ramkumar et al. [35], and Rajkumar et al. [36] underscores their collective effectiveness in tackling complex fuzzy shortest path problems. Each method contributes uniquely

to the field, offering insights into the performance of different fuzzy representations in route selection. This comparative analysis not only highlights the strengths and limitations of the various approaches but also serves as a foundation for future research aimed at refining these methodologies. Consequently, this paper not only enriches the discourse on fuzzy optimization but also paves the way for further advancements in the study of fuzzy shortest path problems.

Table 7: *Comparison table*

Methods	Least Cost Path	Least Cost
Deng et al. [34]	P_3 : ①→④→⑦→⑧	65.00
Nagalakshmi et al. [4]	P_3 : ①→④→⑦→⑧	65.16
Ramkumar et al. [35]	P_3 : ①→④→⑦→⑧	65.17
Rajkumar et al. [36]	P_1 : ①→③→⑥→⑧	65.33
Proposed Method	P_3 : ①→④→⑦→⑧	18.07

The comparison of the least cost values for the generalized hexagonal fuzzy number, derived from our proposed approach and those from [36], is presented in Table 7. The table demonstrates that our results consistently exceed those obtained by [36], underscoring the advantages inherent in our method. This superior performance can be attributed to the innovative aspects of our approach, which distinctly differentiate it from the methodology employed by [36].

To further substantiate this claim, we have implemented the conventional methodology detailed in Section 4.2. The results acquired through this traditional method provide an additional layer of verification, reinforcing the accuracy and reliability of our proposed solution in comparison to existing approaches. This comprehensive evaluation not only highlights the efficacy of our proposed method but also emphasizes its potential for advancing the field of fuzzy optimization in route selection problems. The findings suggest that our approach offers a more effective alternative for practitioners and researchers seeking to optimize least-cost routes using generalized hexagonal fuzzy numbers.

7. Conclusion and Future Research

Fuzzy ranking techniques demonstrate a superior capacity for managing uncertainties and imprecise data compared to deterministic methods, thereby enhancing the robustness of solutions against variations and errors in input parameters. This study presents a numerical case in which cost values are represented by generalized hexagonal fuzzy numbers. We derive these values in both crisp and fuzzy forms utilizing the ranking technique and fuzzy dynamic programming. The adaptability of the ranking function allows for updates or modifications in response to changing conditions or priorities, facilitating dynamic optimization in practical applications.

Our approach outperforms the results of the existing study by [36], showcasing its effectiveness. Additionally, we employ the dynamic programming method as a means of validation, demonstrating the efficacy of our proposed solution. By ranking perceptions according to the relative significance of various factors, our method yields more comprehensible outcomes and promotes transparent decision-making processes.

Future research endeavors may build upon this study by exploring the representation of activity durations using alternative fuzzy number types, further enhancing the applicability and versatility of fuzzy ranking techniques in optimization problems.

Compliance with Ethical Standards:

Declarations

Funding The authors did not receive support from any organization for the submitted work.

Conflict of interest All Authors declare that they have no conflict of interest.

Research involving human participants and/or animals The authors declare that this project does not involve Human Participants and/or animals in any capacity.

References

[1] Bellman, R. E. and Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management science*, 17(4), 141-164.

[2] Hussein, M. L. and Abo-Sinn, M. A. (1995). A fuzzy dynamic approach to the multicriterion resource allocation problem. *Fuzzy sets and systems*, 69(2):115-124.

[3] Baldwin, J. F. and Pilswoth, B. W. (1994). Dynamic Programming for Fuzzy Systems with Fuzzy Environment. *Journal of mathematical analysis and applications*, 85:1-23.

[4] Nagalakshmi, T. and Uthra, G. (2017). An approach of finding an optimal solution for a fuzzy least cost route problem by using generalised trapezoidal fuzzy numbers. *Advances in Fuzzy Mathematics*, 12(3):737-745.

[5] Kim, K. and Park, K. S. (1990). Ranking fuzzy numbers with index of optimism. *Fuzzy sets and Systems*, 35(2):143-150.

[6] Chen, S. H. (1985). Ranking fuzzy numbers with maximizing set and minimizing set. *Fuzzy sets and Systems*, 17(2):113-129.

[7] Liou, T. S. and Wang, M. J. J. (1992). Ranking fuzzy numbers with integral value. *Fuzzy sets and systems*, 50(3):247-255.

[8] Abbasbandy, S. and Asady, B. (2006). Ranking of fuzzy numbers by sign distance. *Information Sciences*, 176(16):2405-2416.

[9] Abbasbandy, S. and Hajjari T. (2009). A new approach for ranking of trapezoidal fuzzy numbers. *Computers & mathematics with applications*, 57(3):413-419.

[10] Abbasbandy, S. and Hajjari, T. (2011). An improvement in centroid point method for ranking of fuzzy numbers. *Journal of Sciences, Islamic Azad University*, 78:109-119.

[11] Wang, Y. J. and Lee, H. S. (2008). The revised method of ranking fuzzy numbers with an area between the centroid and original points. *Computers & Mathematics with Applications*, 55(9):2033-2042.

[12] Chen, S. H. (1985). Operations on fuzzy numbers with function principal. *Journal of Management Science*, 6:13-25.

[13] Yager, R. R. (1980). On a general class of fuzzy connectives. *Fuzzy sets and Systems*, 4(3):235-242.

[14] Rajarajeswari, P. and Sudha, A. S. (2014). Ordering generalised hexagonal fuzzy numbers using rank, mode, divergence and spread. *IOSR Journal of Mathematics*, 10(3):15-22.

[15] Atkinson, D. M., Deadman, P., Dudycha, D. and Traynor, S. (2005). Multi-criteria evaluation and least cost path analysis for an arctic all-weather road. *Applied Geography*, 25(4):287-307.

[16] Almasi, M. H., Mirzapour, M. S., Koting, S. and Karim, M. R. (2014). Analysis of feeder bus network design and scheduling problems. *The Scientific World Journal*, 2014(1):1-10.

[17] Weide, O., Ryan, D. and Ehrgott, M. (2010). An iterative approach to robust and integrated aircraft routing and crew scheduling.*Computers & Operations Research*, 37(5):833-844.

[18] Sivakumar, B., Bhalaji, N. and Sivakumar, D. (2014). A survey on investigating the need for intelligent power-aware load balanced routing protocols for handling critical links in manets. *The Scientific World Journal*, 2014(1):1-12.

[19] Xue, G. (2003). Minimum-cost qos multicast and unicast routing in communication networks. IEEE Transactions on Communications, 51(5):817-824.
[20] Dubois, D.J. (1980). Fuzzy sets and systems: theory and applications 144, Academic press.
[21] Chanas, S, and Kamburowski, J. (1983). The fuzzy shortest route problem. *Interval and fuzzy mathematics, Proc. Polish Symp, Technical, University of Poznan*, 35-41.
[22] Klein, C. M. (1991). Fuzzy shortest paths. Fuzzy Sets Systems. 39(1):27-41.
[23] Lin, K. C. and Chern, M. (1993). The fuzzy shortest path problem and its most vital arcs. *Fuzzy Sets Systems*, 58(3):343-353.
[24] Takahashi, M. T. and Yamakami, A. (2005). On fuzzy shortest path problems with fuzzy parameters: an algorithmic approach. *Annual meeting of the North American fuzzy information processing society*, 654-657.
[25] Okada, S. (2004). Fuzzy shortest path problems incorporating interactivity among paths. *Fuzzy Sets Systems*, 142(3):335-357.
[26] Nayeem, S. M. A. and Pal, M. (2005). Shortest path problem on a network with imprecise edge weight. *Fuzzy Optimization and Decision Making*, 4(4):293-312.
[27] Hernandes, F., Lamata, M. T., Verdegay, J. L. and Yamakami, A. (2007). The shortest path problem on networks with fuzzy parameters. *Fuzzy Sets Systems*. 158(14):1561-1570.
[28] Iraj, M., Rahele, N., Armaghan, H. and Mahdavi, A. N. (2009). A dynamic programming approach for finding shortest chains in a fuzzy network. *Applied soft computing*, 9(2):503-511.
[29] Deng, Y., Chen, Y., Zhang, Y. and Mahadevan, S. (2012). Fuzzy Dijkstra algorithm for shortest path problem under uncertain environment. *Applied Soft Computing*, 12(3):1231-1237.
[30] Hassanzadeh, R., Mahdavi, I., Mahdavi-Amiri, N. and Tajdin, A. (2013). A genetic algorithm for solving fuzzy shortest path problems with mixed fuzzy arc lengths. *Mathematical and Computer Modelling*, 57(1):84-99.
[31] Rajendran, C. and Ananthanarayanan, M. (2018). Fuzzy Critical path Method with Hexagonal and generalised Hexagonal Fuzzy Numbers Using Ranking Method. *International Journal of Applied Engineering Research*, 13(15):11877-11882.
[32] Eddy, S. R. (2004). What is dynamic programming? *Nature Biotechnology*, 22(7):909“910.
[33] Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control*, 8:338-353.
[34] Deng, Y., Chen, Y., Zhang, Y. and Mahadevan, S. (2012). Fuzzy Dijkstra algorithm for shortest path problem under uncertain environment. *Applied Soft Computing*, 12(3):1231-1237.
[35] Ramkumar, S. and Ananthanarayanan, D. M. (2019). A statistical approach to fuzzy shortest travelling path using hexagonal fuzzy numbers.*International Journal of Scientific Development and Research*, 4(3):36-40.
[36] Rajkumar, A., Begum, S. G. and Praveena, N. J. P. (2020). Shortest path using dynamic programming through fuzzy numbers. *AIP Conference Proceedings*, 2261(1):1-9.

RT&A, No 4(80)
Volume 19, December, 2024

REVIEW OF CENSORING SCHEMES: CONCEPTS, DIFFERENT TYPES, MODEL DESCRIPTION, APPLICATIONS AND FUTURE SCOPE

Ninan P Oommen[1], Jiju Gillariose [2]

•

[1,2]Department of Statistics and Data Science, CHRIST (Deemed to be University), Bengaluru, Karnataka, India
ninan.oommen@res.christuniversity.in, jijugillariose@yahoo.com

Abstract

Survival analysis is one of the key techniques utilized in the domains of reliability engineering, statistics, and medical domains. It focuses on the period between the initialization of an experiment and a subsequent incident. Censoring is one of the key aspects of survival analysis, and the techniques created in this domain are designed to manage various censoring schemes with ease, ensuring accurate and insightful time-to-event data analysis. The statistical efficiency of parameter estimates is improved by accurately incorporating censoring information by making use of the available data. This paper reviews the concepts, model descriptions, and applications of conventional and hybrid censoring schemes. The introduction of new censoring schemes from conventional censoring schemes has evolved by rectifying the drawbacks of the previous schemes, which are explained in detail in this study. The evolution of hybrid censoring schemes through the combination of various conventional censoring schemes, the data structures, concepts, methodology, and existing literature works of hybrid censoring schemes are reviewed in this work.

Keywords: bibliometric analysis, hybrid censoring, left and right censoring, random censoring, survival analysis

1. Introduction

In reliability engineering, Survival analysis plays a major role in determining how long an event is estimated to take place, such as the death of a biological creature or a failure of the mechanical system. A collection of statistical methods used for time-to-event data is called survival analysis. The period of time until an event occurs is the relevant outcome variable. Generally, survival data addresses death as an event, but it may also address any occurrence that happens to a person, such as disease, recurrence following remission, and cure. Survival data describes the amount of time that occurs between an origin and an endpoint of interest. The time is considered as the number of years, months, weeks, or days preceding the start of the study enrollment. The survivor function and the hazard function are two related functions typically used to describe and model survival data. The survival function determines the probability that a person will survive from their point of origin to a point after time. Given survival up to that point in time, the hazard function provides the instantaneous potential for an occurrence at that point. Its main applications are in the fields of diagnostics and mathematical model specification for survival analysis.

Investigating the literature within the Scopus database with the search term (TITLE-ABS-KEY ("censoring scheme*") AND TITLE-ABS-KEY ("survival analysis*")) produced 55 documents in English language in the area of decision science, mathematics, statistics, computer science, and

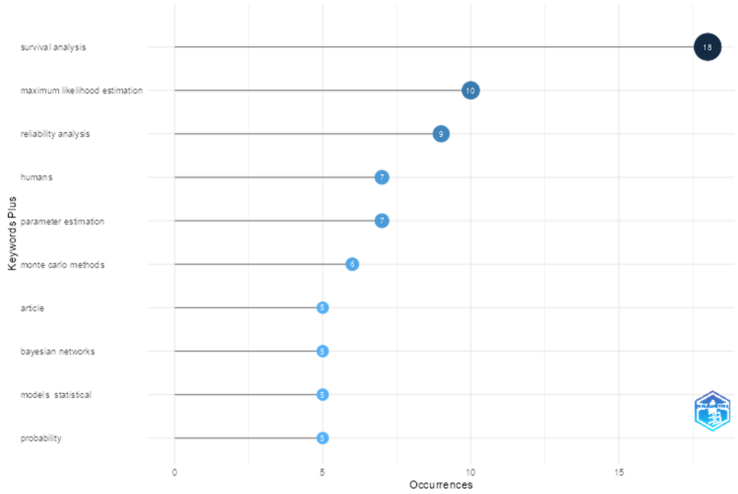


Figure 1: *Most relevant words*

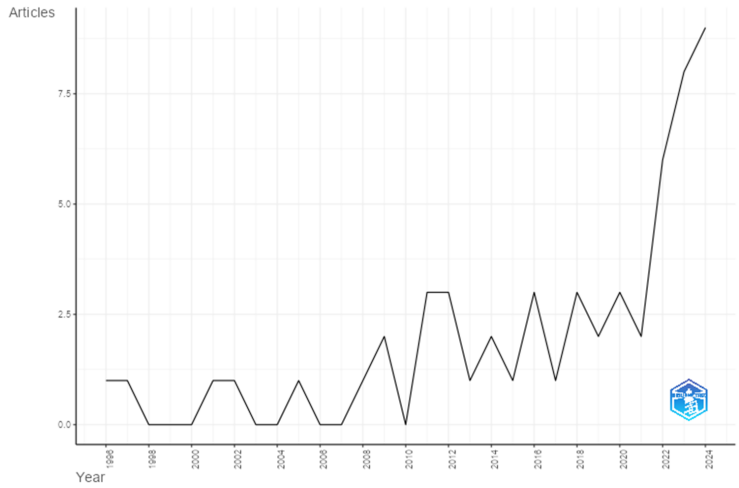


Figure 2: *Annual scientific production*

engineering. The analysis using VOS viewer generated 290 keywords with at least a single occurrence. Furthermore, the visualization of keywords explored using the overlay visualization and the results show that "survival analysis" is the most occurring keyword in the existing literature (Figure 1). The progress of works from 1996 to 2024 is visualized using overlay visualization, and the result interprets that the progress of articles in survival analysis after 2021 is very high (Figure 2). These two results show the importance of survival analysis and its applications in this current scenario.

In survival analysis, censoring schemes are essential to collect the censored data or failure time data of products. The evolution process of various censoring schemes by overcoming the limitations of previous censoring schemes is as follows. By overcoming the limitations of conventional censoring schemes, many hybrid censoring schemes have evolved in the existing literature. Right censoring, left censoring, interval censoring, and random censoring are different wide categories of censoring schemes. Type–I and type–II censoring schemes are the two major conventional censoring schemes. The hybrid censoring scheme, which is a mixture of type–I and type–II censoring schemes, has been introduced by [17]. In the case of type-I hybrid censoring scheme, from the experimenter's perspective, this scheme has the benefit of pre-fixed termination

time. The limitation of type-I hybrid and conventional type-I censoring schemes is that, the inferential results must be developed under the assumption that there is at least one observed failure. In addition, very few failures may occur up to the pre-fixed time T, which leads to low efficiency for the estimators of the model parameters. The type-II hybrid censoring scheme can ensure the observation of a specific number of failures after the experiment compared to the type-I hybrid censoring scheme, which can result in more efficient inferential results. The termination time of the type-II hybrid censoring scheme is a random variable, which is one of its disadvantages. From the experimenter's point of view, this randomness is a drawback as it adds uncertainty to the duration of the experiment.

In order to overcome these drawbacks, [10] introduced generalized type-I and type-II hybrid censoring schemes and obtained the statistical inferences under the conditions of exponential distribution. This censoring scheme can guarantee the number of failures and can provide a limit on the experimental time. In the case of generalized type-I hybrid censoring scheme, one of the drawbacks is there is no guarantee of observing a minimum number of failures if the termination of life test occurs at or before time T. In the case of generalized type–II hybrid censoring scheme, there is a chance of observing zero failures or observing only a few failures until the prefixed time T_2. By combining all the censoring schemes, [8] introduced unified hybrid censoring scheme and derived the exact distribution of the maximum likelihood estimator and exact confidence intervals for the mean of the exponential distribution.

In all the above-mentioned censoring schemes, the units or products cannot be taken at random before the termination of the experiment process. However, in practice, there may be occasions where the units are withdrawn or lost from the experiment before they break down. The loss may happen due to damage to testing facilities or depletion of funds, etc. Hence, the censoring scheme that is suitable for these circumstances will be progressive hybrid censoring scheme, which allows the withdrawal of units from the experiment before the termination time period as well as after the termination of the experiment. A detailed discussion on the theory, methods, and applications of progressive censoring can be seen in [9]. The number of withdrawing units of the censoring scheme is known in advance in progressively censored experiments, which is one of its crucial assumptions. In real life, due to unavoidable circumstances, if the experimenter changes the censoring numbers during the experiment, it will violate the mentioned assumption and hence affect the correctness and efficiency of statistical inference. Hence, the adaptive progressive type-II censoring scheme is an appropriate model proposed by [30], which considers this adaptation.

This paper is beneficial for many experts and stakeholders who are managing censoring schemes and survival analysis from various sectors. Researchers employ survival analysis in clinical trials to assess the efficacy of novel therapies and interventions and to enhance patient care and treatment procedures. Epidemiologists can utilize these survival analysis and censoring techniques to study disease incidence, distribution, and control, facilitating public health decision-making. In engineering, survival analysis aids in the estimation of product and system lifespans and failure rates, which helps with quality control and design enhancements. This paper can be beneficial for such researchers as it explains basic concepts, methodologies, existing literature works, and applications of different conventional and hybrid censoring schemes.

The paper is organized as follows: The concepts and model descriptions of conventional type censoring schemes are given in Section 2. Right censoring, left censoring, interval censoring, random censoring, and type-I and type-II censoring schemes are explained under the conventional censoring scheme. Section 3 provides aspects of various hybrid censoring schemes and their data structures. Section 4 explains the applications and future scope of different censoring schemes. Conclusions about various censoring techniques and applications are given in Section 5.

2. Conventional censoring schemes

2.1. Right censoring

Censoring is the first and most important challenge that survival analysis attempts to address. Censoring occurs when some experiments fail to view the event of interest prior to the termination of the study. It happens when the researcher is not aware of the precise survival time of the entities but has partial survival information. In survival analysis, the time record of events can be collected in two ways. The actual survival time can be collected from those who completed the event of interest. However, one cannot collect the exact failure time of subjects that exceed the censor time or those who lost to follow up the progress. Hence, on a timeline, their true lifetimes will be on the right of their observed censor times; such entities are referred to as right censored. Here, time determines the duration of monitoring, which is either full or partial observation of survival intervals. Survival analysis models generate consistent parameter values with various estimation methods using data from both censored and uncensored observations.

2.1.1 Methodology

According to [23], the methodology for right censoring can be explained as follows. Let Y be the lifetime of an individual under study and S be the fixed censoring time. Let us assume $Y's$ are independent and identically distributed with probability density function $f(Y)$ and survival function $s(Y)$. The two conditions in which the exact lifetime Y of an individual is known are when Y is less than S or equal to S. If Y is greater than S, then the event time is censored at S, and he or she is a survivor. The data obtained from this experiment can be represented by pairs of random variables (T, λ). Here, λ indicates whether the lifetime Y corresponds to an event of interest ($\lambda = 1$) or is censored ($\lambda = 0$).
i.e. For $i = 1, 2, \cdots, n$

$$\lambda = \begin{cases} 1, & \text{if } Y_i \leq S \text{ (unsensored)} \\ 0, & \text{if } Y_i > S \text{ (censored)} \end{cases}$$

T equals Y if the lifetime is observed and S if it is censored. Hence, the observed times are $T = min(Y_i, S)$ for right censoring.

2.2. Left censoring

The definition of left censoring explained by [23] is as follows. Let S be the left censoring time of an individual having a lifetime of Y. To define left censoring, the event of interest must occur for the subject before that person is observed in the study at time S ($T < S$). In other words, a time Y associated with a specific subject in a study is considered to be left censored if it is less than the censoring time S, where S is the left-censoring time. In this case, we know that the event of interest occurred before time S, but we don't know the exact failure time for these subjects. The exact time will be known if and only if Y is greater than or equal to S. As mentioned in the case of right censoring, the data obtained from this experiment can be represented by pairs of random variables (T, δ), where T is equal to Y if the lifetime is observed and δ indicates whether the lifetime Y corresponds to an event of interest ($\delta = 1$), or is censored ($\delta = 0$).

2.2.1 Methodology

Let the survival variables $Y_1, Y_2, ..., Y_n$ are left censored. If the observed sample consists of the ordered pairs (Y_i, δ_i) for $i = 1, 2, \cdots, n$. For each $i : T_i = max(Y_i, T_i)$

$$\delta = \begin{cases} 1, & \text{if } Y_i \geq S \text{ (uncensored)} \\ 0, & \text{if } Y_i < S \text{ (censored)} \end{cases}$$

T is equal to Y if the lifetime is observed, and it is equal to S if it is censored. Hence, the observed times for left censoring are $T = max(Y_i, S)$.

2.3. Interval censoring

Interval censoring often indicates an imperfect data structure or a sampling scheme. A subject is considered interval censored when studied for a while, disappears for follow-up, reappears, and is further investigated. Interval censoring is the practice of only knowing the precise location of a random variable of interest within an interval rather than observing it exactly. Data with interval censoring could be produced by several clinical trials and longitudinal studies. An instance of interval censoring explained by [24] is given as follows. A prevalent instance arises in health or medical research investigations that require regular monitoring. In this case, a patient scheduled for prearranged observations for a clinically evident change in sickness or health status can miss some visits and return with a different state. The true event time, therefore, contains the real (but unseen) time of the change's occurrence, and we only know that it is greater than the most recent observation time at which the change has not occurred and less than or equal to the initial observation time at which the change has been observed to occur, thus giving an interval of data which contains the real-time of occurrence of the change.

2.3.1 Methodology

Let the intervals $I_1, I_2, ..., I_n$ be the observed data in interval censoring for each $i = 1, 2, \cdots, n$, the m^{th} response lies in the interval I_m. An observed interval that consists of a single point, in this instance, corresponds to an uncensored observation of an observed death. Suppose t_1 is the time when we have performed an experiment or test on a subject, and the result of the subject test is negative. However, the subject tested positive on a further checkup on time t_2. In this case, we know that the person was infected with the virus at some point between t_1 and t_2, but the precise time of exposure is unknown. In a clinical trial, for instance, if the time to remission has been determined, then if the m^{th} patient is in remission at the 6^{th} week of the trial but misses subsequent follow-up visits and reappears and was out of remission by the 10^{th} week, then I_m = (6,10] represents the m^{th} patient's censoring interval or length of remission.

2.4. Random censoring

In all the previous methods, it was not explicitly stated as an assumption that the considered censoring time for an individual is a constant value or a value that is known in advance. However, censoring times are usually not fixed constants but rather random variables in real-world contexts. Therefore, one considers censoring as a random variable and the distributions of two random variables (time to event and time to censoring) and how they relate to each other to obtain the influence of censoring on the inferential part. Two distributions need to be generated for the random censoring data in survival analysis. Survival times without censorship are considered to be the first distribution, and the censorship mechanism is controlled by the second distribution. Based on these elements, the censored event is determined by comparing the censored time created in the second distribution with the uncensored survival time generated in the first distribution.

2.4.1 Methodology

In the case of first distribution, consider a random variable Y that generates times until the event of interest occurs. Let $F_Y(y)$ be the cumulative function of the distribution and $S_Y(y)$ be the survival function derived from $F_Y(y)$, which defines the probability of survival after time y. Hence, $F_Y(y) = P(Y \leq y)$ and $S_Y(y) = 1 - P(Y \leq y) = P(Y \geq y)$ and $F_Y(y)$ be the probability density function of the distribution. In the case of the second distribution, it is recommended to define a continuous distribution in the case of censorship mechanism D. And let $f(D)$ be the probability density function of the second distribution. Here, the two distributions are assumed to be independent of each other. Following is the arrangement for the censored samples. We have (T_y, λ_y) where $T_y = min(Y_y, D_y)$ and $\lambda_y = I(Y_y \leq D_y)$ for each Y^{th} observation. Here, when $\lambda = 1$, the lifetime is observed or is uncensored, and $\lambda = 0$, it is censored.

Specific works in the field of interval censoring can be seen in [20]. This work explains the proportional hazards model with case one interval censored data and its maximum likelihood estimator. While the maximum likelihood estimator for the baseline cumulative hazard function only converges at n rate, it is demonstrated that the maximum likelihood estimator for the regression parameter is asymptotically normal with n convergence rate and achieves the information bound. The asymptotic variance matrix estimation for the maximum likelihood estimator of the regression parameter is also taken into account in this article. The work in [21] provides estimates in interval censoring models, including regression model estimation as well as non-parametric distribution function estimation. The asymptotic characteristics and computational processes of the non-parametric maximum likelihood estimators are discussed in the non-parametric case, and the accelerated failure time semi-parametric regression models, the proportional odds, and the proportional hazards in the regression conditions.

In order to estimate density and hazard rate functions from randomly right-censored data, [6] presented a wavelet approach in the study. A non-parametric technique is adopted by assuming that there is no particular parametric form for the density and hazard rate. Simulation of the estimators and two real-life data examples of survival time data for patients with liver metastases from a colorectal primary tumor without other distant metastases and times of unemployment for women are also explained in the study. The methods to estimate the exponential mean lifetime in a random censoring model with insufficient data has been determined by [16], and the simulation results are also demonstrated. For the exponential model, point and interval estimates for mean lifespan of item are obtained using the maximum likelihood method. Using Monte Carlo simulation, it is shown that the large sample approximation to the log likelihood ratio provides precise confidence intervals, and the mean lifetime is positively biased when estimated using the maximum likelihood method. For the data following a generalized inverted exponential distribution, [18] determined the maximum likelihood estimators of the model parameters, expected fisher information, Bayesian estimators under the squared error loss function using Lindley's approximation, and highest posterior density credible intervals of the parameters under the random censoring scheme. The computation of Bayes estimators of Weibull distribution parameters under random censoring scheme has been done by [14]. Bayes estimators are obtained by using Lindley's approximation, importance sampling, and Gibbs sampling techniques. The credible intervals of the estimators and a real life data analysis are also performed.

2.5. Type-I and type-II censoring

In practice, two major conventional censoring schemes are type-I and type-II censoring schemes. Type I censoring data is frequently seen in many applications related to engineering and medical research. The experiment is continuous up to a pre-determined time in the case of type-I censoring scheme, while the experiment is continuous till a pre-determined number of failures of units occurs in the type-II censoring scheme. In the case of a type-I censoring scheme, zero failure units can be produced at times. The stopping time of experiments might be large in the case of a type-II censoring scheme.

2.5.1 Methodology

In type–I, a random number of failure units $x \in (0, n)$ is used in the test, where n is the total number of observations, and a fixed prior test time t is considered. Since many of the products are highly reliable, the experiment may take years to complete, and the product may become outdated. Hence, to overcome this drawback, the final time (t) of the experiment is defined initially, and the products with survival time greater than the fixed time would be termed as a censored observation. In type–II, the test time is randomly monitored, and n units are used in the test until the prefixed m^{th} failure occurs. Initially, the number m of observations is defined out of the total n of observations. After generating the n samples from a distribution, the observations are sorted in increasing order, and the first $n - m$ values are termed as the observed data. The remaining m values correspond to the value of the $m + 1$ position. Hence, we assume the survival

time of these observations is equal to the largest one observed since these observations did not experience the event of interest. The major difference between type–I and type–II is that the number of failures is random in the type-I censoring scheme, and the experimental time is random in the type-II censoring scheme.

Some of the works that have been done in the field of type–I and type–II censoring schemes are mentioned below. Bayes and classical estimators have been derived by [34] for the two-parameter exponentiated-Weibull distribution when the data is taken from a type-II censoring scheme. Using non-informative priors, Bayes estimators have been generated under both the squared error loss function and the LINEX loss function. A comparison of the proposed estimators was also made on the basis of their simulated risks, which were obtained under the two loss functions. The Bayesian estimation procedure under the type-II censoring scheme has been developed by [33] for the data following flexible Weibull distribution. Jeffrey's scale invariant is considered the non informative prior, and gamma prior is considered the informative prior for the model parameters. Using a Monte Carlo simulation analysis, the efficiency of the Bayes estimators has also been compared with the classical estimators of the model parameters. For illustrative purposes, a real data set showing the intervals between secondary reactor pump failures has been examined in the study. The estimation of model parameters of Chen lifetime distributions under partial step stress accelerated life tests has been done by [3] based on the type-II censoring scheme. The parametric bootstrap and the two asymptotic distributions are applied to construct each confidence interval of the model parameters. A Monte Carlo simulation study is used to evaluate the precision outcomes.

3. Hybrid censoring

A mixture of type-I and type-II censoring schemes is known as the hybrid censoring scheme. Assume $Y_{1:n}, ..., Y_{n:n}$, where n be the ordered lifetimes of the sample units, which are independently and identically distributed random variables. The experiment is terminated when a prefixed time T has reached or when a prefixed number $r < n$ has failed. To overcome the disadvantages of the two conventional censoring schemes, [17] introduced type-I hybrid censoring scheme where the experiment gets terminated at time $T* = min(T, Y_{m:n})$ where $Y_{m:n}$ is the failure time of the m^{th} unit. And [13] introduced a type-II hybrid censoring scheme, explaining that the experiment terminates at time $T* = max(T, Y_{m:n})$.

3.0.1 Data form of type–I hybrid censoring

First, we assume the lifetimes of the units are independent and identically distributed random variables. There are two cases in which the data can be type–I hybrid censored. That is, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{r:n}\}$ if $Y_{r:n} \leq T$
Case-II: $\{Y_{1:n} < \cdots < Y_{d:n}\}$ if $Y_{d:n} > T$

Here, d represents the number of failures that occurred before time T.

3.0.2 Data form of type–II hybrid censoring

There are two cases in which the data is type II hybrid censored. That is, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{r:n}\}$ if $Y_{r:n} \geq T$
Case-II: $\{Y_{1:n} < \cdots < Y_{d:n}\}$ if $Y_{d:n} < T$

Here, $r \leq d \leq n$ represents the number of failures that occurred before time T. As assumed earlier, the lifetime of units are independent and identically distributed random variables.

Many studies have been done in the field of censoring schemes in reliability theory so far. The introduction of type-I hybrid censoring scheme, analysis of data under the condition of

exponential life distribution of the experimental units, and two-sided confidence interval of the unknown parameter (without formal proof) have been developed by [17]. Using the conditional moment generating function method, [11] determined the exact distribution of the conditional maximum likelihood estimator of θ and applied it to derive an exact lower confidence bound for θ. A simplified but identical form of the exact distribution of the maximum likelihood estimator of θ as derived by [11] has been obtained by [13]. For a Weibull distributed data, [25] derived the maximum likelihood estimators and approximate maximum likelihood estimators of the distribution parameters under hybrid censoring scheme. The approximate confidence intervals are obtained using the asymptotic distribution of the maximum likelihood estimators and the Bayes estimates of the distribution parameters using Gibbs sampling procedures.

3.1. Generalized hybrid censoring

To overcome the drawbacks of above mentioned censoring schemes, [10] introduced generalized type-I and type-II hybrid censoring schemes and obtained the statistical inferences under the conditions of the exponential distribution.

3.1.1 Methodology – generalized type – I hybrid censoring

Let n be the number of units placed on a life test before the initialization of the experiment. Define $r, k \in 1, 2, \cdots, n$ with $k \leq r \leq n$, and time $T \in (0, \infty)$. The experiment terminates at $min(Y_{r:n}, T)$ if the k^{th} failure occurs before time T. And the experiment terminates at $Y_{k:n}$ if the k^{th} failure occurs after time T. If only a few failures had been reported up to time T, it is clear that this generalized hybrid censoring alters the type-I hybrid censoring by enabling the experiment to continue. Note that the experimenter is equipped to consider a minimum of k failures but prefers to observe r failures under this censoring method.

3.1.2 Data form of generalized type – I hybrid censoring

There are three cases if the data are generalized type–I hybrid censored. That is, the observed data consists of one of the three kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{k:n}\}$ if $Y_{k:n} > T$
Case-II: $\{Y_{1:n} < \cdots < Y_{k:n} < \cdots < Y_{r:n}\}$ if $Y_{r:n} < T$
Case-III: $\{Y_{1:n} < \cdots < Y_{k:n} < \cdots < Y_{D:n}\}$ if $T < Y_{r:n}$

3.1.3 Methodology – generalized type – II hybrid censoring

According to [10], fix T_1, $T_2 \in (0,\infty)$ as time points with $T_1 < T_2$ and $r \in 1, 2, \cdots, n$. The experiment terminates at T_1 if the r^{th} failure occurs before time T_1. The experiment terminates at $y_{r:n}$ if the r^{th} failure occurs between T_1 and T_2. Finally, terminate the experiment at T_2 if the r^{th} failure occurs at T_2. This hybrid censoring scheme modifies the type-II hybrid censoring scheme by ensuring that the experiment will be terminated by time T_2. As a result, T_2 is the maximum duration of time the experiment is allowed to complete.

3.1.4 Data form of generalized type – II hybrid censoring

Three cases exist if the data are generalized type–II hybrid censored. That is, the observed data consists of one of the three kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{k:n}\}$ if $Y_{1:n} < \cdots < Y_{D1:n} < T_1$
Case-II: $\{Y_{1:n} < \cdots < Y_{D1:n} < \cdots < Y_{r:n}\}$ if $T_1 < Y_{r:n} < T_2$
Case-III: $\{Y_{1:n} < \cdots < Y_{D2:n} < T_2\}$ if $Y_{r:n} \geq T_2$

Here, D_1 and D_2 represents the number of failures that occur before T_1 and T_2, respectively.

Several works have been done in the field of generalized hybrid censoring, and mentioning a few of them here. The computation of maximum likelihood estimators and its approximate

confidence intervals for the unknown parameters of Weibull distribution has been done by [5] under type-I generalized hybrid censoring scheme. They also derived its Bayes estimates using the Markov chain Monte Carlo (MCMC) technique. To compute the statistical inferences of the competing risks model with partially observed causes of failure, [4] derived the maximum likelihood estimators, associated confidence intervals, Bayes estimators, and credible intervals of the model parameters when the latent failure times satisfy the assumptions of Lomax life distribution under type-II generalized hybrid censoring scheme. In the presence of partial constant stress acceleration, [4] computed the statistical inference when the Gompertz life products undergo the accelerated life testing experiment. The maximum likelihood and its corresponding confidence intervals for the two samples of exponential distribution have been derived by [2], and hence obtained the Bayes estimators relative to both symmetric and asymmetric loss functions using gamma conjugate priors.

3.2. Unified hybrid censoring

Unified hybrid censoring scheme has been introduced by [8] after combining all the censoring schemes. They derived the exact distribution of the maximum likelihood estimator and exact confidence intervals for the mean of the exponential distribution.

3.2.1 Methodology

According to [8], the methodology for unified hybrid censoring scheme is as follows. Define $r, k \in 1, 2, \cdots, n$ with $k < r$, and fix T_1, $T_2 \in (0,\infty)$ as time points with $T_1 < T_2$. The experiment terminates at $min(max(y_{r:n}, T_1), T_2)$, if the k^{th} failure occurs in prior time T_1. The experiment terminates at $min(y_{r:n}, T_2)$, if the k^{th} failure occurs between T_1 and T_2. Finally, terminate the experiment at $y_{k:n}$ if the k^{th} failure occurs after time point T_2. Therefore, we are certain that the experiment will be completed at most in time T_2 with at least k failures under this censoring scheme. If this is not the case, we can guarantee precisely k failures.

3.2.2 Data form of unified hybrid censoring

There are six cases if the data are unified hybrid censored. That is, the observed data consists of one of the six kinds of observations listed below:

Case-I: $0 < Y_{k:n} < Y_{r:n} < T_1$, the test is terminated at T_1
Case-II: $0 < Y_{k:n} < T_1 < Y_{r:n} < T_2$, the test is terminated at $Y_{r:n}$
Case-III: $0 < Y_{k:n} < T_1 < T_2 < Y_{r:n}$, the test is terminated at T_2
Case-IV: $0 < T_1 < Y_{k:n} < Y_{r:n} < T_2$, the test is terminated at $Y_{r:n}$
Case-V: $0 < T_1 < Y_{k:n} < T_2 < Y_{r:n}$, the test is terminated at T_2
Case-VI: $0 < T_1 < T_2 < Y_{k:n} < Y_{r:n}$, the test is terminated at $Y_{k:n}$

Several works in the field of unified hybrid censoring scheme are as follows. When the data satisfies the assumptions of Burr type-XII distribution, [32] developed maximum likelihood estimators for estimating the unknown parameters using the expectation maximization algorithm (EM) under the for unified hybrid censoring data and obtained Bayesian estimates using Lindley's approximation and MCMC method under the assumption of independent gamma prior and hence constructed the highest posterior density credible interval. Point and interval estimations of the unknown parameters of the inverse Weibull distribution based on a unified hybrid censoring scheme have been studied by [7]. The Bayesian estimations have been obtained based on the squared error loss function and Linex loss function using the MCMC method. Hence, a $(1-\tau) \times 100\%$ approximate, bootstrap-p, credible, and highest posterior density confidence intervals for the parameters have been constructed. In Rayleigh distribution, [22] computed the maximum likelihood estimators of the scale parameter under unified hybrid censoring scheme. Bayes estimator using mean and mode of the posterior distribution and confidence interval, credible interval, and highest posterior density credible interval have also been computed for

the scale parameter. By satisfying the assumptions of Gompertz distribution, [28] computed the statistical inference of unified hybrid censored data under a constant-stress partially accelerated life test model. The stochastic EM algorithm is used to compute the maximum likelihood estimate and shows that it exists uniquely. Bootstrap-p and bootstrap-t methods are used to construct the asymptotic confidence intervals and confidence intervals.

3.3. Progressive hybrid censoring

The introduction of type-I progressive hybrid censoring scheme has been done by [26] and [12].

3.3.1 Methodology - progressive type-I hybrid censoring

According to [26] and [12], the type-I progressive hybrid censoring scheme can be described as follows. Suppose n is the total number of observations throughout the experiment. ($D_1, D_2, ..., D_n$) be the prefixed progressive censoring scheme. Under this censoring scheme, D_1 of the $n-1$ survival units are removed randomly at the time of the first failure from the experiment. Then D_2 of the $n - D_1 - 2$ surviving units are removed at the time of the second failure, and so on. Finally, $D_m = n - D_1 - \cdots D_{m-1} - m$ surviving units are removed from the life test at the time of the m_{th} failure. Let us denote the failure times as $Y_{1:n} < \cdots < Y_{m:n}$ since the D_i's are already fixed. In type-I progressive hybrid censoring, the experiment terminates at $T* = min(y_{m:n}, T)$. If $y_{m:n}$ occurs before time point T, the experiment will stop at $y_{m:n}$ otherwise, it will stop at time point T. The experimental length clearly cannot be longer than T, and this sampling technique will likely provide additional information regarding the tail of the lifetime distribution because of progressive censoring.

3.3.2 Data form of progressive type-I hybrid censoring

Under this censoring scheme, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{m:n}\}$, if $Y_{m:n} \leq T$
Case-II: $\{Y_{1:n} < \cdots < Y_{d:n}\}$, if $Y_{m:n} > T$

Where d is the number of failures that occur prior to time point T.

3.3.3 Methodology - progressive type-II hybrid censoring

The type-II progressive hybrid censoring scheme has been introduced by [12], which overcomes the drawback that the maximum likelihood estimate may not always exist for type-I progressive hybrid censoring scheme. In type-II progressive hybrid censoring, the experiment terminates at time $T* = max(Y_{m:n}, T)$. The experiment terminates at m^{th} failure if $Y_{m:n} > T$ with units being withdrawn following each failure in accordance with the predetermined progressive censoring scheme ($D_1, D_2, ..., D_n$). However, if $Y_{m:n} < T$, then instead of ending the experiment by withdrawing all remaining D units after the m^{th} failure, the experiment will continue to record failures up to time T without withdrawing any units. Thus, in this situation, we have $R_m = R_{m+1} = \cdots = R_d = 0$ where d is the number of failures that occur prior to time point T.

3.3.4 Data form of progressive type-II hybrid censoring

Under this censoring scheme, the observed data consists of one of the two kinds of observations listed below:

Case-I: $\{Y_{1:n} < \cdots < Y_{m:n} < Y_{m+1:n} < \cdots < Y_{d:n}\}$, if $Y_{m:n} < T$
Case-II: $\{Y_{1:n} < \cdots < Y_{m:n}\}$, if $Y_{m:n} \geq T$

Where d is the number of failures that occur prior to time point T.

Some important works under the progressive hybrid censoring scheme are mentioned as follows. The statistical inferences for type-II progressive hybrid censored data under the assumptions of exponential distribution have been derived and analyzed by [26] and [12]. In the presence of constant stress partial acceleration, [15] obtained the maximum likelihood and Bayesian estimates of model parameters for the Nadarajah-Haghighi distribution under progressive type-II censoring and derived the asymptotic confidence intervals through asymptotic variance and covariance matrix and also obtained the Bayesian credible intervals. For Burr XII distribution, [31] derived the estimates using classical and Bayesian approach under progressive type-II hybrid censoring scheme. The EM algorithm is used to compute maximum likelihood estimators, and Lindley's approximation and MCMC techniques are used to compute Bayes estimators. In the case of exponentiated exponential distribution, [1] developed the progressive stress model for type-II progressive hybrid censoring. The Bayes estimates and maximum likelihood estimates have been compared, and normal approximation and bootstrap confidence intervals for the unknown model parameters are developed.

3.4. Adaptive progressive censoring scheme

A mixture of type-I and type-II progressive censoring schemes is known as the adaptive progressive censoring scheme introduced by [30]. In this case, the progressive scheme and effective sample size k are predetermined; however, the number of items gradually eliminated from the experiment at each failure may vary during the experiment being conducted. The experiment can be terminated as soon as possible by adjusting the number of items progressively withdrawn from the experiment at each failure in such a way that the intended level of efficiency of the estimate can be achieved. This will happen if the experimental time exceeds a pre-fixed time T, but the number of observed failures has not yet reached k.

3.4.1 Methodology

Fix integers k, n such that $k < n$ where k is the effective sample size, and n is the total number of observations in a life test. Let $(D_1, D_2, ..., D_n)$ be the progressive censoring scheme, which is fixed in advance before the test, but the values of some D_i's may change in an adaptive way during the test procedure. Let time T be the ideal time of the experiment or the time fixed by the researcher to end the process. Let k be the completely observed failure times by $Y_{i:k:n}$ where $i = 1, \cdots, k$. The experiment is terminated at time $Y_{k:k:n}$ if the k^{th} progressively censored failure time occurs prior to time point T. That is, $Y_{i:m:n} < T$. If not, end the experiment as soon as possible to meet a certain inferential efficiency requirement after the experimental duration has passed time T, but the number of observed failures has not yet reached k. As a result, this arrangement can be considered as a design where the optimal situation would be to have k observed failure times for the efficiency of the inference and, at the same time, ensure the overall test time is slightly similar to the ideal test period T.

Some important works under the adaptive progressive censoring scheme are mentioned as follows. The statistical inferences under adaptive progressive hybrid censoring schemes for the model parameters of exponential and Weibull distributions have been computed by [30] and [27], respectively. In the case of two-parameter exponentiated Weibull distribution, [35] constructed the maximum likelihood estimates, reliability and hazard functions, approximate confidence intervals, Bayes estimates, and credible intervals under adaptive progressive type II censoring samples. Under the inverse Weibull distribution, [29] estimated both the frequentist and Bayesian estimates for the scale parameter λ and shape parameter β using an adaptive type-II progressive hybrid censoring scheme. The statistical inference for log-normal distribution under adaptive type-II progressive hybrid censoring scheme has been derived and analyzed by [19].

4. Applications and future scope

Censored data or various censoring schemes in survival analysis are applicable in a diverse range of fields and practical scenarios. In engineering and industrial scenarios, censoring techniques are used to analyze the reliability of systems or components that are subjected to different stress levels. This helps determine the causes of failure and product lifetimes. Censoring schemes are also used for acceptance sampling and quality control. This allows for the development of the best possible sample plans and appropriate choices based on only partially observed data. Censoring techniques are used in clinical trials and medical research to examine patient survival times under various treatment regimens, which helps assess the efficacy of treatments and patient outcomes. In genetic research, censoring techniques are used to assess the effects of penalized or censored data on genetic parameters, breeding values, and computational effectiveness for duration indicators. In cardiovascular research, censoring techniques are used to evaluate the correlation between offspring and parental age at the onset of cardiovascular events. In order to examine the sustainability and resilience of the environment, censoring schemes are used in environmental studies to analyze the lifespan of ecological systems or environmental components under changing conditions.

5. Conclusions

The purpose of life testing in reliability theory is to examine a wide range of items and collect data relevant to some or all of their lifetimes, which facilitates quality assurance and risk assessment and hence, thereby, improves product satisfaction and warranty estimations. This, in turn, satisfies a customer or buyer who would expect the items to operate as intended for an appropriate length of time. In this paper, the concepts, data forms, methodology, and applications of various censoring schemes are explained in detail. Censoring schemes can improve the quality of goods in various businesses by providing several benefits to customers. Partially observed data can be analyzed through censoring algorithms, which is useful when complete failure data is not obtainable. This renders more accurate statistical analysis and inference on the performance and dependability of products. Censoring techniques can be used to evaluate competing risks and determine the effect of each component on product quality in scenarios where multiple risk variables can lead to product failure. More specific plans for quality enhancement can result from this approach. During the testing process, adaptive censoring methods, including adaptive progressive censoring, enable dynamic changes based on observed data. This flexibility can result in quicker, better-informed decisions that improve final output quality. By utilizing censoring schemes effectively, manufacturers can collect substantial information about product performance, make well-informed decisions to improve quality, and eventually provide customers with more reliable and durable products.

References

[1] Alaa H Abdel-Hamid and Tahani A Abushal. Inference on progressive-stress model for the exponentiated exponential distribution under type-ii progressive hybrid censoring. *Journal of Statistical Computation and Simulation*, 85(6):1165–1186, 2015.

[2] Osama Abo-Kasem and Ahmed Elshahhat. A new two sample generalized type-ii hybrid censoring scheme. *American Journal of Mathematical and Management Sciences*, 41(2):170–184, 2022.

[3] Hanaa H Abu-Zinadah and Neveen Sayed-Ahmed. Competing risks model with partially step-stress accelerate life tests in analyses lifetime chen data under type-ii censoring scheme. *Open Physics*, 17(1):192–199, 2019.

[4] Tahani A Abushal, AA Soliman, and GA Abd-Elmougod. Inference of partially observed causes for failure of lomax competing risks model under type-ii generalized hybrid censoring scheme. *Alexandria Engineering Journal*, 61(7):5427–5439, 2022.

[5] Ali Algarni, Abdullah M Almarashi, and Gamal A Abd-Elmougod. Joint type-i generalized hybrid censoring for estimation two weibull distributions. *Journal of Information Science & Engineering*, 36(6), 2020.

[6] Anestis Antoniadis, Gérard Grégoire, and Guy Nason. Density and hazard rate estimation for right-censored data by using wavelet methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(1):63–84, 1999.

[7] Saieed F Ateya. Estimation under inverse weibull distribution based on balakrishnan's unified hybrid censored scheme. *Communications in Statistics-Simulation and Computation*, 46(5):3645–3666, 2017.

[8] N Balakrishnan, Abbas Rasouli, and N Sanjari Farsipour. Exact likelihood inference based on an unified hybrid censored sample from the exponential distribution. *Journal of Statistical Computation and Simulation*, 78(5):475–488, 2008.

[9] Narayanaswamy Balakrishnan and Rita Aggarwala. *Progressive censoring: theory, methods, and applications*. Springer Science & Business Media, 2000.

[10] B Chandrasekar, A Childs, and N Balakrishnan. Exact likelihood inference for the exponential distribution under generalized type-i and type-ii hybrid censoring. *Naval Research Logistics (NRL)*, 51(7):994–1004, 2004.

[11] Shu-Mei Chen and Gouri K Bhattacharyya. Exact confidence bounds for an exponential parameter under hybrid censoring. *Communications in Statistics-Theory and Methods*, 16(8):2429–2442, 1987.

[12] A Childs, B Chandrasekar, and N Balakrishnan. Exact likelihood inference for an exponential parameter under progressive hybrid censoring schemes. *Statistical models and methods for biomedical and technical systems*, pages 319–330, 2008.

[13] A Childs, B Chandrasekar, N Balakrishnan, and D Kundu. Exact likelihood inference based on type-i and type-ii hybrid censored samples from the exponential distribution. *Annals of the Institute of Statistical Mathematics*, 55:319–330, 2003.

[14] Muhammad Yameen Danish and Muhammad Aslam. Bayesian inference for the randomly censored weibull distribution. *Journal of Statistical Computation and Simulation*, 84(1):215–230, 2014.

[15] Sanku Dey, Liang Wang, and Mazen Nassar. Inference on nadarajah–haghighi distribution with constant stress partially accelerated life tests under progressive type-ii censoring. *Journal of Applied Statistics*, 49(11):2891–2912, 2022.

[16] TI Elperin and IB Gertsbakh. Estimation in a random censoring model with incomplete information: exponential lifetime distribution. *IEEE Transactions on Reliability*, 37(2):223–229, 1988.

[17] Benjamin Epstein. Truncated life tests in the exponential case. *The Annals of Mathematical Statistics*, pages 555–564, 1954.

[18] Renu Garg, Madhulika Dube, Kapil Kumar, and Hare Krishna. On randomly censored generalized inverted exponential distribution. *American Journal of Mathematical and Management Sciences*, 35(4):361–379, 2016.

[19] Fariba Hemmati and Esmaile Khorram. Statistical analysis of the log-normal distribution under type-ii progressive hybrid censoring schemes. *Communications in Statistics-simulation and Computation*, 42(1):52–75, 2013.

[20] Jian Huang. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568, 1996.

[21] Jian Huang and Jon A Wellner. Interval censored survival data: a review of recent progress. In *Proceedings of the first Seattle symposium in biostatistics: survival analysis*, pages 123–169. Springer, 1997.

[22] Young Eun Jeon and Suk-Bok Kang. Estimation of the rayleigh distribution under unified hybrid censoring. *Austrian Journal of Statistics*, 50(1):59–73, 2021.

[23] John P Klein, Melvin L Moeschberger, et al. *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer, 2003.

[24] John P Klein, Hans C Van Houwelingen, Joseph George Ibrahim, and Thomas H Scheike. *Handbook of survival analysis*. CRC Press Boca Raton, FL:, 2014.

[25] Debasis Kundu. On hybrid censored weibull distribution. *Journal of Statistical Planning and Inference*, 137(7):2127–2142, 2007.

[26] Debasis Kundu and Avijit Joarder. Analysis of type-ii progressively hybrid censored data. *Computational Statistics & Data Analysis*, 50(10):2509–2528, 2006.

[27] Chien-Tai Lin, Hon Keung Tony Ng, and Ping Shing Chan. Statistical inference of type-ii progressively hybrid censored data with weibull lifetimes. *Communications in Statistics—Theory and Methods*, 38(10):1710–1729, 2009.

[28] Showkat Ahmad Lone and Hanieh Panahi. Estimation procedures for partially accelerated life test model based on unified hybrid censored sample from the gompertz distribution. *Eksploatacja i Niezawodność*, 24(3), 2022.

[29] Mazen Nassar and OE Abo-Kasem. Estimation of the inverse weibull parameters under adaptive type-ii progressive hybrid censoring scheme. *Journal of computational and applied mathematics*, 315:228–239, 2017.

[30] Hon Keung Tony Ng, Debasis Kundu, and Ping Shing Chan. Statistical analysis of exponential lifetimes under an adaptive type-ii progressive censoring scheme. *Naval Research Logistics (NRL)*, 56(8):687–698, 2009.

[31] M Noori Asl, R Arabi Belaghi, and H Bevrani. On burr xii distribution analysis under progressive type-ii hybrid censored data. *Methodology and Computing in Applied Probability*, 19:665–683, 2017.

[32] Hanieh Panahi and Abdolreza Sayyareh. Estimation and prediction for a unified hybrid-censored burr type xii distribution. *Journal of Statistical Computation and Simulation*, 86(1):55–73, 2016.

[33] Sanjay Kumar Singh, Umesh Singh, and Vikas Kumar Sharma. Bayesian estimation and prediction for flexible weibull model under type-ii censoring scheme. *Journal of Probability and Statistics*, 2013(1):146140, 2013.

[34] Umesh Singh, Pramod K Gupta, and Satyanshu K Upadhyay. Estimation of parameters for exponentiated-weibull family under type-ii censoring scheme. *Computational statistics & data analysis*, 48(3):509–523, 2005.

[35] Mashail M AL Sobhi and Ahmed A Soliman. Estimation for the exponentiated weibull model with adaptive type-ii progressive censored schemes. *Applied Mathematical Modelling*, 40(2):1180–1192, 2016.

PERFORMANCE MODELING OF CRYSTALLIZATION SYSTEM IN SUGAR PLANT USING RAMD APPROACH

Ravi choudhary[1], Vijay Singh Maan[2*], Ashish Kumar[3], Monika Saini[4]

•

[1,2,3,4]Department of Mathematics and Statistics, Manipal University Jaipur, Rajasthan, India
Email: ravifunspace@gmail.com, ashishbarak2020@gmail.com, drmnksaini4@gmail.com
[*]Corresponding Author Email: vsmaan06@gmail.com

Abstract

The aim of the present study is to investigate reliability, availability, maintainability, and dependability (RAMD) of crystallization system of a sugar production plant. Previous studies attentive on the reliability and availability analysis of sugar plants specially its subsystems like evaporation units. This study is focus on the RAMD analysis of the crystallization system of sugar plant having four subsystems with different number of components. Failure and repair rates of all subsystems are taken as exponentially distributed. The transition diagram and Chapman-Kolmogorov differential equations for each subsystem are derived by using Markov birth-death process. For all four subsystems, reliability, availability, mean time between failure (MTBF), mean time to repair (MTTR), and dependability ratio are computed using simple probabilistic concepts. The effect of change in failure rates of subsystem in system performance is also observed. It is shown that the crystallization subsystem found to be more sensitive among four subsystems from reliability point of view. This study can be helpful to system designer for further modeling/designing of reliable systems and enhancement in system's performance through planning efficient maintenance strategies.

Keywords: Mean Time Between Failures, Reliability, Availability, Maintainability, Dependability

I. Introduction

India's economy is primarily agrarian. Agriculture accounts for roughly 20% of the GDP. Agriculture encompasses activities such as cultivating crops, raising poultry, fishing, breeding cattle, and practicing animal husbandry. These actions are critical to our country's survival. The Indian economy has expanded significantly in recent decades. The improvement of agriculture and allied operations to meet international standards has also resulted in a rise in the export of various food products, which has fuelled economic growth. One of the most well-known industries is sugar, of which India is now the world's largest producer, consumer, and exporter. In India, the sugar industry is an agriculture-based industry that has a big influence on the rural economy. Millions of workers and farmers who cultivate sugarcane are impacted by this industry. The sugar manufacturing process involves multiple stages such as extraction, clarification, boiling, crystallization, centrifuging, grading, weighing, and bagging. Achieving maximum production levels and ensuring high system availability are crucial in the manufacturing sector. Industries are configured with several different types of heavy machinery. The primary concern of manufacturing industries is the reliability of these machines. RAMD techniques are widely used by researchers to evaluate the systems performance. This technique effectively assesses the reliability and availability

of individual components in complex systems. Nowadays, systems are becoming increasingly complex in structure. So, it became very necessary to identify the most critical component and carry forward the maintenance strategies on time for a flawless process of production.

Bradley and Dawson [1] Rolls-Royce study discovered that most PC components failed more frequently early in life, resulting in higher beginning operating costs, and suggested a rolling replacement approach to spread costs fairly. Blischke and Murthy [2] investigated dependability, maintenance, maintainability, and quality, focusing on practical challenges within these domains. Each example emphasized reliability and practical project implications. Bhamare et al. [3] evaluated historical achievements in reliability engineering, investigated statistical and fuzzy logic methodologies, and indicated limitations and potential for future research in reliability analysis. Sharma and Kumar [4] highlighted the significance of the RAM approach in modeling engineering systems and enhancing performance. Adhikary et al. [5] investigated RAM indices of 210-megawatt coal-fired thermal power stations to improve the availability of power plants. Sharma and Sharma [6] proposed a MSDM (multi-stage decision-making) model to incorporate a framework for optimizing RAM and cost decisions in a process plant. Kumar [7] created many stochastic computer system models based on the notion of maximum operation and maintenance times. Sharma and Khanduja [8] discussed the efficacy and availability of feeding systems in the sugar industry. Sharma and Vishwakarma [9] emphasized the use of Markov processes and optimization in the refining system of the sugar industry to provide maximum system productivity.

Aggarwal et al. [10] developed a performance model utilizing the RAMD approach for manufacturing skim milk powder systems. This research assisted in identifying the crucial subsystem and its impact on the system's performance under actual operating conditions. Kadyan and Kumar [11] used the SVA (supplementary variable approach) and the Markov process to analyze the availability and profitability of a feeding system in the sugar sector. Ram and Kumar [12] investigated the performability of a system using the 1-out-of-2: G strategy and evaluated the reliability measures for each subsystem. Parida et al. [13] conducted a thorough evaluation of the literature on performance measurement and management in maintenance. Kumar and Saini [14] suggested a sugar plant mathematical model to assess availability using a fuzzy reliability technique. Kadyan and Kumar [15] analyzed the operational behavior of availability and expected profit analysis of a B-Pan crystallization system by using the Markovian technique. Tsarouhas and Besseris [16] provided comprehensive maintainability analysis of the shaving blade section of a high-tech razor manufacturing plant and focused on identifying the areas for improvement. Tsarouhas [17] developed RAM analysis to improve the performance of the wine packaging line by using datasets from the production system. Choudhary et al. [18] examined the effectiveness of RAM analysis for capacity improvement of a cement plant. Dahiya et al. [19] analyzed the performance and profit analysis of the feeding system of sugar plants by using the concept of coverage factor.

Saini et al. [20] derived the reliability, availability, maintainability, and dependability of a microprocessor system made up of seven subsystems, utilizing state transition diagrams and Markov processes to calculate important performance measures. Kumar et al. [21] developed a stochastic model to carry out RAMD analysis and FME (failure mode and effect) analysis of tube-well integrated pipelines. Saini et al. [22] studied failure patterns, best-fit distributions, and suggested maintenance solutions of a sugar plant on the basis of six months data. Gao [23] analyzed a fault-tolerant system with warm standbys, determining its steady-state availability, reliability function, and mean time to first failure using Markov theory and Laplace transforms. Yusuf et al. [24] analyzed two hybrid systems using the RAMD framework and found that evaluating and comparing these systems helped identify improvement opportunities for enhancing operational efficiency and productivity.

This study is focused on the RAMD analysis of the crystallization system of the sugar plant, which has four subsystems with different numbers of components. Failure and repair rates of all subsystems are taken as exponentially distributed. The transition diagram and Chapman-Kolmogorov differential equations for each subsystem are derived by using the Markov birth-death

process. For all four subsystems, reliability, availability, mean time between failure (MTBF), mean time to repair (MTTR), and dependability ratio are computed using simple probabilistic concepts. The effect of changes in failure rates of subsystems on system performance is also observed. It is shown that the crystallization subsystem was found to be more sensitive among four subsystems from a reliability point of view. Findings of this study can assist system designers in developing reliable systems and improving performance by implementing efficient maintenance strategies.

The whole manuscript is divided into five sections. An introduction to the system is appended in Section 1. Materials, methods, and system description are presented in Section 2. Section 3 incorporates the mathematical modeling and RAMD analysis of the system, and results are shown in tabular form. Section 4 covers the discussion and conclusion part of the study.

II. Material and Methods

The techniques used for investigation are described as follows:

I. Notations

The following nomenclature is used to develop the state transition diagram and mathematical modeling of system.

Table 1: *Notations for paint manufacturing plant's sub-system*

○	System is in working condition with full capacity
□	Failure state of system
U, Z, G, W	Fully working states of the subsystems
U_1, U_2, Z_1, Z_2	States where one or two failed unit of subsystem A and B goes under repair
u, z, g, w	Completely failed states of subsystems
$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	Constant failure rates of subsystems A, B, C and D respectively
$\beta_1, \beta_2, \beta_3, \beta_4$	Constant repair rates of subsystems A, B, C and D respectively
$P_1(t)$	Probability of the initial state of the system working with full capacity
P_i ; i=1, 2, 3, 4	Steady state probability of ith state of the system
$f(x) = \begin{cases} \theta e^{-\theta x} & 0 \le x \le \infty \\ 0 & otherwise \end{cases}$	PDF of exponential distribution Here θ =constant rate in failure per unit of measurement
$R(t) = P(T < t) = \int_t^{\infty} f(x)dx \;\; or \;\; R(t) = e^{-\alpha t}$	Reliability function, here α = failure rate
$\frac{MTTF}{MTTF+MTTR} = \frac{Life\ time}{Life\ time+Repair\ time}$	Availability function
$M(t) = P(T < t) = 1 - e^{\frac{-t}{MTTR}}$	Maintainability function
$MTBF = \int_0^{\infty} R(t)dt = \int_0^{\infty} e^{-\alpha t}dt = \frac{1}{\alpha}$	Mean time between failure
$MTTR = \frac{1}{\beta}$	Mean time to repair, here β = Repair rate
$d = \frac{MTBF}{MTTR} = \frac{\beta}{\alpha}$	Dependability ratio
$D_{min} = 1 - \left(\frac{1}{d-1}\right)\left(e^{\frac{-In\ d}{(d-1)}} - e^{\frac{-d\ In\ d}{(d-1)}}\right)$	

II. System Description

This section contains the detailed description of crystallization unit of a sugar production plant. It is a prominent part of process in which raw sugar syrup converted in different size of crystal and impurities are separated by different processes. All subsystems are arranged in a series configuration. The representation of components is in figure 1.

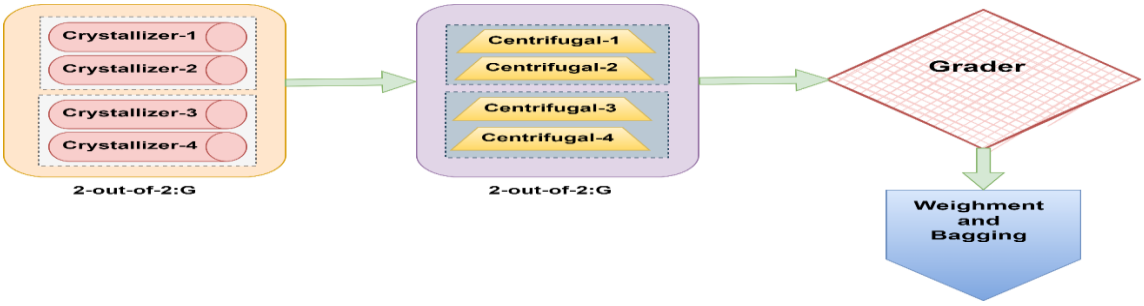


Figure 1: *Configuration block diagram of Crystallization System*

i) Crystallizer (A)

The raw sugar syrup which is produced after evaporation process now transferred to large rotating vessels for cooldown evenly. The next step is seeding, which is done in three steps. The initial crystallization produces crystallized sugar and molasses, or residuals. When the molasses and crystals separate, the liquid is prepared for the next step. This subsystem is configured as 2-out-of-2: G system. It consists of a total of four units, out of which two are in operation and two are in cold standby. The failure and repair rate of all the units are same. The failure of more than two units lead to the complete failure of the system. The differential equations of subsystem crystallizer are calculated by using state transition diagram given in figure 2.

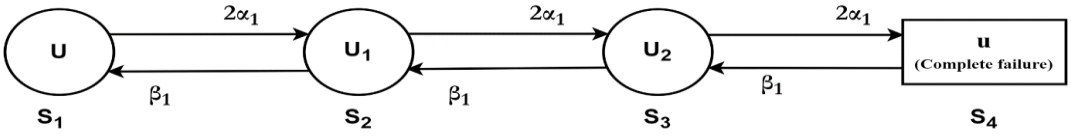


Figure 2: *State transition diagram of crystallizer subsystem*

ii) Centrifugal machine (B)

In this process the liquid syrup is separated from the sugar crystals and then the syrup coating is removed by using fine jet of water. The centrifuged raw sugar contains 97-99% of sucrose and 0.5% of moisture. The amount of molasses left on the crystals determines the type of sugar produced. This can be further stored in bags or bulk. This subsystem is also configured as 2-out-of-2: G system. It consists of a total of four units with same failure and repair rates. Among four units, two are in operation and two are on cold standby. The system faces complete failure if more than two units goes under failure. The differential equations of subsystem centrifugal machine are calculated by using state transition diagram given in figure 3.

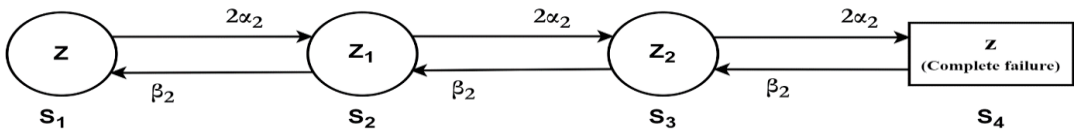


Figure 3: *State transition diagram of centrifugal machine*

iii) Grader (C)

The resulting sugar is made up of heterogeneous crystals and must be sieved and graded before it's packaging. The primary goal is particle size classification, which is accomplished using screens. Typically, classification is accomplished by using wire mesh or perforated plate through which particles smaller than the screen aperture may pass while the biggest fraction is carried over the surface. This subsystem has single unit in operation with constant failure and repair rates. The failure of this unit leads to the failure of entire system. The differential equations of subsystem grader are calculated by using state transition diagram given in figure 4.

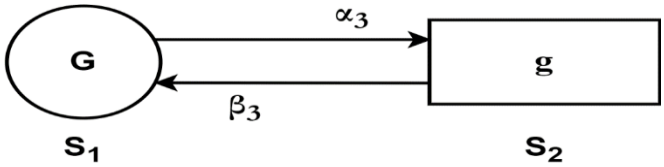


Figure 4: *State transition diagram of grader*

iv) Weighment and bagging (D)

Some of the most difficult aspects of bagging sugar include clean filling, dust reduction, and equipment cleanliness. Raw sugar is carried as both bulk and break-bulk freight. Raw sugar is packaged as break-bulk cargo in bags made of woven natural materials (such as jute) or woven plastic bags with a plastic inner liner that is impermeable to water vapor and provides contamination prevention. This subsystem has single unit in operation with constant failure and repair rates. The failure of this unit leads to the complete failure of system. The differential equations of subsystem grader are calculated by using state transition diagram given in figure 5.

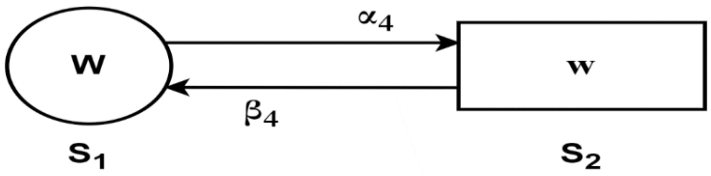


Figure 5: *State transition diagram of weighment and bagging*

III. Assumptions

The system is assumed to worked under the following conditions:

- At time t=0, all subsystems are functioning properly without facing any failure.
- All the failure and repair rates are chosen as arbitrary and distributed exponentially.

- Cold standby redundancy is operated at the component level for subsystem crystallizer and centrifugal machine.
- Repairs are flawless. Failed unit works properly as new after repair.

Table 1: *Failure and repair rates of subsystems*

Sr. No.	Sub-system	Failure-rate (α)	Repair-rate (β)
1	Crystallizer (A)	α_1 = 0.0095	β_1 = 0.53
2	Centrifugal machine (B)	α_2 = 0.0088	β_2 = 0.69
3	Grader (C)	α_3 = 0.0075	β_3 = 0.71
4	Weighment and bagging (D)	α_4 = 0.0091	β_4 = 0.81

III. Mathematical Modeling and RAMD Analysis

In this section, RAMD investigation through development of mathematical models of crystallization subsystem of sugar production plant is investigated. The Markov birth-death process is used to derive the Chapman-Kolmogorov differential equations. Transition diagram for each four subsystems is displayed in figures 2-5. Failure and repair rates of the subsystems are taken as exponential distributed and appended in table 1. The RAMD indices of all four subsystems are shown in table 2. Table 3 shows the variation in reliability with respect of time. The effect of change in failure rates of different subsystems on system's performance, are appended in tables 4-8.

i) RAMD indices for subsystem 1 (SS_1)

This subsystem named crystallizer and is configured as 2-out-of-2: G and has two standby units. If one of the operating units fails, then one standby unite comes in operation and if both working unit goes down then both cold standby units come in operation. Failure of any of the single unit after this leads to complete system failure. The failure and repair rates of units are same. The Chapman-Kolmogorov differential equations are derived by using figure 1 and relations are given below

$$P_1'(t) = -2\alpha_1 P_1(t) + \beta_1 P_2(t) \quad (1)$$

$$P_2'(t) = -(2\alpha_1 + \beta_1)P_2(t) + 2\alpha_1 P_1(t) + \beta_1 P_3(t) \quad (2)$$

$$P_3'(t) = -(2\alpha_1 + \beta_1)P_3(t) + 2\alpha_1 P_2(t) + \beta_1 P_4(t) \quad (3)$$

$$P_4'(t) = -\beta_1 P_4(t) + 2\alpha_1 P_3(t) \quad (4)$$

By using initial condition and $t \to \infty$ we get reduced equations

$$-2\alpha_1 P_1 + \beta_1 P_2 = 0 \quad (5)$$

$$-(2\alpha_1 + \beta_1)P_2 + 2\alpha_1 P_1 + \beta_1 P_3 = 0 \quad (6)$$

$$-(2\alpha_1 + \beta_1)P_3 + 2\alpha_1 P_2 + \beta_1 P_4 = 0 \quad (7)$$

$$-\beta_1 P_4 + 2\alpha_1 P_3 = 0 \quad (8)$$

By using normalization condition

$$P_1 + P_2 + P_3 + P_4 = 1 \quad (9)$$

After putting values of P_2, P_3 and P_4 in terms of P_1 in equation (9), we get

$$P_1 = \frac{1}{1+\frac{2\alpha_1}{\beta_1}+\frac{4{\alpha_1}^2}{{\beta_1}^2}+\frac{8{\alpha_1}^3}{{\beta_1}^3}} \quad (10)$$

Availability of SS_1 will be

$$A_{ss_1} = P_1 + P_2 + P_3$$

$$A_{ss_1} = \frac{1+\frac{2\alpha_1}{\beta_1}+\frac{4{\alpha_1}^2}{{\beta_1}^2}}{1+\frac{2\alpha_1}{\beta_1}+\frac{4{\alpha_1}^2}{{\beta_1}^2}+\frac{8{\alpha_1}^3}{{\beta_1}^3}} \quad (11)$$

After putting values of failure and repair rates

$$A_{ss_1} = 0.9999556 \quad (12)$$

Reliability of SS_1 is driven by using the formula,

$$R_{ss_1}(t) = e^{-0.057\times t} \tag{13}$$

The maintainability of SS_1 is,

$$M_{ss_1}(t) = 1 - e^{-1283.15\times t} \tag{14}$$

Now some different measures of system effectiveness of SS_1 are derived by using equations mention above in notation section are as follows, MTBF= 17.5438596h, MTTR= 0.0007793h, d=22511.94 and $D_{min}(ss_1)$=0.999956

ii) RAMD indices for subsystem 2 (SS_2)

Subsystem 2 is a centrifugal machine and has two operating unit and two standby units. The failure of more than two unit is considered as the complete failure of system. The differential equations are derived by using figure 2 and different reliability measures are derived.

$$P_1'(t) = -2\alpha_2 P_1(t) + \beta_2 P_2(t) \tag{15}$$
$$P_2'(t) = -(2\alpha_2 + \beta_2)P_2(t) + 2\alpha_2 P_1(t) + \beta_2 P_3(t) \tag{16}$$
$$P_3'(t) = -(2\alpha_2 + \beta_2)P_3(t) + 2\alpha_2 P_2(t) + \beta_2 P_4(t) \tag{17}$$
$$P_4'(t) = -\beta_2 P_4(t) + 2\alpha_2 P_3(t) \tag{18}$$

By using initial condition and $t \to \infty$ we get

$$-2\alpha_2 P_1 + \beta_2 P_2 = 0 \tag{19}$$
$$-(2\alpha_2 + \beta_2)P_2 + 2\alpha_2 P_1 + \beta_2 P_3 = 0 \tag{20}$$
$$-(2\alpha_2 + \beta_2)P_3 + 2\alpha_2 P_2 + \beta_2 P_4 = 0 \tag{21}$$
$$-\beta_2 P_4 + 2\alpha_2 P_3 = 0 \tag{22}$$

By using normalization condition

$$P_1 + P_2 + P_3 + P_4 = 1 \tag{23}$$

After putting values of P_2, P_3 and P_4 in terms of P_1 in equation (23), we get

$$P_1 = \frac{1}{1+\frac{2\alpha_2}{\beta_2}+\frac{4{\alpha_2}^2}{{\beta_2}^2}+\frac{8{\alpha_2}^3}{{\beta_2}^3}} \tag{24}$$

Availability expression for SS_2 will be

$$A_{ss_2} = P_1 + P_2 + P_3 \tag{25}$$

$$A_{ss_2} = \frac{1+\frac{2\alpha_2}{\beta_2}+\frac{4{\alpha_2}^2}{{\beta_2}^2}}{1+\frac{2\alpha_2}{\beta_2}+\frac{4{\alpha_2}^2}{{\beta_2}^2}+\frac{8{\alpha_2}^3}{{\beta_2}^3}} \tag{26}$$

After putting values of failure and repair rates

$$A_{ss_2} = 0.9999838 \tag{27}$$

Reliability of SS_2 is given by

$$R_{ss_2}(t) = e^{-0.0528\times t} \tag{28}$$

The maintainability of SS_2 is,

$$M_{ss_2}(t) = 1 - e^{-3264.81\times t} \tag{29}$$

Now, some different measures of system effectiveness of SS_2 are as follow, MTBF= 18.9394h, MTTR= 0.0003063h, d=61833.4 and $D_{min}(ss_2)$=0.999984

iii) RAMD indices for subsystem 3 (SS_3)

Here, grader machine is taken as a subsystem 3. It has only one unit in operation and failure of this unit leads to complete system failure. The differential equations are derived by using figure 3. The relations are as follows,

$$P_1'(t) = -\alpha_3 P_1(t) + \beta_3 P_2(t) \quad (30)$$

$$P_2'(t) = -\beta_3 P_2(t) + \alpha_3 P_1(t) \quad (31)$$

using initial condition $t \to \infty$ and we get

$$-\alpha_3 P_1 + \beta_3 P_2 = 0 \quad (32)$$

$$-\beta_3 P_2 + \alpha_3 P_1 = 0 \quad (33)$$

using normalization condition

$$P_1 + P_2 = 1 \quad (34)$$

After putting values of P_2 in terms of P_1 in equation (34), we get

$$P_1 = \frac{\beta_3}{\beta_3 + \alpha_3} \quad (35)$$

Availability expression for SS_3 will be

$$A_{ss_3} = P_1 \quad (36)$$

$$A_{ss_3} = \frac{\beta_3}{\beta_3 + \alpha_3} \quad (37)$$

After putting values of failure and repair rates

$$A_{ss_3} = 0.989547 \quad (38)$$

Reliability of SS_3 is,

$$R_{ss_3}(t) = e^{-0.0075 \times t} \quad (39)$$

The maintainability of SS_3 is,

$$M_{ss_3}(t) = 1 - e^{-0.71 \times t} \quad (40)$$

Now some different measures of system effectiveness of SS_3 are calculated as follow, MTBF= 133.333h, MTTR= 1.4084507h, d=94.6667 and $D_{min}(ss_3)$=0.990995

iv) RAMD indices for subsystem 4 (SS_4)

Weighment and bagging machine are considered as subsystem 4. It consists single unit in operation and failure of this unit leads to complete system failure. The differential equations are derived by using figure 4. The relations are given below,

$$P_1'(t) = -\alpha_4 P_1(t) + \beta_4 P_2(t) \quad (41)$$

$$P_2'(t) = -\beta_4 P_2(t) + \alpha_4 P_1(t) \quad (42)$$

using initial condition $t \to \infty$ and we get,

$$-\alpha_4 P_1 + \beta_4 P_2 = 0 \quad (43)$$

$$-\beta_4 P_2 + \alpha_4 P_1 = 0 \quad (44)$$

using normalization condition

$$P_1 + P_2 = 1 \quad (45)$$

After putting values of P_2 in terms of P_1 in equation (45), we get

$$P_1 = \frac{\beta_4}{\beta_4 + \alpha_4} \quad (46)$$

Availability expression for SS_4 will be,

$$A_{ss_4} = P_1 \quad (47)$$

$$A_{ss_4} = \frac{\beta_4}{\beta_4 + \alpha_4} \quad (48)$$

After putting values of failure and repair rates

$$A_{ss_4} = 0.9888902 \quad (49)$$

Reliability of SS_4 is driven as,

$$R_{ss_4}(t) = e^{-0.0091 \times t} \quad (50)$$

The maintainability of SS_4 is,

$$M_{ss_4}(t) = 1 - e^{-0.81 \times t} \quad (51)$$

Now some different measures of system effectiveness of SS_4 are derived as, MTBF= 109.89h, MTTR= 1.234567h, d=89.011 and $D_{min}(ss_4)$=0.9904687

v) System's reliability

The proposed subsystems are connected in series combination and failure of any one of them leads to the complete failure of the system. The overall reliability of the system is derived by,

$$R_{System}(t) = R_{ss_1}(t) \times R_{ss_2}(t) \times R_{ss_3}(t) \times R_{ss_4}(t) \quad (52)$$

By putting values in equation (52), we get system reliability as,

$$R_{System}(t) = e^{-0.1264(t)} \quad (53)$$

The variation in reliability with respect to time is derived and shown in table 3.

vi) System's availability

Here four subsystems are connected in series combination and the availability of each one is calculated separately. Now the availability of the entire system is derived and the expression is as follow

$$A_{System}(t) = A_{ss_1}(t) \times A_{ss_2}(t) \times A_{ss_3}(t) \times A_{ss_4}(t) \quad (54)$$

By putting values in equation (54), we get the availability of the system as below,

$$A_{System}(t) = 0.9784944 \quad (55)$$

vii) System's maintainability

The four subsystems are linked in series and the failure of one cause the entire system to fail. The maintainability of entire system is calculated and is given below

$$M_{System}(t) = M_{ss_1}(t) \times M_{ss_2}(t) \times M_{ss_3}(t) \times M_{ss_4}(t) \quad (56)$$

$$M_{System}(t) = 1 - e^{-2409232.47 \times t} \quad (57)$$

viii) System's dependability

The dependability of entire system is derived by multiplying the dependability of each subsystem. It is given by

$$D_{min(system)}(t) = D_{min(ss_1)}(t) \times D_{min(ss_2)}(t) \times D_{min(ss_3)}(t) \times D_{min(ss_4)}(t) \quad (58)$$

$$D_{min(system)}(t) = 0.981491 \quad (59)$$

Table 2: *RAMD indices for subsystems of sugar plant systems*

Subsystems	**Ss1**	**Ss2**	**Ss3**	**Ss4**	**System**
Reliability	$e^{-0.057\times t}$	$e^{-0.0528\times t}$	$e^{-0.0075\times t}$	$e^{-0.0091\times t}$	$e^{-0.1264(t)}$
Availability	0.9999556	0.9999838	0.989547	0.9888902	0.9784944
Maintainability	$1-e^{-1283.15\times t}$	$1-e^{-3264.81\times t}$	$1-e^{-0.71\times t}$	$1-e^{-0.81\times t}$	$1-e^{-2409232.47\times t}$
Dependability	0.999956	0.999984	0.990995	0.9904687	0.981491
MTBF	17.5439h	18.9394h	133.333h	109.89h	279.7059h
MTTR	0.0007793h	0.0003063h	1.408451h	1.234567h	2.644103h
Dependability ratio (d)	d=22511.94	61833.4	94.6667	89.011	

Table 3: *Variation in reliability with respect to time*

Time (days)	**Ss1**	**Ss2**	**Ss3**	**Ss4**	**system**
0	1	1	1	1	1
30	0.7520143	0.7679735	0.7985162	0.7610928	0.3509892

60	0.5655254	0.5897834	0.6376282	0.5792622	0.1231934
90	0.4252832	0.452938	0.5091564	0.4408723	0.0432395
120	0.319819	0.3478444	0.4065697	0.3355447	0.0151766
150	0.2405085	0.2671353	0.3246525	0.2553807	0.0053268

Table 4: *Variation of maintainability with respect time*

Time (days)	M_{ss_1}	M_{ss_2}	M_{ss_3}	M_{ss_4}	M_{system}
0	0	0	0	0	0
30	1	1	0.999999999	1	1
60	1	1	1	1	1
90	1	1	1	1	1
120	1	1	1	1	1
150	1	1	1	1	1

Table 5: *Effect of change in failure rates on subsystem and system reliability*

Time (days)	Subsystem 1				System			
	α_1=0.0075	α_1=0.0085	α_1=0.0095	α_1=0.0105	α_1=0.0075	α_1=0.0085	α_1=0.0095	α_1=0.0105
30	0.7985162	0.7749165	0.7520143	0.7297889	0.3726931	0.3616783	0.3509891	0.3406158
60	0.6376282	0.6004956	0.5655254	0.5325918	0.1389002	0.1308113	0.1231934	0.1160192
90	0.5091564	0.4653339	0.4252832	0.3886796	0.0517671	0.0473116	0.0432396	0.0395180
120	0.4065697	0.3605949	0.319819	0.283654	0.0192933	0.0171116	0.0151766	0.0134604
150	0.3246525	0.279431	0.2405085	0.2070076	0.0071905	0.0061889	0.0053268	0.0045848

Table 6: *Effect of change in failure rates on subsystem and system reliability*

Time (days)	Subsystem 2				System			
	α_2=0.0068	α_2=0.0078	α_2=0.0088	α_2=0.0098	α_2=0.0068	α_2=0.0078	α_2=0.0088	α_2=0.0098
30	0.8154624	0.7913618	0.7679735	0.7452765	0.3726931	0.3616784	0.3509891	0.3406159
60	0.6649789	0.6262535	0.5897834	0.555437	0.1389002	0.1308112	0.1231934	0.1160191
90	0.5422653	0.4955931	0.452938	0.4139542	0.0517671	0.0473116	0.0432395	0.0395180
120	0.4421969	0.3921935	0.3478444	0.3085103	0.0192932	0.0171116	0.0151766	0.0134604
150	0.3605949	0.3103669	0.2671353	0.2299255	0.0071905	0.0061889	0.0053268	0.0045848

Table 7: *Effect of change in failure rates on subsystem and system reliability*

Time (days)	Subsystem 3				System			
	α_3=0.0055	α_3=0.0065	α_3=0.0075	α_3=0.0085	α_3=0.0055	α_3=0.0065	α_3=0.0075	α_3=0.0085
30	0.8478937	0.8228347	0.7985162	0.7749165	0.3726931	0.3616784	0.3509891	0.3406158
60	0.7189237	0.6770569	0.6376282	0.6004956	0.1389001	0.1308112	0.1231934	0.1160191
90	0.6095709	0.5571059	0.5091564	0.4653339	0.0517671	0.0473116	0.0432395	0.0395179
120	0.5168513	0.458406	0.4065697	0.3605949	0.0192933	0.0171116	0.0151766	0.0134604
150	0.438235	0.3771924	0.3246525	0.279431	0.0071905	0.0061889	0.0053268	0.0045848

Table 8: *Effect of change in failure rates on subsystem and system reliability*

Time (days)	Subsystem 4				System			
	α_4=0.0071	α_4=0.0081	α_4=0.0091	α_4=0.0101	α_4=0.0071	α_4=0.0081	α_4=0.0091	α_4=0.0101
30	0.8081561	0.7842715	0.7610928	0.7385991	0.3726931	0.3616783	0.3509891	0.3406158
60	0.6531163	0.6150818	0.5792622	0.5455286	0.1389002	0.1308112	0.1231934	0.1160192
90	0.52782	0.4823911	0.4408723	0.4029269	0.0517671	0.0473116	0.0432395	0.0395180
120	0.426561	0.3783256	0.3355447	0.2976015	0.0192933	0.0171116	0.0151766	0.0134605
150	0.3447279	0.29671	0.2553807	0.2198082	0.0071905	0.0061889	0.0053268	0.0045849

IV. Discussion and conclusion

Reliability analysis of various subsystems and system has been performed for a particular subsystem of sugar production plant. It is observed that the reliability of the system for 60 days is 0.1231934 and its corresponding values for subsystems at time 60 days are R_{SS1}=0.5655254, R_{SS2}=0.5897834, R_{SS3}=0.6376282, and R_{SS4}=0.5792622 respectively. The reliability of subsystem crystallizer is very low at different time point with respect to other subsystems. It needs more attention and maintenance with careful observation. From table 5, 6, 7, and 8, the reliability of subsystems is highly influenced with respect to failure rates. The derived results help maintenance managers, system designers and engineers to properly analyse system performance and plan maintenance strategies

References:

[1] Bradley, M. and Dawson, R. (1998). The cost of unreliability: a case study. *Journal of Quality in Maintenance Engineering*, 4(3):212-218.

[2] Blischke, W.R. and Murthy, D.N.P. (2003). Case Studies in Reliability and Maintenance, John Wiley &Sons.

[3] Bhamare, S.S., Yaday, O.P. and Rathore, A. (2008). Evolution of reliability engineering discipline over the last six decades: a comprehensive review. *International Journal of Reliability and Safety*, 4(1):377-410.

[4] Sharma, R. K. and Kumar, S. (2008). Performance modeling in critical engineering systems using RAM analysis. *Reliability Engineering & System Safety*, 93:891–897.

[5] Adhikary, D. D., Bose, G. K., Chattopadhyay, S., Bose, D. and Mitra, S. (2012). RAM investigation of coal- fired thermal power plants: a case study. *International Journal of Industrial Engineering Computations*, 3:423–434.

[6] Sharma, K. R. and Sharma, P. (2012) Integrated framework to optimize RAM and cost decisions in a process plant. *Journal of Loss Prevention in the Process Industries*, 25(6):883–904.

[7] Kumar, A. (2013). Reliability and cost-benefit analysis of computer systems subject to maximum operation and repair time. Ph.D. thesis, Maharishi Dayanand University, Rohtak, India.

[8] Sharma, G. and Khanduja, R. (2013). Performance evaluation and availability analysis of feeding system in a sugar industry. *International Journal of Research in Engineering and Applied Sciences*, 3(9):38–50.

[9] Sharma, S. P. and Vishwakarma, Y. (2014). Application of Markov process in performance analysis of refining system of sugar industry. *Journal of Industrial Mathematics*, 2014(1):593176.

[10] Aggarwal, A., Kumar, S. and Singh, V. (2015). Performance modeling of the skim milk powder production system of a dairy plant using RAMD analysis. *International Journal of Quality and Reliability Management*, 32(2):167–181.

[11] Kadyan, M. S. and Kumar, R. (2015). Availability and profit analysis of a feeding system in sugar industry. *International Journal of System Assurance Engineering and Management*, 8(1):301–316.

[12] Ram, M. and Kumar, A. (2015). Performability analysis of a system under 1-out-of-2: G scheme with perfect reworking. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 37:1029-1038.

[13] Parida, A., Kumar, U., Galar, D. and Stenström, C. (2015). Performance measurement and management for maintenance: a literature review. *Journal of Quality in Maintenance Engineering*, 21(1):2–33.

[14] Kumar, A. and Saini, M. (2017). Mathematical modeling of sugar plant: a fuzzy approach. *Life Cycle Reliability and Safety Engineering*, 7:11-22.

[15] Kadyan, M. S. and Kumar, R. (2017). Availability based operational behavior of B-Pan crystallization system in the sugar industry. *International Journal of System Assurance Engineering and Management*, 8(2):1450–1460.

[16] Tsarouhas, P. and Besseris, G. (2017). Maintainability analysis in shaving blades industry: a case study. *International Journal of Quality and Reliability Management*, 34(4):581–594.

[17] Tsarouhas, P. (2018). Reliability, availability and maintainability (RAM) analysis for wine packaging production line. *International Journal of Quality and Reliability Management*, 35(3):821–842.

[18] Choudhary, D., Tripathi, M. and Shankar, R. (2019). Reliability, availability and maintainability analysis of a cement plant: a case study. *International Journal of Quality and Reliability Management*, 36(3):298-313.

[19] Dahiya, O., Kumar, A. and Saini, M. (2019). An analysis of feeding system of sugar plant subject to coverage factor. *International Journal of Mechanical and Production Engineering Research and Development*, 9(1):495–508.

[20] Saini, M., Kumar, A. and Shankar, V.G. (2020). A study of microprocessor systems using RAMD approach. *Life Cycle Reliability and Safety Engineering*, 9:181-194.

[21] Kumar, A., Saini, M., Patil, R.B., Al-Dahidi, S. and Mellal, M.A. (2022). Reliability, Availability, Maintainability, and Dependability analysis of Tube-wells Integrated with Underground Pipelines in agricultural fields for irrigation. *Advances in Mechanical Engineering*, 14(8):16878132221115931.

[22] Saini, M., Kumar, A. and Sinwar, D. (2022). Parameter estimation, reliability and maintainability analysis of sugar manufacturing plant. *International Journal of System Assurance Engineering and Management*,1-19.

[23] Gao, S. (2023). Availability and reliability analysis of a retrial system with warm standbys and second optional repair service. *Communications in Statistics-Theory and Methods*, 52(4):1039-1057.

[24] Yusuf, I., Sanusi, A. and Lawan, Z. D. (2024). A Comparative Study of Hybrid Systems Using RAMD Analysis. *Risk Assessment and Management Decisions*, 1(1):75-87.

RELIABILITY ANALYSIS OF OFFSHORE PLATFORM SUPPORT STRUCTURES UNDER EXTREME WAVE LOADS: A CASE STUDY APPROACH

Seymur Bashirzade[1*], Okan Ozcan[2], Rafail Garibov[3]

•

[1]PhD, Dept. of Civil Engineering, Akdeniz University, 07058, Konyaaltı, Antalya, Turkiye,
e-mail: srbashirzade@gmail.com , ORCID: 0000-0002-0870-6345.
[2]Prof. Dr., Dept. of Civil Engineering, Akdeniz University, 07058, Konyaaltı, Antalya, Turkiye,
e-mail: ookan@akdeniz.edu.tr , ORCID: 0000-0001-9905-1657.
[3]Prof. Dr., Institute of Forensic Construction and Technical Expertise, Russia,
garibovr@mail.ru , ORCID: 0000-0001-9500-2874
[*]Corresponding author

Abstract

Wave loads are critical factor for the design and safe operation of offshore structures. The accurate determination of these loads is essential to ensure the structural reliability and operational efficiency of such platforms at sea. This study develops analytical expressions for calculating wave loadings that affect the support of various Condeep-type offshore structures. In this regard, wave load calculations for the Draugen Monopile Condeep platform, previously constructed in Norway, were analyzed in the context of a case study. The results of this assessment provide useful information regarding the characteristics of wave loads and their relevance to the overall structural analysis. Furthermore, the investigation also covers recommendations for design and safety improvements that consider the calculated wave loads and the assessment of the structural reliability. Study is expected to contribute to the knowledge base surrounding offshore engineering practices and improve resilience and functionality against dynamic wave forces.

Keywords: offshore structures, condeep platform, wave load, reliability analysis, failure, probability

I. Introduction

Reliable operation of offshore platforms is essential for ensuring the energy security and sustainable development of the sector. The stability and reliability of operations are two of the cardinal principles in safeguarding a seamless supply of energy, as a technical or structural failure may cause extreme losses, leading to massive economic losses. In addition, the financial efficiency of oil and gas extraction and transportation processes is related to environmental events, such as

hydrocarbon spills, that may occur during an accident. These accidents not only disrupt normal operations and cause financial losses but also lead to serious environmental damage, fire hazards, and explosions. Therefore, safety and effectiveness in the operation of offshore platforms are very important; this would mean enhancing the prevention of such events, reducing environmental harm, and ensuring effective recovery and work continuity [1,2].

Offshore and marine energy applications are mainly influenced by numerous uncertainties that play a fundamental role in both the design process and operational asset management. Such risks range widely from natural to man-made, and may pose a number of challenges for designers and engineers [3]. The typical reliability problem, the limit state in terms of static response, can be comparatively easily and simply evaluated for the probability of failure or the reliability index, because static analysis involves more straightforward calculations involving fewer variables. However, the same situation is far from simple and becomes quite involved if, for support structures, reliability analysis requires the response to be obtained from dynamic analysis. Dynamic reliability analysis requires that the interaction between the support structure and irregular waves, turbulent wind, and nonlinear ground soil be modeled precisely. In other words, such coupling within the dynamics of the structure with environmental conditions requires heavy computation to obtain a reliable set of dynamic responses. Moreover, the computational effort significantly increases because the number of dynamic analyses to be performed in a reliability analysis is proportional to the square of the number of random variables under consideration [4,5].

The probability of failure, commonly known as reliability, is one of the key indicators for evaluating the adequacy of a structure or member by integrating the uncertainties due to the use of loads to counter structural resistance. Failure is defined with respect to various applicable failure modes that are usually designated by the limit states. For instance, the ultimate limit states (ULS) [6] may indicate structural failure to resist the applied load effects, possibly combined with significant inelastic displacements, notable cracking, or punching shear failure in the case of bridge decks. Serviceability limit states (SLS) [7] are essentially defined as failures to meet the standard requirements for the use or durability of the structure [8-10].

Because the forces on structural elements from such a wave vary continuously as they progress along the structure, the internal stresses in those elements vary. Therefore, for a given structural element, because the position of the wave relative to the structure changes with time, so does the element at which failure is likely to occur next. Given the variability in member strengths, this would mean that, depending on the position of the wave, the most likely element to fail next would change. Otherwise, the sequence of member failures calculated for a given position corresponding to a stationary extreme wave condition may not be related to what actually could occur when the extreme wave moves [11]. This study presents an extensive analysis of the implications of variability in wave loading on the comprehensive probability of structural failure. The reaction of the structure to dynamic loads has important relevance, particularly regarding the reliability of structures utilized in offshore settings. In this study, wave loading and its effects on structural reliability were analyzed through a case study focused on the Draugen Monopile Condeep platform established in Norway [12,13]. The calculations of the wave load in the present study were based entirely on the empirical data. Therefore, the evaluation of the reliability of the structure was based on the measured data. This approach permits an extensive understanding of how the platform functions and behaves in response to prevailing wave conditions and simultaneously provides a more realistic and specific judgment of its resistance and continued reliability.

II. Wave load modelling

The wave characteristics should be correctly estimated, and a detailed investigation of the interaction between marine structures and oceanic waves is necessary, as both may have a significant impact on the design and construction of Condeep platforms. The various classes of ocean waves include one class, termed wind-generated waves, which is the most common. The energy of the wind is transferred to the water of the surface as the wind travels over the water, creating wind-generated waves. The heights, periods, and other parameters related to these waves depend on the wind speed, fetch distance, and length of time the wind blown over the water. The large fetch distances available in an open oceanic environment permit very strong winds to generate an exceptionally large size. To understand wave behavior and propagation, it is essential to know the wavelength L, which represents the spatial distance covered by a complete wave cycle, including one peak and one trough (Figure 1). The equation used to characterize a wave profile moving in the positive x-direction is shown in Equation (1):

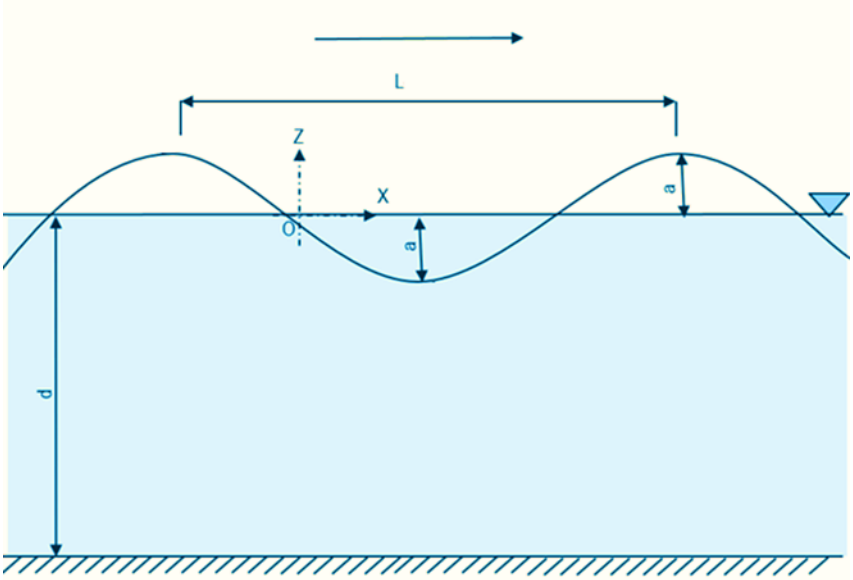


Figure 1. *Sinusoidal wave propagation*

$$\varphi(x,t) = a\,cos(kx - \omega t) \quad \text{whereas:} \quad \begin{cases} k = 2\pi/L \\ \omega = 2\pi/T \\ c = \omega/k = L/T \end{cases} \tag{1}$$

In marine environment, the interrelation among the wavelength of the observed waves (L), wave period (T), and water depth (d) is articulated via a crucial equation (Equation 2).

$$L = \frac{gT^2}{2\pi} tanh\left(\frac{2\pi d}{L}\right) \tag{2}$$

The bathymetric features of the study area, in combination with coastal protection structures and other factors, increase the accuracy and reliability of wave characteristic forecasts. The wave propagation formulas are related to the distance traveled by the wind over its fetch area and the wind speed, as shown in Equation 3. After applying the described transformations of the wind speed, the real wind speed data were converted into wind speeds using the following formulas:

$$\begin{cases} t = \dfrac{1609}{U_f} \\ U_{3600} = \dfrac{U_f}{\left(1.277 + 0.296 tanh\left[0.9 log(\frac{45}{t})\right]\right)} \\ U_A = 0.71 \cdot U^{1.23} \end{cases} \quad (3)$$

In cases where there is a fetch limit, the key parameters include the fetch (F) and wind speed. Based on these parameters, the significant wave heights and periods were like Equation (4):

$$\begin{cases} H_s = \dfrac{U^2}{g} \cdot 0.283\ tanh\left[0.0125\left(\dfrac{g \cdot F}{U_A^2}\right)^{0.42}\right] \\ T_s = \dfrac{2\pi \cdot U}{g} \cdot 1.2\ tanh\left[0.077\left(\dfrac{g \cdot F}{U_A^2}\right)^{0.25}\right] \end{cases} \quad (4)$$

The Morison equation was used to calculate the hydrodynamic forces acting on offshore structures [14,15]. Equation (5) represents the combination of inertial and drag forces acting on submerged offshore structures.

$$F_{total} = F_d + F_i \Rightarrow \begin{cases} F_d = 0.5 \cdot \rho \cdot C_d \cdot A \cdot V_r^2 = 0.5 \cdot \rho \cdot C_d \cdot D \cdot u \cdot |u| \\ F_i = 0.5 \cdot \rho \cdot C_m \cdot A \cdot V_r = 0.25 \cdot \rho \cdot C_m \cdot \pi \cdot D^2 \frac{du}{dt} \end{cases} \quad (5)$$

The Morison equation is a method for determining the combined effects of both drag and inertial forces induced by wave action on a vertically submerged cylindrical structure at a certain depth of submergence (Figure 2). Adapting this equation to a circular column for load calculations is an important part of marine engineering and the design of coastal structures. This equation ensures that structures can operate safely and reliably in the presence of the dynamic forces produced by ocean waves (Equation 6 and 7).

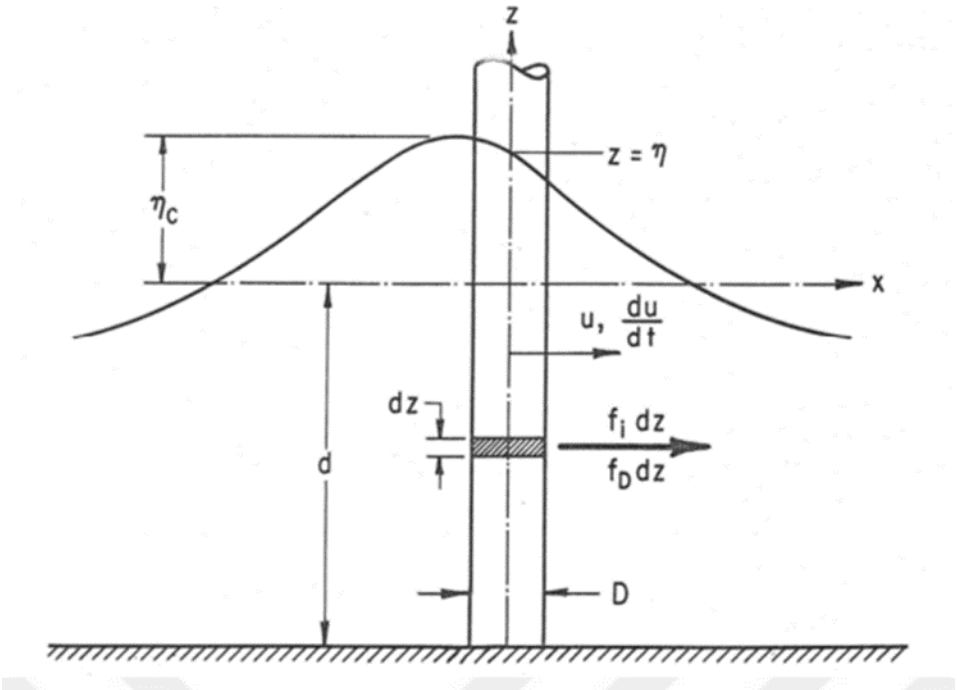


Figure 2. *Wave loads on a vertical cylinder*

$$F_{total} = F_d + F_i = \int_{-d}^{\eta} f_i dz + \int_{-d}^{\eta} f_D dz = -F_i sin\omega t + F_d cos^2 \omega t \quad (6)$$

$$\begin{cases} F_i = \frac{\pi}{4}\rho \cdot g \cdot C_m \cdot D^2 \cdot H \cdot K_i \\ F_D = \frac{1}{2} \cdot \rho \cdot C_d \cdot D \cdot H^2 \cdot K_D \end{cases}; \quad \text{whereas,} \quad \begin{cases} K_i = \frac{1}{2} tanh\, kd \\ K_D = \frac{1}{8}(1 + \frac{2kd}{sinh\, 2kd}) \end{cases} \tag{7}$$

III. Case Study

In this study, the wave load estimates were conducted using data from the Draugen Monopile Condeep platform (Figure 3). This platform was set to a depth of 250 m, with an embedment of 9 m into the seabed. The distance between the mean sea level and bottom surface of the deck of the platform is 30 m. The platform wall exhibits a diameter of 45 m at its most profound section, gradually tapering to 16 m at the elevation of still water. The thickness of the wall is intentionally specified to be 1.9 meters at the foundation of the structure, decreasing to 0.7 meters at its uppermost boundary [16].

Figure 3. *Location plan and general overview of Draugen Monopile Condeep platform [17].*

This parametric analysis takes the annual wind speed of the region for the Draugen platform in Norway as a reference. The wind speed data for the past 10 years were obtained using the Meteoblue dataset API [18], with a focus on specific areas, as shown in Figure 4. Calculations with regard to wave load were therefore carried out using the highest wind speed values corresponding to the four main directions highlighted above. Another important parameter is the frictive wind velocity statistic, represented as U_f , which shows the maximum wind speed recorded within a given period. The data show that a maximum wind speed of 24 m/s was achieved from the north, while 20 m/s was achieved from the west. This was an extreme wind condition for ten years. By accepting the maximum wind speed, the calculations of the wave height and period were obtained using the previously mentioned formulae with consideration of the distance, from the coastline to the desired offshore platform location (Table 1) [19].

Tablo 1. Calculated wave parameters

Direction	F (km)	t (s)	U_{3600} (m/s)	U_A (m/s)	H_s (m)	T_s (s)
North	1380	57.5	19.21	26.91	6.51	10.57
West	985	49.25	15.79	21.14	4.88	8.89

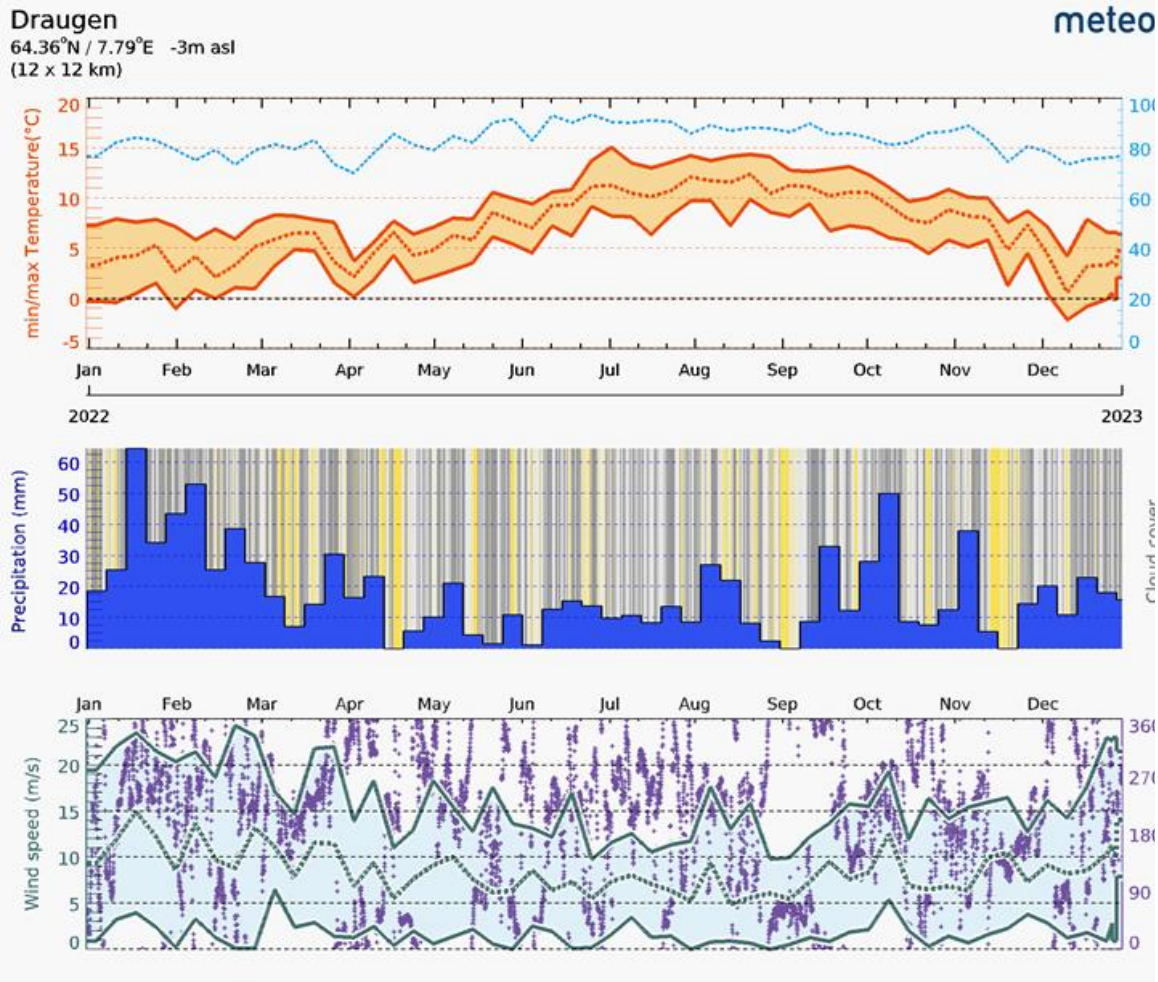


Figure 4. *Metrological statistics of Draugen platform [18]*

In this regard, calculations have primarily been made using small-amplitude wave theory to assess the wave parameters [20,21]. The result from such calculation provides the peak values that allow developing the significant wave height, H_s, and the significant wave period, T_s. Furthermore, in order to consider the design, analysis, and decision-making process through safety-oriented approach, extreme wave events are highlighted representing maximum values obtained from such calculations as H_s and T_s

$$\begin{cases} H_s = H = 6.51\ m \\ T_s = T = 10.57\ s \end{cases} \tag{8}$$

Because the actual situation places the Monopile Condeep platform at a depth of d=250 m, the wavelength was calculated using the proper equations of wave propagation in water.

$$L = \frac{gT^2}{2\pi} tanh\left(\frac{2\pi d}{L}\right) = \frac{9.81 \cdot 10.57^2}{2 \cdot 3.1415} tanh\left(\frac{2 \cdot 3.1415 \cdot 250}{L}\right) \quad (9)$$

Consequently, after many iterations, the wavelength, was found to be 174.52 meters.

$$\frac{d}{L} = \frac{250}{174.52} > 0.5 \quad (10)$$

It was then established that the waves fall into the category of deep-water waves; that is, these waves travel in a water body whose depth is greater than half the wavelength ($d > 0.5L$).

As a result, after evaluating all the necessary parameters, the final wave load for cylindrical support structure has been calculated using the Morison equation, as shown in Equations 6 and 7.

$$F_i = \frac{3.14}{4} \cdot 1025 \cdot 9.81 \cdot 2 \cdot \left(\frac{45 + 16.4}{2}\right)^2 \cdot 6.51 \cdot 0.5 = 4.843 \cdot 10^7 N$$
$$F_D = \frac{1}{2} \cdot 1025 \cdot 9.81 \cdot 0.7 \cdot \left(\frac{45 + 16.4}{2}\right) \cdot 6.51^2 \cdot 0.125 = 0.572 \cdot 10^7\ N \quad (11)$$

$$F_{total} = -F_i sin\omega t + F_d cos^2 \omega t = -F_i sin\omega t = -4.843 \cdot 10^7\ N\ = -48.43\ MN \quad (12)$$
$$cos\omega t = 0 \quad \rightarrow \quad \omega t = \frac{\pi}{2}$$

IV. General reliability assessment of platform under wave load

Structures or their elements fail when subjected to excessive loads or when an interaction of the loads produces a sufficiently intense load effect to create an inappropriate failure mode in the structure. Such a failure condition can be an ultimate failure, where the structure cannot support the applied load, or serviceability failure, where the structure does not function appropriately.To address this problem, the size and consequences of such an event must first be predicted [22]. Furthermore, the strength and deflection under the load of each structural element should be evaluated using the available design information. Such a procedure requires probabilistic models that reflect uncertainties owing to the loading variables, as well as those of the structural resistance [23].

Although the load factor (LF) describes the magnitude and nature of the loads acting on a structure, it is important to determine its load-carrying capacity. The LF must be assessed together with the resistance factor (RF), which is representative of the strength and safety margins of the structure. If the structure is to be safe, LF and RF must be in balance on the right side of the figure so that the structure has a resistance equal to or greater than the expected load (Figure 5). All parameters affecting the structure related to a reliability assessment, such as moment capacity, shear force capacity, stress, and deformation, should be analyzed. Proper modeling and evaluation of these parameters will enable a realistic assessment of the performance of the structure. Proper comparison of the load and resistance parameters allows for more accurate conclusions regarding safety and durability.

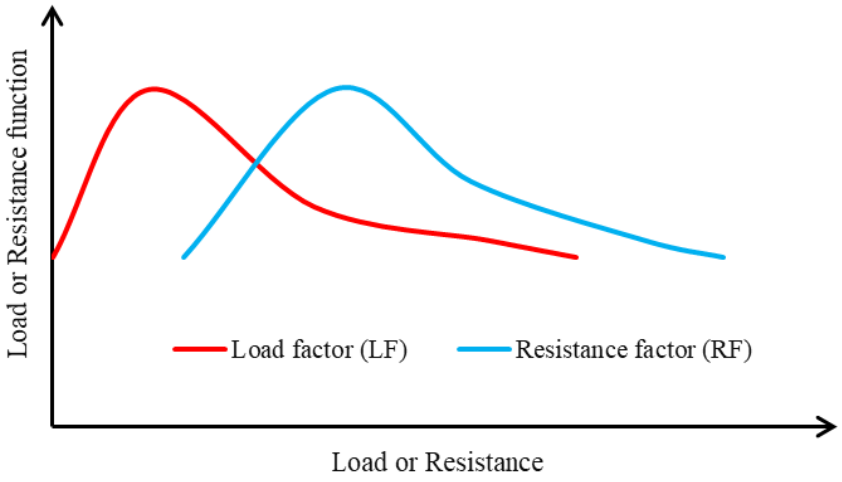


Figure 5. *Relationship between Load and Resistance factors in a reliable structure*

In this respect, in the case study of the Draugen Monopile Condeep platform, during the evaluation of the action of wave loads on the structure, the structural analysis model could be considered as a cantilever beam fixed to the ground. The wave load is lateral in nature and therefore exerts its most extreme effect at the support point, generating a large flexure moment ($M_{flex.}$). The magnitudes of the support moment and moment generated within the structure are related to stability and safety. To ensure safety in this respect, the support moment or the so-called Load Factor-LF of the structure at a certain point should be compared with the strength moment as RF designed by geometrical and material properties of the structure at a similar point. A structure is considered reliable if the support moment is less than or equal to the strength moment.

$$M_{flex.} - M_{wave} \geq 0 \tag{13}$$

In the computation of the unconditional failure probability for platform design, the failure probability corresponding to the platform for different wave heights is multiplied by the probability of occurrence of that particular wave height. This method considers the probabilities of different wave heights and their respective impacts on the platform, which enables failure risk evaluation for every possible scenario [24]. By aggregating all these probabilities of failure, the overall (unconditional) failure probability pertinent to the platform design can be derived. This approach facilitates a more dependable failure analysis by accounting for the potential risks that the platform might encounter under different wave conditions, thereby improving the precision of engineering designs.

$$E[p_F] = \int_H E[\langle p_F|H\rangle p_F]\, f_H(h)dh = \sum_H E[\langle p_F|H\rangle]P_H h \tag{14}$$

To perform these analyses and reliability calculations more accurately, one needs to develop proper probability models of the main load and resistance random variables beforehand. The characterization of the statistical distribution of structural materials and load conditions as accurately as possible is of crucial importance in the reliability study of engineering structures. The modeling of the probability distributions associated with the variables of load and resistance is important for understanding the nature of the uncertainties that structures are subjected to and to

study the implications of these uncertainties on the safety of structures. In this context, the term "model" underlines the fact that the real behavior of structures is very complex, and any computational procedure relies on a number of idealizations to reduce this complexity. Idealization of structures is a basic necessity to make calculations possible; however, the accuracy and generality of the models employed directly influence the reliability of the results derived from the analysis. In addition, the development of 1D, 2D, or 3D models for structures using numerical analysis techniques and the assessment of these models for real loads and environmental effects allow the achievement of more realistic and reliable results [25]. Such models simulate how the structure will behave at both local and overall structural levels, thus enabling more realistic forecasts of structural performance. The execution of these analyses, considering also various loading conditions and environmental influences, provides sharper and more conclusive results regarding structural reliability and durability.

V. Conclusion

This study examined the wave load, representing one of the predominant forces exerted on marine and oceanic structures, and its implications for the reliability of such structures subjected to horizontal loads. This study used the small-amplitude wave theory in the calculation of the wave load, deriving a formula intended to quantify the wave impact on structures with cylindrical cross-sections. This equation was integrated into the standard Morison equation to develop a approach for calculating wave forces acting on cylindrical structures.

A case study evaluated the wave load on a Draugen Monopile Condeep platform installed in Norway. In the structural reliability analysis, the structure was modeled as a cantilever beam anchored to the seabed, for which the most important internal force was considered to be the moment generated at the anchorage to the seabed. This moment was compared with the moment induced by the wave load to quantify the general safety of the entire structure. Moreover, an explicit formula was provided to predict the structural failure due to the wave load. This highlights the need for an in-depth numerical analysis of the structural model to ensure safety. The results show that accurate calculations of wave forces acting on such structures should be performed, along with detailed preliminary scrutiny of the reliability of the structure.

The methodologies and formulae developed in the present study are expected to increase the calculation accuracy of the wave load and provide important guidelines for marine infrastructure design and safety evaluation. It is advisable that subsequent investigations prioritize more extensive numerical simulations to improve the safety of structures subjected to wave forces.

Notation

a : *Wave amplitude*
ω : *Angular frequency of the wave*
T : *Period of the wave*
c : *Singular wave speed*
k : *Wave number*
L : *Wave length*
d : *Water depth*

U_f : *Fastest wind speed*
U_{3600} : *1 hour average wind speed*
U_A : *Wind stress factor*
F : *Fetch distance*
H_s : *Significant wave height*
T_s : *Significant wave period*
F_{total} : *Total wave force*
F_d : *Drag force;*
F_i : *Inertial force;*
ρ : *Fluid density*
A : *Reference area (usually the front surface area of the cylinder)*
V_r : *Horizontal velocity perpendicular to the axis of the fluid particle*
D : *Cylinder diameter*
C_m & C_d : *Hydrodynamic coefficients*
$M_{flex.}$: *Flexural moment capacity*
M_{wave} : *Moment driven by wave forces*
$E[p_F]$: *Platform failure probability*
$f_H(h)$: *Pdf of the wave heights occurrence*
$P_H h$: *Occurrence probability*

Declaration

Acknowledgments: *Not applicable*
Funding: *This paper was made possible with the support of the 'Türkiye Burslari,' funded by the Presidency for Turks Abroad and Related Communities.*
Declaration of Conflicting Interests: *The Authors declare that there is no conflict of interest*
Data availability: *Not applicable*
Authors contribution: *All the authors made substantial contributions to the conception or design of this study.*
Competing Interest: *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.*

Reference

[1] Gafar Ismayilov, Hajar Ismayilova, Hikmet Babirov , and Rashid Jabrayilov. "Assessment of environmental oil spills and economic-environmental risks" Reliability: Theory & Applications, vol. 17, no. SI 4 (70), 2022, pp. 212-217. doi:10.24412/1932-2321-2022-470-212-217

[2] Amiri, Nima, et al. "A comprehensive review on design, monitoring, and failure in fixed offshore platforms." Journal of Marine Science and Engineering 9.12 (2021): 1349. https://doi.org/10.3390/jmse9121349

[3] Leimeister, Mareike, and Athanasios Kolios. "A review of reliability-based methods for risk analysis and their application in the offshore wind industry." Renewable and Sustainable Energy Reviews 91 (2018): 1065-1076. https://doi.org/10.1016/j.rser.2018.04.004

[4] Kim, Dong Hyawn, and Sang Geun Lee. "Reliability analysis of offshore wind turbine support structures under extreme ocean environmental loads." Renewable energy 79 (2015): 161-166. https://doi.org/10.1016/j.renene.2014.11.052

[5] Zhang, Yi, and Jasmine Siu Lee Lam. "Reliability analysis of offshore structures within a time varying environment." Stochastic Environmental Research and Risk Assessment 29 (2015): 1615-1636. https://doi.org/10.1007/s00477-015-1084-7

[6] Meléndez, Carlos, Pedro F. Miguel, and Luis Pallarés. "A simplified approach for the ultimate limit state analysis of three-dimensional reinforced concrete elements." Engineering Structures 123 (2016): 330-340. https://doi.org/10.1016/j.engstruct.2016.05.039

[7] Gilbert, R. I. "The serviceability limit states in reinforced concrete design." Procedia Engineering 14 (2011): 385-395. https://doi.org/10.1016/j.proeng.2011.07.048

[8]. Cheung, Mo Shing, and B. R. Kyle. "Service life prediction of concrete structures by reliability analysis." Construction and Building Materials 10.1 (1996): 45-55. https://doi.org/10.1016/0950-0618(95)00055-0

[9] Alena Rotaru. "Application of non-destructıve testing methods and evaluation of condition of reinforced concrete framing" Reliability: Theory & Applications, vol. 19, no. 2 (78), 2024, pp. 400-407. doi:10.24412/1932-2321-2024-278-400-407

[10] Гарибов, РБ, УМ Гаджиева, and СР Баширзаде. "К вопросу об оценке безопасности и долговечности проектируемых железобетонных конструкций." Вестник Волжского регионального отделения Российской академии архитектуры и строительных наук 21 (2018): 160-164.

[11] Chan, H. Y., and R. E. Melchers. "Wave loading effect in offshore structural reliability." Structural safety 7.1 (1990): 1-10. https://doi.org/10.1016/0167-4730(90)90009-E

[12] Alm, T., et al. "The Draugen platform and subsea structures, installation and foundation aspects." Offshore Technology Conference. OTC, 1995. https://doi.org/10.4043/7670-MS

[13] Austbø, Solmøy. Geotermisk energipotensiale i Norskehavet: En casestudie fra Draugen-området på Trøndelag-plattformen. BS thesis. UIS, 2024. https://hdl.handle.net/11250/3138589

[14] Sarpkava, T. "Morison's equation and the wave forces on off-shore structures." NCEL-CR-82. 008,, 1981, 270 (1981).

[15] Wolfram, Julian. "On alternative approaches to linearization and Morison's equation for wave forces." Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 455.1988 (1999): 2957-2974.. https://doi.org/10.1098/rspa.1999.0434

[16] Faltinsen, O. M., and A. N. Timokha. "Undamped eigenperiods of a sea-based gravity monotower." Applied Mathematical Modelling 40.19-20 (2016): 8217-8243. https://doi.org/10.1016/j.apm.2016.04.003

[17] OljeMuseum, Norsk. "Oil and Gas Fields in Norway: Industrial Heritage Plan." Norsk OljeMuseum (NOM): Stavanger, Norway (2016):

[18] Meteoblue, Dataset. 2023.
https://www.meteoblue.com/en/weather/week/draugen_norway_6296732

[19] Bashirzade, Seymur. Ard Germeli Segmental Açık Deniz Yapıların Tasarımı. PhD dissertation, Akdeniz University Graduate School of Natural and Applied Sciences, 2024. DOI: http://dx.doi.org/10.13140/RG.2.2.23036.14728

[20] Ursell, Fritz, Robert George Dean, and Y. S. Yu. "Forced small-amplitude water waves: a comparison of theory and experiment." Journal of Fluid Mechanics 7.1 (1960): 33-52. https://doi.org/10.1017/S0022112060000037

[21] Le Méhauté, Bernard, and Bernard Le Méhauté. "Linear small amplitude wave theories." An Introduction to Hydrodynamics and Water Waves (1976): 212-238. https://doi.org/10.1007/978-3-642-85567-2_16

[22] Bashirzade, S., & Garibov, R. Comparative Study of the Advantages and Disadvantages of Prestressed and non-Prestressed Concrete Structures for Offshore Applications. International Journal of Scientific Research and Engineering Development vol. 6, no. 2, 2023, pp. 941-947, ISSN: 2581-7175.

[23] Thoft-Cristensen, P., & Baker, M. J. (2012). Structural reliability theory and its applications. Springer Science & Business Media.

[24] De Leon, D., and A. H-S. Ang. "Confidence bounds on structural reliability estimations for offshore platforms." Journal of marine science and technology 13 (2008): 308-315. https://doi.org/10.1007/s00773-008-0019-0

[25] Bashirzade, S., Ozcan, O., Cagdas, I. U. Internal force transfer in segmental RC structures. Research on Engineering Structures & Materials, (2024). Pp 1-24. http://dx.doi.org/10.17515/resm2024.146st0109rs

RT&A, No 4(80)
Volume 19, December, 2024

EXPERIMENTAL AND NUMERICAL INVESTIGATION OF STRESS CONCENTRATION FACTOR FOR POLYGONAL DISCONTINUITIES IN A FINITE PLATE

Rashmiben H. Patel[1,2], Dr. Bhaveshkumar P. Patel[3]

•

1 Ph. D Scholar, Faculty of Engineering and Technology, Ganpat University, Ganpat Vidhyanagar 384012, Mehsana, Gujarat, India. rashmibeme@gmail.com

2 Assistant Professor, Department of Mechanical Engineering, Government Engineering College, Gandhinagar, 382028, Gujarat, India. rashmimech@gecg28.ac.in

3 Professor, Department of Mechanical Engineering, U. V. Patel College of Engineering, Ganpat University, Ganpat Vidhyanagar - 384012, Mehsana, Gujarat, India, bpp01@ganpatuniversity.ac.in

Corresponding Author E-mail. (bpp01@ganpatuniversity.ac.in)

Abstract

Structural steel is widely utilized in the construction engineering sector to build a variety of buildings, including flyovers, skyscrapers, plants, heavy machinery vehicle structures, etc., in different combinations. Due to their wide range of applications, particularly in the automotive and aerospace industries, plates with different kinds of holes are also significant parameters for mechanical design. To satisfy the requirements in the final structure design, these holes are formed into plates. However, these holes concentrate stress, which gradually weakens the structure's mechanical strength. The present study aims to reduce this stress concentration of compressed plates having polygonal holes of varying shapes and sizes. The stress concentration factor around polygonal holes in polycarbonate plates, subject to uniaxial compression loads, is investigated experimentally and numerically. To obtain solutions, three approach are adopted; the finite element method, DOE RSM (Response Surface Methodology) and photoelasticity are used as the experimental method. The study's conclusions are presented here in the form of numerical and graphical data, along with a comparison between the outcomes and the photo-elasticity test results.

Keywords: Finite element method, finite plate, structural discontinuity, polygonal holes, photoelasticity, stress concentration factors

Abbreviation:

RSM	Response Surface Methodology
FEA	Finite Element Analysis
DOE	Design of Experiment
SCF	Stress Concentration Factor
FGM	Finite Graded Material
FEM	Finite Element Method
CNC	Computerized Numerical Control

Notations:

N	Fringe number
f_σ	Stress fringe value (stress fringe constant of model material)
t	Thickness of the plate (mm)
K_t	Stress concentration factor
H	Height of plate (mm)
L	Length of plate (mm)
l	Side of polygonal hole (mm)
l/L	Side ratio
n	Number of side of polygonal hole in a plate
O	Polygonal hole orientation (°)
A	Area of the polygonal hole (mm^2)
P	Applied load (N)
σ_{max}	Maximum stress at the discontinuity (N/mm^2)
σ_{nom}	Nominal or background stress (N/mm^2)
d	Maximum length of polygonal hole along the loading direction (mm)

1. Introduction

Optimization of the structure's shape has been prompted by the growing demand for lightweight, energy-efficient structures in recent years. Structural elements must have various cut-out shapes lighten the system's weight and allow access to other structural components. It is commonly known that highly localized stresses are created in the area around a cut-out or hole in a stressed member. Geometric irregularities like different shoulder fillets, grooves (U, V, Square), keyways, threads, notches, holes, etc., found on plates are essential features because they fulfill specific operational needs. These geometric discontinuities will generate stress concentrations. To design a component or a structure, it is necessary and helpful to estimate the stresses and strains at these geometries. Due to the application of this discontinuities in marine and aerospace structures, thin plates are frequently used. Different kinds of holes or openings are usually made in the plates for practical reasons, like reducing the weight of the system and providing access to its components. When a plate undergoes tension, compression, or shear forces, the presence of holes or openings results in increased stresses in their vicinity. For the best design and resistance to mechanical failure, understanding of the stress distribution surrounding the hole is essential. An experimental methodology can be employed to determine the maximum stress concentration for various complex geometries while taking into account time or cost limitations. The number of sides, orientations of polygonal hole, areas of the polygonal hole, and applied load on a finite plate are four variable parameters considered in the calculation.

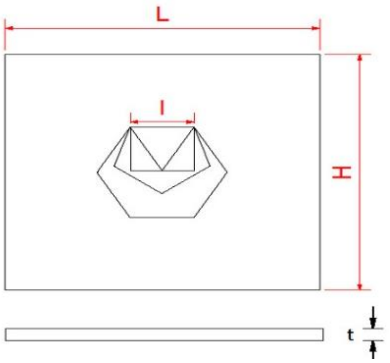


Figure 1: *The geometry of polygonal hole in a plate*

Figure 1 shows flat plate with polygonal hole. A stress concentration is defined as the ratio of the maximum stress within the element to the nominal stress (reference stress).

$$K_t = \frac{\sigma_{max}}{\sigma_{nom}} \quad (1)$$

$$\sigma_{nom} = \frac{P}{(H-\mathrm{d}) \times t} \quad (2)$$

Numerous researchers consistently focus on the stress concentration problem associated with a plate containing a hole, which is a crucial consideration in structural design.

An analysis of stresses and strains in finite and infinite rectangular homogeneous, isotropic elastic plates with a circular hole was provided by M. Ismail et al. under uniform tensile loading [1]. M. Pandey conducted an examination of stress distribution within a plate featuring an inclined circular aperture under varying uniaxial tensile loads. Utilizing finite element analysis (FEA) in ANSYS, stresses were determined for circular holes inclined at angles of 30°, 45°, and 90°. The investigation revealed that the minimum stress concentration factor occurred at a 45° inclination angle, while the maximum SCF was observed for a vertically oriented hole, inclined at 90° [2]. Xiaoli Zhang al. studied the stress generation in infinite skew anisotropic infinite plate with elliptical hexagonal and square holes with the use of the analytical method [3]. A novel analytical-numerical technique is introduced for assessing the combined stresses induced by multiple circular apertures within an infinite elastic plate. This method accounts for both externally applied remote stresses and arbitrarily distributed stresses along the circular boundaries. This method is based on elementary solutions and new integral equations [4]. The elastic stress around a various holes (circle, an ellipse, and overlapping circles) with corners in an infinite plate subjected to biaxial stress was observed by W. Wang et al. Complex Goursat functions are used to formulate the elasticity problem, leading to a set of singular integro-differential equations on the boundary [5]. The distribution of stress in an infinite plate experiencing uniaxial tension with an ellipse or circular hole was calculated using the boundary force method [6]. B. Safaei et al. compared the numerical results with analytical answers after investigating the stress distribution in sheets with circular cutouts and studying stress concentration at the edge of holes [7]. Jaiswal et al. employed a technique to model functionally graded material (FGM) within the finite element method (FEM)-based software ANSYS. They aimed to mitigate stress concentration by implementing an FGM ring around the rounded rectangular slot under tensile loading [8]. Rani et al. further expanded their research by utilizing the extended finite element method to quantitatively analyze the stress concentration around a centrally positioned elliptical inclusion embedded within functionally graded material [9].

Using integral equations based on Green-type solutions, O. Maksymovych et al. determined the impact of stress concentration near cracks, holes, and dies, in the dynamic loading and half plane plate. [10] [11]. M. Patil et al. determined the influence of roundness and orientation of hexagonal cutouts on stress concentration in a plate subjected to tensile loading [12]. M. M. Kumar et al. conducted a study investigating the impact of cut-out orientation and bluntness in an aluminum plate featuring triangular and square cut-outs under uniaxial loading conditions. [13]. The effect of side ratio and height ratio on SCF of polygonal cutout in a finite plate was discussed by R. Patel et al. [14].

The additional auxiliary hole arrangements, different reinforcing thicknesses, hole position, shape, and size, as well as the ratio of the various polygonal holes to the plate size, are all impacted by the stresses that are created in a plate. It demonstrates that the most important design component is the length and that varying the location of the holes can reduce stress concentration [15].

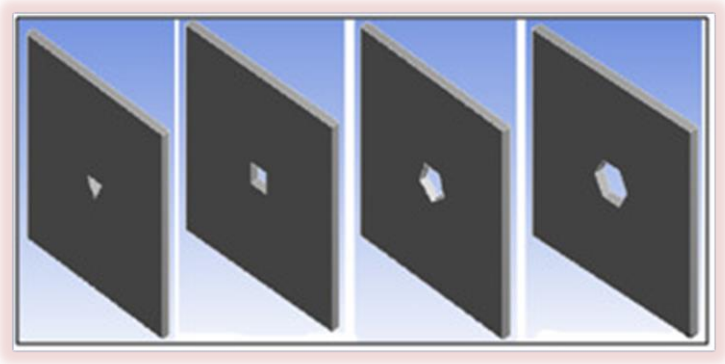
2. Methods

In recent years, the use of computer methods for structural optimization using Finite element analysis (FEA) has grown significantly in the research and development process. The computational cost of these optimization procedures varies according to the problem's complexity. Utilizing DOE (RSM) techniques in conjunction with FEA results in a significant reduction in computing effort without compromising research outcomes.

2.1 Finite Element Analysis

Finite element analysis (FEA) was employed to analyze computational results for different parameters of various flat plate geometries. For all the four parameters (the number of sides, orientations of polygonal hole, areas of the polygonal hole, and applied load on a finite plate) for polygonal hole shapes in a finite flat plate are modelled and analyzed. ANSYS workbench 20 R1 is used for the FEA of all the geometries. Special polygonal-shaped holes are incorporated into a plate to fulfill specific structural requirements, typically using structural steel such as IS 2062, which is commonly utilized across industries. In this study, the plate material conforms to the IS 2062 standard, reflecting prevalent industrial practices. The American equivalent standard is ASME SA 36/ASTM A36 steel.

Figure 2 represent the 2D and 3D model for the plate with hole of various geometric shapes like triangular, square, pentagonal, and hexagonal. The units for geometrical dimensions and force are millimeter and Newton respectively.



(a) Triangular, (b) square, (c) pentagonal, (d) hexagonal holes
Figure 2: *3D model of plate with polygonal hole*

The finite element method analysis is produced and analyzed in ANSYS 2020 R1. The plate featuring various polygonal shapes is meshed with over 396,000 solid elements and encompasses more than 1,716,880 nodes, all meeting acceptable criteria. A mapped meshing technique utilizing quadratic elements (for higher precision) has been employed to mesh the area surrounding the polygonal holes. The model of geometry is divided into small elements. The element has the capabilities having properties like plasticity, creep, stress stiffening, large deflections etc. The lower end of the plate is fixed in the model, while a compressive load is applied to its upper end. Finite element analysis was conducted on the polygonal hole shape within a finite plate under compression loading, with the subsequent calculation of equivalent (von Mises) stresses.

2.2 DOE RSM (Response Surface Methodology)

RSM, or Response Surface Methodology, comprises a collection of statistical and mathematical techniques used to develop empirical models. Its objective is to optimize an output variable, also

known as a response, influenced by several independent variables, or input variables, through careful experimental design. A set of quantitative experimental variables or factors and a response are examined in connection to each other using the Response Surface Method.

The number of sides, orientations of polygonal hole, areas of the polygonal hole, and applied load on a finite plate are four parameters taken into account in the calculation. A series of analyses is performed together the datasets for the RSM model.

The central composite design is utilized to determine the levels of control factors (number of sides of the polygonal hole, orientation, area of the polygonal hole, and applied load to the plate) in order to achieve an optimal response value, namely the stress concentration factor. A standard process of RSM method gives the regression response equation in uncoded units is as follows.

Response = 5.6124 - 0.2361 Polygon Hole + 0.2711 Orientation + 0.5817 Hole Area
+ 0.0428 Applied Load - 0.474 Polygon Hole*Polygon Hole
+ 0.171 Orientation*Orientation + 0.116 Hole Area*Hole Area
+ 0.146 Applied Load*Applied Load - 0.3731 Polygon Hole*Orientation
+ 0.0294 Polygon Hole*Hole Area + 0.0469 Polygon Hole*Applied Load
+ 0.1556 Orientation*Hole Area + 0.0481 Orientation*Applied Load

The results acquired through FEA methods are utilized to derive the response equation via the RSM approach.

The utilization of FEM-based response surface methods has significantly reduced the time and effort needed to assess the design variables of implants. With the hybrid model RSM and FEA, total twenty-five number of experiments are given by RSM method. These experiments are carried out to reduce material and time. These twenty-five numbers experiments are compared with RSM and FEA results.

3. Experimental Work

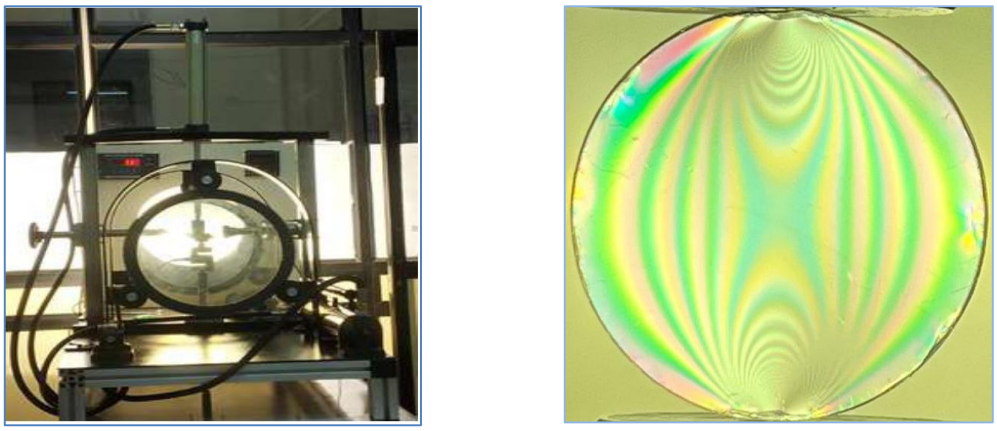
In this experimental study, the plate is assumed to be finite, isotropic, and linearly elastic. A finite plate implies that one side of the polygonal hole should be parallel to the x-axis.

3.1 Experimental Setup

Figure 3: *Photo elasticity experimental apparatus*

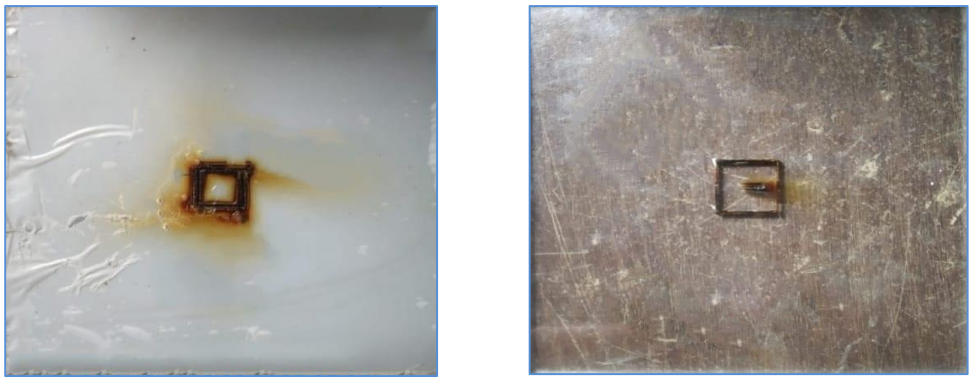
In this study, a photoelastic circular polariscope is utilized to determine the maximum stresses in isotropic finite plates featuring central polygonal holes of various configurations. Figure 3 depicts an image of the photoelastic apparatus used for experimental purposes. This setup was constructed with assistance from Aeolus Aerotech Pvt. Ltd, Bangalore. The purpose

of this experimental unit is to illustrate how materials behave in relation to external forces. Using the photoelasticity theory, specimens can be subjected to forces and/or stresses to observe the stress pattern and calculate the stress concentration factor. Figure 4 shows images of the polycarbonate plate holding on a photoelastic apparatus and a circular plate under compression loading for calibration process.



(a) photoelastic apparatus *(b) calibration*

Figure 4: *Images of the polycarbonate plate having square hole under compression loading and circular disk under compression for calibration process*



(a) Sample plate 1 *(b) Sample plate 2*

Figure 5: *Polycarbonate plate after laser machining*

This photo elasticity unit consists of:

- Analyzer and polarizer,
- Two quarter wave plate,
- Monochromatic and white light source,
- Rotatable analyzer with angular measurements,
- Mechanical load frame with a load capacity of 1.2 kN equipped with a digital load indicator for both tension and compression measurements.

3.2 Material Selection for Experimental Work

In the present work, a polycarbonate (PC) plate is used as workpiece material. It is photoelastic material and widely used for photoelasticity experiments as it demonstrates a birefringent property

essential for conducting photoelastic experiments. In industries variety of polycarbonates are available. This polycarbonate plate is photoelastic material and it exhibits a birefringence property which is needed for any photoelastic experiments. A few samples of acrylic plates are tested to check the results in the form of fringes and tried to minimize the cost of the workpiece as polycarbonate materials are very expensive. But experimentation on acrylic material shows the poor quality of fringes under axial compression loading conditions. So here all experimentations were carried out on a polycarbonate material. Figure 5 (a) and (b) show the specimens made from polycarbonate plates after laser machining. As we can see in the figures, laser cutting in a polycarbonate material is not possible due to the burning of the material as start the machining on polycarbonate with laser cutting. Here, polycarbonate is cut in to necessary shape of polygon shape holes by CNC water jet cutting machine. 16 specimens of 100 × 100 × 5 mm specimens are cut from each sheet for photo elastic experiment subjected to axial compression loading. Figure 6 shows the geometry of the plate having triangular, square, pentagonal and hexagonal holes after water jet machining.

3.3 Calibration of a Material

As shown in Figure 4(b), a standard calibration specimen is the circular disk used in diametral compression. The fringe constant value for selected material can be experimentally determined by the standard calibration procedure for materials. The value of the fringe constant f_σ can be determined experimentally by a model that is made of the same material as the specimen, by observing the corresponding value of N. For this specimen, the diameter D was 63.50 mm, and the load was 1.4 kN. The value of N at the center of the disk, as seen in Figure 4(b), is about 5.0. By following this procedure, we determined that the fringe constant for the selected material is approximately 11 kN/m/fringe [16].

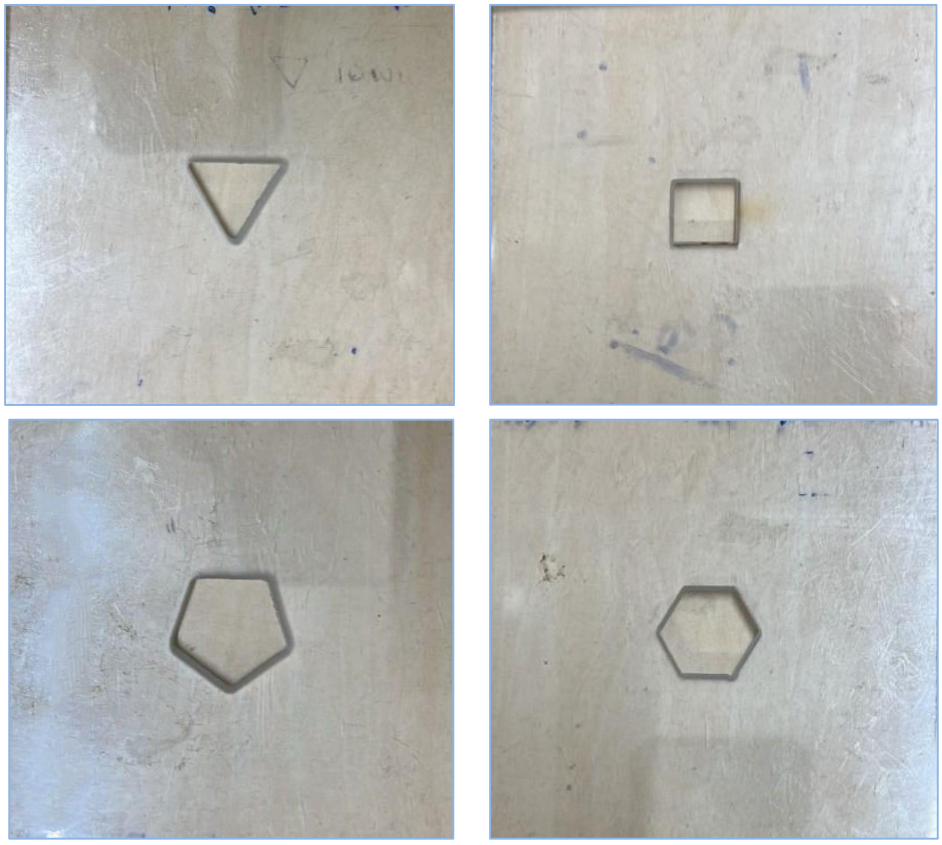


Figure 6: *Polycarbonate plate with various polygonal hole after water jet machining*

3.4 Experimental Procedure and Evaluation Method of SCF

For experiment work, the test specimen is to be held in different positions for different parameters. Figure 4 shows the photoelastic apparatus and the test specimen position for axial compression loading. Total of twenty-five experiments with various polygonal hole side, orientation (hole), constant areas (hole), and applied loads on a plate are performed under axial compression loading. Figure 8 to 10 shows the images captured during photo elastic experimental work for a various parameter. For optimum condition twenty fine experiments of different parameters are performed.

A stress concentration factor is a ratio of the maximum (highest) stress in the element to the nominal (reference) stress.

$$K_t = \frac{\sigma_{max}}{\sigma_{nom}} \tag{3}$$

$$\sigma_{nom} = \frac{P}{(H-\mathrm{d}) \times t} \tag{4}$$

Where, σ_{max} =The maximum stress at the corner of the polygonal hole on the plate occurs in the test specimen during experiments and is determined using a formula derived from the stress optic law theory for the Photoelasticity method.

$$\sigma_1 - \sigma_2 = \sigma_{max} = \frac{Nf_\sigma}{t} \tag{5}$$

σ_2 is zero considered as uniaxial directional loading [7].

$\sigma_{nom} = \frac{P}{(H-\mathrm{d}) \times t}$ for thick plate under axial compression

N = Fringe order

f_σ = The material fringe value is expressed in kN/m and is dependent on the wavelength of the light.

t = Thickness of the plate in mm

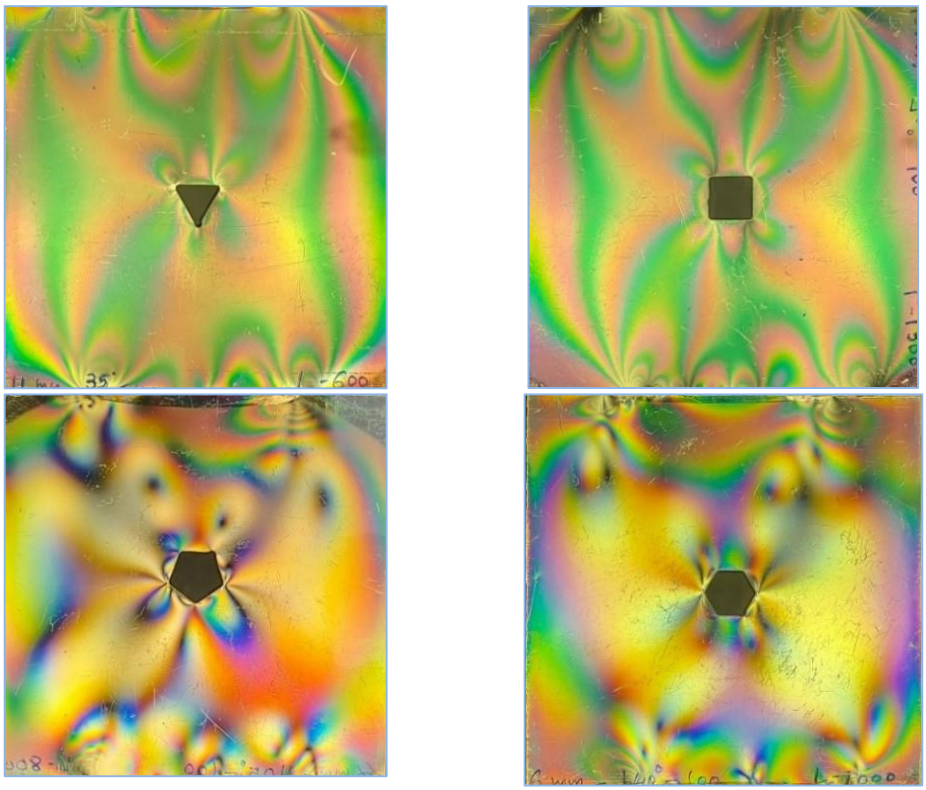
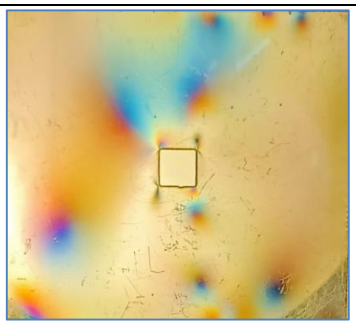
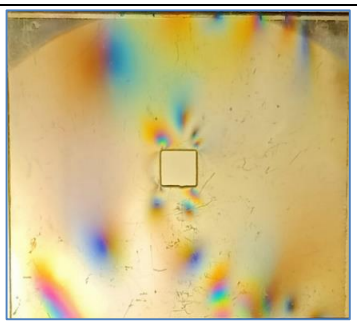


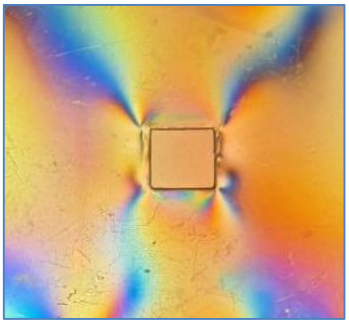
Figure 7: *Fringes generation in a plate with various polygonal holes at 0° orientation*



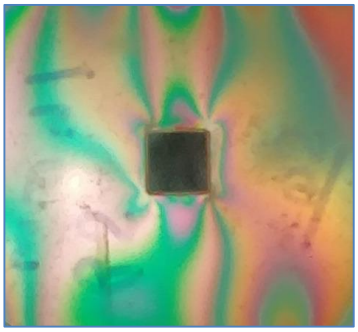
(a) n4 O70 A100 P600



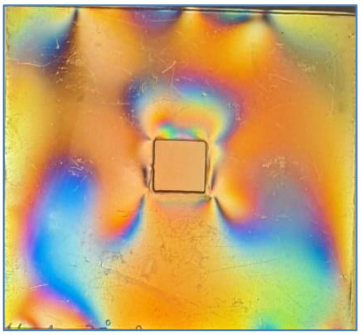
(b) n4 O70 A100 P1000



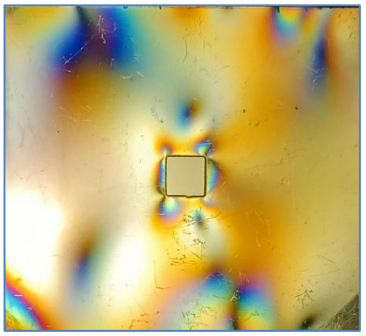
(c) n4 O70 A300 P600



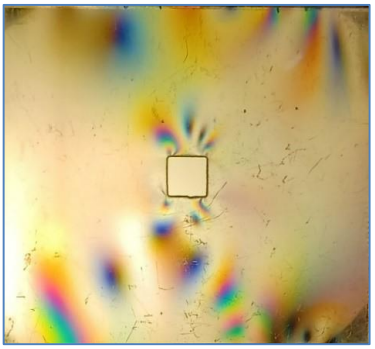
(d) n4 O70 A300 P1000



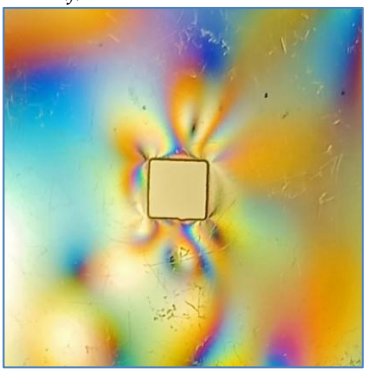
(e) n4 O105 A200 P800



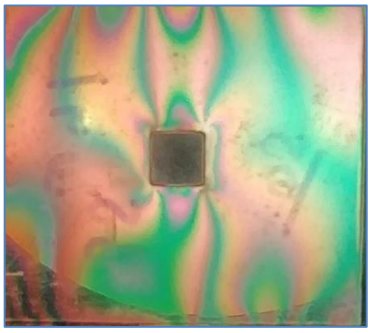
(f) n4 O140 A100 P600



(g) n4 O140 A100 P1000

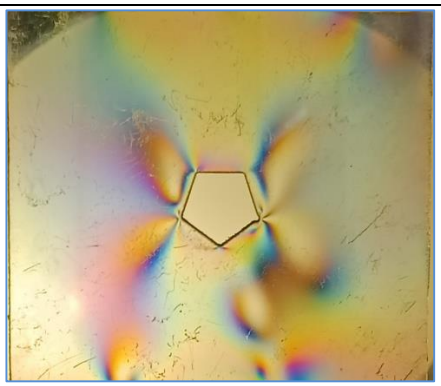


(h) n4 O140 A300 P600

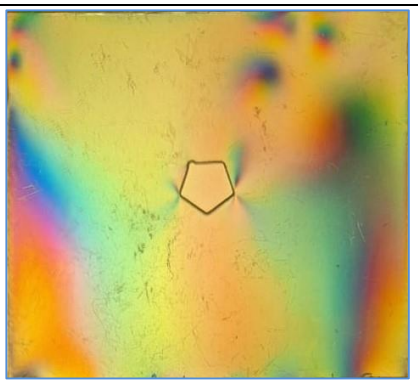


(i) n4 40 A300 P1000

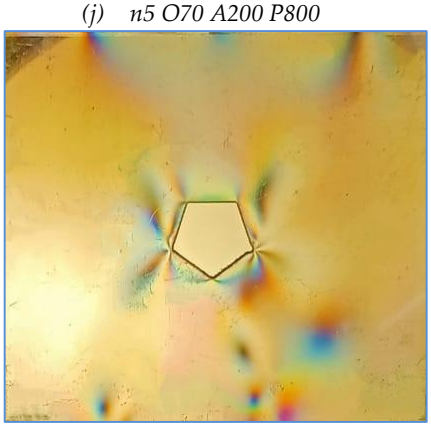
Figure 8: *Images of specimen sample captured during photoelastic experimental work (n = 4)*



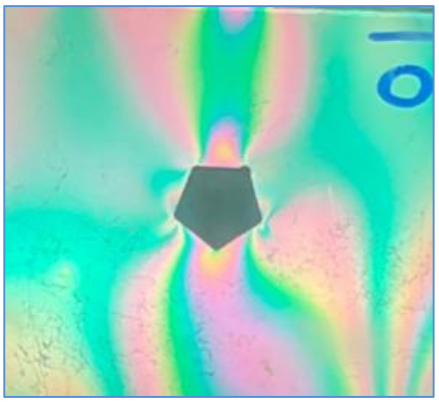
(j) n5 O70 A200 P800



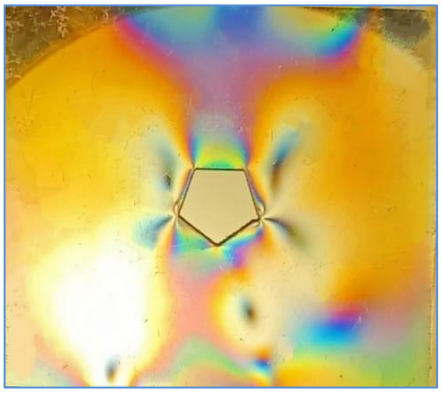
(k) n5 O105 A100 P800



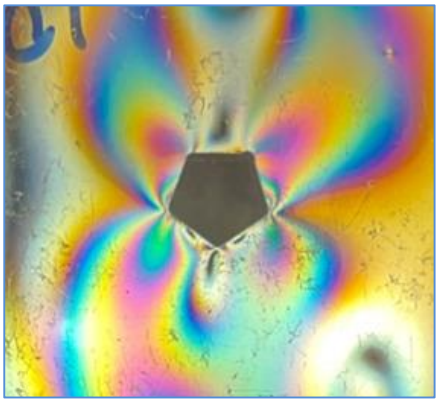
(l) n5 O105 A200 P600



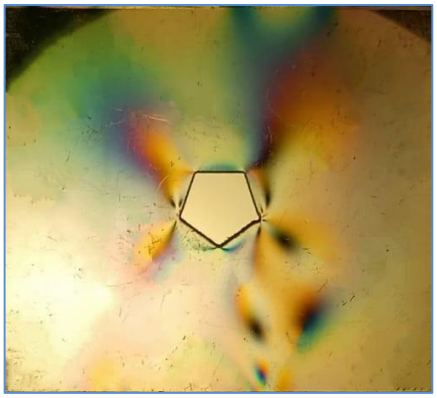
(m) n5 O105 A200 P800



(n) n5 O105 A200 P1000

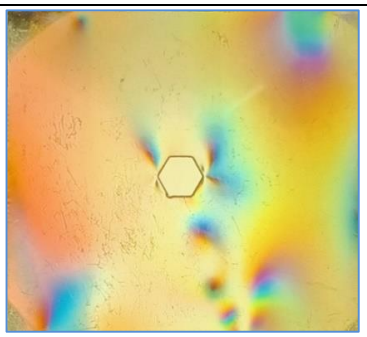


(o) n5 O105 A300 P800

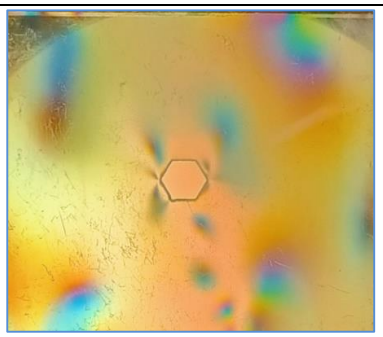


(p) n5 O140 A200 P800

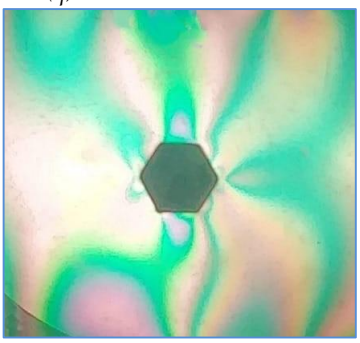
Figure 9: *Images of test specimen captured during photoelastic experimental work (n = 5)*



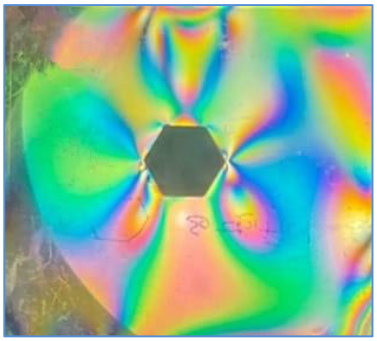
(q) n6 O70 A100 P600



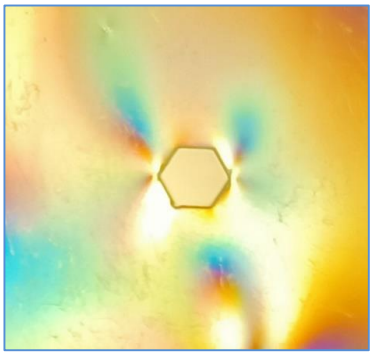
(r) n6 O70 A100 P1000



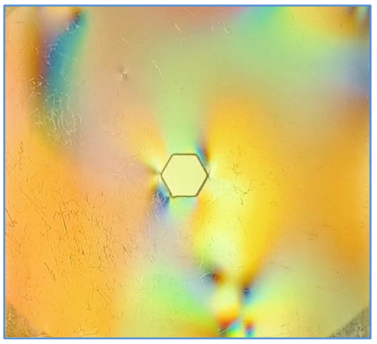
(s) n6 O70 A300 P600



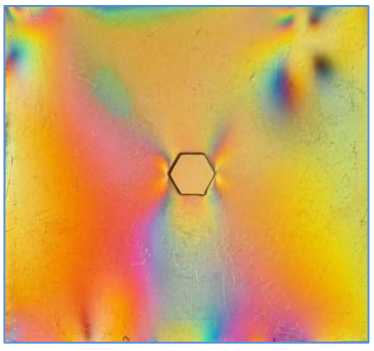
(t) n6 O70 A300 P1000



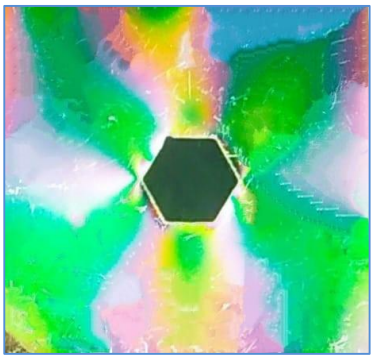
(u) n6 O105 A200 P800



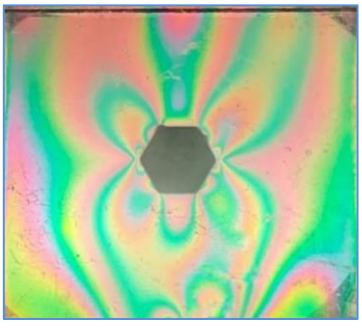
(v) n6 O140 A100 P600



(w) n6 O140 A100 P1000



(x) n6 O140 A300 P600



(y) n6 O140 A300 P1000

Figure 10: *Images of test specimen captured during photoelastic experimental work (n = 6)*

Before the starting of twenty-five experiments given by the DOE (RSM) methods, here are some results presented for the triangular, square, pentagonal and hexagonal holes at orientation angle 0°, hole area 100 mm² and applied load on the plate is 1000N.

For the parametric study, various plate parameters are taken into account, while ensuring that one edge of each polygonal hole remains parallel to the x-axis, with a fiber orientation of 0° (initial position of the plate). The selected load value is determined based on the feasible load range applied through the experimental setup, and the same load range is applied during the experiments. The experimental results in terms of fringes for various parameters combinations given by the RSM method are presented in Figure 8 to 10.

Here, in Figure 8 to 10, n represents the number of sides of the polygonal hole in a plate, O represents polygonal hole orientation (°), A represents the area of the polygonal hole (mm²) and P represents an applied load (N).

Table 1: *Results of photoelastic experimental work of finite plate*

L (mm)	H (mm)	t (mm)	n	O (°)	A (mm²)	P (N)	N	f_σ kN/m	Max. stress (N/mm^2)	Nominal stress (N/mm^2)	SCF
100	100	5	4	70	100	600	3	11	6.6	1.38	4.80
100	100	5	4	70	100	1000	5	11	11.0	2.29	4.80
100	100	5	4	70	300	600	4	11	8.8	1.54	5.71
100	100	5	4	70	300	1000	6.6	11	14.5	2.57	5.65
100	100	5	4	105	200	800	4.13	11	9.1	1.94	4.70
100	100	5	4	140	100	600	4	11	8.8	1.40	6.30
100	100	5	4	140	100	1000	6.1	11	13.4	2.33	5.76
100	100	5	4	140	300	600	5.13	11	11.3	1.59	7.10
100	100	5	4	140	300	1000	8.6	11	18.9	2.65	7.15
100	100	5	5	70	200	800	5	11	11.0	1.92	5.72
100	100	5	5	105	100	800	4	11	8.8	1.82	4.85
100	100	5	5	105	200	600	3.6	11	7.9	1.44	5.49
100	100	5	5	105	200	800	4.6	11	10.1	1.92	5.26
100	100	5	5	105	200	1000	6	11	13.2	2.41	5.49
100	100	5	5	105	300	800	6	11	13.2	2.02	6.55
100	100	5	5	140	200	800	5	11	11.0	1.93	5.71
100	100	5	6	70	100	600	3.1	11	6.8	1.36	5.02
100	100	5	6	70	100	1000	5.13	11	11.3	2.26	4.98
100	100	5	6	70	300	600	4	11	8.8	1.50	5.85
100	100	5	6	70	300	1000	6.6	11	14.5	2.51	5.79
100	100	5	6	105	200	800	4.6	11	10.1	1.93	5.25
100	100	5	6	140	100	600	3	11	6.6	1.37	4.83
100	100	5	6	140	100	1000	4.6	11	10.1	2.28	4.44
100	100	5	6	140	300	600	4	11	8.8	1.52	5.78
100	100	5	6	140	300	1000	7.6	11	16.7	2.54	6.59

Table 1 shows the results of photoelastic experimental work on 5 mm thick plates under axial compression loading. Also, Table 1 presents twenty-five experimental results for various polygonal hole sides, orientations (holes), constant hole areas, and applied loads. Based on the obtained results, the maximum stresses of the polygonal holes within the finite flat plates are identified and utilized for calculating the stress concentration factor.

4. Results and Discussion

There are no methods available for choosing a polygonal hole in a plate that will provide the lowest SCF for a particular geometry following the industry standard for hole design in plates. Therefore, efforts are being made to develop criteria for the selection of a polygonal hole in a plate that will give the least SCF. The FEA, design of experiment (RSM), and experimental method as photoelasticity, are presented in the previous section.

A range of SCF is developed for triangular, square, pentagonal, and hexagonal holes under uniaxial compression loading. Optimal polygonal hole shapes, orientations, areas, and applied loads on the plate are determined from various ranges of SCF. Here, results obtained from the DOE (RSM) and FEA methods are compared with those from photoelasticity experimental methods.

It is observed from Figure 8 to 10 that the vertices are points of high stress concentration in a plate. Hence, the stress field in a finite isotropic plate can be influenced by factors such as the number of vertices, their positions, and the direction of loading along a specific axis.

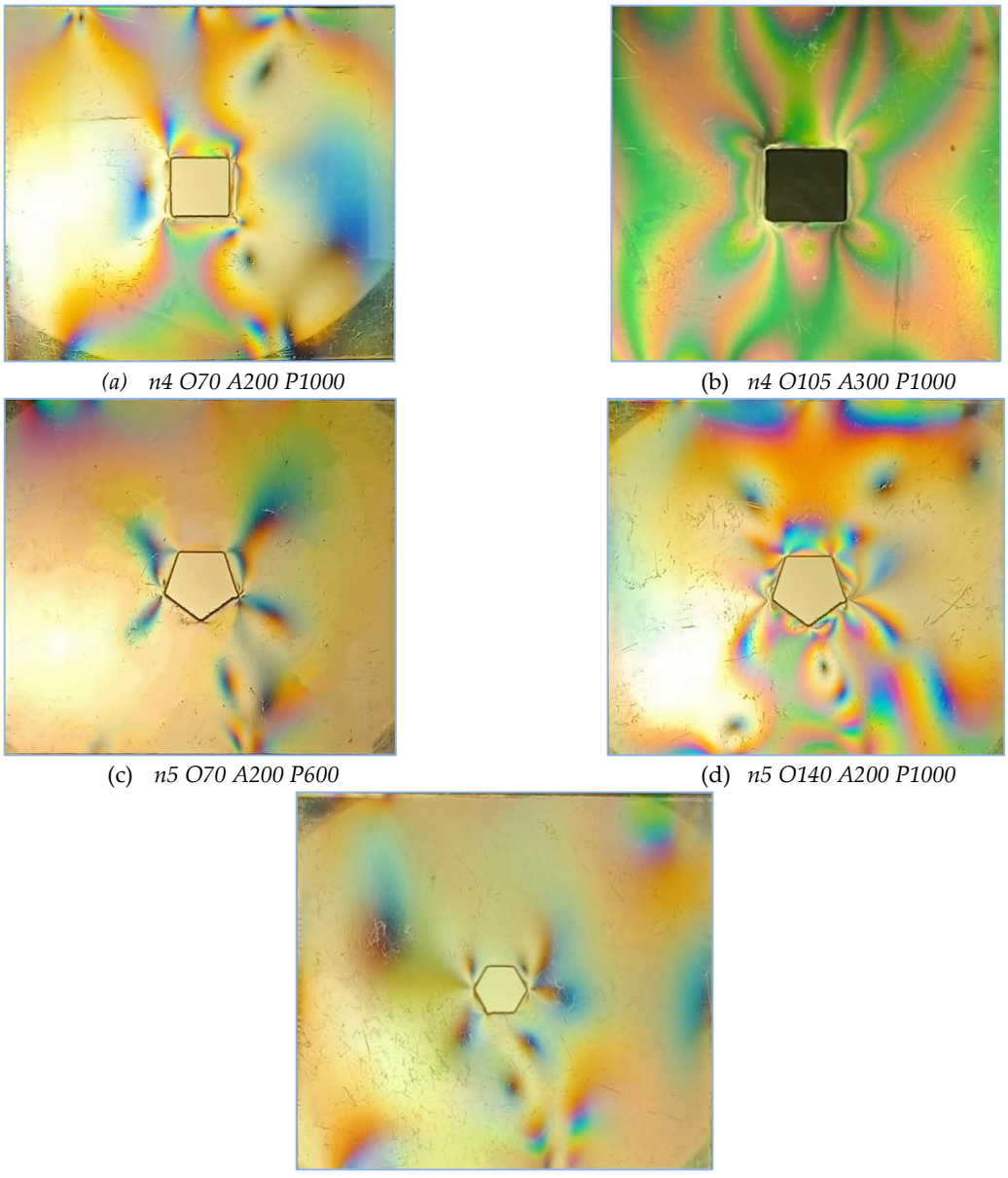
Table 2: *SCF obtained for various polygonal shape hole in a plate from FEA, RSM and Photoelasticity.*

L (mm)	H (mm)	t (mm)	n	O (°)	A (mm^2)	P (N)	FEA SCF	RSM SCF	Experimental SCF	(%) Error FEA vs Experiment	(%) Error RSM vs Experiment
100	100	5	4	70	100	600	4.72	4.95	4.80	-1.6	3.10
100	100	5	4	70	100	1000	4.72	4.95	4.80	-1.6	3.13
100	100	5	4	70	300	600	5.69	5.99	5.71	-0.3	4.74
100	100	5	4	70	300	1000	5.69	5.98	5.65	0.7	5.60
100	100	5	4	105	200	800	4.94	5.00	4.70	4.9	6.09
100	100	5	4	140	100	600	5.9	6.17	6.30	-6.8	-2.02
100	100	5	4	140	100	1000	5.9	6.17	5.76	2.3	6.51
100	100	5	4	140	300	600	7.14	7.49	7.10	0.5	5.22
100	100	5	4	140	300	1000	7.15	7.30	7.15	0	2.05
100	100	5	5	70	200	800	5.55	5.51	5.72	-3.1	-3.83
100	100	5	5	105	100	800	5.08	4.80	4.85	4.6	-0.95
100	100	5	5	105	200	600	5.68	5.72	5.49	3.4	3.97
100	100	5	5	105	200	800	5.68	5.61	5.26	7.4	6.28
100	100	5	5	105	200	1000	5.68	5.60	5.49	3.4	1.99
100	100	5	5	105	300	800	6.22	6.31	6.55	-5.3	-3.76
100	100	5	5	140	200	800	5.86	6.05	5.71	2.5	5.67
100	100	5	6	70	100	600	5.02	5.35	5.02	0	6.13
100	100	5	6	70	100	1000	5.02	5.35	4.98	0.7	6.82
100	100	5	6	70	300	600	5.76	6.15	5.85	-1.6	4.86
100	100	5	6	70	300	1000	5.76	5.85	5.79	-0.6	0.96
100	100	5	6	105	200	800	5.18	5.33	5.25	-1.4	1.45
100	100	5	6	140	100	600	4.36	4.60	4.83	-10.7	-4.96
100	100	5	6	140	100	1000	4.36	4.60	4.44	-1.9	3.42
100	100	5	6	140	300	600	5.69	5.95	5.78	-1.6	2.83
100	100	5	6	140	300	1000	6.45	6.61	6.59	-2.2	0.33

In Table 2, twenty-five results each for FEA, RSM, and photoelasticity are tabulated for various polygonal hole shapes, orientations, constant hole areas, and applied loads on a plate. The maximum

stresses of the plates with holes are determined from the study and employed in the calculation of the stress concentration factor (SCF).

Table 2 show the comparison of SCFs based on FEA, RSM, and photoelasticity experiments for different polygonal hole shapes (N = 4, 5, 6), orientation angles (70°, 105°, 140°), polygonal hole areas (100, 200, 300), and applied loads (600 N, 800 N, 1000 N). Observations from Table 2 indicate, that as the size of the polygonal hole increases, the stress concentration also increases. Figure 8 to 10 shows that the generated stress concentration factor varies at various orientations of the polygonal holes due to its rotation concerning the loading direction.



(a) n4 O70 A200 P1000

(b) *n4 O105 A300 P1000*

(c) *n5 O70 A200 P600*

(d) *n5 O140 A200 P1000*

(e) *n6 O105 A100 P800*

Figure 11: *Images of test sample captured during photoelastic experimental work (intermediate)*

Intermediate results from experiments and finite element analysis are essential for validating the outcomes of various methods. The validation of the obtained RSM equation is carried out by performing experiments and FEA for the intermediate readings (these readings not covered in Table 2) and presented in Table 3 and, the experiment results for various parameter combinations (these readings not covered in Table 2), are presented in Figure 11.

Table 3: *SCF obtained for various polygonal shape hole in a plate from FEA, RSM and Photoelasticity (validation results).*

L (mm)	H (mm)	t (mm)	n	O (°)	A (mm²)	P (N)	FEA SCF	RSM SCF	Experimental SCF	Error FEA vs Experiment (%)	Error RSM vs Experiment (%)
100	100	5	4	70	200	1000	5.33	5.50	5.40	-1.4	1.76
100	100	5	4	105	300	1000	5.24	5.35	5.20	0.7	2.79
100	100	5	5	70	200	600	5.55	5.60	5.49	1.0	1.92
100	100	5	5	140	200	1000	5.86	6.00	6.03	-2.9	0.50
100	100	5	6	105	100	800	4.36	4.60	4.34	0.4	5.55

Among this all parameters, when we are considering three parameters as a constant, and variable as a one parameter, it is shows results as presented in the Figure 12.

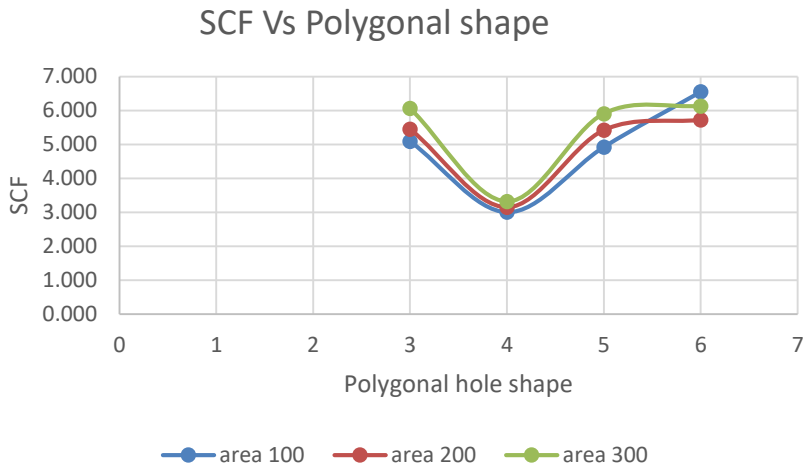


Figure 12: *SCF vs. polygonal hole shape*

4.1 Effect of Polygonal Hole Shape on a Finite Plate

For uniaxial Y loading, the maximum values of SCF are observed for the center positions of holes in the case of n = 3 and n = 5, while the minimum values of SCF are observed for the center positions of holes with n = 4 and n = 6. This may be because triangular and pentagonal holes have two vertices on the X-axis (at 0°), while the rest of the vertices lie in the plane. On the other hand, square and hexagonal holes have one vertex on the X-axis (at 0° and 180°) and the others lie in the plane. It is noted that the locations of high stress concentration are similar in both square (n = 4) and hexagon (n = 6) shapes, as both exhibit symmetry concerning to the X-Y axis. Figure 12 shows a graph of SCF versus polygonal hole shape. It is observed that increasing the number of sides of a polygon affects the observed minimum and maximum values of SCF generation. For example, square and hexagonal holes are symmetrical about the x and y axes, resulting a range of SCF is between 3.01 to 6.16 and 4.01 to 6 respectively. Therefore, the difference between the maximum and minimum values of SCF is larger. Conversely, triangular and pentagonal holes, which are non-symmetrical about both the x and y axes, generate higher SCF values but with a range of SCF is between 4.26 to 4.94 and 5.36 to 6.01 respectively. Consequently, the difference between the maximum and minimum values of SCF is smaller. For unidirectional and static loading, it is advised to use square and hexagonal holes oriented in such a way that generate SCF value 3. The recommendation is to employ triangular and pentagonal holes arranged in a manner that produces minimal variations in SCF, ranging between 4 to 6. These triangular and pentagonal shapes of polygonal holes generate higher SCF, but they can

sustain variable loads. The direction of loading determines the appropriate orientation of polygonal holes during the design of the structure.

4.2 Effect of Orientation of Polygonal Hole

For uniaxial Y loading, if the maximum vertices (corners) are along the X-axis, it shows the highest stresses in all the polygonal holes. The vertex of the highest stress is not necessarily on the X or Y axis but may lie in a plane depending on the geometry of the polygonal shape. As a special case, for square holes with 0°, 90°, and 180° angles, the minimum SCF is obtained. Due to the maximum distance generated between two corners of the square hole when positioned parallel to the loading direction, the maximum SCF shall be at 45° and 135° angles, indicating parallel edges of the square holes in the direction of loading, thereby always reducing the stresses by up to 50%. As we maintain the variable as the orientation of the polygonal hole, while considering other parameters as constant, increasing the orientation of the hole for selected parameters results in an increase in the SCF value up to a certain orientation. After reaching this point, further increases in orientation lead to a decrease in the SCF value.

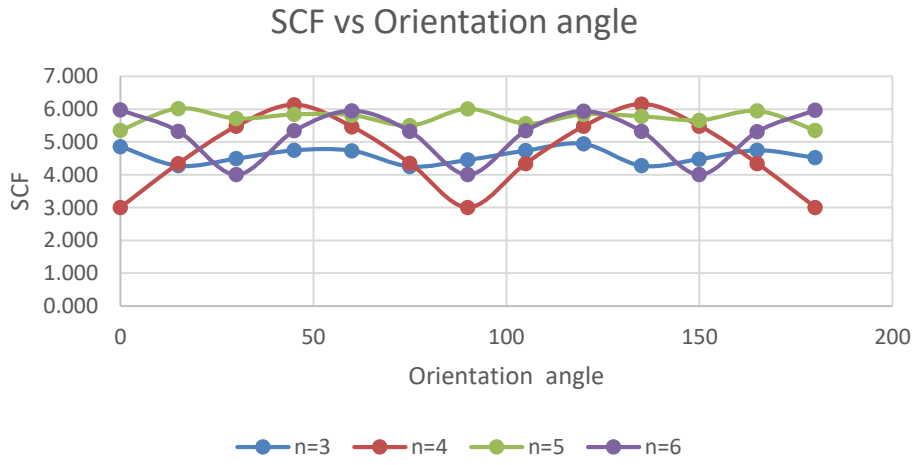


Figure 13: *SCF vs. Orientation angle*

It has been observed that the curve exhibits a repeating nature at certain orientations. Specifically, the curve repeats at intervals of 60°, 90°, 75°, and 60° for triangular, square, pentagonal, and hexagonal shapes respectively. This phenomenon arises due to the geometry of the hole's shape. Figure 13 illustrates a graph of SCF versus the orientation of the polygonal hole shape.

4.3 Effect of the Area of Polygonal Hole

As far as the area of the polygonal hole is concerned, if the area of any polygonal hole increases in a finite plate, the net area will always reduce, generating more stresses and a higher Stress Concentration Factor (SCF). The SCF of all shapes is found to increase as the number of polygon sides increases for a constant area. It is observed that hole and symmetry are more affected by SCF for the same area of all shapes of the polygonal hole. As we keep the variable as the area of the polygonal hole and other parameters constant, an increase in the area of the polygonal hole in a finite plate always results in a reduction in the net area, thereby increasing the stresses and SCF. Figure 14 shows a graph of SCF vs. the constant area of the polygonal hole shape.

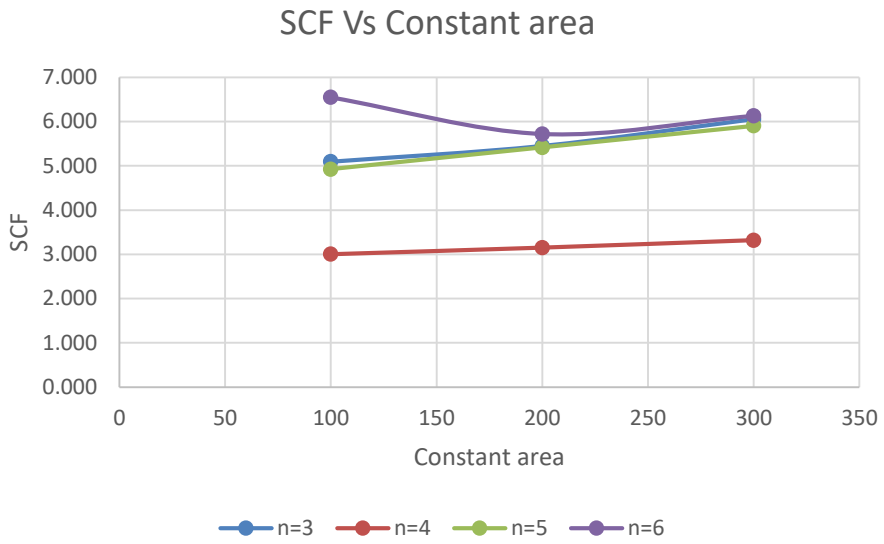


Figure 14: *SCF vs. Constant area of polygonal hole*

4.4 Effect of Applied Load on a Finite Plate

If the applied load on the finite plate is increasing, it does not produce any significant effect on the stress concentration factor for all polygonal shapes in a finite plate. As the load increases on the plate, von Mises stresses increase, and simultaneously nominal stresses also increase, showing a negligible effect on the SCF. Therefore, the applied load is not a significant factor for the SCF for the presented load cases and load range (600 N to 1000 N).

5. Conclusion

A numerical verification of the presented method is carried out using the ANSYS software, Minitab, and a good agreement of stress distributions is found.

It is advised to use square and hexagonal shaped holes at particular orientations when minimum stress concentration factor values are generated is 3.00 for unidirectional and static loading. At points where there is no or negligible variation in load occurs it's advisable to use triangular and pentagonal holes at suitable orientations, where the stress concentration factors generate within the range of 4.00 to 6.00. During the design stage, the direction of loading and arrangement of the orientation of these polygonal shapes are important factors to study.

Among all holes, when the baseline is parallel to the x-axis, the increasing order of stresses and SCF is as follows: square, triangular, pentagonal, and hexagonal.

The stress concentration factor is significantly influenced when one of the corners of the polygonal hole aligns with the loading direction. Consequently, as the side ratio increases, a square hole exhibits the minimum SCF compared to triangular, pentagonal, and hexagonal holes.

A square-shaped hole yields a 40% lower minimum stress concentration factor compared to a hexagonal-shaped hole, primarily because a greater number of edges are parallel to the loading directions

It is preferable to create a hole with the maximum number of edges parallel to the loading direction.

The stress concentration factor rises with an increase in the side ratio for a specific polygonal hole shape, owing to the enlargement of the hole and its area. Larger holes with more edges result in a higher SCF value due to the presence of additional corners.

References

[1] M. Ismail, L. Žiković, N. Čeh, and G. Jelenić. (2022). Experimental and numerical analysis of stress concentration in a plate with a circular hole. *10th International Congress of Croatian Society of Mechanic, 10*:1-2.

[2] Manish Pandey, Aprajita Patel, Kshitiz Jaiswal, Lalit Kirola, and Subodh Kumar Sharma. (2021). Investigation of Variation in Stress Concentration Factor with the Change in Orientation of Central Hole on a Rectangular Plate. *Recent Advances in Mechanical Engineering, Lecture Notes in Mechanical Engineering, 77: 629-635.*

[3] X. Zhang, A. Lu, K. Song, G. Wu, and L. Wang. (2022). Methodology for Determining the Compliance Matrix and Analyzing the Skew Anisotropic Plate with an Arbitrarily Shaped Hole. *Advances in Materials Science and Engineering. 2022:1–14, doi: 10.1155/2022/5949398.*

[4] YI, W., RAO, Q. H., MA, W. B., SUN, D. L., and SHEN, Q. Q. (2020). A new analytical-numerical method for calculating interacting stresses of a multi-hole problem under both remote and arbitrary surface stresses, *Appl. Math. Mech. -Engl. Ed. 41:1539-1560, doi: https://doi.org/10.1007/s10483-020-2653-9.*

[5] W. Wang and B. J. Spencer. (2021). Numerical solution for the stress near a hole with corners in an infinite plate under biaxial loading. *Journal of Eng Math. 127:1-131, doi: 10.1007/s10665-021-10104-8.*

[6] S. Badiger and R. D. S. (2023). Stress distribution in an infinite plate with discontinuities like elliptical or circular hole by boundary force method, *SN Appl. Sci., 5:77, doi: 10.1007/s42452-023-05289-9.*

[7] B. Safaei, Z. Pezeshki, K. Kotrasova, and E. Kormanikova. (2022). Analysis of stress concentration at the edge of hole in plates with different widths by using FEM. *IOP Conf. Ser.: Mater. Sci. Eng. 1252:1-11, doi: 10.1088/1757-899X/1252/1/012067.*

[8] P. Jaiswal, S. Makin, A. D. Dubey, and G. Ghangas. (2022). Analysis of stress concentration reduction around rounded rectangular slot with FGM ring. *Materials Today: Proceedings, vol. 50:1953–1957, doi: https://doi.org/10.1016/j.matpr.2021.09.323.*

[9] P. Rani, D. Varma, and G. Ghangas. (2022). Stress concentration analysis of functionally graded material coated elliptical inclusion under uniaxial tension. *Materials Today: Proceedings, 78:351-358, doi: https://doi.org/10.1016/j.matpr.2022.09.602.*

[10] O. Maksymovych, T. Solyar, A. Sudakov, I. Nazar, and M. Polishchuk. (2021). Determination of stress concentration near the holes under dynamic loadings, *Nauk. visn. nat. hirn. Univ. 3:19–24., doi: 10.33271/nvngu/2021-3/019.*

[11] O. V. Maksymovych and T. Ya. Solyar. (2022). Determination of Stress Concentration Near Dies, Holes, and Cracks in the Half Plane Based on the Method of Integral Equations and Green Solutions. *Journal of Math Sci, 261:162–175, doi: 10.1007/s10958-022-05745-8.*

[12] M. Patil and Ajay More. (2021). Effect of Hexagonal Cutout Orientation and Roundness of Edges on Stress Concentration in Plate. *International Research Journal of Engineering and Technology (IRJET), 08:4431–4435.*

[13] M Mohan Kumar, Rajesh S, Yogesh H, and Yeshaswini B R. (2013). Study on the Effect of Stress Concentration on Cutout Orientation of Plates with Various Cutouts and Bluntness. *International Journal of Modern Engineering Research (IJMER), 3:1295–1303.*

[14] R. H. Patel and B. P. Patel. (2023). Analyzing Stress Concentration Factor in Finite Plate with Different Polygonal Discontinuities Under Uniaxial Compression Using FEM. *Journal of Kejuruteran. 35:1033–1043, doi: 10.17576/jkukm-2023-35(5)-05.*

[15] R. H. Patel and B. P. Patel. (2022). Effect of various discontinuities present in a plate on stress concentration: a review. *Engineering Research Express, 4:1–15. doi: https://doi.org/10.1088/2631-8695/ac8c1b.*

[16] "James W. Phillips "Photoelasticity, Experimental Stress Analysis, TAM 326."

LEHMANN TYPE-II PERK DISTRIBUTION: PROPERTIES AND APPLICATIONS

Venugopal Haridoss[1]

Sudheep Jose[2*]

Thomas Xavier[3]

•

Department of Statistics, Madras Christian College, Chennai - 600 059, India [1],[2]
Principal Biostatistician, Novartis, Hyderabad, 500081, India[3]
haridoss@mcc.edu.in [1]
sudheepezhanikattu@gmail.com [2*]
thomasmxavier@gmail.com [3]
[*]Corresponding author

Abstract

The Lehmann type-II Perk distribution is a flexible statistical model with a wide range of applications in fields such as reliability analysis, survival modeling, and data fitting. This distribution is notable for its distinct properties, including specific patterns in hazard rates and implications for stochastic ordering. Estimating the distribution parameters is essential for effective model fitting and making inferences. The parameters are estimated using the maximum likelihood estimation method, and confidence intervals are determined using normal approximation. To evaluate the performance of these estimation methods, Monte-Carlo simulation studies are conducted, demonstrating their accuracy and efficiency. The Lehmann type-II Perk distribution provides a robust framework for analyzing complex data sets and deriving reliable statistical conclusions.

Keywords: Lehmann type-II Perk distribution, survival function, hazard function, maximumlikelihood estimates, confidence length, Monte-Carlo simulation.

1. Introduction

Researchers in the scientific community have widely embraced Lehmann's [6] type-I (L-I) and type-II (L-II) lifetime models as straightforward and practical tools. Lehmann's type-I (L-I) model is commonly associated with the Power function (PF) distribution, which has garnered significant attention for its simplicity and utility. Additionally, Gupta et al. [8] applied the L-I model to the exponential distribution. The appeal of the PF distribution's simplicity and versatility has led researchers to explore its various applications, extensions, and generalizations across different scientific domains. Furthermore, Cordeiro and De Castro [9] introduced the Lehmann Type-II (L-II) G class through a dual transformation. Cordeiro et al. [5] developed the Lehman type II distribution as a hybrid of the generalized exponentiated distribution. This distribution's closed-form characteristics facilitate the derivation and examination of numerous properties. Researchers have extensively employed both the L-I and L-II approaches in the literature to investigate novel features of classical and modified models. In recent studies, Arshad et al. [12] delved into the development of the L-II G

family, focusing on a bathtub-shaped failure rate model and its application using engineering data. Balogun et al. [14] introduced a potentiated lifetime model exhibiting a bathtub-shaped hazard rate function, termed the Kumaraswamy modified size-biased Lehmann Type-II (Kum–MSBL–II) distribution. Ogunde et al. [1] defined and studied a new generalization of the Frechet distribution, known as the Lehmann Type II Frechet Poisson distribution. Tomazella et al. [23] explored various mathematical properties of the L-II Frechet distribution and its application to aircraft maintenance data, while Awodutire et al. [16] discussed multiple statistical measures of the L-II generalized half logistic distribution, examining its application in sports data. In a different context, Badmus et al. [13] investigated the weighted Weibull distribution via the L-II approach and its application in textile engineering data. Meanwhile, Ogunde et al. [2] extended the Gumbel type-II distribution using the exponentiated L-II G class and applied it to biology data.

The Perks distribution, originally proposed by Perks in [15], is a four-parameter model that also serves as an extension of the Gompertz-Makeham distribution. Researchers in the field have actively explored modifications and generalizations of this model within the literature. Richards, in both [18] and [19], introduced parametric survival models based on the Perks I distribution. Haberman and Renshaw, in [10], conducted a study focused on parametric mortality projection using the Perks distribution as a foundation. Chaudhary and Kumar, in [4], delved into Bayesian analysis of the Perks distribution employing Markov Chain Monte Carlo techniques. In [24], Zeng et al. examined both four and five-parameter Perks mortality equations to model bathtub-shaped failure rates. In another extension of the Perks distribution, Singh and Choudhary [21] introduced the exponentiated Perks distribution, while Chaudhary [3] presented the Perks-II distribution. Most recently, Gonzalez et al. [7] conducted a study on the additive Perks distribution and explored its application in the realm of reliability analysis.

The distribution function of the Lehmann type-II is given by

$$F(x) = 1 - (1 - G(x))^{\alpha}, x > 0, \alpha > 0. \tag{1}$$

The distribtion function of the perk II is given by

$$G(x) = 1 - \left(\frac{1+\alpha}{1+\alpha e^{\beta x}}\right)^{\frac{1}{\beta}}, x > 0, \alpha, \beta > 0. \tag{2}$$

$$F_{LPM}(x) = 1 - \left(\frac{1+\alpha}{1+\alpha e^{\beta x}}\right)^{\frac{\gamma}{\beta}}, x > 0, \alpha, \beta, \gamma > 0. \tag{3}$$

which is a three parameter Lehmann type-II Perk distribution. The corresponding density is given by

$$f_{LPM}(x) = \alpha\gamma\,(1+\alpha)^{\frac{\gamma}{\beta}}\, e^{\beta x} \left[\frac{1}{1+\alpha e^{\beta x}}\right]^{1+\frac{\gamma}{\beta}}, x > 0, \alpha, \beta, \gamma > 0. \tag{4}$$

and $f_{LPM}(x) = 0$ otherwise. A plots of the density function is shown in Figure 1.

2. Properties

This section studies various properties such as the survival and hazard function, quantile function.

2.1. Survival and hazard function

The survival function of the density given in Eq. (5) as

$$S_{LPM}(x) = \left(\frac{1+\alpha}{1+\alpha e^{\beta x}}\right)^{\frac{\gamma}{\beta}} \tag{5}$$

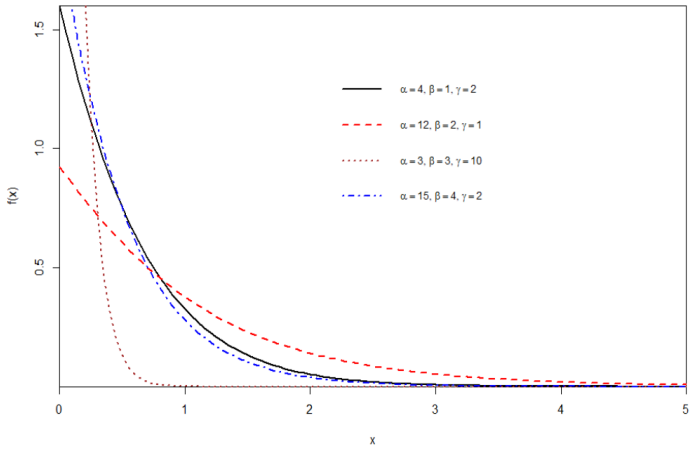


Figure 1: *Few shapes of Lehmann II Perks distribution.*

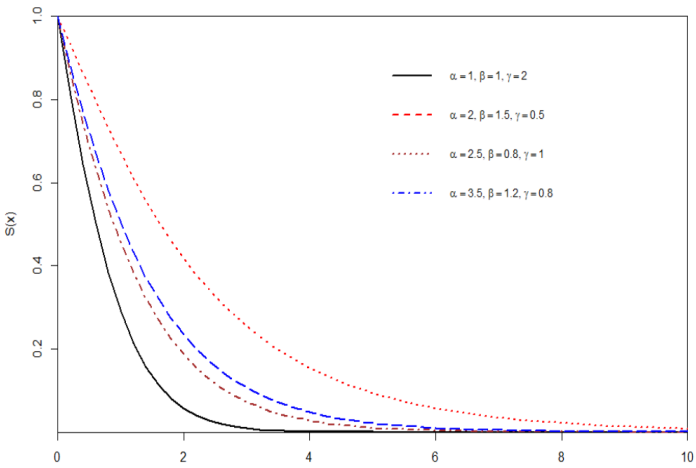


Figure 2: *Survival shapes of Lehmann II Perks distribution*

and the hazard function as $h_{LPM}(x) = \frac{f_{LPM}(x)}{S_{LPM}(x)}$, and hence

$$h_{LPM}(x) = \alpha\gamma \left[\frac{e^{\beta x}}{1 + \alpha e^{\beta x}} \right] \tag{6}$$

The different shapes of the survival and hazard function are given in the Figure 2 and 3.

2.2. The Quantile function

The quantile function $Q(u)$ is defines

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 \leq u \leq 1 \tag{7}$$

The cumulative distribution function $F(x)$ is defined as $F(x) = 1 - S(x)$. Let's consider $F(x) = u$, resulting in the following expression for x:

$$x = \frac{1}{\beta} \ln \frac{1}{\alpha} \left[\frac{1 + \alpha}{(1-u)^{\frac{\beta}{\gamma}}} - 1 \right] \tag{8}$$

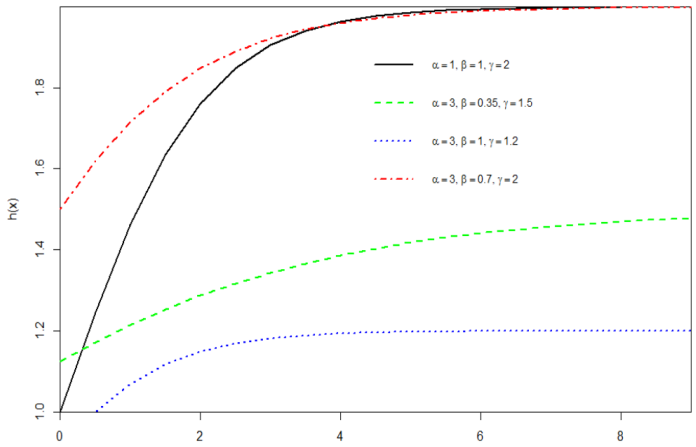


Figure 3: *Hazard shapes of Lehmann II Perks distribution*

2.3. Likelihood ratio ordering

For two random variables X_1 and X_2 where X_1 is larger than X_2 in likelihood ratio ordering, denoted by $X_1 \geq_{lr} X_2$, if $\frac{f_{X_1}(x)}{f_{X_2}(x)}$ is an increasing function in x, where $f_{X_1}(x)$ and $f_{X_2}(x)$ are density function of X_1 and X_2 respectively. Consider that X_1 and X_2 are independent random variables following Lehmann type-II Perks distribution with parameters $(\alpha, \beta, \gamma_1)$ and $(\alpha, \beta, \gamma_2)$ respectively and for $\gamma_1 \leq \gamma_2$. Then

$$\frac{f_{X_1}(x)}{f_{X_2}(x)} = \frac{\gamma_1}{\gamma_2}(1+\alpha)^{\frac{\gamma_1-\gamma_2}{\beta}}\left(1+\alpha e^{\beta x}\right)^{\frac{\gamma_2-\gamma_1}{\beta}}. \tag{9}$$

Further we can have,

$$\frac{\partial}{\partial z}\frac{f_{X_1}(x)}{f_{X_2}(x)} = \alpha\beta e^{\beta x}\frac{\gamma_1}{\gamma_2}\frac{(\gamma_2-\gamma_1)}{\beta}(1+\alpha)^{\frac{\gamma_1-\gamma_2}{\beta}}\left(1+\alpha e^{\beta x}\right)^{\frac{\gamma_2-\gamma_1}{\beta}-1} \tag{10}$$

For $\gamma_1 \leq \gamma_2$, $\frac{f_{X_1}(x)}{f_{X_2}(x)}$ is increasing and hence $X_1 \geq_{lr} X_2$. Shaked and Shanthikumar [20] suggests that likelihood ratio ordering implies hazard rate ordering and which implies stochastic ordering. Thus for $\gamma_1 \leq \gamma_2$, we have $X_1 \geq_{lr} X_2 \Rightarrow X_1 \geq_{hr} X_2 \Rightarrow X_1 \geq_{st} X_2$. This helps to understand one random variable being "bigger" than the other, or in our case when $\gamma_1 \leq \gamma_2$, X_1 is "bigger" than X_2.

2.4. Stress-strength reliability

Suppose X_1 and X_2 are independent random variables following Lehmann type-II Perks distribution with parameters $(\alpha, \beta, \gamma_1)$ and $(\alpha, \beta, \gamma_2)$ respectively. Let $f(x)$ be the density function of X_1 and $g(x)$ be the density function of X_2. Then the stress-strength reliability is

$$\begin{aligned} R &= \Pr\{X_1 > X_2\} = \int_0^\infty f(x_1)\left(\int_0^{x_1} g(x_2)dx_2\right)dx_1 \\ &= 1-\alpha\gamma_1(1+\alpha)^{\frac{\gamma_1+\gamma_2}{\beta}}\int_0^\infty e^{\beta x_1}\left[\frac{1}{1+\alpha e^{\beta x_1}}\right]^{1+\left(\frac{\gamma_1+\gamma_2}{\beta}\right)}dx_1 \\ &= \frac{\gamma_2}{\gamma_1+\gamma_2} \end{aligned} \tag{11}$$

3. Maximum Likelihood Estimation

Let $x_1, x_2, ..., x_n$ be a simple random sample from the density in Eq. (4), then the log-likelihood function is denoted by $L(\alpha, \beta, \gamma)$ is given by

$$\begin{aligned} L(\alpha,\beta,\gamma) &= n\ln\alpha + n\ln\gamma + \frac{n\gamma}{\beta}\ln(1+\alpha) \\ &+ \beta\sum_{i=1}^{n} x_i - \left(1+\frac{\gamma}{\beta}\right)\sum_{i=1}^{n}\ln\left(1+\alpha e^{\beta x_i}\right) \end{aligned} \tag{12}$$

To estimate the parameters α, β and γ by equating the partial derivative of the equation obtained in Eq. (12) with respect to each parameter to zero.

$$\frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} + \frac{n\gamma}{\beta(1+\alpha)} - \left(1+\frac{\gamma}{\beta}\right)\sum_{i=1}^{n}\frac{e^{\beta x_i}}{1+\alpha e^{\beta x_i}} \tag{13}$$

$$\frac{\partial L}{\partial \beta} = \frac{-n\gamma}{\beta^2}\ln(1+\alpha) + \sum_{i=1}^{n} x_i - \left(1+\frac{\gamma}{\beta}\right)\sum_{i=1}^{n}\frac{\alpha x_i e^{\beta x_i}}{1+\alpha e^{\beta x_i}} + \frac{\gamma}{\beta^2}\sum_{i=1}^{n}\ln \tag{14}$$

$$\frac{\partial L}{\partial \gamma} = \frac{n}{\gamma} + \frac{n}{\beta}\ln(1+\alpha) - \frac{1}{\beta}\sum_{i=1}^{n}\ln\left(1+\alpha e^{\beta\sum_{i=1}^{n} x_i}\right) \tag{15}$$

Now setting the above three Equations to zero, the maximum likelihood estimates of α, β, γ can be obtained by solving the non-linear equations numerically using analytical methods like Newton-Raphson algorithm. This can be done by using statistical package, R.

Under certain regulatory conditions, the estimator $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ is asymptotically normally distributed with mean $\theta = (\alpha, \beta, \gamma)$ and variance-covariance matrix $Cov(\hat{\theta}) = I(\theta)^{-1}$ where $I(\theta)$ is the Fisher information matrix or the expected value of $J(\theta) = -\frac{\partial^2 l}{\partial\alpha_i \partial\alpha_j}$ is the (i, j) element of observed information matrix. The normal approximation of the MLEof θ can be used to construct approximate confidence intervals of the parameters. The asymptotic distribution of $\sqrt{n}\left(\hat{\theta} - \theta\right)$ is $N_3\left(0, I^{-1}\left(\hat{\theta}\right)\right)$. The asymptotic multivariate normal distribution can be used to construct asymptotic confidence intervals for α, β and γ. The asymptotic $100\,(1-\eta)\,\%$ confidence intervals for α, β and γ are respectively $\left(\hat{\alpha} \pm z_{\frac{\eta}{2}} S.E(\hat{\alpha})\right)$, $\left(\hat{\beta} \pm z_{\frac{\eta}{2}} S.E(\hat{\beta})\right)$ and $\left(\hat{\gamma} \pm z_{\frac{\eta}{2}} S.E(\hat{\gamma})\right)$, where S.E.(.) is the square root of the diagonal element of $I^{-1}\left(\hat{\theta}\right)$ corresponding to each parameter and $z_{\frac{\eta}{2}}$ is the quantile $\left(1-\frac{\eta}{2}\right)$ of the standard normal distribution.

4. Simulation study

Here, we used a simulation study to investigate the performance of the accuracy of point estimates of the parameters of the Lehmann type-II Perk model (α, β, γ) distribution. The following steps were followed:

- Specify the sample size n and the values of the parameters α, β, γ.
- Generate $Ui \sim \text{Uniform}(0,1); i = 1, 2, 3...., n.$
- Set

$$x = \frac{1}{\beta}\ln\frac{1}{\alpha}\left[\frac{1+\alpha}{(1-u)^{\frac{\beta}{\gamma}}} - 1\right] \tag{16}$$

- Calculate the MLEs of the three parameters.
- Repeat steps 2-3, N times.
- Calculate the mean squared error (MSE) for each parameter.

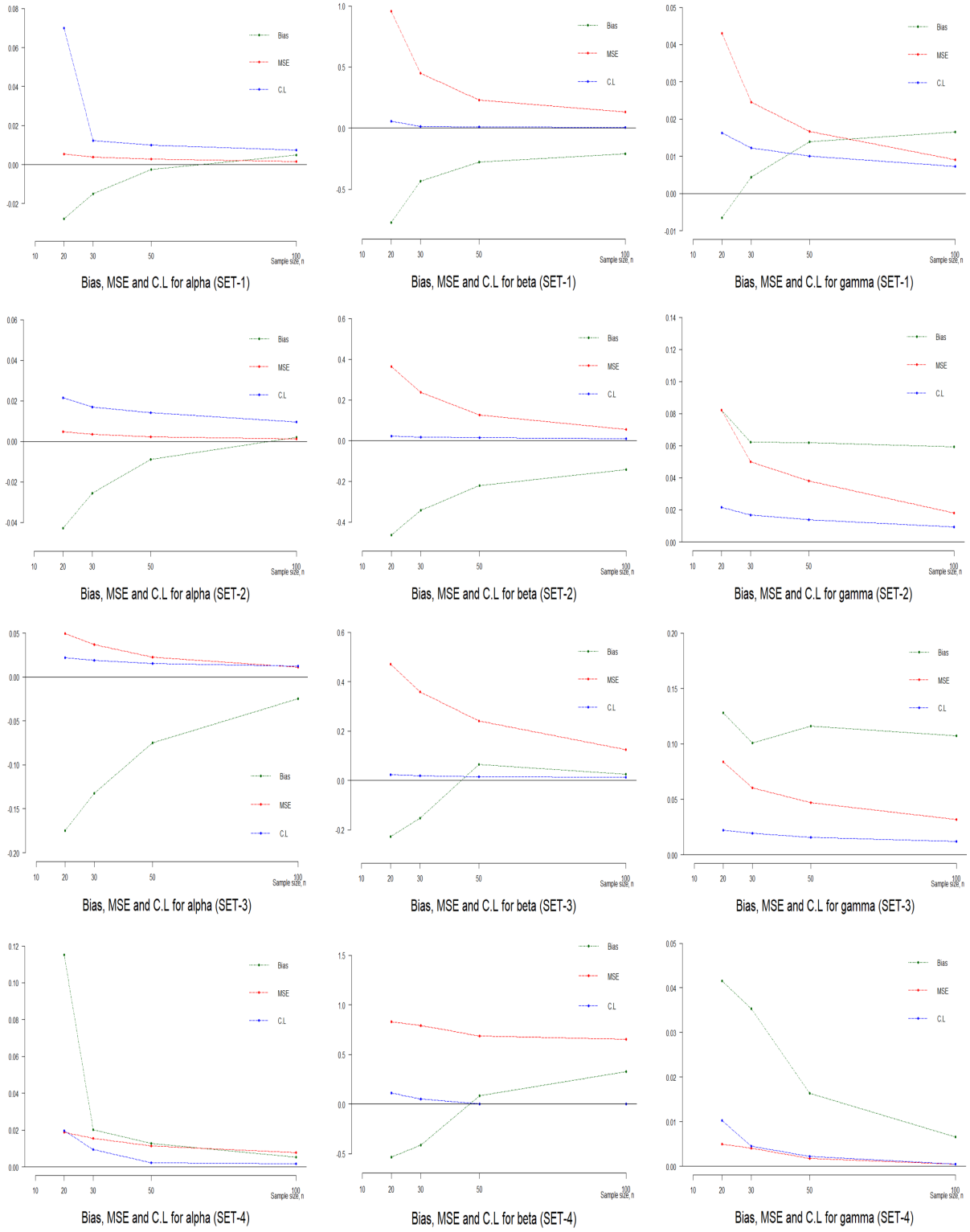


Figure 4: *Bias, mean square error (MSE) and confidence length (C.L)*

Table 1: *Bias, mean square error (MSE) and confidence length (C.L) of different set of parameters*

Sample Size	$\hat{\alpha}$ (Bias)	$\hat{\beta}$ (Bias)	$\hat{\gamma}$ (Bias)	$\hat{\alpha}$ (MSE)	$\hat{\beta}$ (MSE)	$\hat{\gamma}$ (MSE)	$\hat{\alpha}$ (C.L)	$\hat{\beta}$ (C.L)	$\hat{\gamma}$ (C.L)
SET-1									
20	-0.0278	-0.7736	-0.0066	0.0054	0.9552	0.0430	0.0699	0.0544	0.0162
30	-0.0151	-0.4347	0.0043	0.0038	0.4505	0.0245	0.0122	0.0122	0.0122
50	-0.0026	-0.2762	0.0139	0.0027	0.2311	0.0166	0.0100	0.0100	0.0100
100	0.0049	-0.2086	0.0165	0.0016	0.1314	0.0090	0.0073	0.0073	0.0073
SET-2									
20	-0.0428	-0.4639	0.0823	0.0048	0.3628	0.0822	0.0215	0.0215	0.0215
30	-0.0257	-0.3422	0.0623	0.0034	0.2380	0.0500	0.0168	0.0168	0.0168
50	-0.0089	-0.2217	0.0619	0.0021	0.1266	0.0379	0.0140	0.0140	0.0140
100	0.0018	-0.1427	0.0593	0.0011	0.0535	0.0180	0.0094	0.0094	0.0094
SET-3									
20	-0.1749	-0.2284	0.1278	0.0496	0.4697	0.0839	0.0222	0.0222	0.0222
30	-0.1321	-0.1534	0.1008	0.0370	0.3589	0.0605	0.0192	0.0192	0.0192
50	-0.0751	0.0645	0.1161	0.0224	0.2412	0.0468	0.0156	0.0156	0.0156
100	-0.0245	0.0249	0.1072	0.0112	0.1244	0.0319	0.0122	0.0122	0.0122
SET-4									
20	0.1151	-0.5313	0.0416	0.0188	0.8283	0.0049	0.0196	0.1115	0.0102
30	0.0203	-0.4140	0.0353	0.0154	0.7922	0.0040	0.0095	0.0515	0.0045
50	0.0127	0.0826	0.0163	0.0113	0.6858	0.0017	0.0022	0.0022	0.0022
100	0.0052	0.3295	0.0066	0.0076	0.6524	0.0005	0.0016	0.0005	0.0005

The comparison is based on MSEs and C.L. The MSEs and C.L were computed by generating one thousand replications of sample size $n = 20, 30, 50, 100$ from the Lehmann II Perks model with different parameter values. The required results are obtained based on the different combinations of the model parameters place in SET-1 ($\alpha = 0.1$, $\beta = 3$, $\gamma = 1$), SET-2 ($\alpha = 0.1$, $\beta = 2$, $\gamma = 1$), SET-3 ($\alpha = 0.3$, $\beta = 2$, $\gamma = 1$) and SET-4 ($\alpha = 0.1$, $\beta = 3$, $\gamma = 0.2$), which are shown in Tables 1. Figures 4 displays the bias, mean squared error and confidence length for different sample sizes. The assessment based on simulation study is that the MSEs for each parameter decreases with increasing sample size. Moreover, we observe that the confidence lengths also decrease with an increase in the sample size. We used the nlm() package in R to obtain the parameter estimates. All the analyses are conducted using the statistical package, R Studio (version 2023.06.0).

5. Data analysis: Conductors' failure data

This section is devoted to the model comparison between the proposed Lehmann II Perk model (LPM) and some other models. The following data set has been used in order to assess the goodness-offit of the considered models. The data present hours to failure of 59 test conductors of 400 micrometer length, reported by Schafft et al. [19]. All specimens ran to failure at a certain high temperature and current density. All 59 specimens were tested under the same temperature and current density. 6.545, 9.289, 7.543, 6.956, 6.492, 5.459, 8.120, 4.706, 8.687, 2.997, 8.591, 6.129, 11.038, 5.381, 6.958, 4.288, 6.522, 4.137, 7.459, 7.495, 6.573, 6.538, 5.589, 6.087, 5.807, 6.725, 8.532, 9.663, 6.369, 7.024, 8.336, 9.218, 7.945, 6.869, 6.352, 4.700, 6.948, 9.254, 5.009, 7.489, 7.398, 6.033, 10.092, 7.496, 4.531, 7.974, 8.799, 7.683, 7.224, 7.365, 6.923, 5.640, 5.434, 7.937, 6.515, 6.476, 6.071, 10.491, 5.923. We comapre the developed distribution with the following distributions:

1. The Perks distribution (PD) proposed by Perks [15], with density function

$$f_1(x) = \alpha\lambda e^{\lambda x} \frac{(1+\alpha)}{(1+\alpha e^{\lambda x})^2} \tag{17}$$

   where $x, \alpha, \lambda > 0$ and $f_1(x) = 0$ otherwise.

2. Exponentiated Perks distribution (EPD) introduced by Sigh B. and Choudhary N. [21], with density function

Table 2: *Goodness of fit for Conductors' failure data.*

Model	α	β	λ	K-S statistic	p value	AIC
LPM	**0.00028**	**1.2427**	**0.9341**	**0.0507**	**0.9962**	**228.74**
PD	0.0052	-	0.7576	0.1134	0.4043	237.20
EPD	0.0029	2.10	0.9990	0.0673	0.9356	230.03
EED	35.0200	-	0.5797	0.1045	0.5065	235.08
EWD	2.72400	2:92	0.1655	0.0648	0.9518	228.59
GRD	6.4000	-	0.2200	0.0719	0.8990	227.74

$$f_2(x) = \alpha\beta\lambda e^{\lambda x}\frac{(1+\alpha)}{(1+\alpha e^{\lambda x})^2}\left[1-\left(\frac{1+\alpha}{1+\alpha e^{\lambda x}}\right)\right]^{\beta-1} \tag{18}$$

where $x, \alpha, \beta, \lambda > 0$ and $f_2(x) = 0$ otherwise.

3. Exponentiated exponential distribution (EED) introduced by Gupta et.al.[8], with density function

$$f_3(x) = \alpha\lambda\left(1-e^{-\lambda x}\right)^{\alpha-1}e^{-\lambda x} \tag{19}$$

where $x, \alpha, \lambda > 0$ and $f_3(x) = 0$ otherwise.

4. Exponentiated Weibull distribution (EWD) introduced by Mudholkar and Srivastava [11], with density function

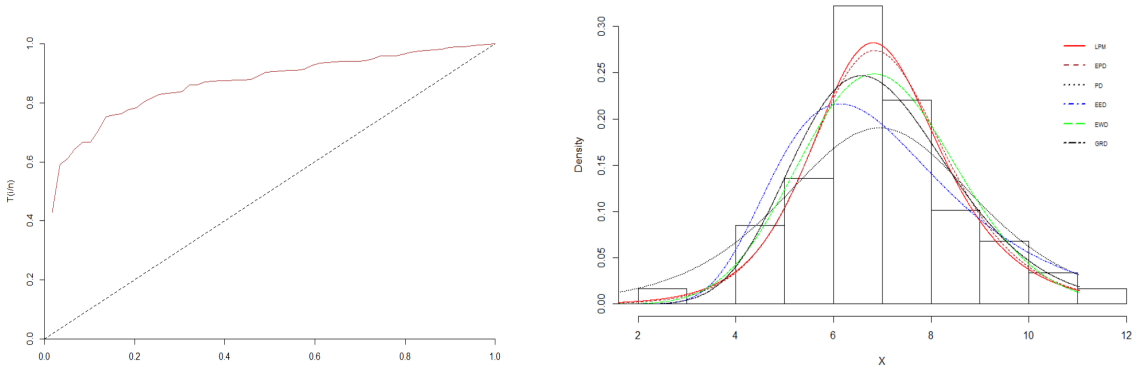
$$f_4(x) = \alpha\beta\lambda^{\beta}x^{\beta-1}\left[1-exp(-\lambda x)^{\beta}\right]^{\alpha-1}exp(-\lambda x)^{\beta} \tag{20}$$

where $x, \alpha, \beta, \lambda > 0$ and $f_4(x) = 0$ otherwise.

5. Generalized Rayleigh distribution (GRD) introduced by surles and padgett [22], with density function

$$f_5(x) = 2\alpha\lambda^2 x e^{-(\lambda x)^2}\left(1-e^{-(\lambda x)^2}\right)^{\alpha-1} \tag{21}$$

where $x, \alpha, \lambda > 0$ and $f_5(x) = 0$ otherwise.



(a) *Total time test plot for conductors failure data.* **(b)** *Histogram and fit for Conductors' failure data*

Figure 5: *Comparison of TTT plot and histogram for conductors' failure data.*

The total time on the test plot (TTT) of the data shown in Figure 5a suggests an increasing hazard rate. Now, we demonstrate the fit of LPM in comparison to other considered distributions for the above data sets. For each distribution, the unknown parameters are estimated by the method of maximum likelihood. Table 2 presents the parameter estimates, Kolmogorov-Smirnov statistic, p-value, and the Akaike information criterion value for the models. The best model is the one with the minimum value for the Akaike information criterion and the Kolmogorov-Smirnov statistic. It is observed that the developed distribution provides the best fit with a Kolmogorov-Smirnov statistic of 0.0507 and a p-value of 0.9962. Additionally, the Akaike information criterion value is the lowest for the developed distribution when compared to the models discussed in Singh B and Chaudhary N. [21] (Table 2). This explains the flexibility of the LPM distribution. Figure 5b displays the histogram and the fit plot for the Conductors' failure data.

6. CONCLUSION

We have studied the properties of the Lehmann II Perk model. Using maximum likelihood method we computed the estimates. To assess the performance of the estimates are evaluated through Monte-Carlo simulation. We have compared the fit of the developed distribution with other models in the literature for the conductor's failure data and found it to be better. Furthermore, for future research, we can consider studying more properties of this distribution and explore different methods of estimation, such as the Bayesian method.

References

[1] A. A. Ogunde, G. A. Olalude, O. E. Adeniji1, K. Balogun . Lehmann type II Frechet Poisson distribution: Properties, inference and applications as a life time distribution, International journal of statistics and probability, (2021).
[2] A.A. Ogunde , S. T. Fayose , B. Ajayi, D. O. Omosigho. Extended gumbel type-2 distribution: properties and applications, Journal of applied Mathematics, (2020) .
[3] Chaudhary A. K.. A study of Perks-II distribution via Bayesian paradigm, Pravaha journal, (2018).
[4] Chaudhary A. K, Kumar and V.. A Bayesian analysis of Perks-I distribution via Markov Chain Monte Carlo simulation, Nepal journal of science and technology, 2013.
[5] Cordeiro, G. M., E., Ortega, E. M. M., and Cunha, D. C. C.. The exponentiated generalized class of distributions, Journal of data science, (2013).
[6] E. L., Lehmann. The power of rank tests, The annals of Mathematical Statistics, (1953).
[7] Gonzalez L. C. M., Picon L. A. R., Olguin J. C. P., Garcia V, Cruz D.L.. The additive Perks distribution and its applications in reliability analysis, Quality technology and quantitative management, (2022).
[8] Gupta RC, Gupta PL, Gupta RD. Modeling failure time data by Lehmann alternatives, Communications Statistics theory methods, (1998).
[9] G. M. Cordeiro and M. De Castro. A new family of generalized distributions, Journal of Statistical computation and simulation, (2011).
[10] Haberman, S., Renshaw, A.. A comparative study of parametric mortality projection models, Insurance: Mathematics and Economics, (2011).
[11] Mudholkar GS, Srivastava DK. Exponentiated Weibull family for analyzing bathtub failure-real data, IEEE transactions on reliability, (1993).
[12] M. Z. Arshad , M. Z. Iqbal , A. Anees , Z. Ahmad and O. S. Balogun. A new bathtub shaped failure rate model: properties and applications to engineering sector, Pakistan journal of Statistics, (2021).
[13] N. I. Badmus , T. A. Bamiduro, S. G. Ogunobi. Lehmann Type II weighted Weibull distribution, International journal of Physical sciences, (2014) .

[14] O. S. Balogun , M. Z. Iqbal , M. Z. Arshad , A. Z. Afify, P. E. Oguntunde. A new generalization of Lehmann type-II distribution: Theory, simulation, and applications to survival and failure rate data, Scientific African, (2021).

[15] Perks W.. On some experiments in the graduation of mortality statistics, Journal of the institute of actuaries, (1932).

[16] P. O. Awodutire , E. C. Nduka, M. A. Ijomah. Lehmann type II generalized half logistic distribution: properties and application, Mathematical theory and modeling, (2020).

[17] Richards. Applying survival models to pensioner mortality data, British actuarial journal, (2008).

[18] Richards, S. J.. A handbook of parametric survival models for actuarial use, Scandinavian actuarial journal, (2012).

[19] Schafft HA, Staton TC, Mandel J, Shott JD, Reproducibility of electro migration measurements. IEEE trans electron devices ED, (1989).

[20] Shaked, M., Shanthikumar, J. G.. Stochastic orders, Springer Science and business Media, (2007).

[21] Singh B, Chaudhary N.. The exponentiated Perks distribution, International journal of systems assurance engineering and management, (2016).

[22] Surles JG, Padgett WJ. Inference for reliability and stressstrength for a scaled Burr type X distribution, Lifetime data analysis, (2001).

[23] V. L. D. Tomazella , P. L. Ramos , P. H. Ferreira , A. L. Mota, F. Louzada. The Lehmann type II inverse Weibull distribution in the presence of censored data, Communications in Statistics - simulation and computation, (2020) .

[24] Zeng, H., Lan, T., Chen, Q.. Five and four-parameter lifetime distributions for bathtub-shaped failure rate using Perks mortality equation, Reliability engineering and system safety, (2016).

RT&A, No 4(80)
Volume 19, December, 2024

STOCHASTIC BEHAVIOUR OF AN ELECTRONIC SYSTEM SUBJECT TO MACHINE AND OPERATOR FAILURE

S. Malik[1], Komal[2], R. K. Yadav[*], Anju[3]

•

[1,2,3]Department of Mathematics, Baba Mastnath University, Rohtak, India
[*]School of Sciences, Christ University, Bengaluru, India

[1]sangeetastat@gmail.com, [2]komalpgt@gmail.com, [*]yadav.ramesh546@gmail.com,
[3]vermaanju198@yahoo.in
[*]Corresponding Author

Abstract

A stochastic model is developed by assuming the human (operator) redundancy in cold standby. For constructing this model, one unit is taken as electronic system which consists of hardware and software components and another unit is operator (human being). The system can be failed due to hardware failure, software failure and human failure. The failed hardware component goes under repair immediately and software goes for upgradation. The operator is subjected to failure during the manual operation. There are two separate service facilities in which one repairs/upgrades the hardware/software component of the electronic system and other gives the treatment to operator. The failure rates of components and operator are considered as constant. The repair rates of hardware/software components and human treatment rate follow arbitrary distributions with different pdfs. The state transition diagram and transition probabilities of the model are constructed by using the concepts of semi-Markov process (SMP) and regenerative point technique (RPT). These same concepts have been used for deriving the expressions (in steady state) for reliability measures or indices. The behavior of some important measures has been shown graphically by taking the particular values of the parameters.

Keywords: Electronic System, Operator Failure, Reliability Measures, Stochastic Analysis, semi-Markov Process, Regenerative Point Technique

I. Introduction

The electronic systems containing h/w and s/w components are playing an impactful role in transforming modern society into a digitalization world. All types of modern industries have become dependent on these systems to furnish various jobs with higher accuracy timely. With electronic systems, the digital world can't ignore the impactful role of manpower in doing work. As these systems require hardware and software, the failure of these components during operation is undeniable. Due to these failures, the job cannot be completed in time or the losses can occur in terms of finances, development lives etc. Therefore, the reliability of these systems becomes very important for the completion of jobs.

Many techniques related to reliability improvement have been described by many researchers. The cold standby redundancy techniques have also been included by the engineers. In the modern society, many research works on reliability modelling of redundant systems have been done to do job accurately. In [14], Sridharan and Mohanavadivu checked the stochastic behavior of two-unit standby system having regular and expert repairmen. The aging properties of the residual life length of k-out-of-n system with independent non-identical components has been discussed in [5]. The idea of switch failure and equipment maintenance has been used in reliability analysis of a two-unit cold standby system [12]. The same type of system has been described in [6] with consideration of human error failure. In [13], Salah and Sherbeny assumed different types of failures in analyzing two non-identical unit system stochastically. The concept of Markov process has been used in reliability analysis of system [11].

Many reliability models for non-identical units have been discussed by engineers and researchers. In [7], Malik and Upma described the non-identical units under preventive maintenance. The reliability measures of two dissimilar units using Gumbel-Hougaard Family Copula have been determined in [1]. The provision of rest and switching device for two non-identical unit standby system have been elaborated [3]. In [2], Kadyan et al. studied non-identical repairable system with working of simultaneous cod standby units. The failure of repairman for system of two-non identical units has been considered by Kumar and Nandal [4] in their research work. During h/w repair, assumption of server failure has been considered in stochastic analysis of computer system by Malik and Yadav [8]. In [9], [10] and [15], the authors described the stochastic analysis of a computer system subject to failure of service facility. However, the human being (operator) has not been considered in redundancy in the above-mentioned research by discussing the reliability modelling of electronic system.

Thus, in this paper, the authors tried the use of manpower (operator) in reliability modelling of electronic system. Here, one unit is taken as electronic system which consists of hardware and software components and another unit is operator (human being). The system can be failed due to hardware failure, software failure and human failure. The failed hardware component goes under repair immediately and software goes for upgradation. The operator is subjected to failure during the manual operation. There are two separate service facilities in which one repairs/upgrades the hardware/software component of the electronic system and other gives the treatment to operator. The failure rates of components and operator are considered as constant. The repair rates of hardware/software components and human treatment rate follow arbitrary distributions with different pdfs. The state transition diagram and transition probabilities of the model are constructed by using the concepts of semi-Markov process (SMP) and regenerative point technique (RPT). These same concepts have been used for deriving the expressions (in steady state) for reliability measures or indices. The behavior of some important measures has been shown graphically by taking the particular values of the parameters.

II. Abbreviations and Notations

MTSF	Mean Time to System Failure
SMP	Semi-Markov Process
RPT	Regenerative Point Technique
MST	Mean Sojourn Time
a/b	Probability of hardware/software failure
$x_1/x_2/\mu$	Hardware/software/ operator failure rates
$\alpha/\beta/\gamma$	Hardware repair/software up-gradation/operator treatment rates
f(t)/F(t)	pdf/cdf of hardware repair time

g(t)/G(t)	pdf/cdf of software repair time
s(t)/S(t)	pdf/cdf of human treatment time
pdf/cdf	Probability density function/Cumulative density function
$q_{ij}(t)/$ $Q_{ij}(t)$	pdf/cdf of passage time from regenerative state S_i to a regenerative state S_j or to a failed state S_j without visiting any other regenerative state in $(0,t]$
$q_{ij.kr}(t)/$ $Q_{ij.kr}(t)$	pdf/cdf of direct transition time from regenerative state S_i to a regenerative state S_j or to a failed state S_j visiting states S_k and S_r once in $(0,t]$
$p_{ij}/p_{ij.kr}$	Steady state probability of transition from state S_i to state S_j directly/via states S_k and S_r once
μ_i	MST in state S_i which is given by $\mu_i = E(T_i) = \int_0^\infty \mathrm{P}(T_i > t)dt$ where T_i denotes the sojourn time in state S_i.
m_{ij}	Contribution to MST(μ_i) in state S_i when system transits directly to state S_j so that $\mu_i = \sum_j m_{ij}$ and $m_{ij} = \int t dQ_{ij}(t) = -q_{ij}^{*\prime}(0)$
$\emptyset_i(t)$	cdf of first passage time from regenerative state S_i to a failed state
$A_i(t)$	Probability that the system is in up-state at instant 't' given that the system entered in regenerative state S_i at $t = 0$
$M_i(t)$	Probability that the system up initially in regenerative state S_i is up at time t without visiting any other state
$R_i(t)$	Expected number of hardware repairs in the interval $(0,t]$ given that the system entered in regenerative state S_i at $t = 0$.
$U_i(t)$	Expected number of software up-gradations in the interval $(0,t]$ given that the system entered in regenerative state S_i at $t = 0$.
$T_i(t)$	Expected number of treatments given to the human in the interval $(0,t]$ given that the system entered in regenerative state S_i at $t = 0$.
Ⓢ/©	Standard notation for Laplace-Stieltjes convolution/Laplace convolution
*/**	Symbol for Laplace Transform (LT)/Laplace Stieltjes Transform (LST)
P	Profit function of system
Z_0	System revenue per unit up-time
Z_1/Z_2	Repair/up-gradation cost per unit time due to hardware failure/software failure
Z_3	Operator Treatment cost of per unit time

III. Assumptions and State Descriptions

To describe the system the following assumptions are made:

a) There are two units in which one is an electronic system made up of hardware and software and other unit is a human being (operator).
b) The components of electronic system and operator are repaired by separate servers.
c) The h/w repairs, s/w up-gradation and treatments are perfect.
d) The failure rates of components and operator are constant.
e) The arbitrary distributions are taken for h/w repair, s/w up-gradation and operator treatment rates.

The description of all states is given in the Table 1

Table 1: *Description of all states in the system model*

States	Description
S_0	Electronic system is in working and operator is in spare
S_1	h/w component is under repair and operator is working manually
S_2	s/w is under up-gradation and operator is working manually
S_3	h/w component is under repair continuously and operator is under treatment
S_4	s/w is under up-gradation continuously and operator is under treatment
S_5	Electronic system is in working and operator is under treatment continuously
S_6	h/w component is under repair continuously and operator is working manually
S_7	s/w is under up-gradation continuously and operator is working manually
S_8	h/w component is under repair and operator is under treatment continuously
S_9	s/w is under up-gradation continuously and operator is under treatment continuously

All the states are presented in the state transition diagram as shown in Figure 1.

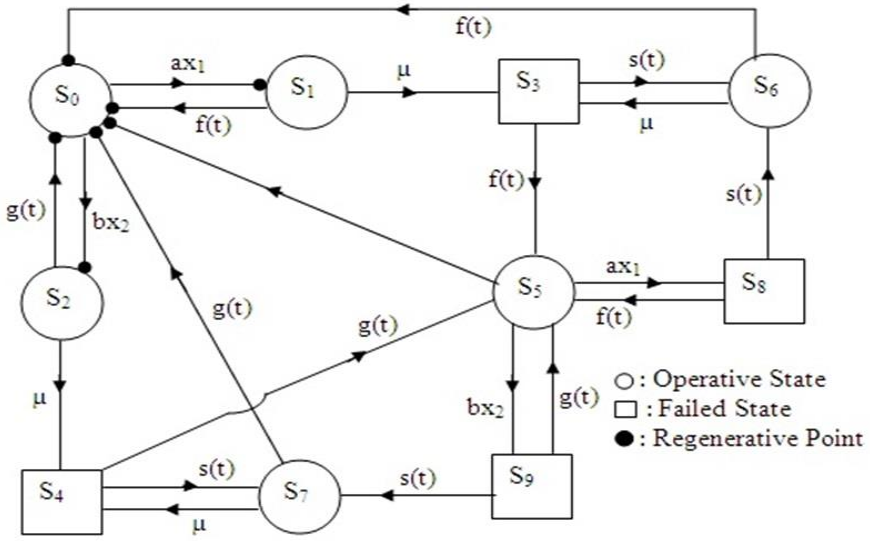


Figure 1: *State transition Diagram of system model*

IV. Reliability Measures

I. Transition Probabilities

The arbitrary distributions are considered as: $f(t) = \alpha e^{-\alpha t}$, $g(t) = \beta e^{-\beta t}$ and $s(t) = \gamma e^{-\gamma t}$.
Using probabilistic arguments, the differential transition probabilities for state S_0 are given by
$dQ_{01}(t) = ax_1 e^{-(ax_1+bx_2)t}dt$, $dQ_{02}(t) = bx_2 e^{-(ax_1+bx_2)t}dt$
The following results: $p_{ij} = \lim_{s\to 0} \emptyset_{ij}^{**}(s) = \emptyset_{ij}^{**}(0) = \int_0^\infty \mathrm{d}Q_{ij}(t) = \int_0^\infty q_{ij}(t)\mathrm{dt}$, have been used to determine the transition probabilities. These are obtained as

$p_{01} = \frac{ax_1}{ax_1+bx_2}$, $p_{02} = \frac{bx_2}{ax_1+bx_2}$, $p_{10} = p_{60} = \frac{\alpha}{\alpha+\mu}$, $p_{13} = p_{63} = \frac{\mu}{\alpha+\mu}$, $p_{20} = p_{70} = \frac{\beta}{\beta+\mu}$,

$p_{24} = p_{74} = \frac{\mu}{\beta+\mu}$, $p_{35} = p_{85} = \frac{\alpha}{\alpha+\gamma}$, $p_{36} = p_{86} = \frac{\gamma}{\alpha+\gamma}$, $p_{45} = p_{95} = \frac{\beta}{\beta+\gamma}$, $p_{47} = p_{97} = \frac{\gamma}{\beta+\gamma}$,

$p_{50} = \frac{\gamma}{ax_1+bx_2+\gamma}$, $p_{58} = \frac{ax_1}{ax_1+bx_2+\gamma}$, $p_{59} = \frac{bx_2}{ax_1+bx_2+\gamma}$, $p_{15.3} = p_{65.3} = p_{13}p_{35}$, $p_{16.3} = p_{66.3} = p_{13}p_{36}$

$p_{25.4} = p_{75.4} = p_{24}p_{45}$, $p_{27.4} = p_{77.4} = p_{24}p_{47}$, $p_{55.8} = p_{58}p_{85}$, $p_{56.8} = p_{58}p_{86}$, $p_{55.9} = p_{59}p_{95}$,
$p_{57.9} = p_{59}p_{97}$

From these, we have
$p_{01} + p_{02} = p_{10} + p_{13} = p_{20} + p_{24} = p_{35} + p_{36} = p_{45} + p_{47} = p_{50} + p_{58} + p_{59} = 1$

$p_{60} + p_{63} = p_{70} + p_{74} = p_{85} + p_{86} = p_{95} + p_{97} = p_{10} + p_{15.3} + p_{16.3} = 1$
$p_{50} + p_{55.8} + p_{56.8} + p_{55.9} + p_{57.9} = p_{60} + p_{65.3} + p_{66.3} = p_{70} + p_{75.4} + p_{77.4} = 1$

II. MST

The MSTs for all states are determined as follows:
$\mu_0 = \frac{1}{ax_1+bx_2}$, $\mu_1 = \frac{1}{\alpha+\mu} = \mu_6$, $\mu_2 = \frac{1}{\beta+\mu} = \mu_7$, $\mu_3 = \frac{1}{\alpha+\gamma} = \mu_8$, $\mu_4 = \frac{1}{\beta+\gamma} = \mu_9$, $\mu_5 = \frac{1}{ax_1+bx_2+\gamma}$,
$\mu_1' = \frac{\alpha+\mu+\gamma}{(\alpha+\mu)(\alpha+\gamma)} = \mu_6'$, $\mu_2' = \frac{\beta+\mu+\gamma}{(\beta+\mu)(\beta+\gamma)} = \mu_7'$, $\mu_5' = \frac{(\beta+\gamma)(\alpha+\gamma+ax_1)+bx_2(\alpha+\gamma)}{(\alpha+\gamma)(ax_1+bx_2+\gamma)(\beta+\gamma)}$

III. Reliability and MTSF

Let $\emptyset_i(t)$ be the c.d.f. of first passage time from regenerative state S_i to a failed state. Regarding the failed state as absorbing state, we have following recursive relations for $\emptyset_i(t)$:
$\emptyset_i(t) = \sum_j Q_{ij}(t)Ⓢ\emptyset_j(t) + \sum_k Q_{ik}(t)$
where S_j is an un-failed regenerative state to which the given regenerative state S_i can transit and S_k is a failed state to which the state S_i can transit directly. Thus, the following equations are obtained as:
$\emptyset_0(t) = Q_{01}(t)Ⓢ\emptyset_1(t) + Q_{02}(t)Ⓢ\emptyset_2(t)$
$\emptyset_1(t) = Q_{10}(t)Ⓢ\emptyset_0(t) + Q_{13}(t)$
$\emptyset_2(t) = Q_{20}(t)Ⓢ\emptyset_0(t) + Q_{24}(t)$
Taking Laplace Stieltjes Transform of above equations, we get
$\emptyset_0^{**}(s) = Q_{01}^{**}(s)\emptyset_1^{**}(s) + Q_{02}^{**}(s)\emptyset_2^{**}(s)$
$\emptyset_1^{**}(s) = Q_{10}^{**}(s)\emptyset_0^{**}(s) + Q_{13}^{**}(s)$
$\emptyset_2^{**}(s) = Q_{20}^{**}(s)\emptyset_0^{**}(s) + Q_{24}^{**}(s)$
Solving for $\emptyset_0^{**}(s)$ by Cramer Rule, we have
$\emptyset_0^{**}(s) = \frac{\Delta_1}{\Delta}$
Where $\Delta = \begin{vmatrix} 1 & -Q_{01}^{**}(s) & -Q_{02}^{**}(s) \\ -Q_{10}^{**}(s) & 1 & 0 \\ -Q_{20}^{**}(s) & 0 & 1 \end{vmatrix}$ and
$\Delta_1 = \begin{vmatrix} 0 & -Q_{01}^{**}(s) & -Q_{02}^{**}(s) \\ Q_{13}^{**}(s) & 1 & 0 \\ Q_{24}^{**}(s) & 0 & 1 \end{vmatrix}$
Now, we have $R^*(s) = \frac{1-\emptyset_0^{**}(s)}{s}$
The reliability of the system model can be obtained by
$$R(t) = L^{-1}[R^*(s)]$$
The MTSF is given by
$MTSF = \lim_{s\to 0} R^*(s) = R^*(0) = \frac{N_1}{D_1}$, where $N_1 = \mu_0 + p_{01}\mu_1 + p_{02}\mu_2$ and
$D_1 = p_{01}p_{13} + p_{02}p_{24}$

IV. Availability

Let $A_i(t)$ be the probability that the system is in up-state at epoch 't' given that the system entered regenerative state S_i at $t = 0$. The recursive relations for $A_i(t)$ are given as
$A_i(t) = M_i(t) + \sum_j q_{ij}^{(n)}(t)©A_j(t)$
where S_j is any successive regenerative state to which the regenerative state S_i can transit through n transitions. Thus, the following equations are obtained as:
$A_0(t) = M_0(t) + q_{01}(t)©A_1(t) + q_{02}(t)©A_2(t)$
$A_1(t) = M_1(t) + q_{10}(t)©A_0(t) + q_{15.3}(t)©A_5(t) + q_{16.3}(t)©A_6(t)$
$A_2(t) = M_2(t) + q_{20}(t)©A_0(t) + q_{25.4}(t)©A_5(t) + q_{27.4}(t)©A_7(t)$
$A_5(t) = M_5(t) + q_{50}(t)©A_0(t) + [q_{55.8}(t) + q_{55.9}(t)]©A_5(t) + q_{56.8}(t)©A_6(t) + q_{57.9}(t)©A_7(t)$

$A_6(t) = M_6(t) + q_{60}(t)©A_0(t) + q_{65.3}(t)©A_5(t) + q_{66.3}(t)©A_6(t)$
$A_7(t) = M_7(t) + q_{70}(t)©A_0(t) + q_{75.4}(t)©A_5(t) + q_{77.4}(t)©A_7(t)$
where $M_0(t) = e^{-(ax_1+bx_2)t}$, $M_1(t) = M_6(t) = e^{-\mu t}\bar{F}(t)$,$M_2(t) = M_7(t) = e^{-\mu t}\bar{G}(t)$, and
$M_5(t) = e^{-(ax_1+bx_2)t}\bar{S}(t)$
Taking LT of above equations and solving for $A_0^*(s)$, the steady state availability is calculated by
$A_0(\infty) = \lim_{s\to 0} sA_0^*(s) = \frac{N_2}{D_2}$
where, $N_2 = p_{01}[p_{57.9}(\mu_1 p_{20} + \mu_2 p_{15.3}) + (1-p_{27.4})\{\mu_1(p_{50}+p_{56.8}) + \mu_5 p_{15.3}\}] + p_{02}[p_{56.8}(\mu_2 p_{10} + \mu_1 p_{25.4}) + (1-p_{16.3})\{\mu_2(p_{50}+p_{57.9}) + \mu_5 p_{25.4}\}] + \mu_0[p_{57.9}p_{20}(1-p_{16.3}) + (1-p_{27.4})\{p_{50}(1-p_{16.3}) + p_{56.8}p_{10}\}]$,

$D_2 = (p_{01}\mu_1' + \mu_0' + p_{02}\mu_2')[p_{57.9}p_{20}(1-p_{16.3}) + (1-p_{27.4})\{p_{50}(1-p_{16.3}) + p_{56.8}p_{10}\}] + \mu_5'\{p_{15.3}p_{01}(1-p_{27.4}) - p_{25.4}p_{02}(1-p_{16.3})\} + \mu_1'[p_{16.3}p_{01}\{p_{57.9}p_{20} + (1-p_{27.4})(p_{50}+p_{58})\} + p_{25.4}p_{02}p_{56.8}] + \mu_2'[p_{27.4}p_{02}\{p_{56.8}p_{10} + (1-p_{16.3})(p_{50}+p_{59})\} + p_{15.3}p_{01}p_{57.9}]$
and $\mu_i = M_i^*(0), i = 0,1,2,5$

V. Expected Number of Hardware Repairs

Let $R_i(t)$ be the expected number of the hardware repairs by the server in the interval $(0,t]$given that the system entered regenerative state S_i at $t = 0$. The expected number of the hardware repairs is given by $R_0(\infty) = \lim_{s\to 0} sR_0^{**}(s)$
The recursive relations for $R_i(t)$ are given as:
$R_i(t) = \sum_j Q_{ij}^{(n)}(t)Ⓢ[\delta_j + R_j(t)]$
Where $S_j$is any regenerative state to which the regenerative state S_i can transit through n transitions and $\delta_j = 1$, if S_j is the regenerative state where server does job afresh, otherwise $\delta_j = 0$. Thus, the following equations are obtained as:
$R_0(t) = Q_{01}(t)ⓈR_1(t) + Q_{02}(t)ⓈR_2(t)$
$R_1(t) = Q_{10}(t)Ⓢ[1 + R_0(t)] + Q_{15.3}(t)Ⓢ[1 + R_5(t)] + Q_{16.3}(t)ⓈR_6(t)$
$R_2(t) = Q_{20}(t)ⓈR_0(t) + Q_{25.4}(t)ⓈR_5(t) + Q_{27.4}(t)ⓈR_7(t)$
$R_5(t) = Q_{50}(t)ⓈR_0(t) + Q_{55.8}(t)Ⓢ[1 + R_5(t)] + Q_{55.9}(t)ⓈR_5(t) + Q_{56.8}(t)ⓈR_6(t) + Q_{57.9}(t)ⓈR_7(t)$
$R_6(t) = Q_{60}(t)Ⓢ[1 + R_0(t)] + Q_{65.3}(t)Ⓢ[1 + R_5(t)] + Q_{66.3}(t)ⓈR_6(t)$
$R_7(t) = Q_{70}(t)ⓈR_0(t) + Q_{75.4}(t)ⓈR_5(t) + Q_{77.4}(t)ⓈR_7(t)$
Taking Laplace Stieltjes Transform of above relation and solving for $R_0^{**}(s)$. The expected number of the hardware repairs is given by
$R_0(\infty) = \lim_{s\to 0} sR_0^{**}(s) = \frac{N_3}{D_2}$
where $N_3 = p_{01}[p_{57.9}p_{20}(1-p_{16.3}) + (1-p_{27.4})\{p_{50}(1-p_{16.3}) + p_{56.8}p_{10}\}] + p_{58}\{p_{15.3}p_{01}(1-p_{27.4}) + p_{25.4}p_{02}(1-p_{16.3})\}$ and D_2 is same as calculated in availability.

VI. Expected Number of Software Up-gradations

Let $U_i(t)$ be the expected number of the software up-gradations by the server in the interval $(0,t]$ given that the system entered regenerative state S_i at $t = 0$. The expected number of the software up-gradations is given by $U_0(\infty) = \lim_{s\to 0} sU_0^{**}(s)$
The recursive relations for $U_i(t)$ are given as:
$U_i(t) = \sum_j Q_{ij}^{(n)}(t)Ⓢ[\delta_j + U_j(t)]$
Where $S_j$is any regenerative state to which the regenerative state S_i can transit through n transitions and $\delta_j = 1$, if S_j is the regenerative state where server does job afresh, otherwise $\delta_j = 0$. Thus, the following equations are obtained as:
$U_0(t) = Q_{01}(t)ⓈU_1(t) + Q_{02}(t)ⓈU_2(t)$
$U_1(t) = Q_{10}(t)ⓈU_0(t) + Q_{15.3}(t)ⓈU_5(t) + Q_{16.3}(t)ⓈU_6(t)$
$U_2(t) = Q_{20}(t)Ⓢ[1 + U_0(t)] + Q_{25.4}(t)Ⓢ[1 + U_5(t)] + Q_{27.4}(t)ⓈU_7(t)$
$U_5(t) = Q_{50}(t)ⓈU_0(t) + Q_{55.8}(t)ⓈU_5(t) + Q_{55.9}(t)Ⓢ[1 + U_5(t)] + Q_{56.8}(t)ⓈU_6(t) + Q_{57.9}(t)ⓈU_7(t)$
$U_6(t) = Q_{60}(t)ⓈT_0(t) + Q_{65.3}(t)ⓈU_5(t) + Q_{66.3}(t)ⓈU_6(t)$

$U_7(t) = Q_{70}(t)\textcircled{S}[1 + U_0(t)] + Q_{75.4}(t)\textcircled{S}[1 + U_5(t)] + Q_{77.4}(t)\textcircled{S}U_7(t)$

Taking Laplace Stieltjes Transform of above relation and solving for $R_0^{**}(s)$. The expected number of the software up-gradationsis given by

$U_0(\infty) = \lim_{s\to 0} sU_0^{**}(s) = \frac{N_4}{D_2}$

where $N_4 = p_{02}[p_{57.9}p_{20}(1 - p_{16.3}) + (1 - p_{27.4})\{p_{50}(1 - p_{16.3}) + p_{56.8}p_{10}\}] + p_{59}\{p_{15.3}p_{01}(1 - p_{27.4}) + p_{25.4}p_{02}(1 - p_{16.3})\}$ and D_2 is same as calculated in availability.

VII. Expected Number of Treatments given to Operator

Let $T_i(t)$ be the expected number of the treatments given to human by the server in the interval $(0, t]$ given that the system entered regenerative state S_i at $t = 0$. The expected number of the treatments given to operator is given by $T_0(\infty) = \lim_{s\to 0} sT_0^{**}(s)$

The recursive relations for $U_i(t)$ are given as:

$T_i(t) = \sum_j Q_{ij}^{(n)}(t)\textcircled{S}[\delta_j + T_j(t)]$

Where $S_j$is any regenerative state to which the regenerative state S_i can transit through n transitions and $\delta_j = 1$, if S_j is the regenerative state where server does job afresh, otherwise $\delta_j = 0''$. Thus, the following equations are obtained as:

$T_0(t) = Q_{01}(t)\textcircled{S}T_1(t) + Q_{02}(t)\textcircled{S}T_2(t)$

$T_1(t) = Q_{10}(t)\textcircled{S}T_0(t) + Q_{15.3}(t)\textcircled{S}T_5(t) + Q_{16.3}(t)\textcircled{S}[1 + T_6(t)]$

$T_2(t) = Q_{20}(t)\textcircled{S}T_0(t) + Q_{25.4}(t)\textcircled{S}T_5(t) + Q_{27.4}(t)\textcircled{S}[1 + T_7(t)]$

$T_5(t) = Q_{50}(t)\textcircled{S}[1 + T_0(t)] + [Q_{55.8}(t) + Q_{55.9}(t)]\textcircled{S}T_5(t) + Q_{56.8}(t)\textcircled{S}[1 + T_6(t)] + Q_{57.9}(t)\textcircled{S}[1 + T_7(t)]$

$T_6(t) = Q_{60}(t)\textcircled{S}T_0(t) + Q_{65.3}(t)\textcircled{S}T_5(t) + Q_{66.3}(t)\textcircled{S}[1 + T_6(t)]$

$T_7(t) = Q_{70}(t)\textcircled{S}T_0(t) + Q_{75.4}(t)\textcircled{S}T_5(t) + Q_{77.4}(t)\textcircled{S}[1 + T_7(t)]$

Taking Laplace Stieltjes Transform of above relation and solving for $R_0^{**}(s)$. The expected number of the treatments given to human is given by

$T_0(\infty) = \lim_{s\to 0} sT_0^{**}(s) = \frac{N_5}{D_2}$

where $N_5 = p_{01}\{p_{57.9}(p_{15.3} + p_{16.3}p_{20}) + p_{16.3}(1 - p_{27.4})(p_{50} + p_{58})\} + p_{02}\{p_{56.8}(p_{24} - p_{27.4}p_{13}) + p_{27.4}(1 - p_{16.3})(p_{50} + p_{59})\}$ and D_2 is same as calculated in availability.

V. Profit Analysis

The profit function in the time t is given by

P(t) = Expected revenue in (0, t] – expected total cost in (0, t]

In steady state, the profit of the system model can be obtained by the following formula:

$P = Z_0A_0(\infty) - Z_1R_0(\infty) - Z_2U_0(\infty) - Z_3T_0(\infty)$

where $Z_0 = ₹\,5000, Z_1 = ₹\,2000, Z_2 = ₹\,1500, Z_3 = ₹\,1000$

VI. Application

The application of the present research work is described in the car washing machines. In modern era, most of the people have at least one vehicle in their home for easy service. Therefore, the Vehicle Washing Shops/Vehicle Service Shops are opened within 2KM circle in most of cities. The vehicle is washed with the help of automatic washing machine as shown in the Figure 2, just standing the vehicle under that machine. Due to short circuit, hardware failure or sudden error in software, the machine can be stopped. A server is facilitated for hardware repair and/or software up-gradation of automatic washing machine. After the failure of the machine, the vehicle is washed by a human. During the washing of vehicle, there is possibility that human can be hurt by any part of vehicle; therefore, another service facility has been given for treatment of human being.

Figure 2: *Automatic Car Washing Machine*

VII. Numerical Illustration

Suppose that in a vehicle washing shop there is an automatic car washing machine and a labor. It is obvious that the vehicle washing machine can fail due to h/w component or s/w component with probabilities 'a' or 'b' respectively. The respective failure rates of h/w, s/w and human being are taken as x_1, x_2 and μ. The repair rates of h/w and s/w are assumed as α and β respectively. The labor undergoes treatment with rate γ.

The reliability measures are determined for arbitrary values of the following parameters:

$x_1 = 0.15, x_2 = 0.003, \mu = 0.002, \alpha = 2, \beta = 3, \gamma = 6, a = 0.6$ and $b = 0.4$

$Z_0 = ₹\,5000, Z_1 = ₹\,2000, Z_2 = ₹\,1500, Z_3 = ₹\,1000.$

The particular values of the transition probabilities are as follows:

$p_{01} = 0.98 = p_{50}, p_{02} = 0.02, p_{10} = 0.99 = p_{20}, p_{13} = 0.01 = p_{24} = p_{58} = p_{59}, \; p_{35} = 0.25$

$p_{36} = 0.75, p_{45} = 0.33, p_{47} = 0.67$

The particular values of the MST's are as follows:

$\mu_0 = 10.96, \mu_1 = 0.5 = \mu_1', \mu_2 = 0.33 = \mu_2', \mu_3 = 0.12, \mu_4 = 0.11, \mu_5 = 0.16, \mu_5' = 0.17$

Thus, MTSF $= 11456.66$, Availability $= 0.999848$ and Profit $= ₹\,4886.86$

VIII. Graphical Study of Reliability Measures

Some important reliability measures such as MTSF, availability and profit have been studied w.r.t h/w failure rate. The graph of MTSF vs h/w failure rate has been shown in the Figure 3. The behavior of availability vs h/w failure rate has been presented in Figure 4. In the similar way, profit analysis has been shown in Figure 5.

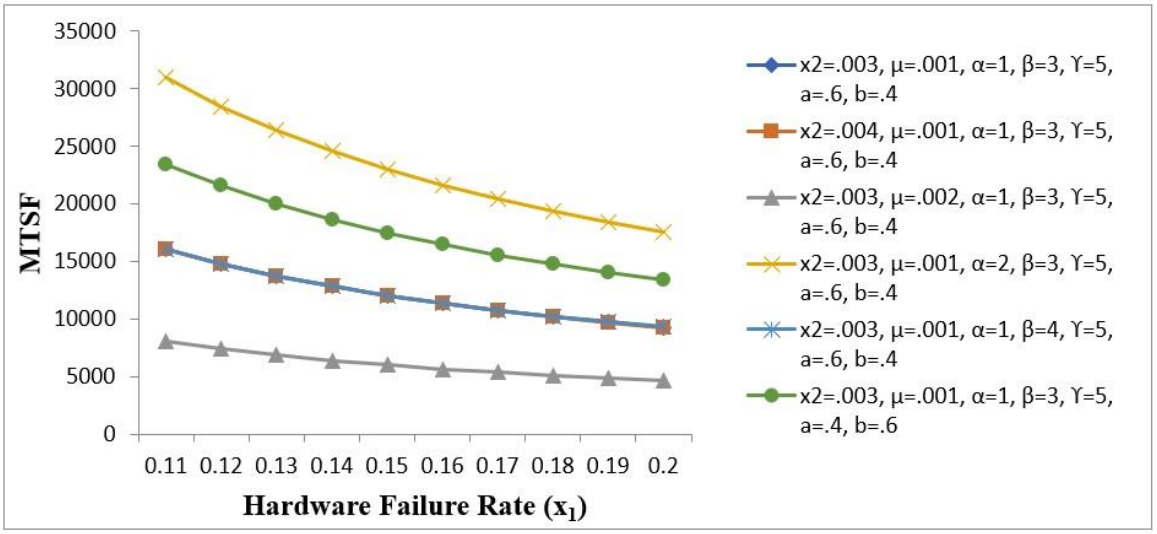


Figure 3: *MTSF Vs Hardware Failure Rate (x_1)*

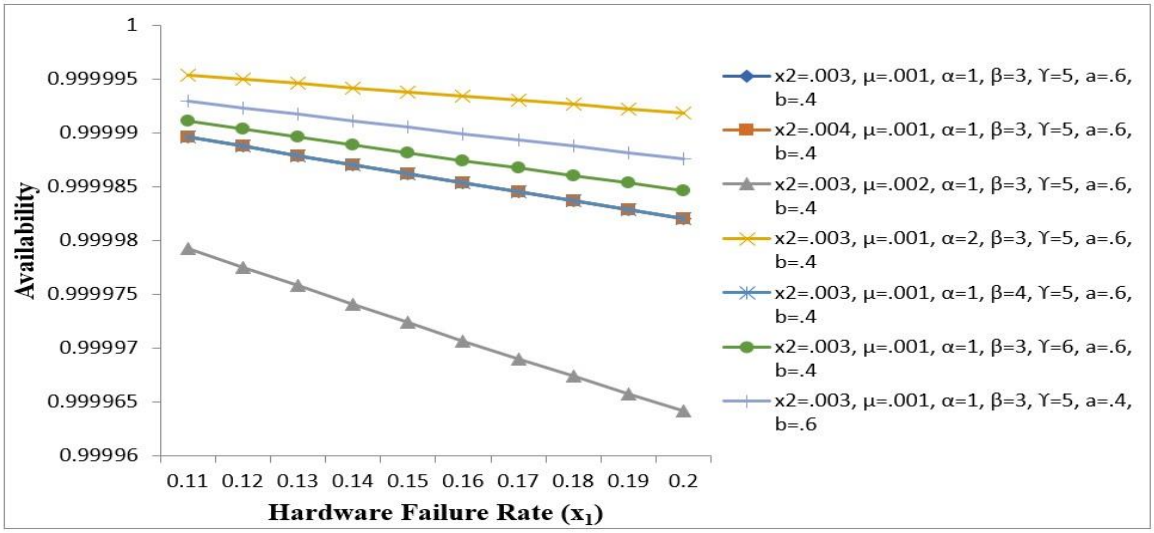


Figure 4: *Availability Vs Hardware Failure Rate (x_1)*

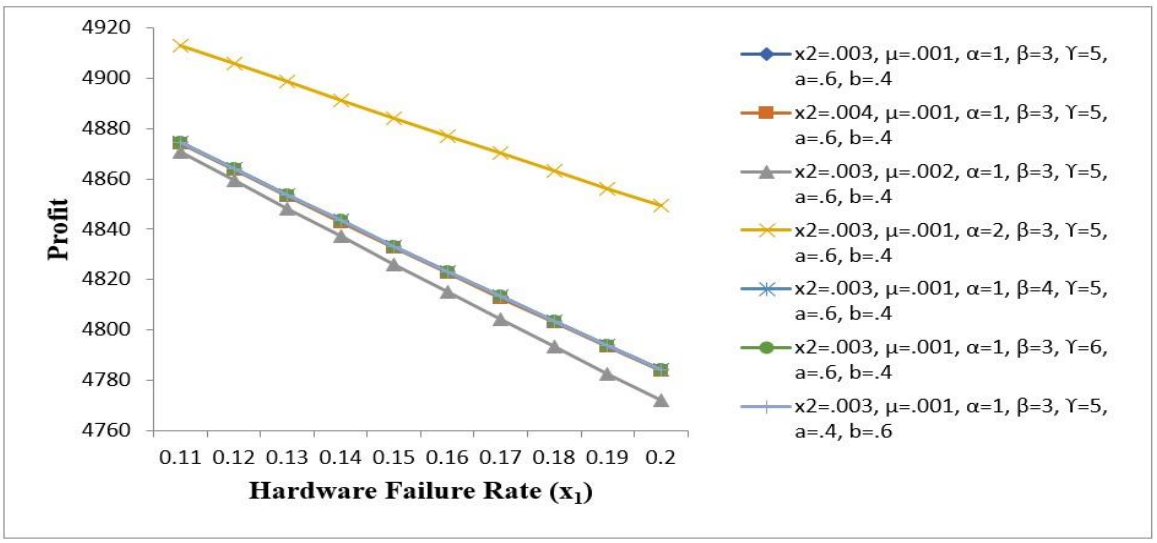


Figure 5: *Profit Vs Hardware Failure Rate (x_1)*

IX. Conclusion

A reliability model of an electronic system with operator failure has been analyzed in present study. The use of SMP and RPT has been incorporated in determining the transition probabilities and various reliability measures. These reliability measures have been explored graphically for the different values of parameters. Figure 3 concludes that MTSF decline according to inclined behavior of components failures and operator failure and MTSF increases with increments in component repair rates and treatment rate. Also, MTSF is very high when hardware repair rate increases from μ=1 to μ=2. Availability and Profit function shows the approximate same behavior as MTSF shows. This nature can be seen if the Figure 4 and Figure 5. There are various future scopes of the present study so that we can make model impactful. The inspection policy for machine can be considered before going under hardware repair as well as software upgradation. The replacement of hardware/software can be done if hardware is not repairable and software is not working. The concept of power failure and timing of power restoration can be considered during the stochastic model. The idea of arbitrary distributions of failure rates of components can be considered.

References

[1] Chopra, G., and Ram, M. (2019). Reliability measures of two dissimilar units parallel system using gumbel-hougaard family copula. International *Journal of Mathematical, Engineering and Management Sciences*, 4(1): 116-130.

[2] Kadyan, S., Malik, S. C. and Gitanjali (2020). Stochastic analysis of a three-unit non-identical repairable system with simultaneous working of cold standby units. *Journal of Reliability and Statistical Studies*, 13(2): 385–400.

[3] Kaur, D., Joorel, J.P.S. and Sharma, N. (2019). Effectiveness analysis of a two non-identical unit standby system with switching device and proviso of rest. *International Journal of Mathematical, Engineering and Management Sciences*, 4(6): 1496-1507.

[4] Kumar, N and Nandal, N. (2020). Stochastic modeling of a system of two non-identical units with priority for operation and repair to main unit subject to conditional failure of repairman. *International Journal of Statistics and Reliability Engineering*, 7(1): 114-122.

[5] Li, X. and Chen, J. (2004). Aging properties of the residual life length of k-out-of-n system with independent but non-identical components. *Applied Stochastic Models in Business and Industry*, 20(2): 143-153.

[6] Mahmaud, M.A.W. and Moshref, M.E. (2010). On a two-unit cold standby system considering hardware, human error failures and preventive maintenance. *Mathematical and Computer Modelling*, 51(5-6): 736-745.

[7] Malik S.C. and Upma (2016). Cost benefit analysis of system of non- identical units under preventive maintenance and replacement. *Journal of Reliability and Statistical Studies*, 9(2): 17-27.

[8] Malik S.C. and Yadav, R.K. (2021). Stochastic analysis of a computer system with unit wise cold standby redundancy and priority to hardware repair subject to failure of service facility. *International Journal of Reliability, Quality and Safety Engineering*, 28(2): 2150013.

[9] Malik, S.C. and Yadav, R.K. (2020). Stochastic analysis of a computer system with unit wise redundancy in cold standby and failure of service facility. *International Journal of Agricultural and Statistical Sciences*, 16(2): 797-806.

[10] Malik, S.C., Yadav, R.K. and Nandal, N. (2023). Reliability Indices of a Computer System with Software Up-gradation Priority and Failure of Service Facility. *Computational Intelligence in Sustainable Reliability Engineering*, Wiley, Editors: S. C. Malik, Deepak Sinwar, Ashish Kumar, S. R. Gadde, Prasenjit Chatterjee and Bui Thanh Hung, Ch. 1: 1-22.

[11] Manglik, M. and Ram, M. (2013). Reliability analysis of a two-unit cold standby system using Markov process. *Journal of Reliability and Statistical Studies*, 6(2), 65-80.

[12] Meng, X-Y, Yuan, L. and Yin, R. (2006). The reliability analysis of a two-unit cold standby system with failable switch and maintenance equipment. International Conference on Computational Intelligence and Security, 2, 941– 944.

[13] Salah and Sherbeny, E.L. (2013). Stochastic analysis of a two non-identical unit parallel system with different types of failures subject to preventive maintenance and repairs. Mathematical Problems in Engineering, 1-10.

[14] Sridharan, V. and Mohanavadivu, P. (1998). Stochastic behavior of two-unit standby system with two types of repairmen and patience time. Mathematical and Computer Modelling, 28(9), 63-71.

[15] Yadav, R.K., Nandal, N., and Malik, S.C. (2022). Stochastic analysis of a cold standby computer system with up-gradation priority and failure of service facility. Reliability: Theory and Applications, 17(4), 132-142.

RT&A, No 4(80)
Volume 19, December, 2024

Unbiased Exponential Type Estimators of Population Mean Using Auxiliary Variable as an Attribute In Double Sampling

Sajad Hussain*

•

Department of Mathematics, School of Engineering (SOE),
Presidency University, Bengaluru, Karnataka, India-560064
sajad.stat321@gmail.com

Abstract

In this paper, unbiased ratio-cum-product exponential type estimators for estimating the population mean have been introduced, specifically within the framework of a double sampling plan. The large sample properties of these estimators are investigated by deriving their bias and mean square error (MSE) expressions. The findings indicate that, under optimal conditions, the proposed estimators are not only unbiased but also more efficient than traditional methods, including the sample mean and the double sampling ratio and product type estimators developed by Naik and Gupta [11] and Singh et al. [17]. To further substantiate the theoretical results, we conducted a numerical study, which demonstrates the practical effectiveness of the proposed estimators in improving estimation accuracy.

Keywords: Exponential Estimator, Auxiliary Attribute, Proposed Estimator, Population, Optimum Value.

1. Introduction

In sampling theory, the precision of an estimate can be enhanced by incorporating information from an auxiliary variable, particularly when this variable is highly correlated with the study variable. The method of utilizing this auxiliary information depends on its form. For quantitative auxiliary information, estimators like Cochran's ratio estimator [4], Robson's product type estimator [12], Bahl and Tuteja's exponential type estimators [2], and other estimators developed by AL-Omari and Bouza [1], Shalabh and Tsai [15], Singh and Vishwakarma [16], Mehta et al. [10], and Hussain et al. [5, 6, 7] are commonly employed.

However, in many practical scenarios, the auxiliary information is qualitative in nature-meaning the auxiliary variable is an attribute correlated with the study variable. For example, Jhajj et al. [8] and Shabir & Gupta [17] illustrated cases such as (a) the height of an individual being dependent on their sex, and (b) crop yield depending on the variety of the crop. In such cases, traditional estimators that rely on quantitative auxiliary information are not suitable due to the point biserial correlation between the study and auxiliary variables. To address this, researchers like Naik and Gupta [11], Jhajj et al. [8], Singh et al. [17], Shabir & Gupta [17, 14], and Abd-Elfattah et al. [14] have developed ratio and product type estimators for the population mean that leverage prior knowledge of the parameters of the auxiliary attribute, thereby improving estimation accuracy in situations where the auxiliary information is qualitative. The proposed exponential estimators typically assume that the population proportion P is known beforehand. However, in real-world scenarios, researchers often encounter situations where this population proportion is not known in advance. To address this, the method of double sampling is employed.

In response to this challenge, Naik and Gupta [11] and Singh et al. [17] introduced double sampling ratio and product type estimators. However, these estimators are biased, meaning they can potentially under or overestimate the true population mean, which introduces inaccuracies in the estimation process.

Taking into account the above discussion, we propose unbiased ratio-cum-product exponential type estimators of the population mean, denoted as T_{de1} and T_{de2}, utilizing auxiliary information in the form of an attribute. Theoretical expressions for the Bias and Mean Square Error (MSE) up to the first order approximation have been derived for these proposed estimators. These results are then compared, both theoretically and empirically, with some existing double sampling ratio and product type estimators.

2. Methodology and Existing Estimators

Use the Simple Random Sampling Without Replacement (SRSWOR) procedure to select a sample of size n from a population consisting of N units. In this context, each unit in the population is associated with two variables: Y_i, which represents the value of the study variable for the i^{th} unit, and ϕ_i, which represents the auxiliary attribute for that same unit $(i = 1, 2, \ldots, N)$. The auxiliary attribute ϕ is assumed to be dichotomous, meaning that it takes on only two possible values: 0 or 1. A value of 1 indicates the presence of the attribute, while a value of 0 indicates its absence. This complete dichotomy within the population allows for a clear classification of units into two distinct categories based on whether the attribute ϕ is present or absent. This classification is crucial for the estimation process, as it allows researchers to leverage the auxiliary information to improve the precision of estimates. For instance, if the attribute ϕ is highly correlated with the study variable Y, this information can be used to develop more efficient estimators of the population mean or other parameters of interest. The SRSWOR method ensures that every unit in the population has an equal chance of being selected, and no unit can be selected more than once, maintaining the randomness and representativeness of the sample. This procedure is fundamental in sampling theory, particularly when auxiliary information is available and can be used to refine estimates and reduce potential bias.

Consider $A = \sum_{i=1}^{N} \phi_i$ and $a = \sum_{i=1}^{n} \phi_i$, the number of units possessing the attribute ϕ in the population and sample respectively. Therefore $P = \frac{A}{N}$ and $p = \frac{a}{n}$ is the proportion of units possessing the given attribute ϕ in the population and sample respectively. When the value of P is unknown, the method of double sampling can be applied. This involves allocating a portion of the budget to gather information on an auxiliary variable. First, a large preliminary sample of size η is collected, with p_1 representing the proportion of units possessing the attribute ϕ in this sample. Then, a smaller, second-phase sample of size n is drawn, which is nested within the first-phase sample ($n < \eta$). In this second sample, p represents the proportion of units with the attribute ϕ, and $\bar{y}$ is the mean of the study variable Y. Some formulas that have been used to compute various measures in this case are presented below as

Population Estimates	**Sample Estimates**
$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} X_i$, is the mean of study variable.	: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} x_i$, is the mean of study variable.
$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$, is the mean square of study variable.	: $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, is the mean square of study variable.
$S_\phi^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\phi_i - P)^2$, is the mean square of auxiliary variable.	: $s_\phi^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\phi_i - p)^2$, is the mean square of auxiliary variable.
$S_{y\phi} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i\phi_i - NP\bar{Y})^2$, is the covariance between study variable and auxiliary attribute.	: $s_{y\phi} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i\phi_i - np\bar{y})^2$, is the covariance between study variable and auxiliary attribute.

Further,

$C_y = \frac{S_y}{\bar{Y}}$ and $C_p = \frac{S_\phi}{P}$, is the coefficient of variation of Y and ϕ respectively.

$\rho_{pb} = \frac{S_{y\phi}}{S_y S_\phi}$, is the correlation between Y and ϕ .

$\hat{\beta} = \frac{s_{y\phi}}{s_{\phi}^2}$, is the sample regression coefficient.
$\gamma = \frac{1-f}{n}, \gamma_1 = \left(\frac{1}{\eta} - \frac{1}{N}\right), \gamma_2 = \left(\frac{1}{n} - \frac{1}{\eta}\right), \gamma_3 = \gamma + \gamma_1$, where $f = \frac{n}{N}$ is the sampling fraction.
To obtain the Bias and MSE expressions when the auxiliary variable is an attribute in single sampling plan, consider

$$e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \quad \text{and} \quad e_\phi = \frac{p - P}{P}.$$

The expected values of quantities e_0, e_ϕ, e_0^2, e_ϕ^2, and $e_0 e_\phi$ are obtained as

$$E(e_0) = E(e_\phi) = 0, \quad E(e_0^2) = \gamma C_y^2, \quad E(e_\phi^2) = \gamma C_p^2, \quad E(e_0 e_\phi) = \gamma C_{yp}.$$

To obtain the Bias and MSE expressions when the auxiliary variable is an attribute in double sampling plan, consider

$$e_0 = \bar{Y}^{-1}(\bar{y} - \bar{Y}), \quad e_\phi = P^{-1}(p - P), \quad e_\phi^{'} = P^{-1}(p_1 - P).$$

Therefore, the following expected values are obtained as

$$E(e_0) = E(e_\phi) = E(e_\phi^{'}) = 0, \quad E(e_0^2) = \gamma C_y^2, \quad E(e_\phi^2) = \gamma C_p^2.$$
$$E(e_\phi^{'2}) = \gamma_1 C_p^2, \quad E(e_0 e_\phi) = \gamma C_{yp}, \quad E(e_0 e_\phi^{'}) = \gamma_1 C_{yp}, \quad E(e_\phi e_\phi^{'}) = \gamma_1 C_p^2.$$

When auxiliary information is unavailable, the sample mean serves as a reliable estimator for the population mean. The sample mean, denoted by $\bar{y}$, is computed by averaging the observations in a sample and provides a practical approach to estimating the population mean

$$t_1 = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The estimator t_1 is unbiased and its variance is as

$$V(t_1) = \gamma \bar{Y}^2 {C_y}^2. \tag{1}$$

Naik and Gupta [11] were the pioneers in proposing double sampling ratio and product type estimators for the population mean, specifically tailored for scenarios where the auxiliary variable is presented as an attribute. Their groundbreaking work introduced a method to effectively utilize categorical or binary auxiliary data to improve the accuracy of population mean estimates

$$T_{ngr} = \bar{y}\left(\frac{p_1}{p}\right).$$

$$T_{ngp} = \bar{y}\left(\frac{p}{p_1}\right).$$

The Bias and MSE of the estimator T_{ngr} and T_{ngp} is as

$$Bias(T_{ngr}) = \bar{Y}(\gamma - \gamma_1)(C_p^2 - C_{yp}).$$
$$MSE(T_{ngr}) = \bar{Y}^2(\gamma {C_y}^2 + \gamma_2 {C_p}^2 - 2\gamma_2 C_{yp}). \tag{2}$$
$$Bias(T_{ngp}) = \bar{Y}(\gamma - \gamma_1) C_{yp}.$$
$$MSE(T_{ngp}) = \bar{Y}^2(\gamma {C_y}^2 + \gamma_2 {C_p}^2 + 2\gamma_2 C_{yp}). \tag{3}$$

Later, Singh et al. [17] introduced a double sampling exponential ratio estimator for the population mean, tailored for situations where the auxiliary information is given as an attribute. This approach enhances estimation accuracy by effectively incorporating categorical or binary auxiliary data

$$T_{sgr} = \bar{y} exp\left(\frac{p_1 - p}{p_1 + p}\right).$$

$$T_{sgp} = \bar{y} exp\left(\frac{p - p_1}{p_1 + p}\right).$$

The Bias and MSE of estimators T_{sgr} and T_{sgp} is as

$$Bias(T_{sgr}) = \frac{1}{2}\gamma_2\bar{Y}\Big(\frac{1}{2}C_p^2 - C_{yp}\Big).$$
$$MSE(T_{sgr}) = \bar{Y}^2\left(\gamma {C_y}^2 + \frac{1}{4}\gamma_2 {C_p}^2 - \gamma_2 C_{yp}\right). \tag{4}$$
$$Bias(T_{sgp}) = \frac{1}{2}\gamma_2\bar{Y}\Big(\frac{1}{2}C_p^2 + C_{yp}\Big).$$
$$MSE(T_{sgp}) = \bar{Y}^2\left(\gamma {C_y}^2 + \frac{1}{4}\gamma_2 {C_p}^2 + \gamma_2 C_{yp}\right). \tag{5}$$

3. Proposed Estimators in Double Sampling

The proposed proportion based unbiased exponential ratio-cum-product estimators in double sampling plan are as

$$T_{rp1} = \bar{y}\left[\omega_1 \exp\left(\frac{p_1 - p}{lp_1}\right) + (1-\omega_1)\exp\left(\frac{p - p_1}{lp_1}\right)\right]. \tag{6}$$

$$T_{rp2} = \bar{y}\left[\omega_2 \exp\left(\frac{p_1 - p}{gp}\right) + (1-\omega_2)\exp\left(\frac{p - p_1}{gp}\right)\right]. \tag{7}$$

Where $l(\neq 0)$, $g(\neq 0)$, ω_1 and ω_2 are constants. The values of l and g are chosen such that the proposed estimators are unbiased and the values of ω_1 & ω_2 are chosen such that MSE (T_{rpi}) $(i = 1, 2)$ is minimum.

Theorem 4.1: The Bias and MSE expressions for the double sampling ratio-cum-product type estimators T_{rp1} and T_{rp2} to the first order of approximation are as

$$Bias(T_{rp1}) = \frac{\gamma_2\bar{Y}}{l}\left[C_{yp} - 2\omega_1 C_{yp} + \frac{C_p^2}{2l}\right]. \tag{8}$$

$$Bias(T_{rp2}) = \frac{\gamma_2\bar{Y}}{g}\left[(2\omega_2 - 1)(C_p^2 - C_{yp}) - \frac{C_p^2}{2g}\right]. \tag{9}$$

and

$$MSE(T_{rp1}) = \bar{Y}^2\left[\gamma C_y^2 + \frac{(2\omega_1 - 1)^2}{l^2}\gamma_2 C_p^2 - \frac{2(2\omega_1 - 1)}{l}\gamma_2 C_{yp}\right]. \tag{10}$$

$$MSE(T_{rp2}) = \bar{Y}^2\left[\gamma C_y^2 + \frac{(2\omega_2 - 1)^2}{g^2}\gamma_2 C_p^2 - \frac{2(2\omega_2 - 1)}{g}\gamma_2 C_{yp}\right]. \tag{11}$$

Proof: Writing the estimator T_{rp1} and T_{rp2} i.e. (6) and (7) in terms of $e^{'s}$ the expression obtaimed are as

$$\begin{aligned} T_{rp1} &= \bar{Y}(1+e_0)\left[\omega_1 \exp\left(\frac{P(1+e_\phi^{'}) - P(1+e_\phi)}{lP(1+e_\phi^{'})}\right)\right. \\ &\qquad \left. +(1-\omega_1)\exp\left(\frac{P(1+e_\phi) - P(1+e_\phi^{'})}{lP(1+e_\phi^{'})}\right)\right] \\ \Rightarrow T_{rp1} &= \bar{Y}(1+e_0)\left[\omega_1 \exp\left(\frac{e_\phi^{'} - e_\phi^{'2} - e_\phi + e_\phi e_\phi^{'} + ...}{l}\right)\right. \\ &\qquad \left. +(1-\omega_1)\exp\left(\frac{e_\phi + e_\phi^{'2} - e_\phi^{'} - e_\phi e_\phi^{'} + ...}{l}\right)\right] \end{aligned} \tag{12}$$

$$T_{rp2} = \bar{Y}(1+e_0)\left[\omega_2 \exp\left(\frac{P(1+e_\phi^{'}) - P(1+e_\phi)}{gP(1+e_\phi)}\right) + (1-\omega_2)\exp\left(\frac{P(1+e_\phi) - P(1+e_\phi^{'})}{gP(1+e_\phi)}\right)\right]$$

$$\Rightarrow T_{rp2} = \bar{Y}(1+e_0)\left[\omega_2 \exp\left(\frac{e_\phi^{'} + e_\phi^2 - e_\phi - e_\phi e_\phi^{'} + ...}{g}\right) + (1-\omega_2)\exp\left(\frac{e_\phi - e_\phi^2 - e_\phi^{'} + e_\phi e_\phi^{'} + ...}{g}\right)\right] \tag{13}$$

Solving (12) & (13) and retaining the terms up to second degree only, we have

$$T_{rp1} = \bar{Y}(1+e_0)\left[1 + \frac{2\omega_1(e_\phi^{'} - e_\phi^{'2} - e_\phi + e_\phi e_\phi^{'})}{l} + \frac{e_\phi^2 + e_\phi^{'2} - 2e_\phi^{'}e_\phi}{2l^2} + \frac{e_\phi + e_\phi^{'2} - e_\phi^{'} - e_\phi e_\phi^{'}}{l}\right] \tag{14}$$

$$T_{rp2} = \bar{Y}(1+e_0)\left[1 + \frac{2\omega_2(e_\phi^{'} + e_\phi^2 - e_\phi - e_\phi e_\phi^{'})}{g} + \frac{e_\phi^2 + e_\phi^{'2} - 2e_\phi^{'}e_\phi}{2g^2} + \frac{e_\phi - e_\phi^2 - e_\phi^{'} + e_\phi e_\phi^{'}}{g}\right] \tag{15}$$

Further solving (14) & (15) and excluding the terms of degree higher than two, we have

$$T_{rp1} - \bar{Y} = \left[e_0 + \frac{2\omega_1(e_\phi^{'} - e_\phi^{'2} - e_\phi + e_\phi e_\phi^{'} + e_\phi^{'}e_0 - e_0 e_\phi)}{l} + \frac{e_\phi^2 + e_\phi^{'2} - 2e_\phi^{'}e_\phi}{2l^2} + \frac{e_\phi + e_\phi^{'2} - e_\phi^{'} - e_\phi e_\phi^{'} + e_0 e_\phi - e_0 e_\phi^{'}}{l}\right] \tag{16}$$

$$T_{rp2} - \bar{Y} = \left[e_0 + \frac{2\omega_2(e_\phi^{'} + e_\phi^2 - e_\phi - e_\phi e_\phi^{'} + e_\phi^{'}e_0 - e_0 e_\phi)}{g} + \frac{e_\phi^2 + e_\phi^{'2} - 2e_\phi^{'}e_\phi}{2g^2} + \frac{e_\phi - e_\phi^2 - e_\phi^{'} + e_\phi e_\phi^{'} + e_0 e_\phi - e_0 e_\phi^{'}}{g}\right] \tag{17}$$

Taking expectation on both sides of equations (16) and (17), the Bias of the estimator T_{rp1} and T_{rp2} is obtained as

$$Bias(T_{rp1}) = \frac{\gamma_2 \bar{Y}}{l}\left[C_{yp} - 2\omega_1 C_{yp} + \frac{C_p^2}{2l}\right]. \tag{18}$$

$$Bias(T_{rp2}) = \frac{\gamma_2 \bar{Y}}{g}\left[(2\omega_2 - 1)(C_p^2 - C_{yp}) - \frac{C_p^2}{2g}\right]. \tag{19}$$

Squaring equations (16) and (17) on both sides and then taking expectation, the mean square error of the proposed estimators is obtained as

$$MSE(T_{rp1}) = \bar{Y}^2\left[\gamma C_y^2 + \frac{(2\omega_1 - 1)^2}{\omega_1^2}\gamma_2 C_p^2 - \frac{2(2\omega_1 - 1)}{l}\gamma_2 C_{yp}\right]. \tag{20}$$

$$MSE(T_{rp2}) = \bar{Y}^2\left[\gamma C_y^2 + \frac{(2\omega_2 - 1)^2}{g^2}\gamma_2 C_p^2 - \frac{2(2\omega_2 - 1)}{g}\gamma_2 C_{yp}\right]. \tag{21}$$

3.0.1 Unbiased Condition

The double sampling exponential estimators T_{rp1} and T_{rp2} are found unbiased, if

$$l = \frac{C_p}{2(2\omega_1 - 1)\rho_{pb}C_y} \quad \text{and} \quad g = \frac{C_p}{2(2\omega_2 - 1)(\rho_{pb}C_y - C_p)} \text{ respectively.}$$

Using the values of l and g in equations (20) and (21) respectively, the expressions obtained are as follows

$$MSE(T_{rp1}) = \bar{Y}^2 \left[\gamma C_y^2 + 4\gamma_3(1-2\omega_1)^4\rho_{pb}^2C_y^2 - 4\gamma_3(1-2\omega_1)^2\rho_{pb}^2C_y^2\right]. \tag{22}$$

$$\begin{aligned} MSE(T_{rp2}) = \bar{Y}^2\Big[\gamma C_y^2 + 4\gamma_2(1-2\omega_2)^4(\rho_{pb}C_y - C_p)^2 - \\ 4\gamma_2(1-2\omega_2)^2(\rho_{pb}C_y - C_p)\rho_{pb}C_y\Big]. \end{aligned} \tag{23}$$

3.0.2 Optimum Value of ω_1 and ω_2

Using the differentation procedure for obtaining the optimal value of ω_1 and ω_2 from equations (22) and (23), we have

$$\omega_{1(opt)} = 0.146,\ 0.854. \tag{24}$$

$$\omega_{2(opt)} = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{\rho_{pb}C_y}{2(\rho_{pb}C_y - C_p)}}. \tag{25}$$

Now substituting the values of $\omega_{1(opt)}$ and $\omega_{2(opt)}$ obtained as equations (24) and (25) in equations (22) and (23) respectively, the minimum value of MSE up to O (n^{-1}) for the estimators T_{rpi} $(i = 1, 2)$ is obtained as

$$MSE_{min}(T_{rpi}) = \gamma\bar{Y}^2{C_y}^2(1-\rho_{pb}^2) + \gamma_1\bar{Y}^2\rho_{pb}^2{C_y}^2. \quad ; i = 1, 2. \tag{26}$$

4. Efficiency Comparison

For comparing the MSE of the proposed double sampling estimators with some existing ratio and product type estimators such as mean per unit estimator (t_m), double sampling ratio and product estimator (T_{ngr} & T_{ngp}) of Naik and Gupta [11], double sampling exponential ratio and product estimators (T_{sgr} & T_{sgp}) of Singh *et al.* [17], we first write the expressions of their mean squared error in double sampling up to the first order of approximation

$$Bias(t_m) = 0.$$

$$MSE(t_m) = \gamma\bar{Y}^2{C_y}^2. \tag{27}$$

$$Bias(T_{ngr}) = \bar{Y}(\gamma - \gamma_1)(C_p^2 - C_{yp}).$$

$$MSE(T_{ngr}) = \bar{Y}^2(\gamma{C_y}^2 + \gamma_2{C_p}^2 - 2\gamma_2C_{yp}). \tag{28}$$

$$Bias(T_{ngp}) = \bar{Y}(\gamma - \gamma_1)C_{yp}.$$

$$MSE(T_{ngp}) = \bar{Y}^2(\gamma{C_y}^2 + \gamma_2{C_p}^2 + 2\gamma_2C_{yp}). \tag{29}$$

$$Bias(T_{sgr}) = \frac{1}{2}\gamma_2\bar{Y}\Big(\frac{1}{2}C_p^2 - C_{yp}\Big).$$

$$MSE(T_{sgr}) = \bar{Y}^2\left(\gamma{C_y}^2 + \frac{1}{4}\gamma_2{C_p}^2 - \gamma_2C_{yp}\right). \tag{30}$$

$$Bias(T_{sgp}) = \frac{1}{2}\gamma_2\bar{Y}\Big(\frac{1}{2}C_p^2 + C_{yp}\Big).$$

$$MSE(T_{sgp}) = \bar{Y}^2\left(\gamma {C_y}^2 + \frac{1}{4}\gamma_2 {C_p}^2 + \gamma_2 C_{yp}\right). \tag{31}$$

The efficiency comparison of the proposed double sampling ratio-cum-product exponential estimators T_{rp1} and T_{rp2} with the other existing estimators is done as

From equations (27), (28), (29), (30) and (31), we get the following conditions under which T_{rp1} and T_{rp2} are more efficient than the estimators t_m, T_{ngp}, T_{ngr}, T_{sgp} and T_{sgr} as

$$MSE_{min}(T_{rpi}) < V(t_m)$$
$$\Rightarrow \gamma\bar{Y}^2{C_y}^2(1-\rho_{pb}^2) + \gamma_1\bar{Y}^2\rho_{pb}^2{C_y}^2 < \gamma\bar{Y}^2{C_y}^2, if$$
$$\gamma_2\rho_{pb}^2\bar{Y}^2 > 0. \tag{32}$$
$$MSE_{min}(T_{rpi}) < MSE(T_{ngr})$$
$$\Rightarrow \gamma\bar{Y}^2{C_y}^2(1-\rho_{pb}^2) + \gamma_1\bar{Y}^2\rho_{pb}^2{C_y}^2 < \bar{Y}^2(\gamma{C_y}^2 + \gamma_2{C_p}^2 - 2\gamma_2 C_{yp}), if$$
$$\gamma_2(C_p - \rho_{pb}C_y)^2 > 0. \tag{33}$$
$$MSE_{min}(T_{rpi}) < MSE(T_{ngp})$$
$$\Rightarrow \gamma\bar{Y}^2{C_y}^2(1-\rho_{pb}^2) + \gamma_1\bar{Y}^2\rho_{pb}^2{C_y}^2 < \bar{Y}^2(\gamma{C_y}^2 + \gamma_2{C_p}^2 + 2\gamma_2 C_{yp}), if$$
$$\gamma_2(C_p + \rho_{pb}C_y)^2 > 0. \tag{34}$$
$$MSE_{min}(T_{rpi}) < MSE(T_{sgr})$$
$$\Rightarrow \gamma\bar{Y}^2{C_y}^2(1-\rho_{pb}^2) + \gamma_1\bar{Y}^2\rho_{pb}^2{C_y}^2 < \bar{Y}^2\left(\gamma{C_y}^2 + \frac{1}{4}\gamma_2{C_p}^2 - \gamma_2 C_{yp}\right), if$$
$$\gamma_2(C_p - 2\rho_{pb}C_y)^2 > 0. \tag{35}$$
$$MSE_{min}(T_{rpi}) < MSE(T_{sgp})$$
$$\Rightarrow \gamma\bar{Y}^2{C_y}^2(1-\rho_{pb}^2) + \gamma_1\bar{Y}^2\rho_{pb}^2{C_y}^2 < \bar{Y}^2\left(\gamma{C_y}^2 + \frac{1}{4}\gamma_2{C_p}^2 + \gamma_2 C_{yp}\right), if$$
$$\gamma_2(C_p + 2\rho_{pb}C_y)^2 > 0. \tag{36}$$

Which is true in all the cases, therefore the proposed estimators are theoretically efficient .

5. Numerical illustration

The populations P1 and P2 have been considered for comparing the efficiency of the proposed ratio-cum-product double sampling exponential estimators T_{de1} and T_{de2} with the existing estimators. The population P1 is from Sukhatme and Sukhatme [18] where the study variable (Y) is the number of villages in the circles and the auxiliary information (ϕ) is a circle consisting of more than five villages. The population P2 is from Mukhopadhyay [9], in which the study variable (Y) is the household size and the auxiliary attribute (ϕ) is a household that availed an agricultural loan from a bank. A first phase sample of size 45 is drawn from the population P1 and a second phase sample which is nested within the first phase sample of size 23 is drawn. Further, a first phase sample of size 13 is drawn from the population P2 and a second phase sample which is nested within the first phase sample of size 7 is drawn.

Table 1: ***Characteristics of population data sets***

Population	N	η	n	$\bar{Y}$	p_1	ρ_{pb}	C_y	C_p	$\beta_2(\phi)$	C_{yp}
P1	89	45	23	2.911	0.067	0.586	0.542	3.782	11.433	1.201
P2	25	13	7	9.462	0.385	− 0.396	0.497	1.317	− 2.056	− 0.259

The description of the population given in Table-1 shows that the correlation between study and auxiliary variable for the population P1 is positive, for population P2 is negative.

Table 2: ***MSE and Bias of various estimators***

Estimator	**Population**	
	P1	
	MSE	**|Bias|**
t_m (Sample mean)	0.080	0.000
T_{ngr} (Naik and Gupta, 1996)	2.224	0.811
T_{sgr} (Singh *et al.*, 2007)	0.508	0.184
T_{rpi} (Proposed)	0.062	0.000

The data of Table-2 clearly shows that the estimators T_{rpi} have the lowest MSE as compared to the other existing estimators t_m, T_{ngr} and T_{sgr} and are also unbiased.

Table 3: ***PRE of*** T_{rpi} ***with respect to the estimators*** t_m, T_{ngr} ***and*** T_{sgr} ***for P6***

Estimator	**Population - P1**		
	Percent Relative Efficiency w.r.t		
	t_m	T_{ngr}	T_{sgr}
t_m (Sample mean)	100.000	2780.000	635.000
T_{ngr} (Naik and Gupta, 1996)	3.597	100.000	22.842
T_{sgr} (Singh *et al.*, 2007)	15.748	437.795	100.000
T_{rpi} (Proposed)	129.032	3587.097	819.355

Perusal of Table-3 shows that the PRE of the proposed estimators T_{rpi} for population P1 with respect to t_m, T_{ngr} and T_{sgr} is 129.032, 3587.097 and 819.355 respectively. The highest PRE value is found with respect to the exponential ratio estimator T_{ngr} followed by the PRE with respect to the exponential ratio type estimator T_{sgr} and the sample mean estimator t_m.

Table 4: ***MSE and Bias of*** t_m, T_{ngp}, T_{sgp} ***and*** T_{rpi}

Estimator	**Population**	
	P2	
	MSE	**|Bias|**
t_m (Sample mean)	2.362	0.000
T_{ngp} (Naik and Gupta, 1996)	9.816	0.162
T_{sgp} (Singh *et al.*, 2007)	3.431	0.190
T_{rpi} (Proposed)	2.125	0.000

The data of Table-4 clearly shows that the estimators T_{rpi} have the lowest MSE as compared to the other estimators t_m, T_{ngp} and T_{sgp} and are also unbiased.

Table 5: ***PRE of*** T_{rpi} ***with respect to the estimators*** t_m, T_{ngp} ***and*** T_{sgp} ***for P2***

Estimator	Population - P2		
	Percent Relative Efficiency w.r.t		
	t_m	T_{ngp}	T_{sgp}
t_m (Sample mean)	100.000	415.580	145.258
T_{ngp} (Naik and Gupta, 1996)	24.063	100.000	34.953
T_{sgp} (Singh *et al.*, 2007)	68.843	286.097	100.000
T_{rpi} (Proposed)	111.153	461.929	161.459

Perusal of Table-5 shows that the PRE of the proposed double sampling product type estimators T_{rpi} for population P2 with respect to t_m, T_{ngp} and T_{sgp} is 111.153, 461.929 and 161.459 respectively. The highest PRE value is found with respect to the exponential product estimator T_{ngp} followed by the PRE with respect to the exponential product type estimator T_{sgp} and the sample mean estimator t_m.

It can be observed from Table-3 and Table-5 that the proposed double sampling ratio estimators T_{rpi} have the highest PRE with respect to the existing estimators t_m, T_{ngr}, T_{ngp}, T_{sgp} and T_{sgr} for both the populations P1 and P2 which concludes that the proposed estimators are efficient.

6. Conclusion

In this paper, we have introduced two unbiased ratio-cum-product exponential estimators of the population mean, denoted as T_{rp1} and T_{rp2}, within the framework of double sampling. The large sample properties of these estimators have been derived to the first order of approximation and theoretically compared with existing ratio and product type estimators. Through numerical evaluation across various datasets, our comparison demonstrates that the proposed estimators consistently outperform existing ones in terms of efficiency while maintaining unbiasedness. These findings underscore the superior performance of the proposed estimators, making them a valuable contribution to the field.

References

[1] AL-Omari, A. and Bouza, C. 2015. Ratio estimators of the population mean with missing values using ranked set sampling, *Environmetrics*, 26(2):67-76.

[2] Bahl, S. and Tuteja, R. K. (1991). Ratio and product type exponential estimator. *Information and Optimization Sciences*, 12(1), 159–163.

[3] Abd-elfattah, A.M., El-Sherpieny, E.A., Mohamed, S.M. and Abdouf, O.F. 2010. Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute. *Applied Mathematics and Computation*, 215(12):4198-4202.

[4] Cochran, W.G. 1940. The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science*, 30:262-275.

[5] Hussain, S., Sharma, M. and Chandra, H., 2021. Modified Exponential Product Type Estimators for Estimating Population Mean Using Auxiliary Information. *In Special Proceedings 23rd Annual Conference*, 24-28.

[6] Hussain, S. and Bhat, V.A., 2023. New Median Based Almost Unbiased Exponential Type Ratio Estimators In The Absence Of Auxiliary Variable. *Reliability: Theory & Applications*, 18 (72):242-249.

[7] Hussain, S., Sharma, M., Bhat, V.A. and Bhat, M.I.J., 2024. Proportion Based Dual Unbiased Exponential Type Estimators of Population Mean. *Thailand Statistician*, 22(1):31-39.
[8] Jhajj, H.S., Sharma, M.K. and Grover, L.K. 2006. A family of estimators of population mean using information on auxiliary attribute. *Pakistan Journal of Statistics* 22(1):43.
[9] Mukhopadhyay, P. 2000. *Theory and methods of survey sampling*, PHI Learning, New Delhi.
[10] Mehta, V., Singh, H. P. and Pal, S. K. 2020. A general procedure for estimating finite population mean using ranked set sampling. *Investigacion Operacional*, 41(1):80-92.
[11] Naik, V.D. and Gupta, P.C. 1996. A note on estimation of mean with known population proportion of an auxiliary character. *Journal of the Indian Society of Agricultural Statistics*, 48(2):151-158.
[12] Robson, D.S. 1957. Applications of multivariate polykays to the theory of unbiased ratio type estimation. *Journal of the American Statistical Association*, 52:511-522.
[13] Shabbir, J. and Gupta, S. 2007. On estimating the finite population mean with known population proportion of an auxiliary variable. *Pakistan Journal of Statistics*, 23(1):1.
[14] Shabbir, J. and Gupta, S. 2010. Estimation of the finite population mean in two phase sampling when auxiliary variables are attributes. *Hacettepe Journal of Mathematics and Statistics*, 39(1):121-129.
[15] Shalabh and Tsai, J.R. 2017. Ratio and product methods of estimation of population mean in the presence of correlated measurement errors. *Communications in Statistics-Simulation and Computation*, 46(7):5566-5593.
[16] Singh, N. and Vishwakarma, G. K. 2019. A generalised class of estimator of population mean with the combined effect of measurement errors and non-response in sample survey. *Investigacion Operacional*, 40(2):275-285.
[17] Singh, R., Chauhan, P., Sawan, N., and Smarandache, F. 2007. Ratio-product type exponential estimator for estimating finite population mean using information on auxiliary attribute. *Renaissance High press USA*, 1:18-32.
[18] Sukhatme, P.V. and Sukhatme, B.V. 1970. *Sampling Theory of Surveys with Applications*, Lowa State University Press, USA.

OPTIMIZING AN INVENTORY MODEL FOR PERISHABLE PRODUCTS WITH PRODUCT RELIABILITY AND TIME DEPENDENT DEMAND USING PENTAGONAL-FUZZY NUMBER

Upasana Rana[1], Tanuj Kumar[1,*]

•

[1]Department of Mathematics, SRM Institute of Science and Technology,
Delhi-NCR Campus, Ghaziabad, India, 201204
ur2288@srmist.edu.in, tanujkuo@srmist.edu.in
*Corresponding author

Abstract

Enhancing inventory control for perishable goods is challenging since their shelf life is short and their demand is constantly changing. The current research examines into a more advanced inventory model for perishable goods, where demand is affected by both time and the reliability the product. The model occupies a pentagonal-fuzzy environment to assist with inherent uncertainties with these kinds of systems. This gives a more accurate picture of how demand fluctuates over time. Using analytical optimization techniques, the model targets to minimize total inventory costs, consisting of ordering cost, holding cost, and deterioration cost while maintaining high service levels. The total cost function is defuzzified using the Graded Mean Integration Representation (GMIR) method. The study's results, which were verified by numerical evaluations, demonstrate that the model is better at cost reduction and boosting dependability than other models using the MATLAB software. This research contributes a robust framework for handling perishable inventory with uncertain situations, which has a major impact on optimizing the supply chain.

Keywords: Inventory Model, Perishable Products, Time and Reliability-Dependent Demand, Pentagonal-Fuzzy Environment, GMIR Method.

1. Introduction and Literature Review

Managing perishable inventory is a vital part of supply chain operations, especially when factoring in product reliability and time-dependent demand. Perishable items inherently have limited lifespans and experience quality deterioration over time, making their inventory management more complex than that of non-perishable goods. Beyond these challenges, the integration of fuzzy logic, particularly within a pentagonal-fuzzy framework, introduces additional complexity to more effectively address the uncertainties present in real-world situations.

Product reliability and demand variability are crucial in shaping effective inventory management strategies for perishable items. The demand for these products frequently changes over time due to factors like seasonality, market trends, and consumer preferences. Additionally, product reliability, which reflects the stability of product quality and lifespan, directly affects inventory decisions by influencing replenishment cycles and stock levels. The work by Yadav, Yadav, and Bansal [25] addresses deteriorating products that deteriorate simultaneously during storage by

incorporating a two-warehouse system into an optimal inventory model to increase management efficiency. Their study extends earlier models by incorporating deterioration dynamics inside a multi-warehouse architecture, which lowers total costs and improves inventory control.

Many studies have examined inventory management for perishable goods, emphasizing the importance of including demand variability and product reliability in inventory models. Nahmias [7] offered an extensive review of perishable inventory models, detailing their features and the mathematical methods used for their management. Subsequent research has expanded on this groundwork, incorporating advanced mathematical techniques and computational tools to address the complexities of managing perishable inventory systems (Bakari et al., [1]). The influence of product reliability on inventory management has been thoroughly examined. Li and Liu [6] investigated how product reliability affects the formulation of optimal inventory policies, noting that increased reliability can reduce waste and lower costs. Additionally, other researchers have underscored the significance of integrating product reliability into perishable inventory models (Goyal & Giri, [5]]). Variations in demand over time are a key factor in inventory management. Various studies have developed models to address time-dependent demand, recognizing that demand for perishable products frequently exhibits specific patterns (Wee, [9]). Bhunia and Maiti [2] incorporated time-dependent demand into their inventory model for deteriorating items, showing how this approach can enhance inventory management. Yadav, A. S., Kumar, A., Yadav, K. K., & Rathee, S. [21], Kumar, K. [20], Mahata, S., & Debnath, B. K. [23], Mahata, S., & Debnath, B. K. [24], Yadav, K. K., Yadav, A. S., & Bansal, S. [22], the researchers, have carried out some recent research that examines deterioration's consequences and highlights the need for degradation measures in order to maximize inventory management.

The application of fuzzy logic in inventory management has become increasingly popular for handling uncertainty and ambiguity in demand and supply conditions. Zadeh [10] introduced fuzzy sets, which have since been utilized in many areas, including inventory management. Researchers have created fuzzy inventory models to better address the uncertainties present in real-world scenarios (Choudhury et al., [3]). Specifically, the pentagonal-fuzzy environment has been demonstrated to improve the flexibility and precision of inventory models for perishable products (Singh & Saxena, [8]). Recent research has concentrated on combining product reliability, time-dependent demand, and fuzzy logic into comprehensive inventory models. Giri et al. [4] created an integrated inventory model that accounts for product deterioration, time-varying demand, and fuzzy parameters. Poswal et al. [26] investigate a fuzzy EOQ model that accounts for price sensitivity and stock dependence while accounting for shortages. Through the use of fuzzy logic, their study enhances EOQ models by more effectively addressing the uncertainty associated with inventory and demand management. Their results suggest that these integrated approaches can greatly enhance inventory management outcomes for perishable products.

The literature highlights an increasing awareness of the significance of product reliability and time-dependent demand in managing perishable inventories. The use of fuzzy logic, especially within a pentagonal-fuzzy framework, presents a promising method for addressing the uncertainties inherent in these systems. This study seeks to advance this knowledge by creating and evaluating an inventory model for perishable products that incorporates these essential factors within a pentagonal-fuzzy context. Adak and Mahapatra [11], Mahapatra, G. S., Adak, S., & Kaladhar, K. [12], Manna, A. K., Dey, J. K., & Mondal, S. K. [13], Adak, S., & Mahapatra, G. S. [14], Rajput, N., Chauhan, A., & Pandey, R. K. [17], Adak, S., & Mahapatra, G. S. [15], Adak, S., & Mahapatra, G. S. [16], Khara, B., Dey, J. K., & Mondal, S. K. [18] and Mahapatra, G. S., Adak, S., Mandal, T. K., & Pal, S. [19] examine how reliability affects inventory systems, highlighting the need for reliability assessments to optimize inventory management.

2. Identified Research Gaps and Our Contributions

Despite extensive research, significant gaps persist in understanding perishable inventory management. Existing models often inadequately integrate product reliability within fuzzy environments and fail to fully address its impact on inventory strategies. Additionally, many models fall short

by not effectively combining time-dependent demand with factors such as product deterioration and reliability. The application of pentagonal-fuzzy logic remains underexplored, and the defuzzification process using the Graded Mean Integration Representation (GMIR) method is not yet well understood.

Our study addresses these gaps by incorporating product reliability into an inventory model within a pentagonal-fuzzy framework, enhancing our understanding of its impact on inventory management. We also develop a model that effectively integrates reliability, product deterioration, and time-dependent demand. Moreover, we utilize the GMIR method for defuzzification and expand the application of pentagonal-fuzzy logic to better manage uncertainties, offering a more precise approach to handling fuzzy data in perishable inventory models.

3. Notations and assumptions

3.1. Notations

The mathematical model was formulated based on the following notations.

Table 1: *Notations*

Notation	Units	Description
x	Constant	Coefficient of demand function
y	Constant	Coefficient of demand function
Z	Constant	Coefficient of deterioration rate.
C_1	\$/unit	Shortage cost.
r	–	product reliability.
A	\$/unit	The ordering cost
$I_1(t)$	Units	Inventory (Stock) level at a time t.
$I_2(t)$	Units	Stock out Inventory (Stock) level at a time t.
C_3	\$/Units	Deterioration cost.
C_4	\$/Units	Holding cost.
Q	Units	The number of orders placed in each cycle.
$ATC(t_1, T)$	\$/Units	The function for average total inventory cost.

Table 2: *Decision-making parameters*

Notation	Units	Description
t_1	–	Time, where shortage is start.
T	Years	Cycle length.

3.2. Assumptions

The mathematical model was formulated based on the following assumptions.

1. The demand rate function is influenced by both time and product reliability, represented by $D(r,t) = xtr^y$, where $x, y > 0$ are constants.
2. Shortages are fully deferred.
3. The filling rate is unlimited, and there is no lead time.
4. The inventory system operates over an infinite time horizon.
5. The deterioration rate remains constant and is denoted by Z; $Z > 0$.

4. Crisp Inventory Model Formulation

Initially, Q units of perishable goods were ordered. During the time interval $t \in [0, t_1]$, the combined effects of deterioration and demand lead to a reduction in the inventory level until it falls to zero. Hence, shortages are permitted to happen within the time interval $t \in [t_1, T]$. The demand that arises during the shortage period $t \in [t_1, T]$ is entirely deferred. (See Figure 1).

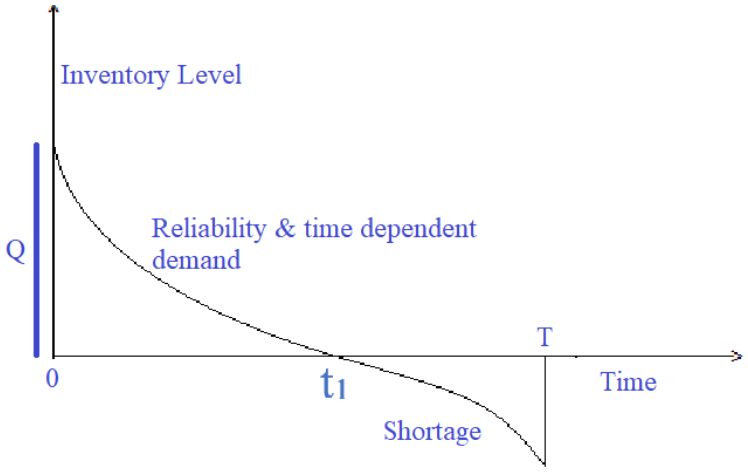


Figure 1: *A visual representation of the inventory system featuring fully backlogging.*

The stock level at $t = 0$ to $t = T$ is characterised in the differential equations as follows:

$$\frac{dI_1(t)}{dt} + ZI_1(t) = -(xtr^y); \quad t \in [0, t_1] \tag{1}$$

with the boundary conditions (B.C.) $I_1(t_1) = 0$ and $I_1(0) = Q$.

$$\frac{dI_2(t)}{dt} = -(xtr^y); \quad t \in [t_1, T] \tag{2}$$

with the boundary conditions (B.C.) $I_2(t_1) = 0$.

The equations (3) and (4) are the solutions of equations (1) and (2), respectively.

$$I_1(t) = \frac{xr^y}{Z^2}\left[1 - e^{Z(t_1-t)(1-Zt_1)}\right] - \frac{xr^y t}{Z} \tag{3}$$

$$I_2(t) = \frac{xr^y}{2}(t_1^2 - t^2) \tag{4}$$

Using the condition $I_1(0) = Q$ in (3), we get (4)

$$Q = \frac{xr^y}{Z^2}\left[1 - e^{Zt_1(1-Zt_1)}\right] \tag{5}$$

The overall cost per cycle comprises the following components:

1. **Opportunity cost:**

$$OC = A \tag{6}$$

2. **Deterioration cost:**

$$DC = ZC_3\left[Q - \int_0^{t_1}(xtr^y)dt\right] = \frac{ZC_3xr^y}{Z^2}\left[1 - e^{Zt_1(1-Zt_1)}\right] - \frac{ZC_3xr^yt_1^2}{2} \tag{7}$$

3. **Holding cost per cycle:**

$$HC = C_4\left[\int_0^{t_1} I_1(t)dt\right] = -\frac{C_4xr^y}{2Z^3}(2e^{Zt_1} + Z^2t_1^2 - 2Zt_1e^{Zt_1} - 2) \tag{8}$$

4. **Shortage cost:**

$$SC = C_1\int_{t_1}^{T} I_2(t)dt = -C_2xr^y\left[\frac{(T-t_1)^2(T+2t_1)}{6}\right] \tag{9}$$

Consequently, the retailer's total relevant inventory costs can be stated as follows:

$$ATC(t_1, T) = \frac{1}{T}[OC + DC + HC + SC]$$

$$ATC(t_1, T) = A + \frac{ZC_3xr^y}{Z^2}\left[1 - e^{Zt_1(1-Zt_1)}\right] - \frac{ZC_3xr^yt_1^2}{2} - \frac{C_4xr^y}{2Z^3}(2e^{Zt_1} + Z^2t_1^2 - 2Zt_1e^{Zt_1} - 2) - C_2xr^y\left[\frac{(T-t_1)^2(T+2t_1)}{6}\right] \tag{10}$$

5. Optimal Solution Approach for Crisp Model

The previously mentioned problem in equation [10] is highly nonlinear. We addressed these problems using an algorithm (described below) with the help of MATLAB software to find an optimal solution to the non linearity. To minimize total inventory cost function per unit time $ATC(t_1, T)$ the optimal values of t_1 and T can be obtained by solving the following equations:

$$\frac{\partial TC(t_1, T)}{\partial t_1} = 0 \ \ and \ \ \frac{\partial TC(t_1, T)}{\partial T} = 0$$

Further, for the total inventory cost function per unit time $ATC(t_1, T)$ to be convex, the following conditions must be satisfied

$$\frac{\partial^2 TC(t_1, T)}{\partial t_1^2} > 0, \ \frac{\partial^2 TC(t_1, T)}{\partial T^2} > 0, \ \ and \ \left(\frac{\partial^2 TC(t_1, T)}{\partial t_1^2} \cdot \frac{\partial^2 TC(t_1, T)}{\partial T^2} - \frac{\partial^2 TC(t_1, T)}{\partial t_1 \partial T}\right).$$

Algorithm:

Step I: Enter the values for all system parameters.
Step II: Construct the functions $ATC(t_1, T)$ as defined by equation [10].
Step III: Calculate the optimal values of the decision variables t_1^* and T^* using the previously mentioned optimal conditions
Step IV: Optimize the objective function $ATC^*(t_1, T)$ by employing the values t_1^* and T^*.
Step V: Determine the optimal values for $ATC^*(t_1, T)$.
Step VI: Finish.

6. Fuzzy Inventory Model Formulation

Given the environmental uncertainty, accurately defining all parameters is difficult. Therefore, we assume that parameters $\widetilde{x}, \widetilde{y}, \widetilde{Z}, \widetilde{C_1}, \widetilde{C_3}$ and $\widetilde{C_4}$ may vary within specific limits. Let pentagonal fuzzy numbers.

$$\widetilde{x} = (x_1, x_2, x_3, x_4, x_5), \quad \widetilde{y} = (y_1, y_2, y_3, y_4, y_5),$$
$$\widetilde{Z} = (Z_1, Z_2, Z_3, Z_4, Z_5), \quad \widetilde{C_1} = (C_{11}, C_{12}, C_{13}, C_{14}, C_{15}),$$
$$\widetilde{C_3} = (C_{31}, C_{32}, C_{33}, C_{34}, C_{35}), \quad \widetilde{C_4} = (C_{41}, C_{42}, C_{43}, C_{44}, C_{45}).$$

Total inventory cost function per unit time in fuzzy sense is given by:

$$\widetilde{ATC}(t_1, T) = A + \frac{\widetilde{Z}\widetilde{C_3}\widetilde{x}r^{\widetilde{y}}}{\widetilde{Z}^2}\left[1 - e^{\widetilde{Z}t_1(1-\widetilde{Z}t_1)}\right] - \frac{\widetilde{Z}\widetilde{C_3}\widetilde{x}r^{\widetilde{y}}t_1^2}{2} - \frac{\widetilde{C_4}\widetilde{x}r^{\widetilde{y}}}{2\widetilde{Z}^3}(2e^{\widetilde{Z}t_1} + \widetilde{Z}^2t_1^2 - 2\widetilde{Z}t_1e^{\widetilde{Z}t_1} - 2)$$
$$- \widetilde{C_2}\widetilde{x}r^{\widetilde{y}}\left[\frac{(T-t_1)^2(T+2t_1)}{6}\right] \quad (11)$$

We convert the fuzzy Total Inventory Cost function per unit time, $\widetilde{ATC}(t_1, T)$, into a crisp value using the Graded Mean Integration Representation (GMIR) method. The Total Inventory Cost function is expressed as follows:

$$\widetilde{ATC}_{dgm}(t_1, T) = \frac{1}{12}\left[\widetilde{ATC}_{dgm1}(t_1, T) + \widetilde{ATC}_{dgm2}(t_1, T) + \widetilde{ATC}_{dgm3}(t_1, T)\right.$$
$$\left. + \widetilde{ATC}_{dgm4}(t_1, T) + \widetilde{ATC}_{dgm5}(t_1, T)\right] \quad (12)$$

Where,

$$\widetilde{ATC_{dgmi}}(t_1, T) = A + \frac{\widetilde{Z_i}\widetilde{C_{3i}}\widetilde{x_i}r^{\widetilde{y_i}}}{\widetilde{Z_i}^2}\left[1 - e^{\widetilde{Z_i}t_1(1-\widetilde{Z_i}t_1)}\right] - \frac{\widetilde{Z_i}\widetilde{C_{3i}}\widetilde{x_i}r^{\widetilde{y_i}}t_1^2}{2} - \frac{\widetilde{C_{4i}}\widetilde{x_i}r^{\widetilde{y_i}}}{2\widetilde{Z_i}^3}(2e^{\widetilde{Z_i}t_1} + \widetilde{Z_i}^2t_1^2$$
$$- 2\widetilde{Z_i}t_1e^{\widetilde{Z_i}t_1} - 2) - \widetilde{C_{2i}}\widetilde{x_i}r^{\widetilde{y_i}}\left[\frac{(T-t_1)^2(T+2t_1)}{6}\right] \quad (13)$$

7. Optimal Solution Approach for Fuzzy Model

The previously mentioned problem in equation [12] is highly nonlinear. We addressed these problems using an algorithm (described below) with the help of MATLAB software to find an optimal solution to the non linearity. To minimize total inventory cost function per unit time $ATC(t_1, T)$ the optimal values of t_1 and T can be obtained by solving the following equations:

$$\frac{\partial TC(t_1, T)}{\partial t_1} = 0 \ \ and \ \ \frac{\partial TC(t_1, T)}{\partial T} = 0$$

Further, for the total inventory cost function per unit time $ATC(t_1, T)$ to be convex, the following conditions must be satisfied

$$\frac{\partial^2 TC(t_1, T)}{\partial t_1^2} > 0, \ \frac{\partial^2 TC(t_1, T)}{\partial T^2} > 0, \ \ and \ \left(\frac{\partial^2 TC(t_1, T)}{\partial t_1^2} \cdot \frac{\partial^2 TC(t_1, T)}{\partial T^2} - \frac{\partial^2 TC(t_1, T)}{\partial t_1 \partial T}\right).$$

Algorithm:

Step I: Enter the values for all system parameters.

Step II: Construct the functions $\widetilde{ATC}_{dgm}(t_1, T)$ as defined by equation [12].

Step III: Calculate the optimal values of the decision variables t_1^* and T^* using the previously mentioned optimal conditions

Step IV: Optimize the objective function $\widetilde{ATC}_{dgm}(t_1, T)$ by employing the values t_1^* and T^*.
Step V: Determine the optimal values for $\widetilde{ATC}_{dgm}(t_1, T)$.
Step VI: Finish.

8. Numerical Illustration

This section presents a numerical example to illustrate the proposed model for both crisp and fuzzy. To evaluate the best ordering strategies, we examine a retailer of perishable goods. The following examples are used to demonstrate the solution process:

8.1. Example

A numerical example is presented in this section to demonstrate the operation of the model. The input parameters used are as follows:

$$x = 0.015,\ y = 0.08,\ Z = 0.48,\ Q = 1674\ Units,\ A = 1320,\ r = 0.087,$$

$$C_1 = 3.1\ \$/unit,\ C_3 = 4.6\ \$/unit,\ C_4 = 0.08\ \$/unit.$$

The optimal solution obtained is as below:

$$t_1^* = 1.2\ Years,\ T^* = 1.6\ Years,\ ATC^* = \$807.4152/cycle.$$

9. Sensitivity analysis

This section utilizes sensitivity analysis to explore how changes in parameter values impact the optimal values. For these studies, each parameter was individually adjusted by $\pm 5\%$ A and $\pm 10\%$ B while keeping the other parameters at their original values. The results of the sensitivity analysis are shown in Table 3.

Table 3: *Effects of changing parameters of various kinds on the total cost and optimal solution.*

Parameters	% Change	% change in optimal value		
		t_1^*	T^*	TC^*
x	-10%	1.1397	1.5216	802.3642
	-5%	1.1525	1.5107	803.9865
	+5%	1.1680	1.5072	808.0856
	+10%	1.1894	1.4941	811.2044
y	-10%	1.2057	1.5762	809.9627
	-5%	1.2584	1.5900	804.8506
	+5%	1.2714	1.6292	799.5823
	+10%	1.3043	1.6536	796.3281
Z	-10%	1.2839	1.6867	803.9647
	-5%	1.2185	1.6228	807.2578
	+5%	1.1835	1.5835	811.6492
	+10%	1.1318	1.5437	816.0247
r	-10%	1.2434	1.5721	810.0624
	-5%	1.2862	1.5937	806.1987
	+5%	1.3048	1.6213	798.1573
	+10%	1.3481	1.6591	792.1557
A	-10%	1.1913	1.6027	804.2684
	-5%	1.2046	1.6348	805.1678
	+5%	1.2761	1.6724	807.1775
	+10%	1.2934	1.6914	813.1576

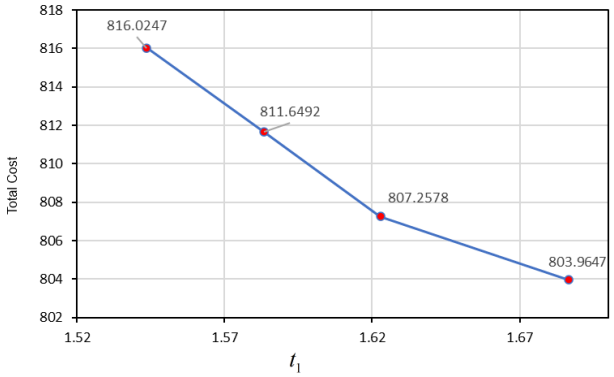


Figure 2: *Variation between* $ATC(r,T)$ *and* t_1

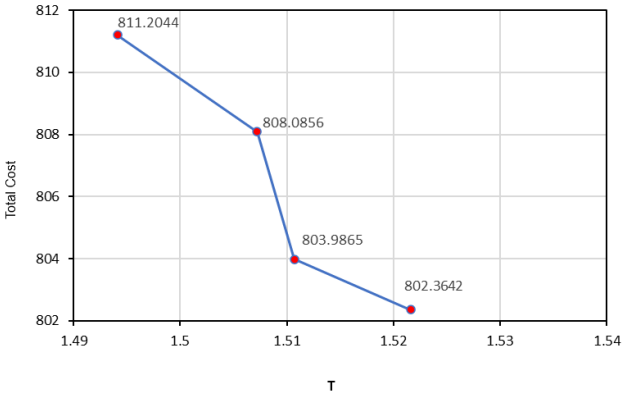


Figure 3: *Variation between* $ATC(r,T)$ *and* T

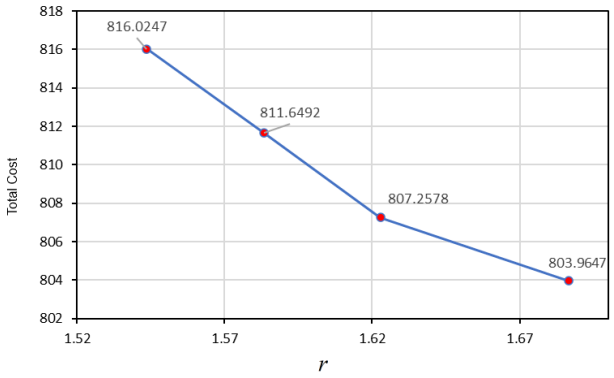


Figure 4: *Variation between* $ATC(r,T)$ *and reliability of the product* (r)

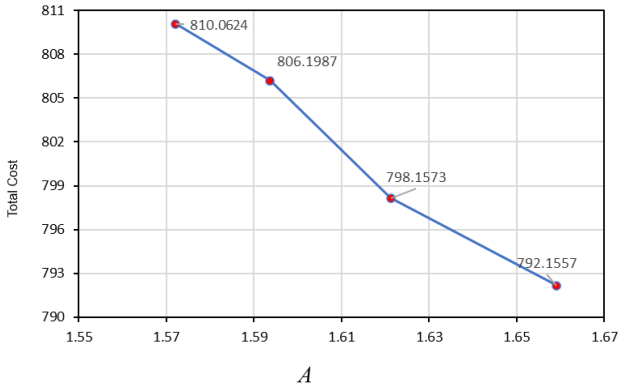


Figure 5: *Variation between $ATC(r,T)$ and Ordering cost (A)*

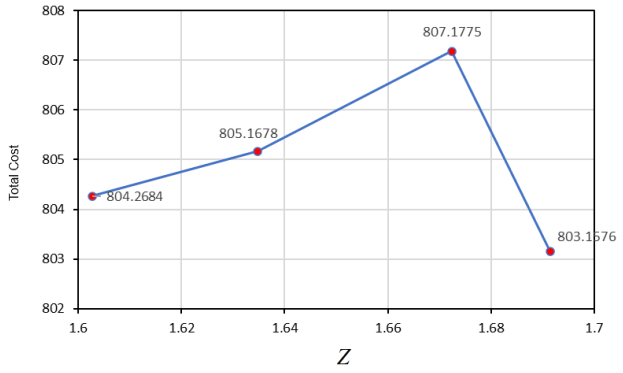


Figure 6: *Variation between $ATC(r,T)$ and deterioration rate Z*

10. Conclusion and Future Directions

This study examined the impact of product reliability and time-dependent demand on an inventory model for perishable items within a pentagonal-fuzzy framework. Using the Graded Mean Integration Representation (GMIR) method for defuzzification, the research delivered an in-depth analysis of the Total Inventory Cost function and its optimal values. The findings show that integrating product reliability and time-dependent demand into the model has a substantial effect on inventory management strategies and cost optimization for perishable products. The numerical examples and sensitivity analysis provide important insights into how changes in parameters influence optimal inventory decisions, thereby increasing the model's practical relevance. This optimization technique is particularly relevant in the food and beverage industry, where managing perishable inventory is crucial to cutting waste, ensuring product quality, and meeting shifting consumer demand. It can also be applied in the pharmaceutical sector, where product shelf life and dependability have a direct bearing on both patient safety and regulatory compliance.

Future research could build on this study by investigating several new areas. For instance, including additional factors like multi-echelon supply chains or different types of perishable products could offer a deeper understanding of inventory dynamics. Applying various defuzzification methods or advanced optimization algorithms might provide more accurate results and insights. Moreover, validating the model with real-world data could strengthen its robustness and applicability. Lastly, adapting the model to consider external factors such as market fluctuations or regulatory changes could increase its relevance and applicability across different contexts.

References

[1] Bakari, A., Bouchard, M., & Ouaret, A. (2020). Advances in perishable inventory management: A literature review. Journal of Industrial Engineering and Management, 13(3), 569-586.
[2] Bhunia, A. K., & Maiti, M. (1998). A two-warehouse inventory model for deteriorating items with a linear trend in demand and shortages. European Journal of Operational Research, 108(3), 346-356.
[3] Choudhury, S. K., Shankar, K., & Tiwari, M. K. (2008). Consensus in group decision making using evolutionary computing: A case of inventory management. Journal of Intelligent Manufacturing, 19(5), 571-582.
[4] Giri, B. C., & Chaudhuri, K. S. (2003). Optimal inventory policies for deteriorating items with time-varying demand and costs. Journal of the Operational Research Society, 54(5), 549-554.
[5] Goyal, S. K., & Giri, B. C. (2001). Recent trends in modeling of deteriorating inventory. European Journal of Operational Research, 134(1), 1-16.
[6] Li, R., & Liu, J. (2006). Inventory management with product quality and reliability. International Journal of Production Economics, 103(2), 631-643.
[7] Nahmias, S. (1982). Perishable inventory theory: A review. Operations Research, 30(4), 680-708.
[8] Singh, S. R., & Saxena, A. (2017). Fuzzy inventory model for deteriorating items with time-dependent demand under trade credit using pentagonal fuzzy numbers. Journal of Intelligent & Fuzzy Systems, 32(3), 2287-2298.
[9] Wee, H. M. (1995). A deterministic lot-size inventory model for deteriorating items with shortages and a declining market. Computers & Operations Research, 22(3), 345-356.
[10] Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338-353.
[11] Adak, Sudip, and G. S. Mahapatra. "Effect of reliability on varying demand and holding cost on inventory system incorporating probabilistic deterioration." Journal of Industrial & Management Optimization 18.1 (2022).
[12] Mahapatra, G. S., Sudip Adak, and Kolla Kaladhar. "A fuzzy inventory model with three parameter Weibull deterioration with reliant holding cost and demand incorporating reliability." Journal of Intelligent & Fuzzy Systems 36.6 (2019): 5731-5744.
[13] Manna, Amalesh Kumar, Jayanta Kumar Dey, and Shyamal Kumar Mondal. "Two layers supply chain in an imperfect production inventory model with two storage facilities under reliability consideration." Journal of Industrial and Production Engineering 35.2 (2018): 57-73.
[14] Adak, Sudip, and G. S. Mahapatra. "An inventory model of time and reliability Dependent demand with deterioration and partial back ordering under fuzziness." International Conference on Information Technology and Applied Mathematics. Cham: Springer International Publishing, 2019.
[15] Adak, Sudip, and G. S. Mahapatra. "Two-echelon imperfect production fuzzy supply chain model for reliability dependent demand with probabilistic deterioration and rework." Journal of Intelligent & Fuzzy Systems 41.2 (2021): 3833-3847.
[16] Adak, Sudip, and G. S. Mahapatra. "Effect of inspection and rework of probabilistic defective production on two-layer supply chain incorporating deterioration and reliability dependent demand." International Journal of System Assurance Engineering and Management 12.3 (2021): 565-578.
[17] Rajput, Neelanjana, Anand Chauhan, and R. K. Pandey. "Optimisation of an FEOQ model for deteriorating items with reliability influence demand." International Journal of Services and Operations Management 43.1 (2022): 125-144.
[18] Khara, Barun, Jayanta Kumar Dey, and Shyamal Kumar Mondal. "Effects of product reliability dependent demand in an EPQ model considering partially imperfect production." International Journal of Mathematics in Operational Research 15.2 (2019): 242-264.
[19] Mahapatra, G. S., et al. "Inventory model for deteriorating items with time and reliability dependent demand and partial backorder." International Journal of Operational Research 29.3 (2017): 344-359.

[20] Kumar, Kamal. "Inventory Policy for Deteriorating Items with Two Warehouse and Effect of Carbon Emission." Reliability: Theory & Applications 16.SI 2 (64) (2021): 156-165.

[21] Yadav, A.S., Kumar, A., Yadav, K.K. et al. Optimization of an inventory model for deteriorating items with both selling price and time-sensitive demand and carbon emission under green technology investment. Int J Interact Des Manuf (2023). https://doi.org/10.1007/s12008-023-01689-8

[22] Yadav, K.K., Yadav, A.S. & Bansal, S. Interval number approach for two-warehouse inventory management of deteriorating items with preservation technology investment using analytical optimization methods. Int J Interact Des Manuf (2024). https://doi.org/10.1007/s12008-023-01672-3

[23] Mahata, Sourav, and Bijoy Krishna Debnath. "A profit maximization single item inventory problem considering deterioration during carrying for price dependent demand and preservation technology investment." RAIRO-Operations Research 56.3 (2022): 1841-1856.

[24] Mahata, Sourav, and Bijoy Krishna Debnath. "The impact of R&D expenditures and screening in an economic production rate (EPR) inventory model for a flawed production system with imperfect screening under an interval-valued environment." Journal of Computational Science 69 (2023): 102027.

[25] Yadav, Krishan Kumar, Ajay Singh Yadav, and Shikha Bansal. "OPTIMIZATION OF AN INVENTORY MODEL FOR DETERIORATING ITEMS ASSUMING DETERIORATION DURING CARRYING WITH TWO-WAREHOUSE FACILITY." Reliability: Theory Applications 19.3 (79) (2024): 442-459. https://doi.org/10.24412/1932-2321-2024-379-442-459

[26] Poswal, Preety, et al. "Investigation and analysis of fuzzy EOQ model for price sensitive and stock dependent demand under shortages." Materials Today: Proceedings 56 (2022): 542-548.

RT&A, No 4(80)
Volume 19, December, 2024

INVENTORY MODEL FOR PROBABILISTIC DETERIORATION WITH RELIABILITY-DEPENDENT DEMAND AND TIME USING CLOUDY-FUZZY ENVIRONMENT

Ashish Negi[1], Ompal Singh[1,*]

•

[1] Department of Mathematics, SRM Institute of Science and Technology,
Delhi-NCR Campus, Ghaziabad, India, 201204
negi.ashish1995@gmail.com, ompalsit@srmist.edu.in
*Corresponding author

Abstract

Inventory control is vital in supply chain management, especially for perishable goods. The paper depicts a probabilistic inventory model for robust products where deterioration and demand change over time and depend on reliability. This paper also talks about the conventional back order reliability inventory model in a fuzzy, cloudy environment. This is because products deteriorate and demand fluctuates all the time. This study shows a novel approach to modeling inventory that deals with these problems. It does this by including uniform distribution deterioration, demand that depends on both time and product reliability, and cloudy-fuzzy numbers to show uncertainty. Although we start with the crisp model and fuzzifying it to obtain a decision under the cloudy fuzzy demand rate (which is an extension of dense fuzzy) demand rate, before putting it to use in practice. For ranking the fuzzy numbers, a new defuzzification method was used. Subsequently, extensive analysis is done to compare the crisp, general fuzzy solutions to the cloudy fuzzy solutions. The numerical examples and graphical are examined to demonstrate that the novel approach is useful in the model itself. The suggested model aims to maintain high service reliability while minimizing the total cost of inventory. Numerical analyses indicate that the model is effective, exhibiting that it can lower costs and improve reliability compared to older models using MATLAB software. This study builds a strong framework for managing inventory in supply lines for perishable goods, which opens up opportunities for more progress in this area.

Keywords: Probabilistic inventory model, Uniform Distribution Deterioration, Time and Reliability-Dependent Demand, Cloudy-Fuzzy Environment, Defuzzification.

1. Introduction and literature survey

Inventory management is an important part of the supply chain, especially for perishable products is an essential field to investigate because of the complexity associated with demand variability, deterioration and uncertainty in system parameters. This literature review carefully investigates at the most significant developments in inventory models, time and reliability-dependent demand, focusing on uniform probabilistic distribution deterioration, and the application of cloudy-fuzzy numbers for handling uncertainties.This study provides an advanced accounting model that takes these complicated issues into account by consisting of three key features: demand that is dependent on both time and product reliability, uniform distribution deterioration over time

and uncertainty represented by cloudy fuzzy numbers. Uniform distribution of deterioration shows how gradually and unpredictable nature the quality of perishable goods declines. Time and reliability dependent demand recognizes that customers choices about purchase are affected by the age the products are and how reliable they are perceived to be. Cloudy-fuzzy numbers provide a robust framework for handling the inherent uncertainties in demand and deterioration rates, combining the concepts of fuzzy logic and probability.

The purpose of this study is to develop up with an mathematically analytical optimization model that minimizes the total cost of inventory, which includes holding costs, deterioration cost, ordering cost, and reliability improvement costs. By making use of the cloudy-fuzzy number approach, we want to make an inventory management tool that is more accurate and flexible. so it can handle uncertainty as well as variations that happen in real-world supply chains.

1.1. Literature survey on Inventory Models with Uniform Probabilistic Distribution Deterioration

Different approaches have been used to study deterioration in inventory models in the conventionally began. In this way, Ghare and Schrader's [1] seminal work presented a model with constant deterioration rates. Real-life situations, on the other hand, frequently include uncertain deterioration, which is why probabilistic models have been developed. Uniform distribution a widely used probability distribution in inventory modelling is particularly relevant due to its simplicity and applicability when little is known about the distribution's parameters. Chang and Dye [2] examined into probabilistic deterioration in inventory systems, using a uniform distribution to model the uncertain the rate of deterioration. This approach was extended by Lo et al. [3] who applied a uniform probabilistic distribution to model deteriorating inventory with partial backlogging. Given that the uniform distribution can be used to model uncertainty in a number of ways, it is often used in inventory research. This is especially true for perishable products that whose rates of deterioration can vary.

Recent studies, such as those by Mahata, G and Mahata, P., [4] have these models were made even better by including uniform probability distributions to demonstrate different rates of degradation. These enhancements have made inventory models more precise and helpful, especially in environments in which there is much of uncertainty. Kumar, P and Dutta, D., [5] a deteriorating Inventory Model with Completely Backlogged Shortages and Uniformly Distributed Random Demand.

In the current market situation, it's hard to get an appropriate choice of the items customers want, so a probabilistic demand technique is more effective to deal with uncertainty. Pioneering the development of such models, Shah, N. H., [6] represent a probabilistic inventory model with allowed payment delays. Contributing to this, Shah, N. H. and Shah, Y. K., [7] included products that are getting worse and trade credit policy over a discrete time interval to the model. Several other researchers, including Shah, N. H., [8] created inventory models into probabilistic uncertain demand, trade credit financing, and shortages. De, L. N. and Goswami, A., [9] dedicated attention to joint inventory costs for customers and retailers, creating an EOQ model for deteriorating goods financed by trade credit. Khedlekar, et al. [10] Studied optimal replenishment decisions considering price, inventory under probabilistic demand and promotion strategies. Aldurgam et al. [11] demonstrated inventory control models for uncertain demand based on probability distributions, with a range of factors and assumptions. Probability distributions are often used to demonstrate how probabilistic demand.

1.2. Literature Survey on Time and Reliability-Dependent Demand in Inventory Models

Inventory management heavily relies on the relationship between demand and product reliability, particularly when it comes to perishable items. Models that take this dynamic into account are necessary since consumer demand frequently decreases as product quality deteriorates over time.

Their research presents the crucial it is to integrating reliability into demand forecasts in order to maximize inventory levels and minimize waste. The work that suggested A model of inventory where demand is a function of both reliability and time represents further advancements in this field. This method offers a more accurate depiction of customer behavior, especially in markets where demand is heavily influenced by product quality. Reliability has been progressively incorporated into demand modelling, with multi-echelon supply chain models containing reliability-dependent demand.

In the current competitive market, producing a high-reliable, dependable product is necessary to maintain a positive reputation in the marketplace. As a result, product quality, which is usually seen as a subset of reliability, is intimately tied to reliability. A product of superior quality is likely to be very reliable in the marketplace for an extended period of time, and vice versa. Reliability is a measure of the probability of failure-free operation over a specified period of time and is concerned with minimizing failure during that period. In actuality, it would be more sensible to view the demand as contingent on both time and reliability. As a result, demand for highly reliable products rises with time; that is, time typically influences a significant portion of customers to purchase high-reliability goods. Abdulla et al. [12] created an inventory model that is optimised while taking reliability into account. Due to the imperfect production that Krishnamoorthi and Panayappan [13] worked on, there are shortages and rework. Sarkar et al. [14] studied optimal reliability in the case of defective items.Pal et al. [15] Customers have many options available to them, but they tend to purchase highly dependable products to save money on future inconveniences and recurrent maintenance costs.

An economic manufacturing model with reliability is investigated by Bag et al. [16]. Sarkar et al. [17] have looked at a study on goods of defective quality and the impact of reliability. Tripathy, P. K., and Pattnaik, M. [18] built a model for inventories that takes reliability into account. Paul et al. [19] have created a production inventory model that takes reliability and unpredictability into account. Bhunia et al.[20] have investigated a production-inventory model that takes flexible reliability into account and a variable demand rate. The interval valued inventory cost criteria have been taken into consideration. An inventory model is examined by Khara et al. [21] who established a reliability model that takes into account the cost factors while accounting for the noninstantaneous deterioration rate of the products. focus of an imperfect production system is the demand rate, which is dependent on reliability. Abdel-Aleem et al. [22] has created an inventory model for a system of production with reliability. Rathore, H. [23] established a reliability model that takes into account the cost factors while accounting for the noninstantaneous deterioration rate of the products. Yadav et al. [24] perceived as a generalization of other research on deterioration and investments in green technologies.

1.3. Literature Survey on Cloudy-Fuzzy Numbers in Inventory Management

For a long time, inventory management has used fuzzy logic to deal with uncertainty in variables like lead times, deterioration rates, and demand. Zadeh [25] created fuzzy sets, which are now widely utilized in inventory models. Nevertheless, conventional fuzzy models frequently fail to fully express the degree of uncertainty present in real-world situations. It must be evident to all inventors that uncertainty gradually leaves the system over time. The traditional understanding of fuzzy set theory, however, makes the assumption that fuzziness is continuous across time. Therefore, De and Mahata [26] developed an inventory model with backorders under cloudy fuzzy and came up with the concept of cloudy fuzzy numbers. De and Mahata, [27] for mediocre quality items, a cloudy fuzzy EOQ model with reasonable proportionate discounts was created. Shah and Patel [28] discussed the best ordering practices for retailers using supplier credits in cases of fuzzy and cloudy fuzzy demand. Padiyar et al. [29] offered a supply chain model for a production process using cloudy fuzzy computing in a few real-world situations.

1.4. Literature Survey on Integration of Uniform Probabilistic Distribution Deterioration, Time and Reliability-Dependent Demand, and Cloudy-Fuzzy Numbers

An important development in inventory modelling is the combination of cloudy-fuzzy numbers, time and reliability-dependent demand, and uniform probabilistic distribution deterioration. Especially for perishable commodities, this method makes inventory systems more realistic and thorough to depict. Many academics have looked at classical inventory models in the study of the inventory system, which have a constant demand rate or that increase or decrease linearly. Numerous researchers have hypothesized that he function of time, stock, and price the demand rate is decreasing. However, the market has seen that, for a while, the demand pattern does not accurately reflect the growth in specific commodities, such as freshly introduced products, hardware devices, cosmetics, stylish clothing, mobile phones, and electronic items. Cash on hand is the most significant and practical factor in the inventory system, and it is a crucial factor in demand patterns. A salaried individual receives their pay at the beginning of each month or fortnight, thus their cash on hand is at its maximum during that period. As a result, they will make an effort to purchase more necessities at the beginning of the cycle. Therefore, the demand will automatically increase. They will purchase only essential and limited products starting in the middle of the time as their financial reserves diminish. As a result of limited purchasing options, demand will naturally decline. Thus, cash on hand is essential to the field of inventory management. Numerous scholars developed inventory models that took into account various demand categories and time-varying demand (Chang, H. J. and Dye, C. Y. [30], Khanra, et al. [31] and Sett, et al. [32]) price-sensitive demand model (Chanda, U. and Kumar, A. [33], Pal, S. and Mahapatra, G. S. [34]). Several researchers (Goli et al. [35] and [36], Shabani et al. [37]) examined a fuzzy deterioration rate and fuzzy demand rate inventory model. Barzegar et al. [38] examined vendor-managed inventory systems with constrained storage space and partial backordering under stochastic demand. . Modak and Kelle [39] demonstrated stochastic demand that is sensitive on pricing and delivery time. Shah and Vaghela [40] examined a time-and effort-dependent demand under-inflation and reliability model using an imperfect production inventory model. Barman et al. [41] offered a cloudy-fuzzy model of backordered inventories with inflation.

To a limited extent, this is accurate; yet, other academics have created various inventory models based on the premise that the holding cost remains constant throughout the inventory cycle. In the current day, buyers are drawn to products with a high degree of reliability; As a result, holding times are shortened and demand is automatically raised. This implies that for a product with high reliability, the holding cost will naturally decreases, resulting in a decreases in the overall cost. A lot of researchers talked about nonlinear holding costs in inventory models (San-Jose, et al. [42] and [43]) and time-varying holding cost (Alfares, H. K. and Ghaithan, A. M. [44], Pervin, et. al. [45]). Pervin et al. [46] talked about a two-echelon inventory model that included variable holding costs for degrading items and stock-dependent demand. Reliability that depends on holding costs in inventory system modelling is quite uncommon, nevertheless. This inventory system research takes reliability into account while assessing holding charges. By creating an analytical optimization model that takes into account cloudy-fuzzy numbers, time and reliability-dependent demand, and uniform probabilistic distribution deterioration, the current research seeks to close this gap. An enhanced framework for handling perishable inventory in uncertain situations is anticipated from this concept. Yadav et al. [47] regarded as interval valued the inventory parameters.This represents uncertainty in another manner.We don't employ any uncertainty elimination techniques in this case. Yadav et al. [48] used price sensitive demand or time sensitive demand. Dutta, Anurag, et al. [49] use of the data by National Aeronautics and Space Administration to train our model.

2. Preliminary Concept

2.1. Normalized General Triangular Fuzzy Number (NGTFN)

Let X be a NGTFN which is the form $\tilde{X} = (X_1, X_1, X_1)$,then its membership function which is defined by

$$\mu(\tilde{X}) = \begin{cases} 0, & \text{if } X < X_1 \text{ and } X > X_2 \\ \frac{X - X_1}{X_2 - X_1}, & \text{if } X_1 \leq X \leq X_2 \\ \frac{X_3 - X}{X_3 - X_2}, & \text{if } X_2 \leq X \leq X_3 \end{cases} \tag{1}$$

Now, the right and left α - cuts of $\mu(\tilde{X})$ are

$$R(\alpha) = X_3 + (X_3 - X_2)\alpha \; and \; L(\alpha) = X_1 + (X_2 - X_1)\alpha \tag{2}$$

It should be noted that the following formula can be used to determine the fuzziness measure:

2.2. Ranking Index

if $R(\alpha)$ and $L(\alpha)$ are the right and left $\alpha-$ cuts of a fuzzy number $\tilde{X}$ then Yager's [50] Ranking Index's defuzzification rule is provided by.

$$I(\tilde{X}) = \frac{1}{2}\int_0^1 [R(\alpha) + L(\alpha)]d\alpha \tag{3}$$

Note that the measure of fuzziness (degree of fuzziness d_f) can be found using the equation $d_f = \frac{U_b - L_b}{2g}$, where L_b and U_b are the lower bounds and upper bounds of the fuzzy numbers respectively and g being their respective mode.

2.3. Cloudy Normalized Triangular Fuzzy Number (CNTFN)

A fuzzy number which is the form $\tilde{A} = (a_1, a_2, a_3)$ is called cloudy triangular fuzzy number. An The set itself converges to a crisp singleton in infinite time. i.e., both $a_1, a_3 \to a_2$. as time $t \to \infty$ Considering the fuzzy number(extension of De and Beg [51]).

$$\tilde{A} = \left\langle a_2\left(1 - \frac{\rho}{1+t}\right), a_2, a_2\left(1 + \frac{\sigma}{1+t}\right)\right\rangle \tag{4}$$

Note that $\lim_{t\to\infty} a_2\left(1 - \frac{\rho}{1+t}\right) = a_2$ and $\lim_{t\to\infty} a_2\left(1 + \frac{\sigma}{1+t}\right) = a_2$ so $\tilde{A} \to a_2$. The following are the membership functions for $0 \leq t$:

$$\mu(x,t) = \begin{cases} 0, & \text{if } X < a_2\left(1 - \frac{\rho}{1+t}\right) \text{ and } X > a_2\left(1 + \frac{\sigma}{1+t}\right) \\ \frac{X - a_2\left(1 - \frac{\rho}{1+t}\right)}{\frac{\rho a_2}{1+t}}, & \text{if } a_2\left(1 - \frac{\rho}{1+t}\right) \leq x \leq a_2 \\ \frac{a_2\left(1 + \frac{\sigma}{1+t}\right) - x}{\frac{\sigma a_2}{1+t}}, & \text{if } a_2 \leq x \leq a_2\left(1 + \frac{\sigma}{1+t}\right) \end{cases} \tag{5}$$

The graphical representation of CNTFN (Figure 1) is obtained in the manner described below:

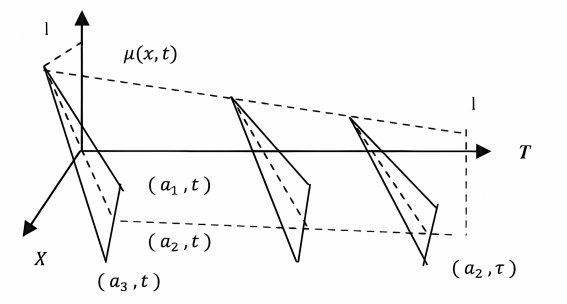


Figure 1: *Membership function of CNTFN*

2.4. Ranking Index on CNTFN

Let consider right and left α-cuts of $\mu(x,t)$ from equation (5) noted as $R(\alpha,t)$ and $L(\alpha,t)$ respectively. After that, under the temporal extension of Yager's ranking index(Extended De and Beg's [51]), the defuzzification formula is provided by

$$I(\tilde{A}) = \frac{1}{2T}\int\int_{\alpha=0,t=0}^{\alpha=1,t=T}[R^{-1}(\alpha,t) + L^{-1}(\alpha,t)]d\alpha dt \tag{6}$$

Here, α and t are independent variables. Let $\tilde{A}$ be a CNTFN stated in equation(4). Then the its membership function equation(5). Now taking the right and left α-cuts of $\mu(x,t)$ from equation(5) we get

$$R^{-1}(\alpha,t) = a_2\left(1 + \frac{\sigma}{1+t} - \frac{\alpha\sigma}{1+t}\right) \; and \; L^{-1}(\alpha,t) = a_2\left(1 - \frac{\rho}{1+t} + \frac{\alpha\rho}{1+t}\right) \tag{7}$$

Thus using equation(6), we have

$$I(\tilde{A}) = \frac{a_2}{2T}\left[2T + \frac{\sigma-\rho}{2}log(1+T)\right] \tag{8}$$

Again equation(8) can be rewritten as

$$I(\tilde{A}) = a_2\left[1 + \frac{\sigma-\rho}{4T}log(1+T)\right] \tag{9}$$

Obviously, $\lim_{T\to\infty}\frac{log(1+T)}{T} = 0$ and therefore $I(\tilde{A}) \to a_2$ as $T \to \infty$ Note that the factor may be referred to as $\frac{log(1+T)}{T}$ cloud index (CI) and in actuality, the time T is measured in days. Figure 2 depicts the cloud index's nature.

3. Assumptions and notations

3.1. Assumptions

1. Demand rate is depends on both time and product reliability, which is $D(r,t) = mtr^n$, where $m, n > 0$ are constants.
2. Shortages are permitted and the backlog is fully..
3. Deterioration rate parameter is depending on uniform distribution. Which is $\alpha = f(\xi) = \frac{1}{\gamma_2-\gamma_1}(\gamma_1 < \xi < \gamma_2)$ similar to Shah, Nita H., et al. [8] as in demand rate.
4. Here $\alpha \sim U(5,10); \alpha = 0.2$ is uniform distribution deterioration parameter.

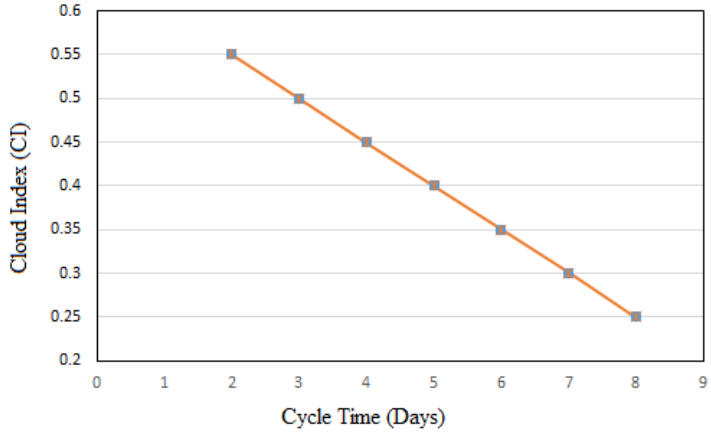


Figure 2: *Character of fuzziness across time*

3.2. Notations

Table 1 a description of the notations used for the created mathematical model is given.

Table 1: *Notations*

Notation	Units	Description
p	\$/unit	Deterioration Cost
h	\$/unit	Holding Cost
q	\$/unit	Shortage cost.
A	\$/unit	ordering cost.
n	Reliability parameter	Which is lies always between $0 < n < 1$
m	–	shape parameter $m > 0$
$I_1(t)$	units	Inventory Level at time t over the ordering cycle $(0, t_1)$
$I_2(t)$	units	Inventory Level at time t over the ordering cycle (t_1, T)
Q	units	Maximum inventory level at time t in $[0, T]$ for ordering cycle i.e.$I_1(0) = Q$.
R	Decision Variable	Shortage Level.
T	Units	Length of total replenishment cycle(days).
$TIC(r, T)$	\$/Units	Total Average inventory cost unit time.
r	–	Reliability of item of the inventory system.
T	Years	Length of the cycle.
$\widetilde{TIC}(r, T)$	\$/Units	Fuzzy Total Average inventory cost unit time.

4. Mathematical Model Formulation

The stock level at $t = 0$ to $t = T$ is characterised as follows in the differential equations:

$$\frac{dI_1(t)}{dt} + \alpha I_1(t) = -(mtr^n); \quad 0 \le t \le t_1 \tag{10}$$

with boundary conditions (B.C.) $I_1(t_1) = 0$ and $I_1(0) = Q$.

$$\frac{dI_2(t)}{dt} = -(mtr^n); \quad t_1 \le t \le T \tag{11}$$

with the boundary conditions (B.C.) $I_2(t_1) = 0$ and $I_2(T) = -R$.
The solutions to equations (10) and equation (11) are equation (12) and equation(13), respectively.

$$I_1(t) = \frac{mr^n}{\alpha^2}\left[1 - e^{\alpha(t_1 - t)}(1 - \alpha t_1)\right]; \quad 0 \le t \le t_1 \tag{12}$$

$$I_2(t) = \frac{mr^n}{2}(T^2 - t_1^2); \quad t_1 \le t \le T \tag{13}$$

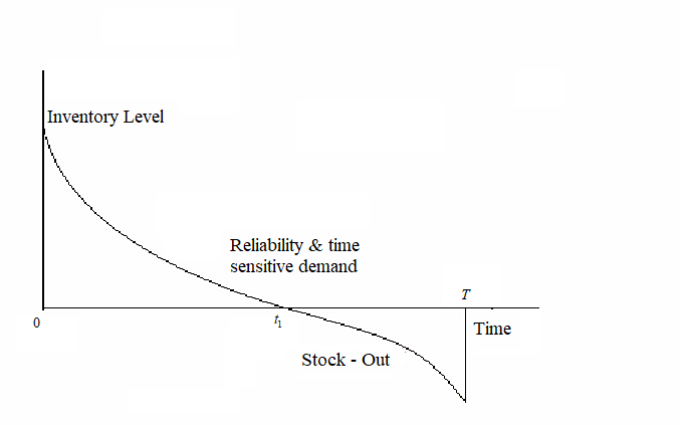


Figure 3: *Mathematical Model between inventory and time*

The following elements make up the inventory system's total cost per unit of time.

(a). **Opportunity Cost :**

$$OC = A \tag{14}$$

(b). **Deterioration cost :**

$$DC = p \cdot \alpha \int_0^{t_1} I_1(t)dt$$

$$DC = -\frac{pmr^n}{2\alpha^2}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2) \tag{15}$$

(c). **Holding Cost :**

$$HC = h \int_0^{t_1} I_1(t)dt$$

$$HC = -\frac{hmr^n}{2\alpha^3}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2) \tag{16}$$

(d). **Shortage Cost :**

$$SC = q \int_{t_1}^{T} [-I_2(t)]dt$$

$$SC = \frac{qmr^n}{6}(T - t_1)^2(T + 2t_1) \tag{17}$$

Hence, the total cost function per unit time is:

(e). **Total Inventory Cost (TIC):**

$$TIC = \frac{1}{T}[OC + DC + HC + SC]$$

$$TIC = \frac{1}{T}[A - \frac{pmr^n}{2\alpha^2}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2) - \frac{hmr^n}{2\alpha^3}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2) + \frac{qmr^n}{6}(T - t_1)^2(T + 2t_1)] \tag{18}$$

Thus, our problem is given by

$$\begin{cases} Minimize\ TIC = \frac{1}{T}[A - \frac{pmr^n}{2\alpha^2}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2) - \frac{hmr^n}{2\alpha^3}(2e^{\alpha t_1} + \alpha^2 t_1^2 \\ -2\alpha t_1 e^{\alpha t_1} - 2) + \frac{qmr^n}{6}(T - t_1)^2(T + 2t_1)] \\ Subject\ to\ Q = \frac{mr^n}{\alpha^2}\left[1 - e^{\alpha t_1}(1 - \alpha t_1)\right], R = \frac{mr^n}{2}(T^2 - t_1^2) \end{cases} \tag{19}$$

5. Developing a Fuzzy Mathematical Model

Over the course of the inventory run period, let the demand rate follow a general fuzzy and cloudy fuzzy. Then taking $\tilde{X}$ as follows

$$\tilde{X} = \begin{cases} \langle X_1, X_2, X_3\rangle\ for\ NGTFN \\ \left\langle X\left(1 - \frac{\rho}{1+t}\right), X, X\left(1 + \frac{\sigma}{1+t}\right)\right\rangle\ for\ CNTFN,\ where\ 0 < \rho, \sigma < 1\ and\ T > 0 \end{cases} \tag{20}$$

And the corresponding fuzzy problem is given by

$$\begin{cases} Minimize\ \widetilde{TIC} = \frac{1}{T}[A - \frac{\tilde{p}\tilde{m}r^{\tilde{n}}}{2\tilde{\alpha}^2}(2e^{\tilde{\alpha}t_1} + \tilde{\alpha}^2 t_1^2 - 2\tilde{\alpha} t_1 e^{\tilde{\alpha}t_1} - 2) - \frac{\tilde{h}\tilde{m}r^{\tilde{n}}}{2\tilde{\alpha}^3}(2e^{\tilde{\alpha}t_1} + \tilde{\alpha}^2 t_1^2 \\ -2\tilde{\alpha} t_1 e^{\tilde{\alpha}t_1} - 2) + \frac{\tilde{q}\tilde{m}r^{\tilde{n}}}{6}(T - t_1)^2(T + 2t_1)] \\ Subject\ to\ Q = \frac{\tilde{m}r^{\tilde{n}}}{\tilde{\alpha}^2}\left[1 - e^{\tilde{\alpha}t_1}(1 - \tilde{\alpha}t_1)\right], R = \frac{\tilde{m}r^{\tilde{n}}}{2}(T^2 - t_1^2) \end{cases} \tag{21}$$

Now, Using equation (3), the order amount and shortage quantity under NGTFN are given by the membership function for the fuzzy objective

$$\mu_1(T\tilde{I}C) = \begin{cases} 0, & \text{if } TIC < TIC_1 \text{ and } TIC > TIC_2 \\ \frac{TIC - TIC_1}{TIC_2 - TIC_1}, & \text{if } TIC_1 \le TIC \le TIC_2 \\ \frac{TIC_3 - TIC}{TIC_3 - TIC_2}, & \text{if } TIC_2 \le TIC \le TIC_3 \end{cases} \tag{22}$$

where,

$$TIC_1 = \frac{1}{T}\Bigg[A - \frac{p_1 m_1 r^{n_1}}{2\alpha_1^2}(2e^{\alpha_1 t_1} + \alpha_1^2 t_1^2 - 2\alpha_1 t_1 e^{\alpha_1 t_1} - 2) - \frac{h_1 m_1 r^{n_1}}{2\alpha_1^3}(2e^{\alpha_1 t_1} + \alpha_1^2 t_1^2 - 2\alpha_1 t_1 e^{\alpha_1 t_1} - 2) + \frac{q_1 m_1 r^{n_1}}{6}(T - t_1)^2(T + 2t_1)\Bigg]$$

$$TIC_2 = \frac{1}{T}\Bigg[A - \frac{p_2 m_2 r^{n_2}}{2\alpha_2^2}(2e^{\alpha_2 t_1} + \alpha_2^2 t_1^2 - 2\alpha_2 t_1 e^{\alpha_2 t_1} - 2) - \frac{h_2 m_2 r^{n_2}}{2\alpha_2^3}(2e^{\alpha_2 t_1} + \alpha_2^2 t_1^2 - 2\alpha_2 t_1 e^{\alpha_2 t_1} - 2) + \frac{q_2 m_2 r^{n_2}}{6}(T - t_1)^2(T + 2t_1)\Bigg]$$

$$TIC_3 = \frac{1}{T}\Bigg[A - \frac{p_3 m_3 r^{n_3}}{2\alpha_3^2}(2e^{\alpha_3 t_1} + \alpha_3^2 t_1^2 - 2\alpha_3 t_1 e^{\alpha_3 t_1} - 2) - \frac{h_3 m_3 r^{n_3}}{2\alpha_3^3}(2e^{\alpha_3 t_1} + \alpha_3^2 t_1^2 - 2\alpha_3 t_1 e^{\alpha_3 t_1} - 2) + \frac{q_3 m_3 r^{n_3}}{6}(T - t_1)^2(T + 2t_1)\Bigg]$$

$$\mu_2(\tilde{Q}) = \begin{cases} 0, & \text{if } Q < Q_1 \text{ and } Q > Q_2 \\ \frac{Q-Q_1}{Q_2-Q_1}, & \text{if } Q_1 \le Q \le Q_2 \\ \frac{Q_3-Q}{Q_3-Q_2}, & \text{if } Q_2 \le Q \le Q_3 \end{cases} \tag{23}$$

where

$$Q_i = \frac{m_i r^{n_i}}{\alpha_i^2}\left[1 - e^{\alpha_i t_1}(1-\alpha_i t_1)\right];\ \ i = 1,2,3$$

$$\mu_3(\tilde{R}) = \begin{cases} 0, & \text{if } R < R_1 \text{ and } R > R_2 \\ \frac{R-R_1}{R_2-R_1}, & \text{if } R_1 \le R \le R_2 \\ \frac{R_3-R}{R_3-R_2}, & \text{if } R_2 \le R \le R_3 \end{cases} \tag{24}$$

where

$$R_i = \frac{m_i r^{n_i}}{2}(T^2 - t_1^2);\ \ i = 1,2,3$$

The fuzzy goal, fuzzy order quantity, and fuzzy shortage quantity index values are obtained, respectively, using equation (2) and equation (3).

$$\begin{aligned} I(\widetilde{TIC}) &= \frac{1}{4}(TIC_1 + 2TIC_2 + TIC_3) \\ &= \frac{(m_1 r^{n_1} + m_2 r^{n_2} + m_3 r^{n_3})}{4}\left[\left(\frac{q_1}{6T}(T-t_1)^2(T+2t_1) + \frac{2q_2}{6T}(T-t_1)^2(T+2t_1)\right.\right. \\ &\quad \left. + \frac{q_3}{6T}(T-t_1)^2(T+2t_1)\right) + \left(-\frac{p_1}{2T\alpha_1^2}(2e^{\alpha_1 t_1} + \alpha_1^2 t_1^2 - 2\alpha_1 t_1 e^{\alpha_1 t_1} - 2)\right. \\ &\quad - \frac{p_2}{2T\alpha_2^2}(2e^{\alpha_2 t_1} + \alpha_2^2 t_2^2 - 2\alpha_1 t_1 e^{\alpha_2 t_1} - 2) - \left.\frac{p_3}{2T\alpha_3^2}(2e^{\alpha_3 t_1} + \alpha_3^2 t_1^2 - 2\alpha_3 t_1 e^{\alpha_3 t_1} - 2)\right) \\ &\quad + \left(-\frac{h_1}{2T\alpha_1^3}(2e^{\alpha_1 t_1} + \alpha_1^2 t_1^2 - 2\alpha_1 t_1 e^{\alpha_1 t_1} - 2) - \frac{h_2}{2T\alpha_2^3}(2e^{\alpha_2 t_1} + \alpha_2^2 t_1^2 - 2\alpha_2 t_1 e^{\alpha_2 t_1} - 2)\right. \\ &\quad \left.\left. - \frac{h_3}{2T\alpha_3^3}(2e^{\alpha_3 t_1} + \alpha_3^2 t_1^2 - 2\alpha_3 t_1 e^{\alpha_3 t_1} - 2)\right)\right] + \frac{4A}{T} \end{aligned} \tag{25}$$

$$I(\widetilde{Q}) = \frac{1}{4}(Q_1 + 2Q_2 + Q_3) = \frac{(m_1 r^{n_1} + m_2 r^{n_2} + m_3 r^{n_3})}{4}\left[\left(\frac{1}{\alpha_1^2}\{1 - e^{\alpha_1 t_1}(1-\alpha_1 t_1)\}\right)\right.$$
$$\left. + \left(\frac{1}{\alpha_2^2}\{1 - e^{\alpha_2 t_1}(1-\alpha_2 t_1)\}\right) + \left(\frac{1}{\alpha_3^2}\{1 - e^{\alpha_3 t_1}(1-\alpha_3 t_1)\}\right)\right]$$

$$I(\widetilde{R}) = \frac{1}{4}(R_1 + 2R_2 + R_3) = \frac{(m_1 r^{n_1} + m_2 r^{n_2} + m_3 r^{n_3})(T^2 - t_1^2)}{2}$$

Particular Case:

(i) if $m_1 r^{n_1} \to m_2 r^{n_2}$ and $m_3 r^{n_3} \to m_2 r^{n_2} \to m r^n$

Then $I(\widetilde{TIC}) \to \frac{(m_1 r^{n_1} + m_2 r^{n_2} + m_3 r^{n_3})}{4}\left[\left(\frac{q_1}{6T}(T-t_1)^2(T+2t_1) + \frac{2q_2}{6T}(T-t_1)^2(T+2t_1)\right.\right.$
$\left. + \frac{q_3}{6T}(T-t_1)^2(T+2t_1)\right) + \left(-\frac{p_1}{2T\alpha_1^2}(2e^{\alpha_1 t_1} + \alpha_1^2 t_1^2 - 2\alpha_1 t_1 e^{\alpha_1 t_1} - 2) - \frac{p_2}{2T\alpha_2^2}(2e^{\alpha_2 t_1} + \alpha_2^2 t_2^2\right.$

$-2\alpha_1 t_1 e^{\alpha_2 t_1} - 2) - \frac{p_3}{2T\alpha_3^2}(2e^{\alpha_3 t_1} + \alpha_3^2 t_1^2 - 2\alpha_3 t_1 e^{\alpha_3 t_1} - 2)\Big) + \Big(- \frac{h_1}{2T\alpha_1^3}(2e^{\alpha_1 t_1} + \alpha_1^2 t_1^2 - 2\alpha_1 t_1 e^{\alpha_1 t_1}$
$- 2) - \frac{h_2}{2T\alpha_2^3}(2e^{\alpha_2 t_1} + \alpha_3^2 t_1^2 - 2\alpha_3 t_1 e^{\alpha_3 t_1} - 2)\Big)\Big]$,
$I(\widetilde{Q}) \to \frac{mr^n}{\alpha^2}\Big[1 - e^{\alpha t_1}(1 - \alpha t_1)\Big]$ and $I(\widetilde{R}) \to \frac{mr^n}{2}(T^2 - t_1^2)$

(ii) if $t_1 \to T, I(\widetilde{TIC}) \to -\frac{pmr^n}{2T\alpha^2}\Big(2e^{\alpha T} + \alpha^2 T^2 - 2\alpha T e^{\alpha T} - 2\Big)$
$- \frac{hmr^n}{2T\alpha^3}\Big(2e^{\alpha T} + \alpha^2 T^2 - 2\alpha T e^{\alpha T} - 2\Big) + \frac{A}{T}$,

On the other hand, applying equation (5), the fuzzy objective, fuzzy order quantity and fuzzy shortage quantity membership functions under the cloudy fuzzy model are provided by

$$\omega_1(TIC, T) = \begin{cases} 0, & \text{if } TIC < TIC_{11} \text{ and } TIC > TIC_{21} \\ \frac{TIC - TIC_{11}}{TIC_{21} - TIC_{21}}, & \text{if } TIC_{21} \le TIC \le TIC_{21} \\ \frac{TIC_{31} - TIC}{TIC_{31} - TIC_{21}}, & \text{if } TIC_{21} \le TIC \le TIC_{31} \end{cases} \tag{26}$$

where,
$TIC_{11} = \frac{1}{T}\Big[\frac{q_1}{6}(T - t_1)^2(T + 2t_1) - \frac{p}{2\alpha^2}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2)$
$- \frac{h}{2\alpha^2}(2e^{\alpha t_1} + \alpha^3 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2)\Big] mr^n\Big(1 - \frac{\rho}{1+T}\Big) + \frac{A}{T}$
$TIC_{21} = \frac{1}{T}\Big[\frac{q_1}{6}(T - t_1)^2(T + 2t_1) - \frac{p}{2\alpha^2}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2)$
$- \frac{h}{2\alpha^2}(2e^{\alpha t_1} + \alpha^3 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2)\Big] mr^n + \frac{A}{T}$
$TIC_{31} = \frac{1}{T}\Big[\frac{q_1}{6}(T - t_1)^2(T + 2t_1) - \frac{p}{2\alpha^2}(2e^{\alpha t_1} + \alpha^2 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2)$
$- \frac{h}{2\alpha^2}(2e^{\alpha t_1} + \alpha^3 t_1^2 - 2\alpha t_1 e^{\alpha t_1} - 2)\Big] mr^n\Big(1 + \frac{\sigma}{1+T}\Big) + \frac{A}{T}$

$$\omega_2(Q, T) = \begin{cases} 0, & \text{if } Q < Q_{11} \text{ and } Q > Q_{21} \\ \frac{Q - Q_{11}}{Q_{21} - Q_{11}}, & \text{if } Q_{11} \le Q \le Q_{21} \\ \frac{Q_{31} - Q}{Q_{31} - Q_{21}}, & \text{if } Q_{21} \le Q \le Q_{31} \end{cases} \tag{27}$$

where
$Q_{11} = \Big(1 - \frac{\rho}{1+T}\Big)\frac{mr^n}{\alpha^2}\Big[1 - e^{\alpha t_1}(1 - \alpha t_1)\Big]$, $Q_{21} = \frac{mr^n}{\alpha^2}\Big[1 - e^{\alpha t_1}(1 - \alpha t_1)\Big]$
$Q_{31} = \Big(1 + \frac{\sigma}{1+T}\Big)\frac{mr^n}{\alpha^2}\Big[1 - e^{\alpha t_1}(1 - \alpha t_1)\Big]$

$$\omega_3(R, T) = \begin{cases} 0, & \text{if } R < R_{11} \text{ and } R > R_{21} \\ \frac{R - R_{11}}{R_{21} - R_{11}}, & \text{if } R_{11} \le R \le R_{21} \\ \frac{R_{31} - R}{R_{31} - R_{21}}, & \text{if } R_{21} \le R \le R_{31} \end{cases} \tag{28}$$

where
$R_{11} = \Big(1 - \frac{\rho}{1+T}\Big)\frac{mr^n}{2}(T^2 - t_1^2)$, $R_{21} = \frac{mr^n}{2}(T^2 - t_1^2)$, $R_{31} = \Big(1 + \frac{\sigma}{1+T}\Big)\frac{mr^n}{2}(T^2 - t_1^2)$

The cloudy fuzzy objective, cloudy fuzzy order quantity and cloudy fuzzy shortage quantity index values are obtained by using equation (6).

$J(\widetilde{TIC}) = \frac{1}{\tau}\int_{T=0}^{\tau}\frac{1}{4}[TIC_{11}+2TIC_{21}+TIC_{31}]dT = \frac{1}{4\tau}\int_{T=0}^{\tau}[TIC_{11}+2TIC_{21}+TIC_{31}]dT$

$$J(\widetilde{TIC}) = \frac{1}{4\tau}\int_{T=0}^{\tau}\left[\frac{4A}{T}+\left\{\left(4+\frac{\sigma-\rho}{1+T}\right)\left(\frac{q_1}{6T}(T-t_1)^2(T+2t_1)-\frac{p}{2T\alpha^2}(2e^{\alpha t_1}\right.\right.\right.$$
$$\left.\left.\left.+\alpha^2t_1^2-2\alpha t_1e^{\alpha t_1}-2)-\frac{h}{2T\alpha^2}(2e^{\alpha t_1}+\alpha^3t_1^2-2\alpha t_1e^{\alpha t_1}-2)\right)\right\}mr^n\right]dT$$

$$J(\widetilde{TIC}) = -\frac{pmr^n}{2\alpha^2}(2e^{\alpha t_1}+\alpha^2t_1^2-2\alpha t_1e^{\alpha t_1}-2)+\frac{mr^nlog\left(\frac{\tau}{\epsilon}\right)}{2\tau}\left\{\frac{2A}{mr^n}+\left(1+\frac{\sigma-\rho}{4}\right)\right.$$
$$\left.(c_1+c_2)t_1^2\right\}+\frac{c_2mr^n}{4}(\tau-4t_1)+\frac{(\sigma-\rho)mr^n}{4}\left[c_2+\left\{-\frac{p}{\alpha^2}(2e^{\alpha t_1}+\alpha^2t_1^2-2\alpha t_1e^{\alpha t_1}-2)\right.\right.$$
$$\left.\left.-c_1t_1^2-c_2(1+t_1)^2\right\}\frac{log|1+\tau|}{\tau}\right] \quad (29)$$

$J(\widetilde{Q}) = \frac{1}{\tau}\int_{T=0}^{\tau}\frac{1}{4}[Q_{11}+2Q_{21}+Q_{31}]dT = \frac{1}{4\tau}\int_{T=0}^{\tau}[Q_{11}+2Q_{21}+Q_{31}]dT$

$J(\widetilde{Q}) = \frac{mr^n}{4\tau}\left\{2\tau^2+(\sigma-\rho)(\tau-log|1+\tau|)\right\}$

$J(\widetilde{R}) = \frac{1}{\tau}\int_{T=0}^{\tau}\frac{1}{4}[R_{11}+2R_{21}+R_{31}]dT = \frac{1}{4\tau}\int_{T=0}^{\tau}[R_{11}+2R_{21}+R_{31}]dT$

$J(\widetilde{R}) = \frac{1}{4\tau}\int_{T=0}^{\tau}\left\{4(T-t_1)+\frac{(\sigma-\rho)}{1+T}(T-t_1)\right\}mr^ndT$

$$J(\widetilde{R}) = \frac{mr^n}{4\tau}\left[2(\tau^2-2t_1\tau)+(\sigma-\rho)\left\{\tau-(1+t_1)log|1+\tau|\right\}\right] \quad (30)$$

Analysis of stability and specific cases:

(i) if $(\sigma-\rho)\to 0$ then $J(\widetilde{TIC}) \to \frac{pmr^n}{2\alpha^2}(2e^{\alpha t_1}+\alpha^2t_1^2-2\alpha t_1e^{\alpha t_1}-2)+\frac{mr^nlog\left(\frac{\tau}{\epsilon}\right)}{2\tau}\left\{\frac{2A}{mr^n}+(c_1+c_2)t_1^2\right\}+\frac{c_2mr^n}{4}(\tau-4t_1), J(\widetilde{Q})\to\frac{mr^n\tau}{2}$ and $J(\widetilde{R})\to\frac{mr^n}{2\tau}\left[(\tau^2-2t_1\tau)\right]$

(ii) If $\sigma\to 0, \rho\to 0$ then the model reduces (i). But the case of the classical back-order EOQ model. So we choose ϵ so that the aforementioned becomes the classical model.Then we take

$$\frac{pmr^n}{2\alpha^2}(2e^{\alpha t_1}+\alpha^2t_1^2-2\alpha t_1e^{\alpha t_1}-2)+\frac{mr^nlog\left(\frac{\tau}{\epsilon}\right)}{2\tau}\left\{\frac{2A}{mr^n}+(c_1+c_2)t_1^2\right\}+\frac{c_2mr^n}{4}(\tau-4t_1)$$
$$=\left[\frac{p}{2\alpha^2}(2e^{\alpha t_1}+\alpha^2t_1^2-2\alpha t_1e^{\alpha t_1}-2)+\frac{log\left(\frac{\tau}{\epsilon}\right)}{2\tau}(c_1+c_2)t_1^2+\frac{c_2}{4}(\tau-4t_1)\right]mr^n+\frac{log\left(\frac{A\tau}{\epsilon}\right)}{\tau}$$

Comparing we get, $log\left(\frac{\tau}{\epsilon}\right)=\frac{1}{T}$ and $\frac{log\left(\frac{\tau}{\epsilon}\right)}{\tau}t_1^2+\frac{(\tau-4t_1)}{4}=(T-t_1)^2$ giving $\tau=2$, and $T=1\Rightarrow\epsilon\to 2e^{-2}\ll 1$.

Hence, model is stable, and the case, we write, $\tau=2T$, and hence, $J(\widetilde{Q})\to mr^nT$ and $J(\widetilde{R})\to mr^nT(T-2t_1)$, which corresponds to the results of the classical back-order model.

6. Numerical Illustration

Crisp Model: Let us consider $A = \$30$ per cycle, $h = \$5$ per unit per year, $p = \$1.6$ per unit per year, $q = \$3.6$ per unit per year, $n = 0.8$, $\alpha = 0.2$, $m = 4.8$ i.e. $D = mtr^n = 550$.
Fuzzy Model: Let us demand rate $\langle D_1, D_2, D_3\rangle = \langle m_1 tr_1^n, m_2 tr_2^n, m_3 tr_3^n\rangle = \langle 500, 550, 650\rangle$ units keeping the other parameter same as crisp.
Cloudy Fuzzy Model: Let us take $\sigma = 0.16$, $\rho = 0.14$, $\epsilon = 0.001$. For the model's numerical illustration.

Take in mind that we determine the model of The following formula can be used to determine the fuzzy demand rate, which is the computation of fuzziness; mean of $\langle 500, 550, 650\rangle = 566.67$, median =550, so mode(m) = 3 $\times$ median - 2$\times$ mean = 3 $\times$ 550- 2 $\times$566.67 = 516.66. Since fuzzy demand components L_b and U_b are the lower bounds and upper bounds, respectively, Then apply degree of fuzziness $d_f = \frac{U_b - L_b}{2g}$ and cloud index $CI = \frac{(1+T)}{T}$.

6.1. Numerical Discussion

Table 2 symbolizes, in a crisp environment, for the 1 month or 30 days (approximate) cycle time with 3.6 days During a shortage, the average optimal inventory cost is assumed to be the value \$18078.52 per cycle, but The fuzzy solution costs more, preserving the total average value of \$19225.62 per cycle. Moreover, it is astonishing that Every time we envision a cloudy, fuzzy environment, the average inventory cost decreases to \$14122.59 per cycle time simply by increasing the of the cycle time to 38.70 days alone. Additionally, we note that by this point, there is less fuzziness from 0.536 to 0.213.

Table 2 Optimal Solution of Model

Model	Cycle time T* (days)	Inventory period t_1^*	Shortage quantity R* units	Ordering quantity A* units	Minimum cost TIC*(\$)	$d_f = \frac{U_b - L_b}{2g}$	$CI = \frac{(1+T)}{T}$
Crisp	14.28	10.92	207.38	708.39	18078.52	—	—
Fuzzy	14.16	10.92	213.26	786.77	19225.62	0.536	—
Clody Fuzzy	36.70	31.68	786.52	1032.56	14122.59	—	0.213

Tables 3 and 4 show that, during the cycle time duration of 33 to 41 days, the average inventory cost is given by both the crisp and general fuzzy models, and both show a rise with cycle time duration; however, the cloudy fuzzy model's objective function assumes a U-shaped curve ™ at 20 days cycle time keeping the minimum possible average inventory cost globally. Furthermore, our observation shows that the hazy fuzzy model's solution is not viable at cycle times shorter than 33 days. In comparison to the results produced in both a crisp and a fuzzy environment, the order quantity and shortage quantity increase in the cloudy fuzzy model. Nonetheless, we can observe that they are naturally increasing in all situations if we look at the trend values of the order quantity and shortage quantity that they take.

Table 3 Objective values under various cycle times

Cycle time	Crisp model				Fuzzy model			
T (days)	t_1^*	R*	q*	TIC**	t_1^*	R*	q*	TIC**
33	27.23	412.56	2186.52	18318.73	27.23	424.29	2245.38	19536.66
34	28.65	445.13	2331.10	18386.04	28.65	456.64	456.64	19605.42
35	30.07	477.70	2475.68	18453.35	30.07	488.99	2532.86	19674.18
36	31.49	510.27	2620.26	18520.66	31.49	521.34	2676.60	19742.94
37	32.91	542.84	2764.84	18587.97	32.91	553.69	2820.34	19811.70
38	34.33	575.41	2909.42	18655.28	34.33	586.04	2964.08	19880.46
39	35.75	607.98	3054.00	18722.59	35.75	618.39	3107.82	19949.22
40	37.17	640.55	3198.58	18789.90	37.17	650.74	3251.56	20017.98

Table 4 Cloudy fuzzy values under several cycle times

Cycle time	Crisp model			
T (days)	t_1^*	R*	q*	TIC**
33	31.28	55.61	1123.54	14307.62
34	31.37	193.39	1194.96	14251.64
35	31.46	331.17	1266.38	14216.28
36	31.55	468.95	1337.80	14208.96
37	**31.64**	**606.73**	**1409.22**	**14207.46**
38	31.73	744.51	1480.64	14218.66
39	31.82	882.29	1552.06	14243.50
40	31.91	1020.07	1623.48	14277.98

7. Sensitivity analysis

In this section, sensitivity analysis is carried out to examine the impact of parameter value changes on the optimal values.To conduct these experiments, one parameter was modified by $\pm$10% and $\pm$20% at a time, while the other parameters were left at their initial values.The results of the sensitivity analysis are shown in Table 5 below.

Table 5 Cloudy Fuzzy Model Sensitivity Analysis

Parameters	Change %	Cycle Time (days) T*	Inventory Period (days) t_1^*	Shortage quantity R*	Order quantity A*	Average Total Cost TIC*	$\frac{(TIC^*-TIC_*)}{TIC_*}100\%$
p	+20	18.71	14.37	335.74	618.81	10536.57	-18.36
	+10	18.65	14.76	383.28	635.76	10416.36	-20.01
	-10	18.39	15.74	406.71	652.37	10308.71	-21.79
	-20	18.33	15.86	448.35	666.71	10216.73	-22.86
h	+20	19.23	15.16	383.57	693.58	10432.61	-19.53
	+10	19.19	15.44	403.82	694.39	10316.79	-20.20
	-10	19.47	16.21	453.36	698.81	10198.71	-22.46
	-20	19.43	16.56	476.89	701.16	10035.16	-23.31
q	+20	16.12	12.71	326.92	597.47	9736.00	-24.76
	+10	17.25	13.69	360.10	647.38	9854.39	-23.53
	-10	23.33	18.57	501.67	855.43	10516.47	-17.77
	-20	27.53	21.51	559.76	995.36	11009.58	-14.83
α	+20	...	...	No	Feasible	...	...
	+10	19.50	18.16	95.23	698.39	13266.71	+0.91
	-10	19.64	12.79	213.36	701.72	7735.46	-41.36
	-20	19.77	10.58	55.11	706.36	5866.73	-54.07
$D = mr^n$	+20	17.56	15.38	221.77	943.46	**15000.61**	+14.65
	+10	18.22	15.38	241.43	851.59	13124.67	+0.28
	-10	21.35	19.76	247.56	533.01	7456.43	-43.13
	-20	24.13	16.65	133.55	418.05	**5539.05**	-57.79
A	+20	22.35	15.86	335.93	718.28	10618.71	-18.76
	+10	21.31	15.79	368.21	687.29	10463.72	-19.54
	-10	18.01	15.64	471.78	593.91	10016.75	-23.37
	-20	16.79	15.57	77.74	543.17	9135.36	-24.81

Parameters	Change %	Cycle Time (days) T*	Inventory Period (days) t_1^*	Shortage quantity R*	Order quantity A*	Average Total Cost TIC*	$\frac{(TIC^*-TIC_*)}{TIC_*}100\%$
ϵ	+20	18.93	15.58	446.76	671.36	10118.91	-22.05
	+10	19.21	15.58	438.87	681.72	10148.63	-21.67
	-10	20.12	15.54	401.77	712.35	10280.25	-20.65
	-20	20.53	15.54	386.13	730.14	10348.72	-20.43
ρ	+20	...	...	No	Feasible	...	...
	+10	...	...	No	Feasible	...	...
	-10	20.76	15.79	381.72	746.56	10450.73	-19.63
	-20	21.53	15.79	358.26	776.18	10605.32	-18.57
σ	+20	21.86	18.63	358.41	788.56	10756.43	-18.36
	+10	20.97	15.84	385.65	755.93	10554.37	-19.84
	-10	...	...	No	Feasible	...	...
	-20	...	...	No	Feasible	...	...

Here are a few insights from the sensitivity analysis findings.

1. If the deterioration cost (p) increases, the total average inventory cost (TIC) rises increases due to the shortage quantity and order quantity r decrease at the constant rate. At the same time, both the cycle length and the product reliability increasing.
2. If the holding cost (h) increase, then total average inventory cost (TIC) increase at constant rate due to the shortage quantity, and order quantity decrease at constant rate. At the sametime, the cycle length decreases.
3. If the parameters q are increased, the total average inventory cost (TIC) decrease at constant rate due to the shortage quantity and order quantity rapidly decrease at constant rate. while the product's reliability also decrease.
4. If the deterioration rate (α) increases, the total average inventory cost (TIC) rises highly rapidly increases due to the shortage quantity Oscillate and order quantity minor decrease. At the same time, the cycle length decreased and the product reliability increasing.
5. If the parameters m and n are increased in demand rate, the total average inventory cost (TIC) rises quickly due to the shortage quantity oscillation and order quantity increase. At the same time, the cycle length decreased and the product reliability increasing.
6. If the ordering cost A are increased in demand rate, the total average inventory cost (TIC) smoothly increasing at constant rate due to the shortage quantity Oscillate and order quantity increase. At the same time, both the cycle length and the product reliability increasing.

8. Graphical Representations of the Model

We will create the cloudy fuzzy model's graphs in order to more clearly justify the recently proposed approach. Figure 4 illustrates that there is a significant disparity in the average inventory expenses of the crisp and fuzzy models in comparison to the cloudy fuzzy model.Furthermore, we've noticed that, everywhere, the fuzzy model yields the highest value of the objective function, while the cloudy fuzzy model yields the lowest value.Therefore, the inventory practitioner, especially the decision maker (DM) should choose the solution in a fuzzy, unclear the environment. The cloudy fuzzy objective function's "U-tern," which appears at 37 days of cycle time and is hence convex, is depicted in Figure 5. Figures 6, 7, and 8 illustrate that the back-order quantity

curves for all three model scenarios intersect at around 38 days of cycle time, resulting in a shortage quantity of approximately 600 units of the model. For crisp and fuzzy models, the back-order curve is probably made up of overlapped lines with lesser gradients, while for hazy fuzzy models, it is a straight line that constantly gets greater gradient values.

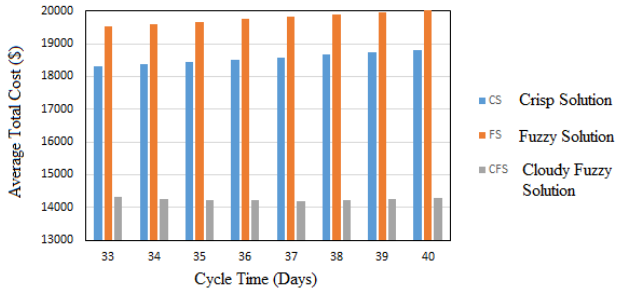


Figure 4: *Total Average cost and Cycle time for all three models*

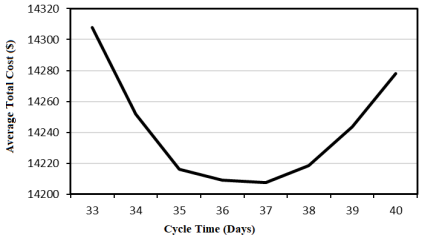


Figure 5: *Total cost variation over cloudy fuzzy model*

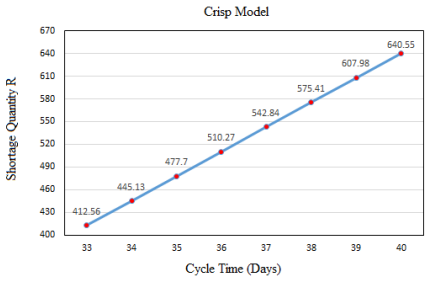


Figure 6: *Shortage Level of Crisp model*

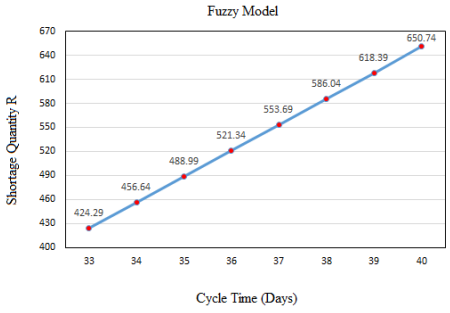


Figure 7: *Shortage Level of Fuzzy model*

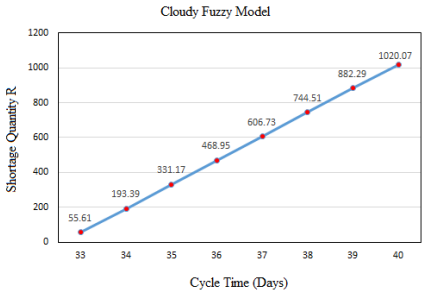


Figure 8: *Shortage Level of Cloudy Fuzzy model*

9. Conclusion

We have covered a back order EOQ model in a cloudy, fuzzy environment in this article. All inventory models are examined using crisp models, fuzzy models, intuitionistic fuzzy models, or fuzzy stochastic environments in the literature. However, the idea of "cloudy fuzzy" in decision-making scenarios is relatively recent. Nonetheless, we observe that the degree of fuzziness and the cloud index decrease as the cycle duration assumes greater values. Lower inventory costs do not equate to less fuzziness. Because, at its optimal, the fuzzy parameters may eventually begin to converge toward a crisp number with decreasing fuzziness. We propose to investigate

the model in a cloudy fuzzy environment by introducing parametric flexibility because the crisp minimization issue yields higher values and is impractical in practice. Therefore, the average inventory cost is likely to converge with the cost derived just from the crisp model if we increase the cycle duration beyond its global optimum. Thus, we are searching the model minimum, which is the primary focus of the model, amid a significant degree of fuzziness.As a result, it is quite simple for any DM to comprehend and decide appropriately.

References

[1] Chare, P, & Schrader, G. (1963). A model for exponentially decaying inventories. Journal of industrial engineering, 15, 238-243.

[2] Chang, H. J., & Dye, C. Y. (1999). An EOQ model for deteriorating items with time varying demand and partial backlogging. Journal of the Operational Research Society, 50(11), 1176-1182.

[3] Lo, S. T., Wee, H. M., & Huang, W. C. (2007). An integrated production-inventory model with imperfect production processes and Weibull distribution deterioration under inflation. International Journal of Production Economics, 106(1), 248-260.

[4] Mahata, G. C., & Mahata, P. (2011). Analysis of a fuzzy economic order quantity model for deteriorating items under retailer partial trade credit financing in a supply chain. Mathematical and Computer Modelling, 53(9-10), 1621-1636.

[5] Kumar, P., & Dutta, D. (2020). A deteriorating inventory model with uniformly distributed random demand and completely backlogged shortages. In Numerical Optimization in Engineering and Sciences: Select Proceedings of NOIEAS 2019 (pp. 213-223). Springer Singapore.

[6] Shah, N. H. (1993). Probabilistic time-scheduling model for an exponentially decaying inventory when delays in payments are permissible. International Journal of Production Economics, 32(1), 77-82.6

[7] Shah, N. H., & Shah, Y. K. (1998). A discrete-in-time probabilistic inventory model for deteriorating items under conditions of permissible delay in payments. International Journal of Systems Science, 29(2), 121-125.

[8] Shah, N. H. (2004). Probabilistic Order Level system when items in inventory deteriorate and delay in payments is permissible. Asia-Pacific Journal of Operational Research, 21(03), 319-331.

[9] De, L. N., & Goswami, A. (2009). Probabilistic EOQ model for deteriorating items under trade credit financing. International Journal of Systems Science, 40(4), 335-346.

[10] Khedlekar, U. K., Kumar, L., & Keswani, M. (2023). A stochastic inventory model with price-sensitive demand, restricted shortage, and promotional efforts. Yugoslav Journal of Operations Research, 33(4), 613-642.

[11] AlDurgam, M., Adegbola, K., & Glock, C. H. (2017). A single-vendor single-manufacturer integrated inventory model with stochastic demand and variable production rate. International Journal of Production Economics, 191, 335-350.

[12] Al Masud, M. A., Paul, S. K., & Azeem, A. (2014). Optimisation of a production inventory model with reliability considerations. International journal of logistics systems and management, 17(1), 22-45.

[13] Krishnamoorthi, C., & Panayappan, S. (2013). An EPQ model for an imperfect production system with rework and shortages. International Journal of Operational Research, 17(1), 104-124.

[14] Sarkar, B., Sana, S. S., & Chaudhuri, K. (2010). Optimal reliability, production lot size, and safety stock in an imperfect production system. International Journal of Mathematics in Operational Research, 2(4), 467-490.

[15] Mahapatra, G. S., Adak, S., Mandal, T. K., & Pal, S. (2017). Inventory model for deteriorating items with time and reliability dependent demand and partial backorder. International Journal of Operational Research, 29(3), 344-359.

[16] Bag, S., Chakraborty, D., & Roy, A. R. (2009). A production inventory model with fuzzy random demand and with flexibility and reliability considerations. Computers & Industrial Engineering, 56(1), 411-416.
[17] Sarkar, B., Sana, S. S., & Chaudhuri, K. (2010). Optimal reliability, production lot size and safety stock: An economic manufacturing quantity model. International Journal of Management Science and Engineering Management, 5(3), 192-202.
[18] Tripathy, P. K., & Pattnaik, M. (2011). Optimal inventory policy with reliability consideration and instantaneous receipt under imperfect production processes. International Journal of Management Science and Engineering Management, 6(6), 413-420.
[19] Paul, S. K., Azeem, A., Sarker, R., & Essam, D. (2014). Development of a production inventory model with uncertainty and reliability considerations. Optimization and Engineering, 15, 697-720.
[20] Bhunia, A. K., Shaikh, A. A., & Crdenas-Barrn, L. E. (2017). A partially integrated production-inventory model with interval valued inventory costs, variable demand, and flexible reliability. Applied Soft Computing, 55, 491-502.
[21] Khara, B., Dey, J. K., & Mondal, S. K. (2017). An inventory model under development cost-dependent imperfect production and reliability-dependent demand. Journal of Management Analytics, 4(3), 258-275.
[22] Abdel-Aleem, A., El-Sharief, M. A., Hassan, M. A., & ElSebaie, M. G. (2018). Optimization of reliability based model for production inventory system. International Journal of Management Science and Engineering Management, 13(1), 54-64.
[23] Rathore, H. (2019). An inventory model with advertisement dependent demand and reliability consideration. International Journal of Applied and computational mathematics, 5(2), 33.
[24] Yadav, A. S., Kumar, A., Yadav, K. K., & Rathee, S. (2023). Optimization of an inventory model for deteriorating items with both selling price and time-sensitive demand and carbon emission under green technology investment. International Journal on Interactive Design and Manufacturing (IJIDeM), 1-17.
[25] Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.
[26] De, S. K., & Mahata, G. C. (2017). Decision of a fuzzy inventory with fuzzy backorder model under cloudy fuzzy demand rate. International Journal of Applied and computational mathematics, 3, 2593-2609.
[27] De, S. K., & Mahata, G. C. (2019). A cloudy fuzzy economic order quantity model for imperfect-quality items with allowable proportionate discounts. Journal of Industrial Engineering International, 15(4), 571-583.
[28] Shah, N. H., & Patel, M. B. (2021). Reducing the deterioration rate of inventory through preservation technology investment under fuzzy and cloud fuzzy environment. In Predictive analytics (pp. 65-80). CRC Press.
[29] Padiyar, S. S., Bhagat, N., Singh, S. R., Punetha, N., & Dem, H. (2023). Production policy for an integrated inventory system under cloudy fuzzy environment. International Journal of Applied Decision Sciences, 16(3), 255-299.
[30] Chang, H. J., & Dye, C. Y. (1999). An EOQ model for deteriorating items with time varying demand and partial backlogging. Journal of the Operational Research Society, 50(11), 1176-1182.
[31] Khanra, S., Ghosh, S. K., & Chaudhuri, K. S. (2011). An EOQ model for a deteriorating item with time dependent quadratic demand under permissible delay in payment. Applied mathematics and computation, 218(1), 1-9.
[32] Sett, B. K., Sarkar, B., & Goswami, A. (2012). A two-warehouse inventory model with increasing demand and time varying deterioration. Scientia Iranica, 19(6), 1969-1977.
[33] Chanda, U., & Kumar, A. (2017). Optimisation of fuzzy EOQ model for advertising and price sensitive demand model under dynamic ceiling on potential adoption. International Journal of Systems Science: Operations & Logistics, 4(2), 145-165.

[34] Pal, S., & Mahapatra, G. S. (2017). A manufacturing-oriented supply chain model for imperfect quality with inspection errors, stochastic demand under rework and shortages. Computers & Industrial Engineering, 106, 299-314.

[35] Goli, A., Zare, H. K., TavakkoliMoghaddam, R., & Sadegheih, A. (2020). Multiobjective fuzzy mathematical model for a financially constrained closedloop supply chain with labor employment. Computational Intelligence, 36(1), 4-34.

[36] Goli, A., Zare, H. K., Tavakkoli-Moghaddam, R., & Sadeghieh, A. (2019). Hybrid artificial intelligence and robust optimization for a multi-objective product portfolio problem Case study: The dairy products industry. Computers & industrial engineering, 137, 106090.

[37] Shabani, S., Mirzazadeh, A., & Sharifi, E. (2016). A two-warehouse inventory model with fuzzy deterioration rate and fuzzy demand rate under conditionally permissible delay in payment. Journal of Industrial and Production Engineering, 33(2), 134-142.

[38] Barzegar, S., Seifbarghy, M., Pasandideh, S. H., & Arjmand, M. (2016). Development of a joint economic lot size model with stochastic demand within non-equal shipments. Scientia Iranica, 23(6), 3026-3034. Logistics, 5(1), 60-68.

[39] Modak, N. M., & Kelle, P. (2019). Managing a dual-channel supply chain under price and delivery-time dependent stochastic demand. European Journal of Operational Research, 272(1), 147-161.

[40] Shah, N. H., & Vaghela, C. R. (2018). Imperfect production inventory model for time and effort dependent demand under inflation and maximum reliability. International Journal of Systems Science: Operations & Logistics, 5(1), 60-68.

[41] Barman, H., Pervin, M., Roy, S. K., & Weber, G. W. (2021). Back-ordered inventory model with inflation in a cloudy-fuzzy environment. Journal of Industrial & Management Optimization, 17(4), 1913-1941.

[42] San-Jos©, L. A., Sicilia, J., & Garca-Laguna, J. (2015). Analysis of an EOQ inventory model with partial backordering and non-linear unit holding cost. Omega, 54, 147-157.

[43] San-Jos©, L. A., Sicilia, J., Crdenas-Barrn, L. E., & Guti©rrez, J. M. (2019). Optimal price and quantity under power demand pattern and non-linear holding cost. Computers & Industrial Engineering, 129, 426-434.

[44] Alfares, H. K., & Ghaithan, A. M. (2016). Inventory and pricing model with price-dependent demand, time-varying holding cost, and quantity discounts. Computers & Industrial Engineering, 94, 170-177.

[45] Pervin, M., Roy, S. K., & Weber, G. W. (2018). Analysis of inventory control model with shortage under time-dependent demand and time-varying holding cost, including stochastic deterioration. Annals of Operations Research, 260, 437-460.

[46] Pervin, M., Roy, S. K., & Weber, G. W. (2017). A Two-echelon inventory model with stock-dependent demand and variable holding cost for deteriorating items. Numerical Algebra, Control and Optimization, 7(1), 21-50.

[47] Yadav, K. K., Yadav, A. S., & Bansal, S. (2024). Interval number approach for two-warehouse inventory management of deteriorating items with preservation technology investment using analytical optimization methods. International Journal on Interactive Design and Manufacturing (IJIDeM), 1-17.

[48] Krishan Kumar Yadav , Ajay Singh Yadav, Shikha Bansal (2024). OPTIMIZATION OF AN INVENTORY MODEL FOR DETERIORATING ITEMS ASSUMING DETERIORATION DURING CARRYING WITH TWO-WAREHOUSE FACILITY. Reliability: Theory Applications, 19 (3 (79)), 442-459. doi: 10.24412/1932-2321-2024-379-442-459.

[49] Dutta, A., Negi, A., Harshith, J., Selvapandian, D., Raj, A. S. A., Patel, P. R. (2023, April). Evaluation modelling of asteroids™ hazardousness using chaosNet. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

[50] Yager, R. R. (1981). A procedure for ordering fuzzy subsets of the unit interval. Information sciences, 24(2), 143-161.

[51] De, S. K., & Beg, I. (2016). Triangular dense fuzzy sets and new defuzzification methods. Journal of Intelligent & Fuzzy systems, 31(1), 469-477.

OPTIMIZATION OF A TWO-WAREHOUSE INVENTORY MANAGEMENT FOR DETERIORATING ITEMS WITH TIME AND RELIABILITY-DEPENDENT DEMAND UNDER CARBON EMISSION CONSTRAINTS

Krishan Kumar Yadav[1], Ajay Singh Yadav[1,*], Shikha Bansal[1]

•

[1]Department of Mathematics, SRM Institute of Science and Technology,
Delhi-NCR Campus, Ghaziabad, India, 201204
ky3686@srmist.edu.in, ajaysiny@srmist.edu.in, shikhab@srmist.edu.in
[*]Corresponding author

Abstract

The main objective of this study is to demonstrate how a company's inventory management can be significantly impacted by its ability to provide reliable, high-quality products and to balance stock availability in order to maintain customer satisfaction. Such measures can ultimately lead to an increase in a company's market share, efficiency, and profitability. In order to analyze the impact of reliability and time-based demand rate on inventory management system, an economic order quantity (EOQ) model with two-warehouse is established. Complete backlog allows for the consequences of constant degradation and shortages. The holding and degradation costs are considered while analyzing the effect of carbon emissions. This study's primary goal is to optimize overall cost while maintaining item reliability and total cycle time. Analytical optimization is used to yield an algorithm for the inventory model that determines the optimal output. A numerical example-based sensitivity analysis using MATLAB Software version R2021b is also presented to illustrate the effect of carbon emission and validation of the model.

Keywords: Two-Warehouse, Inventory Management, Deteriorating Items, Time and Reliability-Dependent Demand, Carbon Emission Constraints, Optimization.

1. Introduction and literature survey

Inventory management is a crucial part of supply chain operations because it ensures that the right amount of items are available when needed to meet demand. However, managing inventories of perishables, chemicals, or other items that deteriorate over time presents unique challenges. Traditional inventory models typically prioritize ordering, holding, and shortfall charges in an effort to keep costs as low as possible. Unfortunately, critical components such as the environmental impact of carbon emissions and the reliability of the supply chain are often overlooked by these models. The need to shift inventory management practices toward more environmentally friendly ones stems from the growing emphasis on sustainability. The carbon emissions resulting from transportation, storage, and other logistical activities significantly augment a business's environmental footprint. Thus, in order to achieve both environmental and economic objectives, carbon emission constraints must now be incorporated into inventory models. Furthermore, the demand for degrading commodities is sometimes influenced by time and the reliability of the inventory system. While time-dependent demand reflects the seasonal or

cyclical character of particular commodities, reliability-dependent demand considers how supply chain performance influences demand and consumer satisfaction. When these components are integrated, inventory models become more realistic and useful.

Two-warehouse inventory models have been extensively studied in the literature, with a particular emphasis on cost optimization and operational performance. The purpose behind this concept is to manage inventory between a primary warehouse (central storage) and a secondary warehouse (regional or local storage). Several significant studies in this sector include: In order to lower holding and transportation costs, Hartley [7] addressed the distribution of stock between two warehouses and developed one of the first two-warehouse inventory models. Demand-dependent costs were added to the two-warehouse model by Banerjee and Gupta [1], providing a foundation for more adaptable inventory control. Jha and Shanker [9] looked at partial backordering in a two-warehouse system and emphasized the trade-offs between holding costs and stockout concerns. These studies laid the groundwork for two-warehouse inventory management, but they neglected to account for item deterioration and environmental influences. For the review, we can assess the research conducted by Yadav, K.K., et al. [26] and Debnath, B.K., et al. [30].

Degrading inventory models address the issues caused by items that eventually become worthless or useless. The primary goal of this research has been to devise strategies to mitigate the impact of degradation on inventory costs. Among the significant inputs are: Ghare and Schrader [5] initially introduced an exponential decay model for inventory systems, providing a theoretical framework for handling deteriorating items. Wee [13] proposed an inventory model with lost sales and partial backordering for degrading items with time-dependent demand. Bakker et al. [2] stressed the need of including degradation rates in procedures for making decisions regarding inventory after analyzing a number of models for degrading inventory. Yadav et al.'s [27] study provides significant insights into the use of optimization algorithms for managing supply chains during the COVID-19 pandemic, proposing a strategic approach to tackling logistical issues. Baloni et al.'s [31] study advances mathematical modeling in oncology by integrating time delays into homogeneous tumor models, offering essential insights to improve predictions of tumor growth and treatment methods. Despite the extensive research on the topic, few studies have combined degraded inventory models with two-warehouse systems and environmental constraints.

Two significant components of demand variability"time- and reliability-dependent demand"are an essential part of inventory management. Time-dependent demand models provide for cyclical, trend-based, and seasonal variations in demand: Silver et al. [12] proposed many inventory control models that account for time-varying demand patterns and provide practical guidance on how to manage these variations. Hill and Omar [8] developed an economic order quantity (EOQ) model for time-dependent demand that highlights the importance of adaptive inventory techniques. When evaluating client demand, demand models that depend on supply chain reliability take this into consideration: Kumar and Uthayakumar [10] investigated inventory systems with reliability-dependent demand, demonstrating how supply chain performance impacts customer ordering decisions. Roy et al. [11] integrated reliability into an inventory model for degraded items to gain insight into the relationship between demand and dependability in supply chain management. However, there is a lack of models in the academic literature that address both time- and reliability-dependent demand simultaneously in a two-warehouse scenario. They previously developed a mathematical inventory model [14-23] that incorporates a demand rate influenced by reliability.

Carbon emissions have only lately been taken into account in inventory management due to growing environmental concerns and legal pressure. Notable contributions include: A framework emphasizing the significance of sustainable practices was developed by Chaabane et al. [4] in order to integrate carbon footprint assessment into supply chain planning. Benjaafar et al. [3] used inventory models that take carbon emissions into account to draw attention to the trade-offs between environmental and economic objectives. Reducing carbon emissions is an essential

component of sustainable supply chain strategies, according to Govindan et al. [6], who studied green supply chain management methodologies. While these studies provide valuable insights into the environmental aspects of supply chain management, they don't specifically address the challenges posed by two-warehouse systems or deteriorating commodities. Several significant and previously published studies [24-25] and [32] address the impact of carbon emissions.

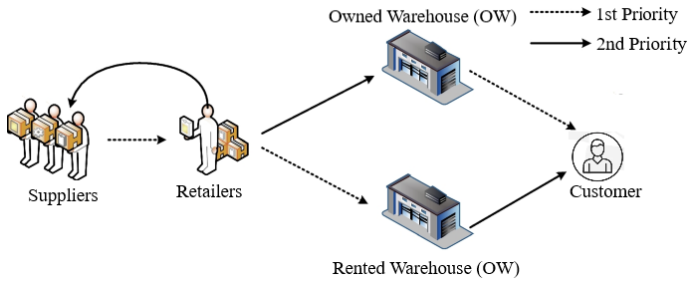


Figure 1: *Sample system framework for managing inventory across two warehouses*

2. Research gap and our contributions

According to the currently published literature, many research studies have been conducted using analytical inventory optimization. They first created a mathematical inventory model [7]. The created problem was then solved in the solution phase using the analytical optimization approach. However, in this study, the inventory model with two-warehouses in a different way. The proposed problem was solved using analytical optimization approach. There are not many studies that use analytical optimization approach for inventory costs. Consequently, Fig. 1 represents the sample system model of the two-warehouse inventory model similar to [33]. The following lists our research's principal contributions:

1. The proposed model is developed considering the impact of carbon emission.
2. When creating the proposed model, inventory criteria such as total cost, and reliability and time-based demand are considered.
3. To manage the inventory in effective manner, then considering the two-warehouses facilities.

The proposed model aims to minimize the whole cost, which includes holding, ordering, degradation, and carbon emission charges. By integrating these elements, the model seeks to provide a comprehensive and durable approach to inventory management for deteriorating items.

Table 1: *The comparison of our current work with previously published work*

Source	Demand	Deterioration	Two-Warehouse	Shortage	Carbon Emission	Optimization Techniques
Kumar, S., and Uthayakumar, R. [10]	Reliability dependent	Yes	No	Yes	No	Analytical
Adak, S., and Mahapatra, G. S. [15]	Time, reliability and advertisement dependent	Yes	No	No	Yes	Analytical
Mahapatra, G. S., et al. [16]	Price, stock, reliability and advertisement dependent	Yes	No	Yes	No	Analytical
Adak, S., and Mahapatra, G. S. [18]	Time and reliability dependent	Yes	No	Yes	No	Analytical
Rajput, N., et al. [21]	Reliability dependent	Yes	No	Yes	No	Analytical
Mahapatra, G. S., et al. [23]	Time and reliability dependent	Yes	No	Yes	No	Analytical
Kumar, K. [24]	Stock dependent	Yes	Yes	Yes	Yes	Analytical
Sharma, A., et al. [32]	Time dependent	Yes	Yes	Yes	Yes	Analytical
Present paper	Time and reliability dependent	Yes	Yes	Yes	Yes	Analytical

3. Presumptions and notations

3.1. Presumptions

The following presumptions were used in the formulation of the mathematical model.

1. The function of demand rate is both time and product reliability dependent, which is $f(t,r) = atr^b$, where $a, b > 0$ are constants.

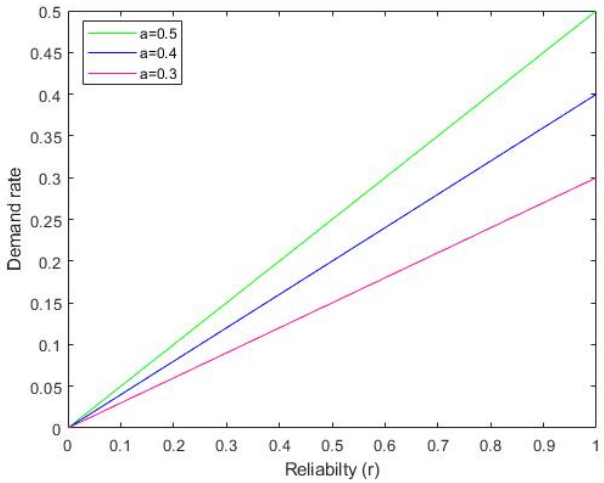


Figure 2: *Correlation between demand rate and product reliability*

2. Shortages is fully backlogged.
3. The filling rate is infinite, and the lead time is zero.
4. The time horizon of the inventory system is infinite.
5. While the rental warehouse (RW) has an infinite capacity, the closely-held warehouse (OW) has a mounted capability of W units.
6. The products of OW are consumed solely once intense, the products unbroken in RW.
7. he rate of deterioration is constant, which is k; $k > 0$.

3.2. Notations

Table 2 is provided a description of the notations utilised for the constructed mathematical model.

Table 2: *Notations*

Notation	Units	Description
a	Constant	Coefficient of demand function
b	Constant	Coefficient of demand function
k	Constant	Coefficient of deterioration rate.
C_s	\$/unit	Shortage cost.
U	capability constraint	The owned warehouse capacity
A	\$/unit	The ordering cost
C_e	\$/unit	Carbon emissions cost for holding items
$I_r(t)$	Units	Rented warehouse Inventory (Stock) level at a time t.
$I_o(t)$	Units	Owned warehouse Inventory (Stock) level at a time t.
$I_s(t)$	Units	Stock out Inventory (Stock) level at a time t.
C_d	\$/Units	Deterioration cost.
h	\$/Units	Holding cost.
Q	Units	The number of orders placed in each cycle.
$TC(r,T)$	\$/Units	The function for total inventory cost.

Table 3: *Decision-makingparameters*

Notation	Units	Description
r	–	Reliability of item of the inventory system.
T	Years	Length of the cycle.

4. Mathematical Model Formulation

In the starting, Q units of deteriorated goods were ordered. Thus, Q it represents the inventory quantity at time zero. At the time interval $t \in [0, t_1]$, the joint effect of deterioration and demand decreases the inventory level in RW to drop until it reaches zero. Also, at the same time interval $t \in [0, t_1]$, only the deterioration effect decreases OW's inventory level. Again, at the time interval $t \in [t_1, t_2]$, the joint impact of deterioration and demand reduces OW's inventory (stock) level to drop until it reaches zero. The period of shortage is $t \in [t_2, T]$ (See Figure 3).

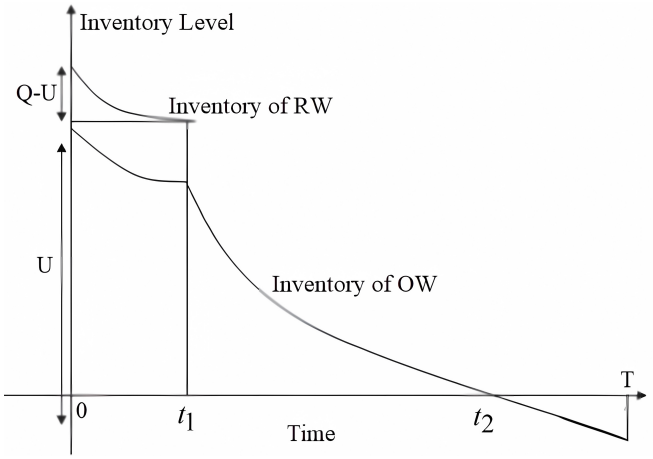


Figure 3: *A graphical depiction of the inventory management system involving two warehouses with complete backlogging.*

The stock level at $t = 0$ to $t = T$ is characterised in the differential equations as follows:

$$\frac{dI_r(t)}{dt} + kI_r(t) = -f(t,r); \quad t \in [0, t_1] \tag{1}$$

with the boundary conditions (B.C.) $I_r(t_1) = 0$ and $I_r(0) = Q - U$.

$$\frac{dI_o(t)}{dt} + kI_o(t) = 0; \quad t \in [0, t_1] \tag{2}$$

with the boundary conditions (B.C.) $I_o(0) = U$.

$$\frac{dI_o(t)}{dt} + kI_o(t) = -f(t,r); \quad t \in [t_1, t_2] \tag{3}$$

with the boundary conditions (B.C.) $I_o(t_2) = 0$.

$$\frac{dI_s(t)}{dt} = -f(t,r); \quad t \in [t_2, T] \tag{4}$$

with the boundary conditions (B.C.) $I_s(t_2) = 0$.

The equations (5), (6), (7) and (8) are the solutions of equations (1), (2), (3) and (4), respectively.

Note that here only taking two terms of exponential functions and remaining are negligible to solve the above four equations (5)-(8).

$$I_r(t) = -\left[\frac{ar^b - ar^b(1-(kt^2))}{2k} + \frac{k^2t^4}{2k}\left(1 + \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right)\right] \tag{5}$$

$$I_o(t) = U\left[1 - \frac{kt^2}{2} + \frac{k^2t^4}{4}\right] \tag{6}$$

$$I_o(t) = -\frac{1}{k}\left[ar^b - ar^b\left(1 - \frac{kt^2}{2} + \frac{k^2t^4}{4}\right)\left(1 + \frac{kt_2^2}{2} + \frac{k^2t_2^4}{4}\right)\right] \tag{7}$$

$$I_s(t) = \frac{ar^b t_2^2}{2} - \frac{ar^b t^2}{2} \tag{8}$$

Using the continuity at $t = t_1$ in equations (6) and (7), we get (9)

$$U = -\frac{1}{k}\left[1 + \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right]\left[ar^b - ar^b\left(1 - \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right)\left(1 + \frac{kt_2^2}{2} + \frac{k^2t_2^4}{4}\right)\right] \tag{9}$$

Using the boundary condition $I_r(0) = Q - U$ in (5), we get (10)

$$Q = U - \frac{1}{k}\left[ar^b - ar^b\left(1 + \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right)\right] \tag{10}$$

From equations (9) and (10), we get (11)

$$Q = -\frac{1}{k}\left[1 + \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right]\left[ar^b - ar^b\left(1 - \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right)\left(1 + \frac{kt_2^2}{2} + \frac{k^2t_2^4}{4}\right)\right] - \frac{1}{k}\left[ar^b - ar^b\left(1 + \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right)\right] \tag{11}$$

The overall cost per cycle comprises the following components:

1. **Ordering cost per cycle :**

$$OC = A \tag{12}$$

2. **Deterioration cost per cycle:**

$$DC = C_d \cdot k\left[\int_0^{t_1} I_r(t)dt + \int_0^{t_1} I_o(t)dt + \int_{t_1}^{t_2} I_o(t)dt\right]$$

$$DC = -C_d \cdot k\left\{\frac{1}{240}\left\{ar^b(t_1 - t_2)\left(3k^3t_1^4t_2^4 + 3k^3t_1^3t_2^5 + 3k^3t_1^2t_2^6 + 3k^3t_1t_2^7 + 3k^3t_2^8 + 6k^2t_1^4t_2^2 + 6k^2t_1^3t_2^3 - 4k^2t_1^2t_2^4 - 4k^2t_1t_2^5 - 4k^2t_2^6 + 12kt_1^4 + 12kt_1^3t_2 - 8kt_1^2t_2^2 - 8kt_1t_2^3 + 52kt_2^4 - 40t_1^2 - 40t_1t_2 + 80t_2^2\right) + ar^bt^3(3k^3t_1^6 - 4k^2t_1^4 + 52kt_1^2 + 80)\right\} - \frac{t_1}{60k}\left[ar^b - ar^b\left(\frac{k^2t_1^4}{4} - \frac{kt_1^2}{2} + 1\right)\left(1 + \frac{kt_1^2}{2} + \frac{k^2t_2^4}{4}\right)\left(1 + \frac{kt_1^2}{2} + \frac{k^2t_1^4}{4}\right)(3k^2t_1^4 - 10kt_1^2 + 60)\right]\right\} \tag{13}$$

3. **Holding cost per cycle:**

$$HC = h\left[\int_0^{t_1} I_r(t)dt + \int_0^{t_1} I_o(t)dt + \int_{t_1}^{t_2} I_o(t)dt\right]$$

$$HC = -h\Bigg\{\frac{1}{240}\Bigg\{ar^b(t_1-t_2)\big(3k^3t_1^4t_2^4+3k^3t_1^3t_2^5+3k^3t_1^2t_2^6+3k^3t_1t_2^7+3k^3t_2^8+6k^2t_1^4t_2^2+6k^2t_1^3t_2^3$$
$$-4k^2t_1^2t_2^4-4k^2t_1t_2^5-4k^2t_2^6+12kt_1^4+12kt_1^3t_2-8kt_1^2t_2^2-8kt_1t_2^3+52kt_2^4-40t_1^2-40t_1t_2+80t_2^2\big)$$
$$+ar^bt^3(3k^3t_1^6-4k^2t_1^4+52kt_1^2+80)\Bigg\}-\frac{t_1}{60k}\Bigg[ar^b-ar^b\bigg(\frac{k^2t_1^4}{4}-\frac{kt_1^2}{2}+1\bigg)\bigg(1+\frac{kt_1^2}{2}+\frac{k^2t_2^4}{4}\bigg)$$
$$(3k^2t_1^4-10kt_1^2+60)\bigg(1+\frac{kt_1^2}{2}+\frac{k^2t_1^4}{4}\bigg)\Bigg]\Bigg\} \quad (14)$$

4. **Carbon Emission cost per cycle:**

$$CEC = C_e\Bigg[\int_0^{t_1} I_r(t)dt+\int_0^{t_1} I_o(t)dt+\int_{t_1}^{t_2} I_o(t)dt\Bigg]$$

$$CEC = C_e\Bigg\{\frac{1}{240}\Bigg\{ar^b(t_1-t_2)\big(3k^3t_1^4t_2^4+3k^3t_1^3t_2^5+3k^3t_1^2t_2^6+3k^3t_1t_2^7+3k^3t_2^8+6k^2t_1^4t_2^2+6k^2t_1^3t_2^3$$
$$-4k^2t_1^2t_2^4-4k^2t_1t_2^5-4k^2t_2^6+12kt_1^4+12kt_1^3t_2-8kt_1^2t_2^2-8kt_1t_2^3+52kt_2^4-40t_1^2-40t_1t_2+80t_2^2\big)$$
$$+ar^bt^3(3k^3t_1^6-4k^2t_1^4+52kt_1^2+80)\Bigg\}-\frac{t_1}{60k}\Bigg[ar^b-ar^b\bigg(\frac{k^2t_1^4}{4}-\frac{kt_1^2}{2}+1\bigg)\bigg(1+\frac{kt_1^2}{2}+\frac{k^2t_2^4}{4}\bigg)$$
$$\bigg(1+\frac{kt_1^2}{2}+\frac{k^2t_1^4}{4}\bigg)(3k^2t_1^4-10kt_1^2+60)\Bigg]\Bigg\} \quad (15)$$

5. **Shortage cost:**

$$SC = C_s\int_{t_2}^{T} I_s(t)dt = -C_sar^b\Bigg[\frac{(T-t_2)^2(T+2t_2)}{6}\Bigg] \quad (16)$$

Consequently, the retailer's total relevant inventory costs can be stated as follows:

$$TC(r,T) = \frac{1}{T}[OC+DC+HC+CEC+SC]$$

$$TC(r,T) = A+(-C_d\cdot k-h-C_e)\Bigg\{\frac{1}{240}\Bigg\{ar^b(t_1-t_2)\big(3k^3t_1^4t_2^4+3k^3t_1^3t_2^5+3k^3t_1^2t_2^6+3k^3t_1t_2^7+3k^3t_2^8$$
$$+6k^2t_1^4t_2^2+6k^2t_1^3t_2^3-4k^2t_1^2t_2^4-4k^2t_1t_2^5-4k^2t_2^6+12kt_1^4+12kt_1^3t_2-8kt_1^2t_2^2$$
$$-8kt_1t_2^3+52kt_2^4-40t_1^2-40t_1t_2+80t_2^2)+ar^bt^3(3k^3t_1^6-4k^2t_1^4+52kt_1^2+80)\Bigg\}-\frac{t_1}{60k}\Bigg[ar^b-ar^b\bigg(\frac{k^2t_1^4}{4}-\frac{kt_1^2}{2}+1\bigg)$$
$$\bigg(1+\frac{kt_1^2}{2}+\frac{k^2t_2^4}{4}\bigg)\bigg(1+\frac{kt_1^2}{2}+\frac{k^2t_1^4}{4}\bigg)(3k^2t_1^4-10kt_1^2+60)\Bigg]\Bigg\}-C_sar^b\Bigg[\frac{(T-t_2)^2(T+2t_2)}{6}\Bigg] \quad (17)$$

Let $t_1 = mt_2,\ 0 < m < 1$, then we get equation (18) from equation (17).

$$TC(r,T) = A + (-C_d \cdot k - h - C_e)\left\{\frac{1}{240}\left\{ar^b(mt_2 - t_2)\big(3k^3(mt_2)^4t_2^4 + 3k^3(mt_2)^3t_2^5 + 3k^3(mt_2)^2t_2^6 + 3k^3mt_2t_2^7 + 3k^3t_2^8 + 6k^2(mt_2)^4t_2^2 + 6k^2(mt_2)^3t_2^3 - 4k^2(mt_2)^2t_2^4 - 4k^2mt_2t_2^5 - 4k^2t_2^6 + 12k(mt_2)^4 + 12k(mt_2)^3t_2 - 8k(mt_2)^2t_2^2 - 8kmt_2t_2^3 + 52kt_2^4 - 40(mt_2)^2 - 40mt_2^2 + 80t_2^2\big) + ar^bt^3(3k^3(mt_2)^6 - 4k^2(mt_2)^4 + 52k(mt_2)^2 + 80)\right\} - \frac{t_1}{60k}\left[ar^b - ar^b\left(\frac{k^2(mt_2)^4}{4} - \frac{k(mt_2)^2}{2} + 1\right)\left(1 + \frac{k(mt_2)^2}{2} + \frac{k^2t_2^4}{4}\right)\left(1 + \frac{k(mt_2)^2}{2} + \frac{k^2(mt_2)^4}{4}\right)(3k^2(mt_2)^4 - 10k(mt_2)^2 + 60)\right]\right\} - C_sar^b\left[\frac{(T-t_2)^2(T+2t_2)}{6}\right] \quad (18)$$

Let $t_2 = nT,\ 0 < n < 1$, then we get equation (19) from equation (18).

$$TC(r,T) = A + (-C_d \cdot k - h - C_e)\left\{\frac{1}{240}\left\{ar^b(mnT - nT)\big(3k^3(mnT)^4nT^4 + 3k^3(mnT)^3(nT)^5 + 3k^3(mnT)^2(nT)^6 + 3k^3m(nT)(nT)^7 + 3k^3(nT)^8 + 6k^2(mnT)^4(nT)^2 + 6k^2(mnT)^3(nT)^3 - 4k^2(mnT)^2(nT)^4 - 4k^2mnT(nT)^5 - 4k^2(nT)^6 + 12k(mnT)^4 + 12k(mnT)^3nT - 8k(mnT)^2(nT)^2 - 8kmnT(nT)^3 + 52k(nT)^4 - 40(mnT)^2 - 40(mnT)^2 + 80(nT)^2\big) + ar^bt^3(3k^3(mnT)^6 - 4k^2(mnT)^4 + 52k(mnT)^2 + 80)\right\} - \frac{t_1}{60k}\left[ar^b - ar^b\left(\frac{k^2(mnT)^4}{4} - \frac{k(mnT)^2}{2} + 1\right)\left(1 + \frac{k(mnT)^2}{2} + \frac{k^2nT^4}{4}\right)\left(1 + \frac{k(mnT)^2}{2} + \frac{k^2(mnT)^4}{4}\right)(3k^2(mnT)^4 - 10k(mnT)^2 + 60)\right]\right\} - C_sar^b\left[\frac{(T-nT)^2(T+2nT)}{6}\right] \quad (19)$$

5. Analytical Optimization Methodology

The optimal values of r and T for minimizing the total inventory costs per unit time is any solution which satisfies simultaneously the equations $\frac{\partial TC(r,T)}{\partial r} = 0$ and $\frac{\partial TC(r,T)}{\partial T} = 0$ for which also satisfies the conditions $\frac{\partial^2 TC(r,T)}{\partial r^2} > 0$, $\frac{\partial^2 TC(r,T)}{\partial T^2} > 0$ and $\left[\frac{\partial^2 TC(r,T)}{\partial r^2} \cdot \frac{\partial^2 TC(r,T)}{\partial T^2} - \frac{\partial^2 TC(r,T)}{\partial r \partial T}\right] > 0$. Using these optimal values of and , the optimal values of can be obtained from above equations.

The aforementioned problems are extremely nonlinear. We have solved the aforementioned problems using an *algorithm* (mentioned below) with the support of MATLAB software (version R2021b) in order to get a solution to non linearity.

Algorithm:

Step 1: Input the values of all the system parameters.
Step 2: Build the functions $TC(r,T)$ given by the equations (10).
Step 3: Find out the optimal values of the decision variables r^* and T^* using the above optimal conditions.
Step 4: Minimize the objective function $TC^*(r,T)$ using r^* and T^*.
Step 5: Calculate the optimal values of $TC^*(r,T)$.
Step 6: Stop.

6. Numerical Illustration

In this part, a numerical example with graphical representations is provided to demonstrate the suggested model. To gauge his best ordering choices, take a look at a retailer of decaying commodities. In order to illustrate the above solution procedure, we consider the following examples:

6.1. Example

A numerical example is given in this section to show how the model works. The following parameter values are used as input:

$$a = 0.005,\ b = 0.02,\ m = 0.5,\ n = 0.5,\ k = 0.2\ Q = 446\ Units,\ A = 2000,$$

$$C_d = 4.0\ \$/unit,\ h = 6.0\ \$/unit,\ C_e = 0.02\ \$/unit,\ C_s = 20.0\ \$/unit.$$

The optimal solution obtained is as below:

$$r^* = 7\%,\ T^* = 14.0835\ weeks,\ TC^* = \$157.5451/cycle.$$

7. Sensitivity analysis

Sensitivity analysis is used in this part to investigate how changes in the parameter's values affect the optimum values. In order to do these studies, one parameter was changed by $\pm 5\%$ and $\pm 10\%$ at a time while remaining at its original value for the other parameters. The following table 3 is demonstrate the outcomes of the sensitivity analysis.

Table 4: *Effects of changing parameters of various kinds on the total cost and optimal solution.*

Parameters	%	% change in optimal value		
	Change	r^*	T^*	TC^*
a	-10%	7.9379	14.2271	155.9346
	-5%	7.9689	14.1524	156.7686
	+5%	7.9785	14.0167	158.3036
	+10%	7.9887	13.9533	159.0294
b	-10%	7.7247	14.1074	157.2758
	-5%	7.5300	14.0957	157.4071
	+5%	7.5144	14.0712	157.6851
	+10%	7.4951	14.0589	157.8240
m	-10%	8.5317	14.4785	153.9983
	-5%	8.0970	14.3002	155.5692
	+5%	7.4332	13.8314	159.9597
	+10%	7.7621	13.5525	162.7882
n	-10%	8.8974	15.6681	140.7720
	-5%	8.0178	14.8333	149.1925
	+5%	7.4691	13.4056	165.8533
	+10%	6.9579	12.7923	174.0883
k	-10%	8.8778	14.6177	151.8048
	-5%	8.1907	14.3412	154.7248
	+5%	7.6089	13.8398	160.3066
	+10%	8.1921	13.6093	163.0060

Table 5: *Effects of changing parameters of various kinds on the total cost and optimal solution (Continue).*

Parameters	% Change	% change in optimal value		
		r^*	T^*	TC^*
h	-10%	7.6383	14.2213	155.8464
	-5%	7.6519	14.1504	156.7162
	+5%	7.7463	14.0185	158.3574
	+10%	7.5684	13.9579	159.1237
C_e	-10%	7.2917	14.0848	157.5303
	-5%	7.3169	14.0845	157.5341
	+5%	7.9907	14.0816	157.5669
	+10%	7.9835	14.0814	157.5694
C_d	-10%	7.5185	14.0928	157.4356
	-5%	7.9942	14.0865	157.5096
	+5%	7.7212	14.0782	157.6081
	+10%	7.5155	14.0743	157.6541
C_s	-10%	8.0570	14.0778	157.7645
	-5%	8.0218	14.0798	157.6643
	+5%	7.4901	14.0855	157.4453
	+10%	7.4331	14.0877	157.3441
A	-10%	7.7104	13.9385	143.2794
	-5%	7.7151	14.0125	150.4348
	+5%	7.9615	14.1490	164.6474
	+10%	7.7081	14.2140	171.6880

Following are a few insights drawn from the sensitivity analysis's observations.

1. If the parameters a and b are increased, the total average inventory cost (TC) rises due to the higher demand rate. At the same time, the cycle length decreases, while the product's reliability increases in relation to a but decreases in relation to b.

2. If the parameters m and n are increased, the total average inventory cost (TC) rises quickly due to the extended running time of the warehouses. Simultaneously, both the cycle length and the product reliability decrease.

3. If the deterioration rate (k) increases, the total average inventory cost (TC) rises rapidly due to the higher amount of product waste. At the same time, the cycle length and the product reliability decrease.

4. If the holding cost (h), deterioration cost (k), and carbon emission cost (C_e) increase, the total average inventory cost (TC) rises due to the higher associated expenses. At the same time, the cycle length decreases, but the product reliability improves gradually.

5. If the ordering cost (A) increases, the total average inventory cost (TC) rises rapidly due to the higher cost per order. At the same time, both the cycle length and product reliability increase.

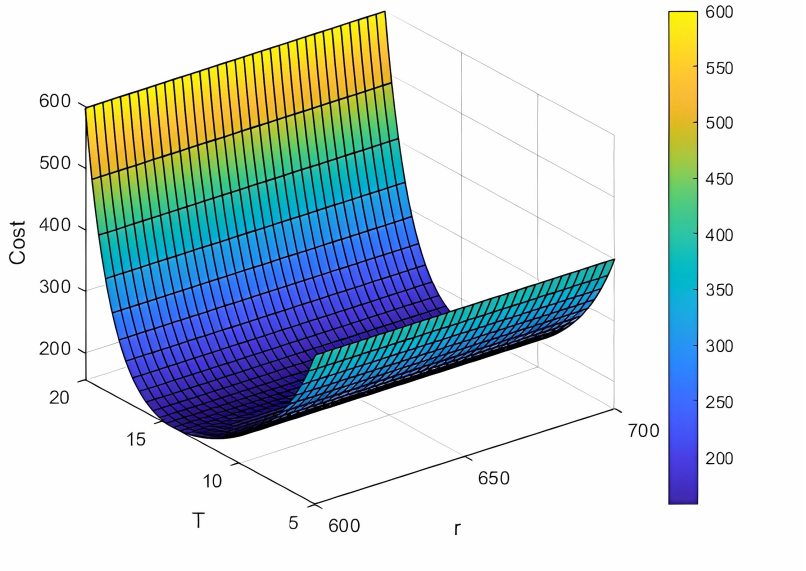


Figure 4: *Convexity of $TC(r, T)$ with respect to r and T*

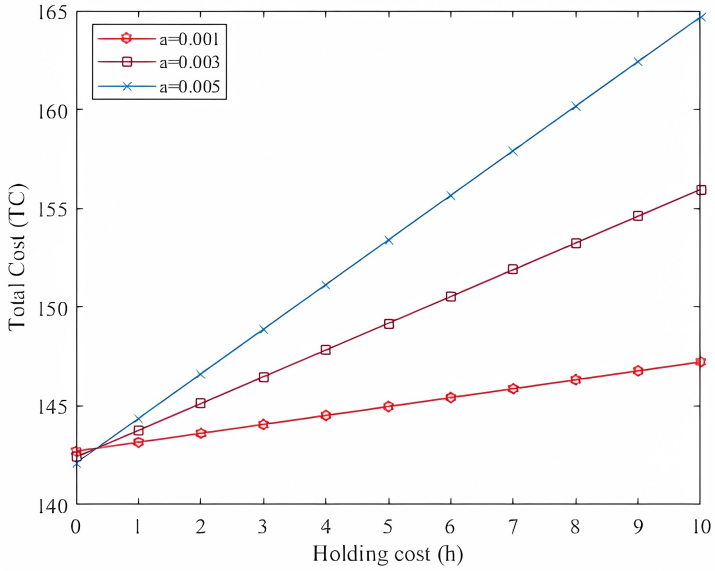


Figure 5: *Relation between Total Cost and Holding Cost using different values of a.*

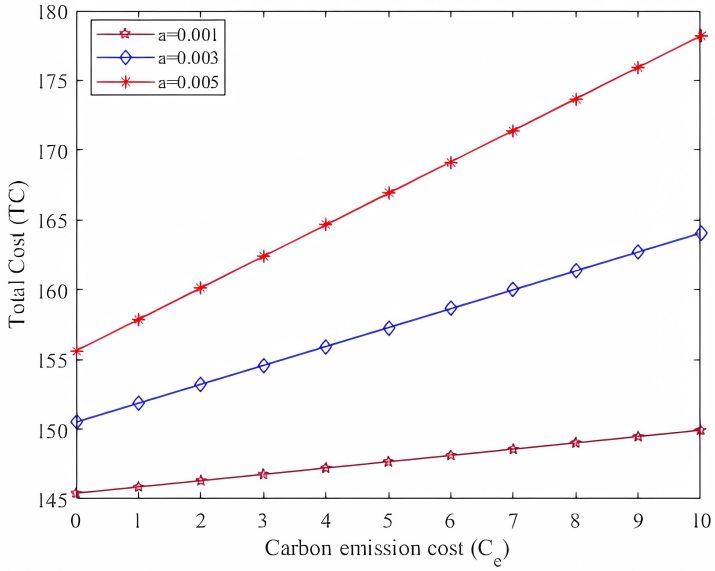


Figure 6: *Relation between Total Cost and Carbon Emission Cost using different values of a.*

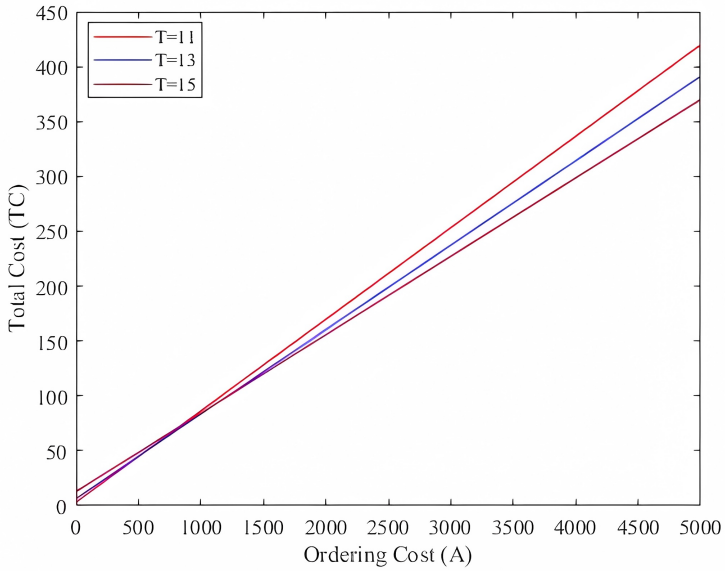


Figure 7: *Relation between Total Cost and Ordering Cost using different values of T.*

8. Conclusion and Future Directions

This research has designed and optimized a two-warehouse inventory management system for deteriorating items, taking into account time and reliability-dependent demand alongside carbon emission constraints. The findings indicate that managing deterioration rates, demand variability, and environmental regulations effectively can notably enhance cost efficiency and inventory reliability. By refining inventory allocation between warehouses and applying effective replenishment strategies, the system can lower total average inventory costs while complying with carbon emission regulations.

The findings emphasize the crucial role of incorporating sustainability into inventory management practices. The model offers important insights into more effective management of deteriorating

items, showing that a thorough evaluation of deterioration rates and carbon constraints can significantly enhance overall performance. Additionally, a sensitivity analysis based on numerical examples using MATLAB Software version R2021b is provided to demonstrate the impact of carbon emissions and validate the model.

Future research could investigate different aspects to enhance the model's relevance and accuracy:

1. Expanded Constraints: Explore additional variables like changing transportation costs, supply chain disruptions, or shifts in regulations that could affect the optimization process.
2. Analytical Optimization Technique: Explore analytical optimization methods to more effectively manage trade-offs between cost, reliability, and environmental impact.
3. Industry-Specific Applications: Implement the developed model in specific industries or sectors to assess its robustness and adaptability across different types of deteriorating items and operational settings.

By exploring these future directions, researchers and practitioners can expand on the groundwork established by this study to create more thorough and flexible inventory management strategies.

References

[1] Banerjee, A., & Gupta, S. M. (1994). Analysis of a two-warehouse inventory model with item-dependent and time-varying demand. European Journal of Operational Research, 77(2), 371-380.
[2] Bakker, M., Riezebos, J., & Teunter, R. H. (2012). Review of inventory systems with deterioration since 2001. European Journal of Operational Research, 221(2), 275-284.
[3] Benjaafar, S., Li, Y., & Daskin, M. (2013). Carbon footprint and the management of supply chains: Insights from simple models. IEEE Transactions on Automation Science and Engineering, 10(1), 99-116.
[4] Chaabane, A., Ramudhin, A., & Paquet, M. (2012). Design of sustainable supply chains under the emission trading scheme. International Journal of Production Economics, 135(1), 37-49.
[5] Ghare, P. M., & Schrader, G. F. (1963). A model for exponentially decaying inventory. Journal of Industrial Engineering, 14(3), 238-243.
[6] Govindan, K., Soleimani, H., & Kannan, D. (2014). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. European Journal of Operational Research, 240(3), 603-626.
[7] Hartley, R. V. (1976). Operations research-a managerial emphasis. Good Year Publishing Company.
[8] Hill, R. M., & Omar, M. (2006). EOQ model for a perishable item with a quadratic time-varying demand. International Journal of Production Economics, 101(2), 369-374.
[9] Jha, J. K., & Shanker, K. (2009). Two-warehouse inventory model with partially backordered shortages and inflation. European Journal of Operational Research, 184(2), 509-518.
[10] Kumar, S., & Uthayakumar, R. (2007). Inventory model for non-instantaneous deteriorating items with reliability in demand and backlogging. International Journal of Production Economics, 106(2), 171-183.
[11] Roy, A., Maiti, M., & Kar, S. (2020). A reliability based inventory model for deteriorating items with linear trend in demand and shortage. Mathematics and Computers in Simulation, 170, 137-148.
[12] Silver, E. A., Pyke, D. F., & Peterson, R. (1998). Inventory management and production planning and scheduling. John Wiley & Sons.
[13] Wee, H. M. (1995). A deterministic lot-size inventory model for deteriorating items with shortages and a declining market. Computers & Operations Research, 22(3), 345-356.

[14] Singh, S. R., and Himanshu Rathore. "An inventory model for deteriorating item with reliability consideration and trade credit." Pakistan Journal of Statistics and Operation Research (2014): 349-360.

[15] Adak, Sudip, and G. S. Mahapatra. "Effect of reliability on varying demand and holding cost on inventory system incorporating probabilistic deterioration." Journal of Industrial & Management Optimization 18.1 (2022).

[16] Mahapatra, G. S., Sudip Adak, and Kolla Kaladhar. "A fuzzy inventory model with three parameter Weibull deterioration with reliant holding cost and demand incorporating reliability." Journal of Intelligent & Fuzzy Systems 36.6 (2019): 5731-5744.

[17] Manna, Amalesh Kumar, Jayanta Kumar Dey, and Shyamal Kumar Mondal. "Two layers supply chain in an imperfect production inventory model with two storage facilities under reliability consideration." Journal of Industrial and Production Engineering 35.2 (2018): 57-73.

[18] Adak, Sudip, and G. S. Mahapatra. "An inventory model of time and reliability Dependent demand with deterioration and partial back ordering under fuzziness." International Conference on Information Technology and Applied Mathematics. Cham: Springer International Publishing, 2019.

[19] Adak, Sudip, and G. S. Mahapatra. "Two-echelon imperfect production fuzzy supply chain model for reliability dependent demand with probabilistic deterioration and rework." Journal of Intelligent & Fuzzy Systems 41.2 (2021): 3833-3847.

[20] Adak, Sudip, and G. S. Mahapatra. "Effect of inspection and rework of probabilistic defective production on two-layer supply chain incorporating deterioration and reliability dependent demand." International Journal of System Assurance Engineering and Management 12.3 (2021): 565-578.

[21] Rajput, Neelanjana, Anand Chauhan, and R. K. Pandey. "Optimisation of an FEOQ model for deteriorating items with reliability influence demand." International Journal of Services and Operations Management 43.1 (2022): 125-144.

[22] Khara, Barun, Jayanta Kumar Dey, and Shyamal Kumar Mondal. "Effects of product reliability dependent demand in an EPQ model considering partially imperfect production." International Journal of Mathematics in Operational Research 15.2 (2019): 242-264.

[23] Mahapatra, G. S., et al. "Inventory model for deteriorating items with time and reliability dependent demand and partial backorder." International Journal of Operational Research 29.3 (2017): 344-359.

[24] Kumar, Kamal. "Inventory Policy for Deteriorating Items with TwoWarehouse and Effect of Carbon Emission." Reliability: Theory & Applications 16.SI 2 (64) (2021): 156-165.

[25] Yadav, A.S., Kumar, A., Yadav, K.K. et al. Optimization of an inventory model for deteriorating items with both selling price and time-sensitive demand and carbon emission under green technology investment. Int J Interact Des Manuf (2023). https://doi.org/10.1007/s12008-023-01689-8

[26] Yadav, K.K., Yadav, A.S. & Bansal, S. Interval number approach for two-warehouse inventory management of deteriorating items with preservation technology investment using analytical optimization methods. Int J Interact Des Manuf (2024). https://doi.org/10.1007/s12008-023-01672-3

[27] Yadav, Ajay Singh, et al. "A study of Covid-19 pandemic on fertilizer supply chain inventory management using travelling salesman problem for Cuckoo Search Algorithms." (2020).

[28] Mahata, Sourav, and Bijoy Krishna Debnath. "A profit maximization single item inventory problem considering deterioration during carrying for price dependent demand and preservation technology investment." RAIRO-Operations Research 56.3 (2022): 1841-1856.

[29] Mahata, Sourav, and Bijoy Krishna Debnath. "The impact of R&D expenditures and screening in an economic production rate (EPR) inventory model for a flawed production system with imperfect screening under an interval-valued environment." Journal of Computational Science 69 (2023): 102027.

[30] Debnath, Bijoy Krishna, Pinki Majumder, and Uttam Kumar Bera. "Two warehouse inventory models of breakable items with stock dependent demand under trade credit policy

with respect to both supplier and retailer." International Journal of Logistics Systems and Management 31.2 (2018): 151-166.

[31] Baloni, J., U. S. Rana, and N. Kumar. "Mathematical Modeling for Homogeneous Tumor with Delay in Time." International Journal of Engineering 22.1 (2009): 49-56.

[32] Sharma, Anupama, et al. "Green inventory model with two-warehouse system considering variable holding cost, time dependent demand, carbon emissions and energy consumption." International Journal of Procurement Management 19.2 (2024): 274-296.

[33] Kumar, Sunil, and Rajendra Prasad Mahapatra. "EPOIM: an advanced optimization method for two warehouse inventory model." Multimedia Tools and Applications (2024): 1-28.

RECTIFYING INSPECTION FOR DOUBLE SAMPLING PLANS WITH FUZZY LOGIC UNDER ZERO-INFLATED POISSON DISTRIBUTION USING IN PYTHON

Kavithanjali S [1], Sheik Abdullah A[*] and Kamalanathan R[2]

•

[1]Research Scholar, Department of Statistics, Salem Sowdeswari College, Salem-10, India
[*2]Assistant Professor, Department of Statistics, Salem Sowdeswari College, Salem-10, India
kavithanjalis2018@gmail.com, sheik.stat@gmail.com, rknstat1@gmail.com

Abstract

Acceptance sampling is a statistical quality control technique used in manufacturing to determine whether to accept or reject a batch of products based on the number of defects obtain in a sample. Among the various sampling plans, the double sampling plan more effective because it often delivers more reliable results in selecting quality lots than other plans. In most of the real life situation, it is not easy found the product as strictly defective or non-defective. In some situation, quality of the product can be classified several types which are expressed as good, almost good, bad, not so bad and so on. This is causes fuzzy logic comes into play. Fuzzy set theory is most powerful mathematical tool, it can deal incomplete and imprecise information. In this paper Double Sampling Plans (DSPs) are derived when non conformities are say imprecise and these imprecisions are model through ZIP distribution. It analyzes, the effectiveness of these sampling plans by comparing vital metrics such as Average Outgoing Quality (AOQ) and Average Total Inspection (ATI) using both fuzzy and crisp environments. These findings are appraised as both numerically and graphically, showing that whether the process quality is either extremely good or very bad, the AOQ curve will be lower, the plan's able to effectively control product quality.

Keywords: Acceptance Double Sampling Plan, FAOQ Curve, FATI curve, ZIP Distribution, Fuzzy Parameter.

I. Introduction

Quality control in industries often meets troubles when handling with data that such as high number of zero values, which usual sampling plans may not manage properly. This paper presents a new approach that combines fuzzy logic and double sampling plan constructed on the basis of Zero-Inflated Poisson (ZIP) distribution. Fuzzy logic enhances the handling of uncertainty and incorrect data, making it easy to deal with the ambiguous limit with zero and non-zero values in ZIP distributions. The method, which was developed by Zadeh and Kosko's (1965) original work, improves quality control by providing a versatile and adaptable method. This technique tries to enhancing decision-making in quality control processes, specifically in circumstances where technological innovation might lead to rare zero defects. Utilizing fuzzy logic, the proposed sampling plan provides an accurate strategy to identify defects in industrial processes, even when

handling complex information. The Zero-Inflated Poisson (ZIP) distribution is used in agriculture, epidemiology, econometrics, public health, process control, medicine, and manufacturing. Especially helpful for studying data containing extra zeros. Bohning, et.al., [1] analyzed dental epidemiology intervention effects using the ZIP model. Buckley [2] developed flexible probability theory under uncertainty. Duncan AJ [3] Quality Control and Industrial Statistics is a most impact guide that combines statistical methods with practical approaches to quality control in industrial circumstances.

Chakraborty [4] gives a fuzzy optimization method for single sample plans that minimizes inspection and manages customer risk using Poisson distribution. Ezzatallah and Bahram [5] suggested fuzzy Poisson-based acceptance double sampling for not clear faulty rates. Kavithanjali, et.al., [6] reviewing a SQC methods in single and double-sampling plans, pointing at possible effects on quality. Kavithanjali and Sheik Abdullah [7] gives an innovative technique by combination of fuzzy logic with Zero-Inflated Poisson distribution for single sampling plans, improving quality control and risk management in ambiguity distribution plans. The Python implementation gives practical value, making it useful for real-world circumstances. Kaviyarasu and Asif T Thottathil [8] deals the application of Zero-Inflated Poisson distribution in designing optimal acceptance sampling plans for quality control in manufacturing with a focus on special type double sampling plans. Lotfi A. Zadeh [9] provide fuzzy sets and the degree of membership, which is basis for the employment of conventional theory of sets in fuzzy controller. Lambert [10] illustrated how ZIP regression might enhance data analysis in manufacturing by removing extra zeros from count data. Malathi and Muthulakshmi [11] studied fuzzy logic in double-sampling plans to overcome uncertainty in quality evaluations.

McLachlan and Peel [12] provided insights into finite mixture models, which are critical for analysing complicated data with diverse populations. Naya and et.al., [13] employed ZIP models to demonstrate that age has a substantial effect on the incidence of black patches. Ridout, Demetrio, and Hinde [14] assess count data model excess zeros by using horticultural examples. Schilling and Neubauer [15] provide a comprehensive and authoritative guide on acceptance sampling plan, offering useful insights for quality control in numerous industries. Tamaki F, Kanagawa and Ohta H [16] deals a unique method applying fuzzy logic to attribute-based sampling inspection plans, improving decision-making under ambiguity. For over-dispersed data, Xie, He, and Goh [17] show that the ZIP distribution is better than the Poisson distribution for statistical process control.

The techniques used will be discussed in the sections that follow, along with the results of our study and an explanation of their importance for quality control practitioners. We did our analysis using Python and powerful libraries like NumPy, Pandas, SciPy, and Matplotlib to help with statistics and data visualization. We hope that our work will make a major contribution to the evolving scene of statistical methods meant to solve the issues raised by demanding distributions in industrial environments.

II. Methodology

2.1 Basic Definitions

2.1.1 Fuzzy Number: A fuzzy number $(\widetilde{N})$ is a fuzzy set on the real line R, characterized by a membership function $\mu_N: R \to [0,1]$, that satisfies the following conditions:

- $(\widetilde{N})$ is normal, meaning there exists some x such that $\mu_N(x) = 1$.
- $(\widetilde{N})$ is convex, meaning for any $x_1, x_2 \in R$ and $\lambda \in [0,1], \mu_N(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_N(x_1), \mu_N(x_2))$.
- The membership function μ_N is upper semi-continuous, meaning the set $\{x \in R \mid \mu_N(x) \geq \alpha\}$ is closed for every $\alpha \in (0,1]$.

- The support of $(\widetilde{N})$, defined as $\text{Sup}(\widetilde{N}) = \{x \in R \mid \mu_N(x) > 0\}$, is bounded.

2.1.2 Triangular Fuzzy Number: A triangular fuzzy number $(\widetilde{N})$ is defined by a triplet (a, b, c), where $a < b < c$. The membership function $\mu_{(\widetilde{N})}(x)$ is given by:

$$\mu_{(\widetilde{N})}(x) = \begin{cases} \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ \dfrac{c-x}{c-b} & \text{if } b < x \le c \\ 0 & \text{otherwise} \end{cases}$$

This function forms a triangular shape with $[a, c]$ as the base and the peak at $x = b$.

2.1.3 α-Cut of Fuzzy : The α-cut of a fuzzy set $\widetilde{N}$ is a crisp set of values where the membership function is at least α. It is defined as:

$$N[\alpha] = \{x \in R \mid \mu_N(x) \ge \alpha\}$$

The fuzzy number $\widetilde{N}[\alpha]$ can be represented by its lower and upper bounds as $N^L[\alpha]$ and $N^U[\alpha]$, where:

$$N^L[\alpha] = \inf\{x \in R \mid \mu_N(x) \ge \alpha\}$$
$$N^U[\alpha] = \sup\{x \in R \mid \mu_N(x) \ge \alpha\}$$

2.1.4 ZIP Distribution: The Zero-Inflated Poisson (ZIP) distribution, define as ZIP (φ, λ), is used when there is an more number of zero counts. The probability mass function (p.m.f.) is found in Lambert [10] and Mclachlan [12]:

$$P(D = d \mid \varphi, \lambda) = \begin{cases} \varphi + (1-\varphi)e^{-\lambda} & \text{if } d = 0 \\ (1-\varphi)\dfrac{e^{-\lambda}\lambda^d}{d!} & \text{if } d = 1,2,\ldots \end{cases}$$

In this distribution:

- φ represents the probability of extra zeros.
- λ is the mean of the underlying Poisson distribution.

The ZIP distribution mean is $(1-\varphi)\lambda$, and the variance is $\lambda(1-\varphi)(1+\varphi\lambda)$.
To extend the ZIP distribution to a fuzzy setting, we replace λ with a fuzzy number $\tilde{\lambda} > 0$. The fuzzy p.m.f. can be represented as:

$$\tilde{P}(d \mid \alpha) = \begin{cases} \varphi + (1-\varphi)e^{-\lambda} & \text{if } d = 0 \\ (1-\varphi)\dfrac{e^{-\lambda}\lambda^d}{d!} & \text{if } d = 1,2,\ldots \end{cases}$$

Where λ belongs to the $\alpha-$ cut of $\tilde{\lambda}$.

2.2 Python Programming

Python was an indispensable part of this study as it helped creating of statistical quality control plans. was specifically used to calculate the upper and lower limits for the Fuzzy AOQ Band and the Fuzzy ATI tables. Moreover, Python's extensive plotting capabilities were utilized to show important metrics, which including the Fuzzy Average Outgoing Quality (AOQ), the fuzzy probability of acceptance, and the Average Total Inspection (ATI) curve. Python was chosen for its strong numerical computation capabilities and a massive library that could otherwise allow for the incorporation of complex statistical strategies into the research environment.

III. Operating procedure for RDSPs

Let us consider a circumstance where we analyse the N- lot size for defects with Zero-Inflated Poisson (ZIP) distribution. These are general steps of the typical double sampling plan.

Step 1:

- Take a random sample of size n_1 and count the number of defective items (D_1).
- c_1 is the acceptance number for the first sample.
- c_2 is the acceptance number for both combined samples.

Step 2:

- Accept the lot if $D_1 \leq c_1$.
- Reject the lot if $D_1 > c_2$.
- If $c_1 < D_1 \leq c_2$, proceed to Step 3 .

Step 3:

- Take a random sample from second sample n_2 and count the number of defective items (D_2).
- Add D_1 and D_2 together.
- Accept the lot if $D_1 + D_2 \leq c_2$, otherwise reject it.

Step 4:

- The random variables D_1 and D_2 follows the ZIP distribution with parameter $\lambda_1 = n_1 p$ and $\lambda_2 = n_2 p$, given a large sample size and a small probability p.
- Let P_a stand for the acceptance probability of the lot based onto the combined samples.
- $\tilde{P}_a^I$is for the acceptance probability after the first sample and $\tilde{P}_a^{II}$ for the second sample.

Thus, the overall probability of acceptance is:

$$\tilde{P}_a = \tilde{P}_a^I . \tilde{P}_a^{II}$$

Using the ZIP distribution p.m.f, the number of nonconforming items in the lot is given by

$$\mathrm{P}(\mathrm{D} = \mathrm{d} \mid \varphi, \lambda) = \tilde{P}(\mathrm{d}) = \begin{cases} \varphi + (1-\varphi)e^{-\lambda}, & \text{Whend } = 0 \\ (1-\varphi)\dfrac{e^{-\lambda}\lambda^d}{d!}, & \text{Whend } = 1,2,\ldots, 0 < \varphi < 1, \lambda > 0 \end{cases}$$

Given a sample size of n_1, the probability of finding no deficiencies will be

$$\tilde{P}(D = 0) = \tilde{P}_a^I(\alpha) = \varphi + (1-\varphi)e^{-n_1 p} \quad (1)$$

Given a sample size of n_2, the probability of finding one deficiencies will be

$$\tilde{P}(D = 1, D_1 + D_2 \leq 1) = \tilde{P}_a^{II}(\alpha) = (1-\varphi)e^{-\mathrm{n}p} n_2 p \quad (2)$$

A DSP only accepts a lot if a sample of size n_1 has no faults and a sample of size n_2 has one defect or less. Thus, DSP's $\hat{P}_a(\alpha)$ will be provided by

$$\tilde{P}_a(\alpha) = \tilde{P}_a^I(\alpha) + \tilde{P}_a^{II}(\alpha)$$

IV. Fuzzy Average Outgoing Quality (FAOQ) under ZIP distribution

In acceptance sampling programs, rectification inspection is used to improve the quality of the lot. When the lot is approved, any faulty items in the sample are replaced with non-defective ones. If the batch is rejected, the whole batch is thoroughly inspected, and any faulty goods are replaced with new ones. Then by the following steps Involved operating procedure for AOQ of double sampling plan the acceptable quality level can be obtained from the double sampling plan as given below.

4.1 Operating Methodology for Fuzzy Average Outgoing Quality(FAOQ)

Step 1: Initial Assumptions

- The lot size is N, and the probability of a faulty item is $\tilde{p}$.
- A sample of size n_1 is taken from the batch.

Step 2: First-Stage Sampling

- The probability of the lot being approved after the first step is $\tilde{P}_a^I$.
- If rejected with probability $1 - \tilde{P}_a^I$, a second stage of sampling is conducted.

Step 3: Results of First Stage Acceptance

- After checking n1 items, the remaining N-n_1 items have an average defect rate of $\tilde{p}(N - n_1)$.

Step 4: Second-stage Sampling

- A second sample of size n_2 is chosen.
- If a lot is rejected beyond this step, all faulty components are replaced to ensure zero defects.
- If passed, the remaining N-n_1-n_2 items have an average of $\tilde{p}(N - n_1 - n_2)$ faulty items.

Step 5: Probabilities of Outcomes

The probabilities of every possible result are:

$\tilde{P}_a^I$: Acceptance lot after first stage with probability $\tilde{p}(N - n_1)$ of having defective products.

$\tilde{P}_a^{II}$: Lot allowed after the second stage, $\tilde{p}(N - n_1 - n_2)$ defective pieces.

$\tilde{P}_a^{III}$: The lots are 100% checked and this means that no defective items get to make their way through the production line. The probability of satisfying: $\tilde{P}_a^I + \tilde{P}_a^{II} + \tilde{P}_a^{III} = 1$.

From using the fuzzy mean definition, the α-cut of FAOQ is as follows.

$$FAOQ(\alpha) = \left\{ \frac{[P_a^I(N - n_1) + P_a^{II}(N - n_1 - n_2)].p}{N} \middle| \; p \in \tilde{p}(\alpha) \right\}$$

$$= [FAOQ^L(\alpha), FAOQ^U(\alpha)]$$

Where

$$FAOQ^L(\alpha) = \min \left\{ \frac{[P_a^I(N - n_1) + P_a^{II}(N - n_1 - n_2)] \cdot p}{N} \middle| \; p \in \tilde{p}(\alpha) \right\} \text{ and}$$

$$FAOQ^U(\alpha) = \max \left\{ \frac{[P_a^I(N - n_1) + P_a^{II}(N - n_1 - n_2)].p}{N} \middle| \; p \in \tilde{p}(\alpha) \right\} \text{ for } 0 \le \alpha \le 1$$

Illustration 1

Take into consideration that $\tilde{P} = (0.01, 0.02, 0.03), \mathrm{N} = 200, \mathrm{n}_1 = 10, \mathrm{n}_2 = 10, \mathrm{c}_1 = 0, \mathrm{c}_2 = 1, \tilde{\lambda} = n\tilde{p}$, $\mathrm{n} = \mathrm{n}_1 + \mathrm{n}_2$, $\varphi = 0.0001$,$\tilde{P}[\alpha] = [0.01 + 0.01\alpha\,, 0.03 - 0.01\alpha]$.

A sample of size n_1, the probability of obtained no defectives will be the equation (1) from

$$\tilde{P}_a^I(\alpha) = \varphi + (1 - \varphi)\mathrm{e}^{-10p}$$

A sample size of n_2, the probability of obtained one defectives will be the equation (2) from

$$\tilde{P}_a^{II}(\alpha) = (1 - \varphi)\mathrm{e}^{-20p}10p$$

$$FAOQ(\alpha) = \left\{ \frac{[\varphi + (1 - \varphi)\mathrm{e}^{-10p}(200 - 10) + (1 - \varphi)\mathrm{e}^{-20p}10p(200 - 10 - 10)] \cdot p}{200} \right\}$$

$$FAOQ(\alpha) = \varphi + (1 - \varphi)\mathrm{e}^{-10p}(0.95)p + (1 - \varphi)\mathrm{e}^{-20p}(9)p^2$$

$$= [FAOQ^L(\alpha), FAOQ^U(\alpha)]$$

Afterward, using examining function

$$f(p) = \varphi + (1 - \varphi)\mathrm{e}^{-10p}(0.95)p + (1 - \varphi)\mathrm{e}^{-20p}(9)p^2$$

We will have the α cut in the following functions

$$FAOQ^{L}(\alpha) = \varphi + (1-\varphi)\mathrm{e}^{-10(0.01+0.01\alpha)}(0.95)(0.01+0.01\alpha) + (1-\varphi)\mathrm{e}^{-20(0.01+0.01\alpha)}(9)(0.01+0.01\alpha)^2$$

and

$$FAOQ^{U}(\alpha) = \varphi + (1-\varphi)\mathrm{e}^{-10(0.03-0.01\alpha)}(0.95)(0.03-0.01\alpha) + (1-\varphi)\mathrm{e}^{-20(0.03-0.01\alpha)}(9)(0.03-0.01\alpha)^2$$

From $\alpha = 0$, find that FAOQ(0) varies between 0.0094 and 0.0256, showing that 94 to 256 products are predicted to be faulty each lot in this process. When $\alpha = 1$, we get FAOQ(1) = 0.0180. Figure 1 represents the FAOQ, which displays increases in the input quality process.

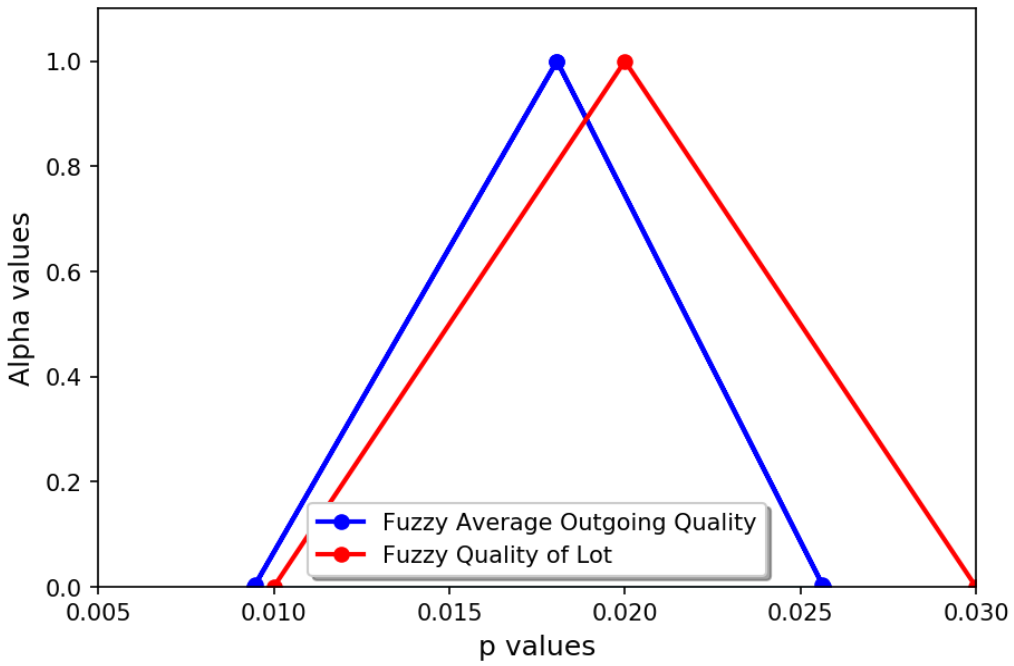


Figure 1: *Comparison of $\tilde{p}$ and fuzzy AOQ with parameters $c_1 = 0$, $c_2 = 1$, and sample sizes $n_1 = n_2 = 10$, $\tilde{P} = (0.01, 0.02, 0.03)$*

FAOQ is a function of lot quality, and when the quality changes, consequently changes the FAOQ. When the FAOQ is plotted against the proportion of bad items in the input lot, the result appears as a diagram with upper and lower boundaries known as the FAOQ band. According upon the established structure for $\tilde{P}$, discussed in the next section.

$$\tilde{P}^{I}_{a,m} = \tilde{\mathrm{P}}_{\mathrm{m}}(\mathrm{D}_1 \le c_1)(\alpha)$$

$$\tilde{P}^{II}_{a,m} = \tilde{\mathrm{P}}_{\mathrm{m}}(c_1 < D_1 \le c_2 \,.\, D_1 + D_2 \le c_2)(0) \text{ and}$$

$$FAOQ(\alpha) = [FAOQ^{L}_{m}(\alpha), FAOQ^{L}_{m}(\alpha)]$$

$$FAOQ(\alpha) = \left\{ \frac{\left[P^{I}_{a,m}(N-n_1) + P^{II}_{a,m}(N-n_1-n_2)\right]p}{N} \mid p \in \tilde{p}(\alpha) \right\}$$

Where

$$\mathrm{FAOQ}^{L}_{m}(\alpha) = \min\left\{ \frac{\left[P^{I}_{a,m}(N-n_1) + P^{II}_{a,m}(N-n_1-n_2)\right]p}{N} \middle|\ p \in \tilde{p}(\alpha) \right\} \text{and}$$

$$\text{FAOQ}_m^U(\alpha) = \max\left\{\frac{\left[P_{a,m}^I(N-n_1)+P_{a,m}^{II}(N-n_1-n_2)\right]p}{N}\middle|\ p\in\tilde{p}(\alpha)\right\}$$

Illustration 2

Considering the following case scenario: $\text{N}=200, \text{n}_1=20, \text{n}_2=20, \text{c}_1=0, \text{c}_2=1, \tilde{\lambda}=n\tilde{p}$, $\text{n}=\text{n}_1+\text{n}_2$, $\text{b}_2=0.01$, $\text{b}_3=0.02, \varphi=0.0001$. From these parameters, the next findings have been generated.

$$\tilde{p}(\alpha)=[m+0.01\alpha, m+0.02-0.01\alpha]$$

$$FAOQ(\alpha)=\varphi+(1-\varphi)\text{e}^{-20p}(0.9)p+(1-\varphi)\text{e}^{-40p}(16)p^2$$

The α-cut of FAOQ at α=0 will be determined as follows

$$\text{FASN}(0)=\begin{cases}\text{FAOQ}^{*},\ \text{FAOQ}^{**} & ,0\le k<0.03\\ \text{FAOQ}^{*},0.021970 & ,0.03\le k<0.0406\\ \text{FAOQ}^{**},0.021970 & ,0.0406\le k<0.05\\ \text{FAOQ}^{**},\ \text{FAOQ}^{*} & ,0.05\le k<0.98\end{cases}$$

$$FAOQ^{*}=\varphi+(1-\varphi)\text{e}^{-20m}(0.9)m+(1-\varphi)\text{e}^{-40m}(16)m^2$$

$$FAOQ^{**}=\varphi+(1-\varphi)\text{e}^{-20(m+0.02)}(0.9)(m+0.02)+(1-\varphi)\text{e}^{-40(m+0.02)}(16)(m+0.02)^2$$

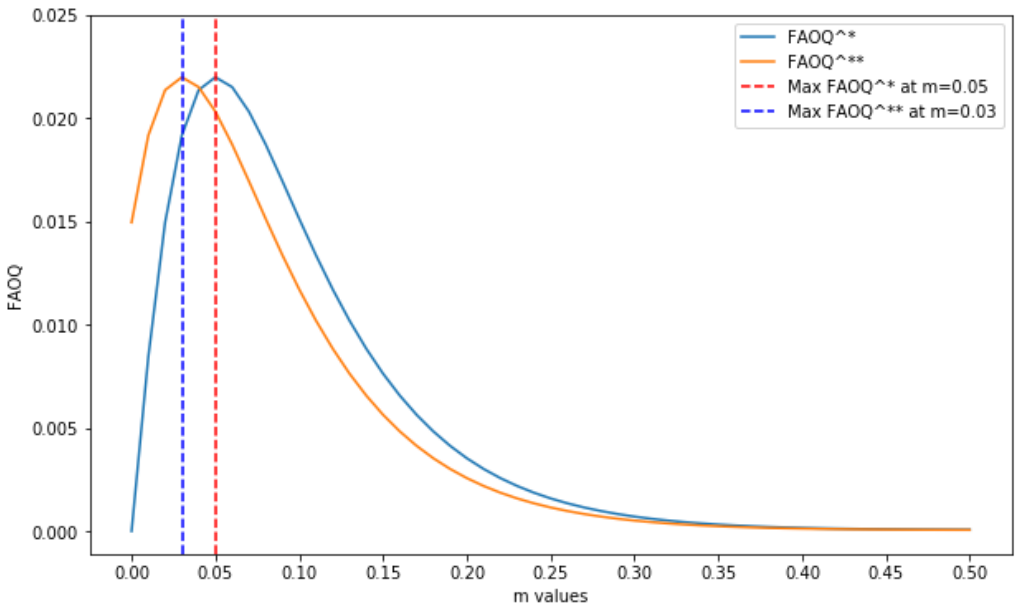


Figure 2: *FAOQ band with the following parameters: N=200, n_1=20, n_2=20, c_1= 0, c_2=1, b_2=0.01, b_3=0.02*

Figure 2 and Table 1 illustrate the FAOQ band for a sampling plan that includes a fuzzy parameter. The outcome show that the FAOQ operates effectively when the proportion of defective items in the incoming lot is either good or bad. The average outgoing quality limit (AOQL) is an important measure of sampling plan performance since it provides the highest expected to percentage of faulty items in inspected lots. The FAOQL, or maximum FAOQ value, is the worst- case scenario for the FAOQ, occurring at a defect level represented as $\tilde{p}^*$. In Instance 2, the results obtained are $\tilde{p}^*=0.03, 0.05$ and FAOQL = 0.02197, 0.02197.

Table 1: *FAOQ with N=200, n_1=20, n_2=20, c_1= 0, c_2=1, b_2=0.01, b_3=0.02*

m	FAOQ*	FAOQ**	m	FAOQ*	FAOQ**
0.00	0.0000	0.01494	0.25	0.00158	0.001146
0.01	0.00844	0.01916	0.26	0.00135	0.000974
0.02	0.01494	0.02135	0.27	0.00115	0.000829
0.03	0.01916	0.02197	0.28	0.00097	0.000705
0.04	0.02135	0.02149	0.29	0.00083	0.0006
0.05	0.02197	0.02031	0.3	0.00071	0.000512
0.06	0.02149	0.01872	0.31	0.0006	0.000437
0.07	0.02031	0.01694	0.32	0.00051	0.000374
0.08	0.01872	0.01512	0.33	0.00044	0.00032
0.09	0.01694	0.01336	0.34	0.00037	0.000275
0.10	0.01512	0.0117	0.35	0.00032	0.000238
0.11	0.01336	0.01019	0.36	0.00028	0.000206
0.12	0.0117	0.00883	0.37	0.00024	0.000179
0.13	0.01019	0.00763	0.38	0.00021	0.000157
0.14	0.00883	0.00656	0.39	0.00018	0.000138
0.15	0.00763	0.00564	0.4	0.00016	0.000123
0.16	0.00656	0.00483	0.41	0.00014	0.00011
0.17	0.00564	0.00413	0.42	0.00012	0.000099
0.18	0.00483	0.00353	0.43	0.00011	0.000091
0.19	0.00413	0.00301	0.44	0.000099	0.000083
0.2	0.00353	0.00257	0.45	0.000091	0.000077
0.21	0.00301	0.00219	0.46	0.000083	0.000072
0.22	0.00257	0.00186	0.47	0.000077	0.000069
0.23	0.00219	0.00158	0.48	0.000072	0.000065
0.24	0.00186	0.00135	0.49	0.000069	0.000063

V. Fuzzy Average Total Inspection (FATI) for DSPs under ZIP distribution

The Fuzzy Average Total Inspection (FATI) is an important technique for rectifying inspection for sampling plans with a fuzzy parameter. The following steps are used to compute the FATI:

- If the lot is accepted in the first inspection stage, the number of inspected items is n_1. The probability of this happening (fuzzy probability) is $\tilde{P}_a^I(\alpha)$.
- If the lot is accepted in the second inspection stage, the total number of inspected items is n_1+n_2 .The probability of this happening is $\tilde{P}_a^{II}(\alpha)$.
- If the lot is not accepted at all, all items (N) are inspected. The probability of this happening is $1-\tilde{P}_a^I(\alpha)-\tilde{P}_a^{II}(\alpha)$.

Finally, the random variable of inspected items has a fuzzy probability function as Table 2 consequently, the FATI according to the definition of a fuzzy mean is as follows:

Table 2: *Fuzzy probability function for the number of inspected items*

Total items under inspection	n_1	n_2	N
Probability with fuzzy	$\tilde{P}_a^I(\alpha)$	$\tilde{P}_a^{II}(\alpha)$	$1-\tilde{P}_a^I(\alpha)-\tilde{P}_a^{II}(\alpha)$

$$\text{FATI}(\alpha) = \{n_1 p_a^I + (n_1 + n_2) p_a^{II} + N p_a^{III}\}$$

$$= \{N - (N - n_1) p_a^I - (N - n_1 - n_2) p_a^{II} \mid p \in \tilde{p}(\alpha)\}$$

$$= [FATI^L(\alpha), FATI^U(\alpha)]$$

In which $FATI^L(\alpha) = \min\{N - (N - n_1) p_a^I - (N - n_1 - n_2) p_a^{II}\}$

$$F_{ATI}{}^L(\alpha) = \max\{N - (N - n_1) p_a^I - (N - n_1 - n_2) p_a^{II}\} \text{ For } 0 \le \alpha \le 1$$

To calculate FATI in illustration 1 using N=200, as follows:

$$FATI = \{200 - \varphi + (1 - \varphi)\mathrm{e}^{-10\mathrm{p}}(190) - (1 - \varphi)\mathrm{e}^{-20\mathrm{p}}(1800p)\}$$

The lower and higher limits of the α cut are shown below

$$FATI^L(\alpha) = \{200 - \varphi + (1 - \varphi)\mathrm{e}^{-(0.1+0.1\alpha)}(190) - (1 - \varphi)\mathrm{e}^{-(0.2+0.2\alpha)}1800(0.01 + 0.01\alpha)\}$$

$$FATI^U(\alpha) = \{200 - \varphi + (1 - \varphi)\mathrm{e}^{-(0.3-0.1\alpha)}(190) - (1 - \varphi)\mathrm{e}^{-(0.6-0.2\alpha)}1800(0.03 - 0.01\alpha)\}$$

For $\alpha = 0$, FATI[0] = [13.36230, 29.6256], and when $\alpha = 1$, FATI[1] = [20.3275, 20.3275]. This implies that we estimate monitoring 20 items from each accepted batch. The figure 3 displays the fuzzy average total inspection using Example 1.

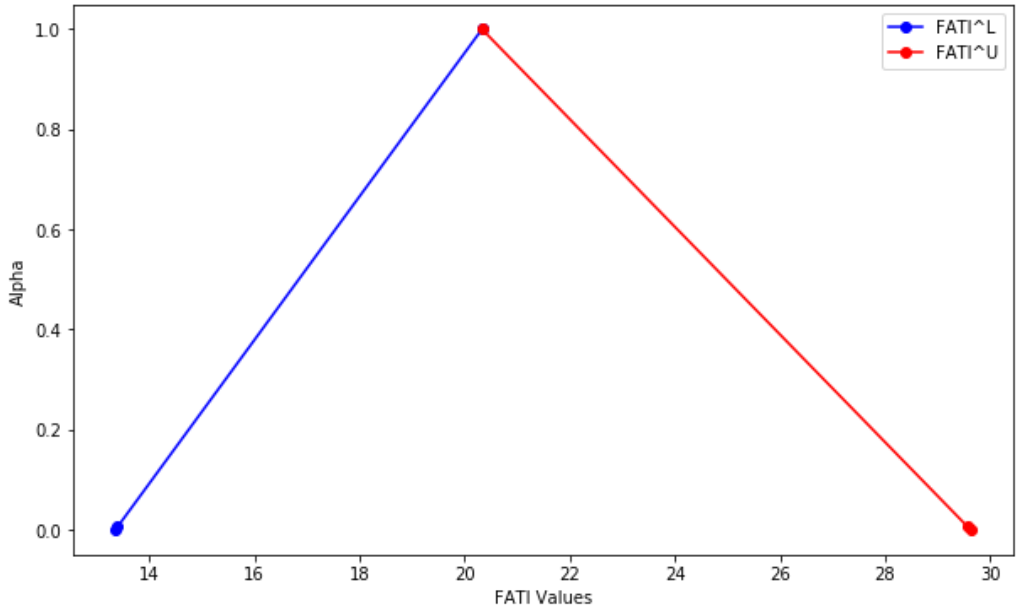


Figure 3: *FATI for DSP parameters N=200, c_1=0, c_2=1, n_1=n_2=10*

As formulated based on the structural definition given for $\tilde{p}$ in this section, the FATI band can be plotted on the basis of $\tilde{p}$ with the range upper and lower boundaries. The width of this band depends on the amount of variation in the proportion parameter as a result of this. A lesser level of uncertainty produces a smaller band, and when the proportion parameter is highly accurate, the upper and lower limits are similar, suggesting that the AOQ and ATI curves return to their classic shape. The FATI band increases in proportion to the number of defective items in the input batch.

Illustration 3

Considering the following case scenario: $N = 200, n_1 = 10, n_2 = 10, c_1 = 0, c_2 = 1, \tilde{\lambda} = n\tilde{p}$, $n = n_1 + n_2$, $b_2 = 0.01$, $b_3 = 0.02, \varphi = 0.0001$. From these parameters, the next findings have been calculating the FATI band.

$$\tilde{p}(\alpha) = [m + 0.01\alpha, m + 0.02 - 0.01\alpha]$$

$$FATI = \{200 - \varphi + (1-\varphi)e^{-10p}(190) - (1-\varphi)e^{-20p}(1800p)\}$$

For $\alpha = 0$ we get

$$FATI^{L}(0) = \{200 - \varphi + (1-\varphi)e^{-10m}(190) - (1-\varphi)e^{-20m}(1800m)\}$$

$$FATI^{U}(0) = \{200 - \varphi + (1-\varphi)e^{-(10\,m+0.2)}(190) - (1-\varphi)e^{-(20\,m+0.4)}1800(m+0.02)\}$$

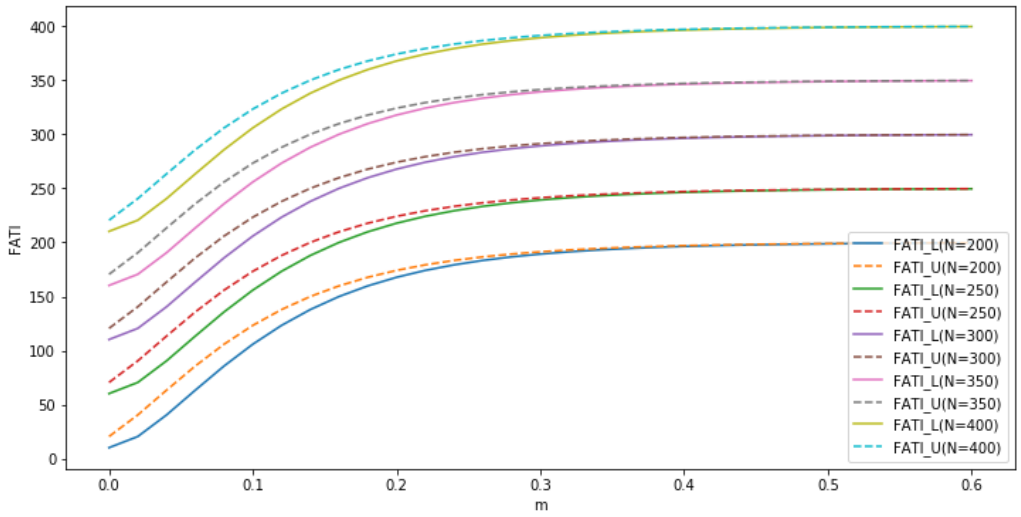


Figure 4: *FATI bands for DSP with fuzzy parameter c_1=0, c_2=1, n_1=n_2=10*

Figure 4 and Table 3 show five FATI bands for N = 200, 250, 300, 350, and 400. These bands illustrate how FATI grows with the amount of defective products. The data demonstrate that when process quality reduces, the FATI band decreases. It has also been found that when process quality is good, FATI approaches the sample size, however when process quality is very low and most lots are rejected, leading FATI to get near the size of the entire lot.

Table 3: *FATI for DSPs with fuzzy parameter c_1=0, c_2=1, n_1=n_2=10*

M	FATI for N = 200	FATI for N = 250	FATI for N = 300	FATI for N = 350	FATI for N = 400
0	10.1890, 20.4883	60.1890, 70.4883	110.1890, 120.4883	160.1890, 170.4883	210.1890, 220.4883
0.02	20.4883, 40.4462	70.4883, 90.4462	120.4883, 140.4462	170.4883, 190.4462	220.4883, 240.4462
0.04	40.4462, 63.3326	90.4462, 113.3326	140.4462, 163.3326	190.4462, 213.3326	240.4462, 263.3326
0.06	63.3326, 85.6678	113.3326, 135.6678	163.3326, 185.6678	213.3326, 235.6678	263.3326, 285.6678

0.08	85.6678, 105.8358	135.6678, 155.8358	185.6678, 205.8358	235.6678, 255.8358	285.6678, 305.8358
0.1	105.8358, 123.2538	155.8358, 173.2538	205.8358, 223.2538	255.8358, 273.2538	305.8358, 323.2538
0.12	123.2538, 137.8836	173.2538, 187.8836	223.2538, 237.8836	273.2538, 287.8836	323.2538, 337.8836
0.14	137.8836, 149.9492	187.8836, 199.9492	237.8836, 249.9492	287.8836, 299.9492	337.8836, 349.9492
0.16	149.9492, 159.7796	199.9492, 209.7796	249.9492, 259.7796	299.9492, 309.7796	349.9492, 359.7796
0.18	159.7796, 167.7240	209.7796, 217.7240	259.7796, 267.7240	309.7796, 317.7240	359.7796, 367.7240
0.2	167.7240, 174.1105	217.7240, 224.1105	267.7240, 274.1105	317.7240, 324.1105	367.7240, 374.1105
0.22	174.1105, 179.2281	224.1105, 229.2281	274.1105, 279.2281	324.1105, 329.2281	374.1105, 379.2281
0.24	179.2281, 183.3220	229.2281, 233.3220	279.2281, 283.3220	329.2281, 333.3220	379.2281, 383.3220
0.26	183.3220, 186.5948	233.3220, 236.5948	283.3220, 286.5948	333.3220, 336.5948	383.3220, 386.5948
0.28	186.5948, 189.2117	236.5948, 239.2117	286.5948, 289.2117	336.5948, 339.2117	386.5948, 389.2117
0.3	189.2117, 191.3058	239.2117, 241.3058	289.2117, 291.3058	339.2117, 341.3058	389.2117, 391.3058
0.32	191.3058, 192.9835	241.3058, 242.9835	291.3058, 292.9835	341.3058, 342.9835	391.3058, 392.9835
0.34	192.9835, 194.3294	242.9835, 244.3294	292.9835, 294.3294	342.9835, 344.3294	392.9835, 394.3294
0.36	194.3294, 195.4108	244.3294, 245.4108	294.3294, 295.4108	344.3294, 345.4108	394.3294, 395.4108
0.38	195.4108, 196.2812	245.4108, 246.2812	295.4108, 296.2812	345.4108, 346.2812	395.4108, 396.2812
0.4	196.2812, 196.9829	246.2812, 246.9829	296.2812, 296.9829	346.2812, 346.9829	396.2812, 396.9829
0.42	196.9829, 197.5494	246.9829, 247.5494	296.9829, 297.5494	346.9829, 347.5494	396.9829, 397.5494
0.44	197.5494, 198.0075	247.5494, 248.0075	297.5494, 298.0075	347.5494, 348.0075	397.5494, 398.0075
0.46	198.0075, 198.3785	248.0075, 248.3785	298.0075, 298.3785	348.0075, 348.3785	398.0075, 398.3785
0.48	198.3785, 198.6793	248.3785, 248.6793	298.3785, 298.6793	348.3785, 348.6793	398.3785, 398.6793
0.50	198.6793, 198.9234	248.6793, 248.9234	298.6793, 298.9234	348.6793, 348.9234	398.6793, 398.9234
0.52	198.9234, 199.1219	248.9234, 249.1219	298.9234, 299.1219	348.9234, 349.1219	398.9234, 399.1219

Conclusion

In conclusion, the integration of acceptance double sampling plans with integrates of fuzzy logic and ZIP distribution greatly improves the majority of methods in quality control. The use of the FAOQ band which has an upper and lower limits captures the fluctuation in defect proportions. The scenario of fuzzy logic resembles that ambiguity in the probability of defects is allowed and that results in most effective and accurate controlling. The ZIP distribution, is suited for higher number of defect free items, very similar to the stated FAOQ and FATI behavior. Acceptance double sampling plans enhance reliability by manage defect rejection and acceptance risks. This combination method gives an adequate basis for monitoring and calculating defect percentage, resulting in enhanced quality control and approaches to decision-making.

References

[1] Bohning, D., Dietz, E., and Schlattmann, P. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society*, 162(2):195–209.

[2] Buckley, J. J. (2003). *Fuzzy Probability*. Heidelberg, Germany.

[3] Duncan, A. J. (1986). *Quality Control and Industrial Statistics*. Homewood: Richard D. Irwin, Inc.

[4] Chakraborty, T. K. (1994). A class of single sampling plan based on fuzzy optimization. *Opsearch*, 63(1):35–43.

[5] Ezzatallah, B., and Bahram, S. G. (2011). Acceptance double sampling plan using fuzzy Poisson distribution. *World Applied Sciences Journal*, 15(12):1692–1702.

[6] Kavithanjali, S., Sheik Abdullah, A., and Kamalanathan, R. (2004a). A review study on various sampling plans. *Journal of Inventive and Scientific Research Studies*, 1(2):15–28.

[7] Kavithanjali, S., and Sheik Abdullah, A. (2024b). Python implementation of fuzzy logic for zero-inflated Poisson single sampling plans. *Reliability: Theory & Applications*, 3(79):275–281.

[8] Kaviyarasu, and Thottathil, A. T. (2018). Designing STDS plan for zero-inflated Poisson distribution through various quality level. *International Journal of Statistics and Applied Mathematics*, 3(4):44–53.

[9] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 38(4):338–353.

[10] Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

[11] Malathi, and Muthulakshmi. (2012). Special double sampling plan with fuzzy parameter. *Indian Journal of Applied Research*, 2(1):117–120.

[12] McLachlan, G., and McLachlan, G. (2000). *Finite Mixture Models*. New York: John Wiley & Sons.

[13] Naya, H., et al. (2008). A comparison between Poisson and zero-inflated Poisson regression models with an application to number of black spots in Corriedale sheep. *Genetics Selection Evolution*, 40(1):379–393.

[14] Ridout, M., et al. (1998). Models for count data with many zeros. *International Biometric Conference*, Cape Town.

[15] Schilling, E. G., and Neubauer, D. V. (2009). *Acceptance Sampling in Quality Control*. Boca Raton: CRC Press.

[16] Tamaki, F., Kanagawa, H., and Ohta, H. (1991). A fuzzy design of sampling inspection plans by attributes. *Japanese Journal of Fuzzy Theory and Systems*, 315–327.

[17] Xie, M., et al. (2001). Zero-inflated Poisson model in statistical process control. *Computational Statistics & Data Analysis*, 38(2):191–201.

ACCEPTANCE SAMPLING PLAN BASED ON TRUNCATED LIFE TESTS FOR RAYLEIGH DISTRIBUTION

C. Geetha[1*], Pachiyappan D[2], and Srividhya K[3]

•

[1]Associate Professor, Department of Statistics, Government Arts college,
Salem-7, Tamil Nadu, India.
Correspondence Email: geethanisha1@gmail.com
[2]Department of Statistics, Periyar University,
Salem-11, Tamil Nadu, India.
d.pachiyappan321@gmail.com
[3]Associate Professor of Biostatistics, School of Allied Health Sciences,
Vinayaka Mission's Research Foundation-DU, Salem, Tamil Nadu, India.
srividhyastat@gmail.com

Abstract

This paper addresses the problem of designing an acceptance sampling plan for a truncated life test where the lifetime of the product follows a generalized Rayleigh distribution. The study identifies the minimum sample sizes needed to ensure the specified mean life for various acceptance numbers, confidence levels, and ratios of the fixed experiment time to the specified mean life. The operating characteristic values of the sampling plans, along with the producer's risk, are discussed. Additionally, tables are provided to facilitate the application of these sampling plans, and a numerical example is included to illustrate the use of these tables.

Keywords: Consumer's risk; normal distribution, Rayleigh distribution, operating characteristic curve.

I. Introduction

Lifetime is an important quality variable of a product. Sampling plans used to determine the acceptability of a product, with respect to its lifetime, are known as reliability sampling plans. In most life test sampling plans, a common constraint is the duration of the total time spent on the test. When the lifetime of a product is expected to be high, it might be exceedingly time consuming to wait until all items fail. Therefore, it is usual to terminate a life test by a pre-assigned time t and note the number of failures. One purpose of these tests is to set a confidence limit on the mean life. If a confidence limit on the mean life is set, it is then desired to establish a specified mean life, μ_0, with a given probability of at least P^*. The decision to accept the specified mean life occurs if and only if the observed number of failures at the end of the pre-determined time t does not exceed a given acceptance number c. That is, if the number of failures exceeds c, one can terminate the test before the time t and reject the lot. Such a test is called the truncated life test. The problem considered is that of finding the smallest sample size necessary to assure a certain mean life based on the truncated life test. The sampling plan consists of the number of items n on test, the acceptance number c, and the ratio t/μ_0. For a fixed P^*, such a sampling plan is characterized by the triplet $(n, c, t/\mu_0)$.

Acceptance sampling plans based on truncated life tests were developed by [1] for exponential

distribution; by [2] for Weibull distribution; by [3] for gamma distribution; by [4] for half logistic distribution, and by [5] for log-logistic distribution. The present paper extends these to the generalized Rayleigh distribution.

The rest of this paper is organized as follows. In the next section we introduce the generalized Rayleigh distribution briefly. In the third section, an acceptance sampling plan for the truncated life test based on the generalized Rayleigh distribution is developed, and some tables are then established. In the fourth section, a numerical example is provided to illustrate the use of the sampling plan. Some conclusions are made in the final section.

2. Generalized Rayleigh Distribution

The Rayleigh distribution was originally derived by Rayleigh (1880) in connection with a problem in the field of acoustics, and was used to be a lifetime distribution in reliability for recent years. The Rayleigh distribution is a special case of the Weibull distribution and has wide applications, such as in communication engineering [6] in life testing of electro-vacuum devices [7]. The probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of the Rayleigh distribution, respectively, are given by

$$f(t;b) = \frac{t}{b^2} e^{-\frac{t^2}{2b^2}} \tag{1}$$

and

$$F(t;b) = 1 - e^{-\frac{t^2}{2b^2}},\ t > 0 \tag{2}$$

where b>0 is the scale parameter. An important characteristic of the Rayleigh distribution is that its failure rate is an increasing linear function of time. This property makes it a suitable model for components that possibly have no manufacturing defects but age rapidly [4] with time. [8] considered a generalized version of the Rayleigh distribution. The generalized Rayleigh distribution has p.d.f.

$$f_k(t;\beta) = \frac{2}{\Gamma(k+1)\beta^{k+1}} t^{2k+1} e^{\frac{t^2}{\beta}},\ t > 0 \tag{3}$$

where $k \geq 0$ is the shape parameter and $\beta > 0$ is the scale parameter. When $k = 0$ and set $\beta = 2b^2$, the generalized Rayleigh distribution reduces to the Rayleigh distribution in equation (1). The c.d.f. of T is given by

$$F_k(t;\beta) = \Gamma_{I,k+1}\left(\frac{t^2}{\beta}\right) \tag{4}$$

where $\Gamma_{I,k}(v) = 1/\Gamma(\kappa)\int_0^v x^{\kappa-1}e^{-x}dx$ is known as the incomplete gamma function. If k is an integer, the c.d.f. in equation (3) reduces to

$$F_k(t;\beta) = 1 - \sum_{j=0}^{k} \frac{(t^2/\beta)^j e^{(-t^2/\beta)}}{j!} \tag{5}$$

The rth moment of the generalized Rayleigh distribution is

$$E(T^r) = \frac{\Gamma(k+r/2+1)}{\Gamma(k+1)} \beta^{r/2},\ r = 1,2,\ldots \tag{6}$$

Then, the mean and variance of the generalized Rayleigh distribution are $\mu = m\sqrt{\beta}$ and $\sigma^2 = (k+1-m^2)\beta$, respectively, where $m = \Gamma(k+3/2)/\Gamma(k+1)$.

3. Design of Acceptance Sampling Plan

In the postulated sampling plan, a lot is sentenced to be bad if the true mean life of items, μ, is below the specified value μ_0. In other words, a lot is sentenced to be good if the true mean life of items is at least μ_0. Therefore, the consumer's risk is the probability of accepting a bad lot, and the producer's risk is the probability of rejecting a good lot.

In this study, we fixed the consumer's risk not to exceed $1 - P^*$. It should be noted here that we consider a lot of infinitely large size so that binomial distribution theory can be applied and the acceptance or rejection of the lot are equivalent to the acceptance or rejection of the hypothesis $\mu \geq \mu_0$. We assume that the lifetime follows a generalized Rayleigh distribution. The hypothesis $\mu \geq \mu_0$ is then equivalent to $\beta \geq \beta_0$. Given a value of $P^*(0 < P^* < 1)$, a value of t/μ_0, and a value of the acceptance number c, then the required n is the smallest positive integer satisfying the inequality.

$$\sum_{i=0}^{c} \binom{n}{i} p_0^i (1 - p_0)^{n-i} \leq 1 - P^* \tag{7}$$

where $p_0 = F_k(t; \beta_0)$ is given by equation (3) and it is the probability that the lifetime does not exceed t if the true mean life is $\mu_0 = m\sqrt{\beta_0}$. Since the c.d.f. $F_k(t, \beta)$ depends only on the ratio $t/\sqrt{\beta}$, the experimenter needs to specify only this ratio. If the number of observed failures is less than or equal to c, then from equation (4) we can make the confidence statement that $F_k(t; \beta) \leq F_k(t; \beta_0)$ with probability P^*. Since the first derivative of $F_k(t; \beta)$ with respect to $t/\sqrt{\beta}$ is positive, the c.d.f $F_k(t; \beta)$ is a monotonically increasing function of $t/\sqrt{\beta}$. It follows that

$$F_k(t; \beta) \leq F_k(t; \beta_0) \iff \beta \geq \beta_0 \ (\text{or } \mu \geq \mu_0) \tag{8}$$

The minimum values of n satisfying equation (4) can be obtained and are given in Table 1 for $k = 0, P^* = 0.75, 0.90, 0.95$, and $t/\mu_0 = 0.4, 0.6, 0.8, 1.0, 1.5$, and 3.0. This choice of t/μ_0 is consistent with the corresponding tables of [3] for a gamma distribution and [9] for a half logistic distribution. For the cases of small values of p_0 and large values of n, the binomial distribution is approximated by Poisson distribution with parameter $\lambda_0 = np_0$. Thus, equation (4) can be rewritten as

$$\sum_{i=0}^{c} \frac{e^{-\lambda_0} \lambda_0^i}{i!} \leq 1 - P^* \tag{9}$$

The operating characteristic function of the sampling plan $\left(n, c, t/\sqrt{\beta_0}\right)$ is the probability of accepting a lot and is given by

$$L(p) = \sum_{i=0}^{c} \binom{n}{i} p^i (1 - p)^{n-i} \tag{10}$$

Table 1: *Minimum sample size n to be tested for a time t in order to assert with probability P^* and acceptance number c that μ≥μ_0 (k=0)*

P^*	c	t/μ_0					
		0.4	0.6	0.8	1.0	1.5	2.0
0.75	0	12	5	3	2	1	1
	1	22	11	6	4	3	2
	2	33	15	9	7	4	3
	3	43	20	12	9	5	4
	4	52	25	15	11	7	5
	5	62	29	18	13	8	6
	6	72	34	21	15	9	8
0.90	0	19	9	5	3	2	1
	1	32	15	9	6	3	2
	2	44	20	12	8	5	4
	3	55	26	15	11	6	5
	4	66	31	19	13	8	6
	5	77	36	22	15	9	7
	6	87	41	25	17	10	8
0.95	0	24	11	6	4	2	1

1	39	18	10	7	4	3
2	52	24	14	10	5	4
3	64	29	18	12	7	5
4	75	35	21	14	8	6
5	87	40	24	17	10	7
6	98	46	27	19	11	8

where $p = F_k(t;\beta) = \Gamma_{I,k+1}\left((t^2/\beta_0)\beta_0/\beta\right)$. If the producer's risk α is given and a sampling plan ($n, c, t/\mu_0$) is adopted, one interesting question is what value of $\mu/\mu_0 (> 1)$ will insure the producer's risk equal to or less than α. For a given producer's risk, say $\alpha = 0.05$, the smallest values of μ/μ_0 satisfying the inequality $L(p) \geq 0.95$ were found and are given in Table 2 for $k = 0, P^* = 0.75,0.90,0.95,0.99$ and the selected values of t/μ_0. Tables 1 and 2 can be generated for any other value of k. A computer program is available with the authors. A numerical example is provided in the next section to illustrate the use of these tables.

4. An Example

Assume that the lifetime distribution of the test items is a generalized Rayleigh distribution with shape parameter $k = 0$ and we wish to establish that the mean life μ_0 is at least 1000 hours with probability $P^* = 0.90$. Suppose the experimenter desires to terminate the life test at $t = 600$ hours and set an acceptance number of $c = 3$. The required n is the entry in Table 1 corresponding to the values of $P^* = 0.90, c = 3$ and $t/\mu_0 = 600/1000 = 0.6$. This number is $n = 26$. That is, 26 items have to be put on test. If no more than three failures are observed during 600 hours, then the experimenter can assert that the mean life of the items is at least 1000 hours with a confidence level of 0.90.

Table 2: *Minimum ratio of μ/μ_0 for the acceptability of a lot with producer's risk of* $0.05 (k = 0)$

P^*	c	t/μ_0					
		0.4	0.6	0.8	1.0	1.5	2.0
0.75	0	5.43	5.25	5.43	5.54	5.87	7.83
	1	2.76	2.89	2.79	2.77	3.49	3.53
	2	2.22	2.20	2.22	2.39	2.49	2.62
	3	1.96	1.96	1.96	2.07	2.06	2.22
	4	1.79	1.82	1.82	1.88	2.06	1.99
	5	1.70	1.70	1.72	1.77	1.86	1.84
	6	1.63	1.64	1.66	1.68	1.72	2.05
0.90	0	6.83	7.05	7.00	6.78	8.31	7.83
	1	3.34	3.40	3.47	3.48	3.49	3.53
	2	2.58	2.57	2.60	2.59	2.91	3.32
	3	2.22	2.26	2.23	2.33	2.37	2.74
	4	2.03	2.04	2.08	2.09	2.28	2.41
	5	1.90	1.91	1.94	1.93	2.05	2.20
	6	1.73	1.74	1.76	1.74	1.89	1.93
0.95	0	7.67	7.79	7.67	7.83	8.31	7.83
	1	3.69	3.74	3.67	3.79	4.15	4.65
	2	2.80	2.82	2.83	2.94	2.91	3.32
	3	2.40	2.39	2.47	2.45	2.64	2.74
	4	2.16	2.18	2.20	2.18	2.28	2.41
	5	2.02	2.02	2.03	2.08	2.22	2.20
	6	1.91	1.93	1.92	1.95	2.03	2.05

In general, almost all the values of the required number n tabulated by us are found to be less than those tabulated by [4] for a gamma distribution and for a half logistic distribution. The table 2 gives the values of μ/μ_0 in order that the producer's risk may not exceed $\alpha = 0.05$. Thus, for the above example, the values of μ/μ_0 for $c = 0,1,2,3,4,5$ are $7.05,3.40,2.57,2.26,2.04$ and 1.91, respectively. The consideration of the actual mean life necessary in order to ship 95% of the lots will

play the key role in deciding which c to be selected. For example, if the actual mean life μ is about 2000 hours (that is, the value of μ/μ_0 is about 2.0) based on the production conditions, and the experimenter desires to terminate the test at $t = 400$ hours under the producer's risk 0.05 and the consumer's risk 0.10 (or $P^* = 0.90$), it follows that $c = 4$ from Table 2. The required n in Table 1 corresponding to the values of $P^* = 0.90, c = 4$ and $t/\mu_0 = 0.4$ is $n = 66$. Hence, a sampling plan $(n, c, t/\mu_0) = (66,4,0.4)$ is taken.

5. Conclusion

In this paper develops an acceptance sampling plan based on a truncated life test, assuming the life distribution of test items follows a generalized Rayleigh distribution. The provided tables facilitate the practical application of the suggested plans, enabling practitioners to implement them conveniently. The real time applications of the proposed sampling plan are given using the industrial data. The proposed sampling plan can also be used in testing of software.

Discloser statement

The authors declare no potential conflict of interest.

References

[1] Kantam, R.R.L., Rosaiah, K., & Rao, G. S. (2001) Acceptance sampling based on life tests: log-logistic model, Journal of Applied Statistics, 28, 121-128.

[2] Goode, H. P., and Kao, J. H. K., (1961). Sampling plans based on the Weibull distribution, Proceedings of Seventh National Symposium on Reliability and Quality Control, Philadelphia, Pennsylvania. pp. 24-40.

[3] Gupta, S. S., and Groll, P. A., (1961). Gamma distribution in acceptance sampling based on life tests. Journal of the American Statistical Association, 56, pp. 942-970.

[4] Baklizi, A., El Masri, Q.E.A., (2004). Acceptance sampling plan based on truncated life test in the Birnbaum-Saunders model. Risk Anal 24 (6), 1453-1457.

[5] Balakrishnan, N., Leiva, V., Lopez, G., (2007). Acceptance sampling plan based on the truncated life test in generalized BirnbaumSaunders distribution. Commun. Stat. -Simul. Comput. 36, 643656.

[6] Vodă, V. Gh., (1976). Inferential procedures on a generalized Rayleigh variate, I, Aplikace Matematiky, 21, pp. 395-412.

[7] Jun, C.-H., Balamurali, S., Lee, S.-H., (2006). Variables sampling plans for Weibull distributed lifetimes under sudden death testing. IEEE Trans. Reliab. 55 (1), 53-58.

[8] Kantam, R.R.L., Rosaiah, K. & Rao, G. S., (2001). Acceptance sampling based on life tests: log-logistic model, Journal of Applied Statistics, 28, pp. 121-128.

[9] Vodă, V. Gh., (1976). Inferential procedures on a generalized Rayleigh variate, I, Aplikace Matematiky, 21, pp. 395-412.

USING SENSORS TO MONITOR THE CONDITION AND SAFETY OF ELECTRICAL EQUIPMENT

I.N. Rahimli[1], A.L. Bakhtiyarov[2], G.K. Abdullayeva[3]

•

Azerbaijan State Oil and Industry University, Baku, Azerbaijan
[1] ilham.raqimli@asoiu.edu.az ; [3] gulshan.abdullayeva@asoiu.edu.az

Abstract

The article explores the important role that modern sensors play in ensuring the safety and efficient operation of electrical equipment. Advances in sensor technologies make it possible to effectively monitor the condition of equipment, identify potential problems and prevent accidents. The article examines the principles of operation of sensors, their diversity and application in various areas of the electric power industry. Particular attention is paid to predictive maintenance technologies, which allow optimizing resources and increasing the reliability of electrical equipment. Ultimately, the use of sensors to monitor the condition and safety of electrical equipment leads to reduced risk of accidents and increased efficiency of electrical power systems.

Keywords: sensor, vibration, electrical equipment monitoring.

I. Introduction

In the modern world, the electric power industry plays a key role in meeting the energy needs of society. However, as the complexity of electrical power systems increases, their requirements for reliability and safety increase. Failures and accidents in electrical equipment can lead to serious consequences for both electricity consumers and the environment. In this context, the use of modern sensors to monitor the condition and safety of electrical equipment becomes a necessity.

Sensors play an important role in ensuring the efficient and safe operation of electrical equipment. They allow you to continuously monitor various parameters such as temperature, humidity, pressure, vibration and electrical parameters, allowing you to quickly identify any anomalies or deviations from the norm [1-3]. Thanks to modern technologies for collecting and analyzing data, the information obtained can be used to predict the likelihood of equipment failure and take appropriate measures to maintain or replace it.

The purpose of this article is to consider the role of sensors in ensuring the safety and efficiency of electrical equipment. We will discuss how sensors work, their variety and application in various applications in the power industry, as well as the benefits they bring in predictive maintenance and optimization of power systems. As a result, we will be able to understand how the use of sensors helps improve the safety and reliability of electrical equipment, as well as reduce the risks of emergency situations.

II. Formulation of the problem

Technical diagnostics are intended to improve the quality, reliability and increase the service life of mechanisms, machines and equipment. The need to use technical diagnostics is caused by

the constant increase in complexity and increase in the level of automation of modern enterprises, factories, plants, thermal and nuclear power plants, sea, air, railway and other modes of transport, etc.

According to the international confederation IMECO, the introduction of continuous technical diagnostics into rotary equipment of companies reduces labor intensity and repair time by more than 40%, reduces fuel consumption of energy companies by 4% and increases the technical utilization rate of equipment by more than 12%. Experience in operating similar systems at a nuclear power plant at the end of the 1990s. allowed to reduce costs by 3 million US dollars within one year and generate additional income in the amount of 19 million by reducing equipment downtime.

Taken together, the introduction of technical diagnostic tools allows:

- prevent accidents;
- prevent and increase the reliability of machinery and equipment;
- increase their durability, reliability and service life;
- increase productivity and production volume;
- predict residual life;
- reduce the duration of downtime with the assessment of additional profits;
- reduce time spent on repair work;
- reduce operating costs;
- reduce the number of service personnel;
- optimize the number of spare parts;
- reduce insurance costs.

The sensors themselves are practically useless - they simply collect huge amounts of information. To store this data, structure, process, analyze and use it for production management, you need special software. It can be installed on company servers or deployed in the cloud.

Examples of standard production monitoring software would be:

1. MES systems. These are production process control systems. They monitor production capacity utilization, help track defects, and alert personnel to production problems, such as material shortages or production process irregularities.

2. Scada. This is software that collects and visualizes information about the operation of equipment. With its help, you can monitor the performance of individual machines or the entire workshop as a whole, control equipment remotely and create reports for management and analysts.

There are also other platforms with additional features that can replace or complement MES and Scada. For example, VK Cloud (formerly MCS) has a ready-made platform for building an equipment monitoring system. It is deployed in the cloud and allows you to rent any capacity needed for monitoring, as well as develop your own solutions to optimize production. The platform can collect and store information from sensors, visualize data and notify about events [8-11] .

With the development of microelectronics, small-sized measuring and diagnostic systems, integrated and installed directly into equipment, have developed. Work in this area is carried out by the world's leading companies: Schneider Electric, Siemens, General Electric, BALLUFF BCM, etc.

III. Problem solution

In production, especially in mechanical engineering, there are extremely many points where monitoring of wear or contamination of components is required. Indicators of these parameters are

usually vibration and temperature.

It is difficult to implement a system for precise control of vibration and temperature with 100% coverage of all points. There is a need for inexpensive mass indicators that can identify points that require close attention.

At the same time, there is a need for a convenient presentation of a large amount of data for further analysis. The BALLUFF company has a solution in this direction.

Objective information about the degree of wear and contamination of components is important for optimal planning of equipment maintenance, as well as preventing unscheduled shutdowns.

Measuring vibration and temperature parameters in this case does not require high absolute accuracy, which today is provided by modern highly specialized vibration control systems.

Such control often comes together with temperature control. With a large number of control points, problems arise with the placement of sensors and wiring of sensors. Then the task of miniaturizing sensors and optimizing signal transmission lines arises.

Vibration control involves measuring about 20 parameters that require additional processing to be analyzed. Usually this function is performed by an external controller, which also complicates the solution.

BALLUFF has brought to market a revolutionary solution by combining vibration and temperature, humidity and atmospheric pressure measurement, as well as a microprocessor in one package. All this is represented by a series of BCM sensors.

The BCM sensor is a complex device that includes individual parameter monitoring modules.

The vibration measurement is made using a microelectromechanical system (MEMS) element, a technology now widely used, including in smartphones, industrial inclination sensors and accelerometers. Microelectronic technologies have also been applied to measure other parameters, making it possible to combine all measurements into one miniature device.

Data processing is carried out by a built-in processor, which creates a data array convenient for analysis and transmission. In addition to the values of certain parameters, this array also includes the event log.

The problem of transferring a large amount of data from one sensor was solved within the framework of the standard IO-Link digital interface, which is gaining momentum in mechanical engineering, designed to connect sensors to a master module in a point-to-point manner (Fig. 1). Communication between the master module and upper-level equipment is carried out using industrial network interfaces, such as, for example, Ethernet/IP and Profinet.

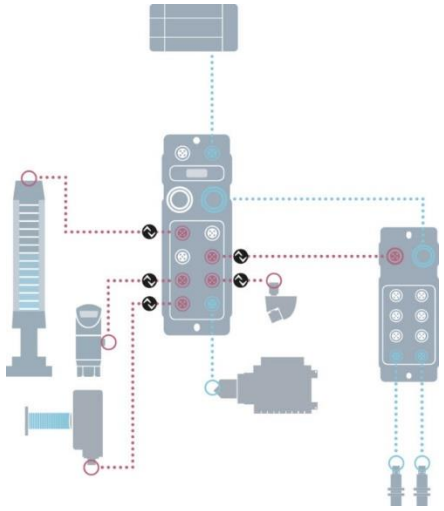


Figure 1: *Connecting equipment with IO-Link interface*

Using the BCM sensor, monitoring an increase in the average values of vibration parameters over a period of time will tell you when it is necessary to replace a bearing or change the oil in a critical mechanical unit; temperature will indicate a possible malfunction of the cooling systems or the presence of parasitic friction; relative humidity will indicate the need to make changes to the technical process for which it may be important. Today, an equipment condition monitoring system can be effectively used in almost every production facility.

Thus, in addition to the transparency of production processes (Fig. 2), a reduction in downtime is achieved both for manual inspection of critical components and for performing repair work.

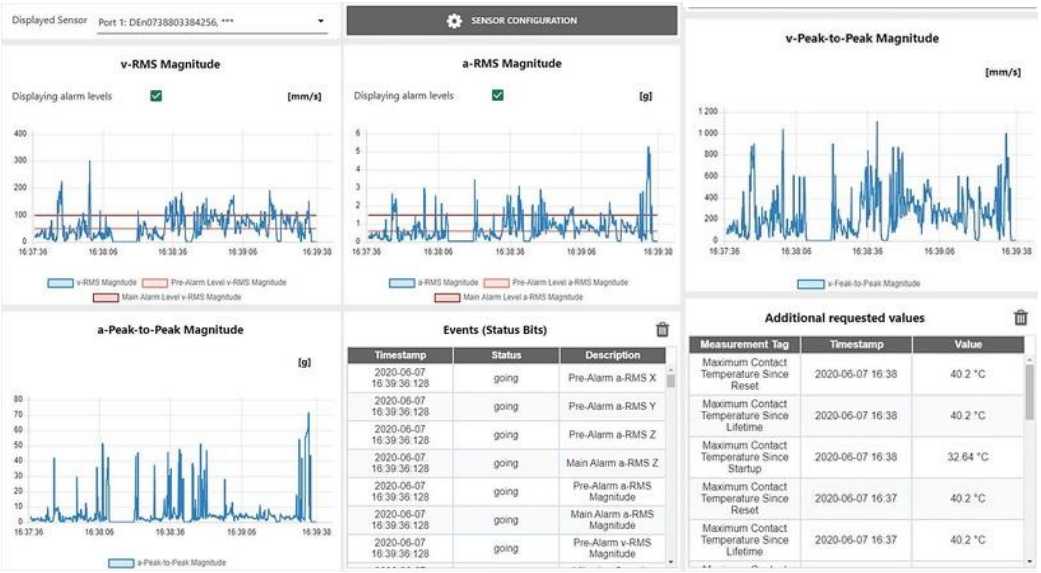


Figure 2: *Visualization of equipment condition monitoring*

Sharp changes in certain parameters or cumulative trends in increasing values can not be analyzed in the control system, but a corresponding logical alarm signal of two levels (conditional "yellow" and "red" alarm) can be received from the BCM sensor, thereby optimizing the program code of the monitoring system and ensuring convenience for the operator, who can quickly make decisions about the need to intervene in the technical process.

Thus, in addition to the transparency of production processes, a reduction in downtime is achieved both for manual inspection of critical components and for repair work.

In addition, there is the possibility of implementing a completely autonomous condition monitoring system, which can automatically create a schedule of repair and maintenance work based on sensor readings and thereby simplify and optimize the work of relevant specialists and departments at enterprises.

IV. Conclusions

In conclusion, we can emphasize the importance of using sensors to monitor the condition and safety of electrical equipment in the modern power industry. Our analysis showed that modern sensor technologies make it possible to quickly monitor various parameters of equipment operation, which in turn helps prevent accidents and failures, increase system efficiency and reduce maintenance and repair costs.

In addition, the use of sensors also contributes to the development of the concept of predictive maintenance, which allows optimizing the process of equipment maintenance and repair, reducing the time and financial costs of its maintenance.

Overall, the use of sensors to monitor the condition and safety of electrical equipment is an important step towards improving the reliability and efficiency of electrical power systems, which is key to ensuring the stable functioning of energy infrastructure and meeting the needs of modern society.

References

[1] Cease T.W. and Johnston P., "A magneto-optic current transducer", IEEE Trans. on PowerDelivery, vol. 5, 548-555, 1990.
[2] Bohnert K., Gabus G., Nehring J., and Brändle H., "Temperature and vibration insensitive fiber-optic current sensor", J. of Lightw.Technol., vol. 20, 267-276, 2002.
[3] Nicatti P.A. and Robert P., "Stabilized current sensor using a Sagnac interferometer", J. Phys.E: Sci. Instrum., vol. 21, 791-796, 1988.
[4] Frosio G., Hug K., and Dändliker R., "All-fiber Sagnac current sensor", in Opto'92 (ESIPublications, Paris 1992), 560-564.
[5] S.V. Rzayeva, N.S. Mammadov and N.A. Ganiyeva "Overvoltages During Single-Phase Earth Fault in Neutral-Isolated Networks (10÷35) kV". Journal of Energy Research and Reviews 13 (1), pp. 7-13, 2023
[6] S.V. Rzayeva, N.S. Mammadov, N.A. Ganiyeva "Neutral grounding mode in the 6-35 kv network through an arcing reactor and organization of relay protection against single-phase ground faults". Deutsche internationale Zeitschrift für zeitgenössische Wissenschaft / German International Journal of Modern Science №42(22) pp.31-34
[7] Enokihara A., Izutsu M., and Sueta T., "Optical fiber sensors using the method of polarizationrotatedreflection", J. Lightw.Technol., vol. 5, 1584-1590, 1987.
[8] Short S.X., Tselikov A.A., J.U. de Arruda, and Blake J.N., "Imperfect quarter-waveplatecompensation in Sagnac interferometer-typecurrent sensors", J. Lightw.Technol., vol. 16,1212-1219, 1998.
[9] Blake J., Tantaswadi P., and R.T. de Carvalho,"In-line Sagnac interferometer current sensor",IEEE Trans. Power Delivery, vol. 11, 116-121,1996.
[10] Tang D., Rose A. H., Day G. W., and Etzel S. M., "Annealing of linear birefringence insingle-mode fiber coils: applications to opticalfiber current sensors", J. Lightw. Technol., vol., 1031-1037, 1991.
[11] Short S.X., Tselikov A.A., J.U. de Arruda, and Blake J.N., "Imperfect quarterwaveplatecompensation in Sagnac interferometer-type current sensors", J. Lightw. Technol., vol. 16,1212-1219, 1998.
[12] Bohnert K. and Nehring J., "Fiber-optic sensing of voltages by line integration of the electricfield", Opt. Lett., vol. 14, 290-292, 1989.
[13] V. D. Gavrichev, A. L. Dmitriev, Fiber-optic sensors magnetic field / Study Guide. - St. Petersburg: SPbNIU ITMO, 2013.p. 83
[14] E. Udd. Fiber optic sensors. M.: Technosfera, 2008. - 520s.
[15] R.G. Jackson. Latest sensors. M: Technosphere, 2007.- 384p.
[16] S. H. Horowitz and A. G. Phadke, "Current and voltage transformers", in Power System Relaying, 4th ed. John Wiley and Sons, 2014, ch. 3, pp. 46–73.
[17] Instrument transformers - Part 2: Additional requirements for current transformers, IEC 61869-2:2012, 2012.
[18] Jackson, D.A. An optical system with potential for remote health monitoring of subsea machinery.Meas. Sci. Technol. 2009, 20, 1–8.
[19] Perciante, C.D.; Ferrari, J.A. Magnetic crosst alk minimization in optical current sensors. IEEETrans. Instrum. Meas. 2008, 57, 2304–2308
[20] Drexler, P.; Fiala, P. Utilization of Faraday mirror in fiber optic current sensors. Radioengineering2008, 17, 101–107.

DIAGNOSTICS OF ELECTRICAL EQUIPMENT AT THERMAL PLANTS

R.K. Karimova[1], H.S. Piriyev[2]

•

Azerbaijan State Oil and Industry University, Baku, Azerbaijan
[1]rashida.karimova@asoiu.edu.az , [2]hamid.piriyev.s@asoiu.edu.az

Abstract

Diagnostics of electrical equipment at thermal power plants plays a key role in ensuring reliable operation of power systems. This article examines methods and technologies for diagnosing electrical equipment at thermal power plants and their significance for ensuring the reliability of power systems. The work analyzes the main approaches to diagnostics, including non-destructive methods, equipment condition monitoring and the use of modern technical means, such as infrared thermography and ultrasound diagnostics. Particular attention is paid to the importance of these methods for ensuring the uninterrupted operation of thermal power systems and minimizing the likelihood of emergency situations, which is important for ensuring energy security and economic efficiency.

Keywords: infrared thermography, ultrasound diagnostics, thermal imagers, monitoring system, diagnostics

I. Introduction

Electrical equipment plays a key role in the operation of thermal power plants, providing power supply, process automation, safety and equipment control. It is used to supply electricity to all components of the installation, including security, control and monitoring systems. Automation of processes such as fuel regulation, temperature and pressure control is carried out using electrical systems. Electrical equipment also plays a role in ensuring plant safety through emergency shutdown, fire extinguishing and equipment condition monitoring systems [1]. In modern installations, energy-saving technologies and electrical-based energy management systems help reduce energy consumption and improve overall system efficiency. Thus, electrical equipment is an integral part of thermal power plants, ensuring their stable, safe and efficient operation.

Equipment failures in thermal power plants can have a serious impact on production processes and safety. Firstly, they can lead to a decrease in productivity or a complete shutdown of the installation, which in turn can cause losses in energy production and a decrease in the efficiency of the entire system. This can result in increased downtime, lost revenue, and increased equipment recovery costs.

The impact of equipment failures on safety is also extremely serious. Failures can disrupt the normal functioning of safety systems such as fire extinguishing systems, emergency lighting, gas leak prevention systems, etc. This creates potential dangers for workers, the environment and

society as a whole. In some cases, equipment failures may be associated with emergency situations, such as accidents with the release of harmful substances or fires, which threaten not only the safety of personnel, but also the surrounding area and population.

Thus, equipment failures in thermal power plants have serious consequences for both production processes and safety, and require careful monitoring, maintenance and updating of equipment to prevent negative consequences [2-5].

II. Problem setting

Visual methods for checking equipment condition.

Visual methods for checking the condition of equipment play an important role in ensuring proper operation and process safety. These methods include visually inspecting the equipment for signs of wear, damage or other abnormalities, as well as assessing its environment. For example, engineers may check for cracks, corrosion, leaks, or other damage on the surface of equipment.

Another important visual method is to compare the current condition of the equipment with its standard or recommended condition. This may include comparison of indicators such as color, shape, size and position with specified standards or criteria for safety and effectiveness. For example, when visually inspecting a piping system, engineers can check for play, cracks, or leaks and compare them to the limits set for the system.

Visual methods may also involve the use of specialized equipment, such as infrared thermal cameras or ultrasonic flaw detectors, to detect potential problems that may not be visible during a normal visual inspection. These methods help to quickly identify possible malfunctions and prevent their development before serious problems or accidents occur.

Thus, visual methods for checking the condition of equipment play an important role in ensuring proper operation and process safety, allowing potential problems to be quickly identified and corrected before they lead to serious consequences.

Using Thermal Imaging Cameras to Detect Overheating.

Thermal imaging cameras are an effective tool for detecting overheating in a variety of systems and equipment, including thermal power plants. These cameras operate using infrared radiation, which allows them to visualize temperature differences on the surface of objects. Overheating can be an indicator of problems or malfunctions in the equipment, such as poor connections, overload, worn components or improper operation of the cooling system.

When using thermal imaging cameras, engineers can quickly scan equipment and identify areas of elevated temperature that may indicate problems. For example, overheating at electrical connections may indicate possible overloads or poor contacts, while overheating at the surface of pipes may indicate problems with the heat transfer or cooling system. With thermal imaging cameras, engineers can quickly identify potential problems, allowing them to take corrective action before serious damage or accidents occur.

Thus, the use of thermal imaging cameras to detect overheating is an effective method for diagnosing the condition of equipment in thermal power plants, allowing us to quickly identify potential problems and prevent their negative consequences.

Application of ultrasonic flaw detectors to detect cracks and defects.

Ultrasonic flaw detectors play an important role in detecting cracks and defects in various materials and structures, including thermal power plant equipment. These instruments use ultrasonic waves to penetrate a material and measure the time it takes for the waves to reflect from internal defects. The use of ultrasonic flaw detectors makes it possible to detect cracks, including the smallest ones, as well as other defects such as inclusions, pores or areas of altered density.

In the context of thermal power plants, ultrasonic flaw detectors are used to test the integrity of materials used in boilers, piping, tanks and other key components. They can detect defects that are not always visible visually, allowing you to quickly identify potential problems such as fatigue cracks, corrosion or structural changes.

The advantages of ultrasonic flaw detectors include high sensitivity to various defects, the ability to conduct deep investigations, and non-impact on the test object, which is especially important for maintaining the integrity and safety of equipment. Such defect detection methods are an important element of maintenance and ensure extended service life and safe operation of equipment in thermal power plants.

Monitoring and diagnostic systems based on IoT (Internet of Things) and data collection.

Monitoring and diagnostic systems based on IoT (Internet of Things) and data collection play an important role in improving the efficiency, reliability and safety of thermal power plants. Using sensors, data acquisition devices and information networks, these systems can continuously monitor equipment performance, environmental conditions and other key indicators.

IoT systems allow you to collect data in real time and transmit it to remote servers for analysis and processing. This makes it possible to quickly respond to changes in equipment condition, identify potential problems and prevent accidents. For example, using IoT systems you can monitor temperature, pressure, vibration and other parameters of equipment operation, as well as control the levels of various substances in cooling systems or fuel tanks.

In addition, IoT-based monitoring and diagnostic systems make it possible to predict the condition of equipment using machine learning algorithms and big data analysis. This helps optimize maintenance schedules, provide early warning of faults and reduce plant downtime.

Thus, monitoring and diagnostic systems based on IoT and data collection play an important role in modern thermal power plants, providing a prompt response to changes in the condition of equipment, increasing its reliability and safety, as well as optimizing maintenance and operation processes.

III. Examples of practical application

Diagnostics of turbines, generators and transformers.

Diagnostics of turbines, generators and transformers in thermal power plants is an important maintenance step and ensures the reliable operation of these key components. For turbines, such diagnostics include checking the condition of the blades, rotor, casing, seals and bearings to identify wear, damage or other problems that may affect their operation and efficiency. For generators, it is important to diagnose the stator and rotor windings, insulation, cooling system, bearings and other key components to identify potential problems such as short circuits, insulation defects or wear. And for transformers, diagnostics include checking the condition of the windings, insulation, cooling system, oil level and other parameters to identify problems such as short circuits, oil leaks or thermal anomalies.

Various methods and technologies can be used to diagnose these components, including visual inspection, temperature measurement, oil analysis, ultrasonic testing, vibration analysis, thermal imaging and others. For example, thermal imaging can be used to detect overheating in the internal components of turbines, generators and transformers, while ultrasonic testing can detect hidden defects in windings or bearings [6-8].

When checking electric motors, you need to pay maximum attention to the following elements:

- bearings - assess their defectiveness by temperature;
- ventilation ducts - check their permeability;
- windings - make sure that there are no turn short circuits.

An example of a thermogram of electric motors is shown in Figure 1.

Inspecting a generator with a thermal imager includes the following steps:

- checking the stator steel for defects;
- determining the temperature of the device and checking abnormal heating zones;
- determination of the temperature of the solder insulation surface;
- determination of brush heating temperature;
- determination of the thermal state of excitation system devices.

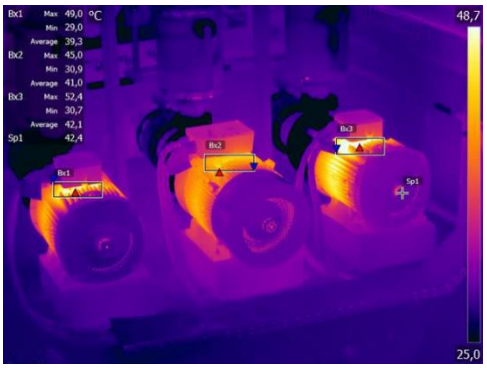


Figure 1: Example of thermogram of electric motors

The thermogram of the electric generator is shown in Figure 2.

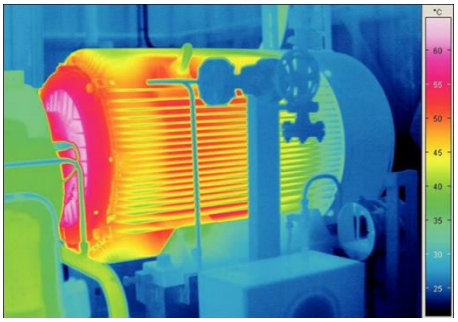


Figure 2: Electric generator thermogram

It is important to note that regular diagnostics and maintenance of these components helps prevent accidents, increase equipment life and ensure reliable operation of thermal power plants in general.

Detection and analysis of defects in electrical circuits and connections.

Detection and analysis of defects in electrical circuits and connections in thermal power plants is essential to ensure the safety and reliability of electrical equipment. Electrical circuits and connections can be subject to various types of defects, such as overheating, corrosion, insulation defects, breaks, short circuits and others. These defects can occur due to various reasons, including improper installation, wear and tear of materials, exposure to external factors, etc.

Various methods and technologies are used to detect and analyze defects in electrical circuits and connections. This includes visual inspection, resistance measurements, thermographic and thermal imaging studies, ultrasonic testing, insulation testing methods and others. For example, thermal imaging cameras can detect overheating in connections or circuit components, which may indicate improper contact or overload. Ultrasonic inspection can be used to identify defects such as cracks or corrosion that may not be visible by visual inspection [9, 10].

DJI's thermal imaging drones are helping users around the world improve productivity and safety (Figure 3). A thermal imaging camera consists of a special lens that transmits infrared frequencies. The camera is also equipped with a thermal sensor and an image processor, which are located in a protective housing. The camera is usually mounted on a drone gimbal, which rotates 360 degrees and helps stabilize the image. As the drone flies over objects, the camera's thermal sensor detects infrared wavelengths and converts them into electronic signals. After receiving the signals, the image processor creates what is called a thermogram or thermographic image, which consists of a color map showing various different temperature values.

The temperature sensor is also called a microbolometer. The structure of this sensor is very

complex, absorbs infrared energy and then creates a thermogram based on its measurements.

Analysis of data obtained as a result of diagnostics of electrical circuits and connections allows you to identify problem areas, assess their severity and take corrective action. Regular diagnostics and maintenance of electrical circuits and connections are important to prevent accidents, ensure personnel safety and reliable operation of thermal power plants [11-13]. Monitoring the insulation and thermal conditions of equipment is an important aspect of maintenance and safety in thermal power plants. Insulating electrical systems is critical to preventing short circuits and accidents, and thermal control of equipment helps prevent overheating and damage to components.

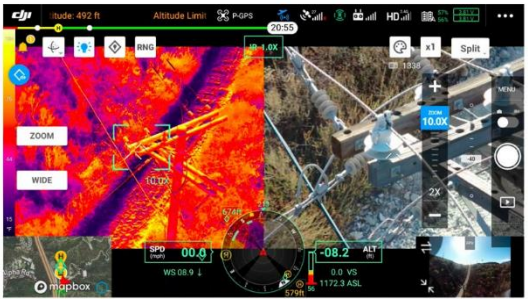


Figure 3: Drones with thermal imaging from DJMonitoring of insulation and thermal conditions of equipment

Insulation monitoring usually involves regularly measuring insulation resistance using specialized testers. This allows you to identify potential insulation defects, such as damage, moisture penetration or contamination, which can lead to current leaks and accidents. Carrying out such measurements allows you to quickly identify problems and take measures to eliminate them, for example, replacing damaged sections of insulation.

When it comes to equipment thermal monitoring, this involves continuously measuring and analyzing the temperatures of critical system components and assemblies. For this purpose, thermal imaging cameras, thermocouples, thermistors and other means of measuring temperature can be used. Continuous thermal monitoring allows you to identify potential problems, such as overheating, insufficient cooling, or thermal imbalance, and take appropriate action, such as improving the cooling system or changing the operating mode of equipment.

In general, monitoring the insulation and thermal conditions of equipment in thermal power plants is an important tool for preventing accidents, ensuring personnel safety and reliable operation of equipment.

Preventing emergency situations and reducing risks in thermal power plants plays a decisive role in ensuring personnel safety, equipment safety and continuity of production processes. For this, various approaches and technologies are used.

Firstly, monitoring and diagnostic systems, such as IoT systems, thermal imaging cameras, ultrasonic flaw detectors and others, make it possible to quickly identify potential problems and deviations in equipment operation before they lead to an emergency. Regular maintenance and diagnostics allow you to detect and correct potential problems at an early stage and prevent them from developing further.

Secondly, strict safety regulations and standards, as well as safety regulations, are the basis for preventing accidents. This includes providing regular instructions and training to staff on the safe handling of equipment, as well as developing and maintaining workplace safety procedures [14, 15].

Another important aspect of preventing accidents is the use of modern technologies and equipment that meet high safety standards and ensure reliable operation. Regular updating and modernization of equipment helps to minimize the risks of emergency situations and ensures a high level of safety in thermal power plants.

IV Conclusions

Increasing the service life of equipment and reducing its maintenance costs are important tasks to ensure efficient and reliable operation of thermal power plants. To achieve these goals, various strategies and methods are used.

Firstly, regular maintenance and preventive maintenance help keep equipment in good condition and prevent malfunctions from occurring. Carrying out regular inspections, replacing worn parts, lubrication and adjustment of mechanisms help prevent premature wear and increase the service life of equipment.

Secondly, the use of modern technologies and innovative solutions makes it possible to increase the efficiency of equipment and reduce the cost of its operation. For example, the use of IoT-based monitoring and diagnostic systems allows you to quickly identify and eliminate potential problems, which helps to increase the service life of equipment and reduce maintenance costs.

In addition, personnel training and compliance with operating and safety regulations help prevent equipment damage and reduce the likelihood of accidents, which also helps to increase its service life and reduce maintenance costs.

Thus, increasing the service life of equipment and reducing the cost of its maintenance is achieved through regular maintenance, the use of modern technologies, personnel training and compliance with operating rules. These measures help keep equipment in good condition, prevent premature wear and improve its efficiency, which ultimately reduces maintenance costs and extends its service life.

Optimizing production processes and improving energy system efficiency in thermal power plants play a key role in ensuring economic efficiency and competitiveness. To achieve these goals, various strategies and technologies are used.

Firstly, the introduction of modern control and automation systems makes it possible to optimize production processes, manage equipment operating modes and make the most efficient use of resources. Automation can reduce energy, raw material and labor costs, and improve the accuracy and reliability of production processes.

Secondly, the use of modern technologies and equipment, such as cogeneration units, solar panels, wind generators and other renewable energy sources, allows us to increase energy efficiency and reduce dependence on traditional energy sources. This reduces energy costs and reduces environmental impact. In addition, optimization of production processes and increased efficiency of power systems is achieved through the implementation of modern methods of management and production planning, as well as data analysis to identify potential for improving production processes and reducing costs.

Thus, optimization of production processes and increasing the efficiency of energy systems in thermal power plants is achieved through the introduction of modern technologies, automation of production processes, use of renewable energy sources and improvement of management and production planning methods. These measures reduce costs, increase the productivity and efficiency of plants, and improve their competitiveness in the energy market.

References

[1]. Lina Y, Lia C, Yanga Y, Qina JF, Sub X, Zhanga H, Zhangao W, “Automatic Display Temperature Range Adjustment for Electrical Equipment Infrared Thermal Images”, The 4th International Conference on Power and Energy Systems Engineering 2017, CPESE 2017, 454-459, Berlin, Germany, 25-29 September 2017.

[2]. Afonin AV, Newport RK, Polyakov VS, et al., “Fundamentals of Infrared Thermography”, Mining Science and Technology, p. 240, Moscow, Russia, 2004.

[3]. S.A. Bazhanov, “Infrared Diagnostics of Electrical Equipment of Switchgears”, Mining Science and Technology, p. 76, Moscow, Russia, 2000.

[4]. Y. Luo, Zh. Li, H. Wang, "A Review of Online Partial Discharge Measurement of Large Generators", MDPI Energies, Issue 10, Vol. 11, No. 1694, pp. 1-32. Basel, Switzerland, October 2017.

[5]. M. Sasic, C. Chan, "Using Measurement Results to Diagnose the Condition of High-Voltage Rotating Machines Part 1 and 2", Industry Topics, Iris Power, pp. 1-9, Canada, 2019

[6]. Mohamed ES. Development and analysis of a variable position thermostat for smart cooling system of a light duty diesel vehicles and engine emissions assessment during NEDC. Applied Thermal Engineering 2016; 99: 358–372. https://doi.org/10.1016/j.applthermaleng.2015.12.099

[7]. Rzayeva SV, Ganiyeva NA, Piriyeva NM. Modern methods of diag-nostics of electric power equipment. The 19th International Conference on" Technical and Physical Problems of Engineering 2023; T. 31. 105-110.

[8]. Rahimli I, Bakhtiyarov A, Abdullayeva G, Rzaeva S, Przeglad Elektrotechniczny 2024; 2, 132-134 DOI: 10.15199/48.2024.02.26

[9]. Rahimli NI, Rzayeva SV, Aliev HZ. Diagnostics and monitoring of power transformers. Universum: technical sciences 2022; (11-6 (104)): 32-35.

[10]. Alimamedova SJ. Some issues in diagnostics of electric motors. Vestnik nauki 2024; 4.1(70): 528-535. DOI 10.24412/2712-8849-2024-170-528-535

[11]. Mammadov N. Analysis of systems and methods of emergency braking of wind turbines. International Science Journal of Engineering & Agriculture 2023; 2(2): 147-152/ doi: 10.46299/j.isjea.20230202.14

[12]. Pirieva NM, Guseinov ZH. Fault Analysis in Power Transformers. Vestnik nauki 2023; T4, 7(64), 297-304. DOI 10.24412/2712-8849-2023-764-297-304

[13]. Mamedova GV, Pirieva NM, Shirinova MCh Diagnostics of power transformers. Internauka: electron. scientific magazine 2023; 6(276), 31-34 DOI:10.32743/26870142.2023.6.276.352720

[14]. Walentynowicz J, Krakowski R. Modeling of the higher pressure cooling system for transport vehicles engines. Transport Problems 2010; 5(4): 39-47. https://doi.org/10.20858/tp.2010.5.4.5.

[15]. Safiyev E.S., Rzayeva S.V., Karimova R.K. The importance of diagnostics of electrical equipment at thermal power plants for ensuring the reliability of power systems. Przeglad elektrotechniczny 2024; R.100 NR 9/2024

RT&A, No 4(80)
Volume 19, December, 2024

ENHANCING REDUNDANT SYSTEM PERFORMANCE: A STOCHASTIC MODEL FOR OPTIMIZED INSPECTION STRATEGIES POST-FAILURE

Purnima Sonker[1*], R.K. Bhardwaj[1]

•

[1]Department of Statistics, Punjabi University Patiala-147002, India
pani.sonkar@gmail.com, rkb_mstates@rediffmail.com

Abstract

This paper delves into the strategic utilization of inspections to determine the appropriate action for components within redundant systems following unit and switch failures. Post-failure, the timely execution of repair and replacement procedures is paramount for restoring system functionality. By assigning inspection tasks to servers, this paper aims to evaluate the condition of system components and make informed decisions regarding repair or replacement. It addresses the standardization of inspection processes and subsequent repair/replacement protocols for industrial systems encountering failures. Introducing a model, the study endeavors to bolster system reliability and availability by addressing failures caused by faults through inspection and subsequent repair/replacement actions. Employing a quantitative approach, it provides insights into maintaining system reliability and availability via a stochastic framework. By integrating unit and switch inspections into the analysis, the paper proposes a strategic approach to optimizing redundant system operations, facilitating effective decision-making concerning repair and replacement strategies post-failure.

Keywords: switch, server, inspection, replacement, and repair.

1. Introduction

Ensuring quality, dependability, and efficiency is paramount in today's industrial landscape; spanning sectors like manufacturing, aerospace, and automotive, where meeting precision standards and exceeding customer expectations are non-negotiable. Central to achieving these goals is the implementation of efficient inspection methods within industrial systems. Inspection procedures are indispensable for defect detection, ensuring product integrity, and upholding safety standards. The evaluation of the operational status is facilitated through inspections. Standby systems typically employ two approaches to inspection policies: routine checks or inspections triggered by failures. Periodic inspection policies aim to optimize standby system performance by conducting inspections at predetermined intervals [1]. Timing periodic inspections optimally can enhance standby system dependability while minimizing costs [2]. Alternatively, the expense and downtime associated with directly repairing a failed unit can be mitigated through post-failure inspections [3]. When a unit malfunctions, the options of repair or replacement are

considered based on its operational state [4]. Replacement involves substituting a defective unit with a new one, which can be reused after the switch [5]. The responsibility for all repair and replacement tasks lies with a single server, emphasizing the critical role of skilled personnel in maintaining system reliability. However, over time, server degradation may occur, leading to increased downtime and necessitating updates [6].

In the event of a main unit failure in a backup system, the switch plays a crucial role in activating the standby unit to ensure system availability and reliability. Switching may occur at predetermined intervals or when the operational unit fails [7]. Failures during switching can affect system functionality, necessitating reboots and repairs [8] and [9]. Switching reliability varies, with standby systems more prone to faulty switches [10]. Post-switch failure inspections dictate repair or replacement based on utility [11]. Insufficient switching can compromise warm standby system reliability, requiring intervention from repair personnel [12]. Server or switch failures during tasks can degrade system performance and significantly affect expenses [13] and [14]. Probabilistic models offer insights into system performance, availability, and failure detection [15]. Maintaining high availability hinges on ensuring the reliability and efficiency of switches and servers, with probabilistic models providing the most accurate assessment of standby system profitability. The Weibull distribution is particularly suitable for simulating random failures and evaluating standby system profitability [16]. Additionally, post-failure economic evaluations can be conducted for switches [17].

This paper introduces a probabilistic model of a standby system comprising two identical units, one serving as the primary operating unit and the other as the cold standby unit. In the event of the main unit failure, the switch activates the cold standby. Subsequently, the server initiates an inspection of the unit to determine whether repair or replacement is warranted, following similar procedures for switches after failure. Replacement is preferred only when repair costs are deemed prohibitive. The server oversees all inspection, repair, and replacement tasks, but can only handle one task at a time. Repairing switches takes time, whereas replacement is expedient. Numerical simulations in this study follow the Weibull distribution for accuracy.

2. Notations

O	The unit is in operative mode.
Cs	The unit is kept as cold standby.
St	The switching mechanism is good.
Se	The server is good.
Csw	The cold standby unit is under waiting.
p/q	The switch is under operation/failed.
F_{ur} / F_{UR}	The unit is under repair/under repair continuously from previous state.
F_{wr} / F_{WR}	The failed unit is waiting for repair/waiting for repair continuously from previous state.
St_{ur} / St_{UR}	The switch mechanism is under repair/under repair continuously from previous state.
St_{wr} / St_{WR}	The switch mechanism is failed and waiting for repair/under treatment continuously from previous state.
St_{wi} / St_{WI}	The switch mechanism is waiting for inspection /continuously waiting for inspection from previous state.
Se_{ut} /Se_{UT}	The server is failed and under treatment/under treatment continuously from previous state.

z(t)/Z(t)	pdf/cdf of failure rate of the unit.
r(t)/R(t)	pdf/cdf of failure rate of the server.
f(t)/F(t)	pdf/cdf of repair time of the failed unit.
h(t)/H(t)	pdf/cdf of repair time of the failed switch.
n(t)/N(t)	pdf/cdf of switch inspection time of the server.
m(t)/M(t)	pdf/cdf of unit inspection time of the server.
s(t)/S(t)	pdf/cdf of the treatment time of the server.
c/d	Probability of Switch repair /replacement feasibility after inspection.
a/b	Probability of Unit repair /replacement feasibility after inspection.
$M_i(t)$	Probability that the system is up initially in state Si ∈ E and remains up at time t without visiting any other regenerative state.
$W_i(t)$	Probability that the unit, switch, and server remain busy in state Si up to time 't' without transitioning to another regenerative state or returning to the same state via non-regenerative states.
○/ ▭	Representation of regenerative states and failed regenerative states in a diagram.
●	Regenerative points.

3. Development of Model

3.1 Assumptions

- One unit is initially powered on and the other is placed in cold standby.
- The unit is repaired directly after a failure, but the switch is checked to see if it can be repaired/replaced.
- Switch switching is instantaneous.
- If the main unit fails then it goes under the inspection process, to check the feasibility of its repair or replacement.
- If switch fails then it goes under the inspection process, to check the feasibility of its repair or replacement.
- The switch is prioritized for repair or replacement after inspection and failure.
- The server can fail while doing its job, but not in an idle state.
- After a failure, the server moves directly to the recovery phase.
- Replacement is instantaneous.
- Inspection/ repair priority is given to switch after failure.
- All repairs and treatments are perfect.
- Random variables are statistically independent.

3.2 States of the System

The following are possible transition states of the system model.

The regenerative states:

$$S_0=(O,Cs,St,Sv), S_1=(F_{ui},O), S_2=(F_{ur},O),\ S_3=(O,F_{wi},Sv_{ut}), S_4=(O,F_{wr},Sv_{ut}).$$

The failed regenerative states:

$$S_5=(F_{wi},Cs_w,St_{ui}),$$

The non-regenerative states:

$S_6=(F_{WI},Cs_w,St_{wi},Sv_{ut}),S_7=(F_{UI},F_{wi}),S_8=(F_{wi},F_{WI},Sv_{ut}),$

$S_9=(F_{wi},F_{WR},Sv_{UT}),S_{10}=(F_{UR},F_{wi}),S_{11}=(F_{wi},F_{WI},Sv_{UT}),$

$S_{12}=(F_{WI},Cs_w,St_{ui}),S_{13}=(F_{WI},Cs_w,St_{ur}),S_{14}=(F_{UI},F_{wi}),$

$S_{15}=(F_{WI},Cs_w,St_{wr},Sv_{ut}),S_{16}=(F_{ur},F_{WI}),S_{17}=(F_{wr},F_{WI},Sv_{ut}).$

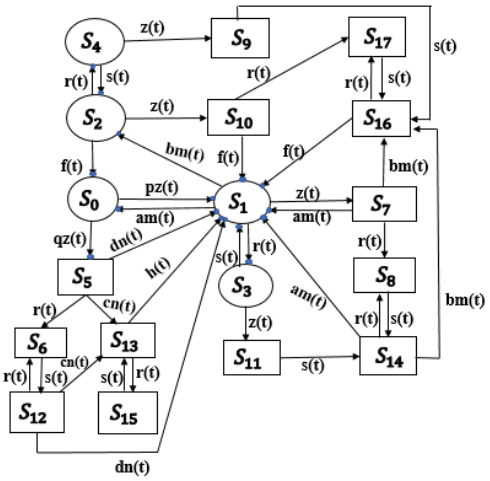


Figure 1: *State transition diagram of model*

3.3 Transition Probabilities and Mean Sojourn Times

Simple probabilistic considerations yield the following expressions for the non-zero element

$$p_{ij} = Q_{ij}(\infty) = \int_0^\infty q_{ij}(t)dt \tag{1}$$

The Mean Sojourn time μ_i in state S_i are given by:

$$\mu_i = E(t) = \int_0^\infty P(T > t)dt \tag{2}$$

We get

$$p_{0,1}=\int_0^\infty pz(t)dt,\ p_{0,5}=\int_0^\infty qz(t)dt,\ p_{1,0}=\int_0^\infty am(t)\bar{Z}(t)\bar{R}(t)dt,$$

$$p_{1,2}=\int_0^\infty bm(t)\bar{Z}(t)\bar{R}(t)dt,\ p_{1,3}=\int_0^\infty r(t)\bar{M}(t)\bar{Z}(t)dt,\ p_{1,4}=\int_0^\infty z(t)\bar{R}(t)\bar{M}(t)dt,$$

$$p_{2,0}=\int_0^\infty f(t)\bar{Z}(t)\bar{R}(t)dt,\ p_{2,4}=\int_0^\infty r(t)\bar{F}(t)\bar{Z}(t)dt,\ p_{2,10}=\int_0^\infty z(t)\bar{R}(t)\bar{F}(t)dt,$$

$$p_{3,1}=\int_0^\infty s(t)\bar{Z}(t)dt,\ p_{3,11}=\int_0^\infty z(t)\bar{S}(t)dt,\ p_{4,2}=\int_0^\infty s(t)\bar{Z}(t)dt,$$

$$p_{4,9}=\int_0^\infty z(t)\bar{S}(t)dt,\ p_{5,1}=\int_0^\infty dn(t)\bar{R}(t)dt,\ p_{5,6}=\int_0^\infty r(t)\bar{N}(t)dt,$$

$$p_{5,13}=\int_0^\infty cn(t)\bar{R}(t)dt,\ p_{6,12}=\int_0^\infty s(t)dt,\ p_{7,1}=\int_0^\infty am(t)\bar{R}(t)dt,$$

$$p_{7,8}=\int_0^\infty r(t)\bar{M}(t)dt,\ p_{7,16}=\int_0^\infty bm(t)\bar{R}(t)dt,\ p_{8,14}=\int_0^\infty s(t)dt,$$

$$p_{9,16}=\int_0^\infty s(t)dt,\ p_{10,1}=\int_0^\infty f(t)\bar{R}(t)dt,\ p_{10,17}=\int_0^\infty r(t)\bar{F}(t)dt,$$

$$p_{11,14}=\int_0^\infty s(t)dt,\ p_{12,1}=\int_0^\infty dn(t)\bar{R}(t)dt,\ p_{12,6}=\int_0^\infty r(t)\bar{N}(t)dt,$$

$$p_{12,13}=\int_0^\infty cn(t)\bar{R}(t)dt,\ p_{13,1}=\int_0^\infty h(t)\bar{R}(t)dt,\ p_{13,15}=\int_0^\infty r(t)\bar{H}(t)dt,$$

$$p_{14,1}=\int_0^\infty am(t)\bar{R}(t)dt,\ p_{14,8}=\int_0^\infty r(t)\bar{M}(t)dt,\ p_{14,16}=\int_0^\infty bm(t)\bar{R}(t)dt,$$

$$p_{15,13}=\int_0^\infty s(t)dt,\ p_{16,1}=\int_0^\infty f(t)\bar{R}(t)dt,\ p_{16,17}=\int_0^\infty r(t)\bar{F}(t)dt,$$

$$p_{17,16}=\int_0^\infty s(t)dt,$$

The expressions for mean sojourn times are as follows:

$$\mu_0=\int_0^\infty \bar{Z}(t)dt,\ \mu_1=\int_0^\infty \bar{Z}(t)\bar{R}(t)\bar{M}(t)dt,\ \mu_2=\int_0^\infty \bar{Z}(t)\bar{F}(t)dt,\ \mu_3=\int_0^\infty \bar{Z}(t)\bar{S}(t)dt,\ \mu_4=\int_0^\infty \bar{Z}(t)\bar{S}(t)dt.$$

4. System's Performance Measures

4.1 Mean Time to System Failure (MTSF)

Let $\phi_i(t)$ be the c.d.f of the first passage time from regenerative state S_i to a failed state. Regarding the failed state as absorbing state, we have the following recursive relations for $\phi_i(t)$:

$$\phi_i(t)=\sum_j Q_{i,j}(t)[s]\phi_j(t)+\sum_k Q_{i,k}(t) \qquad i=0,1,3,4,5 \tag{3}$$

Taking LST of Eq. (3) and solving for $\tilde{\phi}_0(s)$, we have

$$R^*(s)=\frac{1-\tilde{\phi}_0(s)}{s} \tag{4}$$

The reliability R(t) can be obtained by taking inverse Laplace transition of Eq.(4) and MTSF is given by

$$MTSF(t)=\lim_{s\to 0}R^*(s)=\lim_{s\to 0}\frac{1-\tilde{\phi}_0(s)}{s}=\frac{N_1}{D_1} \tag{5}$$

Where

$$N_1=\mu_0[1-p_{1,3}p_{3,1}][1-p_{2,4}p_{4,2}]+p_{0,1}\left[\mu_1+p_{1,3}\mu_3\right]+p_{0,1}p_{1,2}\left[\mu_2+p_{2,4}\mu_4\right]$$

$$D_1=[1-p_{1,3}p_{3,1}-p_{1,0}p_{0,1}][1-p_{2,4}p_{4,2}]-p_{2,0}p_{1,2}p_{0,1}.$$

4.2 Steady State Availability

$M_i(t)$ is the probability that the system is up initially in state $S_i \in E$ is up at time t without visiting to any other regenerative state, we have

$$M_0=\int_0^\infty \bar{Z}(t)dt,\ M_1=\int_0^\infty \bar{Z}(t)\bar{R}(t)\bar{M}(t)dt,\ M_2=\int_0^\infty \bar{Z}(t)\bar{F}(t)dt,\ M_3=\int_0^\infty \bar{Z}(t)\bar{S}(t)dt,\ M_4=\int_0^\infty \bar{Z}(t)\bar{S}(t)dt,$$

Let $A_i(t)$ be the probability that the system is in up-state at an instant 't' given that the system entered regenerative state S_i at t=0. The recursive relations for $A_i(t)$ are as follows:

$$A_i(t)=M_i(t)+\sum_j q_{i,j}^{(n)}(t)[c]A_j(t) \qquad i=0,1,2,3,4,5 \tag{6}$$

Where S_j is any successive regenerative state to which the regenerative state S_i can transit through n transitions. Taking LT of Eq. (6) and solving, the steady state availability is given by

$$A_0=\lim_{s\to 0} sA_0^*(s)=\frac{N_2}{D_2} \tag{7}$$

Where

$$N_2=(p_{1,0}\mu_0+\mu_1+p_{1,3}\mu_3)(1-p_{2,4}p_{4,2})+p_{1,2}(p_{2,0}\mu_0+\mu_2+p_{2,4}\mu_4)$$

$$D_2=(p_{1,0}\mu_0+\mu_1'+p_{1,3}\mu_3'+p_{0,5}\mu_5')(1-p_{2,4}p_{4,2})+p_{1,2}(p_{2,0}\mu_0+\mu_2'+p_{2,4}\mu_4').$$

4.3 Busy Period Analysis for Server

4.3.1 Due to Inspection

Let $B_i^I(t)$ be the probability that the server is busy in inspection at an instant t given that the system entered regenerative state S_i at t = 0. The recursive relations for $B_i^I(t)$ are as follows:

$$B_i^I(t)=W_i^I(t)+\sum_j q_{i,j}^{(n)}(t)[c]B_j^I(t) \qquad i=0,1,2,3,4,5 \tag{8}$$

Where S_j is any successive regenerative state to which the regenerative state S_i can transit through n transitions. $W_i^I(t)$ be the probability that the server is busy in state S_i due to repair of the unit up to time 't' without making any transition to any other regenerative state or returning to the same via one or more non-regenerative state

$$W_1^I(t)=\bar{Z}(t)\bar{M}(t)\bar{R}(t)+(z(t)\bar{R}(t)[c]1)\bar{M}(t)+(z(t)\bar{M}(t)\bar{R}(t)[c]r(t)\bar{M}(t)[c]1)\bar{S}(t)+(z(t)\bar{M}(t)\bar{R}(t)[c]r(t)\bar{M}(t)[c]s(t)[c]1)M(t)$$

$$W_5^I(t)=\bar{N}(t)\bar{R}(t)+(r(t)\bar{N}(t)[c]1)\bar{S}(t)+(r(t)\bar{N}(t)[c]s(t)[c]1)\bar{N}(t)+(cn(t)\bar{R}(t)[c]1)\bar{H}(t)\bar{R}(t)+(cn(t)\bar{R}(t)[c]r(t)\bar{H}(t)[c]1)\bar{S}(t)+(cn(t)\bar{R}(t)[c]r(t)\bar{H}(t)[c]s(t)[c]1)\bar{H}(t)$$

Using LT, of Eq. (8) and solving for $B_0^{I*}(s)$, we have

$$B_0^I=\lim_{s\to 0} sB_0^{I*}(s)=\frac{N_3^I}{D_3^I} \tag{9}$$

Where

$$N_{13}^I=(1-p_{2,4}p_{4,2})W_1^I(0)+p_{0,5}(p_{1,0}(1-p_{2,4}p_{4,2})+p_{1,2}p_{2,0})W_5^I(0)$$

$$D_3^I=(1-p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1'+\mu_3'p_{1,3})+p_{1,0}(\mu_0+p_{0,2}\mu_2'+p_{0,4}\mu_4'+\mu_5'p_{4,5}).$$

4.3.2 Due to Repair

Let $B_i^R(t)$ be the probability that the server is busy in repair of the switch or unit due to inspection at an instant 't' given that the system entered state S_i at time t=0. The recursive relations for $B_i^R(t)$ are as follows:

$$B_i^R(t)=W_i^R(t)+\sum_j q_{i,j}^{(n)}(t)[c]B_j^R(t) \qquad i=0,1,2,3,4,5 \qquad (10)$$

Where S_j is any successive regenerative state to which the regenerative state S_i can transit through n transitions. $W_i^R(t)$ be the probability that the server is busy in state S_i due to preventive maintenance of server up to time 't' without making any transition to any other regenerative state or returning to the same via one or more non-regenerative state

$W_2^R(t)=\bar{Z}(t)\bar{F}(t)\bar{R}(t)+(z(t)\bar{R}(t)[c]1)\bar{F}(t)+(z(t)\bar{F}(t)\bar{R}(t)[c]r(t)\bar{F}(t)[c]1)\bar{S}(t)+(z(t)\bar{F}(t)\bar{R}(t)[c]r(t)\bar{F}(t)[c]s(t)[c]1)\bar{F}(t)$

Using LT, of Eq. (10) and solving for $B_0^{R*}(s)$, we have

$$B_0^R=\lim_{s\to 0} sB_0^{R*}(s)=\frac{N_3^R}{D_3^R} \qquad (11)$$

Where

$N_3^R = p_{12}W_2^{R*}(0)$,

$D_3^R = (1-p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1'+\mu_3'p_{1,3})+p_{1,0}(\mu_0+p_{0,2}\mu_2'+p_{0,4}\mu_4'+\mu_5'p_{4,5})$.

4.4 Expected Number of Treatment of the Server

Let $T_i(t)$ be the expected number of treatment of the failed server in (0, t] given that the system entered the regenerative state S_i at t=0. The recursive relations for $T_i(t)$ are given as:

$$T_i(t)=\sum_j Q_{i,j}(t)[s][\varphi_j+T_j(t)] \qquad i=0,1,2,3,4,5 \qquad (12)$$

where

$$\varphi_j=\begin{cases}1 & \text{if the server performs the task in state } S_j.\\ 0 & \text{otherwise}\end{cases}$$

Using LT, of Eq. (12) and solving for $\tilde{T}_0(s)$, we get

$$T_0=\lim_{s\to 0} s\tilde{T}_0(s)=\frac{N_5}{D_5} \qquad (13)$$

Where

$N_4 = p_{1,3}p_{3,1}(1-p_{2,4}p_{4,2})+p_{1,2}p_{2,4}p_{4,2}$

$D_4 = (1-p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1'+\mu_3'p_{1,3})+p_{1,0}(\mu_0+p_{0,2}\mu_2'+p_{0,4}\mu_4'+\mu_5'p_{4,5})$.

4.5 Expected Number of Inspections

Let $I_i^m(t)$ be the expected number of inspections of the unit and switch in (0, t] given that the system entered regenerative state S_i at time t=0. The recursive relations for $I_i^m(t)$ are as follows:

$$I_i^m(t)=\sum_j Q_{i,j}(t)[s][\varphi_j+I_j^m(t)] \qquad i=0,1,2,3,4,5 \qquad (14)$$

where

$$\varphi_j=\begin{cases}1 & \text{if the server performs the task in state } S_j.\\ 0 & \text{otherwise}\end{cases}$$

Using LT, of Eq. (14) and solving for $\tilde{I}_0^m(s)$, we get

$$I_0^m=\lim_{s\to 0} s\tilde{I}_o^m(s)=\frac{N_k}{D_k} \qquad (15)$$

Expected number of unit inspections $I_i^u(t)$ when m=u and k=5 in Eq. (15), we get

$N_5 = p_{1,2}p_{0,5}p_{2,0}+(1+p_{1,0}p_{0,5})(1-p_{2,4}p_{4,2})$,

$D_5 = (1-p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1'+\mu_3'p_{1,3})+p_{1,0}(\mu_0+p_{0,2}\mu_2'+p_{0,4}\mu_4'+\mu_5'p_{4,5})$.

Expected number of switch inspections $I_i^s(t)$ when m=s and k=6 in Eq. (15), we get

$N_6 = p_{1,2}p_{0,5}p_{2,0} + p_{1,0}p_{0,5}(1-p_{2,4}p_{4,2})$,

$D_6 = (1- p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1' + \mu_3' p_{1,3}) + p_{1,0}(\mu_0 + p_{0,2}\mu_2' + p_{0,4}\mu_4' + \mu_5' p_{4,5})$.

4.6 Expected Number of Replacements

Let $P_i^m(t)$ be the expected number of replacements of unit and switch in (0, t] given that the system entered the regenerative state S_i at t=0. The recursive relations for $P_i^m(t)$ are as follows:

$$P_i^m(t) = \sum_j Q_{i,j}(t)[s][\varphi_j + P_i^m(t)] \qquad i = 0,1,2,3,4,5 \qquad (16)$$

Where

$$\varphi_j = \begin{cases} 1 & \text{if the server performs the task in state } S_j. \\ 0 & \text{otherwise} \end{cases}$$

Using LT, of Eq. (16) and solving for $\tilde{P}_0^m(s)$,we get

$$P_0^m = \lim_{s\to 0} s\tilde{P}_0^m(s) = \frac{N_k}{D_k} \qquad (17)$$

Expected number of unit replacements $\tilde{P}_0^u(s)$ when m=u and k=7 in Eq. (17), we get

$N_7 = (1-p_{2,4}p_{4,2})(p_{1,0} + p_{1,1.7} + p_{1,1.7,(8,14)^n} + p_{1,3}(p_{3,1.11,14} + p_{3,1.11,(14,8)^n}))$,

$D_7 = (1- p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1' + \mu_3' p_{1,3}) + p_{1,0}(\mu_0 + p_{0,2}\mu_2' + p_{0,4}\mu_4' + \mu_5' p_{4,5})$

Expected number of switch replacements $\tilde{P}_0^s(s)$ when m=s and k=8 in Eq. (17), we get

$N_8 = p_{0,5}\{p_{1,0}(1-p_{2,4}p_{4,2}) + p_{2,0}p_{1,2}\}(p_{5,1} + p_{5,1.(6,12)^n})$,

$D_8 = (1- p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1' + \mu_3' p_{1,3}) + p_{1,0}(\mu_0 + p_{0,2}\mu_2' + p_{0,4}\mu_4' + \mu_5' p_{4,5})$

4.7 Expected Number of Repairs

Let $O_i^m(t)$ be the expected number of repairs of the unit or switch by the server in (0, t] given that the system entered regenerative state S_i at time t=0. The recursive relations for $O_i^m(t)$ are as follows:

$$O_i^m(t) = \sum_j Q_{i,j}(t)[s][\varphi_j + O_j^m(t)] \qquad i = 0,1,2,3,4,5 \qquad (18)$$

Where

$$\varphi_j = \begin{cases} 1 & \text{if the server performs the task in state } S_j. \\ 0 & \text{otherwise} \end{cases}$$

Using LT, of Eq. (18) and solving for $\tilde{O}_0^m(s)$,we get

$$O_0^m = \lim_{s\to 0} s\tilde{O}_0^m(s) = \frac{N_k}{D_k} \qquad (19)$$

Expected number of unit repairs $\tilde{O}_0^u(s)$ when m=u and k=9 in Eq. (19), we get

$N_9 = (1- p_{1,0} - p_{1,1.7} - p_{1,1.7,(8,14)^n} - p_{1,3}(p_{3,1} + p_{3,1.11,14} + p_{3,1.11,(14,8)^n}))(1- p_{2,4}p_{4,2})$,

$D_9 = (1- p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1' + \mu_3' p_{1,3}) + p_{1,0}(\mu_0 + p_{0,2}\mu_2' + p_{0,4}\mu_4' + \mu_5' p_{4,5})$

Expected number of switch repairs $\tilde{O}_0^m(s)$ when m=s and k=10 in Eq. (19), we get

$N_{10} = p_{0,5}\{p_{1,0}(1-p_{2,4}p_{4,2}) + p_{2,0}p_{1,2}\}(1- p_{5,1} - p_{5,1.(6,12)^n})$,

$D_{10} = (1- p_{0,4}(p_{4,0}+p_{4,5}p_{5,0}))(\mu_1' + \mu_3' p_{1,3}) + p_{1,0}(\mu_0 + p_{0,2}\mu_2' + p_{0,4}\mu_4' + \mu_5' p_{4,5})$

5. The Profit

The Profit incurred to the system model in steady state is obtained after the deduction of all expenses from the total revenue generated from all sources.

$$P(t) = C_0 A_0(t) - \sum_{j=1}^{9} C_j L_j(t) \tag{20}$$

Where

j :	1	2	3	4	5	6	7	8	9
$L_j(t)$:	B_0^I	B_0^R	T_0	I_0^u	I_0^s	P_0^u	P_0^s	O_0^u	O_0^s

Where

C_0 = Revenue per unit up-time of the system

C_1 = Cost per unit time for which server is in inspection

C_2 = Cost per unit time for which server is busy in repair of failed unit or switch

C_3 = Cost per treatment of the server

C_4 = Cost per unit inspection of unit by the server

C_5 = Cost per unit inspection of unit by the server

C_6 = Cost per unit replacement of the unit

C_7 = Cost per unit replacement of the switch

C_8 = Cost per unit repair of the unit

C_9 = Cost per unit repair of the switch

And $A_0, B_0^I, B_0^R, T_0, I_0^u, I_0^s, P_0^u, P_0^s, O_0^u$, and O_0^s, are already defined.

6. Particular Case

The particular cases are calculated for Weibull density function with common shape parameter and different scale parameters. The probability density function for various random variables included in the model for Weibull case are as follows:

$z(t) = \lambda\eta t^{\eta-1}\exp(-\lambda t^{\eta})$, $f(t) = \alpha\eta t^{\eta-1}\exp(-\alpha t^{\eta})$, $r(t) = \mu\eta t^{\eta-1}\exp(-\mu t^{\eta})$, $h(t) = \gamma\eta t^{\eta-1}\exp(-\gamma t^{\eta})$,

$s(t) = \beta\eta t^{\eta-1}\exp(-\beta t^{\eta})$, $m(t) = \theta\eta t^{\eta-1}\exp(-\theta t^{\eta})$, $n(t) = \xi\eta t^{\eta-1}\exp(-\xi t^{\eta})$,

Where $t \geq 0$ and $\eta, \alpha, \lambda, \beta, \gamma, \mu, \theta, \xi > 0$

Changes made in various parameters of failure rate and repair rate -

Unit's repair rate α vary from 0.4 to 0.7.

Server's failure rate μ vary from o.3 to 0.002.

Server's repair rate β from 0.6 to 0.9.

Switch's repair rate γ from 0.5 to 0.7.

Switch's operating/failure probability p/q from 0.8/0.2 to 0.7/0.3.

Unit's inspection rate θ from 0.6 to 0.8.

Unit's repair/replacement probability after inspection a/b from 0.7/0.3 to 0.9/0.1.

Switch's inspection rate ξ from 0.7 to 0.9.

Switch's repair/replacement probability after inspection c/d from 0.6/0.4 to 0.8/0.2.

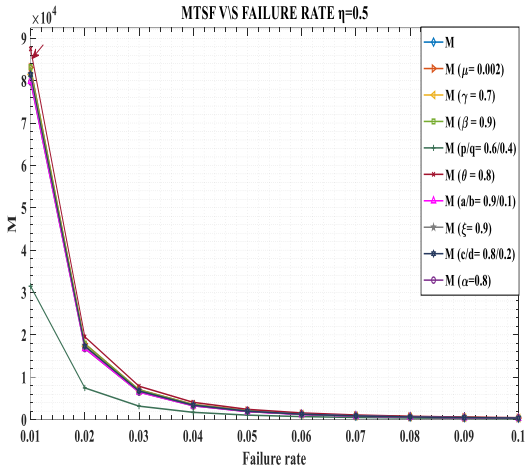


Figure 2(a): *MTSF w.r.t. failure rate when η=0.5*

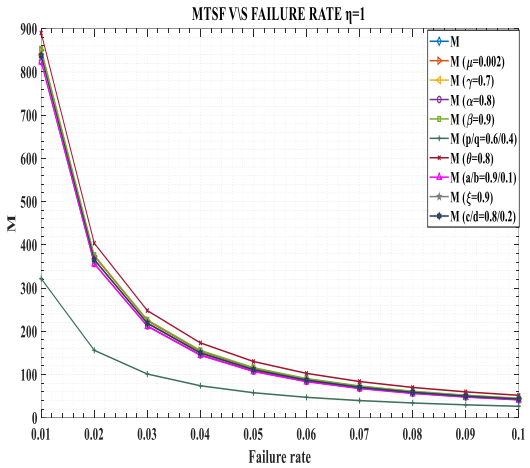


Figure 2(b): *MTSF w.r.t. failure rate when η=1*

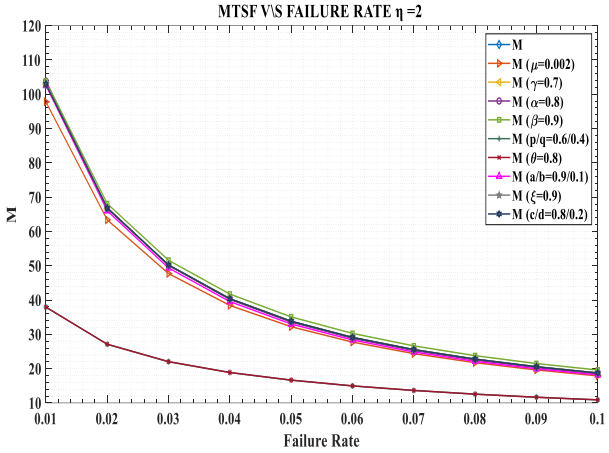


Figure 2(c): *MTSF w.r.t. failure rate when η=2*

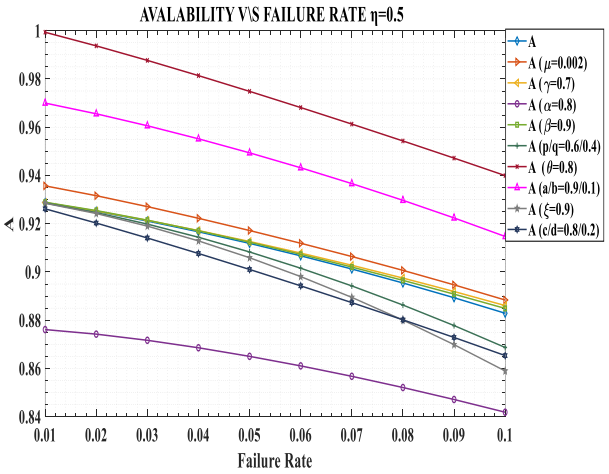


Figure 3(a): *Availability w.r.t. failure rate when η=0.5*

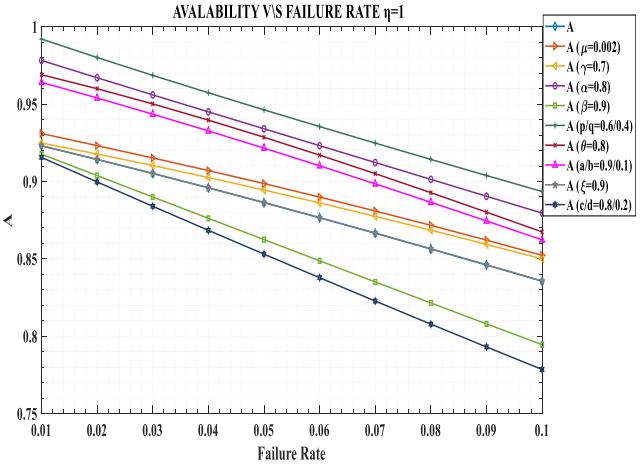


Figure 3(b): *Availability w.r.t. failure rate when η=1*

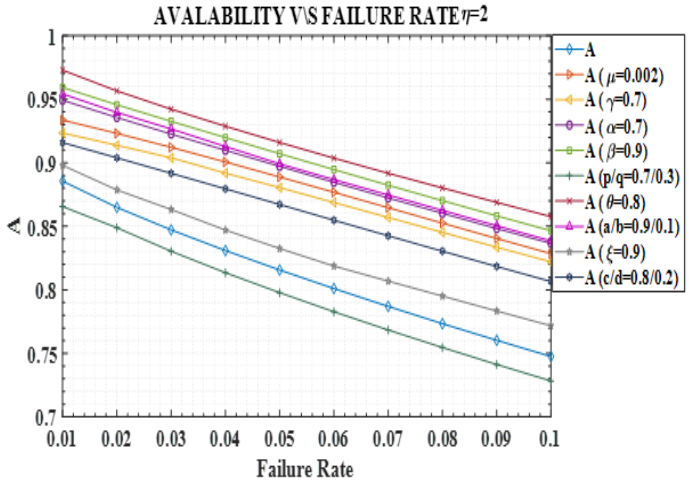


Figure 3(c): *Availability w.r.t. failure rate when η=2*

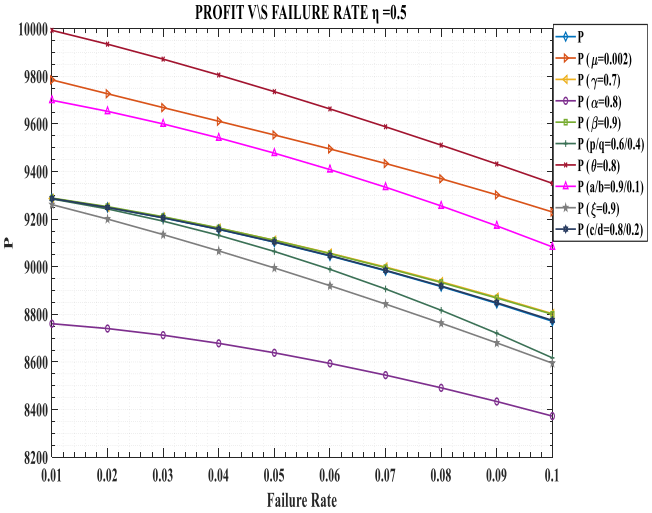


Figure 4(a): *Profit w.r.t. failure rate when η=0.5*

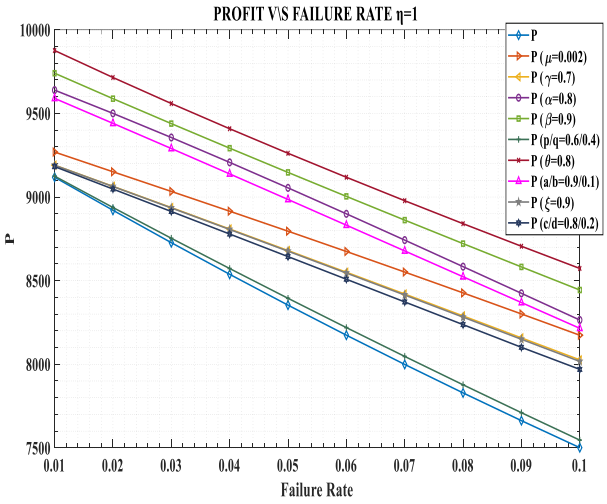


Figure 4(b): *Profit w.r.t. failure rate when η=1*

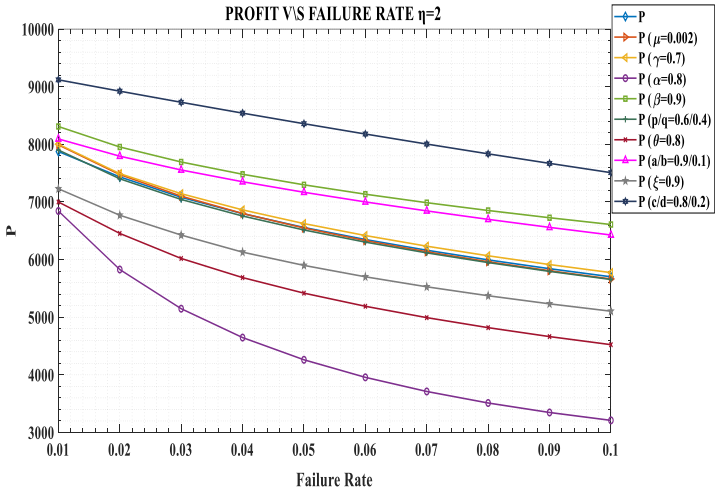


Figure 4(c): *Profit w.r.t. failure rate when η=2*

Mean time to failure (MTSF) is a crucial part in assessing system performance, especially in fields such as engineering and technology. In this context, Figures 2(a), 2(b), and 2(c) depict the behavior of MTSF for different values of failure rate and size parameters. A notable observation is that as the failure rate (λ) increases, the MTSF decreases. This trend is important because it implies that higher failure rates correspond to lower mean time to failure, which can affect the overall efficiency of the system. The failure time distribution is another important aspect affected by parameters such as magnitude (η). Figures 3(a), 3(b), and 3(c) depict similar behavior in terms of availability and failure rate. When the magnitude parameter is less than one, the output decays rapidly, indicating a decrease in performance and reliability over time. In contrast, when the magnitude parameter is greater than one, it suggests an increasing availability trend with higher inspection rates. Figures 4(a), 4(b), and 4(c) show the gain, which exhibits a decreasing trend as the shape parameter (η) increases. This implies that the gain reduces as the inspection rate increases, suggesting a saturation point beyond which further inspection does not significantly improve system performance. However, it is important to note that even though the gain reduces, inspection still has a positive impact on the overall system performance due to its high efficiency and availability.

7. Conclusion

Our research leads us to the conclusion that the evaluation of a malfunctioning device and the subsequent determination of whether to repair or replace it play a crucial role in influencing system performance and availability. While repair may incur greater costs and time commitments, our study highlights that replacing the defective component post-failure stands out as the most effective solution, particularly in terms of enhancing system availability and reliability within server environments. These principles hold significant relevance in critical and hard-to-reach systems, where the failure of a single module can trigger substantial operational disruptions. Illustrative examples include DSLAM networks, wind power plants, hydra production facilities, and automatic plastering machines, underscoring the widespread application of these concepts across diverse industrial settings.

Acknowledgement

The authors would like to acknowledge the anonymous authors for their valuable feedback, which greatly enhanced a previous version of this paper.

References

[1] Lu, X., Wang, W., Yang, H., Zuo, M. J., & Zhou, D. (2012). Optimizing the Periodic Inspection Interval for a 1-out-of-2 Cold Standby System Using the Delay-Time Concept. *Quality and Reliability Engineering International, 28*(6):648-662.

[2] Zhang, J., Huang, X., Fang, Y., Zhou, J., Zhang, H., & Li, J. (2016). Optimal inspection-based preventive maintenance policy for three-state mechanical components under competing failure modes. *Reliability Engineering & System Safety, 152*:95-103.

[3] Wang, W., Wu, Z., Xiong, J., & Xu, Y. (2018). Redundancy optimization of cold-standby systems under periodic inspection and maintenance. *Reliability Engineering & System Safety, 180*:394-402.

[4] Wang, W. (2009). An inspection model for a process with two types of inspections and repairs. *Reliability Engineering & System Safety, 94*(2):526-533.

[5] Levitin, G., Finkelstein, M., & Dai, Y. (2020). Optimal preventive replacement policy for homogeneous cold standby systems with reusable elements. *Reliability Engineering & System Safety, 204*:107135.

[6] Kumar, A., Garg, R., & Barak, M. S. (2023). Reliability measures of a cold standby system subject to refreshment. *International Journal of System Assurance Engineering and Management, 14*(1):147-155.

[7] Jia, X., Chen, H., Cheng, Z., & Guo, B. (2016). A comparison between two switching policies for two-unit standby system. *Reliability Engineering & System Safety, 148*:109-118.

[8] Ke, J. C., Liu, T. H., & Yang, D. Y. (2016). Machine repairing systems with standby switching failure. *Computers & Industrial Engineering, 99:*223-228.

[9] Hsu, Y. L., Ke, J. C., Liu, T. H., & Wu, C. H. (2014). Modeling of multi-server repair problem with switching failure and reboot delay and related profit analysis. *Computers & Industrial Engineering, 69*:21-28.

[10] Levitin, G., Xing, L., & Dai, Y. (2014). Cold vs. hot standby mission operation cost minimization for 1-out-of-N systems. *European Journal of Operational Research, 234*(1):155-162.

[11] Sonker, P., & Bhardwaj, R. K. (2023). Performance indices of a cold-standby system with an unreliable switch and server using Weibull failure model. *International Journal of Statistics and Reliability Engineering, 10*(2):488–495.

[12] Sadeghi, M., & Roghanian, E. (2017). Reliability analysis of a warm standby repairable system with two cases of imperfect switching mechanism. *Scientia Iranica, 24*(2):808-822.

[13] Kuo, C. C., & Ke, J. C. (2016). Comparative analysis of standby systems with unreliable server and switching failure. *Reliability Engineering & System Safety, 145*:74-82.

[14] Ke, J. C., Liu, T. H., & Yang, D. Y. (2018). Modeling of machine interference problem with unreliable repairman and standbys imperfect switchover. *Reliability Engineering & System Safety, 174*:12-18.

[15] Bhardwaj, R. K., Sonker, P., & Singh, R. (2022). A semi-Markov model of a system working under uncertainty. In *System Assurances,* 175-187.

[16] Gupta, R., Kumar, P., & Gupta, A. (2013). Cost benefit analysis of a two dissimilar unit cold standby system with Weibull failure and repair laws. *International Journal of System Assurance Engineering and Management,* 4:327-334.

[17] Kumar, A., Saini, M., & Devi, K. (2018). Stochastic modeling of non-identical redundant systems with priority, preventive maintenance, and Weibull failure and repair distributions. *Life Cycle Reliability and Safety Engineering*, 7:61-70.

PROFIT ANALYSIS OF REPAIRABLE WARM STANDBY SYSTEM UNDER IMPERFECT SWITCH

[1]Nishant Yadav, [2*]Shiv Kant, [3]Shashi Kant, [4]Arunita Chaukiyal, [5]Bindu Jamwal

[1]Department of Mathematics, Yaduvanshi Degree College, Narnaul, Haryana
[2*]Department of Computer Science and Engineering (AI & DS), GNIOT, Greater Noida
[3]Department of Computer Science and Engineering, Sharda University, Greater Noida
[4]Department of Computer Science, Acharya Narendra Dev College, University of Delhi
[5]School of Engineering and Technology, Raffles University, Neemrana
nishantchitosiya@gmail.com, kaushikshivkant@gmail.com, shashi.kant1@sharda.ac.in, arunita@yahoo.com, bindujamwal1810@gmail.com

*Corresponding Author

Abstract

In this paper, the performance of two non identical units repairable system are analyzed by using regenerative point graphical technique. Generally, the system has one operative unit and one warm standby unit. Fuzzy concept is used to find the reliability measures under imperfect switch. Regenerative point graphical technique and semi markov process are used to evaluate the reliability measures. Primary, secondary and tertiary circuits are used to describe the base state. The system is repaired by the available technician when any unit is failed or switch is under imperfect mode. The priority in repair is given to switch before working units. In this paper, the failure time and repair time follow general distributions. The tables are used to explore the reliability measures such that mean time to system failure, availability and profit values.

Keywords: Base state, warm standby, imperfect switch, availability and profit values.

I. Introduction

To meet the increasing demand for their products, manufacturers need to constantly improve their products which can be achieved by improving their production procedures. The MTSF, availability and profitability of two non identical units system with priority in switch repair are discussed by using the regenerating point graphical technique under specific circumstances. In the real time system, it is impossible to ignore the concept of spare unit switching failure when it is utilized to replace a failed operative unit.

Balagurusamy [1] described the terms related to the system meantime, failure, repair, redundancy, maintainability and availability. Gopalan and Bhanu [3] examined the behaviour of two unit repairable system under online preventive maintenance using regenerative point technique. Gupta and Singh [4] threw light on the profit and availability values of redundant system with imperfect switch. Lim [7]. Kumar and Sirohi [6] examined the reliability, availability and profit values of two unit system under cold standby approach with delays in repair of partially failed unit. Kumar and Goel [5] evaluated the behaviour of two unit cold standby system under general distribution. Taneja et al. [10] discussed on the comparative study of profit values two reliability models under varying demands. Bhardwaj et al. [2] analyzed the reliability measures of two unit cold standby system under standby failure and arbitrary distribution for

repair and replacement times. Sadeghi and Roghanian [8] discussed on the availability and profit values of the two dissimilar unit repairable system under two cases of imperfect switch. Singh et al. [9] evaluated the behaviour of the complex repairable system with two subsystems connected in series with switch facility.

II. System Assumptions

To describe the system, there are following assumptions

- The system has two distinct units where unit *A* is in operative mode and unit *B* is in warm standby mode.
- There is an imperfect switching.
- The sequence of repair is switch, operative main unit and warm standby unit.
- A technician is available to repair the failed unit.
- Failure time and repair time follow the general distribution.
- The repaired unit functions just like a brand-new one.

III. System Notations

To explain the juice plant, there are following notations

$i \xrightarrow{Sr} j$	r^{th} directed simple path from state '*i*' to state '*j*' where '*r*' takes the positive integral values for different directions from state '*i*' to state '*j*'.
$\xi \xrightarrow{sff} i$	A directed simple failure free path from state ξ to state '*i*'.
$m-cycle$	A circuit (may be formed through regenerative or non regenerative / failed state) whose terminals are at the regenerative state '*m*'.
$m-\overline{cycle}$	A circuit (may be formed through the unfailed regenerative or non regenerative state) whose terminals are at the regenerative '*m*' state.
$U_{k,k}$	Probability factor of the state '*k*' reachable from the terminal state '*k*' of '*k*' cycle.
$U_{\overline{k,k}}$	The probability factor of state '*k*' reachable from the terminal state '*k*' of $k\ \overline{cycle}$.
μ_i	Mean sojourn time spent in the state '*i*' before visiting any other states.
μ_i'	Total unconditional time spent before transiting to any other regenerative state while the system entered regenerative state '*i*' at t=0.
η_i	Expected waiting time spent while doing a job given that the system entered to the regenerative state '*i*' at t=0.
A/a	The first unit is in the operative state/failed state.
$B/\overline{B}/b$	The second unit is in the operative state/ standby state/ failed state.
	The third unit is in the operative state/reduced state/failed state.
λ_1, λ_2	Fixed failure rate of the unit A/B respectively.
p/q	Probability of switch properly working/not working.
S/s	The switch is in perfect mode/Imperfect mode.
θ_1, θ_2	Fixed repair rate of the unit A/B after respectively.
γ	Fixed repair rate of the switch.
○ ⬭ ▭	Upstate/ reduced state/ failed state.

IV. Circuits Descriptions

Primary, secondary and tertiary circuits are used to find the base state such that

Table 1: *Circuit Descriptions*

i	(C1)	(C2)	(C3)
0	(0, 2, 0) (0, 1, 2, 0) (0, 2, 4, 3, 0)	Nil Nil (4, 3, 4) (3, 4, 3)	Nil Nil Nil Nil
1	1, 2, 0, 1	(0,2,0) (0,2,4,3,0) (2,0,2) (2,4,3,0,2)	Nil (4,3,4) (3,4,3) Nil (4,3,4) (3,4,3)
2	(2,0,2) (2,0,1,2) (2,4,3,0,2)	Nil Nil (3,4,3) (4,3,4)	Nil Nil Nil Nil
3	(3,0,2,4,3)	(0,2,0) (0,1,2,0) (2,0,2)	Nil Nil Nil
4	(4, 3, 0, 2, 4)	(0,2,0) (2,0,2)	Nil Nil

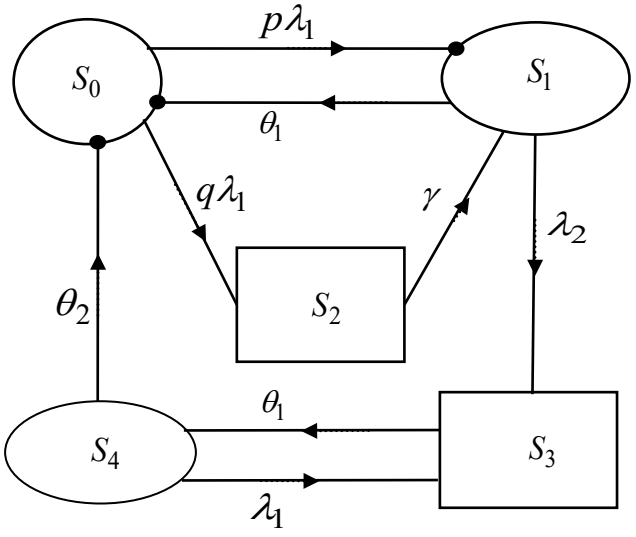


Figure 1 *State Transition Diagram*

where, $S_0 = A\bar{B}S$, $S_1 = aBS$, $S_2 = a\bar{B}s$, $S_3 = abS$, $S_4 = AbS$

V. Transition Probabilities

There are following transition probabilities

$$p_{0,1} = p\,,\ p_{0,2} = q\,,\ p_{1,0} = \theta_1/(\theta_1+\lambda_2)\,,\ p_{1,3} = \lambda_2/(\theta_1+\lambda_2)$$
$$p_{2,1} = p_{3,4} = 1,\ p_{4,0} = \theta_2/(\theta_2+\lambda_1)\,,\ p_{4,3} = \lambda_1/(\theta_2+\lambda_1) \quad (1)$$

It has been conclusively established that

$$p_{0,1} + p_{0,2} = 1,\ \ p_{1,0} + p_{1,3} = 1,\ p_{2,1} = 1,\ p_{3,4} = 1,\ p_{4,0} + p_{4,3} = 1$$

VI. Mean Sojourn Time

For the particular state, it becomes

$$\mu_0 = 1/\lambda_1,\ \mu_1 = 1/(\theta_1+\lambda_2)\ ,\ \mu_2 = 1/\gamma\ ,\ \mu_3 = 1/\theta_1$$
$$\mu_4(t) = 1/(\theta_2+\lambda_1) \quad (2)$$

VII. Evaluation of Parameters

Using the circuit table, '0' is used as the base state to calculate the reliability using the regenerative point graphical technique. The probability factors of all the reachable states from the base state '0' are given below

$$U_{0,0} = (0,1,0) = 1\,,\ U_{0,1} = p_{0,1}\,,\ U_{0,2} = p_{0,2}$$
$$U_{0,3} = p_{0,1}p_{1,3} + p_{0,2}p_{2,1}p_{1,3}\,,\ U_{0,4} = p_{0,1}p_{1,3}p_{3,4} + p_{0,2}p_{2,1}p_{1,3}p_{3,4}$$

I. Mean Time to System Failure

The regenerative un-failed states (i=0, 1) to which the system can transit (with initial state 0) before entering to any failed state (using base state ξ=0) then MTSF becomes

$$T_0 = \left[\sum_{i=0}^{1} Sr\left\{\frac{\left\{pr(0\xrightarrow{Sr(sff)} i)\right\}.\mu_i}{\Pi_{k_1\neq 0}\left\{1-V_{k_1k_1}\right\}}\right\}\right] \div \left[1-\sum Sr\left\{\frac{\left\{pr(0\xrightarrow{Sr(sff)} 0)\right\}}{\Pi_{k_2\neq 0}\left\{1-V_{k_2k_2}\right\}}\right\}\right]$$
$$T_0 = \frac{U_{0,0}\mu_0 + U_{0,1}\mu_1}{[1-(1,0,1)]} \quad (3)$$

II. Availability of the system

The system is available for use at regenerative states j=0, 1, 4 with ξ=0 then the availability of system is defined as

$$A_0 = \left[\sum_{j} Sr\left\{\frac{\left\{pr(0\xrightarrow{Sr} j)\right\}.f_j.\mu_j}{\Pi_{k_1\neq 0}\left\{1-V_{k_1k_1}\right\}}\right\}\right] \div \left[\sum_{i=0}^{4} Sr\left\{\frac{\left\{pr(0\xrightarrow{Sr} i)\right\}.\mu_i'}{\Pi_{k_2\neq 0}\left\{1-V_{k_2k_2}\right\}}\right\}\right]$$
$$A_0 = \frac{U_{0,0}\mu_0 + U_{0,1}\mu_1 + U_{0,4}\mu_4}{U_{0,0}\mu_0 + U_{0,1}\mu_1 + U_{0,2}\mu_2 + U_{0,3}\mu_3 + U_{0,4}\mu_4} \quad (4)$$

III. Busy Period of the Technician

The Technician is busy due to repair of the failed unit at regenerative states j= 1, 2, 3, 4 with ξ = 0 then the fraction of time for which the server remains busy is defined as

$$B_0 = \left[\sum_j Sr \left\{ \frac{\left\{ pr(0 \xrightarrow{Sr} j) \right\}.\eta_j}{\prod_{k_1 \neq 0} \left\{ 1 - V_{k_1 k_1} \right\}} \right\} \right] \div \left[\sum_{i=0}^{4} Sr \left\{ \frac{\left\{ pr(0 \xrightarrow{Sr} i) \right\}.\mu_i'}{\prod_{k_2 \neq 0} \left\{ 1 - V_{k_2 k_2} \right\}} \right\} \right]$$

$$B_0 = \frac{U_{0,1}\mu_1 + U_{0,2}\mu_2 + U_{0,3}\mu_3 + U_{0,4}\mu_4}{U_{0,0}\mu_0 + U_{0,1}\mu_1 + U_{0,2}\mu_2 + U_{0,3}\mu_3 + U_{0,4}\mu_4} \tag{5}$$

IV. Estimated number of visits made by the Technician

The technician visits at regenerative states j= 1, 2 with ξ=0 then the number of visits by the repairman is defined as

$$V_0 = \left[\sum_j Sr \left\{ \frac{\left\{ pr(0 \xrightarrow{Sr} j) \right\}}{\prod_{k_1 \neq 0} \left\{ 1 - V_{k_1 k_1} \right\}} \right\} \right] \div \left[\sum_{i=0}^{4} Sr \left\{ \frac{\left\{ pr(0 \xrightarrow{Sr} i) \right\}.\mu_i'}{\prod_{k_2 \neq 0} \left\{ 1 - V_{k_2 k_2} \right\}} \right\} \right]$$

$$V_0 = \frac{U_{0,1}\mu_1 + U_{0,2}\mu_2}{U_{0,0}\mu_0 + U_{0,1}\mu_1 + U_{0,2}\mu_2 + U_{0,3}\mu_3 + U_{0,4}\mu_4} \tag{6}$$

V. Profit Analysis

The profit of the system is analyzed by

$$P = E_0 A_0 - E_1 B_0 - E_2 V_0 \tag{7}$$

where, $E_0 = 15000$ (Revenue per unit uptime of the system)

$E_1 = 500$ (Cost per unit time for which technician is busy due to repair)

$E_2 = 100$ (Cost per visit of the technician)

VIII. Discussion

Regenerative point graphical technique plays an important role to determine the reliability metrics of the repairable system. Here, λ_i are the failure rate and θ_j are the repair rate. Let all $\lambda_i=\lambda$ (i=1, 2) and $\theta_j=\theta$ (j=1, 2) then tables 2, 3 and 4 describe the nature of mean time to system failure, availability and profit values of the two unit system under imperfect switch having an increasing trend corresponding to increment in repair rate (θ).

In these tables, the values of parameters change such that λ=0.3, 0.4, 0.5 and θ=0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 respectively. When λ=0.3 changes into 0.4, 0.5 then MTSF, availability and profit values have decreasing trends.

Table 2: *MTSF vs. Repair Rate* (θ)

θ ↓	λ=0.3	λ=0.4	λ=0.5
0.5	3.628692	3.333333	3.090278
0.55	3.675035	3.368794	3.132184
0.6	3.720609	3.403509	3.173516
0.65	3.765432	3.4375	3.214286
0.7	3.809524	3.47079	3.254505
0.75	3.852901	3.503401	3.294183
0.8	3.895582	3.535354	3.333333
0.85	3.937583	3.566667	3.371965
0.9	3.97892	3.59736	3.410088
0.95	4.019608	3.627451	3.447712

Table 3: *Availability vs. Repair Rate* (θ)

θ ↓	λ=0.3	λ=0.4	λ=0.5
0.5	0.623324	0.604782	0.542904
0.55	0.628307	0.609813	0.547959
0.6	0.633159	0.614717	0.552904
0.65	0.637887	0.619499	0.557741
0.7	0.642494	0.624164	0.562476
0.75	0.646985	0.628716	0.56711
0.8	0.651365	0.633159	0.571646
0.85	0.655637	0.637497	0.576089
0.9	0.659806	0.641734	0.58044
0.95	0.663876	0.645873	0.584703

Table 4: *Profit vs. Repair Rate* (θ)

θ ↓	λ=0.3	λ=0.4	λ=0.5
0.5	2438.338	2386.076	2019.52
0.55	2467.262	2415.876	2049.467
0.6	2495.431	2444.927	2078.759
0.65	2522.874	2473.257	2107.417
0.7	2549.618	2500.892	2135.461
0.75	2575.691	2527.857	2162.912
0.8	2601.117	2554.178	2189.787
0.85	2625.919	2579.875	2216.104
0.9	2650.121	2604.972	2241.881
0.95	2673.744	2629.49	2267.135

IX. Conclusion

The performance of the two non identical unit system is discussed using the regenerative point graphical technique. The above tables explore that when the repair rate increases then the MTSF, availability and profit values also increase but when the failure rate increases then the MTSF, availability and profit values decrease.

It is clear that RPGT is helpful for industries to analyze the behaviour of the products and components of a system.

X. Future Scope

It is observed that the role of the regenerative point graphical technique for the two non identical unit system will be beneficial and also used by the management, manufacturers and the persons engaged in reliability engineering and working on analyzing the nature and performance analysis of the system like soft drink, paper industry.

References

[1] Balagurusamy, E. (1984). *Reliability Engineering*. *Tata McGraw-Hill Education*.

[2] Bhardwaj, R. K., Kaur, K., and Malik, S. C. (2017). Reliability indices of a redundant system with standby failure and arbitrary distribution for repair and replacement times. *International Journal of System Assurance Engineering and Management*, 8, 423-431.

[3] Gopalan, M. N., and Bhanu, K. S. (1995). Cost analysis of a two unit repairable system subject to on-line preventive maintenance and/or repair. *Microelectronics Reliability*, 35, 251-258.

[4] Gupta, V. K., and Singh, J. (2007). Behavior and profit analysis of a redundant system with imperfect switch. *Journal of Mathematics and System Science*, 3, 16–35.

[5] Kumar, J. and Goel, M. (2016). Availability and profit analysis of a two-unit cold standby system for general distribution. *Cogent Mathematics*, 3(1): 1262937.

[6] Kumar, P., & Sirohi, A. (2015). Profit analysis of a two-unit cold standby system with the delayed repair of the partially failed unit and better utilization of units. *International Journal of Computer Applications*, 117(1), 41-46.

[7] Lim, T. (2000). Analysis of System Reliability with Dependent Repair Modes. *IEEE Transaction on Reliability*, 49(2), 153-162.

[8] Sadeghi, M., and Roghanian, E. (2017). Reliability analysis of a warm standby repairable system with two cases of imperfect switching mechanism. *Scientia Iranica*, 24(2), 808-822.

[9] Singh, V.V., Poonia, P.K., and Adbullahi, A.H. (2020). Performance analysis of a complex repairable system with two subsystems in series configuration with an imperfect switch. *Journal of Mathematical and Computational Science*, 10 (2), 359–383.

[10] Taneja, G., Malhotra, R., and Chitkara, A.K. (2016). Comparative profit analysis of two reliability models with varying demand. *Arya Bhatta Journal of Mathematics and Informatics*, 8(2), 305-314.

BAYESIAN INFERENCE OF WEIBULL-PARETO DISTRIBUTION UNDER DOUBLE TYPE *I* HYBRID CENSORED DATA

Khawla Boudjerda

•

Departement of mathematics, Mohammed Seddik Ben Yahia university, Jijel, Algeria
khawla.boudjerda@univ-jijel.dz

Abstract

This paper investigates the estimation of parameters, reliability, and failure rate functions of the Weibull-Pareto distribution using double type I hybrid censored data. We begin by applying the maximum likelihood method to derive point estimates for the distribution parameters. Subsequently, we explore Bayesian estimation techniques, obtaining Bayesian estimators under various loss functions to enhance robustness. To compute these estimators, we utilize Markov Chain Monte Carlo (MCMC) methods, facilitating effective sampling from complex posterior distributions. We employ Pitman closeness criteria to compare the performance of Bayesian estimators against those derived from maximum likelihood estimation, providing a comprehensive evaluation of their accuracy and efficiency. Additionally, a real data example is presented to illustrate the practical application of our methodologies. The results underscore the advantages of the Bayesian approach, particularly in scenarios characterized by hybrid censoring, while also contributing to the broader understanding of reliability analysis in statistical modeling.

Keywords: Weibull-Pareto distribution, double type *I* hybrid censored data, MCMC methods, Pitman closeness.

1. Introduction

Historically, it has been believed that the Pareto distribution and its associated generalizations are appropriate for modeling income and wealth distributions. A notable generalization is the New Weibull-Pareto distribution (NWPD), defined by [1]. This distribution is particularly useful in modeling real-life scenarios and can simulate data with a bathtub-shaped hazard rate, which is important for feature engineering in reliability analysis. The cumulative distribution function (CDF) of the NWPD, characterized by the shape parameter β and the scale parameters α and λ is given by:

$$F(x;\alpha,\beta,\lambda) = 1 - exp(-\alpha(\frac{x}{\lambda})^{\beta}), \;\; x > 0, \alpha, \beta, \lambda > 0, \tag{1}$$

its probability density function (PDF) is given

$$f(x;\alpha,\beta,\lambda) = \frac{\alpha\beta}{\lambda}(\frac{x}{\lambda})^{\beta-1} exp(-\alpha(\frac{x}{\lambda})^{\beta}), \;\; x > 0. \tag{2}$$

The Reliability and the failure rate function of the NWPD are respectively given by

$$R(x;\alpha,\beta,\lambda) = 1 - F(x;\alpha,\beta,\lambda) = exp(-\alpha(\frac{x}{\lambda})^{\beta}), \;\; x > 0, \tag{3}$$

and

$$h(x;\alpha,\beta,\lambda) = \frac{f(x;\alpha,\beta,\lambda)}{R(x;\alpha,\beta,\lambda)} = \frac{\alpha\beta}{\lambda}(\frac{x}{\lambda})^{\beta-1}, \quad x > 0. \tag{4}$$

The representation graphics of the PDF and reliability functions of the NWPD for differents values of α, β and λ are given by

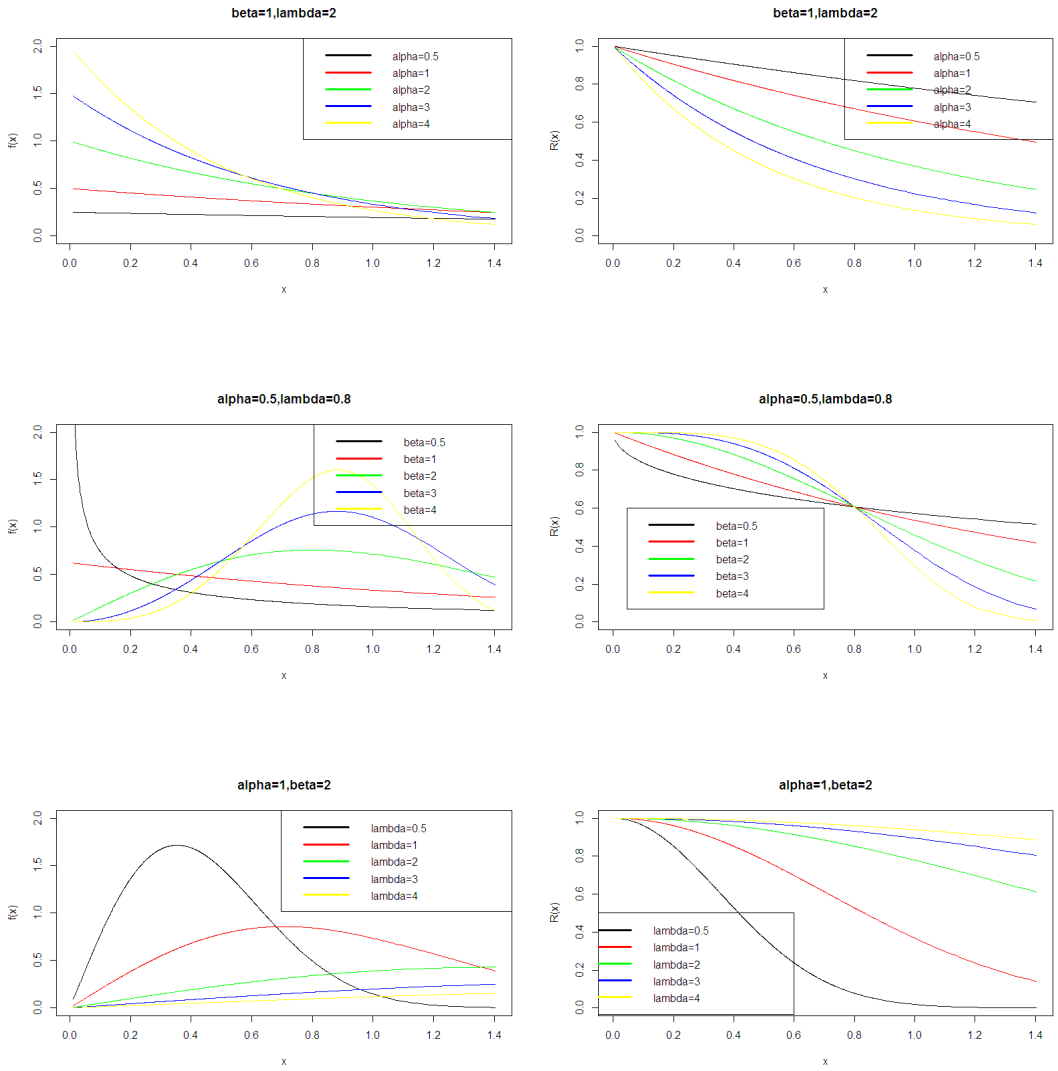


Figure 1: *The PDF and reliability functions of the NWPD for differents values of α, β and λ*

Many authors have discussed various properties of the NWPD and have obtained the maximum likelihood estimators for its parameters (see, [2], [1], [3]).

In this article, we obtain the maximum likelihood and Bayesian estimators of the parameters, as well as the reliability and failure rate functions, using double type I hybrid censored data. The hybrid censored scheme (HCS) is the mixture of type I and type II censored test has appeared. In double type I hybrid censored sheme, we place n units identically independent distributed in a test. k is un integer verify $k > 0$ and $k < n$ where k is given at the beginning. Let t_1 and t_2 be the censored time checked $0 < t_1 < t_2$. The test carried out until time t_1 and k_1 are the number of units which fail. If $k_1 > k$, the test will end at time t_1 and $n - k_1$ units are removed from the test. If $k_1 < k$, the test will end at the time t_2 and k_2 is the number of failed components. The two previous cases can be summarized as follows:

Case 1: $X_{1;n}$, $X_{2;n}$,...,$X_{k_1;n}$ if $k_1 > k$.
Case 2: $X_{1;n}$, $X_{2;n}$,...,$X_{k_2;n}$ if $k_1 < k$.

Hybrid censored data appears in many works. The two-sided confidence interval using double Type I hybrid censored schemes is introduced by [4]. The exact confidence bounds for an exponential parameter under hybrid censoring data are proposed by [5]. Then, [6] obtained the exact two-sided confidence interval of θ following the approach of [5]. Recently, hybrid censored schemes have gained significant popularity in reliability and life-testing experiments. The estimation of parameters for the log-normal distribution based on hybrid censored data is studied by [7]. Additionally, the estimation and prediction of the generalized Lindley distribution using hybrid censored data are provided by [8], while the estimation of unknown parameters of the inverted linear exponential distribution based on Type I hybrid censored data is investigated by [6].

The reste of this paper is organaised as follows, In section 2, we estimate the parameters, reliability and failure rate functions of the NWPD distribution using the maximum likelihood mathod. Bayesian estimation under different loss functions is presented in section 3. Monte-Carlo simulation results are applied in section 5. Analysis of one real data set is introduced in section 6.

2. Maximum likelihood estimation

In this section, maximum likelihood estimators of parameters, reliability function and failure rate function of the NWPD based on double type I hybrid censored data is considered. Let $x = (x_{1:n}, x_{2:n}, ..., x_{m,n})$ be double type I hybrid censored sample from the NWPD, then the likelihood function is given by

$$L(\alpha, \beta, \lambda|x) = \prod_{i=1}^{m} f(x_{i:n}; \alpha, \beta, \lambda)[1 - F(t; \alpha, \beta, \lambda)]^{n-m}, \tag{5}$$

where

$$t = \begin{cases} t_1, & case\,1 \\ t_2, & case\,2 \end{cases}$$

and

$$m = \begin{cases} k_1, & case\,1 \\ k_2, & case\,2 \end{cases}$$

$$L(\alpha, \beta, \lambda|x) = \alpha^m \beta^m \lambda^{-m\beta} \prod_{i=1}^{m} x_i^{\beta-1} exp[-\alpha \sum_{i=1}^{m} (\frac{x_i}{\lambda})^\beta - \alpha(n-k)(\frac{t}{\lambda})^\beta]. \tag{6}$$

Then the log-likelihood function is

$$l(\alpha, \beta, \lambda|x) = mln\alpha + mln\beta - m\beta ln\lambda + (\beta - 1) \sum_{i=1}^{m} lnx_i - \alpha \sum_{i=1}^{m} (\frac{x_i}{\lambda})^\beta - \alpha(n-m)(\frac{t}{\lambda})^\beta. \tag{7}$$

The log-likelihood function derivatives are given by

$$\begin{cases} \frac{\partial l(\alpha,\beta,\lambda|x)}{\partial \alpha} = \frac{m}{\alpha} - \sum_{i=1}^{m} (\frac{x_i}{\lambda})^\beta - (n-m)(\frac{t}{\lambda})^\beta. \\ \frac{\partial l(\alpha,\beta,\lambda|x)}{\partial \beta} = \frac{m}{\beta} - mln\lambda + \sum_{i=1}^{m} lnx_i - \alpha \sum_{i=1}^{m} ln(\frac{x_i}{\lambda})(\frac{x_i}{\lambda})^\beta - \alpha(n-m)ln(\frac{t}{\lambda})(\frac{t}{\lambda})^\beta. \\ \frac{\partial l(\alpha,\beta,\lambda|x)}{\partial \lambda} = -\frac{m\beta}{\lambda} + \frac{\alpha\beta}{\lambda}(\frac{x_i}{\lambda})^\beta + \frac{\alpha\beta}{\lambda}(n-m)(\frac{t}{\lambda})^\beta. \end{cases}$$

We can't obtain the explicit analytical form of the estimators, several methods are applied to find the approximate values of the maximum likelihood estimators $\widehat{\alpha}_{ml}$, $\widehat{\beta}_{ml}$ and $\widehat{\lambda}_{ml}$. In this article, we use the R package BB ([10]), (the BBsolve function).

To obtain the maximum likelihood estimators of the reliability function $R(x, \alpha, \beta, \lambda)$ and the failure rate function $h(x, \alpha, \beta, \lambda)$ we replace α, β and λ by $\widehat{\alpha}_{ml}$, $\widehat{\beta}_{ml}$ and $\widehat{\lambda}_{ml}$ in the formulas 3 et 4.

3. Bayesian estimation

In this section, we consider the Bayesian estimators of parameters, reliability function and failure rate function of the NWPD based on double type I hybrid censored data. We suppose, that the parameters α, β and λ of the NWPD have independent gamma distributions:

$$\pi(\alpha) \propto \alpha^{a_1-1} exp(-b_1\alpha), \ \ \alpha > 0. \tag{8}$$

$$\pi(\beta) \propto \beta^{a_2-1} exp(-b_2\beta), \ \ \beta > 0. \tag{9}$$

$$\pi(\lambda) \propto \lambda^{a_3-1} exp(-b_3\alpha), \ \ \lambda > 0. \tag{10}$$

where the hyper parameters $a_1 > 0$, $b_1 > 0$, $a_2 > 0$, $b_2 > 0$, $a_3 > 0$ and $b_3 > 0$.
Thus, the joint prior distribution of (α, β, λ) is given by

$$\pi(\alpha, \beta, \lambda) \propto \alpha^{a_1-1}\beta^{a_2-1}\lambda^{a_3-1} exp(-b_1\alpha - b_2\beta - b_3\lambda)$$

Using the Bayesian formula, the posterior probability density function of (α, β, λ) is

$$\pi(\alpha, \beta, \lambda|x) = K^{-1}\alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x) \tag{11}$$

where K is the normalisation constant given by

$$K = \int_0^{+\infty}\int_0^{+\infty}\int_0^{+\infty} \alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda.$$

$\zeta(x) = \prod_{i=1}^{m} x_i^{\beta-1}$ and $\phi(x) = exp[-\alpha\sum_{i=1}^{m}(\frac{x_i}{\lambda})^\beta - \alpha(n-m)(\frac{t}{\lambda})^\beta - b_1\alpha - b_2\beta - b_3\lambda]$

3.1. Bayesian estimation under squared loss function

Under squared loss function $L_1(\theta, \delta) = (\theta - \delta)^2$, the Bayesian estimators is the posterior mean

$$\widehat{\delta}_{B1} = E_\pi(\delta|x) = \int\int\int \delta\pi(\alpha, \beta, \lambda|x)d\alpha d\beta d\lambda \tag{12}$$

In the case of the NWPD, we obtain the Bayesian estimators of the parameter α, β and λ

$$\begin{aligned}\hat{\alpha}_{B1} &= E_\pi(\alpha|x)\\ &= \int\int\int \alpha\pi(\alpha, \beta, \lambda|x)d\alpha d\beta d\lambda\\ &= K^{-1}\int\int\int_0^{+\infty} \alpha^{m+a_1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\end{aligned}$$

$$\begin{aligned}\hat{\beta}_{B1} &= E_\pi(\beta|x)\\ &= \int\int\int \beta\pi(\alpha, \beta, \lambda|x)d\alpha d\beta d\lambda\\ &= K^{-1}\int\int\int_0^{+\infty} \alpha^{m+a_1-1}\beta^{m+a_2}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\end{aligned}$$

$$\begin{aligned}\hat{\lambda}_{B1} &= E_\pi(\lambda|x)\\ &= \int\int\int \lambda\pi(\alpha, \beta, \lambda|x)d\alpha d\beta d\lambda\\ &= K^{-1}\int\int\int_0^{+\infty} \alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3}\zeta(x)\phi(x)d\alpha d\beta d\lambda.\end{aligned}$$

The Bayesian estimator of the reliability function under the squared loss function is given by

$$\widehat{R}_{B1}(x) = E_\pi(R(x)|x) = \int\int\int R(x)\pi(\alpha, \beta, \lambda|x)d\alpha d\beta d\lambda$$

$$= K^{-1} \int\int\int_0^{+\infty} \alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)exp\left[-\alpha\left(\frac{x}{\lambda}\right)^{\beta}\right] d\alpha d\beta d\lambda. \quad (13)$$

The Bayesian estimator of the failure rate function under the squared loss function is given by

$$\widehat{h}_{B1}(x) = E_\pi(h(x)|x) = \int\int\int h(x)\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda$$

$$= K^{-1} \int\int\int_0^{+\infty} \alpha^{m+a_1}\beta^{m+a_2}\lambda^{-m\beta+a_3-2}\zeta(x)\left(\frac{x}{\lambda}\right)^{\beta-1}\phi(x)d\alpha d\beta d\lambda. \quad (14)$$

3.2. Bayesian estimation under entropy loss function

Under Entropy loss function $L_2(\theta,\delta) = (\frac{\delta}{\theta})^p - plog(\frac{\delta}{\theta}) - 1$, the Bayesian estimator is given by

$$\widehat{\delta}_{B_2} = E_\pi(\delta^{(-p)}|x)^{\frac{-1}{p}} = \left[\int\int\int \delta^{-p}\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}, \ p \neq 0 \quad (15)$$

In the case of the NWPD, we obtained

$$\hat{\alpha}_{B2} = E_\pi(\alpha^{-p}|x)^{\frac{-1}{p}}, \ p \neq 0$$

$$= \left[\int\int\int \alpha^{-p}\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}$$

$$= \left[K^{-1}\int\int\int_0^{+\infty} \alpha^{m+a_1-1-p}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}.$$

$$\hat{\beta}_{B2} = E_\pi(\beta^{-p}|x)^{\frac{-1}{p}}, \ p \neq 0$$

$$= \left[\int\int\int \beta^{-p}\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}$$

$$= \left[K^{-1}\int\int\int_0^{+\infty} \alpha^{m+a_1-1}\beta^{m+a_2-1-p}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}.$$

$$\hat{\lambda}_{B2} = E_\pi(\lambda^{-p}|x)^{\frac{-1}{p}}, \ p \neq 0$$

$$= \left[\int\int\int \lambda^{-p}\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}$$

$$= \left[K^{-1}\int\int\int_0^{+\infty} \alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1-p}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}.$$

The Bayesian estimator of the reliability function under the entropy loss function is given by

$$\widehat{R}_{B2}(x) = E_\pi(R(x)^{-p}|x)^{\frac{-1}{p}} = \left[\int\int\int R(x)^{-p}\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}, \ p \neq 0$$

$$= \left[K^{-1}\int\int\int_0^{+\infty} \alpha^{k+a_1-1}\beta^{k+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)exp\left[p\alpha(\frac{x}{\lambda})^{\beta}\right] d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}. \quad (16)$$

The Bayesian estimator of the failure rate function under the entropy loss function is given by

$$\widehat{h}_{B2}(x) = E_\pi(h(x)^{-p}|x)^{\frac{-1}{p}} = \left[\int\int\int h(x)^{-p}\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\right]^{\frac{-1}{p}}, \ p \neq 0$$

$$= \left[K^{-1}\int\int\int_0^{+\infty} \alpha^{k+a_1-p-1}\beta^{k+a_2-p-1}\lambda^{-m\beta+a_3+p-1}\zeta(x)(\frac{x}{\lambda})^{p(\beta-1)}\phi(x)d\alpha d\beta d\lambda\right]^{\frac{-1}{p}}. \quad (17)$$

3.3. Bayesian estimation under Linex loss function

Under Linex loss function $L_3(\delta,\theta) = exp(a(\delta-\theta)) - (\delta-\theta) - 1$, the Bayesian estimator is given by

$$\widehat{\delta}_{B_3} = \frac{-1}{a} log E_\pi(exp(-a\delta)) = -\frac{1}{a} log\left[\int\int\int exp(-a\delta)\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right],\ a \neq 0 \qquad (18)$$

In the case of the NWPD, the Bayesian estimators of the parameters are

$$\begin{aligned}\hat{\alpha}_{B3} &= \frac{-1}{a} log E_\pi(exp(-a\alpha)),\ a \neq 0\\ &= \frac{-1}{a} log\left[\int\int\int exp(-a\alpha)\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]\\ &= \frac{-1}{a} log\left[K^{-1}\int\int\int_0^{+\infty} exp(-a\alpha)\alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right].\end{aligned}$$

$$\begin{aligned}\hat{\beta}_{B3} &= \frac{-1}{a} log E_\pi(exp(-a\beta)),\ a \neq 0\\ &= \frac{-1}{a} log\left[\int\int\int exp(-a\beta)\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]\\ &= \frac{-1}{a} log\left[K^{-1}\int\int\int_0^{+\infty} exp(-a\beta)\alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right].\end{aligned}$$

$$\begin{aligned}\hat{\lambda}_{B3} &= \frac{-1}{a} log E_\pi(exp(-a\lambda)),\ a \neq 0\\ &= \frac{-1}{a} log\left[\int\int\int exp(-a\lambda)\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right]\\ &= \frac{-1}{a} log\left[K^{-1}\int\int\int_0^{+\infty} exp(-a\lambda)\alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right].\end{aligned}$$

The Bayesian estimator of the reliability function under Linex loss function is given by

$$\widehat{R}_{B3}(x) = -\frac{1}{a} log E_\pi(exp(-aR(x))) = -\frac{1}{a} log\left[\int\int\int exp(-aR(x))\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right],\ a \neq 0$$

$$= \frac{-1}{a} log\left[K^{-1}\int\int\int_0^{+\infty} \alpha^{k+a_1-1}\beta^{k+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)exp(-aexp(-\alpha(\frac{x}{\lambda})^\beta))d\alpha d\beta d\lambda\right]. \qquad (19)$$

The Bayesian estimator of the failure rate function under the Linex loss function is given by

$$\widehat{h}_{B3}(x) = -\frac{1}{a} log E_\pi(exp(-ah(x))) = -\frac{1}{a}\left[\int\int\int exp(-ah(x))\pi(\alpha,\beta,\lambda|x)d\alpha d\beta d\lambda\right],\ a \neq 0$$

$$= \frac{-1}{a} log\left[K^{-1}\int\int\int_0^{+\infty} exp\left(\frac{-a\alpha\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1}\right)\alpha^{m+a_1-1}\beta^{m+a_2-1}\lambda^{-m\beta+a_3-1}\zeta(x)\phi(x)d\alpha d\beta d\lambda\right] \qquad (20)$$

We can't obtain the explicit analytical form of all these estimators, for this we will use the MCMC methods in the next section.

4. Simulation

In this section, we generate $N = 1000$ samples of differents sizes $(n = 20, 40, 50)$ of the NWPD. We pose the $\alpha = \beta = \lambda = 1$, then we use the R package called BB (especially the function BBsolve) to solve the system of equations and find the maximum likelihood estimators and their mean squared errors, the results are in the next table

Table 1: *Maximum likelihood estimators of the parameters, reliability function and the failure rate function (mean squared errors)*

(n,m)	(t_1, t_2)	$\widehat{\alpha}_{ml}(MSE)$	$\widehat{\beta}_{ml}(MSE)$	$\widehat{\lambda}_{ml}(MSE)$	$\widehat{R}_{ml}(MSE)$	$\widehat{h}_{ml}(MSE)$
(20,16)	(6,7)	0.9216	0.8325	1.0967	0.5108	0.7456
		(0.0016)	(0.0280)	(0.0093)	(0.0014)	(0.0647)
	(8,11)	0.9333	0.7632	1.1417	0.5077	0.6896
		(0.0043)	(0.0560)	(0.0200)	(0.0012)	(0.0963)
	(10,15)	0.9570	0.6774	1.1749	0.4935	0.6378
		(0.0018)	(0.1040)	(0.0306)	(0.0004)	(0.1311)
(40,34)	(20,25)	0.9981	0.6733	1.1490	0.4728	0.6723
		$3.4 * 10^{-6}$	(0.1067)	(0.0222)	$(2.7 * 10^{-7})$	(0.1073)
	(22,30)	1.0054	0.6481	1.1590	0.4728	0.6723
		$(3 * 10^{-5})$	(0.1238)	(0.0259)	$(1.5 * 10^{-5})$	(0.1188)
	(24,33)	1.0041	0.6330	1.1702	0.4687	0.6395
		$(1.7 * 10^{-5})$	(0.1346)	(0.0289)	$(1.2 * 10^{-5})$	(0.1299)
(50,40)	(25,32)	1.0061	0.6712	1.1436	0.4686	0.6784
		$(3.8 * 10^{-5})$	(0.1080)	(0.0208)	$(1.4 * 10^{-5})$	(0.1034)
	(30,35)	1.0099	0.6511	1.1579	0.4671	0.6608
		$(9.9 * 10^{-5})$	(0.1217)	(0.0249)	$(2.7 * 10^{-5})$	(0.1150)
	(32,38)	1.0087	0.6420	1.1665	0.4678	0.6502
		$(7.6 * 10^{-5})$	(0.1281)	(0.0277)	$(2 * 10^{-5})$	(0.1222)

We consider the hyper parameters of the prior distribution

$$a_1 = a_2 = a_3 = 1, \quad b_1 = b_2 = b_3 = 2.$$

We present the Bayesian estimators of the parameters, the reliability function, the failure funtion and their posterior risks under the squared loss function, the entropy loss function and Linex loss function in the followings tables

Table 2: *Bayesian estimators of the parameters, reliability function and the failure rate function under squared loss function (posterior risk)*

(n,m)	(t_1, t_2)	$\widehat{\alpha}_{B_1}(PR)$	$\widehat{\beta}_{B_1}(PR)$	$\widehat{\lambda}_{B_1}(PR)$	$\widehat{R}_{B_1}(PR)$	$\widehat{h}_{B_1}(PR)$
(20,16)	(6,7)	1.0199	0.9702	1.0497	0.4889	0.9500
		(0.0022)	(0.0027)	(0.0043)	(0.0002)	(0.0024)
	(8,11)	1.0213	0.9716	1.0511	0.4891	0.9510
		(0.0022)	(0.0025)	(0.0043)	(0.0002)	(0.0023)
	(10,15)	1.0208	0.9711	1.0506	0.4890	0.9508
		(0.0021)	(0.0025)	(0.0042)	(0.0002)	(0.0024)
(40,34)	(20,25)	1.0217	0.9721	1.0506	0.4892	0.9510
		(0.0025)	(0.0028)	(0.0047)	(0.0002)	(0.0023)
	(22,30)	1.0221	0.9725	1.0519	0.4892	0.9510
		(0.0027)	(0.0030)	(0.0049)	(0.0002)	(0.0023)
	(24,33)	1.0221	0.9724	1.0519	0.4892	0.9508
		(0.0029)	(0.0032)	(0.0051)	(0.0002)	(0.0024)
(50,40)	(25,32)	1.0213	0.9717	1.0510	0.4892	0.9507
		(0.0025)	(0.0029)	(0.0047)	(0.0002)	(0.0024)
	(30,35)	1.0235	0.9739	1.0533	0.4895	0.9520
		(0.0029)	(0.0030)	(0.0051)	(0.0002)	(0.0022)
	(32,38)	1.0255	0.9759	1.0553	0.4898	0.9532
		(0.0031)	(0.0031)	(0.0055)	(0.0003)	(0.0021)

Table 3: *Bayesian estimators of the parameters, reliability function and the failure rate function under entropy loss function (posterior risk)*

(n,m)	(t_1, t_2)	$\widehat{\alpha}_{B_2}(PR)$	$\widehat{\beta}_{B_2}(PR)$	$\widehat{\lambda}_{B_2}(PR)$	$\widehat{R}_{B_2}(PR)$	$\widehat{h}_{B_2}(PR)$
(20,16)	(6,7)	1.0182	0.9685	1.0481	0.4795	0.9493
		(0.0181)	(0.0319)	(0.0469)	(0.0151)	(0.0520)
	(8,11)	1.0197	0.9699	1.0495	0.4796	0.9504
		(0.0195)	(0.0304)	(0.0488)	(0.0153)	(0.0508)
	(10,15)	1.0192	0.9695	1.0491	0.4796	0.9501
		(0.0191)	(0.0309)	(0.0479)	(0.0151)	(0.0511)
(40,34)	(20,25)	1.0198	0.9701	1.0496	0.4798	0.9502
		(0.0196)	(0.0303)	(0.0484)	(0.0156)	(0.0510)
	(22,30)	1.0201	0.9703	1.0499	0.4798	0.9502
		(0.0199)	(0.300)	(0.0487)	(0.0157)	(0.0003)
	(24,33)	1.0199	0.9701	1.0497	0.4799	0.9499
		(0.0197)	(0.0302)	(0.0485)	(0.0159)	(0.0532)
(50,40)	(25,32)	1.0194	0.9697	1.0492	0.4797	0.9499
		(0.0192)	(0.0307)	(0.0480)	(0.0155)	(0.0513)
	(30,35)	1.0214	0.9717	1.0512	0.4800	0.9512
		(0.0212)	(0.0286)	(0.0499)	(0.0161)	(0.0500)
	(32,38)	1.0233	0.9736	1.0531	0.4803	0.9523
		(0.0230)	(0.0267)	(0.0517)	(0.0166)	(0.0488)

Table 4: *Bayesian estimators of the parameters, reliability function and the failure rate function under Linex loss function (posterior risk)*

(n,m)	(t_1, t_2)	$\widehat{\alpha}_{B_3}(PR)$	$\widehat{\beta}_{B_3}(PR)$	$\widehat{\lambda}_{B_3}(PR)$	$\widehat{R}_{B_3}(PR)$	$\widehat{h}_{B_3}(PR)$
(20,16)	(6,7)	1.0190	0.9693	1.0488	0.4796	0.9496
		(0.0008)	(0.0008)	(0.0008)	(0.0093)	(0.0003)
	(8,11)	1.0204	0.9707	1.0502	0.4797	0.9507
		(0.0008)	(0.0008)	(0.008)	(0.0094)	(0.0003)
	(10,15)	1.0200	0.9703	1.0498	0.4796	0.9505
		(0.0008)	(0.0008)	(0.0008)	(0.0093)	(0.0003)
(40,35)	(20,25)	1.0207	0.9710	1.0505	0.4798	0.9506
		(0.0010)	(0.0010)	(0.0010)	(0.0094)	(0.0003)
	(22,30)	1.0210	0.9714	1.0508	0.4799	0.9506
		(0.0011)	(0.0011)	(0.0010)	(0.0094)	(0.0003)
	(24,33)	1.0209	0.9712	1.0507	0.4800	0.9504
		(0.0011)	(0.0011)	(0.0011)	(0.0094)	(0.0004)
(50,40)	(25,32)	1.0203	0.9707	1.0500	0.4798	0.9503
		(0.0010)	(0.0010)	(0.0010)	(0.0093)	(0.0003)
	(30,35)	1.0224	0.9728	1.0521	0.4801	0.9516
		(0.0011)	(0.0011)	(0.0011)	(0.0094)	(0.0004)
	(32,38)	1.0243	0.9747	1.0541	0.4803	0.9528
		(0.0012)	(0.0012)	(0.0012)	(0.0094)	(0.0004)

Comparaison of estimators

We use [9] criterion to make a comparaison between the Bayesian and the maximum likelihood estimators.

Definition 1. An estimator $\widehat{\delta}_1$ of a parameter θ is better than author estimator $\widehat{\delta}_2$ in the sens of Pitman's criterion if for all values of θ

$$P(|\widehat{\delta}_1 - \theta| < |\widehat{\delta}_2 - \theta|) > \frac{1}{2}$$

We conclude that for all values of (n, m, t_1, t_2), the Bayesian estimators of the parametrs α, β and λ are better than the maximum likelihood estimators in the case of the NWPD.

Table 5: *Pitman criteriom*

(n,m)	(t1,t2)	parameters	squared	entropy	Linex
$(20,16)$	$(6,7)$	α	0.848	0.855	0.855
		β	0.897	0.902	0.899
		λ	0.920	0.926	0.921
	$(8,11)$	α	0.843	0.844	0.804
		β	0.891	0.904	0.897
		λ	0.907	0.910	0.909
	$(10,15)$	α	0.798	0.809	0.804
		β	0.865	0.877	0.872
		λ	0.883	0.890	0.887
$(50,40)$	$(25,32)$	α	0.885	0.887	0.886
		β	0.883	0.918	0.906
		λ	0.881	0.885	0.881
	$(22,30)$	α	0.876	0.881	0.878
		β	0.891	0.902	0.897
		λ	0.881	0.884	0.882
	$(24,33)$	α	0.880	0.887	0.886
		β	0.907	0.913	0.910
		λ	0.878	0.893	0.881
$(50,40)$	$(25,32)$	α	0.885	0.887	0.886
		β	0.899	0.908	0.904
		λ	0.870	0.872	0.871
	$(30,35)$	α	0.911	0.917	0.915
		β	0.901	0.913	0.905
		λ	0.898	0.902	0.901
	$(35,38)$	α	0.900	0.906	0.903
		β	0.898	0.903	0.900
		λ	0.905	0.911	0.909

Discussion of results
We notice that all the probabilities are greater than 0.5. So, the Bayesian estimators of the parameters of the NWPD are more consistent than its maximum likelihood estimators.

5. Real data analysis

In this section, we use the real data represents the runoff amounts at Jug Bridge, Maryland

Table 6: *Data set of the rumoff amounts at Jug Bridge, Maryland*

0.17	0.19	0.23	0.33	0.39	0.39	0.40	0.45	0.52	0.56	0.59	0.64	0.66
0.70	0.76	0.77	0.78	0.95	0.97	1.02	1.12	1.24	1.59	1.74	2.92	

- We apply the Kolmogorov-Smirnov (K-S) test on this data set, we obtain

Table 7: *K-S test*

K-S value	0.1601
P-value	0.9062

So, we observe that the NWPD model fit quite well to this data set.
Using this real data set, we calculate the differents estimators of the parameters, reliability and failure rate functions with the maximum likelihood and the Bayesian methods, the results are:

Table 8: *Maximum likelihood estimators*

m	t_1	t_2	$\widehat{\alpha}_{ml}$	$\widehat{\beta}_{ml}$	$\widehat{\lambda}_{ml}$	$\widehat{R}_{ml}$	$\widehat{h}_{ml}$
15	0.8	1.2	1.0014	0.9842	1.0969	0.5021	0.9039
			$(2 * 10^{-6})$	(0.0024)	(0.0094)	(0.0008)	(0.0092)
	1	1.8	1.0963	0.9641	0.9665	0.4237	1.1036
			(0.0092)	(0.0012)	(0.0011)	(0.0023)	(0.0107)
18	0.9	1.2	1.1100	0.9241	0.9736	0.4182	1.0767
			(0.0121)	(0.0054)	(0.0007)	(0.0029)	(0.0059)
	1	1.5	1.0963	0.9641	0.09665	0.4237	1.1036
			(0.0092)	(0.0012)	(0.0011)	(0.0023)	(0.0107)
20	1.1	1.6	1.0915	0.9802	0.9673	0.4271	1.1117
			(0.0083)	(0.0004)	(0.0010)	(0.0020)	(0.0124)
	1.5	1.9	1.0816	0.9924	0.9764	0.4349	1.1015
			(0.0066)	$(5 * 10^{-5})$	(0.0005)	(0.0013)	(0.0103)
23	1.3	1.7	1.0801	1.0087	0.9671	0.4330	1.1198
			(0.0064)	$(1 * 10^{-5})$	0.001	0.0015	0.0123
	1.6	2	1.0643	0.9859	0.9644	0.4357	1.0919
			(0.0041)	(0.0002)	(0.0013)	(0.0013)	(0.0084)

Table 9: *Bayesian estimators of the parameters α, β and λ*

m	t_1	t_2	$\widehat{\alpha}_{B_1}$	$\widehat{\beta}_{B_1}$	$\widehat{\lambda}_{B_1}$	$\widehat{\alpha}_{B_2}$	$\widehat{\beta}_{B_2}$	$\widehat{\lambda}_{B_2}$	$\widehat{\alpha}_{B_3}$	$\widehat{\beta}_{B_3}$	$\widehat{\lambda}_{B_3}$
15	0.8	1.2	1.1644	1.1158	1.1547	1.0598	1.0101	1.0498	1.0768	1.0262	1.0668
			(0.0888)	(0.0717)	(0.0850)	(0.1438)	(0.1443)	(0.1439)	(-0.0740)	(-0.0265)	(0.0617)
	1	1.8	1.2226	1.1727	1.2126	1.0801	1.0303	1.0701	1.1131	1.0634	1.1032
			(0.0797)	(0.0599)	(0.0753)	(0.1586)	(0.1590)	(0.1587)	(0.1072)	(0.0615)	(0.0982)
18	0.9	1.2	1.2547	1.2048	1.2447	1.1332	1.0833	1.1232	1.1446	1.0963	1.1349
			(0.0921)	(0.0692)	(0.0871)	(0.1055)	(0.1066)	(0.1056)	(0.1351)	(0.0919)	(0.1266)
	1	1.5	1.2435	1.1937	1.2335	1.1639	1.1140	1.1539	1.1769	1.1271	1.1670
			(0.0928)	(0.0710)	(0.0881)	(0.0748)	(0.0753)	(0.0749)	(0.1629)	(0.1197)	(0.1544)
20	1.1	1.6	1.2507	1.2011	1.2408	1.1732	1.1238	1.1633	1.1838	1.1339	1.1738
			(0.1009)	(0.0784)	(0.0960)	(0.0655)	(0.0656)	(0.0654)	(0.1687)	(0.1257)	(0.1603)
	1.5	1.9	1.2581	1.2083	1.2408	1.1351	1.0854	1.1252	1.1769	1.1271	1.1670
			(0.1025)	(0.0791)	(0.0974)	(0.1036)	(0.1040)	(0.1037)	(0.1629)	(0.1197)	(0.1544)
23	1.3	1.7	1.2360	1.1863	1.2260	1.1460	1.0966	1.1361	1.1964	1.1466	1.1864
			(0.0952)	(0.0741)	(0.0906)	(0.0927)	(0.0896)	(0.0899)	(0.1793)	(0.1368)	(0.1709)
	1.6	2	1.2294	1.1795	1.2194	1.1772	1.1273	1.1672	1.2037	1.1541	1.1938
			(0.0904)	(0.0700)	(0.0859)	(0.0615)	(0.0620)	(0.0616)	(0.1874)	(0.1433)	(0.1771)

Table 10: *Bayesian estimators of the reliability and failure rate functions*

m	t_1	t_2	$\hat{R}_{B_1}$	$\hat{h}_{B_1}$	$\hat{R}_{B_2}$	$\hat{h}_{B_2}$	$\hat{R}_{B_3}$	$\hat{h}_{B_3}$
15	0.8	1.2	0.5012	1.0586	0.4903	0.9725	0.5004	0.9720
			(0.0008)	(0.0034)	(0.0096)	(0.0774)	(-0.0577)	(0.0283)
	1	1.8	0.4966	1.0486	0.4911	0.8969	0.4923	0.9755
			(0.0005)	(0.0023)	(0.0088)	(0.0630)	(-0.0413)	(0.0248)
18	0.9	1.2	0.4950	1.0511	0.4973	1.0252	0.4919	1.0130
			(0.0005)	(0.0026)	(0.0026)	(0.0247)	(-0.0406)	(-0.0129)
	1	1.5	0.5067	1.0502	0.4914	1.0194	0.4893	1.0011
			(0.0012)	(0.0025)	(0.0085)	(0.0305)	(-0.0353)	(-0.0011)
20	1.1	1.6	0.4946	1.0458	0.4950	1.0130	0.4924	1.0058
			(0.0004)	(0.0021)	(0.0003)	(0.0169)	(-0.0342)	(-0.0045)
	1.5	1.9	0.4947	1.0453	0.4950	1.0130	0.4924	1.0058
			(0.0005)	(0.0020)	(0.0049)	(0.0369)	(-0.0415)	(-0.0057)
23	1.3	1.7	0.4913	1.0376	0.4931	1.0019	0.4917	1.0033
			(0.0003)	(0.0014)	(0.0068)	(0.0480)	(-0.0402)	(-0.0032)
	1.6	2	0.4975	1.0526	0.5019	0.0330	0.4940	1.0170
			1(0.0006)	(0.0027)	(-0.0019)	(0.0169)	(-0.0449)	(-0.0169)

6. Conclusion

In this article, we have estimated the parameters and reliability functions of the Weibull-Pareto model, a valuable tool in engineering applications. We employed two estimation methods: the classical maximum likelihood method and the Bayesian approach, utilizing double type I hybrid censored data. Due to the complexity of the classical case, we relied on numerical methods to

derive the estimators, as explicit analytical forms were not attainable. For the Bayesian method, we assumed a gamma prior distribution for the parameters and calculated estimators under squared, entropy, and Linex loss functions. Since the estimators remained in integral form, we applied the Metropolis-Hastings algorithm to obtain their numerical values.

Our comparative analysis using Pitman's criterion revealed that the Bayesian estimators consistently outperformed the maximum likelihood estimators. To illustrate the applicability of our methodologies, we also conducted a study using real data. The findings underscore the effectiveness of the Weibull-Pareto model and highlight the advantages of the Bayesian approach in reliability analysis, providing valuable insights for future research and practical applications.

References

[1] Nasiro, S. (2015). The New Weibull-Pareto Distribution. *Pak.j.stat.oper.res*, XI:103–114.

[2] Alzatreeh, A. Lee, C and Famoye, F. (2013). A new method for generating families of continuous distributions. *METRON*, 71:63–79.

[3] Almetwally, E. and El mongy, H. (2019). Estimation methods for the new weibull-Pareto distribution: simulation and application. *Journal of Data Science*, 17:613–632.

[4] Epstein, B. (1954). Truncated life tests in the exponential case. *Annals of Mathematical Statistics*, 25:555–564.

[5] Chen, S. and Bhattacharya, G.K. (1998). Exact confidence bounds for an exponential parameter under hybrid censoring. *Communications in Statistics-Theory and Methods*, 17:1857–1870.

[6] Kundu, D. and Gupta, D. (1998). Hybrid censoring schemes with exponantial failure distribution. *COMMUN. STATIST.-THEORY METH*, 27:3065–3083.

[7] Habibi Rad, A. (2010). Analysis of Hybrid Censored Data from the Lognormal Distribution. *Statistical Research and Training Center*, 7:37–46.

[8] Habibi Rad, A. (2016). Estimation and prediction for Type-I hybrid censored data from generalized Lindley distribution. *Journal of Statistics and Management Systems*, 19:367–396.

[9] Pitman, E. (1937). The closest estimates of statistical parameters. *Mathematical proceeding of the Cambridge philosophical society*, 33.

[10] Varadhan, R. and Gilbert, P. (2010). An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of statistical software*, 32.

ESTIMATION OF RELIABILITY ON SEQUENTIAL ORDER STATISTICS FROM (k, n) SYSTEM

K. Glory Prasanth[1] and A.Venmani[2]

•

[1,2] Department of Mathematics and Statistics, Faculty of Science and Humanities,
SRM Institute of Science and Technology, Kattankulathur- 603203,
Chengalpattu district, Tamil nadu, India.
gk9617@srmist.edu.in[1] , venmania@srmist.edu.in[2]

Abstract

The focus of this paper is to introduce a reliability model for differently structured independent sequential (k, n) systems. In such a system, the failure of any component possibly influences the other components such that their underlying failure rate is parametrically adjusted with respect to the number of preceding failures. The system works if and only if at least k out of the n components works. By considering the different models of sequential (k, n) system, we obtain the reliability assuming that the system failure time belongs to exponential/gamma distribution with location and scale parameters. These results are important because the distributions can model diverse time-to-failure behavior. As the result it is found that the reliability decreases with increase in time by shifting location and scale parameters. This indicates that the reliability for different models of sequential (k, n) system are as expected.

Keywords: Sequential (k, n) systems, Reliability function, Exponential/Gamma distribution, Location and Scale parameters.

1. Introduction

Components designed to carry out a specific task make up a system. The component failure times are commonly assumed to be independent and identically distributed in the (k, n) system. The remaining operational components are supposed to remain unaffected by any component failure in the system. However, this is practically not applicable. Any failure in one of the system's components will impact the system as a whole. This places additional strain on the remaining active components, leading to an increase in stress levels. This leads to a rise in failure rates, a fall in efficiency, or both. As a result, an alternative flexible model was developed by taking into account variations in the active component's lifelengths distribution. In this model, as and when each component fails, the remaining active components take on the stress and change their distribution. The extensible model, developed specifically for this reason, is termed as a sequential (k, n) system. Sequential order statistics are the results of the sequential (k, n) system's order statistics. In this case, we suppose that the failure rate for the surviving active components changes with each component failure. A (k, n) system is said to be a sequential (k, n) system wherein lifelengths distribution of the active components changes if any one of the components fails. The r^{th} sequential order statistics model the life length of a sequential $(n - r + 1, n)$ system.

Cramer and Kamps [7] derived the basic results pertaining to the sequential order statistics. Bain [2], Barlow [4], and Meeker et al. [15] provide the most useful statistical techniques in the areas of reliability and life testing models. Basu and Mawaziny [3] identified the minimum variance unbiased estimator of the system's reliability at a given mission time. Baratnia and Doostparast [5] describe the lifetime of engineering systems when component lifetimes are dependent. Pham [18] discussed the most likely estimates of reliability and the uniformly minimal variance unbiased estimator for k-out-of-n systems. These systems consist of n independent, identically distributed components with exponential lifetimes. Méndez-González et al. [16] used the inverse power law and the exponentiated Weibull model to examine the reliability of an electronic component. Using the exponentiated Weibull distribution, Chaturvedi and Pathak [6] were able to get the reliability function's Maximum likelihood estimator. Demiray and Kizilaslan [9] looked in k-out-of-n system stress-strength reliability using point and interval estimates where stress and strength variable follows the proportional hazard rate model. Kalaivani and Kannan [12] introduced the concept of mean time to system failure and reliability function for (k,n) system using weibull distribution. Shi et al. [20] with Burr XII components used both classical and Bayes approaches to investigate how well the m consecutive (k, n): F system worked. Alghamdi and Percy [1] looked at the equivalence and reliability factors of a series-parallel system following exponentiated Weibull distribution. According to Hong and Meeker [11], the system's reliability is determined by the components and system structure.

This article aims to determine the reliability function based on sequential order statistics from different models of sequential (k,n) systems, under the assumption that the failure time distribution follows an exponential/gamma location-scale family. The article is organized as follows: In Section 2, the distribution function, marginal and joint density function of different models of sequential (k,n) systems are given. Section 3 calculates the reliability function of (1, 3) and (2, 3) systems. . Section 4 calculates the reliability function of (1, 4) and (2, 4) systems. In Section 5, numerical illustrations are presented for analysis, and Section 6 gives the conclusion about the result.

2. Sequential Order Statistics

According to Kamps [13], the joint density function of the first $r, (1 \le r \le n)$ sequential order statistics $X_*^{(1)}, X_*^{(2)}, \dots, X_*^{(r)}$ based on absolutely continuous distribution functions $F_1, F_2, \dots, F_n$ with respective density functions $f_1, f_2, \dots, f_n$ is given by

$$f^*(x_1, x_2, \dots, x_r) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} \left[\left\{ \frac{1-F_i(x_i)}{1-F_i(x_{i-1})} \right\}^{n-i} \frac{f_i(x_i)}{1-F_i(x_{i-1})} \right],$$

$$-\infty = x_0 < x_1 < \cdots < x_n < \infty.$$

Revathy and Chandrasekar [19] provide certain reliability metrics and equivariant parameter estimates based on sequential $(1,3)$ and $(2,3)$ systems. By introducing sequential $(1,4)$ and $(2,4)$ systems, Glory Prasanth and Venmani [10] computed the minimum risk equivariant estimator of location, scale, and location-scale families. In this section, we discuss about the distribution function, marginal and joint density function of system failure time for different models of sequential (k,n) system.

The density function of the random variable X with Gamma distribution (δ, τ, k) where δ is the location parameter, τ is the scale parameter and k is the shape parameter is defined as

$$f(x;\delta,\tau,k) = \frac{1}{(k-1)!\,\tau^k}(x-\delta)^{k-1} e^{-\frac{1}{\tau}(x-\delta)}\ ;\ \delta > 0,\ \tau > 0, k > 0, x > 0.$$

2.1. Model 1

Consider the sequential (1, 3) system which are absolutely continuous with the lifelength distributions F_1 , F_2 , F_3 having the respective density functions f_1 , f_2 , f_3 . Let f_1 and f_2 be the density function of Gamma distribution $(\delta, \tau, 2)$ and f_3 the density function of Gamma distribution $(\delta, \tau, 1)$. The sequential order statistics have the failure times as $X_*^{(1)}, X_*^{(2)}$ and $X_*^{(3)}$.

Suppose $F_1(x) = F_2(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)}\left[1 + \frac{x-\delta}{\tau}\right],\ \ \tau > 0,\ x > \delta,\ \ \delta \in R$

and $F_3(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \ \delta \in R.$

Then $f_1(x) = f_2(x) = \frac{1}{\tau^2}(x-\delta)e^{-\frac{1}{\tau}(x-\delta)},\ \ \tau > 0,\ x > \delta,\ \ \delta \in R$

and $f_3(x) = \frac{1}{\tau}e^{-\frac{1}{\tau}(x-\delta)},\ \ \tau > 0,\ x > \delta,\ \ \delta \in R.$

Thus the joint probability density functions of $X_*^{(1)}, X_*^{(2)}$ and $X_*^{(3)}$ is

$$f^*(x_1, x_2, x_3) = 6\frac{1}{\tau^5}(x_1 - \delta)(x_2 - \delta)\left[1 + \frac{x_2 - \delta}{\tau}\right]e^{-\frac{1}{\tau}(x_1 + x_2 + x_3 - 3\delta)},$$
$$\delta < x_1 < x_2 < x_3 < \infty,\ \delta \in R,\ \tau > 0. \qquad (1)$$

2.2. Model 2

Consider the sequential (2, 3) system which are absolutely continuous with the lifelength distributions F_1 , F_2 having the respective density functions f_1 , f_2 . Let f_1 be the density function of Gamma distribution $(\delta, \tau, 2)$ and f_2 be the density function of Gamma distribution $(\delta, \tau, 1)$. Let $X_*^{(1)}$ and $X_*^{(2)}$ be the failure times which are called sequential order statistics.

Suppose $F_1(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)}\left[1 + \frac{x-\delta}{\tau}\right],\ \ \tau > 0,\ x > \delta,\ \ \delta \in R$

and $F_2(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \ \delta \in R.$

Then $f_1(x) = \frac{1}{\tau^2}(x-\delta)e^{-\frac{1}{\tau}(x-\delta)},\ \ \tau > 0,\ x > \delta,\ \ \delta \in R$

and $f_2(x) = \frac{1}{\tau}e^{-\frac{1}{\tau}(x-\delta)},\ \ \tau > 0,\ x > \delta,\ \ \delta \in R.$

Thus the joint probability density function of $X_*^{(1)}$ and $X_*^{(2)}$ is

$$f^*(x_1, x_2) = 6\frac{1}{\tau^3}(x_1 - \delta)\left[1 + \frac{x_1 - \delta}{\tau}\right]^2 e^{-\frac{1}{\tau}(x_1 + 2x_2 - 3\delta)}, \delta < x_1 < x_2 < \infty, \delta \in R, \tau > 0. \qquad (2)$$

2.3. Model 3

Consider the sequential (1, 4) system which are absolutely continuous with the lifelength distributions F_1 , F_2 , F_3 , F_4 having the respective density functions f_1 , f_2 , f_3, f_4 . Let f_1 & f_2 be the density function of Gamma distribution $(\delta, \tau, 2)$ and f_3 & f_4 be the density function of Gamma distribution $(\delta, \tau, 1)$.

Suppose $F_1(x) = F_2(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)}\left[1 + \frac{x-\delta}{\tau}\right],\ \tau > 0,\ x > \delta,\ \ \delta \in R$

and $F_3(x) = F_4(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \delta \in R.$

Then $f_1(x) = f_2(x) = \frac{1}{\tau^2}(x-\delta)e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \delta \in R$

and $f_3(x) = f_4(x) = \frac{1}{\tau}e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \delta \in R.$

Thus the joint probability density function of $X_*^{(1)}, X_*^{(2)}, X_*^{(3)}$ and $X_*^{(4)}$ is

$$f^*(x_1,x_2,x_3,x_4) = \frac{24}{\tau^6}(x_1 - \delta)(x_2 - \delta)\left[1 + \frac{x_2-\delta}{\tau}\right]^2 e^{-\frac{1}{\tau}(x_1+x_2+x_3+x_4-4\delta)},$$
$$\delta < x_1 < x_2 < x_3 < x_4 < \infty,\ \delta \epsilon R,\ \tau > 0. \quad (3)$$

2.4. Model 4

Consider the sequential (2, 4) system which are absolutely continuous with the lifelength distributions F_1, F_2, F_3 having the respective density functions f_1, f_2, f_3. Let f_1 be the density function of Gamma distribution $(\delta, \tau, 2)$ and f_2 & f_3 be the density function of Gamma distribution $(\delta, \tau, 1)$. Let $X_*^{(1)}, X_*^{(2)}$ and $X_*^{(3)}$ be the failure times which are called sequential order statistics.

Suppose $F_1(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)}\left[1 + \frac{x-\delta}{\tau}\right],\ \tau > 0,\ x > \delta,\ \delta \in R$

and $F_2(x) = F_3(x) = 1 - e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \delta \in R.$

Then $f_1(x) = \frac{1}{\tau^2}(x-\delta)e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \delta \in R$

and $f_2(x) = f_3(x) = \frac{1}{\tau}e^{-\frac{1}{\tau}(x-\delta)},\ \tau > 0,\ x > \delta,\ \delta \in R.$

Thus the joint probability density function of $X_*^{(1)}, X_*^{(2)}$ and $X_*^{(3)}$ is

$$f^*(x_1,x_2,x_3) = 24\frac{1}{\tau^4}(x_1 - \delta)\left[1 + \frac{x_1 - \delta}{\tau}\right]^3 e^{-\frac{1}{\tau}(x_1+x_2+2x_3-4\delta)},$$
$$\delta < x_1 < x_2 < x_3 < \infty,\ \delta \epsilon R,\ \tau > 0. \quad (4)$$

3. Reliability Measure of (1, 3) and (2, 3) systems

In this section, using the joint density function given in (1) & (2), the reliability function for different models of sequential (1,3) and (2,3) systems are obtained.

The Reliability function for (1, 3) system is

$$R(t) = \int\int\int f^*(x_1,x_2,x_3)\,dx_3dx_2\,dx_1,\ t < x_1,x_2,x_3 < \infty$$
$$= \int\int\int 6\frac{1}{\tau^5}(x_1 - \delta)(x_2 - \delta)\left[1 + \frac{x_2 - \delta}{\tau}\right]e^{-\frac{1}{\tau}(x_1+x_2+x_3-3\delta)}\,dx_3dx_2\,dx_1,\ t < x_1,x_2,x_3 < \infty$$
$$= \frac{6}{\tau^3}(t-\delta)^3e^{-\frac{3}{\tau}(t-\delta)} + \frac{24}{\tau^2}(t-\delta)^2e^{-\frac{3}{\tau}(t-\delta)} + \frac{36}{\tau}(t-\delta)e^{-\frac{3}{\tau}(t-\delta)} + 18e^{-\frac{3}{\tau}(t-\delta)},$$
$$t < x_1,x_2,x_3 < \infty, \delta > 0, \tau > 0.$$

The Reliability function for (2, 3) system is

$$R(t) = \int\int f^*(x_1,x_2)dx_2dx_1\,, t < x_1, x_2, < \infty$$

$$= \int\int 6\frac{1}{\tau^3}(x_1 - \delta)\left[1+\frac{x_1-\delta}{\tau}\right]^2 e^{-\frac{1}{\tau}(x_1+2x_2-3\delta)}dx_2dx_1\,, t < x_1, x_2, < \infty$$

$$= \frac{3}{\tau^3}(t-\delta)^3e^{-\frac{3}{\tau}(t-\delta)} + \frac{15}{\tau^2}(t-\delta)^2e^{-\frac{3}{\tau}(t-\delta)} + \frac{24}{\tau}(t-\delta)e^{-\frac{3}{\tau}(t-\delta)} + 24e^{-\frac{3}{\tau}(t-\delta)},$$

$$t < x_1, x_2 < \infty, \delta > 0, \tau > 0.$$

4. Reliability Measure of (1, 4) and (2, 4) systems

In this section, using the joint density function given in (3) & (4), the reliability function for different models of sequential (1, 4) and (2, 4) systems are obtained.

The Reliability function for (1, 4) system is

$$R(t) = \int\int\int\int f^*(x_1,x_2,x_3,x_4)dx_4dx_3dx_2dx_1\,, t < x_1, x_2, x_3, x_4 < \infty$$

$$= \int\int\int\int \frac{24}{\tau^6}(x_1 - \delta)(x_2 - \delta)\left[1+\frac{x_2-\delta}{\tau}\right]^2 e^{-\frac{1}{\tau}(x_1+x_2+x_3+x_4-4\delta)}dx_4dx_3dx_2dx_1,$$

$$t < x_1, x_2, x_3, x_4 < \infty$$

$$= \frac{24}{\tau^4}(t-\delta)^4e^{-\frac{4}{\tau}(t-\delta)} + \frac{72}{\tau^3}(t-\delta)^3e^{-\frac{4}{\tau}(t-\delta)} + \frac{312}{\tau^2}(t-\delta)^2e^{-\frac{4}{\tau}(t-\delta)} + \frac{528}{\tau}(t-\delta)e^{-\frac{4}{\tau}(t-\delta)}$$
$$+ 264e^{-\frac{4}{\tau}(t-\delta)}, \quad t < x_1, x_2, x_3, x_4 < \infty, \delta > 0, \tau > 0.$$

The Reliability function for (2, 4) system is

$$R(t) = \int\int\int f^*(x_1,x_2,x_3)\,dx_3dx_2\,dx_1\,,\ t < x_1, x_2, x_3 < \infty$$

$$R(t) = \int\int\int 24\frac{1}{\tau^4}(x_1 - \delta)\left[1+\frac{x_1-\delta}{\tau}\right]^3 e^{-\frac{1}{\tau}(x_1+x_2+2x_3-4\delta)}\,dx_3dx_2\,dx_1\,,\ t < x_1, x_2, x_3 < \infty$$

$$= \frac{12}{\tau^4}(t-\delta)^4e^{-\frac{4}{\tau}(t-\delta)} + \frac{84}{\tau^3}(t-\delta)^3e^{-\frac{4}{\tau}(t-\delta)} + \frac{288}{\tau^2}(t-\delta)^2e^{-\frac{4}{\tau}(t-\delta)} + \frac{588}{\tau}(t-\delta)e^{-\frac{4}{\tau}(t-\delta)}$$
$$+ 588e^{-\frac{4}{\tau}(t-\delta)}, \quad t < x_1, x_2, x_3 < \infty, \delta > 0, \tau > 0.$$

5. Numerical illustration

In this section, numerical illustration is presented for sequential (k,n) systems. The reliability R(t) for different time (t) with various location parameter and scale parameter when the failure time of the system follows exponential/gamma distribution are calculated. The reliability for different time by shifting location and scale parameters for different models of sequential (k,n) systems are

presented in Table 1,2,3,4,5,6,7 and 8.

Table 1: *Reliability for (1, 3) system versus time for $\delta = 0.60$ and $\tau = 0.25, 0.50, 0.75, 1$*

S. No	δ=0.60 , τ=0.25		δ=0.60 , τ=0.5		δ=0.60 , τ=0.75		δ=0.60 , τ=1	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	0.89	0.97948	1.47	0.99987	1.91	0.98624	2.34	0.99987
2	0.91	0.77831	1.49	0.92062	1.93	0.93340	2.36	0.95948
3	0.93	0.61688	1.51	0.84726	1.95	0.88323	2.38	0.92062
4	0.95	0.48776	1.53	0.77941	1.97	0.83558	2.4	0.88323
5	0.97	0.38479	1.55	0.71669	1.99	0.79035	2.42	0.84726
6	0.99	0.30290	1.57	0.65873	2.01	0.74743	2.44	0.81267
7	1.01	0.23795	1.59	0.60522	2.03	0.70670	2.46	0.77941
8	1.03	0.18656	1.61	0.55583	2.05	0.66808	2.48	0.74743
9	1.05	0.14600	1.63	0.51027	2.07	0.63144	2.5	0.71669
10	1.07	0.11405	1.65	0.46827	2.09	0.59671	2.52	0.68713
11	1.09	0.08894	1.67	0.42956	2.11	0.56379	2.54	0.65873
12	1.11	0.06925	1.69	0.39391	2.13	0.53259	2.56	0.63144
13	1.13	0.05384	1.71	0.36109	2.15	0.50303	2.58	0.60522
14	1.15	0.04179	1.73	0.33088	2.17	0.47503	2.6	0.58003
15	1.17	0.03239	1.75	0.30310	2.19	0.44852	2.62	0.55583
16	1.19	0.02508	1.77	0.27756	2.21	0.42341	2.64	0.53259
17	1.21	0.01939	1.79	0.25409	2.23	0.39965	2.66	0.51027
18	1.23	0.01497	1.81	0.23252	2.25	0.37716	2.68	0.48884
19	1.25	0.01154	1.83	0.21272	2.27	0.35588	2.7	0.46827
20	1.27	0.00889	1.85	0.19455	2.29	0.33575	2.72	0.44852

Table 2: *Reliability for (1, 3) system versus time for $\tau = 0.65$ and $\delta = 0.2, 0.4, 0.6, 0.8$*

S. No	δ=0.2 , τ=0.65		δ=0.4 , τ=0.65		δ=0.6 , τ=0.65		δ=0.8 , τ=0.65	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	1.34	0.97174	1.54	0.97174	1.74	0.97174	2.15	0.98387
2	1.36	0.91186	1.56	0.91186	1.76	0.91186	2.17	0.93498
3	1.38	0.85544	1.58	0.85544	1.78	0.85544	2.19	0.88813
4	1.40	0.80230	1.6	0.80230	1.8	0.80230	2.21	0.84328
5	1.42	0.75227	1.62	0.75227	1.82	0.75227	2.23	0.80037
6	1.44	0.70518	1.64	0.70518	1.84	0.70518	2.25	0.75933
7	1.46	0.66088	1.66	0.66088	1.86	0.66088	2.27	0.72011
8	1.48	0.61921	1.68	0.61921	1.88	0.61921	2.29	0.68265
9	1.50	0.58003	1.7	0.58003	1.9	0.58003	2.31	0.64690
10	1.52	0.54320	1.72	0.54320	1.92	0.54320	2.33	0.61278
11	1.54	0.50859	1.74	0.50859	1.94	0.50859	2.35	0.58026
12	1.56	0.47608	1.76	0.47608	1.96	0.47608	2.37	0.54926
13	1.58	0.44555	1.78	0.44555	1.98	0.44555	2.39	0.51973
14	1.60	0.41689	1.80	0.41689	2.00	0.41689	2.41	0.49162
15	1.62	0.38998	1.82	0.38998	2.02	0.38998	2.43	0.46487
16	1.64	0.36474	1.84	0.36474	2.04	0.36474	2.45	0.43943
17	1.66	0.34106	1.86	0.34106	2.06	0.34106	2.47	0.41525
18	1.68	0.31885	1.88	0.31885	2.08	0.31885	2.49	0.39227
19	1.70	0.29802	1.9	0.29802	2.1	0.29802	2.51	0.37044
20	1.72	0.27850	1.92	0.27850	2.12	0.27850	2.53	0.34972

.

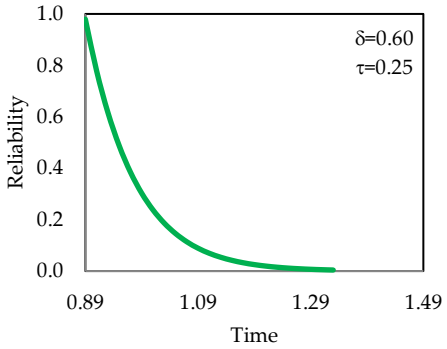


Figure 1: *Reliability versus time with $\delta = 0.60, \tau = 0.25$ for (1, 3) system.*

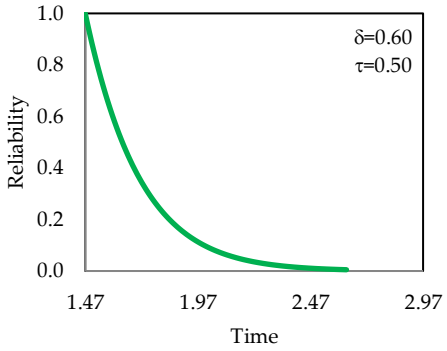


Figure 2: *Reliability versus time with $\delta = 0.60, \tau = 0.50$ for (1, 3) system.*

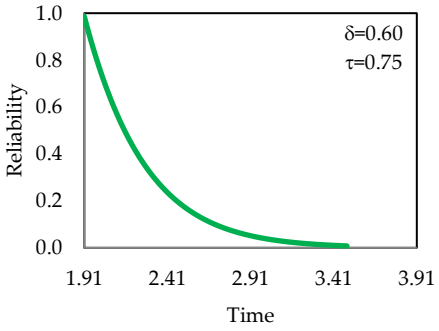


Figure 3: *Reliability versus time with $\delta = 0.60, \tau = 0.75$ for (1, 3) system.*

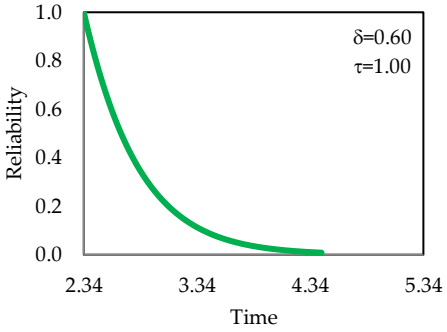


Figure 4: *Reliability versus time with $\delta = 0.60, \tau = 1.00$ for (1, 3) system.*

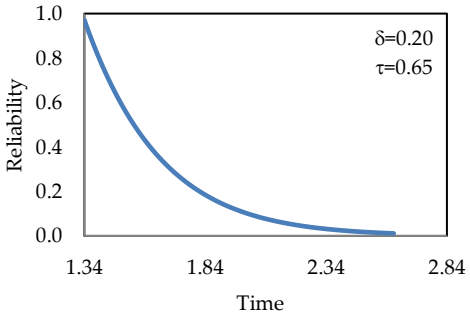


Figure 5: *Reliability versus time with $\delta = 0.20, \tau = 0.65$ for (1, 3) system.*

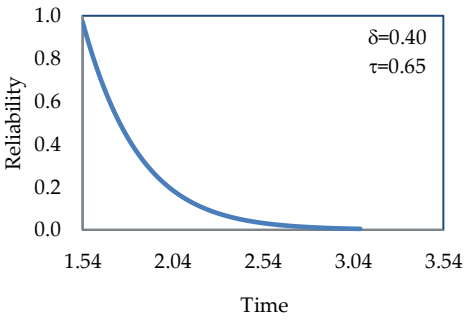


Figure 6: *Reliability versus time with $\delta = 0.40, \tau = 0.65$ for (1, 3) system.*

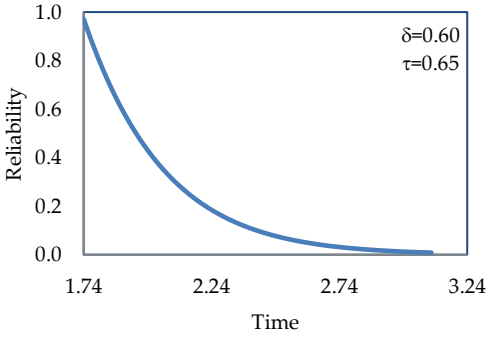


Figure 7: *Reliability versus time with $\delta = 0.60, \tau = 0.65$ for (1, 3) system.*

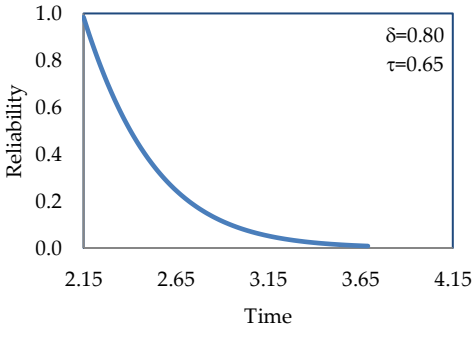


Figure 8: *Reliability versus time with $\delta = 0.80, \tau = 0.65$ for (1, 3) system*

From Table 1 & 2, it is observed that the sequential (1, 3) system's reliability decreases with increasing time and is also clear that time increases along with location and scale parameters.

Table 3: *Reliability for (2, 3) system versus time for $\delta = 0.60$ and $\tau = 0.25, 0.50, 0.75, 1$*

S. No	δ=0.60 , τ=0.25		δ=0.60 , τ=0.5		δ=0.60 , τ=0.75		δ=0.60 , τ=1	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	0.99	1.00000	1.51	1.00000	1.77	1.00000	2.34	0.99987
2	1.01	0.85354	1.53	0.93633	1.79	0.95786	2.36	0.95948
3	1.03	0.71726	1.55	0.86702	1.81	0.90425	2.38	0.92062
4	1.05	0.60204	1.57	0.80246	1.83	0.85354	2.4	0.88323
5	1.07	0.50476	1.59	0.74235	1.85	0.80556	2.42	0.84726
6	1.09	0.42272	1.61	0.68642	1.87	0.76018	2.44	0.81267
7	1.11	0.35363	1.63	0.63441	1.89	0.71726	2.46	0.77941
8	1.13	0.29551	1.65	0.58607	1.91	0.67668	2.48	0.74743
9	1.15	0.24669	1.67	0.54116	1.93	0.63831	2.5	0.71669
10	1.17	0.20573	1.69	0.49948	1.95	0.60204	2.52	0.68713
11	1.19	0.17140	1.71	0.46080	1.97	0.56776	2.54	0.65873
12	1.21	0.14266	1.73	0.42493	1.99	0.53537	2.56	0.63144
13	1.23	0.11863	1.75	0.39169	2.01	0.50476	2.58	0.60522
14	1.25	0.09855	1.77	0.36090	2.03	0.47584	2.6	0.58003
15	1.27	0.08180	1.79	0.33239	2.05	0.44852	2.62	0.55583
16	1.29	0.06784	1.81	0.30601	2.07	0.42272	2.64	0.53259
17	1.31	0.05621	1.83	0.28161	2.09	0.39835	2.66	0.51027
18	1.33	0.04654	1.85	0.25905	2.11	0.37535	2.68	0.48884
19	1.35	0.03850	1.87	0.23821	2.13	0.35363	2.70	0.46827
20	1.37	0.03183	1.89	0.21896	2.15	0.33313	2.72	0.44852

Table 4: *Reliability for (2, 3) system versus time for $\tau = 0.65$ and $\delta = 0.2, 0.4, 0.6, 0.8$*

S. No	δ=0.2 , τ=0.65		δ=0.4 , τ=0.65		δ=0.6 , τ=0.65		δ=0.8 , τ=0.65	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	1.22	0.99454	1.42	0.99454	1.72	1.00000	1.88	0.99508
2	1.24	0.93069	1.44	0.93069	1.74	0.97174	1.90	0.93731
3	1.26	0.87078	1.46	0.87078	1.76	0.91186	1.92	0.88257
4	1.28	0.81458	1.48	0.81458	1.78	0.85544	1.94	0.83074
5	1.30	0.76188	1.50	0.76188	1.80	0.80230	1.96	0.78167
6	1.32	0.71246	1.52	0.71246	1.82	0.75227	1.98	0.73525
7	1.34	0.66614	1.54	0.66614	1.84	0.70518	2.00	0.69135
8	1.36	0.62272	1.56	0.62272	1.86	0.66088	2.02	0.64985
9	1.38	0.58203	1.58	0.58203	1.88	0.61921	2.04	0.61064
10	1.40	0.54391	1.60	0.54391	1.90	0.58003	2.06	0.57361
11	1.42	0.50820	1.62	0.50820	1.92	0.54320	2.08	0.53865
12	1.44	0.47476	1.64	0.47476	1.94	0.50859	2.10	0.50567
13	1.46	0.44344	1.66	0.44344	1.96	0.47608	2.12	0.47455
14	1.48	0.41413	1.68	0.41413	1.98	0.44555	2.14	0.44522
15	1.50	0.38669	1.70	0.38669	2.00	0.41689	2.15	0.43119
16	1.52	0.36101	1.72	0.36101	2.02	0.38998	2.17	0.40436
17	1.54	0.33698	1.74	0.33698	2.04	0.36474	2.19	0.37908
18	1.56	0.31450	1.76	0.31450	2.06	0.34106	2.21	0.35528
19	1.58	0.29348	1.78	0.29348	2.08	0.31885	2.23	0.33289
20	1.60	0.27381	1.80	0.27381	2.10	0.29802	2.25	0.31182

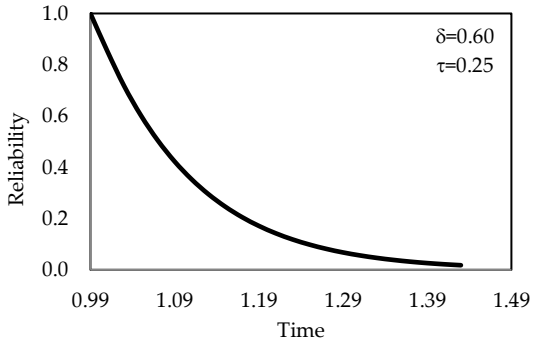


Figure 9: *Reliability versus time with $\delta = 0.60, \tau = 0.25$ for (2, 3) system.*

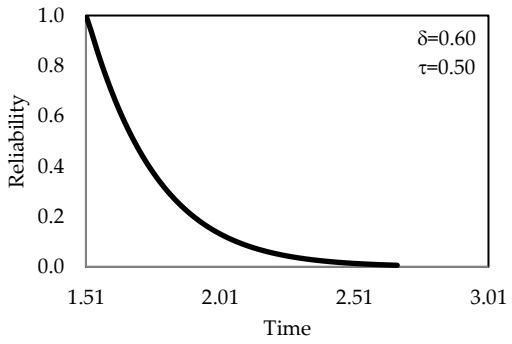


Figure 10: *Reliability versus time with $\delta = 0.60, \tau = 0.50$ for (2, 3) system.*

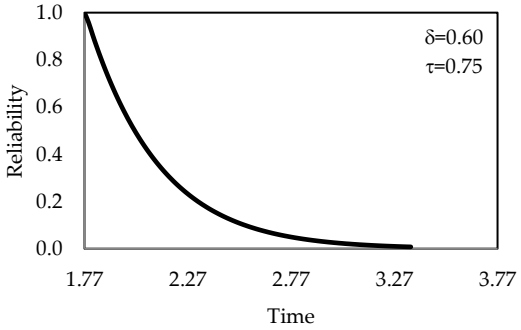


Figure 11: *Reliability versus time with $\delta = 0.60, \tau = 0.75$ for (2, 3) system.*

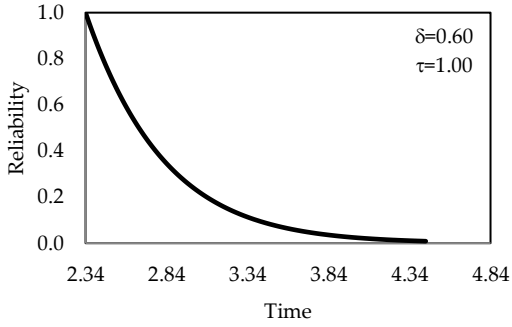


Figure 12: *Reliability versus time with $\delta = 0.60, \tau = 1$ for (2, 3) system.*

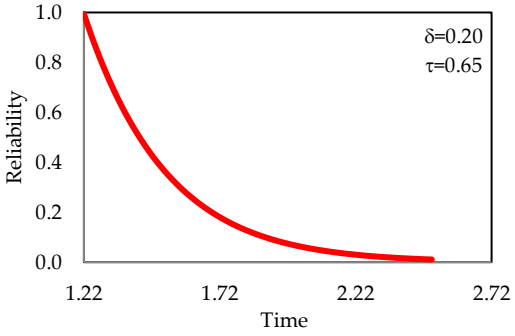


Figure 13: *Reliability versus time with $\delta = 0.2, \tau = 0.65$ for (2, 3) system.*

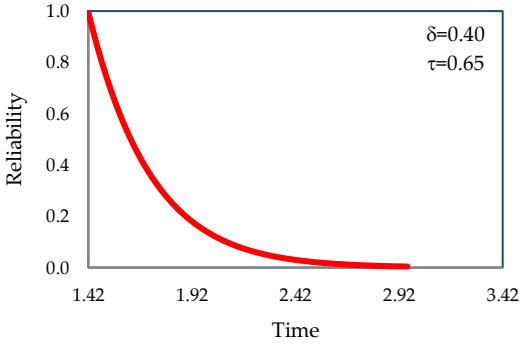


Figure 14: *Reliability versus time with $\delta = 0.40, \tau = 0.65$ for (2, 3) system.*

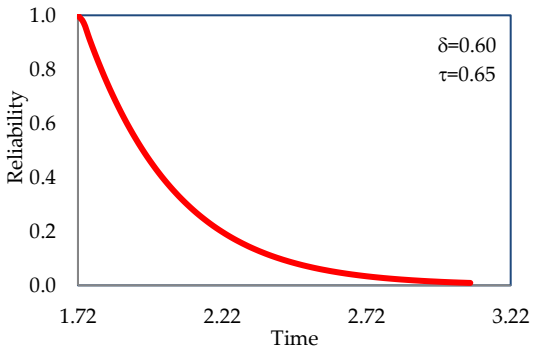


Figure 15: *Reliability versus time with $\delta = 0.6, \tau = 0.65$ for (2, 3) system.*

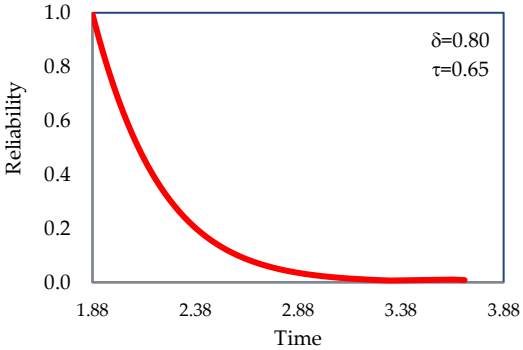


Figure 16: *Reliability versus time with $\delta = 0.8, \tau = 0.65$ for (2, 3) system.*

From Table 3 & 4, it is observed that the sequential (2, 3) system's reliability declines with increasing time and is also clear that time increases along with location and scale parameters.

Table 5: *Reliability for (1, 4) system versus time for δ=0.60 and τ=0.25, 0.50, 0.75, 1*

S. No	δ=0.60 , τ=0.25		δ=0.60 , τ=0.5		δ=0.60 , τ=0.75		δ=0.60 , τ=1	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	1.12	0.92791	1.64	0.92791	2.14	1.00000	2.66	0.98621
2	1.14	0.72668	1.66	0.82127	2.16	0.92791	2.68	0.92791
3	1.16	0.56846	1.68	0.72668	2.18	0.85541	2.70	0.87299
4	1.18	0.44422	1.70	0.64281	2.20	0.78847	2.72	0.82127
5	1.20	0.34677	1.72	0.56846	2.22	0.72668	2.74	0.77256
6	1.22	0.27042	1.74	0.50258	2.24	0.66965	2.76	0.72668
7	1.24	0.21068	1.76	0.44422	2.26	0.61702	2.78	0.68348
8	1.26	0.16397	1.78	0.39253	2.28	0.56846	2.80	0.64281
9	1.28	0.12751	1.80	0.34677	2.30	0.52366	2.82	0.60451
10	1.30	0.09906	1.82	0.30626	2.32	0.48233	2.84	0.56846
11	1.32	0.07689	1.84	0.27042	2.34	0.44422	2.86	0.53452
12	1.34	0.05963	1.86	0.23871	2.36	0.40906	2.88	0.50258
13	1.36	0.04621	1.88	0.21068	2.38	0.37665	2.90	0.47251
14	1.38	0.03577	1.90	0.18589	2.40	0.34677	2.92	0.44422
15	1.40	0.02767	1.92	0.16397	2.42	0.31922	2.94	0.41759
16	1.42	0.02139	1.94	0.14461	2.44	0.29382	2.96	0.39253
17	1.44	0.01652	1.96	0.12751	2.46	0.27042	2.98	0.36895
18	1.46	0.01275	1.98	0.11240	2.48	0.24885	3.00	0.34677
19	1.48	0.00984	2.00	0.09906	2.50	0.22898	3.02	0.32589
20	1.50	0.00758	2.02	0.08728	2.52	0.21068	3.04	0.30626

Table 6: *Reliability for (1, 4) system with time for τ= 0.65 and δ= 0.2, 0.4, 0.6, 0.8*

S. No	δ=0.2 , τ=0.65		δ=0.4 , τ=0.65		δ=0.6 , τ=0.65		δ=0.8 , τ=0.65	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	1.54	0.98160	1.74	0.98160	1.94	0.98160	2.14	0.98160
2	1.56	0.89373	1.76	0.89373	1.96	0.89373	2.16	0.89373
3	1.58	0.81358	1.78	0.81358	1.98	0.81358	2.18	0.81358
4	1.60	0.74050	1.80	0.74050	2.00	0.74050	2.20	0.74050
5	1.62	0.67388	1.82	0.67388	2.02	0.67388	2.22	0.67388
6	1.64	0.61315	1.84	0.61315	2.04	0.61315	2.24	0.61315
7	1.66	0.55780	1.86	0.55780	2.06	0.55780	2.26	0.55780
8	1.68	0.50737	1.88	0.50737	2.08	0.50737	2.28	0.50737
9	1.70	0.46143	1.90	0.46143	2.10	0.46143	2.30	0.46143
10	1.72	0.41958	1.92	0.41958	2.12	0.41958	2.32	0.41958
11	1.74	0.38147	1.94	0.38147	2.14	0.38147	2.34	0.38147
12	1.76	0.34677	1.96	0.34677	2.16	0.34677	2.36	0.34677
13	1.78	0.31517	1.98	0.31517	2.18	0.31517	2.38	0.31517
14	1.80	0.28642	2.00	0.28642	2.20	0.28642	2.40	0.28642
15	1.82	0.26025	2.02	0.26025	2.22	0.26025	2.42	0.26025
16	1.84	0.23643	2.04	0.23643	2.24	0.23643	2.44	0.23643
17	1.86	0.21477	2.06	0.21477	2.26	0.21477	2.46	0.21477
18	1.88	0.19506	2.08	0.19506	2.28	0.19506	2.48	0.19506
19	1.90	0.17714	2.10	0.17714	2.30	0.17714	2.50	0.17714
20	1.92	0.16084	2.12	0.16084	2.32	0.16084	2.52	0.16084

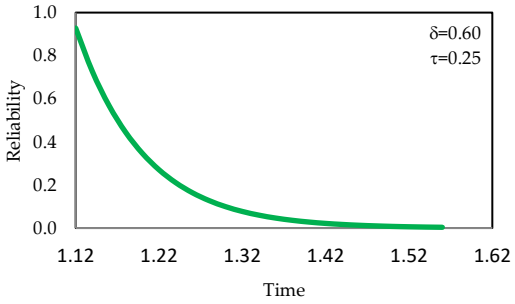


Figure 17: *Reliability versus time with $\delta = 0.60, \tau = 0.25$ for (1, 4) system.*

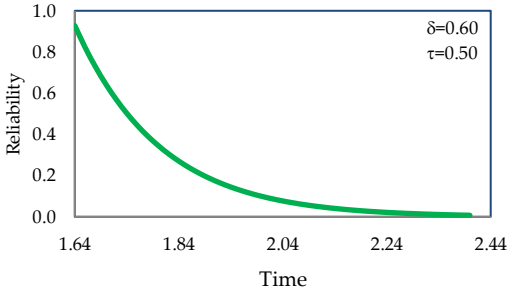


Figure 18: *Reliability versus time with $\delta = 0.60, \tau = 0.50$ for (1, 4) system.*

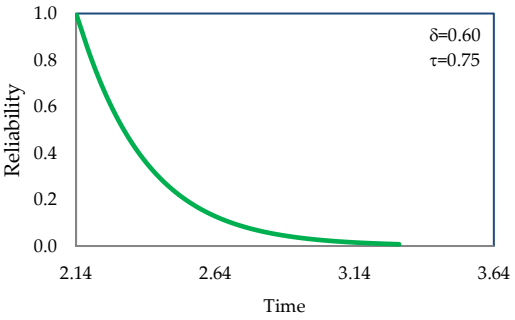


Figure 19: *Reliability versus time with $\delta = 0.60, \tau = 0.75$ for (1, 4) system.*

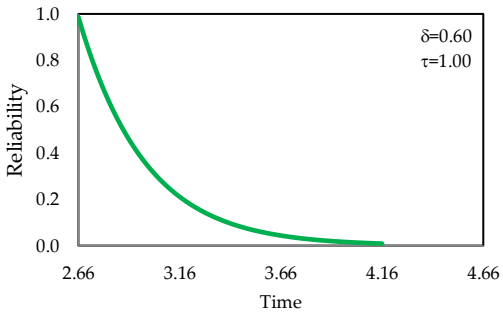


Figure 20: *Reliability versus time with $\delta = 0.60, \tau = 1.00$ for (1, 4) system.*

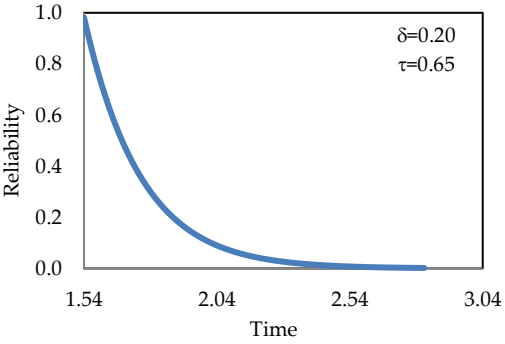


Figure 21: *Reliability versus time with $\delta = 0.20, \tau = 0.65$ for (1, 4) system.*

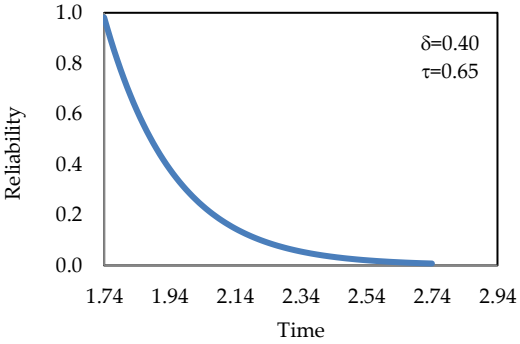


Figure 22: *Reliability versus time with $\delta = 0.40, \tau = 0.65$ for (1, 4) system.*

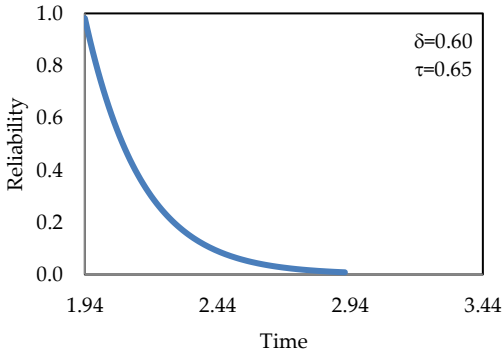


Figure 23: *Reliability versus time with $\delta = 0.60, \tau = 0.65$ for (1, 4) system.*

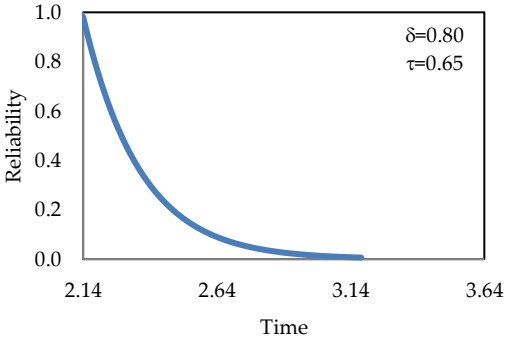


Figure 24: *Reliability versus time with $\delta = 0.8, \tau = 0.65$ for (1, 4) system.*

From Table 5 & 6, it can be observed that for sequential (1, 4) system's reliability declines with increasing time and it is also clear that time increases along with location and scale parameters.

Table 7: *Reliability for (2, 4) system with time for δ=0.60 and τ=0.25, 0.50, 0.75, 1*

S. No	δ=0.60 , τ=0.25		δ=0.60 , τ=0.5		δ=0.60 , τ=0.75		δ=0.60 , τ=1	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	1.12	0.9835	1.64	0.9835	2.16	0.98353	2.68	0.9835
2	1.14	0.7623	1.66	0.8660	2.18	0.90352	2.70	0.9229
3	1.16	0.5904	1.68	0.7623	2.20	0.82994	2.72	0.8660
4	1.18	0.4569	1.70	0.6709	2.22	0.76230	2.74	0.8125
5	1.20	0.3534	1.72	0.5904	2.24	0.70011	2.76	0.7623
6	1.22	0.2731	1.74	0.5194	2.26	0.64294	2.78	0.7152
7	1.24	0.2109	1.76	0.4569	2.28	0.59039	2.80	0.6709
8	1.26	0.1628	1.78	0.4019	2.30	0.54210	2.82	0.6294
9	1.28	0.1255	1.80	0.3534	2.32	0.49771	2.84	0.5904
10	1.30	0.0967	1.82	0.3107	2.34	0.45692	2.86	0.5538
11	1.32	0.0745	1.84	0.2731	2.36	0.41944	2.88	0.5194
12	1.34	0.0573	1.86	0.2400	2.38	0.38501	2.90	0.4872
13	1.36	0.0441	1.88	0.2109	2.40	0.35337	2.92	0.4569
14	1.38	0.0339	1.90	0.1853	2.42	0.32431	2.94	0.4285
15	1.40	0.0260	1.92	0.1628	2.44	0.29761	2.96	0.4019
16	1.42	0.0200	1.94	0.1429	2.46	0.27309	2.98	0.3768
17	1.44	0.0153	1.96	0.1255	2.48	0.25057	3.00	0.3534
18	1.46	0.0118	1.98	0.1102	2.50	0.22989	3.02	0.3313
19	1.48	0.0090	2.00	0.0967	2.52	0.21090	3.04	0.3107
20	1.50	0.0069	2.02	0.0849	2.54	0.19346	3.06	0.2913

Table 8: *Reliability for (2, 4) system with time for τ= 0.65 and δ= 0.2, 0.4, 0.6, 0.8*

S. No	δ=0.2 , τ=0.65		δ=0.4 , τ=0.65		δ=0.6 , τ=0.65		δ=0.8 , τ=0.65	
	t	R(t)	t	R(t)	t	R(t)	t	R(t)
1	1.56	0.99910	1.75	0.99321	1.95	0.99321	2.15	0.99321
2	1.58	0.90750	1.77	0.90057	1.97	0.90057	2.17	0.90057
3	1.60	0.82423	1.79	0.81649	1.99	0.81649	2.19	0.81649
4	1.62	0.74852	1.81	0.74017	2.01	0.74017	2.21	0.74017
5	1.64	0.67971	1.83	0.67092	2.03	0.67092	2.23	0.67092
6	1.66	0.61716	1.85	0.60808	2.05	0.60808	2.25	0.60808
7	1.68	0.56032	1.87	0.55107	2.07	0.55107	2.27	0.55107
8	1.70	0.50866	1.89	0.49935	2.09	0.49935	2.29	0.49935
9	1.72	0.46173	1.91	0.45243	2.11	0.45243	2.31	0.45243
10	1.74	0.41908	1.93	0.40988	2.13	0.40988	2.33	0.40988
11	1.76	0.38033	1.95	0.37129	2.15	0.37129	2.35	0.37129
12	1.78	0.34514	1.97	0.33630	2.17	0.33630	2.37	0.33630
13	1.80	0.31317	1.99	0.30458	2.19	0.30458	2.39	0.30458
14	1.82	0.28414	2.01	0.27581	2.21	0.27581	2.41	0.27581
15	1.84	0.25777	2.03	0.24974	2.23	0.24974	2.43	0.24974
16	1.86	0.23382	2.05	0.22611	2.25	0.22611	2.45	0.22611
17	1.88	0.21208	2.07	0.20469	2.27	0.20469	2.47	0.20469
18	1.90	0.19234	2.09	0.18529	2.29	0.18529	2.49	0.18529
19	1.92	0.17443	2.11	0.16770	2.31	0.16770	2.51	0.16770
20	1.94	0.15816	2.13	0.15177	2.33	0.15177	2.53	0.15177

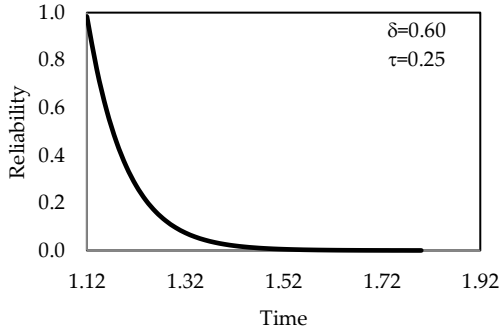


Figure 25: *Reliability versus time with $\delta = 0.60,\ \tau = 0.25$ for (2, 4) system.*

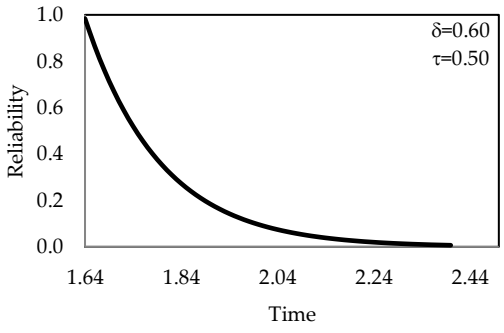


Figure 26:*Reliability versus time with $\delta = 0.60,\ \tau = 0.50$ for (2, 4) system.*

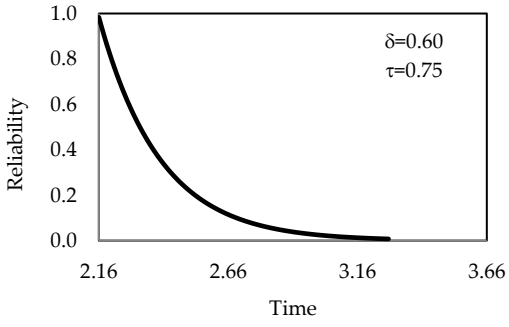


Figure 27: *Reliability versus time with $\delta = 0.60,\ \tau = 0.75$ for (2, 4) system.*

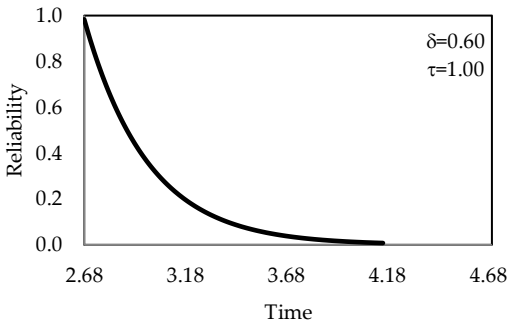


Figure 28: *Reliability versus time with $\delta = 0.60,\ \tau = 1$ for (2, 4) system.*

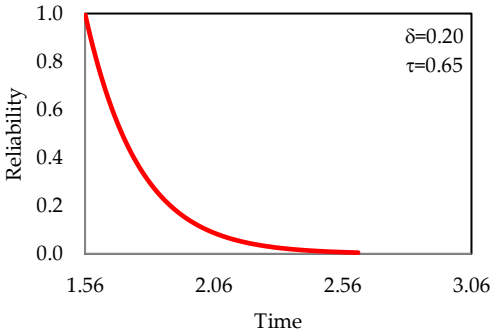


Figure 29: *Reliability versus time with $\delta = 0.2, \tau = 0.65$ for (2, 4) system.*

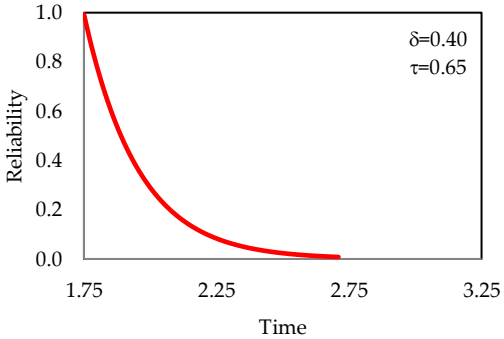


Figure 30: *Reliability versus time with $\delta = 0.40, \tau = 0.65$ for (2, 4) system.*

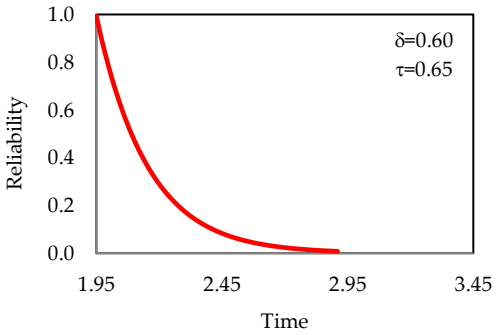


Figure 31: *Reliability versus time with $\delta = 0.6, \tau = 0.65$ for (2, 4) system.*

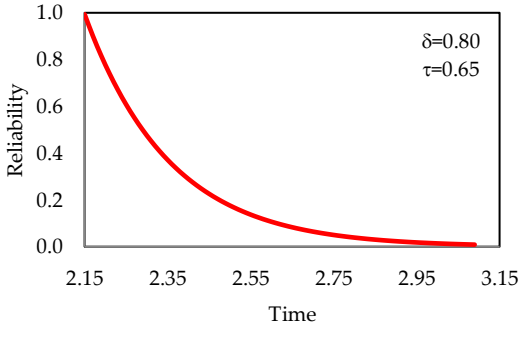


Figure 32: *Reliability versus time with $\delta = 0.8, \tau = 0.65$ for (2, 4) system.*

From table 7 and 8, it can be observed that for sequential (2, 4) system's reliability declines with increasing time and is also clear that time increases along with location and scale parameters.

From table 1,2,3,4,5,6,7,8 and figures, it is clear that as time increases, the different models of sequential (k, n) system's reliabilty decreases. It is also clear that time increases with increase in location and scale parameters. The results of numerical illustration clearly show that.

6. Conclusion

In this paper, different models of Sequential (k, n) system having exponential/gamma distribution with location and scale parameters are considered. The reliability function for different models of sequential (k, n) system's reliability decreases have been determined. Based on the findings in Table 1, 2, 3, 4, 5, 6, 7 & 8 and figures, we note the following

- By shifting the location and scale parameters, the reliability for different time are estimated for the suggested systems. As expected, the reliability decreases as the time increases.
- By monitoring reliability measures through shifting of parameters we can plan updates or patches to prevent system failures.
- It would be of interest to emulate this work for generalised gamma distribution and to extend this work for other continuous distributions, including Weibull distributions and Pareto distribution.

References

[1] Alghamdi, S. M., & Percy, D. F. (2014). Reliability equivalence factors for a series-parallel system assuming an exponentiated Weibull distribution.

[2] Bain, L. (2017). *Statistical analysis of reliability and life-testing models: theory and methods*. Routledge.

[3] Basu A.P., El Mawaziny A.H. (1978). Estimates of reliability of *k*-out-of-*m* structures in the independent exponential case. *Journal of American Statistical Association*, 73(364):850–854.

[4] Barlow, R. B. (1975). Statistical theory of reliability and life testing. *Holt*.

[5] Baratnia, M., & Doostparast, M. (2019). Sequential order statistics from dependent random variables. *Communications in Statistics-Theory and Methods*, *48*(18), 4569-4580.

[6] Chaturvedi, A., & Pathak, A. (2012). Estimation of the Reliability Function for Exponentiated Weibull Distribution. *Journal of Statistics & Applications*, *7*.

[7] Cramer, E., and Kamps, U. (1996).Sequential order statistics and k-out-of-n systems with sequentially adjusted failure rates. *Annals of the Institute of Statistical Mathematics*, *48*, 535-549.

[8] Cramer, E., Kamps, U. (2001). Estimation with Sequential Order Statistics from Exponential Distributions. *Annals of the Institute of Statistical Mathematics* 53, 307–324.

[9] Demiray, D., & Kızılaslan, F. (2022). Stress–strength reliability estimation of a consecutive k-out-of-n system based on proportional hazard rate family. *Journal of Statistical Computation and Simulation*, *92*(1), 159-190.

[10] Glory Prasanth, K., & A. Venmani. (2023). (1, 4) and (2, 4) systems based on sequential order statistics for equivariant parameters estimation. *Utilitas Mathematica*, *120*, pp. 12–24.

[11] Hong, Y., & Meeker, W. Q. (2014). Confidence interval procedures for system reliability and applications to competing risks models. *Lifetime data analysis*, *20*(2), 161-184.

[12] Kalaivani, M., & Kannan, R. (2022). Estimation of reliability function and mean time to system failure for k-out-of-n systems using Weibull failure time model. *International Journal*

of System Assurance Engineering and Management, *13*(5), 2195-2207.

[13] Kamps, U. (1995). A concept of generalized order statistics. Teubner, Stuttgart.

[14] Katzur, A., Kamps, U. (2020). Classification using sequential order statistics. *Advances in Data Analysis and Classification*, 14, 201–230.

[15] Meeker, W. Q., Escobar, L. A., & Pascual, F. G. (2022). *Statistical methods for reliability data*. John Wiley & Sons.

[16] Méndez-González, L. C., Rodríguez-Picón, L. A., Valles-Rosales, D. J., Alvarado Iniesta, A., & Carreón, A. E. Q. (2019). Reliability analysis using exponentiated Weibull distribution and inverse power law. *Quality and Reliability Engineering International*, *35*(4), 1219-1230.

[17] Pesch, T., Polpo, A., Cripps, E., & Cramer, E. (2023). Reliability inference with extended sequential order statistics. *Applied Stochastic Models in Business and Industry*, *39*(4), 520-535.

[18] Pham, H. (2010).On the estimation of reliability of k-out-of-n systems. *International Journal of Systems Assurance Engineering and Management*, 1, 32-35.

[19] Revathy, S.A., and Chandrasekar, B. (2007). Equivariant Estimation of Parameters based on sequential order statistics from (1, 3) and (2,3) Systems, *Communication in Statistics-Theory and it Methods*, 36(3), 541-548.

[20] Shi, Y.M.,Gu, X., & Sun, Y.D.(2011),Reliability evaluation for m-consecutive-k-out-of-n:F system with Burr XII components. In *2011 International Conference on Multimedia Technology* (pp.2314-2317).IEEE.

RT&A, No 4(80)
Volume 19, December, 2024

PREVENTIVE MAINTENANCE POLICIES WITH RELIABILITY THRESHOLDS FOR TABLE SAW MACHINE

[1]Udoh, Nse; [2]Etim, Andrew and [3]Uko, Iniobong

[1&3]Department of Statistics, University of Uyo, Nigeria; [2]Department of Statistics, Akwa Ibom State University, Nigeria
nsesudoh@uniuyo.edu.ng, andrewetim@aksu.edu.ng, iniobonguko144@gmail.com

Abstract

Preventive maintenance policies are essential practical guide for effective maintenance of industrial machines. In this study, system reliability is estimated and used as the condition variable on reliability-based preventive maintenance models to formulate preventive maintenance policies for Table Saw machine which has an increasing hazard rate. Inventory holding cost is introduced as part of the repair cost to complement the actual cost of maintenance. The inter-failure times of the machine was modeled as Weibull distribution and the shape parameters estimate were obtained. Three preventive maintenance policies were obtained for the machine from respective preventive maintenance models with predetermined fixed level of reliability, variable reliability and a combination of both. Result from the third policy with critical reliability level which combines both fixed and unfixed reliability levels is noted as the optimal preventive maintenance policy for the machine in terms of extended lifespan and minimum maintenance cost.

Keywords: Reliability threshold, Preventive Maintenance, Weibull distribution, Table Saw machine, failure distribution.

I. Introduction

Maintenance encompasses a range of activities carried out on facilities or equipment to either restore them to good working condition or to ensure they remain in an acceptable working state. It includes technical procedures aimed at achieving/maintaining satisfactory operation of machines or parts. It serves the primary purpose of ensuring that equipment and facilities can perform their designated tasks as scheduled and under specified conditions. It also helps in preventing unexpected failures and disruptions in operations. Maintenance is vital for the operational efficiency and reliability of facilities and equipment. Neglecting maintenance can lead to increased downtime, decreased performance, and higher repair costs, among others. Therefore, it is essential to prioritize maintenance in the overall strategy of any facility or system, noting that industrial facilities and equipment could be preventively or correctively maintained. *Preventive Maintenance* is proactive and aims to prevent breakdowns and failures. It involves scheduled inspections, repairs, and replacements to keep equipment in good condition while *Corrective Maintenance (Repair)* is reactive and focuses on fixing equipment after a failure or breakdown has occurred. Its goal is to restore equipment to working condition as quickly as possible. Maintenance can be classified as *Perfect Maintenance (As-Good-as-New)* which suggests that after maintenance, equipment is restored to a state identical to when it was new. In reality, achieving this level of restoration is either impracticable

or costly. *Imperfect Maintenance* which includes *Better-than-Old or Worse-than-New* acknowledges that maintenance efforts generally result in equipment being in a better condition than it was before the maintenance but not as good as when it was brand new, *same-as-old* ensures that equipment are restored to performance level before maintenance. While *Worse-than-Old* implies that maintenance does not improve the equipment's condition, but left it in a rather worse state than it was before maintenance. Hence, maintenance falls between perfect maintenance and Worse-than-Old, [1]. This is because the failure nature of repairable systems depends on the repair history of the system, [2]. In summary, maintenance is a critical aspect of facility and equipment management that can significantly impact the overall performance and longevity of assets. The choice between preventive and corrective maintenance strategies depends on the specific needs and goals of a facility. Accordingly, [3] proposed a reliability assessment approach based on multi-deterioration measurement and failure analysis for effective maintenance while [4] proposed a reliability centered predictive maintenance policy using reliability threshold.

PM models are utilized to obtain two kinds of maintenance policies according to their maintenance criteria; the time-dependent PM policy, which determines a PM schedule based on the system age and as concepts of minimal repair and imperfect maintenance; [5] [6] and [7]. For instance, [8] constructed a PM model and applied to selective PM manufacturing system to ensure reliability and minimize the total cost of maintenance and failure losses. While the applicability of the competitive failure model with multiple shock types in the degradation process based on threshold variation was undertaken by [9]. In [10], the lifetime-reward-maximizing maintenance policies under perfect and imperfect maintenance conditions was analyzed and the tradeoff between the system's virtual age and the decision maker's reward rate was investigated. The concept of age reduction factor to formulate imperfect PM policies has been widely used; see, [8], [12]and [13]. Accordingly, [14] investigated the failure data of a marine diesel engine to estimate the reliability of the cylinder liner which was then utilized to develop a reliability-based PM strategy for the diesel engine. The use of estimated reliability as the condition variable to develop three reliability-based PM models with consideration of different scenarios which can assist in evaluating the maintenance cost for each scenario was undertaken by [13]. The proposed approach provides optimal reliability thresholds and PM schedules in advance by which the system availability and quality can be ensured and the organizational resources can be well prepared and managed. In [15], a sequential preventive maintenance was obtained for 8hp-pml gold engine cassava grinding machine with a two parameter Weibull failure distribution. The resulting PM and replacement plan provide effective maintenance schedule that guarantees optimum performance of the machine at specified cost levels., Furthermore, [16] formulated a geometric imperfect preventive maintenance and replacement (GIPMAR) model for aging repairable systems due to age and prolong usage that would meet users need in three phases: within average life span, beyond average life span and beyond initial replacement age of system. The work extended the PM model of [17] to provide PM/replacement schedules for aging repairable systems which was not provided for in earlier models. In another development, [18] proposed a knowledge-based framework that exploits fuzzy logic to generate precise cost implication decisions from an optimal maintenance and replacement schedule using data from a locally fabricated 8HP-PML Gold engine cassava grinding machine whose failure distribution followed the Weibull distribution function, while in the same year, [19] relaxed the assumption of an information-symmetric system, where both the manufacturer's expected profit and the system's expected profit are maximize. Also, [20] studied the stress–strength reliability of a failure profile in which the components of the system are affected by the internal environmental factors and their effect under various scenarios.

Following from [9] and [13], this work seeks to formulate PM and replacement policies for mechanically repairable systems with increasing hazard rate considering three cases with a view to determining an optimal policy that minimizes total cost and extended life cycle of the machine.

II. Methods

2.1 Notations and meaning

θ - shape parameter of Hazard intensity function, $\theta > 0$ of the Weibull distribution
λ - Deteriorating parameter of hazard intensity function, where $\lambda > 0$ of the Weibull distribution
α_j - Age reduction factor, where $\alpha \in (0,1)$
$R(t)$- Reliability function without PM at time, t
y_j - PM interval between successive PM actions, where $j = 1, 2, \dots, N-1$
Z_j - Effective age of the system after jth PM action, $j = 1, 2, \dots, N$
T_w - System lifetime until replacement, where $T_w = \sum_{j=1}^{N} y_j$
R_T^* - The optimal reliability threshold for performing the jth PM action in policy 2
N - The number of scheduled PM actions for models 1, 2 and 3
N_r - Number of minimal repair actions until system replacement
C_r - Cost of minimal repair action
C_h - Cost of holding spare parts (inventory cost)
C_{r^*} - Joint cost for minimal repair action and holding cost of spare part
C_m - Cost of preventive maintenance action
C_i - Cost of system replacement

2.2 Assumptions of the model

- The machine fails randomly
- Failure occurs at the end of time, t given the lifetime distribution of the machine, f(t).
- Spare parts of the machine are readily available in the warehouse with holding cost, C_h
- Failure process is an increasing failure rate
- The required time for PM activities and minimal repairs is negligible
- The cost parameters C_i, C_r, C_m are constants
- PM activities restore the system to "better-than-old" state.
-

2.3 Choice of failure distribution function

There are several probability distributions for modeling the failure rate of repairable systems depending on the failure mechanism. Theoretical consideration is given to both the probabilistic arguments of the failure mode and failure mechanism and practically, the success of modeling empirical failure data through Goodness-of -fit test.

2.3.1 Goodness- of- fit test

There are various tests for assessing the goodness- of- fit of a probability distribution to sample data. These include Chi-square test, kolmogorov Smirnov test, Anderson-Darling test, Cramer-Von Mises test and Mann-Scheuer-Fertig test, etc. This study uses Easyfit software to perform the Chi-squared Goodness-of-fit test because of its simplicity and adaptability. The Easyfit software result in Table 2.1 shows that the 2-parameter Weibull distribution with rank 1 is the best-fit model for the data set.

Table 2.1: Extract of Goodness-of-fit test of Probability Distributions of Interest

Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-squared	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Weibull	0.10264	2	0.87318	1	0.62162	1
Gen. Pareto	0.11014	1	0.87596	2	0.7244	2
Uniform	0.11057	3	0.90513	3	N/A	
Levy (2P)	0.11093	4	0.9133	4	0.9851	3

2.4 The two-parameter Weibull distribution

2.4.1 The density function

The probability distribution function of a two-parameter Weibull distribution is given by;

$$f(z;\theta,\lambda) = \frac{\lambda}{\theta}\left(\frac{z}{\theta}\right)^{\lambda-1} e^{-\left(\frac{z}{\theta}\right)^{\lambda}},\ \lambda > 0,\ \theta > 0,\ Z > 0 \quad (1)$$

Whereθ = scale parameter, λ =shape parameter

2.4.2 Reliability function, $R(t)$

The reliability function of the two-parameter Weibull distribution is given by;

$$R(t) = -P(T \le t) = P(T > t) = 1 - \int_0^t f(t)dt \quad (2)$$

Substituting for $f(t)$ in Eq (2), we have;

$$R(t) = 1 - \int_0^t \left(\frac{\lambda}{\theta}\right)\left(\frac{t}{\theta}\right)^{\lambda-1} e^{-\left(\frac{t}{\theta}\right)^{\lambda}} dt = 1 - \left(\frac{\lambda}{\theta}\right)\int_0^t \left(\frac{t}{\theta}\right)^{\lambda-1} e^{-\left(\frac{t}{\theta}\right)^{\lambda}} dt$$

Since

$$F(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^{\lambda}}$$

$$R(t) = 1 - F(t) = 1 - \left(1 - e^{-\left(\frac{t}{\theta}\right)^{\lambda}}\right)$$

$$\therefore R(t) = e^{-\left(\frac{t}{\theta}\right)^{\lambda}} \quad (3)$$

2.4.3 Failure rate, *h(t)* of two-parameter Weibull distribution

The failure rate, *h(t)* during a given interval of time $t = [t_1, t_2]$ shows the probability that a failure per unit time occurs in the interval (t_1, t_2), conditioned on the event that no failure has occurred at or before time, t_1. This means that $T > t_1$. The failure rate can be defined as follows:

$$h(t) = \frac{R(t_1)-R(t_2)}{(t_2-t_1)R(t_1)} = \frac{F(t_2)-F(t_1)}{(t_2-t_1)R(t_1)} \quad (4)$$

Taking the limit of the failure rate at the interval, $(t_1, \Delta t + 1)$ as Δt approaches zero, where $t = t_1$and $(t + \Delta t) = t_2$ gives the hazard function, $h(t)$, as follows;

$$h(t) = \lim_{\Delta t \to 0} \frac{R(t)-R(t+\Delta t)}{(\Delta t)\times R(t)} \quad (5)$$

$$= \lim_{\Delta t \to 0} \frac{F(t+\Delta t)-F(t)}{(\Delta t)} \times \frac{1}{R(t)} \quad (6)$$

But

$$h(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t)-F(t)}{(\Delta t)} = f(t)$$

$$\therefore h(t) = \frac{f(t)}{R(t)}$$

Where

$$f(t) = \left(\frac{\lambda}{\theta}\right)\left(\frac{t}{\theta}\right)^{\lambda-1} e^{-\left(\frac{t}{\theta}\right)^{\lambda}}$$

$$R(t) = e^{-\left(\frac{t}{\theta}\right)^{\lambda}}$$

$$h(t) = \frac{\left(\frac{\lambda}{\theta}\right)\left(\frac{t}{\theta}\right)^{\lambda-1} e^{-\left(\frac{t}{\theta}\right)^{\lambda}}}{e^{-\left(\frac{t}{\theta}\right)^{\lambda}}}$$

$$\therefore h(t) = \lambda\theta^{-\lambda}t^{\lambda-1} \quad (7)$$

And the cumulative hazard rate is given as;

$$H(t) = \int_0^t h(t)dt = \int_0^t \lambda\theta^{-\lambda}t^{\lambda-1}dt = \theta^{-\lambda}t^{\lambda} \quad (8)$$

2.4.4 Estimating the parameters of Weibull distribution

Easyfit software will be used to determine the appropriate failure function from the observed data and also obtain the parameters estimates of the fit distribution.

2.4.5 Evaluation of the Weibull parameters

As part of the preliminary analysis, the scale and shape parameters estimate of the Weibull distribution were also obtained for the failure data of Table Saw machine with the aid of the Easyfit software as; $\lambda = 3.8548$ and $\theta = 202.92$.

2.5 Preventive maintenance policies

We seek to modify and adapt the models in [13] to obtain suitable maintenance policies for mechanically repairable systems with increasing hazard rate using Table Saw machine as case study.

2.5.1 Policy 1: System undergoes PM activity whenever the reliability reaches the predetermined threshold

Let the effective age of the system just before the jth PM activity be given as;

$$Z_j = y_j + \alpha_{j-1}Z_{j-1};\ j = 1, 2, \dots, N \quad (9)$$

Where,

Z_j is the effective age of the system

y_j is the interval between successive PM activities

α_{j-1} is the age reduction factor; $0 < \alpha_0 < \alpha_1 < \cdots < \alpha_N$

$\alpha_j Z_j$ denotes the effective age of the system immediately after the jth PM activity.

Let the associated expected cost rate per unit time for performing PM activities according to [13] be given as;

$$C(N) = \frac{1}{T_w}[C_i + C_r N_r + (N-1)C_m] \quad (10)$$

$$T_w = \sum_{j=1}^{N} y_j = \sum_{j=1}^{N} Z_j - \sum_{j=1}^{N} \alpha_{j-1}Z_{j-1} = \left(Z_N + \sum_{j=1}^{N-1} Z_j\right) - \sum_{j=1}^{N-1} \alpha_j Z_j = Z_N + \sum_{j=1}^{N-1}(1-\alpha_j)Z_j$$

and $$N_r = \sum_{j=1}^{N} \int_{\alpha_{j-1}Z_{j-1}}^{Z} h(t) = \sum_{j=1}^{N} [H(t)]_{\alpha_{j-1}Z_{j-1}}^{Z_j}$$

Therefore,

$$C(N) = \frac{C_i + (N-1)C_m + C_r \sum_{j=1}^{N} \int_{\alpha_{j-1}Z_{j-1}}^{Z_j} h(t)}{Z_N + \sum_{j=1}^{N-1}(1-\alpha_j)Z_j} \quad (11)$$

Let the reliability threshold at the end of each PM cycle be:

$$R_T = R(\alpha_{j-1}Z_{j-1})R_m\left(Z_j \middle| Z_{j-1}\right)$$

This is on the assumption that the system age can be proportionally reduced by imperfect PM actions to $\alpha_{j-1}Z_{j-1}$immediately after the *(j-1)*th PM activity, [21]. $R_m\left(Z_j\middle|Z_{j-1}\right)$ is the reliability of *jth* PM cycle at age, t given that the system was maintained at Z_{j-1}. This is equivalent to the product of probability of survival until $\alpha_{j-1}Z_{j-1}$.
The optimization problem is to minimize:

$$C(N) = \frac{C_i+(N-1)C_m+C_r\sum_{j=1}^{N}\int_{\alpha_{j-1}Z_{j-1}}^{Z_j} h(t)}{Z_N+\sum_{j=1}^{N-1}(1-\alpha_j)Z_j} \tag{12}$$

Subject to: $$R\left(\alpha_{j-1}Z_{j-1}\right)R_m\left(Z_j\middle|Z_{j-1}\right).$$

According to [22], the reliability function of an initial system is given as; $R(t) = e^{-\int_0^t h(t)dt}$.
The failure density, $f(t)$ and the reliability function, $R(t)$ can be derived from the knowledge of the hazard function, $h(t)$. Then, $R(t) = exp\left[-\int_0^t h(t)dt\right]$ and $f(t) = h(t) \times R(t)$. Since $R(t) = 1 - F(t)$, $R'(t) = -F'(t)$ (by differentiating both sides). Therefore,

$$h(t) = \frac{f(t)}{R(t)} = \frac{F'(t)}{R(t)} = \frac{-R'(t)}{R(t)}$$

Integrating both sides, we have;
$\int_0^t h(t)dt = -\int_0^t \frac{R'(t)}{R(t)}dt = -[ln\,R\,(t) - ln\,R\,(0)]$, under the boundary condition, $R(0) = 1$, since the component will not fail before time $t = 0$, once it is put into operation.
Since $\ln R(0) = \ln 1 = 0$, we see that $-\int_0^t h(t)dt = ln\,R\,(t) \Rightarrow \int_0^t h(t)dt = ln\,R\,(t)$. Taking exponent of both sides,

$$e^{-\int_0^t h(t)dt} = e^{ln\,R(t)} \Rightarrow e^{-\int_0^t h(t)dt} = R(t)$$

Since the cumulative hazard intensity at time, t is given as $H(t) = \int_0^t h(t)dt$, we can say that

$$R(t) = R_T = e^{-H(Z_j)} \tag{13}$$

Substitute Eq (8) in (13), we have; $R_T = e^{-\theta^{-\lambda}Z^{\lambda}}$
Taking the ln of both sides, we have,

$$ln(R_T) = -\theta^{-\lambda}Z^{-\lambda} \implies \left[\frac{-ln(R_T)}{\theta^{-\lambda}}\right]^{1/\lambda} = \theta[ln(R_T)]^{1/\lambda} \tag{14}$$

Considering the jth PM action,

$$Z_j = y_j + \alpha_j Z_j \Rightarrow y_j = Z_j - \alpha_j Z_j \Rightarrow y_j = Z_j\left(1-\alpha_j\right) \tag{15}$$

2.5.2 Policy 2: System with non-fixed Reliability Threshold considered as a decision variable
This policy states that a reliability threshold is not predetermined but considered as a decision variable with additional inventory holding cost
Let R be the reliability threshold which is a decision variable, then the cost function is;

$$C(N,R) = \frac{C_i+(N-1)C_m+C_r\sum_{j=1}^{N}\int_{\alpha_{j-1}Z_{j-1}}^{Z_j} h(t)}{\sum_{j=1}^{N} y_j} \tag{16}$$

We introduce inventory cost, C_h, which is the cost of holding spare parts of the machine readily

available in the warehouse. This will reduce the time of ordering spare part which increases downtime and reduces the operational time of the machine, hence, productivity. Therefore, our new cost of minimal repair is $C_{r^*} = C_r + C_h$. Eq (16) becomes;

$$C(N,R) = \frac{C_i + (N-1)C_m + C_{r^*}\sum_{j=1}^{N}\int_{\alpha_{j-1}Z_{j-1}}^{Z_j} h(t)}{\sum_{j=1}^{N} y_j} = \frac{C_i + (N-1)C_m + C_{r^*}\sum_{j=1}^{N}\left[H(Z_j) - H(\alpha_{j-1}Z_{j-1})\right]}{Z_N + \sum_{j=1}^{N-1}(1-\alpha_j)Z_j} \quad (17)$$

The term under summation sign in the numerator of (Eq 17) can be written as follows;

$$= \theta^{-\lambda}\sum_{j=1}^{N}\left[Z_j^{\lambda} - \left(\alpha_{j-1}^{\lambda}Z_{j-1}^{\lambda}\right)\right] = \theta^{-\lambda}\left[\sum_{j=1}^{N} Z_j^{\lambda} - \sum_{j=1}^{N}\left(\alpha_{j-1}^{\lambda}Z_{j-1}^{\lambda}\right)\right] = \theta^{-\lambda}\left[\left(Z_N^{\lambda} + \sum_{j=1}^{N-1} Z_j^{\lambda}\right) - \sum_{j=1}^{N-1}\left(\alpha_j^{\lambda}Z_j^{\lambda}\right)\right]$$

$$= \theta^{-\lambda}\left[\left(Z_N^{\lambda} + \sum_{j=1}^{N-1} Z_j^{\lambda}\right) - \sum_{j=1}^{N-1}\left(\alpha_j^{\lambda}Z_j^{\lambda}\right)\right] = \theta^{-\lambda}Z_N^{\lambda} + \sum_{j=1}^{N-1}\theta^{-\lambda}Z_j^{\lambda} - \theta^{-\lambda}\sum_{j=1}^{N-1}\left(\alpha_j^{\lambda}Z_{lj}^{\lambda}\right)$$

$$= \theta^{-\lambda}Z_N^{\lambda} + \sum_{j=1}^{N-1} H(Z_j) - \theta^{-\lambda}\sum_{j=1}^{N-1}\left(\alpha_j^{\lambda}Z_j^{\lambda}\right)$$

Since the effective age, Z_j at the replacement point, $Z_N = 0$, then, Z_N vanishes in the numerator and the denominator. Hence, Eq (17), becomes;

$$C(N,R) = \frac{C_i + (N-1)C_m + C_{r^*}\left\{\sum_{j=1}^{N-1}\left[-ln(R_T) + \theta^{-\lambda}\alpha_j^{\lambda}Z_j^{\lambda}\right]\right\}}{\sum_{j=1}^{N-1}(1-\alpha_j)Z_j} = \frac{C_i + (N-1)C_m + C_{r^*}\left\{\sum_{j=1}^{N-1}\left[-ln(R_T) + \alpha_j^{\lambda}H(Z_j)\right]\right\}}{\sum_{j=1}^{N-1}(1-\alpha_j)Z_j}$$

$$= \frac{C_i + (N-1)C_m + C_{r^*}\left\{\sum_{j=1}^{N-1}\left[-ln(R_T) + \alpha_j^{\lambda}\,ln(R_T)\right]\right\}}{\sum_{j=1}^{N-1}(1-\alpha_j)Z_j} = \frac{C_i + (N-1)C_m - C_{r^*}\,ln(R_T)\left[\sum_{j=1}^{N-1}(1-\alpha_j^{\lambda})\right]}{\sum_{j=1}^{N-1}(1-\alpha_j)Z_j}$$

$$C(N,R) = \frac{C_i + (N-1)C_m - C_{r^*}\,ln(R_T)\left[\sum_{j=1}^{N-1}(1-\alpha_j^{\lambda})\right]}{\left[\frac{-ln(R_T)}{\theta^{-\lambda}}\right]^{1/\lambda}\sum_{j=1}^{N-1}(1-\alpha_j)} \quad (18)$$

Differentiating Eq (18) w.r.t. R_T, equating it to zero and writing R_T with respect to other terms, we have;

$$R_T^{\bullet} = exp\left\{\frac{C_i + (N-1)C_m}{(1-\lambda)C_{r^*}\sum_{j=1}^{N-1}(1-\alpha_j^{\lambda})}\right\} \quad (19)$$

2.5.3 Policy 3: System with optimal combination of reliability threshold in policies 1 and 2
The aim of this policy is to combine the PM policy with fixed reliability threshold, R_N and unfixed reliability thresholds, R_j with a view to obtaining the total minimum cost per unit time compared to the previous two models
By substituting Eq (14) in Eq (16), we have;

$$(C, R_N, R_j) = \frac{C_i + (N-1)C_m + C_{r^*}\left\{\sum_{j=1}^{N}\left[-ln(R_T) - \alpha_j^{\lambda}H(Z_j)\right]\right\}}{Z_N + \sum_{j=1}^{N-1}(1-\alpha_j)Z_j} = \frac{C_i + (N-1)C_m + C_{r^*}\left\{\sum_{j=1}^{N}\left[-ln(R_T) - \alpha_j^{\lambda}\,ln(R_T)\right]\right\}}{Z_N + \sum_{j=1}^{N-1}(1-\alpha_j)Z_j}$$

$$= \frac{C_i+(N-1)C_m-C_{r^*}\sum_{j=1}^{N}\left(1-\alpha_j^{\lambda}\right)ln(R_T)}{Z_N+\sum_{j=1}^{N-1}(1-\alpha_j)Z_j} \tag{20}$$

Since $$\sum_{j=1}^{N}\left(1-\alpha_j^{\lambda}\right)ln(R_T) = \left[ln(R_N) + \sum_{j=1}^{N-1}\left(1-\alpha_j^{\lambda}\right)ln(R_T)\right]$$

and $$Z_j = \left[\frac{-ln(R_T)}{\theta^{-\lambda}}\right]^{1/\lambda} = \theta[-ln(R_T)]^{1/\lambda}$$

Eq (20) will now become;

$$C\left(N, R_N, R_j\right) = \frac{C_i + (N-1)C_m - C_{r^*}\left[ln(R_N) + \sum_{j=1}^{N-1}\left(1-\alpha_j^{\lambda}\right)ln(R_T)\right]}{Z_N + \sum_{j=1}^{N-1}(1-\alpha_j)\left[\frac{-ln\left(R_j\right)}{\theta^{-\lambda}}\right]^{1/\lambda}}$$

$$= \frac{C_i+(N-1)C_m-C_{r^*}\left[ln(R_N)+\sum_{j=1}^{N-1}\left(1-\alpha_j^{\lambda}\right)ln(R_j)\right]}{\left[\frac{-ln(R_N)}{\theta^{-\lambda}}\right]^{1/\lambda}+\sum_{j=1}^{N-1}(1-\alpha_j)\left[\frac{-ln\left(R_j\right)}{\theta^{-\lambda}}\right]^{1/\lambda}} \tag{21}$$

Differentiating Eq (21) w.r.t. R_j, equating it to zero and making C the subject, we have;

$$C = \frac{\lambda C_{r^*}\sum_{j=1}^{N-1}\left(1-\alpha_j^{\lambda}\right)}{\sum_{j=1}^{N-1}(1-\alpha_j)\left[\frac{-ln\left(R_j\right)}{\theta^{-\lambda}}\right]^{1/\lambda-1}} = \frac{\lambda C_{r^*}\left(1-\alpha_j^{\lambda}\right)}{\theta(1-\alpha_j)[-ln(R_j)]^{1/\lambda-1}} \tag{22}$$

Similarly, differentiating Eq (21) w.r.t. R_N and making C the subject, we have;

$$C = \frac{\lambda C_{r^*}}{\theta[-ln(R_N)]^{1/\lambda-1}} \tag{23}$$

Equating Eqs (22) and (23), we have;

$$\frac{\lambda C_{r^*}\left(1-\alpha_j^{\lambda}\right)}{\theta\left(1-\alpha_j\right)\left[-ln\left(R_j\right)\right]^{1/\lambda-1}} = \frac{\lambda C_{r^*}}{\theta[-ln(R_N)]^{1/\lambda-1}}$$

$$[-ln(R_N)]^{1/\lambda-1}\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right] = \left[-ln\left(R_j\right)\right]^{1/\lambda-1}$$

$$-ln(R_N)\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}} = -ln\left(R_j\right)$$

$$ln\left(R_j\right) = ln(R_N)\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}} \tag{24}$$

$$\therefore R_j^* = R_N\, exp\left\{\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}}\right\} \tag{25}$$

From Eq (21), we have;

$$C\left\{\left[\frac{-ln(R_N)}{\theta^{-\lambda}}\right]^{1/\lambda} + \sum_{j=1}^{N-1}\left(1-\alpha_j\right)\left[\frac{-ln(R_j)}{\theta^{-\lambda}}\right]^{1/\lambda}\right\} = C_i + (N-1)C_m - C_{r^*}\left[ln(R_N) + \sum_{j=1}^{N-1}\left(1-\alpha_j^{\lambda}\right)ln\left(R_j\right)\right]$$

Note that: $$\left[-ln\left(R_j\right)\right]^{1/\lambda-1} = [-ln(R_N)]^{1/\lambda-1}\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]$$

$$\left[-ln\left(R_j\right)\right]^{1/\lambda} = [-ln(R_N)]^{1/\lambda-1}\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right] \times \left[-ln\left(R_j\right)\right]$$

Substituting for $ln\, R_j$, we have;

$$\left[-ln(R_j)\right]^{1/\lambda} = \left[-ln(R_N)\right]^{1/\lambda - 1}\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right] \times \left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}} \left[-ln(R_N)\right]$$

$$\left[-ln(R_j)\right]^{1/\lambda} = \left[-ln(R_N)\right]^{1/\lambda}\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{1}{1-\lambda}} \tag{26}$$

Substituting Eqs (22), (23) and (24) in (26) where necessary, we have;

$$\frac{\lambda C_{r^*}}{\theta\left[-ln(R_N)\right]^{1/\lambda - 1}}\left\{\left[\frac{-ln(R_N)}{\theta^{-\lambda}}\right]^{1/\lambda} + \sum_{j=1}^{N-1}(1-\alpha_j)\left[\frac{-ln(R_N)}{\theta^{-\lambda}}\right]^{1/\lambda}\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{1}{1-\lambda}}\right\} =$$

$$C_i + (N-1)C_m - C_{r^*}\left[ln(R_N) + ln(R_N)\sum_{j=1}^{N-1}(1-\alpha_j^{\lambda})\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}}\right]$$

$$\Rightarrow C_i + (N-1)C_m - C_{r^*}ln(R_N)\left[1 + \sum_{j=1}^{N-1}(1-\alpha_j^{\lambda})\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}}\right]$$

$$+\lambda C_{r^*}\left[1 + \sum_{j=1}^{N-1}(1-\alpha_j)\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{1}{1-\lambda}}\right]\left[-ln(R_N)\right] = 0$$

$$C_i + (N-1)C_m - C_{r^*}ln(R_N^*)\left\{1 + \sum_{j=0}^{N-1}\left[\frac{(1-\alpha_j)^{\lambda}}{1-\alpha_j^{\lambda}}\right]^{\frac{1}{\lambda-1}} - \lambda\left[1 + \sum_{j=0}^{N-1}\left[\frac{1-\alpha_j^{\lambda}}{(1-\alpha_j)^{\lambda}}\right]^{\frac{1}{1-\lambda}}\right]\right\} = 0$$

$$C_i + (N-1)C_m - C_{r^*}ln(R_N^*)\left\{(1-\lambda)\sum_{j=0}^{N-1}\left[\frac{(1-\alpha_j)^{\lambda}}{1-\alpha_j^{\lambda}}\right]^{\frac{1}{\lambda-1}}\right\} = 0$$

$$ln(R_N^*) = \frac{C_i + (N-1)C_m}{C_{r^*}(1-\lambda)\sum_{j=0}^{N-1}\left[\frac{(1-\alpha_j)^{\lambda}}{1-\alpha_j^{\lambda}}\right]^{\frac{1}{\lambda-1}}}$$

$$\therefore R_N^* = exp\left\{\frac{C_i + (N-1)C_m}{C_{r^*}(1-\lambda)\sum_{j=0}^{N-1}\left[\frac{(1-\alpha_j)^{\lambda}}{1-\alpha_j^{\lambda}}\right]^{\frac{1}{\lambda-1}}}\right\} \tag{27}$$

Substituting Eq (27) in Eq (25), we have;

$$R_j^* = exp\left\{\frac{C_i + (N-1)C_m}{C_{r^*}(1-\lambda)\sum_{j=0}^{N-1}\left[\frac{(1-\alpha_j)^{\lambda}}{1-\alpha_j^{\lambda}}\right]^{\frac{1}{\lambda-1}}}\right\} \times exp\left\{\left[\frac{1-\alpha_j^{\lambda}}{1-\alpha_j}\right]^{\frac{\lambda}{1-\lambda}}\right\} \tag{28}$$

III. Results

Table 3.1: Values of y_j with fixed reliability threshold

N	α_j	y_j	R_t
1	0.3333	0.8281	0.9000
2	0.4000	0.4047	0.9000
3	0.4286	0.3101	0.9000
4	0.4444	0.2852	0.9000
5	0.4545	0.2583	0.9000
6	0.4615	0.2189	0.9000
7	0.4667	0.1943	0.9000
8	0.4706	0.1744	0.9000
9	0.4737	0.1626	0.9000
10	0.4762	0.1610	0.9000
11	0.4783	0.1587	0.9000
12	0.4800	0.1516	0.9000
13	0.4815	0.1511	0.9000
14	0.4828	0.1508	0.9000
15	0.4839	0.1491	0.9000
16	0.4848	0.1489	0.9000
17	0.4857	0.1484	0.9000
18	0.4865	0.1475	0.9000
		4.2035	

Table 3.2: Values of y_j with variable reliability threshold

N	$1-\alpha_j^{\lambda}$	y_j	R_t
1	0.5943	0.5340	0.8586
2	0.5287	0.4897	0.8979
3	0.5013	0.4824	0.8970
4	0.4862	0.4048	0.8985
5	0.4766	0.3426	0.8947
6	0.4700	0.3199	0.8783
7	0.4652	0.3044	0.8662
8	0.4615	0.2902	0.8569
9	0.4586	0.2689	0.8496
10	0.4562	0.2376	0.8437
11	0.4543	0.2220	0.8389
12	0.4526	0.2131	0.8368
13	0.4513	0.1923	0.8320
14	0.4501	0.1901	0.8312
15	0.4490	0.1827	***0.8307***
16	0.4481	0.1792	0.8992
17	0.4473	0.1749	0.8989
18	0.4466	0.1721	0.8987
		5.1920	

Table 3.3: Values of y_j under the combined policies

S/N	$\frac{1-\alpha)^{\lambda}}{1-\alpha^{\lambda}}^{\frac{1}{\lambda-1}}$	R_n	$\frac{1-\lambda}{1-\alpha^{\lambda}}^{\frac{\lambda}{\lambda-1}}$	R_j	y_j
1	0.3507	0.1650	5.4455	0.8987	0.5340
2	0.2960	0.1505	5.9674	0.8981	0.4816
3	0.2748	0.1448	6.1969	0.8975	0.4833
4	0.2636	0.1418	6.3260	0.8970	0.4057
5	0.2566	0.1399	6.4088	0.8963	0.3436
6	0.2518	0.1385	6.4664	0.8957	0.3208
7	0.2483	0.1381	6.5088	0.8988	0.3053
8	0.2457	0.1364	6.5413	0.8923	0.2912
9	0.2437	0.1348	6.5671	0.8850	0.2698
10	0.2420	0.1334	6.5879	0.8791	0.2385
11	0.2406	0.1324	6.6052	0.8743	0.2229
12	0.2395	0.1320	6.6197	0.8735	0.2140
13	0.2385	0.1311	6.6321	***0.8695***	0.1932
14	0.2377	0.1354	6.6428	0.8997	0.1910
15	0.2370	0.1351	6.6521	0.8985	0.1836
16	0.2363	0.1348	6.6603	0.8980	0.1801
17	0.2358	0.1346	6.6675	0.8976	0.1758
18	0.2353	0.1344	6.6740	0.8972	0.1730
					5.2077

Table 3.4: Optimal solution of system performance under policies 1, 2 and 3

System performance	Policy 1	Policy 2	Policy 3
System Lifespan	4.2035	5.192	5.2077
Minimum reliability	0.9000	0.8310	0.8700
Maintenance cost	6.3248	6.1146	5.5316
Number of PM actions	18	18	18

3.1 Application of PM policies with reliability threshold to the maintenance of Table Saw machine
The distribution of the inter failure times of the Table Saw machine was modeled as a two parameter Weibull distribution as given in Eq (1), the shape and scale parameters of the distribution were obtained in section 2.4.5 as $\theta = 202.92$; and $\lambda = 3.8548$, respectively. The results of policies 1, 2 and 3 are contained in Tables 3.1, 3.2 and 3.3. The y_j values are the length of operating time of the machine before PM or replacement maintenance (as the case may be) at respective policy.

3.2 Expected maintenance cost per unit time for policies 1, 2 and 3

The different maintenance costs ratios were obtained as; $\frac{C_{r^*}}{C_m} = 8.0$ and $\frac{C_i}{C_m} = 2.8$. Eqs (18) was used to obtain the expected maintenance cost for each policy as; Policy 1: $C(N) = 6.3248$, Policy 2**:** $(N) = 6.1146$, Policy 3 **:** $C(N) = 5.5316$

IV. Discussion

I. Policy 1: When the Table Saw machine has a fixed reliability threshold

The fixed reliability threshold for proper functioning of the machine was set at 0.9000. Eqs (9) and (14) were used to obtain the operating time (life span) of the machine, y_j before next PM activity. The age improvement factor, $\alpha_j = \frac{j}{(2j+1)}$, ([23] and [8]), increases marginally with the frequency of PM. The values of α_j, y_j, and R_t are shown in Table 3.1. The bold last value of y_j in column 3 is the total lifespan of the machine over 18 PM cycles equal 4.2035 years. The column of R_t is the fixed reliability threshold for all maintenance actions. It took 0.8281unit of the operating time before the first PM activity, 0.4047 unit of working time before the next PM, and so on. Generally, it is observed that the lifespan of the machine keeps decreasing in spite of the fact that the fixed reliability limit is not violated. Hence, maintenance engineers need not rely on this policy.

II. Policy 2: When the Table Saw machine has an unfixed reliability threshold

Eqs (14), (15) and (19) were used to obtained the lifespan of the machine, y_j after each PM and the corresponding reliability index, R(t) which serves as a check for replacement. The total lifespan of the machine over 18 PM cycles equal 5.1920 years. The minimum reliability threshold set at $R_T = 0.8310$ is always set a value below the fixed reliability (0.9000) used in policy 1, see [13]. The y_j column which is the operational time before PM, shows a gradual decrease of PM intervals from 0.5340 unit time in the first cycle of operation before next PM to 0.4897 unit of operational time before the 2nd PM to 0.4824 units of operational time before the third PM activity and so on. The 15th PM calls for replacement maintenance of the failed component of the machine because its reliability value is below the threshold (0.8310) of the machine. After the replacement maintenance, it is observed that the values of the machine's reliability increase to within the tolerance level again illustrating the import of PM actions.

III. Policy 3: A Combination of policies 1 and 2 for the Table Saw machine

Eqs (9) and (28) were used to obtain the lifespan of the machine, y_j. The minimum reliability threshold, $R_T = 0.8700$ while the lifespan equals 5.2077 years. Eq (27) gives columns 2 and 3, Eq (28) gives columns 4 and 5, while Equations (14) and (15) give the last column, all in Table 3.3. The y_j column has 0.5340 unit of operating time before first PM and 0.4816 unit of operating time before the 2nd PM, and 0.4833 unit of operating time before the 3rd PM activity, and so on. Replacement maintenance is required at the 13th PM because its reliability value is below the threshold (0.8700). It is observed that after the appropriate replacement maintenance action, the critical reliability level rises to 0.8997, even above the initial level at the beginning of operation. This underscores the need for PM and a justification to recommend policy 3 for effective maintenance management.

The lifespan of the machine with regard to Policies 1, 2 and 3 in Table 3.4 are 4.2035, 5.1920, and 5.2077, respectively. The minimum reliability values of these three policies are 0.900, 0.8310 and 0.8700, respectively, and the associated respective cost of maintaining this machine in a lifecycle of 18 PM's are 6.3248, 6.1146 and 5.5316, respectively. These results agree with [22] who showed that PM model based on the unfixed reliability threshold led to a lower expected maintenance cost and longer system lifespan.

Therefore, policy 3 which combines fixed predetermined reliability threshold and the variable threshold values as decision variable is the recommended optimum maintenance policy for the Table Saw machine. It yielded the minimum cost of maintenance as well as the longest lifespan of the machine over the PM cycles under consideration.

References

[1] Doyen L. Repair Efficiency Estimation in the ARI1 Imperfect Repair Model. Modern Statistical and Mathematical Methods in Reliability, Wilson, A. G. (Ed.). World scientific, USA, 2005.

[2] Muhammad M, Majid, A. A, Ibrahim, N. A, (2009). A case study of reliability assessment for centrifugal pumps in a petrochemical plant. Proceeding of the 4th World Congress on Engineering Asset Management, September 28-30, Athens, Greece: 398-404.

[3] Liang, Z., Jun, Y., Haifeng, L., Zhenglian, S. (2011). Reliability Assessment Based on Multivariate Degradation Measures and Competing Failure Analysis. Modern Applied Science, 5(6): 232-236.

[4] Zhou, X., Xi, L. and Lee, J. (2007). Reliability-centered predictive maintenance scheduling for a continuously monitored system subject to degradation, Reliability Engineering & System Safety, 92(4): 530-534.

[5] Pham, H., H. Wang. (1996). Imperfect maintenance. European Journal of Operational Research 94(3): 425-438.

[6] Rausand, M., A. Høyland. System Reliability Theory: Models, Statistical Methods, and Applications, 2nd ed. Wiley, New York, 2004.

[7] Chang, C. C. (2014). Optimum preventive maintenance policies for systems subject to random working times, replacement, and minimal repair. Computers & Industrial Engineering 67: 185-194.

[8] Lin, D., Zuo, M. J. and Yam, R. C. (2001). Sequential imperfect preventive maintenance models with two categories of failure modes. Naval Research Logistics, 48(2):172-183.

[9] Ge, X., Bi, J., Qi, X., Wang, Y., Liuying W. and Liu, Fei. (2023). Reliability Analysis and Preventive Maintenance Strategy for Competitive Failure Process of Wind Turbine Blade with Hard Failure-Caused Threshold Variation. International Journal of Reliability, Quality and Safety Engineering. ISSN (online): 1793-6446.

[10] Khojandi, A., Lisa M. Maillart, Oleg A. Prokopyev, (2014). Optimal planning of life-depleting maintenance activities. IIE Transactions 46(7): 636-652.

[11] Zhu, H., F. Liu, X. Shao, Q. Liu, Y. Deng. (2011). A cost based selective maintenance decision making method for machining line. Quality and Reliability Engineering International 27(2): 191-201.

[12] Lin, Z. L and Huang, Y. S. (2010). Nonperiodic preventive maintenance for repairable systems, Naval Research Logistics (NRL), 57(7): 615-625.

[13] Lin, Z. L., Huang, Y. S. and Fang, C. C. (2014). Non-periodic preventive maintenance with reliability thresholds for complex repairable systems," Reliability Engineering & System Safety, 136:145-156.

[14] Giorgio, M., M. Guida, G. Pulcini. (2007). A wear model for assessing the reliability of cylinder liners in marine diesel engines. IEEE Transactions on Reliability 56(1): 158-166.

[15] Udoh, N. and Ekpenyong, E. (2019). Sequential Imperfect preventive maintenance and replacement schedule for 8hp-pml Gold Engine Cassava Grinding Machine. International Journal of Statistics and Reliability Engineering, 6(1):13-18.

[16] Udoh, N. and Effanga, E. (2023). Geometric imperfect preventive maintenance and replacement (GIPMAR) model for aging repairable systems. International Journal of Quality & Reliability Management 40(2): 566-581.

[17] Lin, D., Zuo, M. J. and Yam, R. C. M. (2000). General sequential imperfect preventive maintenance models. International Journal of Reliability, Quality and Safety Engineering 7(3): 253-266.

[18] Udoh, N & and Ekpenyong, M. (2023). A Knowledge-Based Framework for Cost Implication Modeling of Mechanically Repairable Systems with Imperfect Preventive Maintenance and Replacement Schedule. Journal of Applied Science and Engineering, 26(2): 221-234.

[19] Alexander, A., Li, Y., Plante, E. (2023). Comparative study of two menus of contracts for outsourcing the maintenance function of a process having a linear failure rate. IISE Transactions, 55:12, 1230-1241.

[20] Pakdaman, Z., Shekari, M., and Zamani, H. (2023). Stress–Strength Reliability of a Failure Profile with Components Operating under the Internal Environmental Factors. International Journal of Reliability, Quality and Safety Engineering, 30(06): 2350019.

[21] Martorell, S., Sanchez, A. and Serradell, V. (1999). Age-dependent reliability model considering effects of maintenance and working conditions. Reliability Engineering & System Safety, 64(1):19-31.

[22] Lee, J., Park, J. and Ahn, S. (2018). Cumulative Hazard Intensity Based on Sequential Preventive Maintenance Strategy for Ensuring Minimum Reliability. International Journal of Computing, Communications and Networking 7(2): 1-7.

[23] El-Ferik, S. and Ben-Daya, M. (2006). Age-based hybrid model for imperfect preventive maintenance. IIE Transactions, 38(4): 365-375.

RT&A, No 4(80)
Volume 19, December, 2024

OPTIMIZATION OF THE TWO UNIT SYSTEMS WITH DEGRADATION AND PREVENTIVE MAINTENANCE IN ONE UNIT USING DEEP LEARNING ALGORITHMS

Shakuntla Singla[1], Komalpreet Kaur[2*]

•

[1,2] Department of Mathematics and Humanities, MMEC, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India

[1]shakus25@gmail.com,[2*]komalsandha659@gmail.com[2]

Abstract

This study presents a comprehensive behavioral examination of a two-unit organization integrating preventive maintenance strategies and the introduction of degradation in single unit following complete failure. The research explores the intricate dynamics influencing the system's reliability, availability, and performance. The impact of preventive maintenance on reducing unexpected failures and enhancing overall system robustness is investigated, alongside the added complexity introduced by degradation modeling using three methods ADAM, SGD and RMS Prop. The interplay between preventive maintenance and degradation is analyzed, emphasizing the critical role of optimization in achieving effective system performance. Trade-off analysis reveals the delicate balance between maintenance costs and savings from avoiding failures, guiding decision-makers in determining the most cost-effective strategies. Sensitivity analysis identifies key parameters influencing system behavior, aiding in informed decision-making and robust system design. Consideration of life-cycle costs provides a holistic economic perspective, evaluating both short-term and long-term implications of maintenance and operational choices. This model is train in three methods (ADAM, SGD, and RMS Prop), In MTSFof Adam is better than other two methods. In Expected Number of Inspections by repair man of SGD is better than other two methods. In Recall (Busy Period) of Adam is better than other two methods. In Precision (Availability of the System) of RMS Prop is better than other two method.

Keywords: Optimization, RPGT, Deep learning, Adam, SGD, RMS prop

I. Introduction

Plants and all of its industries is not one unique unit. Different processing techniques and businesses assemble many units. This chapter analyses a system made up of two units that incur degradation upon complete failure and that only receive preventive maintenance on one unit prior to a partial failure. A variety of devices coupled in series, parallel, or mixed mode make up a system. The failure of each individual unit determines the system's overall failure. Preventive maintenance is applied in many process industries, including those where a two-unit system is common. In these sectors, one unit is more critical than the other and needs greater care. The system as a whole fails if that one unit malfunctions.

Agrawal et al. [1] this research study uses the RPGT to analyzed water treatment RO plant's profitability under particular structure parameter settings. Chen and Hsieh [3] in this research, we transform the two-dimensional uninterrupted - k - out - of - n: F structure by using fake perfect components. The effectiveness of the strategy put forth is demonstrated by numerical results. A sensitivity study of a urea fertiliser production system with multiple sub-systems of different types is shown in Garg et al. [2]. System managers, exercise supervisors, engineers, and trustworthiness analysts in the industrial sector can all benefit from the analysis and findings presented in this paper. Kumar [4] the author of this study has examined a system known as the linear consecutive 2-ot-of-4: F structure, which has a unique kind of k-out-of-n redundancy. Sensitivity analysis is also used to determine the system's critical units. Raghav et al. [5] studied the convenience and cost function of a continuous functioning series-parallel classification in a fixed time environment are assessed in this study. A strong statistical test is castoff to associate the outcomes and the PSO comes out on top. Singla et al. [6] this research uses supplementary variable technique to calculate the dependability of a four-unit Polytube manufacturing factory. Singla et al. [7] this study presents a scientific model based on the Chapman Kolmogorov approach for determining availability under limited capacity, with the aid of transition diagrams linked to different conceivable combinations of probability. The analysis found that the most significant influence on the overall system availability of some subsystems is the subsystem extruder. Using RPGT, Singla et al. [8] examined the Rice Plant Cost Optimization and Mathematical Modeling. The mathematical modelling and optimisation of the feed plant's system parameters using machine learning technology was examined by Singla et al. [9].Taking into account the significance of each unit in the structure

Therefore, in this study, we require analyzed a two-unit structure under PM in the main unit previously complete disappointment and degradation after comprehensive failure, bearing in mind the relevance of each individual unit in the overall system. Every time the unit deteriorates more, there will come a point at which it can no longer be repaired or it might not be wise to do so because doing so would increase maintenance costs or cause production to be lost. The quality of products may not be up to the mark there may be difficulty in selling the products in the market, resulting into a loss in market share and repeat order of product. Keeping all these in mind, an individual unit needs more care hence preventive maintenance as per schedule or when the need arises is carried out by a repairman or server. A system transition diagram is created by accounting for different scenarios and path probabilities. Since the failure rate of most units is exponential, the failure rate of an individual is also exponential. The unit gets replaced with a new one if the server detects that it cannot be repaired. To ascertain whether the unit is operating at full capacity or at a reduced capacity, fuzzy logic may be employed.

II. Assumptions, Notations and Transformation Diagram

- The system is discussed for long run means for time is infinite
- Preventive Maintenance is available for main unit A only not in other units.
- The backup unit is activated as soon as the primary unit fails, provided that the switch is operational and unbreakable.
- The main unit is switched on as online if the standby unit is online while the main unit is being serviced.
- There is only one repairman facility.
- Switching over to connect devices is considered perfect.
- Unit A can fail wholly and over partial failure in both ways whereas unit B can fail completely.

- The failure rate from A to $\bar{A}$ is λ_1, $\bar{A}$ to a is λ_2, $\bar{A}_1$ to a_1 is λ_3 and A to a is λ.
- The failure rates of unit B from B to b is λ_4.
- The repair rates of unit $\bar{A}$ to A is w_1, a to $\bar{A}_1$ is w_2, a_1 to $\bar{A}_1$ is w_3 respectively.
- The repair rates of unit B from b to B is w_4.
- g(t) : probability density function that a new unit replaces unrepeatable unit A.
- g*(w) is Laplace Transform of probability density function g(t).
- States S_0, S_1, S_3 and S_4 are regenerative states.
- S_0=AB, S_1=$\bar{A}$B, S_2=Ab, S_3=$\bar{A}_1$B, S_4=a_1B, S_5=$\bar{A}_1$b, S_6=$\bar{A}$b, S_7=Ab

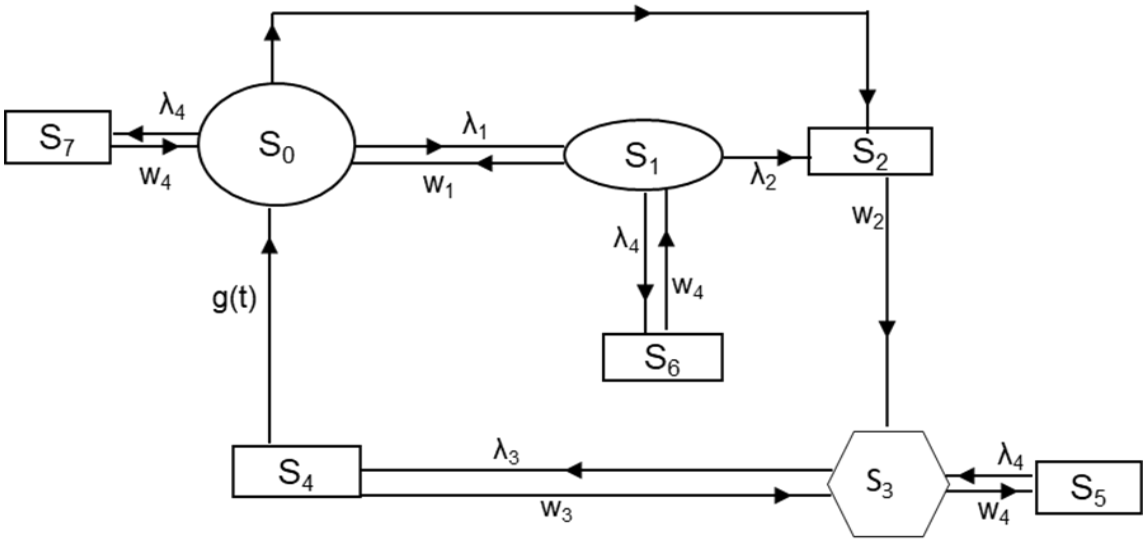


Figure 1: *Transition Diagram*

III. Description of Model

In the system under study there are two units A and B operational in full measurements initially in state S_0. Unit A need subunits in parallel so it can work in reduced capacity $\bar{A}$ when some of its subunits fail with failure rate λ_1, hence the arrangement work in reduced state S_1 in which unit B is functioning in full capacity. When unit A is in reduced state $\bar{A}$ then it is provided preventive maintenance and restored to its original state A by the repairman / server with repair rate w_1. Unit A may also fail directly from initial state at full capacity (A) to failed state a with a failure rate λ or from reduced state $\bar{A}$ to state a with failure rate $\lambda_2 > \lambda_1$ if it is not repaired within a reasonable time to state S_2, state S_2 is the state from which unit A can be repaired to its initial state and upon repair by the server it works in degraded state $\overline{A_1}$ denoted by the state S_3 while unit B is in good state. Upon failure of unit A with failure rate λ_3 to a_1 to state S_4 which on subsequent repair is restored to state S_3 each time with degraded capacity. As the repair is not perfect or it is beyond repair to bring the unit to its previous working state with repair rate w_3, there is a situation when the unit A elsewhere repair and it is substituted by a new unit with probability density function g (t). As the unit B have subunits in series so on failure of any one subunit in B it fails completely with failure rate λ_4 hence whole of the system fail from state S_0, S_1, S_3 to the state S_7, S_6, S_5 respectively in figure 1.Regarding the aforementioned symbols, the structure may be in any of the subsequent states. A Markov process, as we all know, is a stochastic process whose operational behavior is such that the probability

distribution for its future development depends only on its current state, not on the process's path to get there. The Markov process is referred to as a Markov chain if the state space is discrete, that is, finite and countable infinite. The Markov chain's parameter could be continuous or discrete. The chain is referred to as a discrete parameter Markov chain if the parameter space (Index set) is likewise discrete. The duration of the system's stay in any state, as measured by the probability density function, is exponential. The next transition from state i to state j will occur because it was in state i at a specific point in time (time 't'). This transition is dependent just on i and j, not on the history of the process that led to the state i in the past. Certain issues related to queuing, reliability, and inventory theories can be tackled by utilizing Markov renewal processes. Rather than the depth of their theoretical advancements, the Markov renewal theory's significance resides mostly in their broad range of applications. Stated differently, a semi-Markov process is one that follows a Markov chain to change its state, but it deviates from the Markovian property that the future is self-determining of the past specified the existing state and has a random time interval between changes. An activity is any portion of the project which consumes time or resources and has a drainable beginning and ending the dummy activity, represented by a dashed line in project, is to confirm that all the succeeding activities can begin after the completion of all the preceding activities before the dummy. Dummy activity caries a zero time. A node or event is a instantaneous point in time, expressing the beginning and ending of activities. The nodes or events are to be numbered in ascending order. Each activity's successor node number must be larger than its predecessor node number. The complexity of a project network depends on two factors namely the number of activities in the project and secondly on the precedence-relations among the activities. A decomposition technique is developed here to facilitate to identify to which of the above four classes, a given general project network belongs.

IV. Determination of base-state

Four, two, two four, three, one, one, and one primary circuit are located at the vertices 0, 1, 2, 3, 4, 5, 6, and 7, respectively, in the transition diagram (Figure 1). Every vertex 0 and 3 has a primary circuit linked with it. Therefore, any of them could constitute the system's initial state. As of right now, there are the fewest distinct secondary circuits along each of the principal lines that lead from vertex '0' to every vertex. The pathways leading from the vertex "0" do not contain any tertiary or higher level circuits. Therefore, there are 3 primary circuits sideways all paths as of the vertex '0'. And alike there are four, two and one primary, secondary and tertiary circuits individually as of the vertex '2'. Since, there is biggest number (four) of primary circuits at vertex '0' by less integer of secondary, tertiary and higher level circuits, consequently, '0' stands a base-state of model. This indicates that the principal circuits in states '0' and '3' are identical. However, there are fewer secondary and tertiary circuits in state "0", S0 so '0' base state $\xi = 0$.

Table 1: *Primary, Secondary, Tertiary Circuits w.r.t. the Simple Paths (Base-State '0')*

Vertex j	$\mathbf{(0 \to j)}$**: (P0)**	**(P1)**	**(P2)**
1	$(0 \to 1)$:{0,1}	{1,6,1}	-
2	$(0 \to 2)$:{0,2} $(0 \to 2)$:{0,1,2}	- {1,6,1}	- -
3	$(0 \to 3)$:{0,2,3} $(0 \to 3)$:{0,1,2,3}	{3,5,3}, {3,4,3} {1,6,1}, {3,5,3}, {3,4,3}	- -

4	$(0 \to 4)$:{0,2,3,4} $(0 \to 4)$:{0,1,2,3,4}	{3,5,3}, {3,4,3} {1,6,1}, {3,5,3}, {3,4,3}	- -
5	$(0 \to 5)$:{0,1,2,3,5} $(0 \to 5)$:{0,2,3,5}	{1,6,1}, {3,5,3},{3,4,3} {3,5,3}, {3,4,3}	-
6	$(0 \to 6)$:{0,1,6}	{1,6,1}	-
7	$(0 \to 7)$:{0,7}	-	-

V. TRANSITIONAL PROBABILITIES

The transitional Probabilities are the likelihoods through which the system variations its state one to another by the passage of time. In a processing industry generally, A process consists of four components: processing, inspection, transport and storage operations. Of these, only processing adds value; the others can be viewed as waste. Traditional approach has been to reduce waste through improving the activities related to the waste. To reduce inspection, we for example, adopted sampling inspection. Industry's approach is to eliminate inspection altogether by providing mistake-proofing devices. To reduce waste in transportation, we adopted usage of aids such as forklifts. Fundamental improvement in plant layout, however, will eliminate the need for transport altogether. Thus, it is confirmed that the overall state probability for every state is 1.

Table 2: *Transition Probabilities for this system*

$q_{i,j}^{(t)}$	$P_{ij} = q^*_{i,j}{}^{(t)}$
$= \lambda_1 e^{-(\quad_4)t}$ $= \lambda e^{-(\lambda \quad_4)t}$ $= \lambda_4 e^{-(\lambda \quad_4)t}$	$= \lambda_1/(\lambda+\lambda_1+\lambda_4)$ $= \lambda/(\lambda+\lambda_1+\lambda_4)$ $= \lambda/(\lambda+\lambda_1+\lambda_4)$
$= \quad^{-(w \quad_4)t}$ $= \quad^{-(w \quad_4)t}$ $= \quad^{-(w \quad_4)t}$	$= w_1/(w_1+\lambda_2+\lambda_4)$ $= \lambda_2/(w_1+\lambda_2+\lambda_4)$ $= \lambda_4/(w_1+\lambda_2+\lambda_4)$
$=$	$=1$
$= \quad^{-(\lambda \quad_4)t}$ $= \quad^{-(\lambda \quad_4)t}$	$= \lambda_3/(\lambda_3+\lambda_4)$ $= \lambda_4/(\lambda_3+\lambda_4)$
$= g(t)e$ $= \quad \overline{g(t)}$	$= g^*w_3$ $= 1\text{-}g^*w_3$
$=$	$= w_4/w_4 = 1$
$=$	$= w_4/w_4 = 1$
$=$	$= w_4/w_4 = 1$

Table 3: Mean Sojourn Times

$R_i(t)$	$\mu_i = R_i^*(0)$
$(t) = \quad^{-(\lambda \quad)t}$	$\mu_0 = 1/(\lambda+\lambda_1+\lambda_4)$
$(t) = \quad^{-(w \quad)t}$	$\mu_1 = 1/(w_1+\lambda_2+\lambda_4)$
$(t) =$	$\mu_2 = 1/w_2$
$(t) = \quad^{-(\lambda \quad)t}$	$\mu_3 = 1/(\lambda_3+\lambda_4)$
$(t) = \quad \overline{g(t)}$	$\mu_4 = [1\text{-}g^*[(w_3)]/w_3$
$(t) =$	$\mu_5 = 1/w_4$
$(t) =$	$\mu_6 = 1/w_4$
$(t) =$	$\mu_7 = 1/w_4$

5.1 Analyzation of System Parameters

The key parameters of the coordination are appraised by defining a 'base-state' and using Regenrative Point Graphical Technique. The parameter MTSF is unwavering w.r.t. the personalize state '0' and the additional parameters are attained via exhausting base-state. The steady state path probabilities exist provided w.r.t. Base State '0' refers to

$V_{0,0} = \{(0,1,0)/[1-(1,6,1)\}+\{(0,7,0)/1\}+\{(0,1,2,3,4,0)/[1-(1,6,1)1-(3,4,3)1-(3,5,3)\}]$
$+\{(0,2,3,4,0)/[1-(3,5,3)[1-(3,4,3)\}]$
$= (p_{0,1}p_{1,0})/(1-p_{1,6}p_{6,1})+p_{0,7}p_{7,0}+(p_{0,1}p_{1,2}p_{2,3}p_{3,4}p_{4,0})/(1-p_{1,6}p_{6,1})(1-p_{3,4}\ p_{4,3})$
$(1-p_{3,5}p_{5,3})+(p_{0,2}p_{2,3}p_{3,4}p_{4,0})/[(1-p_{3,4}\ p_{4,3})(1-p_{3,5}p_{5,3})]$
$= \{\lambda_1 w_1/(w_1+\lambda_2)(\lambda_1+\lambda+\lambda_4)\}]+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1\lambda_2(\lambda_3+\lambda_4)/\lambda_4(w_1+\lambda_2)(\lambda_1+\lambda+\lambda_4)\}$
$g^*(w_3)+\{\lambda(\lambda_3+\lambda_4)/\lambda_4(\lambda_1+\lambda+\lambda_4)\}.g^*(w_3)= 1$

$V_{0,1} = \{p_{0,1}/1-(1,6,1)\} = \{\lambda/(\lambda_1+\lambda+\lambda_4)\}/1-\lambda_4(\lambda_2+w_1+\lambda_4) = \{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)$
$(\lambda_1+\lambda+\lambda_4)\}$

$V_{0,2} = \{(0,1,2)/1-(1,6,1)\}+(0,2) = \{p_{0,1}p_{1,2}/1-p_{1,6}p_{6,1}\}+p_{0,2}$
$= [\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\lambda_2/(\lambda_2+w_1+\lambda_4)\}/1-\{\lambda_4(w_1+\lambda_2+\lambda_4)\}]$
$= \{\lambda_1\lambda_2/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}+\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\}$
$= \{\lambda_1\lambda_2+\lambda(\lambda_2+w_1)\}/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}$

$V_{0,3} = \{(0,1,2,3)/1-(1,6,1)1-(3,5,3)1-(3,4,3)\}+\{(0,2,3)/1-(3,5,3)1-(3,4,3)\}$
$= \{p_{0,1}p_{1,2}\ p_{2,3}/(1-p_{1,6}p_{6,1})(1-p_{3,5}p_{5,3})(1-p_{3,4}\ p_{4,3})\}+\{p_{0,2}\ p_{2,3}/(1-p_{3,5}p_{5,3})(1-p_{3,4}\ p_{4,3})\}$
$= [\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\ \lambda_2/(\lambda_2+w_1+\lambda_4)\}/\{1-\{\lambda_4/(\lambda_2+w_1+\lambda_4)1-\lambda_4/(\lambda_3+\lambda_4)1-\lambda_3/(\lambda_3+\lambda_4)]$
$+[\{\lambda/(\lambda_1+\lambda+\lambda_4)\}/\{1-\lambda_4/(\lambda_3+\lambda_4)1-\lambda_3/(\lambda_3+\lambda_4)]$
$= \{\lambda_1\lambda_2(\lambda_3+\lambda_4)^2/(\lambda_2+w_1)\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)\}+\{\lambda/(\lambda_3+\lambda_4)(\lambda_1+\lambda+\lambda_4)\}$
$= \{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}$

$V_{0,4} = \{(0,1,2,3,4)/[1-(1,6,1)1-(3,5,3)1-(3,4,3)\}+\{(0,2,3,4)]/[1-(3,5,3)1-(3,4,3)\}$
$= \{p_{0,1}p_{1,2}\ p_{2,3}p_{3,4}/(1-p_{1,6}p_{6,1})(1-p_{3,5}p_{5,3})(1-p_{3,4}\ p_{4,3})\}+\{p_{0,2}\ p_{2,3}p_{3,4}/(1-p_{3,5}p_{5,3})$
$(1-p_{3,4}\ p_{4,3})\}$
$= \{(\lambda_3+\lambda_4)/\lambda_4(\lambda_1+\lambda+\lambda_4)\}\{(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/(w_1+\lambda_2)\}$

$V_{0,5} = \{(0,1,2,3,5)/1-(1,6,1)1-(3,5,3)1-(3,4,3)\}+\{(0,2,3,5)/1-(3,5,3)1-(3,4,3)\}$
$= \{p_{0,1}p_{1,2}\ p_{2,3}p_{3,5}/(1-p_{1,6}p_{6,1})(1-p_{3,5}p_{5,3})(1-p_{3,4}\ p_{4,3})\}+\{p_{0,2}\ p_{2,3}p_{3,5}/(1-p_{3,5}p_{5,3})$
$(1-p_{3,4}\ p_{4,3})\}$
$= [\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\ \lambda_2/(\lambda_2+w_1+\lambda_4)\ \lambda_4/(\lambda_3+\lambda_4)\}/\{1-\{\lambda_4/(w_1+\lambda_2+\lambda_4)1-\lambda_4/(\lambda_3+\lambda_4)$
$1-\lambda_3/(\lambda_3+\lambda_4)]+[\{\lambda/(\lambda_1+\lambda+\lambda_4)\}\{\lambda_4/(\lambda_3+\lambda_4)\}]/[\{1-\lambda_4/(\lambda_3+\lambda_4)1-\lambda_3/(\lambda_3+\lambda_4)]$
$= \{\lambda_1\lambda_2(\lambda_3+\lambda_4)/(w_1+\lambda_2)(\lambda_1+\lambda+\lambda_4)\lambda_3\}+\{\lambda(\lambda_3+\lambda_4)/(\lambda_1+\lambda+\lambda_4)\lambda_3\}$
$= \{(\lambda_3+\lambda_4)/\lambda_3(\lambda_1+\lambda+\lambda_4)\}\{(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/(w_1+\lambda_2)\}$

$V_{0,6} = \{(0,1,6)/1-(1,6,1)\} = p_{0,1}p_{1,6}/(-p_{1,6}\ p_{6,1}) = [\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\}\{\lambda_4/(w_1+\lambda_2+\lambda_4)\}]/$
$[1-\{\lambda_4/(w_1+\lambda_2+\lambda_4)\}]$
$=\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)$
$[\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/$
$(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_2\}\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}$
$\{(1-g^*w_3)/w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}$
$+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]/[\{1/(\lambda_1+\lambda+\lambda_4)\}$
$+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{(\lambda_2+w_1)$
$(\lambda_1+\lambda+\lambda_4)\}\{1/w_2\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}$
$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)(1-g^*w_3)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\ w_3\}\{1-g^*(w_3)/w_3\}+$
$\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}$
$\{1/w_4\}+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]$

$V_{0,7}= p_{0,7} = \lambda_1/(\lambda_1+\lambda+\lambda_4)$

5.2 Mean Time to System Failure (MTSF) (T_0)

The reformative un-failed states toward which the structures dismiss transit, previously entering first unsuccessful state are: 'i' = 0 -7.

$$\text{MTSF} = \left[\sum_{i,sr}\left\{\frac{\left\{\text{pr}\left(\xi^{-}{}^{sr(sff)}\right)\right\}}{\left\{1-\overline{m_{1m}}\right\}}\right\}\right] \div \left[1-\sum_{sr}\left\{\frac{\left\{\text{pr}\left(\xi^{-}{}^{sr(sff)}\right)\right\}}{\left\{1-\overline{m_{2m}}\right\}}\right\}\right] \quad (1)$$

$$T_0 = [(0,0)\mu_0+(0,1)\mu_1]/[1-(0,1,0)$$
$$= (p_{0,0}\mu_0+p_{0,1}\mu_1)/(1-p_{0,1}p_{1,0})$$
$$= [1/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\}1/(\lambda_2+\lambda_4+w_1)]/[1-\{\lambda_1/(\lambda_1+\lambda+\lambda_4)\}$$
$$\{w_1/(w_1+\lambda_2+\lambda_4)\}]$$
$$= (\lambda_2+\lambda_4+w_1+\lambda_1)/[(\lambda_1+\lambda+\lambda_4)(w_1+\lambda_2+\lambda_4)-\lambda_1 w_1]$$

5.3 Availability of the System:

The reformative states at which structure is presented exist 'j' = 0,1,3 and reformative states exist 'i' = 0 to 4 attractive 'ξ' = '0' then the structure is available is assumed by

$$A_0 = \left[\sum_{j,sr}\left\{\frac{\{\quad(\xi^{sr\rightarrow})\}}{\left\{1-\overline{m_{1m}}\right\}}\right\}\right] \div \left[\sum_{i,s_r}\left\{\frac{\{\quad(\xi^{sr\rightarrow i)}\}}{\left\{1-\overline{m_{2m}}\right\}}\right\}\right] \quad (2)$$

$$A_0 = [\sum \quad] \div [\sum \quad]$$
$$= V_{0,0}f_0\mu_0+V_{0,1}f_1\mu_1+V_{0,3}f_3\mu_3/(V_{0,0}\mu_0{}^1+V_{0,1}\mu_1{}^1+V_{0,2}\mu_2{}^1+V_{0,3}\mu_3{}^1+V_{0,4}\mu_4{}^1+V_{0,5}\mu_5{}^1$$
$$+V_{0,6}\mu_6{}^1+V_{0,6}\mu_7{}^1)$$
$$=(V_{0,0}\mu_0+V_{0,1}\mu_1+V_{0,3}\mu_3)(V_{0,0}\mu_0{}^1+V_{0,1}\mu_1{}^1+V_{0,2}\mu_2{}^1+V_{0,3}\mu_3{}^1+V_{0,4}\mu_4{}^1+V_{0,5}\mu_5{}^1+V_{0,6}\mu_6{}^1$$
$$+V_{0,7}\mu_7{}^1)$$
$$= [\{1/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_3+\lambda_4)^2$$
$$(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4(\lambda_1+\lambda+\lambda 4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}]/\{1/(\lambda_1+\lambda+\lambda_4)\}$$
$$+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{(\lambda_2+w_1)$$
$$(\lambda_1+\lambda+\lambda_4)\}\{1/w_2\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}$$
$$+\{\lambda_3\lambda_4(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1-g^*(w_3)/w_3\}+\{(\lambda_3+\lambda_4)$$
$$(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}$$
$$= [\{(\lambda_2+w_1)/1\}+\lambda_1+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3\lambda_4\}]/[\{\lambda_2+w_1+\lambda_1(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_1\}$$
$$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)(1-g^*(w_3))/\lambda_4 w_3\}$$
$$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3 w_4\}+\{\lambda_1\lambda_4/w_4\}+\{\lambda_4(\lambda_2+w_1)/ w4\}]$$

5.4 Proportional Busy Period of Server

The reformative states where attendant 'j' = 1,2,3,4 attractive ξ = '0', for which the attendant remains hard is

$$B_0 = \left[\sum_{j,sr}\left\{\frac{\{\quad(\xi^{sr\rightarrow}j)\}}{\left\{1-\overline{m_{1m}}\right\}}\right\}\right] \div \left[\sum_{i,s_r}\left\{\frac{\{\quad(\xi^{sr\rightarrow i)}\}}{\left\{1-\overline{m_{2m}}\right\}}\right\}\right] \quad (3)$$

$$B_0 = \left[\sum_j V_{\xi,j}\,, n_j\right] \div [\sum \quad]$$
$$= (V_{0,1}n_1+V_{0,2}n_2+V_{0,4}n_4+V_{0,5}n_5+V_{0,6}n_6+V_{0,7}n_7)/(V_{0,0}\,\mu^1{}_0+V_{0,1}\mu^1{}_1+V_{0,2}\mu^1{}_2$$
$$+V_{0,3}\mu^1{}_3+V_{0,4}\,\mu^1{}_4+V_{0,5}\mu^1{}_5+V_{0,6}\mu^1{}_6+V_{0,7}\mu^1{}_7)$$
$$=(V_{0,1}\mu_1+V_{0,2}\mu_2+V_{0,4}\mu_4+V_{0,5}\mu_5+V_{0,6}\mu_6+V_{0,7}\mu_7)/(V_{0,0}\mu_0+V_{0,1}\mu_1+V_{0,2}\mu_2+V_{0,3}\mu_3$$
$$+V_{0,4}\,\mu_4+V_{0,5}\,\mu_5+V_{0,6}\mu_6+V_{0,7}\mu_7)$$
$$= [\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/$$
$$(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_2\}\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}$$
$$\{(1-g^*w_3)/w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}$$
$$+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]/[\{1/(\lambda_1+\lambda+\lambda_4)\}$$
$$+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{(\lambda_2+w_1)$$
$$(\lambda_1+\lambda+\lambda_4)\}\{1/w_2\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}$$
$$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)(1-g^*w_3)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\ w_3\}\{1-g^*(w_3)/w_3\}+$$
$$\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}$$
$$\{1/w_4\}+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]$$
$$= [(\lambda_1/1)+(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)(1-g^*w_3)/\lambda_4 w_3\}+\{(\lambda_3+\lambda_4)$$
$$(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{\lambda_1\lambda_4/w_4\}+\{\lambda_1(\lambda_2+w_1)/w_4\}]/[\{(\lambda_2+w_1)+\lambda_1(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)$$
$$/w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_4 w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\ (1-g^*w_3)/\lambda_3 w_4\}$$

$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_4 w_3\}+\{\lambda_1\lambda_4/w_4\}\{\lambda_4(\lambda_2+w_1)/w_4\}]$

5.5 Expected Fractional Number of repairman's Visits (V_0)

The reformative states where overhaul man do this job j = 1 Taking 'ξ' = '0', the integer of visit by the overhaul man is specified by

$$V_0=\left[\sum_{j,sr}\left\{\frac{\{pr(\xi^{sr\rightarrow}j)\}}{\{1-\overline{k}\}}\right\}\right]\div\left[\sum_{i,s_r}\left\{\frac{\{pr(\xi^{sr\rightarrow}i)\}}{\{1-\overline{k}\}}\right\}\right] \quad (4)$$

$V_0 = [\sum\ ] \div [\sum\ ]$

$= (V_{0,1}+V_{0,2}+V_{0,4}+V_{0,5}+V_{0,6}+V_{0,7})/(V_{0,0}\,\mu^1{}_0 +V_{0,1}\,\mu^1{}_1 +V_{0,2}\,\mu^1{}_2 +V_{0,3}\,\mu^1{}_3 +V_{0,4}\,\mu^1{}_4 +V_{0,5}\,\mu^1{}_5+V_{0,6}\mu^1{}_6+V_{0,7}\mu^1{}_7)$

Taking $\mu_j{}^1 = \mu_j$, using value we get

$= [\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/\lambda_4+1/\lambda_3\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}]/[\{1/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_4 w_3(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_4(\lambda_1+\lambda+\lambda_4)(w_1+\lambda_2)\}\{1-g^*w_3/(w_3)\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_4(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]$

$= [\{\lambda_1(\lambda_2+w_1+\lambda_4)/1\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/1\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{\lambda_1\lambda_4/1\}+\lambda_4(w_1+\lambda_2)]/[(\lambda_2+w_1)+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3\lambda_4(1)\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)(1-g^*w_3)/\lambda_4\lambda_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{(\lambda_1+\lambda_4)/w_4\}+\{\lambda_4(\lambda_2+w_1)/w_4\}]$

$[\{1/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4(\lambda_1+\lambda+\lambda4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}]/\{1/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+w_1)\}/\{(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/w_2\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}+\{\lambda_3\lambda_4(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1-g^*(w_3)/w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]$

$= [\{(\lambda_2+w_1)/1\}+\lambda_1+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3\lambda_4\}]/[\{\lambda_2+w_1+\lambda_1(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_1\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)(1-g^*(w_3))/\lambda_4 w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3 w_4\}+\{\lambda_1\lambda_4/w_4\}+\{\lambda_4(\lambda_2+w_1)/w4\}]$

5.6 Profit Function of the System

Profit function can be used to analyses the system's profitability.

P_0 = Mean Revenue Earning Rate *Availability of system – mean rate of cost that server is busy *total busy period of the server-mean cost per visit which server charges *number of visits server called in a unit time.

$$P_0 = C_1A_0 - C_2B_0 - C_3V_0 \quad (5)$$

$=C_1[\{(\lambda_2+w_1)/1\}+\lambda_1+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3\lambda_4\}]/[\{\lambda_2+w_1+\lambda_1(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_1\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)(1-g^*(w_3))/\lambda_4 w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3 w_4\}+\{\lambda_1\lambda_4/w_4\}+\{\lambda_4(\lambda_2+w_1)/w4\}]$

$-C_2[(\lambda_1/1)+(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)(1-g^*w_3)/\lambda_4 w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{\lambda_1\lambda_4/w_4\}+\{\lambda_1(\lambda_2+w_1)/w_4\}]/[\{(\lambda_2+w_1)+\lambda_1(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_4 w_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\ (1-g^*w_3)/\lambda_3 w_4\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_4 w_3\}+\{\lambda_1\lambda_4/w_4\}\{\lambda_4(\lambda_2+w_1)/w_4\}]$

$- C_3[\{\lambda_1(\lambda_2+w_1+\lambda_4)/1\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/1\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{\lambda_1\lambda_4/1\}+\lambda_4(w_1+\lambda_2)]/[(\lambda_2+w_1)+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_2\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3\lambda_4(1)\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)(1-g^*w_3)/\lambda_4\lambda_3\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{(\lambda_1+\lambda_4)/w_4\}+\{\lambda_4(\lambda_2+w_1)/w_4\}]$

$[\{1/(\lambda_1+\lambda+\lambda_4)\}+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_3+\lambda_4)^2$

$(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4(\lambda_1+\lambda+\lambda 4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}]/\{1/(\lambda_1+\lambda+\lambda_4)\}$
$+\{\lambda_1(\lambda_2+w_1+\lambda_4)/(\lambda_2+w_1)(\lambda_1+\lambda+\lambda_4)\}\{1/(\lambda_2+w_1+\lambda_4)\}+\{(\lambda_1\lambda_2+\lambda\lambda_2+w_1)\}/\{(\lambda_2+w_1)$
$(\lambda_1+\lambda+\lambda_4)\}\{1/w_2\}+\{(\lambda_3+\lambda_4)^2(\lambda_1\lambda_2+\lambda\lambda_2+w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/(\lambda_3+\lambda_4)\}$
$+\{\lambda_3\lambda_4(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)\}/\{\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1-g^*(w_3)/w_3\}+\{(\lambda_3+\lambda_4)$
$(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)\}/\{\lambda_3\lambda_4(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}+\{\lambda_1\lambda_4/(\lambda_1+\lambda+\lambda_4)(\lambda_2+w_1)\}\{1/w_4\}$
$+\{\lambda_4/(\lambda_1+\lambda+\lambda_4)\}\{1/w_4\}]$

$= [\{(\lambda_2+w_1)/1\}+\lambda_1+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3\lambda_4\}]/[\{\lambda_2+w_1+\lambda_1(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/w_1\}$
$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda\lambda_2+\lambda w_1)/\lambda_3\lambda_4\}+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)(1-g^*(w_3))/\lambda_4 w_3\}$
$+\{(\lambda_3+\lambda_4)(\lambda_1\lambda_2+\lambda w_1+\lambda\lambda_2)/\lambda_3 w_4\}+\{\lambda_1\lambda_4/w_4\}+\{\lambda_4(\lambda_2+w_1)/ w4\}]$

5.7 Model description of Two Unit System with Degradation and PM in Single Unit after Comprehensive Failure in Deep DLA

In the landscape of complex systems, the model describing a 2-Unit Organization with PM and Degradation in Single Unit after Comprehensive Disappointment marks a significant advancement, particularly when enriched by the sophistication of deep learning optimization techniques. This model unfolds as a comprehensive framework, weaving together intricate threads of preventive maintenance, unit degradation, and the repercussions of complete failure, all embedded within the deep learning paradigm. At its core, this model grapples with the intricacies of system reliability and performance optimization in the face of evolving challenges. The dual-unit configuration introduces a nuanced dynamic, wherein the occurrence of degradation in one unit following a complete failure introduces a layer of complexity that demands innovative solutions. The integration of deep learning optimization adds a layer of intelligence to this model, elevating it beyond traditional methodologies. Deep learning algorithms, through their aptitude to separately learn and adapt as of statistics patterns, offer a novel approach to addressing the uncertainties inherent in system behavior. By leveraging this advanced computational intelligence, the model seeks to enhance decision-making processes, predict degradation trajectories, and dynamically optimize preventive maintenance strategies. This model represents a pioneering effort to synergize the principles of reliability engineering, preventive maintenance, and artificial intelligence. It aims not only to mitigate the impact of failures but also to proactively manage system health, thereby contributing to enhanced operational efficiency and longevity. In the broader context of complex systems, where resilience and adaptability are paramount, this model stands at the forefront of innovation. Its potential applications span diverse sectors, from critical infrastructure to manufacturing processes, where optimizing system performance is not merely a goal but a necessity for sustained success. As we delve into the intricacies of this model, we uncover not just a technological advancement but a paradigm shift in how we approach the challenges of system reliability and maintenance in the era of deep learning optimization**.**

- Two Unit System with PM using Deep Learning Optimization: In the ever-evolving landscape of systems engineering, the model of a Two-Unit Structure through PM stands as a testament to the quest for enhanced reliability and efficiency. This sophisticated framework not only grapples with the intrinsic complexities of a dual-unit configuration but also incorporates the cutting-edge paradigm of deep learning optimization, ushering in a new era of intelligent and data-driven maintenance strategies. At its core, the Two-Unit System with Preventive Maintenance embodies a proactive approach to system health, acknowledging the inevitability of wear and degradation. This model recognizes the critical role of preventive maintenance in averting potential failures and ensuring the continuous functionality of both units. The challenge, however, lies not only in devising effective preventive maintenance schedules but in optimizing these strategies dynamically based on the evolving conditions of the system. Enter the realm of deep learning optimization – an advanced paradigm that harnesses the power of artificial intelligence to glean

insights, make predictions, and adapt strategies autonomously. By integrating deep learning into the framework, this model transcends traditional approaches, allowing for a more nuanced understanding of system behavior. Deep learning algorithms, capable of learning intricate patterns from vast datasets, become the intelligent backbone of the maintenance optimization process. This model represents a departure from conventional methodologies, embracing the notion that data-driven decision-making can revolutionize the way preventive maintenance is conceptualized and implemented. The deep learning optimization component adds a layer of adaptability, allowing the system to learn from its own performance, predict potential degradation trends, and optimize preventive maintenance schedules in real-time. In an era where the reliability and longevity of systems are paramount, the Two-Unit System with Preventive Maintenance using Deep Learning Optimization emerges as a trailblazer. Its applications extend across industries where system downtime is not merely an inconvenience but a critical concern. From manufacturing plants to critical infrastructure, this model offers a glimpse into the future of maintenance strategies – one where intelligence and adaptability converge to ensure the seamless operation of complex systems. As we delve into the intricacies of this model, we embark on a journey towards a more resilient, efficient, and intelligent approach to preventive maintenance in the age of deep learning optimization.

- Behavioral Analysis of Two Unit System using Deep Learning Optimization Algorithms: Deep learning algorithms are specialized techniques designed to find the optimal parameters for Behavioral Analysis of Two Unit System during training. These algorithms aim to minimize the failure rate, allowing the Behavioral Analysis of Two Unit System to learn from the data and make accurate predictions. Here are some commonly used deep learning optimization algorithms:
- Stochastic Gradient Descent (SGD): SGD is the substance of numerous optimization algorithms. It appraises the parameters of neural network after processing each mini-batch of training data, making it computationally efficient. However, its noise can result in oscillations around the optimal solution.
- Adam (Adaptive Moment Estimation): Adam stands an adaptive culture rate optimization algorithm combining RMS prop and Momentum ideas. It dynamically adjusts the knowledge rates for both parameter built on their historic and squared gradients, if better convergence properties.
- RMS prop (Root Mean Square Propagation): RMS prop alters the scholarship rates for both parameters exclusively by in-between the wisdom rate by the square root of average of the squared gradients. It helps mitigate the diminishing learning rate problem in traditional SGD.

VI. Data Set

The dataset designed for the optimization of a Two-Unit System through PM and Degradation in Single Unit after Far-reaching Failure represents a comprehensive repository of information capturing the intricate dynamics of the system's behavior over time. This dataset serves as a vital tool for researchers and practitioners seeking to unravel the nuanced interplay between preventive maintenance, degradation, and complete failures within the context of a dual-unit system. At its core, the dataset encapsulates a temporal sequence of events, chronicling the life cycle of the system through detailed timestamps. These temporal markers form the backbone of the dataset, enabling the exploration of patterns, trends, and dependencies within the system's performance. The states of each unit, meticulously recorded over time, offer a granular understanding of the operational landscape. Distinctions between operational states, preventive maintenance intervals, and instances of degradation provide crucial insights into the conditions and transitions that shape the overall

behavior of the two-unit system. Central to the dataset are records of complete failure events, providing a narrative of critical system breakdowns. These records encapsulate not only the timing and affected units but also the consequences that reverberate through the entire system, laying the groundwork for a profound analysis of failure modes. Degradation indicators, quantifying the gradual decline of one unit after complete failure, contribute a quantitative dimension to the dataset. These indicators may encompass various measures such as performance metrics, sensor data, or degradation indices, offering a means to gauge the evolving health of the system. Preventive maintenance records intricately weave into the fabric of the dataset, documenting strategic interventions aimed at preserving system integrity. The time stamped instances of preventive maintenance, coupled with details on the nature of the maintenance conducted, furnish a comprehensive account of proactive efforts to sustain system reliability. Operational parameters, ranging from environmental conditions to load variations, enrich the dataset by providing contextual information influencing the system's performance. These parameters offer a broader perspective on external factors shaping the system's behavior. Performance metrics, representing the overarching effectiveness of the two-unit system, become benchmarks for evaluating reliability and availability. These metrics, woven into the dataset, serve as indicators of the system's health and efficacy. The historical data embedded in the dataset acts as a backdrop, offering a contextual lens through which to interpret current behaviors and anticipate future trajectories. Moreover, the dataset delves into the effectiveness of preventive maintenance actions, gauging their impact on mitigating degradation and preventing failures. External factors, meticulously documented, amplify the dataset's scope by acknowledging influences from the broader operational environment. These factors encompass external disturbances, changes in operating conditions, or any external events that bear significance on the system's presentation. We obligate estimated various execution assessment confusion matrices (Recall, Accuracy Precision, and F1- Measure) to assess the performance of our model's implementation. The goal of the model phase evaluation is to assess the design model's generalization precision and accuracy using a test dataset that has not yet been observed.Here we intended this accuracy thru put on the exactness (Availability of the System), accuracy (Mean Time to System Failure(MTSF)), Recall (Proportional Busy Period of the Server), f_score function (Expected Fractional Number of repairman's Visits (V_0)), that stand imported as of metrics module accessible into Scikit-learn Python archive that be contingent on the subsequent formula.In essence, the dataset unfolds as a narrative of the Two-Unit System's journey, providing a multidimensional exploration of its behavioral landscape. Optimal preventive maintenance scheduling takes center stage, with an inquiry into whether adjustments in timing and types of maintenance actions can yield improvements in system performance. The discussion revolves around the fine-tuning of preventive strategies for maximal impact. A panoramic view of system robustness and vulnerabilities is cast, considering both internal and external influences. As analysts embark on the journey of extracting insights from this dataset, they delve into the intricacies of system dynamics, seeking patterns, correlations, and opportunities to enhance reliability through informed decision-making.It is thinkable to establish which strictures are most important to the system's recital and to find the ideal values on behalf of apiece parameter by changing these and other important limitations one at a time and seeing how the system's output vagaries in response.

- Precision: Between these records of classified posts, messages, and news, precision shows the percentage of correctly identified sarcastic news messages. It describes how effective the suggested techniques are. Equation (6) provides the formula for estimating the Precision.

  ision = ——— (6)

- Recall: Recall displays the percentage of actual activity that includes sarcastic posts, news, and communications. It is also acknowledged that the recall is sensitive. One could estimate the values of Recall by applying the formula shown in Equation (7).

cal = —— (7)

- F-Measure: The F-measure suggests evaluating all performance metrics based on the measured results of accuracy and recall. Equation (8) provides a formulation for estimating the values of the f-measure.

asure = ———— (8)

VII. Results and Discussion

The temporal canvas of the system reveals nuanced patterns over time, portraying the ebb and flow of preventive maintenance actions, degradation events, and minimizes failures. By dissecting the temporal landscape, the analysis aims to uncover temporal trends and potential predictive insights that illuminate periods of heightened vulnerability or resilience. Within this temporal context, the reliability of each unit emerges as a focal point. Disparities in failure rates are scrutinized, and the effectiveness of preventive maintenance interventions is assessed. This unit-specific examination lays the foundation for a deeper understanding of the factors influencing overall system resilience. The quantitative analysis of degradation indicators in the unit post-complete failure provides valuable insights into the rate and nature of deterioration. This exploration not only dissects the degradation process but also unravels its ramifications for the broader health of the system. The evaluation extends to the effectiveness of preventive maintenance strategies. Beyond mere inspection, the analysis dissects the types and intervals of maintenance actions, seeking to unravel their influence on degradation mitigation and the prevention of minimize failures. System-level performance metrics, including reliability, availability, and mean time between failures, serve as critical barometers. These metrics encapsulate the overarching operational efficacy of the Two-Unit System, providing a quantitative lens through which to gauge its performance. The ensuing discussion navigates the terrain of results, offering a qualitative interpretation of the observed patterns and phenomena. It places a spotlight on temporal trends, unraveling their significance and exploring potential correlations with system vulnerabilities or robust periods. Unit-specific considerations form a pivotal part of the dialogue, addressing any observed disparities in failure rates or the efficacy of preventive maintenance. This unit-centric discussion lays the groundwork for tailored strategies to fortify specific components of the system. Strategies for managing and mitigating degradation in the unit post-complete failure are explored, with an emphasis on proactive measures. The discussion contemplates how interventions can be designed to arrest or decelerate the degradation process, enhancing overall system longevity. The discussion unfolds as a narrative that explores how the system responds to unexpected events, disturbances, and external factors. Comparisons with industry standards provide a benchmarking context, allowing for a critical assessment of whether observed behaviors align with established norms. This comparative analysis offers valuable insights into the system's standing within broader industry frameworks. In the synthesis of findings and discussions, practical recommendations emerge as guiding beacons. These recommendations are crafted to empower decision-makers with actionable insights for enhancing system reliability, optimizing preventive maintenance, and fortifying overall operational performance.

Table 4: *Table of parameter*

W(w1,w2,-------wn)	**ƛ(ƛ1,**	***n***)	**S(s,s2,-------sn)**	**p**
(0-.100)	**(0-.100)**		**(0-100)**	**(0-.68)**

Table 5: Performance of model

Model	MTSF	F1 Score(Expected Number of Inspections by the repair man)	Recall(Busy Period)	Precision(Availability of the System)
SGD	0.924	0.925	0.920	0.923
Adam	0.910	0.930	0.915	0.935
RMS Prop	0.923	0.927	0.918	0.930

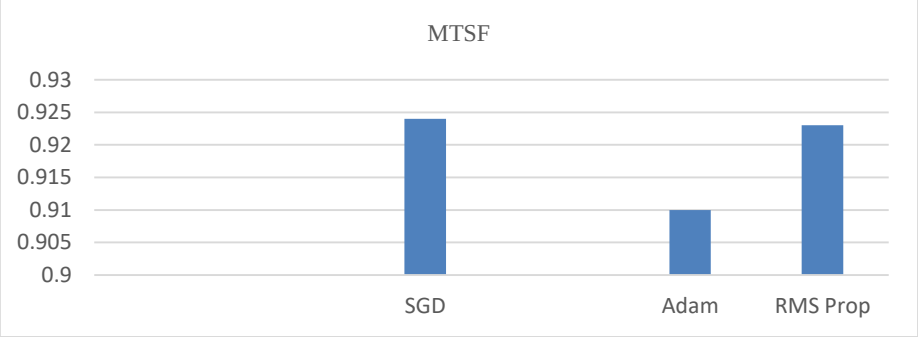


Figure 2: Compare all method using MTSF

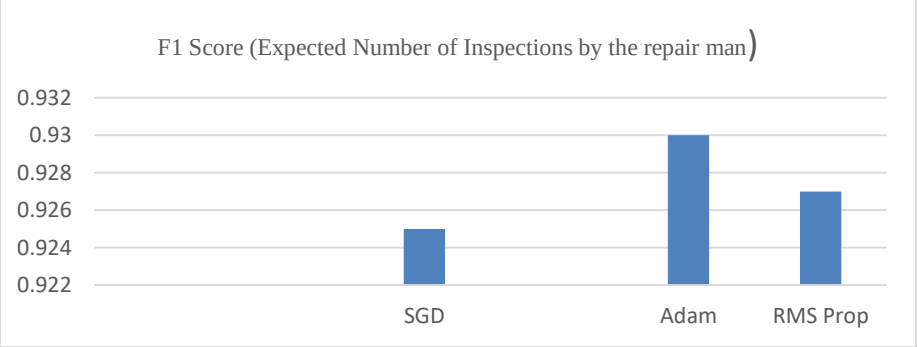


Figure 3: Compare all method using F1 Score

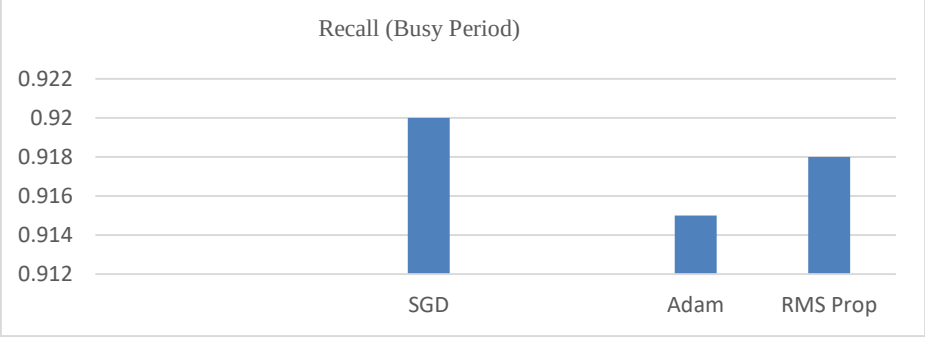


Figure 4: Compare all method using Recall

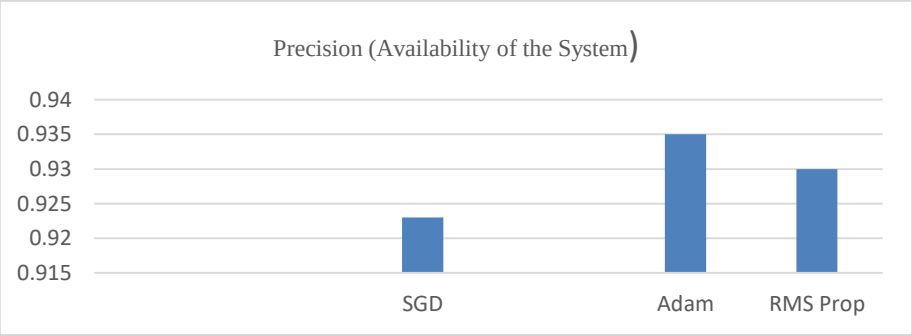


Figure 5: Compare all method using Precision

Using above table 4 and table 5 show draw graph Figure 2, Figure 3, Figure 4 and Figure 5. This model is train in three method (ADAM, SGD, and RMS Prop), in (MTSF of Adam is better than other two method. In Expected Number of Assessments by repair man of SGD is better than other two method. In Recall (Busy Period) of Adam is better than other two methods. In Precision (Availability of the System) of RMS Prop is better than other two method.

VIII. Conclusion

The behavioral analysis of a 2-unit system, incorporating PM and the introduction of degradation in single unit subsequently complete failure, yields comprehensive insights into the system's reliability and performance dynamics, Through a detailed examination, several significant conclusions emerge from the study. Firstly, the inclusion of preventive maintenance demonstrates a positive impact on the overall system reliability. Regular maintenance activities contribute to a reduction in the likelihood of unexpected failures, thereby enhancing the system's overall robustness. This, in turn, leads to increased availability, ensuring that both units are in optimal working condition for extended periods. However, the incorporation of degradation in one unit adds a layer of complexity to the system behavior. Degradation models provide a more realistic representation of the gradual decline in performance before complete failure occurs, allowing on behalf of a more nuanced sympathetic of the structure's behavior. The availability of the system is found to be intricately linked to the interplay between preventive maintenance and degradation. Striking the right balance between these factors becomes crucial in determining the system's ability to meet operational requirements. Optimization of maintenance strategies is paramount to minimizing downtime and maximizing availability effectively. Trade-off analysis reveals a delicate equilibrium between the cost of preventive maintenance and the potential savings derived from avoiding unexpected failures. This model is train in three methods (ADAM, SGD, and RMS Prop. In MTSFof Adam is better than other two methods. In Recall (Busy Period) of Adam is better than other two methods. In Precision (Availability of the System) of RMS Prop is better than other two method. Decision-makers must navigate this balance to determine the most cost-effective strategy for maintaining the system. Sensitivity analysis underscores the system's vulnerability to changes in various parameters such as maintenance intervals, degradation rates, and repair times. Identifying critical factors through sensitivity analysis is essential for informed decision-making and robust system design. Considering life-cycle costs', including maintenance and repair expenses, provides a comprehensive perspective on economic feasibility. Decision-makers must evaluate both short-term and long-term costs associated with chosen maintenance and operational strategies. The analysis also emphasizes the significance of risk mitigation. Identifying potential risks associated with system failures and degradation enables the

implementation of proactive maintenance and monitoring strategies, thereby enhancing overall system resilience.

References

[1] Agrawal, A., Garg, D., Kumar, A., and Kumar, R. (2021). Performance Analysis of the Water Treatment Reverse Osmosis Plant. *Journal of Reliability Theory and Applications,* 16(3): 16-25.

[2] Garg, D., Kumar, A., Joshi, V. K., and Fatima, N. (2022).Sensitivity Analysis of Urea Fertilizer Plant. *Journal of Reliability Theory and Applications,* 17(2), 178-189.

[3] Hsieh, Y., C. and Chen, T., C. (2004). Reliability lower bounds for two-dimensional consecutive-k-out-of-n: F systems. *Computers & Operations Research*, 31(8):1259–1272.

[4] Kumar, A. (2020). Reliability And Sensitivity Analysis of Linear Consecutive 2-out-of- 4: F System. *European Journal of Molecular & Clinical Medicine*, *7*(07): 3791-3804.

[5] Raghav, Y., S., Varshney, R., Modibbo, U. M., Ahmadini, A., A., H., and Ali, I. (2022). Estimation and optimization for system availability under preventive maintenance. *IEEE Access*, *10*: 94337-94353.

[6] Singla, S., Lal, A., K., and Bhatia, S., S. (2021).Reliability analysis of poly tube industry using supplementary variable Technique *Applied Mathematics and Computation* 281: 3981–3992.

[7] Singla, S., Modibbo, U., M., Mijinyawa, M., Malik, S., Verma, S., and Khurana, P. (2022). Mathematical Model for Analysing Availability of Threshing Combine Machine Under Reduced Capacity. *Yugoslav Journal of Operations Research*, *32*(4): 425-437.

[8] Singla, S., Kaur, K., Khurana R., and Garg, V. (2023). Mathematical Modeling and Cost Optimization of Rice Plant (An Application of RPGT). *IEEE Xplore, 694*-698.

[9] Singla, S., and Kaur, K.(2024). Mathematical Modeling and Optimization of System Parameters of Feed Plant Using Machine Learning. *Scite press Digital Library*, 158-162.

RT&A, No 4(80)
Volume 19, December, 2024

A NOVEL APPROACH TO DISTRIBUTION GENERATION WITH APPLICATIONS IN ELECTRICAL ENGINEERING

[1]Nuzhat Ahad; [2*]S.P.Ahmad; [3]J.A.Reshi

•

[1,2] Department of Statistics,University of Kashmir, Srinagar, India
[3] Department of Statistics, Govt. Degree College Pulwama, Srinagar, India
[1]nuzhatahad01@gmail.com, [2*]sprvz@yahoo.com, [3]reshijavaid19@gmail.com

Abstract

Many fields use standard distributions to model lifetime data. However, datasets from areas such as engineering and medical sciences frequently deviate from these standard distributions. This highlights the necessity for developing new distribution models that can accommodate significant variations in data patterns to better align with real-world observations. In this manuscript, we introduce a novel technique called the PNJ Transformation technique (named using the initials of its authors) for generating probability distributions. Using this technique, we developed a new and improved version of the Power function (PF) distribution, named the PNJ Power function (PNJ-PF) distribution. The PNJ-PF distribution offers superior flexibility compared to PF Distribution in terms of probability density function (pdf) and hazard rate function. We investigated the statistical properties of the PNJ-PF distribution and describe the maximum likelihood estimation (MLE) procedure for its parameters. To demonstrate the effectiveness and adaptability of the PNJ-PF distribution, we apply it to a simulated and two real-life datasets and compared proposed model fit with the traditional Power function model and other competitive models based on the various goodness-of-fit measures, such as the Akaike Information Criterion (AIC), Bayesian Information Criterion(BIC), Corrected AIC, Hannan−Quinn Information Criterion (HQIC) and these results are also justified graphically, further demonstrating the superiority and flexibility of the PNJ-PF distribution.

Power function distribution, hazard rate function, survival function, mean residual life, Maximum likelihood estimation

1. Introduction

Choosing the right statistical distribution is crucial for accurate data analysis across various fields such as medicine and engineering. While traditional distributions serve as foundational tools, their limitations in fitting complex real-world data necessities the need for enhanced models. The world of probability distributions encompasses a variety of extensions, from continuous to discrete, symmetric to asymmetric, designed to effectively capture the complexities and variability found in real world datasets. This evolution is paralleled by advancements in data science, which have transformed our capacity to derive insights from vast datasets. Understanding and utilizing data science techniques is crucial for meaningful analysis and applications in the context of probability distributions.

Lee [7] classified transformation techniques as composite methods because they seek to develop new distributions through the combination of existing ones or by integrating additional parameters into existing distributions. By adding extra parameters to base distributions, we aim to boost flexibility and enhance model accuracy, ensuring a better aligning with the characteristics of real-world data. Many innovative transformation techniques, extensively documented in statistical

literature, have been devised to introduce novel distributions. Significant contributions include Exponentiated technique for analyzing bathtub failure rate data using Weibull distribution by [12] , Alpha power transformation by [9], Tangent exponential transformation by [6], innovative transformation applied to Weibull distribution by [8], innovative SMP transformation studied on Lomax distribution by [14], Quadratic Rank Transmutation method (QRTM) proposed by [15]. Despite numerous methods to analyze real-world data, there's a persistent demand to find new ways to create different types of distributions. This shows a continued interest in exploring fresh approaches that can handle the complexities found in real data. This study explores innovative transformation technique for generating probability distributions, with a focus on improving the adaptability and practical utility of the Power function distribution. Power function distribution also called the inverse of Pareto distribution [5] is a simple life time model. And was often employed in the assessment of reliability of semiconductor devices and electrical components. Meniconi [11] were the first who proposed the probability density function (pdf) and cumulative distribution function (cdf) of two parameter PF distribution with scale parameter λ and shape parameter β and are given respectively as,

$$f(x,\lambda,\eta) = \frac{\eta x^{\eta-1}}{\lambda^{\eta}}\ ; \quad 0 < x < \lambda, \lambda > 0, \eta > 0$$

$$F(x,\lambda,\eta) = \left(\frac{x}{\lambda}\right)^{\eta}\ ; \quad 0 < x < \lambda, \lambda > 0, \eta > 0$$

Various extended models of PF distribution has been proposed in literature, some of them include Weibull-PF [16], Transmuted-PF [18], Transmuted Weibull-PF [17], Exponentiated-PF [3] ,among others.

2. PNJ Transformation and its Properties

Let X be a continuous random variable and $F(x)$ be its cumulative distribution function (cdf), then the cdf of PNJ transformation technique for $x \in \mathbb{R}$, is given as follows

$$F_{PNJ}(x) = \begin{cases} \frac{e^{log(\zeta)F(x)} - S(x)}{\zeta}\ ; & \zeta > 1 \\ F(x)\ ; & \zeta = 1 \end{cases} \tag{1}$$

$F_{PNJ}(x)$ is an absolute continuous distribution function, If $F(x)$ is an absolute continuous distribution function. Clearly, $F_{PNJ}(x)$ is a valid cdf, as it satisfied all the properties of valid cdf function , such as,

i $F_{PNJ}(-\infty) = 0$ and $F_{PNJ}(\infty) = 1$

ii $F_{PNJ}(x)$ is a monotonically increasing function of x.

iii $F_{PNJ}(x)$ is right continuous.

iv $0 \leq F_{PNJ}(x) \leq 1$.

The corresponding probability density function (pdf) of $F_{PNJ}(x)$ for $x \in \mathbb{R}$ is given as follows

$$f_{PNJ}(x) = \begin{cases} \frac{f(x}{\zeta}\{log(\zeta)e^{log(\zeta)F(x)} + 1\}\ ; & \zeta > 1 \\ f(x)\ ; & \zeta = 1 \end{cases} \tag{2}$$

The survival function $S_{PNJ}(x)$ for PNJ transformation is given by

$$S_{PNJ}(x) = \frac{(\zeta + S(x)) - e^{log(\zeta)F(x)}}{\zeta}\ ; \qquad \zeta > 1 \tag{3}$$

The hazard rate function $h_{PNJ}(x)$ for PNJ transformation is given by

$$h_{PNJ}(x) = \frac{f(x)(log(\zeta)e^{log(\zeta)F(x)} + 1)}{\zeta + S(x) - e^{log(\zeta)F(x)}}\ ; \qquad \zeta > 1 \tag{4}$$

The $h_{PNJ}(x)$ of PNJ transformation in terms of survival $S(x)$ and hazard rate function $h(x)$ of f can be written as

$$h_{PNJ}(x) = h(x)\bar{F(x)}\frac{(log(\zeta)e^{log(\zeta)F(x)} + 1)}{\zeta + S(x) - e^{log(\zeta)F(x)}}\ ; \qquad \zeta > 1 \tag{5}$$

The PNJ transformation technique, represented by the cumulative distribution function: (1) , maintains consistency with base distributions when $\zeta = 1$, ensuring no added complexity.
One of its primary advantages of PNJ transformation technique is the flexibility and adaptability introduced by the parameter ζ, which allows the transformation to smoothly adjust to different dataset characteristics. This adaptability is crucial in dealing with varied and complex data in applied sciences. Additionally, by manipulating ζ, new distributions can be generated to match specific real-world data characteristics. The innovative parameterization of ζ influences the shape and characteristics of the resulting distribution, making it customizable for specific modeling needs. Theoretical foundation, the technique preserves the essential properties of a cdf, ensuring reliability in statistical analysis.

3. PNJ Power function distribution and its properties

The cdf of PNJ-PF distribution can be obtained from (1) by taking $F(x) = F(x, \lambda, \eta)$, the cdf of the PF distribution, and is given by

$$F_{PNJ}(x, \Theta) = \frac{1}{\zeta}\{e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + \left(\frac{x}{\lambda}\right)^{\eta} - 1\}\ ; \qquad \zeta > 1 \tag{6}$$

and the corresponding pdf of PNJ-PF distribution is given by

$$f_{PNJ}(x, \Theta) = \frac{\eta x^{\eta-1}}{\zeta\lambda^{\eta}}\{log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + 1\}\ ; \qquad \zeta > 1 \tag{7}$$

where, $0 < x < \lambda$, $\Theta = (\zeta, \lambda, \eta)$,$\lambda > 0$ is a scale parameter and $\zeta > 1, \eta > 0$ are shape parameter.

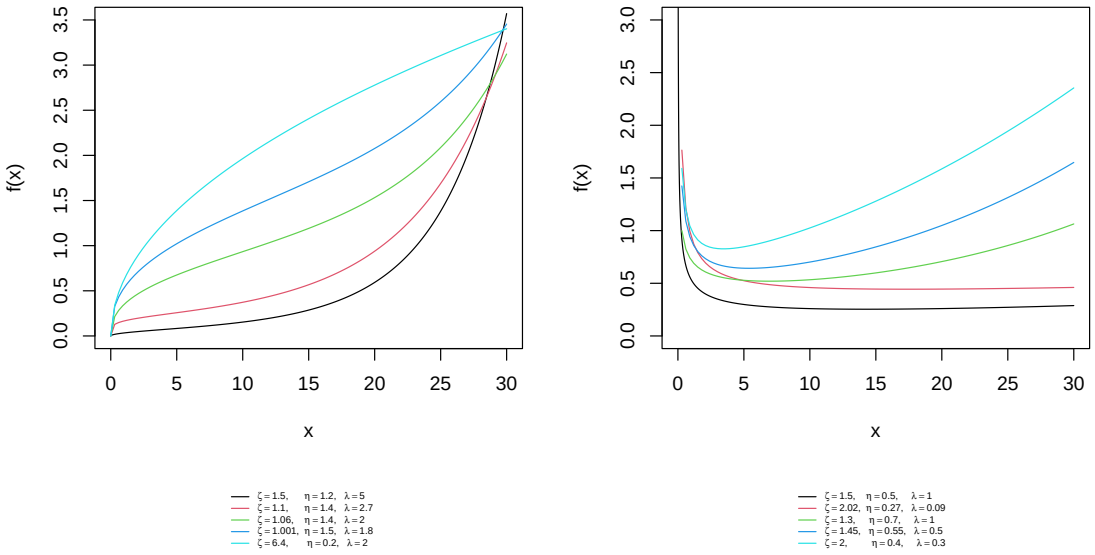


Figure 1: *Plots of the PNJ-PF density for various combinations of ζ and η and λ.*

Figure 1 depicts some different curves of the Pdf for different combination of PNJ-PF parameters ζ, η and λ. It is noted from figure (1) that the density curves for PNJ-PF distribution can be decreasing, decreasing-increasing, and increasing.
The survival function $S_{PNJ}(x,\Theta)$ and the hazard rate function $h_{PNJ}(x,\Theta)$ for $0 < x < \lambda$ are, respectively, given by

$$S_{PNJ}(x,\Theta) = \frac{1}{\zeta}\{1+\zeta-\left(\frac{x}{\lambda}\right)^{\eta} - e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}\}\ ; \qquad \zeta > 1 \tag{8}$$

$$h_{PNJ}(x,\Theta) = \frac{\eta x^{\eta-1}\{log(\zeta)e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}+1\}}{\lambda^{\eta}\{1+\zeta-\left(\frac{x}{\lambda}\right)^{\eta} - e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}\}}\ ;\ \zeta > 1 \tag{9}$$

The Plots of the hazard rate function of the PNJ-PF distribution for selected parameter values are displayed in Figure (2). It is noted that the PNJ-PF distribution possesses increasing, J-shaped, decreasing, constant ,and bathtub shape hazard rate function.

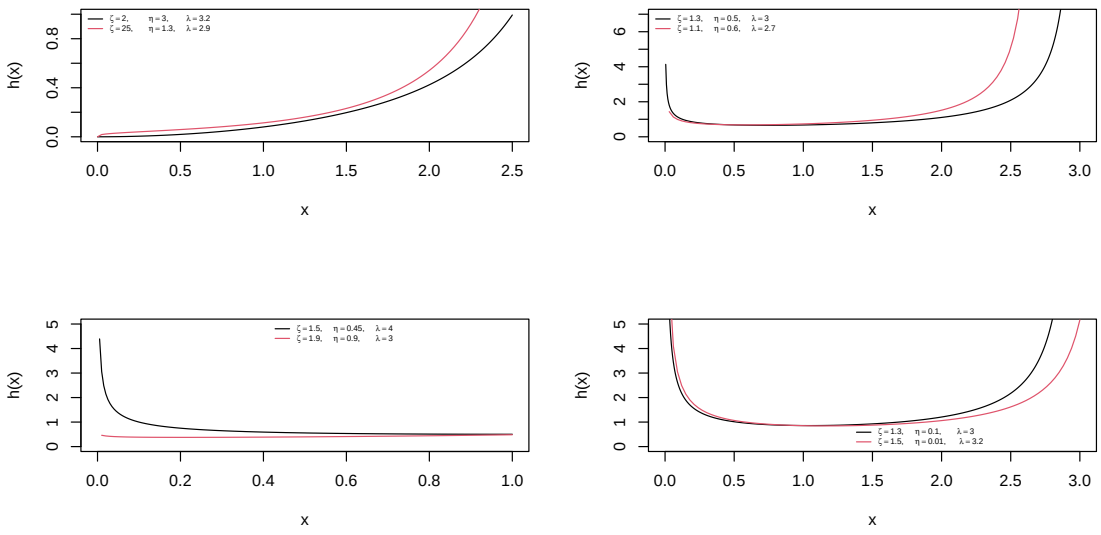


Figure 2: *Plots of the PNJ-PF hazard rate function for various combinations of ζ and η and λ.*

Colorary I: (Stochasting Ordering) The ratio of the densities of transformed variable X_{PNJ} and original random variable (i.e. PF X) is given by

$$R(x) = \frac{1}{\zeta}\{log(\zeta)e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}+1\} \tag{10}$$

and the first order derivative of R(x) is given by

$$R^{'}(x) = \frac{\eta x^{\eta-1}}{\zeta\lambda^{\eta}}\{log(\zeta)e^{(log(\zeta))^2\left(\frac{x}{\lambda}\right)^{\eta}}\} \tag{11}$$

This expression $R^{'}(x)$ is always positive for $x > 0, \zeta > 1, \eta > 0$ and $\lambda > 0$, which implies X_{PNJ} exhibits higher likelihood ratio than the original variable X i.e, $X_{PNJ} \geq_{lr} X$. Based on the chain of implications within various stochastic orders, it can concluded that $X_{PNJ} \geq_{hr} X$ and hence $X_{PNJ} \geq_{mlr} X$.

4. Statistical properties of PNJ-PF distribution

In this section, the essential probabilistic and statistical characteristics of the proposed model are presented.

4.1. Moments and associated measures

Theorem I: Let X follow the PNJ-PF distribution with parameters $\zeta > 1$,$\eta > 0$ and $\lambda > 0$; then, the r-th ordinary moment $E(x^r)$ of X has the form

$$E(X^r) = \frac{\eta\lambda^r}{\zeta}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(r+\eta(j+1))} + \frac{1}{\eta+r}\} \tag{12}$$

Proof:

$$E(X^r) = \int_0^{\lambda} x^r f_{PNJ}(x)dx$$

$$E(X^r) = \frac{\eta}{\zeta\lambda^n}\int_0^{\lambda} x^{r+\eta-1}\{log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + 1\}dx$$

After some algebra, the r-th ordinary moment of X reduces to

$$E(X^r) = \frac{\eta\lambda^r}{\zeta}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(r+\eta(j+1))} + \frac{1}{\eta+r}\}$$

Corollary II: The first and second ordinary moments can be obtained by substituting $r = 1, 2$, in (12), respectively. The expressions for the mean $E(X)$ and variance of PNJ-PF distribution are given, respectively, by

$$E(X) = \frac{\eta\lambda}{\zeta}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\}$$

and,

$$V(X) = \frac{\eta\lambda^2}{\zeta}\left[\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(2+\eta(j+1))} + \frac{1}{\eta+2}\} - \frac{\eta}{\zeta}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\}^2\right]$$

Theorem II: Let $X \sim PNJ - PF(\zeta, \eta, \lambda)$, then r-th incomplete moment $I_r(x)$ of X are

$$I_r(x) = \frac{\eta t^{\eta+r}}{\zeta\lambda^{\eta}}\{(\frac{t}{\lambda})^{\eta j}\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(r+\eta(j+1))} + \frac{1}{\eta+r}\} \tag{13}$$

Proof: The r-th incomplete moments $I_r(x)$ are defined as

$$I_r(x) = \int_0^{t} x^r f_{PNJ}(x)dx$$

$$I_r(x) = \int_0^{t} x^r\frac{\eta x^{\eta-1}}{\zeta\lambda^{\eta}}\{log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + 1\}dx$$

through algebraic simplification, the r-th incomplete moment $I_r(x)$ of X can be expressed as

$$I_r(x) = \frac{\eta t^{\eta+r}}{\zeta\lambda^{\eta}}\{(\frac{t}{\lambda})^{\eta j}\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(r+\eta(j+1))} + \frac{1}{\eta+r}\}$$

Corollary III: The first incomplete moments $I(x)$ can be obtained by substituting $r = 1$, in (18), as

$$I(x) = \frac{\eta t^{\eta+1}}{\zeta\lambda^{\eta}}\{(\frac{t}{\lambda})^{\eta j}\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\}$$

Theorem III: If $X \sim PNJ - PF(\zeta, \eta, \lambda)$, then the MGF $(M_X(t))$ of X is given by

$$M_X(t) = \frac{\eta}{\zeta\lambda^{\eta}}\sum_{k=0}^{\infty}\frac{\lambda^{\eta+k}t^k}{k!}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(k+\eta(j+1))} + \frac{1}{\eta+k}\} \tag{14}$$

Proof: The MGF $(M_X(t))$ is defined as

$$M_X(t) = \int_0^{\lambda} e^{tx} f_{PNJ}(x)dx$$

using 7, we have

$$M_X(t) = \int_0^{\lambda} e^{tx}\frac{\eta x^{\eta-1}}{\zeta\lambda^{\eta}}\{log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + 1\}dx$$

after simplifying by using the series expansion e^{ax} and some algebraic calculations, we get the final expression for MGF as

$$M_X(t) = \frac{\eta}{\zeta\lambda^{\eta}}\sum_{k=0}^{\infty}\frac{\lambda^{\eta+k}t^k}{k!}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(k+\eta(j+1))} + \frac{1}{\eta+k}\}$$

4.2. Mean residual life and mean waiting time

Theorem IV: Let $X \sim PNJ - PF(\zeta, \eta, \lambda)$, then The mean residual life function, say $\mu(t)$ of x , is given by

$$\mu(t) = \frac{\eta}{1+\zeta-(\frac{x}{\lambda})^{\eta}-e^{log(\zeta)(\frac{x}{\lambda})^{\eta}}}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))}(\lambda - \frac{t^{\eta(j+1)+1}}{\lambda^{\eta(j+1)}}) - \frac{1}{\eta+1}(\lambda - \frac{t^{\eta+1}}{\lambda^{\eta}})\} - t \tag{15}$$

Proof: The mean RLF is defined as

$$\mu(t) = \frac{1}{S_{PNJ}(t)}\left(E(t) - \int_0^t x f_{PNJ}(x)dx\right) - t \tag{16}$$

where

$$E(t) = \frac{\eta\lambda}{\zeta}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\} \tag{17}$$

and

$$\int_0^t x f_{PNJ}(x)dx = \frac{\eta t^{\eta+1}}{\zeta\lambda^{\eta}}\{(\frac{t}{\lambda})^{\eta j}\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\} \tag{18}$$

Substituting (8), (17) and (18) in (16), we get final expression for $\mu(t)$ as

$$\mu(t) = \frac{\eta}{1+\zeta-(\frac{t}{\lambda})^{\eta}-e^{log(\zeta)(\frac{t}{\lambda})^{\eta}}}\{\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))}(\lambda - \frac{t^{\eta(j+1)+1}}{\lambda^{\eta(j+1)}}) - \frac{1}{\eta+1}(\lambda - \frac{t^{\eta+1}}{\lambda^{\eta}})\} - t$$

Theorem V: Let $X \sim PNJ - PF(\zeta, \eta, \lambda)$, then the mean waiting time of X, say $\bar{\mu}(t)$, is

$$\bar{\mu}(t) = t - \frac{\eta t^{\eta+1}}{\lambda^{\eta}(e^{log(\zeta)(\frac{t}{\lambda})^{\eta}} + (\frac{t}{\lambda})^{\eta} - 1)}\{(\frac{t}{\lambda})^{\eta j}\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\} \tag{19}$$

Proof: the $\bar{\mu}(t)$, is defined as

$$\bar{\mu}(t) = t - \frac{1}{F_{PNJ}(t)}\int_0^t x f_{PNJ}(x)dx. \tag{20}$$

Substituting (6) and (18) in (20), we get

$$\bar{\mu}(t) = t - \frac{\eta t^{\eta+1}}{\lambda^{\eta}(e^{log(\zeta)(\frac{t}{\lambda})^{\eta}} + (\frac{t}{\lambda})^{\eta} - 1)}\{(\frac{t}{\lambda})^{\eta j}\sum_{j=0}^{\infty}\frac{(log(\zeta))^{j+1}}{j!(1+\eta(j+1))} + \frac{1}{\eta+1}\}$$

4.3. Entropy

Theorem VI: Let $X \sim PNJ - PF(\zeta, \lambda, \eta)$, then the Renyi entropy of X is

$$H_R(x) = \frac{1}{1-\nu}log\left[\left(\frac{\eta}{\zeta\lambda^{\eta}}\right)^{\nu}\frac{\lambda^{\nu(\eta-1)}}{\nu(\eta-1)+1}\sum_{k=0}^{\infty}({}^{\nu}C_k)log^k(\zeta)\sum_{j=0}^{\infty}\frac{(klog(\zeta))^j}{j!}\right]; \qquad \nu > 0, \nu \neq 1 \tag{21}$$

Proof: The Renyi entropy of X is defined as

$$H_R(x) = \frac{1}{1-\nu}log\int_0^{\lambda} f_{PNJ}^{\nu}(x)dx$$

using (7) ,we have

$$H_R(x) = \frac{1}{1-\nu}log\int_0^{\lambda}(\frac{\eta x^{\eta-1}}{\zeta\lambda^{\eta}})^{\nu}\{log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + 1\}^{\nu}dx$$

after integrating, the Renyi entropy reduces to

$$H_R(x) = \frac{1}{1-\nu}log\left[\left(\frac{\eta}{\zeta\lambda^{\eta}}\right)^{\nu}\frac{\lambda^{\nu(\eta-1)}}{\nu(\eta-1)+1}\sum_{k=0}^{\infty}({}^{\nu}C_k)log^k(\zeta)\sum_{j=0}^{\infty}\frac{(klog(\zeta))^j}{j!}\right]$$

Collorary IV: The renyi entropy (21) is useful for computing the entropy measures of Havrda and Charvat $H_H(x)$, as well as Arimoto entropy $H_A(x)$. And the final expressions for these entropy measures are respectively given by

$$H_H(x) = \frac{1}{1-\nu}\left[\left(\frac{\eta}{\zeta\lambda^{\eta}}\right)^{\nu}\frac{\lambda^{\nu(\eta-1)}}{\nu(\eta-1)+1}\sum_{k=0}^{\infty}({}^{\nu}C_k)log^k(\zeta)\sum_{j=0}^{\infty}\frac{(klog(\zeta))^j}{j!} - 1\right]$$

and

$$H_A(x) = \frac{1}{2^{\nu-1}-1}\left[\left(\frac{\eta}{\zeta\lambda^{\eta}}\right)^{\frac{1}{\nu}}\frac{\nu\lambda^{\frac{(\eta-1)}{\nu}}}{(\eta-1)+\nu}\sum_{k=0}^{\infty}({}^{\frac{1}{\nu}}C_k)log^k(\zeta)\sum_{j=0}^{\infty}\frac{(klog(\zeta))^j}{j!} - 1\right]$$

4.4. Order Statistics

Let $X_1, X_2, ..., X_m$ be a random sample of size m from NPJ-PF, and let $X_{r:m}$ denote the rth order statistic, then, the pdf of $X_{r:m}$, say $f_{r:m}(x)$ is given by

$$f_{r:m}(x) = \frac{m!}{(r-1)!(m-r)!} F_{PNJ}(x)^{r-1} f_{PNJ}(x)(1 - F_{PNJ}(x))^{m-r}. \tag{22}$$

Substituting (6) and (7) in (22), we get

$$f_{r:m}(x) = \frac{m!\eta x^{\eta-1\{log(\zeta)e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}+1\}}}{(r-1)!(m-r)!\lambda^{\eta}\zeta^{m}} \left[\{e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + \left(\frac{x}{\lambda}\right)^{\eta} - 1\}\right]^{r-1} \left[\{1+\zeta - \left(\frac{x}{\lambda}\right)^{\eta} - e^{log(\zeta)(\frac{x}{\lambda})^{\eta}}\}\right]^{m-r} \tag{23}$$

The minimum and maximum OS densities are obtained, respectively, by substituting $r = 1$ and $r = m$ in (23), and the expressions are respectively given by

$$f_{1:m}(x) = \frac{m\eta x^{\eta-1\{log(\zeta)e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}+1\}}}{\lambda^{\eta}\zeta^{m}} \left[\{1+\zeta - \left(\frac{x}{\lambda}\right)^{\eta} - e^{log(\zeta)(\frac{x}{\lambda})^{\eta}}\}\right]^{m-1}$$

and

$$f_{m:m}(x) = \frac{m\eta x^{\eta-1\{log(\zeta)e^{log(\zeta)\left(\frac{x}{\lambda}\right)^{\eta}}+1\}}}{\lambda^{\eta}\zeta^{m}} \left[\{e^{log(\zeta)(\frac{x}{\lambda})^{\eta}} + \left(\frac{x}{\lambda}\right)^{\eta} - 1\}\right]^{m-1}$$

4.5. Stress Strength Reliability

Theorem VII: let X_1 and X_2 be independent strength and stress random variables respectively, where $X_1 \sim PNJ - PF(\zeta_1, \eta_1, , \lambda)$ and $X_2 \sim PNJ - PF(\zeta_2, \eta_2, , \lambda)$, then the stress strength reliability defined as the probability that the strength X_1 exceeds the stress X_2 $\mathbb{P}(X_1 > X_2)$, say SSR, is

$$\begin{aligned} SSR = \frac{\eta_1}{\zeta_1\zeta_2} \Bigg[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{(log(\zeta_1))^{j+1}(log(\zeta_2))^{k}}{j!k!(\eta_1(j+1)+\eta_2 k)} - \sum_{l=0}^{\infty} \frac{(log(\zeta_1))^{l+1}}{l!(\eta_1(l+1))} \\ + \sum_{m=0}^{\infty} \frac{(log(\zeta_1))^{m+1}}{m!(\eta_1(m+1)+\eta_2)} + \sum_{n=0}^{\infty} \frac{(log(\zeta_2))^{n}}{n!(\eta_1+\eta_2 n)} + \frac{1}{\eta_1+\eta_2} \end{aligned} \tag{24}$$

Proof: The stress strength reliability $\mathbb{P}(X_1 > X_2)$, say SSR, is defined as

$$SSR = \int_0^{\lambda} f_1(x) F_2(x) dx$$

using (6) , (7), we have

$$SSR = \frac{\eta_1}{\zeta_1\zeta_2(\lambda)^{\eta_1}} \int_0^{\lambda} \left[x^{\eta_1-1}\{log(\zeta_1)e^{log(\zeta_1)\left(\frac{x}{\lambda}\right)^{\eta_1}}+1\}\right] \left[e^{log(\zeta_2)(\frac{x}{\lambda})^{\eta_2}} + \left(\frac{x}{\lambda}\right)^{\eta_2} - 1\right] dx$$

after simplifying and using series expansion e^{ax} we get the final expression for the stress strength reliability SSR as

$$\begin{aligned} SSR = \frac{\eta_1}{\zeta_1\zeta_2(\lambda)^{\eta_1}} \Bigg[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{(log(\zeta_1))^{j+1}(log(\zeta_2))^{k}}{j!k!(\lambda)^{\eta_1 j}(\lambda)^{\eta_2 k}} \frac{(\lambda)^{\eta_1(j+1)+\eta_2 k}}{\eta_1(j+1)+\eta_2 k} - \sum_{l=0}^{\infty} \frac{(log(\zeta_1))^{l+1}}{l!(\lambda)^{\eta_1 l}} \frac{(\lambda)^{\eta_1(l+1)}}{\eta_1(l+1)} \\ + \sum_{m=0}^{\infty} \frac{(log(\zeta_1))^{m+1}}{m!(\lambda)^{\eta_1 m}(\lambda)^{\eta_2}} \frac{(\lambda)^{\eta_1(m+1)+\eta_2}}{\eta_1(m+1)+\eta_2} + \sum_{n=0}^{\infty} \frac{(log(\zeta_2))^{n}}{n!(\lambda)^{\eta_2 n}} \frac{(\lambda)^{\eta_1+\eta_2 n}}{\eta_1+\eta_2 n} + \frac{1}{(\lambda)^{\eta_2}} \frac{\lambda^{\eta_1+\eta_2}}{\eta_1+\eta_2} \end{aligned}$$

$$SSR = \frac{\eta_1}{\zeta_1\zeta_2}\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(log(\zeta_1))^{j+1}(log(\zeta_2))^k}{j!k!(\eta_1(j+1)+\eta_2 k)} - \sum_{l=0}^{\infty}\frac{(log(\zeta_1))^{l+1}}{l!(\eta_1(l+1))} + \sum_{m=0}^{\infty}\frac{(log(\zeta_1))^{m+1}}{m!(\eta_1(m+1)+\eta_2)}\right.$$
$$\left. + \sum_{n=0}^{\infty}\frac{(log(\zeta_2))^n}{n!(\eta_1+\eta_2 n)} + \frac{1}{\eta_1+\eta_2}\right.$$

5. Statistical Inference

5.1. Parameter Estimation

Let $X_1, X_2, ..., X_m$ be a random sample from PNJ-PF distribution, then the logarithm of the likelihood function is

$$l = mlog\eta - mlog(\zeta) - m\eta log(\lambda) + (\eta-1)\sum_{i=1}^{m} log(x_i) + mlog\left[log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^\eta} + 1\right] \tag{25}$$

The MLEs of ζ, λ and η are obtained by partially differentiating (25) with respect to the corresponding parameters and equating to zero, we have

$$\frac{\partial l}{\partial \zeta} = \frac{-m}{\zeta} + \frac{me^{log(\zeta)(\frac{x}{\lambda})^\eta}(log(\zeta)(\frac{x}{\lambda})^\eta + 1)}{\zeta\left[(log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^\eta} + 1\right]} \tag{26}$$

$$\frac{\partial l}{\partial \eta} = \frac{m}{\eta} - mlog\lambda + \sum_{i=1}^{m} log(x_i) + \frac{mx^\eta(log(\zeta))^2 e^{log(\zeta)(\frac{x}{\lambda})^\eta} log(x)}{\lambda^\eta\{(log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^\eta} + 1\}} \tag{27}$$

$$\frac{\partial l}{\partial \lambda} = -\frac{m\eta}{\lambda} - \frac{m\eta(log(\zeta))^2 e^{log(\zeta)(\frac{x}{\lambda})^\eta} x^\eta}{\lambda^{\eta+1}\left[(log(\zeta)e^{log(\zeta)(\frac{x}{\lambda})^\eta} + 1\right]} \tag{28}$$

The above three equations (26),(27) and (28) are not in closed form. Thus, it is difficult to calculate the values of the parameters ζ, η and λ. However, R software can be used to get the MLE.

5.2. Simulation study

A simulation study was conducted using R Software to examine the behavior of the Maximum Likelihood Estimates (MLEs) with varying sample sizes. Three sets of samples (n=25, n=100, and n=500), each replicated 100 times, were generated from the PNJ-PF distribution with different parameter values $\zeta = (1.5, 2)$ $\lambda = (1.5, 2)$, and $\eta = (1.5, 2)$, to effectively check the impact of small $(n = 25)$, medium $(n = 100)$, and large $(n = 500)$ sample sizes on the accuracy and precision of the MLEs, demonstrating improved performance with larger sample sizes. For each configuration, the average MLEs and their corresponding mean squared errors (MSEs) were computed. The results are summarized in Tables 1.

Table 1: *Average values of MLEs and the corresponding MSEs.*

	Parameter			MLE			MSE		
n	ζ	η	λ	$\hat{\zeta}$	$\hat{\eta}$	$\hat{\lambda}$	$\hat{\zeta}$	$\hat{\eta}$	$\hat{\lambda}$
25	1.5	1	1.5	1.4413	0.8666	1.3183	0.9136	0.1296	0.2748
100				1.4621	0.9889	1.4822	0.0992	0.0111	0.0233
500				1.4778	0.9907	1.4845	0.0378	0.0103	0.0228
25			2	1.3489	0.7429	1.5336	1.1309	0.2406	0.9311
100				1.4444	0.933	1.8812	0.2646	0.0625	0.2434
500				1.4963	0.9864	1.9793	0.0515	0.0105	0.0407
25		1.5	1.5	1.4128	1.3403	1.3610	0.4165	0.2194	0.2053
100				1.4768	1.4287	1.4394	0.1824	0.0963	0.0911
500				1.4961	1.483	1.4856	0.0497	0.0245	0.0228
25			2	1.7193	1.2146	1.727	1.9832	0.3490	0.5283
100				1.4868	1.4192	1.9175	0.2886	0.1039	0.1624
500				1.5197	1.4926	1.9991	0.0374	0.0023	0.0001
25	2	1	1.5	1.8251	0.9219	1.3773	0.7054	0.0888	0.1833
100				1.9543	0.9816	1.4667	0.1491	0.0214	0.0457
500				1.9940	0.9887	1.4848	0.0561	0.0104	0.0228
25			2	2.1983	0.9447	2.1166	0.2833	0.0088	0.0378
100				2.0079	0.9652	1.9369	0.2614	0.0325	0.1221
500				2.0035	0.9884	1.9802	0.0686	0.0106	0.0406
25		1.5	1.5	1.8863	1.3938	1.3909	0.7179	0.1897	0.1598
100				1.9681	1.4936	1.4829	0.1917	0.0342	0.0229
500				1.9916	1.5017	1.5001	0.0227	0.0020	0.0000
25			2	2.0949	1.4191	1.9156	1.7986	0.1553	0.1645
100				2.0333	1.4422	1.9402	0.4624	0.0874	0.1218
500				2.0258	1.4913	1.9983	0.0347	0.0034	0.0001

The results of 1 shows that the MLEs are stable and closely approximate the true parameter values. Additionally, as the sample size increases, the MSE consistently decreases across all scenarios, indicating enhanced estimate reliability with larger sample sizes.
Additionally, a random sample of 51 observations has been generated from the PNJ-PF distribution with parameters $\zeta = 15$, $\eta = 1$, and $\lambda = 20$ using R programming language.This sample representing the quantiles of our proposed model, serves to demonstrate theoretical concepts and to compare the fit of the proposed model with baseline and several competitive models. The results are displayed in Table 2 and Table 3 and the resulting simulated dataset, along with its corresponding R code, is provided below.

```
> Data<-function(n,m,zeta,eta,lambda)
+ {set.seed(0)
+ library(zipfR)
+ cdf<-function(x,zeta,eta,lambda)
+ {fn<-((1/(zeta))*(exp(log(zeta)*((x/(lambda))^(eta)))-{1-(x/(lambda))^(eta)}))}
+ data=c() + U=runif(n,0,1)
+ for(i in 1:length(U)){
+ fn<-function(x){cdf(x,zeta,eta,lambda)-U[i]}
+ uni<-uniroot(fn,c(0,100000))
+ data=c(data,uni$root)}
+ return(data)}
> Simulateddata<-Data(51,1,15,1,20)
```

> Simulateddata

[1] 19.216206	10.997038	13.139771	16.053109	19.307662	9.370801	19.229723
[8] 19.590336	17.049291	16.705115	4.028329	9.491038	8.629201	17.322909
[15] 13.347840	18.127826	15.085951	17.630277	19.941461	13.277794	18.197640
[22] 19.514095	9.660749	16.951754	6.882131	11.036435	13.382222	1.028522
[29] 13.318377	18.997202	12.559603	14.868757	16.369590	15.028740	8.922347
[36] 18.640806	17.130323	18.349609	6.183355	17.690021	13.800205	18.585186
[43] 16.901957	18.247629	15.809617	15.512915	18.305730	1.727778	14.800077
[50] 17.773593	17.381194					

Table 2: *Maximum Likelihood Estimates (with standard errors in parentheses) for simulated data set.*

Model	Estimates $\hat{\zeta}$	$\hat{\eta}$	$\hat{\lambda}$
PNJ-PF	29.01068 (0.17446)	0.99048 (0.17443)	19.94143 (0.00003)
ZTP-PF	0.01308 (0.23971)	2.38893 (0.14642)	19.94143 (0.09328)
EP	3.40437 (0.12304)	0.69908 (0.12318)	19.94143 (0.00011)
EPF	5.18796 (0.23553)	1.15104 (0.23563)	19.94171 (0.00010)
PF	-	2.37968 (NaN)	19.94143 (0.00001)

Table 3: *Comparison of PNJ-PF Distribution with other competitive models for simulated data set.*

Model	$-2ll$	AIC	BIC	AICC	HQIC
PNJ-PF	270.8637	276.8637	282.6592	277.3744	279.0783
ZTP-PF	276.0330	282.0330	287.8285	282.5436	284.2476
EP	275.9709	281.9709	287.7664	282.4816	284.1855
EPF	271.3136	277.3136	283.1090	277.8242	279.5282
PF	275.9709	279.9709	283.8346	280.2209	281.4473

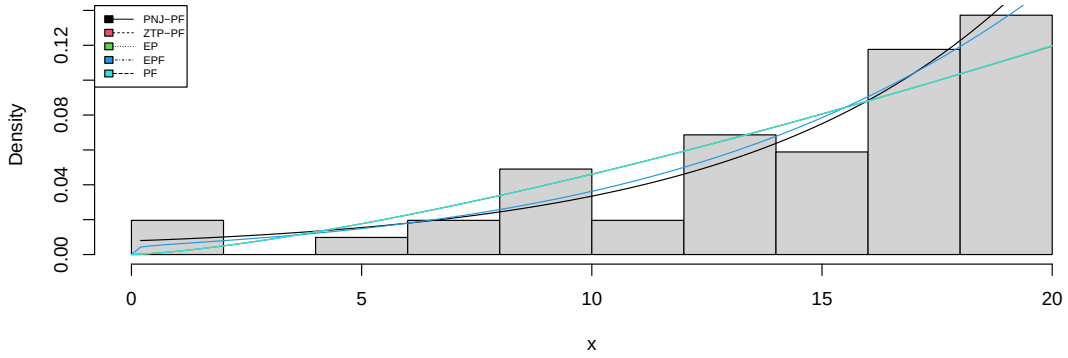


Figure 3: *The relative histogram and the fitted density functions of PNJ-PF and other competing distributions for simulated data set.*

The results of simulated dataset presented in Table 2 and Table 3,clearly demonstrate that the PNJ-PF distribution exhibits the lowest $-2ll$, AIC, BIC, AICC, and HQIC values among all the other competitive models and base line model.These findings are further supported by figure 3. Consequently, our proposed model offers a superior fit and outperforms base model of PF distribution as well as other mentioned competing models.

6. Applications

In this section, we explore two datasets related to electrical engineering to highlight the relevance and versatility of the PNJ-PF distribution. This analysis will demonstrate how the PNJ-PF distribution can be effectively utilized to model and interpret various types of data within the engineering field, illustrating its broad applicability and effectiveness. The data set first represents the times of 30 electronic components exposed to power-line voltage spikes during electric storms published first by [10] and is given as follows: 275,13, 147, 23, 181, 30, 65, 10, 300, 173, 106, 300, 300, 212, 300, 300, 300, 2,307 ,261, 293, 88, 247, 28, 143, 300, 23, 300, 80, 245, 266.
The second data set represents the failure times of first 50 electronic devices. which was originally published by [1], The data is given as follows: 0.1, 0.2, 1.0, 1.0, 1.0, 1.0, 1.0, 2.0, 3.0, 6.0, 7.0, 11.0, 12.0, 18.0, 18.0, 18.0, 18.0, 18.0, 21.0, 32.0, 36.0, 40.0, 45.0, 45.0, 47.0, 50.0, 55.0, 60.0, 63.0, 63.0, 67.0, 67.0, 67.0, 67.0, 72.0, 75.0, 79.0, 82.0, 82.0, 83.0, 84.0, 84.0, 84.0, 85.0, 85.0, 85.0, 85.0, 85.0, 86.0, 86.0.
We compare the fit of the proposed PNJ-PF distribution with its base-model (two parameter Power function (PF) distribution) and with several more related competitive models, namely Exponentiated Power function(EPF) [2], Zero Truncated Poisson Power function(ZTP-PF)[13] and Exponentiated Power(EP) [4] Distribution, their corresponding density functions for $0 < x < \lambda$ are as follows

$$\text{ZTP-PF} \qquad f(x) = \frac{\zeta\eta x^{\eta-1}\exp\left(-\zeta\left(\frac{x}{\lambda}\right)^{\eta}\right)}{\lambda^{\eta}\left(\exp\left(-\zeta\left(\frac{x}{\lambda}\right)^{\eta}\right)\right) - \exp(-\zeta)}$$

$$\text{EPF} \qquad f(x) = \frac{\zeta\eta}{\lambda^{\eta}}\frac{1}{x^{1-\eta}}\left(1+\left(\frac{x}{\lambda}\right)^{\eta}\right)^{-(1-\zeta)}$$

$$\text{EP} \qquad f(x) = \frac{\zeta\eta x^{\zeta\eta-1}}{\lambda^{\zeta\eta}}$$

For comparison, we use criterion of -2log-likelihood(-2ll), along with various information criteria (AIC, BIC,AICc, HQIC). And the results are displayed in Table 4, Table 5, Table 6 and Table 7.

Table 4: *Maximum Likelihood Estimates (with standard errors in parentheses) for first data set.*

Model	Estimates $\hat{\zeta}$	$\hat{\eta}$	$\hat{\lambda}$
PNJ-PF	9.88183	0.67367	307.01898
	(0.10531)	(0.15274)	(0.04681)
ZTP-PF	0.00000	1.01032	307.00001
	(0.00047)	(0.18195)	(0.18121)
EP	1.09892	0.80050	307.00001
	(0.39168)	(0.28722)	(0.10448)
EPF	3.79113	0.66570	307.00001
	(0.15531)	(0.16302)	0.00719
PF	-	1.00657	307.00001
		(NaN)	(0.00173)

Table 5: *Comparison of PNJ-PF Distribution with other competitive models for first data set.*

Model	$-2ll$	AIC	BIC	AICC	HQIC
PNJ-PF	349.4228	355.4228	359.7247	356.3116	356.8251
ZTP-PF	355.0636	361.0636	365.3656	361.9525	362.4660
EP	355.6021	361.6021	365.9040	362.4909	363.0044
EPF	351.7101	357.7101	362.0120	358.5989	359.1124
PF	355.0632	359.0632	361.9312	359.4918	359.9981

Table 6: *Maximum Likelihood Estimates (with standard errors in parentheses) for second dataset.*

Model	Estimates $\hat{\zeta}$	$\hat{\eta}$	$\hat{\lambda}$
PNJ-PF	9.884284	0.478196	86.000000
	(0.078434)	(0.084724)	(0.006096)
ZTP-PF	0.000001	0.739749	86.000003
	0.021726	0.102805	0.081023
EP	0.998401	0.728003	86.000003
	(0.379407)	(0.379788)	0.000197
EPF	3.820468	0.472193	86.000001
	(0.168872)	(0.093803)	(0.075191)
PF	-	0.726714	86.000002
		(NaN)	(0.001117)

Table 7: *Comparison of PNJ-PF Distribution with other competitive models for second data set.*

Model	$-2ll$	AIC	BIC	AICC	HQIC
PNJ-PF	431.2207	437.2207	442.9568	437.7424	439.4050
ZTP-PF	439.7738	445.7738	451.5098	446.2955	447.9581
EP	439.7582	445.7582	451.4943	446.2799	447.9425
EPF	434.5294	440.5294	446.2654	441.0511	442.7137
PF	439.7582	443.7582	447.5822	444.0135	445.2144

From Table 4, Table 5, Table 6 and Table 7, it is clearly evident that PNJ-PF distribution has lowest $-2ll$, AIC, BIC, AICC and HQIC values among all the other competitive models and base line model. Therefore provide superior fit and outperforms base model of PF distribution as well as other mentioned competing models.

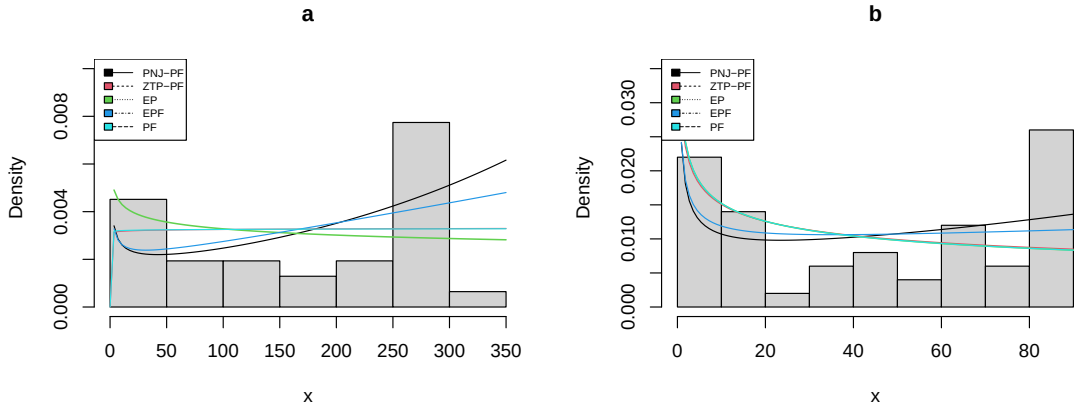


Figure 4: *(a) The relative histogram and the fitted density functions of PNJ-PF and other competing distributions for data set first. (b) The relative histogram and the fitted density functions of PNJ-PF and other competing distributions for data set second.*

The relative histogram and the fitted density functions of PNJ-PF and other competing distributions of the data set first and second are shown in Figures 4. This graphical representation clearly validate the results in Tables 4, Table 5, Table 6 and Table 7.

7. Conclusion

This manuscript introduces an innovative and versatile method known as the PNJ method for generating probability distributions. The PNJ method is specifically tailored to the two-parameter Power function (PF) distribution, resulting in the development of a new three-parameter PNJ-PF distribution. The paper thoroughly examines the various statistical and reliability characteristics of the PNJ-PF model, emphasizing its adaptable and flexible shapes for both density and hazard functions. To illustrate the PNJ-PF model's effectiveness, the study applies it to a simulated and two real-world datasets and conducts a comprehensive comparison with base model and other competing models using goodness-of-fit analysis. The findings clearly demonstrate that

the PNJ-PF model outperforms base model and all other competing models in these datasets, showcasing its superior performance and effectiveness. Also by offering a novel approach that significantly enhances the accuracy and reliability of hazard rate modeling, this manuscript positions the PNJ-PF model as an essential tool for researchers and practitioners in engineering field. The innovative contributions of this study have the potential to bring about substantial advancements in the field, enabling more precise and effective decision-making based on strong statistical foundations.

References

[1] Magne Vollan Aarset. How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, R-36(1):106–108, 1987.

[2] Alya Al Mutairi, Muhammad Z Iqbal, Muhammad Z Arshad, Ahmed Z Afify, et al. A new class of the power function distribution: Theory and inference with an application to engineering data. *Journal of Mathematics*, 2022, 2022.

[3] Muhammad Zeshan Arshad, Muhammad Zafar Iqbal, and Munir Ahmad. Exponentiated power function distribution: Properties and applications. *Journal of Statistical Theory and Applications*, 19:297–313, 2020.

[4] Subramanian C and Aafaq Rather. Exponentiated power distribution: Properties and estimation. *Science, Technology and Development*, VIII:213–223, 09 2019.

[5] AC Dallas. Characterizing the pareto and power distributions. *Annals of the Institute of Statistical Mathematics*, 28:491–497, 1976.

[6] Eslam Hussam, Laxmi Prasad Sapkota, and Ahmed M Gemeay. Tangent exponential-g family of distributions with applications in medical and engineering. *Alexandria Engineering Journal*, 105:181–203, 2024.

[7] Carl Lee, Felix Famoye, and Ayman Y Alzaatreh. Methods for generating families of univariate continuous distributions in the recent decades. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(3):219–238, 2013.

[8] Murtiza Ali Lone, Ishfaq Hassain Dar, and TR Jan. An innovative method for generating distributions: Applied to weibull distribution. *Journal of Scientific Research*, 66(3), 2022.

[9] Abbas Mahdavi and Debasis Kundu. A new method for generating distributions with an application to exponential distribution. *Communications in Statistics-Theory and Methods*, 46(13):6543–6557, 2017.

[10] WQ Meeker and LA Escobar. Statistical methods for reliability data john wiley & sons new york. *New York*, 1998.

[11] M Meniconi and DM Barry. The power function distribution: A useful and simple distribution to assess electrical component reliability. *Microelectronics Reliability*, 36(9):1207–1212, 1996.

[12] Govind S Mudholkar and Deo Kumar Srivastava. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, 42(2):299–302, 1993.

[13] Idika Eke Okorie, Anthony Chukwudi Akpanta, Johnson Ohakwe, David Chidi Chikezie, Chris Uche Onyemachi, and Manoj Kumar Rastogi. Zero-truncated poisson-power function distribution. *Annals of Data Science*, 8:107–129, 2021.

[14] Shamshad Ur Rasool, MA Lone, and SP Ahmad. An innovative technique for generating probability distributions: A study on lomax distribution with applications in medical and engineering fields. *Annals of Data Science*, pages 1–17, 2024.

[15] William T Shaw and Ian RC Buckley. The alchemy of probability distributions: beyond gram-charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv preprint arXiv:0901.0434*, 2009.

[16] MH Tahir, Morad Alizadeh, M Mansoor, Gauss M Cordeiro, and M35248241376 Zubair. The weibull-power function distribution with applications. *Hacettepe Journal of Mathematics and Statistics*, 45(1):245–265, 2016.

[17] Muhammad Ahsan ul Haq, M Elgarhy, Sharqa Hashmi, Gamze Ozel, and Qurat ul Ain. Transmuted weibull power function distribution: Its properties and applications. *Journal of Data Science*, 16(2):397–418, 2018.

[18] Rana Usman, Muhammad Ahsan Ul Haq, Nurbanu Bursa, and Gamze Özel. Exponentiated transmuted power function distribution: Theory & applications. *Gazi University Journal of Science*, 31(2):660–675, 2018.

NEW GENERALIZATION OF INVERTED EXPONENTIAL DISTRIBUTION: PROPERTIES AND ITS APPLICATIONS

TABASUM AHAD[1], S.P.AHMAD[2]

•

[1],[2] Department of Statistics,University of Kashmir, Srinagar, India
tabasumjan1234@gmail.com, sprvz@yahoo.com

Abstract

In this paper, we introduce a new extension of the inverted exponential distribution called as "SMP Inverted Exponential" (SMPIE) distribution through the SMP technique. Various statistical properties of this new distribution have been illustrated, including survival function, hazard function, quantile function, moments, moment generating function, entropy, and order statistics. Method of maximum likelihood estimation is used to evaluate the parameters of the proposed distribution. A simulation study is carried out for illustration of the performance of estimates. Two real-life data sets are incorporated to illustrate the utility and flexibility of the proposed distribution as compared to other existing probability distributions.

Keywords: SMP transformation, inverted exponential, order statistics, Maximum likelihood estimation.

1. Introduction

In Probability theory and statistics, the exponential distribution is a continuous probability distribution that describes the time between events in the Poisson process i.e. a process in which events occur independently and continuously at a constant failure rate. The exponential distribution possesses the constant and bathtub hazard rates. But in real-life problems, there may be a situation where the data shows the inverted bathtub hazard rate (initially increases and then decreases, i.e., uni-modal). So, the exponential distribution becomes unfit for modeling real-life situations. For such data types, another extension of the exponential distribution has been put forward in the statistical literature. That is known as inverted exponential distribution (IED) which possesses the inverted bathtub hazard rate. The inverted exponential distribution was introduced by [5] and has been used by [7] as a lifetime model. It is widely used in biology, medicine, and engineering. Several authors have proposed distributions using inverted exponential (IE) distribution in Statistical literature. The generalized inverted exponential distribution was proposed by [1]. The Bayes estimators of the parameter and reliability function of inverted exponential distribution were obtained by [15]. Exponentiated generalized inverted exponential distribution was proposed by [11]. The exponentiated inverted exponential distribution was proposed by [4]. The alpha power inverted exponential distribution was introduced by [16]. A new Weibull inverted IE distribution was obtained by [3]. The Weibull inverse exponential[Loglogistic] distribution was derived by [8]. New Sine Inverted Exponential distribution was proposed by [2]. The probability density function (PDF) of the IE distribution is given as follows:

$$f(x) = \frac{\lambda}{x^2} e^{\frac{-\lambda}{x}}; \quad x > 0, \lambda > 0, \tag{1}$$

and corresponding cumulative distribution function (CDF) as follows:

$$F(x) = e^{\frac{-\lambda}{x}}; \quad x > 0, \lambda > 0. \tag{2}$$

Where λ is a scale parameter and is greater than 0.
In this paper, we proposed a new extension of inverted exponential (IE) by using the SMP technique. The proposed model is named as SMP inverted exponential distribution (SMPIE). The primary motivation for considering SMP Inverted exponential (SMPIE) distribution is that the proposed model is very efficient and flexible for introducing a new parameter to generalize the existing distributions. The additional parameter can give various desirable properties and is more flexible in the form of hazard and density functions and demonstrates a superior fit compared to other competing models. The rest of this paper is organized as follows: In section 2, we defined the SMP Transformation. In section 3, the PDF and CDF of the proposed model i.e. SMPIE are defined. Section 4 includes the Reliability measures of the model. in section 5, we derive some mathematical properties. Order statistics are investigated in section 6. Maximum likelihood estimation of the model parameters is addressed in section 7. The simulation study and the applicability of the model is discussed in section 8 and 9 respectively. Finally, a concluding remark are addressed in Section 10.

2. SMP Transformation

The SMP transformation was recently proposed by [13] whose CDF and PDF are given by the following equations respectively.

$$G_{SMP}(x) = \begin{cases} \frac{e^{\log(\alpha)\overline{F}(x)} - \alpha}{1-\alpha}, & \alpha \neq 1, \alpha > 0 \\ F(x), & \alpha = 1 \end{cases} \tag{3}$$

where $\overline{F}(x) = 1 - F(x)$
$G_{SMP}(x)$ is a valid CDF, if F(x) is a valid CDF. This is because it satisfies the following properties:

(i) $G_{SMP}(-\infty) = 0$; $G(\infty) = 1$

(ii) $G_{SMP}(x)$ is monotonic increasing function of x

(iii) $G_{SMP}(x)$ is right continous

(iv) $0 \leq G_{SMP}(x) \leq 1$

For $x \in \mathbb{R}$, the PDF of SMP transformation is given as follows:

$$g_{SMP}(x) = \begin{cases} \frac{e^{\log(\alpha)\overline{F}(x)} \log(\alpha) f(x)}{\alpha - 1}, & \alpha \neq 1, \alpha > 0 \\ f(x) \quad , & \alpha = 1 \end{cases} \tag{4}$$

where F(x) and f(x) are the CDF and PDF of the baseline distribution respectively.

3. SMPIE Distribution

A random variable X is said to follow SMPIE distribution with scale parameter $\lambda > 0$ and shape parameter $\alpha > 0$, if its CDF is given as

$$G_{SMPIE}(x; \alpha, \lambda) = \begin{cases} \frac{e^{(\log \alpha)(1 - e^{\frac{-\lambda}{x}})} - \alpha}{1-\alpha}, & \alpha \neq 1, \alpha > 0 \\ e^{\frac{-\lambda}{x}} \quad , & \alpha = 1 \end{cases} \tag{5}$$

Figure (1) displays the CDF plot of the SMPIE distribution for different parameter values of α and λ.

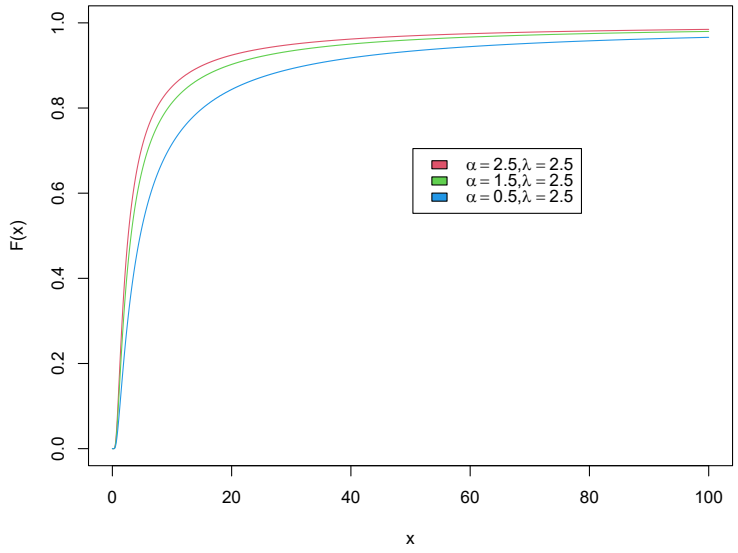


Figure 1: *CDF plot for SMPIE distribution*

The corresponding PDF of SMPIE distribution is given as

$$g_{SMPIE}(x;\alpha,\lambda) = \begin{cases} \frac{\log(\alpha)}{\alpha-1}\frac{\lambda}{x^2}e^{(\log\alpha)(1-e^{\frac{-\lambda}{x}})}e^{\frac{-\lambda}{x}}, & \alpha \neq 1, \alpha > 0 \\ \frac{\lambda}{x^2}e^{\frac{-\lambda}{x}} \quad , & \alpha = 1 \end{cases} \tag{6}$$

Figure (2) displays the pdf plot of the SMPIE distribution for different parameter values.

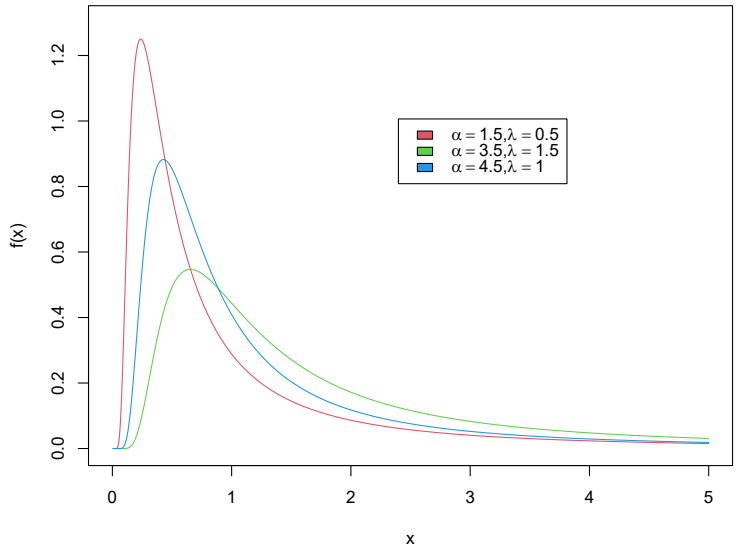


Figure 2: *PDF plot for SMPIE distribution*

4. Relaiability analysis

In this section, we obtain the reliability (survival function), hazard rate (failure rate), reverse hazard function and mills ratio expressions for SMPIE.

4.1. Reliability function

The reliability function is the probability that an item does not fail before time say x and for SMPIE distribution, it is given as

$$R(x;\alpha,\lambda) = 1 - G_{SMPIE}(x;\alpha,\lambda) = \frac{1 - e^{(\log\alpha)(1-e^{\frac{-\lambda}{x}})}}{1-\alpha}, \quad \alpha \neq 1 \tag{7}$$

4.2. Hazard Rate

The hazard rate or failure rate accesses the likelihood of component's failure failure based on how long it has already been in use. Consequently, it has various applications in the analysis of lifetime distributions. The expression for the hazard rate of SMPIE is obtained as

$$h(x;\alpha,\lambda) = \frac{g_{SMPIE}(x;\alpha,\lambda)}{R(x;\alpha,\lambda)} = \frac{\lambda}{x^2}\log\alpha\frac{e^{(\log\alpha)(1-e^{\frac{-\lambda}{x}})}e^{\frac{-\lambda}{x}}}{e^{(\log\alpha)(1-e^{\frac{-\lambda}{x}})}-1} \tag{8}$$

Figure (3) depicts graphs of the hazard rate of the SMPIE distribution for different parameter combinations. It shows that the hazard function exhibits a unimodal increasing, decreasing and constant shapes for different parameter combination. This implies that the SMPIE distribution can be used to describe real life phenomenon with unimodal failure rates.

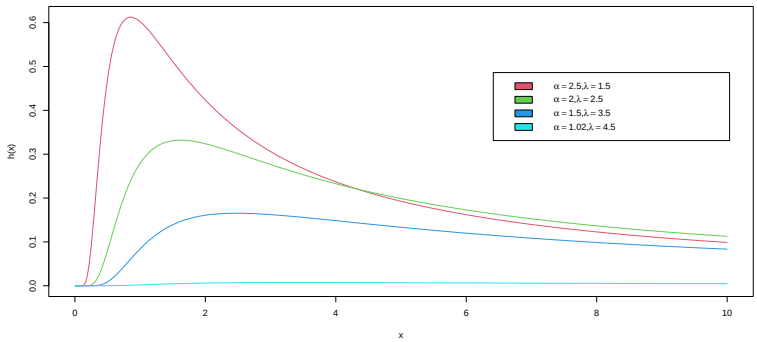


Figure 3: *Hazard plot for SMPIE distribution*

4.3. Reverse Hazard Rate

The concept of reverse hazard rate of a random variable is defined as the ratio between the life probability density to its distribution function and is obtained as

$$h_r(x;\alpha,\lambda) = \frac{g_{SMPIE}(x;\alpha,\lambda)}{G_{SMPIE}(x;\alpha,\lambda)} = \frac{\lambda}{x^2}\log\alpha\frac{e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}e^{-\frac{\lambda}{x}}}{\alpha - e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}} \tag{9}$$

4.4. Mills Ratio

The mills ratio for the SMPIE is defined as

$$M.R = \frac{G_{SMPIE}(x;\alpha,\lambda)}{R_{SMPIE}(x;\alpha,\lambda)} = \frac{e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})} - \alpha}{1 - e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}} \tag{10}$$

4.5. Quantile function

Theorem 1. If $X \sim SMPIE(\alpha,\lambda)$ distribution, then the quantile function of X is given as

$$x = \frac{-\lambda}{\log\left[1 - \frac{\log\{u(1-\alpha)+\alpha\}}{\log\alpha}\right]} \tag{11}$$

where u is a uniform random variable, $0 < u < 1$

Proof. Let $G_{SMPIE}(x;\alpha,\lambda) = u$. The quantile function of SMPIE distribution can be obtained as follows.

$$\Rightarrow \frac{e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})} - \alpha}{1-\alpha} = u$$

$$\Rightarrow e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})} = u(1-\alpha)+\alpha$$

Taking logarithm on both sides and simplifying further, we obtain the required quantile function as

$$x = \frac{-\lambda}{\log\left[1 - \frac{\log\{u(1-\alpha)+\alpha\}}{\log\alpha}\right]} \tag{12}$$

Where u follows a uniform (0,1) distribution. The qth quantile function of SMPIE distribution is given by

$$x_q = \frac{-\lambda}{\log\left[1 - \frac{\log\{u(1-\alpha)+\alpha\}}{\log\alpha}\right]}$$

The median can be obtained as

$$x_{0.5} = \frac{-\lambda}{\log\left[1 - \frac{\log\{0.5(1-\alpha)+\alpha\}}{\log\alpha}\right]}$$

■

5. Statistical properties of smpie distribution

Some of the statistical properties of SMPIE will be discussed in this section.

5.1. Moments

The r^{th} moment for SMPIE distribution can be obtained as

$$\mu_r{}' = E(x^r) = \int_0^\infty x^r g_{SMPIE}(x;\alpha,\lambda)\,dx$$

$$E(x^r) = \frac{\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}}{\alpha-1}\int_0^\infty x^r\left(e^{-\frac{\lambda}{x}}\right)^{k+1}\frac{\lambda}{x^2}\,dx$$

$$\Rightarrow \mu_r' = \frac{\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}}{\alpha-1}\lambda^r(k+1)^{r-1}\gamma(1-r) \tag{13}$$

5.2. Harmonic Mean

Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the data values. The harmonic mean (H) of SMPIE is given as:

$$\frac{1}{H} = E\left(\frac{1}{X}\right) = \int_0^\infty \frac{1}{x}\, g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$\frac{1}{H} = \frac{1}{\alpha-1}\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}\lambda\int_0^\infty \frac{1}{x^{2+1}}\left(e^{-\frac{\lambda}{x}}\right)^{k+1}\, dx$$

$$\frac{1}{H} = \frac{1}{\lambda(\alpha-1)}\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}\frac{1}{(k+1)^2} \tag{14}$$

5.3. Moment Generating Function of SMPIE

Theorem 2. Let X follows the SMPIE distribution, then the moment generating function, $M_X(t)$ of SMPIE distribution is given as

$$M_x(t) = \frac{1}{\alpha-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{t^r}{r!}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}\lambda^r(k+1)^{r-1}\gamma(1-r) \tag{15}$$

Proof. The moment-generating function of SMPIE distribution is defined as

$$M_x(t) = \int_0^\infty e^{tx} g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$M_x(t) = \int_0^\infty \left(1 + tx + \frac{(tx)^2}{2!} + \dots\right) g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$M_x(t) = \sum_{r=0}^{\infty}\frac{t^r}{r!}\int_0^\infty x^r g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$\Rightarrow M_x(t) = \frac{1}{\alpha-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{t^r}{r!}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}\lambda^r(k+1)^{r-1}\gamma(1-r) \tag{16}$$

■

5.4. Characteristic Function of SMPIE distribution

Theorem 3. Let X follows the SMPIE distribution, then the characteristic function, $\phi_X(t)$ of SMPIE distribution is given as

$$\phi_x(t) = \frac{1}{\alpha-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{j}(-1)^k\frac{(it)^r}{r!}\binom{j}{k}(\log\alpha)^{j+1}\lambda^r(k+1)^{r-1}\gamma(1-r)$$

Proof. The characteristic function of SMPIE distribution is defined as

$$\phi_x(t) = \int_0^\infty e^{itx} g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$\phi_x(t) = \int_0^\infty \left(1 + itx + \frac{(itx)^2}{2!} + \dots\right) g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$\phi_x(t) = \sum_{r=0}^{\infty}\frac{(it)^r}{r!}\int_0^\infty x^r g_{SMPIE}(x;\alpha,\lambda)\, dx$$

$$\Rightarrow \phi_x(t) = \frac{1}{\alpha-1}\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{(it)^r}{r!}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}\lambda^r(k+1)^{r-1}\gamma(1-r) \tag{17}$$

■

Lemma 1. Let us suppose a random variable X follows SMPIE (α, λ) with PDF given in equation (6) and let $I_r(t) = \int_0^t x^r g_{SMPIE}(x;\alpha,\lambda)dx$ denotes the r^{th} incomplete moment, then we have

$$I_r(t) = \frac{\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}}{\alpha-1}\lambda^r\gamma(1-r,\lambda/t) \tag{18}$$

where $\gamma(a,b) = \int_b^\infty z^{a-1}e^{-z}\,dz$ denotes the upper incomplete gamma function.
Proof. Using the PDF of SMPIE given in equation (6), we have

$$I_r(t) = \frac{1}{\alpha-1}\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}\int_0^t x^r\left(e^{-\frac{\lambda}{x}}\right)^{k+1}\frac{\lambda}{x^2}\,dx \tag{19}$$

Using substitution ,$\frac{\lambda}{x} = z$ in equation (19), we get

$$I_r(t) = \frac{\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}(\log\alpha)^{j+1}}{\alpha-1}\lambda^r\gamma(1-r,\lambda/t) \tag{20}$$

5.5. Renyi Entropy

The entropy of a random variable is defined as the measure of uncertainty. The Renyi entropy given by [14] is defined as,

$$I_v = \frac{1}{1-v}\log\int_0^\infty g^v(x)\,dx$$

Using PDF given in Equation (6), we have

$$I_v = \frac{1}{1-v}\log\left(\frac{\lambda\log\alpha}{\alpha-1}\right)^v\int_0^\infty\sum_{j=0}^{\infty}\frac{v^j(\log\alpha)^j}{j!}\left(1-e^{-\frac{\lambda}{x}}\right)^j e^{-\frac{\lambda v}{x}}\frac{1}{x^{2v}}\,dx$$

$$I_v = \frac{1}{1-v}\log\left(\frac{\lambda\log\alpha}{\alpha-1}\right)^v\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}\frac{v^j(\log\alpha)^j}{j!}\int_0^\infty\left(e^{-\frac{\lambda}{x}}\right)^{k+v}\frac{1}{x^{2v}}\,dx$$

$$I_v = \frac{1}{1-v}\log\left(\frac{\lambda\log\alpha}{\alpha-1}\right)^v\sum_{k=0}^{j}\sum_{j=0}^{\infty}(-1)^k\binom{j}{k}\frac{v^j(\log\alpha)^j}{j!}(\lambda(k+1))^{1-2v}\gamma(2v-1)$$

which is required expression of Renyi entropy for SMPIE distribution.

6. Order Statistics

Let $X_{(1)},X_{(2)},X_{(3)},\ldots,X_{(n)}$ be the random sample of size n and let $X_{i:n}$ denote the i^{th} order statistics, then the PDF of $X_{i:n}$ is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}F(x)^{i-1}(1-F(x))^{n-i}f(x) \tag{21}$$

Using equation (5) we have,

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!}\left(\frac{e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}-\alpha}{1-\alpha}\right)^{i-1}\left(1-\frac{e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}-\alpha}{1-\alpha}\right)^{n-i}g_{SMPIE}(x;\alpha,\lambda)$$

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!(1-\alpha)^{n-1}}\left(e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}-\alpha\right)^{i-1} \times\left(1-e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}\right)^{n-i}\left(\frac{\log(\alpha)}{\alpha-1}\frac{\lambda}{x^2}e^{(\log\alpha)(1-e^{-\frac{\lambda}{x}})}e^{-\frac{\lambda}{x}}\right) \tag{22}$$

The expression for PDF of the smallest (minimum) order statistics and the largest(maximum) order statistics of SMPIE distribution are respectively obtained by setting i=1 and i=n in above equation.

7. Estimation of Parameters

In this section, we consider the method of maximum likelihood estimation to estimate the unknown parameters of the SMPIE distribution.
Maximum likelihood estimation (MLE) is a technique used for estimating the parameters of a given distribution using some observed data.
Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample of size n having PDF in equation (6) . Then the likelihood function is given by

$$L = \prod_{i=1}^{n} \frac{\log(\alpha)}{\alpha - 1} \frac{\lambda}{x_i^2} e^{(\log \alpha)(1-e^{-\frac{\lambda}{x_i}})} e^{-\frac{\lambda}{x_i}}$$

The log likelihood function is given by;

$$\log L(x : \alpha, \lambda) = -n \log(\alpha - 1) + n \log(\log \alpha) + n \log \lambda - \sum_{i=1}^{n} \frac{\lambda}{x_i} + (\log \alpha) \sum_{i=1}^{n} \left(1 - e^{-\frac{\lambda}{x_i}}\right) - 2 \sum_{i=1}^{n} \log x_i$$

The MLEs of α, λ are obtained by partially differentiating above equation with respect to model parameters and equating to zero, we have

$$\frac{\partial \ell}{\partial \alpha} = -\frac{n}{\alpha - 1} + \frac{n}{\alpha \log \alpha} + \sum_{i=1}^{n} \frac{(1 - e^{-\frac{\lambda}{x_i}})}{\alpha} = 0 \tag{23}$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \frac{1}{\sum_{i=1}^{n} x_i} + \log \alpha \sum_{i=1}^{n} \frac{e^{-\frac{\lambda}{x_i}}}{x_i} = 0 \tag{24}$$

Since the above equations is not in a closed form, therefore we will employ Newton-Raphson method and hence using R software to solve these equations and estimate the parameters.

8. Simulation Study

This section deals with the simulation study using R software to demonstrate the MLE's behavior. The inverse CDF method is used to generate a random sample of size n=25, 75, 150, 300, 500. The process is repeated 1000 times and two different combinations of parameters are chosen as (0.2,0.60) and (0.4, 0.25) with relation to the standard order (α, λ). The average MLE values, bias, and mean square error (MSE) of the ML estimates were calculated for each scenario. Table (1) exhibits the ML estimates, bias, and MSE. The MLEs presented are consistent estimators, which means as the sample size increases, these estimates converge in probability to the true parameter values. The estimates are stable and near the actual parameter values. As the sample size increases the MSE drops for all estimates.

Table 1: *MLE, Bias, and MSE for the parameters α and λ*

Sample size	Parameters		MLE		Bias		MSE	
n	α	λ	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$
25	0.20	0.60	1.15185	0.82595	1.02735	0.33423	5.78128	0.21386
75			0.42503	0.66675	0.32059	0.21102	0.40229	0.07628
150			0.29569	0.62366	0.18561	0.15443	0.07317	0.03772
300			0.24847	0.61015	0.12152	0.11157	0.02855	0.02076
500			0.20553	0.60487	0.09484	0.08943	0.01588	0.01424
25	0.40	0.25	1.53164	0.30428	1.30163	0.10692	10.94223	0.01896
75			0.67186	0.26746	0.43151	0.06582	0.48012	0.00724
150			0.53525	0.25733	0.27963	0.04846	0.18152	0.00392
300			0.46257	0.25416	0.17622	0.03342	0.05949	0.00201
500			0.40278	0.25087	0.13152	0.02502	0.03022	0.00106

9. Application

This section tests the flexibility, adaptability, and suitability of the SMPIE model against a few other existing distributions using two real life data sets. To compare the SMPIE model with other fitted distributions, we compare the fits of the SMPIE distribution with the generalized inverted generalized exponential distribution (GIGE) was proposed by [9], Exponentiated inverted exponential distribution (EIE) was introduced by [4], and transmuted inverted exponential distribution (TIE) was obtained by [10]. Using several goodness of fit criteria such as -2ll, Akaike Information criterion (AIC), Bayesian information criterion (BIC), Akaike Information criterion Corrected (AICC), Kolmogorov Smirnov (KS) and P value. In general, the model is to be considered the best for which these goodness of-fit statistics have the least value.

9.1. Data set 1

The first data set [6] represents the failure times, in minutes, of 15 electronic components in an accelerated life test and they are as follows:
1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2, 23.0, 30.6, 37.3, 46.3, 53.9, 59.8, 66.2

Table 2: *Estimates (standard errors), -2ll, AIC, BIC, AICC, K-S statistic and P-value for Data-set 1.*

Model	$\hat{\alpha}$	$\hat{\lambda}$	$-2ll$	AIC	BIC	AICC	K-S	P-value
SMPIE	0.0218 (0.0610)	3.7371 (2.7082)	134.5441	138.5441	139.9602	139.5441	0.18358	0.6829
GIGE	0.9107 (0.1654)	8.9462 (1.0369)	138.03922	142.0392	143.4553	143.03926	0.24898	0.2633
EIE	3.2399 (913.6348)	2.9504 (832.0125)	138.1101	142.1101	143.5262	143.1101	0.26314	0.2093
TIE	-0.7039 (0.322)	2.0491 (2.1894)	135.973	139.973	141.3891	140.6789	0.20540	0.4886

9.2. Data set 2

The second data set consists of vinyl chloride data (in mg/L) obtained from clean-up-gradient monitoring wells. The data has been previously used by [12].
5.1, 1.2, 1.3, 0.6, 0.5, 2.4, 0.5, 1.1, 8.0, 0.8, 0.4, 0.6, 0.9, 0.4, 2.0, 0.5, 5.3, 3.2, 2.7, 2.9, 2.5, 2.3, 1.0, 0.2, 0.1, 0.1, 1.8, 0.9, 2.0, 4.0, 6.8, 1.2, 0.4, 0.2.

Table 3: *Estimates (standard errors), -2ll, AIC, BIC, AICC, K-S statistic and P-value for Data-set 2.*

Model	$\hat{\alpha}$	$\hat{\lambda}$	$-2ll$	AIC	BIC	AICC	K-S	P-value
SMPIE	0.0468 (0.1145)	0.2523 (0.1782)	114.48888	118.488881	118.8759	121.54161	0.0951	0.9179
GIGE	0.9214 (0.2110)	0.5412 (0.1285)	118.2569	122.2569	122.64403	125.30966	0.1365	0.5499
EIE	0.7060 (49.5443)	0.8108 (56.8950)	118.3860	122.3860	122.7731	125.4388	0.1470	0.4544
TIE	-0.6300 (0.0782)	0.4138 (0.1089)	115.8404	119.8404	120.2275	122.8931	0.10350	0.8592

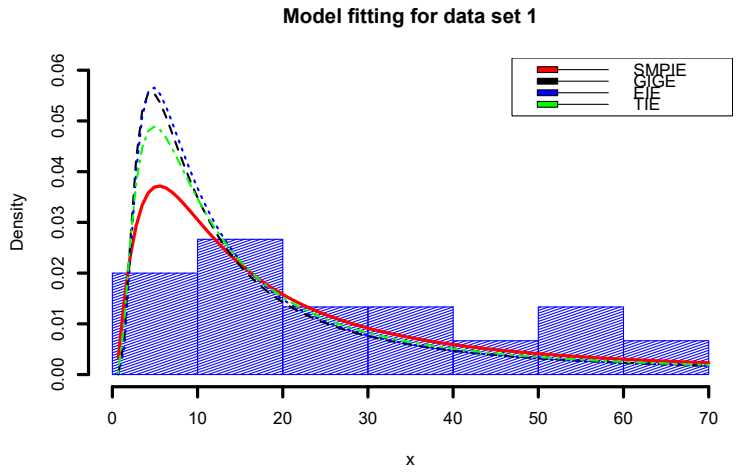


Figure 4: *Fitted density plots for dataset 1*

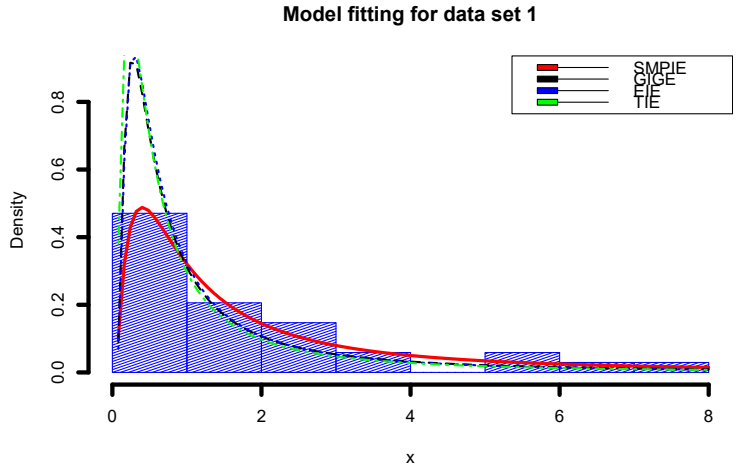


Figure 5: *Fitted density plots for dataset 2*

10. Conclusion

In this manuscript, we have proposed a new distribution called SMP Inverted Exponential (SMPIE) distribution. The model is positively skewed, its shape could be decreasing (depending on the values of the parameters). Various statistical properties of the proposed distribution such as survival function, hazard function, rth moments, quantile function, moment generating function, Renyi entropy, and order statistics were studied. A simulation study was carried out to test the performance of maximum likelihood estimation. The result shows that the mean square error decreases as the sample size increases, i.e. they are consistent estimators. The proposed distribution was applied to two real life data sets and comparing it with well-known standard distributions, and the outcomes are shown in Tables(2) and (3). An application to the real-life data sets shows that the fit of SMPIE distribution is superior to the fits using GIGE, EIE and TIE distributions.

References

[1] A. Abouammoh and A. M. Alshingiti. Reliability estimation of generalized inverted exponential distribution. *Journal of statistical computation and simulation*, 79(11):1301–1315, 2009.

[2] I. David, S. Mathew, and J. Falgore. New sine inverted exponential distribution: Properties, simulation and application. *European Journal of Statistics*, 4:5–5, 2024.

[3] J. Eghwerido. A new weibull inverted exponential distribution: properties and applications. *FUPRE Journal of Scientific and Industrial Research (FJSIR)*, 6(1):58–72, 2022.

[4] K. Fatima and S. P. Ahmad. The exponentiated inverted exponential distribution. *Journal of Applied Information Science*, 5(1), 2017.

[5] A. Keller, A. Kamath, and U. Perera. Reliability analysis of cnc machine tools. *Reliability engineering*, 3(6):449–473, 1982.

[6] J. F. Lawless. *Statistical models and methods for lifetime data*. John Wiley & Sons, 2011.

[7] C. Lin, B. Duran, and T. Lewis. Inverted gamma as a life distribution. *Microelectronics Reliability*, 29(4):619–626, 1989.

[8] M. R. Mahmoud, A. E. Ismail, and M. A. Ahmad. Weibull-inverse exponential [loglogistic] a new distribution. *Asian Journal of Probability and Statistics*, 21(3):33–44, 2023.

[9] P. Oguntunde and A. Adejumo. The generalized inverted generalized exponential distribution with an application to a censored data. *Journal of Statistics Applications & Probability*, 4(2):223, 2015.

[10] P. Oguntunde and A. Adejumo. The transmuted inverse exponential distribution. *International journal of advanced statistics and probability*, 3(1):1–7, 2015.

[11] P. Oguntunde, A. Adejumo, and O. Balogun. Statistical properties of the exponentiated generalized inverted exponential distribution. *Applied Mathematics*, 4(2):47–55, 2014.

[12] H. Okasha and M. Nassar. Product of spacing estimation of entropy for inverse weibull distribution under progressive type-ii censored data with applications. *Journal of Taibah University for Science*, 16(1):259–269, 2022.

[13] S. U. Rasool, M. A. Lone, and S. P. Ahmad. An innovative technique for generating probability distributions: A study on lomax distribution with applications in medical and engineering fields. *Annals of Data Science*, pages 1–17, 2024.

[14] A. Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, volume 4, pages 547–562. University of California Press, 1961.

[15] S. K. Singh, U. Singh, and D. Kumar. Bayes estimators of the reliability function and parameter of inverted exponential distribution using informative and non-informative priors. *Journal of Statistical computation and simulation*, 83(12):2258–2269, 2013.

[16] C. Ünal, S. Cakmakyapan, and G. Özel. Alpha power inverted exponential distribution: Properties and application. *Gazi University Journal of Science*, 31(3):954–965, 2018.

AN ANALYSIS OF RELIABILITY IN MANUFACTURING INDUSTRIES

Shakuntla Singla[1], Sonia[2*], Shilpa Rani[3]

•

[2*]Department of Mathematics, Dr. Bhim Rao Ambedkar Govt. College Kaithal, (Jagdishpura) India
[1,2*,3] Department of Mathematics and Humanities, MMEC, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India
shakus25@gmail.com, sonia_garg99@yahoo.com, gargshilpa46@gmail.com

Abstract

This paper offers an initial evaluation of the organizational and structural relationships among reliability and best warranty programme. In the producing sectors because of automation, plant potential has been extended with inside the system industries. This enables in growing productiveness in addition to the best of the material, but every computerized industry, massive funding remains the top anxiety. Consequently, it additionally anticipated the running structures ought to work for a long time and defective-free. In the time being, it turns into essential to present right care of running machines. Operation of those features and using precise strategies in those regions and the obstacles to their reputation has additionally been discussed. The contemporary paper offers the evaluation of the consistency evaluation. Consistency evaluation of numerous structures is evaluated in specific system productions just like the sugar production, updraft energy productions, milk productions, mining, petroleum productions, etc. The series of reliability and best expenses records and its use with the aid of using pinnacle control in decision-making regarding destiny upgrades have additionally been covered. The specific tactics are used by investigators in numerous grounds to test the overall activity of the running machine. These tactics are genomic procedure, fault tree evaluation, deficiency and impact evaluation, petrify-nets, dependability, accessibility, maintainability and deprivation modeling strategies etc. The growth has been shown from the overall performance of the structures mainly totally depend upon numerous records through the above tactics.

Keywords: Reliability, Consistency, Stochastic.

I. Introduction

In the current business scenario, because of sizable improvement in technology, it turns into very stiff to preserve the workingstructures in properly operating condition. The complicated device and their availability for a long time turn out to be the want for any technical trade. To test the uniformity of any working device, reliability is one of the measurements that is used. There is a possibility that a factor will carry out its supposed characteristic safely for a designated duration beneath Neath the required working situations. Availability is the possibility that an object while used beneath Neath given situations will carry out satisfactorily while required. Maintainability is the possibility that an object may be restored to high-quality working situations inside a designated duration beneath Neath said situations via way of means of employees having prescribed abilitystage, sources and procedures. Then, it turns into very essential to screen renovation plans. The overall presentation of Industry gadget may be better with dependability, accessibility and maintainability evaluation. If the working device is unavailable and unreliable then it lowers the performance of the plant. It ends

in the failure of its manufacturing unit. In the sphere of engineering asset management, studies at the belongings fitness and existence span prediction has been multiplied. Degradation reduces the gadget's existence at distinct duration of time and because of its reliability of the device reduced. During reliability evaluation, it turns into very essential to research the degradation stage of belongings also. The theoretical idea can enhance the existence of belongings however want arises to apply a few mathematical evaluations. Reliability evaluation is a totally fruitful approach which used a few mathematical approaches to discover the supply of the device. This approach facilitates in identifying the priorities of renovation to the gadget. The precedence can be given to the gadget which has a most failure charge. With positive regarded values of failure and restore charge, the optimization strategies are appliedto calculate the high- quality mixture of failure and restore charge for the supply of the device. Later on, this facilitates in making plans renovation strategies/guidelines to hold the belongings in properly operating situations. So that those will make certain more consistency of overall performance for a protracted tenure. Ultimately, the overall performance of a plant may be multiplied on this way. The article ends with a few concluding remarks.

II Literature review

Moses and Stahl (1979) [1] presented different types of member behaviour, arrangements, statistical and mechanical correlation between the elements. This paper presents approach of unzipped structural reliability analysis with illustration. The growing demand for production in huge economic investments and the hostile environment in a single offshore platform suggest that target risks should be considered.

Dhesi & Wardha (1983) [2] mentioned that maximum of the marketers commenced with a bit quantity of preliminary capital and the supply of funding for majority of them became their personal saving or budget from casual supply. House (1984) discovered of them examine that maximum of the owners withinside the casual region has been city citizens of lengthy status and now no longer current migrants. Though, the preliminary capital necessities have been low, but maximum of the marketers had referred to capital scarcity as their fundamental trouble.

Liebl& Roy (2003) [3] have located that a median artisan withinside the handicraft region has insufficient get entry to triumphing facts approximately markets, buyers, tastes and technologies. In addition, additionally they face issues like insufficient capital and absence of herbal uncooked substances.

Avi Ostfeld (2004) [4] talks about reliability is a part of water distribution system layout, operation, design and maintenance. Analysis of reliability is complicated due to affect of many factors like inherent nonlinear behavior and its consumers utility. Although the reliability for water distribution system has more considerable attention for last years without any methodology. For many years two types of reliability evaluation is famous first is 'lumped Supply- Lumped Demand' and second is Stochastic framework.

Singh et al (2008) [5] examine found out that there are numerous issues confronted via way of means of the Bhadohi carpet enterprise inclusive of economic issues, advertising and marketing issues and non- availability of uncooked substances. The examine advised that authorities need to deliver a few economic aid and education to the respondents to begin their gadgets and additionally to replace themselves with the state-of-the-art strategies of carpet weaving. This in flip could growth the manufacturing and earnings stage.

Babel &Choudhary (2009, 2010) [6] try and examine the technique, organizational structure, and layout in addition to the method that's getting used with inside the hand-knotted carpet

production gadgets of Jaipur. The variables recognized for the examine are associated with profile of manufacturers, infrastructural facts of gadgets, uncooked substances and manufacturing technique. The examine concluded that carpet production enterprise meets the economy's wishes of forex in addition to imparting employment to humans in big numbers. The examine highlights the truth that this enterprise calls for professionallytrained, nicely ready and nicely knowledgeable designers, clippers, finishers in addition to manufacturing managers.

Bhavani & Sharada Devi (2010) [7] have made a try and discover the popularity of carpet weaving enterprise of Warangal together with its functioning, the issues confronted via way of means of the weavers and motives for decline in call for for carpets. The examine found out that weavers are dealing with diverse issues associated with procurement of uncooked substances, finance, operating conditions, advertising and marketing and fitness which might be affecting their well-being. The exam ine advised that weavers have to take delivery of right education concerning the state-of-the-art strategies of carpet designing and weaving for up gradation in their gadgets. Government needs to come ahead with a few packages deal of help to reinforce this region because it will assist to enhance the circumstance of weavers and hold the wealthy history of Indian carpet weaving.

Shakuntla et al (2011) [8] purpose of this paper is to calculate the reliability of polytube industry with four elements. Repair and failure rates are vary and all the statistical equations are derived from chapman-Kolmogorov discrepancy equations of the manufacturing plant. Reliability of the system is solved numerically with the help of Runga kutta method of order four and sensitive analysis of reliability improve overall accessability.

Choudhury (2012) [9] advanced a enterprise version for cottage industries primarily based totally on facts conversation era and deliver chain logistics control. The examine highlights the want of organisations and integration of unorganized cottage industries. It is usually recommended that the emphasis need to take delivery of on promotion, improvement and status quo oflatest cottage industries with a purpose to take away nearby imbalances and unfold out the blessings springing up from it.

Azad et al (2012) [10] empirical examine became undertaken to analyze greater approximately demanding situations in Iranian carpet enterprise the use of element analysis. The examine concluded that 8 items (inclusive of specialized relationships, information coordinator, information tool, information organisations, information processes, information chain, information hardware and Knowledge feasibility examine) permit us to construct higher techniques to assist this enterprise develop faster.

Saeedi et al (2012) [11] speak approximately the have an effect on of World Trade Organization (WTO) on Iran's carpet enterprise. The examine predicted that once becoming a member of the WTO, carpet Industry might be greater compatible, home company sources will enhance, seize of greater global markets might be viable and the enterprise's creativity might be strengthened. The examine states that the maximum vital techniques for enhancing the scenario of carpet enterprise are-elevating commercial spending, advertising and marketing studies, aligning manufacturing to worldwide client wishes, making exporters accustomed to fashionable global advertising and marketing strategies and use of e-commerce. Moreover, a few funding needs to be accomplished on studies and improvement and organizational information to make carpet enterprise geared up for competing in worldwide markets.

Saeedi et al (2012) [12] examine examines the effect of marketplace-primarily based totally elements on Iran's carpet Industry. The effects display the fine outcomes of marketplace intensity, client capital and aggressive intelligence on carpet enterprise. According to this examine

marketplace share, client wishes identity and marketplace intelligence are raising because the maximum vital variables in shaping marketplace- primarily based totally view.

Gera (2012) [13] examine is primarily based totally on secondary statistics and is each exploratory and descriptive in nature. The examine concluded that the Indian fabric enterprise has accomplished a large position in phrases of contribution to GDP, exports, employment era and incomes of overseas exchange. Their proposal is to enhance its uncooked cloth base and export excessive fee- introduced merchandise of worldwide standard. At the equal time Indian authorities need to take some time to improve the technology, lessen the taxes and responsibilities and open education centers.

Meenaxi & Sudha (2012) [14] awareness their examine at the fitness issues because of carpet weaving. The examine said that carpet weaving is one of the maximum tedious professions which require lengthy hours of labor and additionally it's far a excessive- chance career for growing musculoskeletal problems as awkward posture, repetitive actions and make contact withstrain are common.

Khan & Misra (2014) [15] studied the Bhadohi carpet enterprise in three phases- possibilities phase (after 1991-92 to 2006-07, there has been average growth with inside the marketplace size), declining phase (brilliant decline at some point of the length 2007-10 because of income and productiveness decline) and marketplace recapturing phase. The primary reasons of financial slowdown in Bhadohi carpets are loss of orders at some point of slowdown length, unfavorable circumstance of exporters, migration and rural employment schemes like MGNREGA. They concluded that the carpet enterprise has gone through fundamental modifications due to the fact conventional markets have been saturated at the same time as new markets are having possibilities for increase best via the version of modifications.

Cristiano Fragassa Ana Pavlovic Salvatore Massimo (2014) [16] in this paper a strategy of Total Quality Management is proposed, civilized and used with the purpose of improving the quality of large-mass industrial commodity far beyond the technical condition demanded at the end-consumer level. This concept combines standard and non- standard tools ill-used for Reliability, Availability and Maintainability analysis. The subroutine also realizes a stricter correlation between experimental evidences and theoretical evaluation methods as part of a modern integrated method for transformation quality in blueprint and process. A commercial Intake Manifold, largely spread in the marketplace, is used as tester for the validation of the methodology. As general additional phenomenon, the research-underlines the impact of Total Quality Management and its tools on the process of invention.

Naga Vamsi Krishna Jasti, Rambabu Kodali (2014) [17], the purpose of this paper is to investigate, validity and reliability analysis on existing Lean manufacturing (LM) frameworks when applied to Indian organizations. LM is one of the best manufacturing strategies that are used by manufacturing plant managers to improve manufacturing capabilities.

Pandey (2014) [18] examine tested the demanding situations dealing with carpet enterprise because of its now no longer being ICT friendly. In current years Indian Institute of Carpet Technology (IICT), Bhadohi has been looking to treatment this via way of means of imparting technical aid/ education on PC primarily based totally designing to the enterprise the use of state-of-the-art software. The examine concluded that via way of means of the use of ICT platform carpet enterprise can turn out tobe able to competing with worldwide carpet manufactures and also can growth the productiveness of the weavers.

.

Singh & Fatima (2015) [19] try to investigate the significance of integrating Handicraft area in

Uttar Pradesh with different sectors for introduction of possibilities of financial boom. This became emphasized specifically with inside the case of hand knotted carpets. Stability of employment (i.e., yr spherical in place of seasonal) in addition to improvement of a marketplace statistics gadget became additionally emphasized.

Jain et al (2015) [20] examine measures India's comparative benefit and competitiveness with inside the carpet and rug enterpriseand compares it with that of fundamental global exporters on this enterprise like China, Belgium, Netherlands and Turkey. The examine famous that India's carpet and rug region has excessive stage of export competitiveness and aggressive benefit, validating its excessive ability to earn forex. India additionally holds a sturdy role with inside the global markets at each -digitstage and 4-digit stage classifications of the product.

Malik & Prasad (2015) [21] tested the limitations and possibilities that have emerged for the micro, small and mediumorganizations in carpet enterprise after change liberalization. The examine concluded that fundamental constraints for the carpet enterprise are- improved price of uncooked fabric, loss of infrastructural facilities, problem in export facilitation, unhelpful legal-regulatory framework and problem in buying budget from neighborhood economic institutions.

Bano (2016) [22] examine made a try to examine the capacity of Indian carpet enterprise in accelerating the boom and improvement of Indian economic system. This examine primarily based totally on secondary statistics located that Indian carpet enterprise skilled fantastic fashion of boom considering the fact that 1961, no matter stiff opposition with inside the international marketplace. The examine concluded that the ever-growing boom in phrases of export fee suggests the excessive capacity of sustainability of this enterprise in coming future.

Bhat &Yadav (2016) [23] attempted to examine the quantity of export of handicrafts, the contribution of handicraft area to Indian economic system and the main overseas markets of export for Indian handicrafts. The examine concluded that handicrafts have the capacity of taking pictures greater overseas markets however promotional efforts with the aid of using the authorities, collaboration to introduce cutting-edge technology, and cutting-edge designs are needed.

Ashraf et al (2016) [24] made a try and acquire the facts approximately the effect of talent up gradation and ability constructing education programme with unique connection with Jammu & Kashmir and additionally to evaluate the extent of development in trainees after present process one of these education programme. The examine concluded that the education helped the artisans at once in connecting them to the marketplace, thereby casting off middlemen with inside the change. It additionally helped them to paintings on marketplace orientated designs, higher satiation combinations, and discovers new re assets of uncooked fabric. Besides this, education software additionally helped the artisans to have fashionable attention concerning distinct schemes of the authorities.

Das et al (2018) [25] try and recognize the popularity of weavers of Bhadohi Carpet enterprise and additionally to discover a viable answer to enhance their life. The examine concluded that there are fundamental motives for go out of weavers from this enterprise- low salary payments, and absence of marketplace talents to enhance their enterprise. The examine advised that distinct institutes need to be setup in order that right talents are supplied to the floor stage people so as to carry a higher earnings and livelihood for them.

Rosmaini Ahmad (2018) [26] presents a case study in his paper which tell us the consistency due to diverse preservation choices for perforating utensil groups that are used in a line-making process. Field preservation numbers for perforating utensils are distributed according to diverse maintenance choices. The numerous data were examined which identify the suitable match for

distribution models. The similar reliability features are then exposed and discussed. The reliability analysis results of four different most important maintenance choices displays that there is no significant difference in mean-time-to-failure (MTTF) of the equipment. However, for each replacement decision package can use the range (lower and upper limit values) of MTTFby a maintenance engineer for equipment spare parts preparation. The conclusion founds that the failure rate of the equipmentwas due to auxiliary decisions packages that are in a failing stage. Thus, the maintenance engineering team observed that the investigation of optimum preventive maintenance time is beneficial and cost effective. This study concludes that it helps the technical team to utilize proper preventive maintenance and spare parts planning.

Mohammad Ali Farsi and Enrico Zio (2019) [27] In this paper, the belief of Industry 4.0 is described and some of these concepts and occasions for reliability engineering are discussed. New instructions for investigation in system modelling, largefacts analysis, health administration, fake-physical system, human-machine contact, ambiguity, jointly analyzation, announcement, and interactions are anticipated. Every matter can be investigated separately, but this paper précises them and prepared a vision about reliability analysis for reflection and discussion by the interested technical community.

Harbhinder Singh, Munish Mehta, Janender Kumar (2020) [28] in the industrial sectors due to computerization, capability has been improved in the developing industries. This supports in growing efficiency as well as the worth of the goods. But forevery mechanical industry, a huge amount of investment is also major anxiety. So, it is also predictable that working system should work for a long time and defective-free. In the interim, it becomes more important to give proper care to working apparatus. They described in their paper overview of the reliability analysis. Reliability examination of various systems is examined in different types in different process industries like the milk industry, mining, sugar industry, thermal power plants, petroleum industries, etc. The different tactics are used by researchers in numerous fields to examine the act of the operating equipment. These tactics are G.A., H.A., PSO, Machine Learning reliability, accessibility, maintainability and deprivation modeling methods etc. The development has been realized in the act of the systems depend on accurate facts with the use of the above methods.

Shakuntla Singla, Sonia 2024 [29], In their paper, discussing reliability distribution of mushroom plant through regenerative point to graphical technique, and examine the reliability in case of constant failure rate and repair rate.

Sonia, S. Singla 2024 [30], presented in their paper, reliability analysis by using the distribution of normal, gamma, exponential and wiebul and observe that normal distribution is best fit for reliability analysis and graphically analysis through minitab Software.

III Conclusions

The boom of any technique enterprise relies upon the provision of its belongings and upkeep strategy. Hence, to get most output from running structures, it's far very a great deal vital that those should be appearance after minutely in order that the extent of failure and restore fee might be minimized. In this way, the general performance of the plant may be optimized. It is concluded that:

- Reliability evaluation allows in figuring out the provision of various components. Which is depend upon that accessibility fact, significances of components gadget may be fix for running system. The reason is to boost the presence of properties.

- In the latest time, the deprivation idea takes hastily unfold in technical businesses as it could boom the dependability of the manufacturer plant.

- The natural stimulated strategies are in trend nowadays like ant colony, PSO and grey wolf optimizations.

- Reliability evaluation allows in lowering the fee of substitute of additives as it isn't always affordable for each time to recover them.

- Deprivation of gadgets decreases the consistency of the device. Reliability examines the facts of failure and restores factswhile deprivation facts assist us in predicting the existence span and fitness of gadgets.

- Reliability evaluation idea is beneficial in computing the device availability and additionally growing the imply time among failures, as unique studies papers are the proof of its success.

- The survey discovered the wide variety of consistency approaches has been pragmatic in productions and high-quality outcomes are discovered in phrases of device accessibility.

- It provides the stage for the medication of running structures through conservation. Different rules or techniques may be deliberate to boom the provision of crucial additives.

References

[1] Moses, F., Stahl, B. (1979). Reliability Analysis Format for Offshore Structures. *J Pet Technol,* 31(03):347-354.

[2] Dhesi, A. S. and Wadhwa, U. (1983). The informal sector in Nangal, 1980 (Punjab): An overview (India). *IndianJournal of Regional Science*. 15(1):1-16.

[3] Liebl, M. and Roy, T. (2003). Handmade in India- Preliminary Analysis of Crafts Producers and Crafts Production. *Economics and Political Weekly*, 38 (51/52):5366- 5376.

[4] Ostfeld, A. (2004). Reliability analysis of water distribution systems. *Journal of Hydroinformatics*, 6(4):281-294.

[5] Singh, K., Sharma, E., and Rukhsana. (2008). Carpet Weaving industry of Bhadohi district, Uttar Pradesh (UP) - AnOverview. Man-made, 51(6): 211-214.

[6] Babel, S., and Choudhary, M. (2009, 2010). Process, Techniques and Designs of Hand Knotted Carpet Manufacturing Units of Jaipur District. *Rajasthan Journal Extension Education*, 17 & 18:198-202.

[7] Bhavani, K., and Devi, A. S. (2010). Carpet weaving industry of Warangal- A field study. *Asian Journal ofHome Science*, 5(2):293-301.

[8] Singla, S., Lal, A. K., Bhatia, S. S. (2011). Reliability analysis of polytube tubr industry using supplimentry variable Technique. *Applied Mathematics and Computation*, 2011.

[9] Choudhury, D. (2012). Organizing the Unorganized Sector: An ICT Enabled Logistics Model in Favour of Cottage and Small-Scale Industries in Northeast India. *Pacific Business Review International*, 5 (2):23-34.

[10] Azad, N., Aliakbar, S. M. S., and Ansari, M. (2012). Investigating knowledge management critical success factors in carpet industry. *Management Science Letters*, 2(8): 2717–2722.

[11] Saeedi, N., Azari, T. and Maleki, T. (2012). Influence world trade organization on Iran‘s carpet industrycompatibility. *African Journal of Business Management*, 6(4):1483-1489.

[12] Saeedi, N., Nabilou, H., Masouleh, S. A., and Beikkhakhian, Y. (2012). A review on Iran ‘s carpet industry situation in international markets. *African Journal of Business Management*, 6 (30): 8902-8909.

[13] Gera, N. (2012). Significance and Future Prospects of Textile Exports in Indian Economy. *IARS InternationalResearch Journal*, 2 (1):1-18.

[14] Meenaxi, T., and Sudha, B. (2012). Causes of Musculo-Skeletal Disorder in Textile Industry.

International ResearchJournal of Social Sciences, 1(4):48- 50.

[15] Khan, W A., and Misra, S. (2014). Impact of Economic Slowdown on Carpet Business in India with specialreference to Bhadohi, UP. *Integral Review Journal of Management*, 1(1):121-128.

[16] Fragassa, C. and Massimo, P. S. (2014). Using a total quality strategy in a new practical approach for improving the product reliability. *International Journal for Quality Research* 8(3):297-310.

[17] Jasti, N. V. K. and Kodali, R., (2014). Validity and reliability of lean manufacturing frameworks: An empirical study in Indian manufacturing industrie. *International Journal of Lean Six Sigma*, 5(4):361-391.

[18] Pandey, S. K. (2014). Need of Effective Marketing Modes for Developing Domestic market for Indian CarpetIndustry. *Journal for Studies in Management and Planning*, 1(6):385-389.

[19] Singh, A. K., and Fatima, S. (2015). Role of Handicraft Sector in the Economic Development of Uttar Pradesh. *International Journal of Research Granthaalayah*, 3(1):58-64.

[20] Jain, M. P., Sharma, S., and Batta, A. (2015). An Empirical assessment of competitiveness of Indian Carpetand Rug Industry. *ELK Asia Pacific Journal of Marketing and Retail Management*, 6 (2):74- 92.

[21] Malik, M. R., and Prasad, R. (2015). Indian Carpet Industry after Trade Liberalization. *Problems and Prospects.Academic Journal of Economic Studies*, 1(3):79–87.

[22] Bano, R. (2016). Role of Handicrafts in Economic Development: A Case Study of Carpet Industry of India. *IRA-International Journal of Management & Social Sciences*, 4(3):512-525.

[23] Bhat, J. A., and Yadav, P. (2016). The Sector of Handicrafts and its Share in Indian Economy. *Arabian Journal ofBusiness and Management Review*, 009(S3):1-6.

[24] Ashraf, S. I., Ashraf, S. N., and Hafiz, S. M. (2016). A Paradigm shift in the Carpet Craft by Skill Up gradation and Capacity Building Training Programmes with special reference to the State of Jammu and Kashmir. *International Journal of Advanced Research*, 4(6):1222-1226.

[25] Das, N. B., Sharma, R. K., Pandey, A., and Narayanan, B. G. (2018). The Existence of Carpet Industry in Bhadohi,India. *Trends in Textile Engineering & Fashion Technology*, 3(3):1-10.

[26] Ahmad, R. (2018). Reliability analysis comparison on punching tool sets due to different maintenance decisions: a case study from the pulp manufacturing industry. *The International Journal of Advanced Manufacturing Technology*, 94(1):1969–1979.

[27] Farsi, M. A., and Zio, E. (2019). Industry 4.0: Some Challenges and Opportunities for Reliability Engineering. *International Journal of Reliability, Risk and Safety: Theory and Applications* 2(1):23-34.

[28] Singh, H., Mehta, M., Bansal, S. A. and Kumar, J. (2020) Reliability Analysis in Process Industries – An Overview. *GIS Sci J*, 7(5):151-168

[29] Singla, S. and Sonia (2024). Stochastic Optimization and Reliability Analysis of Mushroom Plant. *Reliability Theory & Applications* .77(1), 729-743.

[30] Sonia and Singla, S. (2024). Explore the Dynamics of Manufacturing Industries: Reliability Analysis through Stochastic Process Modeling. *Reliability Theory & Applications* .78(2), 467-471.

MODELING OF ARC OVERVOLTAGE DEPENDENCE ON GROUND CIRCUIT RESISTANCE AND PHASE CAPACITANCE

Orijov Najaf İsmail[1], Guliyev Huseyngulu Bayram[1],
Alimammadova Sara Javanshir[3]

•

[1]Baku Engineering University, Khirdalan, Azerbaijan
AZ0101, Hasan Aliyev str. 120
[1]norucov@beu.edu.az
[2]Azerbaijan Technical University, Baku, Azerbaijan
AZ1073, H. Javid avenue 25
[2]huseyngulu@mail.ru, Orcid: 0009-0006-7362-0619
[3]Azerbaijan State University of Oil and Ind**u**stry, Baku, Azerbaijan
AZ1010, Azadlig Avenue 16/21
[3]sara_elimemmedova@mail.ru

Abstract

The need to control the arc overvoltage during the insulation under load test in neutral insulated networks requires the determination of dependencies between single-phase non-stationary ground and parameters characterizing the faults. In most cases, the identification and realization of such dependencies is observed with a number of difficulties. Therefore, for practical conditions, simple mathematical models should be developed that allow knowing the dependencies between these parameters. In this work, the problem of determining the relationship between the overvoltage generated in the neutral isolated network as a result of artificial non-stationary earth faults, the earth fault resistance and the phase capacitance of the network with respect to the earth was considered. For this purpose, using the least squares method, a regression equation was obtained for the dependence of the frequency of overvoltage on the ground fault resistance and the phase capacitance of the network with respect to the ground, and a corresponding 3D image was constructed.

Keywords: neutral isolated network, non-stationary earth fault, earth fault resistance, frequency of overvoltage, regression equation, 3D modeling

I. Introduction

Neutral isolated networks play an important role in uninterrupted supply of electricity to operators. Since such networks mainly carry out the distribution of electric energy between operators, they include a large number of transmission overhead and cable lines. In the process of operation, the insulation of these power lines, especially cable lines, is subjected to long-term electrical, thermal and mechanical effects, so the dielectric properties deteriorate, in other words, the insulation resistance and electrical strength decrease. For this reason, it is important to detect parts of networks with poor insulation before damage occurs and to restore damaged areas.

Therefore, network equipment is periodically taken out of service and its insulation is tested [1,2]. Due to the large number of equipment, such tests are not considered convenient as they require a lot of time and labor. For this reason, various methods have been proposed to test network insulation under load without opening the equipment [3-5].

Insulation under load testing with the above-mentioned methods is based on the principle of creating artificial earth faults in the network. However, the devices used during the application of these methods cannot perform regulation and control functions. For this reason, new test method of insulation in neutral insulated networks was proposed [6]. According to the proposed method, controlled and adjustable artificial non-stationary earth faults are created in the network to test the insulation under load without disconnecting the equipment from the network. It should be noted that the value of the arc overvoltage generated during such earth faults based on Petersen's theory depends on the phase capacity of the network with respect to the earth, the earth fault resistance and the earth fault angle. Therefore, for the determination and control of the value of the test voltage, the preliminary determination of the values of the ground fault resistance, the ground fault angle and the capacitance of the network with respect to the ground is of particular importance and is an urgent issue.

For this purpose, in the previous works of the authors, the frequency of arc overvoltage generated in neutral insulated networks as a result of artificial non-stationary earth faults was determined separately and some pair correlation dependences on earth fault resistance, earth fault angle and phase capacity of the network with respect to earth [7-14]. As a continuation of the studies mentioned in the present study, the correlation dependence of the frequency of artificial earth fault arc overvoltage on the earth fault resistance and the phase capacity of the network with respect to earth is studied, and based on this dependence, an effective, easy-to-realize regression model is proposed for the purpose of controlling the arc overvoltage.

II. Problem statement for arc overvoltage regression model

In order to determine the dependencies of the artificial earth fault arc overvoltage frequency, the earth fault resistance, the earth fault angle and the phase capacity of the network with respect to the earth, using one of the numerical methods of the system of differential equations characterizing the transition process of the non-stationary earth fault created in the neutral isolated networks, with the application of modern computing technologies solution must be performed. However, the solution of the problem by numerical methods becomes much more difficult due to the greater non-linearity of the mentioned differential equations, "stiffness" due to the impossibility of determining certain parametric quantities, due to the mentioned reasons, in certain cases, the stability of the solution of the system of differential equations is violated and the results are distorted. Therefore, to overcome such difficulties, the frequency of arc overvoltage (k) with ground fault resistance (R_0), angle of closure with the ground $(\varphi)$and the phase capacitance of the network with respect to ground (C_f) Obtaining analytical expressions that determine the dependencies between them is one of the important issues.

It should be noted that in relation to the above-mentioned issues, in [7,8], the arc overvoltage ratio is determined from the ground fault resistance, in [9,10], the arc overvoltage ratio is determined from the ground fault angle, and in [11,12], the arc overvoltage ratio is determined. In [13,14], the dependence of the frequency on the phase capacity of the network with respect to the ground, and in [13,14], the analytical expressions for the dependences of the arc overvoltage on the resistance of the network to the ground and the angle to the ground have already been considered.

The compact regression models of arc overvoltage obtained in the mentioned works practically simplify the process of controlling its value during equipment tests. It is important to obtain dependences on other parameters for adjusting the value of arc overvoltage and to continue

research in this direction. As a continuation of the conducted research, the regression model of the dependence of the ground fault resistance of the non-stationary arc overvoltage and the phase capacity of the network with respect to the ground is considered.

III. Problem solving method and algorithm

During single-phase earth faults in the neutral isolated electrical network obtaining an analytical expression for the dependence of the frequency of the resulting arc overvoltage on the ground fault resistance and the phase capacitance of the network with respect to the ground is considered. For this purpose, the results of the experimental studies carried out in the low-voltage model of the neutral-isolated network, given in table 1, are used $\left(\varphi = 90^{\circ}\right)$ [6].

Table 1: $k = f\left(R_0, C_f\right)$ *addiction*

R_0, Om	C_f, mkF						
	1	2	3	4	5	6	8
5	3.30	2.99	2.85	2.74	2.66	2.58	2.47
10	2.96	2.74	2.59	2.47	2.39	2.33	2.21
15	2.77	2.54	2.39	2.28	2.20	2.14	2.08
20	2.62	2.38	2.24	2.13	2.06	2.01	1.92
25	2.49	2.25	2.12	2.02	1.95	1.91	1.84
30	2.38	2.15	2.02	1.93	1.88	1.84	1.80

It should be noted that the proposed forms of regression models of arc overvoltage in [13,14] greatly simplify the issues of its management from the point of view of its technical implementation, and therefore it was proposed to use that form in the work under consideration. The data in Table 1 and the conducted analyzes showed that the dependence between the frequency of the earth fault arc overvoltage and the earth fault resistance and the phase capacity of the network with respect to the earth can be approximated by the hyperbolic regression equation expressing the following theoretical-probability scheme (NES) [7-14,15]:

$$k = \frac{a}{R_0} + \frac{b}{C_f} + c, \tag{1}$$

Here $a, b, c -$ called regression coefficients.

(1) to simplify the matter in $\frac{1}{R_0} = x$ and $\frac{1}{C_f} = y$ if we accept substitutions (1), NES is expressed by the following equation:

$$k = ax + by + c, \tag{2}$$

In other words, the frequency of arc overvoltage (k) with the conductance of the ground fault circuit (x) and the inverse value of the phase capacitance of the network with respect to earth (y) NES between can be approximated by the linear regression equation. The values of the new (x, y) quantities are given in table 2.

Table 2: $k = f(x, y)$ *addiction*

x	y						
	1.00	0.50	0.333	0.250	0.200	0.167	0.125
0.200	3.30	2.99	2.85	2.74	2.66	2.58	2.47
0.100	2.96	2.74	2.59	2.47	2.39	2.33	2.21
0.067	2.77	2.54	2.39	2.28	2.20	2.14	2.08
0.050	2.62	2.38	2.24	2.13	2.06	2.01	1.92
0.040	2.49	2.25	2.12	2.02	1.95	1.91	1.84
0.033	2.38	2.15	2.02	1.93	1.88	1.84	1.80

In this case, the least squares method is used to determine the regression coefficients [15, p.55-59].

It is known that, according to this method, it is required to choose the values of the regression coefficients in such a way that the sum of the squares of the errors between the approximating function and the experimental values is the smallest or minimum:

$$S(a,b,c) = \sum_{i=1}^{n} (ax_i + by_i + c - k_i)^2 \to \min, \tag{3}$$

$S(a,b,c)$for the function to get the smallest (minimum) value a, b and $c-$ according to which the special derivatives given below must be equal to zero:

$$\begin{cases} \dfrac{\partial S(a,b,c)}{\partial a} = 0; \\ \dfrac{\partial S(a,b,c)}{\partial b} = 0; \\ \dfrac{\partial S(a,b,c)}{\partial c} = 0. \end{cases} \tag{4}$$

$S(a,b,c)$ from the function$a, b$and$c$ −taking special derivatives according to (4), taking into account in the system of equations, the following system of linear equations is obtained:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} k_i x_i; \\ a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i^2 + c\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} k_i y_i; \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i + cn = \sum_{i=1}^{n} k_i. \end{cases} \tag{5}$$

To solve the system of equations (5), the correlation table given in table 3 is used. In Table 3, the following notations are adopted:

$$A_i = \left| \frac{ax_i + by_i + c - k_i}{k_i} \right|; \quad B_i = (ax_i + by_i + c - k_i)^2$$

Optional data array: $n = 42;$

$$\sum_{i=1}^{n} x_i = 3{,}43; \ \sum_{i=1}^{n} y_i = 15{,}45; \ \sum_{i=1}^{n} k_i = 97{,}62;$$

$$\sum_{i=1}^{n} x_i^2 = 0{,}4176; \ \sum_{i=1}^{n} y_i^2 = 9{,}0421; \ \sum_{i=1}^{n} k_i^2 = 232{,}157;$$

Table 3: *Table showing correlation report sequence*

i	x_i	y_i	k_i	x_i^2	y_i^2	k_i^2	$x_i y_i$	$k_i x_i$	$k_i y_i$	A_i	B_i
1	0.200	1,000	3.30	0.040	1.000	10.890	0.200	0.660	3,300	0.0139930	0.0021323
2	0.100	1,000	2.96	0.010	1.000	8.762	0.100	0.296	2,960	0.0233497	0.0047769
3	0.067	1,000	2.77	0.004	1.000	7.673	0.067	0.185	2,770	0.0111477	0.0009535
4	0.050	1,000	2.62	0.003	1.000	6.864	0.050	0.131	2,620	0.0165034	0.0018696
5	0.040	1,000	2.49	0.002	1.000	6.200	0.040	0.100	2,490	0.0512890	0.0163098
6	0.033	1,000	2.38	0.001	1.000	5.664	0.033	0.079	2,380	0.0871247	0.0429969
7	0.200	0.500	2.99	0.040	0.250	8.940	0.100	0.598	1,495	0.0086820	0.0006739
8	0.100	0.500	2.74	0.010	0.250	7.508	0.050	0.274	1,370	0.0843982	0.0534770
9	0.067	0.500	2.54	0.004	0.250	6.452	0.033	0.169	1,270	0.0720532	0.0334945
10	0.050	0.500	2.38	0.003	0.250	5.664	0.025	0.119	1,190	0.0415534	0.0097806
11	0.040	0.500	2.25	0.002	0.250	5.063	0.020	0.090	1,125	0.0064117	0.0002081
12	0.033	0.500	2.15	0.001	0.250	4.623	0.017	0.072	1,075	0.0256842	0.0030494
13	0.200	0.333	2.85	0.040	0.111	8.123	0.067	0.570	0.950	0.0046799	0.0001779
14	0.100	0.333	2.59	0.010	0.111	6.708	0.033	0.259	0.863	0.0805520	0.0435263
15	0.067	0.333	2.39	0.004	0.111	5.712	0.022	0.159	0.797	0.0671103	0.0257261
16	0.050	0.333	2.24	0.003	0.111	5.018	0.017	0.112	0.747	0.0385159	0.0074435
17	0.040	0.333	2.12	0.002	0.111	4.494	0.013	0.085	0.707	0.0055683	0.0001394
18	0.033	0.333	2.02	0.001	0.111	4.080	0.011	0.067	0.673	0.0286348	0.0033457
19	0.200	0.250	2.74	0.040	0.063	7.508	0.050	0.548	0.685	0.0120339	0.0010872
20	0.100	0.250	2.47	0.010	0.063	6.101	0.025	0.247	0.618	0.0616676	0.0232011
21	0.067	0.250	2.28	0.004	0.063	5.198	0.017	0.152	0.570	0.0500364	0.0130149
22	0.050	0.250	2.13	0.003	0.063	4.537	0.013	0.107	0.533	0.0187629	0.0015972
23	0.040	0.250	2.02	0.002	0.063	4.080	0.010	0.081	0.505	0.0121316	0.0006005
24	0.033	0.250	1.93	0.001	0.063	3.725	0.008	0.064	0.483	0.0436026	0.0070817
25	0.200	0.200	2.66	0.040	0.040	7.076	0.040	0.532	0.532	0.0281050	0.0055890
26	0.100	0.200	2.39	0.010	0.040	5.712	0.020	0.239	0.478	0.0462479	0.0122174
27	0.067	0.200	2.20	0.004	0.040	4.840	0.013	0.147	0.440	0.0328621	0.0052268
28	0.050	0.200	2.06	0.003	0.040	4.244	0.010	0.103	0.412	0.0039702	0.0000669
29	0.040	0.200	1.95	0.002	0.040	3.803	0.008	0.078	0.390	0.0288678	0.0031688
30	0.033	0.200	1.88	0.001	0.040	3.534	0.007	0.063	0.376	0.0510316	0.0092044
31	0.200	0.167	2.58	0.040	0.028	6.656	0.033	0.516	0.430	0.0501100	0.0167143
32	0.100	0.167	2.33	0.010	0.028	5.429	0.017	0.233	0.388	0.0326216	0.0057773
33	0.067	0.167	2.14	0.004	0.028	4.580	0.011	0.143	0.357	0.0176506	0.0014267
34	0.050	0.167	2.01	0.003	0.028	4.040	0.008	0.101	0.335	0.0081322	0.0002672
35	0.040	0.167	1.91	0.002	0.028	3.648	0.007	0.076	0.318	0.0370767	0.0050150
36	0.033	0.167	1.84	0.001	0.028	3.386	0.006	0.061	0.307	0.0600346	0.0122022
37	0.200	0.125	2.47	0.040	0.016	6.101	0.025	0.494	0.309	0.0839834	0.0430309
38	0.100	0.125	2.21	0.010	0.016	4.884	0.013	0.221	0.276	0.0054964	0.0001476
39	0.067	0.125	2.08	0.004	0.016	4.326	0.008	0.139	0.260	0.0046235	0.0000925
40	0.050	0.125	1.92	0.003	0.016	3.686	0.006	0.096	0.240	0.0388026	0.0055504
41	0.040	0.125	1.84	0.002	0.016	3.386	0.005	0.074	0.230	0.0592238	0.0118749
42	0.033	0.125	1.80	0.001	0.016	3.240	0.004	0.060	0.225	0.0658995	0.0140705
⚜	**3.430**	**15.450**	**97.62**	**0.4176**	**9.0421**	**232.157**	**1.2618**	**8.5982**	**38.477**	**1.5202260**	**0.4483068**

$$\sum_{i=1}^{n} x_i y_i = 1{,}2618;\ \sum_{i=1}^{n} k_i x_i = 8{,}5982;\ \sum_{i=1}^{n} k_i y_i = 38{,}477;$$

$$\sum_{i=1}^{n} \left| \frac{ax_i + by_i + c - k_i}{k_i} \right| = 1{,}520226;\ \sum_{i=1}^{n} \left(ax_i + by_i + c - k_i\right)^2 = 0{,}4483068.$$

(5) the system of linear equations is solved by one of the known methods and the regression coefficients are found, and the following values were obtained in the considered sample:

$$a = 4{,}55;\ b = 0{,}76;\ c = 1{,}67.$$

Thus, after determining the regression coefficients, the NES (2) between the frequency of the arc overvoltage generated in neutral insulated networks as a result of non-stationary earth faults and the conductance of the earth fault circuit and the inverse value of the phase capacitance of the network with respect to earth is written in the following obvious way:

$$k = 4{,}55x + 0{,}76y + 1{,}67 \tag{6}$$

Pairwise linear correlation coefficients are defined by the following well-known expressions:

$$\left.\begin{aligned} r_{xy} &= \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}; \\ r_{kx} &= \frac{\overline{kx} - \bar{k} \cdot \bar{x}}{\sigma_k \sigma_x}; \\ r_{ky} &= \frac{\overline{ky} - \bar{k} \cdot \bar{y}}{\sigma_k \sigma_y}. \end{aligned}\right\} \tag{7}$$

Here $\bar{x},\ \bar{y},\ \bar{k}$ – properly $x,\ y,\ k$ average prices of quantities; $\overline{xy},\ \overline{kx},\ \overline{ky}$ – properly average prices of products; $\sigma_x,\ \sigma_y,\ \sigma_k$ – properly $x,\ y,\ k$ mean square deviations of their quantities are determined according to the correlation table (Table 3):

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 0{,}082;\ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = 0{,}368;\quad \bar{k} = \frac{\sum_{i=1}^{n} k_i}{n} = 2{,}324;$$

$$\overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} = 0{,}03;\ \overline{kx} = \frac{\sum_{i=1}^{n} k_i x_i}{n} = 0{,}205;\quad \overline{ky} = \frac{\sum_{i=1}^{n} k_i y_i}{n} = 0{,}916;$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2}}{n} - \bar{x}^2} = 0{,}057;\ \sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^{\,2}}{n} - \bar{y}^2} = 0{,}283;\ \sigma_k = \sqrt{\frac{\sum_{i=1}^{n} k_i^{\,2}}{n} - \bar{k}^2} = 0{,}354;$$

Then according to statements (7):

$$r_{xy} = 0;\ r_{kx} = 0{,}736;\ r_{ky} = 0{,}611.$$

After that, no-conductivity of the earth fault circuit with the frequency of overvoltage

during stationary earth faults and the inverse value of the phase capacitance of the network with respect to earth it is required to check the adequacy of the NES regression equation obtained between (6). For this, by calculating the multidimensional correlation coefficient, its significance can be checked with the well-known Fisher criterion [15].

The value of the multivariate correlation coefficient is determined by the following well-known expression:

$$R = \sqrt{\frac{r_{kx}^2 + r_{ky}^2 - 2r_{kx}r_{ky}r_{xy}}{1 - r_{xy}^2}} = 0{,}96.$$

The multivariate correlation coefficient is close to unity $(R = 0{,}96 \to 1)$ indicates that the dependence between the frequency of arc overvoltage and the conductance of the ground fault circuit and the inverse value of the phase capacitance of the network with respect to ground can be considered a strong linear correlation relationship.

Determining the significance of the multivariate correlation coefficient and the full adequacy of the model in general is traditional $F-$Fisheris checked with the criterion. It is known that α at the level of significance, the regression equation is considered adequate if $F > F(\alpha, q_1, q_2)$ be paid conditionally [16], here $q_1, q_2 -$ are the degrees of freedom.

$F-$Fisher the reporting value of the criterion is determined as follows based on the data:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}, \tag{8}$$

Here $n-$number of experiments, $n = 42;$ $m-$is the number of factors, $m = 2.$ Then according to statement (8). $F = 229{,}22$.

$F-$Fisher the table value of the criterion is the level of significance (α) and degrees of freedom (q_1, q_2) depending is taken from the table [16]:

$$\alpha = 0{,}05;\ q_1 = m = 2;\ q_2 = n - m - 1 = 42 - 2 - 1 = 39;\ F(\alpha, q_1, q_2) = 3{,}24.$$

$F = 229{,}22 > F(\alpha, q_1, q_2) = 3{,}24$ since multivariate correlation coefficient $(R = 0{,}96)$ and the statistical significance of the regression equation is confirmed.

The mean relative and mean square errors of the obtained approximation are determined by the following well-known expressions, respectively:

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{ax_i + by_i + c - k_i}{k_i}\right| \cdot 100\% = 3{,}62\%;$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(ax_i + by_i + c - k_i)^2}{n}} = 0{,}103.$$

Let's calculate the determination index to determine the degree of change (variability) of the arc overvoltage depending on the mentioned quantities for the particular sample under consideration. The price of the determination index $R^2 = 0{,}96^2 = 0{,}9216$ presence indicates that the frequency of arc overvoltage (k) of change $92{,}16\ \%$ - is the conductance of the ground fault

circuit (x) and the inverse value of the phase capacitance of the network with respect to ground (y) at the expense of change, the rest $7{,}84\ \%$- occurs due to the change of other random factors that are not taken into account, and this case k, x, y It once again confirms that the obtained correlation relationship between the quantities is great and the possibility of being used successfully during management.

We use average conductivity coefficients to separately determine the degree of influence of ground fault circuit conductance and network phase capacitance to ground on the frequency of arc overvoltage. Average elasticity coefficients are determined by the following well-known expressions [16]:

$$\overline{E}_x = \frac{\partial k}{\partial x}\frac{\overline{x}}{\overline{k}} = a\frac{\overline{x}}{\overline{k}} = 0{,}16;$$

$$\overline{E}_y = \frac{\partial k}{\partial y}\frac{\overline{y}}{\overline{k}} = b\frac{\overline{y}}{\overline{k}} = 0{,}121.$$

It follows that the conductivity of the ground fault circuit (x) 1% increase in the frequency of arc overvoltage (k) 0,16 % increase, the inverse value of the phase capacity of the network with respect to the ground (y) 1% increase in the frequency of arc overvoltage causes (k) an increase of 0.121% . $\overline{E}_x = 0{,}16 > \overline{E}_y = 0{,}121$ since the conductance of the ground fault circuit (x) to the inverse value of the phase capacitance of the network with respect to ground (y) relative to the frequency of arc overvoltage (k) has more effect. Average elasticity coefficients less than 1% of the arc extreme stress (k) conductance of the ground fault circuit (x) and the inverse value of the phase capacitance of the network with respect to ground (y) shows that it is not flexible to change. However, the obtained expression (6) can be effectively used during load tests of network isolation, and its validity is not in doubt.

IV. Results of computer modeling of regression dependence

Based on the obtained results, the dependence (1) between the frequency of the arc overvoltage and the ground fault resistance and the phase capacity of the network with respect to the ground in neutral insulated networks as a result of non-stationary earth faults can be written in the following obvious hyperbolic way according to the expression (6):

$$k = \frac{4{,}55}{R_0} + \frac{0{,}76}{C_f} + 1{,}67 \qquad (9)$$

Based on the regression equation (9) obtained using the OriginLab [17] software complex, a 3D (spatial) image of the dependence of the frequency of arc overvoltage on the ground fault resistance and the phase capacity of the network with respect to the ground was constructed (Figure 1). As can be seen from the figure, the value of the arc overvoltage during artificial earth faults, depending on the earth fault resistance and the capacity of the network, for the organization of tests of electrical equipment under load in networks with neutral isolation, is controlled in the (1.9-3.3) U_{nom} interval, depending on the nature of the test problem. can be done.

(9) regression dependence, figure 1 and the dependences obtained in the authors' previous works [7-14] can be used practically and effectively during load tests of electrical equipment and network insulation, while the reliability of the obtained results is ensured.

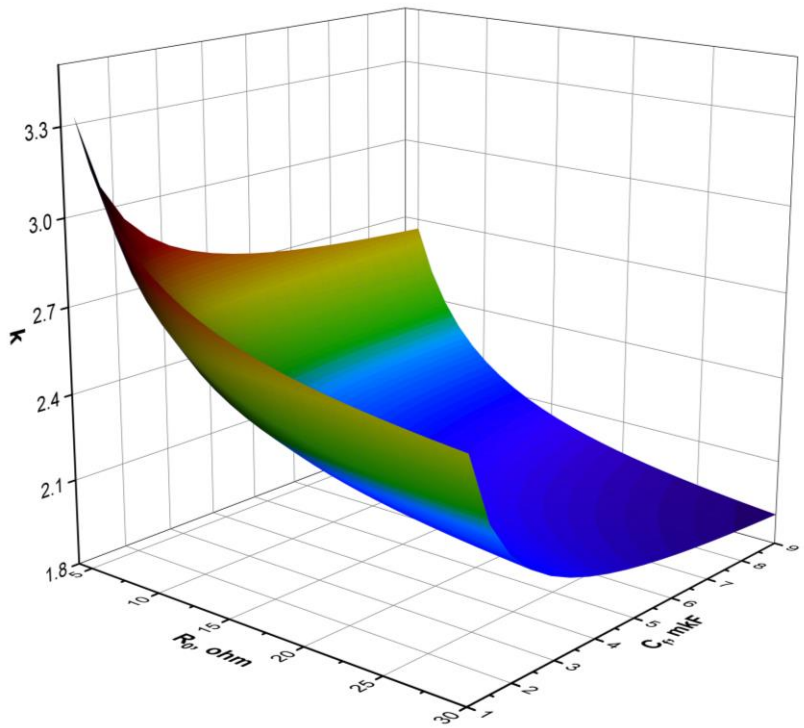


Figure 1: *Earth fault resistance and network overvoltage frequency 3D image of phase capacitance dependence with respect to ground*

IV. Conclusion

1. An adequate hyperbolic regression model, which can be easily implemented in practice, has been obtained between the frequency of arc overvoltage generated in neutral insulated networks as a result of non-stationary earth faults subject to Petersen's theory, the conductivity of the earth fault circuit and the inverse value of the phase capacitance of the network with respect to earth. The model adequacy test results confirm that the mathematical model expressing the proposed analytical dependence between the mentioned location and the closure parameters has a strong linear correlation relationship.

2. The obtained regression dependence and theoretical results can be successfully and easily implemented in the management of non-stationary earthing arc overvoltage during isolation under load in the neutral insulated networks of the Azerbaijani power system in transformation conditions.

References

[1] Khazieva, R.T., Vasiliev, P.I., Aflyatunov, R.R. A study to establish the isolization of electrical equipment in the pierced area. Moscow: Electrical technology in passing, 2022. No. 3(56), c. 65-69.

[2] Patsch, R. Dielectric Diagnostics of Power Transformers and Cables - Return Voltage Measurements, Theory and Practical Results. VDE High Voltage Technology, 2018, ETG Symposium. URL: https://www.vde-verlag.de/proceedingsde/454807130.html. (date of circulation 21.07.2023).

[3] Shalyt G.M. Preventive testing of cable networks 6-10 kV under load. Electric station, 1958, No8, p. 66-70.

[4] Baranov B.M. About preventive tests of cable network isolation. Electric stations, 1961,

No2, p. 55-59.

[5] Dudarev L.E. Preventive tests of the isolation of networks under load by the artificially created overvoltage method. Electricity, 1978, No8, p. 69-72.

[6] Orucov, N.I. Development and research of isolation test methods in neutral isolated networks based on artificial non-stationary earth faults: / Dissertation submitted to receive the degree of Candidate of Technical Sciences, Baku, 1998, 139 p.

[7] Orucov, N.I., Orucov, AO Determining the dependence between the frequency of arc overvoltage and ground fault resistance // - Baku: Scientific works. AzTU, 2013. No. 3, pp. 83-86.

[8] Orucov, N.I. The regression model of the dependence of arc overvoltage on grounding resistance // "The role of engineering in the innovative development of Azerbaijan: goals and perspectives" international scientific-practical conference, Baku: BMU, November 29-30, 2019, p. 285-287.

[9] Orucov, N.I., Orucov, AO Determination of the dependence between the arc extreme voltage and the closing angle with the ground. Baku: Problems of Energetics, Elm, 2014. No. 2, p. 36-39.

[10] Orucov, N.I., Mirili, T. The regression model of the dependence of the arc extreme voltage on the ground contact angle. V international scientific conference of young researchers, Baku, Book 1. BMU, April 29-30, 2021, p.113- 116.

[11] Orucov, N.I. Determining the dependence between the frequency of arc overvoltage and the phase capacity of the network with respect to ground. Materials of the Republican Scientific and Technical Conference on "Building Education – Research, Production Mechanism" dedicated to the 100th anniversary of the Azerbaijan People's Republic, Baku: AzTU, April 4-5, 2018, p. 254-257

[12] Orucov, N. I. The regression model of the dependence of the arc overvoltage on the phase capacity of the network with respect to the ground. II international science and technology conference, Baku: BMU, November 26-27, 2021, p.333-335.

[13] Orucov, N.I., Guliyev, H.B., Alimammadova, S.J. Regression model of arc overvoltage during single-phase non-stationary ground faults in neutral insulated networks. Scientific news of AzTU, No. 1, 2024, p.43-48.

[14] Orucov N.I., Guliyev H.B., Alimammadova S.J. Regression model of arc overvoltage during single-phase non-stationary ground faults in neutral isolated networks. Reliability: Theory & Applications, Vol.19, No.2(78), USA, San Diego, 2024, p.425-433.

[15] Mammadov, N.R., Mammadov B.M. Mathematical processing of the results of the experiment. Baku: Elm, 2005, 160 p.

[16] Orucov, E. Econometrics. Baku: Elm, 2018, 384 p.

[17] Origin(Pro), "Version 2022". OriginLab Corporation, Northampton, MA, USA.

METHODOLOGY OF ASSESSING SOCIAL DAMAGE FROM LONG-TERM SMOKE DURING FIRES IN MOUNTAIN FOREST BELTS OF RUSSIA

D.S. Kovaleva[1], A.A. Dolgov[2]

•

[1,2] Federal state budgetary establishment
"All-Russian scientific research institute for civil defense and emergencies
of the emercom of Russia" (Federal science and high technology center)
(VNII GOChS (FC))
[1] d.tsomseva@ya.ru, [2] dolaa@rambler.ru

Abstract

The article describes a methodology for assessing social damage during fires in mountain forest belts of the Russian Federation, associated with an increase in the overall mortality of the population as a result of long-term and intense smoke in urbanized areas. The relevance of the topic and the demand for the results are associated with the growing number of forest fires, including in mountainous areas, with changing consequences, both for the ecology of regions and human economic activity, and for the life and health of the population, changing consistent long-term smoke pollution of urbanized areas. Recently, many countries have been paying more and more attention to the pollution of the atmosphere of populated areas, namely, air quality is a determining factor for the health and life expectancy of the population. The methodology presented in the article allows us to estimate the concentration of fine particles in space at a given distance from a forest fire and to estimate the possible social damage associated with the formation of general mortality as a result of smoke pollution. An example of testing this methodology is given using the example of long-term smoke in Moscow in 2010.

Keywords: forest fire, social damage, smoke, fine particles.

I Introduction

In the Russian Federation, more than one third of the territory is mountainous - these are the Urals, the North Caucasus, the mountains of Southern Siberia, the mountains of Eastern Siberia, and the Far East. Mountain forests of deciduous and coniferous trees in our country and neighboring countries occupy a huge area – about 45% of the total area of the country's forest fund.

They are characterized by vertical zoning with a slope steepness of ridge slopes with increasing altitude.

Recently, more and more attention has been paid to air pollution in populated areas in our country, since air quality is a key factor influencing the health and life expectancy of the population. Numerous studies have shown that the main contribution to the deterioration of health and the increase in overall mortality of the population is made by finely dispersed substances with an aerodynamic diameter of 10 micrometers or less (hereinafter – PM10) and 2.5 micrometers or less (hereinafter – PM2.5) in the atmosphere of urbanized areas. One of the main sources of fine particles PM10 and PM2.5 in the atmosphere are wildfires. Studies show [1] that about 44% of all fine particle emissions per year come from forest fires. Fires release significant

amounts of smoke and ash containing significant concentrations of these fine particles, which spread over significant distances in the atmosphere. The US National Institute for Health Effects Research analyzed morbidity and mortality statistics and established a quantitative relationship between air pollution by suspended particles with an aerodynamic diameter of less than 10 μm – PM10 and overall mortality. With an increase in the average daily concentration of dispersed particles by every 10 μg/m^3, daily mortality from all causes, except accidents, increased by an average of 0.5% the next day. Similar studies were conducted in 30 regions of Europe. The results of studies in Europe are in good agreement with the results obtained earlier in the USA [2]. In the works of the Chief State Sanitary Doctor of Russia (1996-2013) Onishchenko G.G [3] it was also proposed to calculate the increase in the average daily mortality from all causes, except accidental deaths, by 0.5% with an increase in the average daily concentration of dispersed particles by every 10 μg/m3. Epidemiological studies conducted in Western Europe have shown that the statistical relationship between mortality and the levels of PM2.5 concentration in the air is much stronger than with the levels of PM10 particle concentration (average annual concentration of suspended matter) [4]. Thus, with an increase in the average daily concentration of PM2.5 dispersed particles by every 5 μg/m3, the average daily mortality from all causes, except accidental deaths, increases by 0.7%.

II Main part

Substances emitted as a result of a forest fire in the form of gases and aerosols form a cloud that spreads in the direction of the wind. When gas contamination is detected in advance using the forecasting method, the most probable and closest to real meteorological conditions at different times of the year (day) are used as initial data. The possible type of forest fire is determined. It is recommended to calculate the gas contamination of the territory based on the worst meteorological conditions in terms of the formation of gas contamination zones and the impact of combustion products: wind speed; atmospheric stability; time of day; seasonality; the settlement (economic facility) is located on the axis of the trace of the spread of combustion products. Based on the values of the increase in mortality with an increase in the concentration of fine particles in the atmosphere of urbanized areas obtained in the works [2, 4], it is assumed that:

- with an increase in the average daily concentration of dispersed particles PM10 for every 10 μg/m3, the average daily total mortality from all causes, except accidental deaths, increases by 0.5%.
- with an increase in the average daily concentration of dispersed particles PM2.5 for every 5 μg/m3, the average daily total mortality from all causes, except accidental deaths, increases by 0.7%.

The initial data for the calculations are:

1. the area of the forest fire (hereinafter referred to as FF) upon detection,
2. wind speed as measured at the weather station for a height of 10 m,
3. the stock of forest combustible materials (hereinafter referred to as FCM) (tables 1-2),
4. the height of the forest tiers (tables 1-2),
5. the moisture content of FCM,
6. the population of the urbanized area under consideration.

Assumptions made:

1. The ground FF spreads with the direction of the wind, across the direction of the wind and against the wind.
2. The crown FF spreads only with the direction of the wind.
3. A crown fire exists only in the presence of a ground fire.
4. The contour of the FF is an ellipse. The major axis is directed downwind.
5. Combustion products are released only from the combustion zone along the perimeter of the fire. The width of the combustion zone is different in different zones of the fire perimeter.
6. The flame speed in different zones of the fire perimeter is different.
7. The burning time of the fire perimeter sections is the same for all sections.

Wind speed in a forest at a standard height of 10 m [4]:

— , (m/s) (1)

Where:

h_1 – height of the first tier of forest massif (tables 1-2);

h_3 – height of the third tier of forest massif (tables 1-2);

N – empirical multiplier (table 3);

– wind speed as measured at a weather station for a height of 10 m.

Table 1: *Forest fuel model [5]*

The main conductor of combustion is dried grass	
Model 1. (low grasses)	Areas covered with dried or drying grasses (mainly cereals) about 0.3 m high; the canopy density of trees and shrubs does not exceed 0.3 (meadows, savannas, grassy tundra)
Model 2. (forest grasses and undergrowth)	Areas with a layer of forest fuel material in the form of a mixture of dried grasses and fallen trees and shrubs. These are open shrub landscapes, pine forests, stands of coppice oak; canopy density is 0.3-0.6
Model 3. (tall grasses)	Areas covered with dried or drying grasses about 1 m high. These are areas with wild and cultivated cereals, as well as tall grasses in swamps.
The main conductor of combustion is the canopy of bushes or their litter	
Model 4 (bushes)	Dense thickets of flammable bushes about 2 m high and undergrowth up to 6 m, as well as plantations with a dense flammable layer of undergrowth and undergrowth. There is a significant reserve of ground forest fuel materials (litter, dried grass).
Model 5 (low shrubs)	Thickets of non-flammable (young) bushes or plantations with a dense layer of non-flammable undergrowth and young trees, deciduous young trees. The combustion conductor is litter
Model 6 (drying shrubs and cluttered hardwoods)	Areas similar to those in model 4, but the bushes are not as tall and do not contain much fuel, so they can only burn actively when the wind speed is over 4 m/s
Model 7 (southern weeds)	Thickets of flammable shrubs from 0.5 to 2 m in height, as well as plantations with a low flammable layer of undergrowth and young growth
The main conductor of combustion is forest litter and deadwood	
Model 8 (closed forest litter)	Closed stands of short-needled trees (spruce, fir, white pine, larch) and closed stands of aspen and birch, as well as hardwoods (in summer). The main fuel is compacted litter.
Model 9 (hardwood litter)	Closed stands of long-needle trees (Ponderosa pine, red pine, southern pine, etc.), as well as oak and maple (in winter). The main combustibles are loose litter.
Model 10 (forest litter and undergrowth)	Plantations littered due to damage by insects, wind or due to overmaturity. Old clearings in coniferous plantations.
The main conductor of combustion is logging residues	
Model 11 (light logging residues)	Non-continuous fellings, thinning areas with undeleted logging residues, usually in coniferous and hardwood stands
Model 12	Clear and conditionally clear cuts with a significant stock of logging

(medium logging residues)	residues (up to 16 t/ha with a layer height of 0.7 m). Fires are strong, "spotted"
Model 13 (heavy logging residues)	Clear-cut areas in mature and overmature coniferous stands, heavily littered, mainly with large logging residues (up to 90 t/ha). Fires are strong, "spotted"

Table 2: *Characteristics of forest layers [5, 6]*

Model No.	h (height of the tier of the FCM layer)	m (mass of the FCM stock) kg/m²
1st tier, h_1 – mosses and lichens		
-	0,15	3
2nd tier, h_2 – herbal group		
1	0,3	0,18
2	0,3	0,98
3	0,8	0,74
3rd tier, h_3 – shrub group and undergrowth		
4	6,0	3,92
5	0,6	0,84
6	1,9	1,48
7	0,8	1,39
4th tier, h_4 – crown		
8	20-30	6,3
9	20-30	7,06
10	20-30	8,08
Group of clearings		
11	0,3	2,80
12	0,7	8,49
13	0,9	14,3

Table 3: *N value depending on forest canopy density [6]*

Forest canopy closure	The value of the empirical factor N
0	0,77
0,2	0,62
0,4	0,47
0,6	0,32
0,8	0,165
1	0,09

Wind speed in the forest in the tree crowns (in the fourth tier of the forest) [6]:

, (m/s) (2)

where N is an empirical factor, equal in this case to 0.77;
The burning time at the time of detection of the forest fire:

$$\sqrt{\frac{}{\pi\cdot(\omega \quad _B)\cdot\omega}}\ (s) \tag{3}$$

where:
ω_A – the speed of propagation of the lower FF downwind, m/s, (formulas 4-5);
ω_B – the speed of propagation of the lower FF against the wind, m/s (formulas 4-5);

ω_C – the speed of propagation of the lower FF perpendicular to the wind, m/s (formulas 4-5);

Calculation of the speed of propagation of the lower and upper FF ω_n (ω_A, ω_B, ω_C), which takes into account the steepness of the slope, is calculated using the formula of E.V. Konev [7]:

$$\frac{{}_1(U\ \alpha)\cdot(m_S/m_0)^n}{\left[1\quad\frac{(\rho\quad 0)}{}\right]\cdot[1\quad\cdot(\quad 0)+D\cdot(T\quad 0)]},\ (m/s) \tag{4}$$

$$\begin{array}{l} \left(\alpha\quad -\right)\qquad (0),\quad \left(\alpha\quad -\right)\qquad (\pi) \\ [B\ (0)\qquad\quad (U\qquad)]\qquad\quad (U\qquad) \\ (\nu\quad\alpha)\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \\ \qquad\qquad\qquad (\pi)\qquad\qquad (U\qquad) \end{array} \tag{5}$$

$$\begin{array}{l} \lfloor\qquad\qquad (U\quad\alpha)\qquad\qquad\qquad\qquad - \\ (\beta)\quad \begin{cases} (1\qquad 8°)\ [1\qquad A\,(\beta\quad 8°)] \\ \qquad (1\qquad\beta) \end{cases} \end{array} \tag{6}$$

where:

ω_n – the speed of spread of the ground fire front on the horizontal underlying surface;

m_s – the stock of forest fuel materials on the underlying surface;

m_0 – the stock of forest fuel materials on the underlying surface corresponding to the speed of spread $\omega=\omega_0$;

G_f = 0.3;

ρ_0 – the density of the layer on the horizontal surface at the speed of spread $\omega=\omega_0$;

n = (0.2-0.35) – empirical coefficient;

ν' – wind speed pulsation (ν'~1 m/s);

β – angle of inclination to the horizon of the terrain;

α – angle between the direction of the speed of spread of the fire front $\vec{\omega}$ and the wind speed;

C, D, $B1$, $B2$, $B3$, $B4$, a, b, ω_{n0} – empirical constants [6];

T_0, W_0 – standard initial temperature of the horizontal layer FCM and moisture content, respectively.

A ground fire can transform into a crown fire or a summit crown fire if the following condition is met [6]:

$$\sqrt{m} \tag{7}$$

where:

h_l – height of the lower boundary of the forest canopy;

$H_{h.f.l.}$ – height of the flame of the lower FF (formula 19);

k_e – empirical coefficient equal to 8 for rags, grass and litter.

The area of the front of the lower FF during further spread of the fire at time t is calculated by the formula:

(m²) (8)

where:

L_l – the perimeter of the lower FF contour (ellipse) at time t, m (formula 10);

Δ^l – the width of the combustion front of the lower FF, m (formula 14).

The area of a widespread crown forest fire is determined by the formula [8]:

$$- \ (\omega_A^c + \omega_B) \cdot \omega_C \cdot t^2 \ (m^2) \quad (9)$$

Where ω^c_A is the speed of spread of a crown forest fire by wind.
The perimeter of the lower FF contour at time t:

$$(t \quad t)[0\ 75(\omega \quad) \quad 71\sqrt{(\omega \quad)\omega_C}] \ (m) \quad (10)$$

The width of the combustion front of the lower FF in the direction of the wind is calculated using the formula of professor Albini [9]:

$$\left(\frac{4\cdot(U_l)^2}{}\right) \quad (m) \quad (11)$$

where:
h^* – height of the forest layer involved in the FF, m (tables 1-2);
g – acceleration of gravity.
Width of the combustion front of the lower FF across the wind direction:

$$\left(\frac{c}{}\right) \quad (m) \quad (12)$$

Width of the combustion front of the lower FF against the wind direction:

$$\left(\frac{B}{}\right) \cdot \Delta_a^l \ (m) \quad (13)$$

Average width of the combustion front of the lower FF:

$$\text{———} \ (m) \quad (14)$$

Mass burnout rate at
low FF:

$$\frac{K\cdot(\Sigma_l^I \quad _i)\cdot\omega}{} \ (kg/m^2\cdot s) \quad (15)$$

crown FF:

$$\text{———} \ (kg/m^2\cdot s) \quad (16)$$

where:
n – number of forest layers involved in the combustion process;
K – coefficient of completeness of forest fuel combustion (formula 18);
m_{si} – forest fuel reserve in the l-th forest layer, kg/m² (for crown fire m_s – for the 4th layer, for ground fires m_s – for forest layers involved in the fire) (tables 1-2).
In case of a widespread crown fire, the mass burning rate:

$$(kg/m^2\cdot s) \quad (17)$$

Combustion efficiency coefficient FCM:

$$\text{——} \quad (18)$$

where:
W – moisture content of the fuel;

W_* – critical moisture content of the fuel, upon reaching which the fuel does not burn. For crown FF W_*=0.9.

The height of the torch of the low FF is calculated using the formula [5]:

$$k \cdot \sqrt{\sum_{l=1}^{L} m_{sl} \cdot \omega} \text{ (m)} \quad (19)$$

Where ω is the average speed of propagation of the lower FF:

——————— (m/s) (20)

The height of a crown FF is calculated using the formula:

$$k \cdot \sqrt{\sum_{l=1}^{L} m_{sl} \cdot \omega^{c}} \text{ (m)} \quad (21)$$

Height of convective column during FF [10]:

$$\cdot \sqrt{\frac{}{v \cdot |1 \quad \gamma|}} \quad \text{(m)} \quad (22)$$

where:
T_f – flame temperature, K (table 4);
γ – air temperature gradient, K/m [11];
S_* – area of the combustion front of a ground fire or crown fire (S_B) m²;
V_m – mass burnout rate of a ground fire ($V_{m.l.}$) or crown fire ($V_{m.c.}$).

Table 4: *The value of flame temperature depending on the type of forest fire*

Type of forest fire		Flame temperature T_f (K)
Ground fire	Flame of fire	1104
	Smoldering coals	964
Crown fire		1224

Intensity of emission of the i-th combustion product during a FF:

$$\sum \quad \text{——————} \text{ (kg/s)} \quad (23)$$

where:
k_i – emission coefficient of the i-th pollutant (table 5);
n – number of forest layers involved in the fire.
l – forest layers.

Table 5: *Values of pollutant emission coefficients during forest fires [8, 12]*

Name of the pollutant	Emission coefficient k_i
Carbon monoxide	0,113
Carbon dioxide	1,609
Nitrogen oxides	0,000405
soot	0,0062
Smoke (ultradispersed SiO_2 particles)	0,0345
methane	0,075
Unsaturated hydrocarbons	0,011
ozone	0,001

The calculation of the concentration of the i-th combustion product in a forest fire at ground level at a given distance x from the emission source in the wind direction is carried out on the basis of the Model of the distribution of concentrations of polluting particles in the atmosphere at a constant wind speed, described by the assumption of a double distribution in the Gauss equation:

$$(x \quad 0) \quad \text{———} \quad \left\{exp \quad \left[\text{—}\right]\right\} \text{(kg/m}^3\text{)} \tag{24}$$

where:

U_a – average wind speed in the mixing layer (formula 29);

σ_y and σ_z – dispersion coefficients characterizing the scattering capacity of the atmosphere depending on the stability of the atmosphere.

For expressions of σ_y and σ_z, the Smith-Hosker formulas are currently used [13]:

$$(x) \quad \frac{}{\sqrt{1}} \tag{25}$$

$$(x) \quad F(z \quad x) \cdot g(x) \tag{26}$$

$$F(z \quad x) \quad \begin{cases} \left[c \qquad \left(1 \quad (c \qquad \right) \right)\right], \\ \ln\left[c \qquad \left(1 \quad (c \qquad \right)\right) \right] x \end{cases} \tag{27}$$

$$g(x) \quad \text{———} \tag{28}$$

where:

x – distance from the emission source, m;

z_0 – roughness height for different types of surface microrelief (Table 6).

Table 6: *Roughness height z_0 for different types of surface microrelief*

microrelief	z_0 , cm
Snow, lawn height 1 cm	0,1
Mown and low grass up to 15 cm	0,6-2
Tall grass up to 60 cm	4-9
Uneven surface with alternating patches of grass and shrubs	10-20
Park, forest up to 10m high	20-1000
City buildings	100

Table 7: *Values of parameters used in formulas for calculating σ_y and σ_z*

Atmosphere stability class	c_3	a_1	a_2	b_1	b_2
A	0,22	0,112	$5{,}38 \cdot 10^{-4}$	1,06	0,815
B	0,16	0,130	$6{,}52 \cdot 10^{-4}$	0,95	0,75
C	0,11	0,112	$9{,}05 \cdot 10^{-4}$	0,92	0,718
D	0,08	0,098	$1{,}35 \cdot 10^{-3}$	0,89	0,688
E	0,06	0,061	$1{,}96 \cdot 10^{-3}$	0,89	0,684
F	0,04	0,064	$1{,}36 \cdot 10^{-3}$	0,78	0,672

Table 8: *Values of parameters used in formulas for calculating σ_y and σ_z, depending on surface roughness*

Surface roughness, z_0, м	c_1	d_1	c_2	d_2
0,01	1,56	0,048	6,25·	0,45
0,04	2,02	0,027	7,76·	0,37
0,1	2,72	0	0·	0
0,4	5,16	-0,1	18,6·	-0,23
1,0	7,37	-0,096	4,29·	-0,60
4,0	1,7	-0,13	4,59·	-0,78

The average wind speed across the mixing layer is calculated using the formula [13]:

$$\left(—\right) \quad \text{(m/s)} \tag{29}$$

where:

z_1 – 10 m;

m – parameter depending on the class of atmospheric stability (table 12).

Table 9: *Values of parameter m for different classes of atmospheric stability*

Sustainability category	A	B	C	D	E	F
m	0,1	0,11	0,13	0,14	0,33	0,75

Social damage from forest fires is expressed in an increase in the overall mortality of the population, associated with intense and long-term smoke in urbanized areas.

Social damage from a forest fire is calculated using the formula:

$$\left\{\left(—\quad 005\right)+\left(—\quad 007\right)\right\}\ \frac{}{365\ (day)} \quad \text{(rub)} \tag{30}$$

where:

t_{inh} – inhalation time equal to forest burning time, days;

$C_{PM2.5}$ – PM2.5 concentration, μg/m³ (formula 31);

C_{PM10} – PM10 concentration, μg/m³ (formula 32);

C_{max10} – average daily maximum permissible concentration of suspended particles PM10, equal to 60 μg/m³ [14];

$C_{max2.5}$ – average daily maximum permissible concentration of suspended particles PM2.5, equal to 35 μg/m³ [14];

g_{PM10} – the amount of increase in PM10 concentration in the atmosphere of urbanized areas, equal to 10 μg/m³, at which the average daily mortality rate of the population increases by 0,5%.

$g_{PM2.5}$ – the increase in PM2.5 concentration in the atmosphere of urbanized areas equal to 5 μg/m³, at which the average daily mortality rate of the population increases by 0,7%.

$k_{a.m.}$ – the average annual mortality rate for the subject;

P is the population size in the urbanized area under consideration;

V is the cost estimate of the average statistical life, taken to be equal to 1 million rubles/person [15].

Based on the data on the share of fine fractions in total suspended particles (approximately 55% of PM10 and 65% of PM10 are PM2.5 particles) [16], the concentration of PM2.5 and PM10 is calculated:

$$\left(\mathrm{C}\quad\right)\qquad \text{(mcg/m}^3\text{)} \tag{31}$$

(C) (mcg/m³) (32)

where:
C_C – concentration of soot particles (formula 24);
C_{SiO2} – concentration of ultrafine SiO_2 particles (formula 24).

III Testing

Calculations were made for Moscow for the period from August 1 to 12, 2010 (at the time of prolonged smoke pollution in the capital from large-scale forest and peat fires) using data on the areas of burned forests and peat bogs in the Moscow region in August 2010. The results obtained fully correlate with the data from monitoring the concentrations of suspended particles in the atmosphere and the Civil Registry Office data on mortality in August 2010 in Moscow. According to calculations, the average daily concentration of PM10 from August 1 to 12, 2010 was 550 μm/m³ per day. According to "Air quality monitoring in Moscow, 2011", the average daily concentration of PM10 in the same period was 582 μg/m3. According to the Civil Registry Office of Moscow, the mortality rate in August was 15 016 people/month with an average mortality rate in the same period in other years of about 9 175 people/month (data from 2007 to 2019). The daily mortality rate in August 2010 in Moscow was about 484 people/day. In the calculations, the daily mortality rate was 493 people/day, a slight deviation from the registry office data can be explained by the previous jump in mortality in July, as well as a decrease in the concentration of fine particles in the air of Moscow in the second half of August due to a change in the weather as a result of the departure of the anticyclone.

IV Conclusions

Using this method allows us to take into account the steepness of the slope in the mountain-belt forests of Russia, which accordingly affects the speed of spreading the edge of the forest belt, as well as the thermal and physical characteristics of forest areas typical for the mountainous terrain of the Russian Federation.

References

[1] J. Sacks. Health Effects of Wildfires. Public Information Officers Incident Management Response Roundtable. March 16, 2022.

[2] Strukova E. B., Balbus J., Golub A. A. Health risk and economic assessment of damage from air pollution in Russia. / In the book: Climate. Air quality and health of Muscovites. / Ed. B. A. Revich. Moscow: Adamant Publishing House, 2006.

[3] Onishchenko G. G. et al. "Fundamentals of public health risk assessment when exposed to chemicals polluting the environment. Moscow: Research Institute of Environmental Protection and State Environmental Protection. 2002

[4] Cohen A. J., Anderson H. R., Ostro B. et al., Mortality impacts of urban air pollution. In: Comparative Quantification of Health Risks: Global and Regional Burden of Disease Due to Selected Major Risk Factors (Ezzati M, Lopez AD, Rodgers A, Murray CJL, eds.), vol. 2. World Health Organization, Geneva, Switzerland, 2004.

[5] Volokitina AV, Sofronov MA Classification and mapping of plant combustible materials. - Novosibirsk: Publishing House of the Siberian Branch of the Russian Academy of Sciences, 2002. 314 p.

[6] Grishin AM Physics of forest fires. - Tomsk: Publishing House of Tomsk University, 1999. 207 p.

[7] E.V. Konev Analysis of the process of spreading forest fires and burnings//Thermophysics of forest fires. - Novosibirsk: Institute of Thermal Physics of the Siberian Branch of the USSR Academy of Sciences, 1984. - pp. 99-125.
[8] Methodology for Determining and Calculating Pollutant Emissions from Forest Fires. Grishin A.M., Dolgov A.A., Tsymbalyuk A.F. Put into effect by Order of the State Committee on Ecology of Russia No. 90 dated March 5, 1997. P. 15.
[9] Albini J.A. A physical model for fire spread in brush//11th Symp. (Int.) Combustion, Berkley, Calif. – Pittsburgh, 1967. – P. 553 – 560.
[10] Methodical recommendations for calculating zones of gas contamination by combustion products of large fires. FGBU VNIIPO EMERCOM of Russia. 2007.P. 58
[11] GOST R 12.3.047-2012 Occupational safety standards system. Fire safety of technological processes. General requirements. Control methods.
[12] Hegg D.A., Radke L.F., Hobbs P.V. Emissions of some trace gases from biomass fires - Journ. of Geophys Res., V. 95, N D5, 1990, pp. 5669-5675.
[13] Gusev N.G., Belyaev V.A. Radioactive emissions in the biosphere. Handbook. Moscow: Energoatomizdat, 1991, 256 p
[14] RD 52.04.830-2022 Mass concentration of suspended particles of PM10 and PM2.5 fractions in atmospheric air samples.
[15] Resolution of the Government of the Russian Federation of 28.12.2019 No. 1928.

ANALYSIS OF THE EFFECT OF TEMPERATURE ON SOLAR PANELS AND THEIR COOLING METHODS

I.M. Marufov[1], S.Y. Shikhaliyeva[2]

•

Azerbaijan State Oil and Industry University, Baku, Azerbaijan
[1] ilkinmarifov@mail.ru; [2] saadaasmar@mail.ru

Abstract

One of the most widely used renewable energy sources is solar energy, and it is predicted to continue to be so in the future. Recently, a great increase has been observed both in the study of the working principle of photovoltaic (electricity generated by the effect of light) devices and in increasing their efficiency. Solar cells change as a result of temperature fluctuations. The purpose of the article is the effect of temperature on the efficiency of solar panels and their cooling methods. The novelty. Solar cells change as a result of temperature fluctuations. Methods. Taking into account that the cooling system is implemented by spraying water, we can determine when the cooling starts at the moment when the temperature reaches the maximum by building a mathematical model. Results. In this article, the relationships between solar radiation, efficiency and temperature are determined under different conditions. Practical value. Based on the heating and cooling models, it was determined that starting the cooling process when the temperature of the panels reaches 45 0C is the most convenient method.

Keywords: temperature, photovoltaic elements, efficiency, effect of temperature, physical properties.

I. Introduction

Recent high energy consumption and fuel depletion have increased interest in renewable (solar) energy. Photovoltaic power is a technology that has seen tremendous growth in its use over the past 10 years. One of the most important problems faced by photovoltaic panels is heating caused by solar radiation and excessive temperature. High temperature heating reduces the efficiency of the panels. Depending on the temperature change between 0-75 celsius, the ideal characteristics of the solar panel also change. The characteristic is true as long as solar irradiance E and module temperature Tm are kept constant. If one of these two parameters changes, the entire characteristic changes. As the temperature of the solar panel increases, the maximum power output decreases. The temperature coefficient of the solar panels used in this study is 0.5%/^{0}C. Based on this coefficient, it can be seen that when the temperature rises by 1 Celsius, the efficiency decreases by 0.5%.

A solar cell is a semiconductor structure with p-n conductivity. Direct current is generated when the sunlight hits it. One of the most important problems faced by solar panels is overheating caused by solar radiation and extreme temperatures. It is known that high temperature heating reduces the efficiency of the panels. The ideal p-v characteristic of the solar panel according to the temperature change between 0-75^{0}C is shown in figure 1. In this characteristic, the relationship between the power produced by the solar panel and the voltage is given. The characteristic is paid as long as the solar irradiance E and the module temperature Tm are kept constant. If one of these two parameters changes, the entire characteristic changes [1-3].

The electrical scheme of photovoltaic elements is shown in the figure (Figure 1).

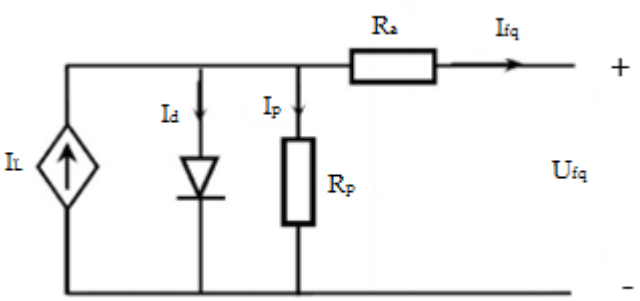


Figure 1: *Electrical scheme of photovoltaic cells*

Photovoltaic (FQ) curves differ with solar insolation and temperature modulus. Let's look at the formula below.

(1)

The change in temperature affects the solar cells and also the power. The voltage generated in FQ changes depending on the temperature [5]. As the temperature increases, the voltage decreases (Figure 2). Figure 2 shows the I-V characteristic of FQ at constant radiation. As the temperature decreases, the current of FQ decreases and its voltage increases [4-6].

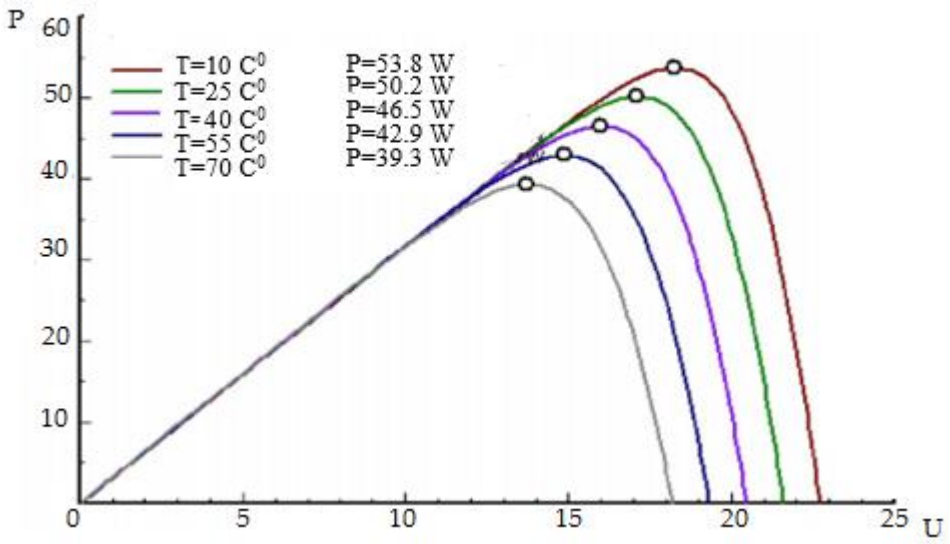


Figure 2: *I-V characteristic of FQ in constant radiation*

II. Materials and methods

The efficiency of solar cells is given by the following formula:

(2)

Here I_{max} and U_{max} are the maximum voltage and current respectively. Based on the air change function, let's take the temperature of the FQ module as T_P, and the temperature of the environment as T_x. Taking into account the effect of temperature on solar modules, we can write its useful efficiency as follows.

(3)

Here, η_x is the useful work coefficient at the adopted temperature. The parameters η_x and β_x are normally supplied by photovoltaic cells. The temperature coefficient depends not only on the material of the FQ module, but also on T_x and is given as follows.

(4)

Here, T_0 is the high temperature that reduces the useful work coefficient to 0. For crystalline silicon solar cells, this temperature is up to 270^0 C. Convection is the main mechanism for spreading heat on the ground. The useful performance factor mainly depends on the installation of the module, the wind speed, the humidity of the environment and the characteristics of the module. The temperature of the module is measured under open circuit. When the temperature is 200 C, the brightness is 0.8 W/m2 and the wind speed is 1 m/s.

Figure 3 shows the dependence of efficiency on temperature in case of solar radiation of 1000 W/m2. As can be seen, there is a linear relationship between temperature and efficiency. As the temperature increases, the efficiency decreases. As a result, we can get the efficiency we want by changing the temperature.

The degree of cooling of solar panels is one of the important factors that have a high impact on their quality. Therefore, we can find the cooling period by determining the cooling rate of the panel. The cooling period is determined based on the energy balance. Since solar cells are mainly made of glass, the physical properties of the glass should be taken as the physical properties of the panels. The temperature of the panel was respectively 45 ^{0}C before cooling and 35 ^{0}C after cooling. It is assumed that the maximum permissible temperature is 45 ^{0}C [7-11].

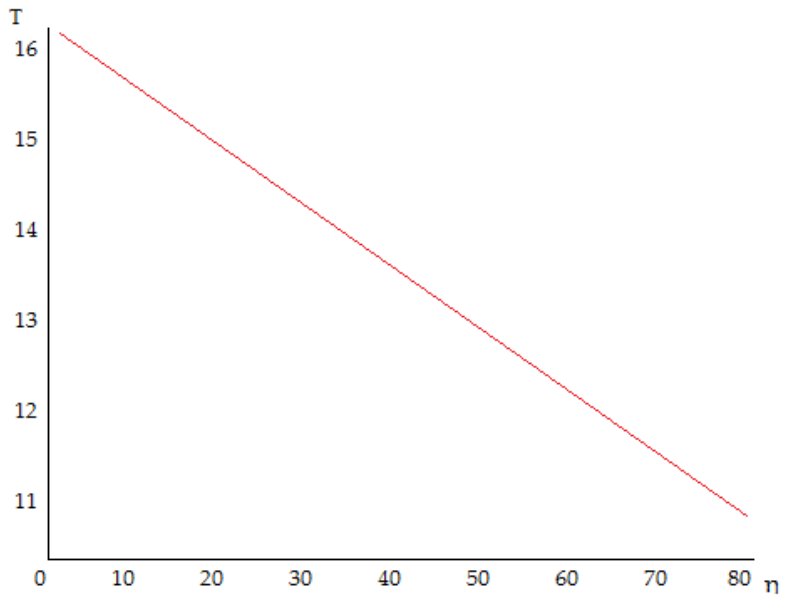


Figure 3: *Temperature dependence of efficiency in case of solar radiation of 1000 W/m2*

Certain results have been obtained from the conducted experiments. (In solar cells, a 10 ^{0}C increase in temperature (from 35 ^{0}C to 45 ^{0}C) causes the efficiency to decrease from 12% to 10.5%. From the above, it can be said that the rise in temperature is a negative thing. That's why, every 5 minutes, the cooling system lowers the temperature by 10 ^{0}C and increases efficiency. The obtained results are shown in Figure 2, and it was observed that the difference between the theoretically obtained result and the result in real life does not exceed 5%. As a result, it can be concluded that formula 1 can be used to determine the module temperature as a function of ambient conditions. Here it is also possible to determine when the panel should be cooled down when the temperature

of the panel reaches its maximum. A linear fit of module temperature is plotted to determine its heating rate [12-15].

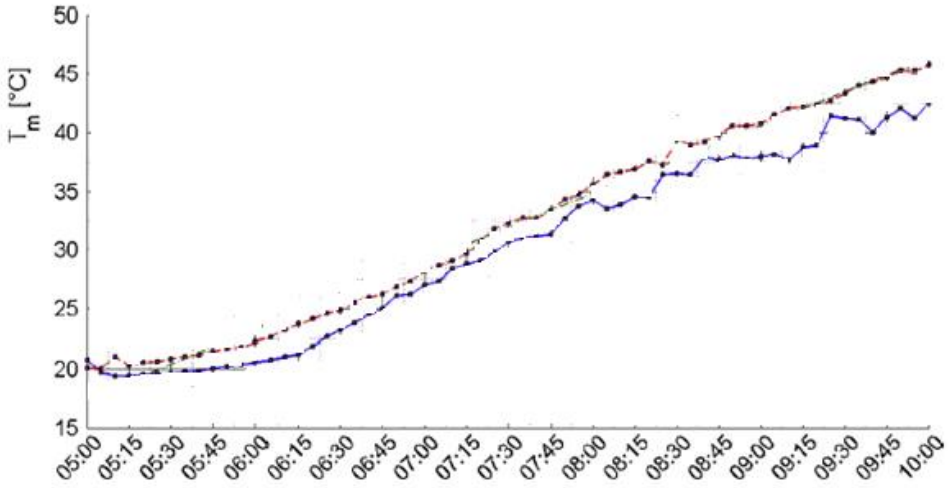


Figure 4: *Time dependence of temperature*

The proposed cooling system can solve the problem of overheating due to solar radiation and maintain efficiency with minimum water consumption. In addition, this method is suitable for use in very hot and sandy areas. The selection of the maximum permissible temperature should be based on the conditions of minimum energy and water consumption and maintaining the efficiency at the accepted level. If we turn on the cooling system for every 1 Celsius increase in temperature, in this case, the energy received from the panel will be used to cool the same panel. Buda means a decrease in efficiency. One of the main goals of this study is to determine when to start the cooling process. Analyzes performed are paid for any panels. The panel is first allowed to heat up to a given limit and then cooled down to normal operating temperature. It can be concluded that the maximum permissible temperature for the panel is 45 ^{0}C, and in this case it is possible to produce the highest energy. As the maximum allowable temperature increases, evaporation of water during cooling will increase and therefore water usage will increase. Therefore, choosing a temperature of 45 ^{0}C is considered the most suitable option [16-20].

III. Conclusions

1. In this article, the temperature effect and photovoltaic sources were investigated. The relationship between solar radiation and temperature was determined using simple formulas. The influence of the temperature effect on the efficiency of photovoltaic sources in cloudy conditions was shown.

2. It was determined that the increase in temperature leads to a decrease in the efficiency of photovoltaic elements. It can be concluded that by reducing the temperature, we can increase the efficiency of photovoltaic elements. At the same time, using comparative analyzes and mathematical models, relationships between temperature, efficiency, and cooling rate were determined.

3. From the obtained results, it can be said that if we take into account the selection of the maximum permissible temperature, the minimum energy and water consumption, and the maintenance of the efficiency at the accepted level, the maximum permissible temperature for the panels is 45 ^{0}C, and in this case it is possible to produce the highest energy.

References

[1] N.S. Mammadov, N.A. Ganiyeva, G.A. Aliyeva, "Role of Renewable Energy Sources in the World". Journal of Renewable Energy, Electrical, and Computer Engineering, Vol. 2, №2, pp. 63-67, Indonesia, 30 September 2022

[2] Haddad, A.; Ramadan, M.; Khaled, M.; Ramadan, H.S.M.; Becherif, M. Triple Hybrid System Coupling Fuel Cell with Wind Turbine and Thermal Solar System. Int. J. Hydrogen Energy, 45, 11484–11491, 2020

[3] Jiao, C.; Li, Z. An updated review of solar cooling systems driven by Photovoltaic–Thermal collectors. Energies, 16, 5331, 2023

[4] Ibrahim, K.A.; Luk, P.; Luo, Z. Cooling of Concentrated Photovoltaic Cells—A review and the perspective of Pulsating flow Cooling. Energies, 16, 2842, 2023

[5] Mammadov N., Mukhtarova K. "Analysis of the smart grid system for renewable energy sources", Journal «Universum: technical sciences", № 2 (107), pp. 64-67, 2023

[6] Herrando, M.; Wang, K.; Huang, G.; Otanicar, T.; Mousa, O.B.; Agathokleous, R.A.; Ding, Y.; Kalogirou, S.A.; Ekins-Daukes, N.J.; Taylor, R.A.; et al. A review of solar hybrid photovoltaic-thermal (PV-T) collectors and systems. Prog. Energy Combust. Sci., 97, 101072, 2023

[7] Sattar, M.; Rehman, A.; Ahmad, N.; Mohammad, A.; Alahmadi, A.; Ullah, N. Performance Analysis and Optimization of a Cooling System for Hybrid Solar Panels Based on Climatic Conditions of Islamabad, Pakistan. Energies, 15, 6278, 2022

[8] Abdelgaied, M.; Kabeel, A.E.; Zeleňáková, M.; Abd-Elhamid, H.F. Floating Photovoltaic Plants as an Effective Option to Reduce Water Evaporation in Water-Stressed Regions and Produce Electricity: A Case Study of Lake Nasser, Egypt. Water, 15, 635, 2023

[9] Kandeal, A.; Algazzar, A.M.; Elkadeem, M.R.; Thakur, A.K.; Abdelaziz, G.B.; El-Said, E.M.; Elsaid, A.M.; An, M.; Kandel, R.; Fawzy, H.E.; et al. Nano-enhanced cooling techniques for photovoltaic panels: A systematic review and prospect recommendations. Sol. Energy, 227, 259–272,2021

[10] Pomares-Hernández, C.; Zuluaga-García, E.A.; Escorcia Salas, G.E.; Robles-Algarín, C.; Sierra Ortega, J. Computational Modeling of Passive and Active Cooling Methods to Improve PV Panels Efficiency. Appl. Sci., 11, 11370, 2021

[11] Mojumder, J.C.; Chong, W.T.; Show, P.L.; Leong, K.W.; Abdullah-Al-Mamoon. An experimental investigation on performance analysis of air type photovoltaic thermal collector system integrated with cooling fins design. Energy Build., 130, 272–285, 2016

[12] Mankani, K.L.; Chaudhry, H.N.; Calautit, J.K. Optimization of an air-cooled heat sink for cooling of a solar photovoltaic panel: A computational study. Energy Build., 270, 112274, 2022

[13] Li, D.; King, M.; Dooner, M.S.; Guo, S.; Wang, J. Study on the cleaning and cooling of solar photovoltaic panels using compressed airflow. Sol. Energy, 221, 433–444, 2021

[14] Nižetić, S.; Jurčević, M.; Čoko, D.; Arıcı, M. A novel and effective passive cooling strategy for photovoltaic panel. Renew. Sustain. Energy Rev., 145, 111164, 2021

[15] Bayrak, F.; Oztop, H.F.; Selimefendigil, F. Experimental study for the application of different cooling techniques in photovoltaic (PV) panels. Energy Convers. Manag., 212, 112789, 2020

[16] I.N. Rahimli, S.V. Rzayeva, E.E. Umudov, "DIRECTION OF ALTERNATIVE ENERGY", Vestnik nauki, Issue 2, Vol. 61, №4, April 2023

[17] Zilli, B.M.; Lenz, A.M.; de Souza, S.N.M.; Secco, D.; Nogueira, C.E.C.; Junior, O.H.A.; Nadaleti, W.C.; Siqueira, J.A.C.; Gurgacz, F. Performance and effect of water-cooling on a microgeneration system of photovoltaic solar energy in Paraná, Brazil. J. Clean. Prod., 192, 477–485, 2018

[18] Tina, G.M.; Scavo, F.B.; Aneli, S.; Gagliano, A. Assessment of the electrical and thermal performances of building integrated bifacial photovoltaic modules. J. Clean. Prod., 313, 127906, 2021

[19] Bei, Y.; Yuan, B.; Wang, Z.; Tang, J. Floating Photovoltaic for the Coal Mining Subsidence Water Area—An Effective Way to Reduce Evaporation. In Proceedings of the 2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE), Chongqing, China, pp. 1126–1130, 22–24 April 2022

[20] Xu, Q.; Ji, Y.; Riggs, B.C.; Ollanik, A.; Farrar-Foley, N.; Ermer, J.; Romanin, V.; Lynn, P.; Codd, D.S.; Escarra, M.D. A transmissive, spectrum-splitting concentrating photovoltaic module for hybrid photovoltaic-solar thermal energy conversion. Sol. Energy, 137, 585–593, 2016

MARKOV CHAIN MODEL FOR COMPARISON OF PRICE MOVEMENT OF FRUITS IN SALEM DISTRICT, TAMILNADU

Kamalanathan R[*1], Sheik Abdullah A[2] and Kavithanjali S[1]

•

[*1& 1]Ph.D Research Scholars, PG and Research Department of Statistics,
Salem Sowdeswari College, Salem- 10, India
[2]Assistant Professor, PG and Research Department of Statistics,
Salem Sowdeswari College, Salem-10, India
rknstat1@gmail.com, sheik.stat@gmail.com, kavithanjalis2018@gmail.com

Abstract

Statistical forecasting requires mathematical models and techniques to predict future outcomes based on historical data. Markov chains are statistical models that can be utilized to analyze the movement of prices in agriculture price, financial market price, business process, fuel prices and etc., They are particularly relevant in the context of price movements because they provide a framework for understanding and predicting the future state of a system based on its current state. In a Markov chain process, there are a set of states and we progress from one state to another based on a fixed probability. In these decades many articles are showed that modeling a market as a random walk was applicable and that a market may be viewed as having the Markov property. The objective of this paper is to construct the Markov chain model for daily fruit price movement in Salem District, Tamil Nadu. Two models are highlighted, where the price movement is considered as being in a state of gain, loss and no change and large gain, or small gain or loss, or large loss and no change. Ten different types of fruits are considered which are cultivated Salem areas and above two models are used to analyze the price movement of each fruit. These models were used to obtain transitional probabilities, steady state probabilities and mean recurrence times. Our results indicate that the pattern of price movement of Banana is similar to price movements of other fruits, in both models. The investor is encouraged to invest in the fruit market at any time in away which leads to a greater chance of getting more gain than loss.

Keywords: Markov chain, Transition probability, Steady state distribution, Ergodic chain, Mean recurrence time, Price movement.

I. Introduction

As fruit is one of the most important consumer goods for residents, the fluctuation of the prices has a direct impact on people's daily life. Especially in recent years, the continuous increase in fruit prices has drawn much attention of the government as well as the people, Salem, District has all a long been one of the Districts in the state with a credit able performance in agricultural production with the farmers relatively more responsive and receptive to changing technologies and market forces. Fruits are produced on aye around basis and a large number of farmers are involved in the

production process. Fruits are being cultivated throughout the year with the help of lift irrigation from the dug wells. Uzhavarsandai marketing system has played an important role in deciding the fruit prices in Salem markets. It ensures that the farmers will get a better price for their produce and to enable the consumers to get fresh fruits at a less or price than the retail market price.

Fruits are seasonal crops, and their supply in town markets is influenced by natural factors. Due to their perishable nature, they must be quickly dispatched to markets. Fruit is the pulpy or dry ripened ovary part of a flowering plant, enclosed by the seed or seeds. Bananas, Custard Apple, Gooseberry, Guava, Mango, Mosambi, Grapes, Sapota, Papaya and Watermelon are the examples of fruits. The cultivation of fruits and its treatment are done by fruit farming. Many animals are attracted by fruits that are pulpy and contain sugar amounts and these animals then disperse the seeds of fruits to new locations. Non-fleshy fruits use different mechanisms for seed dispersal. There are some plants in which fruits can develop without fertilization. This process is called parthenocarpy, and those fruits are seedless. The wall thickens and becomes differentiated into three, more or less distinct, layers during the development of the ovary.

1.1 Importance of Fruits

An apple a day keeps the doctor away! You must have heard this classic saying and understood the importance of fruits in keeping ourselves healthy and keeping the doctors away. Fruits are wholesome food. They are rich in vitamins and other nutrients. It is almost impossible to even think of any doctor who does not recommend fruits for good health and diet. Fruits do not just have a taste, but they also have a health benefits, such as:

- Lower blood pressure
- Reduce the risk of heart disease and stroke
- Prevent some types of cancer
- Lower risk of eye and digestive problem

II. Basics of Markov chain

A random movement (or walk) is said to exhibit the Markov property if the position of the movement at time (n+1) depends only upon the position of the movement at time n. Let Y_n denote the position of the random movement at time n, then equation (1):

$$P\,(Y_{n+1} = j \mid Y_n = i) = p_{ij} \tag{1}$$

is independent of $Y_{n-1}, Y_{n-1}, \ldots, Y_0$ so that the state of Y at time (n+1) depends only upon the state Y of at time n . Here each $p_{ij}$for j = 1, 2,…, N is a probability row vector describing every possible transition from state i to any other existing in N possible states in the process. Then equation (2):

$$\sum_{j=1}^{N} p_{ij} = 1 \tag{2}$$

Generally, a random movement exists in N possible states in the system. Then, in chains, $P\,(Y_{n+1} = j)$ will depend on the whole sequence of random variables starting with the initial value Y_0 (Jones and Smith, 2001); it leads to equation (3):

$$P\,(Y_{n+1}=j \mid Y_n= i\,, Y_{n-1}=i_{n-1},\ldots,Y_0=i_0) = P(Y_{n+1}=j \mid Y_n=i\,) \tag{3}$$

Intuitively, one interprets equation (3) to means that, given the “present” of the process, the

"future" is independent of its "past" found by Parzen E [8]. The random process of moving from one state of the system to another with the associated probabilities of each transition is known as the chain. It is said that every step taken in a chain possessing the Markov property depends only on immediately preceding step.

This expresses the fact that, if the system is in one of the states at one observed value, it will certainly be in one of the states at the next observed value. With these transition probabilities, a N×N matrix, P = (p_{ij}), called the first step transition probability matrix of the Markov chain.

$$P = (p_{ij}) = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}$$

Each row of is the probability distribution relating to a transition from state i to state j.

The probability vectors $p^{(n)}$ for n = 0, 1, 2 … are said to be the state vectors of a Markov chain, where $p_i^{(n)} = P(Y_n = i)$is the probability that the system is in the i^{th}state at the n^{th}step. In particular, the state vector $p^{(0)}$ is called the initial probability or initial state vector of the Markov chain. If P is the transition matrix and $p^{(n)}$ is the state vector at the n^{th}step, one can write,

$$p^{(n+1)} = p^{(n)}P$$

where $p^{(n+1)}$ is the state vector at the$(n + 1)^{th}$ step . From this it follows that

$$p^{(n)} = p^{(0)}P^n \tag{4}$$

i.e., the initial state vector $p^{(0)}$ and the transition matrix P determine the state vector $p^{(n)}$ at the $n^{(th)}$time point. The $n^{(th)}$step transition probabilities are called conditional Probabilities and are denoted by $p_{ij}^{(n)} = P(Y_{(n+k)} = j|Y_k = i)$, where i , j € {1,2,…,N} with $0 \leq P_{ij}^{(n)} \leq 1$ for n = 0,1,2,…and

$$\sum_{j=1}^{N} p_{ij}^{(n)} = 1$$

$p^{(n)} = p^{(0)}P^n$ in matrix notation for the finite state homogeneous Markov chain can be written as;

$$\left(p_1^{(n)}, \ldots, p_N^{(n)}\right) = \left(p_1^{(0)}, \ldots, p_N^{(0)}\right) \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}^n$$

Therefore, the future state vector $P^{(n)}$can be evaluated if the initial state vector and the transition matrix are known. Classifying the states of a finite Markov chain according to certain basic properties, such a classification can be based on the definitions as follows: State j is said to be accessible from state i, if j can be reached from i in a finite number of steps. If two states i and j are accessible to each other, then they are said to communicate. The period of a state i is defined as the greatest common divisor of all integers n≥1, for which $P_{ij}^{(n)}$>0. When the period is one, the state is referred to as a periodic. If all states of a chain is communicate and are a periodic, then the chain is said to be ergodic said to Bhat. U. N [2].

A chain is to have a stationary (or steady state) distribution, if there exists a vector π such that given a transition probability matrix P: $\pi = \pi P$
If a finite Markov chain is Ergodic then

$$\lim_{n\to\infty} p^n = \pi = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \dots & \dots & \dots & \dots \\ \pi_1 & \pi_2 & \dots & \pi_N \end{bmatrix}$$

Where $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ with $0 < \pi_j < 1$ and $\sum_{j=1}^{N} \pi_j = 1$.

This stationary probability vector π can be viewed as the unique distribution of a random variable in the long-run. Consequently expected (or mean) recurrent times μ_j are given by

$$\mu_j = \frac{1}{\pi_j} \tag{5}$$

2.1 Review of Literature

Xiaoxia Zhu and Xiuquan Xu [14] have established the prediction Markov chain model for fruit price fluctuation in China. Many scholars have conducted deep research in the rise of fruit price. Ravi shankar et. al., [11] had a study on effect of price of other seasonal fruits on Mango price in Uttar Pradesh, India. The study examines shifting consumer demand and production patterns towards fruits and vegetables, noting that horticultural commodity production now exceeds food grain production. It specifically analyzes how the prices of other seasonal fruits impact mango prices in major markets of Uttar Pradesh.

Eunice Wacu Mutahi et.al., [4] study was intended to investigate determinants of price fluctuations on fruits and vegetables. Aparna Bairagi and Sarat Kakaty [1] made Markov chain modeling for prediction on future market prices of potatoes with special reference to Nagaon district ,Assam, India. They predicted the arrival market price movement of potatoes in near future. Also found the long term price behavior by using transition probability matrix. Prasenjit et.al., [9] made a case study in price and arrivals of vegetable in different market of south west Bengal, India. They Analysed seasonal and other time series data on whole sale prices and marketing efficie ncy. Tamilselvi et. al., [13] made a study on behaviour of market arrival and price of vegetables in Koyambedu market in Chennai. Also price patterns and correlation between monthly arrivals and price were observed. Khem Chand et.al., [7] made a case study on Fruit marketing. Its efficiency and supply chain constraints in India. This research highlights significant challenges in the marketing of kinnow and aonla and offers practical recommendations to enhance the efficiency and profitability of their supply chains.

Hongyu Yang et.al., [5] had a study on the dynamic impacts of weather changes on vegetable price fluctuations in Shandong Province, China. In this article, the Granger causality test was utilized to demonstrate changes in weather factors and vegetable price fluctuations. Jhade Sunil et.al., [6] had a study on export performance and comparative analysis of Onion with reference to India. In this paper they shown that, inconsistent export, policies, rising domestic onion prices and export bans in India necessitate evaluating export competitiveness and the direction of onion exports. They also noted that the most consistent importer of Indian Onion based on first order Markov chain method. Sachin kumar Verma et.al., [12] had a study on Food Grain Trade Prospect of India .In this study underscores the importance of area expansion over productivity growth in enhancing food grain production in India. Especially the correlation analysis shows a positive relationship between area and output. Also the linear programming analysis identifies stable and unstable export markets.

Divya et.al. [3] done a research on comprehensive analysis of production and import dependency of pulses in India. In this article provides valuable insights for policy makers and industry stakeholders serve as guiding principle in making strategic decisions bolster India's role in the global agricultural sector for contributing economic growth and food security. Priyanka et.al., [10] made a agribusiness performance appraisal on export competitiveness and direction of trade of Banana from India. In this paper, the Markov chain analysis was applied to the export data to the export data and the transitional probability matrix provided a great deal of information on where to sell Indian Banana to get the highest benefits.

III. Results Data Source and Its Properties

The daily price data of Fruit at Salem Uzhavar Santhai 808 days (October 2021 –December 2023) have been collected from the websitehttps://vegetablemarketprice.com/market/salem.

Table 1: *Basic Statistics*

Fruits	*Basic Statistics*						
	Max	*Min*	*Mean*	*Median*	*Mode*	*Variance*	*SD*
Banana	*48*	*22*	*38.62*	*40*	*44*	*37.16*	*6.11*
Custardapple	*65*	*40*	*57.70*	*59*	*60*	*31.54*	*5.62*
Gooseberry	*135*	*67*	*111.61*	*115*	*105*	*222.29*	*14.89*
Guava	*56*	*35*	*45.93*	*45*	*44*	*20.65*	*4.55*
Mango	*160*	*47*	*103.54*	*100*	*145*	*1211.36*	*34.77*
Mosambi	*80*	*41*	*62.62*	*61*	*62*	*79.01*	*8.88*
Grapes	*118*	*26*	*37.00*	*31*	*29*	*489.65*	*22.11*
Sapota	*50*	*21*	*41.19*	*44*	*46*	*47.48*	*6.89*
Papaya	*49*	*18*	*26.24*	*25*	*21*	*37.41*	*6.11*
Watermelon	*28*	*11*	*19.69*	*19*	*22*	*15.10*	*3.88*

IV. Methodology

4.1 Markov Chain

An elementary form of dependence between values of X_n in successive transitions was introduced by Russian mathematician A. A.Markov and it is known as Markov dependence. Markov dependence is a form of dependence which states that X_{n+1} depends only on $X_n$when it is known and is independent of $X_{n-1}, X_{n-2}, \dots, X_0$. This implies that the future of the process depends only on the present, irrespective of the past. This property is known as Markov property. In probabilities terms, the Markov property can be stated as follows:

$$P\left[X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\right] = P\left[X_{n+1} = i_{n+1} | X_n = i_n\right] \quad (6)$$

For all states $i_0, i_1, \dots, i_{n+1}$ and for all n, this is called Markov dependence of the first order.

4.2 Transition Probability Matrix

The state transition probability matrix of a Markov chain gives the probabilities of transitioning from one state to another in a single time unit. It will be useful to extend this concept to longer

time intervals.

The square matrix P consisting of the elements $p_{ij}^{(1)}$ for all possible states i and j is called one – step transition probability matrix of the chain. Therefore

$$P = [p_{ij}^{(1)}] = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

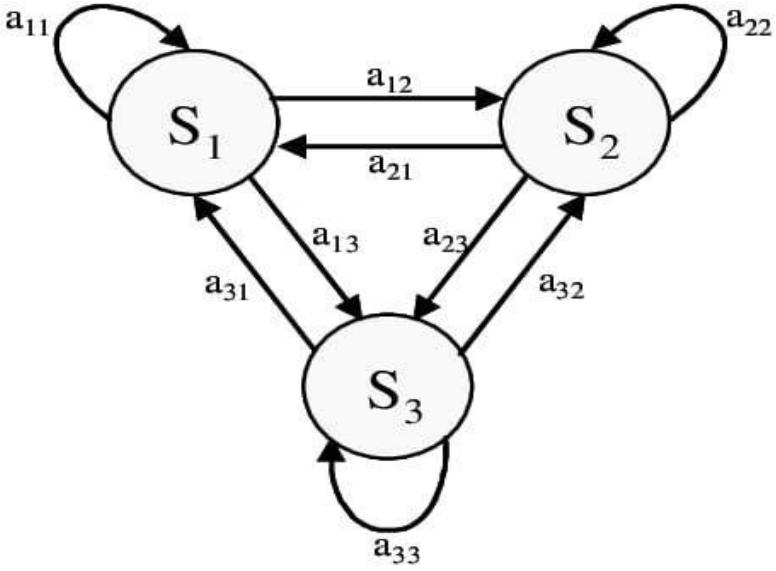


Figure 1: *Three State Transition Probability Matrix*

Similarly, the square matrix $P^{(m)}$ consisting of the elements $p_{ij}^{(m)}$ for all possible values of the states i and j is called them m – step TPM of the chain. Hence, $p^{(m)} = [p_{ij}^{(m)}]$

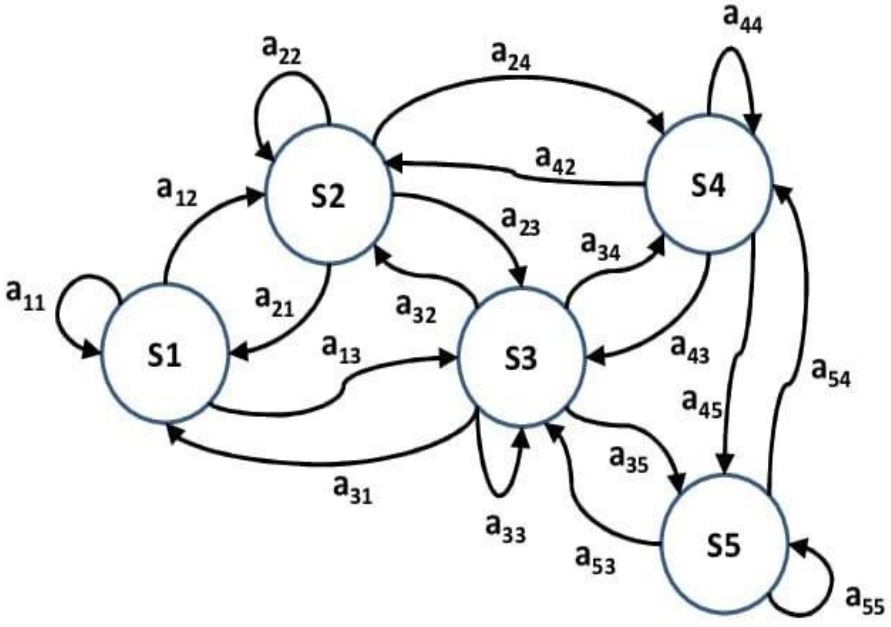


Figure2: *Five State Transition Probability Matrix*

4.3 Chapman – Kolmogorov Equation

We have, considered unit-step or one-step transition probabilities the probability of X_n given X_{n-1}. i.e. the probability of the outcome at the nth step or trial given the outcome at the previous step; P_{jk} gives the probability of unit - step transition from the state j at a trial to the state k at the next following trial. The m-step transition probability is denoted by $P_r\{X_{m+n}=k \mid X_n= j\}=P_{jk}^{(m)}$; $P_{jk}^{(m)}$gives the probability that from the state j at n^{th} trial, the state k is reached at $(m+n)^{th}$ trial in m steps, i.e. the probability of transition from the state j to the state k in exactly m steps. The number 'n' does not occur in the RHS of the relation and the chain is homogeneous. The one-step transition probabilities $P_{jk}^{(1)}$are denoted byP_{jk}. The state k can be reached from the state j in two steps through some intermediate state r. Consider a fixed value of r, we have

$$P\ (X_{n+2}= k.\ X_{n+1}= r \mid X_n=j) = P_{rk}^{(1)}P_{jr}^{(1)}=P_{jr}P_{rk} \tag{7}$$

Since these intermediate states r can assume values r=1, 2…, we have

$$P_{jk}^{(2)}= Pr\ \{X_{n+2}=k \mid X_n=j\}=\sum_r Pr\{X_{n+2}=k,\ X_{n+1}= r \mid X_n=j\} =\sum_r P_{jr}\ P_{rk} \tag{8}$$

(summing over for all the intermediate states)
Similarly, we get

$$P_{jk}^{(m+1)}= \sum_r P_{jr}P_{rk}^{(m)} \tag{9}$$

In general,

$$P_{jr}^{(m+n)}= \sum_r P_{rk}^{(n)}P_{jr}^{(m)}=\sum_r P_{jr}^{(n)}P_{rk}^{(m)} \tag{10}$$

This equation is a special case of Chapman - Kolmogorov equation, which is satisfied by the transition probabilities of a Markov chain, $P_{jr}^{(m+n)}\geq P_{jr}^{(m)}P_{rk}^{(n)}$.

Let P = P_{jk} denote the transition matrix of the unit - step transitions and $P^{(m)}= (P_{jr}^{(m)})$ denoted the transition matrix of the m - step transitions. For m=2, we have the matrix $P^{(2)}$whose elements are given. It follows that the elements of $P^{(2)}$are the elements of the matrix obtained by multiplying the matrix P by itself, i.e, $P^{(2)}$= P . P = P^2 . Similarly, $P^{(m+1)}$ = P. P^m and $P^{(m+n)}= P^m\ .\ P^n$

It should be noted that there exist non – Markovian chain whose transition probabilities satisfy Chapman – Kolmogorov equation.

4.4 Construction of Models

The daily market prices of 10 different types of fruits namely Banana , Custard Apple , Gooseberry, Guava, Mango, Mosambi, Papaya, Sapota, Grapes, Watermelon at https://vegetablemarketprice.com/fruits/tamilnadu/todaywere used in this modeling study. The period February 1st 2022 to November 10th 2023 was chosen in the estimation of the model, but the chain took into account the behavior of the market price for consecutive days, each classified as increase or decrease or remains the same. The data was set up into two types of models and studied separately. One of the main assumptions in Markov chain is stationarity. Let $X_n$denote the price of fruits at the market during the n^{th} day. Then the random variable Z_n is defined as:

$$Z_n = X_n - X_{n-1} \tag{11}$$

Here fruit price changes are taken as the first order of the chain since the nature of the perishable and particular fruit cannot permit selling after a day. Therefore, the first order change is only appropriate for study related to fruit markets. So that, for this study, two models of Markov chain analysis will be considered as follows:

4.5 Construction of Model-I

Each day was classified as having price value higher than or lower than or within "a "rupees from the previous day for this experiment considering the movement from a category of large gain, or small gain/loss, or large loss, let the classification of three states, namely:

- State 1(large gain): Today's price value is more than "a" rupees than the price value of the previous day;
- State 2 (small gain /loss): Today's price value change is within or equal to "a" rupees v is-a-v is the price value of the previous day;
- State 3(large loss): Today's price value is lower than "a" rupees than the price value of the previous day.

The state of this system may be able to for matrinary random variable denoted by

$$Y_n = \begin{cases} 0, & if\ Z_n > a, \\ 1, & if\ |Z_n| \le a, \\ 2, & if\ Z_n < -a. \end{cases}$$

The random variable $\{Y_n\}$ defined by above equation is known as a Markov chain with state space {0, 1, 2}. Here "a" is treated as threshold value of absolute price changes. There is no unique way of determining the threshold in this nature of studies. Generally, researchers used common parameters like measures of central tendency to find the threshold point. In this study, the main focus is only on the changes of the price pattern not on the price values. Here threshold value was fixed by determining the absolute median of the market daily changes of each fruits separately.

4.6 Construction of Model-II

Each day was classified as having price value higher or lower or within "a" than the previous day for this experiment considering the movement from a category of highest gain (or) lowest gain (or) no gain/no loss (or) lowest loss (or) highest loss to the next day, thus letting classification of five states namely:

- State 1 (Highest gain): Today's price value change is greater than "a" rupees than the price value of the previous day.
- State 2 (Lowest gain): Today's price value change is lower than "a" rupees than the price value of the previous day.
- State 3 (No gain / No loss): Today's price value change is equal to "0" rupees than the price value of the previous day.
- State 4 (Lowest loss): Today's price value change is greater than"-a" rupees than the price value of the previous day.
- Step 5(Highest loss): Today's price value change is lower than "-a" rupees than the price value of the previous day.

$$Y_n = \begin{cases} 0, & if\ Z_n < a \\ 1, & if\ Z_n > a \\ 2, & if\ Z_n = a \\ 3, & if\ Z_n < -a \\ 4, & if\ Z_n > -a \end{cases}$$

The random variable $\{Y_n\}$ defined by above equation is known as a Markov chain with state space {0, 1, 2, 3, 4}. Here "a" is treated as threshold value of absolute price changes. There is no unique way of determining the threshold in this nature of studies. Generally, researchers used common parameters like measures of central tendency to find the threshold point. In this study, the main focus is only on the changes of the price pattern not on the price values. Here threshold value was fixed by determining the absolute median of the market daily changes of each fruits separately.

V. Result and Discussion

Probability values for each vector movement of the Model-I is described as follows:

Table 2: *π and μ Values for Model–I*

Model-I	
Banana	*Custard Apple*
State 0 1 2 0 [0 0.9756 0.0243] 1 [0.1659 0.7261 0.1078] 2 [0.0370 0.9629 0] *π = [0.1325 0.7796 0.0875]* *μ = [7.5471 1.2827 11.4285]*	State 0 1 2 0 [0 0.9878 0.0121] 1 [0.4347 0.4239 0.1413] 2 [0.0217 0.9782 0] *π = [0.2760 0.6305 0.0926]* *μ = [3.6231 1.5860 10.7991]*
Gooseberry	*Guava*
State 0 1 2 0 [0 0.9726 0.0273] 1 [0.1540 0.7004 0.1455] 2 [0 0.9726 0.0273] *π = [0.1177 0.7644 0.1177]* *μ = [8.4961 1.3082 8.4961]*	State 0 1 2 0 [0.0123 0.9876 0] 1 [0.1767 0.6241 0.1991] 2 [0.0111 0.9777 0.0111] *π = [0.1309 0.7228 0.1456]* *μ = [7.6394 1.3835 6.8681]*
Mango	*Mosambi*
State 0 1 2 0 [0.0111 0.9666 0.0222] 1 [0.2027 0.5990 0.1981] 2 [0 1 0] *π = [0.1474 0.7102 0.1422]* *μ = [6.7842 1.4080 7.0323]*	State 0 1 2 0 [0 1 0] 1 [0.1834 0.6241 0.1923] 2 [0.0229 0.9655 0.0114] *π = [0.1358 0.7229 0.1408]* *μ = [7.3637 1.3833 7.1022]*
Grapes	*Sapota*
State 0 1 2 0 [0 0.9770 0.0229] 1 [0.1914 0.6193 0.1891] 2 [0.0229 0.9655 0.0114] *π = [0.1406 0.7181 0.1409]* *μ = [7.1123 1.3925 7.0972]*	State 0 1 2 0 [0 0.975 0.025] 1 [0.1741 0.6294 0.1964] 2 [0.0222 0.9777 0] *π = [0.1293 0.7245 0.1455]* *μ = [7.7339 1.3802 6.8728]*

Papaya	*Watermelon*
State 0 1 2 0 [0 0.9888 0.0111] 1 [0.1982 0.6280 0.1737] 2 [0.0126 0.9873 0]	*State* 0 1 2 0 [0.0298 0.9552 0.0149] 1 [0.1371 0.7046 0.1582] 2 [0 0.9870 0.0129]
π = *[0.1455 0.7263 0.1279]*	π = *[0.3840 0.5289 0.0859]*
μ = *[6.8728 1.3768 7.8186]*	μ = *[2.6041 1.8971 11.6414]*

First vector (0, 0.9756, 0.0243), indicates that if a given day has increased price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 0%,97.5%,2.43% respectively. Second vector (0.1659, 0.7261, and 0.1078) indicates that if a given day remains within the threshold limits of price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 16.5%, 72.6%, 10.7% respectively. The last vector (0.0370, 0.9629, and 0) indicates that if a given day has decreased price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 3.7%, 96.2%, 0% respectively.

Based on the probability matrix P_1, all states are communicated and a periodic, so that it is an ergodic chain. Hence, π = [0.1325 0.7796 0.0875] is interpreted as banana's price movement among categories is being a Markov chain with transition probability matrix P_1, after many days, consist of the following proportions in each category: 13.2% in the state-0, 77.9% in the state-1, and 8.7% in the state-2. Further, the mean recurrence time of the corresponding states 0, 1and 2 are [7.5471 1.28271 1.4285] days respectively.

In a similar manner, remaining fruits: Custard Apple, Gooseberry, Guava, Mango, Mosambi, Grapes, Sapota, Papaya, Watermelon results are estimated and which are presented in Table, consist of the transition probability matrix, long-run invariant distribution and mean recurrence time for each model separately.

For the second model, the transition matrix was found to be:

$$P_2 = \begin{array}{c} \textit{State} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 0 & 0.9821 & 0.0178 & 0 \\ 0.0285 & 0 & 0.9142 & 0.0285 & 0.0285 \\ 0.1172 & 0.0774 & 0.6460 & 0.0796 & 0.0796 \\ 0 & 0 & 1 & 0 & 0 \\ 0.0540 & 0 & 0.9459 & 0 & 0 \end{bmatrix} \end{array}$$

$$\pi = [0.0905 \quad 0.0566 \quad 0.7315 \quad 0.0614 \quad 0.0598]$$

$$\mu = [11.0497 \quad 17.6678 \quad 1.3605 \quad 16.2866 \quad 16.7224]$$

Probability values, for each vector movement of the Model 2is described as follows:

First vector (0, 0, 0.9821, 0.0178, 0), indicates that if a given day has increased price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 0%, 0%, 98.2% ,1.7% and 0% respectively. Second vector (0.0285, 0, 0.9142, 0.0285, 0.0285), indicates that if a given day remains within the threshold limits of price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 2.8%, 0%, 91.4%,2.8% and 2.8% respectively. The third vector (0.1172, 0.0774, 0.6460, 0.0796 and 0.0796), indicates that if a given day has decreased price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 11.7%, 7.7%, 64.6%, 7.9% and 7.9% respectively.

The fourth vector (0, 0, 1, 0, 0), indicates that if a given day has decreased price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 0%,0%,1%, 0% and 0%respectively. The last vector (0.0540, 0, 0.9459, 0, 0), indicates

that if a given day has decreased price value, then the next day's price will increase or remain within the threshold limits or decrease in the following percentages 5.4 0%, 94.5%, 0% and 0%respectively.

Based on the probability matrix P_2, all states are communicated and a periodic, so that it is an ergodic chain. Hence, π = [0.0905 0.0566 0.7315 0.0614 0.598] is interpreted as banana's price movement among categories is being a Markov chain with transition probability matrix P_2, after many days, consist of the following proportions in each category:9.0%in the state-0,5.6% in the state1, 73.1% in the state-2, 6.1% in the state-3, 5.9% in the state-4. Further, the mean recurrence time of the corresponding states 0, 1, 2, 3 and 4 are [11.0497 17.667 81.3605 16.2866 16.7224] days respectively.

In a similar manner, remaining fruits: Custard Apple, Gooseberry, Guava, Mango, Mosambi, Papaya, Sapota, Grapes, Watermelon results are estimated and which are presented in Table3, consist of the transition probability matrix, long-run invariant distribution and mean recurrence time for each model separately.

Table 3: *π and μ Values for Model–II*

<table>
<tr><th colspan="2">Model-II</th></tr>
<tr><td>

Banana

State	0	1	2	3	4
0	0	0	0.9821	0.0178	0
1	0.0285	0	0.9142	0.0285	0.0285
2	0.1172	0.0774	0.6460	0.0796	0.0796
3	0	0	1	0	0
4	0.0540	0	0.9459	0	0

π = [0.1325 0.7796 0.0875]

μ = [7.5471 1.2827 11.4285]

</td><td>

Custard Apple

State	0	1	2	3	4
0	0	0	0.9870	0	0.0129
1	0	0	1	0	0
2	0.1689	0.0405	0.6148	0.0630	0.1126
3	0	0	1	0	0
4	0.0392	0	0.9607	0	0

π = [0.1245 0.0291 0.7186 0.0452 0.0824]

μ = [8.0321 34.3642 1.3915 22.1238 12.1359]

</td></tr>
<tr><td>

Gooseberry

State	0	1	2	3	4
0	0	0	0.9726	0	0.0273
1	0	0	1	0	0
2	0.1540	0.0042	0.7004	0.0021	0.1392
3	0	0	1	0	0
4	0	0	1	0	0

π = [0.1181 0.0032 0.7670 0.0016 0.1099]

μ = [8.4674 312.5 1.3037 625 9.0991]

</td><td>

Guava

State	0	1	2	3	4
0	0.0138	0	0.9861	0	0
1	0	0	1	0	0
2	0.1543	0.0581	0.6241	0.0715	0.0917
3	0	0	1	0	0
4	0.0487	0	0.9512	0	0

π = [0.1136 0.0421 0.725 0.0051 0.0664]

μ = [8.8028 23.7529 1.3793 196.0784 15.0602]

</td></tr>
<tr><td>

Mango

State	0	1	2	3	4
0	0.0120	0	0.9879	0	0
1	0	0	0.8888	0.0555	0.0555
2	0.1867	0.0410	0.5990	0.0273	0.1457
3	0	0	1	0	0
4	0	0	1	0	0

π = [0.1342 0.0291 0.7105 0.0209 0.1050]

μ = [7.4515 34.3642 1.4074 47.8468 9.5238]

</td><td>

Mosambi

State	0	1	2	3	4
0	0	0	1	0	0
1	0	0	1	0	0
2	0.1454	0.0648	0.6241	0.0626	0.1029
3	0	0	0.9642	0	0.0357
4	0.0425	0	0.9574	0	0

π = [0.1049 0.0454 0.7007 0.0438 0.0736]

μ = [9.5328 22.0264 1.4271 22.8310 13.5869]

</td></tr>
<tr><td>*Grapes*</td><td>*Sapota*</td></tr>
</table>

State 0 1 2 3 4 0 [0 0.0188 0.9622 0 0.0188] 1 [0 0 0.9777 0.0222 0] 2 [0.1171 0.0990 0.6193 0.0923 0.0720] 3 [0.0232 0 0.9534 0.0232 0] 4 [0 0 1 0 0] π = [0.0857 0.0727 0.7189 0.0695 0.0530] μ = [11.6686 13.7551 1.3910 14.3884 18.8679]	State 0 1 2 3 4 0 [0.0151 0 0.9696 0 0.0151] 1 [0 0 1 0 0] 2 [0.1406 0.0647 0.6294 0.0915 0.0736] 3 [0.0243 0 0.9756 0 0] 4 [0.0294 0 0.9705 0 0] π = [0.1066 0.0469 0.7251 0.0663 0.0549] μ = [9.3808 21.3219 1.3791 15.0829 18.2149]
Papaya State 0 1 2 3 4 0 [0 0 1 0 0] 1 [0 0 0.9705 0 0.0294] 2 [0.1380 0.0757 0.6280 0.0556 0.1024] 3 [0 0 1 0 0] 4 [0.0212 0 0.9787 0 0] π = [0.1018 0.0550 0.7266 0.0403 0.0760] μ = [9.8231 18.1818 1.3762 24.8138 13.1578]	*Watermelon* State 0 1 2 3 4 0 [0 0 0.9729 0 0.0270] 1 [0.0256 0.0256 0.9487 0 0] 2 [0.0759 0.0801 0.7046 0.1181 0.0210] 3 [0 0 0.9824 0.0175 0] 4 [0 0 1 0 0] π = [0.0596 0.0630 0.7672 0.0922 0.0177] μ = [16.7785 15.8730 1.3034 10.8459 56.4971]

Notice that for model II, in the entire transition matrix each row vector contains a more likelihood of probabilities in the first column; indicate that irrespective of what state a day occupies, there is a high probability that the next day will be a day of gain. This is encouraging for investors since the price movement will not be down in a sustained manner, but more likely high.

Conclusion

This paper presents a Markov chain modeling of daily fruit price movement in Salem District, a pioneering study in this context. The method used is valuable for analyzing and predicting price movements in the market. The study's results indicate that the daily price movement pattern for each fruit follows a similar Markov chain model, showing a higher likelihood of gains than consecutive days of loss, leading to slow and steady growth.

However, the functionality of the Salem market is influenced by various factors such as daily demand and investor psychology, making it challenging for any single method, including the Markov chain, to accurately predict daily price changes. While the analysis results are promising, the estimates of mean recurrent times may have limited validity. The study demonstrates that a regional market approach yields better results and suggests that a similar framework could be applied to other markets in the region. Although the current model is used to predict price movement patterns, it can be further developed to forecast the magnitude of price movements.

References

[1] Aparna, B. and Sarat, K. (2017). Markov Chain Modeling for Prediction on Future Market Price of Potatoes With Special Reference to Nagaon District, Assam, India. *IOSR Journal of Business and Management*, 19(12):25–31.

[2] Bhat, U. N. (1972). *Elements of Applied Stochastic Processes*. New York: John Wiley & Sons, Inc.

[3] Divya, K., Prahadeeswaran, M., Malargodi, M. and Uma, M. (2024). *Comprehensive Analysis of Production and Import Dependency of Pulses in India. Legume Research*, 47(6):996–1001.

[4] Eunice, W. M. and Peter, K. N. (2017). Determinants of Price Fluctuations on Vegetables

and Fruits in Karatina Open Air Market. *Scholars Journal of Economics, Business and Management*, 4:454–662.

[5] Hongyu Yang, Yuanxin Cao, Yuemeng Shi, Yuling Wu, and Youzhu Li. (2022). The Dynamic Impacts of Weather Changes on Vegetable Price Fluctuations in Shandong Province, China: An Analysis Based on VAR and TVP-VAR Models. *Agronomy*, 12(11):1–27.

[6] Jhade Sunil, Abhishek Sing, Kuldeep Rajpoot and Manjubala, M. (2023). Export Performance and Comparative Analysis of Onion with Reference to India. *International Journal of Statistics and Applied Mathematics*, 8(3):133–140.

[7] KhemChand, ASuresh, MBD astagiri, and Shalander Kumarand Subhasis Mandal. (2021). Fruit Marketing: Its Efficiency and Supply Chain Constraints in India: A Case Study. *Indian Journal of Agricultural Sciences*, 91(8):1146–1150.

[8] Parzen, E. (1962). *Stochastic Processes*. San Francisco: Holden-Day Inc.

[9] Prasenjit Kundu, Nayan Kishor Adhikary and Abhijit Ghosal. (2019). Vegetables Marketing Scenario and Trends in Prices and Arrivals of Vegetables in Different Markets of South 24 Parganas District of West Bengal, India. *International Journal of Current Microbiology and Applied Sciences*, 8(12):1821–1840.

[10] Priyanka, T., Kerur, N. M., Naik, B. K., Yeladhahalli, R. A. and Venugopal, C. K. (2024). Export Competitiveness and Direction of Trade of Banana from India: A Agribusiness Performance Appraisal. *International Journal of Agriculture Extension and Social Development*, 7(2):438–442.

[11] Ravi Shankar Pardhi, Rakesh Singh, Santosha Rathod and Prashant Kumar Singh. (2016). Effect of Price of Other Seasonal Fruits on Mango Price in Uttar Pradesh. *Economic Affairs*, 61(4):627–631.

[12] Sachin kumar Verma, Shivkumar, Deepshika, Monmoth Gopal and Deep Narayan Yadav. (2024). Food Grain Trade Prospect of India: Markov Chain Analysis. *SSR Institute of International Journal of Life Sciences*, 10(2):5050–5056.

[13] Tamil Selvi, G., Mohan Naidu, M., Murthy, Ramana Murthy, B., and Rajeswari, S. (2021). Behaviour of Market Arrivals and Prices of Vegetables in Koyambedu Market, Chennai. *International Journal of Statistical Science*, 17(1):133–137.

[14] Xiaoxia Zhu and Xiuquan Xu. (2012). Markov Chain Analysis and Prediction on the Fluctuation Cycle of Fruit Price. *Journal of System and Management Sciences*, 2(3):40–49.

OPTIMIZING HIDDEN MARKOV MODELS WITH FUZZIFICATION TECHNIQUES

Vyshnavi. M, Muthukumar. M

•

[1]Research Scholar, Department of Statistics, PSG College of Arts & Science, Coimbatore-641014
[2]Assistant Professor, Department of Statistics, PSG College of Arts & Science, Coimbatore-641014
E-mail: [1]vyshnavimp@gmail.com, [2]muthukumar@psgcas.ac.in

Abstract

This work explores using fuzzified techniques to enhance the performance of Hidden Markov Models (HMMs) in handling uncertainties and imprecise inputs. We construct and evaluate three types of fuzzy HMMs: the Trapezoidal fuzzy HMM, the Sigmoidal fuzzy HMM, and the Gaussian fuzzy HMM. As part of our process, parameter estimations are calculated and models are chosen based on AIC, BIC, AICc, and HQIC criteria. Each state's mean, variance, and stationary distribution are calculated and examined to evaluate the predictability and stability of the models. We use the Viterbi technique to identify the most likely state sequences for the next five years. According to the results, the Gaussian Fuzzy HMM offers superior predicted accuracy and durability when compared to the other models. This paper emphasizes the advantages of using fuzzy membership functions in HMMs and provides the foundation for future research in different areas, such as agricultural data prediction.

Keywords: Trapezoidal, Sigmoidal, Gaussian, Stationary Distribution, Viterbi Algorithm

I. Introduction

HMMs were introduced by Baum and his colleagues in the late 1960s, and they have been widely employed in domains like speech recognition, the biological sciences, and finance. Hidden states, observable emissions, and state transition probabilities characterize an HMM. Hidden Markov Models are a potent tool for modeling sequential data in the fields of statistical modeling and machine learning. Statistical models represent systems having hidden states called Hidden Markov Models. They are composed of observable symbols, a collection of hidden states, transition probabilities between states, and the likelihood that symbols will be emitted from states. The conventional HMM relies on clear and accurate observations, however noise and fuzziness can affect observations in a lot of real-world situations. Traditional statistical methods are sometimes inadequate in addressing difficulties and uncertainties present in agricultural data, such as the oilseed area from 1992 to 2022. Even though standard HMMs are reliable when modeling sequential data, they may not be sufficient when handling fuzziness of this kind. As a result, it is imperative to create and assess models that can manage these uncertainties more skillfully to guarantee greater forecast accuracy and dependability.

Introduced by Zadeh in 1965, fuzzy logic incorporates degrees of membership instead of binary true or false logic, offering a mathematical framework to handle such inaccurate information. Fuzzy logic is an appropriate method for handling uncertainty since it extends the traditional set theory to

incorporate partial membership. The act of utilizing membership functions to convert crisp inputs into fuzzy values is known as "fuzzification." The degrees of membership within the interval [0, 1] are mapped to input values by these membership functions. The resulting fuzzy model's performance and interpretability are greatly influenced by the membership function selection. By adding fuzzy logic to the observation and state representation, FHMMs go beyond conventional HMMs. With this adjustment, the model can now handle observational uncertainty more effectively and generate more reliable predictions.

Ahmed T. Salawuden, et al. [2] introduced a genetic algorithm and hidden Markov model-based FTS forecasting model. Conventional techniques, such as the Baum-Welch algorithm, have trouble capturing the fuzziness present in real data. The model optimizes the HMM parameter estimation problem using GA and an objective function. Qi Deng and Dirk Soffker [11] proposed a driving behavior prediction model based on a novel approach that combines HMM and Fuzzy Logic. The model divides driving scenes into three scenarios: very safe, safe, and risky, with associated HMMs for each. A prefilter reduces collected signals to observed sequences with predefined features. NSGA-II is utilized to maximize model performance, and experimental data from real-world human driving behaviors confirms the efficacy of the fuzzy-based HMM. Ying-Kui Gu, et al. [17] proposed a gearbox vibration experiment, the study gathers vertical vibration signals from three gear fault states: normal, worn, and fractured teeth. A hidden Markov model and fuzzy comprehensive evaluation are used to provide an online diagnosis and performance evaluation model.

This study's main goal is to investigate how fuzzified approaches can be used to improve HMM performance. Our specific purpose is to build and assess three different kinds of fuzzy HMMs: the Gaussian fuzzy HMM, the Sigmoidal fuzzy HMM, and the Trapezoidal fuzzy HMM. We aim to increase the predictability and stability of the models and better handle imprecise inputs by using fuzzy membership functions. The concepts of fuzzy logic, HMMs, and how they combine to form FHMMs form the theoretical basis of this study. Fuzzy logic defines fuzzy sets and membership functions to give a mathematical framework for handling ambiguity and imprecision.

II. Methods

Fuzzy logic is an effective framework for dealing with uncertainty and imprecision in data, making it especially helpful in complicated systems where classic binary logic fails. Unlike classical logic, which uses strict true or false values, fuzzy logic allows for degrees of truth, expressed by values ranging from 0 to 1. This flexibility is done using fuzzy sets and membership functions, such as Sigmoidal, trapezoidal, and Gaussian functions, which determine an element's degree of membership in a set. Fuzzy logic systems use "if-then" rules to process inputs and generate outputs, providing a more sophisticated approach to decision making. The fuzzy inference system (FIS) automates this process by mixing fuzzy rules, and its output is frequently "defuzzified" to convert fuzzy results into crisp, actionable conclusions. Fuzzy logic, which is widely employed in control systems, expert systems, and forecasting, excels in circumstances where data is ambiguous or imprecise, effectively dealing with uncertainty using fuzzy membership functions.

I. Hidden Markov Model

A Statistical model known as a Hidden Markov Model (HMM) can be used to forecast the development of observable events that consist of non-observable internal components. The phenomenon under observation is referred to as a 'state', whereas the witnessed event is called a ‘symbol’. A hidden state invisible process and an observable symbol visible process are the two

stochastic processes that construct an HMM. Within a Markov chain formed by the hidden states, the underlying state determines the probability distribution of the observed symbol. A doubly embedded stochastic process is another name for an HMM [5].

Definition: The triplicate *(S, O, A, B, π)* defines HMM [4] where,

- A set of hidden states is denoted by $S = S_1, S_2, S_3, \ldots, S_n$.
- A set of m observable symbols at each time interval is represented by $O(t) = O_1, O_2, \ldots, O_m$.
- State transition probability is represented by A, which is defined as, $A = a_{ij} = \{ P(X_{t+1} = S_j | X_t = S_i) | 1 \leq i, j \leq n\}$. The probability of changing from state i at time t to state j at time $t+1$ is shown here by the symbol a_{ij}.
- B is the probability of generating a symbol $O(t)$ from state j, and it can be expressed by the formula $B = b_j(t) = \{P(O(t) | X(t) = S_j) | 1 \leq j \leq n\}$.
- The initial state probability is $\pi = \{\pi_i = P(X_1 = S_i) | 1 \leq i \leq n\}$.

Using the forward method, one may determine the likelihood of a sequence of observed symbols, $\alpha_t(i) = P(o_1, o_2, o_3, \ldots, o_m, s_t = s_i) = \sum_{j=1}^{n} \alpha_{t-1}(j) a_{ji} b_i(o_t)$, where $\alpha_t(i)$ indicates the chance of noticing the observations and being in states s_i at time t. The whole observation sequence's likelihood is

$$P(O) = \sum_{i=1}^{n} \alpha_T(i) \quad (1)$$

The Viterbi method determines which hidden state sequence is most likely to exist and HMM parameters are developed using the Baum-Welch algorithm to maximize the likelihood of the observed data.

II. Fuzzy Hidden Markov Model

To more effectively handle uncertainties and imprecisions in the data, a Fuzzy Hidden Markov Model (FHMM) combines fuzzy logic concepts with the conventional HMM architecture. An FHMM's primary concept is to use fuzzy sets to represent the states and observations, which enables more flexible and intricate modeling of the underlying stochastic processes. In a FHMM, states are represented using fuzzy sets, where each state is associated with a membership function that specifies the degree to which an observation belongs to that state. This fuzzy representation allows the model to handle imprecise and noisy data more effectively. Observations in an FHMM can also be fuzzy, paralleling the fuzziness of states, which enhances the model's capability to process uncertain data. Membership functions, such as sigmoidal, Gaussian, and trapezoidal, characterize the degree of fuzziness in both observations and states.

Transition probabilities in an FHMM indicate the likelihood of transitioning from one state to another and can be adjusted to reflect the fuzziness in the states. Similarly, emission probabilities represent the chances of observing a specific output from a given state and can incorporate the uncertainty in observations. Additionally, the initial state probabilities, denoted as π_i, can be fuzzified by describing the likelihood of the system being in a specific state at the beginning of the time step. This fuzzified approach to initial probabilities further enhances the model's ability to handle uncertainty from the onset of the observation sequence.

Hidden Markov Models use fuzzy membership functions to address uncertainty and imprecision in state allocations. Traditional HMMs confine observations to discrete states, which

might be restrictive since real-world state transitions are frequently lagging or confusing. HMMs use fuzzy membership functions to allow observations to belong to various states, capturing data variations and reflecting smoother transitions. This strategy is especially useful for dealing with noisy or ambiguous data since it enhances prediction accuracy and robustness. Furthermore, fuzzy HMMs provide more versatility by allowing the use of various membership function shapes to depict complex relationships. This makes fuzzy HMMs more adaptive and capable of addressing the complexities encountered in real-world applications such as forecasting and pattern recognition.

III. Trapezoidal Fuzzy Hidden Markov Model

Trapezoidal membership functions are used by a Trapezoidal Fuzzy Hidden Markov Model (TrFHMM) to reflect the fuzziness of states and observations. Trapezoidal membership functions, with their flat top that can capture more information, offer a more flexible technique to model uncertainty than triangular functions. The trapezoidal membership function Ψ is defined as:

$$\Psi(x; \check{a}, b, \acute{c}, \hat{d}) = \begin{cases} 0, & if\ x \leq \check{a} \\ \frac{x-\check{a}}{b-\check{a}}, & if\ a < x \leq b \\ 1, & if\ b < x \leq \acute{c} \\ \frac{d-x}{d-\acute{c}}, & if\ \acute{c} < x \leq \hat{d} \\ 0, & if\ x \geq \hat{d} \end{cases} \quad (2)$$

where b and $\acute{c}$ indicate the trapezoid's top and $\check{a}$ and $\hat{d}$, respectively, denote the lower and upper bounds [8].

IV. Sigmoidal Fuzzy Hidden Markov Model

To describe the fuzziness of states and observations, sigmoidal membership functions are used in a Sigmoidal Fuzzy Hidden Markov Model (SFHMM). Sigmoidal membership functions are especially helpful as a more realistic and gradual approach to dealing with uncertainty and imprecision. This is because they smoothly go from 0 to 1. The Sigmoidal membership function Ψ can be defined as:

$$\Psi(x; \check{a}, \acute{c}) = \frac{1}{1+e^{-\check{a}(x-\acute{c})}} \quad (3)$$

where $\acute{c}$ represents the sigmoid curve's center, or midway, and $\check{a}$ determines the function's slope [16].

V. Gaussian Fuzzy Hidden Markov Model

The Gaussian Fuzzy Hidden Markov Model (GFHMM) models the fuzziness of states and observations by using Gaussian membership functions. To effectively depict uncertainty in data, Gaussian functions provide effortless and gradual transitions. The Gaussian membership function Ψ is defined as:

$$\Psi(x; \acute{c}, \sigma) = exp\left(-\frac{(x-\acute{c})^2}{2\sigma^2}\right) \quad (4)$$

where $\acute{c}$ is the center (mean) of the Gaussian function, and σ is the standard deviation [15].

VI. Fuzzification of Observations

Considering a set of $\ddot{O} = \{ \ddot{O}_1, \ddot{O}_2, \ldots, \ddot{O}_{\dot{T}} \}$, the fuzzified observation $\tilde{O}_{\dot{t}}$ at time $\dot{t}$ for a fuzzy state $\check{S}_i$ is

$$\tilde{O}_{\dot{t}}(\check{S}_i) = \Psi_{\check{S}_i}(\ddot{O}_{\dot{t}}) \tag{5}$$

where $\Psi_{\check{S}_i}(\ddot{O}_{\dot{t}})$ is the membership function value of the observation $\ddot{O}_{\dot{t}}$ concerning state $\check{S}_i$.

The likelihood of a state $\check{S}_i$ to state $\check{S}_j$ transition accumulates the same as in the conventional HMM;

$$\hat{A}_{ij} = P(\check{S}_{\dot{t}+1} = \check{S}_j \mid \check{S}_{\dot{t}} = \check{S}_i) \tag{6}$$

Given a state $\check{S}_j$, the emission probability of detecting $\ddot{O}_{\dot{t}}$ is fuzzily defined as,

$$\mathcal{B}_j(\ddot{O}_{\dot{t}}) = \Psi_{\check{S}_i}(\ddot{O}_{\dot{t}}) \tag{7}$$

The forward probability $\acute{\alpha}_{\dot{t}}(i)$ is defined as;

$$\acute{\alpha}_{\dot{t}}(i) = P (\ddot{O}_1, \ddot{O}_2, \ldots, \ddot{O}_{\dot{t}}, \check{S}_{\dot{t}} = \check{S}_i \mid \lambda) \tag{8}$$

with the initial condition,

$$\acute{\alpha}_1(i) = \pi_i \, \mathcal{B}_i(\ddot{O}_1) \tag{9}$$

and the recursive formula,

$$\acute{\alpha}_{t+1}(j) = \left(\sum_{i=1}^{N} \acute{\alpha}_{\dot{t}}(i) \, \hat{A}_{ij}\right) \mathcal{B}_j(\ddot{O}_{\dot{t}+1}) \tag{10}$$

The backward probability $Б_{\dot{t}}(i)$ is defined as;

$$Б_{\dot{t}}(i) = P (\ddot{O}_{\dot{t}+1}, \ddot{O}_{\dot{t}+2}, \ldots, \ddot{O}_{\dot{t}} \mid \check{S}_{\dot{t}} = \check{S}_i \mid \lambda) \tag{11}$$

with the initial condition,

$$Б_T(i) = 1 \tag{12}$$

and the formula for recursion,

$$Б_{\dot{t}}(i) = \sum_{j=1}^{N} \hat{A}_{ij} \, \mathcal{B}_j(\ddot{O}_{\dot{t}+1}) Б_{\dot{t}+1}(j) \tag{13}$$

One well-known method for estimating HMM parameters is the Baum-Welch algorithm. It is a particular kind of Expectation-Maximization (EM) method created with HMMs. The algorithm iteratively refines the model parameters to optimize the likelihood of the observed data. The probabilities that follow are defined for parameter re-estimation [14],

- $\xi_{\dot{t}}(i, j) = P (\check{S}_{\dot{t}} = \check{S}_i, \check{S}_{\dot{t}+1} = \check{S}_j \mid \ddot{O}, \lambda)$
- $\ddot{Y}_{\dot{t}}(i) = P (\check{S}_{\dot{t}} = \check{S}_i \mid \ddot{O}, \lambda)$

The following are the re-estimation formulas [13]:

$$\pi_i = \ddot{Y}_1(i) \tag{14}$$

$$\hat{A}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \ddot{Y}_t(i)} \quad (15)$$

$$B_j(k) = \frac{\sum_{t=1,\ddot{O}_t=k}^{T} \ddot{Y}_t(j)}{\sum_{t=1}^{T} \ddot{Y}_t(j)} \quad (16)$$

The Viterbi method is a dynamic programming technique that determines the most likely sequence of hidden states, or Viterbi path, that leads to a sequence of observable events. Based on the observed sequence of events and the model parameters, the algorithm improves in forecasting, helping to predict future states.

Initialization

$$V_1(i) = \pi_i \ \mathcal{B}_i(\ddot{O}_1) \quad (17)$$

Recursion

$$V_t(j) = \max_i \left(V_{t-1}(i)\, \hat{A}_{ij} \right) \mathcal{B}_j(\ddot{O}_t)$$

$$\text{Path}(t)(j) = \arg\max_i \left(V_{t-1}(i)\, \hat{A}_{ij} \right) \quad (18)$$

Termination

$$\text{Optimal path (T)} = \arg\max_i V_T(i) \quad (19)$$

Backtracking

$$\text{Optimal path } (t) = \text{path}(t+1) \text{ optimal path}(t+1), \text{ for } t = T-1, T-2, \ldots, 1. \quad (20)$$

For every future time step T+k,

$$s_{T+k} = \arg\max_j \hat{A}_{s_{T+k-1},j} \quad (21)$$

The essential elements of the Viterbi algorithm and the procedures required to predict future states can be found in these equations [14].

VII. Data Source

The "Economics & Statistics Division within the Ministry of Agriculture & Farmers Welfare" website, may be accessed at **http://desagri.gov.in.**

VIII. Algorithm

The following are the steps for constructing and assessing Fuzzified HMM:

1. Import historical oilseed area data from 1992 to 2022.
2. Define the fuzzy membership functions: trapezoidal, sigmoidal, and Gaussian.
3. Assign fuzzy states (Low, Medium, and High) to observations using each fuzzification method.
4. Create and assign the sequence according to the highest membership value for each observation.
5. Set up the transition matrix, emission matrix, and initial probabilities for each F-HMM.

6. Using the Baum-Welch algorithm, estimate the transition and emission probability, as well as the initial state distribution.
7. Calculate the AIC, BIC, AICc, and HQIC for each FHMM.
8. Calculate the stationary distribution of each model.
9. Calculate the mean and variance of data for each state.
10. Use the Viterbi method to estimate the most likely state sequence during the next five years.
11. Compare models using AIC, BIC, AICc, HQIC, and prediction metrics to determine the best-performing model.

III. Results and Discussion

The findings from our examination of the three different kinds of FHMM are shown in this section. The main findings are fuzzy values, sequence prediction, parameter estimations, state parameters, and model selection criteria, which are compiled in the tables. Using fuzzy membership functions equations (2), (3), and (4), the fuzzification process transforms crisp values into fuzzy values.

Table 1: Fuzzy Values for oilseed area data.

Trapezoidal Fuzzy Value	Sigmoidal Fuzzy Value	Gaussian Fuzzy Value
0.870	0.224	0.528
1	0.668	0.980
1	0.643	0.995
1	0.692	0.953
1	0.613	0.999
1	0.450	0.782
1	0.802	0.546
1	0.435	0.749
1	0.316	0.928
1	0.345	0.950
0.745	0.624	0.968
1	0.158	0.708
0.52	0.882	0.891
0.86	0.913	0.990
1	0.733	0.878
1	0.766	0.788
0.56	0.886	0.907
1	0.613	0.999
0.78	0.848	0.737
1	0.692	0.953
1	0.727	0.891
1	0.927	0.998
1	0.524	0.923
1	0.643	0.995
1	0.663	0.983
0.745	0.271	0.454
0.79	0.329	0.481
0.86	0.837	0.691

1	0.965	0.708
1	0.975	0.504

The state sequences reported by each model are quite consistent, notably in differentiating times of low (1), medium (2), and high (3) oilseed areas. However, variances in state transitions highlight the complex nature represented by various fuzzy membership functions.

Table 2: FHMM-identified state sequence

Models	Sequences														
TrFHMM	1	2	2	2	2	2	2	2	1	1	1	1	3	3	2
	2	3	2	2	2	2	3	2	2	2	1	2	2	3	3
GFHMM	1	2	2	2	2	2	3	2	1	1	1	1	3	3	2
	2	3	2	3	2	2	3	2	2	2	1	2	3	3	3
SFHMM	2	2	2	2	2	2	2	2	1	1	1	1	2	2	2
	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3

Table 3: Parameter Estimation for FHMMs

Models	Parameter	State-1	State-2	State-2
TrFHMM	TPM	[0.50, 0.33, 0.16]	[0.11, 0.722, 0.16]	[0, 0.5, 0.33]
	EPM	[0.66, 0.16, 0.16]	[0.38, 0.33, 0.27]	[0, 0.16, 0.83]
	π	[0.2, 0.6, 0.2]		
SFHMM	TPM	[0.75, 0.25, 0]	[0.04, 0.92, 0.04]	[0, 0, 1]
	EPM	[0.50, 0.25,0.25]	[0.37, 0.29, 0.33]	[0, 0, 1]
	π	[0.133, 0.8, 0.06]		
GFHMM	TPM	[0.5, 0.33, 0.16]	[0.13, 0.53, 0.33]	[0, 0.62, 0.37]
	EPM	[0.66, 0.16, 0.16]	[0.46, 0.40, 0.13]	[0, 0.11, 0.88]
	π	[0.2, 0.5, 0.3]		

TrFHMM is appropriate for systems in which State-2 dominates and transitions between states are more evenly dispersed. Its balanced transition dynamics and emission probabilities make it adaptable to diverse conditions. GFHMM is extremely stable, particularly in State-2, and has an absorbing State-3, making it ideal for scenarios in which the system is intended to converge to a final state. The deterministic emission probabilities make precise predictions. The sigmoidal model strikes a balance between stability and flexibility, with frequent transitions between states and high emission probabilities for specific data. This model can handle systems with more complex state transitions and moderate stability.

Table 3: Performance of model parameters.

Models	Parameter	State-1	State-2	State-3
TrFHMM	δ	0.134	0.608	0.250
	μ	23.2216	26.165	28.165
	σ	1.1376	0.6439	0.6841
SFHMM	δ	0.138	0.532	0.330
	μ	22.64	26.2208	29

	σ	0.89065	1.0295	0.24041
GFHMM	δ	0.1452	0.5331	0.3214
	μ	23.2216	26.0657	27.805
	σ	1.1376	0.4312	0.768

Systems with moderate variability and a dominant State-2 are most suited for TrFHMM. Strongly preferred State-2 scenarios with very consistent observations in State-3 are well suited for SFHMM. For evenly distributed state preferences, GFHMM is the best option since it provides a balanced stationary distribution and minimal variability in State-2. The optimal model selection for a given prediction task is determined by the desired balance between state stability, variability, and distribution.

Table 4: Measures of Model Performance.

Models	Log-likelihood	AIC	BIC	AICC	HQIC
TrFHMM	-3.44	34.89	54.507	62.891	41.166
GFHMM	-3.367	**34.734**	**54.351**	**62.734**	**41.009**
SFHMM	-3.376	34.752	54.368	62.752	41.027

GFHMM performs better than TrFHMM and SFHMM models according to the evaluation measures. With the highest log-likelihood and the lowest values of AIC, BIC, AICC, and HQIC, the SFHMM offers the best fit and strikes the ideal balance between goodness of fit and model complexity. As a result, the GFHMM is the best model to use with the given data. These measurements also show that the TrFHMM and SFHMM are less efficient, despite their good performance.

The graph illustrates the various degrees of membership for each fuzzy state by representing it over various intervals. Evaluating the model's performance and capacity to capture the variations in the data requires an awareness of how the membership values are spread among the fuzzy states, which is made easier with the aid of this graphic.

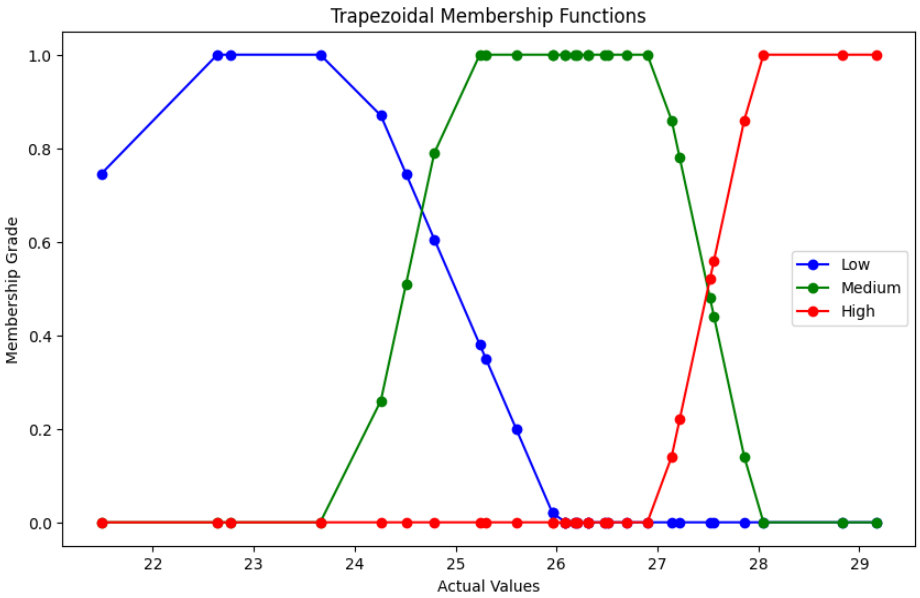


Figure 1: Shows the fuzzy states of trapezoidal membership functions.

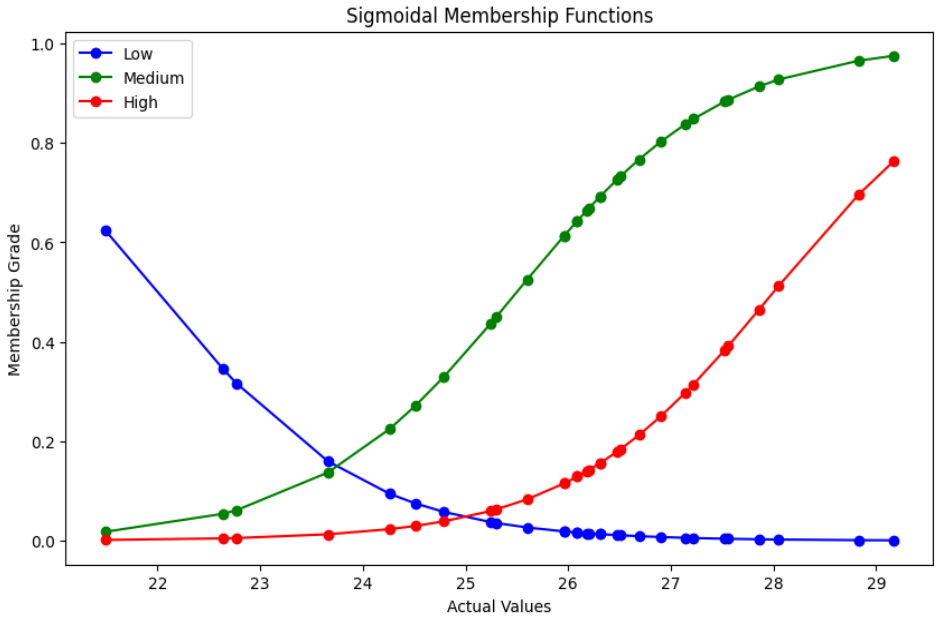


Figure 2: Demonstrates the fuzzy states of sigmoidal membership functions.

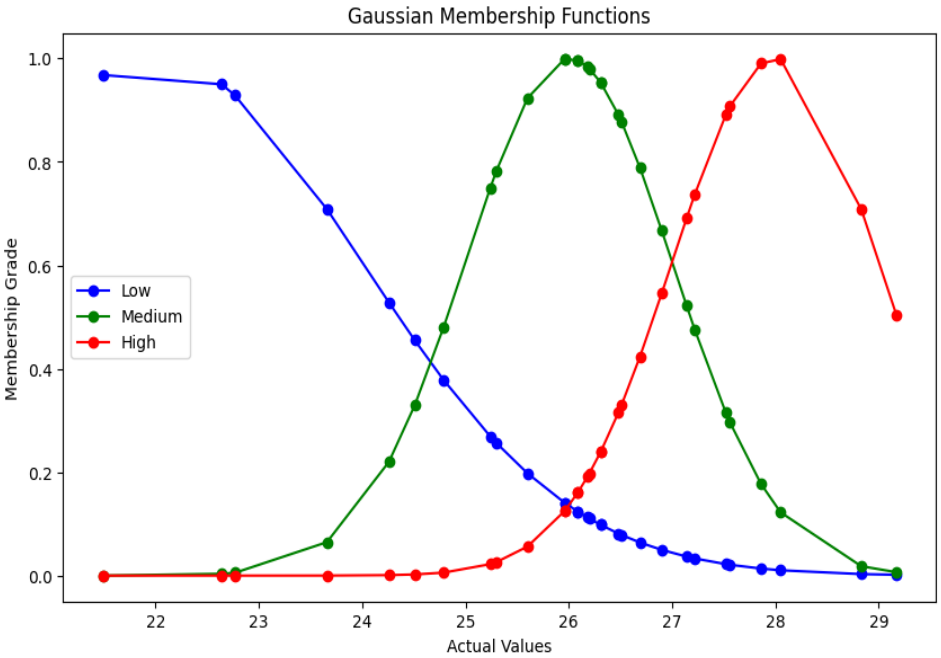


Figure 3: Fuzzy states are shown using Gaussian membership functions.

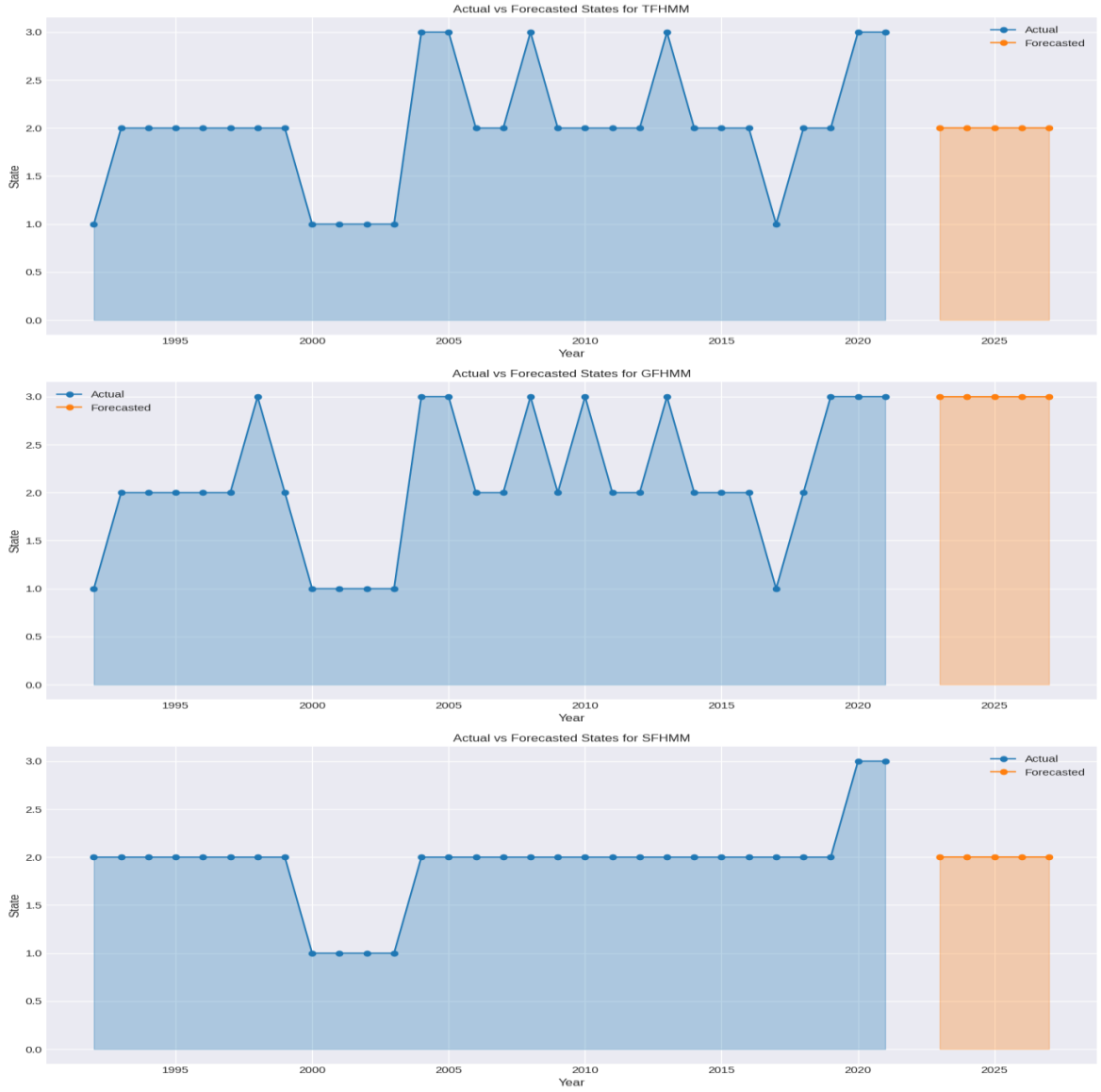


Figure 4: Shows the actual and forecasted states.

IV. Conclusion

The present study developed and evaluated three types of Fuzzy Hidden Markov Models. The study includes evaluating the parameters of each FHMM and comparing their performance against historical data from 1992 to 2022. Among the models, the Gaussian FHMM had the highest prediction accuracy and stability, as evidenced by favorable outcomes in model selection criteria. The estimates for the next five years indicated different results across the models: both the Sigmoidal and Trapezoidal FHMMs predicted that the oilseed area would remain in the "medium" state. In contrast, the Gaussian FHMM predicted that the oilseed area would reach a "high" state. Despite these variations, the Gaussian FHMM remains a reliable instrument for forecasting agricultural trends due to its superior performance. This study provides a significant framework for interpreting and forecasting agricultural commodity data, which will help the industry plan and make better decisions.

References

[1] Ahmed Gamal-Eldin, Salzenstein F and Christophe Collect (2010). Hidden fuzzy Markov chain model with K discrete classes. " 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), Kuala Lumpur, Malaysia, 2010, pp. 109-112.

[2] Ahmed T. Salawuden, Patrick J. Nyabvo, Hussein U. Suleiman, Izuagbe S. Momoh and Emmanuel K. Akut (2021). Heuristic hidden Markov model for fuzzy time series forecasting. International Journal of Intelligent Systems Technologies and Applications, 20(2), 146-166.

[3] Ankit Kumar Kushwaha, Kartik Shinde and Anil Kumar B (2024). Fuzzy probabilistic approach to identify driving behavior of two-wheeler drivers under mixed traffic conditions. IEEE Access.

[4] Bhavya Mor, Sunita Garhwal and Ajay Kumar, A systematic review of Hidden Markov Models and their applications, Archives of computational methods in engineering, 2021, 28: 1429-1448.

[5] Byung-Jun Yoon (2009). Hidden Markov Models and their applications in biological sequence analysis. Current Genomics, 10(6): 402-415.

[6] Ingmar Visser (2011). Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series. Journal of Mathematical Psychology, 55, 403-415.

[7] Joshni George and Seemon Thomas (2019). Modelling of agricultural price data using Hidden Markov Model. International Journal of Computational and Theoretical Statistics, ISSN (2210-1519).

[8] Kiran Khatter (2021). Interval valued trapezoidal neutrosophic set: multi-attribute decision making for prioritization of non-functional requirements. Journal of Ambient Intelligence and Humanized Computing, 12: 1039-1055.

[9] Leite P.B.C, Feitosa R.Q, Formaggio A.R, Costa G.A.O.P, Pakzad K and Sanche I.D.A (2008). Hidden Markov models applied in agricultural crops classification. Proceedings of GEOBIA (GEOgraphic Object-Based Image Analysis for the 21st Century).

[10] Omar Adil M. Ali and Balaswm Salem Sumait (2015). Comparison between the effects of different types of membership functions on fuzzy logic controller performance. International Journal of Emerging Engineering Research and Technology Vol. 3, Issue 3, pp. 76-83.

[11] Qi Deng and Dirk Soffker (2018). Improved driving behaviors prediction based on Fuzzy Logic-Hidden Markov Model (FL-HMM). 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, pp. 2003-2008.

[12] Saeed Jafarzadeh, and Sami Fadali (2014). Fuzzified Viterbi algorithm for hour-ahead wind power prediction. American Control Conference, Portland, OR, USA, pp. 1358-1363.

[13] Satyendra Nath Mandal, Pal Choudhury J, and Bhadra Chaudhuri S.R (2012). In search of suitable fuzzy membership function in prediction of time series data. International Journal of Computer Science, Vol. 9, Issue 3, 1694-0814.

[14] Shearer Robert (2007). Forecasting Israeli-Palestinian conflict with hidden Markov models. Military Operations Research, 5-15.

[15] Shruti Agarwal, Anshika Agarwal and Prabhakar Gupta (2020). Gaussian membership function used for voice recognition in fuzzy logic. International Journal of Recent Technology and Engineering, Vol. 8 Issue-5, ISSN: 2277-3878.

[16] Tamas Jonas (2007). Sigmoid functions in reliability-based management. Periodica Polytechnica social and management sciences, 15(2), 67-72.

[17] Ying-Kui Gu, Bin Xu, Hao Huang and Guang-Qi Qiu (2020). A fuzzy performance evaluation model for a gearbox system using Hidden Markov Model. Jiangxi University of Science and Technology, IEEE Access- Vol. 8, pp. 30400.

A SHORTAGES MULTI WAREHOUSE HAVING IMPERFACT ITEMS AND DIFFERENT DISCOUNT POLICY

Krishan Pal[1], Ajay Singh Yadav[1*], Seema Agarwal[1]

•

[1]Department of Mathematics, SRM Institute of Science and Technology, Delhi-NCR Campus, Ghaziabad, India, 201204
kp8910@srmist.edu.in, ajaysiny@srmist.edu.in, seemas@srmist.edu.in
[*]Corresponding author

Abstract

A multi-warehouse shortage model has been developed where demand is assumed to be deterministic. In reality, machines run for long periods during production and random failures may occur as the system transitions from a controlled to an uncontrolled state. During this time the production system produces defective products. Demand is assumed to be deterministic. Retailers offer a quantity discount per unit on the selling price of an item and in return receive a quantity- based discount on the purchase price of the item. A retailer has limited storage capacity and therefore requires additional space with unlimited storage capacity. This additional space is called a rented warehouse and its storage cost is higher than accompany- owned warehouse. The objective of this model is to study a multiple inventory model of defective items under quantity- based discounts, where defective items can be sorted and sold in a single batch with decision variables set to the optimal order quantity and optimal inventory and shipment quantity to increase overall profits to maximize the value for the retailer. A solution procedure for determining the optimal solution is presented and a numerical example is given to illustrate this study. A sensitivity analysis is also performed to examine the effect of changing parameter values on the optimal solution.

Keywords: Warehouse, Multi-warehouse; Demand, Imperfect items; Shortages; Quantity Discount

I. Introduction and Literature Survey

The essential part of business is Inventory management, as it constitutes a major part of capital expenditure. Hence, inventory management plays a vital role in running a business flexibly. There are several features that affect a business such as customer demand, deterioration of goods, manufacturing defects, stock shortages, quantity based discounts offered by suppliers etc. Managing these features can directly affect a retailer's overall profits. Inventory consists of raw materials, partially finished goods and finished goods. This type of inventory involves ordering costs, carrying costs, and stock out costs. Purchasing costs may also include carrying costs, but most models assume these costs are constant unless volume discounts are involved. The quantity and timing of orders are the most important decisions in inventory management. These decisions help increase a company's overall profits. It is common for suppliers to offer discounts for larger batch sizes. There are several reasons why suppliers offer discounts to retailers. These reasons include increasing cash

flow, reducing product inventory, increasing market share, or simply retaining the retailer. For retailers, inventory costs increase while bulk-ordering costs can be reduced.

These costs can be controlled through mathematical models, allowing retailers to make more profits. It is assumed that products are perfectly manufactured during the manufacturing stage, but in reality, this is not possible due to machine breakdowns after long hours of operation. Therefore, retailers need to inspect defective products and separate defective from good products. In addition, retailers have limited space to store inventory. In this situation, retailers need a separate space to store their inventory, the so-called rental warehouses. Rental warehouses offer unlimited storage space and better preservation options. That is why the storage costs in rental warehouses are higher than in their own warehouses. When demand increases, bottlenecks can occur. Product imperfections can also cause stock outs.

This paper describes a company that sells goods in a deterministic demand environment across multiple warehouses. It assumes discounts on every unit, defects, and shortages. The goal of this study is to optimize quantities and reorders to maximize profits across the inventory. The model becomes more realistic when considering volume discounts, a feature that is important for small retailers. Here, the procedure to obtain the optimal solution is used and a numerical example is provided to verify the results. A sensitivity analysis is also performed to study the effect of the parameters on the optimal solution. This article start with assumptions and notations and end with the conclusion of the article The model is based on a multiple warehouse environment with markdowns and shortages of incomplete items. S. Papachristos, K. Skouri [1] which ensures that the rate of backlogged demand increases as the waiting time to the following replenishment point decreases. Seto et al. [2] described a two-camp model with deterioration and demand growth over time. Jaggi et al. [3] developed a two-camp model with deterioration and inflation, where demand is based on the selling price. Jaggi et al. [4] proposed a two-camp model for deteriorating items with credit financing policy. Furthermore, the model assumes incomplete items. Assuming inventory-dependent demand, Jaggi et al. [5] developed a model with defects and damaged goods in two warehouse environments. This model was based on maximizing retailer profits with deterministic demand. Mandal & Giri [6] proposed a two-inventory model with incomplete items, where demand is based on inventory. An EOQ model was developed by Shah & Naik [7] with price-sensitive demand and discount policy. Volume discounts were offered in this model. Shaikh et al. [8] use a two-camp model under inflationary conditions. . Dhaka et al. [9] developed an inventory model with defects. They assumed inventory-dependent demand due to credit lending policy. By accepting incomplete articles, another two-camp model is proposed by Mashud et al. [10] assuming discountability and non-instantaneous deterioration rates. There were also some bottlenecks in the credit lending policy. Mandeep et al. [11] developed a two-warehouse model that considered the impact of human error on inventory levels and stockouts. Priyanka & Pareek [12] developed a two-inventory model by incorporating non-instantaneous deterioration and stochastic demand. The model also assumes credit financing Gilotra et al. [13] developed another model.

Additionally, we added carbon emissions and human error to the model. The impact of defective items is accounted for in an inventory model developed by Aastha et al. [14]. They assumed shortages in two warehouse environments. Another inventory model was developed by Aastha et al. [15] which is based on volume discounts under credit financing policies. In this model, the authors assumed that non-instantaneous deterioration in demand depends on advertising and selling prices. Mishra et al. [16] studied green technologies and developed a supply chain inventory model in the context of shortage and surplus payment systems. For fresh food products, Jayaswal et al. also developed an EOQ model. [17]. They developed it with a two-stage credit and lending policy under the learning effect and illustrated it with a numerical example. Shah et al. [18] formulated an EPQ model for perishables with reliability and inflation under a two-stage credit policy.

The total cost function was also minimized. Alamri et al. [19] developed an EOQ model that includes learning effect and carbon emissions, taking into account several aspects of naturally spoiled products. The Authors also discussed the implications for various related fields. Kumar et al. [20] Slack inventory ordering policy with pre-pay and post-pay strategies. Mittal et. al. [21] gave a model of the retailer's ordering policy for deteriorating imperfect quality items when demand and price are time-dependent under inflationary conditions and permissible delay in payments Masae et. al. [22] developed a model of order picker routing in warehouse. Ibrahim Hassan [23] gave a model on the role of warehouse layout and operations in warehouse efficiency: a review on it. Here, a mathematical model is develop with the warehouse environment. This include the analysis as well as the sensitivity analysis of the parameters. The given tables discuss the account of the article.

II. Assumptions and Notation

Assumptions

1. A known, and constant demand rate is considered.
2. Replenishment rate is instantaneous.
3. Inventory to be sold in single batch only.
4. Proportion of defective items p follows a uniform distribution$[\alpha, \beta]$where $[0 \leq \alpha \leq \beta \leq 1]$ (Wee et al., 2007)
5. A fully backlogged shortage are allowed.
6. The purchase price is offered under the following, all-units quantity discount scheme: (Papachristos & Skouri, 2003)

$$u_i(\mu) = \begin{cases} u_1 \ for\ m_1 < q \leq m_2 \\ u_2 \ for\ m_2 < q \leq m_3 \\ \vdots \\ u_n \ for\ m_n < q \end{cases}$$

Notation: The notation is used in the present study is describe in Table 1.

Table 1: *Notation and description of the inventory life cycle mode*

Notation	Description
y_1	Lot size units/cycle
W	limited space of OW units/cycle
$y_1 - W$	unlimited space in RW units/cycle
D	demand rate unit / year
j, A	variable cost/ unit and ordering cost / order
p	proportion of imperfect items in y_1
f(p)	p.d.f of p

s	selling price of items with good quality
v	selling price of defective items,$v < j$
b	backordering cost per unit
h_r	holding cost for the items in RW
h_o	holding cost for the items in OW
m_i	ith price breaking point, i=1, 2, 3, …, n
$u_i = u_i(\mu)$	per unit material cost
x	rate of screening
d	cost for screening
t_r	screening rate in RW
t_w	screening rate in OW
t_1	time for RW when all the units used
t_2	time for OW when all the units used
t_3	time period of shortages
B	backorder quantity in units
T	length of a cycle
y_{10pt}	optimal lot size
$TR(B,\mu)$	sale revenue

III. Model Description

I. Problem Definition

Initially, at time $t = 0$, y_1 units come in the system. As the system has rent-warehouse too, W units are stored in own-warehouse and rest of the units are stored in rent-warehouse. But due to high holding costs in rent-warehouse, first units are consumed from rent-warehouse and then from own-warehouse. Generally, it is assumed that all the units come perfectly in the system but practically this is not possible due to improper transport, low quality of raw material. In this case, retailer must apply the screening process with the rate of x units per unit time. This paper assumes the p percent of imperfect items found in y_1 units, where pa random variable is whose p.d.f. is, $f(p)$, and its mean is $E(p) = p$. Thus, py_1 defective items and $(1-p)y_1$ non-defective items found in y_1 units. These imperfect items sold at v per unit salvage value, where $v < j$. After this screening process, the demand is fulfilled from rent warehouse by time t_1 and then upcoming demand is fulfilled from own warehouse. After t_3, shortages occur when both the warehouse is exhausted. (Figure. 1)

II. Mathematical Model

The time horizon of total inventory is given by,

$$T = (1-p)\frac{y_1}{D} \tag{1}$$

Sale revenue ($TR(p, y_1)$) is the sum of sales of imperfect and the perfect items.

$$TR(p, y_1) = sy_1(1-p) + vpy_1 \tag{2}$$

And Total cost ($TC(B, y_1)$) = ordering cost + purchase cost + holding cost + backordering cost + screening cost

$$TC(B, y_1) = u_i y_1 + A + jy_1 + dy_1 + h_r\left[\left(\frac{(t_1(y_1 - W)(1-p)}{2}\right)t_r p(y_1 - W)\right] + h_r\left[Wt_w + W(t_1 - t_w)(1-p) + \frac{(w(1-p) - B(t_2 - t_1)}{2}\right] + \frac{Bbt_3}{2} \tag{3}$$

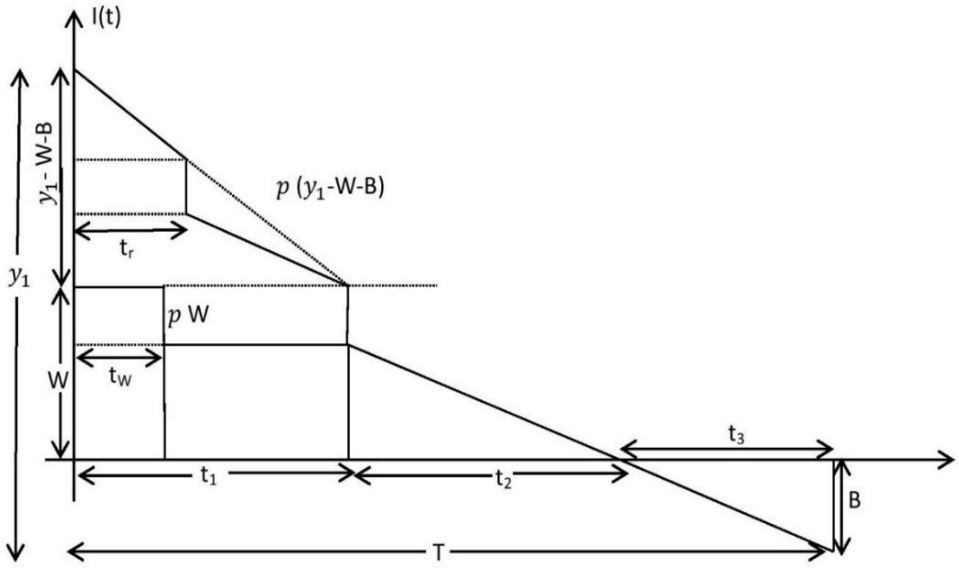


Figure 1: *Multi warehouse Modelling of Inventory*

Since $t_1 = \frac{(y_1-W)(1-p)}{D}, t_r = \frac{y_1-W}{x}, t_W = \frac{W}{x}, t_2 = \frac{W(1-p)-B}{D}, t_3 = \frac{B}{D}$

Then Eq (3) is

$$TC(B, y_1) = u_i y_1 + A + j y_1 + d y_1 + h_r\left(\frac{1}{2D}(y_1 - W)^2(1-p)^2 + \frac{p(y_1-W)^2}{x}\right) + h_o\left(\frac{W(y_1-W)(1-p)^2}{D} + \frac{W^2 p}{x} + \frac{1}{2D}\left((W(1-p)-B)^2 - W(y_1-W)(1-p)^2 + B(1-p)(y_1-W)\right)\right) + \frac{B^2 b}{2D} \quad (4)$$

The total profit per unit of time ($TPU(B, y_1)$) is calculated with the help of total revenue per unit of time and total cost per unit of time. Hence,

$$\mathrm{TPU}(p, y_1) = \frac{TR(p,y_1) - TC(B,y_1)}{T} \quad (5)$$

$$= \mathrm{D}\left(s - v + h_r\frac{(y_1-w)^2}{xy_1} + h_o\frac{W^2}{xy_1}\right) + \left(\frac{D}{1-p}\right)\left(v - u_i - \frac{A}{y_1} - j - d - h_r\frac{(y_1-w)^2}{xy_1} - h_o\frac{W^2}{xy_1}\right)$$

$$-h_r(y_1 - W)^2\left(\frac{1-p}{2y_1}\right) - \mathrm{h}_o\left(\frac{W(y_1-W)(1-p)}{2y_1} + \frac{W^2(1-p)}{2y_1} - \left(\frac{WB}{y_1}\right) + \frac{B(y_1-W)}{2y_1}\right) - \left(\frac{1}{1-p}\right)\left(\frac{\mathrm{h}_o B^2}{2y_1} - \frac{B^2 b}{y_1}\right) \quad (6)$$

Since p is a random variable that follows a uniform distribution with known p.d.f. f (p) then the expected value of total profit ($ETPU(B, y_1)$) is

$$ETPU(B, y_1) = D\left(s - v + \mathrm{h}_r\frac{(y_1-W)^2}{xy_1} + \mathrm{h}_o\frac{W^2}{xy_1}\right) + E\left(\frac{1}{1-p}\right)D\left(v - u_i - \frac{\mathrm{A}}{y_1} - j - d - \mathrm{h}_r\frac{(y_1-W)^2}{xy_1} - \mathrm{h}_o\frac{W^2}{xy_1}\right) - \mathrm{h}_r(y_1-W)^2\left(\frac{1-E(p)}{2y_1}\right)$$

$$-\mathrm{h}_o\left(\frac{W(y_1-W)(1-E(p))}{2y_1} + \frac{W^2(1-E(p))}{2y_1} - \left(\frac{WB}{y_1}\right) + \frac{B(y_1-W)}{2y_1}\right) - E\left(\frac{1}{1-p}\right)\left(\frac{\mathrm{h}_o B^2}{2y_1} - \frac{B^2 b}{y_1}\right) \quad (7)$$

Optimal lot size is $y_1^* = \sqrt{W^2 + \frac{u_i + A + \frac{B^2 b}{D} + h_o\left(\frac{pW^2}{x} + \frac{B^2}{2D} - \frac{3WB(1-p)}{2D}\right)}{h_r\left(\frac{(1-p^2)}{2D} + \frac{p}{x}\right)}}$ (8)

III. Solution Procedure

The objective of this model is to optimize order quantity (y_1^*) for both the warehouses and optimal shortages B^*, from these we get expected total profit $ETPU(B, y_1)$, therefore the necessary condition for expected profit to be optimal are $\frac{\partial ETPU(B,y_1)}{\partial y_1}$ and $\frac{\partial ETPU(B,y_1)}{\partial B}$which can be evaluated by the eq. (7) and equating the result to zero

$$\frac{\partial ETPU(B, y_1)}{\partial B} = -h_o\left(\frac{-W}{y_1} + \frac{1}{2}\frac{y_1 - W}{y_1}\right) - E\left(\frac{1}{1-p}\right)\left(\frac{h_o B}{y_1} - \frac{2Bb}{Dy_1}\right) = 0 \quad (9)$$

$$\frac{\partial ETPU(B, y_1)}{\partial y_1} = D\left(\frac{2h_r(y_1 - W)}{xy_1} - \frac{h_r(y_1 - W)^2}{xy_1^2} - \frac{h_W W^2}{xy_1^2}\right)$$
$$+E\left(\frac{1}{1-p}\right)D\left(\frac{A}{y_1^2} - \frac{2h_r(y_1-W)}{xy_1} + \frac{h_r(y_1-W)^2}{xy_1^2} + \frac{h_o W^2}{xy_1^2}\right) - \frac{h_r(y_1-W)(1-E(p))}{y_1} + \frac{h_r(y_1-W)^2(1-E(p))}{2y_1^2}$$
$$-h_o\left(\frac{1}{2}\frac{W(1-E(p))}{y_1} - \frac{1}{2}\frac{W(y_1-W)(1-E(p))}{y_1^2} - \frac{1}{2}\frac{W^2(1-E(p))}{y_1^2} + \frac{WB}{y_1^2} + \frac{B}{2y_1} - \frac{B(y_1-W)}{2y_1^2}\right)$$
$$-E\left(\frac{1}{1-p}\right)\left(-\frac{h_o B^2}{2y_1^2} + \frac{B^2 b}{y_1^2}\right)$$
$$= 0 \quad (10)$$

Eqs. (9) and (10) can be solved simultaneously for B* and y_1* using Maple. Here, the sufficient condition for expected total profit to be concave. First taking the second derivative.

$$\frac{\partial^2 ETPU(B,y_1)}{\partial B^2} = -E\left(\frac{1}{1-p}\right)\left(\frac{h_0}{y_1} - \frac{2b}{y_1}\right) \quad (11)$$

$$\frac{\partial^2 ETPU(B, y_1)}{\partial y_1^2} = D\left(\frac{2h_r}{xy_1} - \frac{4h_r(y_1 - W)}{xy_1^2} + \frac{2h_r(y_1 - W)^2}{xy_1^3} + \frac{2h_o W^2}{xy_1^3}\right)$$
$$+E\left(\frac{1}{1-p}\right)D\left(-\frac{2A}{y_1^3} - \frac{2h_r}{xy_1} + \frac{4h_r(y_1 - W)}{xy_1^2} - \frac{2h_r(y_1 - W)^2}{xy_1^3} - \frac{2h_o W^2}{xy_1^3}\right)$$
$$-\frac{h_r(1-E(p))}{y_1} + \frac{2h_r(y_1 - W)(1-E(p))}{y_1^2} - \frac{h_r(y_1 - W)^2(1-E(p))}{y_1^3}$$
$$-h_o\left(\frac{-W(1-E(p))}{y_1^2} + \frac{W(y_1-W)(1-E(p))}{y_1^3} + \frac{W^2(1-E(p))}{y_1^3} - \frac{2WB}{y_1^3} - \frac{B}{y_1^2} + \frac{B(y_1-W)}{y_1^3}\right) - E\left(\frac{1}{1-p}\right)\left(\frac{h_o B^2}{y_1^3} - \frac{2B^2 b}{y_1^3}\right) \quad (12)$$

$$\frac{\partial^2 ETPU(B, y_1)}{\partial y_1 \partial B} = -h_o\left(\frac{W}{y_1^2} + \frac{1}{2y_1} - \frac{1}{2}\frac{y_1 - W}{y_1^2}\right) - E\left(\frac{1}{1-p}\right)\left(\frac{-h_o B}{y_1^2} + \frac{2b\,B}{y_1^2}\right) \quad (13)$$

$$\left(\frac{\partial^2 ETPU(B, y_1)}{\partial B^2}\right) \times \left(\frac{\partial^2 ETPU(B, y_1)}{\partial y_1^2}\right) - \left(\frac{\partial^2 ETPU(B, y_1)}{\partial y_1 \partial B}\right)^2 \geq 0 \quad (14)$$

And $\left(\frac{\partial^2 ETPU(B,y_1)}{\partial B^2}\right) \leq 0, \left(\frac{\partial^2 ETPU(B,y_1)}{\partial y_1^2}\right) \leq 0$ (15)

By necessary and sufficient condition, the above equations show that the function *ETPU* (B, y_1) is strictly concave with a negative-definite Hessian matrix with optimal (B^*, y_1^*) values. With the optimal B* and y_1* values known, the net profit can be derived from Eq. (7).

IV. Results & Discussion

I. Numerical Example

In order to assess the mention inventory system, a few parameters are required, and these parameters are acquired from (Aastha. et al., 2020)

$D = 50{,}000\ unit\ /\ year;\ W = 800\ units/cycle;\ A = \$\,100\ /cycle;$
$j = \$\,20\ /unit;\ h_r = \$\,7\ /unit/year;\ h_o = \$\,5/unit/year;\ x = 1\ unit/min;\ d = 0.5/unit;\ u_i = 5/unit;\ s = \$\,50/unit;\ v = \$\,20\ /unit;\ b = 10/units$

In this example, the inventory model works on a 8 $hours/day$ for whole year, Hence, the screening rate per year,$x = 1 * 60 * 8 * 365 = 175200\ units/year$.

The proportion of imperfect random variable, p, follows uniformly distribution with p.d.f. as

$$f(p) = \begin{cases} 25, & 0 \le \rho \le 0.04 \\ 0, & otherwise \end{cases}$$

From this $E(p) = 0.02$ and$E\left(\frac{1}{1-p}\right) = 1.02055$.

The optimal solution is obtained by solving equations (15) and (16).

Table 2: *Optimum values of the model*

$y_1{}^*$	B^*	$ETPU(B^*, y_1{}^*)\$$	$t_2{}^*$(Yr.)	$t_3{}^*$(Yr.)	T^*(Yr.)
1,373.892	167.651	12130.342	0.012	0.001	0.027

II. Sensitivity Analysis

A sensitivity analysis is conducted for the aforementioned case. The effects of fraction defective items ([p]), ordering cost, selling price of defective items and discount on the expected total profit ($ETPU(B^*, y_1{}^*)$) are shown in Figure 1, Figure 2, Figure 3 and Figure 4, respectively.

Table 3: *Sensitivity analysis with different parameters of the inventory system*

Parameter	Value	T (years)	t_1 (years)	t_2 (years)	y_1 (units/cycle)	B (units/cycle)	TP ($)
y_1	0.02	0.027	0.0112	0.012	1,373.80	167.7	12130.4
	0.03	0.026	0.0113	0.0122	1,383.60	164.5	12101.3
	0.04	0.0267	0.0114	0.0121	1,393.50	161.1	12071.4
	0.05	0.0267	0.0115	0.0120	1,403.40	157.8	12040.6
A	100	0.027	0.0112	0.012	1,373.80	167.7	12130.5
	125	0.026	0.0113	0.0122	1,383.60	164.5	12121.6
	150	0.0267	0.0114	0.0121	1,393.50	161.1	12113.5
	175	0.0267	0.0115	0.0120	1,403.40	157.70	12105.6
v	20	0.027	0.0112	0.012	1,373.89	167.7	12130.5
	25	0.026	0.0113	0.0122	1,383.67	164.4	12181.5
	30	0.0267	0.0114	0.0121	1,393.52	161.1	12232.6
	35	0.0267	0.0115	0.0120	1,403.46	157.70	12283.4
u	6	0.027	0.0112	0.012	1,373.89	167.7	11624.5
	5	0.026	0.0113	0.0122	1,383.67	164.4	12130.2
	4	0.0267	0.0114	0.0121	1,393.52	161.1	12640.6
	3	0.0267	0.0115	0.0120	1,403.46	157.70	13150.8

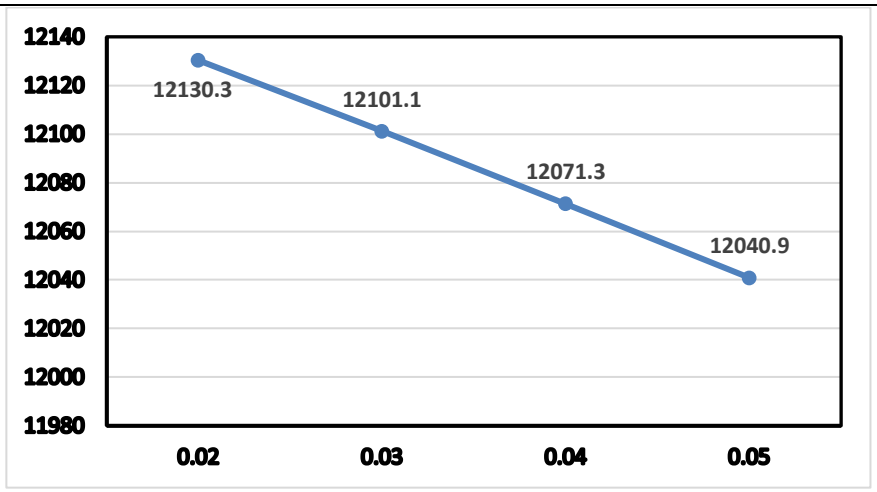


Figure 2: *Plotting of total profit with defective items*

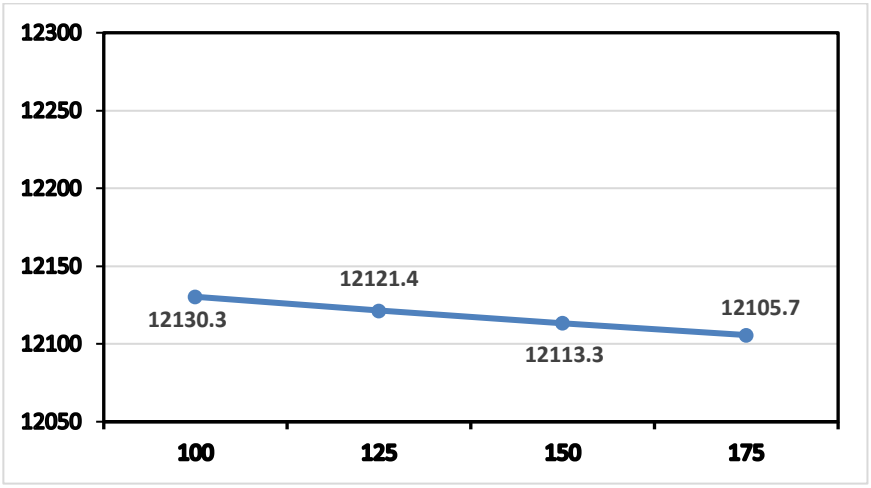


Figure 3: *Plotting of total profit with the ordering cost*

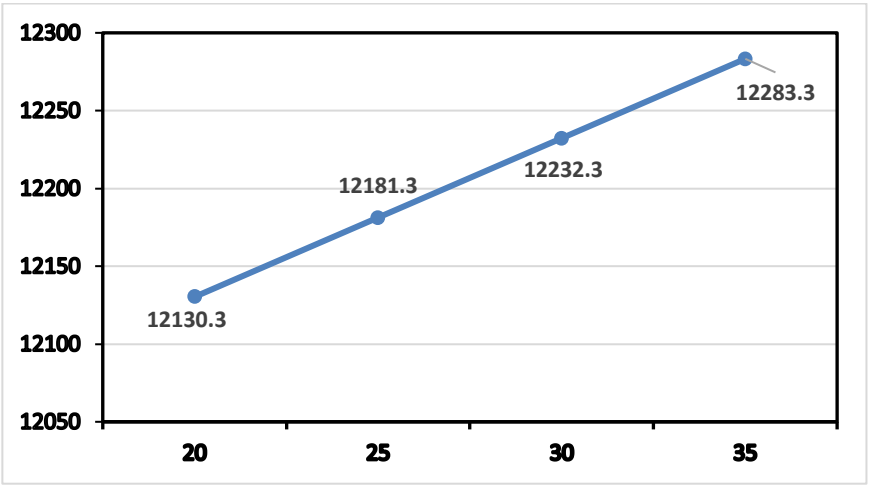


Figure 4: *Plotting of total profit with its selling price of defective items*

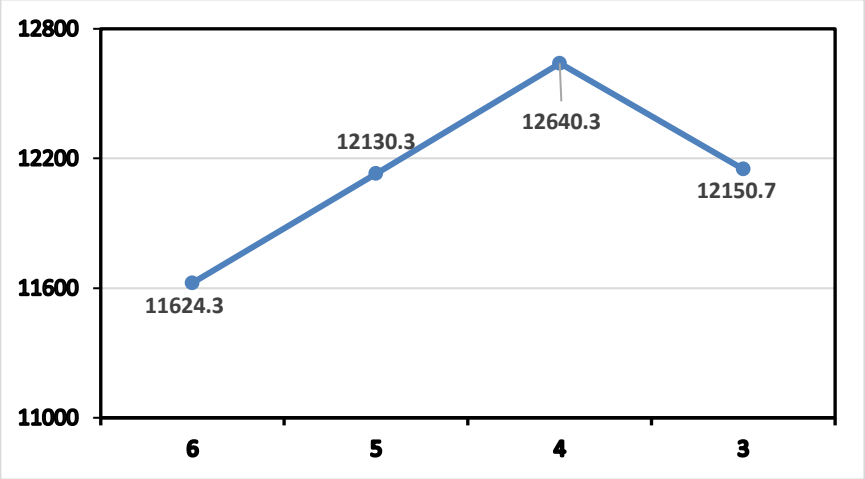


Figure 5: *Plotting of total profit with the offered discounts*

The graph above (see Figure 2) shows that as the percentage of defective goods increases, the expected total profit decreases significantly. (See Figure 3) As the acquisition cost of goods increases, the expected total profit also increases. In this scenario, buyers need to be careful when ordering and sellers need to make their systems more efficient to enhance the quality of goods produced. (See Figure 4) As the selling price of defective goods increases, the expected total profit increases significantly. (See Figure 5) As the discount offered decreases, the total expected profit increases. So, in this scenario, if the buyer receives a defective product, the buyer should pay attention to the sale price of the defective product as well as the discount offered.

V Conclusion

In this article, an EOQ model for defective items with stock outs in a multiple warehouse environment and determine the total expected profit if a discount is offered. It is presented as if the buyer has stock of defective items and can make a profit even with the defective items. To make more profit, a volume discount can also be offered. In this model, businesspersons need to look at the cost of purchasing the goods and the discount offered as these directly affect the overall expected profit. This paper is very useful for retail and manufacturing industries. This is also useful in the business world. The model can be extended in several ways. It can be extended to specify demand types through credit policies under CO_2 emission limit.

References

[1] S. Papachristos*, K. Skouri (2003). An inventory model with deteriorating items, quantity discount, pricing and time-dependent partial backlogging *Int. J. Production Economics 83 (2003) 247–256.*

[2] Seto, B. K., Sarkar, B., & Goswami, A. (2012). A two-warehouse inventory model with increasing demand and time varying deterioration. *Scientia Iranica, 19(6), 1969–1977*

[3] Jaggi, C., Pareek, S., Khanna, A., & Sharma, R. (2015). Two-warehouse inventory model for deteriorating items with price-sensitive demand and partially backlogged shortages under inflationary conditions. *International Journal of Industrial Engineering Computations, 6(1), 59–80.*

[4] Jaggi, C. K., Cárdenas-Barrón, L. E., Tiwari, S., & Shafi, A. (2017). Two-warehouse inventory model for deteriorating items with imperfect quality under the conditions of permissible delay in payments. *Scientia Iranica, 24(1), 390–412.* https://doi.org/10.24200/sci.2017.4042

[5] Mandal, P., & Giri, B. C. (2017). A two-warehouse integrated inventory model with imperfect pro duction process under stock-dependent demand and quantity discount offer.

International Journal of System Science Operations & Logistics

[6] Shah, N. H., & Naik, M. K. (2018). EOQ model for deteriorating item under full advance payment availing of discount when demand is price-sensitive. *International Journal of Supply Chain and Operations Resilience, 3(2), 163–197.*

[7] Chung, K.-J., Liao, J.-J., Ting, P.-S., Lin, S.-D., & Srivastava, H. M. (2018). A unified presentation of inventory models under quantity discounts, trade credits and cash discounts in the supply chain management. *Revista de La Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Mathematics, 112(2), 509–538. https://doi.org/10.1007/s13398-017-*

[8] Dey, B. K., Sarkar, B., Sarkar, M., & Pareek, S. (2019). An integrated inventory model involving discrete setup cost reduction, variable safety factor, selling price dependent demand, and investment. *RAIRO - Operations Research, 53(1), 39–57. https://doi.org/10.1051/ro/2018009*

[9] Dhaka, V., Pareek, S., & Mittal, M. (2019). Stock-Dependent Inventory Model for Imperfect Items Under Permissible Delay in Payments. *Optimization and Inventory Management (pp. 181–194). Springer. https://doi.org/10.1007/978-981-13-9698-4_10*

[10] Agarwal, G. L., & Mandeep, M. (2019). Inventory Classification Using Multi-Level Association Rule Mining. *International Journal of Decision Support System Technology, 11, 1–12. https://doi.org/10.4018/IJDSST.2019040101*

[11] Mashud, A. H. M., Wee, H.-M., Sarkar, B., & Li, Y.-H. C. (2020). A sustainable inventory system with the advanced payment policy and trade-credit strategy for a two-warehouse inventory system. *Kybernetes.*

[12] Mittal, M., Pareek, S. & Aastha, (2020). Effect of Human Errors on an Inventory Model Under Two Warehouse Environments. *Recent Advances in Computer Science and Communications, 13, 1–10.*

[13] Priyanka, & Pareek, S. (2020). Two Storage Inventory Model for Non-Instantaneous Deteriorating Item with Stochastic Demand Under Credit Financing Policy. *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 978–983. https://doi.org/10.1109/ICRITO48877.2020.9197872*

[14] Gilotra, M., Pareek, S., Mandeep, M., & Dhaka, V. (2020). Effect of Carbon Emission and Human Errors on a Two-Echelon Supply Chain under Permissible Delay in Payments. *International Journal of Mathematical, Engineering and Management Sciences, 5,* 225–236. https://doi.org/10.33889/IJMEMS.2020.5.2.018

[15] Aastha, Pareek, S., & Mittal, M. (2020). Non Instantaneous Deteriorating Inventory Model under Credit Financing When Demand Depends on Promotion and Selling Price. *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 973–977. https://doi.org/10.1109/ICRITO48877.2020.9197990*

[16] Dey, B. K., Pareek, S., Tayyab, M., & Sarkar, B. (2021). Autonomation policy to control work-in-process inventory in a smart production system. *International Journal of Production Research, 59(4), 1258–1280. https://doi.org/10.1080/00207543.2020.1722325*

[17] Jayaswal, M. K., Mittal, M., and Sangal, I. (2021). Effect of credit financing on the learning model of perishable items in the preserving environment. In Decision Making in Inventory Management, *pages 49–60. Springer*

[18] Mishra, U., Mashud, A. H. M., Tseng, M.-L., and Wu, J.-Z. (2021). Optimizing a sustainable supply chain inventory model for controllable deterioration and emission rates in greenhouse farm.*Mathematics,9(5):495*

[19] Shah, N. H., Rabari, K., and Patel, E. (2022). Economic production model with reliability and inflation for deteriorating items under credit financing when demand depends on stock displayed. Revista Investigacion *Operacional., 43(2):203–213.*

[20] Alamri, O. A., Jayaswal, M. K., Khan, F. A., and Mittal, M. (2022). An eoq model with carbon emissions and inflation for deteriorating imperfect quality items under learning effect. Sustainability, 14(3):1365.

[21] Kumar, B. A., Paikray, S. K., and Padhy, B. (2022). Retailer's optimal ordering policy for deteriorating inventory having positive lead time under pre-payment interim and post- payment strategy. *International Journal of Applied and Computational Mathematics, 8(4):1–33.*

[22] Ruidas, S.; Seikh, M.R.; Nayak, P.K. (2022). A production inventory model with interval-valued carbon emission parameters under price-sensitive demand. *Computer. Ind. Eng., 154, 107154*

[23] Mittal, M., Khanna, A., & Jaggi, C. K. (2023). Retailer's ordering policy for deteriorating imperfect quality items when demand and price are time-dependent under inflationary conditions and permissible delay in payments. *International Journal of Procurement Management, 10(4), 461–494.*

[24] Masae, M., Glock, C.H., Grosse, E.H. (2023). Order picker routing in warehouses: A systematic literature review. *International Journal of Production Economics, 224: 107564. https://doi.org/10.1016/j.ijpe.2019.107564*

[25] Ibrahim Hassan (2024). The role of warehouse Layout and Operations in Warehouse Efficiency: A literature Review. *Journal Europeen des Systemes Automatises, 56(1), 61-68.*

RT&A, No 4(80)
Volume 19, December, 2024

SELECTION OF BEST ENERGY STORAGE TECHNOLOGY USING ELECTRE III-BWM METHOD UNDER LINGUISTICS NEUTROSOPHIC FUZZY APPROACH

Sasirekha D* and Senthilkumar P[1]

•

PG and Research Department of Mathematics, Government Arts and Science College,
Kangeyam, Tamil Nadu.
rajendranpsk@gmail.com
Correspondance email: sasirekhasaba@gmail.com,

Abstract

Renewable energy provides more environmentally friendly sources of energy, which reduces the demand for fossil fuels and is therefore necessary to reach zero emissions of carbon. But the need for systems that are capable of capturing and storing this energy is expanding as the world gets a growing amount of electricity from these forms of renewable energy. In present-day society, renewable energy storage is widely used, and governments are concentrating on developing suitable storage technologies together with a plan for upcoming energy storage reduction. Energy storage technologies have been proposed as potential solutions for this issue due to their ability to store energy and lower energy consumption. Aspects of technology, economy, society, and environment are the four main criteria used in this study to examine different energy storage techniques. The most effective strategy was identified in this paper. In this study, we use the ELECTRE-III approach to suggest the optimal storage technology under the linguistic neutrosophic fuzzy set. Finally, a numerical example of this area of study is provided. A comparison and sensitivity analysis are shown for the effectiveness of the proposed method.

Keywords: Neutrosophic Fuzzy Set, Linguistic Neutrosophic Fuzzy Set, ELECTRE-III, BWM, Renewable energy storage technology.

1. Introduction

Energy, which is frequently defined as the ability to complete a task, is a crucial concept for both the long-term development of countries and the ongoing growth of the human race. Energy is so vital to the global economic and social progress of the world that an extensive amount of study is focused on guaranteeing its afford-ability, availability, and stability. Renewable energy (RE) and non-renewable energy (NRE) alternatives or sources are the two categories into which types of energy can be separated. Natural sources of energy include wind, sun, hydro, geothermal, biomass, and waves, while NRE sources include coal, oil, nuclear energy, and natural gas. As they emit fewer greenhouse gases (GHGs) than fossil fuels, renewable energy sources (RES) are better for the environment. However, as a result of increasing population and industrialization, there is an increase in worldwide energy demand. To meet this growing need, RES are currently thought to be viable solutions. Nevertheless, these energy sources exhibit unpredictable and intermittent properties. Energy storage technologies (ESTs) were subsequently created to ensure the availability of energy by allowing the storage of extra energy and its utilization when required.

The five kinds of energy technology preservation, often known as alternatives, are electrical, mechanical, chemical-based, electrochemical, and thermo [1, 15].

In the twenty-first century, conventional carbon-based fuels, including coal, oil, and natural gas, are the most widely used sources of energy. Conventional agriculture's fossil fuel consumption has fueled economic growth worldwide, but it has also resulted in significant environmental issues. The primary source of the greenhouse effect is the massive amount of CO_2 produced from the burning of fossil fuels. As a result, there has been a considerable advancement in the development of new and renewable energy sources with plentiful resources and minimal adverse effects on the environment. The creation of RE is essential to lowering CO_2 emissions and resolving environmental issues. It is also a widely accepted solution to resource exhaustion and air pollution. Meanwhile, there have been obstacles and restrictions in the way of the growth of clean energy. The production of energy based on sustainable sources is reliant on the erratic and transient demand for natural assets. These features impact the security and reliability of the power grid, in addition to making it challenging to change and regulate the production of energy [2, 3]. Thus, energy storage is becoming a crucial factor in the advancement of renewable energy sources.

An important idea that makes it possible to assess options in light of numerous competing criteria for a decision-making (DM) challenge is multi-criteria decision-making (MCDM). When the MCDM process is complete, the goal is to have chosen the best option from a variety of options. Despite conventional and fuzzy MCDM techniques being used to achieve this goal, fuzzy models offer a way to deal with the ambiguity inherent in human perspectives, leading to more realistic and feasible outcomes [4]. It is evident that choosing the best EST is an MCDM problem since evaluating various ESTs requires the realization of a thorough assessment in terms of these criteria, and there are numerous competing criteria, including technical, economic, environmental, and social ones. As a result, it makes sense to combine MCDM techniques with Zadeh's fuzzy set theory (FST), which allows for the modeling of human judgment uncertainties during MCDM problem analysis. Furthermore, some recently created extensions of conventional fuzzy sets, like intuitionistic, Pythagorean fuzzy sets, hesitant, and type-2, are useful in addressing vagueness in the MCDM process [5, 6]. The energy storage problem is being solved in this work by employing linguistic neutrosophic fuzzy numbers.

The truth, indeterminacy, and falsity linguistic term values are expressed independently by three specific linguistic variables, l_T, l_I, and l_F, using a linguistic neutrosophic fuzzy number. Researchers have recently focused more on the fascinating study areas of making decisions in linguistic circumstances. Zadeh [24] highlighted how fuzzy logic makes use of language characteristics. To overcome the difficulties in using linguistic information to make judgments, Herrera and Herrera-Viedma [25] and Herrera et al. [26] created linguistic decision studies. According to Chen et al. [27], the linguistic intuitionistic fuzzy number (LIFN) is defined as $s = (l_a, l_b)$, where l_a and l_b represent the linguistics characteristics of membership, and non-membership, respectively. The MAGDM process is then developed using LIFNs. For MADM, Liu and Wang [28] created improved LIFN aggregation methods. The LIFN elucidates the linguistic details of both truth and falsehood degrees and is evidently composed of two linguistic variables, l_a and l_b.

However, LIFNs are not able to describe inconsistent and unclear language data. An integral part of an SVNS, a single-valued neutrosophic number (SVNN) [29] can convey degrees of truth, indeterminacy, and untruth; it also reveals details that are insufficient, inconsistent, and not resolved in SVNN instead of linguistic data. As a result, unlike linguistic variables, it is unable to describe linguistic formation in a linguistic issue of decision-making. Ye [30] introduced the SVNLN, which consists of an SVNN and a syntactic variable. The SVNN indicates the consistency of the given linguistic variable, while the linguistic factor represents the decision-maker's assessment of an object under examination. Within the context of the previously indicated concept, it is necessary to propose an interpretation of a linguistic neutrosophic number (LNN), which is expressed as an LNN $e =< l_4, l_2, l_3 >$. It is clear that LIFN and SVNLN cannot communicate such verbal assessing values; on the other hand, LNN, which combines SVNN and

LIFN into LNN, might just be the minimal language environment. As a result, LNN needs to be used to resolve linguistic DM problems with erratic and indeterminate linguistic information as well as to express incomplete and unresolved linguistic information that corresponds to human fuzzy reasoning regarding complex problems, particularly certain qualitative attribute assessments [31, 32, 33]. As a result, the best appropriate option under a linguistic neutrosophic fuzzy environment has been identified, and ESTs have been assessed employing a fuzzy MCDM approach using the LNFNs technique in this study.

2. Literature Review

Many researchers have identified suitable energy storage methods, which we address in this section on energy storage systems. Examine the energy storage technologies from the standpoint of energy security, as presented by Azzuni and Breyer [7]. Gao and Lu [8] are examining the latest developments in this developing field, particularly the novel ideas, methods, and uses of machine learning technologies for widely utilized energy storage systems and gadgets. The dual-hesitant Pythagorean fuzzy linguistic term sets are defined by Liu and Du [9], who also suggest an MCDM framework for choosing renewable energy storage technologies. Kumar et al. [10] provide a review of different MCDM approaches, advancements made when comparing the techniques to renewable energy usage, and potential future developments. Chen et al. [11] explored Prospect Theory and PROMETHEE in the selection of renewable energy sources. Pamucar et al. [12] proposed a Dombi weighted geometric averaging operator and MAIRCA model under hybrid trapezoidal neutrosophic fuzzy numbers, and then evaluated and rated the available energy storage technology solutions in accordance with the selected specifications. Barin et al. [13] found the most appropriate energy storage system consistent with a power quality priority. Rahman [14] reviews the techno-economic and ecological evaluations of thermal energy, chemical in nature, electro-chemical, and mechanical procedures in order to provide a review of current advancements and create a pertinent database for expenses and emissions. The various energy storage systems, with Aneke and Wang's [16] investigation focusing primarily on the storage system. Ren [17] was working on creating a multi-attribute decision-analysis framework to prioritize energy storage solutions based on sustainability. Large-scale storage of energy systems of life cycle consumption of energy and the release of greenhouse gases were suggested by Denholm and Kulcinski [18].

Further MCDM methods have been used for evaluating the EST problem. Table 1 presents the MCDM methods used and the suggested option.

Table 1: *MCDM methods used various researchers*

Author's	MCDM methodology	Problem
Sengul et al. (2015) [19]	TOPSIS	Renewable energy
Ozkan et al. (2015) [20]	AHP-TOPSIS	Selection of Energy Storage Alternatives
Zhang et al. (2019) [21]	fuzzy MULTIMOORA	Assessment of the energy storage technologies
Ren (2018) [22]	Fuzzy IAHP	Development of renewable energy
Colak and Kaya (2020) [23]	CRITIC-MOORA, TOPSIS, COPRAS	Hydropower

Using various MCDM models, the evaluated research primarily sought to identify the best choices for renewable energy storage technologies, utilizing the constantly evolving MCDM techniques that are essential to the field of sustainable energy management. Because different approaches offer different solutions, it is necessary to perform a thorough, country-specific analysis in order to successfully handle the issue. There is yet unrealized potential for combining ELECTRE III and the Best-Worst Method (BWM) to enhance EST selection. The ability of Linguistic Neutrosophic Fuzzy Sets (LNFSs) to represent not only truth and falsehood but also neutrality provides a strong model for handling difficult decision-making situations; however, more research is needed to determine how well LNFSs integrate with the suggested methodology. The amalgamation of these elements has the potential to greatly expand its relevance in various domains and enhance the effectiveness of decision-making. As such, this study aims to establish a general framework for determining which energy storage technology is most suitable, taking into account social, economic, and technological factors.

3. Preliminaries

Definition 3.1. A Neutrosophic fuzzy set (NFS) ω on R:

$$\omega = \{(o, p_\omega(o), q_\omega(o), r_\omega(o)) : o \in R\} \tag{1}$$

where $p_\omega(o), q_\omega(o), r_\omega(o) \in [0,1], 0 \leq p_\omega(o) + q_\omega(o) + r_\omega(o) \leq 3$ for all $o \in R$, $p_\omega(o)$, $q_\omega(o)$, $r_\omega(o)$ are degrees of membership, indeterminacy and non-membership, respectively.

Definition 3.2. Let $N = n_0, n_1, ..., n_t$ be a LTS with odd cardinality $f+1$. If $e =< n_a, n_b, n_c >$ is defined for $n_a, n_b, n_c \in N$ and $a, b, c \in [0, f]$, where n_a, n_b, and n_c represent the trueness, indeterminacy, and falseness level by linguistic terms and e is called an LNN.

Definition 3.3. Let $u =< n_a, n_b, n_c >$, $u_1 =< n_{a_1}, n_{b_1}, n_{c_1} >$ and $u_2 =< n_{a_2}, n_{b_2}, n_{c_2} >$ be three LNNs in N and $c > 0$, then

- $u_1 \oplus u_2 =< n_{a_1}, n_{b_1}, n_{c_1} > \oplus < n_{a_2}, n_{b_2}, n_{c_2} >=< n_{a_1+b_2-\frac{a_1 a_2}{f}}, n_{\frac{b_1 b_2}{f}}, n_{\frac{c_1 c_2}{f}} >$
- $u_1 \otimes u_2 =< n_{a_1}, n_{b_1}, n_{c_1} > \otimes < n_{a_2}, n_{b_2}, n_{c_2} >=< n_{\frac{a_1 a_2}{f}}, n_{a_1+a_2-\frac{a_1 a_2}{f}}, n_{\frac{c_1 c_2}{f}} >$
- $\phi u = \phi < n_a, n_b, n_c >=< n_{f-f(1-\frac{a}{f})^\phi}, n_{f(\frac{b}{f})^\phi}, n_{f(\frac{c}{f})^\phi} >$
- $u^\phi =< n_a, n_b, n_c >^\phi =< n_{f(\frac{a}{f})^\phi}, n_{f(\frac{b}{f})^\phi}, n_{f-f(1-\frac{c}{f})^\phi} >$

Definition 3.4. Let $e =< n_a, n_b, n_c >$ be an LNN in N. Then the score and accuracy functions of e is given below:

$$S(e) = \frac{(2f + a - b - c)}{3f} \text{ for } S(e) \in [0,1] \tag{2}$$

$$A(e) = \frac{(a-c)}{f} \text{ for } F(e) \in [-1,1] \tag{3}$$

Definition 3.5. Let $u_1 =< s_{a_1}, s_{b_1}, s_{c_1} >$ and $u_2 =< s_{a_2}, s_{b_2}, s_{c_2} >$ be any two LNNs, and let f^* be a linguistic scale function $t \geq 0$. The generalized distance measure of a_i & a_j is

$$d(u_1, u_2) = \left(\frac{1}{3}(|f^*(s_{a_1} - f^*(s_{a_2}))|^t + |f^*(s_{b_1} - f^*(s_{b_2}))|^t + |f^*(s_{c_1} - f^*(s_{c_2}))|^t)^t\right) \tag{4}$$

where $t = 1$ or $t = 2$, and the above equation is reduced to the Hamming distance or the Euclidean distance, respectively.

4. Proposed Method

Let $H = (h_1, h_2, ..., h_t)$ be a collection of criteria for an MCDM issue, and let $G = (g_1, g_2, ..., g_s)$ be a set of alternatives. The effectiveness or assessment of the alternative $g \in G$ for the criterion h_j is represented by $h_j(g_j)$. The degree to which an option satisfies the mentioned criterion depends on if the goal is to optimize or reduce $h_j(g_j)$. The greater or lesser it is, the better. Consequently, the vector $h(g) = (h_1(g), h_2(g), ..., h_t(g))$ will be used to describe the multifaceted assessment of the alternative $g \in G$. The ELECTRE-III model's assessment processes include threshold function establishment, concordance and discordance index disclosure, credibility degree determination, and option ranking. Let the preference and indifference thresholds be represented, respectively, by b(h) and a(h).

Here, we expand the ELECTRA III method with linguistic neutrosophic fuzzy number. Consider a s alternatives $\{G_1, G_2, ..., G_s\}$, t criteria $\{H_1, H_2, ..., H_t\}$ and α decision maker $\{e_1, e_2, ..., e_\alpha\}$ with weighting vector be $\{w_1, w_2, ..., w_\alpha\}$, then, the procedure of decision-making in the LNFN-ELECTRE III model is described in the following steps:.

Step 1: Determine the concordance matrix

The concordance matrix $C(G_p, G_q)$ is obtained for each pair of alternatives and the CCM is calculated by equation (5).

$$CCM(G_p, G_q) = \frac{\sum_{j=1}^{v} w_j \, CCM_j(G_p, G_q)}{\sum_{j=1}^{v} w_j} \tag{5}$$

where $C_i(p, q)$ is the out ranking degree of the alternative p and the alternative q under the criteria i, and $C_i(p, q)$

$$CCM_i(G_p, G_q) = \begin{cases} 0 \;\; if \quad (G_q) - (G_p) > (a_p) \\ 1 \;\; if \quad (G_q) - (G_p) \leq (b_p) \\ \frac{(a_i)+(G_p)-(G_q)}{(a_p)-(b_p)} \qquad otherwise \end{cases}$$

Step 2: Compute the discordance matrix

Determine the discordance matrix $DCM(G_p, G_q)$. The DCM is described in given equation (6).

$$DCM_i(G_p, G_q) = \begin{cases} 0 \;\; if \quad (G_q) - (G_p) \leq (a_p) \\ 1 \;\; if \quad (G_q) - (G_p) > (b_p) \\ \frac{(G_q)-(G_p)-(a_i)}{(b_p)-(a_p)} \qquad otherwise \end{cases} \tag{6}$$

where, $0 \leq DCM_j(G_p, G_q) \leq 1$.

Step 3: Obtain the outranking degree $O(G_p, G_q)$ is in (7),

$$O(G_p, G_q) = \begin{cases} CCM(G_p, G_q) \qquad if \quad DCM(G_p, G_q) \leq CCM(G_p, G_q) \\ CCM(G_p, G_q) \times \prod_{j \in J} \frac{1-DCM_j(G_p, G_q)}{1-CCM_j(G_p, G_q)} \qquad otherwise \end{cases} \tag{7}$$

Step 4: Lastly, outlining the options according to the values of net credibility, discordance credibility, and concordance credibility.

- The benefit of concordance credibility is explained by,

$$\Psi^{+}(G_i) = \sum_{\forall s \in (i=1,2,\ldots,s)} (G_p, G_q) \tag{8}$$

  The outranking character of G_i determines the concordance credibility.

- The following describes the discordance credibility value:

$$\Psi^{-}(G_i) = \sum_{\forall s \in (i=1,2,\ldots,s)} (G_p, G_q) \tag{9}$$

  The outranking character of G_i is portrayed by the discordance credibility.

- The description of the net credibility value is provided by,

$$\Psi(G_i) = \Psi^{+}(G_i) - \Psi^{-}(G_i), \quad \forall \;\; G_i \tag{10}$$

A higher worth indicates a significant engaging quality of G_i. The net credibility value represents the worth capability. On the basis of net credibility, both G_is can be fully placed.

5. The fuzzy BWM method

The $(h_1, h_2, ..., h_n)$ are the criterion to select the best decision. The criteria are H_1-technological, H_2-economic, H_3-environmental aspects, and H_4-social.

Step 1: Obtain the greatest criteria (e.g., beneficial criteria) and the lowest criteria (e.g., non-beneficial).

Step 2: Assign scores ranging from 1 to 9 to reflect their preference for the best overall criterion out of all the others. The vector result for Greatest-to-Others is:

$$H_Y = (h_{Y_1}, h_{Y_2}, ..., h_{Y_t})$$

where h_{Y_j} indicates a preference for criterion j (i.e., $h_{YY} = 1$) above the optimal criterion Y. In our example, a vector shows that H_3-environmental features are preferred over the other factors.

Step 3: Sort all of the criteria by preference over the lowest criterion, using a value between 1 and 9. The next-to-lowest vector that is produced is:

$$H_L = (b_{1D}, b_{2D}, ..., b_{sD})$$

where a preference for criterion j over the worst criterion D is indicated by the symbol h_{jD}. The value of $b_{DD} = 1$ is evident. Here, the vector represents the preferences across all criteria over the time-H_2 criteria.

Step 4: Calculate the optimal weights $(w_1^*, w_2^*, ..., w_n^*)$.
The ideal weight for the criteria is one in which $\frac{w_Y}{w_j} = h_{Yj}$ and $\frac{w_j}{w_D} = h_{jD}$ for every pair of $\frac{w_Y}{w_j}$ and $\frac{w_j}{w_D}$. The solution should minimize $|\frac{w_Y}{w_j} - h_{Yj}|$ and $|\frac{w_j}{w_D} - h_{jD}|$ for every j in order to meet these requirements for all j. Take into consideration the weights' non-negativity and sum criteria as follows:

$$\begin{aligned} &\min \ \max_j \{|\frac{w_Y}{w_j} - h_{Yj}| - |\frac{w_j}{w_{jD}} - h_{jD}|\} \\ &\text{subject to} \end{aligned}$$

$$\sum_j w_j = 1, \ \ w_j \geq 1 \ \text{ forall j} \tag{11}$$

Equation (11) can be applied to the subsequent issue:

$$\begin{aligned} &min \ \ \chi, \\ &subject \ to \\ &\left|\frac{w_Y}{w_j} - h_{Yj}\right| \leq \chi, \ \ \text{forall} \ \ \text{j} \\ &\left|\frac{w_j}{w_D} - h_{jD}\right| \leq \chi, \ \ \text{forall} \ \ \text{j} \\ &\sum_{j=1}^{t} w_j = 1, \forall j \\ &w_j \geq 0, \forall j \end{aligned} \tag{12}$$

Equation (12) must be solved to determine the ideal weights $(w_1^*, w_2^*, ..., w_n^*)$ and χ^*.Next, we provide a consistency ratio (CR) with χ^*. The greater the CR and the less trustworthy the contrasts become, the larger the χ^*.

Table 2: *Consistency index (CI)*

h_{YD}	1	2	3	4	5	6	7	8	9
CI	0.0	0.71	1.05	1.53	2.07	3.00	3.99	4.79	5.80

5.1. Consistency ratio

In this section, we provide a percentage of consistency for the suggested best-worst technique.

In this case, h_{Yj} indicates the preference of the best criteria over j, and h_{jD} is the preference of the worst criteria j, and h_{YD} indicates the preference of the best criteria over the worst criteria. This makes the comparison totally consistent when $h_{Yj} \times b_{jD} = b_{YD} \ \ \forall\, j$.

Furthermore, it is possible for some j to be inconsistent; in this case, we propose a CR to show the consistency of a comparison. In order to achieve this, we start by determining the minimal consistency of a comparison, which is as follows:

As previously stated, where $h_{ij} \in \{1, 2, ..., h_{HD}\}$ the largest possible value of h_{YD} is 9. Consistency decreases when $h_{Yj} \times h_{jD}$ is smaller or larger than h_{YD} or equivalently $h_{Yj} \times h_{jD} \neq 1$, and the largest inequality occurs when h_{Yj} and h_{jD} have the maximum value, which will result in χ. We also know that $(\frac{w_Y}{w_j}) \times (\frac{w_j}{w_D}) = \frac{w_Y}{w_D}$, and given the largest in equality as a results by h_{Yj} and h_{jD}, χ is a value that must be subtracted from h_{Yj} and b_{jD}, then added to b_{YD}, or:

$$(h_{Yj} - \chi) \times (h_{jD} - \chi) = (h_{YD} + \chi) \tag{13}$$

Regarding the minimum degree of consistency $h_{Yj} = h_{jD} = h_{YD}$, we have

$$\begin{aligned} &(h_{YD} - \chi) \times (h_{YD} - \chi) = (h_{YD} + \chi) \\ &\implies \chi^2 - (1 + 2h_{YD})\chi + (b_{YD}^2 - h_{YD}) = 0 \end{aligned} \tag{14}$$

By calculating several values of $h_{YD} \in \{1, 2, ..., 9\}$, we may determine the highest χ that can exist (max χ). As the consistency index in Table 2, these values are utilized.

After that, the CR is computed as follows using χ^* and the appropriate consistency index (CI):

$$\text{CR} = \frac{\chi^*}{\text{CI}} \tag{15}$$

6. Numerical Example

The use of ELECTRE-III based on linguistic neutrosophic fuzzy numbers is discussed in this section. The best energy storage technology will be selected by the proposed method. For this purpose, we choose five alternatives based on four criterion which are give below:

The alternatives of energy storage problem:

G_1-Hydrogen storage

G_2-Electrochemical storage

G_3-Mechanical storage

G_4-Electrical storage

G_5-Thermal storage

The criteria is are:

H_1-Technological

H_2-Economic

H_3-Environmental aspects

H_4-Social.

Table 3: *Vector of pairwise comparisons for the greatest criterion*

Criteria	H_1	H_2	H_3	H_4
H_3	7	3	1	5

Table 4: *Vector of pairwise comparisons for the lowest criterion*

	H_2
H_1	5
H_2	1
H_3	7
H_4	3

6.1. The fuzzy BWM method

Step 1: Construct the criteria list.

Here, we examine the criteria $(H_1, H_2, ..., H_n)$ that is used to make a decision. The criteria are H_1-technological, H_2-economic, H_3-environmental aspects, and H_4-social.

Step 2: Determine which criteria are greatest and lowest. For this particular problem, the greatest and lowest criteria are (H_3)-environmental aspects and H_2-economic.

Step 3: Table 3 provides the pairwise comparison vector for the greatest criterion values.

Step 4: Table 4 provides the pairwise comparison vector for the lowest criterion values.

Step 5: From Table 3 and Table 4 results in equation (8) for this problem, as follows:

$$\begin{aligned}
&min\ \ \chi,\\
&s.t\\
&\left|\frac{w_3}{w_1}-u_{31}\right|\leq\chi,\left|\frac{w_3}{w_2}-u_{32}\right|\leq\chi,\left|\frac{w_3}{w_4}-u_{34}\right|\leq\chi,\ \ \text{forall}\ \ \text{j}\\
&\left|\frac{w_1}{w_2}-u_{12}\right|\leq\chi,\left|\frac{w_3}{w_2}-u_{32}\right|\leq\chi,\left|\frac{w_4}{w_2}-b_{42}\right|\leq\chi,\ \ \text{forall}\ \ \text{j}\\
&w_1+w_2+w_3+w_4=1,\\
&w_1,w_2,w_3,w_4\geq 0,\forall j
\end{aligned}\tag{16}$$

Solving this equation (38), we get the optimal weights $(w_1^*, w_2^*, ..., w_n^*)$ are $w_1 = 0.0853$, $w_2 = 0.1989$, $w_3 = 0.5965$, $w_4 = 0.1193$ and $\chi^* = 0.1858$. For the consistency ratio, as $b_{GP} = b_{34} = 6$, the CI is 3.01 (Table 3), and the CR is $\frac{0.1858}{3.01} = 0.0617$, it suggests excellent consistency.

6.2. The LNFN-ELECTRE-III method

The linguistic scale for linguistic neutrosophic fuzzy number and initial DM are given in Table 5 and Table 6.

Step 1: Construct the initial matrix

Table 5 and Table 6 shows how the first matrix was constructed using the linguistic scale and how decision-makers assessed energy storage technology based on the expert's matrix's chosen criteria.

Step 2: Obtain the concordance matrix (CCM)

The thresholds for alternatives in the concordance matrix are constructed and provided in Table 7.

Using equation (5) to compare the alternatives, the CCM is now calculated; the results are presented in Table 8.

Step 3: Obtain the discordance matrix

Table 5: *The linguistic scale for LNFN*

l_0	extremely low
l_1	very low
l_2	low
l_3	slightly low
l_4	medium
l_5	slightly high
l_6	high
l_7	very high
l_8	extremely high

Table 6: *The initial decision matrix*

Alternatives	H_1	H_2	H_3	H_4
G_1	$< l_6, l_2, l_4 >$	$< l_8, l_3, l_1 >$	$< l_4, l_2, l_5 >$	$< l_1, l_4, l_3 >$
G_2	$< l_3, l_5, l_7 >$	$< l_6, l_1, l_4 >$	$< l_5, l_1, l_3 >$	$< l_4, l_3, l_3 >$
G_3	$< l_6, l_4, l_8 >$	$< l_4, l_3, l_2 >$	$< l_7, l_1, l_1 >$	$< l_3, l_5, l_7 >$
G_4	$< l_3, l_2, l_5 >$	$< l_2, l_1, l_1 >$	$< l_6, l_3, l_4 >$	$< l_1, l_3, l_6 >$
G_5	$< l_5, l_1, l_3 >$	$< l_6, l_4, l_2 >$	$< l_5, l_3, l_1 >$	$< l_7, l_3, l_2 >$

Table 7: *The initial decision matrix*

Alternatives	H_1	H_2	H_3	H_4
G_1	0.6666	0.8148	0.5555	0.4444
G_2	0.3333	0.7037	0.7037	0.5925
G_3	0.4444	0.6296	0.8518	0.3333
G_4	0.5185	0.6666	0.4444	0.3703
G_5	0.7037	0.6666	0.7037	0.7407
a	0.1	0.1	0.1	0.1
b	0.2	0.2	0.2	0.2
v	0.5	0.5	0.5	0.5

Table 8: *The concordance matrix*

	G_1	G_2	G_3	G_4	G_5
G_1	1	0.321	0.404	1	0.881
G_2	0.915	1	1	1	0.915
G_3	0.915	0.881	1	1	0.795
G_4	0.083	0.284	0.404	1	0.284
G_5	0.811	1	0.434	1	1

The discordance matrix is obtained using (6) and the results are given in Table 9.

Step 4: Next, using equation (7), the comparison between the CCM and DCM is computed,

Table 9: *The discordance matrix*

	G_1	G_2	G_3	G_4	G_5
G_1	1	0	0.191	0	0.038
G_2	0.038	1	0	0	0.048
G_3	0.006	0.024	1	0	0.099
G_4	0	0.127	0.412	1	0.48
G_5	0	0	0	0	1

and Table 10 provides the credibility matrix.

Step 5: According to the equations (8)-(10), the ranking results are calculated and given in

Table 10: *The credibility matrix*

	G_1	G_2	G_3	G_4	G_5
G_1	1	0.321	0.191	'1	0.881
G_2	0.915	1	1	1	0.915
G_3	0.915	0.881	1	1	0.795
G_4	0.083	0.284	0	1	0.284
G_5	0.811	1	0.434	1	1

Table 11.

Step 6: The ranking order using the LNFN-ELECTRE-III method is $G_2 > G_3 > G_5 > G_1 > G_4$.

Table 11: *The final ranking results*

Alternatives	results	Ranks
G_1	2.377	4
G_2	3.744	1
G_3	3.462	2
G_4	0.036	5
G_5	3.245	3

Therefore, the G_2- Electrochemical is the best technology for energy storage problem.

7. Comparison and sensitivity analysis

In this part, we compare this proposed method's effectiveness with other approaches that are currently in use, including VIKOR and ARAS in the instance of an LNFN. Sensitivity analysis was employed as well for this investigation.

Table 12: *Comparison analysis results*

ESTs	ARAS	Rank	VIKOR	Rank	Proposed method	Rank
G_1	0.7479	4	0.7299	4	2.377	4
G_2	0.8693	3	0.1588	2	3.744	1
G_3	1.0000	1	0.3378	3	3.462	2
G_4	0.6591	5	1.0000	5	0.036	5
G_5	0.9581	2	0	1	3.245	3

7.1. Comparative analysis

To illustrate the efficacy and performance of the suggested model, it is compared in this section to various MCDM techniques found in the literature. The ARAS model and the VIKOR model are two methods that have already been used to evaluate suggested methodologies. Certain MCDM approaches make use of the suggested criterion weights. The comparison of ranking-order findings is presented in Table 12. The results generated by the suggested ranking deviate considerably from the current VIKOR and ARAS approaches. Consequently, in comparison to previous MCDM models, the suggested method yields more trustworthy results.

Table 13: *Weights in sensitivity analysis*

ESTs	Case 1	Case 2	Case 3
H_1	0.0853	0.5965	0.1193
H_2	0.1989	0.0853	0.5965
H_3	0.5965	0.1193	0.1989
H_4	0.1193	0.1989	0.0853

Table 14: *Sensitivity analysis results*

ESTs	Case 1	Rank	Case 2	Rank	Case 3	Rank
G_1	2.377	4	3.278	2	3.356	3
G_2	3.744	1	2.203	3	3.641	1
G_3	3.462	2	2.070	4	3.483	2
G_4	0.036	5	2.000	5	1.93	5
G_5	3.245	3	3.805	1	3.245	4

7.2. Sensitivity analysis

This approach compares the outcomes of three situations in its sensitivity analysis. The weight values of the properties can be seen in Table 13. The study's result is Case 1, and the additional results found by applying various attribute weights are Cases 2 and 3. According to sensitivity analysis, changing an attribute's weights affects the ranking order, which is shown in Table 14.

8. Conclusion

Based on energy storage requirements, this paper proposes a strategy for choosing an appropriate energy storage technology. In this work, we create a transformation and fusion method to convey the information in an LNFN, and we treat the choice of energy storage technologies as an MCDM problem. The weights of the experts and criteria are then determined using ELECTRE-III and the best-worst technique, respectively. On the basis of this, an appropriate energy storage technology can be chosen. The presented approach can be modified for future research, allowing for even more research to be conducted using it. Type-2 fuzzy sets (T2FSs), IFSs, and PFSs, for instance, can be employed to model distinct types of decision-making environments. Additionally, the model can be modified to incorporate additional criteria. Numerous other energy-related problems can also be resolved using this strategy of supporting decision-making.

References

[1] Albawab. M, Ghenai. C, Bettayeb. M, and Janajreh. I. (2020). Sustainability performance index for ranking energy storage technologies using multi-criteria decision-making model and hybrid computational method. *Journal of Energy Storage*. 32:101820.

[2] Ministry of Power. (2015). Government of India, "Power Sector at a Glance all India", http://powermin.nic.in/power-sector-glance-all-india

[3] Ministry of Power. (2016). Coal and New & Renewable Energy, Governemnt of India " Ujwal Bharat, 2 year Achievements and Initiatives".

[4] Mateo JRSC. (2012). Multi Criteria Analysis in the Renewable Energy Industry. London: Springer.

[5] Colak. M, and Kaya. I. (2017). Prioritization of renewable energy alternatives by using an integrated fuzzy MCDM model: a real case application for Turkey, *Renewable Sustainable Energy Reviews*. 80:840-853.

[6] Ren. J. (2018). Sustainability prioritization of energy storage technologies for promoting the development of renewable energy: a novel intuitionistic fuzzy combinative distance-based assessment approach, *Renewable Energy*. 121:666-676.
[7] Azzuni. A, and Breyer. C. (2018). Energy security and energy storage technologies. *Energy Procedia*. 155:237-258.
[8] Gao. T, and Lu. W. (2020). Machine learning toward advanced energy storage devices and systems. *iScience*. 24:101936.
[9] Liu. Y, Du. J.L. (2020). A multi criteria decision support framework for renewable energy storage technology selection. *Journal of Cleaner Production*. 277:122183.
[10] Kumar. A, Sah. B, Singh. A.R, Deng. Y, He. X, Kumar. P, and Bansal. R. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable Sustainable Energy Reviews*. 69:596-609.
[11] Chen. T, Wang. Y.T, Wang. J.Q, Li. L, and Cheng. P.F. (2020). Multistage Decision Framework for the Selection of Renewable Energy Sources Based on Prospect Theory and PROMETHEE. *International Journal of Fuzzy Systems*. 22:1535-1551.
[12] Pamucar. D, Deveci. M, Schitea. D, Eriskin. L, Iordache. M, and Iordache. I. (2020). Developing a novel fuzzy neutrosophic numbers based decision making analysis for prioritizing the energy storage technologies. *International Journal of Hydrogen Energy*. 45:23027-23047.
[13] Barin. A, Canha. L.N, Abaide. A.R, Magnago. K.F, Wottrich. B, and Machado. R.Q. (2011). Multiple Criteria Analysis for Energy Storage Selection. *Energy and Power Engineering*, 3:557-564.
[14] Rahman. M, Oni. A.O, Gemechu. E, and Kumar. A. (2020). Assessment of energy storage technologies: A review. *Energy Conversion and Management* . 223:113295.
[15] DOE Global Energy Storage Database. (2018). Available online: http://Www.EnergystorageeXchange.Org/Projects/Data_visualization (accessed on 23 January 2021).
[16] Aneke. M, and Wang. M. (2016). Energy storage technologies and real life applications-A state of the art review. *Applied Energy*. 179:350-377.
[17] Ren. J, and Ren. X. (2018). Sustainability ranking of energy storage technologies under uncertainties. *Journal of Cleaner Production*, 170:1387-1398.
[18] Denholm. P, and Kulcinski. G.L, (2004). Life cycle energy requirements and greenhouse gas emissions from large scale energy storage systems. *Energy Conversion and Management*. 45:2153-2172.
[19] Sengul. U, Eren. M, Shiraz. S.E, Gezder. V, and Sengul. A.B. (2015). Fuzzy TOPSIS method for ranking renewable energy supply systems in Turkey. *Renewable Energy*. 75:617-625.
[20] Ozkan. B, Kaya. I, Cebeci. U, and Baesligil. H. (2015). A Hybrid Multi-criteria Decision Making Methodology Based on Type-2 Fuzzy Sets For Selection Among Energy Storage Alternatives. *International Journal of Computational Intelligence Systems*. 8:914-927.
[21] Zhang. C, Chen. C, Streimikiene. D, and Balezentis. T. (2019). Intuitionistic fuzzy MULTIMOORA approach for multi-criteria assessment of the energy storage technologies. *Applied Soft Computing*. 79:410-423.
[22] Ren. J. Sustainability prioritization of energy storage technologies for promoting the development of renewable energy: A novel intuitionistic fuzzy combinative distance-based assessment approach. *Renewable Energy*, 121;666-676.
[23] Colak. M, and Kaya. I.(2020). Multi-criteria evaluation of energy storage technologies based on hesitant fuzzy information: A case study for Turkey. *Journal of Energy Storage*, 28:101211.
[24] Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning Part I. *Information Science*. 8:199-249.
[25] Herrera. F, and Herrera-Viedma. E. (2000). Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets and System*. 115:67-82.
[26] Herrera. F, Herrera-Viedma. E, and Verdegay. L. (1996). A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets and System*. 79:73-87.

[27] Chen. Z.C, Liu. P.H, and Pei. Z. (2015). An approach to multiple attribute group decision making based on linguistic intuitionistic fuzzy numbers. *International Journal of Computational Intelligence Systems*. 8:747-760.

[28] Liu. P, and Wang. P. (2017). Some improved linguistic intuitionistic fuzzy aggregation operators and their applications to multiple-attribute decision making. *International Journal of Information Technology & Decision Making*. 16:817-850.

[29] Smarandache. F. (1998). Neutrosophy: Neutrosophic Probability, Set, and Logic; American Research Press: Rehoboth, DE, USA.

[30] Ye. J. (2015). An extended TOPSIS method for multiple attribute group decision making based on single valued neutrosophic linguistic numbers. *Journal of Intelligent & Fuzzy Systems*. 28:247-255.

[31] Fang. Z, and Ye. J. (2017). Multiple attribute group decision-making method based on linguistic neutrosophic numbers. *Symmetry*. 9(7):111.

[32] Li. Y. Y, Wang. J. Q, and Wang. T. L. (2019). A linguistic neutrosophic multi-criteria group decision-making approach with EDAS method. *Arabian Journal for Science and Engineering*. 44:2737-2749.

[33] Fan. C, Ye. J, Hu. K, and Fan. E. (2017). Bonferroni mean operators of linguistic neutrosophic numbers and their multiple attribute group decision-making methods. *Information*, 8(3):107.

OPTIMALITY PREDICTION OF SECOND ORDER BOX-BEHNKEN DESIGN ROBUST TO MISSING OBSERVATION

A.R. Gokul[1] and M. Pachamuthu[2*]

•

[1]Research Scholar, Department of Statistics, Periyar University, Salem-636011, India
gokulmalar6@gmail.com
[2*]Assistant Professor, Department of Statistics, Periyar University, Salem-636011, India
pachamuthu@periyaruniversity.ac.in

Abstract

The study of robust missing observations has gained prominence in statistical research. In particular, the Response Surface Methodology (RSM), a widely applied approach in experimental design, faces challenges when dealing with missing data. This paper investigates two design variants: the three-level second-order Box-Behnken design (BBD) with one missing observation and the Small Box-Behnken Design (SBBD), which involves fewer experimental runs than the standard BBD. We evaluate prediction performance using a fraction of design space (FDS) plot, revealing the distribution of scaled prediction variance (SPV) values across the design space. Additionally, we assess the efficiency of design model parameters using information-based criteria (A, D, and G relative efficiency). Our analysis spans k factors, ranging from k = 3 to 9. The findings guide practitioners in selecting optimal design points for efficient parameter estimation and accurate prediction within the context of missing observations. This comparative study sheds light on the trade-offs between BBD and SBBD, providing valuable insights for experimental design practitioners.

Keywords: Box-Behnken Design, Fraction of Design Space, Scaled Prediction Variance, Optimality, Small Box-Behnken Design

I. Introduction

Response Surface Methodology is a powerful statistical and mathematical model construction technique blend. It's designed to assess the impact of multiple independent variables and find their optimal values to yield the most desirable outcomes. This methodology benefits scenarios that aim to optimize a product or process. The empirical model is based on data observed from the system or process. *RSM* involves building empirical models using multiple regression and statistical techniques [19]. The second-order model is commonly used in *RSM*, particularly in Central Composite Designs (*CCD*) and *BBD*. Robust missing observation is a critical research area in all statistical methodologies. Even in well-planned experiments, there may be a chaos of missing observations that becomes challenging for estimating parameters in a model. Much of the robust missing observation research in the literature review is performed in different Central Composite Design types using various alpha (α) values rather than second-order Box-Behnken Designs. Draper [7] reviews the research on robust missing observation methods in response surface design and credits the first researcher to develop a parameter estimation formula. Akhtar and Prescott

[2]proposed a minimax loss criterion for handling missing observations, which is now the most used in response surface designs. Akhtar [1] examines a five-factor *CCD* with two missing observations in three different settings. Smucker et al. [20] gave empirical results for the effect of missing observations on various classical and optimal designs and a new type of missing-robust design in screening and response surface settings. Alrweili et al. [4] use the minimax loss criterion to create more robust designs for missing observations by combining the latest *CCDs* from *GSA* and *AEK*, which are new designs. Hayat et al. [8]explore designs from regular and irregular structure subsets, assess how they dealt with missing design points using the minimax loss criterion, and investigate their alphabetic optimality and prediction performance with *FDS* plots of the response difference variance. Alanazi et al.[3] present closed-form expressions for missing two observations as a function of α, the axial value used in *CCDs* with up to 10 factors. Whittinghill [22] explores how Box-Behnken Designs can handle missing observations without losing the ability to estimate all the parameters of interest and the article defines tmax as the maximum number of rows that can be arbitrarily deleted from the design matrix and keep the parameters estimable. Tanco et al.[21] used tmax and D-Efficiency criteria to evaluate three-level second-order polynomial designs, such as Box-Behnken, Face Centered, and other smaller and intermediate designs. Rashid et al.[18] investigate how to deal with one missing observation in Augmented *BBD* (*ABBD*) and Augmented Fractional *BBD* (*AFBBD*) using the minimax loss criterion and relative *D* and *G* efficiency. Rashid et al. [17] examine how a missing observation affects the estimation and prediction abilities of the *ABBD'S* relative *A-*, *D-*, and *G*-efficiencies. Hemavathi et al. [9] explore how sequential third-order rotatable design can handle missing observations without losing much information. Also, the paper measures the information loss due to one or two missing experimental runs at different distances from the center of the design. Park et al. [16] use graphical methods such as variance dispersion graphs, a fraction of the design space plot, and *G-, I-* optimality criteria to examine how different experimental designs perform in spherical and cuboidal regions for three to seven factors. Chigbu et al. [13] compare *CCD*, Small Composite Design (*SCD*), and MinResV designs for spherical regions with $k = 3$ to 7 factors based on the optimality criteria and the Variance Dispersion Graph (*VDG*), and the results show that none of these designs is consistently better for themselves. Li et al. [12] evaluate different *CCD*, *SCD*, and *MinResV* designs for spherical and cuboidal regions with various axial values, and they utilize *FDS* plots and box plots to analyze the prediction variance properties of the designs. Onwuameze et al. [14] use graphical methods such as *VDG* and *FDS* plots to evaluate the prediction variance performance of *CCD*, *SCD*, and *MinResV* in the hypercube region. However, most research on robust missing observations in Box-Behnken design recently focused on third-order designs, called *ABBD*.

This paper conducts a comparative analysis of the classical *BBD* and the *SBBD* focusing on the robustness of these designs when a single observation is missing. The *SBBD* [24] is noted for its advantage of requiring fewer runs than the *BBD*. The paper evaluates these designs using relative *A-*, *D-*, and *G*-efficiency to assess parameter estimation accuracy and explores a fraction of the design space plot in terms of scaled prediction variance. The paper is structured as follows: Section 2 Outlines the methodology used in the study. Section 3.1: Presents the results and discussions on scaled prediction variance and relative efficiencies. Section 3.2: Provides an analysis and discussion using the Fraction of Design Space graph of both *BBD* and *SBBD*. Section 4: Summarizes the findings and conclusions of the study.

II. Methodology

I. Description of Second Order Model

In numerous instances involving response surface methodology, we may not understand the relationship between the predictor variables and the response. A first-order model might not be sufficient to capture the curvature of the response function. Therefore, we often use higher-degree polynomials, such as a second-order model, to better evaluate curvature in optimization

experiments. For k quantitative factors denoted by $x_1, x_2, \dots, x_k$, a second-order model is

$$y = \beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1}^{k}\beta_{ii}x_i^2 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\beta_{ij}x_i x_j + \varepsilon \quad (1)$$

where $\beta_0, \beta_i, \beta_{ii}$, and β_{ij} are the intercept, linear, quadratic, and bilinear terms, respectively, and ϵ_i is a random error with mean zero, variance σ^2 and independent between any pair of runs. The number of unknown parameters to be estimated is denoted as $p = k + k + \binom{k}{2} + 1$, and to have sufficient degrees of freedom to estimate the model coefficients, the number of runs or observations n must be greater than or equal to p.

II. Small Box-Behnken Design

Box and Behnken [6] combined balanced incomplete block design and factorial design to create a three-level factorial design called Box-Behnken design. Box-Behnken Designs are three-level, second-order spherical designs with all points on a sphere. They are typically used for fitting second-order response surface models and are available for 3–12 and 16 factors. This design is widely used for second-order models in analytical chemistry and industrial applications. Small Box-Behnken Design is specially constructed by using a Balanced incomplete block design (*BIBD*) and Partially Balanced incomplete block design (*PBIBD*) and replaces treatments partly by 2^{3-1}_{III} designs and partly by full factorial designs. A unique feature of *SBBD* is that it has minimum runs compared to classical *BBD*. *SBBD* consists of two design point categories: Full Factorial design 2^2 or 2^3 denoted as (*F*), and 2^{3-1}_{III} Fractional Factorial design denoted as (*FF*), which has runs in the form $\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$.

Appendix A of the article outlines the detailed structures of the design point types or called as design matrix X for the Small Box-Behnken Design. Using the design matrix X of *BBD* and *SBBD*, we can calculate further computational analysis. To know about the further construction methods of *SBBD*, refer to article [24].

III. Scaled Prediction Variance

Borkowski [5] gives an analytical form for calculating scaled prediction variance values of central composite design and Box-Behnken design. Scaled prediction variance criteria is an essential tool for selecting response surface designs. It allows for good prediction of response variables at various points of interest throughout the experimental region. This scaling is widely used to facilitate comparisons among designs of various sizes. The prediction variance at a point **x** is given by

$$v(x) = \frac{n \cdot var[\hat{y}_{(x)}]}{\sigma^2} = n \cdot X^{(m)'}(X'X)^{-1}X^{(m)} \quad (2)$$

$x^{(m)}$ is the design point vector in the design space expanded to model form, n is the design size or runs, and σ^2 is the observation error. Desirable designs are those which have the smallest value of scaled prediction variance.

IV. Fraction of Design Space

The *FDS* plot is a useful tool to compare two or more designs, as it shows the *SPV* distributions of designs with a single curve and their *G*-efficiency and *V*-average values. *FDS* plot is [15] constructed by sampling many values, say n, from throughout the design space and obtaining the corresponding *SPV* values. The *FDS* plot informs the experimenter how the *SPV* varies throughout the design space,

including the minimum and maximum *SPV*s. The idea is that the design is better if a larger fraction of the design space is close to the minimum *SPV* value. Moreover, the design is more stable if the line is flatter. The *FDS* plot helps summarize the range and the proportions of *SPV* values in the design space and easily compares designs with a single curve. In addition, [23] it provides the researcher with a single plot to compare designs or study the properties of a specific design. Accordingly, the *FDS* technique could be applied to regular and non-regular design regions.

V. Relative G-, D-, and A-, Efficiency

G-optimality is defined as minimizing the maximum variance of any predicted value over the experimental space. Iwundu[10] investigates how single or multiple missing observations affect cuboidal designs' Relative *A-, D-,* and *G*-efficiency. It is defined as the ratio of the determinant of the information matrix of the design to the determinant of the information matrix of an optimal design.

$$G_{\text{eff}} = \frac{p}{n \cdot MAX_{X \in R} v(x)} \tag{3}$$

Here p is the number of parameters of estimated model, n is the number of observations in the respective design and $MAX_{X \in R} v_{(x)}$ is the maximum value of the variance of predicted response. Thus, relative G-efficiency denoted by RE_G is given as the ratio of G_{eff} of reduced design and of complete design.

$$RE_G = \frac{G_{\text{eff (reduced)}}}{G_{\text{eff}}} = \frac{n \cdot MAX_{X \in R} v(x)}{n_r \cdot MAX_{X \in R} v(x)_{\text{reduced}}} \tag{4}$$

where n is the size of the runs of the complete design, and n_{r} is the size of the runs of the reduced design. According to this definition of RE_G, a design with a higher value of RE_G will be preferred. By utilizing equations (3) and (4), we can compute the relative G-efficiency value. These values are then presented in tables 3 and 4.

D- efficiency is defined as maximizing the determinant of the information matrix or minimizing the determinant of the inverse of the information matrix. Thus, relative *D*-efficiency is given as

$$RE_D = \left(\frac{|X'X|_{\text{reduced}}}{|X'X|} \right)^{\frac{1}{p}} \tag{5}$$

Where, p is the number of parameters of the model to be estimated, $|X'X|_{\text{reduced}}$ is the determinant of the information matrix of reduced design and $|X'X|$ is the determinant of the complete design matrix. A value approaching one will represent a minor loss, whereas a value below one will represent a more significant loss in model estimation. Through the application of equation (5), we are able to determine the relative *D*-efficiency value. These computed values are then listed in tables 3 and 4.

The A-Criterion considers the individual variances of the regression coefficients rather than the covariances among coefficients. Thus, relative *A*-efficiency is given as

$$RE_A = \frac{\text{trace}(X'X)^{-1}}{(\text{trace}(X'X)^{-1})_{\text{reduced}}} \tag{6}$$

where the trace is the sum of the main diagonal values of $(X'X)^{-1}$, where $(\text{trace}\ (X'X)^{-1})_{\text{reduced}}$ is the trace of $(X'X)^{-1}$ of the reduced design, and a design with a higher value of RE_A will be preferable. By utilizing equation (6), we can ascertain the relative *D*-efficiency value. The calculated values are subsequently enumerated in tables 3 and 4.

III. Result and Discussion

I. Scaled Prediction Variance of One Missing Box-Behnken Design

The insignificant difference in the Average Scaled Prediction Variance (*ASPV*) between instances of missing and non-missing observations is evident from the values in Table 1. The smallest prediction value is preferred for optimal prediction performance among each factor's design points. Regarding missing factorial design points, factor k = 8 has the lowest SPV value of 24.08 compared to all other factors k = 3,4,5,6,7 and 9, all of which have *SPV* values ranging from 34 to 45.

Table 1: *SPV values of one missing observation of BBD k = 3 to 9 factors.*

Number of Factors	Missing Design Points	Runs	SPV		
			Min	Avg	Max
k = 3	None		3.2521	7.3627	16.6260
	F	12	3.0656	8.9792	42.2400
	Centre	5	3.6256	7.0704	15.7584
k = 4	None		4.6440	10.4340	17.4960
	F	24	4.4892	10.9765	39.4400
	Centre	6	5.1591	10.1732	16.9157
k = 5	None		6.6424	14.9776	24.9964
	F	40	6.5025	15.3540	40.3875
	Centre	6	7.3260	14.8095	24.5565
k = 6	None		7.7598	13.1220	21.8160
	F	48	7.7009	13.3666	33.7663
	Centre	6	8.5807	13.3030	21.1735
k = 7	None		9.3248	15.3698	20.0942
	F	56	9.1744	15.6709	40.1258
	Centre	6	10.5530	15.5733	19.9958
k = 8	None		12.1800	16.7040	21.0480
	F	112	12.0309	16.7076	24.0856
	Centre	8	12.7449	17.3621	21.5985
k = 9	None		11.8170	24.9730	34.4110
	F	120	11.7648	24.8325	34.8945
	Centre	10	12.7839	25.0647	34.4301

II. Scaled Prediction Variance of One Missing Small Box-Behnken Design

In the case of the Small Box-Behnken Design, there is an upward trend in the Scaled Prediction Variance for most factors when there are non-missing design points and a center. However, factor k = 7 deviates from this increasing trend. When comparing the difference between full factorial and 2^{3-1}_{III} fractional factorial design points based on average and Max SPV, there is a moderate difference in the average SPV of all the factors. Interestingly, factors k = 8 and 9 have similar differences. In contrast, Max SPV for factors k = 4,7 and 9 shows significant differences between *F* and *FF* design points, while other factors such as k = 5,6 and 8 exhibit moderate differences. Therefore, we can infer that full factorial observations perform better in terms of prediction when a factorial type of observation is missing, compared to fractional factorial observations.

Table 2: *SPV values of one missing observation of SBBD k = 4 to 9 factors.*

Number of Factors	Missing Design Points	Runs	SPV		
			Min	Avg	Max
k = 4	None		3.1856	14.1042	44.0000
	F	12	3.0954	18.6249	113.4000
	FF	4	2.8371	27.5100	307.2300
	Centre	6	3.3411	13.5639	43.2600
k = 5	None		4.8030	18.8430	60.6000
	F	8	4.7096	19.9404	65.2500
	FF	16	4.7908	21.9820	123.8300
	Centre	6	5.1881	18.2294	58.2900
k = 6	None		6.0458	21.2116	74.1000
	F	16	5.8793	21.7375	74.0000
	FF	16	5.9385	24.1092	109.1500
	Centre	6	6.6304	21.0974	75.8500
k = 7	None		6.0000	23.5152	67.2000
	F	24	5.8750	25.9534	94.4700
	FF	16	5.7528	29.6429	248.6300
	Centre	8	6.4719	23.0723	70.5000
k = 8	None		7.7760	30.6688	85.1200
	F	28	7.6608	32.4198	102.0600
	FF	28	7.6734	32.8734	121.5900
	Centre	8	8.7129	30.4668	85.0500
k = 9	None		7.0000	40.4110	108.5000
	F	24	6.9000	46.1817	180.7800
	FF	36	6.9000	46.1196	325.6800
	Centre	10	7.6659	40.1373	100.0500

III. Relative G, A, and D efficiency values of Box-Behnken Design

Regarding the impact on relative efficiencies, let's first consider the relative A-efficiency. Table 3 presents the variations in relative A-efficiencies resulting from the absence of a factorial point, a center point, and a non-missing point for all factors from k = 3 to 9. The numerical data indicates that A-efficiency is marginally influenced by the absence of a factorial point for only factor k=3. On the other hand, the absence of a factorial point has a statistically significant effect on all other factors. When a center point is missing, the relative A efficiency is similar to that when no design points are missing. Therefore, estimating the precision of individual variances of regression coefficients of the second-order model performs quite well when either factorial or center run points are missing.

The relative *D*-efficiencies closely mirror the *A*-efficiencies. When a factorial or center run observation is missing, the relative D-efficiencies are similar to the efficiencies of a complete design for factors k = 3 to 9. Furthermore, the relative *D*-efficiency value significantly estimates the covariances among coefficients when some observations are absent.

The relative *G*-efficiencies exhibit notable similarities with the relative *A*- and *D*-efficiencies. The absence of a factorial point significantly impacts the relative G-efficiencies for factors k = 3,4 and 7, moderately affects factors k = 5,6, and is less concerning for factors k = 8,9. Regarding the missing center point, the relative G-efficiency exceeds one compared to when no observations are missing.

Table 3: *Relative G, A, and D efficiency values one missing observation of BBD k = 3 to 9 factors*

Number of Factors	Missing Design Points	G Efficiency	Relative G efficiency	Relative A efficiency	Relative D efficiency
$k = 3$	None	60.1400	1.0000	1.0000	1.0000
	F	23.6900	0.3939	0.7846	0.8706
	Centre	63.4600	1.0552	0.9577	0.9779
$k = 4$	None	85.7300	1.0000	1.0000	1.0000
	F	37.9100	0.4422	0.9181	0.9433
	Centre	88.6700	1.0343	0.9736	0.9879
$k = 5$	None	85.5100	1.0000	1.0000	1.0000
	F	52.0000	0.6081	0.9569	0.9675
	Centre	13.7400	0.1607	0.9780	0.9914
$k = 6$	None	64.17	1.0000	1.0000	1.0000
	F	82.9200	0.6461	0.9607	0.9709
	Centre	66	1.0303	0.9818	0.9935
$k = 7$	None	89.5	1.0000	1.0000	1.0000
	F	89.7100	0.5007	0.9574	0.9731
	Centre	90.025	1.0049	0.9836	0.9949
$k = 8$	None	71.28	1.0000	1.0000	1.0000
	F	93.405	0.8736	0.9853	0.9890
	Centre	69.45	0.9744	0.9853	0.9970
$k = 9$	None	79.84	1.0000	1.0000	1.0000
	F	78.81	0.9871	0.9957	0.9932
	Centre	79.88	1.0004	0.9957	0.9981

IV. Relative G, A, and D efficiency values of SBBD K = 4 to 9 factors.

Relative A and D efficiencies exhibit similar effects for all factors from k = 5 to 9, except for factor k = 4. For factor k = 4, the relative A efficiency for 2^{3-1}_{III} fractional factorial points is 0.3906, while the relative D efficiency is 0.7179. Both full factorial and 2^{3-1}_{III} fractional factorial points demonstrate good accuracy for individual coefficients and covariances among coefficients when observations are missing for factors k = 5 to 9 in terms of relative A and D efficiency. The numerical data indicates that the absence of a center point does not impact all factors' relative A and D efficiency.

Relative G efficiencies are significantly influenced by factor k = 4. However, factors k = 7 and 9 exhibit superior prediction performance compared to factor k = 4. Moreover, factors k = 5,6 and 8 excel in minimizing the maximum prediction variance compared to all other factors when the full factorial observation is missing. When it comes to missing 2^{3-1}_{III} fractional factorial observations, only factors k= 6 and 8 have a marginally better effect than all other factors of SBBD. The absence of a center point does not impact the relative G efficiency for any factor.

Table 4: *Relative G, A, and D efficiency values one missing observation of SBBD K = 4 to 9 factors.*

Number of Factors	Missing Design Points	G Efficiency	Relative G efficiency	Relative A efficiency	Relative D efficiency
$k = 4$	None	34.07	1.0000	1.0000	1.0000
	F	13.2300	0.3883	0.7360	0.8874
	FF	4.5600	0.1338	0.3906	0.7179
	Centre	34.6900	1.0182	0.9855	0.9879

$k = 5$	None	34.6200	1.0000	1.0000	1.0000
	F	32.2000	0.9301	0.9101	0.9399
	FF	16.9700	0.4902	0.8006	0.9127
	Centre	35.9900	1.0396	0.9883	0.9914
$k = 6$	None	37.8100	1.0000	1.0000	1.0000
	F	37.8200	1.0003	0.9553	0.9582
	FF	25.6800	0.6792	0.8147	0.9285
	Centre	36.8300	0.9741	0.9904	0.9935
$k = 7$	None	53.6200	1.0000	1.0000	1.0000
	F	38.0800	0.7102	0.9097	0.9573
	FF	14.0700	0.2624	0.7087	0.8710
	Centre	51.0700	0.9524	0.9959	0.9963
$k = 8$	None	53.0100	1.0000	1.0000	
	F	44.0600	0.8312	0.9426	0.9700
	FF	37.0600	0.6991	0.9013	0.9613
	Centre	52.7300	0.9947	0.9956	0.9971
$k = 9$	None	50.7400	1.0000	1.0000	1.0000
	F	30.4500	0.6001	0.8923	0.9661
	FF	16.8800	0.3327	0.8130	0.9519
	Centre	54.8300	1.0023	0.9864	0.9981

V. Discussion on Box-Behnken Design by FDS plot

If we interpret Figure 1 (a) intending to identify the most effective design point types based on *G*-efficiency and maximum *SPV*, it appears that the center and non-missing design points outperform factorial. They achieve a maximum *SPV* value of 15.98, which equates to a *G*-efficiency of 60.6%; this is considerably better than the factorial design points, which reach a high *SPV* value of 42.44 and a *G*-efficiency of 23.69%.

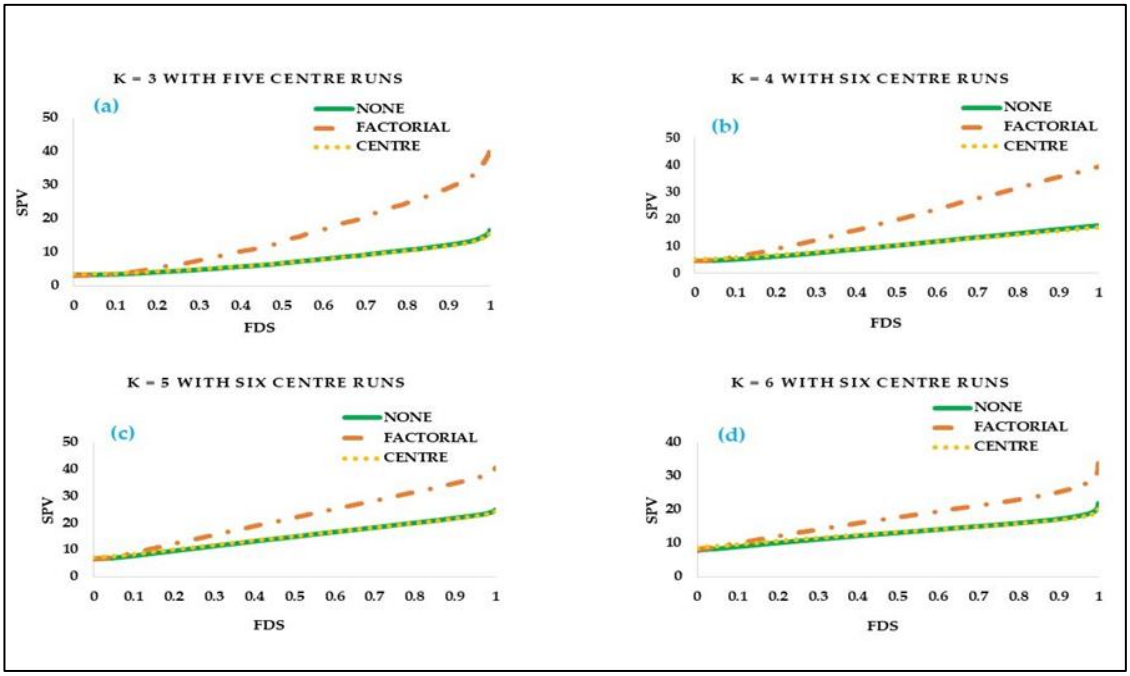


Figure 1: *(a) FDS for BBD K = 3. (b) FDS for BBD K = 4. (c) FDS for BBD K = 5. (d) FDS for BBD K = 6*

When evaluating *BBD k*=4 in Figure 1 (b), we can see a clear difference between the 50th and 75th percentiles of the design points. For the factorial points, the *SPV* at the 50% *FDS* is 19.80, and at the 75% *FDS*, it is 29.92, resulting in a difference of 10.12. On the other hand, for the center and non-missing points, the *SPV* at the 50% *FDS* is 10.29, and at the 75% *FDS*, it is 13.68, yielding a smaller difference of only 3. This percentile-based assessment leads us to conclude that the center and non-missing points demonstrate more consistent and superior performance across the design space than the factorial points.

The *FDS* plot depicted in Figure 1 (c) for *BBD k*=5 suggests that the absence of both the center and factorial design points significantly reduces the likelihood of obtaining a horizontal flat line. The median *FDS* value for the center and non-missing points is 15.21 *SPV*, with a mean or average *SPV* value of 14.8, indicating a lack of symmetry in the *SPV* and FDS distributions [11]. When a factorial point is missing, the maximum *SPV* reaches 40.38, and the average *SPV* is 15.34, suggesting that 50% of the design space exhibits moderate prediction performance.

The *FDS* plots in Figures 1 (d) and (e) reveal that the curves for the center and non-missing design points exhibit similar prediction performance. Both reach a maximum *SPV* value of approximately 20 at *FDS*=1 for factors six and seven. A slight flat line is noticeable in the horizontal curve of the center and non-missing design points for both factors. This increase begins from the median of the *FDS* and extends to roughly 90% of the design space region, suggesting a certain level of stability in the *SPV* distribution of the design points. When a factorial point is missing, factor six achieves a maximum *SPV* value of 33.76, indicating better prediction performance than factor seven, which has an *SPV* value of 40.13.

Based on comparing factors eight and nine from the *FDS* plot in Figures 1 (f) and (g), BBD k=8 exhibits a very similar curve across the entire design space region, with a slightly increasing horizontal line. The *SPV* values range from a minimum of 12 to a maximum of 21. For *BBD* k=9, the curves for the center, factorial, and non-missing design points have similar prediction performance and are very close to each other, with *SPV* values ranging from 12 (*min*) to 34 (*max*) across the entire design space region. Interestingly, each factor's prediction performance is comparable when both design points are missing in factors k = 8 and 9.

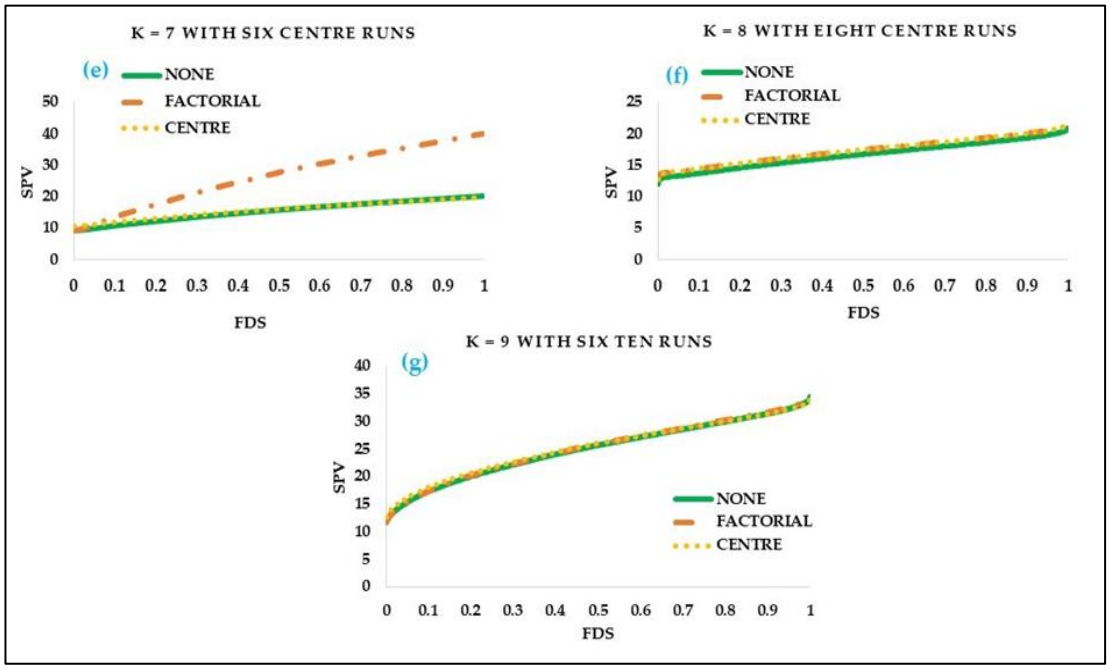


Figure 1: *(e) FDS for BBD K = 7. (f) FDS for BBD K = 8. (g) FDS for BBD K = 9.*

VI. Discussion on Small Box-Behnken Design by FDS plot

Indeed, Figure 2 (a) presents four types of design points: full factorial, 2^{3-1}_{III} fractional factorial, and center and non-missing design points. The point at approximately (0.50,12.15) represents that 50% of the total design space has an *SPV* value at or below 12 for the design point of one missing center run and non-missing run, where these design points exhibit a flat horizontal curve for the maximum design space region up to 1. A flatter curve implies that the maximum and minimum *SPV* values are closer together, indicating a more stable distribution of the *SPV* [11]. The full factorial design points of *SBBD* k = 4 consist of 12 runs. When 2^2 full factorial points are missing, the maximum SPV value is 113 for the maximum (*FDS* = 1) design space region, and 80% of the design region has an *SPV* value of 47 or below. This is similar to missing a center run for the max SPV value. From this, we can infer that 80% of the 2^2 full factorial design space region has moderate *SPV* compared to 100% of *FDS*. The 2^{3-1}_{III} fractional factorial design point in *SBBD* has four runs. When an FF point is missing, it results in a large SPV value of 307 for the maximum (*FDS* = 1) design space region. However, 50% of the total design space has an *SPV* value at or below 76.6, and 75% of the total design space region has an *SPV* value at or below 111.78. This suggests that *FF* has moderate SPV for 75% of the design space region compared to the large *SPV* value for the maximum (*FDS* = 1) region.

The *FDS* plot of the 2^3 full factorial, center, and non-missing design points is depicted in Figure 2 (b). These design points exhibit similar performance, as indicated by their comparable curves and *G*-efficiencies of 32.20, 35.99, and 34.62, respectively. Upon closer inspection, the center and non-missing design points are strikingly similar across the entire design space region, from the minimum (*FDS*=0) to the maximum (*FDS*=1), with SPV values of 60.60 and 58.29, respectively. However, the 2^3 full factorial design point deviates slightly from the other two similarity curves in the *FDS* region of 75%. The average *SPV* values for the center, full factorial, and non-missing points are also similar at 18.84, 19.94, and 18.22, respectively, suggesting that the absence of these design points during experimentation does not significantly impact the prediction performance of *SPV* on average.

Despite having a high SPV value of 123.83 at maximum (*FDS*=1), 2^{3-1}_{III} fractional factorial design points maintain a median *SPV* value of 31.82 in the design space curve and an *SPV* value of 60.42 for 88% of the total design space. This roughly equates to the maximum *SPV* value (*FDS* =1) of the center, factorial, and non-missing runs. Therefore, we can conclude that 2^{3-1}_{III} fractional factorial design points generally provide good prediction performance for most of the design space (90%), except for the maximum region (100%), where their performance is subpar.

As depicted in Figure 2 (c), the design points for the factor k = 6, including centre, 2^3 full factorial, and non-missing, exhibit a similar horizontal curve across the entire design space region, from the minimum (*FDS* = 0) to the maximum (*FDS* = 1). These design points have a *G*-efficiency of 37% and a maximum *SPV* value of 74. The SPV value fluctuates between 21 and 30, covering 55% to 80% of the design space region. It's interesting to note that while both full factorial and fractional factorial design points consist of the same sixteen runs, the absence of a run from the 2^{3-1}_{III} fractional factorial alone results in a significant *SPV* value of 109.45 in the maximum (*FDS* =1) region. The curves for 2^{3-1}_{III} fractional factorial are closely aligned with other design points from the minimum (*FDS* =0) region to 80% of the design space, with an *SPV* of 40.47 or less. All design points demonstrate moderate or average prediction performance up to 80% of the total design space region, indicating satisfactory prediction performance within this *FDS* region.

As shown in Figure 2 (d), 2^{3-1}_{III} fractional factorial design points are the only ones that do not exhibit a flat horizontal curve among all design points. The center and non-missing design points share the same horizontal curve, with an average *SPV* of 22.79 and similar *G*-efficiencies of 51 and 53.62, respectively. The 2^3 full factorial design points of *SBBD* k = 7 consist of 24 runs. If a run is missing from the experiment, the prediction performance of the design remains relatively consistent over approximately 80% of the total design space region, with an *SPV* value of 42.76. The 2^{3-1}_{III} fractional factorial design points have a substantial *SPV* value of 248.63 in the maximum (*FDS*=1) design space region. However, the median of the *FDS* curve has an average *SPV* of 70, where the maximum *SPV* is more than twice the average *SPV* value. Therefore, we can conclude that the

prediction performance of 2^{3-1}_{III} fractional factorial is moderate for only 50% of the total design space region.

As per the *FDS* plot in Figure 2 (e), it's observed that the prediction performance across all shrinkage levels is quite similar for all types of design points, ranging from the lowest to the highest *SPV* value. Approximately half of the *FDS* curves for all design points closely follow the average *SPV* values, while the remaining curves diverge towards the maximum of (*FDS*=1). The global *FDS* curve suggests that less than 25% of the design space has an *SPV* value of 29.61 or lower, and 50% has an *SPV* value of 39.62 or lower for all types of points, including those without missing values. For up to 75% of the design space, the prediction performance of non-missing and center points is significantly better than that of 2^2 or 2^3 full factorial points, and 2^2 or 2^3 full factorial missing points outperform 2^{3-1}_{III}fractional factorial points.

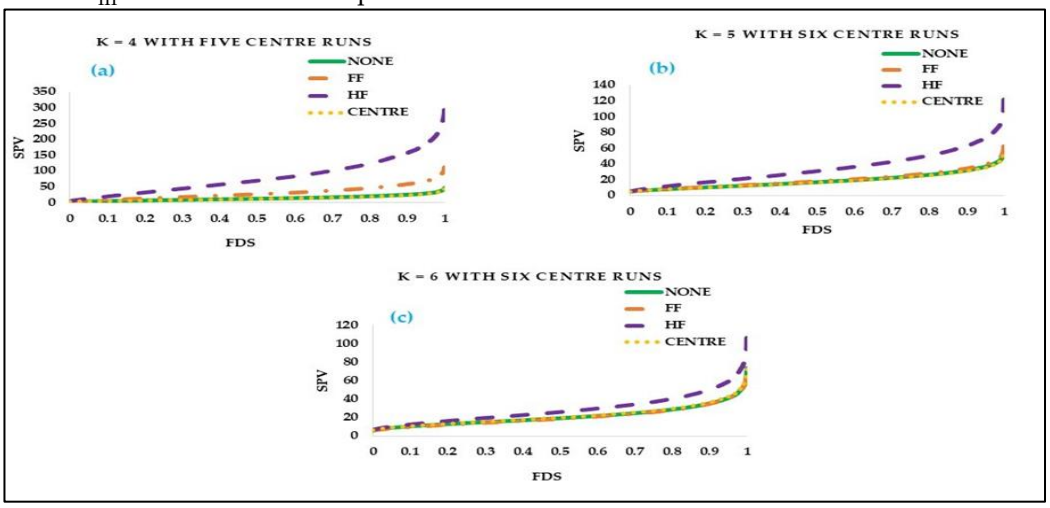


Figure 2: *(a) FDS for SBBD K = 4. (b) FDS for SBBD K = 5. (c) FDS for SBBD K = 6*

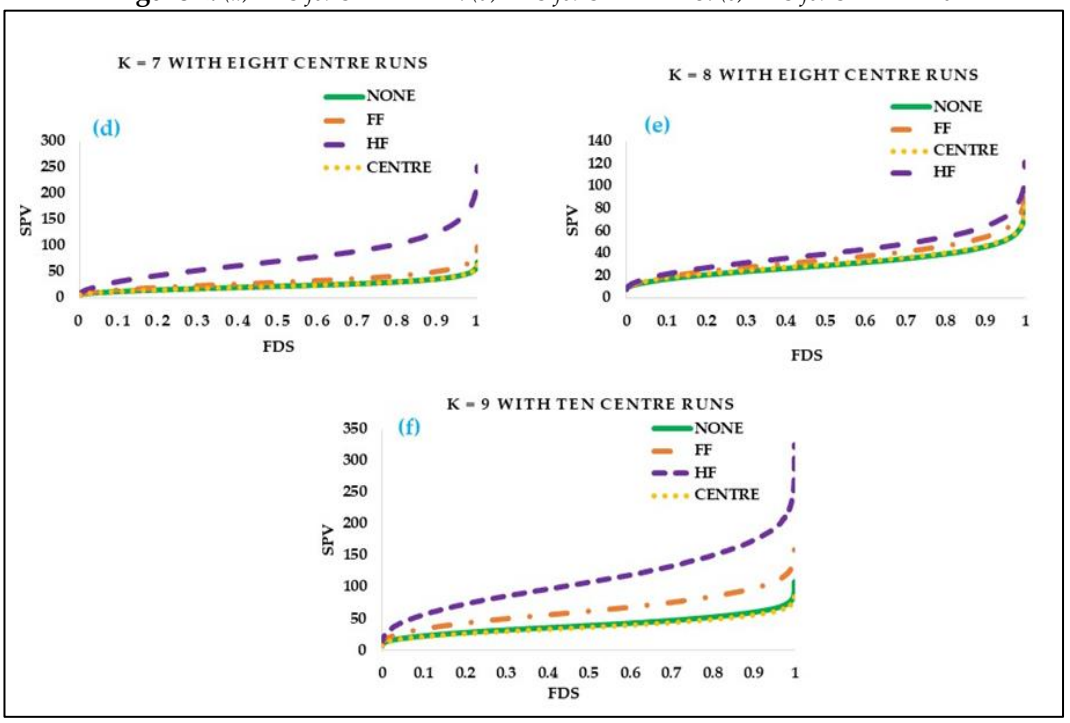


Figure 2: *(a) FDS for SBBD K = 7. (e) FDS for SBBD K = 8. (f) FDS for SBBD K = 9*

IV. Conclusion

The robustness of a single missing observation in *BBD* and *SBBD* is examined to determine which design points offer the best design efficiency and model parameter estimation using an information-based criterion. *BBD* and *SBBD* show good accuracy in estimating individual coefficients and covariances among coefficients when a design point is missing for all factors, except for factor $k = 3$ in *BBD* and $k = 4$ in *SBBD*, which show poor accuracy. The relative *G*- efficiency of *BBD* indicates that factors with increasing numbers have lower maximum variance across the experimental space, except for factor $k = 7$. In *SBBD*, full factorial and center design points have lower maximum variance across the experimental space than fractional factorial design points. The *FDS* plot reveals that missing a center and non-missing run have similar *SPV* values for all factors across the entire design space region for both *BBD* and SBBD. In comparison, the factorial design point type in *SBBD* has higher *SPV* values than BBD despite the fewer runs in *SBBD*. For *BBD*, all factors have similar performance for factorial design points. However, in *SBBD*, full factorial design points outperform 2_{III}^{3-1} fractional factorial design points have a high *SPV* value at the maximum region (FDS =1). If a design point is missing in the 2_{III}^{3-1} fractional factorial design points of *SBBD*, it results in subpar prediction performance.

This research work indeed holds significant potential in identifying robust missing design point types when observations are missing in an experimental situation for both Box-Behnken Design and Small Box-Behnken Design for certain factors. The findings can be beneficial when data may be lost or corrupted during the experimental process. Moreover, the scope for further research is vast. Future studies could extend this work to include more factors and more than one missing observation with multiple combinations. This would allow for a more comprehensive understanding of the robustness of these designs under various conditions.

References

[1] Akhtar, M. (2001). Five-Factor Central Composite Designs Robust To A Pair Of Missing Observations. ▼*J. Res. Sci*, *12*(2), 105–115.

[2] Akhtar, M. and Prescott, P. (1986). Response surface designs robust to missing observations. *Communications in Statistics - Simulation and Computation*, *15*(2), 345–363.

[3] Alanazi, K. Georgiou, S. D. and Stylianou, S. (2023). On the robustness of central composite designs when missing two observations. *Quality and Reliability Engineering International*, *39*(4), 1143–1171.

[4] Alrweili, H. Georgiou, S. and Stylianou, S. (2019). Robustness of response surface designs to missing data. *Quality and Reliability Engineering International*, *35*(5), 1288–1296.

[5] Borkowski, J. J. (1995). Spherical prediction-variance properties of central composite and Box—Behnken designs. *Technometrics*, *37*(4), 399–410.

[6] Box, G. E. P. and Behnken, D. W. (1960). Some New Three Level Designs for the Study of Quantitative Variables. *Technometrics*, *2*(4), 455–475.

[7] Draper, N. R. (1961). Missing Values in Response Surface Designs. *Technometrics*, *3*(3), 389.

[8] Hayat, H. Akbar, A. Ahmad, T. Bhatti, S. H. and Ullah, M. I. (2023). Robustness to missing observation and optimalities of response surface designs with regular and complex structure. *Communications in Statistics - Simulation and Computation*, *52*(11), 5213–5230.

[9] Hemavathi, M. Varghese, E. Shekhar, S. Athulya, C. K. Ebeneezar, S. Gills, R. and Jaggi, S. (2022). Robustness of sequential third-order response surface design to missing observations. *Journal of Taibah University for Science*, *16*(1), 270–279.

[10] Iwundu, M. (2017). Missing observations: The loss in relative A-, D- and G-efficiency. *International Journal of Advanced Mathematical Sciences*, *5*(2), 43.

[11] Khuri, A. I. and Mukhopadhyay, S. (2010, March). Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*.

[12] Li, J. Li, L. Borror, C. M. Anderson-Cook, C. and Montgomery, D. C. (2009). Graphical Summaries to Compare Prediction Variance Performance for Variations of the Central Composite Design for 6 to 10 Factors. *Quality Technology & Quantitative Management*, *6*(4), 433–449.

[13] Chigbu, P. E. Ukaegbu, E. C. and Nwanya, J. C. (2009). On comparing the prediction variances of some central composite designs in spherical regions: A review. *Statistica*, *69*(4), 285-298.

[14] Onwuamaeze, C. U. (2021). Optimal prediction variance properties of some central composite designs in the hypercube. *Communications in Statistics - Theory and Methods*, *50*(8), 1911–1924.

[15] Ozol-Godfrey, A. Anderson-Cook, C. M. and Montgomery, D. C. (2005). Fraction of Design Space Plots for Examining Model Robustness. *Journal of Quality Technology*, *37*(3), 223–235.

[16] Park, Y. J. Richardson, D. E. Montgomery, D. C. Ozol-Godfrey, A. Borror, C. M. and Anderson-Cook, C. M. (2005). Prediction variance properties of second-order designs for cuboidal regions. *Journal of Quality Technology*, *37*(4), 253–266.

[17] Rashid, F. Akbar, A. and Arshad, H. M. (2022). Effects of missing observations on predictive capability of augmented Box-Behnken designs. *Communications in Statistics - Theory and Methods*, *51*(20), 7225–7242.

[18] Rashid, F. Akbar, A. and Zafar, Z. (2019). Some new third order designs robust to one missing observation. *Communications in Statistics - Theory and Methods*, *48*(24), 6054–6062.

[19] Myers, R. H. Montgomery, D. C. and Anderson-Cook, C. M. (2016).Response surface methodology: process and product optimization using designed experiments. *John Wiley & Sons.*

[20] Smucker, B. J. Jensen, W. Wu, Z. and Wang, B. (2017). Robustness of classical and optimal designs to missing observations. *Computational Statistics and Data Analysis*, *113*, 251–260.

[21] Tanco, M. Del Castillo, E. and Viles, E. (2013). Robustness of three-level response surface designs against missing data. *IIE Transactions (Institute of Industrial Engineers)*, *45*(5), 544–553.

[22] Whittinghill, D. C. (1998). A note on the robustness of Box-Behnken designs to the unavailability of data. *Metrika* (Vol. 48). Springer-Verlag.

[23] Zahran, A. Anderson-Cook, C. M. and Myers, R. H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, *35*(4), 377–386.

[24] Zhang, T. F. Yang, J. F. and Lin, D. K. J. (2011). Small Box-Behnken design. *Statistics and Probability Letters*, *81*(8), 1027–1033.

METHODOLOGY FOR ASSESSING THE RELIABILITY OF AGS BASED ON RENEWABLE ENERGY SOURCES

N.S. Mammadov[1], K.M. Mukhtarova[2]

•

Azerbaijan State Oil and Industry University, Baku, Azerbaijan
[1] nicat.memmedov.sa@asoiu.edu.az ; [2] kubra.muxtarova.m@asoiu.edu.az

Abstract

With the introduction of renewable energy sources, in particular wind power and photovoltaic installations, in the autonomous generation systems, the problem of reliability of the equipment used and the entire energy complex becomes one of the main ones. It is necessary to develop and improve methods for analyzing and calculating reliability, which will make it possible at the design stage to take into account the probabilistic characteristics of renewable energy resources, reliability indicators and operating experience of the equipment used. The article discusses the scheme of an autonomous energy complex based on renewable energy sources. A graph of the dependence of failure rate on recovery time is presented. This paper discusses various methods for assessing the reliability of autonomous generation systems based on renewable energy sources: analytical methods, state space method (Markov process theory), Monte Carlo method, fault tree method and state enumeration method. The advantages and disadvantages of these methods are considered.

Keywords: autonomous generation systems, renewable energy sources, energy complex, wind turbine, reliability.

I. Introduction

In recent decades, renewable energy sources (RES) have become a key element of sustainable development strategies. These sources, including solar, wind, hydropower and biomass, have significant potential to reduce dependence on fossil fuels and reduce environmental impacts. One of the most promising applications of renewable energy sources is autonomous generation systems that provide electricity to remote or isolated areas where connection to centralized power grids is not economically or technically feasible.

There are several main configuration options for constructing autonomous generation systems using wind turbines and/or solar panels, diesel generators and batteries:

– a wind turbine operating in conjunction with a diesel generator;

– a wind turbine operating in conjunction with a diesel generator and an energy storage system;

– photoelectric converter operating in conjunction with a diesel generator and battery;

– joint operation of a wind turbine, photoelectric converter and diesel generator;

– joint operation of a wind turbine, photoelectric converter, diesel generator and battery.

The composition of the equipment of autonomous generation systems based on renewable energy sources varies depending on the availability of local energy resources. Figure 1 shows a scheme of an autonomous energy complex using wind and solar energy. This complex includes:

1) generating equipment photoelectric converter, wind turbine, diesel generator and energy storage system;

2) elements of transformation (inverter, converter, rectifier), transmission (cable line) of electricity;

3) switching equipment (automatic switches).

Photoelectric converters generate constant voltage electricity, which depends on external conditions, unlike wind turbines and a diesel generator. To coordinate different voltage levels received from the panels, wind turbines and diesel generators, and to ensure the rated output voltage, when the solar cell operates in the maximum power output mode, a converter is used in the circuit. The batteries serves as a backup power source and helps smooth out possible power fluctuations caused by the variable nature of natural energy resources [1-5].

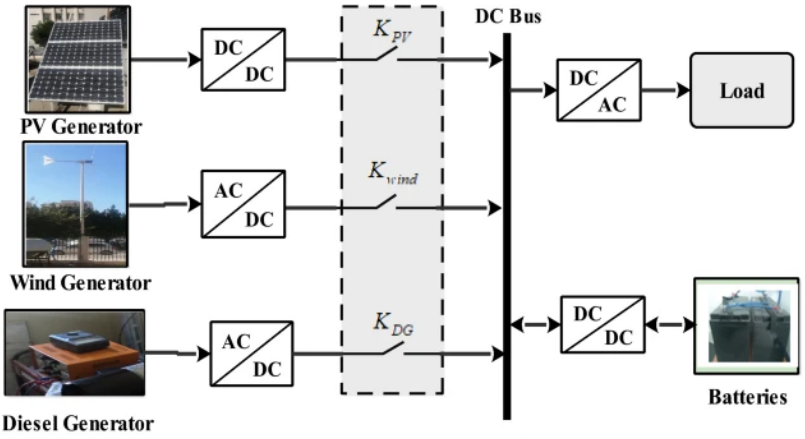


Figure 1: *Scheme of an autonomous energy complex based on renewable energy sources*

II. Formulation of the problem

The reliability of such off-grid generation systems is a critical factor in determining their efficiency and acceptability. Reliability refers to the ability of a system to perform its functions under specified operating conditions over a required period of time. For autonomous generation systems based on renewable energy sources, this means a stable and continuous supply of electricity, despite variable and often unpredictable natural conditions. The main challenges facing such systems include the variability of generated power due to natural factors, the need for efficient energy storage, the integration and management of different types of renewable energy sources, and the durability and wear resistance of system components. Given these challenges, the development of reliability assessment techniques is becoming a key element in the process of designing, operating and improving autonomous generation systems.

Over time, equipment in electrical power plants may fail. Failure is defined as an event in which the operational state of an object is disrupted. Most often, failed elements can be restored. The use of reserve elements allows you to restore equipment without interrupting its operation. The process of equipment restoration and prevention does not completely eliminate the possibility of failures, but significantly reduces their likelihood, which increases reliability. Any object and its properties can be described using various systems of reliability indicators. The choice of a system of indicators depends on the nature of the object, its purpose, general requirements for the process and the results of its operation, as well as the economic efficiency criteria used. In this work, the main indicators characterizing the reliability of electrical equipment of autonomous generation systems (AGS) based on RES are the failure rate $\lambda(t)$, 1/year, the average recovery time τ_{avg}, h, and the recovery rate μ, 1/year. Failure rate is the conditional probability density of a failure occurring at a given point in time, provided that a failure has not occurred before that point. The physical

meaning of the failure probability density is the number of failures of an element over a sufficiently short time interval [6-9].

Figure 2 shows a failure rate vs. operating time curve that is typical for many pieces of electrical and electronic equipment. The figure shows that the entire time interval can be divided into three sections. The first section (run-in period) is characterized by increased values of the function $\lambda(t)$ due to equipment failure shortly after the start of operation, caused by hidden defects during production. The second section (the period of normal operation) is characterized by a constant value of the failure rate. The last section is the aging period.

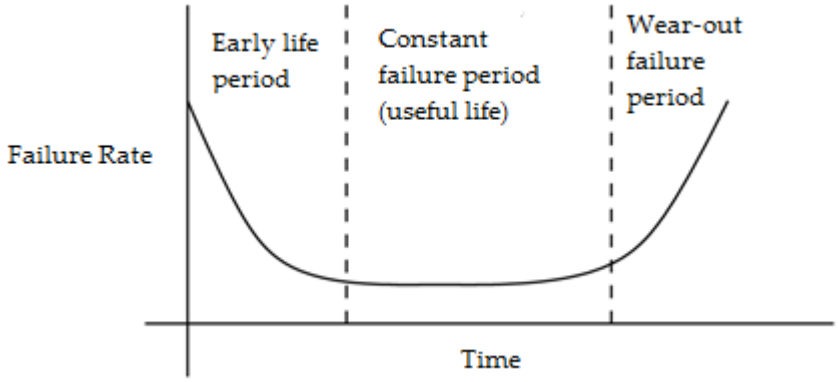


Figure 2: *Dependence of failure rate on recovery time*

In this work, we assume that $\lambda(t)=\lambda=const$, since for most elements in autonomous generation systems based on renewable energy sources, the failure rate remains almost constant for a long period of time. The run-in period can be ignored, assuming that normal operation begins immediately after its completion. Mean time to recovery is the time required to detect and eliminate one failure, during which the equipment is in forced downtime. The recovery time of wind turbines or photovoltaic panels depends on many factors, the main of which is the nature of the failure, the availability of technical diagnostics, as well as the level of qualifications of maintenance and repair personnel. Complex indicators characterizing the reliability of the operation of autonomous generation systems (AGS) based on renewable energy sources (RES) in this work include the forced (emergency) downtime coefficient (q_d) and the availability coefficient (K_a). The forced downtime coefficient is the probability that a system (or its element) will be inoperable at a randomly selected point in time between scheduled repairs. Forced downtime includes periods necessary to detect and eliminate failures, restart the system, as well as downtime caused by the lack of spare parts and maintenance. In this case, downtime associated with scheduled repairs and maintenance is not taken into account [10-12].

III. Problem solution

Analyzing the reliability of autonomous generation systems based on renewable energy sources is a complex practical problem, the solution of which has been the subject of a significant amount of research. Let's consider various methods for assessing the reliability of autonomous generation systems based on renewable energy sources: analytical methods, state space method (Markov process theory), Monte Carlo method, fault tree method and state enumeration method. Most studies have focused on assessing the reliability of generation systems based only on wind or solar energy, while significantly less work has been devoted to assessing the reliability of systems

using renewable energy sources in combination with traditional energy sources such as diesel generators.

The analytical method provides information about the probability of failure of generation facilities to cover the predicted load. In most analytical methods, the generation model is presented in the form of a table of powers and associated outage probabilities. The load model is represented by a daily peak load curve or a constant load curve. Also, the analytical method makes it possible to determine the probability of load loss and the expected shortfall in energy supply from an autonomous system based on wind turbines, diesel generators and an energy storage system. An analytical method can also be proposed for assessing the reliability of an autonomous wind turbine, represented as series-connected elements. The overall failure rate is estimated based on statistical data on the reliability of wind turbine elements, taking into account planned repairs and weather conditions. Another analytical method is a method that takes into account the uncertainties associated with solar radiation, wind speed, electricity consumption and shutdown of various generators. Beta and Weibull distributions are used to model solar radiation and wind speed. The study period is divided into several time periods. In practice, there is also the use of an analytical method to determine the probability of failure-free operation of a system consisting of PV arrays with various converter connection schemes [13-16].

The Monte Carlo method involves simulating hourly data on natural energy resources (wind speed, solar radiation) for various locations using time series models created from historical data collected over several years. Figure 3 shows a model with wind data for the summer season. Hourly data on natural energy resources makes it possible to determine the power output of wind turbines and/or solar power plants. The method also includes modeling the state of the battery based on the generation and load time series of the analyzed system. The difficulty in applying the Monte Carlo method is the need to have accurate and detailed data on wind speed or solar radiation for a specific location. For many regions, meteorological data is either unavailable or available in compressed form.

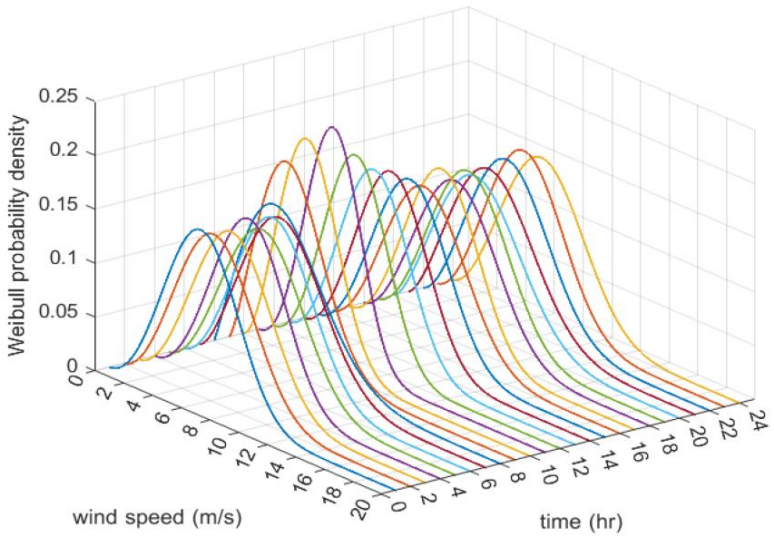


Figure 3: *Model with wind data for the summer season*

To analyze the reliability of renewable energy sources in autonomous generation systems, a state space method (Markov processes) is also proposed. An analysis of the reliability of an autonomous wind farm consisting of two wind turbines shows that all possible transitions in such a system are described by a Markov state graph. To analyze the reliability of an autonomous wind farm, taking into account the stochastic characteristics of wind speed, wind turbine failures and

repairs, a model implemented using Markov chains can be proposed.

The main advantage of the state space method is the clear representation of all states of the system and the transitions between them. Assessing the reliability indicators of technical systems using Markov chains allows one to take into account many factors (partial failures, failures due to common causes, emergency and planned repairs, weather conditions, dependent failures, sequence of failures) affecting the system. However, the method has a number of disadvantages that limit its use: the lack of initial data (the intensity of the transition between states, the probability of states and the duration of stay in each state), the high dimension of the mathematical model when analyzing the reliability of generation systems with a large number of elements.

The fault tree method and state enumeration method are less common for assessing the reliability of autonomous generation systems based on renewable energy sources. The fault tree (FT) method is used to assess the reliability of an autonomous power supply system based on solar cells and batteries, taking into account failures of the inverter and control system. It is also used to assess the reliability of large-scale grid-tied PV systems by taking into account the presence of an energy storage system and a charge controller. The DL method is a powerful tool for assessing the reliability of real technical systems, as it allows you to predict potential failures and improve the reliability of the system at the design stage, visually representing the cause-and-effect relationships between failures and identifying weak points of the system. However, constructing a fault tree requires significant effort, which limits its use in assessing the reliability of autonomous power supply systems based on renewable energy sources.

The state enumeration method is used to quantify the reliability indicators of an autonomous power supply system based on solar cells. The method takes into account the impact of system component failure rates on output power, voltage level and power loss. Each component of a photovoltaic system can be in one of two states: operating or non-operating. The method of enumerating states allows one to take into account changes in the intensity of solar radiation, but is only suitable for systems with the same type of elements [17-20].

Despite the wide variety of methods for analyzing the reliability of AGS using RES, some issues remain unresolved or require more careful study. In particular, the following are not fully taken into account: the influence of weather conditions on the reliable operation of generation systems; emergency failures of generating and additional equipment (inverter, converter, busbars, cable lines), backup equipment and switching equipment. This creates a need and determines the relevance of carrying out research work on the development of scientifically based models and methods for analyzing the reliability of AGS based on RES.

IV. Conclusions

Analyzing the reliability of autonomous generation systems based on renewable energy sources is a complex task that requires taking into account many factors and the use of various methods. The methods discussed, such as analytical methods, state space method, Monte Carlo method, fault tree method and state enumeration method, offer different approaches to assessing the reliability of systems. Each method has its own advantages and disadvantages, which determines their use depending on the specifics of the system being analyzed.

Analytical and Markov process methods can provide accurate estimates when the necessary data are available, whereas the Monte Carlo method requires detailed meteorological data for specific locations. The fault tree method and the state enumeration method are powerful tools for analyzing real technical systems, but require significant effort in their implementation. In general, the use of combined approaches and the development of methods for assessing reliability makes it possible to increase the efficiency and sustainability of autonomous generation systems based on renewable energy sources.

References

[1] Mammadov N., Mukhtarova K. "Analysis of the smart grid system for renewable energy sources", Journal «Universum: technical sciences", № 2 (107), pp. 64-67, 2023
[2] Evans, A.; Strezov, V.; Evans, T.J. Assessment of sustainability indicators for renewable energy technologies. Renew. Sustain. Energy Rev. 2009, 13, 1082–1088
[3] Brellas, K.; Tsikalakis, A.; Kalaitzakis, K. Reducing Solar Dish Park Production Volatility Utilizing Lithium-ion Batteries. Period. Polytech. Electr. Eng. Comput. Sci. 2016, 60, 254–260
[4] Mammadov N.S., Rzayeva S.V., Ganiyeva N.A., "Analisys of synchronized asynchronous generator for a wind electric installation", Przegląd Elektrotechniczny, №5, 2023, pp.37-40, doi:10.15199/48.2023.05.07
[5] Nijat Mammadov, "Analysis of systems and methods of emergency braking of wind turbines". International Science Journal of Engineering & Agriculture, Vol. 2, № 2, pp. 147-152, Ukraine, April 2023
[6] N.M.Piriyeva "Design of electric devices with induction levitation elements", International Journal on "Technical and Physical Problems of Engineering" (IJTPE) Published by International Organization of IOTPE, Vol.14, No.1, pp. 124-129, march 2022.
[7] Singh, C.; Jirutitijaroen, P.; Mitra, J. Electric Power Grid Reliability Evaluation: Models and Methods; John Wiley & Sons: Hoboken, NJ, USA, 2018
[8] Nijat Mammadov, Ilkin Marufov, Saadat Shikhaliyeva, Gulnara Aliyeva, Saida Kerimova, "Research of methods power control of wind turbines", Przeglad elektrotechniczny, R. 100 NR 5/2024, pp. 236-239
[9] N.S. Mammadov, N.A. Ganiyeva, G.A. Aliyeva, "Role of Renewable Energy Sources in the World". Journal of Renewable Energy, Electrical, and Computer Engineering, Vol. 2, №2, pp. 63-67, Indonesia, 30 September 2022
[10] Gusmao, A.; Groissböck, M. Capacity value of photovoltaic and wind power plants in an isolated Mini-grid in the Kingdom of Saudi Arabia. In Proceedings of the Saudi Arabia Smart Grid (SASG), Jeddah, Saudi Arabia, 7–9 December 2015; pp. 1–8
[11] Ilkin Marufov, Aynura Allahverdiyeva, Nijat Mammadov, "Study of application characteristics of cylindrical structure induction levitator in general and vertical axis wind turbines", Przegląd elektrotechniczny, R. 99 NR 10/2023, pp.196-199
[12] Mammadov N.S., "Vibration research in wind turbines", XV International Scientific and Practical Conference «The main directions of the development of scientific research», Helsinki, Finland, 2023, pp. 345-346
[13] N.S. Mammadov, "Methods for improving the energy efficiency of wind turbines at low wind speeds", Vestnik nauki journal, Issue 2, Vol. 61, №4, Russia, April 2023
[14] Ilkin Marufov, Najiba Piriyeva, Nijat Mammadov, Shukufa Ismayilova, "Calculation of induction levitation vertical axis wind generator-turbine system parameters, levitation and influence loop" Przegląd elektrotechniczny – 2024 – No.2 - pp.135-139
[15] Jurasz, J.; Campana, P.E. The potential of photovoltaic systems to reduce energy costs for office buildings in time-dependent and peak-load-dependent tariffs. Sustain. Cities Soc. 2019, 44, 871–879
[16] Stefanakis, J. Crete: An ideal study case for increased wind power penetration in medium sized autonomous power systems. In Proceedings of the IEEE Power Engineering Society Winter Meeting, Conference Proceedings, New York, NY, USA, 27–31 January 2002; Volume 1, pp. 329–334
[17] Mammadov N.S., Aliyeva G.A. "Energy efficiency improving of a wind electric installation using a thyristor switching system for the stator winding of a two-speed asynchronous generator", IJTPE, Issue 55, Volume 55, Number 2 , pp. 285-290, June 2023
[18] Brown, R.E. Electric Power Distribution Reliability; CRC Press: Boca Raton, FL, USA, 2017
[19] Billinton, R.; Allan, R.N. Reliability Assessment of Large Electric Power Systems; Springer Science & Business Media: Berlin, Germany, 2012
[20] I.M. Marufov, N.S. Mammadov, K.M. Mukhtarova, N.A. Ganiyeva, G.A. Aliyeva "Calculation of main parameters of induction levitation device used in vertical axis wind generators". International Journal on technical and Physical Problems of Engineering", Issue 54, Volume 15, Number 1, pp. 184-189, March 2023

RT&A, No 4(80)
Volume 19, December, 2024

A NEW EXTENSION OF KUMARASWAMY DISTRIBUTION FOR IMPROVED DATA MODELING :PROPERTIES AND APPLICATIONS

Mahvish Jan[1], S.P.Ahmad[2]

•

[1,2] Department of Statistics, University of Kashmir, Srinagar, India
wanimahvish@gmail.com, sprvz@yahoo.com

Abstract

In this manuscript, we have introduced a new model of the Kumaraswamy distribution known as SMP Kumaraswamy (SMPK) distribution using SMP technique. The SMPK distribution has the desirable feature of allowing greater flexibility than some of its well-known extensions. A comprehensive account of statistical properties along with the estimation of parameters using classical estimation method is presented. Furthermore, a simulation study is carried out to assess the behavior of estimators based on their biases and mean square errors. Finally, we consider two real-life data sets; we observe that the proposed model outperforms other competing models using goodness of fit measures.

Keywords: Entropy, SMP transformation, Kumaraswamy distribution, Order statistics, Maximum likelihood estimation.

1. Introduction

Probability models offer a decisive role in data analysis, so researchers aim to create novel probability models to handle large data sets in many different domains. Statistical illustration is crucial in real-data studies because novel applications and phenomena is steady, necessitating the continuous construction of probability distributions. Despite the fact that there are many traditional distributions for dealing with data, new distributions are required to overcome inadequacies of these distributions and solve the problems more effectively and precisely. Several techniques of generalizing distributions have been introduced to increase the adaptability of traditional distributions by introducing extra parameters. Suppose the random variable X has the Kumaraswamy distribution with parameters β and λ respectively, then its probability density function (PDF) and cumulative distribution function (CDF) are respectively given by:

$$g(x;\beta,\lambda) = \beta\lambda x^{\lambda-1}(1-x^{\lambda})^{\beta-1}; \quad 0 < x < 1,\ \beta > 0, \lambda > 0 \tag{1}$$

$$G(x;\beta,\lambda) = 1-(1-x^{\lambda})^{\beta}; \quad 0 < x < 1,\ \beta > 0, \lambda > 0 \tag{2}$$

A two-parameter Kumaraswamy distribution for modeling hydrological data was introduced by [9]. Moreover, several new families of probability distributions have been introduced for modeling such type of data based on Kumaraswamy distribution, for example [3] introduced the Kumaraswamy Weibull distribution with application to failure data. A generalization of the Kumaraswamy distribution was proposed by [10] and derived some of its statistical properties and referred to it as the Exponentiated Kumaraswamy distribution and its log-transform. A new distribution using quadratic rank transmutation map was developed by [7] and named

the distribution as Transmuted Kumaraswamy distribution. DUS-Kumaraswamy distribution having same domain as Kumaraswamy distribution introduced by [6] . A new continuous probability density function for a non-negative random variable as an alternative to some bounded domain distributions named Log-Kumaraswamy distribution was introduced by [4] . Kumaraswamy-Gull Alpha Power Rayleigh distribution was proposed by [8] . A generalization of the Exponentiated Kumaraswamy distribution was put forward by [5] and referred to it as the Transmuted Exponentiated Kumaraswamy distribution. A new distribution called Generalized Inverted Kumaraswamy-Rayleigh Distribution was proposed by [11]. A new distribution called the cubic transmuted Log-Logistic distribution was proposed by [15]. An innovative technique for generating probability distributions was proposed by [16] and named it as SMP technique. The CDF and PDF of SMP distribution are respectively given as

$$G_{\text{SMP}}(x) = \begin{cases} \frac{e^{\log \alpha \bar{F}(x)} - \alpha}{1-\alpha}; & \alpha \neq 1, \quad \alpha > 0 \\ F(x); & \alpha = 1 \end{cases} \tag{3}$$

$$g_{\text{SMP}}(x) = \begin{cases} \frac{e^{\log \alpha \bar{F}(x)} \log \alpha f(x)}{\alpha - 1}; & \alpha \neq 1, \quad \alpha > 0 \\ f(x); & \alpha = 1 \end{cases} \tag{4}$$

where $\bar{F}(x) = 1 - F(x)$ and for x $\in$ R, F(x) is the CDF and f(x) is the PDF of the distribution to be extended.
$G_{SMP(x)}$ is a valid CDF. It satisfies the following properties:

1. $G_{SMP}(-\infty) = 0$; $G_{SMP}(\infty) = 1$
2. $G_{SMP}(x)$is monotonic increasing function of x.
3. $G_{SMP}(x)$ is right continuous.
4. $0 \leq G_{SMP}(x) \leq 1$

In the present manuscript, we investigate a novel extension of Kumaraswamy distribution using SMP method. The proposed distribution is named as SMP Kumaraswamy (SMPK) distribution. The primary rationale for contemplating SMPK distribution may be summarized as follows:

- The extension will involve the incorporation of additional parameters to capture more complex data patterns.
- This work will include theoretical derivations, properties of the new distributions, and comparisons with the existing models.
- The proposed model will offer greater flexibility and provide better fit than the other competing models.
- The proposed model offers more flexible shapes of hazard and density plots.
- The proposed model can be used to model various datasets.
- The model considers classical estimation methodologies for parameter estimation.
- Applications of the extended distributions in real-world data analysis will be demonstrated to highlight their practical utility.

The rest of the paper is presented as follows. Section 2 introduces the SMP Kumaraswamy distribution; Section 3 and 4 unfolds reliability analysis and statistical properties, while section 5 focuses on the estimation of unknown parameters using the maximum likelihood approach for the proposed model. Section 6 and 7 presents simulation study and real-life applications. The conclusion of the study is given in Section 8.

2. Smpk Distribution

The CDF and PDF of the proposed SMPK distribution is given by

$$G_{SMPK}(x;\alpha,\beta,\lambda) = \begin{cases} \frac{e^{\log\alpha(1-x^{\lambda})^{\beta}}-\alpha}{1-\alpha}; & \alpha \neq 1, \alpha > 0 \\ 1-(1-x^{\lambda})^{\beta}; & \alpha = 1 \end{cases} \tag{5}$$

$$g_{SMPK}(x;\alpha,\beta,\lambda) = \begin{cases} \frac{\lambda\beta\log\alpha x^{\lambda-1}(1-x^{\lambda})^{\beta-1}e^{\log\alpha(1-x^{\lambda})^{\beta}}}{\alpha-1}; & \alpha \neq 0, \alpha > 0 \\ \lambda\beta x^{\lambda-1}(1-x^{\lambda})^{\beta-1}; & \alpha = 1 \end{cases} \tag{6}$$

The density function plots of SMPK distribution for different combinations of parameters are presented in figure 1. From these plots it is evident that the proposed distribution is unimodal, symmetric, and negatively skewed.

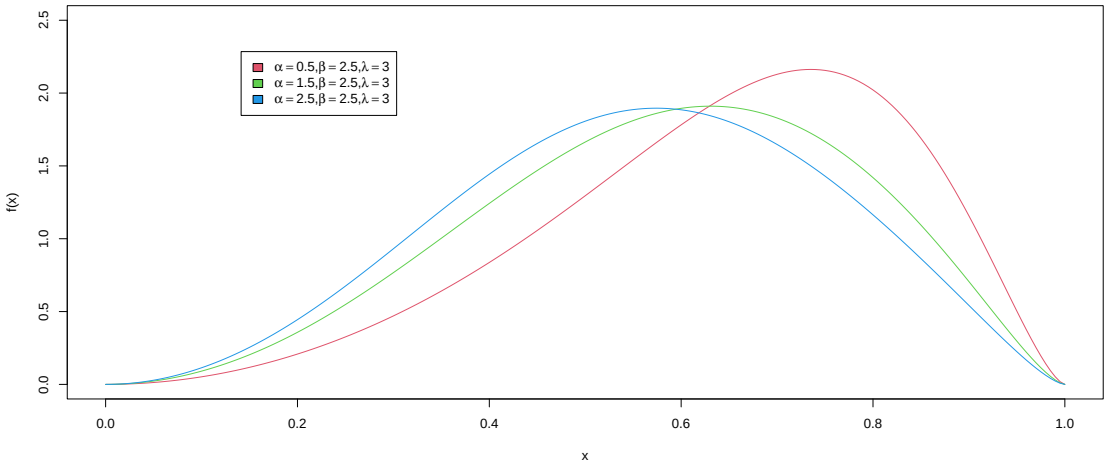


Figure 1: *Plots of the PDF of SMPK distribution.*

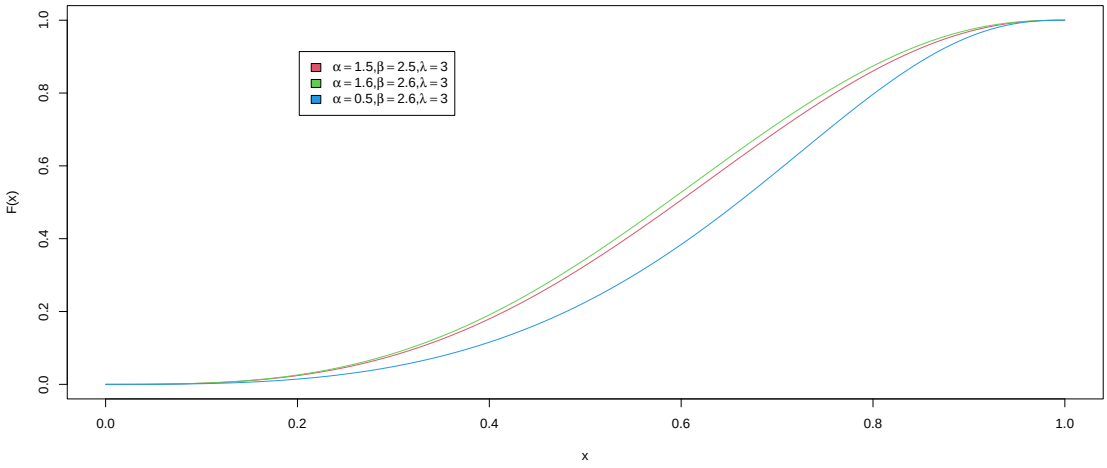


Figure 2: *Plots of the CDF of SMPK DISTRIBUTION.*

3. Reliability Analysis Of the SMPKumaraswamy Distribution

This section focuses on reliability analysis of the SMPK distribution.

3.1. Survival Function

The survival function for the SMPK distribution is given as

$$R(x;\alpha,\beta,\lambda) = 1 - G(x;\alpha,\beta,\lambda) = \frac{1 - e^{\log\alpha(1-x^{\lambda})^{\beta}}}{1-\alpha}; \quad \alpha \neq 1$$

3.2. Hazard Rate

The expression for the hazard rate of the SMPK distribution is obtained as

$$h(x;\alpha,\beta,\lambda) = \frac{g(x;\alpha,\beta,\lambda)}{R(x;\alpha,\beta,\lambda)} = \frac{\log\alpha\lambda\beta x^{\lambda-1}(1-x^{\lambda})^{\beta-1}e^{\log\alpha(1-x^{\lambda})^{\beta}}}{e^{\log\alpha(1-x^{\lambda})^{\beta}} - 1}, \quad \alpha \neq 1$$

From figure 3 it is clear that model has varying shapes like constant, decreasing, increasing and J-shaped for different values of parameters. Accordingly, the proposed model can be used to model datasets with such failure rates.

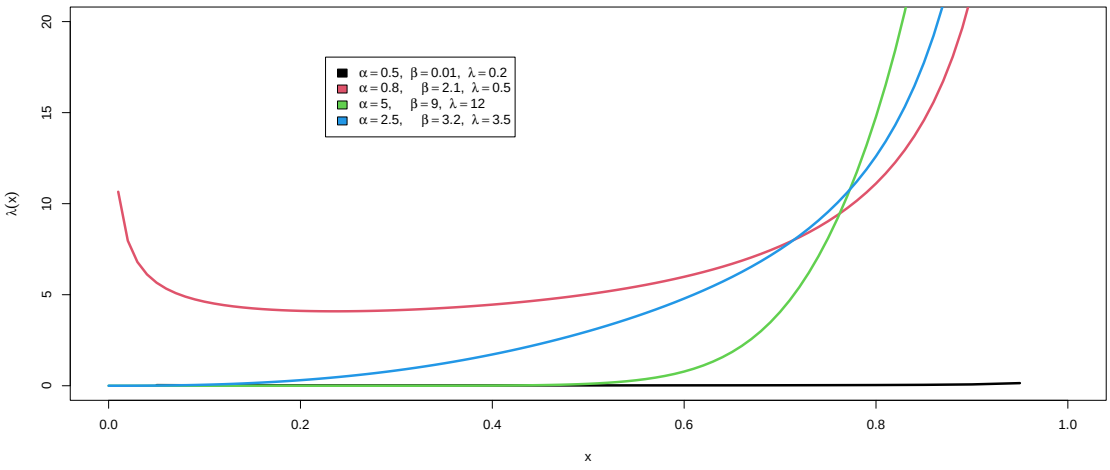


Figure 3: *Plots of the hazard rate of SMPK distribution.*

3.3. Reverse Hazard Function

The reverse hazard rate is defined as the ratio of the probability density function and the corresponding distribution function. It is given as

$$h_r(x;\alpha,\beta,\lambda) = \frac{g(x;\alpha,\beta,\lambda)}{G(x;\alpha,\beta,\lambda)} = \frac{\log\alpha\lambda\beta x^{\lambda-1}(1-x^{\lambda})^{\beta-1}e^{\log\alpha(1-x^{\lambda})^{\beta}}}{\alpha - e^{\log\alpha(1-x^{\lambda})^{\beta}}}, \quad \alpha \neq 1$$

4. Statistical Properties of the SMPK Distribution

In this section, some important statistical properties of the SMPK distribution are presented.

4.1. Quantile function

Theorem 1: If X $\sim$ $SMPK$ (α, β, λ) distribution, then the quantile function of X is given as

$$x = \left[1 - \left(\frac{\log\left[u(1-\alpha)+\alpha\right]}{\log\alpha}\right)^{\frac{1}{\beta}}\right]^{\frac{1}{\lambda}} \tag{7}$$

Where U is a uniform random variable, $0 < u < 1$.
Proof: Let $G(x;\alpha,\beta,\lambda)$ = u

$$\frac{e^{\log\alpha(1-x^{\lambda})^{\beta}} - \alpha}{1-\alpha} = u$$

$$x = \left[1 - \left(\frac{\log\left[u(1-\alpha)+\alpha\right]}{\log\alpha}\right)^{\frac{1}{\beta}}\right]^{\frac{1}{\lambda}}$$

Remark:
The p^{th} quantile is given by

$$x_p = \left[1 - \left(\frac{\log\left[p(1-\alpha)+\alpha\right]}{\log\alpha}\right)^{\frac{1}{\beta}}\right]^{\frac{1}{\lambda}}$$

4.2. Moments

The r^{th} moment about origin of a random variable X having the SMPK distribution is obtained as

$$\mu_{r'} = \int_0^1 x^r \frac{e^{\log\alpha\left(1-x^{\lambda}\right)^{\beta}} \log\alpha\lambda\beta x^{\lambda-1}\left(1-x^{\lambda}\right)^{\beta-1}}{\alpha-1}\,dx$$

Using the expansion,

$$e^x = \sum_{j=0}^{\infty}\frac{x^j}{j!}$$

$$\mu_r' = \frac{1}{\alpha-1}\lambda\beta\sum_{j=0}^{\infty}\frac{(\log\alpha)^{j+1}}{j!}\int_0^1 x^r x^{\lambda-1}(1-x^{\lambda})^{\beta(j+1)-1}\,dx$$

$$\mu_r' = \frac{\beta}{\alpha-1}\sum_{j=0}^{\infty}\frac{(\log\alpha)^{j+1}}{j!}B\left(\frac{r}{\lambda}+1,\beta(j+1)\right) \tag{8}$$

$B[(\frac{r}{\lambda}+1),\beta(j+1)]$ represents the beta function. Substituting r =1, 2, 3, 4 the first four moments about origin of the SMPK distribution are obtained.

Lemma 1. Suppose a random variable $X \sim SMPK(\alpha,\beta,\lambda)$ distribution with PDF given in Eq. (6) and let $I_r(t) = \int_0^t g_{\text{SMPK}}(x;\alpha,\beta,\lambda)\,dx$ denotes the r^{th} incomplete moment, then we have

$$I_r(t) = \frac{\beta}{\alpha-1}\sum_{j=0}^{\infty}\frac{(\log\alpha)^{j+1}}{j!}B\left(t^{\lambda};\frac{r}{\lambda}+1,\beta(j+1)\right) \tag{9}$$

Proof:

$$I_r(t) = \int_0^t x^r\, g_{\text{smpk}}(x;\alpha,\beta,\lambda)\,dx$$

$$I_r(t) = \int_0^t x^r \frac{e^{\log\alpha(1-x^{\lambda})^{\beta}}\log\alpha\lambda\beta x^{\lambda-1}(1-x^{\lambda})^{\beta-1}}{\alpha-1}\,dx$$

$$I_r(t) = \frac{\beta}{\alpha - 1} \sum_{j=0}^{\infty} \frac{(\log \alpha)^{j+1}}{j!} B\left(t^{\lambda}; \frac{r}{\lambda} + 1, \beta(j+1)\right)$$

Where $B[z; a, b] = \int_0^a x^{b-1}(1-x)^{c-1}\, dx$ is incomplete beta function, setting r=1 in Eq.(9) will yield first incomplete moment

$$I_1(t) = \frac{\beta}{\alpha - 1} \sum_{j=0}^{\infty} \frac{(\log \alpha)^{j+1}}{j!} B\left(t^{\lambda}; \frac{1}{\lambda} + 1, \beta(j+1)\right)$$

4.3. Mean Residual life

Mean residual life is the expected remaining life given that a component has survived up to time t and is given by

$$\mu(t) = \frac{1}{R(t)} \left(E(x) - \int_0^t x g_{\text{SMPK}}(x; \alpha, \beta, \lambda)\, dx\right) - t$$

Where

$$E(x) = \frac{\beta}{\alpha - 1} \sum_{j=0}^{\infty} \frac{(\log \alpha)^{j+1}}{j!} B\left(\frac{1}{\lambda} + 1, \beta(j+1)\right)$$

$$\int_0^t x g(x; \alpha, \beta, \lambda)\, dx = \frac{\beta}{\alpha - 1} \sum_{j=0}^{\infty} \frac{(\log \alpha)^{j+1}}{j!} B\left(t^{\lambda}; \frac{1}{\lambda} + 1, \beta(j+1)\right)$$

$$\mu(t) = \frac{1 - \alpha}{1 - e^{\log \alpha \cdot (1 - x^{\lambda})^{\beta}}} \left\{ \frac{\beta}{\alpha - 1} \sum_{j=0}^{\infty} \frac{(\log \alpha)^{j+1}}{j!} \left[B\left(\frac{1}{\lambda} + 1, \beta(j+1)\right) - B\left(t^{\lambda}; \frac{1}{\lambda} + 1, \beta(j+1)\right) \right] \right\} - t$$

4.4. Mean Waiting Time

Mean waiting time is the time elapsed since the failure of an item given that the item has failed in [0, t] and is given by

$$\bar{\mu}(t) = t - \frac{1}{G(t)} \int_0^t x\, g_{\text{SMPK}}(x; \alpha, \beta, \lambda)\, dx$$

$$\bar{\mu}(t) = t - \frac{\beta}{\alpha - e^{\log \alpha (1 - t^{\lambda})^{\beta}}} \sum_{j=0}^{\infty} \frac{(\log \alpha)^{j+1}}{j!} B\left(t^{\lambda}; \frac{1}{\lambda} + 1, \beta(j+1)\right)$$

4.5. Renyi Entropy

The Renyi entropy was introduced by [17] in 1960 and is expressed as

$$I_{\delta}(x) = \frac{1}{1 - \delta} \log \int_{-\infty}^{\infty} g^{\delta}(x)\, dx \quad \delta > 0, \delta \neq 1$$

$$I_{\delta}(x) = \frac{1}{1 - \delta} \log \left\{ \frac{1}{(\alpha - 1)^{\delta}} \sum_{j=0}^{\infty} \frac{\delta^{j} (\log \alpha)^{j+\delta}}{j!} \beta^{\delta} \lambda^{\delta - 1} B\left[\left(1 - \frac{1}{\lambda}\right)(\delta - 1) + 1, \beta(j + \delta) - \delta + 1\right] \right\}$$

Remark: Shannon entropy is a special case of Renyi entropy for $\delta = 1$

4.6. Harvda & Charvat Entropy

The Harvard & Charvat entropy of a random variable X is defined by

$$I_{\delta}(x) = \frac{1}{1 - \delta} \left[1 - \int_0^{\infty} g^{\delta}(x)\, dx\right] \quad \text{where } \delta > 0,\ \delta \neq 1$$

$$I_{\delta}(x) = \frac{1}{1 - \delta} \left[1 - \frac{1}{(\alpha - 1)^{\delta}} \sum_{j=0}^{\infty} \frac{\delta^{j} (\log(\alpha))^{j+\delta}}{j!} \beta^{\delta} \lambda^{\delta - 1} B\left(\left(1 - \frac{1}{\lambda}\right)((\delta - 1) + 1), \beta(j + \delta) - \delta + 1\right)\right]$$

4.7. Moment Generating Function

The moments of distribution are represented by the moment generating function (MGF). The following theorem provides the MGF for the SMPK distribution.

Theorem 2: Let X follows SMPK distribution, then the moment generating function $M_x(t)$ is

$$M_x(t) = \frac{\beta}{\alpha - 1} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \frac{t^r \log(\alpha)^{j+1}}{j!r!} B\left(\frac{r}{\lambda} + 1, \beta(j+1)\right) \tag{10}$$

Proof: The moment generating function of SMPK distribution can be obtained using the relation

$$M_x(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r' \tag{11}$$

Using Eq. (8) in Eq. (11) and after necessary calculations, we get

$$M_x(t) = \frac{\beta}{\alpha - 1} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \frac{t^r \log(\alpha)^{j+1}}{j!r!} B\left(\frac{r}{\lambda} + 1, \beta(j+1)\right)$$

4.8. Order Statistics

Theorem 3: The PDF of the general order statistics of SMPK distribution is given by

$$g_{(t:n)}(x) = \frac{n!}{(t-1)!(n-t)!} \left[\frac{e^{\log\alpha\left(1-x^{\lambda}\right)^{\beta}} \log\alpha\, \lambda\beta x^{\lambda-1}(1-x^{\lambda})^{\beta-1}}{\alpha - 1}\right] \times \sum_{j=0}^{\infty} (-1)^j \binom{n-t}{j} \left[\frac{e^{\log\alpha(1-x^{\lambda})^{\beta}} - \alpha}{1-\alpha}\right]^{j+t-1}$$

Proof: Let $x_{(1)}x_{(2)}, ..., x_{(n)}$ be the order statistics of a random sample derived from SMPK distribution. Then, the PDF of t^{th} order statistics is given by

$$g_{(t:n)}(x) = \frac{n!}{(t-1)!(n-t)!} \left[G(x;\alpha,\beta,\lambda)\right]^{t-1} \left[1 - G(x;\alpha,\beta,\lambda)\right]^{n-t} g(x;\alpha,\beta,\lambda) \tag{12}$$

Prior to incorporating Eq.(5) and Eq.(6) in Eq.(12), we use binomial expansion of $[1 - G(x;\alpha,\beta,\lambda)]^{n-t}$ as

$$[1 - G(x;\alpha,\beta,\lambda)]^{n-t} = \sum_{j=0}^{\infty} (-1)^j \binom{n-t}{j} [G(x;\alpha,\beta,\lambda)]^j$$

Thus, we obtain

$$g_{(t:n)}(x) = \frac{n!}{(t-1)!(n-t)!} g(x;\alpha,\beta,\lambda) \sum_{j=0}^{\infty} (-1)^j \binom{n-t}{j} [G(x;\alpha,\beta,\lambda)]^{j+t-1}$$

$$g_{(t:n)}(x) = \frac{n!}{(t-1)!(n-t)!} \left[\frac{e^{\log\alpha\left(1-x^{\lambda}\right)^{\beta}} \log\alpha\, \lambda\beta x^{\lambda-1}(1-x^{\lambda})^{\beta-1}}{\alpha - 1}\right] \times \sum_{j=0}^{\infty} (-1)^j \binom{n-t}{j} \left[\frac{e^{\log\alpha(1-x^{\lambda})^{\beta}} - \alpha}{1-\alpha}\right]^{j+t-1} \tag{13}$$

The expression for PDF of minimum order statistics $x_{(1)}$ and maximum order statistics $x_{(n)}$ of SMPK distribution are respectively obtained by setting t=1 and t=n in Eq. (13).

5. Estimation of Parameters

We assume $x_1, x_2, x_3, ...x_n$ is a random sample of n observations drawn from the SMPK distribution with unknown parameters α, β, λ. The likelihood and log-likelihood functions are respectively given as

$$L(x;\alpha,\beta,\lambda) = \frac{e^{\log\alpha\sum_{i=1}^{n}\left(1-x_i^{\lambda}\right)^{\beta}}\ (\log\alpha)^n\,\lambda^n\beta^n\prod_{i=1}^{n}x_i^{\lambda-1}\left(1-x_i^{\lambda}\right)^{\beta-1}}{(\alpha-1)^n}$$

$$l = n\log\left[\frac{\log\alpha\,\lambda\beta}{\alpha-1}\right] + \log\alpha\sum_{i=1}^{n}(1-x_i^{\lambda})^{\beta} + \sum_{i=1}^{n}\left[(\lambda-1)\log x_i + (\beta-1)\log\left(1-x_i^{\lambda}\right)\right] - n\log(\alpha-1)$$

$$\frac{\partial l}{\partial\alpha} = n\left[\frac{\log\left(\log\alpha\,\lambda\beta\right)\alpha\log\alpha-(\alpha+1)}{\alpha\log\alpha(\alpha-1)^2}\right] + \frac{\sum_{i=1}^{n}\left(1-x_i^{\lambda}\right)^{\beta}}{\alpha} = 0$$

$$\frac{\partial l}{\partial\beta} = \frac{n}{\alpha-1} + \sum_{i=1}^{n}\log\left(1-x_i^{\lambda}\right)\left[1+\log\alpha\,\left(1-x_i^{\lambda}\right)^{\beta}\right] = 0$$

$$\frac{\partial l}{\partial\lambda} = \frac{n}{\lambda(\alpha-1)} + \sum_{i=1}^{n}\log x_i\left[1-\beta\log\alpha\,\left(1-x_i^{\lambda}\right)^{\beta-1}x_i^{\lambda} - (\beta-1)\frac{x_i^{\lambda}}{1-x_i^{\lambda}}\right] = 0$$

Since, above equations are non-linear, we will use Newton-Raphson method and hence R software to solve these equations and estimate the parameters.

6. Simulation Study

To assess the performance of the proposed estimation method for the parameters of the SMPK distribution, we conducted a simulation study. Table 1 shows the true parameter values α, β, λ fixed at (0.5, 0.20, 0.25) & (1.5, 0.3, 0.4) respectively. Using R software, samples of sizes 20, 50, 125 and 500 were randomly generated based on the quantile function from Eq. (7), with each scenario replicated 1000 times. For each parameter combination, we computed the MLEs along with their corresponding bias and mean squared errors (MSEs). The results are summarized in Table 1.

Table 1: *MLE, Bias, and MSE for the parameters*

Sample size	Parameters			MLE			Bias			MSE		
n	α	β	λ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
20	0.5	0.20	0.25	1.552	0.217	0.411	1.337	0.058	0.241	28.782	0.006	0.357
50				0.893	0.201	0.295	0.599	0.037	0.106	2.219	0.003	0.025
125				0.585	0.200	0.260	0.234	0.026	0.059	0.161	0.003	0.008
500				0.514	0.202	0.249	0.085	0.016	0.024	0.026	0.004	0.002
20	1.5	0.3	0.4	1.000	0.240	0.292	0.500	0.060	0.108	3.781	0.016	0.042
50				1.000	0.248	0.333	0.500	0.052	0.067	1.459	0.008	0.018
125				1.080	0.259	0.378	0.428	0.041	0.022	0.865	0.004	0.008
500				1.489	0.295	0.378	0.010	0.005	0.004	0.356	0.001	0.002

From Table 1 it is clear that the MLEs exhibit stability and closely approximate the true parameter values. As the sample size increases across all parameter combinations, the MSE decreases, indicating enhanced precision in the estimation of model parameters. Additionally, the bias of all parameters consistently decreases with larger sample sizes, demonstrating improved accuracy of the estimation method.

7. Real Life Applications

In this section, we demonstrate the practical applicability of the SMPK distribution with two real life data sets. The potentiality of the proposed model is determined by comparing its performance with several other models, namely Transmuted Kumaraswamy distribution [7], Kumaraswamy Inverse Exponential distribution [14], Weighted Kumaraswamy distribution [1] and Kumaraswamy distribution [9] using goodness-of-fit criterions including -2ll, Akaike Information Criterion (AIC), Akaike Information Criterion Corrected (AICC), Hannan - Quinn information criterion (HQIC) , Kolmogorov-Smirnov (KS) and P value statistics. The distribution with the lowest value of -2ll, AIC, AICC, HQIC, K-S and maximum P value is considered the best fit.

Application 1:Snowfall data

The data set relates to the daily snowfall amounts of 30 observations measured in inches of water, conducted in the vicinity of Climax by [12].

Application 2:Milk Production data

The data set shows the measurements of the proportion of total milk production in the first birth of 107 SINDI cows studied by [13]. The data has been previously studied by [2].

Table 2: *Estimates, -2ll, AIC, AICC, HQIC, K-S statistic, and P-value for Dataset 1.*

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{c}$	-2ll	AIC	AICC	HQIC	K-S	P-value
SMPK	21.014	4.641	1.0648	-	-	-81.891	-75.891	-74.969	-156.437	0.081	0.991
TKD	0.945	-	0.614	5.819	-	-80.794	-74.799	-73.871	-154.244	0.602	0.073
KIED	0.093	0.952	0.294	-	-	-79.227	-73.227	-72.304	-151.110	0.188	0.240
WKD	-	7.810	1.001	-	0.8561	-79.118	-73.118	-72.196	-150.891	0.172	0.002
KUMD	0.861	6.8361	-	-	-	-79.595	-75.595	-75.151	-154.294	0.121	0.774

Table 3: *Estimates, -2ll, AIC, AICC, HQIC, K-S statistic, and P-value for Dataset 2.*

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{c}$	-2ll	AIC	AICC	HQIC	K-S	P-value
SMPK	0.066	3.471	1.439	-	-	-56.842	-50.842	-50.609	-104.434	0.047	0.969
TKD	1.823	-	-0.561	3.436	-	-54.097	-48.098	-47.865	-98.945	0.060	0.836
KIED	0.574	2.256	0.826	-	-	75.946	81.946	82.179	161.143	0.261	0.080
WKD	-	3.931	3.048	-	0.001	-52.816	-46.816	-46.582	-96.380	62.586	0.002
KUMD	2.195	3.436	-	-	-	-50.789	-46.789	-46.674	-95.411	0.076	0.562

From Table 2 and Table 3 it is observed that SMPK distribution has least numerical value of all the comparison criterions and hence fits better to the real dataset as compared to other competing models.The plots of the fitted models are shown in figures 4 and 5.These plots also demonstrate that the SMPK distribution offers a close fit to both data sets.Additionally,Q-Q plots for the two data sets are also given in figures 6 and 7.Furthermore,we extract the shape of the hazard function from the observed data using the total time on test (TTT) plots given in Figures 8 and 9.The TTT plots for the data sets indicate that the data sets decreasing ,increasing hazard rate.

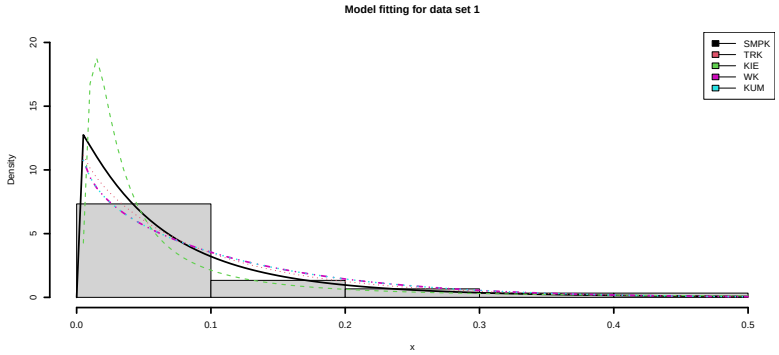


Figure 4: *Fitted density plots for dataset 1*

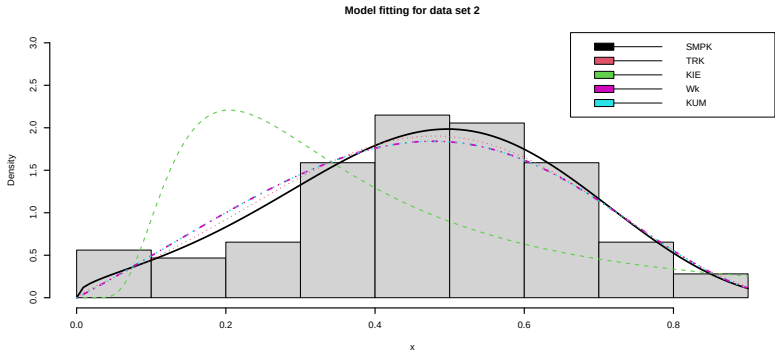


Figure 5: *Fitted density plots for dataset 2*

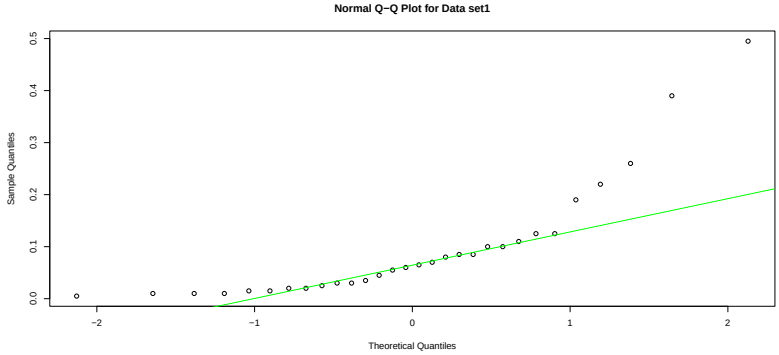


Figure 6: *Q-Q Plot for dataset 1*

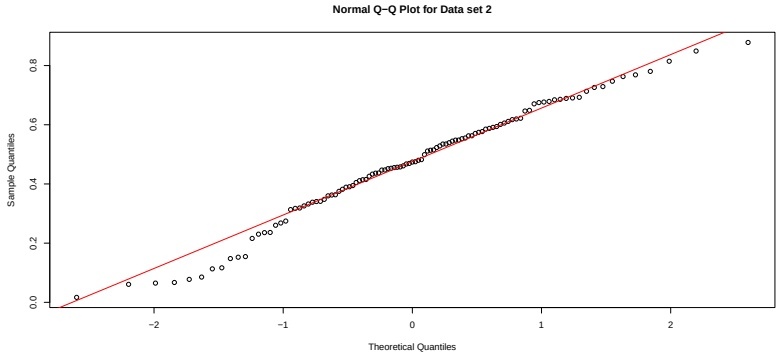


Figure 7: *Q-Q Plot for dataset 2*

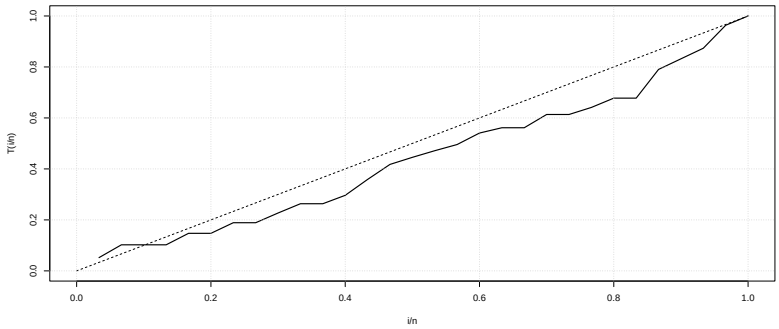


Figure 8: *TTT Plot for data set 1*

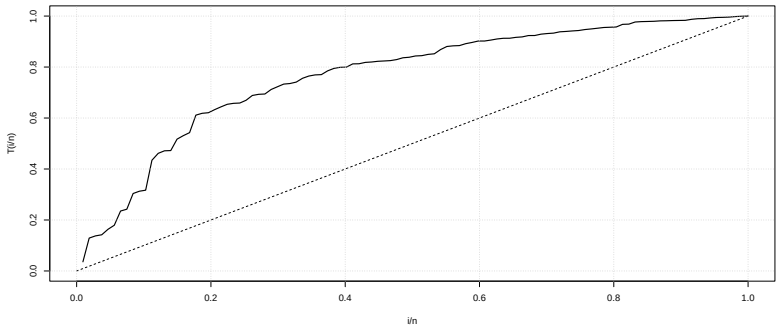


Figure 9: *TTT Plot for data set 2*

8. Conclusion

This study introduces the SMPK distribution as a new extension of the Kumaraswamy distribution, using the SMP approach. We have examined several statistical characteristics of the proposed model, including the survival function, hazard rate function, reverse hazard function, moments, quantile function, mean residual life, mean waiting time, Renyi entropy, Harvda & Charvat entropy, moment generating function and order statistics The parameter estimation is performed using the maximum likelihood estimation method. A simulation study is performed to evaluate

the performance of the maximum likelihood estimator (MLE) in estimating the parameters. The model's performance is evaluated using goodness-of-fit statistics. The proposed distribution is unimodal, symmetric and negatively skewed. Additionally, it displays constant, declining, rising, and J-shaped failure rates across various parameter values. Accordingly, the proposed distribution can be used to model datasets with similar failure rates. For practical applicability, the proposed distribution is applied to two real life datasets and it suggested that the SMPK model outperforms and provides a better fit than the Competitive models.

References

[1] M. Abd El-Monsef and S. Ghoneim. The weighted kumaraswamy distribution. *Information*, 18(8):3289–3300, 2015.

[2] A. Bhat, S. P. Ahmad, E. M. Almetwally, N. Yehia, N. Alsadat, and A. H. Tolba. The odd lindley power rayleigh distribution: properties, classical and bayesian estimation with applications. *Scientific African*, 20:e01736, 2023.

[3] G. M. Cordeiro, E. M. Ortega, and S. Nadarajah. The kumaraswamy weibull distribution with application to failure data. *Journal of the Franklin institute*, 347(8):1399–1429, 2010.

[4] A. I. Ishaq, A. A. Suleiman, H. Daud, N. S. S. Singh, M. Othman, R. Sokkalingam, P. Wiratchotisatian, A. G. Usman, and S. I. Abba. Log-kumaraswamy distribution: its features and applications. *Frontiers in Applied Mathematics and Statistics*, 9:1258961, 2023.

[5] J. Joseph and M. Ravindran. Transmuted exponentiated kumaraswamy distribution. *Reliability: Theory & Applications*, 18(1 (72)):539–552, 2023.

[6] K. Karakaya, İ. Kınacı, C. Kuş, and Y. Akdoğan. On the dus-kumaraswamy distribution. *Istatistik Journal of The Turkish Statistical Association*, 13(1):29–38, 2021.

[7] M. S. Khan, R. King, and I. L. Hudson. Transmuted kumaraswamy distribution. *Statistics in Transition new series*, 2(17):183–210, 2016.

[8] M. Kpangay, L. O. Odongo, and G. O. Orwa. The kumaraswamy-gull alpha power rayleigh distribution: properties and application to hiv/aids data. *International Journal of Scientific Research and Engineering Development*, 6(1):431–442, 2023.

[9] P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of hydrology*, 46(1-2):79–88, 1980.

[10] A. J. Lemonte, W. Barreto-Souza, and G. M. Cordeiro. The exponentiated kumaraswamy distribution and its log-transform. *Brazilian Journal of Probability and Statistics*, 27(1):31–53, 2013.

[11] A. S. Malik and S. P. Ahmad. Generalized inverted kumaraswamy-rayleigh distribution: Properties and application. *Journal of Modern Applied Statistical Methods*, 23, 2024.

[12] P. W. Mielke Jr, L. O. Grant, and C. F. Chappell. An independent replication of the climax wintertime orographic cloud seeding experiment. *Journal of Applied Meteorology (1962-1982)*, pages 1198–1212, 1971.

[13] G. Moutinho Cordeiro and R. dos Santos Brito. The beta power distribution. *Brazilian Journal of Probability and Statistics*, pages 88–112, 2012.

[14] P. Oguntunde, O. Babatunde, and A. Ogunmola. Theoretical analysis of the kumaraswamy-inverse exponential distribution. *International Journal of Statistics and Applications*, 4(2):113–116, 2014.

[15] M. M. Rahman, J. A. Darwish, S. H. Shahbaz, G. Hamedani, and M. Q. Shahbaz. A new cubic transmuted log-logistic distribution: Properties, applications, and characterizations. *Advances and Applications in Statistics*, 91(3):335–361, 2024.

[16] S. U. Rasool, M. A. Lone, and S. P. Ahmad. An innovative technique for generating probability distributions: A study on lomax distribution with applications in medical and engineering fields. *Annals of Data Science*, pages 1–17, 2024.

[17] A. Rényi. On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, volume 4, pages 547–562. University of California Press, 1961.

RT&A, No 4(80)
Volume 19, December, 2024

STOCHASTIC OPTIMIZATION OF PERISHABLE INVENTORY INCORPORATING PRESERVATION, FRESHNESS INDEX, EXPIRY DATE AND OPTIMISING PROMOTIONAL STRATEGIES FOR EFFECTIVE MANAGEMENT

KAJAL SHARMA[1], LALJI KUMAR[2], UTTAM KUMAR KHEDLEKAR[3]

•

[1,2,3]Department of Mathematics and Statistics,
Dr. Harisingh Gour Vishwavidyalaya, Sagar, Madhya Pradesh, India - 470003
[1]*kajal100136@gmail.com*,[2]*laljikumar07@gmail.com*,[3]*uvkkcm@yahoo.co.in*

Abstract

Inventory management is a critical aspect of supply chain efficiency and can be influenced by various factors such as advertising, pricing, and preservation policies. Recent research has proposed a model that considers critical variables such as fluctuations in pricing, advertising tactics, and preservation expenses within uncertain scenarios to improve inventory management. The study provides valuable insights into advertising dynamics, optimal pricing strategies, and the impact of preservation costs on decision-making. Decision-makers can apply these insights to enhance the efficiency of their supply chains in a competitive environment. The study emphasizes the importance of flexibility while aligning inventory practices with corporate sustainability goals. Although the model's applicability may be context-specific, the findings contribute to discussions on inventory management strategies while acknowledging certain assumptions made during the study. Proper advertising, pricing, and preservation policies can increase awareness, attract customers, and maintain quality, influencing product demand. This research proposes a model to improve inventory management, considering variables such as pricing fluctuations, advertising tactics, and preservation expenses in uncertain scenarios. The study provides insights into advertising dynamics, optimal pricing strategies, and how preservation costs influence decision-making. Decision-makers can apply these insights to improve supply chain efficiency. The study stresses the importance of flexibility in a competitive environment and aligning inventory practices with corporate sustainability goals. The findings contribute to discussions on inventory management strategies, but the model's applicability may be context-specific, and the study makes certain assumptions.

Keywords: Inventory, Deterioration, Preservation, Stochastic optimisation, Advertising

1. Introduction

The issue of demand uncertainty is a challenging problem in inventory control that makes it difficult to accurately predict market dynamics and consumer behaviour. This uncertainty arises due to several factors, including evolving consumer preferences influenced by societal changes and trends and market dynamics such as economic conditions and currency fluctuations, which compound the complexity of demand forecasting. For example, during a seasonal sale, a sudden and unpredictable surge in demand for a popular product may occur due to external factors such as economic incentives and promotional activities, in addition to consumer behaviour.

The empirical evaluation section of this paper focuses on revealing the mathematical behaviour and patterns of demand uncertainty by closely examining real-world data to identify trends and deviations. The paper associates these findings with probability distributions, providing businesses with a framework to respond to the uncertainty of market dynamics. It is characterised by Stochastic dynamics, which are characterized by unpredictability and randomness, play a critical role in understanding the dynamics of modern markets. The study aims to fill a gap in existing knowledge by recognizing the stochastic nature of demand and seamlessly incorporating it into the decision-making process.

Traditional models may need to pay more attention to the impact of preservation costs on overall profitability. This research acknowledges preservation costs as a critical factor and efforts to provide decision-makers with insights into optimizing preservation strategies. By recognizing the insubstantial balance between the need for timely replenishment and the financial implications of preserving inventory, businesses can make informed decisions that align with their overarching goals. The element of expiry dates introduces a time-sensitive dimension to inventory management. In industries where products have a limited shelf life, managing stock effectively to avoid losses due to expiry is essential. By doing so, businesses can align their stocking strategies with product lifecycles, minimizing waste and maximizing profitability.

This literature review focuses on the replenishment period, where inventory experiences a continuous decrease due to the combined impact of demand and deterioration. The pioneering work has revolutionized inventory modelling by introducing the concept of no-shortage stock level [1]. An extension of this model, using a two-parametric Weibull distribution for variable deterioration rates, has provided insights into different distribution patterns [2] and, in inventory management, emphasized the significance of demand fluctuations [3]. And an inventory system with deteriorating items, challenging the assumption of a constant demand rate [4].

Mostly, models considers deterioration rate is constant over time. However, in reality, deterioration rate can fluctuating over time due to different factors like environmental factors, storage conditions, and other external factors. To address this problem, the EOQ model was developed with time-varying deterioration and inventory-dependent demand [5]. it can also be applied to declining products with shifting demand based on factors like selling price and advertising frequency. Moving forward, the exploration of time-dependent factors in inventory models will be considered by considering holding costs and demand functions that vary over time [6]. This temporal dimension adds a layer of sophistication, acknowledging the dynamic nature of both costs and demand, further enhancing the applicability of inventory models to real-world scenario [7]. A two-phase inventory model is proposed for non-instantaneous decay items with stock-dependent demand before and constant demand during decay and shortage periods [8].

A pricing strategy known as dynamic pricing was developed for products that have price sensitive demand [9]. A seasonal inventory model proposed with preservation technology reveals profit dynamics and determines optimal solutions for selling price, replenishment cycle, and preservation investment across business setups [10]. Also, the model focuses on optimized price and inventory control for items that deteriorate over time, experience partial backordering, and highlight the need for a comprehensive approach to decision-making [11]. The framework proposed by [12] presents strategies for aligning economic and environmental objectives, leading to cost reductions, lower carbon footprints, and improved operational stability. A comparison of deterministic and stochastic inventory modeling approaches reveals that deterministic models may oversimplify demand dynamics, whereas stochastic models, particularly those utilizing Particle Swarm Optimization (PSO), better address market uncertainties. This highlights the value of stochastic methods for enhancing adaptive decision-making, resource allocation, and risk management in complex supply chains. To maximize profits, [13] introduces a novel dynamic pricing approach that optimizes the number of price adjustments during sales seasons based on stochastic factors and real-time demand fluctuations. Further, A model for demand forecasting and dynamic pricing evaluates Holt-Winters Exponential Smoothing (HWES) and Autoregressive Integrated Moving Average (ARIMA) [14]. The study finds that ARIMA models outperform HWES in minimizing lost sales and mitigating stockouts under varying economic conditions.

Table 1: *Contribution table and highlighting the research gap*

Author(s)	Deterioration	Preservation	Freshness	Pr. Opt.	St. Opt.
[16]	Non-Inst	×	×	✓	×
[17]	Non-Inst	×	×	✓	×
[18]	Non-Inst	×	×	×	×
[19]	Inst	×	×	✓	×
[20]	Non-Inst	✓	×	×	✓
[21]	Non-Inst	✓	×	×	×
[22]	Non-Inst	×	×	✓	✓
[23]	Non-Inst	✓	×	✓	×
[24]	Inst.	✓	×	✓	✓
[25]	×	×	×	✓	✓
This Study	Inst	✓	✓	✓	✓

Non-Inst: Non-instantaneous deterioration; Pro. Opt.: Promotional Optimisation; St. Opt.: Stochastic Optimisation

Integrating promotional efforts, preservation technology, proportionate shortage time, and partial backlogging emphasizes the need for a holistic understanding of inventory dynamics in real-time scenarios. In their subsequent work, [15] extended their research to investigate profit optimization in conversion processes such as milk to butter and debenture to share. By employing Particle Swarm Optimization, the study addresses factors like conversion timing, quantity, cost, and duration, offering insights into strategic decision-making for inventory managers. This paper makes a significant contribution to the modelling of a retailer's business by incorporating promotional optimization and stochastic optimization. Our analysis has identified a research gap in this field, emphasizing the need for further exploration. Table 1 summarizes these contributions and gaps.

2. Problem Definition

In the traditional inventory models, which often rely on deterministic assumptions, fail to address these uncertainties effectively. Another issue like deterioration is particularly critical for businesses dealing with perishable goods. Conventional models often overlook the intricate dynamics of deterioration, leading to suboptimal decisions and financial losses. This research addresses this challenge by integrating strategies that account for the perishable nature of goods, providing insights for optimizing stock levels and minimizing financial impact. Preservation costs further complicate inventory management, requiring a balance between timely replenishment and the costs associated with preserving inventory. This study investigates optimal resource allocation strategies, considering preservation costs to enhance decision-making and align with broader business objectives. Promotional efforts also impact inventory dynamics and profitability, yet traditional models often neglect this aspect. The research examines the effect of promotional activities, including advertising frequency, on inventory management and profitability. By understanding the relationship between advertising strategies and inventory dynamics, businesses can refine their promotional efforts for optimal results.

This research aims to develop a comprehensive inventory management model that incorporates stochastic dynamics, addresses deterioration and preservation costs, integrates expiry dates, and optimizes promotional efforts. The objective is to provide decision-makers with actionable strategies to navigate modern market complexities and enhance overall inventory management performance.

3. NOTATIONS

Notations bearing usual tradition, utilized in the subsequent discussions, are laid down as follows:
Parameters:

τ : Maximum life span of the product,

γ : Advertising impact parameter associated with the advertising frequency,

a : Demand potential parameter,

α : Price sensitive parameter of the demand,

b : Time sensitive parameter of the demand,

ϵ : Continuous random variable associated with demand responsible for demand uncertainty,

λ : Preservation impact parameter,

η : Gain fraction of total demand in case of shortage,

C_h : Holding cost per unit per unit of time,

c_1 : Shortage Cost per unit per unit of time,

c_2 : Opportunity cost per unit of time,

C_0 : Set up cost per set up,

W : Backlogged demand,

Q : Total order quantity of without shortage case,

S : Initial stock level in case of shortage,

$f(u)$: Instantaneous freshness index,

Decision variables:

p : Selling price per unit,

T : Total inventory period of the model including with and without shortage cases,

t_1 : Shortage time period,

A : Advertisement frequency,

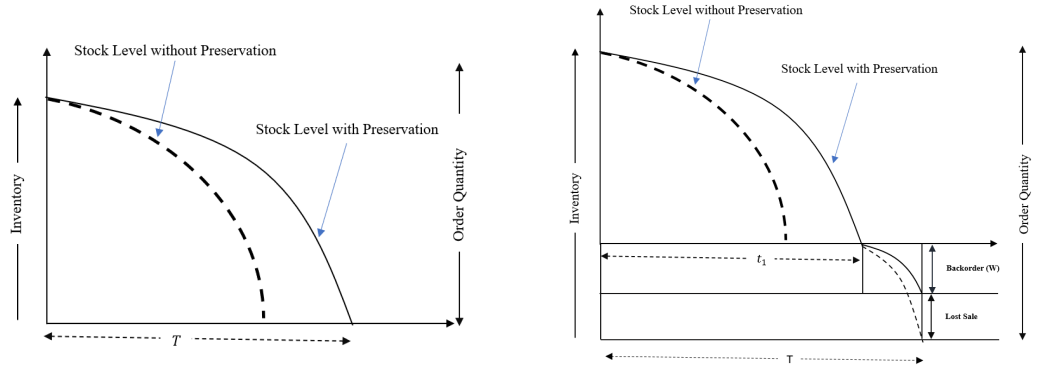
ξ : Per unit preservation cost,

$q(t)$: Instantaneous inventory level at variable time t,

Π_1, Π_2 : Total profit function for the model including without and with shortage respectively,

4. Assumptions

1. The proposed model is for single retailer model dealing with the perishable product having fixed self-life τ,
2. Demand is influenced by the freshness level of the product and the freshness index is determined by $f(t) = \frac{\tau - t}{\tau}, \quad t \in [0, \tau]$
3. The deterioration is considered instantaneous determined by $\theta(t) = \frac{1-m(\xi)}{1+\tau-t}$, where per unit preservation cost ξ is employed to reduce the rate of deterioration in $0 \leq t \leq T \leq \tau$,
4. The preservation effort is a continuous and concave function of per unit preservation cost ξ, and determined as $m(\xi) = 1 - e^{-\alpha\xi}$. Thus, reduced rate of deterioration is $1 - m(\xi)$,
5. Price-sensitive and time dependent demand is employed under the influence of uncertain factor ϵ and determined as $D(t, p) = (a - \alpha p - bt + \epsilon)$, for $a > 0$, $b > 0$ and $p > 0$



(a) Stock level without shortage

(b) Stock level with shortage

Figure 1: *Comparison of stock levels with and without shortage*

6. The promotional effort is incorporated to enhance the demand with promotional efforts parameter A^{γ}. Where, the γ is the frequency of advertisement and gives a freedom to management that how many time it could advertise the product to get optimum profit,

7. The mathematical formulation section is categorised into two categories; 'with' and 'without'shortage. In shortage scenario, the demand is assumed to be partially backlogged with the gain fraction η and Lead time is to be considered zero.

5. Mathematical Formulation

This paper introduces a mathematical framework to help retailers to increase profits when restocking perishable items having a fix self-life τ. The proposed formulation focuses on optimizing profit of retailer who invests in an ordering cost (C_0) and restocking inventory regularly having purchasing cost (C_p) and selling price p. With negligible lead time, This model accounts for the critical relationship between freshness of stock level, demand level, as consumers strongly prefer fresh, perishable items and the advertisement frequency. It is simply to understand that the maximum lifespan of the product determines the maximum length of a business cycle (T) and as well as less deterioration. As a result, products that stay longer their maximum lifespan will be avid to sale for the customers and organisation's welfare. Additionally, consumers always struggle to purchase the best available product at a reasonable price, so the selling price is often the primary factor for customers. Therefore, the proposed formulation optimises the replenishment time T, optimal selling price p, advertising frequency A, and the preservation cost ξ.

The formulation is divided into two separate categories, with and without shortage. The further section focuses on the case of without shortage.

Case I: The model with no shortage

The model (refer to figure 1(a)) starts by examining a situation where a retailer gets the order quantity Q, of a perishable items from a supplier. After some time, the quality and quantity of the products reduces gradually due to both natural decay and consumer demand. Eventually, the quantity will reach zero at time $t = T$. In any instant t, the stock level $q(t)$ follows following differential equation:

$$\frac{dq(t)}{dt} + \frac{1 - m(\xi)}{1 + \tau - t} q(t) = -A^{\gamma}(a - \alpha p - bt + \epsilon)\frac{\tau - t}{\tau}, \quad 0 \leq t \leq T \tag{1}$$

subject to $q(T) = 0$, and $q(0) = Q$. For notational convenience consider $1 - m(\xi) = k$, and

$$\psi_1 = \frac{b}{k-3}; \quad \psi_2 = \frac{-a + p\alpha - \epsilon + b(\tau+1)}{k-1}; \quad \psi_3 = \frac{-a + p\alpha - \epsilon + b(\tau+2)}{k-2};$$

$$\phi_1 = (1+\tau)^3 C_p + \frac{C_h(1+\tau)^4}{4}; \quad \phi_2 = (1+\tau)C_p + \frac{C_h(1+\tau)^2}{2};$$

$$\phi_3 = (1+\tau)^2 C_p + \frac{C_h(1+\tau)^3}{3}; \quad \phi_4 = (1+\tau)^k C_p + \frac{C_h(1+\tau)^{(k+1)}}{k+1}$$

Equation (1), leads to

$$\begin{aligned} q(t) = \quad & \frac{A^\gamma}{\tau}\Big[-(1+\tau-t)\Big\{\psi_1(1+\tau-t)^2 + \psi_2 - \psi_3(1+\tau-t)\Big\} \\ & +(1+\tau-t)^k(1+\tau-T)^{(1-k)}\Big\{\psi_1(1+\tau-T)^2 + \psi_2 - \psi_3(1-T+\tau)\Big\}\Big] \end{aligned} \tag{2}$$

The retailer's lot size Q with $q(0) = Q$, will be

$$\begin{aligned} Q(T) = \quad & \frac{A^\gamma}{\tau}\Big[-\psi_1(1+\tau)^3 - \psi_2(1+\tau) + \psi_3(1+\tau)^2 \\ & +(1+\tau)^k\Big\{\psi_1(1+\tau-T)^{(3-k)} + \psi_2(1+\tau-T)^{(1-k)} - \psi_3(1+\tau-T)^{(2-k)}\Big\}\Big] \end{aligned} \tag{3}$$

From the above inventory level and retailer's lot size, there are several cost functions incorporated in the proposed profit function. The purchase cost (PC) of each cycle for the retailer's is as follows:

$$\begin{aligned} PC(T) = \quad & C_pQ = C_p\frac{A^\gamma}{\tau}\Big[-\psi_1(1+\tau)^3 - \psi_2(1+\tau) + \psi_3(1+\tau)^2 + \\ & (1+\tau)^k\Big\{\psi_1(1+\tau-T)^{(3-k)} + \psi_2(1+\tau-T)^{(1-k)} - \psi_3(1+\tau-T)^{(2-k)}\Big\}\Big] \end{aligned} \tag{4}$$

The cost associated with holding (CC) over the entire cycle

$$\begin{aligned} CC(T) = \quad & C_h\int_0^T q(t)dt \\ & = \frac{C_hA^\gamma}{\tau}\Big[\psi_1(1+\tau-t)^4\left(\frac{1}{4} - \frac{1}{k+1}\right) + \psi_2(1+\tau-T)^2\left(\frac{1}{2} - \frac{1}{k+1}\right) \\ & -\psi_3(1+\tau-t)^3\left(\frac{1}{3} - \frac{1}{k+1}\right) - \frac{(1+\tau)^{(k+1)}}{k+1}\Big[\psi_2(1+\tau-T)^{(1-k)} + \\ & \psi_1(1+\tau-T)^{(3-k)} - \psi_3(1+\tau-T)^{(2-k)}\Big] - \frac{\psi_1}{4}(1+\tau)^4 - \frac{\psi_2}{2}(1+\tau)^2 + \\ & \frac{\psi_3}{3}(1+\tau)^3\Big] \end{aligned} \tag{5}$$

The advertising cost (AC) with advertisement frequency A is

$$AC = G \times A \tag{6}$$

where G is the cost per advertisement. During the interval $[0, T]$, the total sales revenue (SR) is

$$\begin{aligned} SR(T) = \quad & p\int_0^T A^\gamma(a - \alpha p - bt + \epsilon)\frac{\tau - t}{\tau}dt \\ = \quad & \frac{pA^\gamma}{6\tau}\Big[2bT^3 + 6T(a - \alpha p + \epsilon)\tau - 3T^2(a - \alpha p + b\tau + \epsilon)\Big] \end{aligned} \tag{7}$$

Incorporating all the above cost, the profit earned by retailer in period $[0, T]$ is ,

$$\begin{aligned}\Pi_1(A,p,T) = \quad & SR - OC - AC - PC - CC \\ = \quad & \frac{pA^\gamma}{6\tau}\left[2bT^3 + 6T(a-\alpha p+\epsilon)\tau - 3T^2(a-\alpha p+b\tau+\epsilon)\right] - C_0 - GA \\ & + \frac{A^\gamma}{\tau}\left[(\psi_1\phi_1+\psi_2\phi_2-\psi_3\phi_3) + \left\{-\psi_1(1+\tau-T)^{(3-k)} - \psi_2(1+\tau-T)^{(1-k)}\right.\right. \\ & \left.+\psi_3(1+\tau-T)^{(2-k)}\right\}\phi_4 - C_h\left\{\psi_1(1+\tau-T)^4\left(\frac{1}{4}-\frac{1}{k+1}\right)\right. \\ & \left.\left.+\psi_2(1+\tau-T)^2\left(\frac{1}{2}-\frac{1}{k+1}\right) - \psi_3(1+\tau-T)^3\left(\frac{1}{3}-\frac{1}{k+1}\right)\right\}\right] \end{aligned} \tag{8}$$

Our primary objective is to determine the optimal advertisement frequency, per unit selling price, and replenishment time that will lead to the highest expected profit for the retailer. We determined the optimal advertisement frequency using an algorithmic approach, and obtained conditions in the theorem given below.

Theorem 5.1. *The proposed profit function $\Pi_1(A,p,T)$ is strictly concave for price p, for $T < 2\tau$.*
Proof: *Consider the profit function equation (8), we have*

$$\begin{aligned}\frac{\partial\Pi_1(A,p,T)}{\partial p} = \quad & \frac{A^\gamma}{6\tau}\left[2bT^3 + 6T(a-\alpha p+\epsilon)\tau - 3T^2(a-\alpha p+b\tau+\epsilon)\right] + \frac{pA^\gamma}{6\tau}\left[-6T\alpha\tau\right. \\ & \left.+3T^2\alpha\right] + \frac{A^\gamma}{\tau}\left[\left[\frac{\alpha\phi_2}{k-1} - \frac{\alpha\phi_3}{k-2}\right] + \right. \\ & \left[-\frac{\alpha(1+\tau-t_1)^{1-k}}{k-1} + \frac{\alpha(1+\tau-T)^{2-k}}{k-2}\right]\phi_4 \\ & \left.-C_h\left[\frac{\alpha(1+\tau-t_1)^2(\frac{1}{2}-\frac{1}{k+1})}{k-1} - \frac{\alpha(1+\tau-t_1)^3(\frac{1}{3}-\frac{1}{k+1})}{k-2}\right]\right] \end{aligned} \tag{9}$$

Putting $\frac{\partial\Pi_1(A,p,T)}{\partial p} = 0$, we have optimal value of p denoted by p^*

$$\begin{aligned}p^* = \quad & \frac{1}{\alpha T(2-T)}\left[\frac{A^\gamma}{6\tau}\left(2bT^3 + 6T\tau(a+\epsilon) - 3T^2(a+b\tau+\epsilon)\right) + \frac{A^\gamma}{\tau}\left\{\left(\frac{\alpha\phi_2}{k-1} - \frac{\alpha\phi_3}{k-2}\right)\right.\right. \\ & \left.\left.-C_h\left(\frac{\alpha(1+\tau-T)^2(\frac{1}{2}-\frac{1}{k+1})}{k-1} - \frac{\alpha(1+\tau-T)^3(\frac{1}{3}-\frac{1}{k+1})}{k-2}\right)\right\}\right. \\ & \left.+\left(\frac{\alpha(1+\tau-T)^{2-k}}{k-2} - \frac{\alpha(1+\tau-T)^{1-k}}{k-1}\right)\phi_4\right] \end{aligned} \tag{10}$$

differentiating again the above equation, we have

$$\frac{\partial^2\Pi_1(A,p,T)}{\partial p^2} = 2A^\gamma\alpha T\left[-1+\frac{T}{2\tau}\right] < 0, \quad for \quad T < 2\tau \tag{11}$$

Hence, for $T < 2\tau$, the profit function $\Pi_1(A,p,T)$ is concave with respect to selling price p, and attained maxima at p^* (refer to figure 2(a)).

After, finding the condition for the optimal solution of the price, we have shown the concave behaviour of the objective function with respect to the time T.

Theorem 5.2. *The objective function $\Pi_1(A,p,T)$ is concave in $T \in (0,\tau]$.*

Proof: *From the objective function* $\Pi_1(A,p,T)$ *(equation (8)), we have*

$$\begin{aligned}\frac{\partial\Pi_1(A,p,T)}{\partial T} &= \left[\frac{pA^\gamma}{6\tau}\left\{6bT^2+6(a-\alpha p+\epsilon)\tau-6T(a-\alpha p+b\tau+\epsilon)\right\}\right.\\ &+\frac{A^\gamma}{\tau}\left\{(3-k)\psi_1(1+\tau-T)^{(2-k)}+(1-k)\psi_2(1+\tau-T)^{-k}\right.\\ &\left.-(2-k)\psi_3(1+\tau-T)^{(1-k)}\right\}\phi_4\\ &-C_h\frac{A^\gamma}{\tau}\left\{-4\psi_1(1+\tau-T)^3\left(\frac{1}{4}-\frac{1}{k+1}\right)\right.\\ &\left.\left.-2\psi_2(1+\tau-T)\left(\frac{1}{2}-\frac{1}{k+1}\right)+3\psi_3(1+\tau-T)^2\left(\frac{1}{3}-\frac{1}{k+1}\right)\right\}\right] \quad (12)\end{aligned}$$

The above equation leads to,

$$\begin{aligned}\frac{\partial^2\Pi_1(A,p,T)}{\partial T^2} = &\frac{A^\gamma}{\tau}\left[p\left[b(2T-1)+\alpha p-(a+\epsilon)\right]-(1+\tau-T)^{-k}\left[-\psi_1(k-2)\right.\right.\\ &(k-3)(1+\tau-T)+k(k-1)(1+\tau-T)^{-1}\psi_2+\psi_3(k-1)(k-2)\left.\right]\phi_4\\ &-C_h\left[12\psi_1(1+\tau-T)^2\left(\frac{1}{4}-\frac{1}{k+1}\right)\right.\\ &\left.\left.+2\psi_2\left(\frac{1}{2}-\frac{1}{k+1}\right)-6\psi_3(1+\tau-T)\left(\frac{1}{3}-\frac{1}{k+1}\right)\right]\right] \quad (13)\end{aligned}$$

$$\frac{\partial^2\Pi_1(A,p,T)}{\partial T^2}=\frac{A^\gamma}{\tau}(\delta_1-\delta_2)<0$$

for $Max(\delta_1)<Min(\delta_2)$*, where*

$$\begin{aligned}\delta_1 = &\ p\left[b(2T-1)+\alpha p-(a+\epsilon)\right]-(1+\tau-T)^{-k}\left[-\psi_1(k-2)(k-3)(1+\tau-T)\right.\\ &\left.+k(k-1)(1+\tau-T)^{-1}\psi_2+\psi_3(k-1)(k-2)\right]\phi_4\end{aligned}$$

and

$$\delta_2=C_h\left[12\psi_1(1+\tau-T)^2\left(\frac{1}{4}-\frac{1}{k+1}\right)+2\psi_2\left(\frac{1}{2}-\frac{1}{k+1}\right)-6\psi_3(1+\tau-T)\left(\frac{1}{3}-\frac{1}{k+1}\right)\right]$$

Here

$$\frac{\partial\delta_1}{\partial T}>0,\ \frac{\partial\delta_2}{\partial T}<0$$

and thus, δ_1 *is increasing in T and* δ_2 *is decreasing in T, Due to this, maximum and minimum value of* δ_1*and* δ_2 *respectively is obtained on* τ *in the interval* $[0,\tau]$.
From above,

$$\begin{aligned}Max(\delta_1)-Min(\delta_2) = &-\frac{1}{(3-k)(1+k)}\left[(3-k)(1+k)p(a-p\alpha+\epsilon-b(-1+2\tau))\right.\\ &+C_h\left[(a-\alpha p-\epsilon)(5-k)+b(2+5\tau-k\tau)\right]\\ &+(1+\tau)^k(C_h+C_p+C_pk+C_h\tau)(a(-3+5k-2k^2)\\ &\left.+(3-5k+2k^2)(p\alpha-\epsilon)+b(3\tau+2k^2(1+\tau)-k(4+5\tau)))\right]<0\end{aligned} \quad (14)$$

Hence, the profit function $\Pi_1(A,p,T)$ *(equation (8)) is concave with respect to T .*

Using these obtained conditions for optimality of objective function, we have developed an algorithm to obtain the numerical solution of the problem.

Algorithm 1. *Algorithm to determine the optimal solution without shortage*
Input: *Demand, A_{max}, Parameters, Objective function*
Output: *Optimal value of p^*, T^*, A^*, π^**
Define Module 1 taking argument (μ, σ)

1. *Initialize $\epsilon = f(\mu, \sigma)$*
2. *Return ϵ*

Main Body:

1. *Initialize $T^* = 0, p^* = 0, A^* = 0, \xi^* = 0, \pi^* = -\infty$*
2. *for $A = 1, 2, ..., A_{max}$*
3. *Solve Eq. (8) and (11) for p and T.*
4. *Solve $\frac{\partial \pi_1}{\partial \xi} = 0$ for ξ.*
5. *If $(T < 2\tau)$ do*
6. *If $[\delta_1]_{T=\tau} < [\delta_2]_{T=\tau}$ do*
7. *Calculate $\pi_1(A, p, T, \xi)$*
8. *If $\pi_1(.) > \pi_1^*(.)$*
9. *Set $T^* = T, p^* = p, A^* = A, \xi^* = \xi, \pi^* = \pi_1$*
10. *end If*
11. *end If*
12. *end If*
13. *end for*
14. *Get $T^*, p^*, A^*, \xi^*, \pi^*$*

Now, we have another category for the formulation, in which the shortage is allowed. Further discussion have investigated the shortage condition and derived the condition for the optimal solution. where

Case II: The model with shortage

At the beginning of the sale period, the retailer purchases Q units of a perishable products from the supplier. These units are considered fresh and in their original state. Once the stock reaches to zero, the retailer experiences shortages, and the demand is accumulated with a fraction η in interval $[t_1, T]$ (Figure 1(b)). So, the inventory level at any instant is govern in equation 15 and 16.

$$\frac{dq_1(t)}{dt} + \frac{1 - m(\xi)}{1 + \tau - t} q(t) = -A^{\gamma}(a - \alpha p - bt + \epsilon)\frac{\tau - t}{\tau}, \quad 0 \le t \le t_1 \tag{15}$$

and

$$\frac{dq_2(t)}{dt} = -\eta A^{\gamma}(a - bt - \alpha p + \epsilon), \quad t_1 \le t \le T \tag{16}$$

with the help of $q_1(0) = S$,$q_1(t_1) = 0$,$q_2(t_1) = 0$, and $q_2(T) = -W$., the above equations, provide solution given as follows

$$\begin{aligned} q_1(t) = \quad & \frac{A^\gamma}{\tau}\Big[-\psi_1(1+\tau-t)^3 - \psi_2(1+\tau-t) + \psi_3(1+\tau-t)^2 + \\ & (1+\tau-t)^k(1+\tau-t_1)^{(1-k)}\Big\{\psi_1(1+\tau-t_1)^2 + \psi_2 - \psi_3(1-t_1+\tau)\Big\}\Big] \end{aligned} \tag{17}$$

and,

$$q_2(t) = \eta A^\gamma\left[(a-\alpha p+\epsilon)(t_1-t) - \frac{b(t_1^2-t^2)}{2}\right] \tag{18}$$

The given amounts for initial inventory and backlogged shortage respectively are as follows:

$$\begin{aligned} S = \quad & \frac{A^\gamma}{\tau}\Big[-\psi_1(1+\tau)^3 - \psi_2(1+\tau) + \psi_3(1+\tau)^2 + \\ & (1+\tau)^k\Big[\psi_1(1+\tau-t_1)^{(3-k)} + \psi_2(1+\tau-T)^{(1-k)} - \psi_3(1+\tau-T)^{(2-k)}\Big]\Big] \end{aligned} \tag{19}$$

and,

$$W = \eta A^\gamma\left[(a-\alpha p+\epsilon)(T-t_1) - \frac{b(T^2-t_1^2)}{2}\right] \tag{20}$$

Employing the above inventory level and order quantities, there are several costs associated with the inventory control. The cost incurred by the retailer for purchasing cost (PC) is given as,

$$\begin{aligned} PC = \quad & C_p(S+W) \\ = \quad & C_p\frac{A^\gamma}{\tau}\Big[-\psi_1(1+\tau)^3 - \psi_2(1+\tau) + \psi_3(1+\tau)^2 \\ & +(1+\tau)^k\Big[\psi_1(1+\tau-t_1)^{(3-k)} + \psi_2(1+\tau-t_1)^{(1-k)} - \psi_3(1+\tau-t_1)^{(2-k)}\Big] \\ & +C_p\eta A^\gamma\left[(a-\alpha p+\epsilon)(T-t_1) - \frac{b(T^2-t_1^2)}{2}\right]\Big] \end{aligned} \tag{21}$$

The carrying cost (CC) for each cycle $[0, t_1]$ is,

$$\begin{aligned} CC(t_1) = \quad & C_h\int_0^{t_1} q_1(t)dt \\ = \quad & \frac{C_hA^\gamma}{\tau}\Big[\psi_1(1+\tau-t_1)^4\left(\frac{1}{4}-\frac{1}{k+1}\right) + \psi_2(1+\tau-t_1)^2\left(\frac{1}{2}-\frac{1}{k+1}\right) \\ & -\psi_3(1+\tau-t_1)^3\left(\frac{1}{3}-\frac{1}{k+1}\right) - \frac{(1+\tau)^{(k+1)}}{k+1}\Big[\psi_2(1+\tau-t_1)^{(1-k)} \\ & +\psi_1(1+\tau-t_1)^{(3-k)} - \psi_3(1+\tau-t_1)^{(2-k)}\Big] - \frac{\psi_1}{4}(1+\tau)^4 \\ & -\frac{\psi_2}{2}(1+\tau)^2 + \frac{\psi_3}{3}(1+\tau)^3\Big] \end{aligned} \tag{22}$$

The cost of backorder (BC), and lost sales (LSC) per replenishment cycle respectively is given as,

$$\begin{aligned} BC(t_1,T) = \quad & -C_1\int_{t_1}^{T} q_2(t)dt \\ = \quad & -C_1\eta A^\gamma\left[(a-\alpha p+\epsilon)\left(t_1T - \frac{T^2}{2} - \frac{t_1^2}{2}\right) - \frac{b}{2}\left(t_1^2T - \frac{T^3}{3} - \frac{2t_1^3}{3}\right)\right] \end{aligned} \tag{23}$$

and

$$\begin{aligned} LSC(t_1,T) = \quad & C_2(1-\eta)\int_{t_1}^{T} A^{\gamma}(a-\alpha p-bt+\epsilon)dt \\ = \quad & C_2(1-\eta)A^{\gamma}\left[(a-\alpha p+\epsilon)(T-t_1)-b\frac{T^2-t_1^2}{2}\right] \end{aligned} \tag{24}$$

During inventory cycle, the total number of advertisements is A and the cost per advertisement is G. Therefore, the total advertisement cost is GA.
The total sales revenue (SR) during [0, T] is

$$\begin{aligned} SR(p,t_1) = \quad & p\int_0^{t_1} A^{\gamma}(a-\alpha p-bt+\epsilon)\frac{\tau-t}{\tau}dt + pW \\ = \quad & pA^{\gamma}\left[\frac{1}{6\tau}\left\{2bt_1^3+6t_1(a-\alpha p+\epsilon)\tau-3t_1^2(a-\alpha p+b\tau+\epsilon)\right\}+\right. \\ & \left.\eta\left\{(a-\alpha p+\epsilon)(T-t_1)-\frac{b(T^2-t_1^2)}{2}\right\}\right] \end{aligned} \tag{25}$$

Hence, accumulating all the above costs the profit function given is,

$$\begin{aligned} \Pi_2(A,p,t_1,T) = \quad & \frac{pA^{\gamma}}{6\tau}\left[2bt_1^3+6t_1(a-\alpha p+\epsilon)\tau-3t_1^2(a-\alpha p+\tau+\epsilon)\right]-C_0-GA \\ & +\frac{A^{\gamma}}{\tau}\left[(\psi_1\phi_1+\psi_2\phi_2-\psi_3\phi_3)+\left\{-\psi_1(1+\tau-t_1)^{(3-k)}-\psi_2(1+\tau-t_1)^{(1-k)}\right.\right. \\ & \left.+\psi_3(1+\tau-t_1)^{(2-k)}\right\}\phi_4-C_h\left\{\psi_1(1+\tau-t_1)^4\left(\frac{1}{4}-\frac{1}{k+1}\right)\right. \\ & \left.\left.+\psi_2(1+\tau-t_1)^2\left(\frac{1}{2}-\frac{1}{k+1}\right)-\psi_3(1+\tau-t_1)^3\left(\frac{1}{3}-\frac{1}{k+1}\right)\right\}\right] \\ & +\eta A^{\gamma}\left[p\left\{(a-\alpha p+\epsilon)(T-t_1)-\frac{b(T^2-t_1^2)}{2}\right\}\right. \\ & \left.+C_1\left\{(a-\alpha p+\epsilon)\left(t_1T-\frac{T^2}{2}-\frac{t_1^2}{2}\right)-\frac{b}{2}\left(t_1^2T-\frac{T^3}{3}-2\frac{t_1^3}{3}\right)\right\}\right] \\ & -C_2(1-\eta)A^{\gamma}\left[(a-\alpha p+\epsilon)(T-t_1)-\frac{b(T^2-t_1^2)}{2}\right] \end{aligned} \tag{26}$$

Our objective is to analyse the above profit function (equation (26)) and identify the optimal values for the frequency of advertisement, selling price, and replenishment time to achieve maximum profit for the retailer. In order to achieve this, we have formulated the following theorems .

Theorem 5.3. *For $t_1 < 2\tau$, the profit function $\Pi_2(A,p,t_1,T)$ is strictly concave for price p, where $t_1 \in (0,\tau)$.*

Proof: Consider the profit function equation (26), we obtained

$$\begin{aligned}\frac{\partial \Pi_2(A,p,t_1,T)}{\partial p} = \quad & \frac{A^\gamma}{6\tau}\left[2bt_1^3+6t_1(a-\alpha p+\epsilon)\tau-3t_1^2(a-\alpha p+b\tau+\epsilon)\right]+\frac{pA^\gamma}{6\tau}\left[-6t_1\alpha\tau+3t_1^2\alpha\right] \\ & +\frac{A^\gamma}{\tau}\left\{\left(\frac{\alpha\phi_2}{k-1}-\frac{\alpha\phi_3}{k-2}\right)+\left(\frac{\alpha(1+\tau-t_1)^{2-k}-\frac{\alpha(1+\tau-t_1)^{1-k}}{k-1}}{k-2}\right)\phi_4\right. \\ & \left.-C_h\left(\frac{\alpha(1+\tau-t_1)^2(\frac{1}{2}-\frac{1}{k+1})}{k-1}-\frac{\alpha(1+\tau-t_1)^3(\frac{1}{3}-\frac{1}{k+1})}{k-2}\right)\right\}+ \\ & \eta A^\gamma\left[\left((a-\alpha p+\epsilon)(T-t_1)-b\frac{(T^2-t_1^2)}{2}-\alpha p(T-t_1)\right)\right. \\ & \left.-C_1\alpha\left(t_1T-\frac{T^2}{2}-\frac{t_1^2}{2}\right)\right]-C_2(1-\eta)A^\gamma\left[-\alpha(T-t_1)\right] \end{aligned} \tag{27}$$

Putting $\frac{\partial \Pi_2(A,p,T)}{\partial p}=0$, we have optimal value of p is,

$$\begin{aligned} p^* = \quad & \frac{1}{\frac{A^\gamma}{\tau}\left[2t_1\alpha\tau-\alpha t_1^2\right]+\eta A^\gamma\left[2\alpha(T-t_1)\right]}\left[\frac{A^\gamma}{\tau}\left\{\frac{bt_1^3}{3}+t_1(a+\epsilon)\tau-\frac{t_1^2(a+b\tau+\epsilon)}{2}\right.\right. \\ & +\frac{\alpha\phi_2}{k-1}-\frac{\alpha\phi_3}{k-2}+\left(-\frac{\alpha(1+\tau-t_1)^{1-k}}{k-1}+\frac{\alpha(1+\tau-t_1)^{2-k}}{k-2}\right)\phi_4 \\ & \left.\left.-C_h\left(\frac{\alpha(1+\tau-t_1)^2(\frac{1}{2}-\frac{1}{k+1})}{k-1}-\frac{\alpha(1+\tau-t_1)^3(\frac{1}{3}-\frac{1}{k+1}))}{k-2}\right)\right\}\right. \\ & +\eta A^\gamma\left[\left((a+\epsilon(T-t_1)-b\frac{(T^2-t_1^2)}{2}-\alpha p(T-t_1)\right)\right. \\ & \left.\left.-C_1\alpha\left(t_1T-\frac{T^2}{2}-\frac{t_1^2}{2}\right)\right]+C_2(1-\eta)A^\gamma\alpha(T-t_1)\right] \end{aligned} \tag{28}$$

Differentiating again, we have,

$$\frac{\partial^2 \Pi_2(A,p,t_1,T)}{\partial p^2}=2A^\gamma\alpha\left[t_1\left(-1+\frac{t_1}{2\tau}\right)-\eta(T-t_1)\right]<0 \tag{29}$$

Thus, for $T<t_1<2\tau$, the profit function $\Pi_2(A,p,T)$ is concave and attain the maxima on the optimal price p^*.

Preposition 1. *For $a>0$, $b>0$, $p>0$, $Max(\delta_1+\delta_3)<Min(\delta_2)$,*

$$\begin{aligned}\delta_1 = \quad & \frac{A^\gamma}{\tau}\left[p\left[b(2t_1-1)+\alpha p-(a+\epsilon)\right]-(1+\tau-t_1)^{-k}\left[-\psi_1(k-2)(k-3)(1+\tau-t_1)+\right.\right. \\ & \left.\left.k(k-1)(1+\tau-t_1)^{-1}\psi_2+\psi_3(k-1)(k-2)\right]\phi_4\right] \end{aligned}$$

$$\delta_2 = \quad C_h\left[12\psi_1(1+\tau-t_1)^2\left(\frac{1}{4}-\frac{1}{k+1}\right)+2\psi_2\left(\frac{1}{2}-\frac{1}{k+1}\right)-6\psi_3(1+\tau-t_1)\left(\frac{1}{3}-\frac{1}{k+1}\right)\right]$$

$$\delta_3 = \quad -A^\gamma\eta\left[b-C_1((a-\alpha p+\epsilon)-b(T-2t_1))\right]-C_2A^\gamma b(1-\eta)$$

Proof: From theorem (5.2), we have, $Max(\delta_1)<Min(\delta_2)$. So, we have added negative quantity $(\delta_3<0)$ in the RHS, resultant maximum value of $\delta_1+\delta_3$ is still down then we have,

$$Max(\delta_1+\delta_3)<Min(\delta_2)$$

Now, $\frac{dTP(p)}{dp} = 0$ implies a unique value of p using Mathematica software, which is complex in nature. Clearly, $\frac{d^2TP(p)}{dp^2}$ is negative (see figure 2(b))

Theorem 5.4. *Objective function $\Pi_2(A, p, t_1, T)$ concave for t_1 in the interval $[0, \tau]$.*

Proof: From the equation (26), we may write,

$$\begin{aligned}\frac{\partial \Pi_2}{\partial t_1} = \quad & \frac{pA^\gamma}{6\tau}\left[6bt_1^2 + 6(a-\alpha p+\epsilon)\tau - 6t_1(a-\alpha p+b\tau+\epsilon)\right] \\ & + \frac{A^\gamma}{\tau}\left[(3-k)\psi_1(1+\tau-t_1)^{(2-k)} + (1-k)\psi_2(1+\tau-t_1)^{-k}\right. \\ & \left.-(2-k)\psi_3(1+\tau-t_1)^{(1-k)}\right]\phi_4 \\ & - C_h\frac{A^\gamma}{\tau}\left[-4\psi_1(1+\tau-t_1)^3\left(\frac{1}{4}-\frac{1}{k+1}\right) - 2\psi_2(1+\tau-t_1)\left(\frac{1}{2}-\frac{1}{k+1}\right)\right. \\ & \left.+3\psi_3(1+\tau-t_1)^2\left(\frac{1}{3}-\frac{1}{k+1}\right)\right] \\ & + A^\gamma\eta\left[p(-(a-\alpha p+\epsilon)+bt_1) + C_1((a-\alpha p+\epsilon)(T-t_1) - b(t_1T-t_1^2))\right] \\ & - C_2(1-\eta)A^\gamma\left[bt_1-(a-\alpha p+\epsilon)\right] \end{aligned} \tag{30}$$

Above equation leads to,

$$\frac{\partial \Pi_2(A, p, t_1, T)}{\partial t_1} = \delta_1 - \delta_2 + \delta_3$$

where, $\delta_1, \delta_2, \delta_3$ is given in above preposition. From the preposition 1, $\frac{\partial \Pi_2(A,p,t_1,T)}{\partial T} < 0$.

Theorem 5.5. *For $c_2 < \eta\left[b(p+c_2) + c_1D(p,t)\right]$, the objective function is concave for T in the interval $[0, \tau]$.*
Proof: *From the equation (26), we have*

$$\begin{aligned}\frac{\partial \Pi_2(A, p, t_1, T)}{\partial T} = \quad & \eta A^\gamma\left[p\left[(a-\alpha p+\epsilon) - bT\right] + C_1\left\{(a-\alpha p+\epsilon)(t_1-T) - \frac{b}{2}\left(t_1^2-T^2\right)\right\}\right] \\ & - C_2(1-\eta)A^\gamma\left[(a-\alpha p+\epsilon) - bT\right] \end{aligned} \tag{31}$$

or,

$$\frac{\partial \Pi_2^2(A, p, t_1, T)}{\partial T^2} = A^\gamma\eta\left[-bP - C_1(a-\alpha p+\epsilon+bT)\right] + C_2(1-\eta)A^\gamma b \tag{32}$$

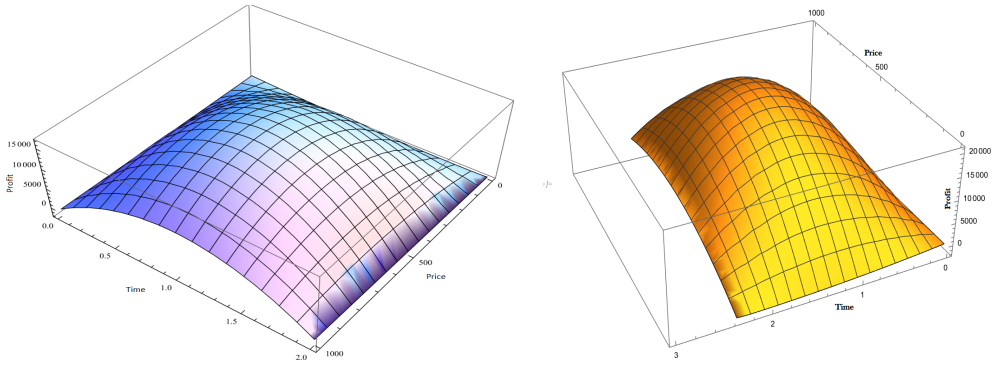
For the given positive values of parameters and $c_2 < \eta\left[b(p+c_2) + c_1D(p,t)\right]$, the objective function is concave in T.

After obtaining the conditions for optimality of the objective function, we developed an algorithm to determine the algorithm to get numerical solution.

Algorithm 2. *Algorithm to determine the optimal solution without shortage*
Input: *Demand, A_{max}, Parameters, Objective function*
Output: *Optimal value of p^*, T^*, A^*, π^**
Define Module 1 taking argument (μ, σ)

1. *Initialize $\epsilon = f(\mu, \sigma)$*
2. *Return ϵ*

Main Body:



(a) Concavity of the profit function with respect to time and price without shortages

(b) Concavity of the profit function with respect to time and price with shortages

Figure 2: *Three dimensional concavity of profit function*

1. *Initialize* $t_1 = 0, T^* = 0, p^* = 0, A^* = 0, \xi^* = 0, \pi^* = -\infty$
2. *for* $A = 1, 2, ..., A_{max}$
3. *Solve Eq. (27), (30) and (31) for* p, t_1 *and* T.
4. *Solve* $\frac{\partial \pi_1}{\partial \xi} = 0$ *for* ξ.
5. *If* $(T < 2\tau)$ *and* $c_2 < \eta\left[b(p + c_2) + c_1 D(p, t)\right]$ *do*
6. *If* $[\delta_1 + \delta_3]_{T=\tau} < [\delta_2]_{T=\tau}$ *do*
7. *Calculate* $\pi_1(A, p, T, \xi)$
8. *If* $\pi_1(.) > \pi_1^*(.)$
9. *Set* $t_1^* = t_1, T^* = T, p^* = p, A^* = A, \xi^* = A, \pi^* = \pi_1$
10. *end If*
11. *end If*
12. *end If*
13. *end for*
14. *Get* $t_1^*, T^*, p^*, A^*, \xi^*, \pi^*$

After obtaining the algorithms, we have derived numerical results for this formulations given in the further sections.

6. Empirical Evaluation

A numerical example illustrates the model's adaptability and efficacy, providing insights into replenishment time, optimal pricing, preservation costs, and overall profitability. This research contributes to evolving inventory management strategies, particularly in industries with volatile demand and deteriorating items.

Example 1. ***Example of model without shortage:*** *In the retail store, research investigates an inventory management model that integrates stochastic demand and advertising techniques. The example incorporates essential parametric values such as the price-sensitive factor* (α) *is 0.1, a setup cost* (C_0) *is \$520, a purchasing cost* (C_p) *is \$5, a potential demand* (a) *is 100 units, a holding cost per unit* (C_h) *is \$1, a time-sensitive parameter* (b) *is 0.1, a maximum product lifetime* (τ) *is 2 unit of time, an advertising frequency* (A) *is 1 per advertisement, an advertising impact parameter* (γ) *is 0.1, an advertisement cost* (G) *is \$50, a stochastic demand factor* (ϵ) *is 20, and a preservation cost parameter* (λ) *is 0.0025.*

The model adapts pricing and preservation strategies, resulting in a replenishment time (T) *is 0.986282 units of time, an optimal price* (p) *is \$607.44, an optimal preservation cost is \$277.259, and an optimal profit is \$16,963.9.*

This example illustrates how the model can adapt in the real world, where dynamic strategies are aligned with market demands and deterioration. It offers valuable insights into inventory management, demonstrating the model's effectiveness in addressing challenges and adapting to market dynamics.

Example 2. ***Example for the model with shortage:*** *The example is formulated precisely to examine a system that includes shortages. Key parameters are as follows: Price sensitivity* (α) *is 0.1, setup cost* (C_0) *is \$520, purchasing cost* (C_p) *is \$5, potential demand* (a) *is 100 units, time-sensitivity parameter* (b) *is 0.1, holding cost* (C_h) *is \$1 per unit per unit of time, maximum product lifetime* (τ) *is 2 unit of time, advertising frequency* (A) *is 1 per advertisement, advertisement impact* (γ) *is 0.1, advertisement cost* (G) *at \$50 per advertisement, stochastic demand factor* (ϵ) *is 20, preservation cost impact parameter* (λ) *is 0.0025, backlogging fraction* (η) *is 0.6, shortage cost* (c_1) *is \$1, and opportunity cost* (c_2) *is \$1.*

The numerical outcomes reveal a replenishment time (T) *is 1.28626 units of time, an optimal shortage time* (t_1) *is 1.09332 units, an optimal price* (p) *is \$606.71, an optimal preservation cost is \$277.26, and an optimal profit \$20,916.3.*

The numerical results show a relationship among these elements, leading to optimal shortage time, price, preservation cost, and profit. This emphasizes the model's versatility in addressing broader challenges, such as unexpected demand fluctuations.

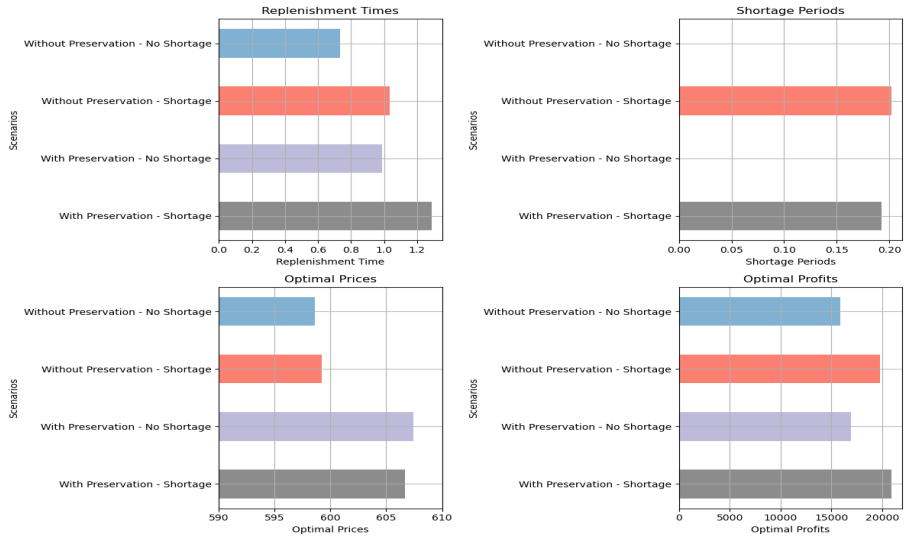
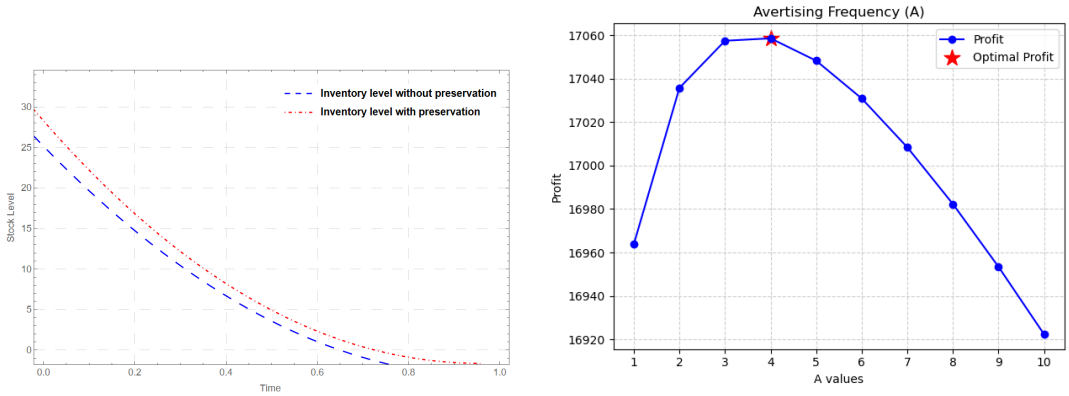


Figure 3: *Market dynamics in different scenarios*

In both scenarios, the model proves adaptable to various operational conditions, integrating multiple parameters for optimal decision-making in inventory management. Beyond numerical outcomes, the discussion provides managerial insights and practical implications.

From a managerial perspective, the model that includes without shortages emphasizes the importance of adopting forward-looking strategies (refer to Figure 3). It is observed that dynamic nature of pricing and strategic advertising effectively enhance overall profitability and competitiveness.It reveals that managers can optimize resource allocation, refine pricing strategies, and



(a) Inventory level and replenishment time

(b) Optimal value of advertisement frequency (A)

Figure 4: *Graphical representation of inventory level and optimal advertisement level*

boost sales efficiently. On the other hand, the example that deals with shortages focuses on mitigating risks and uncertainties. The model helps managers make informed decisions about backlogging, shortage, and opportunity costs. This is particularly relevant in real-world scenarios where unanticipated disruptions can impact inventory levels.

Further investigation considers models with and without shortages and the influence of preservation costs. The analysis reveals that the optimal replenishment time varies significantly across different scenarios. When considering preservation costs, the optimal replenishment time for the model without shortages is approximately 0.986282. This shows that businesses attempt to balance the need for timely replenishment and the financial implications of preserving inventory. However, the counterpart with shortages experiences a longer optimal replenishment time of 1.28626. This extension may be due to the complex interaction between shortage mitigation strategies and the time required to replenish stock, considering stochastic demand and expiration date-related deterioration.

Moreover, analyzing optimal prices provides valuable insights into the pricing strategies that optimize profit. Without shortages with preservation, the optimal price is \$598.624. This reveals that businesses can maximize profit under these conditions by setting a slightly lower price. However, when shortages are introduced into the model, the optimal price increases to \$606.711. This increase in price highlights the difficulty of maintaining a balance between price and shortage. Businesses may use higher prices to make up for potential profit losses caused by shortages. The complexity of determining the best pricing strategy becomes more evident when considering the costs of preserving deteriorating stock levels.

In the scenario with shortages, the model experiences a shortage period, representing a period during which demand exceeds the available stock. The duration of this shortage period is unaffected by preservation costs. The model without preservation costs exhibits a less shortage period of 0.19294 compared to 0.2021 in the model with preservation costs. This indicates the complex relationship between preservation expenses, reordering time, and periods of shortage. Therefore, decision-makers need to balance these factors carefully to optimize overall profit.

The findings also emphasize the strategic significance of pricing decisions in managing perishable inventory. The observed variations in optimal prices highlight the need for businesses to retailer their pricing strategies based on the specific dynamics of their operational environment. Higher prices may mitigate profit losses associated with shortages, but the impact on overall profitability should be carefully assessed. This sensitive understanding of optimal pricing gives managers actionable insights to navigate the delicate balance between pricing and inventory management. It also shows that the different advertising frequencies $(A = 1, 2, 3, ..., 10)$ impact on optimal profit differently (refer to Figure 4(a) and Figure 4(b)). Notably, the highest profit occurs at $A = 4$, emphasizing the crucial role of strategic decision-making in advertising frequency for

businesses aiming to maximize profitability within the given inventory management model. The rise at $A = 4$ suggests an optimal balance in allocating resources to advertising efforts. Increasing advertising frequency may lead to higher costs but declining revenue, while lowering frequency may result in suboptimal market reach. The finding at $A = 4$ indicates an optimal spot where the benefits from increased advertising frequency align with the costs, resulting in the highest profit. From a managerial perspective, this insight provides actionable recommendations for businesses to optimize their advertising strategies. Managers can strategically allocate resources, focusing on the frequency that maximizes profit, aligning with the trend in marketing where data-driven decisions play a vital role in resource allocation and campaign effectiveness.

Additionally, the result underscores the dynamic nature of advertising strategies. Managers should continuously evaluate and adjust advertising frequencies based on real-time data and market conditions. The findings contribute valuable insights to the literature on inventory management and advertising strategies, guiding practitioners in making informed decisions impacting the bottom line. The research proposes that businesses can achieve optimal profits in the given inventory model context by strategically selecting an advertising frequency, with the peak observed at A = 4. This knowledge empowers managers to fine-tune their advertising strategies, balance cost and impact, and ultimately enhance overall business performance.

7. ANALYSIS OF UNCERTAINTY

In the proposed model, the uncertainty in demand (represented by ϵ) is intricately linked to various factors, each of which can be categorized into specific probability distributions. Based on Consumer behaviour and preferences, it is observe that there is in subject to unpredictable shifts and societal influences, making a uniform distribution suitable for capturing this diversity. Due to market dynamics there is fluctuation on economic conditions and currency instability, align with the characteristics of normal distribution by the central limit theorem, providing a logical fit for capturing the aggregate impact of these macroeconomic factors. The rapid pace of technological advances and disruptive innovations reflects the uncertainty inherent in exponential processes, making the exponential distribution a suitable choice for modelling this source of variability. Retail and revenue disruptions, regulatory changes, global events, and competitive actions characterized by various potential outcomes resonate with the triangular distribution, allowing for a flexible representation of uncertainties that span different magnitudes.

Assumes that $F(\epsilon)$ is the distribution function of random variable ϵ, and $f(\epsilon)$ is the density function of ϵ. The presented graph examines the influence of various distributions on a mathematical model representing time-dependent demand (Figure 6). Four distributions—Uniform, Normal, Exponential, and Triangular are analyzed for their impact on the demand function. Each subplot illustrates the demand profile with different levels of randomness, providing insights into the dynamic relationship between time and demand. The graph is a versatile illustration applicable to scenarios ranging from inventory management to supply chain dynamics. Including mean and standard deviation, annotations offer a comprehensive understanding of the slight variations introduced by different distributions, contributing to a deeper understanding of time-dependent demand fluctuations (refer to figure 5).

This study delves into profit maximization in dynamic business models, focusing on the relationship between external factors and the resulting profit function (refer to figure 4(a)). The study reveals a profit function sensitive to changes in external factors represented by epsilon. The objective is to understand how different epsilon distributions, modelled by normal and uniform distributions, impact the profit function. Numerical simulations reveal distinct profit patterns for different epsilon distributions. The deterministic case shows a continuous increase in profit, lacking adaptability to varying external conditions. Introducing uncertainty through normal distributions with different means and standard deviations reveals varying profit landscapes. Higher mean values correlate with increased profitability. Uniform distributions introduce a more comprehensive range of possibilities, with broader and flatter curves suggesting potential volatility in business performance. The results highlight the sensitivity of the profit function to

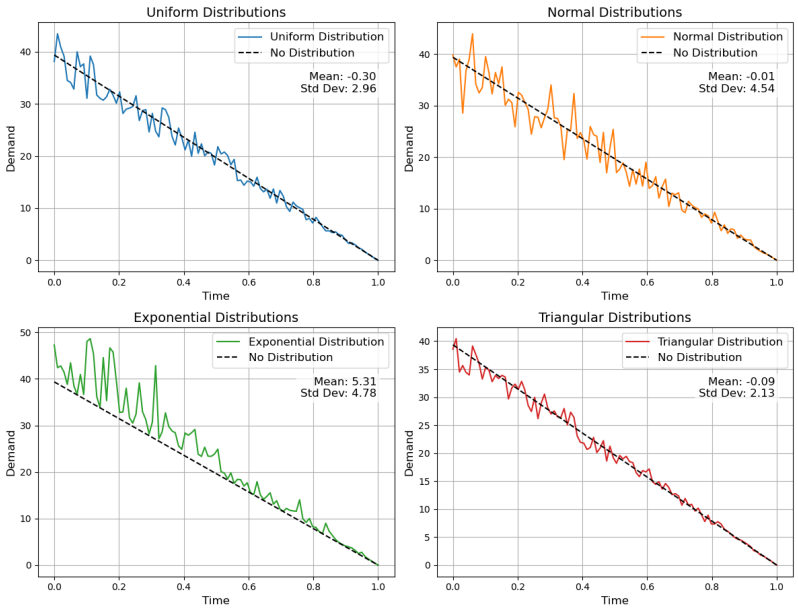


Figure 5: *Demand under the influence of various uncertainty factors*

changes in external factors, emphasizing the need for adaptive strategies to navigate uncertainties. This study provides a nuanced perspective on the impact of epsilon distributions on profit maximization in business models, emphasizing the importance of adaptive strategies to address external influences. It is graphically illustrate in figure 6.

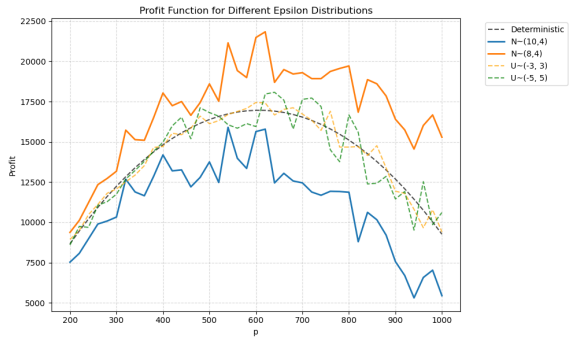


Figure 6: *Optimal profit under the influence of stochastic demand*

8. Conclusion and Further Work

This framework provides decision-makers with a robust approach to navigating the challenges of volatile markets. Through exploring managerial insights and results, this study offers practical guidance, emphasizing the importance of adaptability in decision-making:

1. The strategy for inventory management emphasizes the need for flexible and data-driven approaches to decision-making in pricing, advertising, and for managing retail market scenario. Decision-makers should recognize and adapt to the uncertainties inherent in the business environment.

2. The stochastic demand factor (ϵ) introduces variability, influencing optimal replenishment time and pricing. Different epsilon distributions, modelled by normal, exponential, triangular and uniform distributions, show distinct profit patterns. Higher mean values in

normal distributions correlate with increased profitability, offering valuable insights for decision-makers.

3. The optimal replenishment time (T) exhibits significant variation between scenarios, shows the model's adaptability to different operational conditions.

4. Optimal prices vary based on shortages, emphasizing the delicate balance required to optimize pricing strategies considering both price and shortage considerations. This managerial insight guides decision-makers in shaping pricing policies for overall profitability.

5. The highest profit occurs at an optimal advertising frequency of $A = 4$, indicating an optimal balance in resource allocation to advertising efforts. This result highlights the strategic importance of aligning advertising frequency with cost-effectiveness, offering actionable recommendations for businesses aiming to maximize profitability through advertising strategies. This insight guides decision-makers in making informed choices in deteriorating items, where preservation costs must be carefully weighed against potential benefits.

6. Continuous evaluation and adjustment of advertising frequencies based on real-time data emerge as crucial for optimizing advertising strategies. This insight aligns with contemporary marketing trends, emphasizing the importance of data-driven decision-making in resource allocation and campaign effectiveness. By adopting a dynamic approach, managers should ensure that advertising strategies remain aligned with market conditions.

The future scope outlined in the study suggests promising avenues for further research, ensuring the continued relevance and applicability of the model in addressing evolving business dynamics.

- Explore incorporating machine learning algorithms to enhance the model's adaptability. Machine learning can predict market trends and optimize pricing and advertising strategies, providing valuable insights for decision-makers in dynamic environments.

- Investigate the sustainability aspects of inventory management. Evaluate the environmental impact of different inventory strategies, ensuring alignment with corporate sustainability goals for responsible business practices.

- Examine the influence of real-time data analytics on advertising strategies. Understand how real-time data shapes decision-making in advertising, particularly in dynamic market conditions, offering insights for businesses aiming to stay competitive.

References

[1] Ghare P. M., Schrader G. P. (1963). A model for an exponentially decaying inventory. *Journal of Industrial Engineering*, 14(5):238–243.

[2] Covert R. P., Philip G. S. (1973). An EOQ model for items with Weibull distribution deterioration. *AIIE Transactions*, 5:323–326.

[3] Ogata K., Sanchez G. L. P. (1987). *System Dynamics*. Prentice-Hall, Hispanoamericana, 494–523.

[4] Chung K., Ting P. (1993). A heuristic for replenishment of deteriorating items with a linear trend in demand. *The Journal of the Operational Research Society*, 44:1235–1241.

[5] Sarkar B. (2013). A production-inventory model with probabilistic deterioration in two-echelon supply chain management. *Applied Mathematics and Computation*, 37(5):3138–3151.

[6] Giri B. C., Goswami A., Chaudhuri K. S. (1996). An EOQ model for deteriorating items with time-varying demand and costs. *The Journal of the Operational Research Society*, 47(11):1398–1405.

[7] Mishra V. K., Singh L. S., Kumar R. (2013). An inventory model for deteriorating items with time-dependent demand and time-varying holding cost under partial backlogging. *Journal of Industrial Engineering International*.

[8] Malumfashi M. L., Ali M. K. M. (2023). A production inventory model for non-instantaneous deteriorating items with two-phase production period, stock-dependent demand and shortages. *International Journal of Mathematics in Operational Research*, 24(2):173–193.

[9] Khedlekar U. K., Shukla D. (2013). Dynamic pricing model with logarithmic demand. *Opsearch*, 50(1):1–13.

[10] Khedlekar U. K., Shukla D., Namdeo A. (2016). Pricing policy for declining demand using item preservation technology. *SpringerPlus*, 5:1–11.

[11] Pang Z. (2011). Optimal dynamic pricing and inventory control with stock deterioration and partial backordering. *Operations Research Letters*, 39(5):375–379.

[12] Khedlekar U. K., Kumar L. (2024). Mathematical modelling for convertible items with rework using particle swarm optimisation. *International Journal of Systems Science: Operations & Logistics*, 11(1):2306222.

[13] Kumar L., Sharma K., Khedlekar U. K. (2024a). Dynamic pricing strategies for efficient inventory management with auto-correlative stochastic demand forecasting using exponential smoothing method. *Results in Control and Optimization*, 15:100432.

[14] Kumar L., Khedlekar S., Khedlekar U. K. (2024b). A comparative assessment of Holt Winter exponential smoothing and autoregressive integrated moving average for inventory optimization in supply chains. *Supply Chain Analytics*, 100084.

[15] San-José L. A., Sicilia J., Abdul-Jalbar B. (2021). Optimal policy for an inventory system with demand dependent on price, time and frequency of advertisement. *Computers & Operations Research*, 128:105169.

[16] Dash B., Pattnaik M., Pattnaik H. (2014). The impact of promotional activities and inflationary trends on a deteriorated inventory model allowing delay in payment. *Journal of Business and Management Sciences*, 2(3A):1–16.

[17] Dye C. Y., Hsieh T. P. (2012). An optimal replenishment policy for deteriorating items with effective investment in preservation technology. *European Journal of Operational Research*, 218(1):106–112.

[18] Duary A., Khan M. A. A., Pani S., Shaikh A. A., Hezam I. M., Alraheedi A. F., Gwak J. (2022). Inventory model with nonlinear price-dependent demand for non-instantaneous decaying items via advance payment and installment facility. *AIMS Mathematics*, 7(11):19794-19821.

[19] Khan M. A. A., Shaikh A. A., Konstantaras I., Bhunia A. K., Cárdenas-Barrón L. E. (2020). Inventory models for perishable items with advanced payment, linearly time-dependent holding cost and demand dependent on advertisement and selling price. *International Journal of Production Economics*, 230:107804.

[20] Khedlekar U. K., Tiwari, R. K. (2018). Imperfect production model for sensitive demand with shortage. *Reliability: Theory & Applications*, 13(4(51)):43-54.

[21] Mahapatra A. S., Dasgupta A., Shaw A., Sarkar B. (2022). An inventory model with uncertain demand under preservation strategy for deteriorating items. *RAIRO-Operations Research*, 56(6):4251–4280.

[22] Maihami R., Karimi B. (2014). Optimizing the pricing and replenishment policy for non-instantaneous deteriorating items with stochastic demand and promotional efforts. *Computers & Operations Research*, 51:302–312.

[23] Md Mashud A. H., Hasan M. R., Wee H. M., Daryanto Y. (2020). Non-instantaneous deteriorating inventory model under the joined effect of trade-credit, preservation technology and advertisement policy. *Kybernetes*, 49(6):1645–1674.

[24] Kumar, L., Ghoshi, P. S., Saxena, S., Sharma, K. (2024). A Comparative study of inventory modelling: deterministic over stochastic approach. *Reliability: Theory & Applications*, 19(1(77)):804-818.

[25] Khedlekar, U. K., Kumar, L., Sharma, K., Dwivedi, V. (2024). An optimal sustainable production policy for imperfect production system with stochastic demand, price, and machine failure with FRW policy under carbon emission. *Process Integrations and Optimisations for sustainability*, 8:919–938.

RT&A, No 4(80)
Volume 19, December, 2024

A NEW GENERALIZATION OF AREA BIASED DISTRIBUTION WITH PROPERTIES AND ITS APPLICATION TO REAL LIFE DATA

P. Pandiyan and R. Jothika

•

(1). Professor, Department of Statistics, Annamalai University
(2). Research Scholar, Department of Statistics, Annamalai University
Email: pandianau@gmail.com, rjothika2023@gmail.com

Abstract

This paper proposed a new generalization of the Samade distribution. The term "area biased samade distribution" refers to the recently created distribution model. After studying the various structural features, entropies, order statistics, moments, generating functions for moments, survival functions, and hazard functions were calculated. The parameters of the suggested model are estimated using the maximum likelihood estimation technique. Ultimately, a fitting of an application to a real-life blood cancer data set reveals a good fit.

Keywords: area biased samade distribution, reliability analysis, maximum likelihood estimation, order statistics, entropies.

I. Introduction

Samade distribution is a recently introduced two parametric lifetime distribution. Introduced by [1] has been discussed by samade probability distribution. Its properties and application to real lifetime data and determine the its various mathematical and statistical properties. Additionally, modelling and analysing real life data are essential in many practical fields including engineering, health, finance, and insurance. Area biased samade distribution a new parametric life time distribution is more flexible in the manner it processes lifetime data than the samade distribution. the area biased version of the samade distribution in this study exhibits greater adaptability in handling life time than the samade distribution. after fitting a real-life data set, it was discovered that the area biased samade distribution gives better performance.

In [5] firstly introduced the concept of weighted distributions to model ascertainment biases in the data and later on [14] reformulated it in a unifying theory for problems where the observations fail in non-experimental, non-replicated and non-random manner. It has been observed that when an investigator records observations in the nature according to certain stochastic model, the distribution of the recorded observations will be different from the original distribution unless every observation given an equal chance of being recorded. Suppose the original observation x_0 comes from a distribution having probability density function (pdf)

$f_0(x,\theta_1)$ where θ_1 may be a vector of parameters and the observation x is recorded according to a probability re weighted by weight function $w(x,\theta_2) > 0$, $\theta_2$being a new parameter vector, then x comes from a distribution having pdf

$$f(x;\theta_1,\theta_2) = AW(x;\theta_2)f_0(x,\theta_1)$$

Where A is a normalizing constant if should be noted that such type of distribution is known as weighted distributions. The weighted distribution with weight function $w(x,\theta_2) = x$ are called length biased distribution or size biased distribution. In [12] proposed the weighted pareto type-II distribution as a new model for handling medical science data and studied its statistical properties and applications. In [7] established a new length biased distribution. and a new length biased distribution was introduced by [15]. In [9] presented Area biased quasi-Transmuted uniform distribution. In [4] presented the size biased Zeghdoudi distribution and discuss its various statistical properties and application. In [8] provided a length and area biased exponentiated Weibull distribution. and recently [2] discussed the characterization and estimation of Area biased quasi-Akash distribution. In this study our motives are to prove that area biased version of samade distribution is more flexible and fits better in real life time data.

II. Area-Biased Samade Distribution

Let random variable be x with scale parameter alpha and shape parameter theta then the probability density function (pdf) and cumulative distribution function (cdf) of the area biased Samade distribution.

The probability density function of Samade distribution is given by

$$f(x;\alpha,\theta) = \frac{\theta^4}{\theta^4+6\alpha}(\theta+\alpha x^3)e^{-\theta x} \;; \qquad x,\theta,\alpha > 0 \tag{1}$$

The cumulative distribution function (cdf) of Samade distribution is given by

$$F(x;\alpha,\theta) = 1 - \left(\frac{\theta^4+\alpha\left(6+x\theta\left(6+x\theta(3+x\theta)\right)\right)}{6\alpha+\theta^4}\right)e^{-\theta x};\; x,\theta,\alpha > 0 \tag{2}$$

considered a random variable x with a probability density function $f(x)$. Let $w(x)$ be a non-negative weight function. Denote a new probability density function.

$$f_w(x) = \frac{w(x)f(x)}{E[w(x)]} \quad ; \quad x > 0$$

Where its non-negative weight function be $w(x)$ and $E(w(x)) = \int w(x)f(x)dx < \infty$.

Depending upon the various choices of weight function especially when $w(x) = x^c$, result is called weighted distribution. In this paper, we have to obtain the Area biased version of samade distribution, so we will consider as $w(x) = x^{\,2}$ to obtain the Area biased samade distribution. Then, the probability density function of Area biased version is given as

$$f_a(x) = \frac{x^2 f(x)}{E(X^2)} \tag{3}$$

Where

$$E(X^2) = \int_0^\infty x^2 f(x) dx$$

$$E(X^2) = \frac{2\theta^4 + 120\alpha}{\theta^4(\theta^4 + 6\alpha)} \tag{4}$$

By substituting equations (1) and (4) in equation (3), we will get the probability density function of Area biased samade distribution

$$f_a(x) = \frac{\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3)e^{-\theta x} \tag{5}$$

And the cumulative distribution function of area biased samade distribution can be obtained as

$$F_a(x) = \int_0^x f_a(x)\, dx$$

$$F_a(x) = \int_0^x \frac{\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3)e^{-\theta x}\ dx$$

$$F_a(x) = \frac{\theta^6}{2\theta^4 + 120\alpha} \int_0^x x^2(\theta + \alpha x^3)e^{-\theta x}\ dx$$

$$F_a(x) = \frac{\theta^6}{2\theta^4 + 120\alpha} \int_0^x (x^2\theta + \alpha x^5)e^{-\theta x}\ dx$$

$$F_a(x) = \frac{\theta^6}{2\theta^4 + 120\alpha} \left[\int_0^x x^2\theta\, e^{-\theta x}\, dx + \int_0^x \alpha x^5\, e^{-\theta x}\, dx \right]$$

$$F_a(x) = \frac{\theta^6}{2\theta^4 + 120\alpha} \left[\theta \int_0^x x^2\, e^{-\theta x}\, dx + \alpha \int_0^x x^5\, e^{-\theta x}\, dx \right]$$

Put $\theta x = t, \quad x = \frac{t}{\theta}, \qquad dx = \frac{dt}{\theta}$

When $x \to 0, t \to 0, and\ x \to x,\ \ t \to \theta x$

$$F_a(x) = \frac{\theta^6}{2\theta^4 + 120\alpha} \left[\theta \int_0^{\theta x} \left(\frac{t^2}{\theta^3}\right) e^{-t}\, dt + \alpha \int_0^{\theta x} \left(\frac{t^5}{\theta^6}\right) e^{-t}\, dt \right] \tag{6}$$

After simplification of equation (6), we obtain the cumulative distribution function of area biased samade distribution.

$$F_a(x) = \frac{1}{2\theta^4 + 120\alpha} [\theta^4 \gamma(3, \theta x) + \alpha\gamma(6, \theta x)] \tag{7}$$

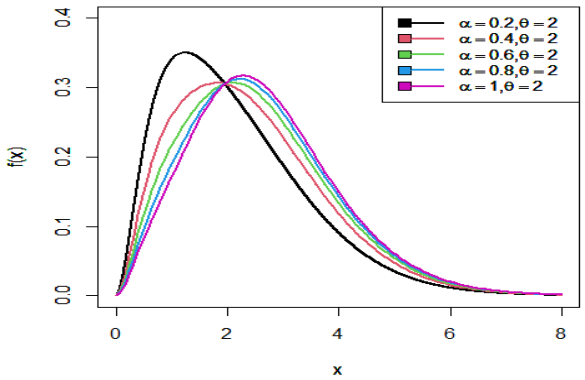


Figure 1: *Pdf plot of area biased samade distribution*

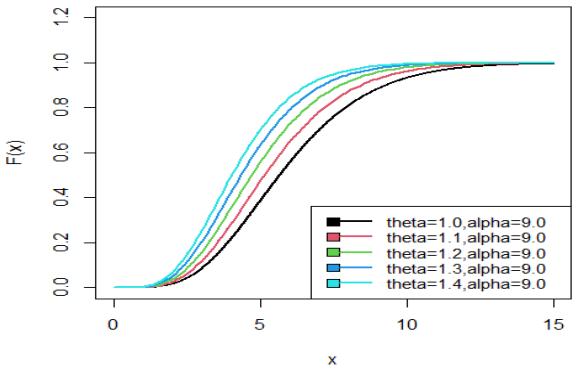


Figure 2: *Cdf plot of area biased Samade distribution*

III. Reliability Analysis

In this section, we will discuss the survival function, hazard function, reverse hazard function, cumulative hazard function, Odds rate, Mills ratio and, Mean Residual function for the proposed area biased samade distribution**.**

I. Survival Function

The survival function or the reliability function of the area biased Samade distribution is given by

$$S(x) = 1 - F_a(x)$$

$$S(x) = 1 - \frac{1}{2\theta^4 + 120\alpha}[\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)]$$

II. Hazard Function

The corresponding hazard function or failure rate of the area biased Samade distribution is given by

$$h(x) = \frac{f_a(x)}{S(x)}$$

$$h(x) = \frac{\frac{\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3)e^{-\theta x}}{1 - \frac{1}{2\theta^4 + 120\alpha}[\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)]}$$

$$h(x) = \frac{x^2\theta^6(\theta + \alpha x^3)e^{-\theta x}}{(2\theta^4 + 120\alpha) - [\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)]}$$

III. Reverse Hazard Function

Reverse hazard function of area biased samade distribution is given by

$$h_r(x) = \frac{f_a(x)}{F_a(x)}$$

$$h_r(x) = \frac{x^2\theta^6(\theta + \alpha x^3)e^{-\theta x}}{\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)}$$

IV. Odds Rate Function

Odds Rate function of area biased samade distribution is given by

$$O(x) = \frac{F_a(x)}{1 - F_a(x)}$$

$$O(x) = \frac{\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)}{2\theta^4 + 120\alpha - \theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)}$$

V. Cumulative Hazard Function

Cumulative hazard function of area biased samade distribution is given by

$$H(x) = -\ln(1 - F_a(x))$$

$$H(x) = -\ln\left(1 - \frac{\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)}{2\theta^4 + 120\alpha}\right)$$

VI. Mills Ratio

$$\text{Mills ratio} = \frac{1}{\mathrm{h_r(x)}}$$

$$\text{Mills ratio} = \frac{\theta^4\gamma(3,\theta x) + \alpha\gamma(6,\theta x)}{x^2\theta^6(\theta + \alpha x^3)e^{-\theta x}}$$

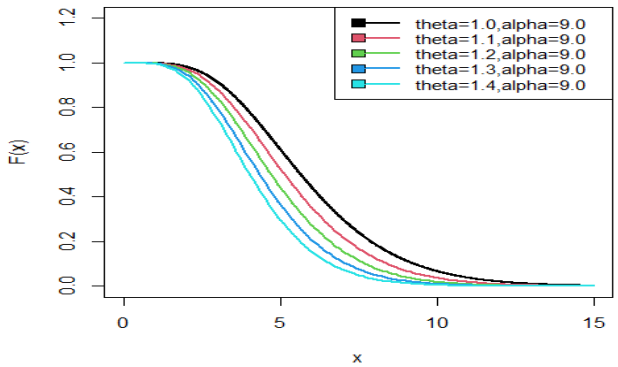


Figure 3: *Survival plot of area biased Samade distribution*

IV. Statistical Properties

In this section, discuss about the different statistical properties of area biased samade distribution especially its moments, harmonic mean, moment generating function and characteristic function.

I. Moments

let the random variable X represents area biased samade distribution with parameters θ and α, then the r^{th} order moment $E(X^r)$ of X about origin can be obtained as

$$E(X^r) = \mu_r{}' = \int_0^\infty x^r f_a(x)dx$$

$$E(X^r) = \int_0^\infty x^r \frac{\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3)e^{-\theta x}dx$$

$$E(X^r) = \frac{\theta^6}{2\theta^4 + 120\alpha} \int_0^\infty x^{r+2}(\theta + \alpha x^3)e^{-\theta x}dx$$

$$E(X^r) = \frac{\theta^6}{2\theta^4 + 120\alpha}\left(\theta \int_0^\infty x^{(r+3)-1}e^{-\theta x}dx + \alpha \int_0^\infty x^{(r+6)-1}e^{-\theta x}dx\right) \quad (8)$$

After simplification of equation (14) get

$$E(X^r) = \mu_r{}' = \frac{\theta^4 \Gamma r + 3 + \alpha \Gamma r + 6}{\theta^r(2\theta^4 + 120\alpha)} \quad (9)$$

Putting r = 1, 2, 3 and 4 in equation (9), obtain the first four moments of Area biased samade distribution.

$$E(X^1) = \mu_1{}' = \frac{6\theta^4 + 720\alpha}{\theta(2\theta^4 + 120\alpha)}$$

$$E(X^2) = \mu_2{}' = \frac{24\theta^4 + 5040\alpha}{\theta^2(2\theta^4 + 120\alpha)}$$

$$E(X^3) = \mu_3{}' = \frac{120\theta^4 + 40320\alpha}{\theta^3(2\theta^4 + 120\alpha)}$$

$$E(X^4) = \mu_4{}' = \frac{720\theta^4 + 362880\alpha}{\theta^4(2\theta^4 + 120\alpha)}$$

$$\text{variance} = \mu_2{}' - (\mu_1{}')^2$$

$$\text{variance} = \frac{(24\theta^4 + 5040\alpha)(2\theta^4 + 120\alpha) - (6\theta^4 + 720\alpha)^2}{\theta^2(2\theta^4 + 120\alpha)^2}$$

Standard Deviation

$$S.\mathrm{D}(\sigma) = \frac{\sqrt{(24\theta^4 + 5040\alpha)(2\theta^4 + 120\alpha) - (6\theta^4 + 720\alpha)^2}}{\theta(2\theta^4 + 120\alpha)}$$

Coefficient Of Variation

$$C.V\left(\frac{\sigma}{\mu}\right)=\frac{\sqrt{(24\theta^4+5040\alpha)(2\theta^4+120\alpha)-(6\theta^4+720\alpha)^2}}{6\theta^4+720\alpha}$$

Dispersion

$$\text{Dispersion}=\frac{\sigma^2}{\mu}$$

$$\text{Dispersion}=\left[\frac{(24\theta^4+5040\alpha)(2\theta^4+120\alpha)-(6\theta^4+720\alpha)^2}{\theta(2\theta^4+120\alpha)(6\theta^4+720\alpha)}\right]$$

II. Harmonic Mean

The harmonic mean for the proposed model can be obtained as

$$H.M=E\left(\frac{1}{x}\right)=\int_0^\infty \frac{1}{x}f_a(x)dx$$

$$H.M=E\left(\frac{1}{x}\right)=\int_0^\infty \frac{1}{x}\frac{\theta^6}{2\theta^4+120\alpha}x^2(\theta+\alpha x^3)e^{-\theta x}dx$$

After simplification, obtain

$$H.M=\frac{\theta^6}{2\theta^4+120\alpha}\left[\theta\int_0^\infty x\,e^{-\theta x}dx+\int_0^\infty x^4e^{-\theta x}dx\right]$$

$$H.M=\frac{\theta^6}{2\theta^4+120\alpha}\left[\theta\left(\frac{1!}{\theta^{1+1}}\right)+\alpha\left(\frac{4!}{\theta^{4+1}}\right)\right]$$

After simplification get

$$H.M=\frac{\theta(\theta^4+6\alpha)}{(2\theta^4+120\alpha)}$$

III. Moment Generating Function and Characteristic Function

Suppose the random variable X follows Area biased samade distribution with parameters θ and α, then the MGF of X can be obtained as:

$$M_X(t)=E(e^{tx})=\int_0^\infty e^{tx}f_a(x)dx$$

Using Taylor's series, can be obtain

$$M_X(t)=E(e^{tx})=\int_0^\infty\left(1+tx+\frac{(tx)^2}{2!}+\cdots\right)f_a(x)dx$$

$$M_X(t)=\int_0^\infty\sum_{j=0}^\infty\frac{t^j}{j!}x^jf_a(x)dx$$

$$M_X(t)=\sum_{j=0}^\infty\frac{t^j}{j!}\mu_j'$$

$$M_X(t)=\sum_{j=0}^\infty\frac{t^j}{j!}\left(\frac{\theta^4\Gamma j+3+\alpha\Gamma j+6}{\theta^r(2\theta^4+120\alpha)}\right)$$

$$M_X(t) = \frac{1}{2\theta^4 + 120\alpha} \sum_{j=0}^{\infty} \frac{t^j}{j!\ \theta^j} \ (\theta^4 \Gamma j + 3 + \alpha \Gamma j + 6)$$

Similarly, the characteristic function area biased samade distribution can be obtained as $\phi_X(t) = M_X(it)$

$$M_X(it) = \frac{1}{2\theta^4 + 120\alpha} \sum_{j=0}^{\infty} \frac{it^j}{j!\ \theta^j} \ (\theta^4 \Gamma j + 3 + \alpha \Gamma j + 6)$$

V. Order statistics

In this section, derived the distributions of order statistics from the area biased Samade distribution.

Let $X_{(1)}, X_{(2)}, X_{(3)}, \ldots . X_{(n)}$ be the order statistics of the random sample$X_1, X_2, X_3, \ldots . X_n$ selected from area biased samade distribution. Then the probability density function of the r^{th}order statistics X(r) is defined as.

$$f_{X(r)}(x) = \frac{n!}{(r-1)!\,(n-r)!} f_X(x)[F_X(x)]^{r-1}[1 - F_X(x)]^{n-r} \tag{10}$$

Using equations (5) and (7) in equation, get the probability density function of r^{th} order statistics of area biased samade distribution.

$$f_{X(r)}(x) = \frac{n!}{(r-1)!\,(n-r)!} \left(\frac{\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3) e^{-\theta x} \right)$$
$$\times \left(\frac{1}{2\theta^4 + 120\alpha} [\theta^4 \gamma(3, \theta x) + \alpha \gamma(6, \theta x)] \right)^{r-1}$$
$$\times \left(1 - \frac{1}{2\theta^4 + 120\alpha} [\theta^4 \gamma(3, \theta x) + \alpha \gamma(6, \theta x)] \right)^{n-r}$$

$$f_{X(1)}(x)$$
$$= \frac{n\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3) e^{-\theta x} \times \left(1 - \frac{1}{2\theta^4 + 120\alpha} [\theta^4 \gamma(3, \theta x) + \alpha \gamma(6, \theta x)] \right)^{n-1}$$

$$f_{X(n)}(x) = \frac{n\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3) e^{-\theta x} \times \left(\frac{1}{2\theta^4 + 120\alpha} [\theta^4 \gamma(3, \theta x) + \alpha \gamma(6, \theta x)] \right)^{n-1}$$

VI. Likelihood Ratio Test

In this section, derive the likelihood ratio test from the area biased samade distribution. Let $X_1, X_2, \ldots, X_n$ be a random sample from the area biased samade distribution. To test the hypothesis.

$$H_0 : f(x) = f(x{:}\,\theta) \text{ against}$$

$$H_1 : f(x) = f_a(x{:}\,\theta)$$

In test whether the random sample of size n comes from the samade distribution or area biased samade distribution, the following test statistics is used.

$$\Delta = \frac{L_1}{L_0} = \prod_{i=1}^{n} \frac{f_a(x{:}\,\theta)}{f(x{:}\,\theta)}$$

$$\Delta = \frac{L_1}{L_0} = \prod_{i=1}^{n} \frac{\theta^2(\theta^4 + 6\alpha)}{2\theta^4 + 120\alpha} {x_i}^2$$

$$\Delta = \frac{L_1}{L_0} = \left(\frac{\theta^2(\theta^4 + 6\alpha)}{2\theta^4 + 120\alpha}\right)^n \prod_{i=1}^{n} {x_i}^2$$

We should reject the null hypothesis, if

$$\Delta = \left(\frac{\theta^2(\theta^4 + 6\alpha)}{2\theta^4 + 120\alpha}\right)^n \prod_{i=1}^{n} {x_i}^2 > k$$

Or equivalently, reject the null hypothesis

$$\Delta^* = \prod_{i=1}^{n} {x_i}^2 > k\left(\frac{2\theta^4 + 120\alpha}{\theta^2(\theta^4 + 6\alpha)}\right)^n$$

$$\Delta^* = \prod_{i=1}^{n} {x_i}^2 > k^* \ \ where\ k^* = k\left(\frac{2\theta^4 + 120\alpha}{\theta^2(\theta^4 + 6\alpha)}\right)^n$$

For large sample size n, 2 log Δ is distribution as chi-square variates with one degree of freedom. Thus, we rejected the null hypothesis, when the probability value is given by $p(\Delta^* > \alpha^*)$,where $\alpha^* = \prod_{i=1}^{n} {x_i}^2$is less than level of significance and $\prod_{i=1}^{n} {x_i}^2$is the observed value of the statistics Δ^*.

VII. Bonferroni and Lorenz Curves

In this section, derived the Bonferroni and Lorenz curves and from the area biased samade distribution.

The Bonferroni and Lorenz curve is a powerful tool in the analysis of distributions and has applications in many fields, such as economies, insurance, income, reliability, and medicine. The Bonferroni and Lorenz cures for a X be the random variable of a unit and $f(x)$ be the probability density function of x. $f(x)dx$ will be represented by the probability that a unit selected at random is defined as

$$B(p) = \frac{1}{p{\mu_1}'} \int_0^q x\, f_a(x)dx$$

And

$$L(p) = \frac{1}{{\mu_1}'} \int_0^q x\, f_a(x)dx$$

Where ${\mu_1}' = E(X) = \frac{6\theta^4 + 720\alpha}{\theta(2\theta^4 + 120\alpha)}$ and $q = F^{-1}(p)$

$$B(p) = \frac{\theta(2\theta^4 + 120\alpha)}{p(6\theta^4 + 720\alpha)} \int_0^q x\, \frac{\theta^6}{2\theta^4 + 120\alpha} x^2(\theta + \alpha x^3)e^{-\theta x}dx$$

$$B(p) = \frac{\theta^7}{p(6\theta^4 + 720\alpha)} \int_0^q x^3(\theta + \alpha x^3)e^{-\theta x}dx$$

$$B(p) = \frac{\theta^7}{p(6\theta^4 + 720\alpha)} \int_0^q x^3\theta e^{-\theta x}dx + \int_0^q \alpha x^6 e^{-\theta x}dx$$

After simplification we get

$$B(p) = \frac{\theta^4\gamma(4, \theta q) + \alpha\gamma(7, \theta q)}{p(6\theta^4 + 720\alpha)}$$

Where $L(p) = pB(p)$

$$L(p) = \frac{\theta^4\gamma(4,\theta q) + \alpha\gamma(7,\theta q)}{(6\theta^4 + 720\alpha)}$$

VIII. Entropies

The concept of entropy is important in various fields such as probability, statistics, physics, communication theory and economics. Entropies quantify the diversity, uncertainty, or randomness of a system. Entropy of a random variable X is a measure of variation of the uncertainty.

I. Shannon Entropy

Shannon entropy of the random variable X such that area biased samade distribution is defined as

$$S_\lambda = -\int_0^\infty f(x)\log(f(x))dx \qquad \lambda > 0, \lambda \neq 1$$

$$S_\lambda = -\int_0^\infty f(x)\log(f_a(x))dx$$

$$S_\lambda = -\int_0^\infty \frac{\theta^6}{2\theta^4 + 120\alpha}x^2(\theta + \alpha x^3)e^{-\theta x}dx \times \log\left(\frac{\theta^6}{2\theta^4 + 120\alpha}x^2(\theta + \alpha x^3)e^{-\theta x}\right)dx$$

II. Renyi Entropy

The Renyi entropy is important in ecology and statistics as an index of diversity. It is also important in quantum information, where it can be used as a measure of entanglement. For a given probability distribution, Renyi entropy is given by

$$R(\beta) = \frac{1}{1-\beta}\log\int_0^\infty ({f_a}^\beta(x))dx$$

Where $\beta > 0\ and\ \beta \neq 1$

$$R(\beta) = \frac{1}{1-\beta}\log\int_0^\infty \left(\frac{\theta^6}{2\theta^4 + 120\alpha}x^2(\theta + \alpha x^3)e^{-\theta x}dx\right)^\beta dx$$

$$R(\beta) = \frac{1}{1-\beta}\log\left(\frac{\theta^6}{2\theta^4 + 120\alpha}\right)^\beta \int_0^\infty x^{2\beta}(\theta + \alpha x^3)^\beta e^{-\theta\beta x}dx$$

Using binomial expansion in above equation and can be obtain

$$R(\beta) = \frac{1}{1-\beta}\log\left(\frac{\theta^6}{2\theta^4 + 120\alpha}\right)^\beta \sum_{j=0}^{\beta}\sum_{k=0}^{\infty}\binom{\beta}{j}\frac{\log(\alpha)^k j^k}{k!}\int_0^\infty x^{2\beta+j}e^{-\theta\beta x}dx$$

$$R(\beta)=\frac{1}{1-\beta}\log\left(\frac{\theta^6}{2\theta^4+120\alpha}\right)^{\beta}\sum_{j=0}^{\beta}\sum_{k=0}^{\infty}\binom{\beta}{j}\frac{\log(\alpha)^k j^k}{k!}\frac{\Gamma 2\beta+j+1}{(\theta\beta)^{2\beta+j+1}}$$

III. Tsallis Entropy

A generalization of Boltzmann-Gibbs (B-G) statistical mechanics initiated by Tsallis has attracted a great deal of attention. This generalization of (B-G) statistics was proposed firstly by introducing the mathematical expression of Tsallis entropy (Tsallis, 1988) for a continuous random variable which is defined as follow.

$$T_\lambda=\frac{1}{\lambda-1}\left(1-\int_0^{\infty}{f_a}^{\lambda}(x)dx\right)$$

$$T_\lambda=\frac{1}{\lambda-1}\left(1-\int_0^{\infty}\left(\frac{\theta^6}{2\theta^4+120\alpha}x^2(\theta+\alpha x^3)e^{-\theta x}dx\right)^{\lambda}dx\right)$$

$$T_\lambda=\frac{1}{\lambda-1}\left(1-\left(\frac{\theta^6}{2\theta^4+120\alpha}\right)^{\lambda}\int_0^{\infty}x^{2\lambda}(\theta+\alpha x^3)^{\lambda}e^{-\theta x}dx\right)$$

Using by binomial expansion in above equation

$$T_\lambda=\frac{1}{\lambda-1}\left(1-\left(\frac{\theta^6}{2\theta^4+120\alpha}\right)^{\lambda}\sum_{j=0}^{\lambda}\sum_{k=0}^{\infty}\binom{\lambda}{j}\frac{\log(\alpha)^k j^k}{k!}\frac{\Gamma 2\lambda+j+1}{(\theta\lambda)^{2\lambda+j+1}}\right)$$

IX. Estimations of Parameter

In this section, the maximum likelihood estimates and Fisher's information matrix of the area biased samade distribution parameter is given.

Maximum Likelihood estimation (MLE)and Fisher's Information Matrix

Consider $(x_1, x_2, x_3, \ldots x_n)$be a random sample of size n from the area biased samade distribution. Then the likelihood function is given by

$$L(x)=\prod_{i=1}^{n}f_a(x)$$

$$L(x)=\prod_{i=1}^{n}\left(\frac{\theta^6}{2\theta^4+120\alpha}{x_i}^2(\theta+\alpha{x_i}^3)e^{-\theta x_i}\right)$$

$$L(x)=\frac{\theta^{6n}}{(2\theta^4+120\alpha)^n}\prod_{i=1}^{n}\left({x_i}^2(\theta+\alpha{x_i}^3)e^{-\theta x_i}\right)$$

The log likelihood function is given by

$$\log L = 6n \log \theta\,(2\theta^4 + 120\alpha) + 2\sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \log(\theta + \alpha {x_i}^3) - \theta \sum_{i=1}^{n} x_i \qquad (11)$$

Now differentiating the log likelihood equation (11) with respect to parameters θ and α we must satisfy the normal equations as

$$\frac{\partial \log L}{\partial \theta} = \frac{6n}{\theta} - n\left(\frac{8\theta^3}{2\theta^4 + 120\alpha}\right) + n\left(\frac{1}{(\theta + \alpha {x_i}^3)}\right) - \sum_{i=1}^{n} x_i = 0 \qquad (12)$$

$$\frac{\partial \log L}{\partial \alpha} = -n\left(\frac{120}{2\theta^4 + 120\alpha}\right) + n\left(\frac{{x_i}^3}{(\theta + \alpha {x_i}^3)}\right) = 0 \qquad (13)$$

The equation (12) and (13) gives the maximum likelihood estimation of the parameters for the area biased samade distribution. However, the equation cannot be solved analytically, thus solved numerically using R programming with data set.

To obtain confidence interval use the asymptotic normality results. have that if $\hat{\lambda} = (\hat{\theta}, \hat{\alpha})$ denotes the MLE of $\lambda = (\theta, \alpha)$. state the results as follows.

$$\sqrt{n}(\hat{\lambda} - \lambda) \longrightarrow N_2(0, I^{-1}(\lambda))$$

Where $I(\lambda)$ is fisher's information matrix. i.e.

$$I(\lambda) = \begin{matrix} E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right] & E\left[\frac{\partial^2 \log L}{\partial \theta \partial \alpha}\right] \\ \left\lfloor E\left[\frac{\partial^2 \log L}{\partial \theta \partial \alpha}\right] \right. & \left. E\left[\frac{\partial^2 \log L}{\partial \alpha^2}\right] \right\rfloor \end{matrix}$$

Here we see

$$E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right] = -\frac{6n}{\theta^2} - n\left(\frac{2880\theta^2\alpha - 16\theta^2}{(2\theta^4 + 120\alpha)^2}\right) + n\left(\frac{(\theta + \alpha {x_i}^3) - 1}{(\theta + \alpha {x_i}^3)^2}\right)$$

$$E\left[\frac{\partial^2 \log L}{\partial \alpha^2}\right] = n\left(\frac{14400}{(2\theta^4 + 120\alpha)^2}\right) - n\left(\frac{\alpha {x_i}^3}{(\theta + \alpha {x_i}^3)^2}\right)$$

$$E\left[\frac{\partial^2 \log L}{\partial \theta \partial \alpha}\right] = -n\left(\frac{960\theta^3}{(2\theta^4 + 120\alpha)^2}\right) - n\left(\frac{\alpha {x_i}^3}{(\theta + \alpha {x_i}^3)^2}\right)$$

Since λ being unknown, estimate $I^{-1}(\lambda)$ by $I^{-1}(\hat{\lambda})$and this can be used to obtain asymptotic confidence interval for θ and α.

X. Applications

In this section, we have fitted a real lifetime data set in area biased Samade distribution to discuss its goodness of fit and the fit has been compared over Samade and Aradhana distributions. The real lifetime data set is given below as

The following real lifetime data set consists of 40 patients suffering from blood cancer (leukemia) reported from one of the ministry of health hospitals in Saudi Arabia (see Abouammah et al.). The ordered lifetimes (in year) is given below in Data set 1 as:

Data set 1: Data regarding the blood cancer (leukemia) patients reported by [3]

(0.315, 0.496, 0.616, 1.145, 1.208, 1.263, 1.414, 2.025, 2.036, 2.162, 2.211, 2.37, 2.532, 2.693, 2.805, 2.91, 2.912, 3.192, 3.263, 3.348, 3.348, 3.427, 3.499, 3.534, 3.767, 3.751, 3.858 ,3.986, 4.049, 4.244, 4.323, 4.381, 4.392, 4.397, 4.647, 4.753, 4.929, 4.973 ,5.074 ,5.381)

To compare to the goodness of fit of the fitted distribution, the following criteria: Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Akaike Information Criteria Corrected (AICC) and $-2\log L$.
AIC, BIC, AICC and $-2\log L$ can be evaluated by using the formula as follows.

$$AIC = 2k - 2\log L\,, \quad BIC = k\log n - 2\log L \ \ and\ AICC = AIC + \frac{2k(k\ +\ 1)}{(n\ -\ k\ -\ 1)}$$

Where, k = number of parameters, n sample size and $-2\log L$ is the maximized value of loglikelihood function.

Table 1: *MLEs AIC, BIC, AICC, and -2log L of the fitted distribution for the given data set 1*

Distribution	ML Estimates	-2 log L	AIC	BIC	AICC
Area biased Samade distribution	$\hat{\alpha} = \mathbf{0.5475209}$ $(\mathbf{0.4673469})$ $\hat{\theta} = \mathbf{1.7122824}$ $(\mathbf{0.1423247})$	**141.1354**	**145.1354**	**148.5131**	**145.4597**
Samade distribution	$\hat{\alpha} = 4.4717985$ (4.5855767) $\hat{\theta} = 1.2041767$ (0.1058379)	144.0608	148.0680	151.4385	148.3923
Aradhana	$\hat{\theta} = 0.74955550$ (0.07008456)	153.1011	155.1011	156.79	155.4254

From table 1 it can be clearly observed and seen from the results that the area biased samade distribution have the lesser AIC, BIC, AICC, -2log L, and values as compared to the Samade and Aradhana, which indicates that the area biased samade distribution better fits than the Samade and Aradhana distributions. Hence, it can be concluded that the area biased samade distribution leads to a better fit over the other distributions.

XI. Conclusion

Research has focused a great deal of attention on selecting an appropriate model for fitting survival data. In this paper, the samade distribution is extended to provide a new distribution called the area biased samade distribution for the model's lifetime data. It has various special cases that have been presented in the paper. The statistical properties of the distribution have been studied, including survival and hazard functions, moments, mean and, median deviations, moment generating functions, Entropies, Bonferroni and Lorenz curve, and order statistics. The Inference of parameters for a area biased Samade distribution was obtained using the method of maximum likelihood estimates. When the parameters have been estimated using the maximum likelihood

method, a good performance is seen. The application of statistical distributions is critical for medical research and can significantly affect public health, especially for cancer patients.
Thus, the usefulness of this distribution is illustrated through its applications to the survival of some cancer patients, including both complete and censored cases. In using various goodness-of-fit criteria, including AIC, BIC, AICC, and -2logL, the results demonstrate the superior performance of the area biased samade distribution. Overall, it is intended that the area biased Samade distribution that is given in table 1 will offer a better fit than other existing distributions for simulating real life data in survival analysis, specifically for cancer data.

References

[1] Aderoju, S. (2021). Samade probability distribution: its Properties and Application to real lifetime data. *Asian Journal of Probability and Statistics*, 11:1-11.

[2] Ade, R. B. et al. (2021). Characterization and estimation of area biased quasi-Akash distribution. *International Journal of Mathematics Trends and Technology*, 67:53-59.

[3] Abouammoh, A. M. Ahmed, R. and Khalique, A. (2000). On new renewal better than used classes of life distribution. *Statistics and Probability Letters*, 48:189-194.

[4] Chouia, S. Zeghdoudi, H. Raman, V. and Beghriche, A. (2021). A new size biased distribution with application. *Journal of Applied Probability and Statistics*, 16:111-125.

[5] Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. Annals of Eugenics, 6:13-25.

[6] Gupta, R. C. and Keating, J. P. (1986). Relations for reliability measures under length biased sampling. *Scandinavian Journal of Statistics*, 13:49-56.

[7] Jayakumar. and Elangovan, R. (2019). A New Length Biased Distribution with Blood Cancer Data. *Journal of Information and Computational Science*, 9:463-471.

[8] Oluwafemi, OS. Olalekan, DM. (2017). Length and area biased exponentiated Weibull distribution based on forest inventories. *Biometrics & Biostatistics International Journal*, 6:311-320.

[9] Osowole, OI. Amaefula, CG. and Ngozi, N. (2020). On the area biased quasi transmuted uniform distribution. *International Journal of Innovative Mathematics, Statistics and Energy Policies*, 8:13-23.

[10] Patil, G. P. and Rao, C. R. (1978). Weighted distributions and Size biased sampling with applications to wildlife populations and human families. Biometrics, 34:179-189.

[11] Perveen, Z. Munir, M. Ahmed, Z. and Ahmad, M. (2016). On area biased weighted weibull distribution. *Sci. Int. (Lahore)*, 28:3669- 3679.

[12] Para, B. A. and Jan, T. R. (2018). On Three Parameter Weighted Pareto Type II Distribution: Properties and Applications in Medical Sciences. *Applied Mathematics & Information Sciences Letters*, 6:13-26.

[13] Shanker, R. and Shukla, K. K. (2017). Weighted Shanker distribution and its applications to model lifetime data‖. *Journal of Applied Quantitative Methods*, 12:1 – 17.

[14] Rao, C. R. and Patil, G. P. (1965). On discrete distributions arising out of methods of ascertainment, in Classical and Contagious Discrete Distribution. *Pergamon Press and Statistical Publishing Society, Calcutta,* 320-332.

[15] Rajagopalan, V. Rashid, A, Ganaie. and Aafaq, A, Rather. (2019). A New Length Biased Distribution with Applications. *Science Technology and Development*, 8:161-174.

[16] Shanker, R. (2022). Uma Distribution with properties and applications, *Biometrics & Biostatistics International Journal*, 11: 165-169.

[17] Sharma, VK. Dey, S. Singh, SK. and Manzoor, U. (2017). On length and area-biased Maxwell distributions. *Communications in Statistics Simulation and Computation*, 47:1506-1528.

RT&A, No 4(80)
Volume 19, December, 2024

ANALYSIS OF AN ENCOURAGED ARRIVAL QUEUING MODEL WITH SERVERS REPEATED VACATIONS AND BREAKDOWNS

Jenifer Princy P *, K Julia Rose Mary

•

[1,2] Department of Mathematics, Nirmala College for Women, Coimbatore, India
* pjeniferprincy2001@gmail.com
juliakuladaisamy@gmailcom

Abstract

The behavior of customers plays a vital role in realizing the nature of a queue. If there is a favor for customers from the side of service facility the arrival rate increases than usual. Also the positive perspective about the service providers also encourages more number of customers to join the system. The arrival rate of the customers follow Poisson distribution. This paper analyses a queuing model with those encouraged customers who urges to join the system. Here the customers are served in batches according to the general bulk service rule along with the phenomenon that the servers undergo repeated vacations until they find minimum number of customers to start the service. In addition this paper interprets the scenario that if there is a breakdown in the service facility, the waiting line of the customers increases which causes a greater impact on the effectiveness of the service providers favoring the customers. On account of this situation the steady state probability solutions and some performance measures are evaluated along with a numerical illustration.

Keywords: Markovian Process, Poisson distribution, repeated vacations, general bulk service rule, breakdowns, encouraged arrival

1. Introduction

In everyday life, we would have come across different circumstances of waiting in a queue. Waiting in checking counter, Ticket booking, for consulting doctor etc. basically in need of a service to be delivered. Queue is built whenever the customers do not receive service instantaneously. Queue size can be reduced by speeding up the service or by increasing the number of servers to meet up the demand. Mathematically we can figure out every queuing situation into a queuing model which becomes the topic of issue. Agner K. Erlang introduced the concept of queuing theory by modeling the system of incoming calls at the Copenhagen telephone exchange company

In past decades queuing models with breakdowns was one of the area of interest for many researchers. White and Christie [19] proposed the concept of queuing systems with server interruption and evaluated the probabilities of performance in means of geometric function approach. Neuts and Lucantoni [11] discussed the M/M/N queuing model along with sever failures. Xiaolan Yang and Attahiru S. Alfa [20] had analyzed a class of multiserver queuing system with server failures due to undesirable servers.

In the similar way research on vacation queuing models has also become a crucial area of study for the past decades. The term vacation points out the behavior of the server who leaves the system for a random period for various reasons. The server takes the vacation whenever there are no customers in the queue. B. T Doshi [4] attempted to provide a methodological overview of the

queuing systems with two vacation models. K.J.R Mary et. al [10] investigated a queuing system with second optional service channel under bi-level control policy and server's single vacation. Wang and Ruiling [18] made a strategic analysis on both single and multiple working vacations. S. Sindhu et.al [15] illustrated a theoretical comparison between single working vacation model with interdependent and independent arrival and service process. William J. Gray et.al. [5] studied the vacation queuing model with service breakdowns. Jeyakumar. S and Senthilnathan. B [6] analysed a single service queue with batch service along with multiple working vacations and server breakdowns where the server works with different rates without stopping the service completely during vacation period. M. Seenivasan and S. Chandiraleka [14] dicussed on the queuing model with multiple working vacation queuing model with breakdown using Matrix Geometric Approach. R.K. Srivastava et.al.[17] analyzed a model with bulk arrival where the service is provided in two categories. Lidiya and K. J. R. Mary [9] aimed to interpret the queuing model M/M (a, b)/1 with multiple working vacations and breakdown where the customers are served in batches.

In multiserver queuing models the servers usually leave the system for vacation whenever they are idle. In order to reduce the waiting time of customers arriving during servers vacation it is better if the system contains some additional servers to provide service so that the customers are served instantaneously. Afthab Begum and Nadarajan [1] had been analysed this situation where the system comprises of atleast r number of servers while the s-r servers are in vacation. Srinivas R. Chakravarthy et.al. [3] depicted a queuing model with a backup server in which the backup server helps the system to run continuously irrespective of the breakdown, repairs, vacation.

On completion of a random vacation period if a server does not find the minimum accessible number of customers to start the service he leaves the system again for another random period. This type of server's behavior can be defined as repeated vacation. S. Palaniammal et.al.[12] analysed M/M (a, b)/ (2,1) queuing model with servers repeated vacation in which at least one server should remain in the system always among the two servers. S.Baskar and Palaniamal [2] discussed the M/M (a, b)/ (2,1) queuing model along with servers repeated and delayed vacation.

Service providers usually try to fascinate customers by providing profitable deals which encourage many customers to take service than usual. Those customers can be called as encouraged customers. Som B.S and Seth. S [16] interprets the concept of encouraged arrivals, Impatient customers and retention of impatient customers to design effective business strategies. Khan I.E and Paramasivam. R [7] studied encouraged arrival Markovian Model with breakdown and numerous vacations where the customers are served in batches in which PGF is determined in Laplace transforms. Khan I.E and Paramasivam. R [8] evaluated the performance of the reduction in waiting time of single server Markovian encouraged arrival queuing model using control chart technique. Prakati and K. J. R. Mary [13] made comparative study on single working vacation and multiple working vacation under encouraged arrival.

Moreover if the encouraged customers join the system then the queue length will increase than usual. Perhaps disruption in the service facility may further extends the queue length leading to an extended waiting time. This paper aims to interpret the steady state probability solution, the average queue length, and the probability measures of idle and busy servers in M/M (a, b)/ (2,1) queuing model with servers repeated vacation,encouraged arrival of customers and breakdown in the system which will be useful to manage the service facility in an effective sense.

2. Mathematical Model Description

This model involves two service facility (servers) where the arrival of customers into the system follows Poisson distribution with parameter λ. Also the encouraged arrival rate follows Poisson distribution with parameter $\lambda(1+\eta_e)$. Both the service and the breakdown rate of the servers follow exponential distribution with parameters μ and β. Service is done in batches according to the General Bulk Service Rule introduced by Neuts. The service is provided, if there are minimum

'a' number of customers in the queue. If the number of customers is more than the maximum limit 'b' only the first b customers are allowed for taking the service.

The servers are in the state of idle if there is not even a minimum number of customers in a queue. In this case a server leaves the system for a random period called vacation since only one server will be allowed to go for a vacation at a time. This follows an exponential distribution with parameter θ. On returning from the vacation if the server finds less than 'a' customers in the queue while the other server is busy or idle, the server leaves the system for another random period called repeated vacation. The server will continue the same activity until he finds minimum 'a' number of customers to start his service.

Breakdown of the system in the busy period will keep the customers to be in the service for some extent which also cause an impact in the waiting line of the system. If this situation happens for the customers who were being encouraged for getting service from service facility the queue length of the system increases furthermore and even it prolongs the waiting time duration of the customers in the queue.

On the state space $(j,n)\,, j=0,1,2; n\geq 0$, the queue is studied as a Markov process where $n\geq 0$ denotes the number of waiting customers in the queue and j denotes the level of the server.

1. State $(0,n)$, where $0\leq n\leq a-1$ represents that one server is idle and the other server is on vacation.
2. State $(1,n)$, where $n\geq 0$ represents that one server is busy and the other server is on vacation.
3. State $(2,n)$, where $n\geq 0$ represents that both the servers are busy.

Defining $P_{j\,n}(t)=$ Prb { At time t, the system is in the state $(j,\ n)j=0,1,2; n\geq 0\}$ and considering that the steady state probabilities are $P_{0\,n}=\lim\limits_{t\to\infty} p_{0\,n}(t)$,$P_{1\,n}=\lim\limits_{t\to\infty} p_{1\,n}(t)$ and $P_{2\,n}=\lim\limits_{t\to\infty} p_{2\,n}(t)$.

The steady state equations are

$$\lambda p_{0\,0}=\mu p_{1\,0} \tag{1}$$

$$\lambda p_{0\,n}=\lambda p_{0\,n-1}+\mu p_{1\,n}\quad (1\leq n\leq a-1) \tag{2}$$

$$(\lambda\,(1+\eta_e)+\beta+\mu)p_{1\,0}=\lambda p_{0\,a-1}+2\mu p_{2\,0}+\mu\sum_{n=a}^{b} p_{1\,n} \tag{3}$$

$$(\lambda\,(1+\eta_e)+\beta+\mu)p_{1\,n}=(\lambda\,(1+\eta_e)+\beta)\,p_{1\,n-1}+2\mu p_{2\,n}+\mu p_{1\,n+b}\quad (1\leq n\leq a-1) \tag{4}$$

$$(\lambda\,(1+\eta_e)+\mu+\beta+\theta)p_{1n}=(\lambda\,(1+\eta_e)+\beta)\,p_{1\,n-1}+\mu p_{1\,n+b}\quad (n\geq a) \tag{5}$$

$$(\lambda\,(1+\eta_e)+\beta+2\mu)p_{2\,0}=\theta\sum_{n=a}^{b} p_{1\,n}+2\mu\sum_{n=a}^{b} p_{2\,n} \tag{6}$$

$$(\lambda\,(1+\eta_e)+2\mu+\beta)p_{2\,n}=(\lambda\,(1+\eta_e)+\beta)\,p_{2\,n-1}+\theta p_{1\,n+b}+2\mu p_{2\,n+b}\quad (n\geq 1) \tag{7}$$

3. Steady State solutions

Let E denote the forward shifting operator defined by $E(p_{1\,n})=p_{1\,n+1}$. Equation (5) implies $[\mu E^{b+1}-(\lambda(1+\eta_e)+\mu+\beta+\theta)E+(\lambda(1+\eta_e)+\beta)]p_{1\,n}=0.\quad (n\geq 1)$.

The corresponding characteristic equation is

$$\mu Z^{b+1}-(\lambda(1+\eta_e)+\mu+\beta+\theta)Z+(\lambda(1+\eta_e)+\beta)=0. \tag{8}$$

Then by Rouche's Theorem., it has only one real root inside the circle $|z|=1$ when $\rho=\dfrac{\lambda(1+\eta_e)+\beta+\theta}{b\mu}<1$. Let r_0 be the root of the above characteristic equation with $|r_0|<1$. Therefore the homogeneous difference equation has the solution of the form

$$p_{1\,n} = A_1 r_0^n \quad (n \geq a-1)$$

so we get

$$p_{1\,n} = r_0^{n-a+1} p_{1\,a-1} \quad (n \geq a) \tag{9}$$

Using equation (7) we get

$$(2\mu E^{b+1} - (\lambda(1+\eta_e) + 2\mu + \beta)E + (\lambda(1+\eta_e) + \beta))p_{2\,n} = -\theta p_{1\,n+b+1} \quad (n \geq 1)$$

The corresponding characteristic equation is

$$(2\mu Z^{b+1} - (\lambda(1+\eta_e) + 2\mu + \beta)Z + (\lambda(1+\eta_e) + \beta))p_{2\,n} = 0 \tag{10}$$

If r_1 is the root of the above characteristic equation with $|r_1| < 1$ which exists when $\rho_1 = \left(\dfrac{\lambda(1+\eta_e)+\beta}{2b\mu}\right) < 1$. This non - homogeneous difference equation (7) has the solution

$$p_{2\,n} = (A_1 r_1^n + K r_0^n) p_{1\,a-1} \quad (n \geq 0) \tag{11}$$

where A_1 is constant and $K = \dfrac{\left(-\theta r_0^{b-a+2}\right)}{(\lambda(1+\eta_e)+\beta+2\theta)\, r_0 - (\lambda(1+\eta_e)+\beta)}$

Using equation (4) and substituting for $p_{2\,n+1}$ and $p_{1\,n+b+1}$ we have after simplification.

$$p_{1\,n} = [A_2 R^n + G_1(r_0) r_0^n + G_2(r_1) r_1^n]\, p_{1\,a-1} \quad (0 \leq n \leq a-1) \tag{12}$$

where $R = \dfrac{\lambda(1+\eta_e)+\beta}{\lambda+\lambda\eta_e+\beta+\mu}$, $G_1(r_0) = \dfrac{\mu k\,(\lambda(1+\eta_e)+\beta)\,(1-r_0)}{\theta(\lambda+\lambda\eta_e+\beta+\mu)r_0 - (\lambda(1+\eta_e)+\beta)}$ and $G_2(r_1) = \dfrac{2\mu A_1 r_1}{(\lambda+\lambda\eta_e+\beta+\mu)r_1 - (\lambda(1+\eta_e)+\beta)}$

Adding equation (2) over $k = 1$ to n and substituting for $p_{1\,k}$ from equation (12) and simplifying, we get

$$p_{0\,n} = \frac{\mu}{\lambda}\left[A_2\frac{1-R^{n+1}}{1-R} + G_1(r_0)\frac{1-r_0^{n+1}}{1-r_0} + G_2(r_1)\frac{1-r_1^{n+!}}{1-r_1}\right] p_{1\,a-1} \quad (0 \leq n \leq a-1) \tag{13}$$

Using equation (9) and (11) in (6) then we obtain

$$A_1 = \frac{(1-r_1)}{(1-r_1^a)\,(1-r_0)}\left[\frac{r_0\theta}{2\mu} - k\,(1-r_0^a)\right] \tag{14}$$

Further, In equation (12) substituting $n = a-1$ we get the value of A_1 as

$$A_2 = \frac{1}{R^{a-1}}\left[1 - G_1(r_0) r_0^{a-1} - G_2(r_1) r_1^{a-1}\right] \tag{15}$$

The value of $p_{1\,a-1}$ is obtained by using the normalizing condition

$$\sum_{n=0}^{\infty} p_{2\,n} + \sum_{n=a}^{\infty} p_{1\,n} + \sum_{n=0}^{a-1}(p_{0\,n} + p_{1\,n}) = 1 \tag{16}$$

substituting for $p_{2\,n}$, $p_{1\,n}$ and $p_{0\,n}$ and simplifying we get,

$$p_{1\,a-1}^{-1} = A_1 J(R) + G_1(r_0) J(r_0) + G_2(r_1) J(r_1) + \frac{A_1}{1-r_1} + \frac{K}{1-r_0} + \frac{r_0}{1-r_0} \tag{17}$$

where $J(x) = \dfrac{1-y^a}{1-y} + \dfrac{\mu}{\lambda}\left(\dfrac{a}{1-y} - \dfrac{y}{1-y}\dfrac{1-y^a}{1-y}\right)$ and the values of $R, G_1(r_0), G_2(r_1)$ are obtained from equation (12).

4. Analysis of the average measures of performance

The following are the performance measures for the effective mechanism of the queuing model M/M(a, b)/(2, 1) during the time of encouraged arrivals with servers repeated vacations and breakdowns.

1. **Expected Queue Length**

   The expected queue length is given by

$$L_q = \sum_{n=1}^{\infty} np_{2\,n} + \sum_{n=a}^{\infty} np_{1\,n} + \sum_{n=1}^{a-1} n\left(p_{0\,n} + p_{1\,n}\right)$$

   Using equations (9) to (15) and simplifying, we have

$$L_q = \left[A_2 S(R) + G_1(r_0)S(r_0) + G_2(r_1)S(r_1) + \frac{ar_0}{1-r_0} + \frac{r_0^2}{(1-r_0)^2} + \frac{A_1 r_1}{(1-r_1)^2} + \frac{Kr_0}{(1-r_0)^2}\right] p_{1\,a-1}$$

   where $S(y) = \left(\dfrac{1-y^a - ay^{a-1}(1-y)}{(1-y)^2}\right)\left(y - \dfrac{y^2\mu}{\lambda(1-y)}\right) + \dfrac{\mu a(a-1)}{2\lambda(1-y)}$

2. Let $P_{2\,B}$ denote the probability that both the servers are busy then

$$P_{2\,B} = \sum_{n=0}^{\infty} p_{2\,n} = \left(\frac{A_1}{1-r_1} + \frac{K}{1-r_0}\right) p_{1\,a-1}$$

3. Let $P_{1\,B}$ denote the probability that one server is busy and the other server is on vacation then

$$P_{1\,B} = \sum_{n=0}^{\infty} p_{1\,n} = \left(A_2\frac{1-R^a}{1-R} + G_1(r_0)\frac{1-r_0^a}{1-r_0} + G_2(r_1)\frac{r_0}{1-r_0} + \frac{r_0}{1-r_0}\right) p_{1\,a-1}$$

4. Let $P_{0\,B}$ denote the probability that one server is idle and one server is on vacation then

$$P_{0\,B} = \sum_{n=0}^{a-1} p_{1\,n} = \left[A_2 U(R) + G_1(r_0)U(r_0) + G_2(r_1)U(r_1)\right] p_{1\,a-1}$$

   where $U(x) = \dfrac{\mu}{\lambda}\left(\dfrac{a}{1-y} + \dfrac{y(1-y^a)}{(1-y)^2}\right)$ and the values of $R, G_1(r_0), G_2(r_1)$ are obtained from equation (12).

5. Numerical Analysis

In view of an effective performance of the system a sample numerical outcome is analysed in this section by considering sample values.

For a batch of minimum size $a = 10$ and maximum size $b = 25$ service is delivered at a constant rate $\mu = 1$ and the arrival rate λ of the customers tends to be $\lambda = 10$. With these parameters the following table shows the expected queue length for various values of breakdown rate β, encouraged arrival rate η_e, and for the mean vacation time $\dfrac{1}{\theta}$.

Table 1: *The Expected Queue Length for $a = 5$ and $b = 25$ with respect to η_e & β*

		L_q				
η_e	$\frac{1}{\theta}$	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0.8$
0.01		5.863294871	5.9182236	6.028200339	6.136426578	6.249321606
0.03	1.25	5.972415244	6.0282003	6.136426578	6.249321606	6.358418869
0.06		6.136426578	6.1923198	6.302708582	6.416655323	6.52759823
0.09		6.302708582	6.3563036	6.472789527	6.583415019	6.697392467
0.01		7.134515555	7.2066632	7.355389687	7.502775938	7.655940074
0.03	2.5	7.28025688	7.3553897	7.502775938	7.655940074	7.807667263
0.06		7.502775938	7.5766104	7.731087036	7.878229335	8.034363193
0.09		7.731087036	7.8057204	7.957402426	8.112777881	8.265590233
0.01		8.383532999	8.485142	8.676987246	8.87656302	9.074707742
0.03	5	8.580062218	8.6769872	8.87656302	9.074707742	9.275671618
0.06		8.87656302	8.9697	9.171598075	9.381842427	9.578596515
0.09		9.171598075	9.2747364	9.484788121	9.685588806	9.894134297
0.01		9.367526009	9.4843712	9.726184801	9.966874674	10.21860956
0.03	10	9.603887979	9.7261848	9.966874674	10.21860956	10.4684483
0.06		9.966874674	10.084971	10.34215714	10.59752669	10.84996592
0.09		10.34215714	10.468027	10.7152536	10.98782382	11.24189079

It has been interpreted that as there is an increase in encouraged arrival and the breakdown rate the queue length also increases. Figure 1 shows that the queue length increases with increase in encouraged arrival rate η and breakdown rate β for the mean vacation time $\frac{1}{\theta} = 10$ by taking the service rate $\mu = 1$ and also the arrival rate $\lambda = 10$.

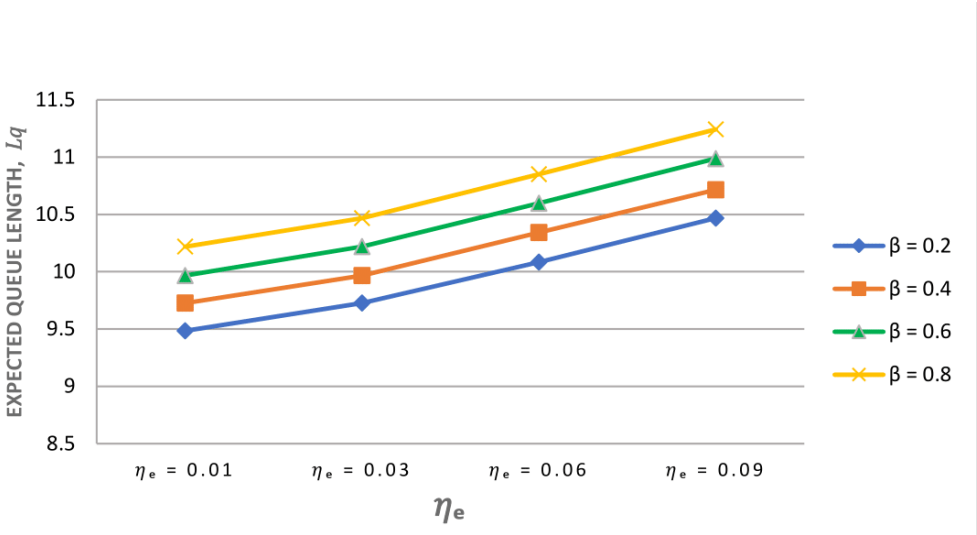


Figure 1: *Expected Queue length L_q for various values of η_e and β*

Figure 2 shows that the queue length increases with increase in encouraged arrival rate η_e and breakdown rate β for the mean vacation time $\frac{1}{\theta} = 2.5$ with the arrival rate $\lambda = 10$.

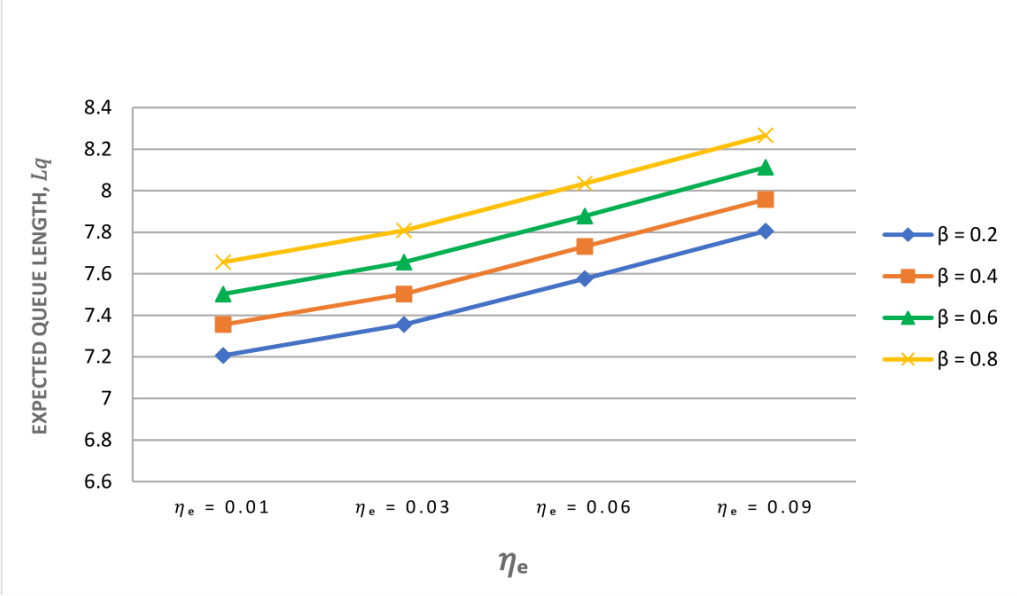


Figure 2: *Expected Queue length L_q for various values of η_e and β*

Thus it has been noted from both Figure 1., and Figure 2., that queue length of the system M/M(a, b)/(2, 1) with repeated vacation, encouraged arrival and breakdown is increased with the mean vacation time, $\frac{1}{\theta} = 10$ when compared to the system with mean vacation time, $\frac{1}{\theta} = 2.5$.

For the batch of maximum size $b = 40$ considering $\eta_e = 0.03$ and $\beta = 0.5$ and opting various values for minimum size a of the batch it has been noted from the following Table 2 that increase in minimum capacity of the batch size increases the queue length which is also depicted graphically in Figure 3.

Table 2: *The Expected Queue Length for $b = 40$ and for various values of a*

λ	$a = 15$	$a = 20$	$a = 25$
10	9.461203	11.03285	13.00848
12	10.81249	12.04423	13.7487
18	13.33766	14.12254	15.40794
20	16.44243	16.88795	17.7878
25	18.84208	19.12746	19.80789
36	26.08044	26.19581	26.50847

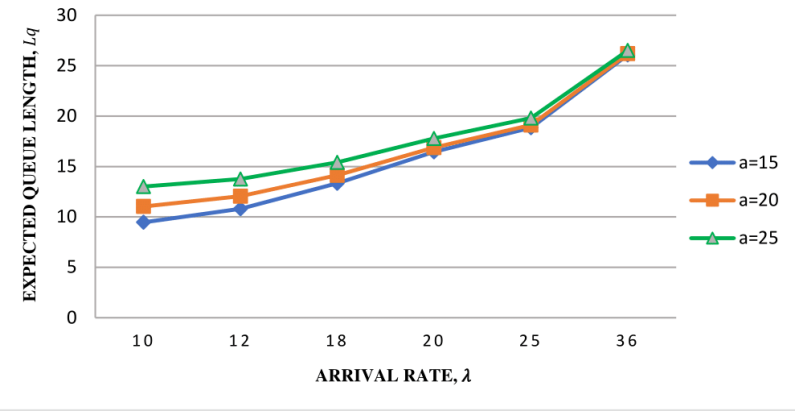


Figure 3: *Expected Queue length L_q for various values of a*

For the batch of minimum size $a = 10$ and opting various values for the maximum size b of the batch and also considering $\eta_e = 0.03$ and $\beta = 0.5$ it has been noted from the following Table 3 that increase in maximum capacity of the batch size decreases the queue length which is also depicted graphically in Figure 4.

Table 3: *The Expected Queue Length for $a = 10$ and for various values of b*

λ	$b = 20$	$b = 30$	$b = 40$
10	9.302594	8.473836	8.266449
12	11.9158	10.3027	9.867244
18	17.06309	13.59169	12.60109
20	23.84891	17.61646	15.73194
25	29.2534	20.72457	18.05442
36	35.34298	24.24377	20.57053

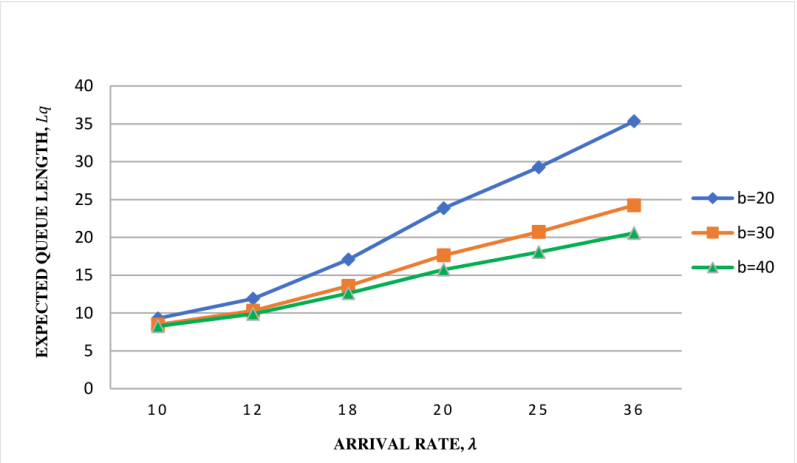


Figure 4: *Expected Queue length L_q for various values of b*

6. Conclusion

This paper analysed the steady state solutions of the M/M(a, b)/(2, 1) queuing model with servers repeated vacations and encouraged arrival of customers together with the breakdown mechanism. The expected queue length, probability measures of the idle and the busy servers are formulated and verified with an numerical illustration which helps to review the system for future effectiveness. Further a comparative study can be made with M/M(a, b)/1 queuing model with servers repeated vacation.

References

[1] Afthab Begum, M. I. and Nadarajan, R. (1997). Multiserver Markovian queueing system with vacation. *Optimization*,41(1), 71-87.

[2] Baskar, S. and Palaniammal, S. (2014). M/M(a,b)/(2,1) queueing system with servers repeated and delayed vacation. *Asian-European Journal of Mathematics*,7(02), 1450032.

[3] Chakravarthy, S. R. and Kulshreshtha, R. (2020). A queueing model with server breakdowns, repairs, vacations, and backup server. *Operations research perspectives*, 7, 100131.

[4] Doshi, B. T. (1986). Queuueing systems with vacations - a survey. *Queueing systems*, 1, 29 - 66.

[5] Gray, W. J. Wang, P. P. and Scott, M. (2000). A vacation queueing model with service breakdowns. *Applied Mathematical Modeling*, 24(5 - 6), 391 - 400.

[6] Jeyakumar, S. and Senthilnathan, B. (2017). Modelling and analysis of a bulk service queueing model with multiple working vacations and server breakdown. *RAIRO - Operations Research - Recherche Operationnelle*, 51(2), 485 - 508.

[7] Khan, I. E. and Paramasivam, R. (2023). Analysis of Batch Encouraged Arrival Markovian Model Due to a Secondary Optional Service, Break-Down and Numerous Vacations. *Mathematical Statistician and Engineering Applications*, 72(1), 1166 - 1177.

[8] Khan, I. E. and Paramasivam, R. (2023). Reduction in waiting time of single server Markovian queuing encouraged arrival model. *Reliability:Theory & Applications*, 18(3(74)), 776 - 784.

[9] Lidiya, P. and Mary, K. J. R (2024). Analysing the Performance of Queuing Model with the Busy period Breakdown. *Baghdad Science Journal*, 22(2).

[10] Mary, K. J. R. and Begum, M. A. and Parveen, M. J. (2011). Bi - level threshold policy of MX/(G 1, G 2)/1 queue with early setup and single vacation. *International Journal of Operational Research*, 10(4), 469 - 493.

[11] Neuts, M. F. and Lucantoni, D. M. (1979). A Markovian queue with N servers subject to breakdowns and repairs. *Management Science*, 25(9), 849 - 861.

[12] Palaniammal, S. and Nadarajan, M. A. B. R. (2005). Servers Repeated Vacation. *Computational Mathematics*, 232.

[13] Prakati, P. (2024). M/M/C Queue with multiple working vacations and single working vacation under encouraged arrival with impatient customers. *Reliability:Theory & Applications*, 19(1(77)) 650-652.

[14] Seenivasan, M. Chandiraleka, S. (2023). Single Server Queueing Model with Multiple Working Vacation and with Breakdown. *Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies(ICAECT)*, (pp. 1-5) IEEE.

[15] Sindhu, S. Krishnamoorthy, A. and Kozyrev, D. (2023). On Queues with working vacation and interdependence in arrival and service processes. *Mathematics*, 11(10), 2280.

[16] Som, B. S. & Seth, S. (2018). Queuing System with Encouraged Arrivals, Impatient Customers and Retention of Impatient Customers for Designing Effective Business Strategies. *AISECT University Journal*, 6(13).

[17] Srivastava, R. K. Singh, S. and Singh, A. (2020). Bulk Arrival Markovian Queueing System with Two Types of Services and MUltiple Vacations. *International Journal of Mathematics and Computer Reserach*, 8(8), 2130 - 2136.

[18] Wang, Y. and Tian, R. (2020). Strategic analysis in Markovian queues with a single working vacation and multiple vacations. *International Journal of Innovative Computing, Information and Control*, 16(1), 349 - 366.

[19] White, H. and Christie, L. S. (1958). Queing with preemptive priorities or with breakdown. *Operations research*, 6(1), 79 - 95.

[20] Yang, X. and Alfa, A. S. (2009). A class of multi - server queueing system with server failures. *Computers and industrial Engineering*, 56(1), 33 - 43.

IMPROVING THE RELIABILITY OF RECOGNIZING POTENTIALLY HAZARD UNDERWATER OBJECTS

Artyukhin Valerii, Vyalyshev Alexander, Zinoviev Sergey, Tuzov Fedor

•

All-Russian Research Institute for Civil Defence of the EMERCOM of Russia (the Federal Science and High Technology Center), Moscow, Russia
ikshot@mail.ru, vialyshev@rambler.ru, golf1972@yandex.ru
Corresponding author: Tuzov Fedor fedor-tuz@mail.ru

Abstract

The process of recognizing an underwater object and detecting potentially hazardous underwater object is very important in underwater operations. To facilitate the work of the side scan sonar operator, this paper proposes to increase the reliability of recognizing hydroacoustic images of potentially hazardous underwater objects in automatic mode. Based on the analysis of sonar images received from the side scan sonar, an image of an object is formed, which is then recognized (classified) as belonging to a certain class of objects. Five classes of recognized objects are defined. A convolutional neural network used to determine whether an underwater potentially dangerous object belongs to one of the classes is described. Filters for initial sonar images for acceleration of neural network operation are defined. Algorithms and software for forming an image of the object and making a decision on its belonging to one or another class are developed. It is shown that the use of convolutional neural network allows to determine the correct class of the object with an accuracy of 91%.

Keywords: reliability, side scan sonar, potentially hazardous underwater object, neural networks, classification, classification features, information

I. Introduction

The reliability of underwater object type detection from side scan sonar image processing is an important task for determining the hazard of a submerged object. The usual method of recognizing objects on the basis of classification features and operator's work requires a long time and with a large number of detected objects has a low probability of correctly determining the type of object. On the bottom of the world ocean there are a large number of objects that may pose a potential danger to the population and the environment -potentially hazardous underwater objects (PHUO) [1]. Such objects primarily include:

- sunken nuclear submarines and their structures with used nuclear fuel;
- sunken and sunken ships and vessels with ammunition on board;
- sunken containers and barrels with hazardous substances;
- ammunition (mines, shells, etc.).

The most difficult problem has been and remains the search for small-sized objects, as well objects partially submerged in the ground and silt. These include bottom and near-bottom mines, emergency containers with toxic and radioactive substances, lost ammunition (bombs, torpedoes, warheads, etc.), qualified by hydroacoustic specialists as “small-sized objects”. The “black boxes” of

aircrafts that have crashed over the sea fall into the same category.

In recent years, various neural network architectures have been actively implemented in the field of object recognition in side scan sonar images. The most common approaches include:

convolutional neural networks (CNN). Converged networks are the basis for many computer vision tasks, including object recognition. They perform well in object identification and classification tasks due to their ability to extract spatial hierarchies of features [2].

YOLO (You Only Look Once). This neural network architecture enables real-time object detection. YOLO divides the image into a grid and predicts object boundaries and class membership probabilities for each cell [3].

SSD (Single Shot MultiBox Detector). Similar to YOLO, SSD is another architecture for real-time object detection. It utilizes a variety of sizes of predicted rectangles to better handle objects of different scales [4].

Faster R-CNN. This model combines the strengths of CNN and regression algorithms for detection. Faster R-CNN utilizes Region Proposed Networks (RPNs), which achieves high accuracy, albeit at a higher resource cost [5].

The universality of the approach proposed in the paper allows the developed system to be used for classification of any objects on the bottom of the water area.

II Methods

For detection of underwater potentially dangerous objects by hydroacoustic means, as a rule, side-scan sonars (SSS) with a range of up to hundreds of meters and designed for operation in shallow water conditions in coastal and shelf zones of seas are used. As a rule, they have relatively high operating frequencies and good range resolution.

Side scan sonar provides a 3D image of the bottom relief with objects located on it. An example of an image with a submerged vessel is shown in figure 1.

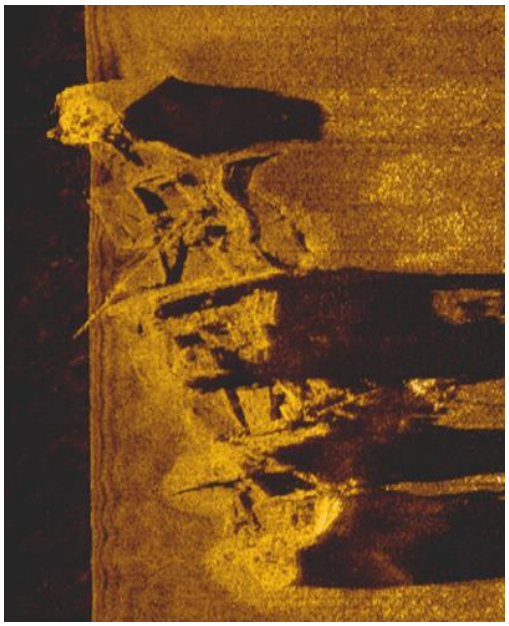


Figure 1: *Submerged vessel on side scan sonar image*

The assessment of the SSS image is usually given visually by the sonar operator, who determines the class of the object. For a number of tasks with a large amount of information it is expedient to recognize images in automatic mode, however, despite the wide application of pattern

recognition methods in various fields, the issue with underwater potentially dangerous objects has not been solved.

The problem of reliability of classification or recognition of PHUO is complicated by a variety of types and sizes of objects on the bottom, as well as their location relative to the line of motion of the ship-tugboat SSS, i.e. the same object at different distances from the ship-tugboat of SSS and at different angles of sighting can have a completely different image (picture) on the SSS screen [6]. It is especially noticeable on extended objects (boats, logs, explosive ordnance, etc.). Therefore, the task has to be solved in two stages.

At the first stage, the method of selecting a single object from the data stream from the side-scan sonar output is used, or in other words, the detection of "candidates" for PHUO and their initial selection by object size ("item" for PCOR should be at least 0.5 meters and not more than 100 meters in size). The image at the SSS output allows to do this, since the resolution of sensors (image pixel sizes) is currently not more than 0.2 meters. The results of the detection algorithm are transmitted to classification in the form of image fragments containing "items" for PHUO. We will call them strobes containing images of smaller volume than the original image from the SSS output.

In the second stage, classification of objects in each strobe obtained in the algorithm for detecting "items" in PHUO is performed.

Thus, we can term primary processing as object detection and strobe selection, and secondary processing as object classification in a strobe.

The SSS output produces an image of the underwater situation to the right and left of the SSS tugboat line of motion, which is processed in the pre-processing block. At the output of the pre-processing block two images are obtained (to the right and to the left of the line of motion of the ship-tugboat), in which linear distortions are eliminated and the overlapping part of the frame, which is under the ship-tugboat, is removed. Further the right and left parts of the frame are processed independently.

All lines are processed according to the same algorithm in order to detect the boundaries of a potential object and its acoustic shadow in each line. For this purpose, the signal in the string is "passed" through a linear matched filter, and the filtered signal is compared with two thresholds: exceeding the upper threshold determines the left and right boundaries of the object in this string, and the value of the filtered signal less than the lower threshold determines the left and right boundaries of the acoustic shadow of the object in this string.

After processing of all frame lines, "stitching" of potential object boundaries between the nearest lines is performed according to the "nearest neighbor" principle. The "stitching" of boundaries continues until there is no (will not be found) continuation of the potential object boundary in two neighboring lines. After that, the lower left and upper right boundaries of the potential object are calculated, and these parameters are used to calculate the strobe coordinates (for further selection of a strobe containing signals from the potential object from the frame, taking into account its potential boundaries in the frame).

The selected strobe is then passed to the classification algorithm.

Two approaches can be used to classify the selected underwater potentially hazardous object:

- Measurement and selection of classification features with subsequent analysis and decision making;

- recognizing the image of an underwater potentially hazardous object as given classification type, formed on the basis of accumulated previous data.

In the first approach, the source of initial information is sonar, which, using spatial and temporal processing, generates a signal space belonging to all detected targets. A priori, it is not known which targets and at what distances are located in the processing space. From the output of the system of spatial and temporal processing all signals arrive in the block of measurement of classification features, in which the classification features are selected and converted into a useful

form for further analysis. It is necessary to ensure measurement accuracy and resolution in time and space for this process.

The results of this transformation give an n - dimensional vector, which is the code of the selected class. This code has a random character, because the signs are measured in conditions of interference and errors of the measurement system, caused by the error of its own motion, as well as the randomness of the appearance of a particular representative of a particular class of objects. From the output of the block of measurement of signs the code of the object comes to the block of decision making, which also receives the codes of standards of all classes.

From output of the decision making unit provides quantitative indicators characterizing the proximity of the presented realization with the reference. The result of comparison is, as a rule, a posteriori probability or correlation coefficient. In the decision making block the object is compared according to the specified criterion and attributed to one of the standards. The database is designed to store the attributes of reference classes and the results obtained at various stages of recognition. The attributes of reference classes are put into it on the basis of initial a priori data.

The system of classification attributes should be previously known and a spatio-temporal processing should be developed for it.

It is possible to go another way: using the existing spatio-temporal processing, analyze the results obtained and on their basis formulate a set of classification features. In the first case, when the set of classification features is known, it is possible to form automatic decision making procedures. In the second case, the decision, as a rule, should be made by the operator on the basis of preliminary processing of initial information.

It is the first case when a classification system, in which algorithms of automatic measurement of classification features and algorithms of automatic decision making were implemented, was developed in relation to a small-size environment illumination complex.

The complex was installed on a specific stationary carrier and had the function of ensuring the safety of navigation. The classification system contained a set of classification features that were automatically measured when a target was detected. Based on the developed classification attributes, an automatic decision on the target class was formed for the generated full group of events. The operator made an independent decision based on the type of information displayed and the measured classification features.

In this paper, the classification of potential hazardous underwater object images in software is proposed to be performed using convolutional neuron network algorithms.

The method of neural networks is based on a simple mathematical model of brain functioning. Neural networks can be used to obtain both numerical and binary solutions. The method is computationally expensive, but the structure of the networks allows capturing (accounting for) complex relationships between predictors and values of the dependent variable. On the other hand, the network can be perceived as a “black box” because the identified relationships are not easy to understand and interpret. Thus, the neural network method is a good choice when accurate forecasts are required, but it is unsuccessful when solving descriptive analysis problems or when it is necessary to investigate the nature of relationships between variables. Forecasts obtained with the help of neural networks can be combined with other forecasts to obtain prognostic “ensembles” [7, 8].

The schematic of an artificial neuron (as a mathematical reflection of a physical neuron in the human brain) can be visualized as shown below in figure 2.

As with a biological neuron, an artificial neuron takes input values $x_1, x_2, \dots, x_n$, multiplies them with weights $w_1, w_2, \dots, w_n$, sums the resulting values (producing the “logit” of the neuron, ($z = \sum_i^n w_i x_i$) and passes them to a function f that calculates the neuron's output $y = f(z)$. The value of y can then be passed to other neurons as input.

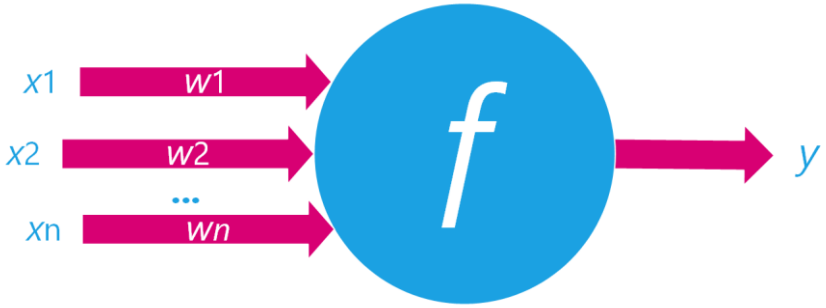


Figure 2: *Scheme of a neuron in an artificial neural network*

In the human brain, neurons are organized into layers. An artificial neural network arises when we connect artificial neurons into layers and link the layers with outputs and inputs, with the first layer receiving as input the raw data of a problem and the last layer producing as output the solution to it (Figure 3).

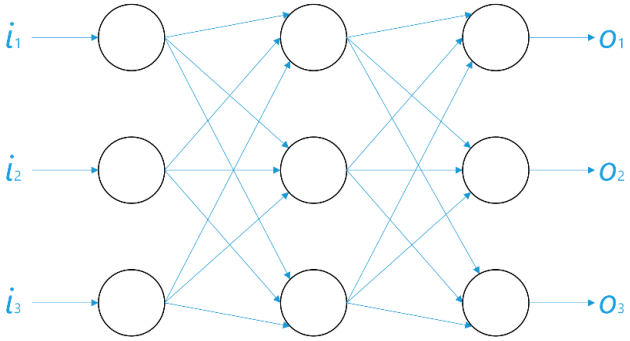


Figure 3: *A simple example of a neural network with three layers*

The logit transformation functions of the neurons in each layer can be different: linear in the simplest case ($f(z) = az + b$), sigmoidal, tangent, or boundedly linear (respectively, $f(z) = \frac{1}{1+e^{-z}}$), $f(z) = \tanh(z)$, and $f(z) = \max(0; z)$ and others. The simplest example of a neural network is a network without hidden layers (there are only input and output layers), equivalent to linear regression. The prediction (solution to the problem) is obtained as a linear combination of inputs. The weights are chosen using a learning algorithm that minimizes a cost function (e.g., MSE - mean squared error). Of course, in such a simple case, there is no sense in a neural network at all, because using a linear regression method directly is much more efficient [9].

In our case, a more complex network will be used; the input of the lower layer of neurons will be image data (pixel intensity values) of underwater objects, and the outputs of the upper layer of neurons will give us probabilities of assigning objects in the images to different classes. The neural network converts image pixel values into a useful representation by extracting high-level image features (such as shapes and edges in the image) that describe complex concepts by combining a large number of small pieces of information [10, 11, 1].

In the task of image classification, it is convenient to obtain the probability distribution of assigning an object in the image to a particular class at the output of a neural network, so in this kind of tasks, a softmax function is used on the output layer of neurons. The radical difference of the function is that its calculation requires not only the logit value of a given neuron, but also the logit values of all other neurons in the layer (1):

$$y_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (1)$$

In the case of a "good", "strong" prediction, the output of one of the neurons will be significantly greater than the outputs of the others, in other words, ideally, the object should be assigned to a particular class with a fair amount of confidence.

Initially, all connection weights in the network are initialized in some random way. During the training phase, images of objects whose classes are known (labeled) are loaded into the model. By comparing the network answers y with the known correct answers t one can obtain the error function E, in particular, in the form of a second-order norm (2):

$$E = \frac{1}{2}\sum_i \left(t^{(i)} - y^{(i)}\right)^2 \quad (2)$$

When E is closer to zero, the more accurate the predictions of our model are, and if E=0, then our model performs predictions with perfect accuracy.

The accuracy of the model during the training phase is improved by applying optimization techniques to the error function E. More precisely, we need to find the minimum value of E as a function of, among other things, the weights of the inputs to the last layer of neurons. Using gradient descent (or another optimization method), we can adjust the weights for the inputs of the last layer. On the other hand, the value of the error function depends not only on the weights of connections in the last layer of neurons, but also on the values of inputs to it. Those, in turn, depend on the weights on the layer before the last one, etc. By spreading the "responsibility" for the magnitude of the error to all layers, we can adjust all the weights in the network, thus training the model. This approach is called the "error back propagation method" [13, 14, 15].

A significant problem arises in image processing by classical neural networks. Because the input image is divided into pixels, it means that the number of inputs (and thus weights) to each neuron of the lower layer of the network will be:

number of pixels in height * number of pixels in width * number of color components, e.g. 200*200*3=120,000.

This number of connections is not only redundant, consumes a large amount of memory and slows down the learning process, it is also likely to overtrain the network with respect to the training data.

A convolutional neural network takes advantage of the knowledge that it is images that we are processing, which means that we can intelligently constrain the network architecture in a way that significantly reduces the number of incoming connections.

The basis of a convolutional neural network is the use of filter layers and max pooling layers.

The filtering layer processes a three-dimensional array of information (in our case, the dimensions of such an array are determined by the length and width of the image, as well as the number of color components in it) and outputs another three-dimensional array of information, with length and width determined by the parameters of the filters, and depth determined by the number of applied filters (figure. 4).

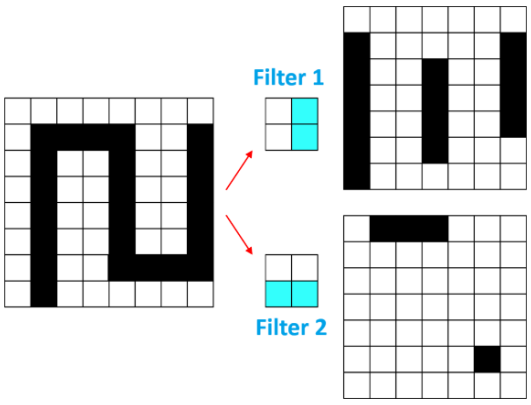


Figure 4: *Application of two filters that select horizontal and vertical lines*

The unifying layer is responsible for reducing the size of the information array (and, as a consequence, the number of inputs to the next layer) by splitting the information array into equal-sized fragments and condensing the values of individual elements in each fragment into a single value (figure 5) [13, 14].

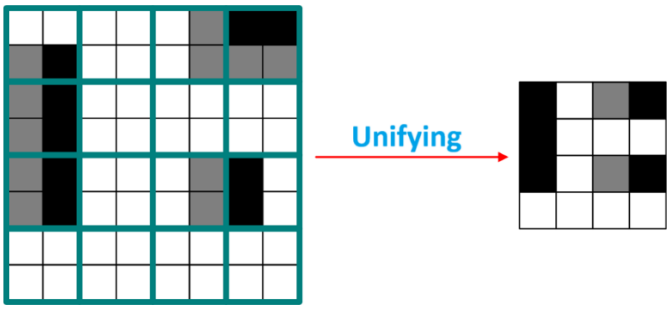


Figure 5: *An example of significant reduction of the number of neural network parameters by applying th When solving the problem of object classification in general terms, a neural algorithm is a computational procedure, the main part of which can be realized on a neural network.*

III. Results and discussion

The basis for the development of a neural algorithm for solving the problem is a systematic approach, in which the process of solving the problem is represented as the functioning in time of some dynamic system, whose input is a set D (initial data), and the output is a set R (objects to be defined and received their values).

Solving the problem with the help of neural algorithm includes the following points:

1) Obtaining a specific neural network structure corresponding to the applied algorithm.

2) Finding the values of weight coefficients, or selecting them from memory, if they were found earlier.

3) Generation of initial approximations of parameters, if it is necessary.

4) Transfer of all numerical values to the neural network and running it.

5) Functioning of the network according to the mode:

- over a single step or a fixed number of steps;

- for a variable number of steps, depending on the required accuracy and/or specific numerical values of the parameters (in it, the process of tuning the input signal takes place).

6) Obtaining the solution.

Points 1 and 2 may be performed once for subsequent repeated application of points 3, 4, 5 and 6.e unification.

In the case of applying the approach based on a neural network to the task of classification of images of underwater objects, the above means that at the stage of training in the neural network with the selected topology loaded pre-arranged data, that is, data directly images and pre-assigned to these images classes of objects, corrected network parameters, and then at the stage of work in the network loaded one by one already unarranged images, for make a decision on the assigned class at the output of the neural network.

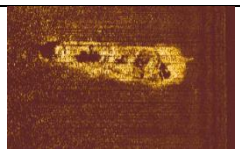
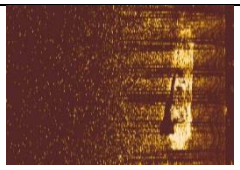
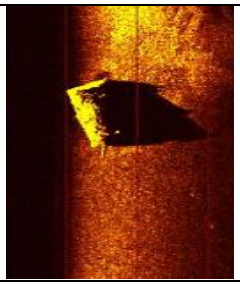
Modern image recognition and classification models may include millions of parameters. Training them from scratch requires large amounts of training marked data (millions of images) and considerable time (hundreds of hours of GPU work - graphics processing units on video cards used for parallel computations, including those associated with neural networks).

Knowledge transfer is a technique that takes a part of a model already trained (trained) on a similar task and uses it in a new model for a different task. This approach does not provide the

accuracy that can be obtained by training a full model, but it shows itself to be surprisingly good on multiple tasks, also allowing thousands rather than millions of training images to be used, and models to be trained in half an hour or an hour rather than hundreds of hours. This approach applied in this paper [16, 17, 18, 19].

After training the network on a large database, 100 images with objects were selected/synthesized to test the accuracy of the recognition program. It was not previously encountered by the neural network, i.e. it was not used in its training. Selected test results are given in the table below.

Table 1: *Results of testing the underwater object classification program on side scan sonar images (Letter designations correspond to the following categories: A - ship/submarine, B - containers, C - barrels, D - ammunition, E - other)*

Test number	Image	The probability of categorizing the object in the image, %				
		A	B	C	D	E
13		83,36	3,78	1,43	9,27	2,16
14		94,57	1,26	1,53	1,22	1,41
26		10,95	10,68	0,76	2,12	75,49

According to the results of the final testing, the program showed the correctness of category determination for 100 new images at the level of 91%.

The developed method can be used to improve the reliability of classification of potentially hazardous underwater objects, both directly at the side scan sonar output and during post-processing of the accumulated records.

References

[1] Vjalyshev A.I. (2017) Emercom of Russia and Potentially Hazardous Underwater Objects [MChS Rossii i podvodnye potencial'no opasnye obekty]. Civil security Technology. Vol. 14 № 1(51) p 4-10. (in Russian).

[2] Du, X., Sun, Y., Song, Y., Sun, H., & Yang, L. (2023). A comparative study of different CNN models and transfer learning effect for underwater object classification in side-scan sonar images. Remote Sensing, 15(3), 593.

[3] Yulin, T., Jin, S., Bian, G., & Zhang, Y. (2020). Shipwreck target recognition in side-scan sonar images by improved YOLOv3 model based on transfer learning. IEEE Access, 8, 173450-173460.

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

[5] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 39(6), 1137-1149.

[6] Merrifield, S. T., Celona, S., McCarthy, R. A., Pietruszka, A., Batchelor, H., Hess, R., ... & Terrill, E. J. (2023). Wide-area debris field and seabed characterization of a deep ocean dump site surveyed by Autonomous Underwater Vehicles. Environmental Science & Technology, 57(46), 18162-18171.

[7] Shmueli Galit, Lichtendahl Jr. Kenneth C. (2016) Practical Time Series Forecasting with R. Axelrod Schnall Publishers.

[8] Woodward Wayne A., Sadler Bivin Philip, Robertson Stephen. (2024) Time Series for Data Science. CRC Press.

[9] Hyndman Rob J., Athanasopoulos George. (2021) Forecasting. Principles and Practice. 3rd ed. OTexts.com.

[10] Hope Tom, Resheff Yehezkel S., Lieder Itay. (2017) Learning Tensor Flow. O'Reilly.

[11] Géron Aurélien. (2022) Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems. 3rd ed. O'Reilly.

[12] Chollet Francois, Kalinowski Tomasz, Allaire J. J. (2022) Deep Learning with R. 2nd ed. Manning.

[13] Galushkin A.I. (2010) Neural networks: fundamentals of theory [Nejronnye seti: osnovy teorii]. Hotline-Telecom Publishing House. (in Russian).

[14] Hagan Martin T., Demuth Howard B., Beale Mark H., De Jesús Orlando. (2014) Neural Network Design. (Martin hagan) 2nd ed.

[15] Nikhil Buduma. (2017). Fundamentals of Deep Learning: Designing Next-generation Artificial Intelligence Algorithms/c Nikhil Buduma. Beijing, Boston, Farnham, Sebastopol, Tokyo: OReilly.

[16] Jiang, J., Shu, Y., Wang, J., & Long, M. (2022). Transferability in deep learning: A survey. arXiv preprint arXiv:2201.05867.

[17] Gutstein, S., Fuentes, O., & Freudenthal, E. (2008). Knowledge transfer in deep convolutional neural nets. International Journal on Artificial Intelligence Tools, 17(03), 555-567.

[18] Passalis, N., Tzelepi, M., & Tefas, A. (2020). Probabilistic knowledge transfer for lightweight deep representation learning. IEEE Transactions on Neural Networks and Learning Systems, 32(5), 2030-2039.

[19] Chen, T., Goodfellow, I., & Shlens, J. (2015). Net2net: Accelerating learning via knowledge transfer. *arXiv preprint arXiv:1511.05641*.

THE INVERSE LOMAX ODD-EXPONENTIATED EXPONENTIAL DISTRIBUTION WITH INDUSTRIAL APPLICATIONS

Jamilu Yunusa Falgore[1], Yahaya Abubakar[2], Sani Ibrahim Doguwa[3], Aminu Suleiman Mohammed[4], and Abdussamad Tanko Imam[5]

•

[1,2,3,4]Department of Statistics, Ahmadu Bello University, Zaria-Nigeria.
[5]Department of Mathematics, Ahmadu Bello University, Zaria-Nigeria.
[1] jamiluyf@gmail.com, [2]ensiliyu2@yahoo.co.uk, [3]sidoguwa@gmail.com,
[4]aminusmohammed@gmail.com, [5]atimam@abu.edu.ng

Abstract

Based on the limitations of the Inverse Lomax distribution and exponential distribution as outlined in the literature, a new extension of the exponential distribution is introduced in this paper. Some statistical properties of the ILOEED such as mean, variance, skewness, quantile function, moment, moment generating function, as well as kurtosis were demonstrated. The shapes of the hazard function of the proposed distribution suggest that it can be used to fit a dataset with increasing and bath-tube shapes. A simulation study for three different cases was also presented. The result of the simulation for three different cases (I, II, and III) indicated that ILOEED's estimates are consistent. Lastly, an application to Industry datasets was demonstrated based on the ILOEED. Having minimum values of the Goodness-of-fit criteria and Goodness-of-fit statistics, the ILOEED can be recommended to fit these three datasets, in preference to other distributions considered in this paper.

Keywords: Inverse Lomax-G family, Exponentiated Exponential Distribution, Weibull Distribution, Heavy Tail Distribution.

I. Introduction

In their ongoing pursuit of adaptive and flexible statistical models, scientists and researchers have been investigating new distributions that may accurately represent a wide range of real-world data patterns. In this quest, the proposed distribution shows great promise as a more versatile model for a range of phenomena. Interestingly, it leverages the advantages of its parent distributions, the Odd-Exponentiated and the Inverse Lomax distribution (ILD), to produce a distribution that can describe a wide range of datasets, especially those with heavy tails and non-monotone failure rates ([1],[2],[3]). Extreme occurrences or outliers are more likely to occur in a heavy-tail distribution because its tails decay more slowly than those of a normal distribution. This suggests that there is more risk or variability in the data. The Pareto II (ILD), Cauchy, and Student's t-distributions are a few instances of heavy-tail distributions.

Inverse Exponentiated Odd Lomax Exponential distribution was proposed by [4], offering a fresh outlook on statistical modeling and analysis. Their research investigates the statistical properties of this distribution, contributing to a better comprehension of its practical applicability.

A four-parameter Exponentiated Odd Lomax Exponential (EOLE) distribution was proposed by [3], combining an exponentiated odd function with Lomax and exponential elements, thereby enhancing the distribution's versatility and providing a more intricate parameterization for modeling. To give researchers a wider range of tools for a variety of applications, Inverse Exponentiated Lomax Power Series distribution was proposed by [2], which expanded the family of distributions that combines Lomax and exponential components. Inverse Lomax distribution has two major drawbacks. These are: Limited Flexibility in Shape i.e. ILD has some difficulty adjusting its probability density function, especially in the peak and tail areas. This can hinder its ability to precisely model data with certain patterns [5]. A potential constraint in modeling hazard rates i.e. while often used to model non-monotone hazard rates (failure rates that vary over time), the Inverse Lomax distribution may not be able to capture all possible hazard rate shapes that can arise in real-world situations [6]. However, exponential distribution has limitations of constant failure rate and memory-less property. Hence, the need to study the proposed distribution to remedy some of the drawbacks of the ILD and exponential distribution.

To tackle the challenge of modeling the lifespans of electronic devices, which is essential for predicting future failures and achieving energy savings, a reliability model that is based on the inverse power law and generalized inverse Weibull distribution [7]. It demonstrates how successful the proposed distribution is in influencing the average time to failure of the examined capacitor, as opposed to standard models such as the inverse Weibull, using an empirical analysis that focuses on the life cycle of a surface-mounted electrolytic capacitor. To analyze COVID-19 death cases in Europe and China, the Exponentiated Transformation of Gumbel Type-II (ETGT-II) model [8]. This model provides a thorough analysis of statistical features and estimates model parameters using maximum likelihood and Bayesian approaches. The ETGT-II model is shown to be efficient through simulation analysis. It exhibits a promising adaptation to the COVID-19 death data sets, perhaps providing a better fit than other models. The new exponential inverted Topp-Leone (NEITL) distribution is presented by [9]. It is an extension of the inverted Topp-Leone distribution with an extra shape parameter. Its features, estimation methods, and application to actual datasets in the engineering and medical domains are all explored. The generalized log-exponential transformation of Gumbel Type-II (GLET-GTII), which was proposed by [10] as a generator for a generalized version of the Gumbel type-II model, increases modeling flexibility by adding a new parameter. Quantiles, survival function, and reliability are among the statistical attributes that are examined. Maximum likelihood and Bayesian approaches are used as parametric estimation methods, and they show consistency through Monte Carlo simulations and outperform other models in practical implementations, especially when it comes to infectious diseases like COVID-19. Using a new power function and a modified Kies generalized transformation, a novel statistical model and discusses its theoretical characteristics, including the density function, quantile function, and stochastic ordering [11]. The moment exponential distribution is extended by the two-parameter alpha power-transformed moment exponential (APTME) distribution, which shows excellent fit and performance through a variety of estimators and simulation studies. Its practical significance is demonstrated by its application to real-world datasets [12].

Moreover, the flexible four-parameter Kumaraswamy extended exponential (KwEE) distribution is presented by [13]. This model shows that the novel distribution may provide a better fit than current models in several COVID-19 spread analysis situations by evaluating COVID-19 mortality rates in nations such as Italy and the United Kingdom. To represent the dependability metrics of a generalized exponential model based on the inverse power law (IPL), was suggested using a multilayer ANN with Bayesian regularization by [14]. The outcomes show how well ANNs operate as a reliable mathematical tool for evaluating lifetime model reliability, and they are backed by a real-world application. Under the generalized type-I progressive hybrid censoring sample (GTI-PHCS), statistical inference for the Kavya-Manoharan generalized exponential distribution was

proposed by [15]. It does this by examining different estimation techniques, such as maximum likelihood and Bayesian approaches, and using real-world data analysis and simulations to illustrate how well the techniques work. Through simulation studies and application to engineering datasets, the half-logistic modified Kies exponential (HLMKEx) distribution as a flexible three-parameter model for modeling real-world data was presented [16]. It provides detailed mathematical features, such as density function forms and estimation methods, and shows its superior fit over competing distributions. Generalized exponentiated unit Gompertz (GEUG), a unique four-parametric model, was introduced by [17] to represent clinical trial data of patients with arthritis. By adding new tuning factors to the unit Gompertz (UG) model, the GEUG model aims to improve the estimate of distribution parameters, hence increasing the model's adaptability.

Motivation: The inability of Inverse Lomax distribution and exponential distribution to adequately capture some intricate data patterns led to the development of the proposed distribution. To introduce a more adaptable model, that expands on the advantages of its parent distributions; the Odd-Exponentiated and the Inverse Lomax. **Key features:** More flexibility in defining its density and hazard functions, Non-monotone hazard rates capable of simulating phenomena with fluctuating risk characteristics over time, as well as heavy tails that capture extremes and outliers well. In this article, a new extension of Exponential distribution is introduced. The proposed distribution is formulated based on the Inverse Lomax Odd Exponentiated-G family of distributions. The most important feature of the proposed distribution, with two shape parameters, a scale, and rate parameters represents its ability to provide different density shapes. This means that the proposed distribution can fit various datasets adequately. The proposed distribution has the following desirable properties. (i) The probability density function (pdf) of the ILOEED proposed distribution has a simple closed form. Then, ILOEED can be used for modeling and analyzing real-life data in Industries; (ii) The shape parameters of the proposed distribution make it very flexible to exhibit increasing and bath-tube failure rate shapes; (iii) Additionally, the density of the proposed distribution can also provide more flexible shapes. The paper is organized into six sections. The proposed distribution is defined in Section 2. The Statistical properties of the proposed distribution are presented in section 3. The estimation of the parameters of the proposed distribution using the method of Maximum Likelihood Estimates (MLEs) is introduced in Section 4. In Section 5, a simulation study based on the properties of the MLEs of the proposed distribution is presented. Applications of the proposed distribution to industry datasets are presented In Section 6. Section 7 concludes the paper.

The Inverse Lomax-Odd Exponentiated G (IL-OEG) family was proposed by [1] based on the T-X generator of [18].The cumulative density function (CDF) and probability density function (PDF) of IL-OEG are given as

$$F(x;\lambda,\gamma,\theta,\Delta)=\left[1+\lambda\left\{\frac{(1-G(x;\boldsymbol{\Delta}))}{G(x;\boldsymbol{\Delta})}\right\}^{\theta}\right]^{-\gamma};\ x>0,\lambda,\gamma,\theta,\boldsymbol{\Delta}>0 \tag{1}$$

And

$$f(x;\lambda,\gamma,\theta,\Delta)=\frac{\theta\gamma\lambda g(x;\Delta)[1-G(x;\boldsymbol{\Delta})]^{\theta-1}}{[G(x;\boldsymbol{\Delta})]^{\theta+1}}\left[1+\lambda\left[\frac{(1-G(x;\boldsymbol{\Delta}))}{G(x;\boldsymbol{\Delta})}\right]^{\theta}\right]^{-(1+\gamma)};\ x>0,\lambda,\gamma,\theta,\boldsymbol{\Delta}>0 \tag{2}$$

Where $\boldsymbol{\Delta}$ is a vector of parameter(s) for the baseline distribution, $G(x;\boldsymbol{\Delta})=1-G(x;\boldsymbol{\Delta})$. Based on equations (1) and (2), Exponential distribution is considered to be the baseline distribution. So, the $G(x;\Delta)$ is equivalent to an exponential density.

II. Inverse Lomax Odd Exponentiated Exponential distribution (ILOEED)

The exponential distribution is the probability distribution of the time between events in a Poisson point process, that is, a process in which events occur continuously and independently at a constant average rate. It is a subset of the gamma distribution. The CDF and PDF of the exponential distribution are presented in equations (3) and (4).

$$G(x;\upsilon) = 1 - e^{-\upsilon x}, x, \upsilon > 0 \tag{3}$$

And

$$g(x;\upsilon) = \upsilon e^{-\upsilon x}; x, \upsilon > 0 \tag{4}$$

Then, the CDF and PDF of the ILOEED can be given as:

$$F(x;\theta,\lambda,\gamma,\upsilon) = \left[1 + \lambda\left\{\frac{e^{-\upsilon x}}{(1-e^{-\upsilon x})}\right\}^{\theta}\right]^{-\gamma}; x, \upsilon, \gamma, \lambda, \theta > 0 \tag{5}$$

And

$$f(x;\theta,\lambda,\gamma,\upsilon) = \frac{\upsilon\theta\gamma\lambda e^{-\upsilon\theta x}\left[1 + \lambda\left\{\frac{e^{-\upsilon x}}{(1-e^{-\upsilon x})}\right\}^{\theta}\right]^{-(1+\gamma)}}{[1-e^{-\upsilon x}]^{\theta+1}}; x, \upsilon, \gamma, \lambda, \theta > 0 \tag{6}$$

Where υ is the rate, λ is the scale, and θ and γ are the shape parameters, respectively. Having this combination of parameters, we hope that ILOEED will fit datasets of different shapes. The reliability, hazard, and cumulative hazard functions of the ILOEED are presented in equations (7), (8), and (9).

$$R(x;\lambda,\gamma,\theta,\upsilon) = \int_t^{\infty} f(x;\lambda,\gamma,\theta,\upsilon)dx$$

$$= 1 - \left[1 + \lambda\left\{\frac{e^{-\upsilon x}}{(1-e^{-\upsilon x})}\right\}^{\theta}\right]^{-\gamma}; x, \lambda, \gamma, \theta, \upsilon > 0, \tag{7}$$

And

$$H(x;\lambda,\gamma,\theta,\upsilon) = \int_{\infty}^{x} h(v)dv = -log(R(x;\lambda,\gamma,\theta,\upsilon))$$

$$= -log\left\{1 - \left[1 + \lambda\left\{\frac{e^{-\upsilon x}}{(1-e^{-\upsilon x})}\right\}^{\theta}\right]^{-\gamma}\right\}. \tag{8}$$

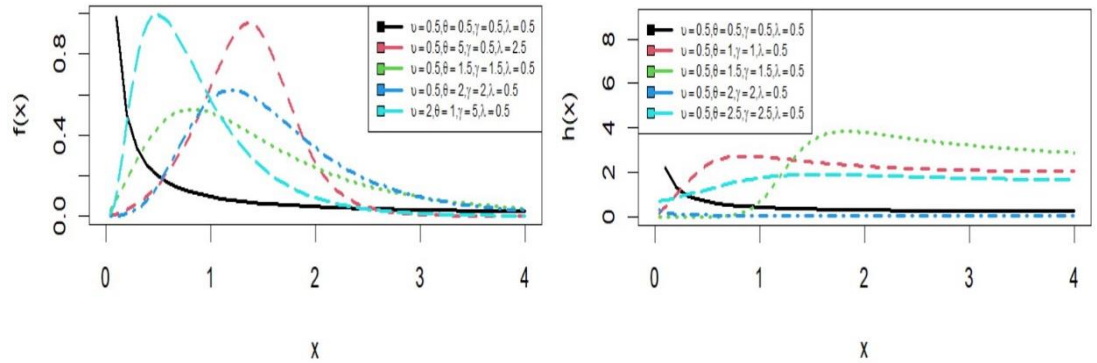


Figure 1: *PDF and CDF plots of ILOEED at various parameter values*

Figure (1) shows the various shapes of the ILOEED's PDF. This includes skewed and symmetry. Figure (1) also indicates the various shapes that the ILOEED can take which include constant, Bathtub, and monotone-increasing hazard shapes.

III. The Statistical Properties of the ILOEED

I. The quantile function of the ILOEED

The quantile function of ILOEED can be derived by inverting the CDF of the ILOEED given in equation (5) as follows:

$$F(x;\lambda,\gamma,\theta,\upsilon)=U=\left[1+\lambda\left\{\frac{e^{-\upsilon x}}{(1-e^{-\upsilon x})}\right\}^{\theta}\right]^{-\gamma} \tag{9}$$

Then,

$$U^{\frac{-1}{\gamma}}-1=\lambda\left\{\frac{e^{-\upsilon x}}{(1-e^{-\upsilon x})}\right\}^{\theta}$$

After simplifying and by making x the subject of the formula, we have

$$x=\frac{-log\left(\frac{k}{1+k}\right)}{\upsilon} \tag{10}$$

Where $k=\left(\frac{U^{\frac{-1}{\gamma}}-1}{\lambda}\right)^{\frac{1}{\theta}}$ and U is uniformly distributed between 0 and 1.

The median of the ILOEED family can be derived by setting U=0.5 in equation (10).

II. The moments of the ILOEED

Let X be a random variable that follows ILOEED with parameters $(\lambda,\gamma,\theta,\upsilon)$, then the C^{th} moment about the origin is given by:

$$\mu_C^{'}=E(X^C)=\int_0^{\infty}x^c f(x;\lambda,\gamma,\theta,\upsilon)dx$$

Using some linear representation, we have the moment of the ILOOED as

$$\mu_C^{'}=\sum_{k_1,k_2=0}^{\infty}\psi\,[k_2-\theta k_1]\int_0^{\infty}x^c\upsilon e^{-\upsilon x}(1-e^{-\upsilon x})^{[k_2-\theta k_1-1]}dx \tag{11}$$

Where $\psi=\frac{(-1)^{k_1+k_2}\Gamma(\gamma+k_1)\Gamma(\theta k_1+k_2)\lambda^{k_1}}{k_1\Gamma(\gamma)k_2!\Gamma(\theta k_1)}$. By considering the binomial expansion of the term $(1-e^{-\upsilon x})^{[k_2-\theta k_1-1]}=\sum_{j=0}^{\infty}(-1)^j\binom{[k_2-\theta k_1-1]}{j}e^{-j\upsilon x}$. Then, equation (11) becomes

$$\mu_C^{'}=\mathrm{Y}_{k,j}\int_0^{\infty}x^c e^{-\upsilon(1+j)x}dx=\frac{\mathrm{Y}_{k,j}c!}{\left[\upsilon(1+j)\right]^{c+1}}. \tag{12}$$

Where $\Upsilon_{k,j} = \sum_{j,k_1,k_2}^{\infty} \psi\upsilon[k_2 - \theta k_1 - 1](-1)^j \binom{[k_2 - \theta k_1 - 1]}{j}$. The mean of the ILOEED can be derived by setting c=1 in equation (12). Moreover, the second moment can also be derived by setting c=2, and then using the relation $Var(X) = \mu_2^{'} - [\mu_1^{'}]^2$ to find the variance.

III. The Characteristic and Moment Generating Functions of the ILOEED

The characteristic function of the ILOEED can be given as

$$L(x;\upsilon,\theta,\gamma,\lambda) = n\log(\lambda\gamma\theta\upsilon) - \upsilon\theta\sum_{i=1}^{n} x_i - (1+\gamma)\sum_{i=1}^{n} \log\left[1+\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}\right] - (1+\theta)\sum_{i=1}^{n}\log\left[1-e^{-\upsilon x_i}\right] \quad (13)$$

And the moment generating function of the ILOEED can be given as

$$\begin{aligned} M_X(t) &= \int_0^{\infty} e^{tx} f(x;\lambda,\gamma,\theta,\upsilon)dx \\ &= \sum_{k_1,k_2}^{\infty} \psi[k_2 - \theta k_1]\int_0^{\infty} e^{tx}\upsilon e^{-\upsilon x}(1-e^{-\upsilon x})^{[k_2-\theta k_1-1]}dx \\ &= \Upsilon_{k,j}\int_0^{\infty} e^{-(t-\upsilon(1+j))x}dx = \frac{\Upsilon_{k,j}}{\upsilon(1+j)-t} \end{aligned} \quad (14)$$

Table 1: *Mean, Variance, Skewness, and Kurtosis of ILOEED at various parameter values*

Parameter	E(X)	Var(X)	SK(X)	KUR(X)
$\upsilon=\theta=\gamma=\lambda=0.5$	1.0694	6.6518	4.1394	26.2247
$\upsilon=1,\theta=0.5,\gamma=1,\lambda=0.5$	0.9794	2.8142	3.325	14.6679
$\upsilon=1.5,\theta=0.5,\gamma=1.5,\lambda=0.5$	0.9061	1.6187	2.4044	10.900
$\upsilon=2,\theta=0.5,\gamma=2,\lambda=0.5$	0.8449	1.0637	2.0973	9.0716
$\upsilon=2.5,\theta=0.5,\gamma=2.5,\lambda=0.5$	0.7927	0.7549	1.8992	8.0158
$\upsilon=3,\theta=0.5,\gamma=3,\lambda=0.5$	0.7476	0.5639	1.7499	7.3406
$\upsilon=3.5,\theta=0.5,\gamma=3.5,\lambda=0.5$	0.7081	0.4372	1.6408	6.8795
$\upsilon=4,\theta=0.5,\gamma=4,\lambda=0.5$	0.6320	0.3486	1.5570	6.5505

Table (1) presents some basic statistics based on the moments of the ILOEED at varying parameter values. It's evident from the table that as the shape parameters increase, all the values of the statistics decrease. The skewness is positive and the tails tend towards the right. Positive kurtosis also indicates the tails are heavy.

IV. Maximum Likelihood Estimates (MLE)

In this section, we used the maximum likelihood method to estimate the parameters of the ILOEED. Let $x_1, x_2, x_3, x_4, ..., x_n$ be a random sample independently drawn from ILOEED family. Then, the log-likelihood function $L(\upsilon,\theta,\gamma,\lambda)$ of equation (6) is given as

$$L(x;\upsilon,\theta,\gamma,\lambda) = n\log(\lambda\gamma\theta\upsilon) - \upsilon\theta\sum_{i=1}^{n} x_i - (1+\gamma)\sum_{i=1}^{n} \log\left[1+\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}\right] - (1+\theta)\sum_{i=1}^{n}\log\left[1-e^{-\upsilon x_i}\right] \quad (15)$$

Taking the partial derivatives of equation (15) with respect to υ, θ, γ, and λ, yields:

$$\frac{\partial L}{\partial \upsilon}=\frac{n}{\upsilon}-\theta\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\frac{(1+\gamma)\theta\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta-1}xe^{\upsilon x}}{\left(1-e^{-\upsilon x}\right)^2\left[1+\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}\right]}. \tag{16}$$

$$\frac{\partial L}{\partial \theta}=\frac{n}{\theta}-\upsilon\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\frac{(1+\gamma)\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}log\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)}{\left[1+\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}\right]}-\sum_{i=1}^{n}\log\left[1-e^{-\upsilon x_i}\right]. \tag{17}$$

$$\frac{\partial L}{\partial \gamma}=\frac{n}{\gamma}-\sum_{i=1}^{n}log\left[1+\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}\right]. \tag{18}$$

$$\frac{\partial L}{\partial \lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}\frac{(1+\gamma)\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}}{\left[1+\lambda\left(\frac{e^{-\upsilon x_i}}{1-e^{-\upsilon x_i}}\right)^{\theta}\right]}. \tag{19}$$

The MLEs of the υ, θ, γ, and λ can be determine by solving the following non-linear equations (16), (17), (18), and (19) with respect to each parameter.

V. The Simulation Studies of the ILOEED

Here, we highlighted five (5) steps on how to do a Monte Carlo simulation study as follows:
Step 1: Clearly state the pseudo-population that can be used in generating random samples usually by writing code in a specific method. In this study, the pseudo-population is the quantile function of ILOEED given in equation (10).
Step 2: Sample from the population of interest (depending on your objective).
Step 3: Estimate the parameter of interest from the sample and keep it in a vector.
Step 4: Repeat the previous steps i.e 2 and 3 n-times (n is the number of trials).
Step 5: Create a relative frequency distribution of resulting values that is a Monte Carlo approximation of the distribution of samples under the conditions defined by the pseudo-population and the procedures of sampling. Based on the above procedure, we carry out a simulation studies as explained below:

i). For known parameter values i.e $\Theta=(\theta,\upsilon,\gamma,\lambda)^T$, we simulated a random sample of size n from the ILOEED using equation (10).

ii). We then estimated the parameters of the ILOEED by the method of maximum likelihood. iii). Perform 1,000 replications of steps i through ii.

iv). For each of the four (4) parameters of the ILOEED, we compute the mean, bias, and Root mean squared error (RMSE) from the 1,000 parameter estimates. The statistics are given by

$$\hat{\Theta}=\frac{1}{1,000}\sum_{i=1}^{1,000}\Theta_i, Bias(\hat{\Theta})=\frac{1}{1,000}\sum_{i=1}^{1,000}(\Theta_i-\Theta), RMSE(\hat{\Theta})=\sqrt{\frac{1}{1,000}\sum_{i=1}^{1,000}(\Theta_i-\Theta)^2} \tag{20}$$

Where $\Theta_i = (\hat{\theta}, \hat{\upsilon}, \hat{\gamma}, \hat{\lambda})$ is the MLE for each iteration (n=10, 20, 30, 50, 70, 90, 150, 170). Table (2) reports Case I. R software by [19] was used for the simulation. 123 was set as the seed for reproducibility. Three cases were considered for the simulation. Case I:(υ = 0.3, θ = 1, γ = 0.9 and λ = 0.4), Case II:(υ = 0.6, θ = 0.5, γ = 0.9 and λ = 0.5), and Case III: (υ = 2, θ = 5, γ = 0.5 and λ = 0.3). Tables (3) and (4) are for Case II and Case III, respectively.

Table 2: *Simulation Results for Case I*

N	Estimates	Bias	RMSE	n	Estimates	Bias	RMSE
10	0.4219	0.1219	0.2651	70	0.3335	0.0335	0.1254
	1.1481	0.1481	0.4772		1.0131	0.0131	0.1942
	1.0841	0.1841	0.4819		0.9849	0.0849	0.305
	0.5721	0.1721	0.5426		0.4487	0.0487	0.2497
20	0.3735	0.0735	0.2062	90	0.3275	0.0275	0.1124
	1.0835	0.0835	0.348		1.0048	0.0048	0.1701
	1.0434	0.1434	0.4297		0.9808	0.0808	0.2909
	0.5046	0.1046	0.4076		0.4373	0.3730	0.2225
30	0.3540	0.0540	0.1713	150	0.3204	0.0204	0.0853
	1.0533	0.0533	0.2797		0.0998	-0.0021	0.1396
	1.0039	0.1039	0.3742		0.9600	0.0600	0.2393
	0.4951	0.0951	0.3606		0.4293	0.0293	0.1763
50	0.3378	0.0378	0.1405	170	0.3174	0.0174	0.0823
	1.0324	0.0324	0.2324		0.9969	-0.0031	0.1315
	0.9905	0.0905	0.3365		0.9549	0.0549	0.2261
	0.4550	0.0550	0.2725		0.4219	0.0219	0.1635

Table 3: *Simulation Results for Case II*

n	Estimates	Bias	RMSE	N	Estimates	Bias	RMSE
10	0.7954	0.1954	0.4988	70	0.6469	0.0469	0.2290
	0.6110	0.1110	0.3080		0.5111	0.0111	0.1061
	1.0250	0.1250	0.4286		0.9809	0.0809	0.2904
	0.6217	0.1217	0.5800		0.5105	0.0105	0.2252
20	0.7301	0.1301	0.3991	90	0.6320	0.0320	0.1907
	0.5496	0.0496	0.1884		0.5084	0.0084	0.0931
	1.0232	0.1232	0.3936		0.9665	0.0665	0.2571
	0.5751	0.0751	0.4643		0.5096	0.0096	0.2034
30	0.7013	0.1013	0.3505	150	0.6243	0.0243	0.1497
	0.5289	0.0289	0.1521		0.5032	0.0032	0.0724
	0.9932	0.0932	0.3403		0.9512	0.0512	0.2258
	0.5607	0.0607	0.3573		0.5070	0.0070	0.1680
50	0.6588	0.0588	0.2572	170	0.6192	0.0192	0.1462
	0.5177	0.0177	0.1177		0.5028	0.0028	0.0718
	0.9869	0.0869	0.3146		0.9469	0.0469	0.2102
	0.5268	0.0268	0.2862		0.5025	0.0025	0.1513

Table 4: *Simulation Results for Case III*

n	Estimates	Bias	RMSE	N	Estimates	Bias	RMSE
10	2.2708	0.2708	0.5049	70	2.1298	0.1298	0.2559
	5.4425	0.4425	1.7522		5.0473	0.0473	0.8753
	0.7649	0.2649	0.6554		0.5417	0.0417	0.1907
	0.5027	0.2027	0.4227		0.4346	0.1346	0.2376
20	2.2121	0.2121	0.3836	90	2.1104	0.1104	0.2116
	5.2441	0.2441	1.3917		5.0473	0.0473	0.7814
	0.6323	0.1323	0.3960		0.5322	0.0322	0.1673
	0.4837	0.1837	0.3432		0.4188	0.1188	0.2100
30	2.1779	0.1779	0.3433	150	2.0906	0.0906	0.1679
	5.2282	0.2282	1.2953		5.0259	0.0259	0.6080
	0.5810	0.0810	0.3174		0.5159	0.0159	0.1181
	0.4822	0.1822	0.3395		0.4053	0.1053	0.1776
50	2.1505	0.1505	0.2937	170	2.0785	0.0785	0.1473
	5.1003	0.1003	1.0516		5.0160	0.0160	0.5512
	0.5593	0.0593	0.2506		0.5135	0.0135	0.1033
	0.4484	0.1484	0.2640		0.3933	0.0933	0.1581

Tables (2), (3), and (4) presents the simulation results. As the value of the sample size (n) increases, the simulation results of the ILOEED show:

- Stability of the MLES,
- The bias of the MLEs approach zero, and
- Decrease in the RMSEs of the MLEs.

VI. Applications of the ILOEED to Industry Datasets

Odd Exponentiated Inverse Lomax Distribution (ILOEED) was fitted to three datasets. This includes datasets with increasing and bathtub hazard shapes. The Goodness-of-fit criteria used are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). A statistical metric called the Akaike Information Criterion (AIC) by [20] was employed in the selection of models. AIC offers a quantitative method for weighing a model's complexity concerning it's goodness of fit, intending to choose the model that most accurately describes the data while preventing overfitting. Bayesian Information Criterion (BIC), a statistical metric for model selection, was developed by [21]. Similar to AIC, BIC takes a Bayesian approach to balancing model fit and complexity. Moreover, three Goodness-of-fir statistics were used. These are the Kolmogorov-Smirnov (K-s), Anderson-Darling (A-D), and Cramer-Von Mises (C-vM). By comparing the cumulative distribution functions (PDFs) of two datasets, the Kolmogorov-Smirnov test determines whether they have the same continuous distribution. It is a widely applicable test that helps to compare theoretical and empirical distributions and is especially helpful when parametric assumptions are not met. The A-D test is useful for determining fit, especially in situations with extreme values or interesting tail behavior. It does this by calculating the goodness-of-fit between the sample's empirical distribution function and the specified distribution's cumulative distribution function. This test is sensitive to deviations in the distribution's tails. The C-vM test, named for

Carl von Mises and Harald Cramer, measures the difference between the sample's empirical distribution function and the specified distribution's cumulative distribution function to determine whether the sample fits the distribution. This test is favored in some circumstances because it is easy to compute and provides a good measure of goodness-of-fit. Finally, The negative log-likelihood (-ll), which is frequently minimized in maximum likelihood estimation, measures how well a model fits observed data by esti-mating the probability of observing the data given the model parameters. It is favored for its stability and ease of use in parameter estimation and is essential to many different disciplines, including biology, econometrics, and machine learning. Fitdistrplus package by [22] in R was used in fitting the three datasets in this section.

I. Application to the Breaking Stress of Carbon Fibres Dataset

These data as reported by the [23], represent the breaking stress of carbon fibres of 50mm length in Gpa. The data is symmetry and has an increasing hazard shape. The dataset is as follows:
0.39, 0.85, 1.08, 1.25, 1.47, 1.57, 1.61, 1.61, 1.69, 1.80, 1.84, 1.87, 1.89, 2.03, 2.03, 2.05, 2.12, 2.35, 2.41, 2.43, 2.48, 2.50, 2.53, 2.55, 2.55, 2.56, 2.59, 2.67, 2.73, 2.74, 2.79, 2.81, 2.82, 2.85, 2.87, 2.88, 2.93, 2.95, 2.96, 2.97, 3.09, 3.11, 3.11, 3.15, 3.15, 3.19, 3.22, 3.22, 3.27, 3.28, 3.31, 3.31, 3.33, 3.39, 3.39, 3.56, 3.60, 3.65, 3.68, 3.70, 3.75, 4.20, 4.38, 4.42, 4.70, 4.90.
ILOEED was fitted alongside the Inverse Exponentiated Odd Lomax Exponential Distribution (IEOLED) by [4], Alpa Power Exponential Distribution (APED) by [24], as well as the Exponential Distribution (ED).

Table 5*: MLEs and Goodness-of-fit Criteria for the fitted ILOEED and other comparators for the comparators for the Breaking Strengths of Carbon Fibres Dataset*

Distributions	Estimates	Standard Error	-ll	AIC	BIC
ILOEED$(\upsilon,\theta,\gamma,\lambda)$	0.6586	1.3019	84.6124	177.2247	185.9833
	3.4545	5.7285			
	0.4797	0.3904			
	1047.5537	4727.8614			
IEOLED$(\upsilon,\theta,\gamma,\lambda)$	0.0854	0.0274	86.1725	180.3251	189.1037
	3.369	0.2963			
	1.03E+07	1.67E+04			
	1.69E+12	2.37E+04			
APED(υ,θ)	2.99E+05	1.19E+04	92.3964	188.7927	193.172
	1.10E+00	5.46E-02			
ED(υ)	3.63E-01	4.46E-02	132.9944	267.9887	270.1784

Table (5) presents the MLEs, log-likelihoods, AICs, and BICs of the ILOEED and others. The Table indicates that ILOEED is the best with minimum values of AIC and BIC. Furthermore, Table (6) indicated that the ILOEED fitted the data well with small values of the Goodness-of-fit statistics. These are the K-S, C-vM, and A-D.

Table 6: *The Goodness-of-fit statistics of the ILOEED and others for the Breaking Strengths of Carbon Fibres Dataset*

Distributions	K-S	C-vM	A-D
ILOEED	0.0565	0.033	0.2298
IEOLED	0.085	0.0884	0.5055
APED	0.1353	0.2826	1.6089
ED	0.3581	2.871	14.0343

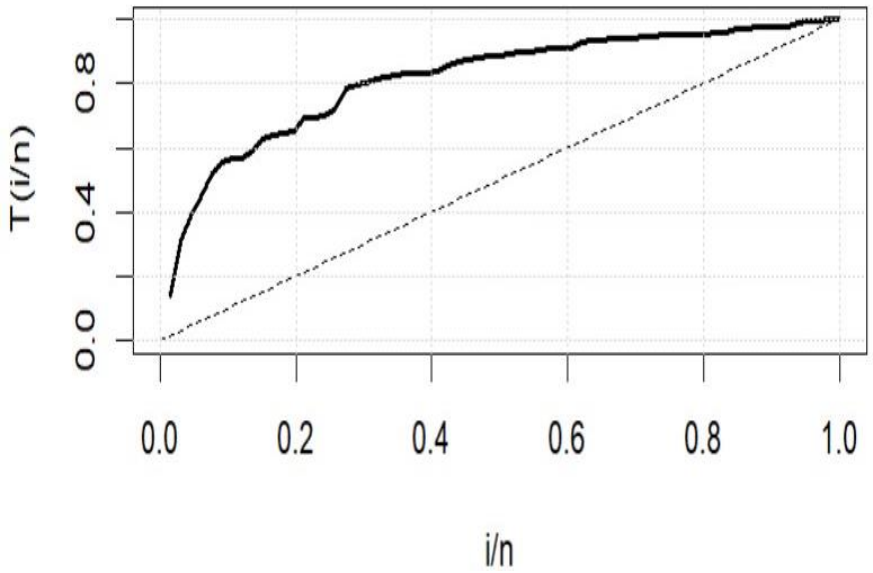


Figure 3: *The TTT-Plot for the Carbon Fibre Dataset*

The Total Time on Test (TTT) plot for the Carbon Fibre Dataset indicates an increasing hazard rate (concave shape), as seen in Figure (3).

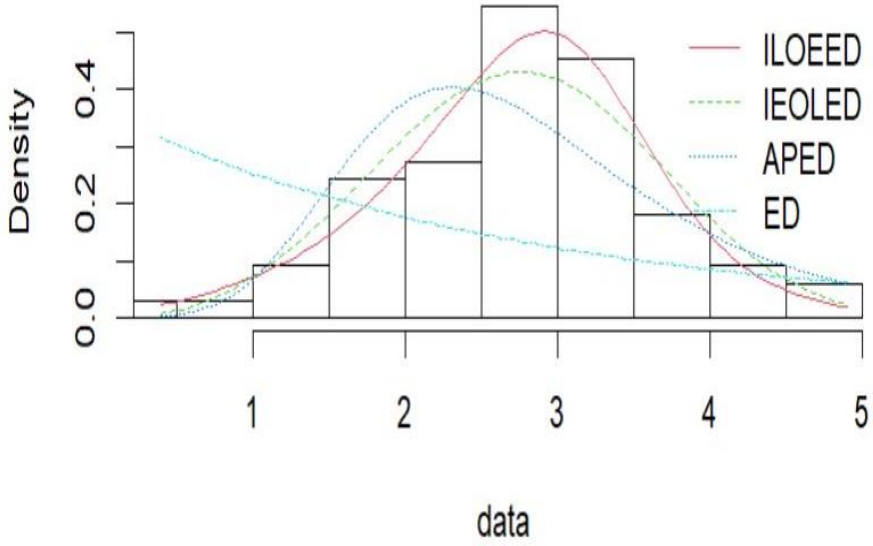


Figure 4: *The Fitted PDFs and CDFs of the ILOEED and others for the Carbon Fibre Dataset*

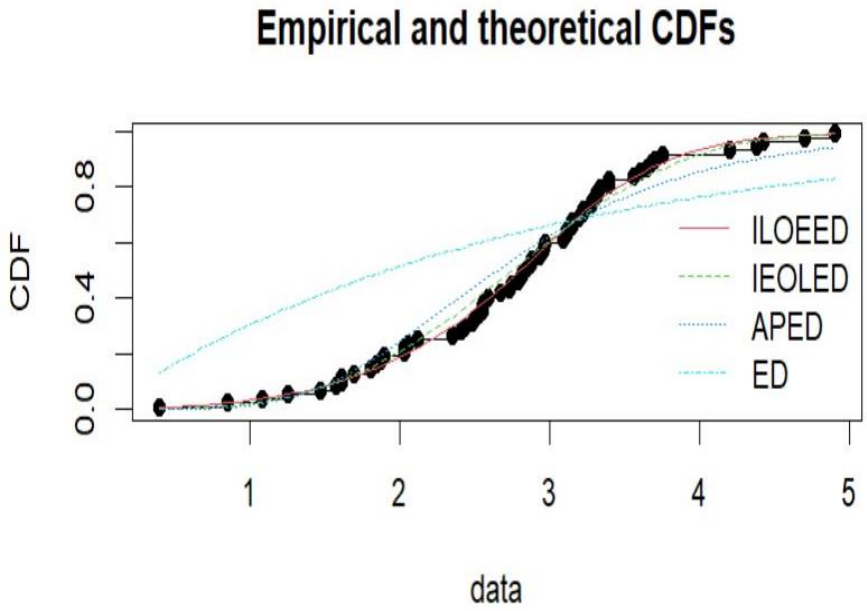


Figure 5: *The Fitted PDFs and CDFs of the ILOEED and others for the Carbon Fibre Dataset*

Figures (4) and (5) indicates that the Carbon Fibre data is symmetrical. ILOEED fitted the data better compared with the other comparators.

II. Application to the Strengths Dataset

The following dataset reported by [25], is the Strengths reading in GPa of individual carbon fibres that were put to stress at 20mm gauges and the values are:
1.359, 1.382, 1.382, 1.426, 1.434, 1.435, 1.478, 1.490, 1.511, 1.514, 1.535, 1.554, 1.566, 1.570, 1.586, 1.629, 1.633, 1.642, 1.648, 1.684, 1.697, 1.726, 1.770, 1.773, 1.800, 1.809, 1.818, 1.821,1.848, 1.880, 1.954, 2.012, 2.067, 2.084, 2.090, 2.096, 2.128, 2.233, 2.433, 2.585, 2.585.

Table 7: *MLEs and Goodness-of-fit Criteria for the fitted ILOEED and other comparators for the comparators for the Strengths Dataset*

Distributions	Estimates	Standard Error	-ll	AIC	BIC
ILOEED$(\upsilon,\theta,\gamma,\lambda)$	0.7639	1.2163	86.0082	180.0163	188.775
	3.1315	4.4416			
	0.4570	0.3044			
	1943.9353	5787.6972			
IEOLED$(\upsilon,\theta,\gamma,\lambda)$	0.1017	0.0311	88.2736	184.5472	193.3058
	3.3417	0.2939			
	786928.5	12792.55			
	6.4E+10	1.79E+04			
APED(υ,θ)	59738.7189	17013.25	95.0118	194.0236	198.4029
	1.0525	0.0556			
ED(υ)	0.5642	0.0881	96.0231	195.0986	197.0178

Table (7) presents the MLEs, log-likelihoods, AICs, and BICs of the ILOEED and others. The

Table indicates that ILOEED is the best with minimum values of AIC and BIC. Furthermore, Table (8) indicated that the ILOEED fitted the data well with small values of the Goodness-of-fit statistics.

Table 8*: The Goodness-of-fit statistics of the ILOEED and others for the Strengths Dataset*

Distributions	K-S	C-vM	A-D
ILOEED	0.0581	0.046	0.2382
IEOLED	0.0876	0.0949	0.6043
APED	0.1461	0.3327	1.9048
ED	0.3447	2.7639	13.5301

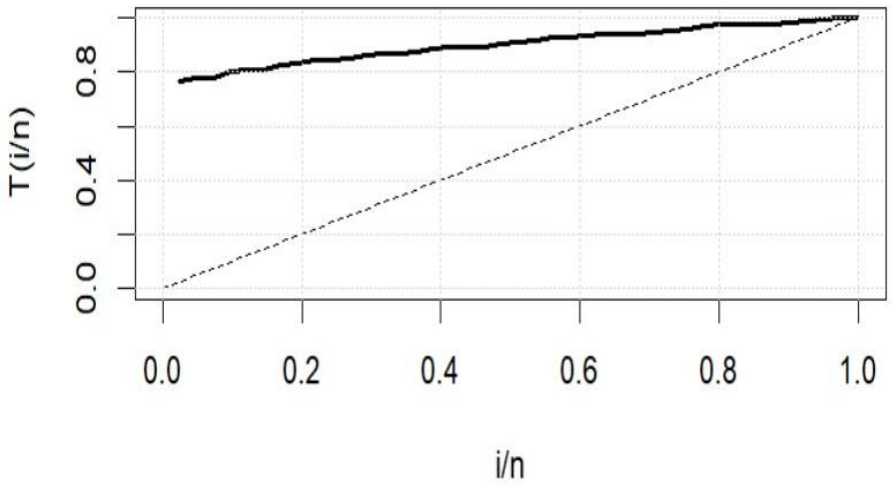


Figure 6: *The TTT-Plot for the Strengths Dataset*

The Total Time on Test (TTT) plot for the Strengths Dataset indicates an increasing hazard rate (concave shape), as seen in Figure (6).

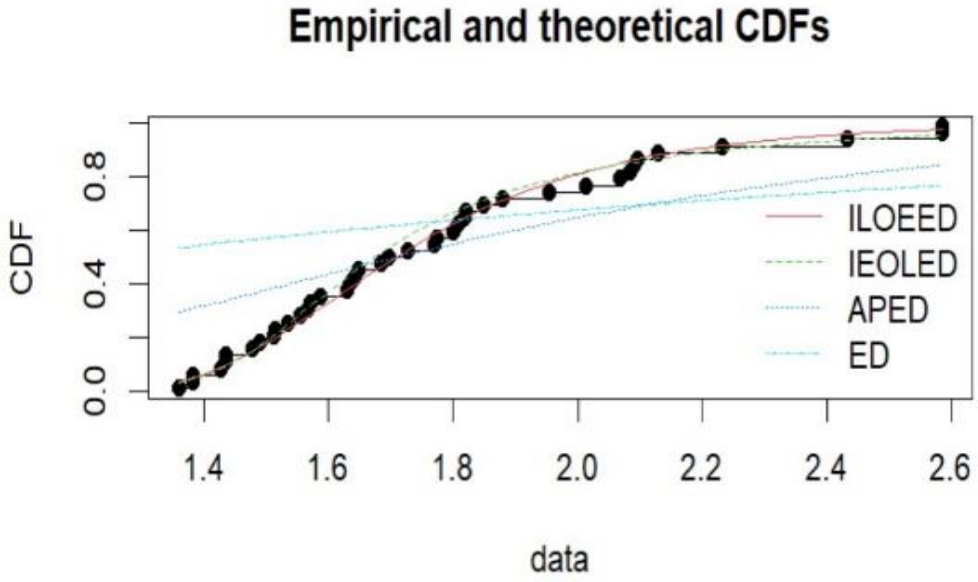


Figure 7: *The Fitted CDFs of the ILOEED and others for the Strengths Dataset*

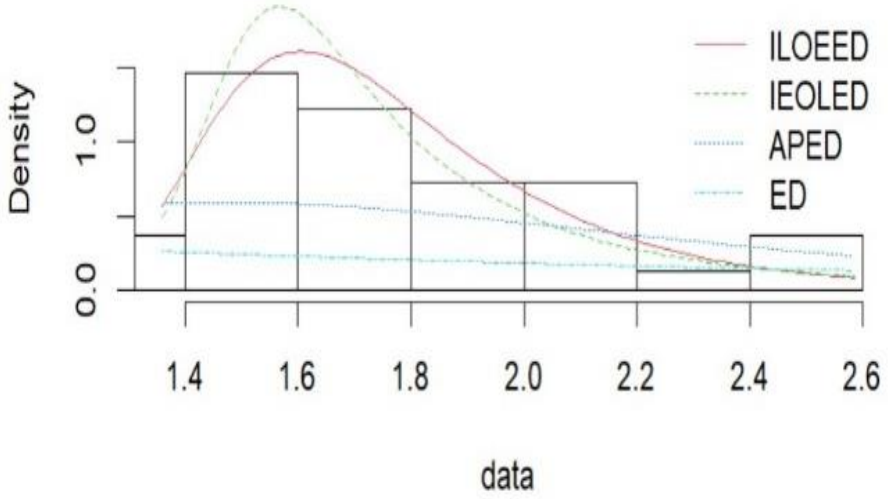


Figure 8: *The Fitted PDFs of the ILOEED and others for the Strengths Dataset*

Figures (7) and (8) indicates that the Strengths data is symmetry. ILOEED fitted the data well compared with the other comparators.

III. Application to the Times to Failure Dataset

This dataset has a Bathtub-shape hazard rate as reported by [26]. The data is about the Times to Failure of 50 Devices that were on life test at time 0. The dataset is as follows:
0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6, 7, 1, 1, 12, 18, 18, 18, 18, 18, 21, 32, 36, 40, 45, 46, 47, 50, 55, 60, 63, 63, 67, 67, 67, 67, 72, 75, 79, 82, 82, 83, 84, 84, 84, 85, 85, 85, 85, 85, 86, 86.

Table 9: *MLEs and Goodness-of-fit Criteria for the fitted ILOEED and other comparators for the comparators for the Times to Failure Dataset*

Distributions	Estimates	Standard Error	-ll	AIC	BIC
ILOEED$(\upsilon,\theta,\gamma,\lambda)$	0.1025	0.0478	234.3902	476.7803	484.5076
	0.3363	0.2143			
	1.4327	1.7004			
	2.3317	4.4526			
IEOLED$(\upsilon,\theta,\gamma,\lambda)$	0.1104	0.2196	243.9899	495.9799	503.7072
	0.8589	0.1104			
	2968.220	5982.6273			
	490198	1.4053			
APED(υ,θ)	1.9690	1.4053	244.2936	492.5873	496.4509
	0.0259	0.0048			
ED(υ)	0.0224	0.0031	244.7001	491.4002	493.3321

Table (9) presents the MLEs, log-likelihoods, AICs, and BICs of the ILOEED and others. The Table indicates that ILOEED is the best with minimum values of AIC and BIC. Furthermore, Table (10) indicated that the ILOEED fitted the data well with small values of the Goodness-of-fit statistics.

Table 10: ***The Goodness-of-fit statistics of the ILOEED and others for the tomes to Failure Dataset***

Distributions	K-S	C-vM	A-D
ILOEED	0.1503	0.2817	1.9053
IEOLED	0.2012	0.5934	3.8944
APED	0.1799	0.5331	4.7647
ED	0.1904	0.5727	4.5609

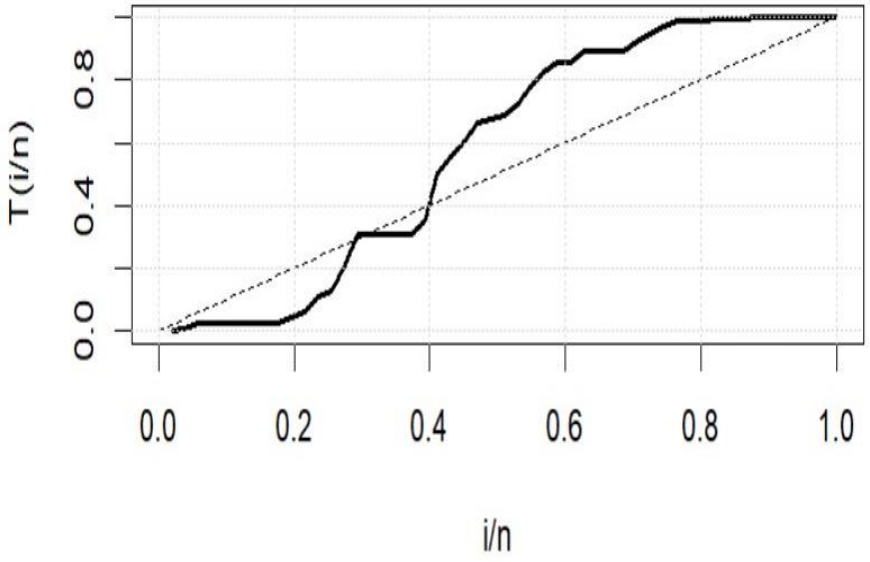


Figure 9: *The TTT-Plot for the Times to Failures Dataset*

The Total Time on Test (TTT) plot for the Fatigue Fracture Dataset indicates a decreasing and then increasing hazard rate (Bath-tub shape), as seen in Figure (9).

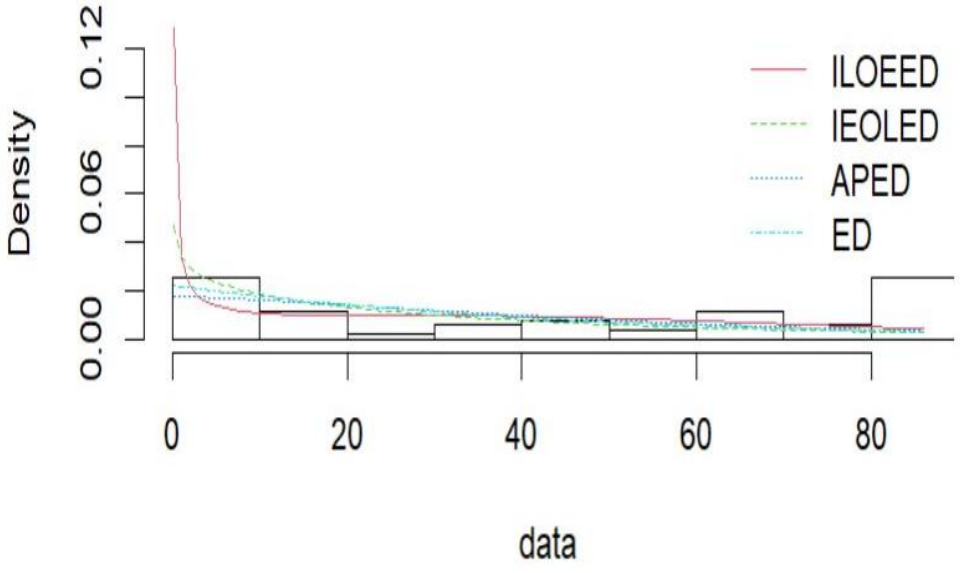


Figure 11: *The Fitted PDFs of the ILOEED and others for the Times to Failure Dataset*

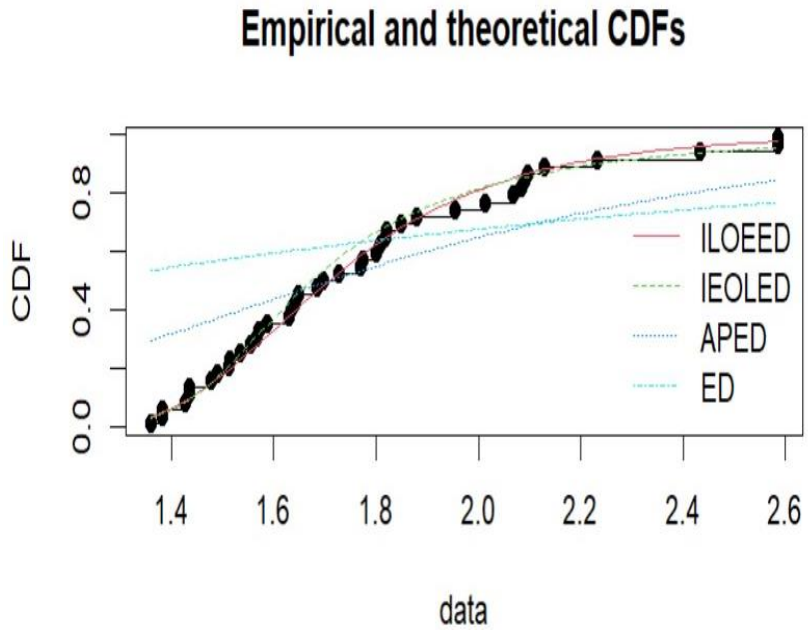


Figure 11: *The Fitted CDFs of the ILOEED and others for the Times to Failure Dataset*

Figures (10) and (11) indicates that the Times to Failure data is skewed to the right. ILOEED fitted the data well compared with the other comparators.

VII. Conclusion

In this research, we suggest and investigate a novel probability distribution that is a combination of the Inverse Lomax and exponential distributions, combining the properties of the two. This merger is required if the data in issue combines both the Inverse Lomax and the exponential distributions' features described in Section 1. We looked into some of its statistical properties, such as moments, the moment generating function, the characteristic function, and the quantile function. The parameters were determined using the maximum likelihood technique. According to the simulation studies, as the sample size grows, the estimations of the Biases and RMSEs approach zero, indicating that the estimates are more accurate. Three cases of parameter combination were considered for the simulation studies. The estimates were stable. Exemplifications of real-world datasets demonstrate the ILOEED's significance. For the three datasets used, the proposed distribution is the best with minimum values of the Goodness-of-fit criteria and Goodness-of-fit statistics. This means ILOEED can be used to fit datasets with increasing and bath-tube hazard rates. Based on these facts, we hope that the ILOEED will be preferred above the other models considered in this study. Only datasets from the industry were considered to fit the proposed distribution. We suggest that other areas should be explored in terms of the application of the proposed distribution. Also, other methods of estimation can be considered in further studies.

References

[1] Falgore, J. Y. and Doguwa, S. I. (2020). Inverse lomax-exponentiated-g (il-eg) family of distributions: properties and applications, *Asian Journal of Probability and Statistics*, 9 (4):48–64.

[2] Hassan, A. S., Almetwally, E. M., Gamoura, S. C., and Metwally, A. S. (2022). Inverse exponentiated lomax power series distribution: Model, estimation, and application, *Journal of Mathematics*, 1-21.

[3] Dhungana, G. P. and Kumar, V. (2022). Exponentiated odd lomax exponential distribution with application to covid-19 death cases of nepal, *PloS one*, 17 (6) e0269450.

[4] Chaudhary, A. K., Telee, L. B. S., and Kumar, V. (2022). Inverse exponentiated odd lomax

exponential distribution: Properties and applications, *Nepalese Journal of Statistics*, 6:29–50.

[5] Al-Marzouki, S., Jamal, F., Chesneau, C., Elgarhy, M. (2021). Half logistic inverse lomax distribution with applications, *Symmetry*, 13 (2): 309.

[6] Ogunde, A. A., Chukwu, A. U., and Oseghale, I. O. (2023). The kumaraswamy generalized inverse lomax distribution and applications to reliability and survival data, Scientific African 19: e01483.

[7] Sindhu, T. N. and Atangana, A. (2021). Reliability analysis incorporating exponentiated inverse weibull distribution and inverse power law, *Quality and Reliability Engineering International*, 37 (6): 2399–2422.

[8] Sindhu, T. N., Shafiq, A., and Al-Mdallal, Q. M. (2021). Exponentiated transformation of gumbel type-ii distribution for modeling covid-19 data, *Alexandria Engineering Journal* 60 (1): 671–689.

[9] Metwally, A. S. M., Hassan, A. S., Almetwally, E. M., Kibria, B. G., and Almongy, H. M. (2021). Reliability analysis of the new exponential inverted topp–leone distribution with applications, *Entropy*, 23 (12):1662.

[10] Sindhu, T. N., Shafiq, A., and Al-Mdallal, Q. M. (2021). On the analysis of number of deaths due to covid- 19 outbreak data using a new class of distributions, *Results in Physics* 21: 103747.

[11] Shafiq, A., Sindhu, T. N., and Alotaibi, N. (2022). A novel extended model with versatile shaped failure rate: Statistical inference with covid-19 applica-tions, *Results in Physics*, 36: 105398.

[12] Shrahili, M., Hassan, A. S., Almetwally, E. M., Ghorbal, A. B., and Elbatal, I. (2022). Alpha power moment exponential model with applications to biomedical science, *Scientific Programming*, *2022*(1):6897405.

[13] Hassan, A. S., Mohamed, R. E., Kharazmi, O., and Nagy, H. F. (2022). A new four parameter extended exponential distribution with statistical properties and applications, *Pakistan Journal of Statistics and Operation Research*, 18(1):179–193.

[14] Sindhu, T. N., Colak, A. B., Lone, S. A., and Shafiq, A. (2023). Reliability study of generalized exponential distribution based on inverse power law using artificial neural network with bayesian regularization, Quality and Reliability Engineering International 39 (6): 2398–2421.

[15] Abdelwahab, M. M., Ghorbal, A. B., Hassan, A. S., Elgarhy, M., Almetwally, E. M., and Hashem, A. F. (2023). Classical and bayesian inference for the kavya–manoharan generalized exponential distribution under generalized progressively hybrid censored data, *Symmetry* 15 (6):1193.

[16] Alghamdi, S. M., Shrahili, M., Hassan, A. S., Gemeay, A. M., Elbatal, I., and Elgarhy, M. (2023). Statistical inference of the half logistic modified kies exponential model with modeling to engineering data, *Symmetry*, 15 (3):586.

[17] Sindhu, T. N., Shafiq, A., and Huassian, Z. (2024). Generalized exponentiated unit gompertz distribution for modeling arthritic pain relief times data: classical approach to statistical inference, *Journal of Biopharmaceutical Statistics*, 34 (3):323–348.

[18] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions, *Metron*, 71 (1):63–79.

[19] Team, R. C. (2013). R: A language and environment for statistical computing. r foundation for statistical computing, (No Title).

[20] Akaike, H. (1974). A new look at the statistical model identification, *IEEE transactions on automatic control*, 19 (6):716–723.

[21] Schwarz, G. (1978). Estimating the dimension of a model, *The annals of statistics*, 461–464.

[22] Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An r package for fitting distributions, *Journal of statistical software* 64:1–34.

[23] Yousof, H. M., Alizadeh, M., Jahanshahi, S., Ramires, T. G., Ghosh, I., and Hamedani, G. (2017). The transmuted topp-leone g family of distributions: theory, characterizations and

applications, *Journal of Data Science,* 15 (4):723–740.

[24] Mahdavi, A. and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution, *Communications in Statistics-Theory and Methods* 46 (13):6543–6557.

[25] Nichols, M. D. and Padgett, W. (2006). A bootstrap control chart for weibull percentiles, *Quality and reliability engineering international*, 22 (2):141–151.

[26] Aarset, M. V. (1987). How to identify a bathtub hazard rate, *IEEE Transactions on Reliability*, 36 (1):106–108.

REGRESSION-TYPE IMPUTATION SCHEME UNDER SUBSAMPLING WITH EQUAL CHANCE OF RANDOM NON-RESPONSE AT FIRST STAGE.

Abubakar Ibrahim[1], Yahaya Abubakar [2], Garba Jamilu[3] and Aliyu Yakubu[4]

•

[1]Department of Mathematical Sciences, Kaduna State University Kaduna, Nigeria
[2, 3, 4]Department of Statistics, Ahmadu Bello University, Zaria, Kaduna, Nigeria
[1]ibrahim.abubakar@kasu.edu.ng. [2]ensiliyu2@yahoo.co.uk, [3]gyjamilu@gmail.com, [4]yakubualiyu@abu.edu.ng

Abstract

The study addresses the challenges of estimating the population mean in two-stage cluster sampling, where there is an equal chance of random non-response at the first-stage unit. The researchers propose some regression-type imputation schemes and regression-type estimators that incorporate measurement error parameters for both the study and supplementary variables. The properties of the proposed estimators were derived and numerically compared using a simulated sample population. The proposed estimators outperformed the existing estimators consider in the study. The researchers conclude that their proposed methodology can be practically applied, using the actual responses of the respondents and including the measurement error parameters to estimate the finite population mean.

Keywords: First Stage Unit,Regression-type imputation scheme,Regression-type estimators,Random Non-Response

I. Introduction

In field surveys frequently indicate that it is not always possible to obtain a complete list of everyone who is a member of the research population, indicating that selecting a simple random sample is difficult. Cluster sampling can be used to collect data in this scenario because it is usually less expensive and does not require a list of all observations in the population [1].
Clusters are produced in cluster sampling by dividing the survey area into smaller sub-areas. Then using simple random sampling, some of these areas are chosen, and all elements of the chosen clusters are counted.

Assuming we are interested in the academic performance of all 400 level students in a given city. Because there is no sampling frame for such units, obtaining a complete list of everyone in the research population is extremely difficult. However, a list of university each student attends should be available. In these cases it is recommended to select a simple random sample of 400 level students from each university. The technique used in this study is two-stage cluster sampling. In cluster sampling, better precision is achieved by first selecting a cluster and then enumerating a specific number of elements from each cluster. Two-stage cluster sampling refers to the process of first picking clusters, which are the sampling units in the first stage, and then selecting a predetermined number of elements from each selected cluster, which are the sampling units in the second stage. The clusters that constitute the sampling units in the first stage are referred to as First

Stage Units (FSUs) or Primary Stage Units. The elements inside these clusters that form the sampling units in the second stage are referred to as Second Stage Units (SSUs).The key advantage of this two-stage cluster sampling approach is that it can provide better precision in the estimates compared to simpler random sampling methods.

In a sample survey, it is usually assumed that all information is obtained from the study population's unit and that the observed variables are obtained without error. Such an assumption is not always met because researchers face the issues of non-response and measurement error. Most human population surveys face the problem of non-response, where some units of the study population fail to provide the requested information for various reasons, such as refusal, absence, lack of interest, or adherence to ethical standards. This non-response causes issues during data collection, calculation, and estimation. The problem of non-response in estimating the finite population mean was first address by [2]. The typical approach is to return to the field and collect the missing values through a call-back method, but this requires additional resources like time, people, and money. Three concepts related to non-response: Missing at Random (MAR), Observed at Random (OAR), and Parameterized Distribution (PD), were discussed by [3]. According to [3], data are MAR when the probability of missing data does not depend on the value of the unobserved data. Missing completely at random (MCAR) and missing at random (MAR) was distinguished by [4]. Various imputation schemes have been used over time to address the problem of estimating unknown parameters in the presence of missing values in sample surveys. Imputation involves filling in missing values with specific substitutes so that standard data analysis methods can be applied. The regression imputation method is used to replace missing value with a linear function. The approach also predicts missing values using regression models using the other variables, and the fitted values are entered into the model. It is assumed that the value of one variable varies linearly with other variables. Several researchers, including [5-14], and many others have proposed imputation methods to handle missing data. However, drawing simple random samples is impractical without a full list of every unit in the population. As a result, the imputation schemes and their estimators suggested by the previous literature are not applicable when the complete list of all population units is not available (as is the case in non-response). Cluster sampling is a common sampling method when there is no complete list of all population units in a survey. Hence the adoption of two-stage cluster sampling method used in this study.

In addition to non-response, survey researchers face the issue of measurement error (ME). Several survey researchers work under the assumption that the information they acquire from respondents is correct, and some of the estimator attributes (biases and mean square errors) are derived from this assumption. The assumption that the observed data accurately represents the true values is not always correct, as researchers often face the problem of measurement error. Measurement error refers to the discrepancy between the observed values gathered from respondents and the real, underlying values. In other words, the observed data may not perfectly reflect the true information, and there is an element of error or inaccuracy introduced during the data collection process. This measurement error can be problematic and needs to be accounted for in the analysis, as relying solely on the observed data may lead to biased or inaccurate results. Let's assume we want to collect information on the cumulative grade point average (CGPA) of some students; some may report a CGPA that is lower or higher than their real CGPA. As a result, the observed value remains erroneous, because the students did not provide their real CGPA. When the measurement error is insignificant, the inferences drawn on the observed value may be correct; nevertheless, when the measurement is not insignificantly small, the inferences taken on the observed value may have some unanticipated and unpleasant implications.

Several researchers have examined the issue of measurement error separately in their work, including [15-24].

Typically, non-response and ME are investigated separately using known supplementary variables. In reality, survey sampling results to simultaneous measurement errors and non-response. However, in this research, we will study both non-response and measurement errors.

II. Methods

2. Construction of sample structure

Suppose U is a finite population be divided into N FSU represented by $\left(U_1, U_2, \cdots U_N\right)$ in such a way that the quantity of SSU in every first stage unit is M. Assume y_{ij}, x_{1ij}, and x_{2ij} be the actual values for the character under study Y, first supplementary variable X_1 and second supplementary variable X_2, respectively. Also, assume that $y_{ij(e)}$, $x_{1ij(e)}$, $x_{2ij(e)}$ be the observed values for Y, X_1, X_2 on the j^{th} second stage units $\left(j = 1,2,\cdots M\right)$ in the i^{th} first stage units $\left(i = 1,2,\cdots N\right)$. Let U_{ij}, V_{ij} and W_{ij} be the ME parameters associated with the study variable, first supplementary variable, and second supplementary variable respectively. The MEs associated with these variables are thus defined as:

The ME associated with the character under study be

$$U_{ij} = y_{ij} - y_{ij(e)},\ U_{ij} \sim N\left(0, \sigma_U^2\right) \tag{1}$$

The ME associated with the first supplementary variable be

$$V_{ij} = x_{1ij} - x_{1ij(e)}, V_{ij} \sim N\left(0, \sigma_V^2\right) \tag{2}$$

The ME associated with the second supplementary variable be

$$W_{ij} = x_{2ij} - x_{2ij(e)},\ W_{ij} \sim N\left(0, \sigma_W^2\right) \tag{3}$$

However, in this research work, we take into account a scenario in which the information on the first supplementary variable X_1 is not known at the level of first stage unit. Hence, information on the first supplementary variable x_1 can be gathered using the following strategy:

STRATEGY: At the level of first stage units, information on the first supplementary variable x_1 is gathered, and a sample of the first stage unit is chosen using the SRSWOR procedure.

Moreover, the above-mentioned strategy will be discussed under clusters with equal chance of random non-response which is detailed below.

2.1. Clusters with equal chance of random non-response

2.1.1. Strategy: When supplementary information is gathered at Level of First Stage Unit

We take into account a situation where the population mean $\bar{X}_{1..}$ of the first supplementary variable x_1 is unknown at the level of first stage unit, so we used a two-phase or double sampling strategy to furnish the estimate. However, information on the second supplementary variable x_2 is known for every unit of the population. In order to estimate the population mean of Y, a first phase sample $S_{n'}\left(S_{n'} \subset U\right)$ of size n' first stage unit is taken out of N FSU from the population using SRSWOR method followed by a second phase sample S_1 of size n FSU $\left(n < n'\right)$ taken based on the subsequent two cases by using the SRSWOR technique to observe the character under study Y.

Case A: To create a S_1, a subsample of $S_{n'}\left(S_1 \subset S_{n'}\right)$ is taken.

Case B: In this case, S_1 is drawn independently of $S_{n'}$.

Furthermore, in order to estimate the population mean of Y, a second stage sample S_2 is

obtained by selecting a portion of m second stage unit from M second stage unit for every one of the n chosen first stage unit in S_1 utilizing SRSWOR scheme.

At the second stage, it is assumed that the study variable y and the first supplementary variable x_1 have random failure to respond, but the sampled unit has full response for the second supplementary variable x_2. For such random non-response conditions, we consider the following probability model shown in section (2.1.1.1).

2.1.1.1. Probability of Non-Response Model

Since, we assume the occurrence of random non-response conditions on the study variable y and the first supplementary variable x_1 from the second stage sample; therefore, we are going to investigate random non-response conditions from the second stage sample S_2. Let $r\left\{r=0,1,\cdots,(m-2)\right\}$ represents the number of second stage sampling units that did not respond. Accordingly, we write A_r and A_r^c to represent the collection of respondent unit and non-respondent unit, respectively. The observations of the corresponding variables in which random non-response occurs could be obtained from the rest of the $(m-r)$ unit of each of the n first stage unit of the second stage sample (SSS).

We further suppose that if p represents the probability of random failure to respond among $(m-2)$ possible failure to respond cases, and then r follows the probability distribution shown in equation (4):

$$P(r)=\frac{m-r}{mq+2p}\,{}^{(m-2)}C_r\,p^r q^{m-2-r}\ ;\ r=0,1,\cdots(m-2) \tag{4}$$

For example, see the work of [25-27], where $q=1-p$ and ${}^{(m-2)}C_r$ represent the overall possible methods to provide r failure to respond from $(m-2)$ total non-response.

Henceforth, the following notations will be used:

$\bar{Y}_{..}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}Y_{ij}$, Population average of the study variable y .

$\bar{X}_{1..}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}X_{1ij}$, Population average of the first supplementary variable x_1 .

$\bar{X}_{2..}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}X_{2ij}$, Population average of the second supplementary variable x_2 .

$\bar{y}^{*}_{i(e)}=\frac{1}{m}\sum_{j=1}^{m}y_{ij(e)}$, Sample average of the character under study on i^{th} FSU in S_2.

$\bar{y}^{*}_{i(m-r)(e)}=\frac{1}{m-r}\sum_{j=1}^{m-r}y_{ij(e)}$, Sample mean of y based on the respondent region of i^{th} FSU in S_2.

$\bar{x}^{*}_{1i(e)}=\frac{1}{m}\sum_{j=1}^{m}x_{1ij(e)}$, Sample average of the first supplementary variable on i^{th} FSU in S_2.

$\bar{x}^{*}_{1i(m-r)(e)}=\frac{1}{m-r}\sum_{j=1}^{m-r}x_{1ij(e)}$, Sample mean of x_1 based on the respondent region of i^{th} FSU in S_2.

$\bar{x}^{*}_{2i(e)}=\frac{1}{m}\sum_{j=1}^{m}x_{2ij(e)}$, Sample average of x_2 on i^{th} FSU in S_2.

$\bar{y}^{**}_{n(m-r)(e)}=\frac{1}{n}\sum_{i=1}^{n}\bar{y}^{*}_{i(e)}$, Sample average of the n FSU of the character under study.

$\bar{X}^{**}_{1n(m-r)(e)} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}^{*}_{1i(e)}$, Sample average of the n FSU of the first variable.

$\bar{X}^{**}_{2nm(e)} = \frac{1}{n}\sum_{i=1}^{n} \bar{X}^{*}_{2i(e)}$, Sample average of the n FSU of the second supplementary variable.

$C_{X_2} = \frac{S_{X_2}}{\bar{X}_{2..}}$, Coefficient of variation of the second supplementary variable.

S_{X_2} , Standard deviation of the second supplementary variable.

$B_1\left(X_2\right)$, Population coefficient of skewness of the second supplementary variable.

$B_2\left(X_2\right)$, Population coefficient of skewness of the second supplementary variable.

2.1.1.2 Proposed Imputation Schemes and Estimators

We assumed that the second supplementary variable was generally accessible all through the population U . Inspired by the imputation schemes given by [28], we propose the following regression-type imputation schemes based on responding and non-responding units of the second stage sample S_2 to estimate population parameter under study $\bar{Y}_{..}$ as:

$$y_{ij(e)} = \begin{cases} y_{ij(e)}, & if\ j \in A_r \\ \dfrac{\bar{y}_{i(m-r)(e)} + b_{yx_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right), & if\ j \in A_r^c \end{cases} ;\ (i = 1, 2, \ldots, n) \tag{5}$$

where A_{x_2} and B_{x_2} are available functions of supplementary variable like coefficient of skewness, kurtosis, variation, standard deviation, $b_{yx_2(e)} = \sum_{j\in A_r} y_{ij(e)} \Big/ \sum_{j\in A_r} x_{2ij(e)}$, $b_{x_1x_2(e)} = \sum_{j\in A_r} x_{1ij(e)} \Big/ \sum_{j\in A_r} x_{2ij(e)}$.

Remark 1: Note that $A_{x_2} \neq B_{x_2}$ and $A_{x_2} \neq 0$

Under this approach, we derived the sample means of y on the i^{th} first stage units in S_2 denoted by $\bar{y}^{*}_{i(e)}$ as:

$$\bar{y}^{*}_{i(e)} = \frac{1}{m}\sum_{j=1}^{m} y_{ij(e)} = \frac{1}{m}\left[\sum_{j\in A_r} y_{ij(e)} + \sum_{j\in A_r^c} y_{ij(e)}\right] \tag{6}$$

$$\bar{y}^{*}_{i(e)} = \frac{1}{m}\sum_{j=1}^{m} y_{ij(e)} = \left(1 - \frac{r}{m}\right)\bar{y}_{i(m-r)(e)} + \frac{r}{m}\left(\frac{\bar{y}_{i(m-r)(e)} + b_{yx_2(e)}\left(\bar{X}_{2□} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2□} + B_{x_2}\right)\right) \tag{7}$$

In S_2 , the mean of n first stage unit of y is now:

$$\bar{y}^{**}_{n(m-r)(e)} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}^{*}_{i(e)} = \left(1 - \frac{r}{m}\right)\bar{y}_{n(m-r)(e)} + \frac{r}{m}\left(\frac{\bar{y}_{n(m-r)(e)} + b_{yx_2(e)}\left(\bar{X}_{2□} - \bar{x}_{2mn(e)}\right)}{A_{x_2}\bar{x}_{2mn(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2□} + B_{x_2}\right)\right) \tag{8}$$

Likewise, for each unit in the second stage

$$x_{1ij(e)}=\begin{cases} x_{1ij(e)}, & j\in A_r \\ \dfrac{\bar{X}_{1i(m-r)(e)}+b_{x_1x_2(e)}\left(\bar{X}_{2\square}\text{-}\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2\square}+B_{x_2}\right), & j\in A_r^c \quad ;(i=1,\,2,\ldots,n) \end{cases} \tag{9}$$

Under this approach, we derived the sample means of x_{1i} on the i^{th} first stage units in S_2 denoted by $\bar{x}^{*}_{1i(e)}$ as:

$$\bar{x}^{*}_{1i(e)}=\frac{1}{m}\sum_{j=1}^{m}x_{1ij(e)}=\frac{1}{m}\left[\sum_{j\in A_r}x_{1ij(e)}+\sum_{j\in A_r^c}x_{1ij(e)}\right] \tag{10}$$

$$\bar{X}^{*}_{1i(e)}=\frac{1}{m}\sum_{j=1}^{m}x_{1ij(e)}=\left(1-\frac{r}{m}\right)\bar{x}_{1i(m-r)(e)}+\frac{r}{m}\left(\frac{\bar{x}_{1i(m-r)(e)}+b_{x_1x_2(e)}\left(\bar{X}_{2..}\text{-}\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right)\right) \tag{11}$$

In S_2, the mean of n first stage unit of $x_{1(e)}$ is now:

$$\bar{x}^{**}_{1n(m-r)(e)}=\frac{1}{n}\sum_{i=1}^{n}\bar{x}^{*}_{1i(e)}=\left(1-\frac{r}{m}\right)\bar{x}_{1n(m-r)(e)}+\frac{r}{m}\left(\frac{\bar{x}_{1n(m-r)(e)}+b_{x_1x_2(e)}\left(\bar{X}_{2..}\text{-}\bar{x}_{2mn(e)}\right)}{A_{x_2}\bar{x}_{2mn(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right)\right) \tag{12}$$

Hence the proposed estimator denoted by τ^{*} under the above proposed imputation scheme is obtained as:

$$\tau^{*}=\bar{y}^{**}_{n(m-r)(e)}+b^{*}_{(e)}\left(\bar{x}_{1n'M(e)}-\bar{x}^{**}_{1n(m-r)(e)}\right) \tag{13}$$

Where $b^{*}_{(e)}$ is a suitable constant chosen to minimize the mean square error of the proposed estimator τ^{*}.

2.1.1.3 Properties of the Proposed Estimators

Since τ^{*} is regression-type estimator, it is biased for $\bar{Y}_{..}$ the bias and mean square of τ^{*} up to the first order of approximations are derived under large sample approximations (ignoring f.p.c) using the following assumptions:

$$\bar{y}_{n(m-r)(e)}=\bar{Y}_{..}\left(1+\Delta_{0(e)}\right),\ \bar{x}_{1n(m-r)(e)}=\bar{X}_{1..}\left(1+\Delta_{1(e)}\right),\ \bar{x}_{1n'M(e)}=\bar{X}_{1..}\left(1+\Delta_{2(e)}\right),\ \bar{x}_{2nm(e)}=\bar{X}_{2..}\left(1+\Delta_{3(e)}\right)$$

Such that $E\left(\Delta_i\right)=0$ and $\left|\Delta_i\right|<1$ for all $i=0,1,2,3$.

Express $\left(\tau^{*}\right)$ in terms of errors $\Delta_{0(e)}$, $\Delta_{1(e)}$, $\Delta_{2(e)}$ and $\Delta_{3(e)}$.

$$\begin{aligned}\tilde{\tau}_1=\bar{Y}_{\square}\Bigg[1+\Delta_{0(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{3(e)}+\frac{r}{m}\left(\vartheta_X^2+\vartheta_X+1\right)\Delta_{3(e)}^2-\left(\vartheta_X+1\right)\Delta_{0(e)}\Delta_{3(e)}\Bigg]+\\ b^{*}_{(e)}\Bigg[\left(1+\Delta_{2(e)}\right)\bar{X}_{1\square}-\bar{X}_{1\square}\left(1+\Delta_{1(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{3(e)}+\frac{r}{m}\left(\vartheta_X^2+\vartheta_X+1\right)\Delta_{3(e)}^2-\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{1(e)}\Delta_{3(e)}\right)\Bigg]\end{aligned} \tag{14}$$

$$\begin{aligned}\hat{\tau}_1-\bar{Y}_{\square\square}&=\bar{Y}_{\square\square}\left[\Delta_{0(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{3(e)}+\frac{r}{m}\left(\vartheta_X^2+\vartheta_X+1\right)\Delta_{3(e)}^2-\left(\vartheta_X+1\right)\Delta_{0(e)}\Delta_{3(e)}\right]+\\ b_{(e)}^*\bar{X}_{1\square\square}&\left[\Delta_{2(e)}-\Delta_{1(e)}+\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{3(e)}-\frac{r}{m}\left(\vartheta_X^2+\vartheta_X+1\right)\Delta_{3(e)}^2+\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{1(e)}\Delta_{3(e)}\right]\end{aligned}\tag{15}$$

Where, $\vartheta_X=\left(\frac{A_{x_2}\bar{X}_{2..}}{A_{x_2}\bar{X}_{2..}+B_{x_2}}\right)$

We have separately derived the bias and mean square error of the estimator τ^* for Cases A and B of the two-phase sampling structure defined in Section 2, and these are presented below.

Case A: To create a S_1, a subsample of $S_{n'}\left(S_1\subset S_{n'}\right)$ is taken.

To obtain the expressions for the bias and mean square error in this situation, we will consider the following expected values of the sample statistics.

$$E\left(\Delta_{0(e)}^2\right)=\varphi_{n,N}\frac{S_y^{*2}+S_u^{*2}}{\bar{Y}_{\square\square}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_y^2+\bar{S}_u^2}{\bar{Y}_{\square\square}^2},\ E\left(\Delta_{1(e)}^2\right)=\varphi_{n,N}\frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1\square\square}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{x_1}^2+\bar{S}_v^2}{\bar{X}_{1\square\square}^2},$$

$$E\left(\Delta_{2(e)}^2\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1\square\square}^2},\ E\left(\Delta_{3(e)}^2\right)=\varphi_{n,N}\frac{S_{x_2}^{*2}+S_w^{*2}}{\bar{X}_{2\square\square}^2}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_2}^2+\bar{S}_w^2}{\bar{X}_{2\square\square}^2},$$

$$E\left(\Delta_{0(e)}\Delta_{1(e)}\right)=\varphi_{n,N}\frac{S_{yx_1}^*+S_{uv}^*}{\bar{Y}_{\square\square}\bar{X}_{1\square\square}}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{yx_1}+\bar{S}_{uv}}{\bar{Y}_{\square\square}\bar{X}_{1\square\square}},\ E\left(\Delta_{0(e)}\Delta_{2(e)}\right)=\varphi_{n',N}\frac{S_{yx_1}^*+S_{uv}^*}{\bar{Y}_{\square\square}\bar{X}_{1\square\square}},$$

$$E\left(\Delta_{0(e)}\Delta_{3(e)}\right)=\varphi_{n,N}\frac{S_{yx_2}^*+S_{uw}^*}{\bar{Y}_{\square\square}\bar{X}_{2\square\square}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{yx_2}+\bar{S}_{uw}}{\bar{Y}_{\square\square}\bar{X}_{2\square\square}},\ E\left(\Delta_{1(e)}\Delta_{2(e)}\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1\square\square}^2}=E\left(\Delta_{2(e)}^2\right),$$

$$E\left(\Delta_{1(e)}\Delta_{3(e)}\right)=\varphi_{n,N}\frac{S_{x_1x_2}^*+S_{vw}^*}{\bar{X}_{1\square\square}\bar{X}_{2\square\square}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_1x_2}+\bar{S}_{vw}}{\bar{X}_{1\square\square}\bar{X}_{2\square\square}},\ E\left(\Delta_{2(e)}\Delta_{3(e)}\right)=\varphi_{n',N}\frac{S_{x_1x_2}^*+S_{vw}^*}{\bar{X}_{1\square\square}\bar{X}_{2\square\square}}$$

For simplicity we let

$$\left.\begin{aligned}&\zeta_{0(e)}=E\left(\Delta_{0(e)}^2\right),\zeta_{1(e)}=E\left(\Delta_{1(e)}^2\right),\zeta_{2(e)}=E\left(\Delta_{2(e)}^2\right)=E\left(\Delta_{1(e)}\Delta_{2(e)}\right),\\&\zeta_{3(e)}=E\left(\Delta_{3(e)}^2\right),\zeta_{4(e)}=E\left(\Delta_{0(e)}\Delta_{1(e)}\right),\zeta_{5(e)}=E\left(\Delta_{0(e)}\Delta_{2(e)}\right),\\&\zeta_{6(e)}=E\left(\Delta_{0(e)}\Delta_{3(e)}\right),\zeta_{7(e)}=E\left(\Delta_{1(e)}\Delta_{3(e)}\right),\zeta_{8(e)}=E\left(\Delta_{2(e)}\Delta_{3(e)}\right)\end{aligned}\right\}\tag{16}$$

The following notations and expectation will be use under this strategy, when measurement error is not taking into account.

$$E\left(\Delta_0^2\right)=\varphi_{n,N}\frac{S_y^{*2}}{\bar{Y}_{\square\square}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_y^2}{\bar{Y}_{\square\square}^2},\ E\left(\Delta_1^2\right)=\varphi_{n,N}\frac{S_{x_1}^{*2}}{\bar{X}_{1\square\square}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{x_1}^2}{\bar{X}_{1\square\square}^2},\ E\left(\Delta_2^2\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}}{\bar{X}_{1\square\square}^2},$$

$$E\left(\Delta_3^2\right)=\varphi_{n,N}\frac{S_{x_2}^{*2}}{\bar{X}_{2\square\square}^2}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_2}^2}{\bar{X}_{2\square\square}^2},\ E\left(\Delta_0\Delta_1\right)=\varphi_{n,N}\frac{S_{yx_1}^*}{\bar{Y}_{\square\square}\bar{X}_{1\square\square}}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{yx_1}}{\bar{Y}_{\square\square}\bar{X}_{1\square\square}},\ E\left(\Delta_0\Delta_2\right)=\varphi_{n',N}\frac{S_{yx_1}^*}{\bar{Y}_{\square\square}\bar{X}_{1\square\square}},$$

$$E\left(\Delta_0\Delta_3\right)=\varphi_{n,N}\frac{S_{yx_2}^*}{\bar{Y}_{\square\square}\bar{X}_{2\square\square}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{yx_2}}{\bar{Y}_{\square\square}\bar{X}_{2\square\square}},\ E\left(\Delta_1\Delta_2\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}}{\bar{X}_{1\square\square}^2}=E\left(\Delta_2^2\right),$$

$$E\left(\Delta_1\Delta_3\right)=\varphi_{n,N}\frac{S_{x_1x_2}^*}{\bar{X}_{1\square\square}\bar{X}_{2\square\square}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_1x_2}}{\bar{X}_{1\square\square}\bar{X}_{2\square\square}},\ E\left(\Delta_2\Delta_3\right)=\varphi_{n',N}\frac{S_{x_1x_2}^*}{\bar{X}_{1\square\square}\bar{X}_{2\square\square}}$$

Similarly, for simplicity we let

$$\left.\begin{aligned}&\zeta_0=E\left(\Delta_0^2\right),\ \zeta_1=E\left(\Delta_1^2\right),\zeta_2=E\left(\Delta_2^2\right)=E\left(\Delta_1\Delta_2\right),\zeta_3=E\left(\Delta_3^2\right),\ \zeta_4=E\left(\Delta_0\Delta_1\right),\\&\zeta_5=E\left(\Delta_0\Delta_2\right),\zeta_6=E\left(\Delta_0\Delta_3\right),\ \ \zeta_7=E\left(\Delta_1\Delta_3\right),\ \zeta_8=E\left(\Delta_2\Delta_3\right)\end{aligned}\right\}\tag{17}$$

Where

$$\bar{Y}_{\square\square}=\frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{i\square},\ \bar{Y}_{i\square}=\frac{1}{M}\sum_{j=1}^{M}y_{ij},\ \bar{X}_{1\square\square}=\frac{1}{N}\sum_{i=1}^{N}\bar{X}_{1i\square},\ \bar{X}_{1i\square}=\frac{1}{M}\sum_{j=1}^{M}x_{1ij},\ \bar{X}_{2\square\square}=\frac{1}{N}\sum_{i=1}^{N}\bar{X}_{2i\square},\ \bar{X}_{2i\square}=\frac{1}{M}\sum_{j=1}^{M}x_{2ij},$$

$$S_y^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i\square}-\bar{Y}_{\square\square}\right)^2,\ \bar{S}_y^2=\frac{1}{N}\sum_{i=1}^{N}S_{y_i}^2,\ S_{y_{i.}}^2=\frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij}-\bar{Y}_{i\square}\right)^2$$

$$S_{x_1}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{1i\square}-\bar{X}_{1\square\square}\right)^2,\ \bar{S}_{x_1}^2=\frac{1}{N}\sum_{i=1}^{N}S_{x_{1i}}^2,\ S_{x_{1i}}^2=\frac{1}{M-1}\sum_{j=1}^{M}\left(x_{1ij}-\bar{X}_{1i\square}\right)^2$$

$$S_{x_2}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{2i\square}-\bar{X}_{2\square\square}\right)^2,\ \bar{S}_{x_2}^2=\frac{1}{N}\sum_{i=1}^{N}S_{x_{2i}}^2,\ S_{x_{2i}}^2=\frac{1}{M-1}\sum_{j=1}^{M}\left(x_{2ij}-\bar{X}_{2i\square}\right)^2$$

$$S_{yx_1}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i\square}-\bar{Y}_{\square\square}\right)\left(\bar{X}_{1i\square}-\bar{X}_{1\square\square}\right),\ \bar{S}_{yx_1}=\frac{1}{N}\sum_{i=1}^{N}S_{yx_{1i}},\ S_{yx_{1i}}=\frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij}-\bar{Y}_{i\square}\right)\left(x_{1ij}-\bar{X}_{1i\square}\right)$$

$$S_{yx_2}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i\square}-\bar{Y}_{\square\square}\right)\left(\bar{X}_{2i\square}-\bar{X}_{2\square\square}\right),\ \bar{S}_{yx_2}=\frac{1}{N}\sum_{i=1}^{N}S_{yx_{2i}},\ S_{yx_{2i}}=\frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij}-\bar{Y}_{i\square}\right)\left(x_{2ij}-\bar{X}_{2i\square}\right)$$

$$S_{x_1x_2}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{1i\square}-\bar{X}_{1\square\square}\right)\left(\bar{X}_{2i\square}-\bar{X}_{2\square\square}\right),\ \bar{S}_{x_1x_2}=\frac{1}{N}\sum_{i=1}^{N}S_{x_1x_{2i}},\ S_{x_1x_{2i}}=\frac{1}{M-1}\sum_{j=1}^{M}\left(x_{1ij}-\bar{X}_{1i\square}\right)\left(x_{2ij}-\bar{X}_{2i\square}\right)$$

$$\bar{U}_{\square\square}=\frac{1}{N}\sum_{i=1}^{N}\bar{U}_{i\square},\bar{U}_{i\square}=\frac{1}{M}\sum_{j=1}^{M}u_{ij},\bar{V}_{\square\square}=\frac{1}{N}\sum_{i=1}^{N}\bar{V}_{i.},\bar{V}_{i\square}=\frac{1}{M}\sum_{j=1}^{M}v_{ij},\bar{W}_{\square\square}=\frac{1}{N}\sum_{i=1}^{N}\bar{W}_{i.},\bar{W}_{i\square}=\frac{1}{M}\sum_{j=1}^{M}w_{ij}$$

$$S_u^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{U}_{i\square}-\bar{U}_{\square\square}\right)^2,\ \bar{S}_u^2=\frac{1}{N}\sum_{i=1}^{N}S_{u_i}^2,\ S_{u_{i.}}^2=\frac{1}{M-1}\sum_{j=1}^{M}\left(u_{ij}-\bar{U}_{i\square}\right)^2$$

$$S_v^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{V}_{i\square}-\bar{V}_{\square\square}\right)^2,\ \bar{S}_v^2=\frac{1}{N}\sum_{i=1}^{N}S_{v_i}^2,\ S_{v_i}^2=\frac{1}{M-1}\sum_{j=1}^{M}\left(v_{ij}-\bar{V}_{i\square}\right)^2$$

$$S_w^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{W}_{i\square}-\bar{W}_{\square\square}\right)^2,\ \bar{S}_w^2=\frac{1}{N}\sum_{i=1}^{N}S_{w_i}^2,\ S_{w_i}^2=\frac{1}{M-1}\sum_{j=1}^{M}\left(w_{ij}-\bar{W}_{i\square}\right)^2$$

$$S_{uv}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{U}_{I\square}-\bar{U}_{\square\square}\right)\left(\bar{V}_{i\square}-\bar{V}_{\square\square}\right),\ \bar{S}_{uv}=\frac{1}{N}\sum_{i=1}^{N}S_{uv_i},\ S_{uv_i}=\frac{1}{M-1}\sum_{j=1}^{M}\left(u_{ij}-\bar{U}_{i\square}\right)\left(v_{ij}-\bar{V}_{I\square}\right)$$

$$S_{uw}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{U}_{i\square}-\bar{U}_{\square\square}\right)\left(\bar{W}_{i\square}-\bar{W}_{\square\square}\right),\ \bar{S}_{uw}=\frac{1}{N}\sum_{i=1}^{N}S_{uw_i},\ S_{uw_i}=\frac{1}{M-1}\sum_{j=1}^{M}\left(u_{ij}-\bar{U}_{i\square}\right)\left(w_{ij}-\bar{W}_{i\square}\right)$$

$$S_{vw}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{V}_{i\square}-\bar{V}_{\square\square}\right)\left(\bar{W}_{i\square}-\bar{W}_{\square\square}\right),\ \bar{S}_{vw}=\frac{1}{N}\sum_{i=1}^{N}S_{vw_i},\ S_{vw_i}=\frac{1}{M-1}\sum_{j=1}^{M}\left(v_{ij}-\bar{V}_{i\square}\right)\left(w_{ij}-\bar{W}_{i\square}\right)$$

$$S_b^2=\frac{1}{N}\sum_{i=1}^{N}\left(\bar{y}_i-\bar{\bar{Y}}\right)^2,\ S_w^2=\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{M-1}\sum_{j=i}^{M}\left(\bar{y}_{ij}-\bar{y}_i\right)^2\right\},\ S_{b(U)}^2=\frac{1}{N}\sum_{i=1}^{N}\left(\bar{U}_i-\bar{\bar{U}}\right)^2,\ S_{w(U)}^2=\frac{1}{N}\left\{\sum_{i=1}^{N}\frac{1}{M-1}\sum_{j=i}^{M}\left(U_{ij}-\bar{U}_i\right)^2\right\}$$

$$\varphi_{n,N}=\left(\frac{1}{n}-\frac{1}{N}\right),\ \varphi_{n',N}=\left(\frac{1}{n'}-\frac{1}{N}\right),\varphi_{m,M}=\left(\frac{1}{m}-\frac{1}{M}\right),\varphi_{m,r}=\left(\frac{1}{mq+2p}-\frac{1}{M}\right)$$

Taking expectation on both sides of (15) and applying the results of (16) we obtain the bias of τ^* as:

$$B\left(\hat{\tau}_1\right)_I=\frac{r}{m}\left[\left(\vartheta_X^2+\vartheta_X+1\right)\left(\bar{Y}_{\square\square}-b_{(e)}^*\bar{X}_{1\square\square}\right)\zeta_{3(e)}-\left(\vartheta_X+1\right)\left(\bar{Y}_{\square\square}-b_{(e)}^*\bar{X}_{1\square\square}\right)\left(\zeta_{6(e)}-\zeta_{7(e)}\right)\right] \tag{18}$$

The mean square error (MSE) of τ^* is obtained by taking expectation and square on both sides of (15) and applying the results of (16)

$$\begin{aligned}MSE\left(\tau^*\right)_I=\bar{Y}_{\cdot\cdot}^2&\left[\zeta_{0(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{6(e)}\right]-\\&2b_{(e)}^*\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}+\zeta_{4(e)}-\zeta_{5(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}-\zeta_{8(e)}\right)\right]+\\&b_{(e)}^{*2}\bar{X}_{1\cdot\cdot}^2\left[\zeta_{1(e)}-\zeta_{2(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{7(e)}-\zeta_{8(e)}\right)\right]\end{aligned} \tag{19}$$

To obtain expression for that minimize $MSE\left(\tau^*\right)_I$, differentiate (19) partially with respect to $b_{(e)}^*$ and equate the result to zero.

$$b_{(e)opt}^*=\frac{\bar{Y}_{\square\square}\left(\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}+\zeta_{4(e)}-\zeta_{5(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}-\zeta_{8(e)}\right)\right)}{\bar{X}_{1\square\square}\left(\zeta_{1(e)}-\zeta_{2(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{7(e)}-\zeta_{8(e)}\right)\right)} \tag{20}$$

Substituting the value of $b_{opt(e)}^*$ in (19), gives the minimum value of $MSE\left(\tau^*\right)_I$ as:

$$\begin{aligned}MSE_{\min}\left(\tau^*\right)_I=\bar{Y}_{\cdot\cdot}^2&\left[\zeta_{0(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{6(e)}\right]-\\&2b_{opt(e)}^*\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}+\zeta_{4(e)}-\zeta_{5(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}-\zeta_{8(e)}\right)\right]+\\&b_{opt(e)}^{*2}\bar{X}_{1\cdot\cdot}^2\left[\zeta_{1(e)}-\zeta_{2(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{7(e)}-\zeta_{8(e)}\right)\right]\end{aligned} \tag{21}$$

The mean square error without measurement error is given by:

$$\begin{aligned}MSE_{\min}\left(\tau\right)_I=\bar{Y}_{\cdot\cdot}^2&\left[\zeta_0+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3-2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_6\right]-\\&2b_{opt}^*\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3+\zeta_4-\zeta_5-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_6+\zeta_7-\zeta_8\right)\right]+\\&b_{opt}^{*2}\bar{X}_{1\cdot\cdot}^2\left[\zeta_1-\zeta_2+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_7-\zeta_8\right)\right]\end{aligned} \tag{22}$$

where,

$$b_{opt}^*=\frac{\bar{Y}_{\square\square}\left(\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3+\zeta_4-\zeta_5-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_6+\zeta_7-\zeta_8\right)\right)}{\bar{X}_{1\square\square}\left(\zeta_1-\zeta_2+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_7-\zeta_8\right)\right)} \tag{23}$$

Case B: In this case, S_1 is drawn independently of $S_{n'}$.

$E\left(\Delta_{0(e)}\Delta_{2(e)}\right)=E\left(\Delta_{1(e)}\Delta_{2(e)}\right)=E\left(\Delta_{2(e)}\Delta_{3(e)}\right)=0$, and other expectation are the same as stated in Case A.

Following the procedure used in Case A, we have obtained the minimum mean square error of τ^* as:

$$MSE_{\min}\left(\tau^{*}\right)_{II}=\bar{Y}_{..}^{2}\left[\zeta_{0(e)}+\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_{X}+1\right)\zeta_{6(e)}\right]-2b_{opt(e)}^{*}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3(e)}+\zeta_{4(e)}-\frac{r}{m}\left(\vartheta_{X}+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}\right)\right]+b_{opt(e)}^{*2}\bar{X}_{1..}^{2}\left[\zeta_{1(e)}+\zeta_{2(e)}+\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_{X}+1\right)\zeta_{7(e)}\right] \tag{24}$$

Where,

$$b_{opt(e)}^{*}=\frac{\bar{Y}_{..}\left(\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3(e)}+\zeta_{4(e)}-\frac{r}{m}\left(\vartheta_{X}+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}\right)\right)}{\bar{X}_{1..}\left(\zeta_{1(e)}+\zeta_{2(e)}+\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_{X}+1\right)\zeta_{7(e)}\right)} \tag{25}$$

The mean square error without measurement error is given by:

$$MSE_{\min}\left(\tau\right)_{II}=\bar{Y}_{..}^{2}\left[\zeta_{0}+\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3}-2\frac{r}{m}\left(\vartheta_{X}+1\right)\zeta_{6}\right]-2b_{opt}^{*}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3}+\zeta_{4}-\frac{r}{m}\left(\vartheta_{X}+1\right)\left(\zeta_{6}+\zeta_{7}\right)\right]+b_{opt}^{*2}\bar{X}_{1..}^{2}\left[\zeta_{1}+\zeta_{2}+\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3}-2\frac{r}{m}\left(\vartheta_{X}+1\right)\zeta_{7}\right] \tag{26}$$

Where,

$$b_{opt}^{*}=\frac{\bar{Y}_{..}\left(\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3}+\zeta_{4}-\frac{r}{m}\left(\vartheta_{X}+1\right)\left(\zeta_{6}+\zeta_{7}\right)\right)}{\bar{X}_{1..}\left(\zeta_{1}+\zeta_{2}+\frac{r^{2}}{m^{2}}\left(\vartheta_{X}+1\right)^{2}\zeta_{3}-2\frac{r}{m}\left(\vartheta_{X}+1\right)\zeta_{7}\right)} \tag{27}$$

2.1.1.4 Efficiency comparison

To evaluate the efficiency of the proposed estimators, we compare them with the usual mean per unit estimator without supplementary information and with the [29] estimators of the population mean in a two-stage cluster sampling scheme, using the strategy discussed in Section 2.1.1.

The mean per unit estimator τ_{0}^{*} and its variance in the presence of measurement error are given by:

$$\tau_{0}^{*}=\bar{y}_{nm(e)} \tag{28}$$

$$V\left(\tau_{0}^{*}\right)=\varphi_{n,N}\left(S_{b}^{2}+S_{b(U)}^{2}\right)+\frac{1}{n}\varphi_{m,M}\left(S_{w}^{2}+S_{w(U)}^{2}\right) \tag{29}$$

The mean per unit estimator τ_{0} and its variance in the absence of measurement error are given by:

$$\tau_{0}=\bar{y}_{nm} \tag{30}$$

$$V\left(\tau_{0}\right)=\varphi_{n,N}S_{b}^{2}+\frac{1}{n}\varphi_{m,M}S_{w}^{2} \tag{31}$$

The following estimators of population and their mean square error under case A and case B in the absence of measurement error were proposed by [29].

$$\tau_{MSB}=\bar{y}_{n(m-r)(e)}^{*}+B\left(\bar{x}_{1n'M(e)}-\bar{x}_{1n(m-r)(e)}^{*}\right) \tag{32}$$

The $MSE_{\min}\left(\tau_{MSB}\right)$, for both Case A and Case B are given by:

Case A

$$MSE_{\min}\left(\tau_{MSB}\right)_I = \bar{Y}_{..}^2\left(\zeta_0+\frac{1}{4}\zeta_3-\zeta_6\right)-2B_{opt}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6+\frac{1}{2}\zeta_8-\frac{1}{2}\zeta_7-\zeta_5+\zeta_4\right)+ B_{opt}^2\bar{X}_{1..}^2\left(\frac{1}{4}\zeta_3+\zeta_1-\zeta_2+\zeta_8-\zeta_7\right) \tag{33}$$

Where,

$$B_{opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6+\frac{1}{2}\zeta_8-\frac{1}{2}\zeta_7-\zeta_5+\zeta_4\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_3+\zeta_1-\zeta_2+\zeta_8-\zeta_7\right)} \tag{34}$$

Case B

$$MSE_{\min}\left(\tau_{MSB}\right)_{II} = \bar{Y}_{..}^2\left(\zeta_0+\frac{1}{4}\zeta_3-\zeta_6\right)-2B_{1opt}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6-\frac{1}{2}\zeta_7+\zeta_4\right)+ B_{1opt}^2\bar{X}_1^2\left(\frac{1}{4}\zeta_3+\zeta_1+\zeta_2-\zeta_7\right) \tag{35}$$

Where,

$$B_{opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6-\frac{1}{2}\zeta_7+\zeta_4\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_3+\zeta_1+\zeta_2-\zeta_7\right)} \tag{36}$$

To compare it with our proposed estimator in the presence of measurement, we include the contribution of measurement error parameters in the [29] estimators.

The mean square error, accounting for measurement error, for both Case A and Case B, is as follows:

Case A

$$MSE_{\min}\left(\tau_{MSB}^*\right)_I = \bar{Y}_{..}^2\left(\zeta_{0(e)}+\frac{1}{4}\zeta_{3(e)}-\zeta_{6(e)}\right)-2B_{1opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}+\frac{1}{2}\zeta_{8(e)}-\frac{1}{2}\zeta_{7(e)}-\zeta_{5(e)}+\zeta_{4(e)}\right)+ B_{opt(e)}^2\bar{X}_{1..}^2\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}+\zeta_{8(e)}-\zeta_{7(e)}\right) \tag{37}$$

where,

$$B_{(e)opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}+\frac{1}{2}\zeta_{8(e)}-\frac{1}{2}\zeta_{7(e)}-\zeta_{5(e)}+\zeta_{4(e)}\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}+\zeta_{8(e)}-\zeta_{7(e)}\right)} \tag{38}$$

Case B

$$MSE_{\min}\left(\tau_{MSB}^*\right)_{II} = \bar{Y}_{..}^2\left(\zeta_{0(e)}+\frac{1}{4}\zeta_{3(e)}-\zeta_{6(e)}\right)-2B_{opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}-\frac{1}{2}\zeta_{7(e)}+\zeta_{4(e)}\right)+ B_{opt(e)}^2\bar{X}_1^2\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}-\zeta_{7(e)}\right) \tag{39}$$

Where,

$$B_{(e)opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}-\frac{1}{2}\zeta_{7(e)}+\zeta_{4(e)}\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}-\zeta_{7(e)}\right)} \tag{40}$$

To demonstrate the performance of our suggested estimators, we compared their percentage relative efficiency (PRE) to the traditional mean per unit estimator, which is based on the normal two-stage design technique without supplementary information, as well as [29] estimators. The empirical study was carried out employing simulated population data sets.

The PRE of an estimator τ^* relative to the natural mean per unit estimator τ_0^* is defined as:

$$PRE = \frac{\left(V\left(\tau_0^*\right)\right)}{MSE_{\min}\left(\tau^*\right)} \times 100 \tag{41}$$

III. Results

3.1. Study Using Artificially Generated Population

An important aspect of simulation is that one builds a simulation model to replicate the actual system. Simulation allows comparison of analytical techniques and helps in concluding whether a newly developed technique is better than the existing ones. Motivated by [6], [30], and [29], who used artificial population generation techniques.

3.1.1. Simulation Results

This simulation exercise consists of the following steps:

1. Six independent variables (normally distributed) are simulated (a total of N times M)
2. The simulated data is then split into M distinct clusters each of size N. The variables Y, X_1 and X_2 are constructed following the relationship defined in the work of [29] only that here, the error component is added.
3. A random sample of m (or m' then m) clusters is selected out of the M total clusters. This is called the first sample units (fsu).
4. A random sample of n (or n' then n) units are sampled from each of m selected clusters. This is called the second sample units (ssu).
5. All the different estimators of the Mean Square Error are calculated based on the observed data and compared.
6. Steps 3 to 5 are repeated a hundred times for each specific case and the estimates of Mean Square Error are all saved in arrays after which the means are calculated and compared.

3.2. Numerical Illustration using Artificial Population

Population 1

$Y = \mu_y + \sigma_y\left(\rho_{x_1y} \times pop\left[,2\right] + \sqrt{1-\rho_{x_1y}^2} \times pop\left[,1\right]\right) + U$, $X_1 = \mu_{x_1} + \sigma_{x_1} \times pop\left[,2\right] + V$

$X_2 = \mu_{x_2} + \sigma_{x_2}\left(\rho_{x_1x_2} \times pop\left[,2\right] + \sqrt{1-\rho_{x_1x_2}^2}\, pop\left[,1\right]\right) + W$,

$U \sim N\left(0,3\right), V \sim N\left(0,8\right)$, $W \sim N\left(0,12\right)$, $\rho_{x_1y} = 0.7$, $\rho_{x_1x_2} = 0.6$, $\sigma_y^2 = 5$, $\sigma_{x_1}^2 = 12$, $\sigma_{x_2}^2 = 9$, $\mu_y = 20$, $\mu_{x_1} = 50$, $\mu_{x_2} = 40$, $N = 10$, $M = 10$, $n' = 9$, $n = 5$, $m = 7$.

Table 1: *Percentage Relative Efficiency (PRE) of Estimators in the Presence of Measurement Error, under Case A.*

Estimators	Auxiliary parameter	P=0.05 q=0.95	P=0.1 q=0.9	P=0.15 q=0.85	P=0.2 q=0.8
τ_0^*	Not applicable	100.00	100.00	100.00	100.00
τ_{MSB}^*	Not applicable	199.88	192.13	184.52	177.02
τ_1^*	$A_{x_2}=1, B_{x_2}=0$	201.78	194.51	186.67	178.33
τ_2^*	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	201.78	194.51	186.66	178.32
τ_3^*	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	201.77	194.5	186.67	178.34
τ_4^*	$A_{x_2}=1, B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33
τ_5^*	$A_{x_2}=1, B_{x_2}=S_{X_2}$	201.76	194.5	186.69	178.41
τ_6^*	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	201.79	194.52	186.68	178.33
τ_7^*	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33
τ_8^*	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	201.68	194.09	186.05	177.58
τ_9^*	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	201.78	194.51	186.66	178.32
τ_{10}^*	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33
τ_{11}^*	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	201.77	194.51	186.69	178.38
τ_{12}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	201.77	194.48	186.62	178.24
τ_{13}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	201.70	194.44	186.72	178.59
τ_{14}^*	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	201.60	194.40	186.88	179.10
τ_{15}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	201.78	194.51	186.67	178.33
τ_{16}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	201.78	194.51	186.67	178.33
τ_{17}^*	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33

Table 2: *Percentage Relative Efficiency (PRE) of Estimators in the Presence of Measurement Error, under Case B.*

Estimators	Auxiliary parameter	P=0.05 q=0.95	P=0.1 q=0.9	P=0.15 q=0.85	P=0.2 q=0.8
τ_0^*	Not applicable	100.00	100.00	100.00	100.00
τ_{MSB}^*	Not applicable	214.13	206.07	198.13	190.28
τ_1^*	$A_{x_2}=1, B_{x_2}=0$	217.60	209.49	200.87	191.83
τ_2^*	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	217.60	209.49	200.87	191.83
τ_3^*	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	217.60	209.49	200.88	191.85
τ_4^*	$A_{x_2}=1, B_{x_2}=C_{X_2}$	217.60	209.49	200.87	191.83
τ_5^*	$A_{x_2}=1, B_{x_2}=S_{X_2}$	217.60	209.50	200.92	191.94
τ_6^*	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	217.63	209.53	200.91	191.85
τ_7^*	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	217.60	209.50	200.87	191.83
τ_8^*	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	217.64	209.42	200.49	191.14
τ_9^*	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	217.60	209.49	200.87	191.83
τ_{10}^*	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	217.60	209.49	200.87	191.83
τ_{11}^*	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	217.60	209.50	200.90	191.89
τ_{12}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	217.61	209.50	200.85	191.78
τ_{13}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	217.56	209.50	201.03	192.21
τ_{14}^*	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	217.51	209.56	201.33	192.87
τ_{15}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	217.60	209.49	200.87	191.83
τ_{16}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	217.60	209.49	200.87	191.83
τ_{17}^*	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	217.60	209.49	200.87	191.83

Table 3: *Percentage Relative Efficiency (PRE) of Estimators in the Absence of Measurement Error, under Case A.*

Estimators	Auxiliary parameter	P=0.05 q=0.95	P=0.1 q=0.9	P=0.15 q=0.85	P=0.2 q=0.8
τ_0	Not applicable	100.00	100.00	100	100
τ_{MSB}	Not applicable	226.48	217.5	208.73	200.14
τ_1	$A_{x_2}=1, B_{x_2}=0$	229.00	220.21	211.02	201.47
τ_2	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	229.00	220.21	211.02	201.47
τ_3	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	229.00	220.21	211.03	201.49
τ_4	$A_{x_2}=1, B_{x_2}=C_{X_2}$	229.00	220.21	211.02	201.47
τ_5	$A_{x_2}=1, B_{x_2}=S_{X_2}$	228.99	220.22	211.06	201.56
τ_6	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	229.01	220.23	211.04	201.48
τ_7	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	229.00	220.22	211.02	201.47
τ_8	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	228.90	219.73	210.29	200.6
τ_9	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	229.00	220.21	211.02	201.47
τ_{10}	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	229.00	220.21	211.02	201.47
τ_{11}	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	229.00	220.22	211.05	201.53
τ_{12}	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	228.98	220.18	210.96	201.39
τ_{13}	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	228.96	220.20	211.13	201.76
τ_{14}	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	228.93	220.27	211.41	202.38
τ_{15}	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	229.00	220.21	211.02	201.47
τ_{16}	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	229.00	220.21	211.02	201.47
τ_{17}	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	229.00	220.21	211.02	201.47

Table 4: *Percentage Relative Efficiency (PRE) of Estimators in the Absence of Measurement Error, under Case B.*

Estimators	Auxiliary parameter	*P*=0.05 *q*=0.95	*P*=0.1 *q=0.9*	*P*=0.15 *q*=0.85	*P*=0.2 *q*=0.8
τ_0	Not applicable	100.00	100.00	100.00	100.00
τ_{MSB}	Not applicable	242.19	232.93	223.84	214.91
τ_1	$A_{x_2}=1, B_{x_2}=0$	246.09	236.61	226.72	216.51
τ_2	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	246.09	236.61	226.72	216.51
τ_3	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	246.08	236.62	226.74	216.54
τ_4	$A_{x_2}=1, B_{x_2}=C_{X_2}$	246.09	236.61	226.72	216.51
τ_5	$A_{x_2}=1, B_{x_2}=S_{X_2}$	246.08	236.63	226.78	216.63
τ_6	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	246.11	236.65	226.77	216.54
τ_7	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	246.09	236.61	226.73	216.51
τ_8	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	246.77	237.29	227.05	216.41
τ_9	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	246.09	236.61	226.72	216.51
τ_{10}	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	246.09	236.61	226.72	216.51
τ_{11}	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	246.09	236.63	226.76	216.57
τ_{12}	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	246.10	236.62	226.71	216.47
τ_{13}	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	246.07	236.66	226.93	216.94
τ_{14}	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	246.07	236.80	227.32	217.68
τ_{15}	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	246.09	236.61	226.72	216.51
τ_{16}	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	246.09	236.61	226.73	216.51
τ_{17}	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	246.09	236.61	226.72	216.51

IV. Discussion

From table 1, table 2, table 3 and table 4, it can be observed that our proposed estimator, which utilizes the second supplementary variable parameter, is more efficient, with higher percentage relative efficiencies (PREs) than the usual mean per unit estimator without supplementary information and the [29] estimator in both cases scenario for all choices of probabilities. Therefore, it can be concluded that our proposed methodology can be practically applied, utilizing the actual responses of the respondents and including the measurement error parameters in estimating the finite population mean.

References

[1] Sandarl, C. E., Swensson, B., and Wretman, J. (2003). Model assisted survey sampling. Springer Science & Business Media.

[2] Hansen, M. H. And Hurwitz, W. N. (1946).The problem of non-response in sample surveys.Journal of the American Statistical Association, 41(236):517-529.

[3] Rubin, D. B. (1976). Inference and missing data. Biometrika, 63 (3):581–92.

[4] Heitjan, D. F., and Basu, S. (1996). Distinguishing 'missing at random' and 'missing completely at random. American Statistician, 50 (3):207–13.

[5] Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. Metrika 51(3):267-276.

[6] Singh, S and Deo, B. (2003).Imputation by power transformation. Statistical Papers, 44(4):555-579.

[7] Kadilar, C. and Cingi, H. (2008). Estimators for the population mean in the case of missing data, Communications in Statistics- Theory and Methods, 37, 2226-2236.

[8] Singh, H. P and Karpe, N. (2009). On the estimation of ratio and product of two population means using supplementary information in presence. Statistica, 69(1): 27-47.

[9] Singh, A. K., Singh, P. and Singh, V. K. (2014a). Exponential-Type Compromised Imputation in Survey Sampling, Journal of Statistics & probability, 3(2):211-217.

[10] Gira, A. A. (2015). Estimation of population mean with a New Imputation Methods. Applied Mathematical Sciences, 9(34), 1663-1672.

[11] Singh, G. N., Maurya, S. Khetan, M. and Kadilar, C. (2016). Some imputation methods for missing data in sample surveys.Hacettepe Journal of Mathematics and Statistics, 45(6):1865-1880.

[12] Bhushan, S. and Pandey, A. P. (2016). Optimal imputation of missing data for estimation of population mean. Journal of Statistics and Management Systems, 19 (6), 755-769.

[13] Pandey, A. K., Singh, G. N., Bhattacharyya, D., Ali, A. Q., Al-Thubaiti, S., and Yakout, H. (2021). Some classes of logarithmic-type imputation techniques for handling missing data. Computational Intelligence and Neuroscience, pages11.

[14] Singh, G., Jaiswal, A. K., Singh, C., and Usman, M. (2022). An improved alternative method of imputation for missing data is survey sampling. Hacettepe Journal of Statistics Applications and Probability, 11(2):535-543.

[15] Cochran, W. G. (1968). Errors of Measurement in Statistics. Technometrics, 10(4):637-666.

[16] Fuller, W. A. (1995). Estimation in the presence of measurement error. International Statistical Review/Revue Internationale de Statistique, 121-141.

[17] Shalabh, S.(1997). Ratio method of estimation in the presence of measurement errors.Journal of the Indian Society of Agricultural Statistics, 52:150-155.

[18] Manisha and Singh, K. R. (2001). An estimation of population mean in the presence of measurement errors. Journal of the Indian Society of Agricultural Statistics, 54(1):13-18.

[19] Allen, J., Singh, H. P., and Smarandache, F. (2003). A family of estimators of population mean using multiauxiliary information in presence of measurement errors. International Journal of Social Economics, 30(7):837-848.

[20] Singh, S. (2009). A new method of imputation in survey sampling. Statistics, 43(5):499-511.

[21] Salas, C. and Gregoire, T. G. (2010).Statistical analysis of ratio estimators and their estimators of variances when the auxiliary variate is measured with error. European Journal of Forest Research, 129:847-861.

[22] Kumar, M., Rajesh, S., Singh, A. K., and Smarandache, F. (2011).Some ratio type estimators under measurement errors.WorldApplied Sciences Journal, 14(2):272.

[23] Shukla, D., Pathak, S., and Thakur, N. S. (2012).An estimator for mean estimation in presence of measurement error.A Journal of Statistics, 1(1):1-8.

[24] Singh, R., Malik, S., and Kohshnivesan, M. (2014b). An alternative estimator for estimating the finite population mean in presence of measurement errors with the view to financial modelling. Science Journal of Applied Mathematics and Statistics, 2:107-111.

[25] Singh, S., and A. H. Joarder. (1998). Estimation of finite population variance using random non-response in survey sampling. Metrika, 47 (1):241–9.

[26] Singh, S., A. H. Joarder, and D. S. Tracy. (2000). Regression type estimators of random non-response in survey sampling. Statistica, 60(1):39-44.

[27] Singh, H. P., R. Tailor, J. M. Kim, and S. Singh. (2012). Families of estimators of finite population variance using a random non-response in survey sampling. The Korean Journal of Applied Statistics, 25 (4):681–95.

[28] Audu, A., Ishaq, O. O., Abubakar, A., Akintola, K. A., Isah, U., Rashida, A., and Muhammed, S. (2021). Regression-type Imputation Class of Estimators using Auxiliary Attributes. Asian Research Journal of Mathematics, 17(5), 1-13.

[29] Maji, R., Singh, G. N. and Bandyopadhyay, A. (2018). Estimation of population mean in presence of random non response in two-stage cluster sampling, Communications in Statistics - Theory and Methods, 48(14):3586-3608.

[30] Singh, S., A. H. Joarder, and D. S. Tracy. 2001. Median estimation using double sampling. Australian & New Zealand Journal of Statistics 43 (1):33–46.

RT&A, No 4(80)
Volume 19, December, 2024

M/M(A,B)/1 MULTIPLE WORKING VACATIONS QUEUING SYSTEM WITH HETEROGENOUS ENCOURAGED ARRIVAL

PRAKATI.P*, JULIA ROSE MARY.K

•

Department of Mathematics, Nirmala College for Women, India
*prakatidhanam@gmail.com
juliakulandaisamy@gmail.com

Abstract

The concept of Queuing system is most commonly used in our everyday life. It is essential to characterize the practical queuing characteristics in order to improve the performance of the queuing model. This study investigates M/M(a,b)/1/MWV queuing model with heterogeneous encouraged arrival occurring in the regular busy period. The considered model follows General bulk service rule and if the system is not in use, or when it is vacant, the server goes on vacation, thus there occurs multiple working vacations which are exponentially distributed. In this study, a model of multiple working vacation queues in which with heterogeneous encouraged arrivals following Poisson process is examined. With the mentioned conditions, the explicit formulations for the steady state probabilities and the performance measures of the proposed model is derived. Also,some particular cases have been developed and compared with existing models. Finally, the numerical impact of various parameters on performance attributes are also analysed.

Keywords: Heterogeneous Encouraged Arrival, General Bulk Service Rule, Multiple Working Vacations (MWV), Poisson Distribution, Mean Queue Length

1. Introduction

Several fields including telecommunications, traffic signals, the medical sector, inventory and control etc., widely use queuing theory. In a vacations queuing system, there may be a chance of unavailability of a server from a primary service center for a period of time. A vacation is a period of time spent away from the main service location and it can be caused by a variety of circumstances. Neuts'[7] general bulk service rule (GBSR), states that the server processes the clients in batches. Many Researchers have contributed their findings in Queuing theory with server vacations.

Initially, a general type of bulk queues is discussed by Neuts, who also examines the length of the queue and its busiest times. Further,Y. Levy and U. Yechiali[4] discussed about the vacations in queuing model. A class of semi-vacations policies were first presented by L.D Servi and S.G Finn[9], in which the servers operate at a reduced pace rather than suspending all primary service altogether while on vacations and such vacations is termed as Working Vacations.

De-An Wu and Hideaki Takagi[1] examined M/G/1 queue with multiple working vacations and evaluated its performance measures. Liu et al.,[5] stated the queue size probabilities of M/M/1 Multiple Working Vacation Queuing model. Further, K. Julia Rose Mary and M.I. Afthab Begum[3] analysed a single server with bulk service queue with general arrival pattern following multiple working vacations period. Later, K. Santhi and S. Pazhani Bala Murugan[8] examined on the concept of heterogeneous service on M/G/1 queue with two-stage service under single

working vacation. Numerous Researchers analysed the concept of multiple working vacations. Further, O.C. Ibe, O.A. Isijola [2] also discussed on M/M/1 multiple vacations queuing systems with differentiated vacations.

While B.K. Som, S. Seth [10] evaluated M/M/1/N Queuing system with Encouraged Arrivals. Recently, S. Malik, R. Gupta [6] analysed the Finite Capacity Queuing System with Multiple Vacations and Encouraged customers. This paper discusses the concept of heterogeneous arrival specifically encouraged arrival in the busy period state of M/M(a,b)/1/MWV and obtained with performance measures.Also, Prakati and Julia Rose Mary[**?**] discussed the concept of encouraged arrival with both single and multiple working vacations performing with single server. Further, the numerical analysis of the considered model is evaluated.

2. Methods

An M/M(a,b)/1 queuing model of multiple working vacations with heterogeneous encouraged arrival is considered with the following assumptions:

The three different states of arrival is assumed to be heterogeneous and are uniquely denoted. The heterogeneous arrival process in this model is considered to follow Poisson distribution with parameter λ_{wi} in the idle state , λ_{wv} in the vacation state, whereas the heterogeneous encouraged arrival process also follows Poisson distribution with parameter $\lambda_{wb}(1+\delta)$. According to Neuts (1967) general bulk service rule (GBSR), the server processes the clients in batches. This rule states that the server will only begin providing service if at least "a" customers is present. After completing a service, if the server discovers "a" (or) more clients but not more than "b" clients in the system, he serves them all at once; if he discovers more than b, he serves the first b-customers in turn while the others wait.

As a result, there are minimum of "a" units and maximum of "b" units in each batch for service. The assumption is that the service time of batches of size s$(a \leq s \leq b)$ is an independent random variable with identical distribution and a parameter with an exponential distribution "μ_w". When a service is finished and there are less than 'a' clients in the queue, the server starts vacations, which is an exponentially distributed random variable with parameter η' .If the system length is still less than "a" after finishing one vacations, the server takes another vacations, and so on, until the server detects at least "a" customers in the queue (i.e., multiple vacations is used).

During vacations, if the queue size reaches at least "a" customers, then the server begins providing service at a service rate "μ_{wv}" which is different from the regular service rate. The size of the batch in service is "k" with $a \leq k \leq b$ and the service rates are independent of the size of the batch in service, thus when the vacations is over the server will shift the service rate from μ_{wv} to μ_w, when the server is operating. In this model, there is an increase in the arrival rate, i.e., encouraged arrival occurs in the regular busy period and then the server still continues to serve following GBSR. The above queuing model is denoted as M/M(a,b)/1/MWV with heterogeneous encouraged arrival.

3. Steady State Equations

Let $N_K(t)$ = number of customers waiting at the time, "t"
and M(t) = 0,1 or 2 denotes that the server is idle during the vacations or working during vacations or in the regular busy period respectively.

Let $IR_n(t) = Pr\{N_{k(t)} = n, M(t) = 0\} \quad ; \quad 0 \leq n \leq a-1$

$VQ_n(t) = Pr\{N_{k(t)} = n, M(t) = 1\}; n \geq 0$

$BP_n(t) = Pr\{N_{k(t)} = n, M(t) = 2\}; n \geq 0$

when M(t) = 0, the size of the queue and the system are same,

when M(t) = 1 or 2, then the size of the system is the sum of total no. of customers waiting in the queue or the size of the service batch containing $a \leq k \leq b$ customers.

The Steady State Probabilities satisfying the Chapman Kolmogrov equations are assumed as

follows:

$$VQ_n = t \xrightarrow{lt} \infty VQ_n(t);$$

$$IR_n = t \xrightarrow{lt} \infty IR_n(t);$$

$$BP_n = t \xrightarrow{lt} \infty BP_n(t)$$

For the specified model, the steady state equations are expressed below:

$$\lambda_{wi} IR_0 = \mu_w BP_0 + \mu_{wv} VQ_0 \tag{1}$$

$$\lambda_{wi} IR_n = \lambda_{wi} IR_{n-1} + \mu_w BP_n + \mu_{wv} VQ_n \quad ; \quad 1 \leq n \leq a-1 \tag{2}$$

$$(\lambda_{wv} + \dot{\eta} + \mu_{wv}) VQ_0 = \lambda_{wi} IR_{a-1} + \mu_{wv} \sum_{n=a}^{b} VQ_n \tag{3}$$

$$(\lambda_{wv} + \dot{\eta} + \mu_{wv}) VQ_n = \lambda_{wv} VQ_{n-1} + \mu_{wv} VQ_{n+b} \quad ; \quad n \geq 1 \tag{4}$$

$$(\lambda_{wb}(1+\delta) + \mu_w) BP_0 = \mu_w \sum_{n=a}^{b} BP_n + \dot{\eta} VQ_0 \tag{5}$$

$$(\lambda_{wb}(1+\delta) + \mu_w) BP_n = (\lambda_{wb}(1+\delta)) BP_{n-1} + \mu_w BP_{n+b} + \dot{\eta} VQ_n) \quad ; \quad n \geq 1 \tag{6}$$

In the above mentioned steady state equations, Eq (1) and Eq (2) represent idle state, working vacations are defined in Eq (3) and Eq (4). Furthermore, encouraged arrival occurs during the regular busy period i.e., in Eq (5) & Eq (6)

3.1. Steady State Solution

The concept of forward shifting operator (E) is introduced on BP_n and VQ_n to solve the above defined steady state equations,

$$E(BP_n) = BP_{n+1}; E(VQ_n) = VQ_{n+1}; for (n \geq 0)$$

The homogeneous difference equation is obtained from Eq (4)

$$(\mu_{wv} E(b+1) - (\lambda_{wi} + \dot{\eta} + \mu_{wv}) E + \lambda_{wi}) VQ_n = 0; n \geq 0 \tag{7}$$

The characteristic equation of the difference equation is expressed as follows
$h(w) = (\mu_{wv} w(b+1) - (\lambda_{wi} + \dot{\eta} + \mu_{wv}) w + \lambda_{wi}) = 0.$

By assuming $f(w) = (\lambda_{wi} + \dot{\eta} + \mu_{wv}) w$ and $g(w) = \mu_{wv} w(b+1) + \lambda_{wi}$, it is known that if $|g(w)| \leq |f(w)|$ on $|w| = 1$, then by Rouche's theorem, h(w) has only one root z_v inside the contour $|w|$. As the root is real, solution of the homogeneous difference equation is obtained as

$$VQ_n = z_v^n VQ_0; n \geq 0 \tag{8}$$

Similarly, Eq(6) can be written as

$$[\mu_w E(b+1) - (\lambda_{wb}(1+\delta) + \mu_w) E + \lambda_{wb}(1+\delta)] BP_n = -\dot{\eta} z_v^{(n+1)} VQ_0; n \geq 0 \tag{9}$$

Again by Rouche's Theorem, Eq (9) is obtained as

$[\mu_w w^{(b+1)} - (\lambda_{wb}(1+\delta) + \mu_w)w + \lambda_{wb}(1+\delta)] = 0$ has a unique root 'z' with $|z| < 1$ provided $\frac{\lambda_{wb}(1+\delta)}{b\mu} < 1$

The non-homogeneous difference equation is solved and the solution obtained is given by

$$BP_n = (Xz^n + Yz_v^n)VQ_0 \tag{10}$$

where

$$Y = \frac{\dot{\eta} z_v}{(\lambda_{wb}(1+\delta))(z_v - 1) + \mu_w z_v (1 - {z_v}^b)}; if z_v \neq z \tag{11}$$

by adding Eqs (1) & (2) over 0 to n, and substituting VQ_n & BP_n , IR_n is obtained as follows

$$IR_n = [\frac{\mu_w}{\lambda_{wi}} \left(\frac{X(1-z^{n+1})}{(1-z)} + \frac{Y(1-{z_v}^{n+1})}{(1-z_v)} \right) + \frac{\mu_{wv}}{\lambda_{wi}} \frac{(1-{z_v}^{n+1})}{(1-z_v)}]VQ_0$$

Hence, the steady state queue size probabilities are expressed in terms of the unknowns X and VQ_0.

Now to calculate X , consider Eq (5) and on substituting the value of BP_n, it is found that

$$X(\lambda_{wb}(1+\delta) + \mu_w) - \mu_w \frac{(z^a - z^{b+1})}{(1-z)} = \dot{\eta} - Y((\lambda_{wb}(1+\delta) + \mu_w) - \mu_w \frac{({z_v}^a - {z_v}^{b+1})}{(1-z_v)}) \tag{12}$$

which can be simplified as

$$\frac{X\mu_w(1-z^a)}{(1-z)} = \left(\left(\frac{\dot{\eta}}{(1-z_v)} \right) - \frac{Y\mu_w(1-{z_v}^a)}{(1-z_v)} \right) \tag{13}$$

Moreover, Eq (3) is also verified and the steady state queue size probabilities are expressed in terms of VQ_0 and obtained as

$$VQ_n = {z_v}^n VQ_0 \quad ; \quad n \geq 0 \tag{14}$$

$$BP_n = (Xz^n + Y{z_v}^n)VQ_0 \quad ; \quad n \geq 0 \tag{15}$$

$$IR_n = [\frac{\mu_w}{\lambda_{wi}} \left(\frac{X(1-z^{n+1})}{(1-z)} + \frac{Y(1-{z_v}^{n+1}}{(1-z_v)} \right) + \frac{\mu_{wv}}{\lambda_{wi}} \frac{(1-{z_v}^{n+1})}{(1-z_v)}]VQ_0; \quad 0 \leq n \leq a-1 \tag{16}$$

where

$$X = \frac{(1-z)}{\mu_w(1-z^a)} \left(\left(\frac{\dot{\eta}}{(1-z_v)} \right) - \frac{Y\mu_w(1-{z_v}^a)}{(1-z_v)} \right) \tag{17}$$

The value for VQ_0 is evaluated by using the normalizing condition
$\sum_{n=0}^{\infty}(VQ_n + BP_n) + \sum_{n=0}^{a-1} IR_n = 1$

$$\sum_{n=0}^{\infty}({z_v}^n)VQ_0 \; + (Xz^n + Yz_v^n)VQ_0 + \sum_{n=0}^{a-1}[\frac{\mu_w}{\lambda_{wi}} \left(\frac{X(1-z^{n+1})}{(1-z)} + \frac{Y(1-{z_v}^{n+1}}{(1-z_v)} \right) + \frac{\mu_{wv}}{\lambda_{wi}} \frac{(1-{z_v}^{n+1})}{(1-z_v)}]VQ_0 = 1$$

Thus $(VQ_0^{-1}) = F(z_v, \mu_{wv}) + XF(z, \mu_w) + YF(z_v, \mu_w)$;

where $F(r,t) = \frac{1}{1-r}(1 + \frac{t}{\lambda_{wi}}(a - \frac{r(1-r^a)}{(1-r)}))$

4. Performance Measures

One of the objectives of this paper is to deduce the expected queue length of the considered model. The expected queue length is also known as Mean Queue length and denoted as L_q

4.1. Mean Queue Length

The expected queue length L_q is calculated as

$$L_q = \sum_{n=1}^{\infty} n(VQ_n + BP_n) + \sum_{n=0}^{a-1} nIR_n \tag{18}$$

by substituting the values of VQ_n, BP_n and IR_n , L_q is simplified as

$$L_q = XH(z, \mu_w) + YH(z_v, \mu_w) + H(z_v, \mu_{wv})$$

where

$$H(r,t) = \frac{r}{(1-r)^2} + \frac{t}{\lambda_w(1-r)}\left(\frac{a(a-1)}{2} + \frac{ar^{a+1}(1-r) - r^2(1-r^a}{(1-r)^2}\right)$$

as X and Y are given by Eq (15) & Eq (16)

Also,the other characteristics of the queuing model includes if P_v, P_{busy} and P_{idle} respectively which denotes the probability that the server is in working vacations state, in regular and is idle in vacations state, then

$$P_v = \sum_{n=0}^{\infty}(VQ_n) = \frac{VQ_0}{(1-z_v)} \tag{19}$$

$$P_{busy} = \sum_{n=0}^{\infty}(BP_n) = \left(\frac{X}{(1-z)} + \frac{Y}{(1-z_v)}\right)VQ_0 \tag{20}$$

$P_{busy} = \left(\frac{X}{(1-z)} + \frac{Y}{(1-z_v)}\right)VQ_0$,where X & Y are specified in Eq (17) & Eq (11)

$$P_{idle} = \sum_{n=0}^{a-1} IR_n \tag{21}$$

where IR_n is given by Eq (16). Thus, the above characteristics helps in deciding the efficiency of the queuing model.

5. Particular Cases

5.1. Case 1: M/M/1 Multiple Working Vacations Model

The M/M/1 Multiple working vacations queuing model's steady state queue size probabilities are calculated.

By letting in equations Eq (14) to Eq (16) at a= b = 1 and $\lambda_{wb}(1+\delta) = \lambda_w$, it is observed that

$$\begin{aligned} VQ_n &= z_v^n VQ_0 \quad ; n \geq 0 \\ BP_n &= \frac{y}{z_v}\left(z_v^{n+1} - z_v^{n+1}\right) \quad ; n \geq 0 \\ IR_n &= \frac{VQ^0}{z_v} \text{where 'z'} = \frac{\lambda_{wi}}{\mu} = \rho, \quad ; n \geq 0 \end{aligned}$$

where

$$X = -\frac{Y\rho}{z_v} \quad \text{and} \quad Y = \frac{\dot{\eta} z_v}{\mu_w(1-z_v) + (z_v - \rho)}$$

which results in the queue size probabilities of $M/M/1$ Multiple Working Vacations Queuing model (Liu et al.2007)

5.2. Case 2: M/M/(a,b)/1 Multiple Working Vacations Model

Letting $\lambda_{wb}(1+\delta) = \lambda_w$,it is obtained as $M/M(a,b)/1$ under working vacations.

$$\begin{aligned} VQ_n &= z_v^n VQ_0 \quad ; n \geq 0 \\ BP_n &= (Xz^n + Yz_v^n)VQ_0 \quad ; n \geq 0 \\ IR_n &= \left[\frac{\mu_w}{\lambda_w i}\left(\frac{X(1-z^{n+1})}{(1-z)} + \frac{Y(1-z_v^{n+1})}{(1-z_v)} + \frac{\mu_{wv}}{\lambda_w i}\frac{(1-z_v^{n+1})}{(1-z_v)}\right]\right. VQ_0; 0 \leq n \leq a-1 \end{aligned}$$

By substituting the values of VQ_n, BP_n and IR_n in (L_q) , it is observed that (L_q)= X H(z,μ_w) + YH(z_v,μ_w) + H(z_v, μ_{wv})

As $H(s,t)$= $\frac{s}{(1-s)^2} + \frac{t}{\lambda_w i(1-s)}\left(\frac{a(a-1)}{2} + \frac{as^{a+1}(1-s)-s^2(1-s^a)}{(1-s)^2}\right)$;

where $X = \frac{(1-z)}{\mu_w(1-z^a)}\left(\left(\frac{\dot{\eta}}{(1-z_v)}\right) - \frac{Y\mu_w(1-{z_v}^a)}{(1-z_v)}\right)$

$Y = \frac{\dot{\eta}z_v}{\lambda_{wi}(z_v-1)+\mu_w z_v(1-{z_v}^b)}$ if $z_v \neq z$

which coincides with the queue size probabilities of M/M(a,b)/1 Multiple Working Vacations (K.Julia Rose Mary,2011)

6. Sensitivity Analysis

The numerical values of the mean queue length are derived for $M/M(a,b)/1$ MWV queuing system with heterogeneous encouraged arrival during busy period. The obtained numerical values are examined to study the impact of the vacation parameter $\dot{\eta}$ and also various other parameters like encouraged arrival rate $\lambda_w b(1+\delta)$, regular service rate (μ_w), service rate during vacation (μ_{wv}).

Based on the above parameters, the mean queue length (L_q) i.e., the expected system size of batch arrival has been calculated. Further, a comparative study is made between L_q of $M/M(a,b)/1$ MWV Encouraged Arrival and L_q of $M/M(a,b)/1$ MWV under Heterogeneous Encouraged arrival.

Thus the impact of the mentioned parameters on the system size probabilities and on the expected system size during the idle state, working vacation state and the regular busy period are also analysed by considering z_v and z as the roots of the following characteristic equations

$$(\mu_{wv}w^{(b+1)} - (\lambda_{wv} + \dot{\eta} + \mu_{wv})w + \lambda_{wv}) = 0 \quad \&$$

$$(\mu_w w^{(b+1)} - (\lambda_{wb}(1+\delta) + \mu_w)w + \lambda_{wb}(1+\delta) = 0$$

respectively.By considering homogeneous arrival rate during idle and vacation state as $\lambda_w = 4.05$ and during encouraged arrival $\lambda_w(1+\delta) = 5.051$, also, by assuming for heterogeneous arrival queuing system arrival rate as $\lambda_{wi} = 3.9; \lambda_{wv} = 4.0$& during encouraged arrival $\lambda_{wb}(1+\delta) = 5.051$ and further by assuming $\mu_w = 0.9$ as constant and by varying $\dot{\eta}$ & μ_{wv}, the mean queue length (L_q) is calculated and tabulated below.

Table 1: *L_q of $M/M(a,b)/1$ MWV Queuing system with Heterogeneous EA and $M/M(a,b)/1$ MWV Queuing system under Homogeneous EA*

μ_{wv}	$\dot{\eta}$	L_q (Heterogeneous EA)	L_q (Homogeneous EA)
0.05	0.02	211.4142	211.5021
	0.04	159.4418	159.5500
	0.06	89.8687	89.9613
	0.08	64.5690	64.6545
	0.1	48.6767	48.7549
0.1	0.02	127.0639	127.1303
	0.04	105.3322	105.4188
	0.06	74.3123	74.3973
	0.08	52.9088	52.9864
	0.1	46.01627	46.0943
0.15	0.02	100.1150	100.1765
	0.04	83.3706	83.44902
	0.06	64.1551	64.2356
	0.08	45.1962	45.2689
	0.1	41.5132	41.5887
0.2	0.02	75.1551	75.2117
	0.04	72.1117	72.1873
	0.06	52.6958	52.7703
	0.08	41.8486	41.9209
	0.1	39.3368	39.41266

From the above tabulation, it is evident that the queue length of the heterogeneous encouraged arrival is comparatively less than that of the homogeneous encouraged arrival, as the arrival rate differs in each state of the heterogeneous encouraged arrival model. The obtained numerical values are represented graphically.

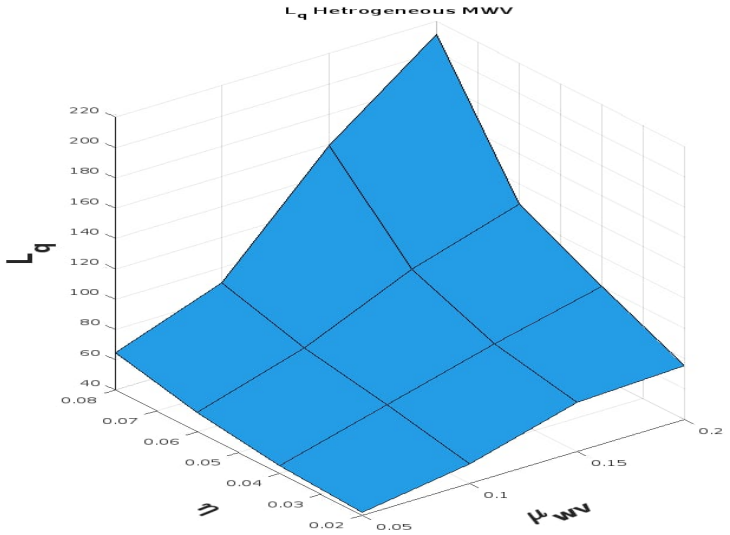


Figure 1: *$M/M(a,b)/1$ MWVwith Heterogeneous Encouraged Arrival*

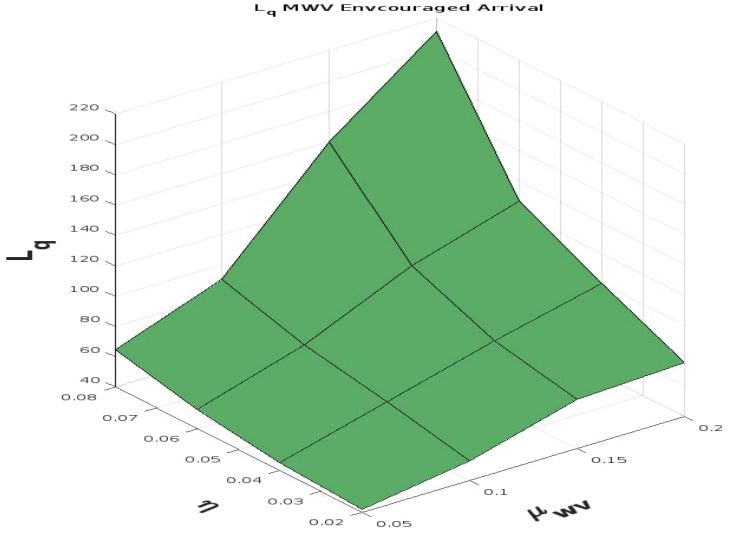


Figure 2: *$M/M(a,b)/1$ MWV with HomogeneouEncouraged Arrival*

The above graphical representations demonstrate that the mean queue length reaches its peak at the minimum values of $\dot{\eta}$&μ_{wv} and the mean queue length reduces at a great margin level when $\dot{\eta}$&μ_{wv} are maximum. For example if μ_{wv} = 0.05 , and $\dot{\eta}$= 0.02 , the mean queue length of M/M(a,b)/1 MWV with heterogeneous encouraged arrival is 211.4142 while the mean queue length of M/M(a,b)/1 MWV queuing model with encouraged arrival is 211.5021. Thus with the varied arrival rates there arises a gradual variation in the queue length.

As discussed earlier, the encouraged arrival is an impact factor in deciding the queue length. Thus on varying the encouraged arrival rate during the regular busy period as $\lambda_{wb}(1+\delta)$ = 5.051 ; $\lambda_{wb}(1+\delta)$ = 5.053 ; $\lambda_{wb}(1+\delta)$ = 5.055 and by having λ_{wi}= 3.9, λ_{wv}= 4.0 &μ_w = 0.9 as constant and on varying μ_{wv} from 0.05 to 0.2 and $\dot{\eta}$ from 0.02 to 0.1 , the mean queue length is computed and tabulated below

Table 2: *L_q of $M/M(a,b)/1$ MWV Heterogeneous Queuing system with varied EA*

μ_{wv}	$\dot{\eta}$	$L_q(\lambda_{wb}(1+\delta)=5.051)$	$L_q(\lambda_{wb}(1+\delta)=5.053)$	$L_q(\lambda_{wb}(1+\delta)=5.053)$
	0.02	127.0639	127.06636	127.0687
	0.04	105.3322	105.3360	105.3399
0.1	0.06	74.3123	74.3164	74.3205
	0.08	52.9088	52.91281	2.968
	0.1	46.01627	46.02055	46.0248
	0.02	100.1150	100.1170	100.11900
	0.04	83.3706	83.37881	83.37708
0.15	0.06	64.1551	64.1588	64.1624
	0.08	45.1962	45.1998	45.2034
	0.1	41.5132	41.5172	41.5212
	0.02	75.1551	75.1566	75.1582
	0.04	72.1117	72.1146	72.1174
0.2	0.06	52.6958	52.6990	52.7022
	0.08	41.8486	41.8520	41.85550
	0.1	39.3368	39.3407	39.3446

Thus from the above tabulations, it is clear that as the encouraged arrival rate increases gradually from $\lambda_{wb}(1+\delta)$= 5.051 to $\lambda_{wb}(1+\delta)$= 5.055 the queue length increases gradually from 211.4142705 to 211.4180277. Similarly, when the encouraged arrival rate increases from $\lambda_{wb}(1+\delta)$= 5.051 to $\lambda_{wb}(1+\delta)$= 5.055 the queue length increases from **211.4142705 to 211.4217896** . Thus it is clear that as the encouraged arrival rate increases,the queue length increase. Hence the encouraged arrival rate significantly decides the queue length of the considered model. The above tabulations is represented graphically below

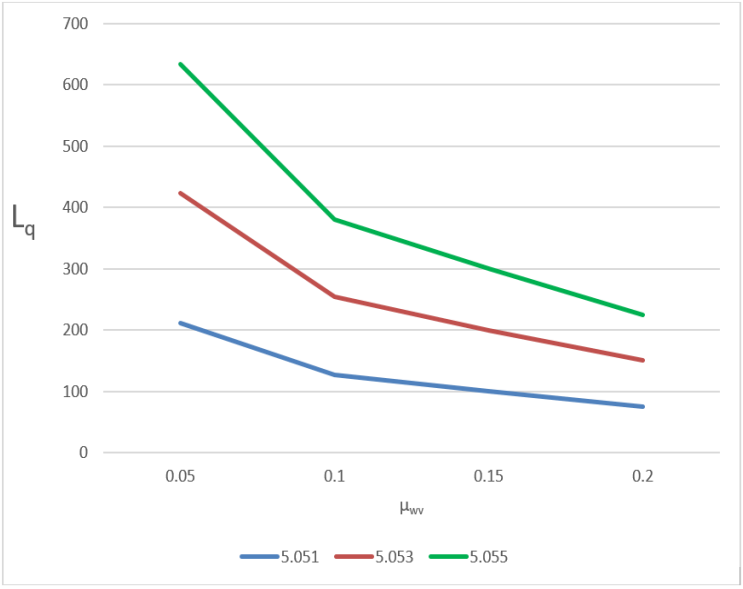


Figure 3: *$M/M(a,b)/1$ MWV Heterogeneous Queuig system with varied Encouraged Arrival*

Thus for the vacation service rate μ_{wv} varying from 0.05 to 0.2, with the increased encouraged arrival rate, the corresponding mean queue length is denoted in the above graph.

Now, the other performance measures of the discussed model like P_{idle}. $P_{vacation}$, P_{busy} are computed by using Eq (19), Eq (20) & Eq (21) and tabulated below:

Table 3: *Other Characteristics of $M/M(a,b)/1$ MWV Heterogeneous Queuing system under encouraged arrival*

μ_{wv}	$\dot{\eta}$	P_{idle}	$P_{vacation}$	P_{busy}
0.1	0.02	0.01438	0.7468	0.2387
	0.04	0.02275	0.6294	0.3477
	0.06	0.0319	0.5935	0.3745
	0.08	0.04156	0.5786	0.37973
	0.1	0.04845	0.54449	0.4070
0.15	0.02	0.0170	0.7791	0.2038
	0.04	0.0261	0.6685	0.3053
	0.06	0.0352	0.6168	0.3478
	0.08	0.04602	0.6010	0.3529
	0.1	0.05232	0.5584	0.3892
0.2	0.02	0.02104	0.8105	0.1683
	0.04	0.02933	0.6895	0.2811
	0.06	0.0400	0.6451	0.3147
	0.08	0.0496	0.61013	0.3402
	0.1	0.05560	0.5644	0.3799

From the above table, as the working vacation service rate (μ_{wv}) increases and with the vacation parameter varying from 0.02 to 0.1, the probability values during the idle state and busy period increases and the probability during the vacation period decreases.

7. Conclusion

M/M(a,b)/1 MWV queue under heterogeneous encouraged arrival of customers is studied with steady state and the steady state solutions are derived. Additionally, the mean queue length and various other performance measures were also calculated and numerically examined. Moreover, other queuing models were deduced as particular cases.

From the numerical examples computed for the mean queue length of the considered model and compared with M/M(a,b)/1/MWV with encouraged arrival, evidently concluded that many factors like vacation parameter, service rate plays a role in deciding the queue length, the key factor is the encouraged arrival rate that makes a notable impact in deciding the queue length of the considered batch service heterogeneous queuing model with single server.

Conflict of Interest: The authors declare no conflict of interest

8. Acknowledgement

The corresponding author would like to acknowledge the UGC Scholarship team for approving and providing (SJSGC) Scholarship Grants, which is more likely helpful and beneficial to complete this study.

References

[1] Wu, D. A. and Takagi, H. (2006). M/G/1 queue with multiple working vacations. *Performance Evaluation*, 63:654–681.

[2] Ibe, O. C. and Isijola, O. A. (2014). M/M/1 multiple vacation queueing systems with differentiated vacations.*Modelling and Simulation in Engineering*, 2014:158247.

[3] Mary, J. R. K. and Afthab Begum, MI (2009). Closed form Analytical solution of the General Bulk service queuing model M/M (a,b)/1 under working vacation. *International conference on Mathematical and Computational models, PSG College of Tech*, 1:92–100.

[4] Levy, Y. and Yechiali, U. (1976). An M/M/s queue with servers vacations. *INFOR: Information Systems and Operational Research*, 14:153–163.

[5] Liu, W. Y., Xu, X. L. and Tian, N. S. (2007). Stochastic decompositions in the M/M/1 queue with working vacations. *Operations Research Letters*, 35:595–600.

[6] Malik, S., and Gupta, R. (2022). Analysis of Finite Capacity Queuing System with Multiple Vacations and Encouraged Customers. *International Journal of Scientific Research in Mathematical and Statistical Sciences*, 9:17–22.

[7] Neuts, M. F. (1967). A general class of bulk queues with Poisson input. *The Annals of Mathematical Statistics*, 38:759–770.

[8] Santhi, K. and Pazhani Bala Murugan, S. (2013). An M/G/1 queue with two-stage heterogeneous service and single working vacation. *International Mathematical Forum*, 8:1323–1336.

[9] Servi, L. D. and Finn, S. G. (2002). M/M/1 queues with working vacations (M/M/1/WV). *Performance Evaluation*, 50:41–52.

[10] B.K. Som, S. Seth.(2017). An M/M/1/N Queuing system with Encouraged Arrivals. *Global Journal of Pure and Applied Mathematics*, 7:3443–3453.

[11] Prakati,P. and Julia Rose Mary.K (2024). M/M/C Queue With Multiple Working Vacations and Single Working Vacation Under Encouraged Arrival With Impatient Customers. *Reliability:Theory and Applications*, 19:650–662.

RT&A, No 4(80)
Volume 19, December, 2024

TRIANGULAR AND SKEW-SYMMETRIC SPLITTING METHOD FOR SOLVING FUZZY STOCHASTIC LINEAR SYSTEM

A. Shivaji [1], B. Harika [2], D. Rajaiah [3], L.P. Rajkumar [4]

•

Department of Mathematics, Kakatiya University, Telangana, India [1,2,4]
Department of Mathematics, KITS Warangal, Telangana, India [3]
sivajiarepalli@gmail.com, hbolledla@gmail.com , dr.mh@kitsw.ac.in , ladalla.raj@gmail.com

Abstract

Based on the Triangular and Skew Symmetric (TSS) splitting method, a novel iterative approach is proposed to solve a class of fuzzy regularized linear system of equations with fuzzy coefficient stochastic rate matrix. The non-homogeneous fully fuzzy linear system is same as the non-homogeneous linear system which is derived from the homogeneous linear system with stochastic rate matrix and steady state vector. An iterative procedure is developed for finding a unique non-trivial solution. Numerical results shown that the proposed method is effective and efficient when compared with the existing classical methods.

Keywords: Fuzzy Stochastic Rate Matrix; Triangular and Symmetric Splitting Method; Parametric Form of Fuzzy Number; Fully Fuzzy System of Liner Equations; Error Analysis.

1. Introduction

The system of fuzzy linear equations has a variety of applications in the areas of information, engineering, statistics, mathematics, etc. In several applications, the system's parameters and measurements are performed by fuzzy values rather than crisp ones. Thus, it is key to expand mathematical models and numerical mechanisms that would be treated as an ordinary fuzzy linear system (FLS) and solve them using different techniques. The mathematical modelling of the problem is considered as fuzzy system of linear equations. The homogeneous system $\pi Q = 0$, where π is the stationary probability vector and Q is the stochastic rate matrix is transformed into a non-homogeneous system $Ax = b$, where $A = Q^T + \epsilon I$ with small perturbation $\epsilon > 0$. The regularized linear system $Ax = b$ where the parameters are uncertain and vague is transformed into a fully fuzzy linear system (FFLS) $\Phi X = \Psi$, with Φ and Ψ are fuzzy matrices and X is an unknown fuzzy vector to be determined for a unique non-zero solution.

Friedman [1] et al designed a model with an embedding technique for computing a class of $n \times n$ (FLS). LU decomposition method was developed by Abbasbandy [2] et al, Steepest descent method by Abbasbandy and Jafarian [3], The Jacobi, Gauss-Seidel and SOR iterative methods are used by Allahviranloo [4]. Adomian decomposition method was suggested by Allahviranloo [5], inherited LU factorization method by M. A. Fariborzi Araghi and A. Fallahzadeh [6] for solving a fuzzy system of linear equations. A few numerical methods were developed and discussed by the general model [7, 8]. A.N.A. Koam [9] et al LU decomposition scheme is used for solving m-polar fuzzy system of linear equations. Block SOR method was proposed by S. X. Miao [10] et

al, the QR-decomposition method was developed by S.H. Nasseri [11] et al, K. Wang and Y. Wu introduced the Uzawa-SOR method [12], Symmetric Successive Over Relaxation method, Block iterative method, and Splitting iterative methods were established by K. Wang B. Zheng and J. F. Yin[13, 14, 15]. Y.R. Wang and Y. L. Chen suggested a modified Jacobi iterative method for large-size linear systems [16], and a new method based on Jacobi iteration was proposed for solving the fuzzy linear systems by Zhen Huang [17] et al. The traditional TSS method is easy to execute and applicable to compute stationary probability vector and the performance measures in understanding many real-time systems. In this research work, an advanced iterative method is deployed based on TSS iteration method which brings the solution for fuzzy linear systems [18].

The rest of the paper is organized as follows. Section 2 gives some fundamentals of FLS. In section 3, the new method is established. Numerical examples are presented in section 4 and the conclusions are drawn in section 5.

2. Fundamentals

Fuzzy number : A pair of functions $(\underline{u}(r), \overline{u}(r))$, $0 \leq r \leq 1$, which satisfies the conditions,

- $\underline{u}(r)$ is a bounded left continuous nondecreasing function over [0, 1]
- $\overline{u}(r)$ is a bounded left continuous nonincreasing function over [0, 1]
- $\underline{u}(r) \leq \overline{u}(r)$, $0 \leq r \leq 1$

is known as a fuzzy number.

Arithmetic Operations : The arithmetic operations involving in arbitrary fuzzy numbers $x = (\underline{x}(r), \overline{x}(r))$, $y = (\underline{y}(r), \overline{y}(r))$, $0 \leq r \leq 1$, and for $k \in R$, are defined by:

- $x = y$ if and only if $\underline{x}(r) = \underline{y}(r)$ and $\overline{x}(r) = \overline{y}(r)$
- $x + y = (\underline{x}(r) + \underline{y}(r), \overline{x}(r) + \overline{y}(r))$ and
- $kx = \begin{cases} (k\underline{x}(r), k\overline{x}(r)), k > 0, \\ (k\overline{x}(r), k\underline{x}(r)), k < 0. \end{cases}$

Fuzzy System of Linear Equations : The $n \times n$ fuzzy linear system (FLS) may be written as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdot + a_{nn}x_n &= b_n \end{aligned}$$

the matrix form of the above linear system is

$$Ax = b \tag{1}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11} & a_{nn} & \cdots & a_{nn} \end{bmatrix} \text{ is a crisp matrix}$$

$b = [b_1, b_2, \cdots, b_n]^T$ is a fuzzy vector and $x = [x_1, x_2, \cdots, x_n]^T$ is unknown.

Solution of Fuzzy Linear System : The solution of the fuzzy linear system is a fuzzy vector $x = (x_1, x_2, \ldots, x_n)^T$
given by

$$x_i = (\underline{x}_i(r), \overline{x}_i(r)), 1 \leq i \leq n, 0 \leq r \leq 1$$

if

$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij}x_j = \sum\limits_{j=1}^{n} \underline{a_{ij}x_j} = \underline{b}_i \\ \overline{\sum\limits_{j=1}^{n} a_{ij}x_j} = \sum\limits_{j=1}^{n} \overline{a_{ij}x_j} = \overline{b}_i \end{cases} \tag{2}$$

The extended FLS (1) into the $2n \times 2n$ crisp linear system is defined as,

$$\Phi X = \Psi \tag{3}$$

where, $\Phi = (\phi_{kl})$, ϕ_{kl} are determined as follows

$$a_{ij} > 0 \Rightarrow \phi_{kj} = a_{ij}, \phi_{n+i,n+j} = a_{ij}$$
$$a_{ij} < 0 \Rightarrow \phi_{i,n+j} = a_{ij}, \phi_{n+i,j} = a_{ij}, 1 \leq i, j \leq n,$$

and any ϕ_{kl} which is not determined by the above items is zero, $1 \leq i, j \leq 2n$, and

$$X = \begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \\ \overline{x}_1 \\ \vdots \\ \overline{x}_n \end{bmatrix}, \Psi \begin{bmatrix} = \begin{matrix} \underline{b}_1 \\ \vdots \\ \underline{b}_n \\ \overline{b}_1 \\ \vdots \\ \overline{b}_n \end{matrix} \end{bmatrix}$$

Moreover, the matrix Φ has the form $\begin{bmatrix} \Phi_1 & \Phi_2 \\ \Phi_2 & \Phi_1 \end{bmatrix}$, $A = \Phi_1 + \Phi_2$, and (2) can be written as

$$\begin{cases} \Phi_1 \underline{X} + \Phi_2 \overline{X} = \underline{\Psi} \\ \Phi_2 \underline{X} + \Phi_1 \overline{X} = \overline{\Psi} \end{cases}$$

where

$$\underline{X} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix} \quad \overline{X} = \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_n \end{bmatrix}$$

$$\underline{\Psi} = \begin{bmatrix} \underline{b}_1 \\ \underline{b}_2 \\ \vdots \\ \underline{b}_n \end{bmatrix} \quad \overline{\Psi} = \begin{bmatrix} \overline{b}_1 \\ \overline{b}_2 \\ \vdots \\ \overline{b}_n \end{bmatrix}$$

Stochastic Rate Matrix : A matrix $Q = [q_{ij}]$ is known as stochastic rate matrix, if is satisfies the following conditions

- $q_{ij} \geq 0, 1 \leq i, j \leq n$
- $q_{ii} = -\sum\limits_{j \neq i} q_{ij}$

3. Triangular and Skew Symmetric Splitting Iterative Method for Regularized Linear System

In this section, the stationary probability vector π of $\pi Q = 0$ can be found using the Fuzzy Triangular and Skew-Symmetric (FTSS) iterative method of a regularized linear system (3). We establish the TSS splitting method for the stochastic rate matrix as follows,
The matrix A of the system (1) is split into the form,

$$\begin{aligned} A =& (L + D + U^T) + (U - U^T) \\ =& T + S, \end{aligned}$$

where,
$T = L + D + U^T$, $S = (U - U^T)$ are triangular and skew symmetric matrices.

Thus the regularized system (2) can take the form $(T + S)X = Y$.

Consider $D = \begin{bmatrix} D_1 & 0 \\ 0 & D_1 \end{bmatrix}$, $L = \begin{bmatrix} L_1 & 0 \\ -S_2 & L_1 \end{bmatrix}$, $U = \begin{bmatrix} U_1 & -S_2 \\ 0 & U_1 \end{bmatrix}$

$$\Rightarrow U^T = \begin{bmatrix} U_1 & 0 \\ -S_2 & U_1 \end{bmatrix}$$

Then

$$\begin{aligned} L + D + U^T &= \begin{bmatrix} L_1 & 0 \\ -S_2 & L_1 \end{bmatrix} + \begin{bmatrix} D_1 & 0 \\ 0 & D_1 \end{bmatrix} + \begin{bmatrix} U_1 & -S_2 \\ 0 & U_1 \end{bmatrix} \\ &= \begin{bmatrix} L_1 + D_1 + U_1 & 0 \\ -2S_2 & L_1 + D_1 + U_1 \end{bmatrix} \end{aligned}$$

and

$$U - U^T = \begin{bmatrix} 0 & -S_2 \\ S_2 & 0 \end{bmatrix}$$

where D_1, L_1, and U_1 are diagonal, lower, and upper triangular matrices respectively.
The method of triangular and skew-symmetric splitting iterative is as follows,

$$X^{(k+1)} = H(\alpha)X^{(k)} + G(\alpha)b, \quad \text{for} \quad k = 0, 1, 2, \ldots,$$

where,

$$\begin{aligned} X^{(k+1)} &= \begin{bmatrix} \frac{x^{k+1}}{x^{k+1}} \end{bmatrix} \\ H(\alpha) &= (\alpha I_n + S)^{-1}(\alpha I_n - T)(\alpha I_n + T)^{-1}(\alpha I_n - S) \end{aligned}$$

and

$$G(\alpha) = 2\alpha(\alpha I_n + S)^{-1}(\alpha I_n + T)^{-1}$$

We have

$$\alpha I_n + S = \begin{bmatrix} \alpha I_n & -S_2 \\ S_2 & \alpha I_n \end{bmatrix}$$
$$\alpha I_n - S = \begin{bmatrix} \alpha I_n & S_2 \\ -S_2 & \alpha I_n \end{bmatrix}$$
$$\alpha I_n + T = \begin{bmatrix} \alpha I_n + T_1 & 0 \\ -2S_2 & \alpha I_n + T_1 \end{bmatrix}$$
$$\alpha I_n - T = \begin{bmatrix} \alpha I_n - T_1 & 0 \\ 2S_2 & \alpha I_n - T_1 \end{bmatrix}$$

$$\Rightarrow H(\alpha) = \frac{1}{[(\alpha I_n)^2 + S_2^2][\alpha I_n + T_1]^2} \begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \end{bmatrix}$$

where

$$\begin{aligned} H_1 =& (\alpha I_n)^4 - (\alpha I_n)^2 T_1^2 + 3(\alpha I_n)^2 S_2^2 + T_1^2 S_2^2 \\ H_2 =& 2(\alpha I_n)^3 S_2 - 2\alpha I_n T_1^2 S_2 + 4\alpha I_n S_2^3 \\ H_3 =& 2(\alpha I_n)^3 S_2 + 2\alpha I_n T_1^2 S_2 \\ H_4 =& (\alpha I_n)^4 - (\alpha I_n)^2 T_1^2 + 3(\alpha I_n)^2 S_2^2 + T_1^2 S_2^2 \end{aligned}$$

and

$$G(\alpha) = \frac{2\alpha}{[(\alpha I_n)^2 + S_2^2][\alpha I_n + T_1]^2} \begin{bmatrix} (\alpha I_n)^2 + \alpha I_n T_1 + 2S_2^2 + T_1^2 S_2^2 & \alpha I_n S_2 + T_1 S_2 \\ \alpha I_n S_2 - T_1 S_2 & (\alpha I_n)^2 + \alpha I_n T_1 \end{bmatrix}$$

The next section involves the numerical solution of the non-homogeneous regularized fuzzy linear system (1). It is clear that the solution to the system (1) may give a small error due to the membership value r.

Theorem: The iterative solution of the FLS (1) is convergent if $\rho(H(\alpha)) < 1$.

4. Numerical Results

In this section, the adequacy of the FTSS iterative method for the numerical solution of the stochastic rate matrices in a fuzzy nature is inspected. The convergence analysis of the stationary probability vector of FTSS method is compared with traditional Jacobi and TSS iterative methods. For validation, we take the following 3×3 stochastic rate matrix

$$Q = \begin{bmatrix} 0.7 & -0.55 & -0.15 \\ -0.15 & 0.7 & -0.55 \\ -0.55 & -0.15 & 0.7 \end{bmatrix}$$

The above system is transformed into a regularized linear system (1). The regularized linear system is remodelled into 6×6 fully fuzzy linear system

$$\Phi X = \Psi$$

where

$$\Phi = \begin{bmatrix} 0.7+r & 0 & 0 & 0 & -0.55 & -0.15 \\ 0 & 0.7+r & 0 & -0.15 & 0 & -0.55 \\ 0 & 0 & 0.7+r & -0.55 & -0.15 & 0 \\ 0 & -0.55 & -0.15 & 0.7+r & 0 & 0 \\ -0.15 & 0 & -0.55 & 0 & 0.7+r & 0 \\ -0.55 & -0.15 & 0 & 0 & 0 & 0.7+r \end{bmatrix}$$

The initial distribution $x^{(0)}$ and Ψ are taken as $x^{(0)} = [0\ 0\ 0\ 0\ 0\ 1]^T$ and $\Psi = [0\ 0\ r\ 0\ 0\ -r]$, where, r is the membership value which varies from 0 to 1. In this work, only one case $A = (L + D + U^T) + (U - U^T) = T_1 + S_1$ of TSS splitting iterative method is taken and the remaining methods would obey the same. The result for the incident of contraction factor $\alpha = 0.7 + r$ is numerically the same as the diagonal elements of the matrix Q for the diverse values of r. The absolute error and relative errors of the fuzzy linear system are computed and the same are compared with the traditional Jacobi and TSS splitting iterative method which is depicted in Figures (1-4).

Figure 1 depicts the convergence of iterative solutions for the classical Jacobi, TSS and FTSS splitting methods. Figure 2 shows the absolute error in different cases of r for the Fuzzy TSS iterative method. Figure 3 displays the relative error for various values of r for the newly established method, and Figure 4 illustrates the comparison of absolute and relative errors of the TSS splitting iterative model in a fuzzy environment.

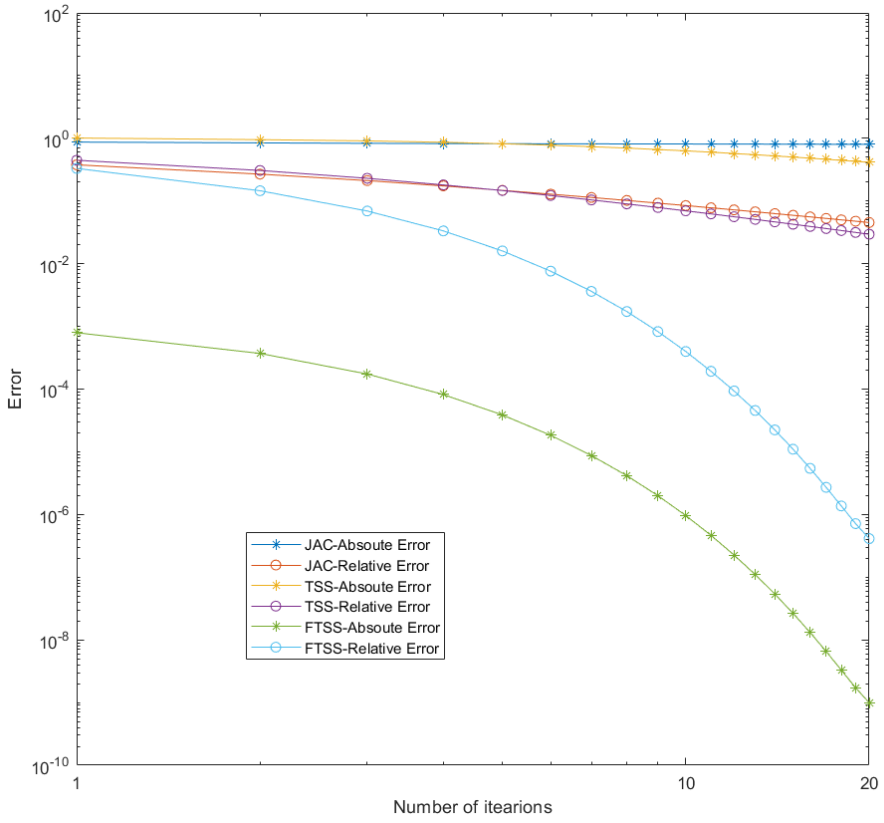


Figure 1: *Absolute and Relative Errors of the Jacobi, TSS and FTSS iterative methods*

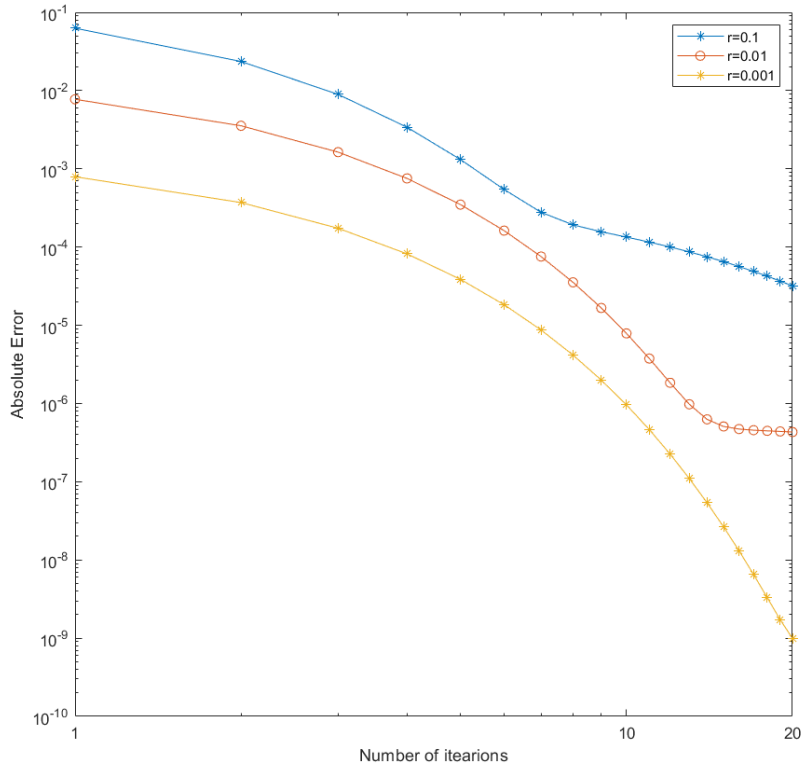


Figure 2: *Absolute Error of FTSS for different values of r*

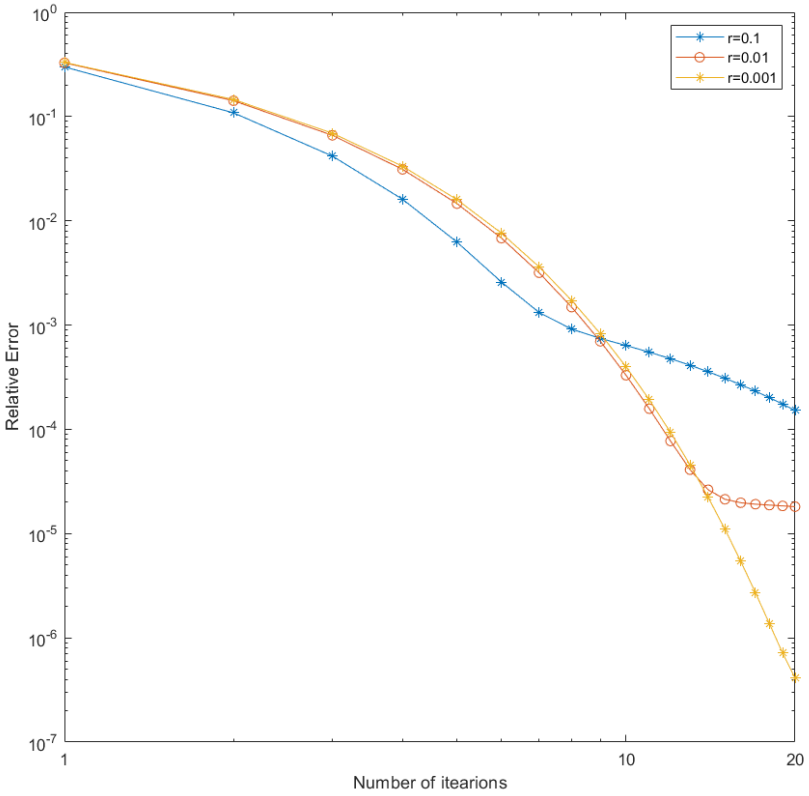


Figure 3: *Relative Error of FTSS for different values of r*

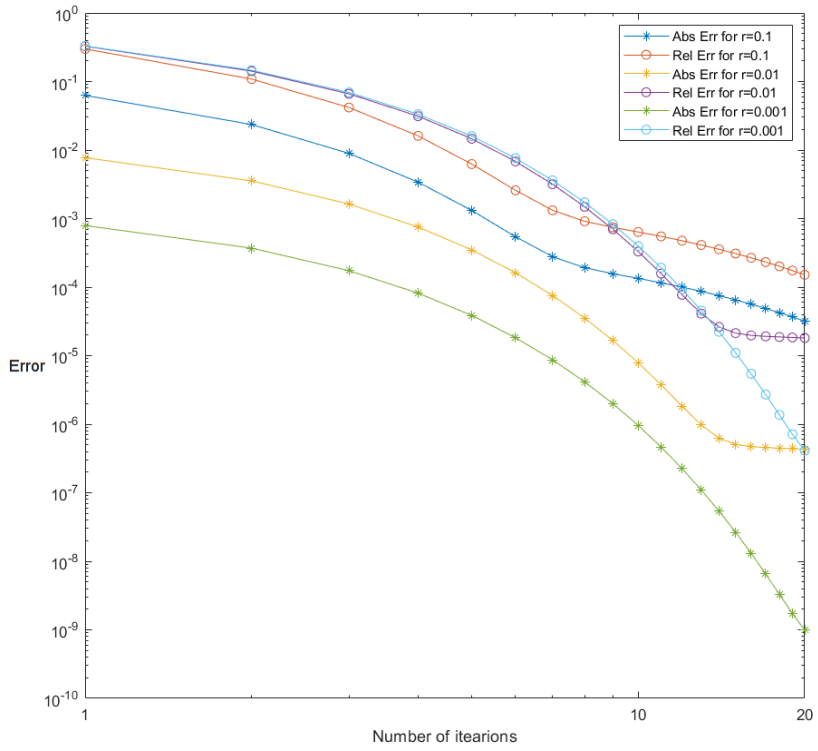


Figure 4: *Absolute and Relative Errors of FTSS for different values of r*

5. Conclusions

In this research work, a new iterative method is formulated on Triangular and Skew-Symmetric iteration for solving a class of fuzzy linear systems of equations with crisp valued stochastic rate matrix. The iterative method is presented and the solution is compared with the traditional methods. The numerical example portrays that the proposed method is effective and competent when compared with traditional iterative methods. We conclude that the suggested method converges to a unique solution and the rate of convergence is faster than the existing traditional methods.

Aknowledgement

One of the Authors A. Shivaji wishes to acknowledge the Ministry of Social Justice & Empowerment, Government of India (MOSJE, GoI) for their funding under the scheme NFSC (National Fellowship for Scheduled Caste) (UGC-Ref. No.: 4545/(CSIR-UGCNETJUNE2019).

References

[1] M. Friedman, M. Ming and A. Kandel [1998], Fuzzy linear systems , *Fuzzy Sets and Systems*, 96:201–209.
[2] S. Abbasbandy, R. Ezzati and A. Jafarian [2006], LU decomposition method for solving fuzzy system of linear equations , *Appl. Math. Comput.*, 172:633–643.
[3] S. Abbasbandy and A. Jafarian [2006], Steepest descent method for system of fuzzy linear equations , *Appl. Math. Comput.*, 175:823–833.

[4] T. Allahviranloo [2004], Numerical methods for fuzzy system of linear equations , *Appl. Math. Comput.*, 155:493–502.
[5] T. Allahviranloo [2005], The Adomian decomposition method for fuzzy system of linear equations , *Appl. Math. Comput.*, 163:553–563.
[6] M.A. Fariborzi Araghi and A. Fallahzadeh [2014], Inherited LU factorization for solving fuzzy system of linear equations , *Soft Comput.*, 17:159–163.
[7] M. Dehghan and B. Hashemi [2006], Iterative solution of fuzzy linear systems , *Appl. Math. Comput.*, 175:645–674.
[8] R. Ezzati [2011], Solving fuzzy linear systems , *Soft Comput.*, 15:193–197.
[9] A.N.A. Koam, M. Akram, G. Muhammad and N. Hussain [2020], LU Decomposition Scheme for Solving m-Polar Fuzzy System of Linear Equations , *Math. Probl. Eng.* Article ID 8384593.
[10] S. X. Miao, B. Zheng and K. Wang [2008], Block SOR methods for fuzzy linear systems , *J. Appl. Math. Comput.*, 26:201–218.
[11] S. H. Nasseri, M. Matinfar and M. Sohrabi [2009], QR-decomposition method for solving fuzzy system of linear equations , *Int. J. Math. Comput.*, 4:129–136.
[12] K. Wang and Y. Wu [2012], Uzawa-SOR method for fuzzy linear system , *International Journal of Information and Computer Science*, 1:36–39.
[13] K. Wang and B. Zheng [2006], Symmetric successive over relaxation methods for fuzzy linear systems , *J. Appl. Math. Comput.*, 175:891–901.
[14] K. Wang and B. Zheng [2007], Block iterative methods for fuzzy linear systems , *J. Appl. Math. Comput.*, 25:119–136.
[15] J. F. Yin and K. Wang [2009], Splitting iterative methods for fuzzy system of linear equations , *Comput. Math. Model.*, 20:326–335.
[16] Y. R. Wang and Y. L. Chen [2020], A modified Jacobi iterative method for large-size linear systems , *Journal of Shandong University (Natural Science)*, 55:122–126.
[17] Zhen Huang, Zhuo Chen, Shjun Zhang, Shiheng Wang, and Ke Wang [2023], A new method based on Jacobi iterative method for fuzzy linear systems , *Thai Journal of Mathematics*, 21:29–37.
[18] Chun Wen, Ting-Zhu Huang and Chao Wang [2011], Triangular and skew-symmetric splitting method for numerical solutions of Markov chains , *Computers and Mathematics with Applications*, 62:4039–4048.

IMPACT OF PREVENTIVE MAINTENANCE AND FAILURE RATE ON A COMPLEXLY CONFIGURED SYSTEM: A SENSITIVE ANALYSIS

Shakuntla Singla[1], Diksha Mangla[2*], Shilpa Rani[3*], Umar Muhammad Modibbo[4]

•

[1] Associate Professor (Mathematics), M.M.E.C.,Maharishi Markandeshwar(Deemed to be University), Mullana, Ambala, Haryana, India.
[2,3] Research Scholar (Mathematics), M.M.E.C.,Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India.
[4] Associate Professor, Department of Operation Research, Faculty of Physical Sciences, Modibbo Adama University, PMB 2076,Yola-NIGERIA

[1] shaku25@gmail.com, [2*] dikshamangla1995@gmail.com, [3*] gargshilpa46@gmail.com, [4]umarmodibbo@mau.edu.ng

Abstract

The availability of uninterrupted performance time has become essential for any industry seeking to maximize profits while incurring minimal maintenance costs. However, the system's components become weary as a result of the constant burden, resulting in decreased system efficiency and automatic full failure in the end. Complete failure is not always manageable; it might result in a significant loss of profit or productivity. In this regard, preventative maintenance is critical to ensuring that the industry runs smoothly, even with lower efficiency. Preventive maintenance is required in any sector to satisfy the demands of maximum profit and low cost for good output. This study examines the reliability of a complexly organized system of three units, A, B and C in order to determine its sensitivity to the effects of deteriorated rate and preventative maintenance rate over time. The three units are further made up of subunits which are in series or parallel configuration. The mathematical design work is based on the Markov process and the Laplace transformation. Different system parameters such as mean time to system failure, Available performance time, reliability, and profit, are analysed with respect to time and various rates. Further, A sensitivity analysis is used to explore how the rate of deterioration and preventative maintenance affects the system over time. Various malfunction and repair rates effect the system parameters in increasing or decreasing manner and sensitive analysis evaluated the impact of one unit on another or whole system. Here is a numerical example generated with the help of an appropriate model; the results are visually represented which concluded that with the passage of time reliability and other system parameters of system decreased under the influence of different rates. Utilizing the service cost, Profit is analysed which help to estimate the overall gain by the presented system. Also, by sensitive analysis it is concluded that out of three units A, B and C, Unit C has more effect as compared to B and C which is shown graphically. The purposed study can elaborate the profit after examined the reliability indices which become a key point for different industries like as diary plant ,fertilizer plant etc to have good outcomes with less maintenance cost.

Keywords: Preventive maintenance, Laplace Transformation, Relaibility and Sensitive Analysis .

I. Introduction

The availability and dependability of manufacturing machines determine the capacity to create any commodity. To have the most, quality output that meets demand is required. The highest level of

system dependability may be reached if all components perform well. However, it is not possible to operate at 100% efficiency for a lengthy period of time since certain system components wear down due to friction or other factors, even if they fail completely. rare systems' full failure cannot be fixed, resulting in a loss of money and, in rare cases, life. As a result, increasingly reliable systems require daily gains in loss minimization, which may be accomplished by performing the appropriate preventive or corrective maintenance methods. Improving dependability necessitates excessive preventative maintenance. Preventive maintenance conducted on a regular basis increases availability and profits. Creating an appropriate mathematical programming model has proven useful in system work. Numerous studies have been conducted in this topic.

Singla et al. [1], determined the plant's production, which comprises of four polytube units. The Chapman-Kolmogorov differential equations are solved using the method of extra variables and total performance time, thus reliability has been calculated using the likelihood of malfunctions and the number of repairs in a certain period of time. Naithani et al. [2], Using semi-Markov processes and the effects of cold standby, the dependability of induced draft fans in thermal plants was investigated, and it was determined that how the reliability and availability of the whole industry changes over time, as well as the impact of induces fans on them. The issue of assessing a banking server while considering a warm standby system was addressed Kumar and Goel, [3]. The reliable server has been derived by overcoming the effect of failure concept with the help of maintenance strategy. The lucrative idea for industry takes into account the appropriate choice of failure and repair rate. Employing computational models, the Markov method, the complementary variable technique, and the Laplace transform of the power plant's gas turbine system, Ram and Nagiya,[4], explored a number of reliability metrics, including mean time before failure, accessibility, reliability, and anticipated revenue of the system of the gas power plant and inspected out that which portion of the entire structure accomplished the reliability and MTTF more and also which component required greater scrutiny in order to have a growing profit using the sensitive analysis. According to Mohajan [5], the dependability of a measuring device is measured by how much one can trust the data produced by the instrument or how effectively any measuring apparatus corrects for the odd inaccuracy. The validity of a measuring instrument is determined by two factors: what it measures and how well it measures. Issue characteristics and problem-solving strategies continue to emphasize the reliability of a series-parallel system. Many research has also been undertaken in an attempt to improve these systems. Kumar et al. [6], have focused on the operational efficiency of a wiping unit used in the paper sector that employs Regenerative point graphical technique (RPGT) to analyse reliability characteristics, as well as the influence of failure and repair rates on reliability parameters. Yang and Tsao [7], explored a matrix-analytic technique to examine the dependability and availability of backup systems with operational escapes. They used a sensitivity analysis with the Laplace transformation to evaluate the MTTF and reliability function. The results show that increasing the number of spare components and maintenance frequency can increase system dependability. Tyagi et al. [8], conducted sensitivity analysis and reliability modeling on a flood alerting system (FAS) based on the Internet of Things (IoT). The authors employed a Markov method to calculate the probability of state change, which were then corrected using Laplace transformation. Kumar et al. [9], investigated the dependability of tripod turnstile machines operating in parallel configuration for extremely secure considerations utilizing the Laplace solver and the Markov idea. The sensitivity analysis for reliability has been investigated to determine the influence of one machine on another, and hence on the entire system. Modibbo et al. [10], offered two separate strategies to maximize system dependability under degradation in order to have lower costs for component maintenance while increasing profit. This work introduced a hybrid idea combining estimate and optimization theory. Reliability, availability, and maintainability of a threshing machine are three critical metrics described by Anchal et al. [11], and are highly useful in agriculture to get high reliability results. The influence of working units over time on one another and total productivity has been examined in order to make a profit in the agricultural industry. Particle swarm optimization, a nature-inspired algorithm, is being addressed to optimize the cost of rubber

cultivation, and negotiations about availability and other reliability parameters have been graphically depicted to understand the effect of failure and repair rate on overall system performance, Singla et al.,[12]. Shakuntla and Pooja, [13] devised the mathematical analysis of the Regenerative Point Graphical Technique (RPGT) to examine the reliability metrics and total profit benefit of the proposed system. The variation of failure and repair rates on reliability metrics has been provided, and the value of each rate is computed to determine which rate is best for increased output and profit. Shubham et al.'s analysis [14], concentrated on executing competence evaluation, validation, and optimization activities for the steam production system of a coal-fired thermal power plant to understand the influence of one component's performance rate on the other, and therefore on the entire system. Saini et al. [15] discussed the availability of a steam turbine plant, the exponential failure time behavior, and the arbitrary repair time behavior. The Particle swarm optimization with Genetic Algorithm approach was used to solve the differential equation and get various system metrics at several places during the system's malfunction and recovery, and the system's overall profit and efficiency were calculated. According to Khan et al.,[16], this study bridged the gap that was existing by employing the BLLP-Bi-level programming plan to handle the optimization problem concerning the liability of a system undergoing chosen maintenance.

A numerical demonstration of the Khun-Trucker approach with linear constraints is shown. This study examines how changes to reliability indices affected overall dependability and performance. Singla et al. [17], investigated a deep learning technique in order to maximize reliability characteristics, increase industrial revenues, and manufacture a 2:3 good system. Deep learning algorithms are compared to one another based on their availability. Singla et al. [19], study a failing system using a genetic algorithm to maximize the mean time to system failure and availability acquired by RPGT, as well as to determine the reliability metrics driven by degradation rate and preventative maintenance rate. Ahmadini et al. [20], conducted a study on dependability metrics under the influence of preventative maintenance using a heuristic method and an artificial bee colony algorithm to determine the impact of degraded rate and system failure. The analysis of poly-tube manufacturing plant has been done to discuss the availability regarding each unit of plant with the optimizing tool PSO by Singla et. al. [21].

Different research studies have counted the variation in various reliability attributes related to failure and repair parameters, but the Laplace transformation methodology, which accounts for time variation, enables the industry to operate for an extended duration. Additionally, the sensitivity analysis has revealed the following: The current project's purpose is to increase the dependability of a presentation system while keeping it operational throughout time and at variable rates. The purpose of this research is to focus on how the sensitivity of system units affects the reliability, accessibility, and profitability of complex unit arrangements under preventive maintenance which was studied in less content in previous year. The paper is organized as follows: Section 2 describes the model's characteristics, including the state overview, assumptions, notations, and model frame. Section 3 discusses the methodology behind mathematical simulations. The MTTF, available performance time, reliability, and profit analysis are among the primary topics covered in Section 4. Section 5 addresses the system's sensitivity to parameters and reliability. Section 6 concludes with the outcome discussion. Section 7 has the conclusion.

II. Model description

I. System description

For the sake of this study, the complete system is represented by a complicated configuration made up of three major units (A, B, and C) plus subunits with mixed configurations stacked in series as seen in many plants like fertilizer plant, yarn mill, Soap industry, soft drink plant etc. Figure 1 depicts how the system is organized with its units and subunits. Unit A can run at low efficiency and be restored to full efficiency with a single preventive maintenance treatment; however, it will

shut down after a second decline. There is no such situation for Unit C, which means it may undergo a catastrophic failure, whereas the Unit B component functions in parallel, causing it to collapse when both fail. The Markov process is used to generate the Champman-Kolmogorov differential equations. These equations can be solved using the concept of the Laplace transformation. There is a constructed mathematical model whose functioning is determined by the rates of failure, degradation, preventative maintenance, and corrective maintenance. Following a sensitivity analysis, the different metrics are calculated using the Laplace concept.

II. Assumptions

- At first, the system can execute its mission perfectly and efficiently, as if it were a new system.
- When a system degrades to a certain point, it loses part of its efficiency or becomes less efficient at accomplishing the task.
- After repair, the system operates as new, with each subcomponent taking the same amount of time to correct.
- Subunits A_1 and A_2 are assumed to have the same deterioration and failure rates. Components C_1, C_2, and C_3 have identical failure rates, as do components B_1 and B_2.
- It is assumed that failure and deterioration rates follow exponential distributions.

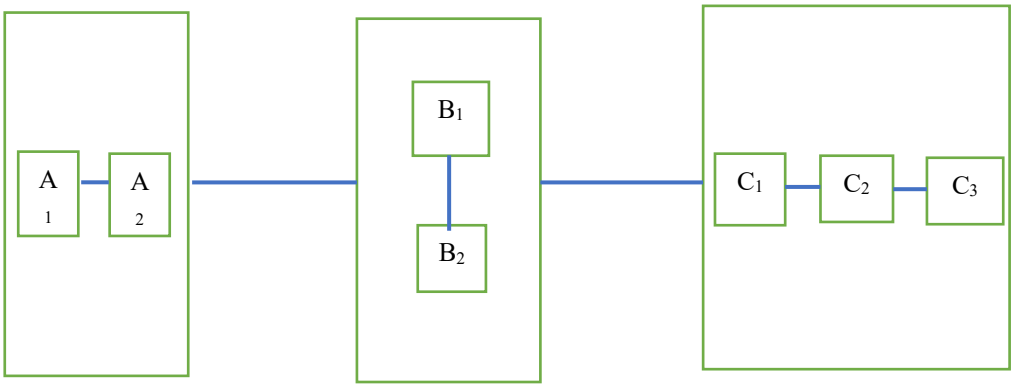


Figure 1: *System configuration*

III. Notations:

Table 1: *Various phrases pertaining to the work being presented.*

⬭ ⬠ ▭	Represent full functional state, reduced state and down state respectivly
MTTF	Mean time to system failure
Av(t)	Performance time of system over t
R(t)	The reliability of system over t
U_i	Represent transitions state where , $0 \le i \le 11$.
A	Represent 100% functional state of system
$\overline{A}$	The unit is reduced to some low efficiency due to some external or internal cause and need preventive maintenance.
A	The system is demised completely.
λ_j	Failure rate from excellent to malfunctioned state for Units B and C or from deteriorated to destroyed state For Unit A, where j = A, B, C
α_A	It is the degraded rate going from good condition to reduced condition for the unit A i.e. from A to $\overline{A}$.

μ_A	It represents the preventive maintenance rate to reversed the condition from reduced state to good state, only one time applicable for unit A.
M	Repair rate for taking back state from failed to good working state by applying corrective maintenance
$P_i(t)$	Represent the probability of various changing states where $0 \le i \le 11$.
P(t)	Represent the overall probability vector and its associated differential vector.
S	Laplace transformation variable
k_1	Revenue cost
k_2	Service cost

The potential state transition diagram for the model that is being presented can be seen as in Figure 2.

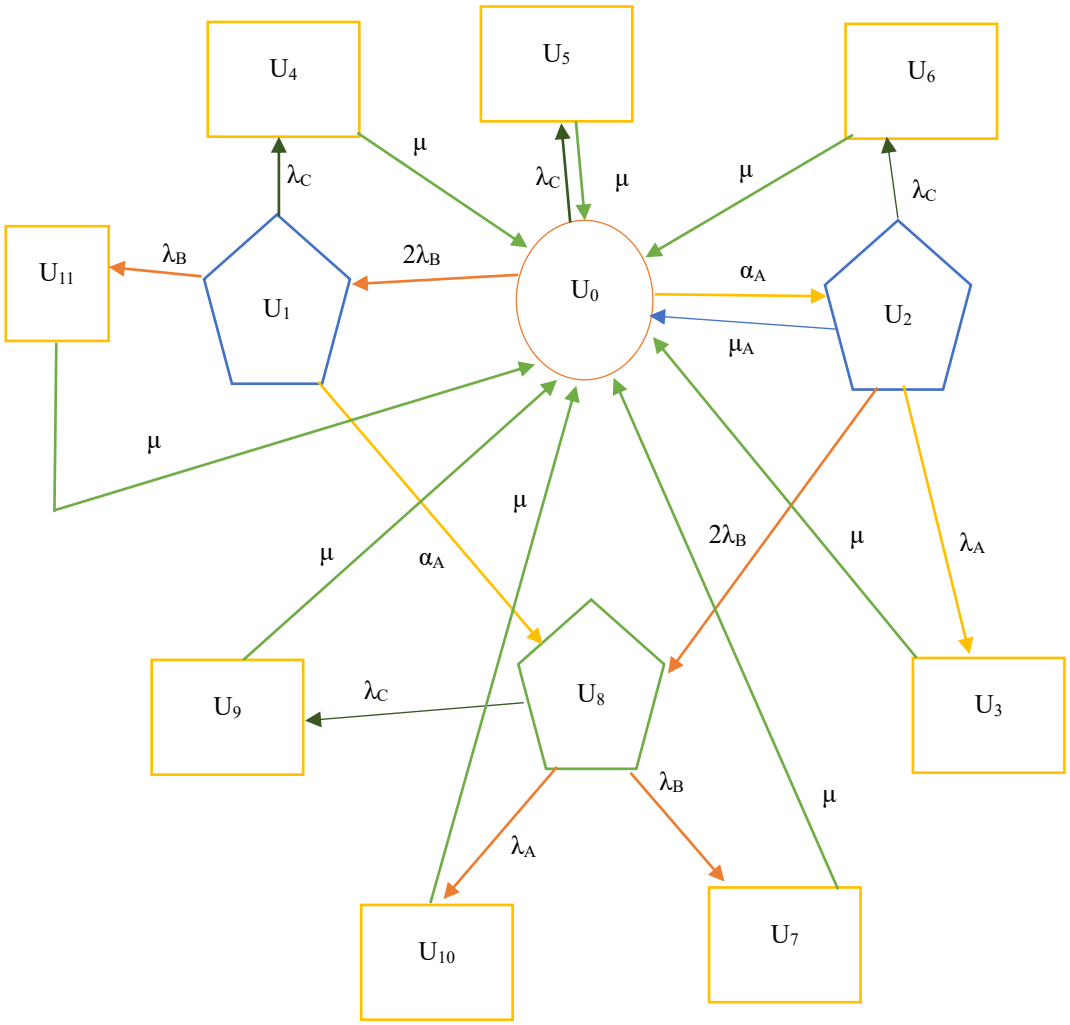


Figure 2: *The state transition diagram for different possibilities of working of three unit A,B,C.*

IV. State Description:

Table 2: *Various descriptions of states associated with the transition diagram.*

U_0	(A,B,C)	This is the full functional state with 100% efficiency.

U_1	(A,B(b),C)	One of the component of B is failed but 100% efficient state due to parallel arrangement of parts of unit B
U_2	($\bar{A}$,B,C)	Reduced state due to deteioration of unit A and efficiency of whole system reduced to some level and undergo preventive maintenance
U_3	(a,B,C)	Represting a breakdowned state due to complete breakdown of Unit A
U_4	(A,B(b),c)	It is a down state as Unit C breakdowned.
U_5	(A,B,c)	Breakdowned state as Unit C failed.
U_6	($\bar{A}$,B,c)	Down state as unit c malfunctioned fully while A is in reduced state and B is functional
U_7	($\bar{A}$,b,C)	Breakdown state as both component of unit B stop working while A is in reduced state and C is functional
U_8	($\bar{A}$,B(b),C)	whole condition refer to reduced state.
U_9	($\bar{A}$,B(b),c)	It is a down state as Unit C stopped while A is in reduced state and B is functional
U_{10}	(a,B(b),C)	Breakdowned state due to Unit A while other two are functional.
U_{11}	(A,b,C)	Breakdowned state due to Unit B while other two are functional

III. Mathematical Modelling of the Presented Model

The computational foundation for the mechanism being discussed incorporates the concept of the Markov process. The creation of first-order Chapman-Kolmogorov differential equations, which correspond to the numerous stable states depicted in the transition diagram, assists in the determination of reliability parameters. Pi(t) is the chance that the system will be in state Ui at time t ≥0. Furthermore, let P(t) be the probability vector at time t in hours given a starting condition.

$$P_i(0) = \begin{cases} 1 & if\ i = 0 \\ 0 & if\ i \neq 0 \end{cases} \quad (1)$$

The differentioal equations associated to Figure 2 are:

$$P'_0(t) = -(\alpha_A + \lambda_C + 2\lambda_B)P_0(t) + \mu_A P_2(t) + \mu\left(\sum_{i=3}^{11} P_i(t) - P_8(t)\right) \quad (2)$$

$$P'_1(t) = -(\alpha_A + \lambda_C + \lambda_B)P_1(t) + 2\lambda_B P_0(t) \quad (3)$$

$$P'_2(t) = -(\mu_A + 2\lambda_B + \lambda_A + \lambda_C)P_2(t) + \alpha_A P_0(t) \quad (4)$$

$$P'_3(t) = -\mu P_3(t) + \lambda_A P_2(t) \quad (5)$$

$$P'_4(t) = -\mu P_4(t) + \lambda_C P_1(t) \quad (6)$$

$$P'_5(t) = -\mu P_5(t) + \lambda_C P_0(t) \quad (7)$$

$$P'_6(t) = -\mu P_6(t) + \lambda_C P_2(t) \quad (8)$$

$$P'_7(t) = -\mu P_7(t) + \lambda_B P_8(t) \quad (9)$$

$$P'_8(t) = -(\lambda_A + \lambda_C + \lambda_B)P_8(t) + 2\lambda_B P_2(t) + \alpha_A P_1(t) \quad (10)$$

$$P'_9(t) = -\mu P_9(t) + \lambda_C P_8(t) \quad (11)$$

$$P'_{10}(t) = -\mu P_{10}(t) + \lambda_A P_8(t) \quad (12)$$

$$P'_{11}(t) = -\mu P_{11}(t) + \lambda_B P_1(t) \quad (13)$$

Equations (2) to (13) can be transformed using the Laplace transform, and the initial condition, i.e., equation (1), That is what we understand.

$$(s + \alpha_A + \lambda_C + 2\lambda_B)\bar{P}_0(s) = 1 + \mu_A \bar{P}_2(s) + \mu\left(\sum_{i=3}^{11} \bar{P}_i(s) - P_8(t)\right) \quad (14)$$

$$(s + \alpha_A + \lambda_C + \lambda_B)\bar{P}_1(s) = 2\lambda_B \bar{P}_0(s) \quad (15)$$

$$(s + \mu_A + 2\lambda_B + \lambda_A + \lambda_C)\bar{P}_2(s) = \alpha_A \bar{P}_0(s) \quad (16)$$

$$(s + \mu)\bar{P}_3(s) = \lambda_A \bar{P}_2(s) \quad (17)$$

$$(s + \mu)\bar{P}_4(s) = \lambda_C \bar{P}_1(s) \quad (18)$$

$$(s + \mu)\bar{P}_5(s) = \lambda_C \bar{P}_0(s) \quad (19)$$

$$(s + \mu)\bar{P}_6(s) = \lambda_C \bar{P}_2(s) \quad (20)$$

$$(s+\mu)\bar{P}_7(s) = \lambda_B \bar{P}_8(s) \quad (21)$$
$$(s+\lambda_A + \lambda_C + \lambda_B)\bar{P}_8(s) = 2\lambda_B \bar{P}_2(s) + \alpha_A \bar{P}_1(s) \quad (22)$$
$$(s+\mu)\bar{P}_9(s) = \lambda_C \bar{P}_8(s) \quad (23)$$
$$(s+\mu)\bar{P}_{10}(s) = \lambda_A \bar{P}_8(s) \quad (24)$$
$$(s+\mu)\bar{P}_{11}(s) = \lambda_B \bar{P}_1(s) \quad (25)$$

Now solving equation from (14) –(25) , We get the trasition state probabilities:

$$\bar{P}_0(s)=\frac{1}{\left[(s+C_1)-\left\{\frac{\mu_A\alpha_A}{(s+C_3)}+\frac{\mu}{(s+\mu)}\left(\frac{(\lambda_A+\lambda_C)\,\alpha_A}{(s+C_3)}+\frac{2(\lambda_B+\lambda_C)\lambda_B}{(s+C_2)}+\lambda_C+\frac{2\lambda_B\ \alpha_A}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\right)\right\}\right]} \quad (26)$$

$$\bar{P}_1(s) = \left(\frac{2\lambda_B}{s+C_2}\right)\bar{P}_0(s) \quad (27)$$
$$\bar{P}_2(s) = \left(\frac{\alpha_A}{s+C_3}\right)\bar{P}_0(s) \quad (28)$$
$$\bar{P}_3(s) = \left(\frac{\lambda_A\alpha_A}{s+C_3}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (29)$$
$$\bar{P}_4(s) = \left(\frac{2\lambda_B\,\lambda_C}{s+C_2}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (30)$$
$$\bar{P}_5(s) = \left(\frac{\lambda_C}{s+\mu}\right)\bar{P}_0(s) \quad (31)$$
$$\bar{P}_6(s) = \left(\frac{\alpha_A\lambda_C}{s+C_3}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (32)$$
$$\bar{P}_7(s) = \frac{2\lambda_B\,\lambda_B\ \alpha_A}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (33)$$
$$\bar{P}_8(s) = \left(\frac{2\lambda_B\ \alpha_A}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\right)\bar{P}_0(s) \quad (34)$$
$$\bar{P}_9(s) = \frac{2\lambda_B\,\lambda_C\ \alpha_A}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (35)$$
$$\bar{P}_{10}(s) = \frac{2\lambda_B\,\lambda_A\ \alpha_A}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (36)$$
$$\bar{P}_{11}(s) = \left(\frac{2\lambda_B\,\lambda_B}{s+C_2}\right)\left(\frac{1}{s+\mu}\right)\bar{P}_0(s) \quad (37)$$

Subsequently, the Laplace transformation of the system's up-state probability looked like this:

$$\bar{P}_{up}(s) = \bar{P}_0(s) + \bar{P}_1(s) + \bar{P}_2(s) + \bar{P}_8(s)$$
$$\bar{P}_{up}(s) = \left(1 + \left(\frac{2\lambda_B}{s+C_2}\right) + \left(\frac{\alpha_A}{s+C_3}\right) + \frac{2\lambda_B\ \alpha_A}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\right)\bar{P}_0(s) \quad (38)$$

Furthermore, the Laplace transformation is used to change the system's down-state probability as follows:

$$\bar{P}_{down}(s) = \sum_{i=3}^{11}\bar{P}_i(s) - \bar{P}_8(s)$$
$$\bar{P}_{down}(s) = \left[\begin{array}{c}\left(\frac{\lambda_A\alpha_A}{s+C_3}\right)\left(\frac{1}{s+\mu}\right) + \left(\frac{2\lambda_B\,\lambda_C}{s+C_2}\right)\left(\frac{1}{s+\mu}\right) + \left(\frac{\lambda_C}{s+\mu}\right) + \left(\frac{\alpha_A\lambda_C}{s+C_3}\right)\left(\frac{1}{s+\mu}\right) + \\ \frac{2\lambda_B\ \alpha_A C_4}{(s+C_4)}\left(\frac{1}{s+C_3}+\frac{1}{s+C_2}\right)\left(\frac{1}{s+\mu}\right) + \left(\frac{2\lambda_B\,\lambda_B}{s+C_2}\right)\left(\frac{1}{s+\mu}\right)\end{array}\right] \quad (39)$$

IV. Mathematical Measure of the presented Model

I. Availability Analysis (Av)

A system's availability may be summarized as the frequency of problems and the speed with which they are resolved. By using the different rate values as $\alpha_A = 0.01 = \mu_A$, $\lambda_C = 0.03, \lambda_B = 0.02, \mu = 0.04$ and $\lambda_A = 0.01$ in (38) and applying inverse Laplace transformation. The system's available performance time is expressed as follow:

$$\text{Av} = 0.1520e^{-0.1353t} + 0.0370e^{-0.0523t} + 0.4988 + 0.3122e^{-0.0712t}\cos(0.0463t) + 0.3184e^{-0.0712t}\sin(0.0463t) \quad (40)$$

Equation (40) allows us to change the time from 0 to 30 hours, resulting in numerical availability metrics, as shown in Table 3 and Figure 3.

Table 3: *Variation in Availability w.r.t time*

Time	Availability
0	1.0000
1	0.9708
2	0.9432
3	0.9170
4	0.8922
5	0.8686
6	0.8461
7	0.8248
8	0.8045
9	0.7852
10	0.7669
11	0.7494
12	0.7329
13	0.7172
14	0.7023
15	0.6881
16	0.6748
17	0.6622
18	0.6502
19	0.6390
20	0.6284
21	0.6184
22	0.6090
23	0.6003
24	0.5920
25	0.5843
26	0.5771
27	0.5703
28	0.5641
29	0.5582
30	0.5528

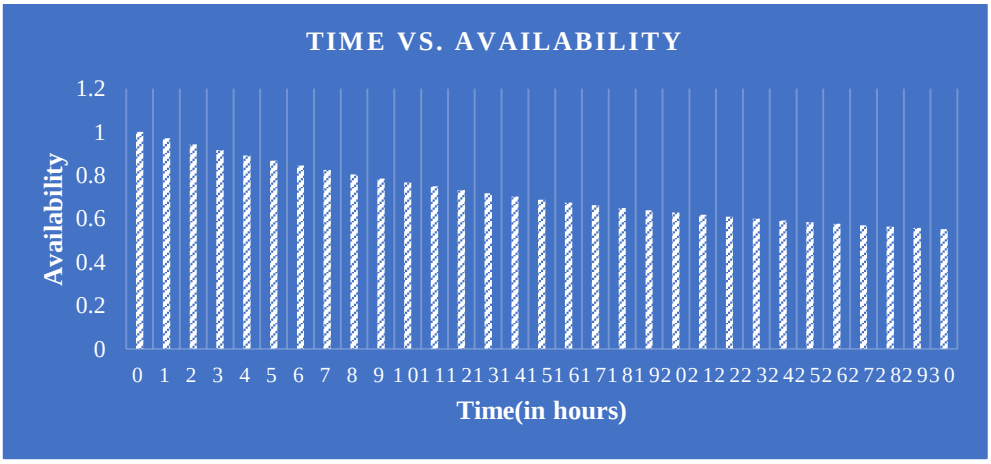


Figure 3: *The variation in Availability with variation in time.*

II. Mean time to system failure (MTTF) Analysis

MTTF calculates the average time projected before an initial malfunction, i.e. the length of system operation before the first failure. Using μ=0 and the limit approaching zero in (38), we may obtain MTTF as

$$\text{MTTF}=\lim_{s\to 0}\overline{P_{up}}(s)$$

$$\text{MTTF}=\left(\frac{1+\frac{2\lambda_B}{\alpha_A+\lambda_C+\lambda_B}+\frac{\alpha_A}{\mu_A+2\lambda_B+\lambda_A+\lambda_C}+\frac{2\lambda_B\ \alpha_A}{(\lambda_B+\lambda_A+\lambda_C)(\alpha_A+\lambda_C+\lambda_B)}+\frac{2\lambda_B\ \alpha_A}{(\lambda_B+\lambda_A+\lambda_C)(\mu_A+2\lambda_B+\lambda_A+\lambda_C)}}{(\alpha_A+\lambda_C+2\lambda_B)-\frac{\mu_A\alpha_A}{\mu_A+2\lambda_B+\lambda_A+\lambda_C}}\right) \quad (41)$$

Putting the value of $\alpha_A = 0.01$ and $\mu_A = 0.01$ and then varying the both one by one, respectively, from going 0.1 to 0.9 in (41), we get the values represented in Table 4 and Figure 4.

Table 4 : *Variation in MTTF with variation in degraded rate and preventive maintenance rate.*

Variable in α_A and μ_A	**MTTF**	
	α_A	μ_A
0.1	21.7817	25.1244
0.2	21.5087	25.2179
0.3	21.4242	25.2636
0.4	21.3577	25.2907
0.5	21.2619	25.3086
0.6	21.1956	25.3214
0.7	21.1471	25.3309
0.8	21.1101	25.3383
0.9	21.0809	25.3442

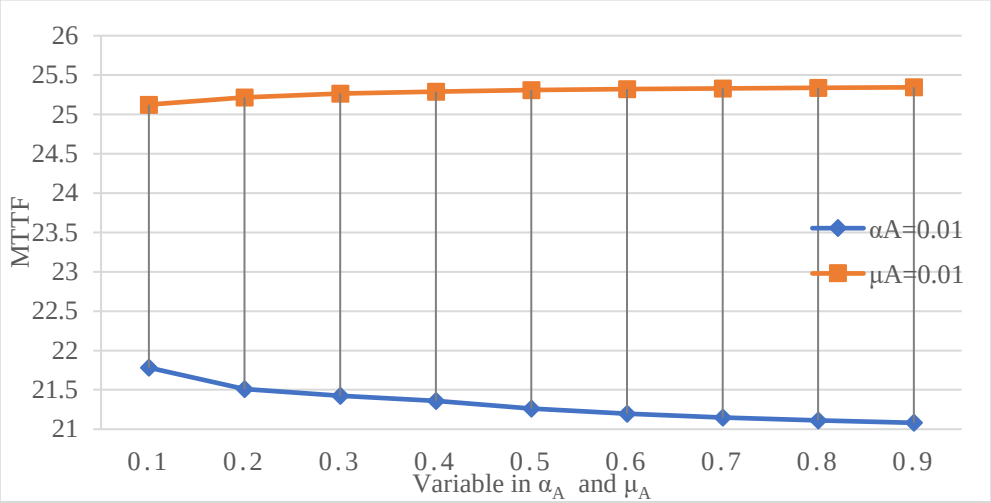


Figure 4: *Variation in MTTF with degraded and preventive maintenance rate*

III. Reliability Analysis

Reliability refers to working time without any failure in a given period of time. By using the different rate values as $\alpha_A = 0.01 = \mu_A, \lambda_C = 0.03, \lambda_B = 0.02$, λ_A =0.01 and recovery rate μ=0 in (38) and by taking inverse Laplace transformation, Reliability may be described as

$$R(t) = 0.1025e^{-0.0962t} - 0.7025\, e^{-0.0738t} + 1.6\, e^{-0.06t} \quad (42)$$

Varying time from 0 to 30 hours, the following data is obtained, depicted by Table 5 and represented graphically in Figure 5.

Table 5: *Variation in Reliability vs. time*

Time	Reliability
0	1
1	0.9474
2	0.8975
3	0.8503
4	0.8054
5	0.7629
6	0.7227
7	0.6845
8	0.6483
9	0.6140
10	0.5814
11	0.5506
12	0.5214
13	0.4937
14	0.4674
15	0.4425
16	0.4189
17	0.3966

18	0.3754
19	0.3553
20	0.3363
21	0.3183
22	0.3012
23	0.2851
24	0.2698
25	0.2552
26	0.2415
27	0.2285
28	0.2162
29	0.2045
30	0.1934

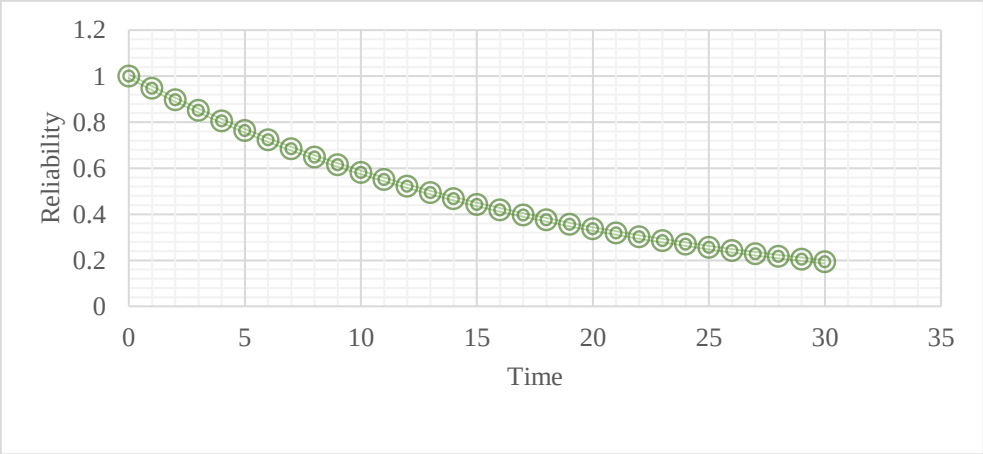


Figure 5: *Time vs. Reliability*.

IV. Profit Analysis

If the facility for service is assumed to have been readily available at all times, the expected profit is as follows, accounting for maintenance expenditures for the range [0, t].

$$\text{Profit} = k_1 \int_0^t P_{up}(t)dt - tk_2 \quad (43)$$

Using (38) and Having $k_1 = 1$ and $k_2 = 0.1, 0.2, 0.3, 0.4, 0.5$ respectively we get data which is depicted by the Table 6 and Graph 6.

Table 6: *Profit analysis with respect to time*

Time (t)	$k_2 = 0.1$	$k_2 = 0.2$	$k_2 = 0.3$	$k_2 = 0.4$	$k_2 = 0.5$
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.8853	0.7853	0.6853	0.5853	0.4853
2	1.7422	1.5422	1.3422	1.1422	0.9422
3	2.5722	2.2722	1.9722	1.6722	1.3722
4	3.3767	2.9767	2.5767	2.1767	1.7767
5	4.1570	3.6570	3.1570	2.6570	2.1570
6	4.9143	4.3143	3.7143	3.1143	2.5143

7	5.6497	4.9497	4.2497	3.5497	2.8497
8	6.3642	5.5642	4.7642	3.9642	3.1642
9	7.0590	6.1590	5.2590	4.3590	3.4590
10	7.7350	6.7350	5.7350	4.7350	3.7350

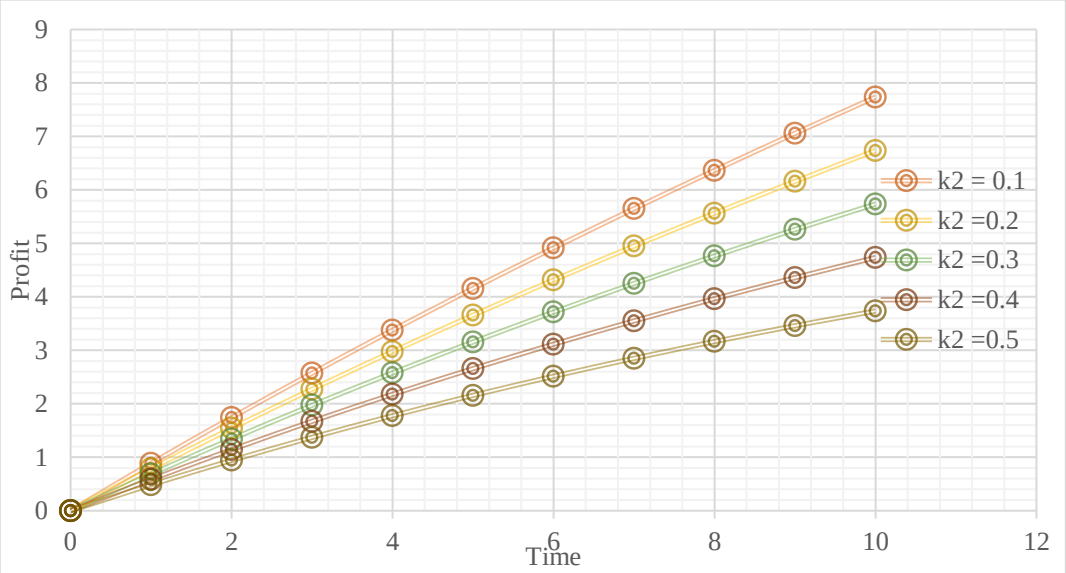


Figure 6: *Time vs. profit*

V. Sensitive Analysis

I. Sensitivity of Availability

The inverse Laplace equation(38) is differentiated to assess the sensitivity of availability, and then time is varies from 0 to 30 hours with respect to, λ_A,λ_B and λ_C one by one, as demonstrated by tabular form using Table 7 and graphically by Figure 7.

Table 7: *Sensitivity Analysis of System Availability*

	Reliability's	**Sensitivity**	
Time(t)	λ_A	λ_B	λ_C
0	0	0	0
1	-0.0979	-0.9856	-1.4892
2	-0.1903	-1.4804	-1.6677
3	-0.2690	-1.9190	-1.8259
4	-0.3360	-2.3066	-1.9654
5	-0.3926	-2.6480	-2.0877
6	-0.4405	-2.9473	-2.1940
7	-0.4807	-3.2086	-2.2858
8	-0.5143	-3.4355	-2.3643
9	-0.5424	-3.6313	-2.4305
10	-0.5656	-3.7990	-2.4856

11	-0.5848	-3.9414	-2.5306
12	-0.6005	-4.0612	-2.5665
13	-0.6132	-4.1604	-2.5940
14	-0.6235	-4.2412	-2.5665
15	-0.6317	-4.3060	-2.5940
16	-0.6383	-4.3560	-2.6140
17	-0.6433	-4.3931	-2.6274
18	-0.6473	-4.4186	-2.6346
19	-0.6502	-4.4339	-2.6337
20	-0.6525	-4.4403	-2.6266
21	-0.6541	-4.4387	-2.6017
22	-0.6552	-4.4303	-2.5848
23	-0.6560	-4.4160	-2.5655
24	-0.6565	-4.3965	-2.5442
25	-0.6568	-4.3725	-2.5211
26	-0.6569	-4.3449	-2.4966
27	-0.6570	-4.3140	-2.4709
28	-0.6571	-4.2806	-2.4442
29	-0.6571	-4.2451	-2.4169
30	-0.6571	-4.2079	-2.3892

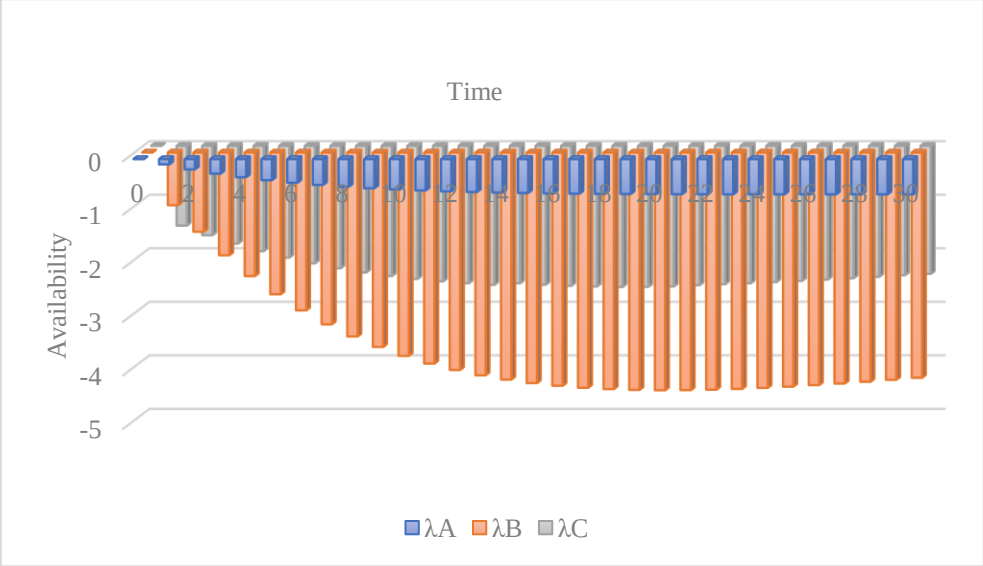


Figure 7: *Time vs. Availability 's sensitivity*

II. Sensitivity of MTTF

By differentiating equation (41) with respect to three failure rates, λ_A ,λ_B and λ_C one by one, repectively, then varing the value form 0.1 to 0.9 .

Table 8: *Sensitivity of MTTF*.

Variable in λ_A, λ_B and λ_C	**MTTF**

	λ_A	λ_B	λ_C
0.1	-4.7269	-70.4246	-80.1074
0.2	-1.9436	-24.8599	-23.1077
0.3	-1.0526	-12.5387	-10.6718
0.4	-0.6587	-7.5156	-6.0982
0.5	-0.4506	-4.9840	-3.9340
0.6	-0.3276	-3.5294	-2.7444
0.7	-0.2488	-2.6153	-2.0220
0.8	-0.1953	-2.0020	-1.5509
0.9	-0.1574	-1.5691	-1.2269

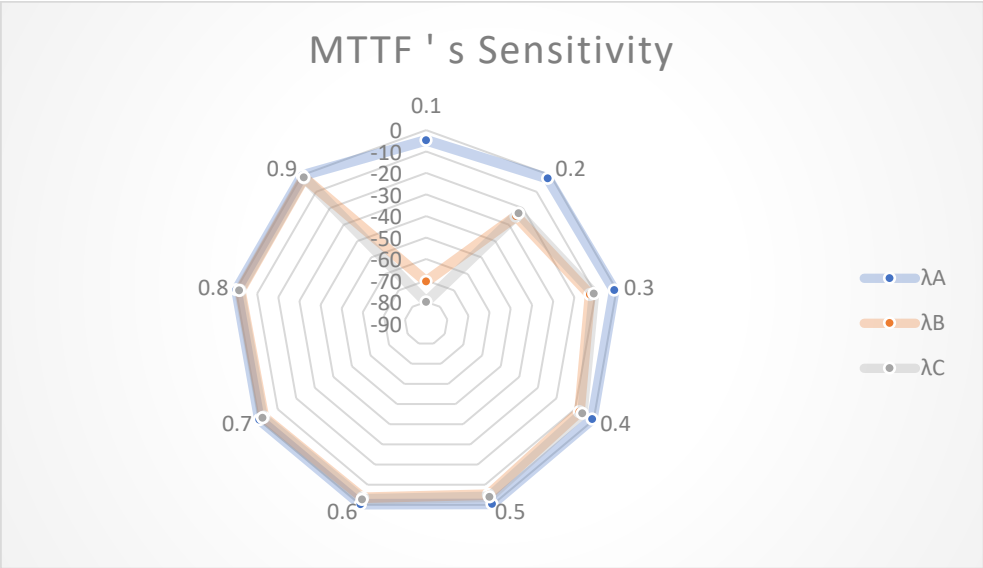


Figure 8: *MTTF's Sensitivity*

III. Sensitivity of Reliability

By taking inverse Laplace of (38) with value of μ=0, and differentiating equation with respect to three failure rates, λ_A ,λ_B and λ_C one by one, repectively, then by changing time from 0 to 30 hours. we get the Figure 9 and Table 9.

Table 9: *Reliability Sensitivity*

	Reliability's Sensitivity		
Time(t)	λ_A	λ_B	λ_C
0	0	0	0
1	-0.0046	-0.0557	-0.9696
2	-0.0169	-0.2071	-1.8775
3	-0.0351	-0.4328	-2.7226
4	-0.0574	-0.7149	-3.5048
5	-0.0825	-1.0379	-4.2245
6	-0.1094	-1.3889	-4.8829
7	-0.1371	-1.7569	-5.4812

8	-0.1649	-2.1329	-6.0215
9	-0.1922	-2.5092	-6.5057
10	-0.2185	-2.8797	-6.9361
11	-0.2435	-3.2395	-7.3152
12	-0.2669	-3.5847	-7.6455
13	-0.2886	-3.9120	-7.9296
14	-0.3084	-4.2192	-8.1702
15	-0.3263	-4.5046	-8.3699
16	-0.3421	-4.7671	-8.5314
17	-0.3560	-5.0060	-8.6572
18	-0.3679	-5.2211	-8.7498
19	-0.3779	-5.4122	-8.8117
20	-0.3861	-5.5798	-8.8454
21	-0.3925	-5.7244	-8.8529
22	-0.3973	-5.8466	-8.8367
23	-0.4005	-5.9473	-8.7986
24	-0.4023	-6.0273	-8.7409
25	-0.4027	-6.0877	-8.6652
26	-0.4019	-6.1296	-8.5735
27	-0.3999	-6.1540	-8.4673
28	-0.3969	-6.1621	-8.3484
29	-0.3930	-6.1549	-8.2181
30	-0.3882	-6.1336	-8.0779

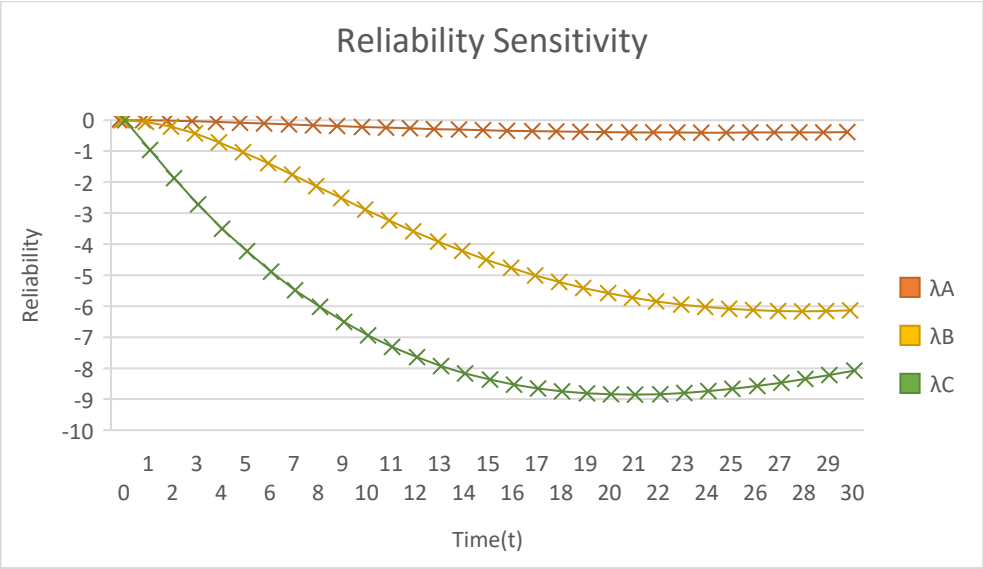


Figure 9: *Reliability's Sensitivity*

VI. Result and Discussion

The authors of the current study investigated the sensitivity analysis and reliability metrics of a multi-configuration complex system that incorporates failure rates, degradation, and preventative maintenance. The choice of parameters used in work is appropriate to have a good reliable system, Shivani et al. [18]. The framework's essential outlines are given below once the approach has been implemented.

Using Table 3 and Figure 3, the authors demonstrate how a system's available performance time evolves over time. After setting factors such as failure rate, repair rate, degradation rate, and preventive maintenance rate, availability drops with time as the chance of failure increases. After a long period, it becomes consistent and, to a lesser extent, continuous, affecting many portions of the system.

The MTTF is determined while accounting for system variances. This demonstrates that the MTTF of the system rises with the use of preventative maintenance (as depicted graphically in figure 4) and falls with regard to the deteriorated rate.

Figure 5 depicts the system's behavior, illustrating how dependability degrades over time and finally approaches zero. Table 5 displays the system's influence on reliability over time with fixed settings.

Figure 6 depicts an estimated profit analysis based on service cost and time. Critical inspection of the graphs reveals that the system creates a bigger profit over time; nevertheless, this profit decreases when service expenses grow.

In terms of failure rate, Table 7 and Figure 7 focus on the system's sensitivity analysis. The visual portrayal shows that the availability value decreases with time until t = 10 units, at which point it stabilizes. According to the analysis, the value declines until t = 23 units, at which point it climbs somewhat. When considering time, the availability value decreases until t = 21 units, then increases from t = 22 units to 30 units.

A critical examination of Figure 8 reveals that at a failure rate of 0.1 to 0.2 units, the sensitivity of the MTTF increases with some difference in value across all failure rates. As the failure rate climbed, so did the value of MTTF.

Figure 9 depicts how the system's dependability is affected by each of the three failure rates. As time passes, the value of dependability decreases until t = 28, at which point it increases, as opposed to initially declining until t = 25. This demonstrates that variation in the failure rate of unit A has the greatest impact on system dependability, followed by unit B.

VII. Conclusion

A complicated three-unit system with numerous configurations is investigated, as well as the sensitivity of dependability measurements. In this work, the reliability indices are obtained using the Laplace transformation after the system's equation has been created using the Champman-Kolmogorov differential approach and the Markov process notion. Many academic papers have been produced regarding series-parallel systems, but none of them have considered a complex system that employs preventative maintenance. These findings lead us to the following conclusions: as time passes, the framework's dependability and available performance time decline, and as failure rates grow, MTTF reduces as well. This emphasizes the importance of preventive maintenance and its significance in the choice of strategy to be used in the maintenance field to reduce the cause of degradation, as well as in identifying the unit with the highest failure rate that should be repaired as soon as possible to minimize loss.

A sensitivity investigation of the three units indicated that unit C is more sensitive than the other two in starting level, thus a maintenance concept is applied to it. However, when compared to

unit A, it is clear that unit B is more susceptible to failure rates. As a result, the system becomes incredibly lucrative as the failure rate is reduced through preventative and corrective maintenance. By taking care of units B and C, the researchers may improve productivity while requiring less maintenance, considerably increasing profit. This article offers research findings that highlight the importance of parameter and unit selection for engineers and designers in developing more lucrative and low-maintenance systems.

To design cost-effective systems in the future, writers might develop mathematical models that optimize dependability while decreasing cost and sensitivity. Another element pushing designers to reduce service costs is the study's use of preventative maintenance. A meta-heuristic technique may be used to maximize reliability and other elements of the system by creating a model containing the reliability function.

References

[1] Singla, S., Lal, A. K., Bhatia, S. S. and Singh, J. (2011). Reliability Analysis of Polytube Industry using Supplementary Variable Technique. *Applied mathematics and computation* ,3981-3992.

[2] Naithani, A., Parashar, B., Bhatia, P.K. and Taneja, G. (2013). Cost Benefit Analysis of a 2-Out -of -3 Induced Draft Fans System with Priority for Operation to Cold Standby over Working at Reduced Capacity. *Advanced Modelling and optimization*, 499-509.

[3] Kumar, R. and Goel, P. (2016). Availability analysis of banking server with warm standby. *International Journal of Statistics and Applied Mathematics*, 24-35.

[4] Ram, M. and Nagiya, K.(2017). Gas turbine power plant performance evaluation under key failures. *J. Eng. Sci. Technol.*,1871-1886.

[5] Mohajan, H.K (2017). Two criteria for good measurements in research: validity and reliability. *Ann. Spiru Haret Univ. Econ. Ser. 17*, 59–82.

[6] Kumar, A., Garg, D. and Goel, P. (2019). Mathematical modelling and behavioral analysis of a washing unit in a paper mill. *International Journal of System Assurance Engineering and Management*, 1639-1645.

[7] Yang, D.Y. and Tsao, C.L. (2019). Reliability and availability analysis of standby systems with working vacations and retrial of failed components. *Reliab. Eng. Syst. Saf.*,46–55.

[8] Bansal, S., Tyagi, S. L. and Verma, V. K. (2021). Performance modelling and availability analysis of A screw manufacturing plant. *Materials Today: Proceedings*, 1985-1988.

[9] Kumar, P. and Kumar, A. (2021). Tripod turnstile machines performance analysis for the system safety and security without considering simultaneous failures using reliability approach. *Int. J. Math. Eng. Manage. Sci*..

[10] Modibbo, U. M., Arshad, M., Abdalghani, O. and Ali, I. (2021). Optimization and estimation in system reliability allocation problem. *Reliability Engineering & System Safety*.

[11] Sharma, A., Ailawalia, P. and Singla, S. (2021). RAM(Reliability, Availability, Maintainability) of Threshing Machine in Agriculture. *Agriculture and Natural Resources*, 1057–1061.

[12] Kumari, S., Singla, S. and Khurana, P. (2022). Particles Swarm Optimization for Constrained Cost Reliability of a Rubber Plant. *Life Cycle Reliability and Safety Engineering*.

[13] Singla, S. and Pooja (2022). Mathematical Analysis of Regenerative Point Graphical Technique (RPGT). *A mathematical analysis and contemporary applications*, 49–56.

[14] Singla, S., Verma, S., Malik, S. and Gupta, A. (2022). Perform ability evaluation, validation and optimization for the steam generation system of a coal-fired thermal power plant. *Method X.*

[15] Saini, M., Gupta, N., Shankar, V. G. and Kumar, A. (2022). Stochastic modeling and availability optimization of condenser used in steam turbine power plants using GA and PSO. *Quality and Reliability Engineering International*, 2670-2690.

[16] Khan, M. F., Modibbo, U. M., Ahmad, N. and Ali, I. (2022). Nonlinear optimization in bi-level selective maintenance allocation problem. *Journal of King Saud University – Science*, 1-8.

[17] Singla, S., Rani, S., Modibbo, U. M. and Ali, I. (2023). Optimization of System Parameters of 2:3 Good Serial System using Deep Learning. *Reliability Theory and Applications*, 670-679.

[18] Shivani, Ram, M., Goyal, N. and Kumar, A. (2023). Analysis of series systems sensitivity in context of components failures. *RAIRO Operations Research*, 2131-2149.

[19] Singla, S., Mangla, D., Panwar, D. and Taj, S. Z. (2024). Reliability Optimization of a Degraded System under Preventive Maintenance using Genetic Algorithm. *Journal of Mechanics of Continua and Mathematical Sciences*, 1-14.

[20] Ahmadini, A. A. H., Singla, S., Mangla, D., Modibbo, U. M., Rani, S. (2024). Reliability Assessment and Profit Optimization of Multi-unit Mixed Configured System using ABC Algorithm under Preventive Maintenance. *IEEE Access*, 85190-85202.

[21] Singla, S., Mangla, D., Modibbo, U. M. and Lal, A. K. (2024). Critical analysis of failure and repair rates of poly-tube manufacturing plant using PSO. *Reliability Theory and Applications,* 351-364.

ENHANCING PRECISION IN STRATIFIED SAMPLING USING MATHEMATICAL PROGRAMMING APPROACH

Mushtaq A. Lone[1], S. A. Mir[2], Kaisar Ahmad[3], Aafaq A. Rather[4,*], Danish Qayoom[5], S. Ramki[6]

•

[1, 2]SKUAST-Kashmir, India
[3]Department of Statistics, University of Kashmir, Srinagar, J&k, India
[4*, 5]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[6]Department of Community Medicine, Dhanalakshmi Srinivasan Institute of Medical Sciences and Hospital, Thuraiyur road, Perambalur-621212, India
[1]lonemushtaq11@gmail.com, [2]mir_98@msn.com, [3]ahmadkaisar31@gmail.com, [4*]aafaq7741@gmail.com, [5]danishqayoom11@gmail.com, [6]ramki.stat24@gmail.com

Abstract

This article addresses the challenges of determining the optimal allocation of sample sizes in stratified sampling design to minimize the cost function. Researchers employed the iterative procedure of Rosen's Gradient projection method and obtained optimal allocation of non-linear programming problem through manual calculation, which are often susceptible to human errors, such as rounding or arithmetic mistakes especially for complex nonlinear programming problems. R software performs calculations with high precision and consistency. In this paper, we demonstrate how to solve the non-linear programming problem by using iterative based procedure of Rosen's Gradient projection method through R software.

Keywords: Stratified random sampling**,** Optimal allocation, Gradient project method, Nonlinear programming problems

1. Introduction

Stratified sampling is widely utilized statistical method across various fields of scientific research, aimed enhancing the accuracy of estimates by reducing heterogeneity among population units. This is accomplished through a process known as stratification, where the entire population is segmented into distinct sub populations referred as strata. These strata are typically formed based on factors like administrative classifications, geographic locations and additional characters, ensuring they are non-overlapping and collectively they encompass the entire population. These strata are made to be homogeneous within and heterogeneous between. Once the strata established, samples are independently drawn from each stratum. The key challenge in stratified sampling is the determination of optimal allocation of sample sizes within each stratum, which can either aim to minimize the variance while adhering the cost or minimize cost while maintaining the variance. Thus, the problem of optimally selecting these the sample sizes is known as the optimal allocation problem, first addressed by Neyman [19] with further contributions by Cochran

[5], Sukhatme et al. [22] and Thompson [23]. The allocation problem of distribution becomes more difficult in in many studies, because an allocation optimal for one characteristic may not be optimal/suitable for others. Various researchers, including Wywial [24], Bethel [4], kreienbrock [13], Khan et al. [14, 15], Kozak [16], Ghosh [10], Yates [25], Aoyama [1], Hartley [12], Folks and Antle [9], Gren [11], Chatterjee [6], Ansari et al. [2], Chromy [7], have explored compromise allocations that suit multiple characteristics. The optimal allocation is characterized as a non-mathematical programming problem, the objective function being the variance subject to a cost constraint, or vice versa. This problem is solved using the Lagrange multiplier method, see Sukhatmeh et al [22] or the Cauchy-Schwarz inequality, see Cochran [5] for univariate case and Arthanari and Dodge [3] for multivariate one, both from deterministic point of view.

Dalenius [8] proposed a graphical solution for the problem involving two characteristics. Kokan and khan [17] demonstrated the existence and uniqueness of the solution and have given the optimal solution through iterative procedure. Chatterjee [6] developed an algorithm to solve the problem. In 1960, Rosen [20] developed the Gradient Projection method for linear constraints and later Rosen [21] in 1961, generalized it for nonlinear constraints. It uses the projection of the negative gradient in such a way that improves the objection function and maintains feasibility. In this paper, objective to determine the optimal allocation of sample sizes using Rosen's [20, 21] Gradient projection method through R software instead of using manual calculations. Lone et al [18] employed the same iterative procedure of Rosen's [20, 21] Gradient projection method and obtained optimal allocation of non-linear programming problem through manual calculation. Manual procedure might rely on a simplified or less robust version of an optimization algorithm. Performing a sufficient number of iterations manually to reach the optimal allocation is a challenging task due to time constraints or computational limits which may sometime provide less accurate results and involves iterative calculations in case of complex problems. R software provides a more reliable and accurate approach to solving complex optimization problems, explaining the difference in optimal allocation compared to manual procedures with high precision and consistency.

2. Formulation of the problem

Assume that there are p characteristics under study, with Y_j being the jth characteristic considered.

$$\bar{y}_{ij} = \frac{1}{n_i}\sum_{h=1}^{n_i} y_{ijh} \quad \text{for all } i = 1, 2, 3, ..., L \text{ and } j = 1, 2, 3, ..., p \tag{1}$$

Where y_{ijh} is the observed value for Y_j in the i^{th} stratum for the h^{th} sample unit.

Then $\bar{y}_j(st) = \frac{1}{N}\sum_{i=1}^{L} N_i \bar{y}_{ij}$ is an unbiased estimate of population mean $\bar{Y}$.

$$V(\bar{y}(st)) = \frac{1}{N}\sum_{i=1}^{L} N_i \bar{y}_{ij} = \frac{1}{N^2}\sum_{i=1}^{L} N_i{}^2 V(\bar{y}_{ij}) = \sum_{i=1}^{L} W_i{}^2 S_{ij}^2 x_i \tag{2}$$

Where

$$W_i = \frac{N_i}{N}; S_{ij}^2 = \frac{1}{N_i - 1}\sum_{h=1}^{N_i}\left(y_{ijh} - \bar{Y}_{ij}\right)^2 \text{ and } x_i = \frac{1}{n_i} - \frac{1}{N_i}$$

Let $a_{ij} = W_i S_{ij}^2$. Also, let C_i be the cost of sampling all the p characteristics on a single unit in the i^{th} stratum. The total variable cost of the survey assuming linearity is $C = \sum_{i=1}^{L} C_i n_i$. Assume that $a_{ij}, C_{i.} > 0$ for all $i = 1, 2, 3, ..., L$ and $j = 1, 2, 3, ..., p$.

In this context, the challenge of deriving statistical information about the population characteristics using sample data, this can be framed as an optimization problem, where we aim to determine optimum allocation of sample size n_i for all $i = 1, 2, 3, ..., L$ to minimize the survey cost is minimized. The multivariate sample design and its optimization are approached as a mathematical programming problem as discussed by Arthanari and Dodge [3]. Therefore, the allocation problem is defined accordingly, following the work of Sukhatme*et al*. [22] and Arthanari and Dodge [3].

$$\text{Minimize} \sum_{i=1}^{L} C_i n_i \text{ subject to } \sum_{i=1}^{L} a_{ij} x_i \le v_j \text{ and } 0 \le x_i \le 1 - \frac{1}{N_i} \tag{3}$$

Where v_j is the allowable error in the estimate of the j^{th} characteristics. The problem (3) can be equivalently written as

$$\text{Minimize} \sum_{i=1}^{L} \frac{C_i}{X_i} \text{ subject to } \sum_{i=1}^{L} a_{ij} X_i \le v_j \text{ and } \frac{1}{N_i} \le X_i \le 1 \tag{4}$$

Since N_i, are given, it is sufficient to minimize $\sum_{i=1}^{L} \frac{a_{ij}}{n_i}$. Where $X_i = \frac{1}{n_i'}$ and $\frac{C}{X_i}$ is strictly convex for $C_i > 0$ because of this objective function is strictly convex and the set of constraints provides a bounded convex feasible region and an optimal solution will also exit. Although the method has been described by Rosen for a general non-linear programming problem, its effectiveness is confined primarily to problems in which the constraints are all linear. The procedure involved in the application of the gradient projection method can be described in the following Algorithm.
The formulated non- linear programming model has been taken from Lone *et al.* [18].

$$\text{Minimize} = \frac{3}{X_1} + \frac{4}{X_2} \tag{5}$$

Subject to

$$0.36 X_1 + 3.24 X_2 \le 0.30$$
$$0.81 X_1 + 8.12 X_2 \le 0.60$$
$$0.09 X_1 + 9.92 X_2 \le 0.50$$
$$\frac{1}{180} \le X_1 \le 1$$
$$\frac{1}{270} \le X_2 \le 1$$

It is also assumed that, the variance of the estimate for each character cannot be greater than the specified limit i.e.

$$V_1 \le 0.30,\ V_2 \le 0.60,\ \textit{and}\ V_3 \le 0.50.$$

The solution of the above NLLP is obtained from R software through Rosen's Gradient Projection method.

3. Results

The optimal allocation obtained through R software using Gradient Projection method is $X_1 = 7$ and $X_2 = 20$ and optimal Value of the objective function is 103. The optimal solution using the same method through manual calculation is $X_1 = 4.0$ and $X_2 = 21$ and Value of the objective function is 96.

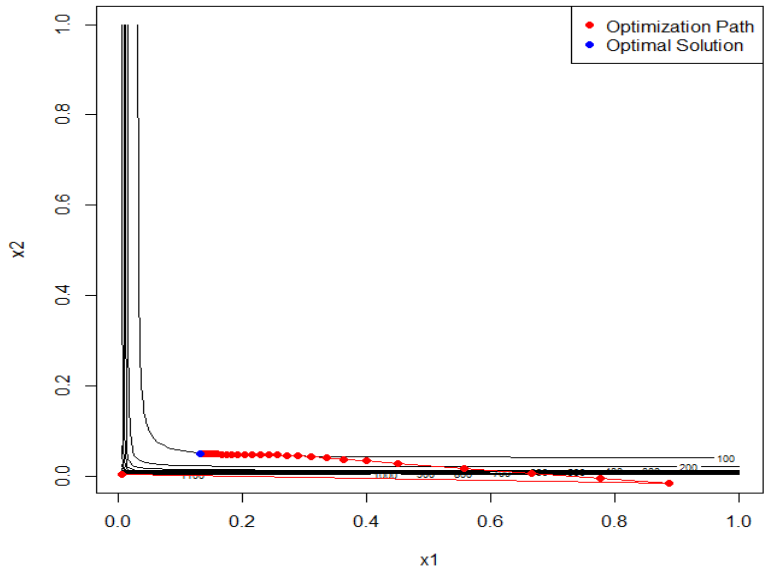


Figure 1: *Contour plot with optimization path*

4. Conclusion

This article highlights the complexities and potential for human error by calculation manually the optimal allocation of sample sizes using Gradient Project method in stratified sampling design. This study successfully demonstrates the use of R software to solve the NLPP for optimal allocation which shows significant improvements in precision and efficiency compared to manual calculation.

References

[1] Aoyama, H. (1963). Stratified random sampling with optimum allocation for multivariate populations. *Annals of the Institute of Statistical Mathematics*, 14, 251–258.

References

[1] Aoyama, H. (1963). Stratified random sampling with optimum allocation for multivariate populations. *Annals of the Institute of Statistical Mathematics*, 14, 251–258.

[2] Ansari, A. H., Najmussehar, and Ahsan, M. J., (2009). On multiple response stratified random sampling design.*International Journal of Statistical Sciences*, 1(1), 45-54.

[3] Arthanari, T. S., and Dodge, Y. (1981). Mathematical programming in statistics. *New York, NY: John Wiley.*

[4] Bethel, J. (1989). Sample allocation in multivariate surveys. *Survey Methodology*, 15, 40-57.

[5] Cochran, W. G. (1977). Sampling techniques (3rd ed.). *New York, NY: John Wiley & Sons.*

[6] Chatterjee, S. (1968). Multivariate stratified surveys. *Journal of the American Statistical Association*, 63, 530-535.

[7] Chromy, J. R. (1987). Design optimization with multiple objectives. *In Proceedings of the*

Survey Research Methods Section, 194–199, Alexandria, VA: American Statistical Association.

[8] Dalenius, T. (1957). Sampling in Sweden: Contributions to the methods and theories of sample survey practice. *Stockholm, Sweden: Almqvist & Wicksell.*

[9] Folks, J. K., and Antle, C. E. (1965). Optimum allocation of sampling units to the strata when there are R responses of interest. *Journal of the American Statistical Association,* 60, 225-233.

[10] Ghosh, S. P. (1958). A note on stratified random sampling with multiple characters. *Calcutta Statistical Bulletin,* 8, 81-89.

[11] Gren, J. (1966). Some application of non-linear programming in sampling methods. *Przeglad Statystyczny, 13, 203–217 (in Polish).*

[12] Hartley, H. O. (1965). Multiple purpose optimum allocation in stratified sampling. *In Proceedings of the American Statistical Association, Social Statistics Section, 258-261, Alexandria, VA: American Statistical Association.*

[13] Kreienbrock, L. (1993). Generalized measures of dispersion to solve the allocation problem in multivariate stratified random sampling. *Communications in Statistics: Theory and Methods,* 22(1), 219-239.

[14] Khan, M. G. M., Ahsan, M. J., and Jahan, N. (1997). Compromise allocation in multivariate stratified sampling: An integer solution. *Naval Research Logistics*, 44, 69-79.

[15] Khan, M. G. M., Ahsan, M. J., and Jahan, N. (2003). An optimal multivariate stratified sampling using dynamic programming. *Australian & New Zealand Journal of Statistics,* 45(1), 107-113.

[16] Kozok, M. (2006). On sample allocation in multivariate surveys. *Communications in Statistics - Simulation and Computation,* 35, 901-910.

[17] Kokan, A. R., and Khan, S. U. (1967). Optimum allocation in multivariate surveys: An analytical solution. *Journal of the Royal Statistical Society, Series B,* 29, 115-125.

[18] Lone, M.A., Mir, S. A., Khan, I., and Wani, M. S. (2017). Optimal allocation of stratified sampling design using Gradient Projection method. *Oriental Journal of Computer Science and Technology,* 10(1), 11-17.

[19] Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society,* 97, 558-625.

[20] Rosen, J. B. (1960). The gradient projection method for nonlinear programming, part I, linear constraints. *SIAM Journal of Applied Mathematics,* 8, 181-219.

[21] Rosen, J. B. (1961). The gradient projection method for nonlinear programming, part II, nonlinear constraints. *SIAM Journal of Applied Mathematics,* 9, 514-553.

[22] Sukhatme, P. V., Sukhatme, B. V., & Sukhatme, C. (1984). Sampling theory of surveys with applications.*Ames, IA: Iowa State University Press.*

[23] Thompson, M. E. (1997). Theory of sample surveys. *London, England: Chapman & Hall.*

[24] Wywial, J. (1988). Minimizing the spectral radius of means vector from sample variance-covariance matrix sample allocation between strata. *Prace Naukowe Akademii Ekonomicznej we Wroclawiu,*404, 223–235 (in Polish).

[25] Yates, F. (1960). Sampling methods for censuses and surveys (3rd ed.). *London, England: Charles Griffin.*

SAMPLE SIZE DETERMINATION PROCEDURES IN CLINICAL TRIALS: A COMPARATIVE ANALYSIS FOR RELIABLE AND VALID RESEARCH RESULTS

Faizan Danish[1], G.R.V. Triveni[2], Rafia Jan[3], Aafaq A. Rather[4,*], Danish Qayoom[5], Kaiser Ahmad[6]

•

[1,2] Department of Mathematics, School of Advanced Sciences, VIT-AP University, Inavolu, Beside AP Secretariat, Amaravati AP-522237, India
[3]Department of Statistics, Government Degree College Bejbehara, Anantnag, JandK, India,
[4*, 5]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[6]Department of Statistics, University of Kashmir, Srinagar, J&K, India
[1]Email: danishstat@gmail.com,[2]Email: trivenigullinkala@gmail.com, [3]Email: rafiajan836@gmail.com,
[4,*] Corresponding author: aafaq7741@gmail.com, [5] Email: danishqayoom11@gmail.com,
[6]Email: ahmadkaisar31@gmail.com

Abstract

Accurate sample size determination is paramount in clinical trials assuring the consistency and validity of research studies. This comparative analysis delves into the various procedures employed for sample size estimation in clinical trials and assesses their effectiveness in producing reliable results. By numerous formulaes and methods, this study seeks to identify best practices for optimizing sample sizes, thereby enhancing the statistical power of clinical trials. This research paper aims to conduct a comparative analysis of different formulae commonly employed in determining sample sizes evaluating their strengths, limitations, and applicability across various research scenarios. Several formulae have been considered with varying parameters, and the sample size was calculated and presented in different graphs.

Keywords: Sample Size, Population estimation, Sensitivity, Specificity, Experimental design

1. Introduction

Sample size refers to the volume of observations, units, or individuals included in a study,and the Sample size is a crucial step while doing experiments like Small-scale experiments, large-scale experiments, pilot studies, clinical trials, etc. A large sample size gives accurate and reliable outcomes and boosts statistical test precision. The test can detect an actual effect. We need to consider some important components in finding the exact sample size, including population size, variability of the data, desired level of precision, confidence level, research objectives, types of data, and analysis and resources available. Generally, a larger population indicates the least variability in the data. So, if the population is large, the sample should also be large to give the desired level of accuracy. The spread of the data points increases with the degree of data dispersion. To eradicate the problem of variability we have to use appropriate Statistical techniques. The level of precision specifies the closeness of the

sample estimate and population estimate. In this process, the discrepancy between the sample and the actual population is referred to as the sampling error. We can add or subtract 3% from the value in a survey if the sampling error is ±3%. For instance, 77% of students in a district get distinction marks in an exam, we can say, that between 74% and 80% of students get distinction marks. The reliability level tells us the probability that the target population parameters fall inside a scope of values. It is expressed in percentages like 90%, 95%, or 99%. The greater level of certainty in the results depends on a higher confidence level. Proper Research objectives should be defined clearly. The type of data and analysisare important aspects in the selection of a sample. The study also heavily weighs resources like personnel, money, and time. Therefore, it is crucial to use the resources at our disposal to achieve the appropriate degree of accuracy in our outcomes.

The sample size is considered an important factor when designing experiments or conducting clinical trials. Here few authors expressed their views in the nail down the difficulty of the preference of sampling size in the medical field. Halperin*et al.* [11] and Brown *et al.* [2] by utilizing the values of alpha and beta errors, found the size of a sample. Morse [21] described which factors need to be considered in the determination of sample size. Lenth [20] clearly described different aspects that lead to findinganappropriate sample size and he studied the power of a test of hypothesis to identifyaneffective sample size. Tatiana [34] explained the importance of the selection of samples and provided books to estimate study sizes like Fliess and Bland, and also provided software programs like EpiInfo and nQuery. Eng [8] elaborated on the essentiality of sample size, how to minimize sample size, and parameters that need to be considered and provided sample size formulas for comparative along with descriptive studies. Sakpal [31] calculated sample size by comparing two proportions and two means and gave basic rules for estimating sample size. Rohrig *et al.* [30] compared two drugsusing a t-test and explained the sample size principles,which require the power and rejection region of the statistical test. Ahmad *et al.* [1] suggested formulas for single and two ratios and took one instance from medical research to find an adequate sample size. Further, several authors have contributed towards the sample size determination size such as Lachenbruch [19], Obuchowski [26], Wittes [35], Whitley and Ball [36], Williamson [37], Pezeshk [27], Julius and Patterson [17], Chadha [5], Brutti *et al.* [3], Hertzog [12], Willan and Kowgier [38], Charles *et al.* [6]. Prajapati *et al.* [28], Mason, M. [22], Hajian-Tilaki [13], Burmeister and Aitken [4], Pourhoseingholi *et al.* [29], Hajian-Tilaki [15], Juneja and Sharma [18], Sami *et al.* [33], Negida *et al.* [23], Greene [10], Nanjundeswaraswamy and Divakar [25]. Hayat [14] and Singh and Masuku [32] clearly explained the significance of sample size and interpreted the key concepts involved in the establishment of sample size. Charan and Biswas [7] rationalized the effectiveness of sample size and provided information on different methods of sample size calculations for different studies and gave formulas accordingly. Negida [24] justified the sample size calculation in clinical studies by applying a software called StatsDirect to identify sample size using a correlation coefficient between two variables. We go through in detail of the following research papers.

In this paper, we have comprised several ways to estimate the sample size for any study. Glenn D. Israel [9] explained the criteria for concluding the sample size, which is affected by factors like thestudy's objective, population size, the problem of choosing an inappropriate sample, and Bias in sampling. He reviewed the accuracy, confidence, and variability levelsbriefly, which are very important points in judging the sample size. It also explains different proposals for resolving sample size issues.

2. Different approaches for estimation

The formula for computing a sample for ratios (Cochran1963:75)

$$n_0 = \frac{Z^2 pq}{e^2} \tag{1}$$

where,
n_0 = sample size
Z^2 = Normal curves abscissa
p = Estimated prevalence rate
q = 1- estimated prevalence rate

The Finite population correction for ratios is

$$n = \frac{n_0}{1+\frac{n_0 - 1}{N}} \tag{2}$$

Also, the Simplified formula for ratios (Yamane 1967:886):

$$n = \frac{N}{1+N(e)^2} \tag{3}$$

where,
n = sample number
N = Population count
e = Level of accuracy
Furthermore, the sample size formula for the mean is

$$n_0 = \frac{Z^2 \sigma^2}{e^2} \tag{4}$$

where,
σ^2 = variance of an attribute in the population
The following table 1 presents the sample size estimation at different values of related parameters namely sample proportions, the finite population correction, the streamlined formula for proportions, and samples for the mean, sample size estimate is performed for various values of the unknowns.

Table 1: *Sample size estimation at different values of related parameters*

p	0.99	0.93	0.88	0.85	0.77	0.73	0.69	0.63	0.61	0.59	0.56	0.46	0.54	0.49	0.5
q	0.01	0.07	0.12	0.15	0.23	0.27	0.31	0.37	0.39	0.41	0.44	0.54	0.46	0.51	0.5
N1	15	100	162	196	272	303	329	358	366	372	379	382	382	384	384
N2	14	83	134	164	224	252	277	304	315	324	333	339	349	357	361
N	250	500	750	1000	1250	1500	1750	2000	2250	2500	2750	3000	4000	5000	6000
N3	154	222	261	286	303	316	326	333	340	345	349	353	364	370	375
$\boldsymbol{\sigma^2}$	0.3	0.37	0.41	0.49	0.53	0.59	0.61	0.68	0.71	0.77	0.80	0.84	0.90	0.96	0.99
N4	138	210	258	369	432	535	572	711	775	911	983	1084	1245	1416	1506

Where,
N = Population count

p = estimated proportion of an attribute
$q = 1 - p$
N_1 = Sample for ratios
N_2 = finite population correction for ratios
N_3 = stream lined formula for proportions
N_4 = Sample for the mean

Baoliang Zhong [39] explained the sample size assessment in a Randomized controlled trial (RCT). The gold standard for establishing whether a treatment has a meaningful impact is to analyze the results of two groups using RCT. In RCT design, he explained four statistical conceptions, including H_0 and H_1, size of the critical region, and False negative. The Table 2 presents several formulae for dichotomous and continuous variable behaviour.

Table 2: *Basic General formulae for sample size calculation are as follows*

	Dichotomous variable	**Continuous variable**
Non – inferiority design*	$N = 2 \times \left[\frac{Z_{1-\alpha} + Z_{1-\beta}}{\gamma_0} \right]^2 \times t(1-t)$	$N = 2 \times \left[\frac{Z_{1-\alpha} + Z_{1-\beta}}{\gamma_0} \right]^2 \times s^2$
Equivalence design**	$N = 2 \times \left[\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\gamma_0} \right]^2 \times t(1-t)$	$N = 2 \times \left[\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\gamma_0} \right]^2 \times s^2$
Statistical superiority design***	$N = \frac{1}{2} \times \left[\frac{Z_{\frac{\alpha}{2}} + Z_{\beta}}{\arcsin\sqrt{t} - \arcsin\sqrt{t_0}} \right]^2$	$N = 2 \times \left[\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\gamma} \right]^2 \times s^2$
Clinical superiority design****	$N = 2 \times \left[\frac{Z_{1-\alpha} + Z_{1-\beta}}{d - \gamma_0} \right]^2 \times t(1-t)$	$N = 2 \times \left[\frac{Z_{1-\alpha} + Z_{1-\beta}}{\gamma - \gamma_0} \right]^2 \times s^2$

Where,
N represents the size per group
t represents the response rate of the control group receiving the standard treatment
t_0 represents the response rate of the experiment group receiving the new drug treatment
d signifies the actual difference between the treatment effects of the two groups
γ refers to the clinically acceptable margin for non inferiority, equivalence or superiority
γ_0 represents a pre determined clinically acceptable margin
S^2 represents the pooled standard deviation of the comparison groups

*Non-inferiority design aims to prove that a novel treatment or intervention is not significantly inferior to an established standard treatment by a pre-determined margin.

**It states that within a pre-determined margin of difference, an equivalence design in RCT seeks to express that a novel treatment or intervention is equally effective to a recognized standard treatment.

***It is used when we need to demonstrate that the new therapy is more effective than currently available treatments or no treatment at all and in order to support regulatory approval of the new intervention and to direct clinical practice, superiority trial findings are used.

****it is a study designed in RCT where the mainaim is to expressthat one intervention is better than another in clinical outcomes.

We may have the response variable as binary variable and in such situations the sample size can be calculated for several response rates is presented in Table 3 by assuming some values of the parameters. However when the response variable is continuous in nature the sample size can be calculated for different response rates is presented in Table 4 for several population sizes.

Table 3: *Sample size calculations for different response rates when the outcome measure is a dichotomous variable*

t	$1-t$	t_0	$N1$	$N2$	$N3$	$N4$
0.4	0.6	0.58	82	105	143	607
0.45	0.55	0.6	85	108	232	626
0.47	0.53	0.63	86	109	220	630
0.48	0.52	0.66	86	109	184	632
0.49	0.51	0.68	86	109	173	632
0.52	0.48	0.7	86	109	208	632
0.54	0.46	0.72	85	108	220	629
0.57	0.43	0.75	84	107	240	620
0.59	0.41	0.79	83	105	211	612
0.9	0.1	0.8	31	39	1304	228

Table 4: *Sample size calculations for different response rates when the outcome measure is a continuous variable*

s^2	$N1$	$N2$	$N3$	$N4$
36	112	142	63	446
49	152	193	86	608
64	198	252	112	794
81	251	319	142	1004
100	310	393	175	1240
121	375	476	212	1500

Here,
t represents the response rate of the reference treatment group
t_0 denotes the response rate of the experimental group
S^2 represents the pooled standard deviation of the comparison groups
N_1 represents the sample size for the non inferiority study design
N_2 represents the sample size for the equivalence study design
N_3 represents the sample size for the statistical study design
N_4 represents the sample size for the clinical superiority study design

Helio Amante Miot [16] explained how to assessthe alpha level of the estimate, the utmostacceptable sample error (in units of the average value), and the population standard deviation of the quantitative variable (discrete or continuous) in order to characterize thepopulation estimates.One should ascertain the population frequency of the variable outcomes, the alpha level of the estimate, and the utmost acceptable sample error to characterize the population estimate represented by a categorical variable. Sample size considerations should be made for the percentage of each category that makes up a qualitative variable when it is not dichotomous. This subgroup's performance should be considered as the population estimate when the variable's population standard deviation or

frequency is unclear and the literature does not contain comparable data.A population's quantitative and qualitative characteristics can be described using formulas for sample sizing as

2.1 Quantitative variable

For infinite population,

$$n = \left(\frac{(Z_{\frac{\alpha}{2}}) \times \delta}{E} \right)^2 \tag{5}$$

For finite population,

$$n = \frac{N * \delta^2 * \left(Z_{\frac{\alpha}{2}} \right)^2}{(N-1) * E^2 + \left(\delta * Z_{\frac{\alpha}{2}} \right)^2} \tag{6}$$

where,
n = sample number
N = finite population size
δ = population standard deviation of the variable
$Z_{\frac{\alpha}{2}}$ = value of error α
E = standard error

2.2 Qualitative variable

For infinite population,

$$n = \left(\frac{(Z_{\frac{\alpha}{2}}) \times \sqrt{pq}}{E} \right)^2 \tag{7}$$

For finite population,

$$n = \frac{N * pq * \left(Z_{\frac{\alpha}{2}} \right)^2}{(N-1) * E^2 + pq \left(Z_{\frac{\alpha}{2}} \right)^2} \tag{8}$$

where,
p = ratio of desirable outcomes of the variable in a population
q = ratio of undesirable outcomes in a population

2.3 Formulae for sample sizing to evaluate two groups

The formulae for sample sizing to evaluate two groups based on quantitative and qualitative variables and according to the pairing of cases are obtained as

For Numeric variables: In case of independent sample, we have

$$n = (S_a^2 + S_b^2) * \left(\frac{Z_{\frac{\alpha}{2}} + Z_\beta}{d} \right)^2 \quad (9)$$

And for dependent sample,

$$nP = \left(\frac{Z_{\frac{\alpha}{2}} + Z_\beta * sd}{\overline{D}} \right) \quad (10)$$

Where,

Np = number of pairs

Z_β = value of error β

d = minimum difference between the average values

Sa^2 & Sb^2 = standard deviation of variable in each group

sd = standard deviation of dissimilarity within paired observations

$\overline{D}$ = average value of the dissimilarity with the paired observations

For Qualitative variable: In case of qualitative variable with non-paired sample,

$$n = (p_1 q_1 + p_2 q_2) * \left(\frac{Z_{\frac{\alpha}{2}} + Z_\beta}{p_1 - p_2} \right)^2 \quad (11)$$

And for paired sample we have,

$$nP = \left(\frac{Z_{\frac{\alpha}{2}} + 2Z_\beta * p_a * q_a}{4 * p_d * (p_a - 0.5)^2} \right)^2 \quad (12)$$

where,

p_1 and p_2 represents the ratio of desirable outcomes in subgroups 1 and 2 respectively

q_1 and q_2 represents the ratio of undesirable outcomes in subgroups 1 and 2 respectively

p_a = ratio of unrelated pairs of group 1

q_a = ratio of unrelated pairs of group 1

p_d = the two group's sum of the ratio of unrelated pairs

2.4 Quantitative variables with linear correlation

The sample-size formula for quantitative variables with linear correlation is given by

$$n = 4 + \left(\frac{Z_{\frac{\alpha}{2}} + Z_\beta}{0.5 * \ln\left(\frac{1+r}{1-r}\right)} \right)^2 \quad (13)$$

Where

r = represents the linear correlation coefficient

2.5 Sample size based on sensitivity and specificity

Zaidi *et al.* [40] have done a comprehensive study on estimating the sample size of diagnostic studies in health sciences and concluded that Sensitivity and specificity are two effective measures used in diagnostic testing and statistical analysis.

Sensitivity depicts a diagnostic test's capacity to accurately recognize people who have the condition it is planned to diagnose. In other words, it calculates the percentage of people who are truly positive for the disease out of all those who have the condition. A test with high sensitivity will have a low incidence of false negative results, making it unlikelythat those with the condition will go undiagnosed. A sensitivity rating of 100% would be ideal.

The capability of a diagnostic test to accurately identify individuals who do not possess the disease being tested for is measured by the term "specificity." In other terms, it counts the percentage of people who are truly negative about the condition. People who do not have the disease are unlikely to receive a false diagnosis when a test has a high specificity because of its low rate of false positives. A precision rating of 100% would be ideal.

The Various measures of diagnostic accuracy are

$$\text{Sensitivity} = \frac{(\text{True Positive})}{(\text{True Positive} + \text{False Negative})} * 100 \tag{14}$$

$$\text{Specificity} = \frac{(\text{True Negative})}{(\text{True Negative} + \text{False Positive})} * 100 \tag{15}$$

$$\text{Positive Predicative Values} = \frac{(\text{True Positive})}{(\text{True Positive} + \text{False Positive})} * 100 \tag{16}$$

$$\text{Negative Predicative Values} = \frac{(\text{True Negative})}{(\text{True Negative} + \text{False Negative})} * 100 \tag{17}$$

where,

True Positive (TP) cases are those where a test has accurately identified them as positive, indicating that the patient has the disease and the test has done its job.

The number of cases in which a test accurately identified them as negative while the patient did not have the disease is referred to as "True Negative" (TN).

False Positive (FP) cases are instances where a test result was incorrectly reported as positive even though the patient did not actually have the illness.

False Negative (FN) cases are those in which a test incorrectly reports a result as negative despite the fact that the patient actually has the illness.

The measurement of diagnostic effectiveness is given as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{18}$$

In a particular population, prevalence denotes the count (percentage) of cases of disease present in that population. It is given by the following formula

$$P(\text{percentage}) = \frac{\text{Number of cases of disease}}{\text{Total number of Population}} \tag{19}$$

So, the formula for sensitivity and specificity of sample size are expressed as,

Based on sensitivity, sample size (n) is given by

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} * S_N * (1 - S_N)}{L^2 * \text{Prevalence}} \tag{20}$$

Based on specificity, sample size (n) is given by

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} * S_P * (1 - S_P)}{L^2 * (1 - \text{Prevalence})} \tag{21}$$

Where

$n = \text{requisite sample size}$

$S_N = \text{expected sensitivit y}$

$S_P = \text{expected specificit y}$

$\alpha = \text{Type1 error rate}$

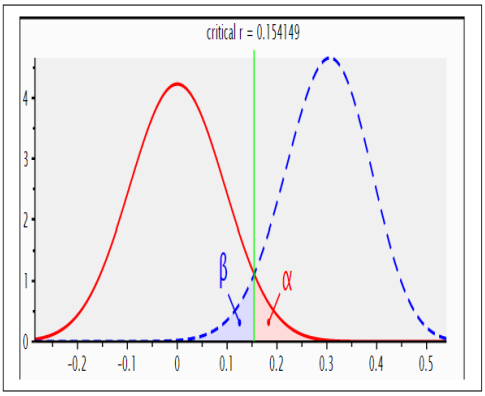
$Z_{1-\frac{\alpha}{2}} = \text{Normalized value}$

$L = \text{Exact accuracy}$

3. Graphical representation

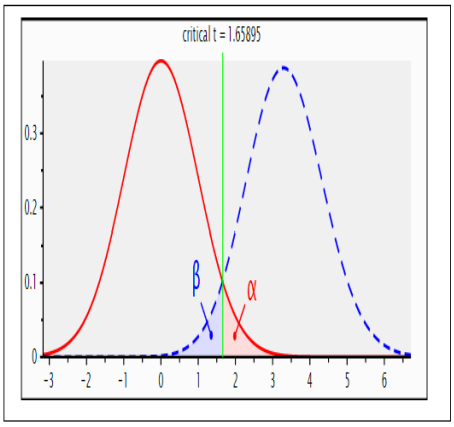
Incorporating the findings from our comparative analysis into the practical realm of clinical trials holds significant promise. Researchers and practitioners can leverage this knowledge to refine their study designs, ensuring that sample sizes align with research goals and statistical power requirements. Our research also emphasizes the importance of transparent reporting, promorting credibility in the scientific community. By embracing these applications, we can evaluate the quality of clinical trials and in turn, contribute to more reliable and valid research outcomes. Using the above several formulae in different circumstances we estimated the sample size using different parameters and the same has been presented in graphs as follows:

3.1 Correlation and Regression



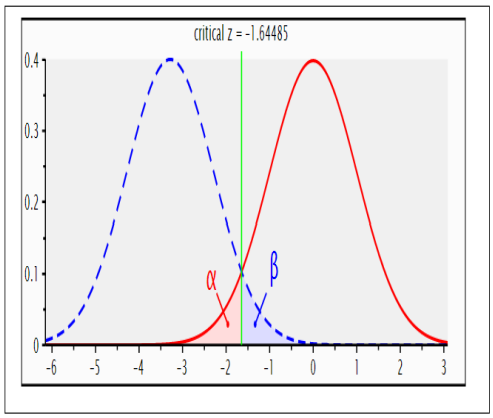
Input:	Tail(s)	=	One
	Correlation ρ H1	=	0.3
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Correlation ρ H0	=	0
Output:	Lower critical r	=	0.1541491
	Upper critical r	=	0.1541491
	Total sample size	=	115
	Actual power	=	0.9501151

Fig 1: *Correlation: Bivariate normal model for exact distribution*



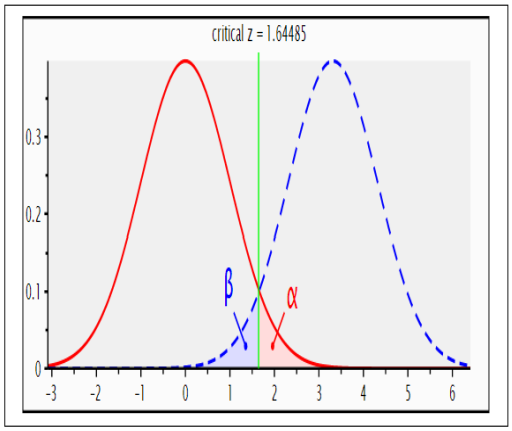
Input:	Tail(s)	=	One
	Effect size |ρ|	=	0.3
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
Output:	Non centrality parameter δ	=	3.3133098
	Critical t	=	1.6589535
	Df	=	109
	Total sample size	=	111
	Actual power	=	0.9503016

Fig 2: *t tests Correlation: Point biserial model*



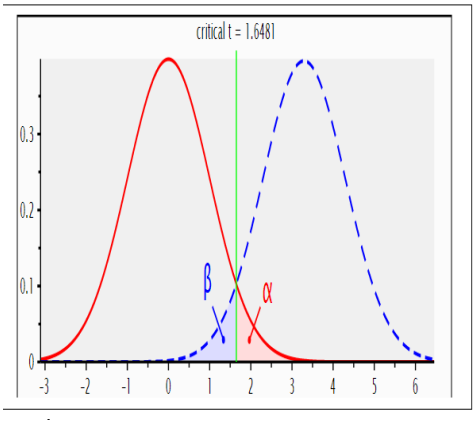
Input:	Tail(s)	=	One
	H1 corr ρ_ac	=	-0.1
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	H0 corr ρ_ab	=	0.1
	Corr ρ_bc	=	-0.1
Output:	Critical z	=	-1.6448536
	Sample size	=	595
	Actual power	=	0.9502184

Fig 3: *Z test correlations : Two dependent Pearson r's (common index)*



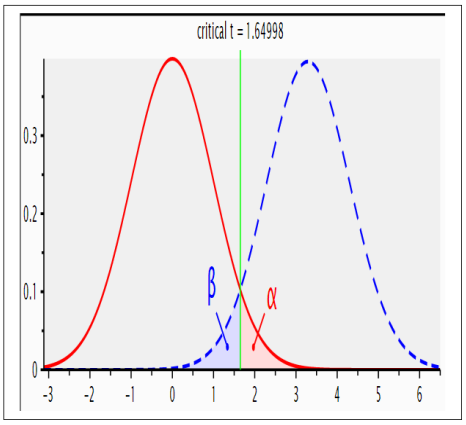
Input:	Tail(s)	=	One
	Effect size q	=	0.6
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Allocation ratio N2/N1	=	1
Output:	Critical z	=	1.6448536
	Sample size group 1	=	64
	Sample size group 2	=	64
	Total sample size	=	128
	Actual power	=	0.9524170

Fig 4: *Z test-Correlations: two independent Pearson r's*



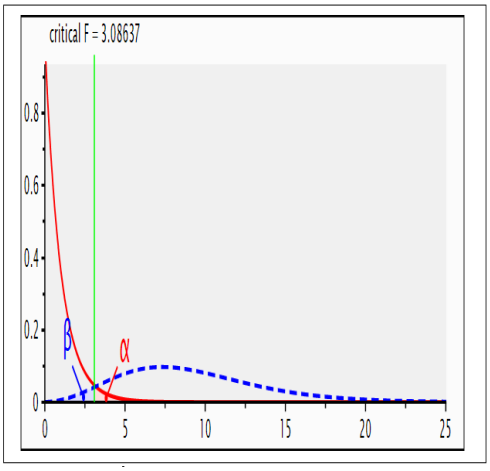
Input:	Tail(s)	=	One
	Slope H1	=	0.15
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Slope H0	=	0
	Std dev σx	=	1
	Std dev σy	=	1
Output:	Non centrality parameter δ	=	3.2961265
	Critical t	=	1.6481021
	Df	=	470
	Total sample size	=	472
	Actual power	=	0.9501720

Fig 5: *t tests-Linear bivariate regression: One group, size of slope*



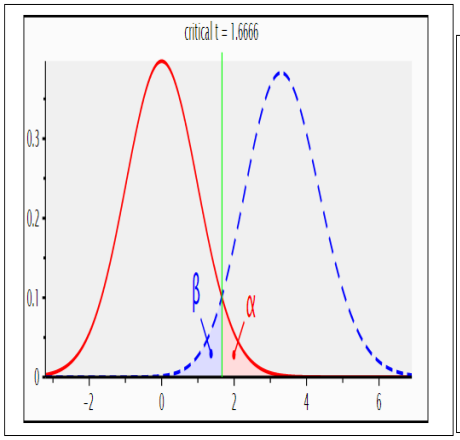
Input:	Tail(s)	=	One
	|Δ slope|	=	0.015
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Allocation ratio N2/N1	=	1
	Std dev residual σ	=	0.5
	Std dev σ_x1	=	20
	Std dev σ_x2	=	10
Output:	Non centrality parameter δ	=	3.2972716
	Critical t	=	1.6499830
	Df	=	298
	Sample size group 1	=	151
	Sample size group 2	=	151
	Total sample size	=	302
	Actual power	=	0.9500068

Fig 6: *t tests-Linear bivariate regression: two groups, difference between slopes*



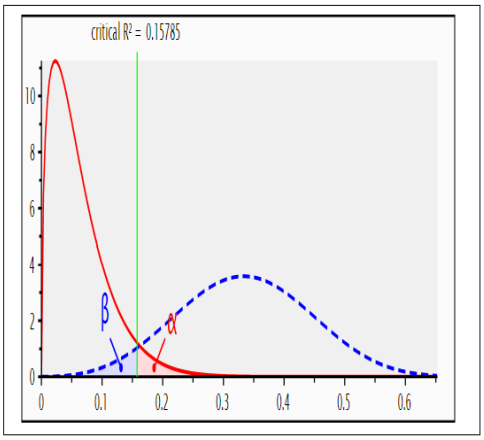
Input:	Effect size f^2	=	0.15
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Number of tested predictors	=	2
	Total number of predictors	=	5
Output:	Non centrality parameter λ	=	16.0500000
	Critical F	=	3.0863712
	Numerator df	=	2
	Denominator df	=	101
	Total sample size	=	107
	Actual power	=	0.9516875

Fig 7: *F tests-Linear multiple regression: Fixed model, R^2 increase*



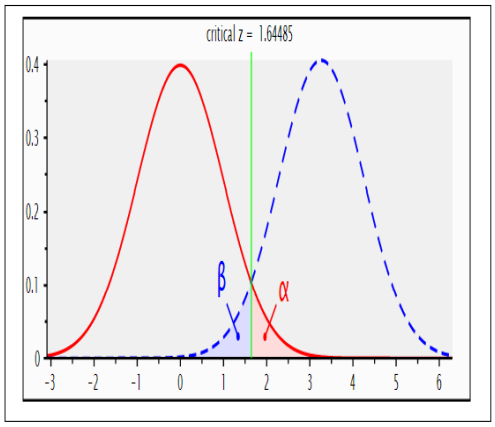
Input:	Tail(s)	= One
	Effect size f^2	= 0.15
	α err prob	= 0.05
	Power (1−β err prob)	= 0.95
	Number of predictors	= 2
Output:	Non centrality parameter δ	=3.3316662
	Critical t	=1.6665997
	Df	= 71
	Total sample size	= 74
	Actual power	=0.9510185

Fig 8: *t tests-Linear multiple regression: Fixed model, single regression coefficient*



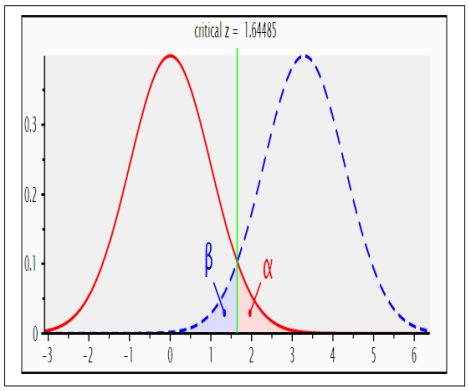
Input:	Tail(s)	= One
	H1 ρ^2	= 0.3
	H0 ρ^2	= 0
	α err prob	= 0.05
	Power (1−β err prob)	= 0.95
	Number of predictors	= 3
Output:	Lower critical R^2	=0.1578495
	Upper critical R^2	=0.1578495
	Total sample size	= 49
	Actual power	=0.9526746

Fig 9: *Exact-Linear multiple regression: Random model*



Input:	Tail(s)	= One
	Odds ratio	= 1.3
	Pr(Y=1|X=1) H0	= 0.2
	α err prob	= 0.05
	Power (1−β err prob)	= 0.95
	R^2 other X	= 0
	X distribution	= Normal
	X parm μ	= 0
	X parm σ	= 1
Output:	Critical z	=1.6448536
	Total sample size	= 988
	Actual power	=0.9501283

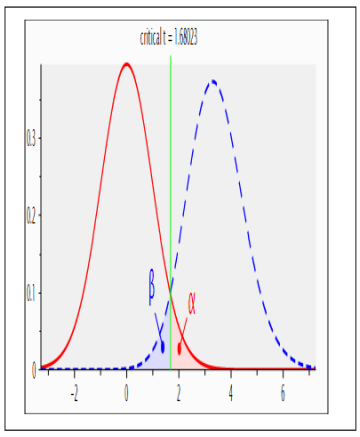
Fig 10: *Z tests-Logistic regression*



Input:	Tail(s)	=	One
	Exp(β1)	=	1.3
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
	Base rate exp(β0)	=	0.85
	Mean exposure	=	1
	R^2 other X	=	0
	X distribution	=	Normal
	X parm μ	=	0
	X parm σ	=	1
Output:	Critical z	=	1.6448536
	Total sample size	=	179
	Actual power	=	0.9502785

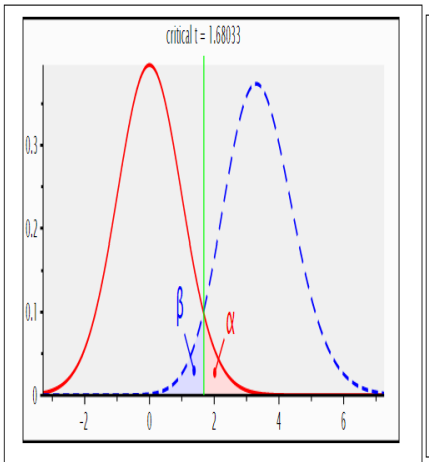
Fig 11: *Z tests-Poisson regression*

3.2 Means



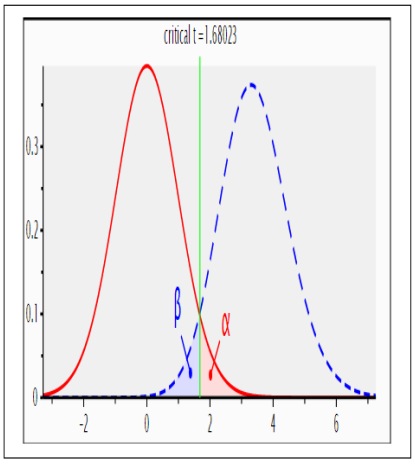
Input:	Tail(s)	=	One
	Effect size d	=	0.5
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
Output:	Non centrality parameter δ	=	3.3541020
	Critical t	=	1.6802300
	Df	=	44
	Total sample size	=	45
	Actual power	=	0.9512400

Fig 12: *t tests-Means: Difference from constant (One sample case)*



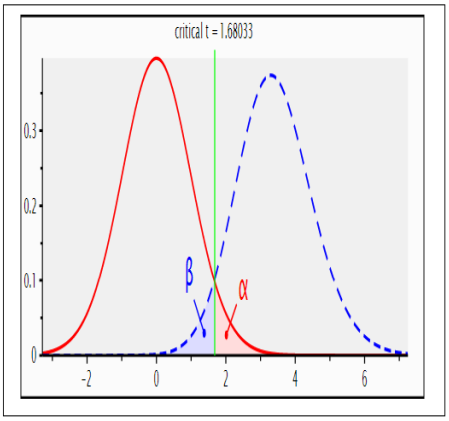
Input:	Tail(s)	=	One
	Parent distribution	=	Normal
	Effect size d	=	0.5
	α err prob	=	0.05
	Power (1−β err prob)	=	0.95
Output:	Non centrality parameter δ	=	3.3496901
	Critical t	=	1.6803274
	Df	=	43.8816940
	Total sample size	=	47
	Actual power	=	0.9507851

Fig 13: *t tests-Means: Wilcoxon signed-rank test (One sample case)*



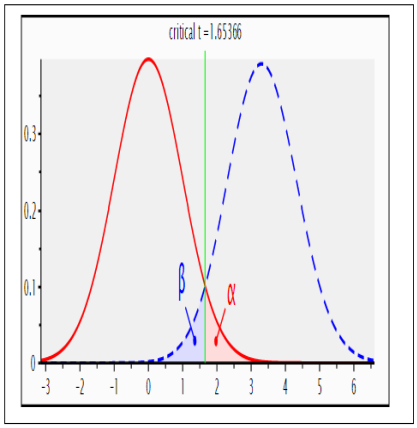
Input:	Tail(s)	=	One
	Effect size dz	=	0.5
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
Output:	Non centrality parameter δ	=	3.3541020
	Critical t	=	1.6802300
	Df	=	44
	Total sample size	=	45
	Actual power	=	0.9512400

Fig 14: *t tests-Means: Difference between two dependent means (matched pairs)*



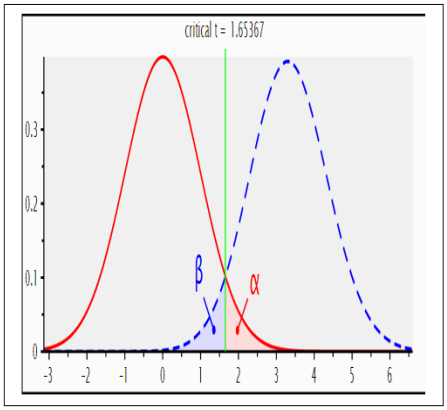
Input:	Tail(s)	=	One
	Parent distribution	=	Normal
	Effect size dz	=	0.5
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
Output:	Non centrality parameter δ	=	3.3496901
	Critical t	=	1.6803274
	Df	=	43.8816940
	Total sample size	=	47
	Actual power	=	0.9507851

Fig 15: *t tests-Means: Wilcoxon signed-rank test(matched pairs)*



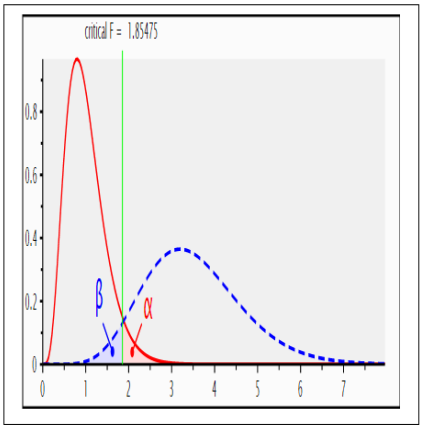
Input:	Tail(s)	=	One
	Effect size d	=	0.5
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Allocation ratio N2/N1	=	1
Output:	Non centrality parameter δ	=	3.3166248
	Critical t	=	1.6536580
	Df	=	174
	Sample size group 1	=	88
	Sample size group 2	=	88
	Total sample size	=	176
	Actual power	=	0.9514254

Fig 16: *t tests-Means: Difference between two independent means (two groups)*



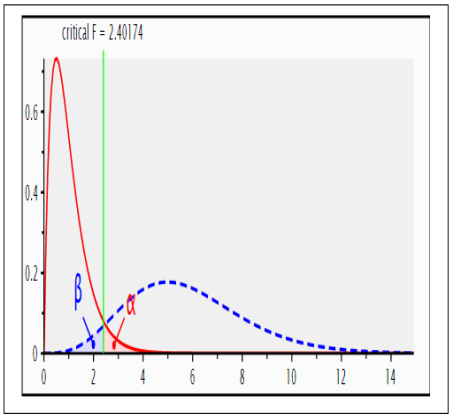
Input:	Tail(s)	=	One
	Parent distribution	=	Normal
	Effect size d	=	0.5
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Allocation ratio N2/N1	=	1
Output:	Non centrality parameter δ	=	3.3138635
	Critical t	=	1.6536729
	Df	=	173.7071
	Sample size group 1	=	92
	Sample size group 2	=	92
	Total sample size	=	184
	Actual power	=	0.9511454

Fig 17: *t tests-Means: Wilcoxon-Mann-Whitney test (two groups)*



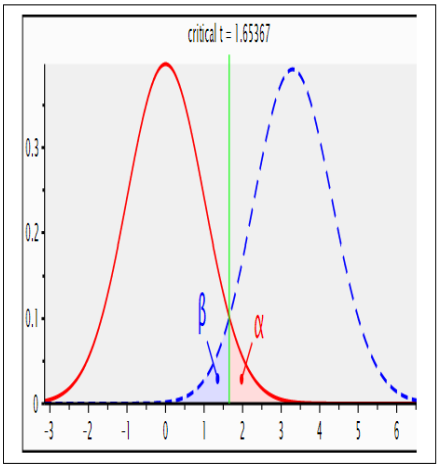
Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Numerator df	=	10
	Number of groups	=	5
	Number of covariates	=	1
Output:	Non centrality parameter λ	=	25.0000000
	Critical F	=	1.8547531
	Denominator df	=	394
	Total sample size	=	400
	Actual power	=	0.9504058

Fig 18: *F tests-ANCOVA: Fixed effects, main effects and interactions*



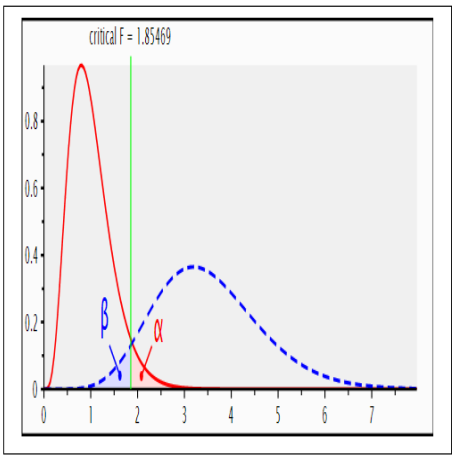
Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	5
Output:	Non centrality parameter λ	=	19.0625000
	Critical F	=	2.4017397
	Numerator df	=	4
	Denominator df	=	300
	Total sample size	=	305
	Actual power	=	0.9521313

Fig 19: *F tests-ANOVA: Fixed effects, omnibus, one-way*



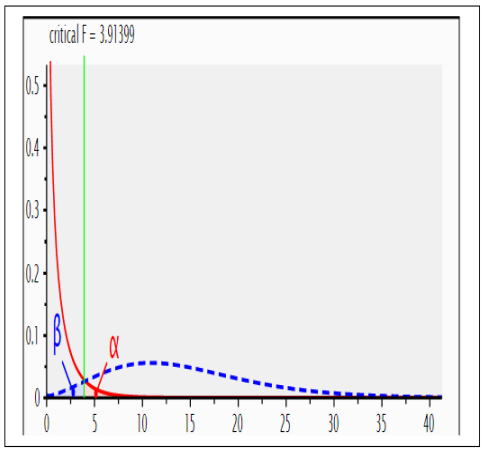
Input:	Tail(s)	=	One
	Parent distribution	=	Normal
	Effect size d	=	0.5
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Allocation ratio N2/N1	=	1
Output:	Non centrality parameter δ	=	3.3138635
	Critical t	=	1.6536729
	Df	=	173.7071
	Sample size group 1	=	92
	Sample size group 2	=	92
	Total sample size	=	184
	Actual power	=	0.9511454

Fig 20: *t tests-Means: Wilcoxon-Mann-Whitney test (two groups)*



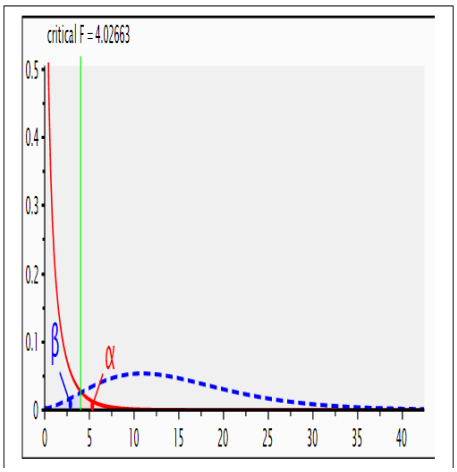
Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Numerator df	=	10
	Number of groups	=	5
Output:	Non centrality parameter λ	=	25.0000000
	Critical F	=	1.8546920
	Denominator df	=	395
	Total sample size	=	400
	Actual power	=	0.9504191

Fig 21: *F tests-ANOVA: Fixed effects, special, main effects and interactions*



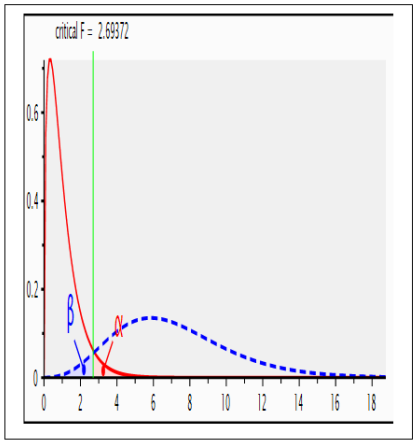
Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	2
	Number of measurements	=	4
	Corr among rep measures	=	0.5
Output:	Noncentrality parameter λ	=	13.2000000
	Critical F	=	3.9139890
	Numerator df	=	1.0000000
	Denominator df	=	130
	Total sample size	=	132
	Actual power	=	0.9501436

Fig 22: *F tests-ANOVA: Repeated measures, between factors*



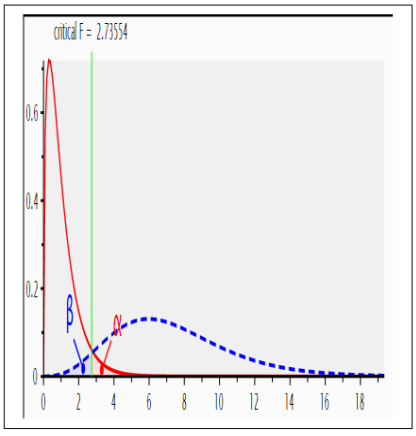
Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	2
	Number of measurements	=	4
	Corr among rep measures	=	0
Output:	Noncentrality parameter λ	=	13.5000000
	Critical F	=	4.0266314
	Numerator df	=	1.0000000
	Denominator df	=	52.0000000
	Total sample size	=	54
	Actual power	=	0.9500773
	Pillai V	=	0.2000000

Fig 23: *F tests-MANOVA: Repeated measures, between factors*



Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	2
	Number of measurements	=	4
	Corr among rep measures	=	0.5
	Non sphericity correction ϵ	=	1
Output:	Noncentrality parameter λ	=	18.0000000
	Critical F	=	2.6937209
	Numerator df	=	3.0000000
	Denominator df	=	102
	Total sample size	=	36
	Actual power	=	0.9517650

Fig 24: *F tests-ANOVA: Repeated measures, within factors*



Input:	Effect size f	=	0.25
	α err prob	=	0.05
	Power (1–β err prob)	=	0.95
	Number of groups	=	2
	Number of measurements	=	4
	Corr among rep measures	=	0
Output:	Noncentrality parameter λ	=	18.5000000
	Critical F	=	2.7355415
	Numerator df	=	3.0000000
	Denominator df	=	70.0000000
	Total sample size	=	74
	Actual power	=	0.9536728
	Pillai V	=	0.2000000

Fig 25: *F tests-MANOVA: Repeated measures, within factors*

The above sample size estimation in clinical trials has been done using different formulae at different situations and graphs have been drawn accordingly such as for correlation and regression, we need to know some given information such as level of significance, power of test and correlation coefficient under null and alternative hypothesis. Once the information is available we can substitute the available information in the formula and estimate the sample size along with actual power and lower

and upper critical value of correlation coefficient. Further, we can plot the same information in a graph such as given in Fig. 1 to Fig. 11. We can proceed in same passion for non-parametric tests too. The same can be concluded for Means, Proportions, Variance, and Generic etc. The Fig 1 to Fig. 11 presents the sample size estimation in graph for the data set concerning about the correlation and regression however if the concentration is on the mean the sample size graphs using the means are presented from Fig. 12 to 25 which is actually one of the mostly used statistic for the estimation of sample size. Thus, we can observe the behaviour of the sample size along with rest related parameters which is shows in the graphs from Fig. 1 to Fig. 25 for several characteristics of the study under consideration.

4. Conclusion

The comparative analysis of sample size determination procedures in clinical trials has shed light on the critical importance of this process in ensuring the reliability and validity of research results. It describes how choosing the right sample size based on the study's design and metrics may be done by researchers. It also offers advice on the data needed when requesting professional advice on evaluatingsample size in clinical trials. It is often recommended to presume that the sensitivity and specificity of the reference test are not exactly known when evaluating a novel diagnostic test. As a result, estimating them using the available methodologies is recommended. Unfortunately, sample size estimates for diagnostic investigations are infrequently disclosed by clinical investigators, and many doctors are ignorant of their significance. Instead, researchers frequently choose the sample size randomly, either out of convenience or in reliance on prior research. The precise interests of the researcher will ultimately determine the final sample size. Separate sample sizes should be chosen for each sensitivity and specificity if they are both equally crucial. The ultimate sample size would then be computed by selecting the greater of the two sample sizes. In instances where the researcher values sensitivity over specificity, the sample size would be determined solely by sensitivity. It is standard procedure for researchers to raise the sample size by 10% for the sake of accommodating people whoareunreachable.

Furthermore, the sample size is frequently raised by 30% to account for non-response. In order to get the appropriate degree of confidence and precision, more surveys or scheduled interviews may be conducted than is necessary. It follows that the determination of sample size in clinical trialsdepends upon the issue being addressed.

Ultimately, this research serves as a call to action for researchers, clinicians and educators alike. It emphasizes the imperative of elevating the discourse on sample size determination, promoting transparency and enhancing the quality of clinical trials to ensure that the research outcomes we generate can be trusted and applied to improve patient care and advance medical knowledge.

References

[1] Ahmad, W. M. A. W., Amin, W. A. A. W. M., Aleng, N. A. and Mohamed, N. (2012). Some practical guidelines for effective sample-size determination in observational studies. *Aceh International Journal of Science and Technology*, *1*(2), 51-53.

[2] Brown, C. G., Kelen, G. D., Ashton, J. J. and Werman, H. A. (1987). The beta error and sample size determination in clinical trials in emergency medicine. *Annals of emergency medicine*, *16*(2), 183-187.

[3] Brutti, P., De Santis, F. and Gubbiotti, S. (2008). Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine*, *27*(13), 2290-2306.

[4] Burmeister, E. and Aitken, L. M. (2012). Sample size: How many is enough? *Australian Critical Care*, *25*(4), 271-274.

[5] Chadha, V. K. (2006). Sample size determination in health studies. *NTI bulletin*, *42*(3and4), 55-62.

[6] Charles, P., Giraudeau, B., Dechartres, A., Baron, G. and Ravaud, P. (2009). Reporting of sample size calculation in randomized controlled trials. *Bmj*, *338*.

[7] Charan, J. and Biswas, T. (2013). How to calculate sample size for different study designs in medical research? *Indian journal of psychological medicine*, *35*(2), 121-126.

[8] Eng, J. (2003). Sample size estimation: how many individuals should be studied? *Radiology*, *227*(2), 309-313.

[9] Glenn, D. I. (1992). Determining sample size. *A series of the Program Evaluation and Organizational Development. University of Florida, Publication date: November*.

[10] Greene, T. (2021). Randomized Controlled Trials 6: Determining the Sample Size and Power for Clinical Trials and Cohort Studies. *Clinical Epidemiology: Practice and Methods*,2249, 281-305.

[11] Halperin, M., Rogot, E., Gurian, J., andEderer, F. (1968). Sample sizes for medical trials with special reference to long-term therapy. *Journal of chronic diseases*, *21*(1), 13-24.

[12] Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in nursing and health*, *31*(2), 180-191.

[13] Hajian-Tilaki, K. (2011). Sample size estimation in epidemiologic studies. *Caspian journal of internal medicine*, *2*(4), 289-298.

[14] Hayat, M. J. (2013). Understanding sample size determination in nursing research. *Western Journal of Nursing Research*, *35*(7), 943-956.

[15] Hajian-Tilaki, K. (2014). Sample size estimation in diagnostic test studies of biomedical informatics. *Journal of biomedical informatics*, *48*, 193-204.

[16] Miot, H. A. (2011). Sample size in clinical and experimental trials. *Jornal Vascular Brasileiro*, *10*, 275-278.

[17] Julious, S. A. and Patterson, S. D. (2004). Sample sizes for estimation in clinical research. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, *3*(3), 213-215.

[18] Juneja, A. and Sharma, S. (2015). Issues of sample size in sensitivity and specificity analysis with special reference to oncology. *Journal of Cancer Research and Therapeutics*, *11*(2), 482-484.

[19] Lachenbruch, P. A. (1991). Sample size determination in health studies: a practical manual. *Journal of the American Statistical Association*, *86*(416), 1149-1150.

[20] Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, *55*(3), 187-193.

[21] Morse, J. M. (2000). Determining sample size. *Qualitative health research*, *10*(1), 3-5.

[22] Mason, M. (2010). Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, *11*(3). https://doi.org/10.17169/fqs-11.3.1428.

[23] Negida, A., Fahim, N. K. and Negida, Y. (2019). Sample size calculation guide-part 4: how to calculate the sample size for a diagnostic test accuracy study based on sensitivity, specificity, and the area under the roc curve. *Frontiers in Emergency Medicine*, *3*(3), e33-e33.

[24] Negida, A. (2020). Sample size calculation guide-part 7: how to calculate the sample size based on a correlation. *Frontiers in Emergency Medicine*, *4*(2), e34-e34.

[25] Nanjundeswaraswamy, T. S., andDivakar, S. (2021). Determination of sample size and sampling methods in applied research. *Proceedings on Engineering Sciences*, *3*(1), 25-32.

[26] Obuchowski, N. A. (1998). Sample size calculations in studies of test accuracy. *Statistical Methods in Medical Research*, *7*(4), 371-392.

[27] Pezeshk, H. (2003). Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical Methods in Medical Research*, *12*(6), 489-504.

[28] Prajapati, B., Dunne, M., and Armstrong, R. (2010). Sample size estimation and statistical power analyses. *Optometry today*, *16*(7), 10-18.

[29] Pourhoseingholi, M. A., Vahedi, M. and Rahimzadeh, M. (2013). Sample size calculation in medical studies. *Gastroenterology and Hepatology from bed to bench*, *6*(1),14-17.

[30] Rohrig, B., Du Prel, J. B., Wachtlin, D., Kwiecien, R. and Blettner, M. (2010). Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *DeutschesÄrzteblatt International*, *107*(31-32), 552-556.

[31] Sakpal, T. V. (2010). Sample size estimation in clinical trial. *Perspectives in clinical research*, *1*(2), 67-69.

[32] Singh, A. S. and Masuku, M. B. (2014). Sampling techniques and determination of sample size in applied statistics research: An overview. *International Journal of economics, commerce and management*, *2*(11), 1-22.

[33] Sami, W., Alrukban, M. O., Waqas, T., Asad, M. R. and Afzal, K. (2018). Sample size determination in health research. *Journal of Ayub Medical College Abbottabad*, *30*(2), 308-311.

[34] Tatiana, M. (2003). Sample size determination for research projects. *British Orthodontic Society*.30,99-100.

[35] Wittes, J. (2002). Sample size calculations for randomized controlled trials. *Epidemiologic reviews*, *24*(1), 39-53.

[36] Whitley, E. and Ball, J. (2002). Statistics review 4: sample size calculations. *Critical care*, *6*, 1-7.

[37] Williamson, Owen D. (2003). "Determining the sample size in a clinical trial." *Medical Journal of Australia* 178(7), 358-358.

[38] Willan, A. and Kowgier, M. (2008). Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clinical Trials*, *5*(4), 289-300.

[39] Zhong, B. (2009). How to calculate sample size in randomized controlled trial? *Journal of thoracic disease*, *1*(1), 51-54.

[40] Zaidi, S. M. H., Waseem, H. F., Ansari, F. A., Irfan, M., Fahim, S. and Ahmad, M. (2016). Sample size estimation of diagnostic test studies in health sciences. In *Proceedings of 14th International Conference on Statistical Sciences*,29, 239-246.

OPTIMAL AND ECONOMIC DESIGN OF CHAIN SAMPLING PLAN FOR ASSURING MEDIAN LIFE UNDER NEW COMPOUNDED BELL WEIBULL LIFE TIME MODEL

M. Muthumeena[1] and S. Balamurali[2*]

•

Department of Mathematics
Kalasalingam Academy of Research and Education
Krishnankoil, 626126, Tamilnadu, India
muthumeena.m@klu.ac.in[1] and sbmurali@rediffmail.com[2*]

Abstract

The methodology to design, one of the cumulative results plans called chain sampling plan, is proposed in this paper which ensures the median lifetime of the products under the complementary bell Weibull model. For costly and destructive testing, usually single sampling plan with zero acceptance number is used. But chain sampling plan is an alternative to zero acceptance number single sampling plans. A comparative analysis of proposed plan's OC curve outperforms in discrimination between the lots of varying quality, when compared to the single sampling plan. The advantages of the proposed plan by comparing the performance of the OC curve with other lifetime distributions are also discussed. Tables are constructed to select the optimal parameters for the various combinations of lifetime distributions. The implementation of the proposed plan in industrial scenarios is also explained by using a real time data. Finally, an economic design of the proposed sampling plan is discussed by considering some cost models to minimize the total cost.

Keywords: Consumer's risk, Complementary Bell Weibull distribution, Chain sampling plan, Economic design, Median life, Producer's risk, Truncated life test.

I. Introduction

In today's global marketplace, quality is no longer merely an option or a goal for companies. Instead, it has become an essential requirement for businesses to thrive and succeed. Statistical quality control or SQC uses the technique of statistical approaches like reliable product quality, monitoring, maintenance, and assurance in a variety of sectors. Due to the requirement of testing and destroying every unit in a lot, a 100% inspection plan with destructive testing is not feasible option for both the buyer and the producer when the lot is destroyed, there will be a perpetual shortage for the buyer and also the producers may not achieve the sufficient return on investment. Acceptance sampling plan (ASP) plays a major role in SQC by aiding in the process of decision-making by assessing the quality of a batch and determining whether it should be accepted or rejected. This tool is highly significant in ensuring that only lots with satisfactory quality are accepted. Making decisions about the lot allocation based on samples can increase the chances of errors occurring. Probability of Type I error, also mentioned to as the producer's risk (α), and happens when there is possibility of rejecting a good lot. On the other hand, Probability of Type II error is referred to as the consumer's risk (β) occurs when there is a chance of accepting a defective lot.

ASP's are very important in ensuring the quality and conformity of submitted lot as well as adherence with its acceptance criteria, the utmost number of flawed items may be present in the sample, accepting the entire lot priory. The performance of an ASP's effectiveness lies in the investigation of its operating characteristic (OC) curve. This graphical representation allows the ability of ASP to differentiate between acceptable and unacceptable lots. For costly and destructive testing, an SSP with zero acceptance number is more preferable. But SSP with zero-acceptance number, exhibits limited discriminatory capability. Consequently, the OC curve of these plans lacks effectiveness and does not display any inflection point.

Dodge [5] introduced the idea of chain sampling plan, which is denoted as ChSP-1; to address the drawback of SSP with acceptance number is zero. ChSP-1 applies the accumulated outcomes of multiple samples to determine the acceptance or rejection of a submitted lot. A ChSP-1 aims to enhance the shape of the OC curve compared to SSP with a zero acceptance number, also minimizing the necessary sample size. It is also known as a conditional sampling plan as it leverages information from previous lots to determine how the current lot should be disposed off. The ChSP-1 plan is designed to minimize the risk of rejecting an entire lot, due to the presence of one nonconforming item was detected in the sample. Currently, literature of ASP contains several studies on ChSP-1 across diverse conditions. The procedure outlined by Soundararajan [15] helps practitioners in constructing the ChSP-1 based on specific requirements. Govindaraju and Subramani [6] proposed an alternative method for selecting the ChSP-1 plan that minimizes both producer's and consumer's risk. For further information regarding the concept of ChSP-1, one may refer to Soundararajan and Govindaraju ([16], [17]), Raju [13], Govindaraju and Balamurali [7], Raju and Raghottam [14], Jeyadurga and Balamurali [10] and so on.

The process of developing a product is incomplete without conducting a life testing experiments, because it focuses how to determine the reliability of products over an extended period. Overall, the implementation of a reliability test plan that includes time truncated life (TTL) testing can greatly enhance the product quality assurance efforts and help to ensure that customers receive reliable and high-quality products. Acceptance sampling involves the use TTL, to ascertain, that either the average or percentile lifetime of products meets certain criteria. Several sampling plans have been investigated by using the TTL test, (see for example Aslam and Jun [1], Jeyadurga *et al.* [11], Vijayaraghavan and Saranya [19], Jayalakshmi and Veerakumari [12], Vijayaraghavan and Pavithra [20]). In this paper, we attempt to design of ChSP-1 plan with the intention of assuring complementary Bell Weibull distributed median lifetime of products. The literature survey reveals that there is no prior research has been explored on ChSP-1 for complementary Bell Weibull distribution or compared its performance with other lifetime models. Implementation of the proposed plan is explained with an industrial example of fatigue fracture of Kevlar 373/epoxy composite material, and also the advantages of economic aspect of proposed plan with reduced cost of inspection and total cost for assuring complementary Bell Weibull distribution with different median lifetime models are also discussed.

II. Determination of Failure Probability with Complementary Bell Weibull Model

Combining continuous lifetime distributions with discrete distribution is a powerful approach to create flexible and versatile models that capture the complexity of lifetime data and leading to more accurate predictions and better analysis. Tahir and Cordeiro [18] discussed a comprehensive review of compounding techniques for univariate distributions. Their study focused on generalization classes and explored various complementary compound models. Castellares *et al.* [4] have proposed a single-parameter discrete bell distribution as an alternative to the widely used Poisson distribution. Castellares *et al.* [4] had found that the Poisson model cannot be directly combined into the Bell model, although the bell distribution converges to the Poisson distribution

when the parameter value decreases. In other words, the Poisson distribution can be viewed as a special case of the Bell distribution that emerges when the Bell parameter approaches to zero. Recently, Algarni [2] proposed a new compounded model known as complementary Bell Weibull distribution (CBell- WD) that aims to ensure the median lifetime based on group acceptance sampling plan. The cumulative distribution function (CDF) of the CBell- WD is given by

$$F(t)=\frac{exp\left[e^{\theta\left\{1-exp\left[-\left(t/\delta\right)^{\eta}\right]\right\}}-1\right]-1}{exp\,[e^{\theta}-1]-1}\quad 0<t<\infty;\ \delta,\eta>0 \tag{1}$$

where θ denoted as strictly positive Bell model parameter also η and δ stands for shape and scale parameters respectively. The median of CBell- W distribution is discussed by Algarni [2] and it is represented as

$$m=\delta\left[-ln\left(1-\left\{\theta^{-1}ln\left[1+ln\left\{1+0.5\left[exp\left(e^{\theta}-1\right)-1\right]\right\}\right]\right\}\right)\right]^{1/\eta} \tag{2}$$

and now let us take

$$\xi=\left[-ln\left(1-\left\{\theta^{-1}ln\left[1+ln\left\{1+0.5\left[exp\left(e^{\theta}-1\right)-1\right]\right\}\right]\right\}\right)\right]^{1/\eta}$$

The failure probability of an item prior to the time $t_0$is represented by $p=F(t_0)$ and can be amended as

$$p=\frac{exp\left[e^{\theta\left\{1-exp\left[-\left(t_0/\delta\right)^{\eta}\right]\right\}}-1\right]-1}{exp\,[e^{\theta}-1]-1} \tag{3}$$

In the context of developing sampling plans for lifetime assurance, it is common to express the test duration (t_0) as a fixed multiple of the specified median life (m_0). This mathematical relationship can be represented by the equation as $t_0 = am_0$ where *'a'* symbolizes the experiment termination ratio. The probability that an item fails prior to a given experiment time 't_0' as determined by the CBell - W distribution, is calculated as follows:

$$p=\frac{exp\left[e^{\theta\left\{1-exp\left[-\left(\frac{a\xi}{m/m_0}\right)^{\eta}\right]\right\}}-1\right]-1}{exp[e^{\theta}-1]-1} \tag{4}$$

I. Survey of Few Other Lifetime Distributions

In this paper, we have chosen a few lifetime distributions to compare the ChSP-1 for assuring the median lifetime of the products such as extended odd Weibull exponential distribution (EOW-ED), exponentiated Weibull distribution (EWD), exponentiated Frechet distribution (EFD) and Burr XII distribution (BXIID). Our analysis revolves around, analysing the performance of the ChSP-1 using CBell-WD with other lifetime distributions. Additionally, we will utilize these models to compare the OC curve performance of the proposed sampling plan. Table 1 provides the description of these lifetime distributions and their failure probabilities based on median lifetime.

Table 1: *Failure probability based on median life of some selected life time models*

S. No	Models	Median lifetime ($q = 0.5$)	Failure probability
1.	EOW-ED	$m = -\frac{1}{\lambda} log\left[1-\left(\frac{-1+(1-q)^{-\eta}}{\eta+\left[-1+(1-q)^{-\eta}\right]}\right)^{1/\delta}\right]$ where $\omega = log\left[1-\left(\frac{-1+(1-q)^{-\eta}}{\eta+\left[-1+(1-q)^{-\eta}\right]}\right)^{1/\delta}\right]$	$p = 1-\left(1+\eta\left[e^{-\omega\, a\,(1/r_2)}-1\right]^{\delta}\right)^{-\frac{1}{\eta}}$ where η and δ are the two shape parameters.
2.	BXIID	$m = \sigma * \omega$ where $\omega = \lambda * \left[(1/(1-q))^{1/\delta}-1\right]^{1/\eta}$	$p = 1-\left[1+\left(\frac{\omega * a}{r_2}\right)^{\delta}\right]^{-\eta}$ where η and δ are the two shape parameters.
3.	EWD	$m = \lambda\left[-ln\left(1-q^{1/\eta}\right)\right]^{1/\delta}$	$p = \left[1-exp\left((r_2)^{\delta}\left(a\ ln\left(1-q^{1/\eta}\right)\right)^{\delta}\right)\right]^{\eta}$ where η and δ are the two shape parameters.
4.	EFD	$m = \sigma * \omega$ where $\omega = \left(-ln\left(1-(1-q)^{1/\eta}\right)^{-1/\delta}\right)$	$p = 1-\left[1-exp\left\{-\left(\frac{a*\omega}{r2}\right)^{-\eta}\right\}^{\delta}\right]$ where η and δ are the two shape parameters.

III. Conditions for Application and Operating Procedure of ChSP-1

The following are the conditions of application and the operating procedure of the proposed ChSP-1 plan.

1. The product to be inspected consists of a sequence of consecutive lots, either material or individual units, produced through a continuous and steady flow.
2. Chain sampling plans rely on a level of trust between the consumer and the producer. The producer will not intentionally supply inferior lots because the plan uses a combination of small sample size and information from previous samples to decide on the current lot's acceptance.
3. The production process should operate in a stable and continuous manner to ensure consistency and reliability. The product originates from a trusted source in which the consumer has high level of confidence.

The operating procedure of ChSP-1 based on TTL test is described in Figure 1 as shown below.

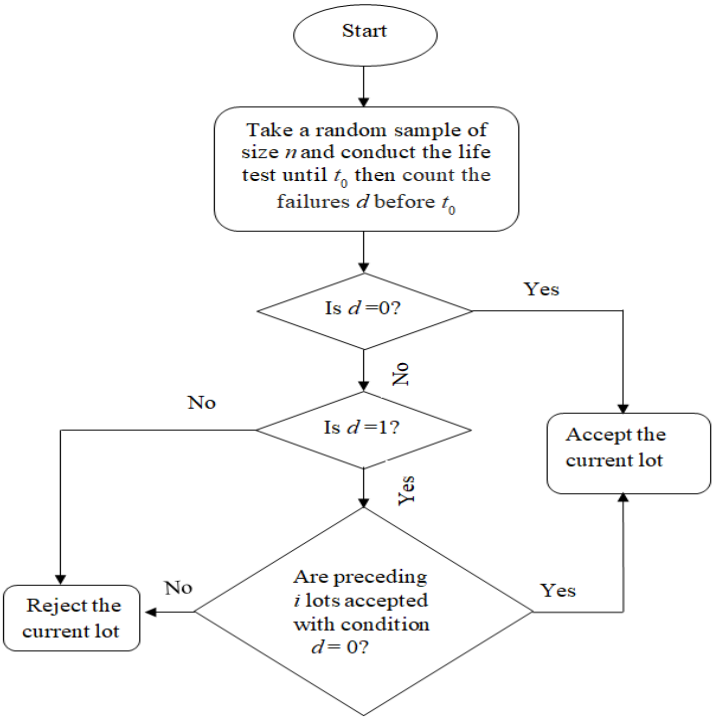


Figure 1: *Operating procedure of ChSP-1*

The ChSP-1 has two parameters namely n and i where n is the sample size, and i denoted as the number of preceding lots that are taken into account when making a decision about the current lot. When the limit of i approaches infinity $(i \to \infty)$, the operating characteristic (OC) function of a ChSP -1 simplifies to the OC function of a SSP with acceptance number zero.

The OC function of ChSP-1, as referenced in Dodge [5], can be expressed using the following equation.

$$L_A(p) = P(d=0) + [P(d=1) \times (P(d=0))^i] \tag{5}$$

where the probability $P(d = 0)$ indicates the possibility that no items in a given sample are non-conforming out of n items. Conversely, $P(d = 1)$ is used to represent the probability of a current sample containing a single non-conforming item also i denoted as the number of preceding samples.

Under the binomial distribution,

$$L_A(p) = (1-p)^n + np(1+p)^{n+ni-1} \tag{6}$$

IV. Designing Methodology

This section presents the methodology for designing the ChSP-1, which aims to ensure the product has a median lifetime that adheres to the CBell - WD. The primary objective of using the two-points on the OC curve approach is to fulfil both the producer's risk and the consumer's risk simultaneously. The optimal plan parameters are determined by passing through the two designated points on the OC curve such as (PQL, 1-α) and (CQL, β). By solving the following

optimization problem, the optimal parameters of the proposed plan are determined.

$$\begin{aligned} &\text{Minimize } ASN(p) = n \\ &\text{Subject to } L_A(p_1) \geq 1-\alpha \\ &\qquad L_A(p_2) \leq \beta \\ &\qquad n > 1,\ i \geq 1 \end{aligned} \tag{7}$$

The failure probabilities of both the producer's and consumer's risk are represented as p_1 and p_2 respectively. The probability acceptance of the lot at producer quality Level (PQL) and consumer quality level (CQL) under ChSP-1 is as follows:

$$L_A(p_1) = (1-p_1)^n + np_1(1+p_1)^{n+ni-1} \tag{8}$$

$$L_A(p_2) = (1-p_2)^n + np_2(1+p_2)^{n+ni-1} \tag{9}$$

The quality level can be expressed as the median proportionate between product's actual median lifetime and its specified median lifetime. This ratio is commonly denoted as m/m_0.The failure probability, denoted as p_1, can be determined by considering the median ratios m/m_0= 4, 6, 8 and 10. These median ratios are noted as PQL. Let p_2 be the failure probability and its median ratio m/m_0= 1 is considered as CQL.

V. Algorithm

The following algorithm can be used to compute the optimal parameters of the proposed ChSP-1 plan, with the objective of assuring a median lifetime of the submitted products for inspection.

Step 1 : Specify α, β, a, η, θ and median ratios.

Step 2 : For the specified input parameters α,β, a, η, θ and median ratios to calculate the failure probability (p_1) using an equation (4) denoted as PQL similarly to get the failure probability (p_1) when the median ratio is equal to 1 labelled it as CQL.

Step 3 : Set n =1 and i =1

Step 4 : Using an equation (6), replace n and i to calculate the lot acceptance probability at p_1 denoted as $L_A(p_1)$. Likewise, substitute n and i to calculate lot acceptance probability at p_2 denoted as $L_A(p_2)$.

Step 5 : Determine the largest sample size n, say n_A, and satisfying the condition $L_A(p_1) \geq 1-\alpha$ for all $n \leq n_A$.

Step 6 : Determine the largest sample size n, say n_B, and satisfying the condition $L_A(p_1) \leq \beta$ for all $n_B \leq n$.

Step 7 : If the condition, $n_B \leq n_A$ exists, then the corresponding set of plan parameter (n_B and i) is obtained. If not, Steps 4 to 5 must be repeated with various combination of n and i until an optimum plan parameter is successfully acquired.

By using the above algorithm, Tables 1 and 3 are constructed. These tables display the optimal plan parameters of ChSP-1, which have been determined by considering different combinations of the parameters (η, θ) = (2, 1.25), (2, 1.50) and (1, 1.20) for the CBell - WD. We have also considered

the producer's risk α = 0.05 and consumer's risk β = 0.25, 0.10, 0.05 and 0.01 and the values of median ratios m/m_0= 4, 6, 8, 10 at the termination ratios a = (0.5 and 1.0). The tables are presented for the calculated performance measures, including the OC functions, at specific failure probabilities of p_1 and p_2. From the tables, we analysed that if any value ofβ,a, m/m_0 and η increases then the value of sample size leads to decrease. Therefore, consumers will be safeguarded from accepting the lots of poor quality and producer's risks are also decreased while reaching the decision for accepting the current lot.

Table 1: *Optimal parameters of the proposed ChSP-1 under the CBell - W model with $\eta = 2$ and $\theta = 1.25$*

		a = 0.5				a = 1.0			
β	r_2	n	i	$L_A(p_1)$	$L_A(p_2)$	n	i	$L_A(p_1)$	$L_A(p_2)$
0.25	4	19	3	0.9876	0.2401	3	1	0.9976	0.1718
	6	↑	↑	0.9974	0.2401	↑	↑	0.9995	0.1718
	8	↑	↑	0.9991	0.2401	↑	↑	0.9998	0.1718
	10	↑	↑	0.9996	0.2401	↑	↑	0.9999	0.1718
0.10	4	31	2	0.9772	0.0965	4	1	0.9956	0.0781
	6	↑	↑	0.9951	0.0965	↑	↑	0.9992	0.0781
	8	↑	↑	0.9984	0.0965	↑	↑	0.9997	0.0781
	10	↑	↑	0.9993	0.0965	↑	↑	0.9999	0.0781
0.05	4	40	2	0.9639	0.0479	5	1	0.9932	0.0361
	6	↑	↑	0.9921	0.0479	↑	↑	0.9987	0.0361
	8	↑	↑	0.9974	0.0479	↑	↑	0.9996	0.0361
	10	↑	↑	0.9989	0.0479	↑	↑	0.9998	0.0361
0.01	4	89	1	0.9633	0.0091	7	1	0.9869	0.0082
	6	61	2	0.9825	0.0096	↑	↑	0.9975	0.0082
	8	↑	↑	0.9941	0.0096	↑	↑	0.9992	0.0082
	10	↑	↑	0.9975	0.0096	↑	↑	0.9996	0.0082

Table 2: *Optimal parameters of the proposed ChSP - 1 under the CBell - W model with $\eta = 2$ and $\theta = 1.50$*

		a = 0.5				a = 1.0			
β	r_2	n	i	$L_A(p_1)$	$L_A(p_2)$	n	i	$L_A(p_1)$	$L_A(p_2)$
0.25	4	30	3	0.9916	0.2449	3	1	0.9993	0.1718
	6	↑	↑	0.9983	0.2449	↑	↑	0.9998	0.1718
	8	↑	↑	0.9994	0.2449	↑	↑	0.9999	0.1718
	10	↑	↑	0.9997	0.2449	↑	↑	0.9999	0.1718
0.10	4	49	2	0.9845	0.0994	4	1	0.9987	0.0781
	6	↑	↑	0.9968	0.0994	↑	↑	0.9997	0.0781
	8	↑	↑	0.9989	0.0994	↑	↑	0.9999	0.0781
	10	↑	↑	0.9995	0.0994	↑	↑	0.9997	0.0781
0.05	4	63	4	0.9601	0.0499	5	1	0.9980	0.0361
	6	↑	↑	0.9911	0.0499	↑	↑	0.9996	0.0361
	8	↑	↑	0.9970	0.0499	↑	↑	0.9998	0.0361
	10	↑	↑	0.9987	0.0499	↑	↑	0.9999	0.0361
0.01	4	106	1	0.9598	0.0067	7	1	0.9961	0.0082
	6	97	2	0.9882	0.0099	↑	↑	0.9993	0.0082
	8	↑	↑	0.9961	0.0099	↑	↑	0.9997	0.0082
	10	↑	↑	0.9983	0.0099	↑	↑	0.9999	0.0082

(↑) use the plan above

Table 3: *Optimal parameters of the proposed ChSP-1 under the CBell - W model with $\eta = 1$ and $\theta = 1.20$*

β	r_2	a = 0.5				a = 1.0			
		n	i	$L_A(p_1)$	$L_A(p_2)$	n	i	$L_A(p_1)$	$L_A(p_2)$
0.25	4	*	*	*	*	*	*	*	*
	6	7	2	0.9556	0.2297	3	1	0.9755	0.1718
	8	7	2	0.9747	0.2297	3	1	0.9870	0.1718
	10	7	2	0.9837	0.2297	3	1	0.9920	0.1718
0.10	4	*	*	*	*	*	*	*	*
	6	*	*	*	*	4	1	0.9576	0.0781
	8	12	1	0.9571	0.0853	4	1	0.9772	0.0781
	10	11	2	0.9627	0.0898	4	1	0.9859	0.0781
0.05	4	*	*	*	*	*	*	*	*
	6	*	*	*	*	*	*	*	*
	8	22	1	0.9528	0.0498	4	1	0.9653	0.0361
	10	15	1	0.9582	0.0412	4	1	0.9783	0.0361
0.01	4	*	*	*	*	*	*	*	*
	6	*	*	*	*	*	*	*	*
	8	*	*	*	*	*	*	*	*
	10	*	*	*	*	7	1	0.9592	0.0082

(*) refers plan doesn't exist

VI. Comparison Using OC Curve

The first subsection, initiates the analysis by comparing the performance of proposed plan with SSP based on CBell-WD in terms of OC function. The subsequent subsection explores the OC curve performance of proposed plan based on CBell-WD with multiple median lifetime distributions.

I. Performance Analysis on OC Curve of Proposed ChSP-1 with SSP

In this section, we will briefly analyze the proposed plan's performance in comparison to SSP with CBell-WD. This comparison will be done using the performance of OC curve. Then the OC curve of proposed plan parameter is drawn for n = 3, i= 1 and the SSP with Ac = 0 is drawn for the same sample size with η = 2 and θ = 1.50 at the level of α = 0.05 and β = 0.25 which are gratified with m/m_0 = 2 and a =1.0 as shown in Figure 2. Also the comparison of acceptance probabilities L_A (p_1) and L_A (p_2) and the optimal plan parameters of ChSP-1 and SSP are calculated in Table 4. From Figure 2, it is clear that the OC curve of the proposed plan surpasses the SSP plan for low failure probabilities. The OC curve of a zero acceptance number (Ac = 0) SSP lacks a point of inflection. This means the probability of accepting a lot drops sharply even for the slightest increase in the proportion of nonconforming items (p). Thus, the proposed plan offers higher probability of acceptance compared to SSP. The proposed plan gives the protection for producers, where their products are exhibiting the best quality simultaneously; this plan also ensures the consumer protection in cases where the product's quality falls on poor quality. Therefore, based on the discussion, it can be concluded that the proposed plan will distinguish the various lots in terms of their quality, in comparison to the discriminative capacity of SSP with Ac = 0.

Table 4: *Comparison of acceptance probabilities proposed ChSP-1 with SSP under the CBell - W model with* $\eta = 2$ *and* $\theta = 1.50$

β	r_2	a = 1.0					
		SSP			ChSP-1		
		(n, c)	$L_A(p_1)$	$L_A(p_2)$	(n, i)	$L_A(p_1)$	$L_A(p_2)$
0.25	4	(3,0)	0.9771	0.125	(3,1)	0.9993	0.1718
	6	(3,0)	0.9905	0.125	(3,1)	0.9998	0.1718
	8	(3,0)	0.9948	0.125	(3,1)	0.9999	0.1718
	10	(3,0)	0.9967	0.125	(3,1)	0.9999	0.1718
0.10	4	(4,0)	0.9697	0.0625	(4,1)	0.9987	0.0781
	6	(4,0)	0.9874	0.0625	(4,1)	0.9997	0.0781
	8	(4,0)	0.9931	0.0625	(4,1)	0.9999	0.0781
	10	(4,0)	0.9956	0.0625	(4,1)	0.9999	0.0781
0.05	4	(5,0)	0.9622	0.0313	(5,1)	0.9980	0.0361
	6	(5,0)	0.9843	0.0313	(5,1)	0.9996	0.0361
	8	(5,0)	0.9914	0.0313	(5,1)	0.9998	0.0361
	10	(5,0)	0.9945	0.0313	(5,1)	0.9999	0.0361
0.01	4	*	*	*	(7,1)	0.9961	0.0082
	6	(7,0)	0.9781	0.0078	(7,1)	0.9993	0.0082
	8	(7,0)	0.9879	0.0078	(7,1)	0.9997	0.0082
	10	(7,0)	0.9924	0.0078	(7,1)	0.9999	0.0082

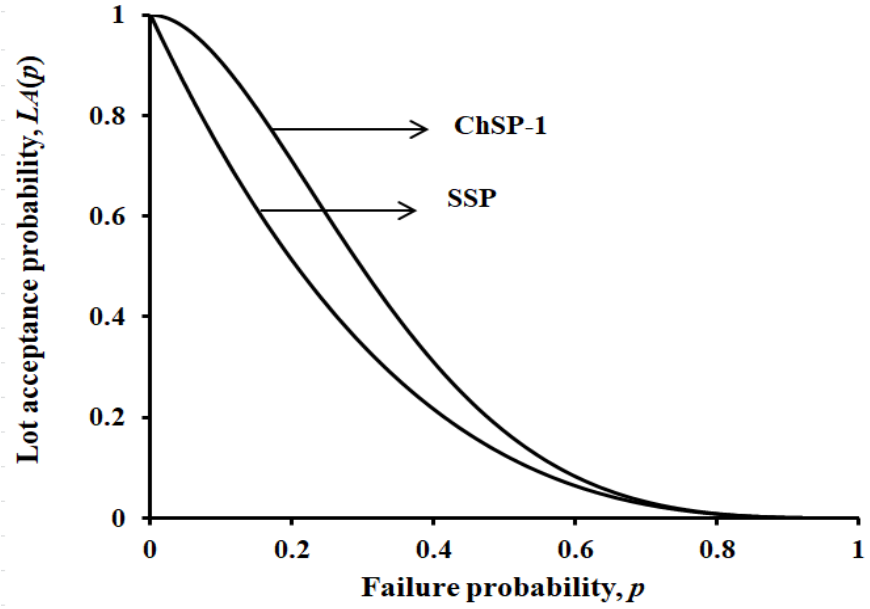


Figure 2: *OC curves of the proposed ChSP-1 and SSP with Ac = 0*

II. OC curve Comparison of Proposed Plan Based on Distinct Lifetime Distributions

In this analysis, we compare the OC curve performance of ChSP- 1 using CBell- WDwith other few other lifetime distributions such as EOW-ED, EWD, EFD, BXIID and CBell- WD. To draw the OC curve of ChSP- 1, the optimal parameters are chosen under each distribution when α = 0.05, β = 0.25, a = 0.5, r_2 = 6 and considering the model's shape parameter as η = 1, θ = 2 for EOW-ED (n = 9, i = 3), and fixing η = 1, θ = 3 for EWD (n = 17, i= 2), fixing η = 1, θ = 3 for EFD (n = 11, i= 3), fixing η = 1, θ = 4 for BXIID (n = 23, i= 4) and fixing the parameter η = 1, θ = 1.2 for CBell- WD (n = 7, i=2). Figure 3 show that the OC curve of the proposed plan under the CBell- WD is dominating the same under all the other lifetime distributions. In simpler terms, when using the ChSP-1 with CBell-WD, is accurately distinguishes between the good batches (when the chances of accepting high probability acceptance, $L_A(p)$) and reject the bad batches (when the failure probability, p is high).Therefore it is clear that the discriminating power of the proposed ChSP- 1 based on the CBell- WD is better than the other lifetime models.

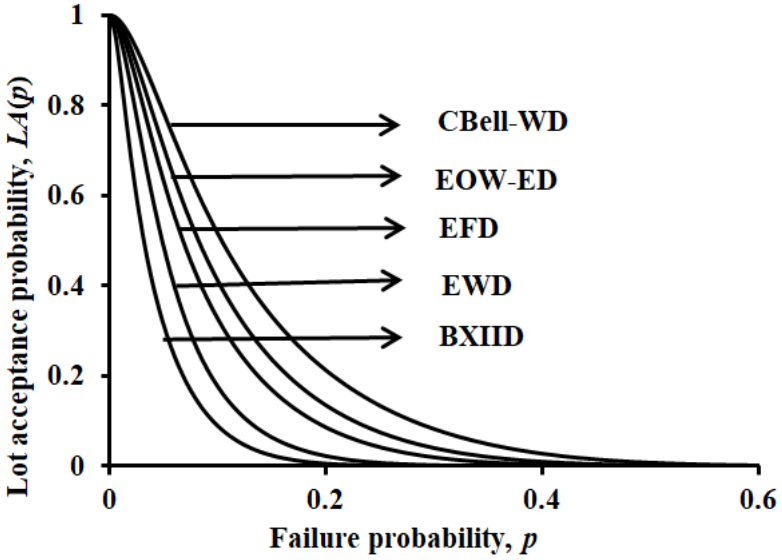


Figure 3: *OC curves of proposed ChSP-1 based on different life time models*

VII. Industrial Application

In this section, we address the execution of the proposed ChSP- 1 plan based on CBell-WD, utilizing the lifetime of Kevlar 373/epoxy fatigue fracture is the well fitted dataset analysed from the study of Algarni [2]. Kevlar 373, is a species of aramid fiber known for its strength, it is intricately combined with epoxy resin, which serves as a binding agent to coalesce the fibers within the composite material. The epoxy resin, which acts as a binder for the fibers, is susceptible to cracking due to repeated stress. Setting a constant pressure at 90% of the stress level is a unique feature. This is likely near to the material's endurance limit, where the stress level of failure is minimal even after testing a high number of cycles. By Understanding the material's behaviour under sustained 90% stress level can be informative for design and facilitate the prediction of long-term performance characteristics. The Kevlar 373/epoxy fatigue fracture dataset specifically measures the lifetime, in terms of number of cycles that a sample can endure before failing, with the constant fatigue stress level set at 90% of the material's limit.To illustrate the implementation of the proposed ChSP- 1 for assuring the median lifetime of Kevlar 373/epoxy fatigue fracture are used in industrial machinery, specifically their lifetime are measured in number of cycles. The

dataset contains 76 observations which are given by Gómez-Déniz*et al.* [8].

According to Algarni [2] for the above data set, the maximum likelihood estimator of the model parameters under CBell-WD are found to be η = 1.6316, θ = 0.6734(i.e., $\hat{\eta}$ =1.6, $\hat{\theta}$ =0.7). The manufacturer aims to investigate how the proposed ChSP- 1 can be used to reliably assure the median lifetime for Kevlar 373/epoxyfatigue fracture (measured in cycles) under 90% stress level. Suppose that the median number of cycles in the fatigue fracture of Kevlar 373/epoxy is specified (m_0) as 1.257 cycles but the quality inspector decide to run the experiment before complete the testing for only 0.6285 cycles (t_0 = 0.628 cycles). Therefore, the test termination ratio of a = 0.5.The producer's risk is specified as 5% and the consumer risk's as 25%, when the true median number of cycles (m) in Kevlar 373/epoxy is 7.542 cycles at least thrice the specified number of cycles as 1.257 cycles therefore, the median ratio becomes m/m_0 = 6. It is estimated that the model parameters of CBell - WD using the fatigue fracture of Kevlar 373/epoxy as $\hat{\eta}$ = 1.6, $\hat{\theta}$ = 0.7. Based on the above values, the optimal plan parameters n = 7 and i= 2 are depicted from Table 5. Suppose the simulated breaking strength of Kevlar 373/epoxyfatigue fracture (measured in cycles) of 7 materials are as given below.

0.900	1.5728	2.920	3.399	5.230	7.443	9.061

The implementation of the proposed ChSP- 1 using Kevlar 373/epoxyfatigue fracture (measured in cycles) under 90% stress level is explained below.

Step 1: A random sample of 7 items is selected from the current lot of Kevlar 373/epoxyand count the number of defective items before complete the life testing 0.628 cycles.

Step 2: Accept the current lot if no sample item is found to be defective and rejects the current lot if more than one sample item is defective before 0.628 cycles.

Step 3: If the defective item is equal to one then accept the current lot reaching before 0.628 cycles under the condition that preceding 2 lots have been accepted before 0.628 cycles.

VIII. Economic Design of ChSP-1 Using CBell-WD

The implementation of a cost-effective ChSP-1 has been proactively undertaken, driven by industry recognition of its potential to significantly reduce overall product inspection costs. By implementing the proposed plan, producers can avoid disrupting inspection costs and reworking defective products, thereby ensuring high-quality products reach customers. This approach not only maintains customer satisfaction but also enhances the producers' reputation in the market by offering quality products at minimal cost. Ailor *et al.* [3] determined the economic designing of sampling plan based on the combination of attributes and variables. Hsu and Hsu [9] proposed an economic design for a two-stage supply chain, which utilizes a single acceptance sampling plan to optimize efficiency and cost. Despite the existing literature, no prior work has ventured into an economic model for a chain sampling plan based on different lifetime distribution. Our work bridges this gap by introducing an economic design based on the CBell -WD which outperforms other lifetime distribution in terms of Average total inspection (ATI) and Total Cost (TC). In order to calculate the *TC*, we need to consider few economic measures, such as *ATI*, detection of failure

items (D_d) and non- detection of failure items (D_n). These measures are defined as follows. (see Hsu and Hsu [9]).

Table 5: *Optimal parameters of the proposed ChSP-1 under the CBell - W model with* $\hat{\eta} = 0.7$ *and* $\hat{\theta} = 1.6$

β	r_2	a = 0.5				a = 1.0			
		n	i	$L_A(p_1)$	$L_A(p_2)$	n	i	$L_A(p_1)$	$L_A(p_2)$
0.25	4	*	*	*	*	*	*	*	*
	6	7	2	0.9577	0.2192	3	1	0.9774	0.1718
	8	7	2	0.9748	0.2192	3	1	0.9881	0.1718
	10	7	2	0.9830	0.2192	3	1	0.9926	0.1718
0.10	4	*	*	*	*	*	*	*	*
	6	*	*	*	*	4	1	0.9609	0.0781
	8	12	1	0.9574	0.0787	4	1	0.9791	0.0781
	10	11	2	0.9611	0.0838	4	1	0.9869	0.0781
0.05	4	*	*	*	*	*	*	*	*
	6	*	*	*	*	8	0	0.9510	0.0351
	8	*	*	*	*	5	1	0.9680	0.0361
	10	14	1	0.9611	0.0478	5	1	0.9798	0.0361
0.01	4	*	*	*	*	*	*	*	*
	6	*	*	*	*	*	*	*	*
	8	*	*	*	*	*	*	*	*
	10	*	*	*	*	7	1	0.9620	0.0082

(*) refers plan doesn't exist

$$ATI(p) = n + \{1\text{-}L_A(p)\}(N\text{-}n) \quad (10)$$

$$D_d = np + (1\text{-}\ L_A(p)) \times (N - n) \times p \quad (11)$$

$$D_n = p \times L_A(p) \times (N - n) \quad (12)$$

where N is the lot size and $\{1-L_A(p)\}$ is the probability of rejection of the lot with failure probability p and $L_A(p)$ has been given in equation (5). The probability that the lot will be accepted ($L_A(p)$) under the economic ChSP-1 at the specified failure probability p in equations (10), (11), and (12) which is calculated for the median ratios of PQL and CQL, respectively. To calculate the TC, we consider three cost measures such as C_o, C_i and C_f as used by Hsu and Hsu [9] and are defined as C_o - cost of an outgoing failure item, C_i - life testing cost per item and C_f - replacement cost. In order to determine the economic ChSP-1 which minimizes the TC, we use the following optimization model;

$$\begin{aligned} &Minimize\ TC = C_i \cdot ATI + C_f \cdot D_d + C_o \cdot D_n \\ &Subject\ to\ L_A(p_1) \geq 1\text{-}\ \alpha \\ &\qquad L_A(p_2) \leq \beta \\ &\qquad n > 1,\ i \geq 1 \end{aligned} \quad (13)$$

where $L_A(p_1)$ and $L_A(p_2)$ are the probabilities of lot acceptance at PQL and CQL and that can be obtained from an equation (5) by substituting p_1 and p_2 respectively instead of p. The plan parameters namely (n, i) and the economic performance measure includes $ATI(p)$, $L_A(p)$, D_d, D_n, and TC have been estimated from the input values of model parameters using median life assurance which are given as (a, η, θ) = (0.5, 3, 1.25) and (0.5, 2, 1) by fixing the lot size N =1000 and cost parameters (C_i, C_f and C_o) = (1.0, 2.0 and 10). The values of failure probability (p) are defined to the

ratios 2.5, 3.5, 4.5, 5.5 which are corresponds to the average ratios of PQL (r_2 = 4, 6, 8 and 10) and CQL (r_1=1) which are reported in Table 6 and 7. From these tables we observe that if the consumers risk β decreases, then the values of $ATI(p)$, D_d and TC get increased while the value of D_n decreases. Similarly, when the median ratio (r_2) increases, then the values of $ATI\ (p)$ and overall TC get decreased.

Table 6: *Optimal parameters of economic ChSP-1 under the CBell- WD for assuring median life when $\eta = 3$ and $\theta = 1.25$, $a = 0.5$*

β	r_2	$L_A(p)$	n	i	D_d	D_n	$ATI(p)$	TC
0.25	4	0.9860	47	2	0.10	1.60	60.32	76.55
	6	0.9969	44	4	0.02	0.58	46.94	52.88
	8	0.9929	44	4	0.01	0.27	44.67	47.47
	10	0.9997	44	4	0.01	0.15	44.20	45.73
0.10	4	0.9768	79	1	0.17	1.53	100.27	115.96
	6	0.9951	74	2	0.04	0.56	78.45	84.24
	8	0.9985	73	3	0.02	0.26	74.38	77.10
	10	0.9995	73	3	0.01	0.14	73.42	74.91
0.05	4	0.9651	99	1	0.22	1.48	130.41	145.70
	6	0.9922	95	2	0.06	0.55	102.02	107.70
	8	0.9981	95	2	0.02	0.26	96.63	99.31
	10	0.9994	95	2	0.01	0.14	95.50	96.96
0.01	4	0.9301	147	1	0.35	1.35	206.62	220.86
	6	0.9889	147	1	0.10	0.52	156.42	161.82
	8	0.9974	147	1	0.04	0.24	149.20	151.75
	10	0.9982	145	3	0.23	0.13	146.50	147.90

Table 7: *Optimal parameters of economic ChSP-1 under the CBell- WD for assuring median life when $\eta = 2$ and $\theta = 1$, $a = 0.5$*

β	r_2	$L_A(p)$	n	i	D_d	D_n	$ATI(p)$	TC
0.25	4	0.9441	16	1	0.98	12.82	70.99	201.22
	6	0.9840	16	1	0.22	6.72	31.71	99.37
	8	0.9939	16	1	0.09	4.08	21.97	62.98
	10	0.9972	16	1	0.05	2.73	18.72	46.16
0.10	4	*	*	*	*	*	*	*
	6	0.9661	24	1	0.39	6.54	57.05	123.31
	8	0.9867	24	1	0.15	4.02	36.89	77.40
	10	0.9938	24	1	0.08	2.70	29.96	57.15
0.05	4	*	*	*	*	*	*	*
	6	0.9494	30	1	0.54	6.40	79.00	144.03
	8	0.9799	30	1	0.20	3.96	49.48	89.57
	10	0.9906	30	1	0.99	2.67	3.10	66.09
0.01	4	*	*	*	*	*	*	*
	6	*	*	*	*	*	*	*
	8	0.9578	45	1	0.35	3.81	85.22	124.12
	10	0.9798	45	1	0.17	2.60	64.26	90.68

(*) refers plan doesn't exist

IX. Comparative Study of ChSP-1 Performance In Terms of ATI and TC Curves

This section offers a comprehensive analysis of the *ATI* and *TC* curve of ChSP-1 plan using the CBell-WD under median lifetime assurance. The following section consists of two subsections. In Section 7.1, we present the *ATI* and *TC* curves and analyse the performance of the proposed plan using CBell-WD based on median lifetime with SSP. In Section 7.2, we proceed to compare the *ATI*and *TC* curve of the proposed plan assuring the product's median lifetime based on CBell-WD and the same based on some other distinct lifetime distributions.

I. *ATI* and *TC* Comparison of the Proposed ChSP-1 Plan with SSP

This section conducts a comparison analysis of the proposed plan by using the cost metrics such as *ATI*and *TC*. Table 8 compares the measures of *ATI* and *TC* curve based on the economic perspective of SSP and ChSP-1 with the input parameters are η = 3 and θ = 1.250 and the termination ratio a = 0.5 for assuring the median lifetime of the product under CEBT-ED. As shown in Table 8, the proposed ChSP-1 plan consistently outperforms SSP in terms of *ATI* and *TC* with the (%) decrement between these two plans across all combinations of β and r_2 indicating the better decision-making capabilities while reducing the quality inspection costs. Consider the scenarios, where the cost parameters (C_i= 1.0, C_o = 10.0 and C_f= 2.0), N = 1000, α = 0.05, β = 0.05, r_2 = 6. In this case, the *ATI* of SSP is 146.56 and ChSP-1 is 102.02, the percentage reduction of these two plan is 30.40% and finally the TC of SSP is 152.00 and ChSP-1 is 107.70, resulting in a 29.14% reduction. Therefore, Table 8 illustrates that the proposed ChSP-1 plan consistently offers better performance than SSP, in terms of *ATI*and *TC* is minimum, across various input scenarios, confirming our earlier expectations. Additionally, in order to draw the TC curve by following the above specified input parameters. Figure 4 portray the economic ChSP-1 parameter with n = 95, i= 2 and SSP parameters with n = 95 and c = 0. From this figure, it is apparent that a lot inspection has been performed excellent when the TC curves of economic ChSP-1 is more effective than SSP.

Table 8: *Comparison of ATI(p) and TC of economic SSP and ChSP-1 under CBell-WD for median life assurance when η = 3 and θ = 1.250, a = 0.5*

β	r_2	*ATI*(*p*)		*Reduction* in (%)	*TC*		*Reduction* in (%)
		SSP	*ChSP -1*		*SSP*	*ChSP -1*	
0.25	4	94.83	60.32	36.40%	110.60	76.55	30.78%
	6	69.83	46.94	32.78%	75.60	52.88	30.05%
	8	56.11	44.67	20.38%	58.88	47.47	19.38%
	10	50.64	44.20	12.71%	52.16	45.73	12.32%
0.10	4	139.70	100.27	28.22%	154.85	115.96	25.11%
	6	113.85	78.45	31.10%	119.47	84.24	29.49%
	8	92.41	74.38	19.51%	95.10	77.10	18.93%
	10	83.67	73.42	12.25%	85.15	74.91	12.02%
0.05	4	173.40	130.41	24.80%	188.09	145.70	22.53%
	6	146.56	102.02	30.40%	152.00	107.70	29.14%
	8	119.58	96.63	19.19%	122.20	99.31	18.73%
	10	108.53	95.50	12.00%	109.98	96.96	11.84%
0.01	4	249.96	206.62	17.33%	263.60	220.86	16.21%
	6	216.07	156.42	27.61%	221.17	161.82	26.83%
	8	180.19	149.20	17.20%	182.67	151.75	16.92%
	10	164.43	146.50	10.90%	165.81	147.90	10.80%

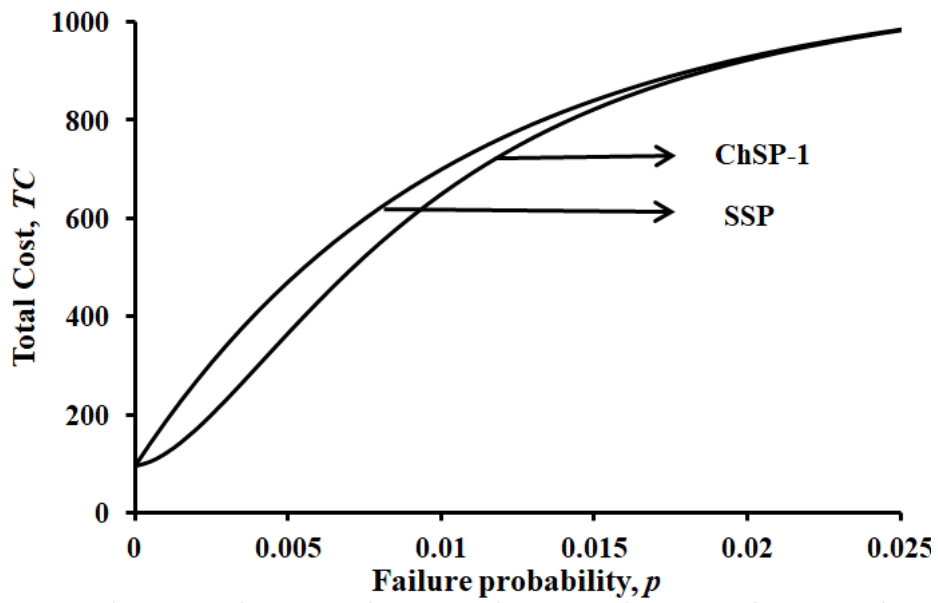


Figure 4: *TC curves of proposed ChSP-1 and SSP*

II. Comparison of *ATI* and *TC* of Proposed ChSP-1 Plan Based on Distinct Lifetime Distributions

This section conducts a comparison analysis of the proposed plan by using the cost metrics such as *ATI* and *TC*. Here, we focus on, how the ChSP-1 plan performs, in terms of *ATI*and *TC* by using the different distributions such as EOW-ED, EWD, EFD, BXIID and CBell- WD respectively. To draw the *ATI*and *TC* curve of ChSP-1, the plan parameters are chosen under each distribution when α = 0.05, β = 0.25, a = 0.5, r_2 = 6 and following the same shape parameter as η = 2, θ = 4 for EOW-ED (n = 35, i = 4), and fixing η = 2, θ = 1.9 for EWD (n = 18, i= 3), fixing η = 2, θ = 1 for EFD (n = 22, i= 3), fixing η = 2, θ = 4 for BXIID (n = 28, i= 3) and fixing η = 2, θ = 1 for CBell- WD (n = 16, i=1) respectively. Figure 5 and 6 shows that economic *ATI*and *TC*curve of the proposed plan under the CBell- WD outperforms other lifetime distributions. Additionally, Table 9 compares the measures of *TC* curve based on the economic perspective of different lifetime models by fixing the same shape parameters as η = 2 and the termination ratio a = 0.5. Consider the scenarios, where N = 1000, α = 0.05, β = 0.25, r_2 = 6. In this case, the *TC*of EFD is 810.73, the *TC*of EOW-ED is 221.11, the *TC*of BXIID is 186.91, the *TC*of EWD is 170.91, and the *TC* of CBell- WD attains 99.37. Table 9 reveals that the *TC*of ChSP-1 plan is minimum based on CBell- WD when compared with other lifetime models.

X. Advantages of ChSP-1 Based on CBell-WD

The following are the main advantages of the proposed ChSP-1 plan based on CBell-WD.

❖ The CBell-W distribution's pdf and cdf can be articulated as a linear combination of Weibull distribution also containing the unique features of Bell numbers. Therefore, CBell-WD is a derivative of the Bell-WD. This distribution has well-defined theoretical properties, making it easier to work and analyse in practical applications.

❖ ChSP-1 based on CBell-WD can be particularly useful for industrial applications like Kevlar 373/epoxy fatigue fracture, where early detection of potential failures may be crucial. ChSP-1 allows early termination of the life testing process (in this case we fix, t_0 = 0.628 cycles for materials) if the sample meets the acceptance criteria.For example, if the previous two lots were good, accepting a current lot with no defective item might be a reasonable decision.

❖ This plan will significantly reduce testing time and the associated costs compared to other sampling plans where all items are tested until failure.

❖ The shape parameter of CBell-WD provides a stable and flexible distribution when compared to other lifetime models also the CBell-WD observed high probability of acceptance when the failure probability is decreased.

❖ By implementing this economic approach, inspection costs can be significantly reduced when compare to other lifetime models and leading to more effective and efficient inspection processes.

Table 9: *Comparing TC of economic ChSP-1 under different lifetime models for median life assurance when a = 0.5*

β	r_2	*TC*				
		EOW-ED	EWD	*EFD*	*BXIID*	*CBell-WD*
		$\eta = 2,\ \delta=3$	$\eta = 2,\ \delta=1$	$\eta = 2,\ \delta=0.5$	$\eta = 2,\ \delta=1$	$\eta = 2,\ \theta=1$
0.25	4	443.59	343.65	*	372.92	201.22
	6	221.11	170.91	810.73	186.91	99.37
	8	132.28	103.37	534.81	112.89	62.98
	10	87.88	70.87	363.77	76.96	46.16
0.10	4	488.84	309.45	*	428.84	*
	6	242.02	196.49	843.76	216.44	123.31
	8	143.26	118.22	562.83	129.15	77.40
	10	95.15	80.62	382.76	87.43	57.15
0.05	4	528.45	430.28	*	463.52	*
	6	261.46	215.88	*	233.64	144.03
	8	155.51	132.25	583.65	140.69	89.57
	10	102.99	89.60	397.30	94.70	66.09
0.01	4	*	*	*	*	*
	6	305.59	263.80	*	281.60	194.62
	8	179.99	160.67	639.67	170.30	124.12
	10	121.41	110.24	438.47	117.14	90.68

(*) refers plan doesn't exist

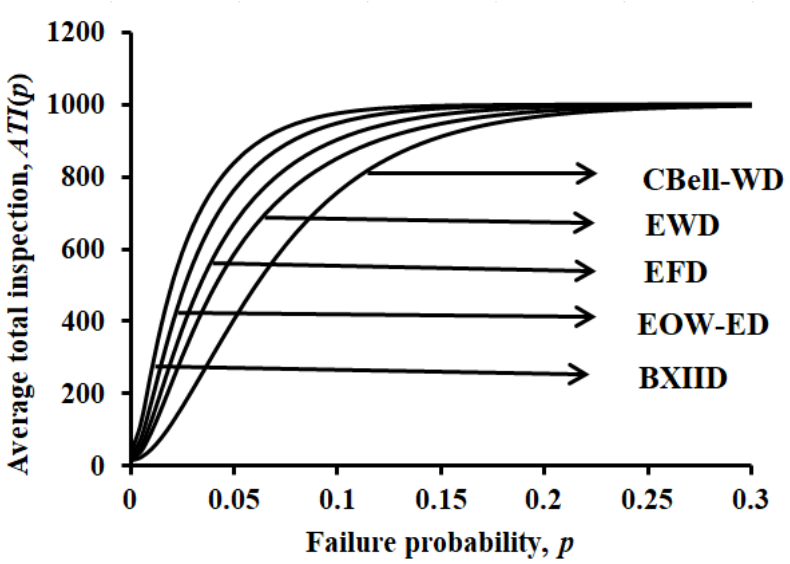


Figure 5: *ATI curves of proposed economic ChSP-1 under different lifetime models*

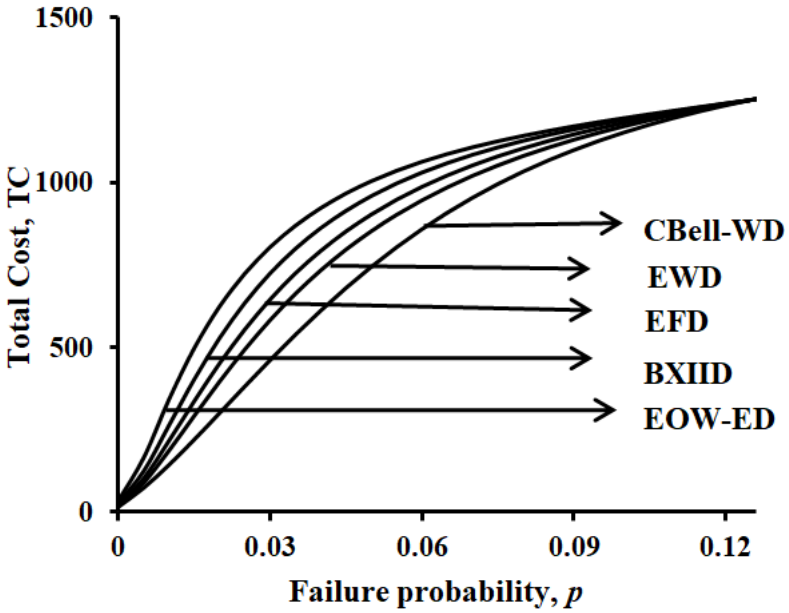


Figure 6. *TC curves of proposed economic ChSP-1 under different lifetime models*

XI. Conclusions

In this article, the proposed ChSP -1 presented for TTL test under median lifetime of the CBell -W model. By considering performance of two specific points on the OC curve, this approach allowed to determine the optimal parameters. The main comparison of the OC curves for both the proposed plan and SSP with Ac = 0 indicated that the proposed plan have higher level of discriminatory power when compared to the SSP. Also, the proposed plan under CBell-WD performed as the better model when compared to ChSP-1 based on other lifetime models. The proposed ChSP-1 plan under CBell-WD is applied in industry for fatigue fracture testing of Kevlar 373/epoxy composites at a 90% stress level, which is striking an optimal balance between test efficiency and result accuracy, thereby ensuring both reliable and efficient testing outcomes. The idea of quality control from an economic perspective may not be new idea of research; still it is relevant and worth considering, as it can lead to cost savings and efficiency gains in production processes by finding the right balance between inspection efforts and production outputs. An economic design of ChSP-1 is more effective than SSP when reducing *ATI* and TC, achieving the best performance at a lower inspection cost per unit. It has been also shown that the proposed sampling plan using CBell-WD is more economical than the existing other lifetime models. Therefore, the proposed plan will not only safeguard the producer needs but also guarantee the consumer satisfaction through the good product quality.

Acknowledgement

The first author wishes to convey her gratitude to the Kalasalingam Academy of Research and Education for granting her a fellowship opportunity.

References

[1] Aslam, M.and Jun, C. H. (2009). A group acceptance sampling plan for truncated life test having Weibull distribution.*Journal of Applied Statistics*, 36(9): 1021-1027.

[2] Algarni, A. (2022). Group acceptance sampling plan based on new compounded three-parameterWeibull model.*Axioms*, 11(9): 438.

[3] Ailor, R. H., Schmidt, J. W. and Bennett, G. K. (1975). The design of economic acceptance sampling plans for a mixture of variables and attributes. *AIIE Transactions*, 7 (4): 370-378.

[4] Castellares, F., Ferrari, S. L.and Lemonte, A. J. (2018). On the bell distribution and its associated regression model for count data. *Applied Mathematical Modelling*, 56: 172-185.

[5] Dodge, H. F. (1955). Chain sampling inspection plan. *Industrial Quality Control*, 11: 10-13.

[6] Govindaraju, K. and Subramani, K. (1993). Selection of chain sampling plans ChSP-1 and ChSP- (0, 1) for given acceptable quality level and limiting quality level. *American Journal of Mathematical and Management Sciences*, 13 (1-2): 123- 136.

[7] Govindaraju, K. and Balamurali, S. (1998). Chain sampling plan for variables inspection. *Journal of Applied Statistics*, 25 (1): 103- 109.

[8] Gomez–Deniz, E., Iriarte, Y. A., Gomez, Y. M., Chamorro, I. B. and Gomez, H. W.(2021). Statistical inference for a general family of modified exponentiated distributions. *Mathematics*, 9: 3069.

[9] Hsu, L. F. and Hsu, J. T. (2012). Economic design of acceptance sampling plans in a two-stage supply chain. *Advances in Decision Sciences*, 2012: 1-14.

[10] Jeyadurga, P. and Balamurali, S. (2019). Designing of chain sampling plan under gamma-Poisson distribution. *International Journal of Engineering and Advanced Technology*, 9(1S4): 2249- 8958.

[11] Jeyadurga, P., Usha, M. and Balamurali, S. (2018). Modified chain sampling plan for assuring percentile life under Weibull distribution and generalized exponential distribution. *International Journal of Quality and Reliability Management*, 35 (9): 1989- 2005.

[12] Jayalakshmi, S. and Veerakumari, K. P. (2022). Skip lot sampling plan of type SkSP-T with group acceptance sampling plan as reference plan under Burr Type XII distribution. *Reliability: Theory and Applications*, 17 (1(67)): 240-250.

[13] Raju, C. (1990). Designing chain sampling plans (ChSP-1) with fixed sample size. *International Journal of Quality and Reliability Management*, 7 (3): 59-64.

[14] Raju, C. and Raghottam, H. A. (2012). On designing chain sampling plan (ChSP-1) for average quality protection. *Model Assisted Statistics and Applications*, 7 (2):125- 130.

[15] Soundararajan, V. (1978). Procedures and tables for construction and selection of chain sampling plans (ChSP-1) Part 2: Tables for selection of chain sampling plans. *Journal of Quality Technology*, 10 (3): 56- 60.

[16] Soundararajan, V.and Govindaraju, K. (1982). A note on designing chain sampling plans ChSP-1, *The QR Journal*, 9 (3): 121-123.

[17] Soundararajan, V. and Govindaraju, K. (1983). Construction and selection of chain sampling plans ChSP-(0, 1). *Journal of Quality Technology*, 15 (4): 180- 185.

[18] Tahir, M. H.and Cordeiro, G. M. (2016). Compounding of distributions: A survey and new generalized classes. *Journal of Statistical Distribution and Applications*, 3(1): 1-35.

[19] Vijayaraghavan, R. and Saranya, C. R. (2021). Reliability single sampling plans under the assumption of Burr Type XII distribution. *Reliability: Theory and Applications*, 16 (4(65)): 308-322.

[20] Vijayaraghavan, R. and Pavithra, A. (2022). Reliability criteria for designing life test sampling inspection plans based on Lomax distribution. *Reliability: Theory and Applications*, 17 (2(68)): 245-258.

TWO-STAGE GROUP ACCEPTANCE SAMPLING PLAN FOR HALF-NORMAL DISTRIBUTION

C. Geetha[1], S. Jayabharathi[2], Mohammed Ahmar Uddin[3], and Pachiyappan D[4,*]

•

[1]Associate Professor, Department of Statistics, Government Arts college,
Salem-7, Tamil Nadu, India.
geethanisha1@gmail.com
[2]Associate Professor, Department of Mathematics, Sona college of technology,
Salem-5, Tamil Nadu, India.
jayaajay2005@yahoo.com
[3]Department of Finance and Economics, College of Commerce and Business Administration,
Dhofar University, Salalah, Dhofar, Oman.
Ahmar@du.edu.om
[4]Research Assistant, Department of Finance and Economics, College of Commerce and Business Administration, Dhofar University, Salalah, Dhofar, Oman
* Correspondence Email: d.pachiyappan321@gmail.com

Abstract

This paper proposes a time-truncated life test based on a two-stage group acceptance sampling plan for the percentile lifetime following a half-normal distribution. The optimal parameters for this plan are determined to simultaneously satisfy both producer's and consumer's risks for a given experimentation time and sample size. The efficiency of the proposed sampling plan is evaluated by comparing the average sample number with that of existing sampling plans. Industrial examples are provided to illustrate the application of the proposed sampling plan.

Keywords: Two-stage group sample, producer and consumer risk, normal distribution**.**

I. Introduction

In today's modern industrial environment, producing high-quality products using modern statistical quality control techniques has become essential. Product quality is a crucial factor contributing to business success, growth, and competitive advantage [1]. SQC techniques are now vital for any manufacturing process, as their application helps improve product quality by reducing process and product variability. SQC plays a significant role in the success of any industry. SQC involves a set of operational activities that an enterprise implements to ensure that its products meet the required quality levels set by consumers. According to [2], product quality can be evaluated based on various dimensions, including durability, serviceability, performance, aesthetics, features, reliability, and conformance to standards. These dimensions collectively determine the overall quality of a product. Therefore, quality has become the most significant factor in consumer satisfaction when selecting among competing products and services.

SQC involves a set of operating activities that an enterprise implements in order to get certified that the quality of its products is at required levels of the consumers. According to [3] one can evaluate the quality of the product in terms of its durability, serviceability, performance, aesthetics,

features, reliability, quality and its conformance to standards collectively and they are termed as the dimensions of quality. Hence, we can say that in selecting among competing products and services, the quality has become the most significant factor for consumer's satisfaction.

The quality of any finished product can be judged by inspecting a few items taken randomly from a lot of products and the process of taking such samples is called the sampling. In quality management, the acceptance sampling plans are vital tools in making a decision about the product whether to accept or reject based on the inspection of sampled items and the sampling plans prescribe the experimenter how many items in the sample should be selected from the submitted lot for inspection and how many defectives can be allowed in this sample in order to satisfy both the producer and the consumer. The probability of rejecting a good lot is called the producer's risk and the chance of accepting bad lot is called the consumer's risk. The cost of any life test experiment is directly proportional to the sample size. Therefore, sampling plans that provide smaller sample size for inspection and minimize two risks are considered as efficient sampling plans.

In general, the acceptance sampling plans can be classified into two types namely attribute sampling plans and variables sampling plans. The attribute sampling plans are implemented for quality characteristics which are expressed on a ''go, no go'' basis whereas the variables sampling plans can be applied where the quality characteristics of interest are measured on a numerical scale [4]. Various types of sampling plans such as single sampling plan, double sampling plan, multiple sampling plan, and sequential sampling plan are available in the literature [5]. Group acceptance sampling plan is one of the types of sampling plans which involves a number of testers to be used for testing so that cost and time can be reduced. The inspection of multiple items simultaneously can be made easy to the experimenter for testing. Two stage group acceptance sampling plan is the extension of GASP which involves two groups. The GASP is more advantageous than the conventional sampling plans in terms of minimum inspection so that the considerable testing time and cost can be reduced [6]. The advantage of two stage group sampling plan is that it reduces the average sample number as compared to the GASP.

Several authors have investigated the sampling plans under various life time distributions, which are available in the literature of acceptance sampling [7,8,9,10,11,12] proposed the SSP based on half normal distribution. By exploring the literature on two stage group sampling plans, it can be seen that no work is available based on the half normal distribution. In this paper, we will present the designing of two stage group sampling plan when the life time of an item follows the half normal distribution. The structure of proposed plan is presented and efficiency is compared with the existing sampling plan. The application of the proposed sampling plan is explained with the help of industrial illustrative examples.

2. Half-normal distribution

As far as the variables sampling plans are concerned, the normal distribution is the most preferred statistical distribution. But for life testing problems, normal distribution is not preferred because of its range [-∞, ∞]. However, one of the normal family distributions called the half-normal distribution is the widely used probability distribution for nonnegative data modeling, particularly, in life time testing. [13] investigated the properties of half normal distribution. [14] investigated the maintenance performance of the system under half-normal failure lifetime model as well as a repair-time model. The probability density function of a half normal distribution with 0 mean and its parameter θ with domain y ∈ [0, ∞] is given by

$$f(x) = \frac{2\theta}{\pi} e^{-\frac{y^2\theta^2}{\pi}}, y > 0, \theta > 0 \quad (1)$$

Its cumulative distribution function is given by

$$F(\mathrm{y}) = \mathrm{erf}\left(\frac{\theta y}{\sqrt{\pi}}\right),\ y > 0, \theta > 0 \quad (2)$$

Here erf is the "Error Function" defined by

$$\text{errf}\,(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt \tag{3}$$

Consider that life time of product follows a half-normal distribution with σ as a scale parameter. Its cdf is given by

$$F(y) = \text{erf}\left(\frac{\theta\left(\frac{t}{\sigma}\right)}{\sqrt{\pi}}\right),\ t > 0, \theta > 0, \sigma > 0 \tag{4}$$

The 100th percentile of the half normal distribution with $0 < p < 1$, is defined as

$$t_p = \frac{\sigma\sqrt{\pi}}{\theta}\text{erf}^{-1}\,(\text{p}),\ \text{ for } 0 < p < 1 \tag{5}$$

Here erf^{-1} is inverse function of error function. The Maclaurin series of $\text{erf}^{-1}\,(.)$ is given by

$$erf^{-1}(y) = \sqrt{\pi}\left(\frac{1}{2}y + \frac{1}{24}\pi y^3 + \frac{7}{960}\pi^2 y^5 + \frac{127}{80640}\pi^3 y^7 + \cdots\right) \tag{6}$$

According to Pewsey, "If Z is a standard normal random variable, Z~N(0,1), then Y=|Z| follows a standard positive half normal distribution and -Y=-|Z|. follows a standard negative half-normal distribution. The half-normal distribution is a central chi-square distribution with one degree of freedom and a special case of truncated and folded normal distributions". The half-normal distribution is also a limiting case of skewed normal distribution [15]. The applications of half-normal in reliability analysis can be seen in Ayman and Kristen [15]. As half-normal distribution has positively skewed shape, there is a need to model monotone hazard rates. The half-normal distribution is very widespread model to describe the lifetime process of any device under fatigue (see [16].

3. Two Stage Group Acceptance Sampling Plan

Naveed et al. [17] proposed the two-stage group sampling plan. The operating procedure of this plan is explained below.

3.1 Stage one

- Extract the first random sample of size n_1 from a lot submitted for inspection.
- Randomly assign r items to each of g_1 groups or testers so that $n_1 = rg_1$ and set them on test for the duration of t_0 units of time.
- Accept the lot if the total number, Y_1, of failures from g_1 groups is smaller than or equal to c_{1a}.
- Truncate the test and reject the lot as soon as the number of failures Y_1 reaches $c_{1r} (> c_{1a})$ during the test. Otherwise, go to stage two.

3.2 Stage two

- Extract a second random sample of size n_2 from the same lot.
- Randomly assign r items to each of g_2 groups so that $n_2 = rg_2$ and set them on test for the duration of t_0 units of time again.
- Let the total number of failures from the second sample be Y_2.
- Accept the lot if the total number, $Y_1 + Y_2$, of failures from g_1 and g_2 groups is smaller than or equal to $c_{2a} (> c_{1a})$. Otherwise, truncate the test and reject the lot.

The design parameters of above two stage group sampling plans are g_1, g_2, c_{1a}, c_{1r}, and c_{2a}. The acceptance number, c_{2a}, in the second stage is larger than the acceptance number, c_{1a}, in the first stage, since in two stage sampling total number of failures from both stages is used in decision making.

4. Designing of the proposed sampling plan

Based on the operating procedure of two-stage group sampling plan, the lot acceptance probability at the first stage is given as

$$P_a^{(1)} = P\{Y_1 \leqslant c_{1a}\} = \sum_{j=0}^{c_{1a}} \binom{rg_1}{j} p^j (1-p)^{rg_1-j} \tag{7}$$

The lot rejection probability at the stage one is as follows (see Aslam et al. (2012)).

$$P_r^{(1)} = \sum_{j=c_{1r}}^{rg_1} \binom{rg_1}{j} p^j (1-p)^{rg_1-j} = 1 - \sum_{j=0}^{c_{1r}-1} \binom{rg_1}{j} p^j (1-p)^{rg_1-j} \tag{8}$$

To accept the lot based on stage two, the total number, $Y_1 + Y_2$, of failures from both groups g_1 and g_2 must be smaller than or equal to c_{2a}. So, the lot acceptance probability at this stage under the proposed two stage sampling plan is as follows

$$P_a^{(2)} = P\{c_{1a} + 1 \leqslant Y_1 \leqslant c_{1r} - 1, Y_1 + Y_{2.} \leqslant c_{2a}\} \tag{9}$$

Thus, under the proposed two stage group acceptance sampling plan, the lot acceptance probability is as follows.

$$L(p) = P_a^{(1)} + P_a^{(2)} \tag{10}$$

Stated that in order to implement an acceptance sampling plan to assure the percentile lifetime in a truncated life test, it is convenient to determine the experiment time in terms of the specified percentile lifetime as $t_0 = \delta_{\mathrm{p}} t_{\mathrm{p}}^0$, where δ_{p} is called the termination time ratio. The probability of failure of an item before time t_0 is given as

$$p = F(t_0) = F\left(\delta_{\mathrm{p}} t_{\mathrm{p}}^0\right) \tag{11}$$

or

$$p = F(t_0) = \mathrm{erf}\left(\delta_{\mathrm{p}} erf^{-1}(pd)/\left(t_{\mathrm{p}}/t_{\mathrm{p}}^0\right)\right) \tag{12}$$

Where t_{p} = true unknown population pth percentile. The erf (.) and erf^{-1} (.) functions have been defined in Eqs. (3) and (6) respectively. The above expression is represented in terms of $t_{\mathrm{p}}/t_{\mathrm{p}}^0$ and is also called failure probability. Let α be the producer's risk and β be the consumer's risk. The producer wishes that the lot acceptance probability should be larger than $1-\alpha$ at various values of percentile ratio $t_{\mathrm{p}}/t_{\mathrm{p}}^0$ and the consumer wishes that it should be smaller than β at $t_{\mathrm{p}}/t_{\mathrm{p}}^0 = 1$. Therefore, the plan parameters of the proposed plan will be determined by minimizing ASN at consumer's risk using the following non-linear optimization problem [16]. Minimize ASN $(p_2) = \mathrm{rg}_1 + \mathrm{rg}_2\left(1 - P_a^{(1)} - P_r^{(1)}\right)$, Subject to $L(p_2) \geqslant 1-\alpha$

$$L(p_1) \leqslant \beta \tag{13}$$

$$g_1 > 1, g_2 > 1, r > 1, c_{1r} > c_{1a} > 0, c_{2a} > c_{1a} > 0 \tag{14}$$

5. Description of tables and industrial examples

Tables for the selection of optimal parameters of the proposed sampling plan are given for various specified requirements such as consumer's risk ($\beta = 0.25,0.10,0.05,0.01$), producer's risk ($\alpha = 0.05$), percentile ratio ($d_2 = 2,4,6,8$), ($r = 5$ or 10), $\delta_{\mathrm{p}} = 0.5$ or 1.0 and $d_1 = 1$. The tables are presented for average life ($p = 0.5$ and $p = 0.25$). The tables can be made for any other specified parameters. The R codes are available with the authors upon request.

The optimal parameters of the proposed sampling plan along with the ASN and the lot acceptance probability are presented in Tables 1-4. Tables 1 and 2 provide the optimal parameters of the two-stage group acceptance sampling plan to ensure median life time of the products and Tables 3 and 4 show the optimal parameters of the two-stage group acceptance sampling plan to

ensure lower percentile life. From Tables 1-4, we observe that when the ratio d_2 increases the number of groups for the experiment and the acceptance numbers decrease. Tables 2-4 will be made available on request.

6. Case study

For the implementation of proposed sampling plan, we will consider the data from a leading ball-bearing manufacturing company in Korea. The failure data are well fitted to the half normal distribution. Suppose that the company is interested to test the product using the proposed sampling plan. Let $\alpha = 5\%, \beta = 5\%$, and $p = 0.5$. The number of testers of the product is limited to $r = 5$ From Table 1, we have $c_{1r} = 3$, $c_{1a} = 0, c_{2a} = 2, g_I = 3$ and $g_2 = 2$. A sample of size 15 items are selected and distributed into three groups. The failure times for each of three groups are given as follows

Table 1: *Optimal parameters of the two-stage group sampling plan under half-normal distribution with δq=0.5, q=0.5*

β	d_2	$r = 5$							$r = 10$						
		c_{1r}	c_{1a}	c_{2a}	g_1	g_2	ASN	$L(p_2)$	c_{1r}	c_{1a}	c_{2a}	g_1	g_2	ASN	$L(p_2)$
0.25	2	9	6	13	7	5	41.34	0.953	10	7	13	4	2	44.54	0.956
	4	4	2	3	3	1	16.06	0.9671	4	2	5	2	1	21.14	0.9558
	6	3	0	2	2	1	12.19	0.9732	4	2	3	2	1	21.14	0.9717
	8	2	0	3	2	1	10.84	0.9574	3	1	9	2	1	20.53	0.9717
0.1	2	13	0	17	11	6	63.20	0.9554	16	10	17	6	3	72.45	0.9552
	4	4	2	5	4	3	21.72	0.9505	5	1	4	2	1	23.42	0.9534
	6	3	1	3	3	2	16.36	0.9644	4	2	4	2	1	21.14	0.9857
	8	3	1	2	3	1	15.68	0.9753	3	1	7	2	1	20.53	0.9717
0.05	2	14	4	20	12	9	71.26	0.9502	15	10	26	7	6	77.64	0.9529
	4	5	2	6	5	4	27.96	0.9614	5	3	8	3	2	30.92	0.9514
	6	4	1	3	4	2	21.67	0.9598	4	1	3	2	1	21.67	0.9598
	8	3	0	2	3	2	16.90	0.9519	3	1	4	2	1	20.53	0.9713
0.01	2	19	8	30	18	15	97.62	0.9521	19	2	31	9	8	98.13	0.9540
	4	6	1	8	7	5	36.74	0.9634	7	3	7	4	2	41.28	0.9528
	6	4	0	4	5	3	26.08	0.9580	4	0	6	3	2	30.51	0.9547
	8	3	0	3	4	3	21.03	0.9569	4	2	3	3	1	30.19	0.9648

Group-1	Group-2	Group-3
0.6825	1.5650	0.9232
1.8024	0.8981	0.0607
0.0509	0.7322	0.4541
1.2080	2.1866	1.0035
0.4275	0.4223	0.6611

Let $\delta_{\mathrm{p}} = 0.5$ and $t_{\mathrm{p}}^0 = 1.5$, these leads $t_0 = 0.075$. We note one failure from group 1, no failure from group 2 and 1 failure from group 3 before time $t_0 = 0.075$. So, the total number of failures from three groups is 2. As the total number of failures lies between $c_{1a} = 0$ and $c_{Ir} = 3$, a decision about the disposition of lot will be made on the basis of second sample. A second sample of size 10 is selected from the lot and distributed into two groups. The number of failures from two groups is given in the following table

Group-1	Group-2
0.8472	0.0701
0.7845	0.4341
0.5452	0.1104
0.1316	0.7054
0.2624	0.8239

From group 1, we note that no failure occurs before experiment time and from group 2, we note that only one failure occurs before the experiment time. Since, the total number of failures from both samples is larger than $c_{2a} = 2$, the lot of products will be rejected.

7. Comparison

In this section, a comparison is made between the GASP and the proposed two-stage group sampling plan. This comparison is made based on the ASN needed for both sampling plans. It is to be pointed out that the two-stage group sampling plan could have a chance to use the sample from the first stage or combined samples from both stages to make a decision. For example, if the product lifetime has a half-normal distribution, the next step would be to decide whether to use a GASP or to use the proposed two stage group sampling plan which will have a minimum ASN. Here we compare both plans for experiment termination time ($\delta = 0.50$ and 1.0) with $p = 0.5$ median life time quality level.

Tables 5 and 6 provide the ASN for both acceptance sampling plans, where presumed producer's risk was set as $\alpha = 0.05$. It has been witnessed from tables that the ASN for the proposed two-stage group sampling plan is much smaller than the GASP at lower ARL mean ratios with termination time $\delta_{\mathrm{p}} = 0.5$ and 1.0. So, it is concluded that the two-stage group acceptance sampling plan is better than the GASP as it provides lesser ASN at lower percentile ratios for accepting or rejecting a lot in case of median life time quality.

As in Table 5, when $\delta_{\mathrm{p}} = 0.5$ with $p = 0.5$ and consumer's risk is set as 0.05 with ratio 2, ASN for GASP when $r = 5$ is 110 and the ASN of the proposed two stage group acceptance sampling plan reduces to 71.26. Likewise, from Table 6, when $\delta_{\mathrm{p}} = 1.0$ and $p = 0.5$ with same consumer's risk and ratio, the ASN for GASP is 50 but for proposed plan, it reduces to 38.50. Table 6 will be made available on request.

8. Concluding remarks

In this paper, two stage sampling plan has been proposed for the inspection of products whose life time follows a half-normal distribution. Tables have been presented for industrial applications of the proposed sampling plan. The efficiency of the proposed sampling has been compared with the existing single stage group sampling plan. It is concluded that the proposed sampling plan is more efficient in reducing the ASN for the life test experiment. The real time applications of the proposed sampling plan are given using the industrial data. The proposed sampling plan can also be used in testing of software

Table 5: *ASN for GASP and two-stage group sampling plan under half-normal distribution* $\delta_p = 0.5$ *when* $p = 0.5$.

B	d_2	GASP with $r = 5$	GASP with $r = 10$	Proposed two-stage plan	
				$r = 5$	$r = 10$
0.25	2	65	70	41.34	44.54
	4	20	*	16.06	*
	6	15	*	12.19	*
0.1	2	95	100	63.20	72.45
	4	30	30	21.72	23.42
	6	25	30	16.36	21.14
	8	20		15.68	
0.05	2	110	40	71.26	77.64
	4	40	30	27.96	30.92
	6	30	160	21.67	21.67
	8	30	60	16.90	20.53
0.01	2	155	40	37.62	98.13

4	60	40	26.74	41.28
6	40		21.03	30.51
8	35			

Optimal plan does not exist. reliability, testing/lot sizing of electronic product, automobile industry and mobile manufacturing industry. The efficiency of the proposed sampling plan using a cost model can be considered as future research.

Discloser statement

The authors declare no potential conflict of interest.

References

[1] Naveed, M., Azam, M., Khan, N., Aslam, M., Saleem, M., & Saeed, M. (2024). Control charts using half-normal and half-exponential power distributions using repetitive sampling. Scientific Reports, 14(1), 1-20.

[2] Ayman, A., Kristen, K., (2013). On the gamma half-normal distribution and its applications. J. Mod. Appl. Stat. Methods 12 (1), 103-119.

[3] Baklizi, A., El Masri, Q.E.A., (2004). Acceptance sampling plan based on truncated life test in the Birnbaum-Saunders model. Risk Anal 24 (6), 1453-1457.

[5] Balakrishnan, N., Leiva, V., Lopez, G., (2007). Acceptance sampling plan based on the truncated life test in generalized BirnbaumSaunders distribution. Commun. Stat. -Simul. Comput. 36, 643656.

[6] Balamurali, S., Jun, C.-H., (2006). Repetitive group sampling procedure for variables inspection. J. Appl. Stat. 33 (3), 327-338.

[7] Jun, C.-H., Balamurali, S., Lee, S.-H., (2006). Variables sampling plans for Weibull distributed lifetimes under sudden death testing. IEEE Trans. Reliab. 55 (1), 53-58.

[8] Kantam, R.R.L., Rosaiah, K., Rao, G.S., (2001). Acceptance sampling based on life tests: Log-logistic models. J. Appl. Stat. 28 (1), 121128.

[9] Khan, M.A., Islam, H., (2012). Bayesian analysis of system availability with half-normal life time. Qual. Technol. Quant. Manage. 9 (2), 203-209

[10] Lio, Y.L., Tsai, T.-R., Wu, S.-J., (2010). Acceptance sampling plan for truncated life test based on the Burr Type XII percentiles. J. Chin Inst. Ind. Eng. 27 (4), 270-280.

[11] Lu, X., Gui, W., Yan, J., (2013). Acceptance sampling plans for half normal distribution under truncated life tests. Am. J. Math Manage. Sci. 32 (2), 133-144.

[12] Luis, M.C., Hector, W.G., Maria, V., (2012). Epsilon half-normal model: properties and inference. Comput. Stat. Data Anal. 56, 4338-4347.

[13] Montgomery, D.C., (2009). Introduction to Statistical Quality Control, Sixth ed. John Wiley & Sons, Inc.

[14] Pewsey, A., (2004). Improved likelihood-based inference for the general half-normal distribution. Commun. Stat. Theory Methods 33 (2), 197-204.

[15] Rao, S.G., (2009). A group acceptance sampling plan for lifetimes following a generalized exponential distribution. Econ. Qual. Control 24 (1), 75-85.

[16] Rao, S.G., Kantam, R.R.L., (2010). Acceptance sampling plans from truncated life tests based on log-logistic distribution for percentiles. Econ. Qual. Control 25 (2), 153-167.

[17] Rodrigo, R.P., Clarice, G.B.D., Gauss, M.C., Edwin, M.M.O. Mariana, R.U., (2010). The beta generalized half-normal distribution. Comput. Stat. Data Anal. 54 (4), 945-957.

RT&A, No 4(80)
Volume 19, December, 2024

A NEW COMPACT DETECTION MODEL FOR LINE TRANSECT DATA SAMPLING

Ishfaq S. Ahmad[1], , Rameesa Jan[2]*,

•

[1] Department of Mathematical Sciences,
Islamic University of Science and Technology, Awantipora, J&K, India;
[2] Department of Statistics,
Government Degree College Sopore, Baramulla, J&K, India;
[1] peerishfaq007@gmail.com [2] bhatrumaisa2@gmail.com

Abstract

A new parametric model is proposed in line transect sampling for perpendicular distances density functions. It is simple, compact and monotonic non increasing with distance from transect line and also satisfies the shoulder condition at the origin. Numerous interesting statistical properties like shape of the probability density function, moments, and other related measures are discussed. Method of Moments and Maximum Likelihood Estimation is carried out .Applicability of the model is demonstrated using a practical data set of perpendicular distances and compared with other models using some goodness of fit tests.

Keywords: Line transect; shoulder conditions; detection function; maximum likelihood estimation; perpendicular distance.

1. Introduction

Line transect approach is a key technique for determining the population abundance(D) or density of objects in a study region(A). These objects may be species of animals, birds or plants that are easily visible at close range (Buckland et al. [**?**], Buckland et al. [**?**] and Barabesi [**?**]). It is the easiest, most useful and inexpensive of all the population abundance estimation. In a typical application of Line transect method, an observer walks a straight path of length L, noting all the animals seen (n) and their right-angle distances from the transect line (x).

In order to determine D from this data, a model is required which can be mathematically represented by conditional function $h(x)$ knows as detection function which is defined as :

$$h(x) = \Pr(\text{an object is detected given its perpendicular distances x from line})$$

where $0 < x < \zeta$, and ζ is the limiting value of perpendicular distance at which the observations are made. To demonstrate the fruitfulness of a detection model, numerous assumptions are to be made [Buckland et al. [**?**] and Miller and Thomas [**?**]]. It is logical to say that objects which are far away from transect line have least chances of detection and therefore in mathematical terms we can say that h(x) is assumed to be monotonically non increasing with respect to x. Furthermore, $h(0) = 1$, implies "with probability 1 objects on the path will be spotted" and the detection probability should approach to 1 at a distance approaching to 0. Additionally, tangent slope is 0 at $x = 0$ (i.e.,$h'(0) = 0$, indicating flat at zero distance) depicting horizontal tangent thereby h(x) satisfying the shape rule. These are the shoulder conditions which must exist in any

detection model. Buckland [**?**] and Buckland et al. [**?**] mentioned some other prominent features of line transect sampling :

- N entities are randomly distributed over A with $D = N/A$.
- $h(0) = 1$.
- Entities are found at the initial observing place.
- No entity is counted included twice.
- Perpendicular distances are noted without errors.
- Entities are distributed independently of the line.
- Some, perhaps many, entities will be missed.

The elementary relation for evaluating the density of entities in a particular area [Burnham and Anderson [**?**] and Seber [**?**]] can be stated as

$$D = \frac{E(n)j(0)}{2M}, \tag{1}$$

where $E(n)$ is the expected value of the number of spotted entities. Burnham and Anderson [**?**] showed the general estimate for D as:

$$\hat{D} = \frac{n\hat{j}(0)}{2M}, \tag{2}$$

$\hat{J}(0)$ is a suitable sample estimator of $j(0)$ based on 'n' examined distances $x_1, x_2, \ldots, x_n$. When objects are observed from a line transect with a detection function $h(x)$, the distance X to the observed object from a randomly placed transect will tend to have a pdf j(x) of the same shape as h(x) , but scaled so as the area under j(x)=1 i.e,

$$j(x) = \frac{h(x)}{\kappa}, \tag{3}$$

where $\kappa = \int_0^\zeta h(x)dx$ is the normalizing constant and ζ is taken to as ∞. $j(x)$ satisfies the shoulder conditions iff $f^{'}(0) = 0$ and and $f(x)$ is monotonically non increasing[Eberhardt [**?**]]. This condition is one of the most important criteria for a robust estimation of $j(0)$ which is related to the properties of the proposed model for $j(x)$ [Crain et al. [**?**]]. Numerous parametric and non parametric methodologies have been proposed to estimate j(0). This article focuses on the parametric method to estimate the parameters using MLE. Hence, an estimator of j(0) and D is obtained.

The layout of the article is outlined as: In Section **??**, a new single-parameter detection model (SPDM) satisfying the shoulder conditions has been introduced. Some intriguing properties have been discussed in Section **??**. All the related expressions of this model have closed forms and hence easy to work out. Section **??** deals with estimation of the parameters and the practical application of the model is being described in Section **??** . Lastly, the article is completed with some remarks in Section **??** .

2. The Proposed Model

Suppose the detection function of SPDM with parameter β $(\beta > 0)$ is given by

$$h(x;\beta) = \left(3 - 2e^{-\frac{x}{\beta}}\right) e^{-\frac{2x}{\beta}}, \qquad 0 \le x < \infty, \quad \beta > 0. \tag{4}$$

The detection function (**??**) satisfies all the shoulder conditions; $h(0) = 1$ making it impeccable for detection on the transect line path. The first derivatives of (**??**) w.r.t x are, respectively, given by

$$h'(x) = -\frac{6}{\beta}\left(e^{\frac{x}{\beta}} - 1\right) e^{-\frac{3x}{\beta}},$$

$\implies h'(0) = 0 \; \forall \beta$. Since $e^{-\frac{3x}{\beta}} > 0 \quad \forall \quad x \quad \in (0, \infty)$, implies that $h'(x) = -\frac{6}{\beta} < 0 \quad \forall \quad \beta > 0$ which means that (**??**) is monotonically decreasing $\forall x \in (0, \infty)$. Figure **??** confirms all the shoulder conditions of the detection function.

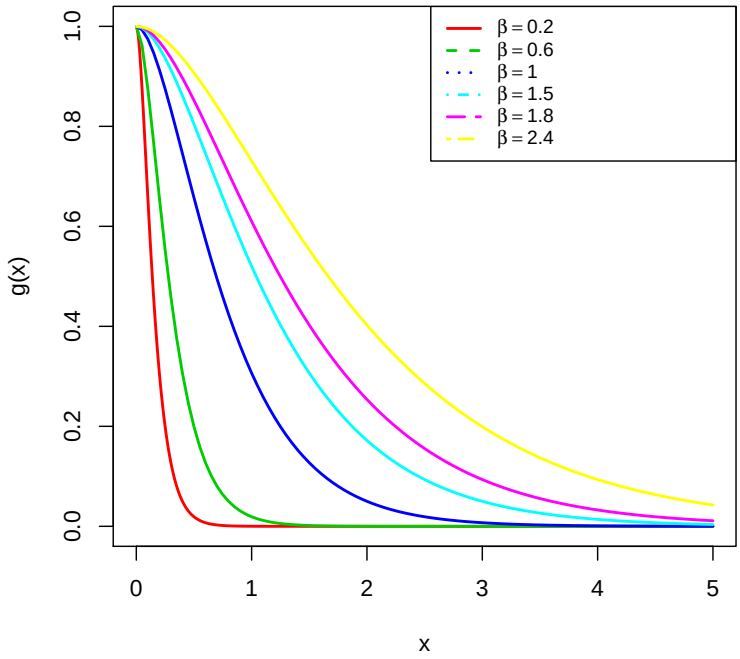


Figure 1: *Plot of detection function for different choices of parameter.*

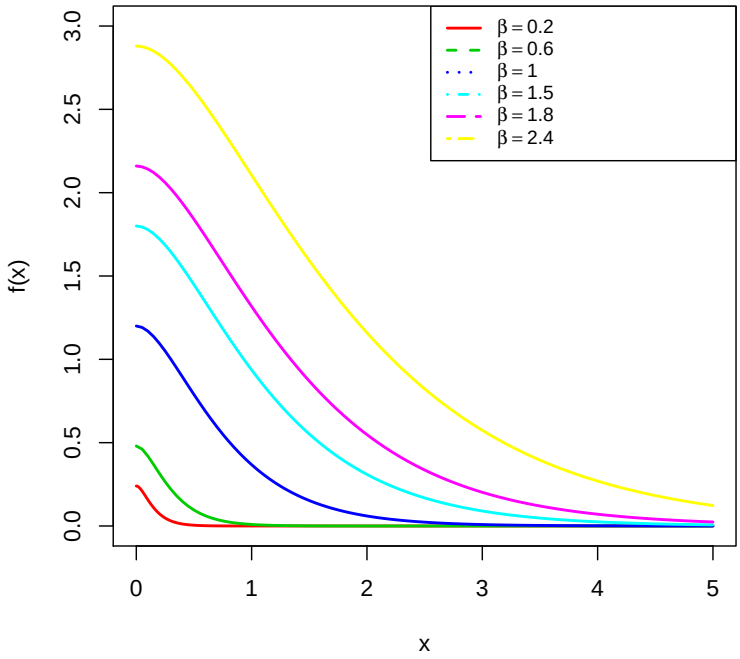


Figure 2: *pdf plot for several choices of parameter.*

Now the corresponding pdf is obtained by substituting (**??**) in (**??**) as:

$$j(x;\beta) = \frac{6}{5\beta}\left(3 - 2e^{-\frac{x}{\beta}}\right)e^{-\frac{2x}{\beta}}, \quad x > 0, \quad \beta > 0. \tag{5}$$

The pdf plot of SPDM model for different choices of parameter β is exhibited in Figure **??**. The cummulative distribution function (cdf) corresponding to pdf (**??**) is :

$$J(x;\beta) = 1 - \frac{1}{5}\left[9 - 4e^{-\frac{x}{\beta}}\right]e^{-2\frac{x}{\beta}}, \quad x > 0, \quad \beta > 0. \tag{6}$$

Since $h(0) = 1$ and if we substitute $x = 0$ in (**??**), we will get

$$j(0) = \frac{6}{5\beta}, \quad \beta > 0. \tag{7}$$

In the light of above expression, the pdf of SPDM can be phrased as

$$f(x;\beta) = f(0)\left(3 - 2e^{-\frac{x}{\beta}}\right)e^{-\frac{2x}{\beta}}, \quad x > 0, \quad \beta > 0, \tag{8}$$

which is a function of β and $j(0)$ and will serve as the base of the MLE maximum of β and $j(0)$.

The first derivative of pdf (**??**) w.r.t x is

$$\frac{\partial j(x;\beta)}{\partial x} = -\frac{36e^{-\frac{3x}{\beta}}\left(e^{x/\beta} - 1\right)}{5\beta^2}, \quad x \geq 0, \quad \beta > 0. \tag{9}$$

Since from expression (**??**), it is clear that h(x) $\propto$ j(x). Therefore, j(x) possess some attributes similar to h(x), such as $j'(0) = 0 \quad \forall \quad \beta > 0$ and the property of being monotonically decreasing. These characteristics are displayed in Figures **??** and **??**, and in turn the proposed model introduces a robust estimator for $f(0)$, named as "Shape Criterion" []Burnham et al. [**?**]]. It is also evident from the plots of pdf and detection function that as we move away from transect line (i.e., $x \to \infty$) , the probability of observing an object diminishes (i.e., all plots decays slowly to 0), that is one of the preferred character of a detection model.

Besides

$$D = \frac{E(n)j(0)}{2M},$$

substituting (**??**) in above expression, we obtain

$$\hat{D} = \frac{3\hat{n}}{5M\hat{\beta}}. \tag{10}$$

For estimating β, we will use MLE technique. Thereupon, we estimate j(0) and D which will be addressed at length in the subsequent section.

3. Statistical Properties

For SPDM model it is easy to prove the following properties:

1. The moment generating function (mgf): $M_X(t) = \frac{6(5-\beta t)}{5(\beta^2 t^2 - 5\beta t + 6)}$.
2. The rth moments: $E(X^r) = \frac{\Gamma(r+1)\left(3^{r+2} - 2^{r+2}\right)}{5(6\beta)^r}$.
3. $E(X) = \frac{19\beta}{30}$, $Var(X) = \frac{289\beta^2}{900}$ and coefficient of variation (C.V)=0.89 and skewness=1.69. The mean and variance for different choice of parameter β are exhibited in Table **??**. However, the C.V, skewness and kurtosis are independent of parameter β.

Table 1: *Mean and variance for different choices of parameter of SPDM Model*

Parameter $\beta\downarrow$	Mean $\downarrow$	Variance $\downarrow$
0.2	0.1267	0.0128
0.6	0.3800	0.1156
1.2	0.7600	0.4624
2	1.2667	1.2844
2.6	1.6467	2.1707
3.5	2.2167	3.93362

4. Assume a random sample $X_1, X_2, \ldots, X_n$ of size n drawn from SPDM pdf (**??**), then the Fisher information measure about the parameter β is given by

$$\mathbf{I}(\beta, n) = -nE\left[\frac{\partial^2 \log j}{\partial \beta^2}\right] = \frac{1.7957n}{\beta^2}.$$

- If $\hat{\beta}_{MVUE}$ is the MVUE for the parameter β, then

$$Var(\hat{\beta}_{MVUE}) = \frac{\beta^2}{1.7957n}. \tag{11}$$

Note that, this is the lower limit of Cramér?Rao inequality related to $\mathbb{NDM}(\beta)$.

4. Estimation

Here we will consider two methods of estimation: MOM and MLE for estimating the parameters of SPDM Model which are being discussed one by one in the following subsections.

4.1. Method of Moments

Suppose $x_1, x_2, \ldots, x_n$ be the observed values of a random sample (r.s) taken from model (**??**). Moment estimators consists of equating first m sample moments with corresponding m population moments, and solving the resulting system of simultaneous equations. Thus

$$m_1 = \frac{19}{25j(0)},$$

and

$$m_2 = \frac{13\beta}{15j(0)},$$

where m_1 and m_2 are first and second sample moments. Solving for $f(0)$ and β, we get

$$\hat{j}(0) = \frac{19}{25m_1}, \tag{12}$$

and

$$\hat{\beta} = \frac{57m_2}{45m_1}. \tag{13}$$

By substituting values of m_1 and m_2 from the sample, we can calculate the parameter estimates of $f(0)$ and β directly without involving any non-linear approximation. Both the estimates derived here can be taken as initial guesses for parameters to be estimated via MLE method.

4.2. Maximum Likelihood Estimation

Assume $\underline{\mathbf{x}} = \{x_1, x_2, \cdots, x_n\}$ be a r.s of size n from (**??**). The likelihood function is obtained as

$$L(j(0), \beta|\underline{\mathbf{x}}) = \prod_{i=1}^{n} j(x_i) = \prod_{i=1}^{n} j(0) \left(3 - 2e^{-\frac{x_i}{\beta}}\right) e^{-\frac{2x_i}{\beta}}. \tag{14}$$

The log-likelihood function analogous to (**??**) is obtained as

$$\log L(j(0), \beta|\underline{\mathbf{x}}) = n \log[j(0)] + \sum_{i=1}^{n} \log\left[3 - 2e^{-\frac{x_i}{\beta}}\right] - 2\sum_{i=1}^{n} \frac{x_i}{\beta}. \tag{15}$$

The ML Estimates $\hat{j}(0)$ of $j(0)$ and $\hat{\beta}$ of β, can be derived as:

$$\frac{\partial \log L}{\partial j(0)} = 0, \quad \text{and} \quad \frac{\partial \log L}{\partial \beta} = 0.$$

where

$$\frac{\partial \log L}{\partial j(0)} = \frac{n}{j(0)},$$

and

$$\frac{\partial \log L}{\partial \beta} = 2\sum_{i=1}^{n} \frac{x_i}{\beta^2} - \sum_{i=1}^{n} \frac{3x_i e^{x_i/\beta}}{\beta^2\left(3e^{x_i/\beta} - 2\right)}.$$

As the above equation is not in closed form, hence cannot be solved explicitly. Using an iterative procedure to find the estimates of β through `maxLik()` function in `R` would do the job.

The Fisher information matrix is given as

$$\mathbf{V_x} = \begin{bmatrix} -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial j(0)^2}\right) & -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial j(0)\partial \beta}\right) \\ -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial \beta \partial j(0)}\right) & -\mathbb{E}\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{bmatrix}$$

which can be approximated and written as

$$\mathbf{V_x} \approx \begin{bmatrix} V_{j(0)j(0)} & V_{j(0)\beta} \\ V_{\beta j(0)} & V_{\beta\beta} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \log L}{\partial j(0)^2}\big|_{\hat{j}(0),\hat{\beta}} & \frac{\partial^2 \log L}{\partial j(0)\partial \beta}\big|_{\hat{j}(0),\hat{\beta}} \\ \\ \frac{\partial^2 \log L}{\partial \beta \partial j(0)}\big|_{\hat{j}(0),\hat{\beta}} & \frac{\partial^2 \log L}{\partial \beta^2}\big|_{\hat{j}(0),\hat{\beta}} \end{bmatrix}$$

where $\hat{j}(0)$ and $\hat{\beta}$ are the ML estimators of $j(0)$ and β respectively. Hence, when n is large and under some mild regularity conditions, we have

$$\sqrt{n}\begin{pmatrix} j(0) - \hat{j}(0) \\ \beta - \hat{\beta} \end{pmatrix} \overset{a}{\sim} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, {\mathbf{V_x}}^{-1}\right),$$

${\mathbf{V_x}}^{-1}$ is the inverse of $\mathbf{V_x}$. The approximate confidence intervals for the parameters are; $\hat{j}(0) \pm z_{1-\alpha/2} se(\hat{j}(0))$ and $\hat{\beta} \pm z_{1-\alpha/2} se(\hat{\beta})$ for $j(0)$ and β. Here, se is the asymptotic standard error of the parameters that can be derived as a square root of the diagonal element of ${\mathbf{V_x}}^{-1}$, $z_{(1-\alpha/2)}$ indicate the $(1 - \alpha/2)$ quantile of standard normal distribution.

5. Numerical Illustration

To check the practical potentiality of the suggested model, it has been analyzed with already existing models using some goodness of fit tests. The existing models with their detection functions h(x) and pdfs j(x) over the support $0 \leq x \leq \infty$, are given as under :

1. Two parameter model (NDM) (Bakouch et al [**?**]):

$$h(x) = (1+\lambda x^{\beta})e^{-\lambda x^{\beta}}, \quad j(x) = \frac{\beta^2 \lambda^{1/\beta}}{(\beta+1)\Gamma(1/\beta)}(1+\lambda x^{\beta})e^{-\lambda x^{\beta}}; \quad \lambda, \beta \geq 0.$$

2. Negative exponential model (NEM) (Gates et al. [**?**]):

$$h(x) = e^{-\lambda x}, \quad j(x) = \lambda e^{-\lambda x}; \quad \lambda \geq 0.$$

3. Exponential power series model (EPSM) (Pollock [**?**]):

$$h(x) = e^{-(x/\lambda)^{\beta}}, \quad j(x) = \frac{e^{-(x/\lambda)^{\beta}}}{\lambda\Gamma(1+1/\beta)}; \quad \lambda, \beta \geq 0.$$

4. Reverse logistic model (RLM) (Eberhardt [**?**]):

$$h(x) = \frac{(1+\gamma)e^{-\alpha x}}{1+\gamma e^{-\alpha x}}, \quad j(x) = \frac{\alpha\gamma(1+\gamma)e^{-\alpha x}}{(1+\gamma)\log(1+\gamma)(1+\gamma e^{-\alpha x})}; \quad \alpha, \gamma \geq 0.$$

5. Weighted exponential model (WEM) [Ababneh and Eidous [**?**]]:

$$h(x) = \left(2 - e^{-\theta x}\right)e^{-\theta x}, \quad j(x) = \frac{2\theta}{3}\left(2 - e^{-\theta x}\right)e^{-\theta x}; \quad \theta \geq 0.$$

The data set here has been reported by Burnham et al. [**?**], Barabesi [**?**], Bakouch et al. [**?**] and corresponds to a number of perpendicular distances, assumed to be in meter(mtr), of wooden stakes in a sagebrush meadow east of Logan with $D = 0.00375$ stake/mtr. Walking a single path of length L=1000 meters, out of population size N=150 stakes, a number (sample) of objects n=68 stakes are detected and their corresponding perpendicular distances are recorded, constituting the data $x_1, x_2, \ldots, x_n$. The data are: 2.02, 2.90, 11.82, 4.85, 3.17, 15.24, 1.27, 9.10, 1.23, 4.97, 0.45, 8.16, 14.23, 1.47, 7.10, 3.47, 13.72, 3.25, 1.67, 3.17, 10.40, 6.47, 2.44, 18.60, 10.71, 3.05, 6.25, 8.49, 4.53, 7.67, 3.61, 5.66, 1.61, 0.41, 3.86, 7.93, 3.59, 6.08, 3.12, 18.16, 0.92, 2.95, 31.31, 0.40, 6.05, 18.15, 9.04, 0.40, 3.05, 4.08, 1.00, 3.96, 6.50, 0.20, 6.42, 10.05, 7.68, 9.33, 6.06, 3.40, 0.09, 8.27, 11.59, 3.79, 4.41,4.89, 0.53, 4.40.

The ML estimates of the data for all the given detection models have been obtained and presented in Table **??**. As displayed previously, the functioning of the detection model is directly proportional with its pdf. For checking the performance of the given models, different tests such as Akaike's Information Criterion (AIC) [**?**], Bayesian information criterion (BIC) [**?**], Kolmogorov-Smirnov statistics (K-S) and associated p-value (p-value) have been carried out and results have been shown in Table **??**.

Table 2: *ML Estimates and LL values*

Model	ML Estimates	LL
$SPDM(\beta)$	$\hat{\beta} = 9.594$	-190.009
$NDM(\beta, \lambda)$	$\hat{\beta} = 1.00941$, $\hat{\lambda} = 0.239$	-190.021
$NEM(\lambda)$	$\hat{\lambda} = 0.164$	-190.967
$EPSM(\beta, \lambda)$	$\hat{\beta} = 1.313$, $\hat{\lambda} = 8.306$	-190.22
$RLM(\alpha, \gamma)$	$\hat{\alpha} = 0.221$,$\hat{\gamma} = 2.292$	-190.048
$WEM(\theta)$	$\hat{\theta} = 0.192$	-190.044

From these Tables , it has been found that the proposed model outbeats the models in comparison in terms of Log-Likelihood (LL) values, AIC, BIC, K-S and p-values. Thus, the proposed model can be considered as a powerful competitor among other detection models.

Table 3: *Goodness of fit values*

Model	AIC	BIC	K-S Value	p-value
$SPDM(\beta)$	382.018	381.851	0.10437	0.4493
$NDM(\beta,\lambda)$	384.042	383.707	0.1115	0.4137
$NEM(\lambda)$	383.934	383.767	0.14306	0.1236
$EPSM(\beta,\lambda)$	384.44	384.105	0.1530	0.035
$RLM(\alpha,\gamma)$	384.096	383.761	0.1502	0.3703
$WEM(\theta)$	382.900	381.921	0.1917	0.0135

6. Conclusion

This manuscript focuses on the introduction of new one-parameter detection model which satisfies the shoulder conditions of the detection model and has more flexible shapes of detection model. Methods like MOM and MLE are used to estimate the parameters of model. Applicability of this model has been tested using perpendicular distances data set, therefore can be expected to appeal wide range of real life situations.

References

[1] Ababneh, F.; Eidous, O. M. (2012). A weighted exponential detection function model for line transect data. *Journal of Modern Applied Statistical Methods*, 11(1) : 475–478.

[2] Akaike, H.(1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* 19(6), 716–723.

[3] Buckland, S. T.; Anderson, D. R.; Burnham, K. P.; Laake, J. L.; Borchers, D. L.; Thomas, L.(2001).*Introduction to distance sampling: estimating abundance of biological populations* Oxford university press

[4] Buckland, S. T.; Anderson, D. R.; Burnham, K. P.; Laake, J. L.; Borchers, D. L.; Thomas, L. 2004. . *Advanced distance sampling,* Oxford University Press,

[5] Barabesi, L. Environmetrics. 2000. The official journal of the International Environmetrics Society. *Off. Jour. of Int. Evs. 11*, 413–422.

[6] Miller, D. L.; Thomas, L. 2015. Mixture models for distance sampling detection functions. *Plos One*. *10*, 413–422.

[7] Burnham, K. P.; Anderson, D. R. 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics*, 325–336.

[8] Seber, G. A. F. 1982. *Estimation of animal abundance and related parameters,* 2nd ed.; Publisher: London: Griffin.

[9] Bakouch, H. S.; Chesneau, C.; Abdullah, R. I.2022. A pliant parametric detection model for line transect data sampling. *Communications in StatisticsTheory and Methods*, 51 (21), 7340–7353.

[10] Buckland, S. T.; Rexstad, E. A,; Marques, T. A.; Oedekoven, C. S. 2015. *Distance sampling: methods and applications*, **431**, Publisher: Springer.

[11] Buckland, S. 1985. Perpendicular distance models for line transect sampling. *Biometrics*, 41 (1), 177–195.

[12] Burnham, K. P.; Anderson, D. R.; Laake, J. L.. (1980). Estimation of density from line transect sampling of biological populations. *Wildlife Monographs,* (72):3–202.

[13] Eberhardt, L. Transect methods for population studies. 1978. *The Journal of Wildlife Management*, 42(1): 1–31.

[14] Crain, B.; Burnham, K.; Anderson, D.; Lake, J. (1979). Nonparametric estimation of population density for line transect sampling using fourier series. *Biometrical Journal* 21 (8): 731–48.

[15] Gates, C. E.; Marshall, W. H.; Olson, D. P.1968. Line transect method of estimating grouse population densities. *Biometrics*, 135–145.

[16] Pollock, K. H. 1978. A family of density estimators for line-transect sampling. *Biometrics*, 475–478.

[17] Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2): 461–464.

A NEW ATTRIBUTE CONTROL CHART BASED ON EXPONENTIATED EXPONENTIAL DISTRIBUTION UNDER ACCELERATED LIFE TEST WITH HYBRID CENSORING

Gunasekaran Munian
•
Department of Applied Research, The Gandhigram Rural Institute (Deemed to be University)
Gandhigram, Dindigul – 624 302, Tamil Nadu, India
gunasekaranm1988@gmail.com

Abstract

In this article, we propose a new attribute np control chart for monitoring the median lifetime of the products under accelerated life test with hybrid censoring scheme assuming that the lifetime of the products follows an exponentiated exponential distribution. The optimal parameters for constructing the proposed control chart are determined so that the average run length for the in-control process is as closest to the prescribed average run length as possible. The control chart parameters are estimated for various set of values, and the developed control chart's performance is analysed using the average run length. The proposed control chart is illustrated with numerical examples, and its applicability is demonstrated with simulated data.

Keywords: exponentiated exponential distribution, accelerated life test, hybrid censoring scheme, control chart, average run length

I. Introduction

A control chart is a graphical description of data collected from manufacturing industries. The information may apply to the measurement of quality characteristics or the evaluation of quality characteristics of a sample. It monitors the consistency of the process and raises an alarm if it finds any deviations from the specified tolerance limit. Control charts are effective tool for monitoring the behaviour of a production process in order to determine whether it is stable or not. Control charts visualize the data throughout the order in which it occurred.

In order to monitor the production process in a various circumstances, many new control chart techniques have been developed. A control chart is a type of graph that can be used to analyse how a process varies over time. Every control chart has a minimum of two control limits, known as the lower control limit (LCL) and upper control limit (UCL). It is claimed that the process is in-control if the control statistic falls within the control limits. There are two different kinds of control charts, which are referred to as attribute control charts and variable control charts. The attribute control charts are used to distinguish between conforming and non-conforming items. When the industrial data collected from the measurement process, the variable control charts are implemented.

Lifetime is regarded as a quality characteristic for some products. Life tests are used to monitor the manufacturing process for these products. The product may be classed as conforming or non-conforming based on the results of the life test. This form of product testing takes a considerable amount of time because the duration of the test is lengthy. In this circumstance, the censoring technique is essential and cannot be ignored. Some of the censoring techniques used in life testing include Type-I censoring, Type-II censoring, and hybrid censoring. According to the hybrid censoring, the life test terminates at the earliest of the specified test time t or the time at which the $(UCL + 1)^{th}$ failure is discovered. When the number of failures observed during the life test is between the UCL and LCL at time t, the production process is in control; otherwise, the production process is out of control. Accelerated life testing (ALT) is a test and analytical technique for figuring out how failures will probably happen in the future. As a result of its capacity to "speed up time," ALT is a widely used testing technique. When we cannot wait for failures to occur at their regular rate but need to know how failures are expected to occur in the future, we frequently employ ALT. In an ALT, items are subjected to accelerated conditions including stress, strain, temperature, pressure, etc.; as a result, the products fail faster compared to their performance in a conventional life test. Therefore, ALT under hybrid censoring can minimise total time to failure and inspection cost.

The fraction non-conforming is based on the attribute control chart, such as the *np* control chart, which is constructed by assuming that the quality characteristic follows the normal distribution. The distribution of the quality characteristics may not be normal in reality. Industrial engineers may be misled and the number of non-conforming products may increase if the current control chart is used in this scenario. Numerous authors, including [1-9] and [14-24] have contributed to the literature and worked on attribute control charts for various lifetime distributions. According to a review of the available literature, no research on control charts with a life test for a non-normal distribution, such as an exponentiated exponential distribution with hybrid censoring under accelerated life test, have been attempted.

In reliability and life testing in recent years, exponentiated models have drawn more attention. By exponentiating the corresponding cumulative distribution function with a parameter, the probability distributions are constructed. The exponentiated exponential distribution was first introduced by [13]. It has been revealed that this distribution is more suitable for particular lifetime data than other commonly used lifetime distributions. It has been noted that lifetime data analysis has been quite effective when using the exponentiated exponential distribution. The exponentiated exponential distribution is found to fit data in many cases better than the Weibull, gamma, log-normal, and generalized Rayleigh distributions (see more information on this distribution from [10-12].

This work presents an attribute control chart based on an exponentiated exponential distribution under an accelerated life test with a hybrid censoring scheme. The control chart coefficient was determined, and the performance of the proposed control chart was discussed in relation to the average run length (ARL). When the process scale parameter (median) is changed, the proposed control chart is designed. With the use of simulated data, the proposed control chart's applicability is described. Simulated data is used to illustrate the proposed chart. Exponentiated exponential distribution is introduced in Section 2 along with an explanation of the control chart's design. A simulation study is described in Section 3. The last Section described the overall summary.

II. Design of proposed control chart based on Exponentiated Exponential Distribution under Accelerated Life Test

Let "T" be the product's lifetime, which is distributed to the Exponentiated Exponential Distribution (EED (λ, θ)) with scale parameter "λ" and shape parameter "θ". The probability density function for the variable "T" is given by,

$$f(t|\lambda;\theta) = \left(\frac{\theta}{\lambda}\right) e^{-\left(\frac{t}{\lambda}\right)} \left[1 - e^{-\left(\frac{t}{\lambda}\right)}\right]^{\theta-1} ; t, \lambda, \theta > 0 \quad (1)$$

The underlying distribution's cumulative distribution function is defined by,

$$F(t|\lambda;\theta) = \left[1 - e^{-\left(\frac{t}{\lambda}\right)}\right]^{\theta} ; t, \lambda, \theta > 0 \quad (2)$$

With this distribution, the product's mean as well as median lifetime are given as,

$$\mu = Mean(T) = \lambda[\Psi(\theta + 1) - \Psi(1)]$$

Here, $\Psi(x) = \frac{d}{dx}\Gamma(x)$ and $\Gamma(x) = \int_0^\infty x^{u-1}e^{-x}dx$

$$M = Median(T) = -\lambda\, ln\left[1 - \left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right]$$

The product's lifetime under ordinary conditions is represented by the symbol t_U, and it follows an exponentiated exponential distribution with specified shape (θ) and scale (λ_U) parameters. As a result, the following equation may be used to determine the product's lifetime under ordinary conditions.

$$F_U(t_U|\lambda_U;\theta) = \left[1 - e^{-\left(\frac{t_U}{\lambda_U}\right)}\right]^{\theta} \quad (3)$$

Similar to this, the lifetime of the product under accelerated conditions is denoted by t_A, and it is assumed that t_A follows an exponentiated exponential distribution with specified shape (θ) and scale (λ_A) parameters. As a result, the product's lifetime under accelerated conditions is stated as follows.

$$F_A(t_A|\lambda_A;\theta) = \left[1 - e^{-\left(\frac{t_A}{\lambda_A}\right)}\right]^{\theta} \quad (4)$$

The product's mean and median lifetime under accelerated conditions as

$$\mu_A = \lambda_A[\Psi(\theta + 1) - \Psi(1)]\ and\ M_A = -\lambda_A\, ln\left[1 - \left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right] \quad (5)$$

By using the Acceleration Factor (AF) definition, we can write

$$\lambda_A = \frac{\lambda_U}{AF}$$

Therefore, Equation (4) becomes

$$F_A(t_A) = \left[1 - e^{-\left(\frac{t_A \times AF}{\lambda_U}\right)}\right]^{\theta}$$

The failure probability of the product, denoted by p_0 while its process is under control, is characterized by the probability that the product will fail before the censoring time t_A.

$$p_0 = \left[1 - e^{-\left(\frac{t_A \times AF}{\lambda_U}\right)}\right]^{\theta} \quad (6)$$

We express the censoring time as a constant multiple of the product's median lifetime, such as $t_A = a*M_A$, where 'a' is referred to as a test termination ratio. Equation (6) may be rewritten by substituting values for t_A and M_A.

$$p_0 = \left[1 - e^{\left(\frac{\lambda_A}{\lambda_U} \times a \times AF \times ln\left[1-\left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right]\right)}\right]^{\theta} \quad (7)$$

It is assumed that there is no change in scale parameter when the process is in-control. i.e., $\lambda_A = \lambda_U$. Hence, Equation (7) becomes

$$p_0 = \left[1 - e^{\left(a \times AF \times ln\left[1-\left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right]\right)}\right]^{\theta} \quad (8)$$

We can determine the failure probability of the product under accelerated life test using the above equation. The process is considered as out-of-control if there is a change in scale parameter, under this situation the scale parameter shifts from λ_U to λ_A. The ratio between the actual scale parameter and the shifted scale parameter as accelerated life test is denoted by 'f' and is called as shift constant. That is, $\frac{\lambda_U}{\lambda_A} = f$. Therefore, the failure probability of the product under accelerated life test when the process is out-of-control is denoted by p_1 and determined by

$$p_1 = \left[1 - e^{\left(\frac{a \times AF}{f} \times ln\left[1-\left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right]\right)}\right]^{\theta} \quad (9)$$

Based on the number of products in each subgroup, we propose the following *np* control chart for EED (λ, θ) under accelerated life test with hybrid censoring:

Step 1: Select a set of n products randomly from the production process.

Step 2: Conduct the life test on the selected products considering t_A as the test termination time under the fixed acceleration factor. Observe the number of failed items (D, say)

Step 3: Terminate the life test either after reached at time t_A or D > UCL before reaching time t_A, whichever is earlier.

Step 4: Declare the process as out of control if $D>UCL$ or $D<LCL$. Declare the process as in control if $LCL \le D \le UCL$.

The above is considered to as a *np* control chart since it shows the number of failures (D) instead of the proportion nonconforming (p). While the process is in control, the random variable D follows a binomial distribution with parameters n and p_0, where p_0 is the probability that an item fails before time t_A. As a result, the control limits for the control process are as follows:

$$UCL = np_0 + k\sqrt{np_0(1-p_0)} \quad (10a)$$

$$LCL = max\left[0, np_0 - k\sqrt{np_0(1-p_0)}\right] \quad (10b)$$

Where, k is the control limit coefficient. The fraction nonconforming in the control process (p_0) can be determined using equation (8).

$$p_0 = \left[1 - e^{\left(a \times AF \times ln\left[1-\left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right]\right)}\right]^{\theta}$$

The control limits for the practical application are provided as, because in practise, probability p_0 is generally unknown.

$$UCL = \bar{D} + k\sqrt{\bar{D}\left(1 - \frac{\bar{D}}{n}\right)}$$
$$LCL = max\left[0, \bar{D} - k\sqrt{\bar{D}\left(1 - \frac{\bar{D}}{n}\right)}\right]$$

Here, $\bar{D}$is the average number of failures over the subgroups.

The following is the probability of declaring as in control for the proposed control chart:

$$p_{in}^{0} = P(LCL \le D \le UCL|p_0)$$
$$p_{in}^{0} = \sum_{d=[LCL]+1}^{[UCL]} \binom{n}{d} p_0^d (1-p_0)^{n-d} \quad (11)$$

The control chart's performance is measured by the ARL. The ARL for the process of in control as follows:

$$ARL_0 = \frac{1}{1-\left[\sum_{d=[LCL]+1}^{[UCL]} \binom{n}{d} p_0^d (1-p_0)^{n-d}\right]} \quad (12)$$

I. ARL when the scale parameter is shifted:

Assume the process median has been changed from M_0 to M_1. The probability in (9) now becomes

$$p_1 = \left[1 - e^{\left(\frac{a \times AF}{f} \times ln\left[1-\left(\frac{1}{2}\right)^{\left(\frac{1}{\theta}\right)}\right]\right)}\right]^{\theta}$$

The probability that the process is declared to be in control after shifting to M_1 is now determined by

$$p_{in}^{1} = P(LCL \le D \le UCL|p_1)$$
$$p_{in}^{1} = \sum_{d=[LCL]+1}^{[UCL]} \binom{n}{d} p_1^d (1-p_1)^{n-d} \quad (13)$$

The ARL for the shifted process is given as follows:

$$ARL_1 = \frac{1}{1-\left[\sum_{d=[LCL]+1}^{[UCL]} \binom{n}{d} p_1^d (1-p_1)^{n-d}\right]} \quad (14)$$

To construct the tables for the proposed control chart, we applied the following algorithm.

Step 1: Specify the values of ARL, say $r_0$and sample size n.

Step 2: Determine the control chart parameters and truncated time constant a values for which the ARL from equation (8) approach r_0.

Step 3: By using the values of the control chart parameters obtained in step 2, determine the ARL_1 in accordance with shift constant f by using equation (14).

For various values of r_0 and n, we determine the control chart parameters and ARL_1, which are shown in Tables 1 & 2. Table 1 & Table 2 shows that the ARLs tend to get smaller when the shift constant f gets less.

Table 1: *ARLs when the process average (median) shifted when the sample size fixed*

	n = 25								
	AF = 1			AF = 1.5			AF = 2		
	r_0 = 300	r_0 = 350	r_0 = 450	r_0 = 300	r_0 = 350	r_0 = 450	r_0 = 300	r_0 = 350	r_0 = 450
UCL	14	16	18	14	16	18	14	16	18
LCL	1	2	3	1	2	3	1	2	3
a	0.6444	0.7689	0.8298	0.4296	0.5126	0.5532	0.3222	0.3844	0.4149
k	2.6331	2.8254	2.9472	2.6331	2.8254	2.9668	2.6331	2.8249	2.9668
Shift (f)	ARL								
0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.40	1.04	1.03	1.08	1.04	1.03	1.08	1.04	1.03	1.08
0.50	1.42	1.33	1.69	1.42	1.33	1.69	1.42	1.33	1.69
0.60	2.99	2.64	4.46	2.99	2.64	4.46	2.99	2.64	4.46
0.70	8.97	7.68	17.28	8.97	7.68	17.21	8.97	7.68	17.21
0.80	33.01	28.40	82.95	33.01	28.40	82.95	33.01	28.39	82.95
0.90	128.99	117.81	391.28	128.99	117.81	391.28	128.99	117.81	391.28
1.00	300.01	350.01	450.01	300.01	350.01	450.01	300.01	350.01	450.01

Table 2: *ARLs when the process average (median) shifted when r_0 fixed*

	n = 500								
	AF = 1			AF = 1.5			AF = 2		
	n = 20	n = 25	n = 30	n = 20	n = 25	n = 30	n = 20	n = 25	n = 30
UCL	12	15	19	12	15	19	12	15	19
LCL	0	1	3	0	1	3	0	1	3
a	0.6161	0.6926	0.7439	0.4258	0.46176	0.4959	0.3081	0.3463	0.3890
k	2.9166	2.8550	2.9467	2.7402	2.8547	2.9471	2.9158	2.8550	2.9581
Shift (f)	ARL								
0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.30	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.40	1.20	1.04	1.02	1.14	1.04	1.02	1.20	1.04	1.01
0.50	2.25	1.43	1.34	1.94	1.43	1.34	2.25	1.43	1.21
0.60	6.28	3.08	2.97	4.92	3.08	2.98	6.27	3.08	2.29
0.70	22.20	9.56	10.51	16.00	9.57	10.51	22.17	9.56	6.86
0.80	87.50	36.94	49.86	59.10	36.97	49.90	87.35	36.93	28.19
0.90	317.92	157.16	261.95	218.69	157.30	262.19	317.42	157.16	137.85
1.00	500.05	500.13	500.34	500.06	500.35	500.16	500.24	500.13	500.08

II. Illustration 1:

Suppose that the lifetimes of the products follow an exponentiated exponential distribution with parameters λ, θ = 2 and AF = 1. Consider the following values for the products: M_0 = 1000 hours, r_0 = 300 and n = 25. Table 1 provided the following control chart parameters: k = 2.6331, a = 0.6444, LCL = 1 and UCL = 14. As a consequence, the control chart was established in the following manner:

Step 1: Select a sample of 25 products from each subgroup and submit them to the life test for 644 hours. During the testing, count the number of failed items (D).

Step 2: Declare that the process is in control if the value of D is between 1 and 14; otherwise, declare that it is out of control.

III. Illustration 2:

Assume that the product lifetimes follow an exponentiated exponential distribution with parameters λ, $\theta = 2$ and AF = 2. Consider the following product values: M_0 is 1000 hours, r_0 is 500, and n is 25. The following control chart parameters were presented in Table 2: k = 2.8550, a = 0.3463, LCL = 1 and UCL = 15. As a result, the control chart was established in the following manner:

From each subgroup, choose a sample of 25 products, and put them through a life test that lasted at least 346 hours while being accelerated. Count the number of failed items during testing (D). If the value of D is between 1 and 15, the process is in control; otherwise, it is out of control.

III. Simulation Study

This section discusses the application of the produced control chart with simulated data. The data was generated and shown in Table 3, using an exponentiated exponential distribution with an average (median) lifetime (M_0) of 1000 hours, the value of acceleration factor (AF) is 1, and the value of the shape parameter (θ) is 2. Assuming that the sample size (n) is 20 and the specified ARL (r_0) is 500. At M_0 = 1000 hours and f = 1, the process is considered to be in control. In-control parameters are employed to construct the first 15 observations of subgroup size 20. Let's now assume that since the median of the exponentiated exponential distribution has shifted, the process has shifted as well. The value of the shift constant f is 0.5. When f is set to 0.5, the shifted median is used to produce the next 15 observations (shown in Table 4).

Take into account that the experiment was conducted at t = 616 hours. Table 3 and Table 4 shows the number of failures, indicated by the letter D, for each subgroup. Figure 1 shows the calculated LCL = 0 and UCL = 12 for simulated data.

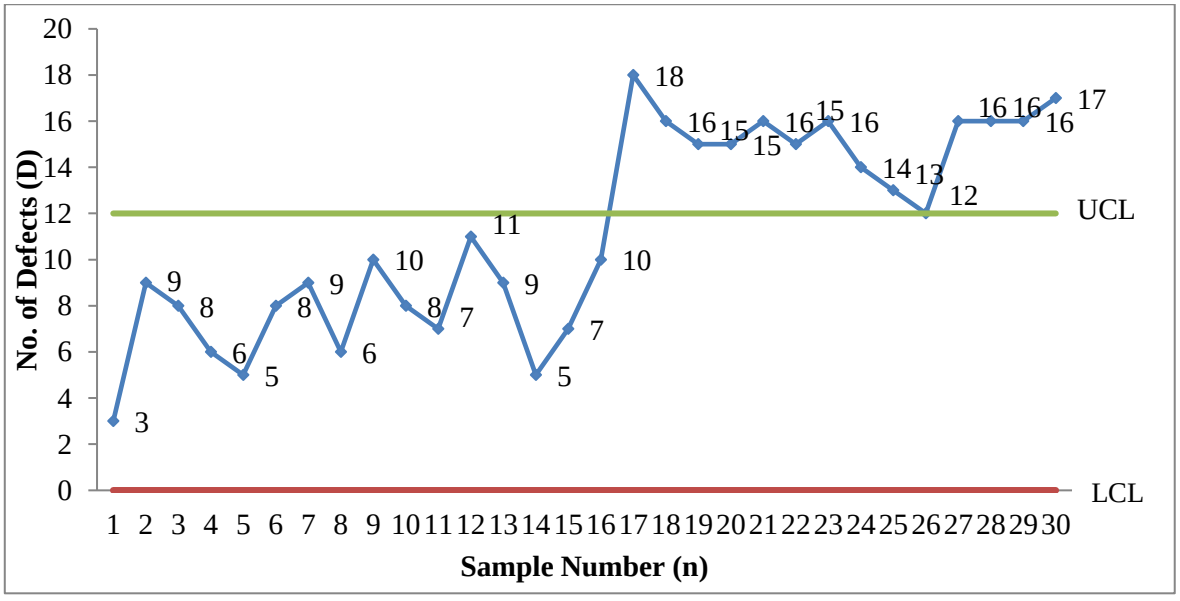


Figure 1: *Control Chart for Simulated Data*

Table 3: *Simulated Data when M_0 = 1000 hours, AF = 1, θ = 2, ARL_0 = 500, n = 20 and f = 1*

S. No	Sample															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	D
1	1075	558	1334	520	832	479	545	1683	361	1862	639	756	1169	814	365	7
2	1379	881	976	335	324	523	1392	901	918	2901	223	1208	894	751	681	11
3	1253	763	1096	487	2703	386	918	824	560	860	370	203	450	685	769	9
4	673	273	486	356	3318	1049	499	1264	1242	1090	888	113	2837	822	1560	5
5	913	920	377	510	984	1963	2969	990	424	543	1393	4303	2023	2015	1517	10
6	1465	2041	1026	704	2708	1940	707	3080	390	1055	1259	107	394	356	873	8
7	492	351	264	347	451	220	805	1409	1396	1010	961	1878	426	1017	494	6
8	693	389	796	1169	1397	697	665	2177	678	144	614	1023	485	807	2407	12
9	1304	913	1562	1483	371	1681	401	225	188	594	501	547	394	132	1439	4
10	887	585	255	1079	963	300	2463	257	203	996	2092	2414	1720	446	1265	13
11	750	851	359	1852	1296	621	430	901	1458	780	1279	2128	2741	235	2100	3
12	681	898	571	1736	1333	796	317	1031	253	1016	1420	354	302	1093	2160	14
13	2263	1382	733	1879	394	483	2328	345	223	504	819	413	126	750	1618	2
14	420	599	162	955	1614	712	1346	513	404	2911	1403	1682	1689	539	2929	15
15	3150	582	756	1415	882	1231	577	3107	2142	228	440	124	777	1828	478	1
16	971	675	1097	980	364	2052	667	1841	1615	1094	234	2067	1822	1869	2107	16
17	931	545	2520	983	1169	491	548	584	977	204	1917	324	335	691	521	17
18	1902	535	665	1069	719	453	676	1232	1004	477	724	187	1049	829	1469	18
19	272	990	2264	1353	820	914	117	223	702	463	2186	198	1838	779	471	19
20	3368	1858	454	744	1285	897	462	1030	382	2883	3209	470	521	1019	449	20

Table 4: *Simulated Data when M_0 = 1000 hours, AF = 1, θ = 2, ARL_0 = 500, n = 20 and f = 0.5*

S. No	Sample																				D
1	159	303	792	933	1168	436	298	191	364	320	622	173	361	536	462						10
2	709	209	480	447	959	87	3035	244	926	887	130	555	174	511	333						18
3	1661	361	1165	183	177	590	141	1460	433	315	291	447	688	633	553						15
4	183	310	213	545	106	867	496	285	539	177	904	511	1064	691	893						16
5	927	130	426	558	345	144	102	1030	595	1030	292	252	325	1029	478						15
6	948	233	223	245	192	202	399	247	300	1718	303	193	163	435	530						16
7	170	112	421	132	231	923	740	134	480	338	1811	211	865	94	460						12
8	935	205	534	1121	327	1070	550	94	1545	186	185	216	279	343	84						13
9	1308	278	423	325	778	126	714	267	575	664	591	237	204	360	747						14
10	498	481	510	662	628	397	1214	108	490	942	137	382	416	613	282						16
11	216	571	66	568	499	506	509	329	54	1065	503	699	25	437	466						15
12	1119	42	255	519	602	423	538	606	783	447	367	1063	263	317	330						16
13	1544	861	108	238	674	323	441	621	42	531	1951	442	147	330	685						16
14	373	322	125	948	189	750	391	107	952	252	315	366	27	162	611						15
15	1727	131	135	193	324	139	315	162	537	897	872	121	1424	352	135						16
16	1336	307	368	160	122	573	135	192	893	611	125	78	560	730	292						16
17	537	313	784	207	131	344	1086	457	856	171	725	217	251	459	288						16
18	76	73	398	162	192	312	332	150	109	274	626	712	219	233	158						16
19	306	116	438	827	258	199	477	105	60	453	280	1128	180	178	58						16
20	367	620	1294	215	463	226	447	637	202	208	639	443	149	302	272						16

The constructed control chart, as shown in Figure 1, reveals a shift at the 17th (2nd observation following the shift) observation, with a calculated ARL of 2.25. As a consequence, the developed control chart efficiently identifies the shift in the production process.

IV. Conclusion

In this article, a new attribute control chart based on exponentiated exponential distribution under accelerated life test with hybrid censoring scheme is proposed to ensure the median lifetime of the product as the quality criterion. The newly established control chart is highly adaptable and may be used to monitor the lifetime of quality products. The tables are provided for industrial usage and are explicated with the use of simulated data. It is generated using *R* software from an exponentiated exponential distribution. The performance of the proposed control chart is expressed in terms of ARLs for various shift constants (*f*). It should be noted that if the hybrid censoring scheme is used to carry out the life test, executing the sampling inspection will reduce the time and cost of conducting the life test. The developed attribute control chart has the potential to be extended for use with various other statistical distributions as part of a continuing research study.

References

[1] Adeoti, O. A. and Ogundipe, P. (2021). A Control Chart for the generalized exponential distribution under time truncated life test. Life Cycle Reliability and Safety Engineering, 10:53-59.

[2] Al-Marshadi, A. H., Shafqat, A., Aslam, M. and Alharbey, A. (2021). Performance of a New Time-Truncated Control Chart for Weibull Distribution under Uncertainty. *International Journal of Computational Intelligence Systems,* 14(1):1256-1262.

[3] Arif, O. H., Aslam, M. and Butt, K. A. (2018). Attribute Control Chart for a lognormal distribution under Accelerated Time-Censoring. *Journal of Computational and Theoretical Nanoscience,* 15(3):919 – 923.

[4] Aslam, M. and Jun, C. (2015). Attribute Control Charts for the Weibull Distribution under Truncated Life Tests. *Quality Engineering,* 27(3):283 – 288.

[5] Aslam, M., Arif, O. H. and Jun, C. (2016b). An Attribute Control Chart based on the Birnbaum-Saunders Distribution using Repetitive Sampling. *IEEE Access,* 5:9350 – 9360.

[6] Aslam, M., Nasrullah Khan and Chi – Hyuck (2016a). A control chart for time truncated life tests using Pareto distribution of second kind. *Journal of Statistical Computation and Simulation,* 86(11):2113 – 2122.

[7] Baklizi, A. and Ghannam, S. A. (2022). An attribute control chart for the inverse Weibull distribution under truncated life tests. *Heliyan,* 8(12).

[8] Balamurali, S. and Jeyadurga, P. (2019). An attribute np control chart for monitoring mean life using multiple deferred state sampling based on truncated life tests. *International journal of Reliability, Quality and Safety Engineering,* 26(1):1-18.

[9] Barbosa, E. P. and Joekes, S. (2013). An improved attribute control chart for monitoring non-conforming proportion in high quality process. *Control Engineering Practice,* 21:407 – 412.

[10] Gupta, R. D. and Kundu, D. (2001a). Exponentiated exponential distribution: An alternative to gamma and Weibull distributions. *Biometrical Journal,* 43(1):117 – 130.

[11] Gupta, R. D. and Kundu, D. (2001b). Generalized exponential distributions: different methods of estimations. *Journal of Statistical Computation and Simulation,* 69(4):315 – 338.

[12] Gupta, R. D. and Kundu, D. (2002). Generalized exponential distribution: Statistical Inferences. *Journal of Statistical Theory and Applications,* 1(2):101 – 118.

[13] Gupta, R. D. and Kundu, D. (1999). Generalized Exponential Distributions. *Australian and New Zealand Journal of Statistics,* 41(2):173 – 188.

[14] Hsu, L. F. (2004). Note on Design of Double and Triple Sampling Control Charts Using Genetic Algorithms. *International Journal Production Research,* 42(5):1043 – 1047.

[15] Hsu, L. F. (2007). Note on Construction of Double Sampling s – Control Charts for Agile Manufacturing. *Quality and Reliability Engineering International,* 23:269-272.

[16] Jayadurga, P., Balamurali, S. and Aslam, M. (2018). Design of an attribute np control chart for process monitoring based on repetitive group sampling under truncated life tests. *Communication in Statistics – Theory and Methods,* 47(24):5934 – 5955.

[17] Kavitha, T. and Gunasekaran, M. (2020). Construction of Control Chart based on Exponentiated Exponential Distribution. *Journal of Engineering Sciences,* 11(4):979 – 985.

[18] Nanthakumar, C. and Kavitha, T. (2017). Design of Attribute Control Chart based on Inverse Rayleigh Distribution under Type – I Censoring. *International Journal of Statistical Sciences,* 4(6):17 – 22.

[19] Shafqat A., Hussain J., Al-Nasser, A. D. and Aslam, M. (2018). Attribute Control Chart for some popular distributions. *Communications in Statistics – Theory and Methods,* 47(8):1978 – 1988.

[20] Srinivasa Rao, G. and Edwin Paul (2020). Time Truncated Control Chart using Log-Logistic Distribution. *Biometrics & Biostatistics International Journal,* 9:76- 82.

[21] Srinivasa Rao, G., Fulment, A. K. and Josephat, P. K. (2019). Attribute control chart for the Dagum distribution under truncated life tests. *Life Cycle Reliability and Safety Engineering,* 8:329 – 335.

[22] Srinivasa Rao, G. (2018). A Control Chart for Time Truncated Life tests Using Exponentiated Half Logistic Distribution. *Applied Mathematics & Information Sciences: An International Journal,* 12(1):125 – 131.

[23] Wu, Z. and Wang, Q. (2007). An np control chart using double inspections. *Journal of Applied Statistics,* 34(7):843 – 855.

[24] Wu, Z., Luo, H. and Zhange, X. (2006). Optimal np control chart with curtailment. *European Journal of Operational Research,* 174:1723 – 1741.

CHARACTERIZATION OF GENERALIZED DISTRIBUTIONS BASED ON CONDITIONAL EXPECTATION OF ORDER STATISTICS

Abu Bakar[1], Haseeb Athar[2] and Mohd Azam Khan[3]

•

Department of Statistics and Operations Research, Aligarh Muslim University
Aligarh – 202002, India
[2]haseebathar.st@amu.ac.in, [3]khanazam2808@gmail.com
Corresponding author: [1]gh2022@myamu.ac.in

Abstract

Characterization of probability distributions plays a significant role in the field of probability and statistics and attracted many researchers these days. Characterization refers to the process of identifying distributions uniquely based on certain statistical properties or functions. The various characterization results have been established by using different methods. The paper aims to characterize two general forms of continuous distributions using the conditional expectation of order statistics. Further, the results obtained are applied to some well-known continuous distributions. Finally, some numerical calculations are performed.

Keywords: Order statistics, conditional expectation; truncated moments, characterization, continuous distributions.

I. Introduction

The characterization of probability distributions is indeed crucial in statistical studies as it allows us to understand and utilize various distributions effectively. The distributions can be characterized using different statistical properties like moments, truncated moments, order statistics, record values, reliability functions, characteristic function etc. Each of these approaches leverages different statistical properties to uniquely identify or characterize specific probability distributions. This allows statisticians and researchers to model data effectively, understand underlying processes, and make informed decisions based on statistical analyses.

The various characterization results using truncated moment and conditional expectation of order statistics are available in the literature. Grudzien and Szynal [15] characterized the uniform distribution in terms of moments of order statistics when the sample size is random whereas Balasubramanian and Beg [11] focused on distribution characterizations using moments and order statistics. Khan and Abu-Salih [18] characterized a general class of distribution through conditional expectations of order statistics. Further, Khan and Abouammoh [17] extended the results of Khan and Abu-Salih [18] by characterized the general form of distribution for higher order gap. Khan and Athar [23] characterized some continuous distributions by examining the linearity of regression when conditioned on a pair of order statistics. Nassar [27] characterized a mixture of two generalized power function distributions based on the conditional expectation of order statistics. Similarly, Lee

et al. [25] provided a characterization of mixtures of Weibull and Pareto distributions through the conditional expectation of order statistics and upper record values. The recent interest on characterizing probability distributions via truncated moments has led to significant contributions from various authors. For example, Ahsanullah *et al*. [3] explore the characterization of Lindley distribution based on a relation between truncated moments and failure rate function. Kilany [24] established the characterization of the Lindley distribution using the truncated moments of order statistics. Athar and Abdel-Aty [6] characterized a class of continuous distributions based on left and right truncated moments whereas Athar *et al.* [9] studied the characterization of some generalized continuous distribution by doubly truncated moments. Bashir and Khan [12] employed a range of techniques to characterize the weighted power function distribution, using mean inactivity times, mean residual function, conditional moments, conditional variance, doubly truncated mean, incomplete moments, and the reverse hazard function. For more studies on characterization, one can refer to Khan and Khan [21], Khan and Masoom Ali [22], Franco and Ruiz [14], Ali and Khan [4], Khan and Athar [20], Khan and Alzaid [19], Athar *et al*. [10], Ahsanullah and Hamdani [1], Huang and Su [16], Ahsanullah and Shakil [2], Athar and Akhtar [7], Athar *et al.* [8], Ansari *et al.* [5] and references given therein.

Let $X_1, X_2, ..., X_n$ be a random sample of size $n(>2)$ from a continuous population having probability density function (*pdf*) $f(x)$ and cumulative distribution function (*cdf*) $F(x)$ and its corresponding order statistics be the $X_{1:n} \le X_{2:n} \le ... \le X_{n:n}$ respectively. It is well Known that the conditional *pdf* of r^{th} order statistic $(X_{r:n} = x)$ given that s^{th} order statistic $(X_{s:n} = y)$ for $s > r$, is the same as the distribution of the r^{th} order statistics obtained from a sample of size $(s-1)$ from a population whose distribution is truncated on the right side at y while the conditional *pdf* of s^{th} order statistic $(X_{s:n} = y)$ given that r^{th} order statistic $(X_{r:n} = x)$ for $r < s$, is the same as the distribution of the $(s-r)^{th}$ order statistics from a sample of size $(n-r)$ from a population whose distribution is simply truncated on the left side at x (David and Nagaraja [13]).

Therefore, the conditional *pdf* of r^{th} order statistic $(X_{r:n} = x)$ given that s^{th} order statistic $(X_{s:n} = y), 1 \le r < s \le n$ is,

$$f(x \mid y) = \frac{(s-1)!}{(r-1)!(s-r-1)!} \frac{\left(F(x)\right)^{r-1}\left(F(y)-F(x)\right)^{s-r-1}}{\left(F(y)\right)^{s-1}} f(x), x \le y \tag{1.1}$$

and the conditional *pdf* of s^{th} order statistic given r^{th} order statistic is

$$f(y \mid x) = \frac{(n-r)!}{(s-r-1)!(n-s)!} \frac{\left(F(y)-F(x)\right)^{s-r-1}\left(1-F(y)\right)^{n-s}}{\left(1-F(x)\right)^{n-r}} f(y), x \le y. \tag{1.2}$$

Also note that for any monotonic and differentiable function $\xi(x)$ of X over the support (α, β). Here α may be $-\infty$ and β may be $+\infty$.

$$E[\xi(X_{n:n}) \mid X_{n-1:n} = x] = E[\xi(X) \mid X \ge x] \tag{1.3}$$

and

$$E[\xi(X_{1:n}) \mid X_{2:n} = x] = E[\xi(X) \mid X \le x]. \tag{1.4}$$

In this paper, first we have characterized two general form of distributions $F(x)=e^{-a\xi(x)}$, $a\neq 0$ and $F(x)=1-e^{-a\xi(x)}$, $a\neq 0$ through the conditional expectation of order statistics $E\left[\xi(X)\mid X_{r:n}=x\right]$, where $X_{r:n}$ is r^{th} order statistics. Further, these results are applied to some well-known continuous distributions.

II. Characterization Theorems

Before presenting the main result, we will discuss the following propositions that were established by Athar *et al*. [10]:

Proposition 2.1: *Let* X *be a random variable and* $E\left[\xi(X)\mid X_{r:n}=x\right]=\mu_r(x)$, *then for* $1\le r<s\le n$,

$$\frac{f(x)}{1-F(x)}=\frac{n(s-1)\mu_r'(x)-n(r-1)\mu_s'(x)-(s-r)\xi'(x)}{n(s-1)\mu_r(x)-n(r-1)\mu_s(x)-n(s-r)\xi(x)}, \tag{2.1}$$

where $\xi(x)$ is a monotonic and differentiable of $x\in(\alpha,\beta)$.

Proposition 2.2: *Under the conditions as stated in Proposition 2.1,*

$$\frac{f(x)}{F(x)}=-\frac{n(n-r)\mu_s'(x)-n(n-s)\mu_r'(x)-(s-r)\xi'(x)}{n(n-r)\mu_s(x)-n(n-s)\mu_r(x)-n(s-r)\xi(x)}. \tag{2.2}$$

Theorem 2.1: *Let* X *be a absolutely continuous* (*w.r.t Lebesgue measure*) *random variable with* CDF $F(x)$ *and* PDF $f(x)$ *on the support* (α,β), *where* α *and* β *may be finite or infinite. Then for* $1\le r<s\le n$

$$E\left[\xi(X)\mid X_{r:n}=x\right]=\mu_r\left(x\right)=T_0(x)+\frac{r-1}{n}T_1(x)+\frac{n-r}{n}T_2(x), \tag{2.3}$$

if and only if

$$\bar{F}(x)=e^{-a\xi(x)},\ x\in(\alpha,\beta),\ a\neq 0, \tag{2.4}$$

where,

$$\bar{F}(x)=1-F(x)$$

$$T_0(x)=\frac{\xi\left(x\right)}{n},$$

$$T_1(x)=E\left[\xi(X)\mid X\le x\right]=T_2(x)-\frac{\xi(x)}{1-e^{-a\xi(x)}},$$

$$T_2(x)=E\left[\xi(X)\mid X\ge x\right]=\xi(x)+\frac{1}{a},$$

and $\xi(x)$ *is a monotonic and differentiable function of x, such that* $\xi(x)\to 0$ *as* $x\to\alpha$ *and* $\xi(x)\bar{F}(x)\to 0$ *as* $x\to\beta$.

Proof. To prove necessary part, the value of $T_1(x)$ and $T_2(x)$ can be obtained easily by integrating by parts and noting that

$$1-F(x)=\frac{1}{a\xi'(x)}\,f(x).$$

To prove the sufficiency part, from (2.1) we have

$$\frac{f(x)}{1-F(x)}=\frac{n(s-1)\mu_r'(x)-n(r-1)\mu_s'(x)-(s-r)\xi'(x)}{n(s-1)\mu_r(x)-n(r-1)\mu_s(x)-n(s-r)\xi(x)}=\frac{N}{D}.$$

Now,

$$n(s-1)\mu_r(x)-n(r-1)\mu_s(x)=(s-r)\xi(x)+(n-1)(s-r)T_2(x).$$

Thus,

$$D=\frac{(n-1)(s-r)}{a}$$

and $N=(n-1)(s-r)\xi'(x)$.

Therefore,

$$\frac{f(x)}{1-F(x)}=a\xi'(x).$$

Integrating both the sides with respect to *x,* we get

$$1-F(x)=e^{-a\xi(x)},$$

and hence the sufficiency part.

Theorem 2.2: *Under the conditions as stated in Theorem 2.1.*

$$E\left[\xi(X)\mid X_{r:n}=x\right]=\lambda_r(x)=T_0(x)+\frac{r-1}{n}T_1(x)+\frac{n-r}{n}T_2(x), \tag{2.5}$$

if and only if

$$F(x)=e^{-a\xi(x)};\ \ x\in(\alpha,\beta),\ \ a\neq 0, \tag{2.6}$$

where,

$$T_1(x)=E\left[\xi(X)\mid X\le x\right]=\xi(x)+\frac{1}{a},$$

and

$$T_2(x)=E\left[\xi(X)\mid X\ge x\right]=T_1(x)-\frac{\xi(x)}{1-e^{-a\xi(x)}},$$

such that $\xi(x)F(x)\to 0$ *when* $x\to\alpha$ *and* $\xi(x)\to 0$ *when* $x\to\beta$.

Proof. Necessary part can be proved on the lines of Theorem 2.1 after noting the relation

$$F(x)=e^{-a\xi(x)}=-\frac{f(x)}{a\xi'(x)}$$

To prove sufficiency part, in view of (2.2), we have

$$\frac{f(x)}{F(x)}=-\frac{N}{D}.$$

Now,

$$n\left[(n-r)\lambda_s(x)-(n-s)\lambda_r(x)\right]=(s-r)\xi(x)+(n-1)(s-r)E\left[\xi(x)\mid X\le x\right].$$

Differentiating the above equation *w.r.t.* x, and after rearranging, we get,

$$N=(n-1)(s-r)\xi'(x)$$

$$D=(n-1)(s-r)\left\{\frac{1}{a}\right\}$$

$$\frac{f(x)}{F(x)}=-\frac{N}{D}=-a\xi'(x).$$

Now, integrating both the sides *w.r.t* x, we get

$$F(x)=e^{-a\xi(x)},\ a>0.$$

Hence the result is proved.

III. Examples and Applications

I. Examples Based on Theorem 2.1

1. Weibull distribution

Let the CDF of Weibull distribution is given as

$$F(x)=1-e^{-\lambda x^p}\ ,\ \ x>0,\lambda>0,p>0. \tag{2.7}$$

On comparison of (2.7) with (2.4), we get

$$a=\lambda \text{ and } \xi(x)=x^p.$$

Thus,

$$T_1(x)=E\left[X^p\mid X\le x\right]=\frac{1}{\lambda}-\frac{e^{-\lambda x^p}}{1-e^{-\lambda x^p}}x^p.$$

$$T_2(x)=E\left[X^p\mid X\ge x\right]=x^p+\frac{1}{\lambda}.$$

Therefore,

$$E\left[X^p\mid X_{r:n}=x\right]=\mu_r(x)=\frac{x^p}{n}+\frac{r-1}{n}\left\{\frac{1}{\lambda}-\frac{e^{-\lambda x^p}}{1-e^{-\lambda x^p}}x^p\right\}+\frac{n-r}{n}\left\{x^p+\frac{1}{\lambda}\right\}.$$

$$=x^p+\frac{n-1}{n\lambda}-\frac{r-1}{n}\frac{x^p}{1-e^{-\lambda x^p}},$$

if and only if

$$F(x)=1-e^{-\lambda x^p},\quad 0\le x<\infty;\ \lambda,p>0.$$

2. Pareto Distribution

Suppose random variable X follows Pareto distribution with CDF given as

$$F(x) = 1 - \nu^p x^{-p},\ \nu \le x < \infty,\ p > 0. \tag{2.8}$$

By comparing (2.8) with (2.4), we get

$$a = -p, \xi(x) = \log\left(\frac{\nu}{x}\right).$$

Therefore,

$$T_2(x) = E\left[\log\left(\frac{\nu}{X}\right) \mid X \ge x\right] = -\frac{1}{p} + \log\left(\frac{\nu}{x}\right),$$

or $$T_2(x) = E\left[\log X \mid X \ge x\right] = \log x + p,$$

$$T_1(x) = E\left[\log\left(\frac{\nu}{X}\right) \mid X \le x\right] = T_2(x) - \frac{\log\left(\frac{\nu}{x}\right)}{1 - \nu^p x^{-p}},$$

or $$T_1(x) = E\left[\log X \mid X \le x\right] = \log\left(\frac{\nu}{x}\right)\left[\frac{2 - \nu^p x^{-p}}{1 - \nu^p x^{-p}}\right] - p,$$

and $$\mu_r(x) = \frac{1}{n}\log\left(\frac{\nu}{x}\right) + \frac{r-1}{n}\left\{\log\left(\frac{\nu}{x}\right)\left[\frac{2 - \nu^p x^{-p}}{1 - \nu^p x^{-p}}\right] - p\right\} + \frac{n-r}{n}\left(\log x + p\right).$$

Hence,

$$E\left[\log\left(\frac{\nu}{X}\right) \mid X_{r:n} = x\right] = \frac{r}{n}\log\left(\frac{\nu}{x}\right)\left(\frac{2 - \nu^p x^{-p}}{1 - \nu^p x^{-p}}\right) + \frac{n-1}{n}p + \frac{n-r}{n}\log x,$$

or $$E\left[\log X \mid X_{r:n} = x\right] = \log \nu - \frac{r}{n}\log\left(\frac{\nu}{x}\right)\left(\frac{2 - \nu^p x^{-p}}{1 - \nu^p x^{-p}}\right) - \frac{n-1}{n}p - \frac{n-r}{n}\log x,$$

if and only if

$$F(x) = 1 - \nu^p x^{-p},\quad \nu \le x < \infty.$$

3. Gumbel Extreme Value I

Let the CDF of Gumbel extreme value I distribution is given as

$$F(x) = 1 - e^{-e^x},\ -\infty < x < \infty. \tag{2.9}$$

Now on comparison of (2.9) with (2.4), we get

$$a = 1, \xi(x) = e^x.$$

Therefore,

$$E\left[e^X \mid X \ge x\right] = T_2(x) = e^x + 1$$

$$E\left[e^X \mid X \le x\right] = T_1(x) = 1 - \frac{e^x e^{-e^x}}{1 - e^{-e^x}},$$

and $$E\left[e^{X} \mid X_{r:n}=x\right]=\mu_r(x)=\frac{n-r+1}{n}e^{x}+\frac{r-1}{n}\left(1-\frac{e^{x}e^{-e^{x}}}{1-e^{-e^{x}}}\right)+\frac{n-r}{n}.$$

if and only if

$$F(x)=1-e^{-e^{x}},\ x>0.$$

Similarly, with proper choice of a and $\xi(x)$ several other distributions can be characterized using Theorem 2.1. For more distribution belonging to this class, one may refer to Khan and Abu-Salih [18] and Noor and Athar [28].

II. Examples Based on Theorem 2.2

1. Inverse Weibull Distribution

Let random variable *X* follows inverse Weibull distribution with CDF

$$F(x)=e^{-\lambda x^{-p}},0\le x<\infty. \tag{2.10}$$

On comparison of (2.10) with (2.6), we get

Here, $a=\lambda$ and $\xi(x)=x^{-p}$.

Therefore,

$$E\left[X^{-p} \mid X\le x\right]=T_1(x)=x^{-p}+\frac{1}{\lambda},$$

$$E\left[X^{-p} \mid X\ge x\right]=T_2(x)=\frac{1}{\lambda}-\frac{x^{-p}e^{-\lambda x^{-p}}}{1-e^{-\lambda x^{-p}}},$$

and $$E\left(X^{-p} \mid X_{r:n}=x\right)=\lambda_r(x)=\frac{n-1}{n\lambda}+\left(\frac{r}{n}-e^{-\lambda x^{-p}}\right)\left(\frac{x^{-p}}{1-e^{-\lambda x^{-p}}}\right),$$

if and only if

$$F(x)=e^{-\lambda x^{-p}},0\le x<\infty.$$

2. Power Function Distribution

Let the CDF of power function distribution is

$$F(x)=\nu^{-p}x^{p},0<x<\nu;\nu,\ p>0. \tag{2.11}$$

Now on comparing (2.11) with (2.6), we get

$$a=-p \text{ and } \xi(x)=\log\left(\frac{x}{\nu}\right).$$

Therefore,

$$E\left[\log\left(\frac{X}{\nu}\right)\mid X\le x\right]=T_1(x)=\log x-\log\nu-\frac{1}{p}$$

or $$E\left[\log X \mid X\le x\right]=T_1(x)=\log x-\frac{1}{p}.$$

$$E\left[\log\left(\frac{X}{\nu}\right)\mid X\ge x\right]=T_2(x)=\log x-\frac{1}{p}-\frac{\log(x/\nu)}{1-(x/\nu)^{p}}$$

or $$E\left[\log X \mid X\ge x\right]=T_2(x)=\log\nu+\log x-\frac{1}{p}-\frac{\log(x/\nu)}{1-(x/\nu)^{p}},$$

and hence

$$E\left[\log\left(\frac{X}{\nu}\right)\mid X_{r:n}=x\right]=\lambda_r(x)=\frac{\log(x/\nu)}{n}+\frac{r-1}{n}\left(\log x-\frac{1}{p}\right)$$
$$+\frac{n-r}{n}\left(\log \nu x-\frac{1}{p}-\frac{\nu^p\log(x/\nu)}{(\nu^p-x^p)}\right)$$

or $$E\left[\log X\mid X_{r:n}=x\right]=\log x+\frac{2n-r-1}{n}\log\nu-\frac{n-r}{n}\frac{\nu^p}{(\nu^p-x^p)}\log(x/\nu)-\frac{n-1}{np}.$$

if and only if

$$F(x)=\nu^{-p}x^p, 0<x<\nu;\ \nu, p>0.$$

3. Burr type II

Let the CDF of Burr type II distribution is given as

$$F(x)=(1+e^{-x})^{-k}\qquad -\infty<x<\infty. \tag{2.12}$$

Now on comparing (2.12) with (2.6), we get

$$a=k \text{ and } \xi(x)=\log(1+e^{-x}).$$

Therefore,

$$E\left[\log(1+e^{-X})\mid X\le x\right]=T_1(x)=\log\left(1+e^{-x}\right)+\frac{1}{k},$$

$$E\left[\log(1+e^{-X})\mid X\ge x\right]=T_2(x)=\log(1+e^{-x})+\frac{1}{k}-\frac{\log(1+e^{-x})}{1-(1+e^{-x})^{-k}},$$

and

$$E\left[\log(1+e^{-X})\mid X_{r:n}=x\right]=\lambda_r(x)=\frac{n-1}{nk}+\log(1+e^{-x})-\frac{n-r}{n}\frac{\log(1+e^{-x})}{1-(1+e^{-x})^{-k}}.$$

IV. Numerical computation

In this section, we have carried out some numerical computation. In Table 1 estimated value of X using Theorem 2.1 and based on Weibull, Pareto, and Gumbel distributions for different randomly chosen truncation points are listed while Table 2 is based on Theorem 2.2 for inverse Weibull, power function and Burr type II. A random number is used to choose the various random truncation points. For the purpose of computation work, we have taken the real data set represents the failure times of 50 components (per 1000 hours) [Merovei et al. [26]].

0.036	0.058	0.061	0.074	0.078	0.086	0.102	0.103	0.114	0.116
0.148	0.183	0.192	0.254	0.262	0.379	0.381	0.538	0.570	0.574
0.590	0.618	0.645	0.961	1.228	1.600	2.006	2.054	2.804	3.058
3.076	3.147	3.625	3.704	3.931	4.073	4.393	4.534	4.893	6.274
6.816	7.896	7.904	8.022	9.337	10.940	11.020	13.880	14.730	15.080

For the power function distribution, the values in the given data set are not within an interval of [0, 1]. Thus, the original values are divided by the maximum value (15.080) and changed into the interval [0, 1]. Similarly, for the Pareto distribution, the original values are divided by the minimum value (0.036) of the data set and shifted them in the interval of (1, ∞).

Table 1. *Estimated values of X for different parameters*

Distribution	Parameters	r	x	$T_0(x)$	$T_1(x)$	$T_2(x)$	$\mu_r(x)$	$\hat{X}$
Weibull	$p=0.5, \lambda=0.5$	3	0.061	0.005	0.048	1.292	1.221	1.492
		8	0.103	0.006	0.099	1.330	1.138	1.294
		13	0.192	0.009	0.168	1.387	1.076	1.157
		18	0.538	0.015	0.317	1.515	1.092	1.193
		22	0.618	0.016	0.341	1.536	1.019	1.038
	$p=0.5, \lambda=1.0$	3	0.061	0.005	0.046	1.118	1.058	1.119
		8	0.103	0.006	0.094	1.182	1.013	1.026
		13	0.192	0.009	0.158	1.282	0.996	0.992
		18	0.538	0.015	0.292	1.524	1.089	1.187
		22	0.618	0.016	0.313	1.566	1.024	1.048
	$p=1.0, \lambda=0.5$	3	0.061	0.001	0.020	2.052	1.931	1.931
		8	0.103	0.002	0.045	2.093	1.767	1.767
		13	0.192	0.004	0.091	2.182	1.640	1.640
		18	0.538	0.011	0.256	2.523	1.714	1.714
		22	0.618	0.012	0.292	2.606	1.594	1.594
	$p=1.0, \lambda=1.0$	3	0.061	0.001	0.019	1.061	0.999	0.999
		8	0.103	0.002	0.044	1.103	0.935	0.935
		13	0.192	0.004	0.089	1.192	0.907	0.907
		18	0.538	0.011	0.243	1.538	1.078	1.078
		22	0.618	0.012	0.276	1.618	1.034	1.034
Pareto	$p=1.5$	5	2.167	0.016	0.314	1.438	1.334	0.137
		10	3.222	0.023	0.422	1.832	1.565	0.172
		15	7.278	0.040	0.560	2.636	2.042	0.277
		25	34.111	0.071	0.649	4.041	2.402	0.398
		45	259.361	0.111	0.665	2.959	0.993	0.097
	$p=4.5$	5	2.167	0.015	0.198	0.996	0.927	0.091
		10	3.222	0.023	0.216	1.392	1.176	0.117
		15	7.278	0.040	0.222	2.207	1.647	0.187
		25	34.111	0.071	0.222	3.752	2.053	0.281
		45	259.361	0.111	0.222	5.057	0.812	0.081
	$p=7.5$	5	2.167	0.016	0.131	0.907	0.842	0.084
		10	3.222	0.023	0.133	1.303	1.090	0.107
		15	7.278	0.040	0.133	2.126	1.565	0.172
		25	34.111	0.071	0.133	3.663	1.966	0.257
		45	259.361	0.111	0.133	5.522	0.781	0.079
Gumbel		4	0.074	0.022	0.022	2.077	1.934	0.659
		8	0.103	0.022	0.034	2.108	1.798	0.587
		13	0.192	0.024	0.091	2.212	1.683	0.520
		21	0.590	0.036	0.312	2.804	1.787	0.581
		29	2.804	0.330	0.722	17.511	8.089	2.091
		33	3.625	0.751	0.722	38.602	14.338	2.663
		47	11.020	1221.674	0.722	0.000	1222.338	7.108

Table 2. *Estimated values of X for different parameters*

Distribution	Parameters	r	x	$T_0(x)$	$T_1(x)$	$T_2(x)$	$\lambda_r(x)$	$\hat{X}$
Inverse Weibull	$p=1.0$ $\lambda=1.0$	4	0.074	0.270	14.513	0.998	2.059	0.486
		15	0.262	0.076	4.817	0.912	2.063	0.485
		26	1.600	0.013	1.625	0.276	0.957	1.045
		37	4.393	0.005	1.228	0.099	0.914	1.094
		48	13.880	0.002	1.072	0.005	1.009	0.991
	$p=1.0$ $\lambda=1.5$	4	0.074	0.001	14.180	0.664	1.463	0.684
		15	0.262	0.001	4.484	0.651	1.713	0.584
		26	1.600	0.013	1.292	0.259	0.783	1.278
		37	4.393	0.005	0.894	0.097	0.674	1.485
		48	13.880	0.001	0.739	0.005	0.696	1.437
	$p=1.5$ $\lambda=1.0$	4	0.074	0.001	50.677	0.999	3.962	0.399
		15	0.262	0.149	8.457	0.996	3.214	0.459
		26	1.600	0.010	1.494	0.226	0.867	1.101
		37	4.393	0.002	1.109	0.052	0.814	1.147
		48	13.880	0.001	1.0193	0.002	0.960	1.028
Power Function	$p=1.0$	7	0.008	-0.097	-4.025	-0.961	-1.406	3.696
		19	0.038	-0.065	-3.891	-0.871	-2.006	2.029
		28	0.136	-0.040	-2.889	-0.686	-1.902	2.251
		39	0.325	-0.023	-2.080	-0.459	-1.704	2.744
		46	0.726	-0.006	-1.296	-0.150	-1.185	4.611
	$p=1.5$	7	0.008	-0.097	-4.635	-0.663	-1.223	4.438
		19	0.038	-0.065	-3.854	-0.642	-1.851	2.369
		28	0.136	-0.040	-2.649	-0.561	-1.718	2.707
		39	0.325	-0.023	-1.787	-0.411	-1.471	3.463
		46	0.726	-0.006	-0.986	-0.147	-0.905	6.098
	$p=3.5$	7	0.008	-0.097	-5.022	-0.400	-1.043	5.313
		19	0.038	-0.065	-3.666	-0.399	-1.633	2.947
		28	0.136	-0.040	-2.395	-0.386	-1.503	3.354
		39	0.325	-0.023	-1.524	-0.328	-1.253	4.308
		46	0.726	-0.006	-0.720	-0.139	-0.666	7.750
Burr Type II	$k=0.5$	8	0.103	0.013	0.011	0.304	0.270	1.172
		16	0.379	0.010	0.044	0.249	0.193	1.545
		27	2.006	0.003	0.093	0.062	0.080	2.489
		38	4.534	0.000	0.092	0.005	0.069	2.633
	$k=2.0$	8	0.103	0.013	0.041	0.254	0.232	1.341
		16	0.379	0.010	0.157	0.216	0.205	1.482
		27	2.006	0.003	0.234	0.061	0.152	1.806
		38	4.534	0.000	0.201	0.005	0.149	1.827
	$k=3.5$	8	0.103	0.013	0.071	0.210	0.199	1.513
		16	0.379	0.010	0.246	0.186	0.211	1.451
		27	2.006	0.003	0.271	0.058	0.171	1.683
		38	4.534	0.000	0.203	0.005	0.151	1.811

Table 1 presents the estimated values of X for various parameters under three different distributions: Weibull, Pareto, and Gumbel. These distributions exhibit different behaviors as their parameters change. For the Weibull distribution, increasing the shape parameters $(p=0.5,1.0)$ leads to an increase in the estimated values of X. Conversely, increasing the scale parameters $(\lambda=0.5,1.0)$ and the $r-th$ order statistics results in a decrease in the estimated values of X. In the

Pareto distribution, increasing the parameters causes a decrease in the estimated values of X, while an increase in the $r-th$ order statistics results in higher estimated values of X. For the Gumbel distribution, no specific pattern is observed in the behavior of the estimated values of X. In Table 2, the estimated values X for the inverse Weibull, power function, and Burr type II distributions increase as the parameter values increase.

V. Summary

The paper contributes to the field of probability and statistics by proposing a method to characterize continuous distributions based on the conditional expectation of order statistics. This approach not only enhances theoretical understanding but also provides practical insights into modelling and analyzing data using well-known distributional forms. Further, the results are applied to some well-known continuous distributions, like Weibull, Pareto, extreme value I, inverse Weibull, power function, Burr type II and Lindley distribution. One may utilize our results to characterize more distributions belonging to these classes. The numerical computations serve to support the theoretical findings and demonstrate the applicability of the method.

References

[1] Ahsanullah, M. and Hamedani, G. G. (2007). Certain characterizations of power function and beta distributions based on order statistics. J. Stat. Theory Appl. 6: 220-226.

[2] Ahsanullah, M. and Shakil, M. (2013). Characterizations of Rayleigh distribution based on order statistics and record values. Bull. Malays. Math. Sci. Soc., 36(3), 625-635

[3] Ahsanullah, M., Shakil, M., and Kibria, B. M. (2016). Characterizations of continuous distributions by truncated moment. Journal of Modern Applied Statistics Methods, 15(1), 316-331

[4] Ali, M. A. and Khan, A. H. (1998). Characterization of some types of distributions. International Journal of Information and Management Sciences, 9(2), 1 – 9.

[5] Ansari, A., Faizan, M., & Khan, R. (2023). Characterization of new quasi lindley distribution by truncated moments and conditional expectation of order statistics. Reliability: Theory & Applications, 18(4 (76)), 168-177.

[6] Athar, H. and Abdel-Aty, Y. (2020). Characterization of general class of distribution by truncated moment. Thailand Statistician, 18(2), 95-107.

[7] Athar, H. and Akhter, Z. (2015). Some Characterization of continuous distributions based on order statistics. International Journal of Computation and Theoretical Statistics, 2(1), 31-36.

[8] Athar, H., Abdel-Aty, Y., Ali, M. A. (2021). Characterization of generalized distributions by doubly truncated moment. STATISTICA, 81(1), 25 – 44.

[9] Athar, H., Ahsanullah, M., and Ali, M. (2023). Characterisation of some generalised continuous distributions by doubly truncated moments. Operations Research and Decisions, 33(1).

[10] Athar, H., Islam, H. M. and Khan, R. U. (2007). On characterization of distributions. Journal of Applied Probability & Statistics, 2, 179-187.

[11] Balasubramanian, K. and Beg, M. I. (1992). Distributions determined by conditioning on a pair of order statistics. Metrika, 39, 107-112.

[12] Bashir, S. and Khan, H. (2023). Characterization of the weighted power function distribution by reliability functions and moments. Research in Mathematics, 10(1), 2202023.

[13] David, H. A., & Nagaraja, H. N. (2003). Order statistics. John Wiley & Sons.

[14] Franco, M. and Ruiz, J. M. (1995). On characterizations of continuous distributions with adjacent order statistics. Statistics, 26(4), 375-385.

[15] Grudzień, Z., & Szynal, D. (1995). Characterizations of distributions by moments of order statistics when the sample size is random. Applicationes Mathematicae, 23(3), 305-318.

[16] Huang, W. J. and Su, N. C. (2012). Characterizations of distributions based on moments of residual life. Comm. Statist. Theory Methods, 41(15), 2750-2761.

[17] Khan, A. H. and Abouammoh (2000). Characterizations of distributions by conditional expectation of order statistics. J. Appl. Statist. Sci., 9, 159-167.

[18] Khan, A. H. and Abu-Salih, M. S. (1989). Characterizations of probability distributions by conditional expectation of order statistics. Metron, 47, 171-181.

[19] Khan, A. H. and Alzaid, A. A. (2004). Characterization of distributions through linear regression of non-adjacent generalized order statistics. J. Appl. Statist. Sci., 13, 123-136.

[20] Khan, A. H. and Athar, H. (2004). Characterization of distributions through order statistics. Journal of Applied Statistical Science, 13(2), 147 – 154.

[21] Khan, A. H. and Khan, I. A. (1987). Moment of Order Statistics from Burr Distribution and its Characterizations. Metron, 45(1), 21-29.

[22] Khan, A. H. and Masoom Ali, M. (1987). Characterization of probability distributions through higher order gap. Comm. Statist. Theory Methods, 16(5), 1281-1287.

[23] Khan, A. H., & Athar, H. (2002). On characterization of distributions by conditioning on a pair of order statistics. The Aligarh Journal of Statistics, 22, 63-72.

[24] Kilany, N. M. (2017). Characterization of Lindley distribution based on truncated moments of order statistics. Journal of Statistics Applications & Probability, 6(2), 355-360.

[25] Lee, W. C., Wu, J. W., & Li, C. T. (2013). Characterization of mixtures of Weibull and Pareto distributions based on conditional expectation of order statistics. Communications in Statistics-Theory and Methods, 42(3), 413-428.

[26] Merovei, F., Yousof, H. M. and Hamedani, G. G. (2020). The Poisson Topp Leone generator of distributions for lifetime data: Theory, characterizations and applications. Pakistan Journal of Statistics and Operation Research, 16(2), 343–355.

[27] Nassar, M. M. (2006). On the Characterization of a mixture of generalized power function distributions by conditional expectation of order statistics. Communications in Statistics-Theory and Methods, 35(11), 1957-1962.

[28] Noor, Z. and Athar, H. (2014). Characterization of probability distributions by conditional expectation of function of record statistics. J. Egypt. Math. Soc., 22(2), 275 – 279.

A STOCHASTIC RELIABILITY MODELING APPROACH FOR MULTIPLE SYSTEM SUBSCALES

Alena Breznická[1], Ľudmila Timárová[2], Pavol Mikuš[3]

•

Faculty of Special Technology, Alexander Dubček University of Trenčín,
Ku kyselke 469, 911 06, Trenčín, Slovakia [1,2,3]
alena.breznicka@tnuni.sk
ludmila.timarova@tnuni.sk
pavol.mikus@tnuni.sk

Abstract

The article discusses the approach of stochastic simulation of the reliability of technical systems. Stochastic simulation works with variables that are expected to change with a certain probability. A stochastic model creates a projection of a model that is based on a set of random outputs. These are recorded, then the projection is repeated with a new set of random variables. Repetition takes place many times, which can be thousands or more repetitions. At the end of the process, the distribution of these outputs shows not only the most probable values and estimates, but also their limits, which are reasonable to expect. The presented paper presents the possibilities of simulation using the Matlab software package and illustrates the simulation experiment on a specific case of monitored reliability variables.

Keywords: stochastic simulation, Matlab software, reliability, failure

1. Principles of reliability modeling

The efficiency of reliability management and the use of suitable simulation methods is closely related to the perfect knowledge of all described phenomena and processes in the monitored area. Analytical activity is also such a source of research and learning. Such a method of analysis requires the merging of knowledge from various scientific fields and disciplines in the so-called system approach. The system approach describes phenomena comprehensively and focuses on their internal and external connections. The inclusion and active use of applied and theoretical system disciplines, which have their own suitable research methods, but their own research subject may be missing, is assumed [1]. The research method used is simulation modeling. From the research possibilities of the given methods, it follows that the primary position is given to computer simulation and modeling of systems, which allows us to:

- describe and express processes whose analytical solution we do not know,
- to simplify the solution of complex mathematical problems,
- shorten time-consuming experiments,
- carry out experiments of new projects,
- quickly and efficiently analyze changes and assess their consequences,
- examine a large number of failure states and assess their variants [2].

The theoretical basis for the management of such a specific area is knowledge from various scientific disciplines, which enter into the research process and thus enable the research of the given area [3]. The development of these disciplines and the application of analysis to all areas of science was conditioned in the past by the use of mathematics and the use of computer technology for the needs of modeling and simulation. Mathematical analysis of various scientific fields made it possible to apply various research tools such as modeling and simulation of systems [4,5]

Of course, it is important to include external environmental influences in the structure and quality of material inputs in these monitored processes, in the form of random deviations in quantity, etc., about which we obtain objective statistical values and characteristics. With the complexity, we cannot specify the prediction, e.g. where, when, on which machine and technique a malfunction, downtime, what material will be needed. It follows from the conclusions that these are complex, dynamic and stochastic systems [6].

The indicated properties describe the facts that we must take into account:

- some phenomena and states we do not know or cannot express stochastically, they can only be predicted with difficulty,
- connections and ties between elements and the environment are complex, they most often occur between several components. This can lead to the fact that phenomena and the probabilities of their occurrence are always conditional,
- complicated investigation of stimuli and relationships in the system,
- limited possibilities for carrying out the experiment,
- definition, or the description of the system is possible only with an extensive mathematical model, sometimes difficult to achieve, due to the bias of the input data,
- when the level of investigation is reduced, a part is torn out of the whole, distorted outputs occur [7].

In the process of investigating objective processes and phenomena, models are a suitable aid, representing the state and structure and behavior of the system in a way similar to reality, i.e. it is very convenient to consume a model and a suitable simulation.

The concept of systems simulation is based on the following facts:

1) Systems simulation is a specific form of the research process, which is used as a tool or method of researching existing or proposed new systems, we understand it as a means of creating knowledge, supporting the decision-making process, a means of describing and establishing hypotheses or forecasts.

2) The subject of simulation is systems focused on objects of knowledge and their movement, so the subject of monitoring should be dynamic systems. Simulation of systems is possible for deterministic systems as well as for stochastic systems.

3) The basic principle for system simulation is the derivation of assumptions about the simulated system using experiments. Experiments have a physical or mathematical nature, most often in the form of a simulation model. Systems must respond to a stimulus of the same input quantity in such a way that the output quantities of both systems are equal.

4) We can understand the simulation model as a system that defines the current idea of the simulated system and the behavior and reporting of the processes in it. All known knowledge, design and technical conditions, rules of operation of the investigated system must be displayed in the simulation model in such a way as to preserve the stochastic nature of events and the real arrangement of system changes over time. In the case of computer simulation with a program and its implementation on a computer.

Simulation assumes the creation of a model using a computer program that provides information about the analyzed system. The simulation approach to system analysis is different from the analytical approach, the analysis method is purely theoretical. Therefore, this reliability analysis

approach provides more flexibility and convenience. Model activities consist of events that are activated at certain points in time and thus affect the overall state of the system. The use of simulation modeling is also very advantageous when predicting system reliability. System reliability is determined or influenced by random phenomena and agents [8].

It has not only a probabilistic approach, but we express the reliability indicators:

- Probability densities
- Distribution functions
- Additional functions
- Intensities of phenomena.

The mathematical apparatus used is probability theory and mathematical statistics.

II. Methods of modeling the probabilistic behavior of random variables

We can describe and express the law of probability distribution:

1. expressing a complete description of the probability distribution of the relevant random variable by the type of probability distribution and its parameters:
   - Distribution function - CDF, F(t)
   - By complementary function - R(t)=1 - F(t)
   - Probability density - PDF, f(t)

2. numerical characteristics of the relevant random variable:
   - Location characteristics (e.g. Mean value)
   - Characteristics of variability (e.g. variance and standard deviation)
   - Quantiles (eg Median, Mode, p-quantile)

Simulation models use the following terminological basis for creating models:

- An attribute is an expression of a certain characteristic of components or elements, e.g. working capacity, type, operation cost, device number and process speed. Each element that enters the model is defined by a set number of attributes.
- A moment is a specified, prescribed value of simulation time in which at least one attribute value of a model element changes. Most often, the realization of the event occurs.
- An interval is a specified period of time between two moments.
- An event is a change in the state of an element that occurs at the moment of starting the necessary activity to change the state or the occurrence of equipment damage, the start of service or inclusion in a queue.
- Activity is the state of an element between two subsequent events describing specific changes in the state of an element such as performing an operation, waiting in a queue, transport to a service location.
- A process is a sequence of states of elements expressed by activities in a certain time span, such as what happens to the part from its removal to assembly [9].

Simulation modeling is very advantageous for analysis, modeling and prediction, especially for the possibilities of monitoring the dynamics of the plot with graphic outputs that give a clearer idea of random processes. The output quantities can be quickly compared and based on the conclusions, changes to the parameters of the system model can be implemented so that the target behavior of the system is adequate. The results of other experiments showed that the level of fault-freeness of individual groups is at a comparable, high level.

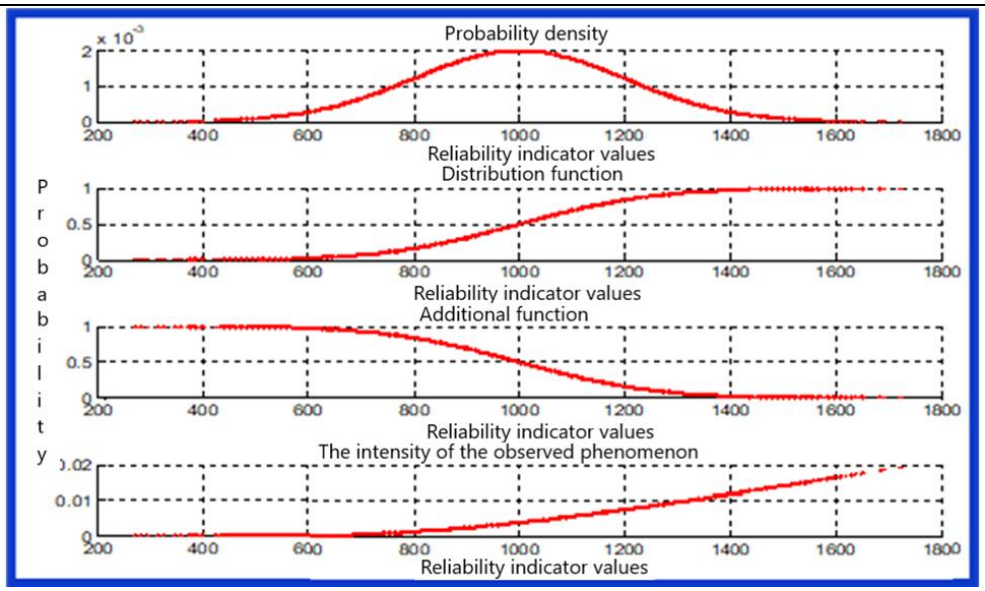


Figure 1: *Numerical characteristics of the relevant random variable*

III. The variable time step principle

We solve simulation models intended for analysis and reliability management by shifting the simulation time, basically in two ways. Constant time step - this defines a simple way of organizing the time structure. The simulation time is increased by a unit constant predetermined step from the initial time to the final time. This method is applicable for mathematical modeling, where we model events by calculating in the time of the time step shift.

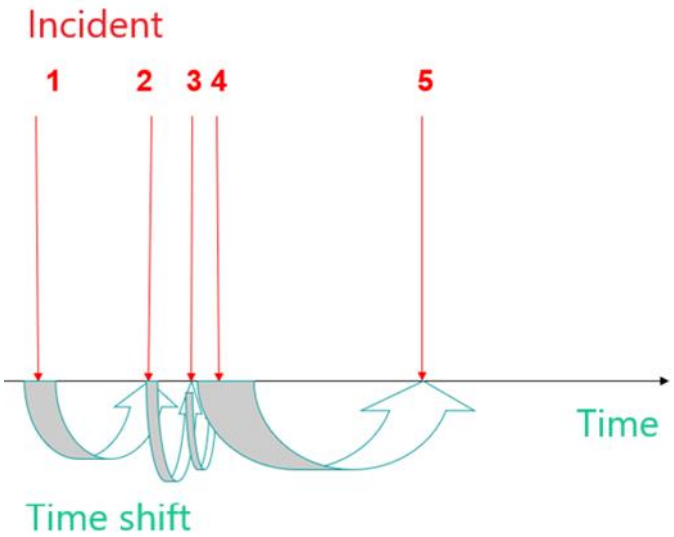


Figure 2: *The variable time step principle*

It assumes that the failure rate of the system is dependent on time, we have data on all object events in the form of a type of probability distribution and distribution parameters. System failure rate analysis is performed with the aim of obtaining the failure distribution of the entire system based on the probability distributions of its part failures.

Each object is characterized by the probability of failure - failure in the form of a statistical expression of the distribution of a random variable in the form of a function. Stochastic analysis

can be carried out by mathematical modeling. MATLAB makes it possible to determine the probability of failure in the entire range of the distribution function F(t), the probability density f(t) of each element of the system, given the knowledge of the distribution of the randomly variable time between TTF failures and the parameters of the distribution [10].

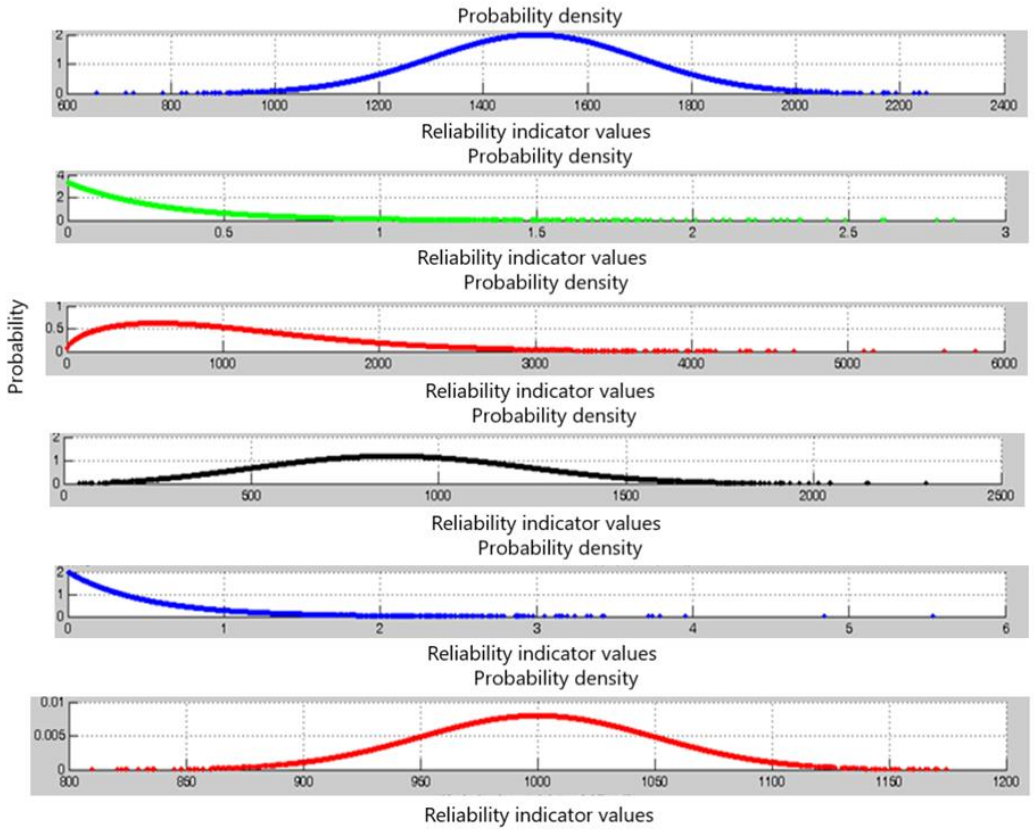


Figure 3: *Plots of simulated element probability densities*

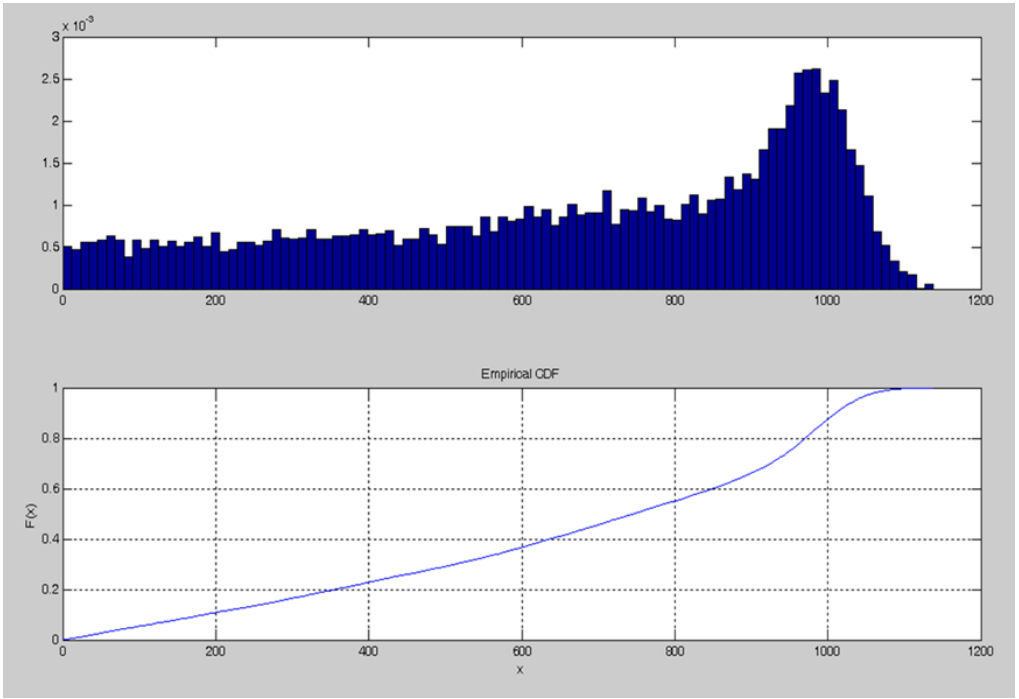


Figure 4: *Histogram of simulated system failure events*

According to the theory, the malfunction rate of the system will reach at any time of operationuntil probability of malfunction 1 is reached, a higher probability than the probability of the most malfunctioning element of the system. The probability of system failures will exceed the quantile value of 0.99.

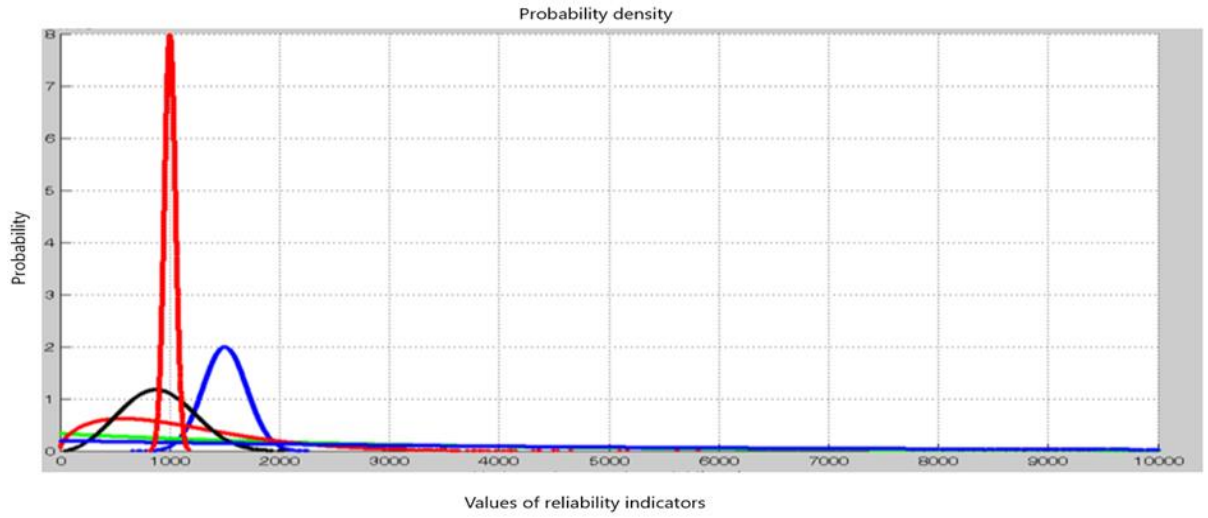


Figure 5: *Results of simulated event times between element failures*

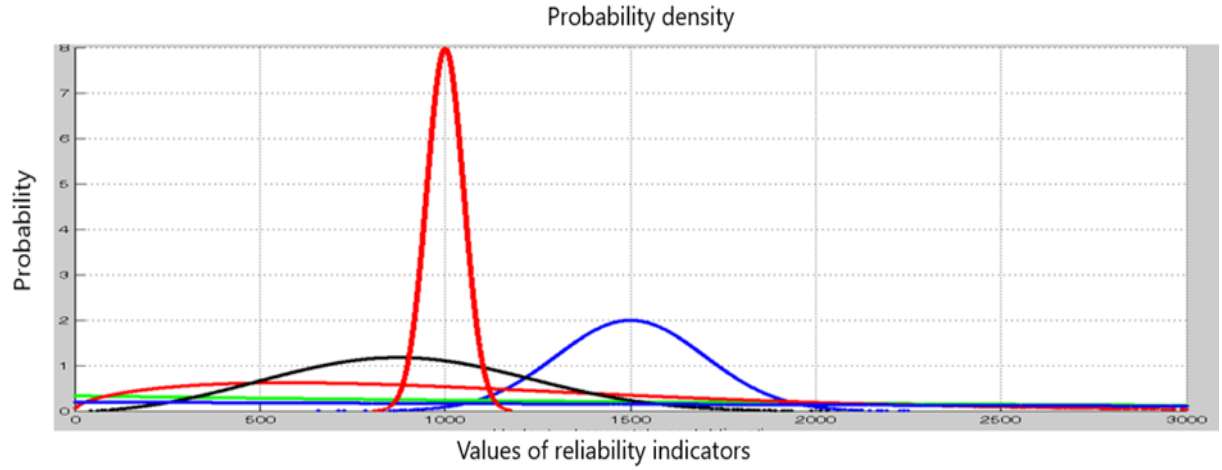


Figure 6: *Results of simulated events of operating units values*

Note the different ranges of values of operating units (times between failures of elements) on the x axis. Elements have a different assumption of occurrence of failures in terms of probability and time.

The generated times to failure of selected elements are shown in the form of probability densities and distribution functions. Static and stochastic quantitative analysis ensures the calculation (estimation) of quantitative numerical values of selected reliability indicators. The numerical value of the indicator is obtained by experimenting with the model using computer technology, considering the elementary phenomena that the model structurally connects to the behavior and analyzed states of the system [11]. The model and all input quantities are stochastic in nature, the result of the analysis is also stochastic, burdened with a certain degree of uncertainty, which can be reduced, but not completely eliminated.

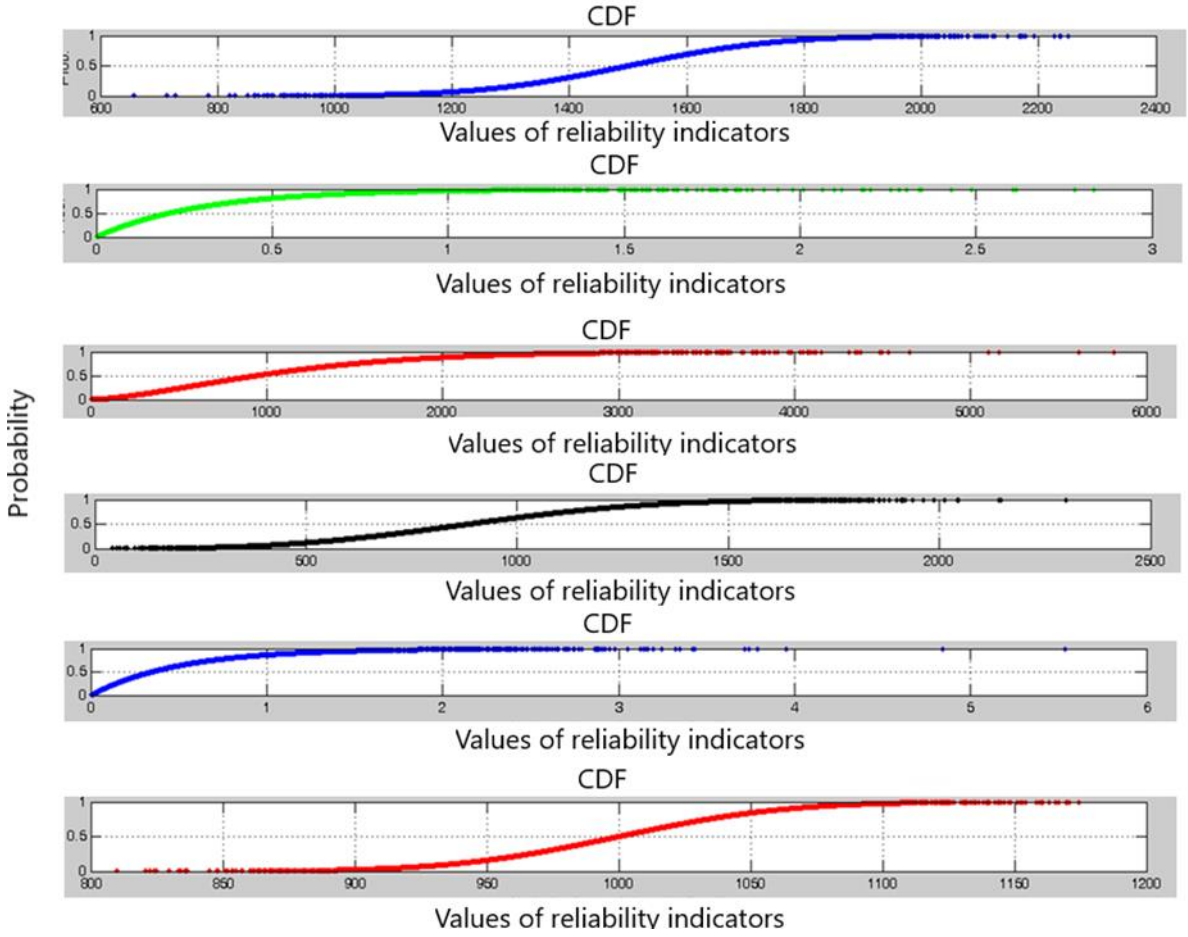


Figure 7: *Results of simulated events assuming the occurrence of faults result 1*

After the unification of the axes, we see that elements 2 and 7 have a range of assumptions for the occurrence of faults up to 30,000 - 55,000 operating units. Elements 1, 3, 4, 5, 6, 8 contribute to the failure rate of the system only in the intervals between failures 1 approx. 2400 km, 3.4- approx. 6000 km, 5.6- approx. 2500 km, 8- approx. 1200 km.

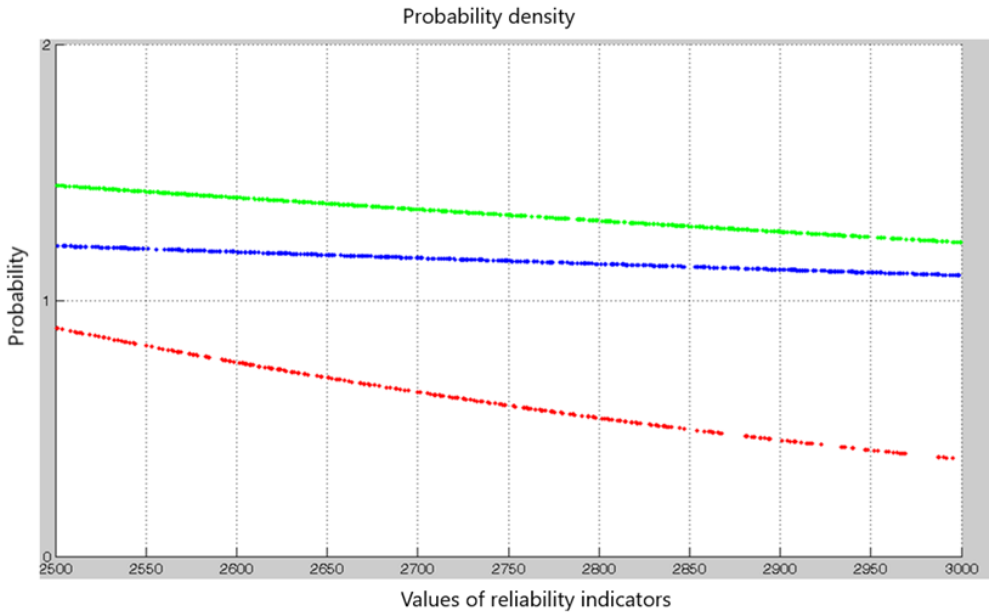


Figure 8: *Analysis of CDF results*

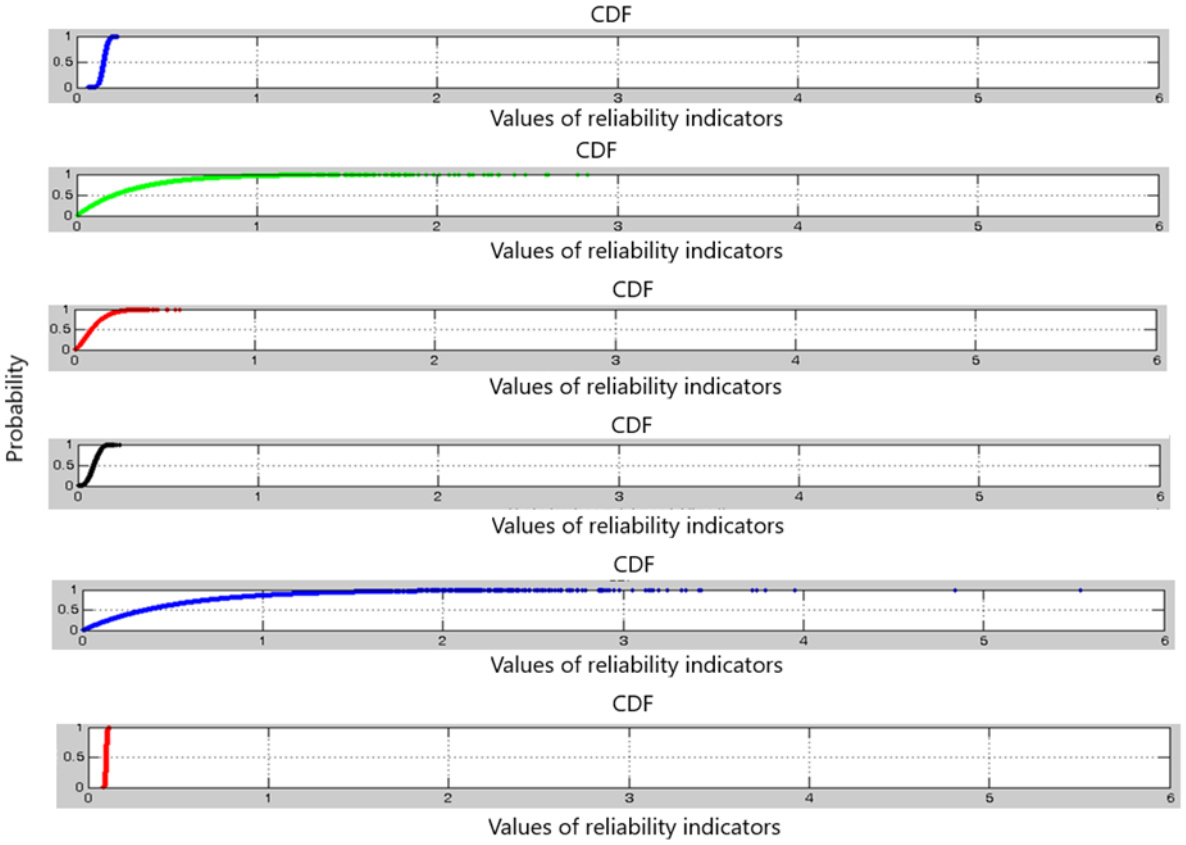


Figure 9: *Results of simulated events assuming the occurrence of faults result 2*

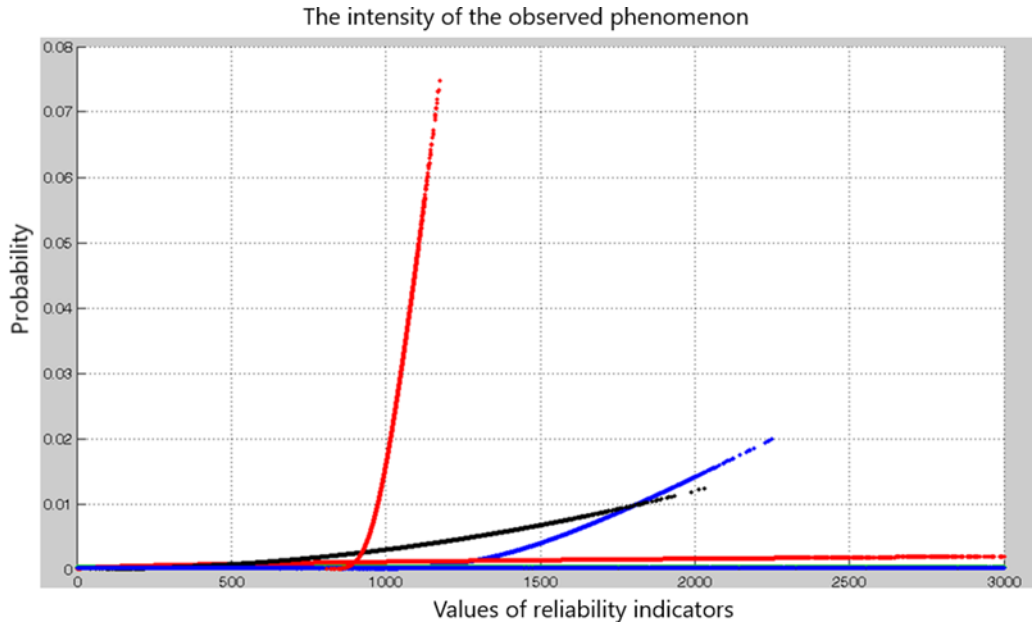


Figure 10: *Analysis of the results of element failure intensities*

IV. Results

The presented simulation is based on the principle of calculation with the possibility of using information about the values of the mean values of failure rate of individual elements from real

operation. However, it does not offer the idea of predicting the course of failures of elements and systems over time. However, stochastic simulation with the use of mathematical modeling sufficiently removes these shortcomings, and that is why the authors decided to use just such a modeling approach [12]. In the event that there are different types of distribution of the probability of failure of the elements, the graphical expression of the statistical functions is a significant and important tool to determine the probability of failure for a specific time/time of operation. We can determine the interval in which specific elements can participate in the failure rate and thus make decisions about the method of evaluating the system by changing the reliability properties of the elements [13]. The stochastic mathematical modeling approach is more suitable compared to options with deterministic calculation using mean values, which runs into several shortcomings [14,15]. The stochastic approach offers us an idea of the failure rate of specific elements and systems at a specific time.

References

[1] FABIÁNEK, D. and LEGÁT, V. and ALEŠ, Z. (2021). Weibull's analysis of the dependability of critical components of selected agricultural machinery. *In. Manufacturing Technology*. Vol. 21, No 5. pp. 605-615.

[2] RAUSAND, M. and ROYLAND, A. (2003). System Reliability Theory: Models, Statistical Methods, and Applications, 2nd Edition (*Wiley Series in Probability and Statistics*), ISBN-13: 978-0471471332.

[3] LEITNER, B. and LUSKOVÁ, M. (2019). Quantified probability estimation of traffic congestion as source of societal risks. *In: Transport Means - Proceedings of the International Conference*, Volume 2019. pp. 118 – 123, ISSN 1822296X.

[4] Al-Faris, R. Q. and Khan, S. (2008). Sine square distribution: a new statistical model based on the sine function. *Journal of Applied Probability and Statistics*, 3(1), 163-173.

[5] BREZNICKÁ, A. and MIKUŠ, P. (2024). The use of experimental modelling in the prediction of product reliability. *Reliability: Theory & Applications*, No 1 (77), Vol. 19, (77): ISSN 1932-2321, 310-319.

[6] Ashour, S. K. and Afif , M. (2007). Statistical analysis of exponentiated Weibull family
under type I progressive interval censoring with random remo vals. *Journal of Applied Sciences Research*, 3(12), 1851–1863.

[7] Gupta, R. C. and Gupta, L. and Gupta, R. D. (1998). Modeling failur e time data by Lehman alter nativ es. *Communications in Statistics-Theory and methods*, 27(4), 887–904.

[8] Pal, M, Ali, M, M, Woo, J, (2006) Exponentiated weibull distribution. *Statistica*, 66(2), 139–147.

[9] BREZNICKÁ, A. and MIKUŠ, P. (2023). Modelling of reliability indicators by means of the interference method, *Reliability: Theory and Applications*, 18(72), ISSN 1932-2321, pp. 425–431.

[10] RYKOV, V.and IVANOVA, N. and FARHADOV, M. (2022). On principles of risk analysis with a practical example, Issue № 3 (66), Volume 17, *Reliability: Theory and Applications,* ISSN 1932-2321, pages 38,44.

[11] Henley, E. J. and Kumamoto, H. (1991). Probabilistic Risk Assessment: Reliability Engineering, Design, and Analysis, *IEEE Press*, New York.

[12] DURAN, J, M. (2020) What is a Simulation Model? Minds and Machines, *Journal for Artificial Intelligence, Philosophy and Cognitive Science*, (2020). Pages 301–323.

[13] ARSHAM, H. (2018) Performance Extrapolation in Discrete-event Systems Simulation, *Journal of Systems Science*, ISBN 199627(9), Pages 863-869.

[14] SMITH, D. (2017). Reliability Maintainability and Risk Practical Methods for Engineers Including Reliability Centred Maintenance and Safety, *Taylor & Francis Inc*, ISBN 9781498785174.

[15] BIROLINI, A. (2010). Reliability Engineering, Provides design guidelines for reliability, maintainability, and software quality, *Springer*, ISBN 978-3-642-14952-8.

RT&A, No 4(80)
Volume 19, December, 2024

A PRODUCTION INVENTORY MODEL WITH TIME-DEPENDENT DEMAND, PRODUCTION AND DETERIORATION OVER A FINITE PLANNING HORIZON WITH TWO STORAGES

Neha Chauhan[1], Ajay Singh Yadav[1,*]

•

[1]Department of Mathematics, SRM Institute of Science and Technology,
Delhi-NCR Campus, Ghaziabad, India, 201204
nehac9133@gmail.com, ajaysiny@srmist.edu.in
[*]Corresponding author

Abstract

The complexities of time-dependent demand, production rates, and deterioration over a limited planning horizon are taken into consideration in our comprehensive production inventory model, which has two distinct storage facilities. In our approach, these elements work together to provide a unified framework that maximizes inventory management strategies while staying within realistic bounds. Specifically, considering the effects of both short- and long-term deterioration, we look into how various demand trends and production capacity affect stock levels and storage decisions. Organizations can lower the risk of rotting and enable dynamic modifications to production schedules by employing a dual-storage method to assess inventory allocation in greater detail. Our model makes use of advanced optimization techniques to offer useful insights into how to meet fluctuating demand while controlling the expenses of manufacturing, storage, and inventory. We demonstrate the model's efficacy and adaptability through numerical simulations and sensitivity analyses, offering managers a valuable instrument to enhance operational efficiency in scenarios including time-varying variables. This research improves the field by offering a strong solution framework for inventory management in complex scenarios with dual storage considerations, paving the way for more reliable and effective production strategies.

Keywords: Production Inventory Optimization; Time-Dependent Demand; Deterioration Management; Dual Storage Systems; Finite Planning Horizon.

1. Introduction and Review of Existing Research

One of the important concerns of the management is to decide when and how much to order or to manufacture so that the total cost associated with the inventory system should be minimum. This is somewhat more important, when the inventory under goes decay or deterioration. Most of the researchers in inventory system were directed towards non-deteriorating products. However, there are certain substances, whose utility do not remain same with the passage of time. Deterioration of these items plays an important role and items cannot be stored for a long time. Deterioration of an item may be defined as decay, evaporation, obsolescence, loss of utility or marginal value of an item that results in the decreasing usefulness of an inventory from the original condition. When the items of the commodity are kept in stock as an inventory for fulfilling the future demand,

there may be the deterioration of items in the inventory system, which may occur due to one or many factors i.e. storage conditions, weather conditions or due to humidity.

Commodities such as fruits, vegetables and foodstuffs suffer from depletion by direct spoilage while kept in store. Highly volatile liquids such as alcohol, gasoline, etc. undergo physical depletion over time through the process of evaporation. Electronic goods, photographic film, grain, chemicals, pharmaceuticals etc. deteriorate through a gradual loss of potential or utility with the passage of time. Thus deterioration of physical goods in stock is very realistic feature. As a result, while determining the optimal inventory policy of that type of products, the loss due to deterioration cannot be ignored. In the literature of inventory theory, the deteriorating inventory models have been continually modified so as to accumulate more practical features of the real inventory systems.

The analysis of deteriorating inventory began with Ghare and Schrader [4], who established the classical no-shortage inventory model with a constant rate of decay. However, it has been empirically observed that failure and life expectancy of many items can be expressed in terms of Weibull distribution. This empirical observation has prompted researchers to represent the time to deterioration of a product by a Weibull distribution. Covert and Philip [1] extended Ghare and Schrader™s [4] model and obtain an economic order quantity model for a variable rate of deterioration by assuming a two-parameter Weibull distribution. Philip [12] presented an EOQ model for items with Weibull distribution deterioration rate. An order level inventory model for a system with constant rate of deterioration was presented by Shah and Jaiswal [14]. Roychowdhury and Chaudhuri [13] formulated an order level inventory model for deteriorating items with finite rate of replenishment. Hollier and Mak [8] developed inventory replenishment policies for deteriorating item with demand rate decreases negative exponentially and constant rate of deterioration. Datta and Pal [3] investigated an order level inventory model with power demand pattern with a special form of Weibull function for deterioration rate, considering deterministic demand as well as probabilistic demand. An EOQ model for deteriorating items with a linear trend in demand was formulated by Goswami and Chaudhuri [7]. Inventory models for perishable items with stock dependent selling rate were suggested by Padmanabhan and Vrat [11]. The selling rate was assumed to be a function of current inventory level and rate of deterioration was taken to be constant with complete, partial backlogging and without backlogging. Su et al. [15] formulated a deterministic production inventory model for deteriorating items with an exponential declining demand over a fix time horizon. A production inventory model for deteriorating items with exponential declining demand was discussed by Kumar and Sharma [9]. Time horizon was fixed and the production rate at any instant was taken as the linear combination of on hand inventory and demand. A single-vender and multiple-buyers production-inventory policy for a deteriorating item was formulated by Yang and Wee [17]. Production and demand rates were taken to be constant. A mathematical model incorporating the costs of both the vender and the buyers was developed. Goyal and Giri [6] considered the production-inventory problem in which the demand, production and deterioration rates of a product were assumed to vary with time. Shortages of a cycle were allowed to be partially backlogged. An order-level inventory model for a deteriorating item with Weibull distribution deterioration, time-quadratic demand and shortages was suggested by Ghosh and Chaudhuri [5]. An economic production quantity model for deteriorating items was discussed by Teng and Chang [16]. Demand rate was taken as dependent on the display stock level and the selling price per unit. An order level inventory system for deteriorating items has been discussed by Manna and Chaudhuri [10]. The demand rate was taken as ramp type function of time and the finite production rate was proportional to the demand rate at any instant. The deterioration rate was time proportional and the unit production cost was inversely proportional to the demand rate. A note on the inventory models for deteriorating items with ramp type demand was developed by Deng et al. [2]. They have proposed an extended inventory model with ramp type demand rate and its optimal feasible solution. In their study, Yadav et al. [18] explore an inventory optimization model for decaying items that reduces carbon emissions through investments in environmentally friendly technologies. Inventory management aims to achieve a balance between environmental and financial objectives.

Yadav, Yadav, and Bansal [19] recommend utilizing an interval number strategy that combines investments in preservation technology with analytical optimization techniques to optimize the two-warehouse inventory management of degrading commodities. Their methodology enhances inventory efficiency and makes better decisions in the face of uncertainty. Mahata and Debnath [20] study a profit-maximizing inventory model that accounts for storage deterioration for a single item. Their study focuses on profit maximization but also considers how preservation techniques impact inventory management. Mahata and Debnath [21] look at the effects of RD investment and screening on an interval number unknown Economic Production Rate (EPR) inventory model. Their objective is to optimize inventory control while accounting for manufacturing defects and research expenses. Yadav, Yadav, and Bansal's [22] study optimizes a model for degrading products that considers storage degradation in a two-warehouse system with the goal of improving inventory management. Their approach builds on prior research by combining complex deterioration dynamics with multi-warehouse logistics to enhance inventory efficiency and management.

In this paper a production model is developed for deteriorating items with two-storage facility and inflation. We assume that the deterioration rates of the items stored are different in the two warehouses due to the difference in the environment conditions or preserving conditions. We assume that demand rate, production rate and deterioration rates all are functions of time. Planning horizon is taken finite and the parameters of demand function are assumed to remain constant during the finite planning horizon, rather than being reset at the beginning of each production cycle.

2. Identified Gaps and Our Contributions

The existing literature on inventory management often addresses time-dependent demand, production rates, and deterioration rates in isolation instead of combining them into a single model. Moreover, most models ignore the real significance of time restrictions in industries with discrete product life cycles by assuming infinite planning horizons. Furthermore, previous research has a propensity to ignore the complex interactions that occur between different types of storage in dynamic situations and oversimplify dual storage systems. Furthermore, insufficient research has been conducted to investigate the interplay of these time-dependent variables in various storage systems, leading to less-than-optimal decisions. To close these gaps, our work creates a comprehensive production inventory model that takes into account two distinct storage strategies and time-dependent output, deterioration, and demand across a limited planning horizon. We also propose feasible methods to handle the related complex inventory problems, closing the gap between theoretical models and practical application.

3. Assumptions and Notations

3.1. Assumptions

The following assumptions are used in this study:

1. Lead time is zero and the initial inventory level is zero.
2. Shortages are not allowed.
3. Deterioration is considered only after the inventory is stored in the warehouse.
4. The OW has a fixed capacity of w units and the RW has unlimited capacity.
5. The inventory cost (including carrying cost and deterioration cost) in RW are higher than those in OW.
6. The difference between transportation costs from OW to customers and that from RW to customers is negligible.

7. Production rate is time-dependent, given by $p = a + bf(t)$.

8. Deterioration rates in OW and RW are considered as a linear function of t, as follows $\alpha = a_1 + b_1 t$ and $\beta = a_2 + b_2 t$.

3.2. Notations

The notation used in this model are shown as follows

Table 1: *Notations*

Notation	Units	Description
H	–	total planning horizon
$p = a + bf(t)$	–	production rate
$f(t)$	–	demand rate at time t with 0<f(t)<p.
W	unit	fixed capacity level of OW.
$\alpha = a_1 + b_1 t$	–	deterioration rate of inventory items in OW with $a_1, b_1 > 0$
$\beta = a_2 + b_2 t$	–	deterioration rate of inventory items in RW with $a_2, b_2 > 0$
r	–	inflation rate.
n	constant	number of production cycles during the entire horizon H.
L_1	–	a category of production cycle that only OW is used within the cycle.
L_2	–	a category of production cycle that both OW and RW are used.
m	–	the index of production cycle whose type switches from L_2 to L_1 or from L_1 to L_2.
t_{i0}	Time	the time at the beginning of the ith production cycle belonging to L_2.
t_{i1}	Time	the time at which the inventory level in OW first reaches W units within the ith production cycle.
t_{i2}	Time	the time at the end of production of the ith production cycle.
t_{i3}	Time	the time at which all inventory units in RW are depleted within the ith production cycle.
$I_{i1}(t)$	–	inventory level in OW at time t with $t \in [t_{i0}, t_{i1}]$.
$I_{i2}(t)$	–	inventory level in RW at time t with $t \in [t_{i1}, t_{i2}]$.
$I_{i3}(t)$	–	inventory level in RW at time t with $t \in [t_{i2}, t_{i3}]$.
$I_{i4}(t)$	–	inventory level in OW at time t with $t \in [t_{i3}, t_{i+1,0}]$.
$I_{i5}(t)$	–	inventory level in OW at time t with $t \in [t_{i1}, t_{i3}]$.
t_{j0}	Time	the time at the beginning of the jth production cycle belonging to L_1.
t_{j2}	Time	the time at the end of production of the jth production cycle.
U_j	–	the maximum inventory level during the jth production cycle.
D_i	–	the quantity of deteriorated items during the ith production cycle.
D_j	–	the quantity of deteriorated items during the jth production cycle.
C_1	Cost	setup cost per production run.
C_2	Cost	cost of a deteriorated unit.
C_{OW}	Cost	carrying cost per inventory unit held in OW per unit time.
C_{RW}	Cost	carrying cost per inventory unit held in RW per unit time..
TC	Cost	total system cost during H.

4. The Mathematical model

The behavior of inventory level with time-dependent demand and production rate per deteriorating items can be represented as shown in figure1. Fig.1 (a) portrays the inventory level during a production cycle in which both OW and RW are used. Within any arbitrary production cycle i for L_2-system the cycle starts from t_{i0}, at which production, demand and deterioration occur simultaneously. The amount of stock is zero at $t_{i+1,0}$. At t_{i1} the OW is filled to its capacity and then excess of the items are stored in the RW. During $[t_{i2}, t_{i3}]$ the inventory level in the RW gradually decreases due to demand and deterioration, and it becomes zero at t_{i3}. In the OW, the inventory decreased during $[t_{i1}, t_{i3}]$ due to deterioration only, and during $[t_{i3}, t_{i+1,0}]$ the decrease in inventory is both due to demand and deterioration. At $t_{i+1,0}$ all inventory in OW will be fully exhausted.

Fig.1 (b) depicts the inventory level during a production cycle in which only OW is used. Within any arbitrary cycle j for L_1-system, the model can be considered according to two time intervals $[t_{j0}, t_{j1}]$ and $[t_{j1}, t_{j+1,0}]$. During $[t_{j0}, t_{j1}]$ the level of inventory in OW gradually increases but always remains less than W and during $[t_{j1}, t_{j+1,0}]$ the stocks in OW gradually decreases due to demand and deterioration and will be exhausted at $t_{j+1,0}$.
For L_2-system, the differential equations describing the inventory level within any production cycle i are given as follows:

$$\frac{dI_{i1}(t)}{dt} + (a_1 + b_1 t) I_{i1}(t) = a + (b-1) f(t); \quad t_{i0} \le t \le t_{i1} \tag{1}$$

$$\frac{dI_{i2}(t)}{dt} + (a_2 + b_2 t)I_{i2}(t) = a + (b-1)f(t); \quad t_{i1} \le t \le t_{i2} \tag{2}$$

$$\frac{dI_{i3}(t)}{dt} + (a_2 + b_2 t)I_{i2}(t) = -f(t); \quad t_{i2} \le t \le t_{i3} \tag{3}$$

$$\frac{dI_{i4}(t)}{dt} + (a_1 + b_1 t)I_{i4}(t) = -f(t); \quad t_{i3} \le t \le t_{i+1,0} \tag{4}$$

$$\frac{dI_{i5}(t)}{dt} + (a_1 + b_1 t)I_{i5}(t) = 0; \quad t_{i1} \le t \le t_{i3} \tag{5}$$

With the boundary conditions

$$I_{i1}(t_{i0}) = 0,\ I_{i2}(t_{i1}) = 0,\ I_{i3}(t_{i3}) = 0,\ I_{i4}(t_{i+1,0}) = 0 \ \ and \ \ I_{i5}(t_{i1}) = W.$$

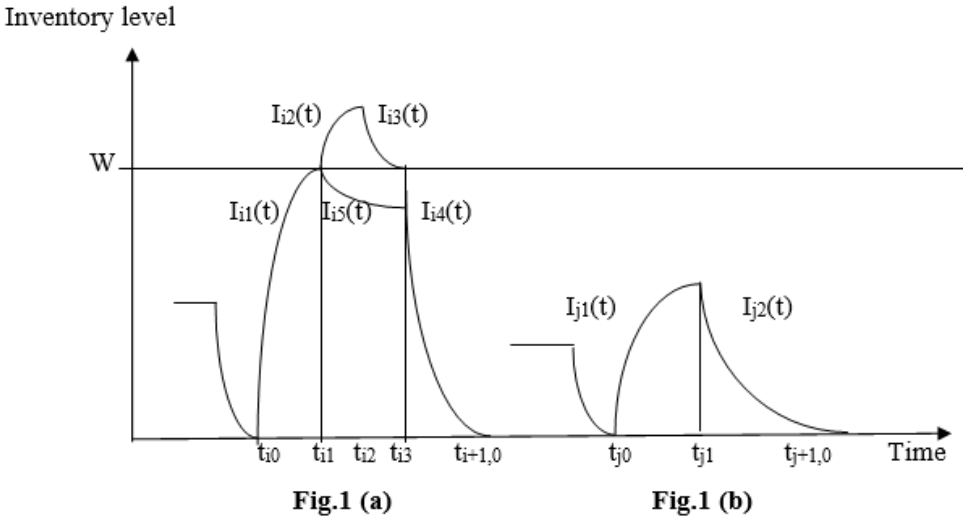


Figure 1: *Inventory level in a production system for deteriorating items with time-dependent demand.*

The above equations can be solved, respectively, as follows:
Therefore solution of equation (1) is

$$I_{i1}(t) = \int_{t_{i0}}^{t} e^{-a_1(t-u)-\frac{b_1}{2}(t^2-u^2)}[a + (b-1)f(u)]du \tag{6}$$

Similarly, solutions of others equations are

$$I_{i2}(t) = \int_{t_{i1}}^{t} e^{-a_2(t-u)-\frac{b_2}{2}(t^2-u^2)}[a + (b-1)f(u)]du \tag{7}$$

$$I_{i3}(t) = \int_{t}^{t_{i3}} e^{-a_2(t-u)-\frac{b_2}{2}(t^2-u^2)}f(u)du \tag{8}$$

$$I_{i4}(t) = \int_{t}^{t_{i+1,0}} e^{-a_1(t-u)-\frac{b_1}{2}(t^2-u^2)}f(u)du \tag{9}$$

$$I_{i5}(t) = W \cdot e^{a_1(t_{i1}-t)-\frac{b_1}{2}(t_{i1}^2-t^2)} \tag{10}$$

Now, the inventory level in RW can be calculated as

$$I_{RW,i}(t) = \int_{t_{i1}}^{t_{i2}} e^{-rt}I_{i2}(t)dt + \int_{t_{i2}}^{t_{i3}} e^{-rt}I_{i3}(t)dt \tag{11}$$

$$I_{RW,i}(t) = \left\{ \int_{t_{i1}}^{t_{i2}} e^{a_2 t + \frac{b_2}{2}t^2}[a + (b-1)f(t)]\left[(t_{i2} - t) - \frac{(a_2+r)}{2}(t_{i2}^2 - t^2) - \frac{b_2}{6}(t_{i2}^3 - t^3)\right]dt \right.$$
$$\left. \int_{t_{i2}}^{t_{i3}} e^{a_2 t + \frac{b_2}{2}t^2} f(t)\left[(t_{i2} - t) - \frac{(a_2+r)}{2}(t_{i2}^2 - t^2) - \frac{b_2}{6}(t_{i2}^3 - t^3)\right]dt \right\}$$

$$= \left(t_{i2} - \frac{(a_2+r)}{2}t_{i2}^2 - \frac{b_2}{6}t_{i2}^3\right)\left\{ \int_{t_{i2}}^{t_{i3}} e^{a_2 t + \frac{b_2}{2}t^2}[a + (b-1)f(t)]dt - \int_{t_{i2}}^{t_{i3}} e^{a_2 t + \frac{b_2}{2}t^2} f(t)dt \right\}$$
$$+ \int_{t_{i2}}^{t_{i3}} e^{a_2 t + \frac{b_2}{2}t^2}[t - \frac{(a_2+r)}{2}t^2 - \frac{b_2}{6}t^3]f(t)dt$$
$$- \int_{t_{i1}}^{t_{i2}} e^{a_2 t + \frac{b_2}{2}t^2}[t - \frac{(a_2+r)}{2}t^2 - \frac{b_2}{6}t^3][a + (b-1)f(t)]dt$$

$$= -\int_{t_{i1}}^{t_{i2}} \left(1 + a_2 t + \frac{b_2}{2}t^2\right)\left[t - \frac{(a_2+r)}{2}t^2 - \frac{b_2}{6}t^3\right](a + (b-1)f(t))dt$$
$$+ \int_{t_{i2}}^{t_{i3}} \left(1 + a_2 t + \frac{b_2}{2}t^2\right)\left[t - \frac{(a_2+r)}{2}t^2 - \frac{b_2}{6}t^3\right]f(t)dt$$

By using the relation $I_{i2}(t_{i2}) = I_{i3}(t_{i2})$, i.e.,

$$\int_{t_{i1}}^{t_{i2}} e^{-a_2(t_{i2}-t) - \frac{b_2}{2}(t_{i2}^2 - t^2)}\Big(a + (b-1)f(t)\Big)dt = \int_{t_{i2}}^{t_{i3}} e^{-a_2(t_{i2}-t) - \frac{b_2}{2}(t_{i2}^2 - t^2)} f(t)dt$$

$$= \int_{t_{i1}}^{t_{i3}} \left[t + \frac{(a_2 - r)}{2}t^2 + \frac{b_2}{3}t^3 + \right]f(t)dt - b\int_{t_{i1}}^{t_{i2}} \left[t + \frac{(a_2 - r)}{2}t^2 + \frac{b_2}{3}t^3 + \right]f(t)dt$$
$$- a\left[\frac{1}{2}(t_{i2}^2 - t_{i1}^2) + \frac{(a_2 - r)}{6}(t_{i2}^3 - t_{i1}^3) + \frac{b_2}{12}(t_{i2}^4 - t_{i1}^4)\right] \quad (12)$$

Similarly, in OW, the inventory level can be derived as

$$I_{OW,i} = \int_{t_{i0}}^{t_{i1}} e^{-rt} I_{i1}(t)dt + \int_{t_{i3}}^{t_{i+10}} e^{-rt} I_{i4}(t)dt + \int_{t_{i1}}^{t_{i3}} e^{-rt} I_{i5}(t)dt \quad (13)$$

Now

$$I_{OW,i} = \int_{t_{i0}}^{t_{i1}} e^{a_1 t + \frac{b_1}{2}t^2}\left[a + (b-1)f(t)\right]\left[(t_{i1} - t) - \frac{(a_1+r)}{2}(t_{i1}^2 - t^2) - \frac{b_1}{6}(t_{i1}^3 - t^3)\right]dt$$
$$- \int_{t_{i3}}^{t_{i+1,0}} e^{a_1 t + \frac{b_1}{2}t^2} f(t\left[(t_{i3} - t) - \frac{(a_1+r)}{2}(t_{i3}^2 - t^2) - \frac{b_1}{6}(t_{i3}^3 - t^3)\right]dt$$
$$+ we^{a_1 t_{i1} + \frac{b_1}{2}t_{i1}^2}\left[(t_{i3} - t_{i1}) - \frac{(a_1+r)}{2}(t_{i3}^2 - t_{i1}^2) - \frac{b_1}{6}(t_{i3}^3 - t_{i1}^3)\right]$$

$$= \left[\left(t_{i1} - \frac{(a_1+r)}{2}t_{i1}^2 - \frac{b_1}{6}t_{i1}^3\right)\right]\left\{\int_{t_{i0}}^{t_{i1}} e^{a_1 t + \frac{b_1}{2}t^2}\left[a + (b-1)f(t)\right]dt - we^{a_1 t_{i1} + \frac{b_1}{2}t_{i1}^2}\right\} - \int_{t_{i0}}^{t_{i1}} e^{a_1 t + \frac{b_1}{2}t^2}$$
$$\left[t - \frac{(a_1+r)}{2}t^2 - \frac{b_1}{6}t^3\right]\left[a + (b-1)f(t)\right]dt - \left[\left(t_{i3} - \frac{(a_1+r)}{2}t_{i3}^2 - \frac{b_1}{6}t_{i3}^3\right)\right]\left\{\int_{t_{i3}}^{t_{i+1,0}} e^{a_1 t + \frac{b_1}{2}t^2} f(t)dt\right.$$
$$\left. - we^{a_1 t_{i1} + \frac{b_1}{2}t_{i1}^2}\right\} + \int_{t_{i3}}^{t_{i+1,0}} e^{a_1 t + \frac{b_1}{6}t^2}\left[t - \frac{(a_1+r)}{2}t^2 - \frac{b_1}{6}t^3\right]f(t)dt$$

On using the conditions $I_{i4}(t_{i3}) = I_{i5}(t_{i3})$

$$\Rightarrow \int_{t_{i3}}^{t_{i+1,0}} e^{-a_1}(t_{i3}-t) + \frac{b_1}{6}(t_{i3}^2 - t^2)f(t)dt = we^{a_1(t_{i1}-t_{i3})+\frac{b_1}{2}(t_{i1}^2-t_{i3}^2)}$$

$$\int_{t_{i3}}^{t_{i+1,0}} e^{a_1} + \frac{b_1}{6}t^2 f(t)dt = we^{a_1+\frac{b_1}{2}t_{i1}^2}$$

and

$$I_{i1}(t_{i1}) = W$$

$$\Longrightarrow \int_{t_{i0}}^{t_{i1}} e^{-a_1(t_{i1}-t^2)+\frac{b_1}{2}(t_{i1}^2-t^2)}\Big[a+(b-1)f(t)\Big]dt = W$$

$$\Longrightarrow \int_{t_{i0}}^{t_{i1}} e^{a_1 t+\frac{b_1}{2}(t_{i1}^2-t^2)}\Big[a+(b-1)f(t)\Big]dt = We^{a_1 t_{i1}+\frac{b_1}{2}t_{i1}^2}$$

$$I_{OW,i} = \int_{t_{i3}}^{t_{i+1,0}} e^{a_1 t+\frac{b_1}{2}t^2}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]f(t)dt + \int_{t_{i0}}^{t_{i1}} e^{a_1 t+\frac{b_1}{2}t^2}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]\Big[a+(b-1)f(t)\Big]f(t)dt$$

$$I_{OW,i} = \int_{t_{i3}}^{t_{i+1,0}} \left\{1+a_1 t+\frac{b_1}{2}t^2\right\}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]f(t)dt - \int_{t_{i0}}^{t_{i1}} \left\{1+a_1 t+\frac{b_1}{2}t^2\right\}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]\Big[a+(b-1)f(t)\Big]f(t)dt$$

$$I_{OW,i} = \int_{t_{i3}}^{t_{i+1,0}} \left[t+\frac{(a_1-r)}{2}t^2+\frac{b_1}{6}t^3\right]f(t)dt - \int_{t_{i0}}^{t_{i1}} \left[t+\frac{(a_1-r)}{2}t^2+\frac{b_1}{6}t^3\right]\Big[a+(b-1)f(t)\Big]f(t)dt$$

$$I_{OW,i} = \int_{t_{i3}}^{t_{i+1,0}} \left[t+\frac{(a_1-r)}{2}t^2+\frac{b_1}{6}t^3\right]f(t)dt - (b-1)\int_{t_{i0}}^{t_{i1}} \left[t+\frac{(a_1-r)}{2}t^2+\frac{b_1}{6}t^3\right]f(t)dt - a\left[\frac{1}{2}(t_{i1}^2-t_{i0}^2)+\frac{(a_1-r)}{2}(t_{i1}^3-t_{i0}^3)+\frac{b_1}{12}(t_{i1}^4-t_{i0}^4)\right] \quad (14)$$

Additionally, the inventory units deteriorated during the production cycle i,

$$D_i = \int_{t_{i0}}^{t_{i1}} (a_1+b_1 t)e^{-rt}I_{i1}(t)dt + \int_{t_{i3}}^{t_{i+1,0}} (a_1+b_1 t)e^{-rt}I_{i4}(t)dt + \int_{t_{i1}}^{t_{i3}} (a_1+b_1 t)e^{-rt}I_{i5}(t)dt + \int_{t_{i1}}^{t_{i2}} (a_2+b_2 t)e^{-rt}I_{i2}(t)dt + \int_{t_{i2}}^{t_{i3}} (a_2+b_2 t)e^{-rt}I_{i3}(t)dt$$

$$D_i = a_1 I_{OW,i} + a_2 I_{RW,i} + b_2\left[\int_{t_{i1}}^{t_{i2}} te^{-rt}I_{i2}(t)dt + \int_{t_{i2}}^{t_{i3}} te^{-rt}I_{i3}(t)dt\right] + b_1\left[\int_{t_{i0}}^{t_{i1}} te^{-rt}I_{i1}(t)dt + \int_{t_{i3}}^{t_{i+1,0}} te^{-rt}I_{i4}(t)dt + \int_{t_{i1}}^{t_{i3}} te^{-rt}I_{i5}(t)dt\right]$$

$$D_i = a_2\left[\int_{t_{i1}}^{t_{i3}}\left(t-\frac{r}{2}t^2\right)f(t)dt - b\int_{t_{i1}}^{t_{i2}}\left(t-\frac{r}{2}t^2\right)f(t)dt - a\left\{\frac{1}{2}(t_{i2}^2-t_{i1}^2)-\frac{r}{6}(t_{i2}^3-t_{i1}^3)\right\}\right]$$
$$a_1\left[\int_{t_{i3}}^{t_{i+1,0}}\left(t-\frac{r}{2}t^2\right)f(t)dt - (b-1)\int_{t_{i0}}^{t_{i1}}\left(t-\frac{r}{2}t^2\right)f(t)dt - a\left\{\frac{1}{2}(t_{i1}^2-t_{i0}^2)-\frac{r}{6}(t_{i2}^3-t_{i1}^3)\right\}\right]$$

$$D_i = \int_{t_{i1}}^{t_{i3}}\left\{a_2t+\frac{(b_2-a_2r)}{2}t^2-\frac{b_2r}{3}t^3\right\}f(t)dt - b\int_{t_{i1}}^{t_{i2}}\left\{a_2t+\frac{(b_2-a_2r)}{2}t^2-\frac{b_2r}{3}t^3\right\}f(t)dt$$
$$+\int_{t_{i3}}^{t_{i+1,0}}\left\{a_1t+\frac{(b_1-a_1r)}{2}t^2-\frac{b_1r}{3}t^3\right\}f(t)dt-(b-1)\int_{t_{i0}}^{t_{i1}}\left\{a_1t+\frac{(b_1-a_1r)}{2}t^2-\frac{b_1r}{3}t^3\right\}f(t)dt$$
$$+a\left\{\frac{a_2}{2}(t_{i2}^2-t_{i1}^2)+\frac{(b_2-r)}{6}(t_{i2}^3-t_{i1}^3)-\frac{b_2r}{12}(t_{i2}^4-t_{i1}^4)\right\}$$
$$-a\left\{\frac{a_1}{2}(t_{i1}^2-t_{i0}^2)+\frac{(b_1-r)}{6}(t_{i1}^3-t_{i0}^3)-\frac{b_1r}{12}(t_{i1}^4-t_{i0}^4)\right\} \quad (15)$$

Since W is known and $p > f$, a unique solution of t_{i1} can be obtained by using the condition $I_{i1}(t_{i1}) = W$, for a given t_{i0}

$$\int_{t_{i0}}^{t_{i1}} e^{-a_1(t_{i1}-t)-\frac{b_1}{2}(t_{i1}^2-t^2)}\left[a+(b-1)f(t)\right]dt = W \quad (16)$$

Furthermore, from $I_{i4}(t_{i3}) = I_{i5}(t_{i3})$ a relationship between $t_{i3}, t_{i+1,0}$ and t_{i1} can be derived as

$$\int_{t_{i3}}^{t_{i+1,0}} e^{a_1t+\frac{b_1}{6}t^2}f(t)dt = we^{a_1t_{i1}+\frac{b_1}{2}t_{i1}^2} \quad (17)$$

For $f(t) > 0$, a unique solution of t_{i1} can be obtained from (6) for given $t_{i+1,0}$ and t_{i1}. .Additionally, from the condition $I_{i2}(t_{i2}) = I_{i3}(t_{i2})$, we have

$$\int_{t_{i1}}^{t_{i2}} e^{a_2t+\frac{b_2}{2}t^2}[a+bf(t)]dt = \int_{t_{i1}}^{t_{i3}} e^{a_2t+\frac{b_2}{2}t^2}f(t)dt \quad (18)$$

Then, for given t_{i1} and t_{i3}, corresponding value of t_{i2} can be determined from (18).Again, for those cycles using the L_1 -system, the above results can be simplified to express the inventory level in the one-warehouse model. For any arbitrary production cycle j in L_1 -system, I_{j1} is similar to I_{i1} behavior and I_{j2} is similar to I_{i4}. Therefore, by using the analogous results, we have

$$I_{j1}(t) = \int_{t_{j0}}^{t} e^{-a_1(t-u)-\frac{b_2}{2}(t^2-u^2)}[a+(b-1)f(u)]du, t_{j0} \le t \le t_{j1}$$

and

$$I_{j2}(t) = \int_{t}^{t_{j+1,0}} e^{-a_1(t-u)-\frac{b_2}{2}(t^2-u^2)}[a+(b-1)f(u)]du, t_{j1} \le t \le t_{j+1,0}$$

Then the inventory levels for cycle j can be calculated as

$$I_{RW,j} = 0 \quad (19)$$
$$I_{OW,j} = \int_{t_{j0}}^{t_{j1}} e^{-rt}I_{j1}(t)dt + \int_{t_{j1}}^{t_{j+1,0}} e^{-rt}I_{j2}(t)dt$$

and

$$I_{OW,j} = \left[t_{j1}-\frac{(a_1+r)}{2}t_{j1}^2-\frac{b_1}{6}t_{j1}^3\right]\left\{\int_{t_{j0}}^{t_{j1}} e^{a_1t+\frac{b_1}{2}t^2}\left[a+(b-1)f(t)\right]dt - \int_{t_{j1}}^{t_{j+1,0}} e^{a_1t+\frac{b_1}{2}t^2}f(t)dt\right\}$$
$$-\int_{t_{j0}}^{t_{j1}} e^{a_1t+\frac{b_1}{2}t^2}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]\left[a+(b-1)f(t)\right]f(t)dt$$
$$+\int_{t_{j1}}^{t_{j+1,0}} e^{a_1t+\frac{b_1}{2}t^2}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]f(t)dt$$

By using the condition $I_{j1}(t_{j1}) = I_{j2}(t_{j1})$

$$\int_{t_{j0}}^{t_{j1}} e^{-a_1(t_{j1}-t)-\frac{b_1}{2}(t_{j1}^2-t^2)}\left[a+(b-1)\right]f(t)dt = \int_{t_{j1}}^{t_{j+1,0}} e^{-a_1(t_{j1}-t)-\frac{b_1}{2}(t_{j1}^2-t^2)}f(t)dt$$

$$\Longrightarrow \int_{t_{j0}}^{t_{j1}} e^{a_1t+\frac{b_1}{2}t^2}\left[a+(b-1)\right]f(t)dt = \int_{t_{j1}}^{t_{j+1,0}} e^{a_1t+\frac{b_1}{2}t^2}f(t)dt$$

$$I_{OW,j} = \int_{t_{j0}}^{t_{j+1,0}} e^{a_1t+\frac{b_1}{2}t^2}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]f(t)dt$$

$$+\int_{t_{j0}}^{t_{j1}} e^{a_1t+\frac{b_1}{2}t^2}\left[t-\frac{(a_1+r)}{2}t^2-\frac{b_1}{6}t^3\right]\left[a+bf(t)\right]f(t)dt$$

$$I_{OW,j} = \int_{t_{j0}}^{t_{j+1,0}}\left[t+\frac{(a_1-r)}{2}t^2+\frac{b_1}{3}t^3\right]f(t)dt - \int_{t_{j0}}^{t_{j1}}\left[t+\frac{(a_1-r)}{2}t^2+\frac{b_1}{6}t^3\right]f(t)dt$$
$$-a\left[\frac{1}{2}(t_{j1}^2-t_{j0}^2)+\frac{(a_1-r)}{6}(t_{j1}^3-t_{j0}^3)+\frac{b_1}{12}(t_{j1}^4-t_{j0}^4)\right] \quad (20)$$

Now, the amount of deteriorated items during the production cycle j is

$$D_j = \int_{t_{j0}}^{t_{j1}}(a_1+b_1t)e^{-rt}I_{j1}(t)dt + \int_{t_{j1}}^{t_{j+1,0}}(a_1+b_1t)e^{-rt}I_{j2}(t)dt$$

$$D_j = a_1.I_{OW,j} + b_1\left[\int_{t_{j0}}^{t_{j1}} te^{-rt}I_{j1}(t)dt + \int_{t_{j1}}^{t_{j+1,0}} te^{-rt}I_{j2}(t)dt\right]$$

$$D_j = a_1\left[\int_{t_{j0}}^{t_{j+1,0}}\left(t-\frac{r}{2}t^2\right)f(t)dt - b\int_{t_{j0}}^{t_{j1}}\left(t-\frac{r}{2}t^2\right)f(t)dt\right] - aa_1\left[\frac{1}{2}(t_{j1}^2-t_{j0}^2)+\frac{r}{6}(t_{j1}^3-t_{j0}^3)\right]$$
$$+b_1\left[\int_{t_{j0}}^{t_{j+1,0}}\left(t-\frac{r}{2}t^2\right)f(t)dt - b\int_{t_{j0}}^{t_{j1}}\left(t-\frac{r}{2}t^2\right)f(t)dt\right] - ab_1\left[\frac{1}{6}(t_{j1}^3-t_{j0}^3)+\frac{r}{12}(t_{j1}^4-t_{j0}^4)\right]$$

$$D_j = (a_1+b_1)\left[\int_{t_{j0}}^{t_{j+1,0}}\left(t-\frac{r}{2}t^2\right)f(t)dt - b\int_{t_{j0}}^{t_{j1}}\left(t-\frac{r}{2}t^2\right)f(t)dt\right]$$
$$-a\left[\frac{a_1}{2}(t_{j1}^2-t_{j0}^2)-\frac{a_1r}{6}(t_{j1}^3-t_{j0}^3)+\frac{b_1}{6}(t_{j1}^3-t_{j0}^3)-\frac{b_1r}{12}(t_{j1}^4-t_{j0}^4)\right] \quad (21)$$

The condition $I_{j1}(t_{j1}) = I_{j2}(t_{j1})$ gives a relation among t_{j0}, t_{j1} and $t_{j+1,0}$ as follows:

$$\int_{t_{j0}}^{t_{j1}} e^{a_1t+\frac{b_1}{2}t^2}\left[a+(b-1)f(t)\right]dt = \int_{t_{j0}}^{t_{j+1,0}} e^{a_1t+\frac{b_1}{2}t^2}f(t)dt$$

Clearly, within a cycle j, at t_{j1}, maximum inventory level occurs and then

$$U_j = I_{j1}(t_{j1})$$

The total system cost, which consists of carrying cost, setup cost and deterioration cost incurred in each production cycle within the planning horizon H can be given as

$$TC = nC_1 + C_{RW}\sum_i I_{RW,i} + C_{OW}\sum_i I_{OW,i} + C_{OW}\sum_j I_{OW,i} + C_2\sum_i D_i + C_2\sum_j D_j$$

$$TC = nC_1 + C_{RW}\sum_i \Bigg[\int_{t_{i1}}^{t_{i3}} \left[t + \frac{(a_2 - r)}{2}t^2 + \frac{b_2}{3}t^3 + \right] f(t)dt - b\int_{t_{i1}}^{t_{i2}} \left[t + \frac{(a_2 - r)}{2}t^2 + \frac{b_2}{3}t^3 + \right] f(t)dt$$
$$- a\left[\frac{1}{2}(t_{i2}^2 - t_{i1}^2) + \frac{(a_2 - r)}{6}(t_{i2}^3 - t_{i1}^3) + \frac{b_2}{12}(t_{i2}^4 - t_{i1}^4) \right] \Bigg] + C_{OW}\sum_i \Bigg[\int_{t_{i3}}^{t_{i+1,0}} \left[t + \frac{(a_1 - r)}{2}t^2 + \frac{b_1}{6}t^3 \right] f(t)dt$$
$$- (b-1)\int_{t_{i0}}^{t_{i1}} \left[t + \frac{(a_1 - r)}{2}t^2 + \frac{b_1}{6}t^3 \right] f(t)dt - a\left[\frac{1}{2}(t_{i1}^2 - t_{i0}^2) + \frac{(a_1 - r)}{2}(t_{i1}^3 - t_{i0}^3) + \frac{b_1}{12}(t_{i1}^4 - t_{i0}^4) \right] \Bigg]$$
$$+ C_{OW}\sum_j \Bigg[\int_{t_{j3}}^{t_{j+1,0}} \left[t + \frac{(a_1 - r)}{2}t^2 + \frac{b_1}{6}t^3 \right] f(t)dt - (b-1)\int_{t_{j0}}^{t_{j1}} \left[t + \frac{(a_1 - r)}{2}t^2 + \frac{b_1}{6}t^3 \right] f(t)dt$$
$$- a\left[\frac{1}{2}(t_{j1}^2 - t_{j0}^2) + \frac{(a_1 - r)}{2}(t_{j1}^3 - t_{j0}^3) + \frac{b_1}{12}(t_{j1}^4 - t_{j0}^4) \right] \Bigg] + C_2\sum_i \Bigg[\int_{t_{i1}}^{t_{i3}} \left\{ a_2 t + \frac{(b_2 - a_2 r)}{2}t^2 - \frac{b_2 r}{3}t^3 \right\} f(t)dt$$
$$- b\int_{t_{i1}}^{t_{i2}} \left\{ a_2 t + \frac{(b_2 - a_2 r)}{2}t^2 - \frac{b_2 r}{3}t^3 \right\} f(t)dt + a\left\{ \frac{a_2}{2}(t_{i2}^2 - t_{i1}^2) + \frac{(b_2 - r)}{6}(t_{i2}^3 - t_{i1}^3) - \frac{b_2 r}{12}(t_{i2}^4 - t_{i1}^4) \right\}$$
$$- a\left\{ \frac{a_1}{2}(t_{i1}^2 - t_{i0}^2) + \frac{(b_1 - r)}{6}(t_{i1}^3 - t_{i0}^3) - \frac{b_1 r}{12}(t_{i1}^4 - t_{i0}^4) \right\} \Bigg] + C_2\sum_j \Bigg[(a_1 + b_1)\Bigg[\int_{t_{j0}}^{t_{j+1,0}} \left(t - \frac{r}{2}t^2 \right) f(t)dt$$
$$- b\int_{t_{j0}}^{t_{j1}} \left(t - \frac{r}{2}t^2 \right) f(t)dt \Bigg] - a\left[\frac{a_1}{2}(t_{j1}^2 - t_{j0}^2) - \frac{a_1 r}{6}(t_{j1}^3 - t_{j0}^3) + \frac{b_1}{6}(t_{j1}^3 - t_{j0}^3) - \frac{b_1 r}{12}(t_{j1}^4 - t_{j0}^4) \right] \Bigg] \tag{22}$$

Note that when the demand is increasing with time the type of production cycle may switch from L_1 to L_2. On the contrary, when the demand is decreasing with time, the switch may have the opposite direction.Therefore, the utilization of warehouse is affected by not only the capacity of OW but also the characteristics of demand.

5. Solution Process

Due to the complexity of (22) which is immensely increased by the adoption of a general demand function and production rate, it is extremely difficult to establish the property of convexity analytically, for the optimal solution. However the optimum values of t_{i0}, t_{i1}, t_{i2} and t_{i3} which minimize the cost function TC are the solutions of the equations

$$\frac{\partial TC}{\partial t_{i0}} = 0,\ \frac{\partial TC}{\partial t_{i1}} = 0,\ \frac{\partial TC}{\partial t_{i2}} = 0,\ \ and\ \ \frac{\partial TC}{\partial t_{i3}} = 0$$

Provided these values of $T_{ik},\ \ (k = 0, 1, 2, 3)$ satisfy the conditions $D_k > 0,\ \ (k = 0, 1, 2, 3)$ where D_k is the Hessian determinant of order k given by

$$D_k = \begin{vmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ C_{21} & C_{22} & \cdots & C_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ C_{k1} & C_{k2} & \cdots & C_{kk} \end{vmatrix}$$

Where $C_{kl} = \frac{\partial^2 TC}{\partial t_k \partial t_i},\ \ (k, l = 0, 1, 2, 3)$.
Also optimum values of t_{j0} and t_{j1} for TC are the solutions of the equations

$$\frac{\partial TC}{\partial t_{j0}} = 0,\ \ and\ \ \frac{\partial TC}{\partial t_{j1}} = 0$$

Provided they satisfy the sufficient conditions

$$\frac{\partial^2 TC}{\partial t_{j0}^2} > 0,\ \frac{\partial^2 TC}{\partial t_{j1}^2} > 0,\ \ and\ \ \frac{\partial TC}{\partial t_{j0}}\frac{\partial TC}{\partial t_{j1}} - \left(\frac{\partial^2 TC}{\partial t_{j0}\partial t_{j1}} \right)^2 > 0.$$

6. Numerical Example

The application of our model is illustrated by the following example:

H=10,	p=1800,	$\alpha = 0.02$,
$\beta = 0.01$,	$C_1 = 10,000\ Rs$,	$C_2 = 750\ Rs\ per\ unit$,
$C_{OW} = 200\ Rs$,	$C_{RW} = 300\ Rs$,	W=500 units,
f(t)=a e^{bt},	a=200,	b=0.1,

Solution of above example is illustrated in the table as follows: From above table it is observed that the total system cost is convex of n. The minimum of TC is achieved when seven inventory cycles ($n^* = 7$) are involved and both OW and RW are used in each cycle. However if we exceed the number of cycles more than seven then we see that total cost increases, which is verified by the table.

n	t_{i0}	TC
1	$t_{10} = 0.00$	152871.89
2	$t_{10} = 0.00,\ t_{20} = 6.4312$	87012.47
3	$t_{10} = 0.00,\ t_{20} = 4.3719,\ t_{30} = 7.0432$	64243.77
4	$t_{10} = 0.00,\ t_{20} = 3.4379,\ t_{30} = 5.7524,\ t_{40} = 7.2845$	43174.42
5	$t_{10} = 0.00,\ t_{20} = 2.3765,\ t_{30} = 4.9432,\ t_{40} = 6.4369,\ t_{50} = 8.2784$	32114.74
6	$t_{10} = 0.00,\ t_{20} = 1.9827,\ t_{30} = 4.4326,\ t_{40} = 5.5274,\ t_{50} = 7.3239,\ t_{60} = 8.5768$	27332.57
7	$t_{10} = 0.00,\ t_{20} = 1.7292,\ t_{30} = 3.4662,\ t_{40} = 4.8829,\ t_{50} = 6.0021,\ t_{60} = 7.5341,\ t_{70} = 8.7598$	26998.69
8	$t_{10} = 0.00,\ t_{20} = 1.5492,\ t_{30} = 2.9793,\ t_{40} = 4.2324,\ t_{50} = 5.1789,\ t_{60} = 6.7249,\ t_{70} = 7.9842,\ t_{80} = 8.9598$	27124.71
9	$t_{10} = 0.00,\ t_{20} = 1.3247,\ t_{30} = 2.6744,\ t_{40} = 3.9896,\ t_{50} = 4.7249,\ t_{60} = 5.9215,\ t_{70} = 7.4889,\ t_{80} = 8.3647,\ t_{90} = 9.2343$	27843.29
10	$t_{10} = 0.00,\ t_{20} = 1.0421,\ t_{30} = 2.1257,\ t_{40} = 3.9653,\ t_{50} = 4.1257,\ t_{60} = 5.3237,\ t_{70} = 7.0187,\ t_{80} = 7.9839,\ t_{90} = 8.7698,\ t_{10,0} = 9.1430$	29834.23

7. Graphical Representations

The graphical depictions in this part provide a comprehensive grasp of the dynamics within the production inventory model. Figure 2 and Figure 3 explains the link between deterioration rates and time, while Figure 4 illustrates how demand impacts inventory management usnig different type of functions of time together with the adjustments needed to effectively balance production, demand, and degradation.

1. In an inventory model, the rate of deterioration often increases as the time period is extended. This is due to the fact that goods that are exposed to the environment for extended periods of time degrade or decay more quickly. Therefore, longer time horizons usually result in larger quantities of depreciating inventory, which might impact inventory costs overall and management strategies.

2. If the demand is time-dependent and increases over time, then extending the time horizon is probably going to result in more demand. In a time-dependent demand model, the demand rate often varies with time, often increasing over time. A longer planning horizon therefore suggests that the model will account for this increasing demand, which could affect order quantities, inventory levels, and overall plans for inventory management.

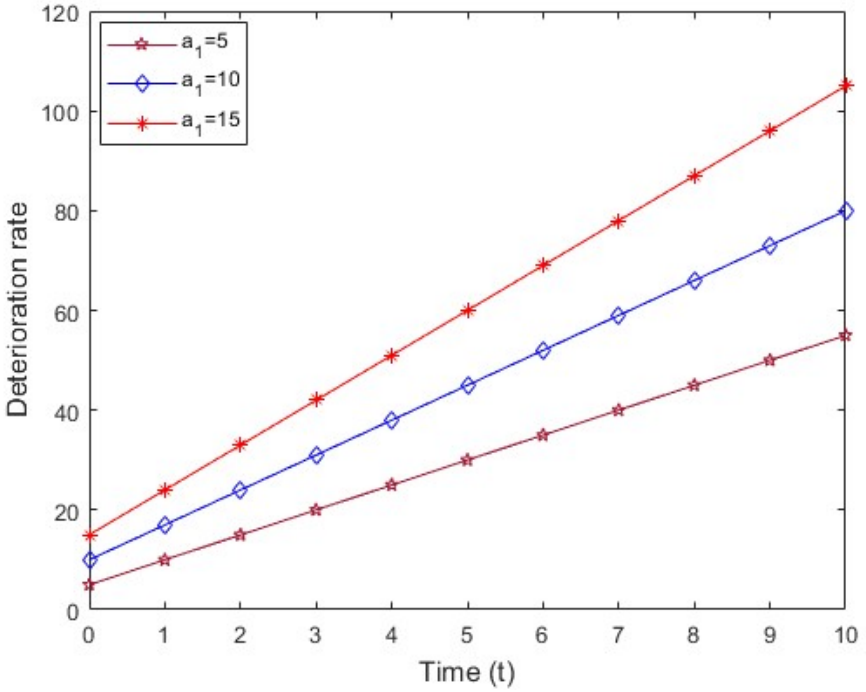


Figure 2: *Relation between deterioration rate and time, where b_1 is constant.*

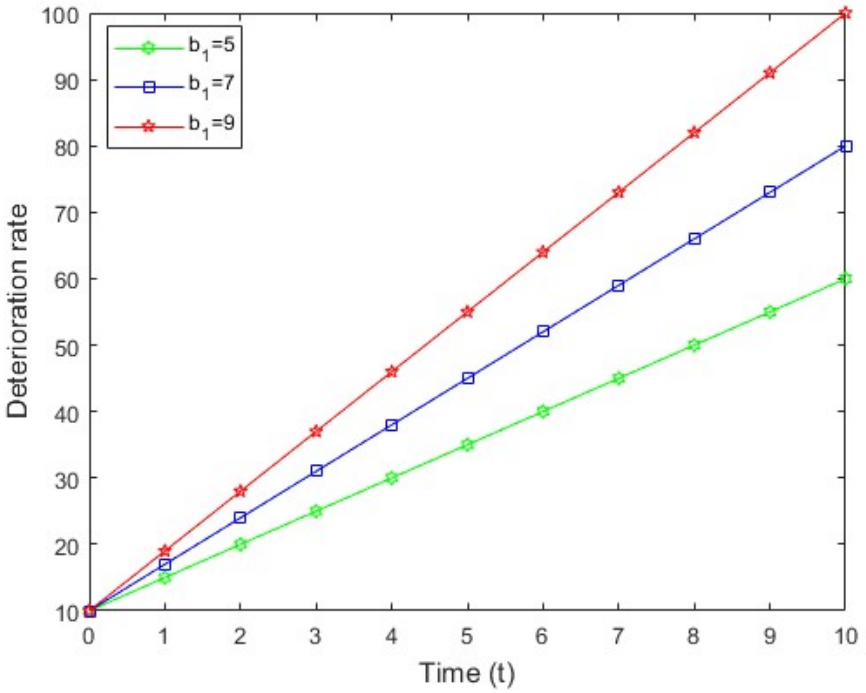


Figure 3: *Relation between deterioration rate and time, where a_1 is constant.*

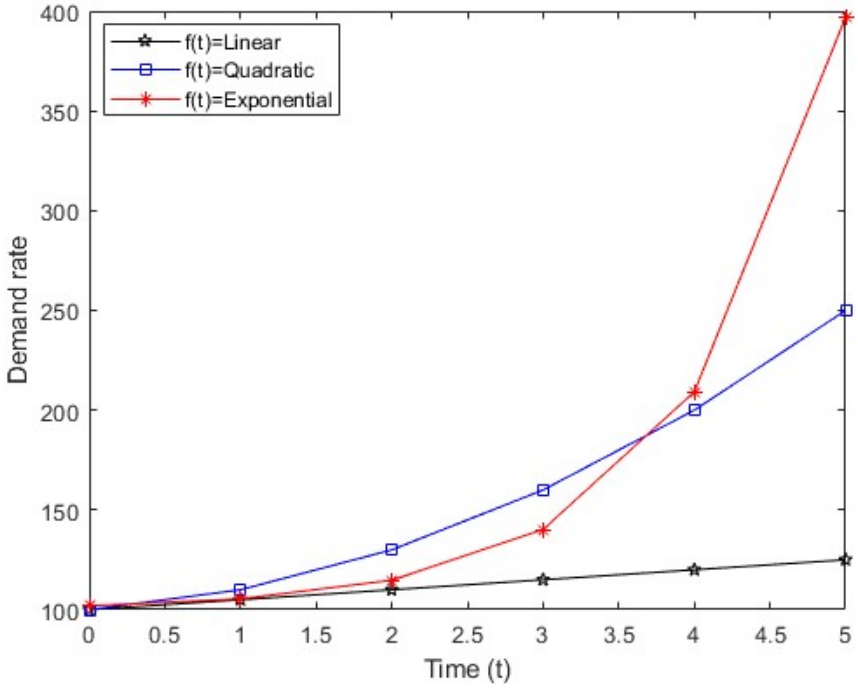


Figure 4: *Relation between demand rate and time using different functions of t.*

8. Conclusions

Most products experience a period of rapid demand increase during the introduction phase of product life cycle, level off in demand after reaching their maturity period, and will enter a period of sales decline due to new competing products or changes in consumer preference. The two-warehouse inventory control is an intriguing yet practicable issue of decision science when time-dependent demand is involved. As most of the work on the inventory model with time-dependent market demand has to determine the planning horizon before a meaningful inventory policy can be implemented, we rectify the two-warehouse inventory model with time-dependent demand, production and deterioration. This problem is different from that with constant demand case where keeping a consistent inventory level in the rented warehouse is the best solution. Due to the complexity of modeling, previous studies either adopted a heuristic approach of equal production cycle times or made ample assumptions to achieve the solution.

This general model can be applied to the inventory problem of either time increasing or time decreasing market demand.Numerical analyses indicate that both the total system cost and the number of production runs are significantly sensitive to the variation of length of planning horizon, setup cost, and demand parameters.

References

[1] Covert, Richard P., and George C. Philip. "An EOQ model for items with Weibull distribution deterioration." AIIE transactions 5.4 (1973): 323-326.

[2] Deng, Peter Shaohua, Robert H-J. Lin, and Peter Chu. "A note on the inventory models for deteriorating items with ramp type demand rate." European Journal of Operational Research 178.1 (2007): 112-120.

[3] Datta, T. K., and A. K. Pal. "A note on an inventory model with inventory-level-dependent demand rate." Journal of the Operational Research Society 41.10 (1990): 971-975.

[4] Ghare, P. M. "A model for an exponentially decaying inventory." J. ind. Engng 14 (1963): 238-243.
[5] Ghosh, S. K., and K. S. Chaudhuri. "An order-level inventory model for a deteriorating item with Weibull distribution deterioration, time-quadratic demand and shortages." Advanced modeling and optimization 6.1 (2004): 21-35.
[6] Goyal, Suresh K., and Bibhas Chandra Giri. "The production“inventory problem of a product with time varying demand, production and deterioration rates." European Journal of Operational Research 147.3 (2003): 549-557.
[7] Goswami, Adrijit, and K. S. Chaudhuri. "An EOQ model for deteriorating items with shortages and a linear trend in demand." Journal of the Operational Research Society 42.12 (1991): 1105-1110.
[8] Hollier, R. H., and K. L. Mak. "Inventory replenishment policies for deteriorating items in a declining market." International Journal of Production Research 21.6 (1983): 813-836.
[9] Kumar, Naresh, and Anil Kumar Sharma. "On deterministic production inventory model for deteriorating items with an exponential declining demand." ACTA CIENCIA INDICA MATHEMATICS 26.4 (2000): 305-310.
[10] Manna, Swapan Kumar, and K. S. Chaudhuri. "An EOQ model with ramp type demand rate, time dependent deterioration rate, unit production cost and shortages." European Journal of Operational Research 171.2 (2006): 557-566.
[11] Padmanabhan, G., and Prem Vrat. "EOQ models for perishable items under stock dependent selling rate." European Journal of Operational Research 86.2 (1995): 281-292.
[12] Philip, George C. "A generalized EOQ model for items with Weibull distribution deterioration." AIIE transactions 6.2 (1974): 159-162.
[13] Roychowdhury, M., and K. S. Chaudhuri. "An order level inventory model for deteriorating items with finite rate of replenishment." Opsearch 20 (1983): 99-106.
[14] Shah, Y. K., and M. C. Jaiswal. "An order-level inventory model for a system with constant rate of deterioration." Opsearch 14.3 (1977): 174-184.
[15] Su, Chao-Ton, Change-Wang Lin, and Chih-Hung Tsai. "A deterministic production inventory model for deteriorating items with an exponential declining demand." Opsearch 36 (1999): 95-106.
[16] Teng, Jinn-Tsair, and Chun-Tao Chang. "Economic production quantity models for deteriorating items with price-and stock-dependent demand." Computers Operations Research 32.2 (2005): 297-308.
[17] Yang, Po-Chung, and Hui-Ming Wee. "A single-vendor and multiple-buyers production“inventory policy for a deteriorating item." European Journal of Operational Research 143.3 (2002): 570-581.
[18] Yadav, A.S., Kumar, A., Yadav, K.K. et al. Optimization of an inventory model for deteriorating items with both selling price and time-sensitive demand and carbon emission under green technology investment. Int J Interact Des Manuf (2023). https://doi.org/10.1007/s12008-023-01689-8
[19] Yadav, K.K., Yadav, A.S. Bansal, S. Interval number approach for two-warehouse inventory management of deteriorating items with preservation technology investment using analytical optimization methods. Int J Interact Des Manuf (2024). https://doi.org/10.1007/s12008-023-01672-3
[20] Mahata, Sourav, and Bijoy Krishna Debnath. "A profit maximization single item inventory problem considering deterioration during carrying for price dependent demand and preservation technology investment." RAIRO-Operations Research 56.3 (2022): 1841-1856.
[21] Mahata, Sourav, and Bijoy Krishna Debnath. "The impact of R&D expenditures and screening in an economic production rate (EPR) inventory model for a flawed production system with imperfect screening under an interval-valued environment." Journal of Computational Science 69 (2023): 102027.
[22] Yadav, Krishan Kumar, Ajay Singh Yadav, and Shikha Bansal. "OPTIMIZATION OF AN INVENTORY MODEL FOR DETERIORATING ITEMS ASSUMING DETERIORATION DURING CARRYING WITH TWO-WAREHOUSE FACILITY." Reliability: Theory & Applications 19.3 (79) (2024): 442-459. DOI: https://doi.org/10.24412/1932-2321-2024-379-442-459

A CLASS OF CONTROL CHARTS FOR PROCESS LOCATION PARAMETER OF EXPONENTIAL DISTRIBUTION

*Sharada V. Bhat and Shradha Patil

•

Department of Statistics, Karnatak University, Dharwad-580003, India
bhat_sharada@yahoo.com, shradhaspatil97@gmail.com

Abstract

Control charts are essential in production processes to maintain quality of the products. Inspite of numerous control charts existing for process location under normal model, there is a need for developing control charts when situations demand production process under other distributions. In this paper, a class of control charts based on various midranges is proposed for monitoring location parameter of an ongoing process when process variables follow exponential distribution. The midranges are defined and their distributions are obtained. The performance of some members of the proposed class are evaluated in terms of their power, average run length (ARL), median run length (MRL) and standard deviation of run length (SDRL). Also, optimality and effectiveness of members of the class are discussed along with their illustration through an example.

Keywords: ARL, control limits, exponential distribution, location parameter, process control, r^{th} midrange

1. Introduction

Control charts are indispensable tools in statistical process control (SPC) for monitoring and optimizing ongoing manufacturing processes. They provide a visual representation of process variation over time, enabling practitioners to distinguish between chance cause of variation and assignable cause of variation. The monitoring of process location parameter helps in maintaining ongoing process and gives an indication about when corrective actions are required. Many control charts are developed with the assumption that process variables are taken from normal distribution, which usually do not represent some real-world scenarios. For example, chemical process, lifetime process and cutting tool wear process do not follow a normal distribution as highlighted in [3]. Therefore, studies focusing on non-normal distributions, particularly skewed distributions are of importance in decision making in process control.

In manufacturing operations, downtime often follows a two parameter exponential distribution, characterized with probability density function (pdf)

$$f(x) = \frac{1}{\lambda} e^{-\frac{x-\mu}{\lambda}}, \quad x \geq \mu, \quad -\infty < \mu < \infty, \quad \lambda > 0 \tag{1}$$

where μ represents the average downtime duration. Monitoring μ is crucial for understanding and managing production interruptions effectively. Focusing on μ, which signifies the central tendency of downtime durations, allows companies to implement proactive measures such as predictive maintenance scheduling and process optimization. Statistical process control techniques enable the monitoring of μ through tools like process average control charts. These charts aid

in identifying deviations from expected downtime durations and facilitate corrective actions to maintain operational efficiency and minimize disruptions. They enhance production reliability, resource utilization and support continuous improvement efforts aimed at reducing overall downtime resulting in improvement of productivity.

Numerous control charts have been developed for process location and scale parameters based on the assumption of normal distribution. The control charts for process location based on various approaches under normal model are discussed in [11], [13], [15], [18], [21] and [22]. The nonparametric control charts offer an alternative method to monitor ongoing processes suitable for scenarios where the distributional form of process variables is not known or it is non-normal. The nonparametric control charts based on distribution-free statistics are deliberated in [2],[4], [7], [8], [16],[17] and [23].

The quantile based statistics offering robust tools are helpful in assessing the location parameter. Various approaches of defining quantiles are discussed in [5] , [12] and [20] elaborates on r^{th} midrange and discuss its asymptotic variance. The sampling distribution of quantiles is explored in [10] along with providing foundational insights into their statistical properties. The control charts whose control limits are depending on quantiles of non-normal distributions are discussed in [3].

In recent years, the prime focus has been towards developing control charts for exponential distribution. For instance, a control chart to jointly monitor both process location and scale is proposed by [9]. An in-depth examination of the theoretical foundations and practical techniques associated with exponential distribution is provided in [1]. A control chart to monitor process stability using exponential distribution modelling for event times is studied in [19]. The control charts for joint monitoring of origin and scale parameters for ongoing processes are proposed in [14]. A median control chart is suggested to monitor process median under non normality including exponential distribution in [6].

The motivation for proposing a class of control charts based on r^{th} midrange when process variables are taken from exponential distribution is to intensify either sensitivity or robustness of control charts. Also, identifying an optimal control chart among the proposed class of control charts that rationalizes the application of the control chart under exponential model.

In this paper, we propose a class of control charts based on r^{th} midrange and obtain the optimal control chart to detect shift in the process location parameter. Section 2 elaborates on the features of r^{th} midrange. Section 3 deals with the proposed class of control charts and their evaluation. Section 4 and 5, respectively, deal with the illustration of control charts and conclusions based on our observations.

2. Role of r^{th} Midrange and its Distribution

Suppose $X_1,\ X_2, \ldots, X_n$ is a random sample of size n taken from exponential distribution $E\left(\mu,\ \lambda^{-1}\right)$ whose pdf is given in (1) and μ, λ are location and scale parameters respectively. The r^{th} midrange due to [20] is given by

$$M_r = \frac{1}{2}\left(X_{(r)} + X_{(n-r+1)}\right), \quad r = 1, 2, \ldots, \left\lfloor \frac{n}{2} + \frac{1}{2} \right\rfloor \tag{2}$$

where $X_{(r)}$ is the r^{th} order statistic and [y] is the largest integer $\leq\ y$.
The r^{th} midrange is known for its wide range of sensitivity as well as robustness to outliers. It also provides a reliable measure of central tendency which is ideal for skewed or heavy-tailed distributions. Its flexibility allows it to adapt to various statistical measures, including median, midquartile and midpercentiles. Its minimal data requirement and computational simplicity compared to measures like arithmetic mean enhances its utility. These features make it important for real time applications and statistical analysis that need quick-reliable summaries of the data.

$$\text{For } r = 1, \text{ we have,} \quad M_1 = \frac{1}{2}\left(X_{(1)} + X_{(n)}\right) \tag{3}$$

which is midrange given by the average of extreme order observations of a sample. It is a sensitive measure and offers a simple, yet effective summary of the location.
For $r = \frac{n+1}{2}$, $\frac{n+1}{4}$, $\frac{n+1}{10}$ and $\frac{n+1}{100}$ we get respectively, the

$$\text{median}, \quad M_d = X_{\left(\frac{n+1}{2}\right)} \tag{4}$$

$$\text{midquartile or midhinge}, \quad M_h = \frac{X_{\left(\frac{n+1}{4}\right)} + X_{\left(3\frac{n+1}{4}\right)}}{2}, \tag{5}$$

$$\text{middecile}, \quad M_{D_i} = \frac{X_{\left(i\frac{n+1}{10}\right)} + X_{\left((10-i)\frac{n+1}{10}\right)}}{2}, \quad i = 1, 2, \ldots, 5 \tag{6}$$

$$\text{and midpercentile}, \quad M_{P_i} = \frac{X_{\left(i\frac{n+1}{100}\right)} + X_{\left((100-i)\frac{n+1}{100}\right)}}{2}, \quad i = 1, 2, \ldots, 50. \tag{7}$$

Also, M_r is called as quasi midrange for $r \geq 2$.

Further, the p^{th} sample quantile is given by $X_{(r)}$ with

$$r = \begin{cases} np, & \text{if } np \text{ is an integer} \\ [np] + 1, & \text{if } np \text{ is not an integer} \end{cases} \tag{8}$$

where 0 < p < 1. Here, we get the median at $p = \frac{1}{2}$.

When $r = [np] + 1$, the midquantile is given by

$$M_q = \frac{X_{([np]+1)} + X_{(n-[np])}}{2}$$

$$= \frac{X_{(p(n+1))} + X_{((1-p)(n+1))}}{2} \tag{9}$$

As given in [10], suppose z_p is a sample quantile, then

$$z_p \sim N\left(\xi_p, \frac{p(1-p)}{nf^2\left(\xi_p\right)}\right) \tag{10}$$

where ξ_p is a population quantile given by $\xi_p = F^{-1}(p)$ and $f\left(\xi_p\right)$ is the pdf evaluated at ξ_p. Hence,

$$M_q = \frac{z_p + z_{1-p}}{2} \sim N\left(\frac{\xi_p + \xi_{1-p}}{2}, \sigma^2_{M_q}\right)$$

where

$$\sigma^2_{M_q} = \frac{1}{4}\left[\frac{p(1-p)}{nf^2\left(\xi_p\right)} + \frac{(1-p)p}{nf^2\left(\xi_{1-p}\right)} + \frac{2p^2}{nf\left(\xi_p\right)f\left(\xi_{1-p}\right)}\right]. \tag{11}$$

Under exponential distribution,

$$\xi_p = \mu - \lambda \log(1-p) \tag{12}$$

and

$$f\left(\xi_p\right) = \frac{1-p}{\lambda}. \tag{13}$$

Therefore,

$$E\left(M_q\right) = E\left(\frac{z_p + z_{1-p}}{2}\right)$$
$$= \frac{\mu - \lambda\log(1-p) + \mu - \lambda\log(p)}{2}$$

$$= \mu - \frac{\lambda}{2}\log(p(1-p)) \tag{14}$$

Defining $M_q^* = M_q + \frac{\lambda}{2}\log(p(1-p))$,

$$\text{we get} \quad E\left(M_q^*\right) = \mu \quad \text{and} \quad \sigma^2_{M_q^*} = \sigma^2_{M_q}. \tag{15}$$

$$\begin{aligned}
\text{Therefore, } \sigma^2_{M_q^*} &= \frac{1}{4}\left[\frac{p(1-p)\lambda^2}{n(1-p)^2} + \frac{(1-p)p\lambda^2}{np^2} + \frac{2p^2\lambda^2}{n(1-p)p}\right] \\
&= \frac{\lambda^2}{4n}\left[\frac{p}{(1-p)} + \frac{(1-p)}{p} + \frac{2p}{(1-p)}\right] \\
&= \frac{\lambda^2}{4n}\left[\frac{p^2+(1-p)^2+2p^2}{p(1-p)}\right] \\
&= \frac{\lambda^2}{4n}\left[\frac{4p^2-2p+1}{p(1-p)}\right], \quad 0 < p \leq \frac{1}{2}.
\end{aligned} \tag{16}$$

However, to obtain the mean and variance of M_1, we consider the following results due to [1].

$$E\left(X_{(r)}\right) = \mu + \lambda\sum_{j=1}^{r}\frac{1}{n-j+1}, \tag{17}$$

$$\text{Var}\left(X_{(r)}\right) = \lambda^2\sum_{j=1}^{r}\frac{1}{(n-j+1)^2}, \quad r = 1,2,\ldots,n \tag{18}$$

$$\text{and Cov}\left(X_{(r)}, X_{(s)}\right) = \lambda^2\sum_{j=1}^{r}\frac{1}{(n-j+1)^2}, \quad 1 \leq r \leq s \leq n. \tag{19}$$

Now,

$$E\left(X_{(1)}\right) = \mu + \frac{\lambda}{n} \quad \text{and} \quad E\left(X_{(n)}\right) = \mu + \lambda\log(n).$$

Hence,

$$\begin{aligned}
E\left(M_1\right) &= \frac{E\left(X_{(1)}\right) + E\left(X_{(n)}\right)}{2} \\
&= \frac{\mu + \frac{\lambda}{n} + \mu + \lambda\log(n)}{2} \\
&= \mu + \frac{\lambda}{2}\left[\frac{1}{n} + \log(n)\right].
\end{aligned}$$

$$\therefore E\left(M_1^*\right) = E\left(M_1 - \frac{\lambda}{2}\left[\frac{1}{n} + \log(n)\right]\right) = \mu. \tag{20}$$

$$\text{Also, } \text{Var}\left(X_{(1)}\right) = \frac{\lambda^2}{n^2}, \quad \text{Var}\left(X_{(n)}\right) = \lambda^2, \quad \text{and} \quad \text{Cov}\left(X_{(1)}, X_{(n)}\right) = \frac{\lambda^2}{n^2}.$$

$$\begin{aligned}
\text{Therefore, } \text{Var}\left(M_1^*\right) = \sigma^2_{M_1^*} &= \frac{1}{4}\left[\text{Var}\left(X_{(1)}\right) + \text{Var}\left(X_{(n)}\right) + 2\cdot\text{Cov}\left(X_{(1)}, X_{(n)}\right)\right] \\
&= \frac{1}{4}\left[\frac{\lambda^2}{n^2} + \lambda^2 + \frac{2\lambda^2}{n^2}\right] \\
&= \frac{\lambda^2}{4}\left[\frac{3}{n^2} + 1\right].
\end{aligned} \tag{21}$$

3. Control charts and their evaluation

In this section, we propose a class of shewhart type control charts based on r^{th} midrange, using the mean and standard deviation (sd) of the appropriate statistics. The control limits of M_r^* control charts are given by

$$UCL_{M_r^*} = E\left(M_r^*\right) + 3\sigma_{M_r^*}, \quad CL_{M_r^*} = E\left(M_r^*\right), \quad LCL_{M_r^*} = E\left(M_r^*\right) - 3\sigma_{M_r^*}. \tag{22}$$

where $UCL_{M_r^*}$, $CL_{M_r^*}$ and $LCL_{M_r^*}$ is the upper control limit, center line and lower control limit of M_r^*control chart. $E\left(M_r^*\right)$ and $\sigma_{M_r^*}$ represent respectively the mean and sd of M_r^* control chart. The variance of some members of the proposed class of control charts, M_r^* are furnished in Exhibit 1.

Exhibit 1: Variance of various members of r^{th} midrange

M_r	M_r^*	$\sigma^2_{M_r^*}$
M_1	$M_1 - \frac{\lambda}{2}\left[\frac{1}{n} + \log(n)\right]$	$\frac{\lambda^2}{4}\left[\frac{3}{n^2} + 1\right]$
M_{D_1}	$M_{D_1} - 1.2040\lambda$	$2.3333\frac{\lambda^2}{n}$
$M_{D_2} = M_{P_{20}}$	$M_{D_2} - 0.9163\lambda$	$1.1875\frac{\lambda^2}{n}$
M_h	$M_h - 0.8370\lambda$	$\frac{\lambda^2}{n}$
$M_{D_3} = M_{P_{30}}$	$M_{D_3} - 0.7803\lambda$	$0.9048\frac{\lambda^2}{n}$
$M_{D_4} = M_{P_{40}}$	$M_{D_4} - 0.7136\lambda$	$0.8750\frac{\lambda^2}{n}$
$M_{D_5} = M_{P_{50}} = M_d$	$M_{D_5} - 0.6931\lambda$	$\frac{\lambda^2}{n}$

Since $E\left(M_d^*\right) = E(M_h^*)$ and $\sigma^2_{M_d^*} = \sigma^2_{M_h^*}$, the M_d^* and M_h^* control charts and their performances will be the same.

The performance of proposed M_r^* control chart is evaluated using some performance measures viz. power, $P_{M_r^*}$; average run length, $ARL_{M_r^*}$; median run length, $MRL_{M_r^*}$ and sd of run length, $SDRL_{M_r^*}$. These measures are defined as

$$P_{M_r^*} = 1 - \beta_{M_r^*} \tag{23}$$

$$\text{where,} \quad \beta_{M_r^*} = P\left(LCL_{M_r^*} < M_r^* < UCL_{M_r^*} \mid \mu'\right) \tag{24}$$

is the operating characteristic (OC) function,

$$ARL_{M_r^*} = \frac{1}{P_{M_r^*}} \tag{25}$$

$$MRL_{M_r^*} = \frac{\log\left(0.5\right)}{\log\left(1 - P_{M_r^*}\right)} \tag{26}$$

$$\text{and} \quad SDRL_{M_r^*} = \sqrt{ARL_{M_r^*}\left(1 - ARL_{M_r^*}\right)} \tag{27}$$

The $\beta_{M_r^*}$ represents the probability of not detecting the shift a in the process location parameter in the first subsequent sample if the process location shifts from μ to $\mu^{'} = \mu + a$. The $P_{M_r^*}$ indicates the effectiveness of control chart in detecting a shift in the process location parameter. The $ARL_{M_r^*}$ represents the average number of samples needed to detect the shift and measures how quickly the control chart responds to process shifts, with a lower ARL indicating faster detection. $MRL_{M_r^*}$ provides the median number of samples required to detect a shift and it complements the ARL by offering a central tendency measure that is less influenced by extreme values. The $SDRL_{M_r^*}$ measures the variability in the number of samples needed to detect a shift and asses the consistency of control chart's performance, with lower SDRL indicating higher consistency. The values of $P_{M_r^*}$, $ARL_{M_r^*}$, $MRL_{M_r^*}$ and $SDRL_{M_r^*}$ are computed by setting $\lambda^2 = 1$. The values of $P_{M_r^*}$, $ARL_{M_r^*}$ are presented in Table 1 and $MRL_{M_r^*}$, $SDRL_{M_r^*}$ in Table 2.

When a random sample is taken from exponential distribution, the distributional form of midrange is not known. Hence using R program, we evaluate various performance measures of M_1^* for different values of n and a. The computed values of these performance measures are presented in Table 3. All the tables are given in appendix.

For various values of a and $n = 10$, we plot Figure 1 using Table 1 and 2. Figure 2 is plotted using Table 3 for different values on n.

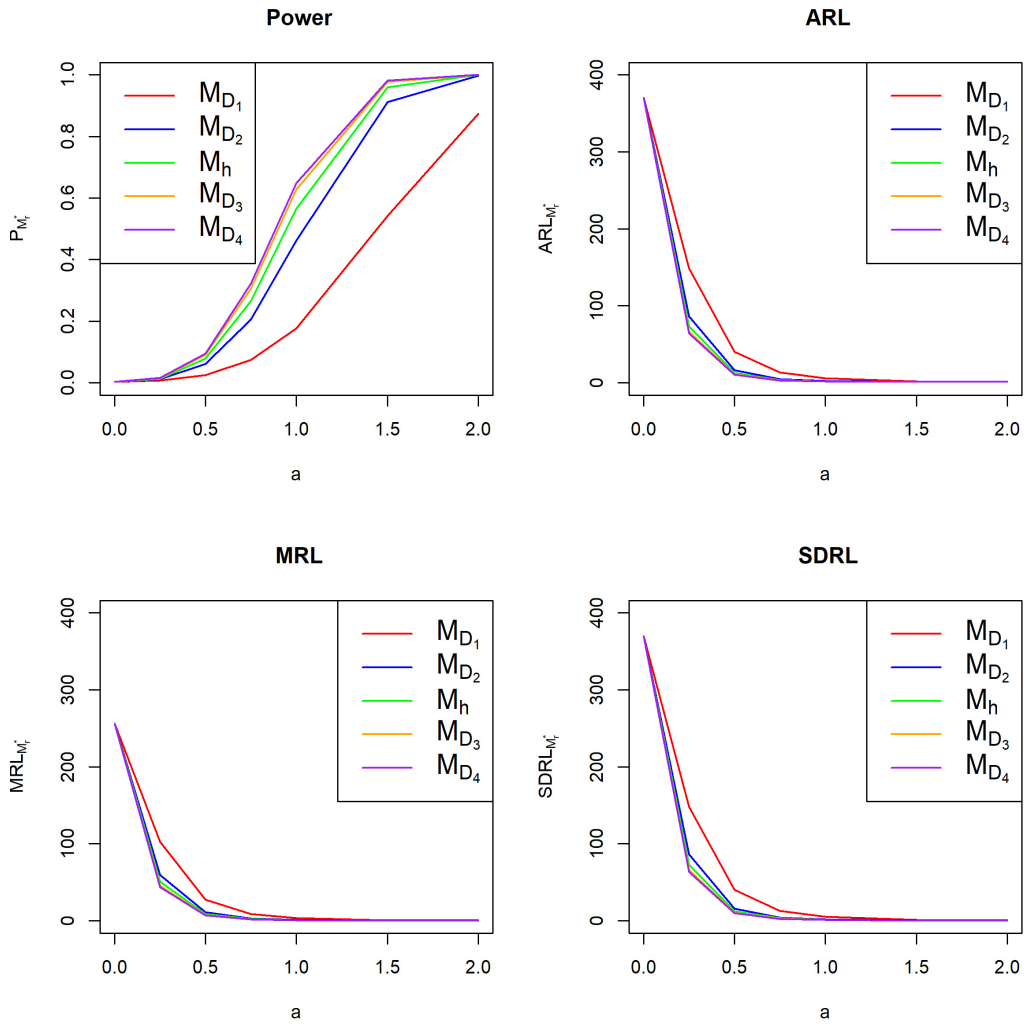


Figure 1: *Performance measures of M_r^* control charts*

From Table 1, Table 2 and Figure 1, we observe that, for fixed n and increasing a, $P_{M_r^*}$ increases, where as $ARL_{M_r^*}$, $MRL_{M_r^*}$ and $SDRL_{M_r^*}$ decreases. At $a = 0$ for various values of n, $P_{M_r^*}$ is 0.0027, $ARL_{M_r^*}$, $MRL_{M_r^*}$ and $SDRL_{M_r^*}$ are approximately 370, 256 and 369 respectively. Additionally, across different control charts, it is observed that, the $M_{D_2}^*$ control chart performs better than $M_{D_1}^*$, M_h^* is better than $M_{D_2}^*$ and $M_{D_3}^*$ outperforms M_h^* indicating a progressive improvement in performance of control charts as decile value increases.

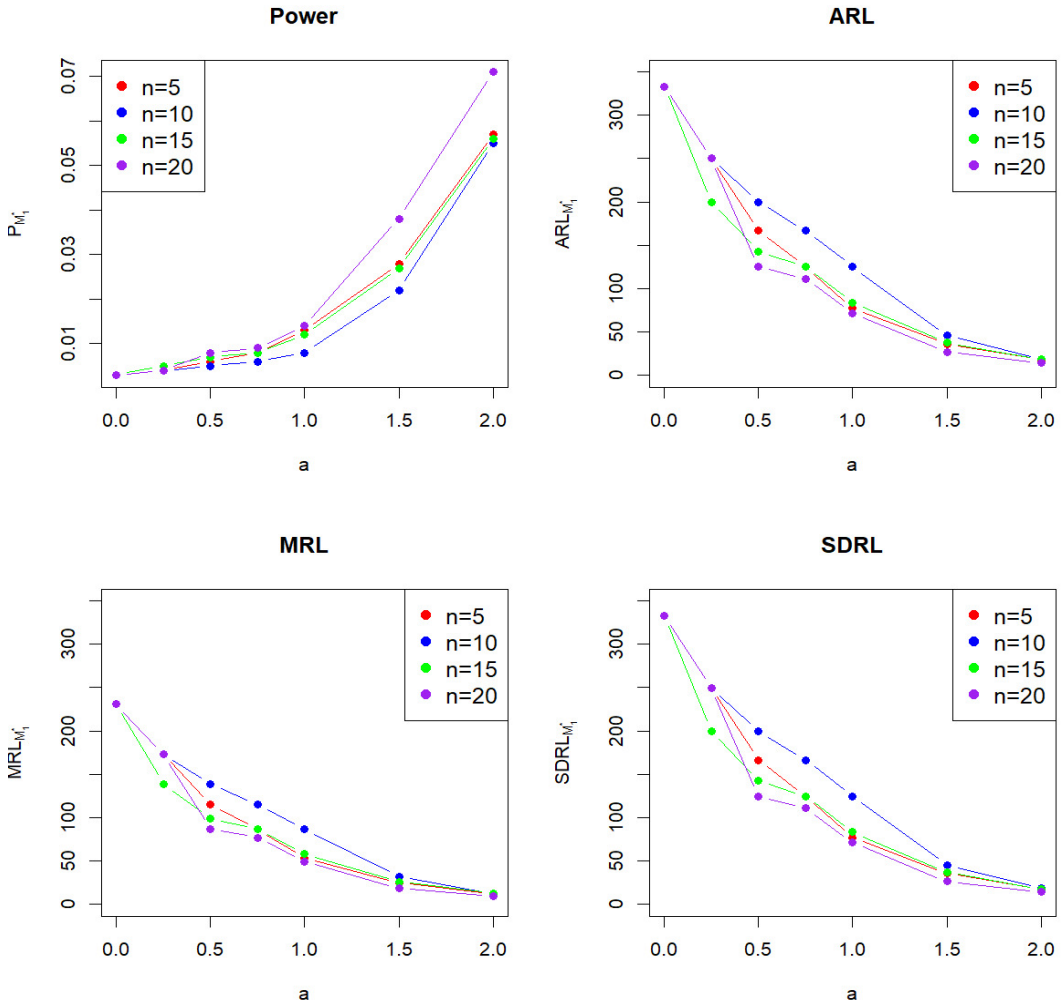


Figure 2: *Performance measures of M_1 control chart for various values of n and a*

From Table 3 and Figure 2, we observe that, as n increases, for a specified shift a, $P_{M_1^*}$ increases, $ARL_{M_1^*}$, $MRL_{M_1^*}$ and $SDRL_{M_1^*}$ decrease.
We also observe from exhibit 1 that, $\sigma^2_{M^*_{P_{30}}} > \sigma^2_{M^*_{P_{40}}}$, which reflects that there is a decreasing trend in the variance of 30^{th} percentile to 40^{th} percentile. Hence, we obtain $\sigma^2_{M^*_{P_{35}}} = 0.8681\frac{\lambda^2}{n} > \sigma^2_{M^*_{P_{40}}}$. Also, we evaluate the values of 36^{th} to 38^{th} percentile and are given by $\sigma^2_{M^*_{P_{36}}} = 0.8663\frac{\lambda^2}{n}$, $\sigma^2_{M^*_{P_{37}}} = 0.8662\frac{\lambda^2}{n}$, $\sigma^2_{M^*_{P_{38}}} = 0.8676\frac{\lambda^2}{n}$. We see that, $\sigma^2_{M^*_{P_{36}}} > \sigma^2_{M^*_{P_{37}}} < \sigma^2_{M^*_{P_{38}}}$ yielding minimum variance for 37^{th} percentile. Hence, we compute various performance measures of $M^*_{P_{37}}$ control chart for various values of n and present in Exhibit 2.

Exhibit 2: Performance measures for the $M^*_{P_{37}}$ control chart

a \ n	5	10	15	20	5	10	15	20
	$P_{M^*_{P_{37}}}$				$\text{ARL}_{M^*_{P_{37}}}$			
0.00	0.0027	0.0027	0.0027	0.0027	370.3983	370.3983	370.3983	370.3983
0.25	0.0084	0.0158	0.0250	0.0360	119.4651	63.2330	39.9287	27.7430
0.50	0.0360	0.0966	0.1790	0.2751	27.7430	10.3510	5.5877	3.6349
0.75	0.1154	0.3257	0.5482	0.7270	8.6621	3.0699	1.8243	1.3755
1.00	0.2751	0.6546	0.8773	0.9645	3.6349	1.5277	1.1399	1.0368
1.50	0.7270	0.9820	0.9994	1.0000	1.3755	1.0183	1.0006	1.0000
2.00	0.9645	0.9999	1.0000	1.0000	1.0368	1.0001	1.0000	1.0000
	$\text{MRL}_{M^*_{P_{37}}}$				$\text{SDRL}_{M^*_{P_{37}}}$			
0.00	256.3938	256.3938	256.3938	256.3938	369.8980	369.8980	369.8980	369.8980
0.25	82.4598	43.4823	27.3284	18.8813	118.9640	62.7310	39.4255	27.2384
0.50	18.8813	6.8223	3.5151	2.1544	27.2384	9.8383	5.0630	3.0948
0.75	5.6505	1.7586	0.8725	0.5339	8.1468	2.5208	1.2262	0.7186
1.00	2.1544	0.6521	0.3304	0.2077	3.0948	0.8978	0.3994	0.1954
1.50	0.5339	0.1726	0.0933	0.0616	0.7186	0.1367	0.0244	0.0036
2.00	0.2077	0.0728	0.0413	0.0281	0.1954	0.0086	0.0002	0.0000

From Exhibit 2 and Tables 1, 2 and 3, we observe that, $M^*_{P_{37}}$ control chart displays the highest power and the lowest ARL, MRL and SDRL when compared to other control charts within the proposed class. Hence, we consider $M^*_{P_{37}}$ as the optimal control chart among the class of M^*_r control charts.

4. Illustration

In this section, we provide an example to illustrate the class of M_r control charts using partial data from [3]. The dataset comprises the failure times of light bulbs, recorded in units across 10 samples, each consisting of 10 observations. Various values of M_r for each sample is computed in Exhibit 3 (a). Here, $n = 10$, μ is estimated by $\overline{M}_r = \frac{1}{10}\sum_{i=1}^{10} M_{r_i}$ and σ_{M_r} by $\widehat{\sigma}_{M_r} = \delta\widehat{\lambda}$, where $\delta = \sqrt{\frac{n-1}{2}}\left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right)$, $\widehat{\lambda} = \frac{1}{10}\sum_{i=1}^{10} s_i$, $s_i^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x})^2$.

Exhibit 3 (a): CL_{M_r} of various control charts

Sl. no	M_1	$M_{(D_1)}$	$M_{(D_2)}$	M_h	$M_{(D_3)}$	$M_{(P_{37})}$	$M_{(D_4)}$	$\bar{x}$	s_i
1	0.6600	1.0025	0.8960	0.8438	0.7965	0.7155	0.6750	0.8550	0.5355
2	0.6500	1.1275	1.0340	1.0112	1.0135	0.9474	0.8920	1.0140	0.4709
3	1.1700	1.0020	0.9840	0.9725	0.9590	0.9368	0.9260	0.9720	0.5548
4	1.6100	0.6885	0.6320	0.5925	0.5340	0.4851	0.4770	0.6380	0.4327
5	0.6500	0.9640	0.9200	0.8887	0.8505	0.7970	0.7740	0.8870	0.4436
6	0.9500	0.7370	0.7460	0.7688	0.7935	0.8578	0.8970	0.8420	0.4411
7	0.9800	0.7125	0.6460	0.6512	0.6895	0.7150	0.7150	0.7010	0.3990
8	1.1700	1.0760	0.9140	0.8700	0.8700	0.8485	0.8310	0.9170	0.4607
9	1.6100	0.8310	0.8360	0.7888	0.6965	0.6136	0.5960	0.7590	0.5363
10	1.2500	1.0230	0.9890	1.0100	1.0640	1.0719	1.0490	1.0250	0.5394
$CL_{M_r} = \bar{M}_r$	1.0700	0.9164	0.8597	0.8398	0.8267	0.7989	0.7832	$\bar{x} = 0.8610$	$\hat{\lambda} = 0.4814$

Exhibit 3 (b): σ_{M_r}, UCL_{M_r}, CL_{M_r}, LCL_{M_r} and w_{M_r} of various control charts

Statistic	σ_{M_r}	UCL_{M_r}	CL_{M_r}	LCL_{M_r}	w_{M_r}
M_1	0.2511	1.8234	1.0700	0.3166	1.5069
M_{D_1}	0.2391	1.6336	0.9164	0.1992	1.4344
M_{D_2}	0.1706	1.3714	0.8597	0.3480	1.0233
M_h	0.1565	1.3093	0.8398	0.3703	0.9391
M_{D_3}	0.1489	1.2733	0.8267	0.3801	0.8932
$M_{P_{37}}$	0.1457	1.2359	0.7989	0.3619	0.8740
M_{D_4}	0.1464	1.2224	0.7832	0.3440	0.8784

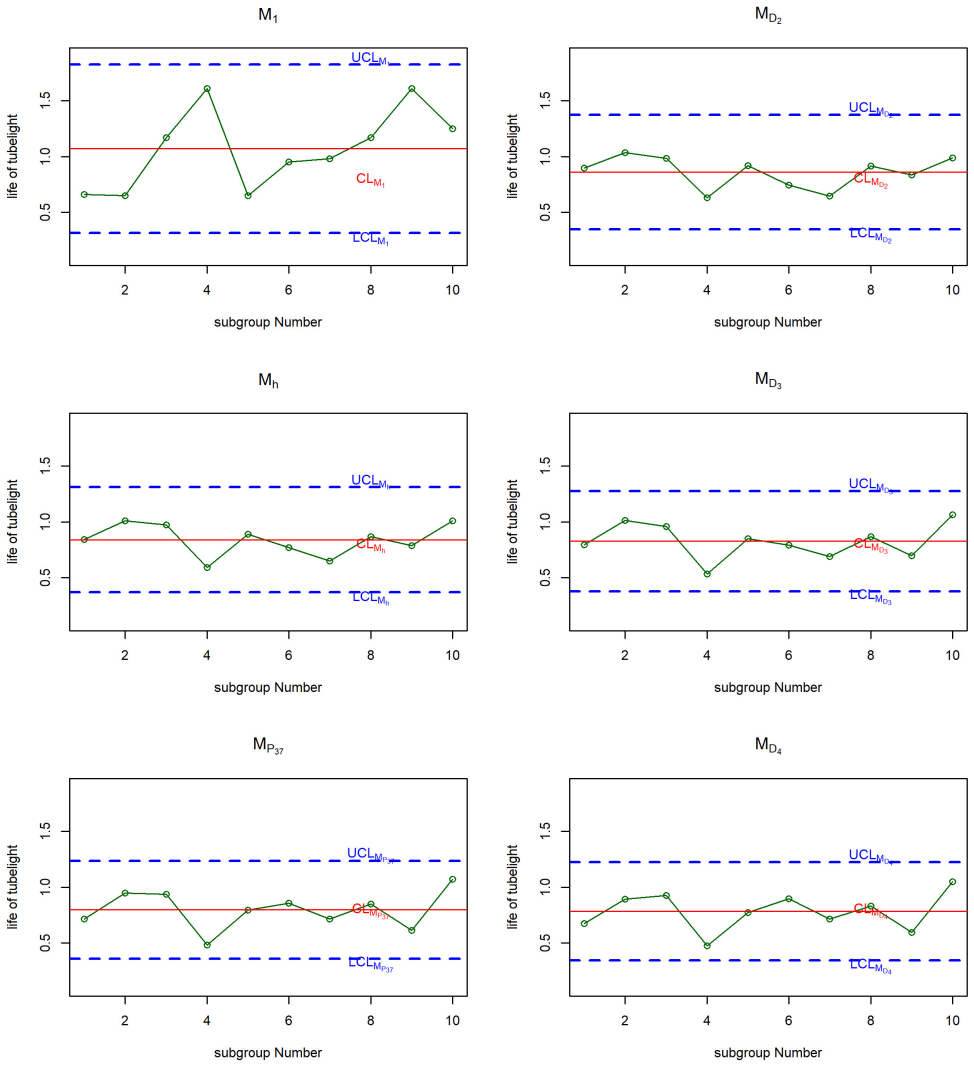


Figure 3: *M_1, M_{D_2}, M_h, M_{D_3}, $M_{P_{37}}$, M_{D_4} control charts*

From Exhibit 3 (b) and Figure 3 it is observe that, the M_1, M_{D_2}, M_h, M_{D_3}, $M_{P_{37}}$ and M_{D_4} control charts show that the process is in control. Further, the w_{M_r} is largest for M_{D_1} control chart and smallest for $M_{P_{37}}$ control chart.

5. Conclusions

In this section, we record our conclusions about the proposed class of M_r control charts based on our findings.

- A class of control charts based on r^{th} sample midrange is proposed for location parameter when process is under exponential model.
- The proposed class of control charts includes midrange, mid quantile, middecile, midpercentile, midhinge and median control charts as its members.
- M_r is biased estimator. Hence, after adjusting the bias, it is renamed as M_r^*.
- When bias of M_h and M_d estimators are adjusted, the M_h^* and M_d^* control charts are the same.
- The power of the proposed class increases for smaller shifts as the sample size increases, exhibiting greater ability to detect shifts in process location.
- ARL and MRL decrease as sample size increases indicating improved performance of the control charts.
- As sample size increases, the performance of the control charts stabilizes.
- Among the various members of the proposed class, the control chart based on 37$^{\text{th}}$ percentile, $M_{P_{37}}^*$ outperforms other control charts establishing its optimality.

Appendix

Table 1: *$P_{M_r^*}$ and $ARL_{M_r^*}$ of various control charts*

Statistic	n	5	10	15	20	5	10	15	20
	a	$\mathbf{P_{M_{r^*}}}$				$\mathbf{ARL_{M_{r^*}}}$			
M_{D_1}	0.00	0.0027	0.0027	0.0027	0.0027	370.3983	370.3983	370.3983	370.3983
	0.25	0.0046	0.0067	0.0091	0.0118	217.3808	148.3252	109.5630	85.0522
	0.50	0.0118	0.0247	0.0416	0.0623	85.0522	40.4195	24.0249	16.0627
	0.75	0.0286	0.0739	0.1360	0.2106	34.9655	13.5312	7.3520	4.7475
	1.00	0.0623	0.1762	0.3211	0.4712	16.0627	5.6741	3.1139	2.1223
	1.50	0.2106	0.5419	0.7891	0.9180	4.7475	1.8452	1.2673	1.0894
	2.00	0.4712	0.8729	0.9808	0.9979	2.1223	1.1455	1.0196	1.0022
M_{D_2}	0.00	0.0027	0.0027	0.0027	0.0027	370.3983	370.3983	370.3983	370.3983
	0.25	0.0067	0.0116	0.0174	0.0242	150.0911	86.4698	57.4181	41.2916
	0.50	0.0242	0.0607	0.1107	0.1716	41.2916	16.4773	9.0355	5.8291
	0.75	0.0720	0.2051	0.3690	0.5311	13.8877	4.8759	2.7098	1.8830
	1.00	0.1716	0.4609	0.7102	0.8652	5.8291	2.1695	1.4080	1.1558
	1.50	0.5311	0.9119	0.9901	0.9992	1.8830	1.0966	1.0100	1.0008
	2.00	0.8652	0.9975	1.0000	1.0000	1.1558	1.0025	1.0000	1.0000
M_h	0.00	0.0027	0.0027	0.0027	0.0027	370.3983	370.3983	370.3983	370.3983
	0.25	0.0075	0.0136	0.0211	0.0299	133.1594	73.2735	47.3362	33.4008
	0.50	0.0299	0.0780	0.1438	0.2225	33.4008	12.8251	6.9553	4.4953
	0.75	0.0929	0.2649	0.4621	0.6384	10.7611	3.7749	2.1643	1.5665
	1.00	0.2225	0.5645	0.8087	0.9295	4.4953	1.7716	1.2366	1.0758
	1.50	0.6384	0.9594	0.9975	0.9999	1.5665	1.0424	1.0025	1.0001
	2.00	0.9295	0.9996	1.0000	1.0000	1.0758	1.0004	1.0000	1.0000
M_{D_3}	0.00	0.0027	0.0027	0.0027	0.0027	370.3983	370.3983	370.3983	370.3983
	0.25	0.0081	0.0151	0.0238	0.0340	123.5647	66.1805	42.0789	29.3723
	0.50	0.0340	0.0905	0.1675	0.2581	29.3723	11.0505	5.9706	3.8744
	0.75	0.1081	0.3062	0.5214	0.7006	9.2538	3.2656	1.9177	1.4273
	1.00	0.2581	0.6272	0.8581	0.9556	3.8744	1.5943	1.1654	1.0465

	1.50	0.7006	0.9765	0.9991	1.0000	1.4273	1.0240	1.0009	1.0000
	2.00	0.9556	0.9999	1.0000	1.0000	1.0465	1.0001	1.0000	1.0000
	0.00	0.0027	0.0027	0.0027	0.0027	370.3983	370.3983	370.3983	370.3983
	0.25	0.0083	0.0156	0.0247	0.0356	120.4118	63.9092	40.4202	28.1145
	0.50	0.0356	0.0952	0.1762	0.2711	28.1145	10.5096	5.6742	3.6889
$\mathbf{M_{D_4}}$	0.75	0.1137	0.3211	0.5419	0.7210	8.7961	3.1140	1.8453	1.3870
	1.00	0.2711	0.6483	0.8729	0.9625	3.6889	1.5426	1.1456	1.0389
	1.50	0.7210	0.9808	0.9993	1.0000	1.3870	1.0196	1.0007	1.0000
	2.00	0.9625	0.9999	1.0000	1.0000	1.0389	1.0001	1.0000	1.0000

Table 2: *$MRL_{M_r^*}$ and $SDRL_{M_r^*}$ of various control charts*

Statistic	**n**	**5**	**10**	**15**	**20**	**5**	**10**	**15**	**20**
	a	**$\mathbf{MRL_{M_{r^*}}}$**				**$\mathbf{SDRL_{M_{r^*}}}$**			
	0.00	256.3938	256.3938	256.3938	256.3938	369.8980	369.8980	369.8980	369.8980
	0.25	150.3300	102.4642	75.5962	58.6064	216.8802	147.8243	109.0618	84.5507
	0.50	58.6064	27.6686	16.3037	10.7835	84.5507	39.9164	23.5195	15.5546
M_{D_1}	0.75	23.8880	9.0281	4.7410	2.9305	34.4619	13.0216	6.8337	4.2179
	1.00	10.7835	3.5752	1.7895	1.0879	15.5546	5.1498	2.5657	1.5433
	1.50	2.9305	0.8878	0.4454	0.2772	4.2179	1.2489	0.5820	0.3120
	2.00	1.0879	0.3360	0.1753	0.1128	1.5433	0.4083	0.1412	0.0465
	0.00	256.3938	256.3938	256.3938	256.3938	369.8980	369.8980	369.8980	369.8980
	0.25	103.6882	59.5891	39.4516	28.2732	149.5902	85.9683	56.9159	40.7885
	0.50	28.2732	11.0710	5.9096	3.6830	40.7885	15.9695	8.5208	5.3056
M_{D_2}	0.75	9.2753	3.0199	1.5052	0.9153	13.3784	4.3472	2.1525	1.2895
	1.00	3.6830	1.1218	0.5596	0.3459	5.3056	1.5929	0.7579	0.4244
	1.50	0.9153	0.2853	0.1501	0.0972	1.2895	0.3254	0.1004	0.0283
	2.00	0.3459	0.1159	0.0640	0.0429	0.4244	0.0504	0.0045	0.0003
	0.00	256.3938	256.3938	256.3938	256.3938	369.8980	369.8980	369.8980	369.8980
	0.25	91.9521	50.4419	32.4632	22.8033	132.6585	72.7718	46.8335	32.8970
	0.50	22.8033	8.5384	4.4655	2.7548	32.8970	12.3150	6.4359	3.9639
M_{lh}	0.75	7.1068	2.2523	1.1180	0.6815	10.2489	3.2365	1.5874	0.9420
	1.00	2.7548	0.8340	0.4191	0.2613	3.9639	1.1692	0.5409	0.2856
	1.50	0.6815	0.2164	0.1155	0.0756	0.9420	0.2101	0.0499	0.0102
	2.00	0.2613	0.0898	0.0503	0.0340	0.2856	0.0211	0.0010	0.0000
	0.00	256.3938	256.3938	256.3938	256.3938	369.8980	369.8980	369.8980	369.8980
	0.25	85.3015	45.5254	28.8189	20.0108	123.0637	65.6786	41.5759	28.8680
	0.50	20.0108	7.3076	3.7814	2.3217	28.8680	10.5387	5.4477	3.3371
M_{D_3}	0.75	6.0611	1.8959	0.9405	0.5747	8.7396	2.7201	1.3267	0.7809
	1.00	2.3217	0.7024	0.3550	0.2226	3.3371	0.9734	0.4390	0.2205
	1.50	0.5747	0.1847	0.0995	0.0655	0.7809	0.1569	0.0307	0.0050
	2.00	0.2226	0.0776	0.0438	0.0297	0.2205	0.0115	0.0004	0.0000
	0.00	256.3938	256.3938	256.3938	256.3938	369.8980	369.8980	369.8980	369.8980
	0.25	83.1160	43.9510	27.6691	19.1388	119.9107	63.4073	39.9171	27.6100
	0.50	19.1388	6.9323	3.5753	2.1922	27.6100	9.9971	5.1500	3.1495
$\mathbf{M_{D_4}}$	0.75	5.7435	1.7896	0.8878	0.5431	8.2810	2.5657	1.2489	0.7327
	1.00	2.1922	0.6634	0.3360	0.2110	3.1495	0.9149	0.4083	0.2011
	1.50	0.5431	0.1753	0.0947	0.0625	0.7327	0.1412	0.0258	0.0039
	2.00	0.2110	0.0739	0.0419	0.0285	0.2011	0.0092	0.0003	0.0000

Table 3: *$MRL_{M_r^*}$ and $SDRL_{M_r^*}$ of various control charts*

n / a	5	10	15	20	5	10	15	20
	$\mathbf{P_{M_{r^*}}}$				$\mathbf{ARL_{M_{r^*}}}$			
0.00	0.0030	0.0030	0.0030	0.0030	333.3333	333.3333	333.3333	333.3333
0.25	0.0040	0.0040	0.0050	0.0040	250.0000	250.0000	200.0000	250.0000
0.50	0.0060	0.0050	0.0070	0.0080	166.6667	200.0000	142.8571	125.0000
0.75	0.0080	0.0060	0.0080	0.0090	125.0000	166.6667	125.0000	111.1111
1.00	0.0130	0.0080	0.0120	0.0140	76.9231	125.0000	83.3333	71.4286
1.50	0.0280	0.0220	0.0270	0.0380	35.7143	45.4545	37.0370	26.3158
2.00	0.0570	0.0550	0.0560	0.0710	17.5439	18.1818	17.8571	14.0845
	$\mathbf{MRL_{M_{1^*}}}$				$\mathbf{SDRL_{M_{1^*}}}$			
0.00	230.7023	230.7023	230.7023	230.7023	332.8330	332.8330	332.8330	332.8330
0.25	172.9400	172.9400	138.2826	172.9400	249.4995	249.4995	199.4994	249.4995
0.50	115.1776	138.2826	98.6741	86.2964	166.1659	199.4994	142.3563	124.4990
0.75	86.2964	115.1776	86.2964	76.6693	124.4990	166.1659	124.4990	110.6100
1.00	52.9717	86.2964	57.4150	49.1631	76.4214	124.4990	82.8318	70.9268
1.50	24.4070	31.1588	25.3240	17.8919	35.2107	44.9518	36.5336	25.8109
2.00	11.8105	12.2528	12.0277	9.4118	17.0365	17.6747	17.3499	13.5753

Acknowledgement

The second author acknowledges Department of Backward Classes Welfare for the award of Ph.D Fellowship.

References

[1] Ahsanullah, M. and Hamedani, G. Exponential Distribution: Theory and Methods. Nova Science Publishers, (2010).
[2] Amin, R. W., Reynolds, M. R. Jr and Saad, B. (1995). Nonparametric Quality Control Charts Based on the Sign Statistic. *Communications in Statistics-Theory and Methods*, 24(6): 1597–1623.
[3] Arif, O. H. Statistical Process Control by Quantile Approach. Sheffield Hallam University, (2000).
[4] Bakir, S. T. (2004). A Distribution-Free Shewhart Quality Control Chart Based on Signed-Ranks. *Quality Engineering*, 16(4): 613–623.
[5] Bhat, S. V. and Patil, B. A. (2009). A New Approach to the Notion of Sample Quantiles. *International Journal of Agricultural and Statistical Sciences*, 5(2): 335–340.
[6] Bhat, S. V. and Patil, S. S. (2024). Shewhart Type Control Chart Based on Median for Location Parameter. *International Journal of Agricultural & Statistical Sciences*, 20(1).
[7] Chakraborti, S. and Graham, M. A. (2019). Nonparametric (Distribution-Free) Control Charts: An Updated Overview and Some Results. *Quality Engineering*, 31(4): 523–544.
[8] Chakraborti, S. and Graham, M. Nonparametric Statistical Process Control. John Wiley & Sons, (2019).
[9] Chen, G. and Cheng, S. W. (1998). Max Chart: Combining X-bar Chart and S Chart. *Statistica Sinica*, pp. 263–271.
[10] Cramer, H. Mathematical Methods of Statistics. Princeton University Press, 43, (1946).
[11] Duncan, A. J. (1956). The Economic Design of X Charts Used to Maintain Current Control of a Process. *Journal of the American Statistical Association*, 51(274): 228–242.
[12] Gibbons, J. and Chakraborti, S. (1992). Nonparametric Statistical Inference. Marcel Dekker Inc, New York.
[13] Montgomery, D. C. Introduction to Statistical Quality Control. John Wiley & Sons, (1996).

[14] Mukherjee, A., McCracken, A. K. and Chakraborti, S. (2015). Control Charts for Simultaneous Monitoring of Parameters of a Shifted Exponential Distribution. *Journal of Quality Technology*, 47(2): 176–192.
[15] Page, E. S. (1961). Cumulative Sum Charts. *Technometrics*, 3(1): 1–9.
[16] Pawar, V. Y. and Shirke, D. T. (2010). A Nonparametric Shewhart-Type Synthetic Control Chart. *Communications in Statistics-Simulation and Computation*, 39(8): 1493–1505.
[17] Pawar, V. Y., Shirke, D. T. and Khilare, S. K. (2018). Nonparametric Moving Average Control Charts Using Sign and Signed-Rank Statistics. *Int. J. Sci. Res. in Mathematical and Statistical Sciences*, 5: 4.
[18] Roberts, S. W. (2000). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 42(1): 97–101.
[19] Santiago, E. and Smith, J. (2013). Control Charts Based on the Exponential Distribution: Adapting Runs Rules for the T Chart. *Quality Engineering*, 25(2): 85–96.
[20] Sen, P. K. (1961). On Some Properties of the Asymptotic Variance of the Sample Quantiles and Mid-Ranges. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 23(2): 453–459.
[21] Shewhart, W. A. (1931). Statistical Method from an Engineering Viewpoint. *Journal of the American Statistical Association*, 26(175): 262–269.
[22] Wu, Z. and Spedding, T. A. (2000). A Synthetic Control Chart for Detecting Small Shifts in the Process Mean. *Journal of Quality Technology*, 32(1): 32–38.
[23] Zombade, D. M. and Ghute, V. B. (2019). Shewhart-Type Nonparametric Control Chart for Process Location. *Communications in Statistics-Theory and Methods*, 48(7): 1621–1634.

RT&A, No 4(80)
Volume 19, December, 2024

GENERALIZATION OF RAYLEIGH DISTRIBUTION THROUGH A NEW TRANSMUTATION TECHNIQUE

Aliya Syed Malik And S.P. Ahmad

•

Department of Statistics, University of Kashmir, Srinagar, India
aaliyasayeed2@gmail.com, sprvz@yahoo.com

Abstract

In our research paper, we introduce an innovative statistical distribution known as the New Transmuted Rayleigh Distribution. This distribution serves as a versatile expansion of the traditional Rayleigh distribution and has been developed using a novel transmutation technique. We provide an in-depth analysis of several statistical properties of this new distribution. The resulting model has the ability to represent complex shapes, making it suitable for a wide range of applications. Our manuscript thoroughly examines the fundamental characteristics of the new model, outlining the methodology for estimating its unknown parameters through maximum likelihood estimation. Additionally, we demonstrate the practical significance of the model by applying it to an empirical dataset and conclusively establishing its superiority over some existing prominent models.

Keywords: New Transmuted Rayleigh distribution, Rayleigh distribution, reliability, stress strength reliability, parameter estimation.

1. Introduction

Innumerable methods to extend the probability models are available in the literature. A prominent technique called Quadratic Rank Transmutation map was put forth by Shaw and Buckley [11] that has been in use for a very long time. It enables us to obtain extended models with improved flexibility. Lately, some new ways to obtain the transmuted version of distribution are being introduced by various researchers. The introduction of these methods is triggered by the inability of the conventional method to model data that arise in different fields that are of random nature.

A novel transmutation technique introduced by Mansour et.al . [5] to derive an extended version of a probability model. Let the cdf of baseline distribution be the $G(x)$, then the new cdf, $M(x)$ is defined as

$$M(x) = (1+\lambda)G^{\delta}(x) - \lambda G^{\alpha}(x) \quad ; x > 0, \tag{1}$$

where $\alpha, \delta > 0$ if $0 > \lambda > -1$ and $\alpha > 0, \frac{\alpha}{2} \leq \delta \leq \alpha + \frac{\alpha}{2}$ if $0 < \lambda < 1$.
The pdf associated with equation (1) is presented in (2).

$$m(x) = [\delta(1+\lambda)G^{\delta-1}(x) - \alpha\lambda G^{\alpha-1}(x)]g(x). \tag{2}$$

Rayleigh Distribution (RD) was propounded by Rayleigh [7]. Balakrishnan [1] approximated the likelihood function to derive an exact estimator to estimate the scale parameter of RD. Sarti et.al. [10] studied maximum likelihood segmentation of ultrasound images with RD. The

distribution followed by the product of independent Rayleigh variates was studied by EL-Sallabi and Vainikainen [9]. The cdf and pdf are respectively given by equations (3) and (4).

$$G(x) = 1 - e^{-\frac{x^2}{2\theta^2}} \quad ; x, \theta > 0, \tag{3}$$

$$g(x) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}. \tag{4}$$

The failure rate of RD accelerates with time. Hence can be used in situations where the components have no manufacturing defect but intense aging. However, this distribution fails to provide an acceptable fit when the failure rate takes complex shapes which prove to be a motivation for extending this distribution. Some researchers who have obtained flexible extensions of RD are : Merovci [6], Reshi et.al. [8], Malik and Ahmad [4], Bhat and Ahmad [2] etc. In this paper, we exploit the generalization suggested by Mansour et.al. [5] to define a new generalization of RD. The new model so obtained is named as New Transmuted Rayleigh Distribution. The main motivation for considering this generalization is given as the hazard rate of NTRD exhibits several complex shapes such as constant, increasing-decreasing, decreasing-increasing, etc. thus, overcoming the shortcomings of RD. Also, the new distribution outperforms the base distribution and some well-known models in view of a real-life data set. The remaining manuscript is arranged as: in section 2, the pdf and cdf of the NTRD are defined and its special cases are discussed. The alternative form of the pdf and cdf of NTRD are given in section 3 and the prominent properties of the NTRD are expounded in section 4. In section 5, the parameter estimation is executed utilizing a very powerful method. Finally, the applicability of NTRD and its conclusion are given in sections 6 and 7 respectively.

2. New Transmuted Rayleigh Distribution (NTRD)

The cdf of NTRD can be obtained upon substituting equation (3) in equation (1) and is given as

$$M(x) = (1+\lambda)\left[1 - e^{-\frac{x^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1 - e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha} \quad ; x > 0. \tag{5}$$

Also, the pdf corresponding to equation (5) is presented below:

$$m(x) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} \left(\delta(1+\lambda)\left[1 - e^{-\frac{x^2}{2\theta^2}}\right]^{\delta-1} - \alpha\lambda\left[1 - e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha-1}\right), \tag{6}$$

where $\alpha, \delta, \theta > 0$ if $0 > \lambda > -1$ and $\alpha, \theta > 0$, $\frac{\alpha}{2} \leq \delta \leq \alpha + \frac{\alpha}{2}$ if $0 < \lambda < 1$.

2.1. Special cases of NTRD

The special cases of NTRD are as follows:

- RD: For $\lambda = 0$ and $\delta = 1$, equation (5) reduces to equation (3), which is the cdf of RD.
- Transmuted Rayleigh Distribution (TRD): For $\alpha = 2$ and $\delta = 1$, equation (5) becomes equation (7)

$$M(x) = (1+\lambda)\left[1 - e^{-\frac{x^2}{2\theta^2}}\right] - \lambda\left[1 - e^{-\frac{x^2}{2\theta^2}}\right]^2 \quad ; \theta, x > 0, -1 \leq \lambda \leq 1, \tag{7}$$

  which is the cdf of TRD.

- Transmuted Exponentiated Rayleigh Distribution (TERD): For $\delta = \frac{\alpha}{2}$, equation (5) becomes equation (8)

$$M(x) = (1+\lambda)\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\frac{\alpha}{2}} - \lambda\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha} \quad ;\theta,\alpha,x>0,-1\leq\lambda\leq 1, \tag{8}$$

  which is the cdf of TERD

- Exponentiated Rayleigh Distribution (ERD): For $\lambda = 0$, equation (5) becomes equation (9)

$$M(x) = \left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\delta} \quad ;\theta,\delta,x>0, \tag{9}$$

  which is the cdf of ERD.

The graphical overview of possible shapes of the pdf of NTRD is shown in Figure 1. It is clear from the Figure 1 that PDF plots of NTRD are unimodal, Decreasing, Increasing, Symmetric and postively skewed

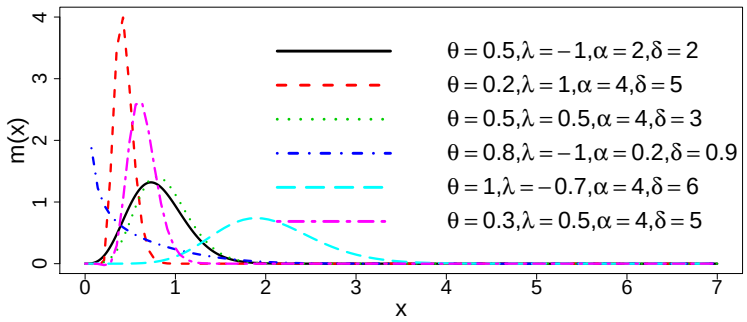


Figure 1: *Pdf plots of NTRD*

The survival function $(S(x))$, hazard rate function $(h(x))$(hrf) and cummulative hrf $(H(x))$ are respectively given as:

$$S(x) = 1-\left((1+\lambda)\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha}\right),$$

$$h(x) = \frac{\frac{x}{\theta^2}e^{-\frac{x^2}{2\theta^2}}\left(\delta(1+\lambda)\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\delta-1} - \alpha\lambda\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha-1}\right)}{1-\left((1+\lambda)\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha}\right)},$$

$$H(x) = -ln\left[1-\left((1+\lambda)\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{x^2}{2\theta^2}}\right]^{\alpha}\right)\right].$$

The shapes of hrf for different parameter values are displayed in Figure 2. Figure 2 clearly shows that the proposed model is flexible and can exhibit different types of shapes such as U-Shaped, J shaped, Increasing and Bathtub over the parameter space.

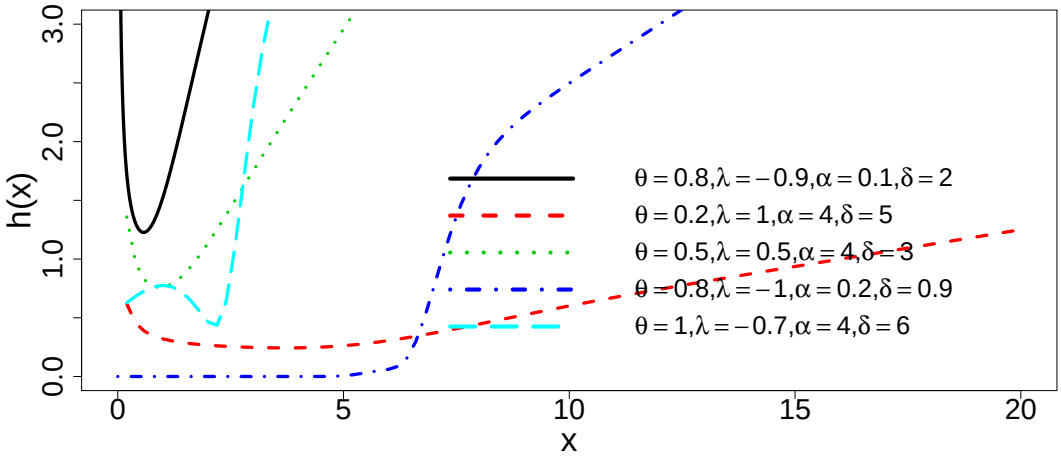


Figure 2: *Hrf plots of NTRD*

3. Mixture representation of NTRD

Using the expansions

$$(a-b)^n = \sum_{j=0}^{n} \binom{n}{j} (-1)^j b^j a^{n-j},$$

in equation (6), we get

$$m(x) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} \sum_{j=0}^{1} (-1)^j \{\delta(1+\lambda)\}^{1-j} \{\lambda\alpha\}^j \left(1 - e^{-\frac{x^2}{2\theta^2}}\right)^{(\delta-1)(1-j)+(\alpha-1)j}. \tag{10}$$

Now, using

$$(1-y)^k = \sum_{m=0}^{\infty} \frac{(-1)^m \Gamma(k+1)}{\Gamma(k-m+1)m!} y^m \quad ; \quad |y| < 1, k > 0,$$

in equation (10), we obtain the alternative form of pdf given as

$$\begin{aligned} m(x) &= \sum_{j=0}^{1} \sum_{m=0}^{\infty} \binom{(\delta-1)(1-j)+(\alpha-1)j}{m} (-1)^{j+m} \{\delta(1+\lambda)\}^{1-j} \\ &\qquad \{\lambda\alpha\}^j \frac{x}{\theta^2} e^{-(m+1)\frac{x^2}{2\theta^2}}, \\ &= \sum_{m=0}^{\infty} \eta(j) \frac{x}{\theta^2} e^{-(m+1)\frac{x^2}{2\theta^2}}, \end{aligned} \tag{11}$$

where

$$\eta(j) = \sum_{j=0}^{1} \binom{(\delta-1)(1-j)+(\alpha-1)j}{m} (-1)^{j+m} \{\delta(1+\lambda)\}^{1-j} \{\lambda\alpha\}^j.$$

Similarly, the cdf of NTRD takes the alternative expression given in equation (12).

$$M(x) = \sum_{s=0}^{\infty} \eta'(p) e^{-s\frac{x^2}{2\theta^2}}, \tag{12}$$

where $\eta'(p) = \sum_{p=0}^{1} (-1)^{p+s} \lambda^p (1+\lambda)^{1-p} \binom{\lambda p+\delta(1-p)}{s}$.

4. Statistical Properties of NTRD

This section brings forth several prominent belonging such as moments, Generating functions, mean deviation about mean and median, mean residual life and mean waiting time and Stress Strength Reliability to NTRD .

4.1. Moments of NTRD

The r^{th} moment about origin can be expounded as

$$\mu_r^{'} = \int_0^{\infty} x^r m(x) dx.$$

Using equation (11), we get

$$= \sum_{m=0}^{\infty} \eta(j) \int_0^{\infty} \frac{x^{r+1}}{\theta^2} e^{-(m+1)\frac{x^2}{2\theta^2}} dx,$$

which upon simplification yields equation (13).

$$\mu_r^{'} = \sum_{m=0}^{\infty} \eta(j) \frac{(2\theta^2)^{\frac{r}{2}} \Gamma\left(\frac{r}{2}+1\right)}{(m+1)^{\left(\frac{r}{2}+1\right)}}. \tag{13}$$

The mean and variance are respectively given by equations (14) and (15) respectively.

$$Mean = \sum_{m=0}^{\infty} \eta(j) \frac{(2\theta^2)^{\frac{1}{2}} \Gamma\left(\frac{3}{2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}, \tag{14}$$

$$Variance = \left(\sum_{m=0}^{\infty} \eta(j) \frac{2\theta^2}{(m+1)^2} \right) - \left(\sum_{m=0}^{\infty} \eta(j) \frac{(2\theta^2)^{\frac{1}{2}} \Gamma\left(\frac{3}{2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}} \right)^2. \tag{15}$$

Also, for NTRD, the n^{th} incomplete moment about origin is

$$\begin{aligned} \psi_n(s) &= \int_0^{s} x^n m(x) dx, \\ &= \sum_{m=0}^{\infty} \eta(j) \int_0^{s} \frac{x^{n+1}}{\theta^2} e^{-(m+1)\frac{x^2}{2\theta^2}} dx, \\ &= \sum_{m=0}^{\infty} \eta(j) \frac{(2\theta^2)^{\frac{n}{2}} \gamma\left(\frac{n}{2}+1, \frac{(m+1)s^2}{2\theta^2}\right)}{(m+1)^{\left(\frac{n}{2}+1\right)}}. \end{aligned} \tag{16}$$

4.2. Generating functions of NTRD

The moment, characteristic and cummulant generating function are respectively given by equations (17), (18) and (19) respectively.

$$M_X(t) = \sum_{r=0}^{\infty} \sum_{m=0}^{\infty} \frac{t^r \eta(j)}{r!} \frac{(2\theta^2)^{\frac{r}{2}} \Gamma\left(\frac{r}{2}+1\right)}{(m+1)^{\left(\frac{r}{2}+1\right)}}, \tag{17}$$

$$\phi_X(t) = \sum_{r=0}^{\infty} \sum_{m=0}^{\infty} \frac{(it)^r \eta(j)}{(i)^r r!} \frac{(2\theta^2)^{\frac{r}{2}} \Gamma\left(\frac{r}{2}+1\right)}{(m+1)^{\left(\frac{r}{2}+1\right)}}, \tag{18}$$

$$K_X(t) = ln \left[\sum_{r=0}^{\infty} \sum_{m=0}^{\infty} \frac{t^r \eta(j)}{r!} \frac{(2\theta^2)^{\frac{r}{2}} \Gamma\left(\frac{r}{2}+1\right)}{(m+1)^{\left(\frac{r}{2}+1\right)}} \right]. \tag{19}$$

4.3. Mean Deviation about Mean and Median of NTRD

The mean Deviation about mean of NTRD is discussed below

$$\begin{aligned} D(\mu) =& E[|X-\mu|], \\ =& 2\mu M(\mu) - 2\psi_1(\mu), \\ =& 2\mu\left((1+\lambda)\left[1-e^{-\frac{\mu^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{\mu^2}{2\theta^2}}\right]^{\alpha}\right) - \\ & 2\left(\sum_{m=0}^{\infty}\eta(j)\frac{(2\theta^2)^{\frac{1}{2}}\gamma\left(\frac{3}{2},\frac{(m+1)\mu^2}{2\theta^2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}\right). \end{aligned}$$

Also, for NTRD the mean deviation about Median is

$$\begin{aligned} D(M) =& E[|X-M|], \\ =& \mu - 2\psi_1(M), \\ =& \mu - 2\left(\sum_{m=0}^{\infty}\eta(j)\frac{(2\theta^2)^{\frac{1}{2}}\gamma\left(\frac{3}{2},\frac{(m+1)M^2}{2\theta^2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}\right). \end{aligned}$$

4.4. Mean Residual Life (MRL) and Mean Waiting Time (MWT) of NTRD

The formula for computing MRL is

$$MRL = \frac{E(t)-\psi_1(t)}{1-M(t)} - t. \tag{20}$$

Using equation (16) for $n=1$, equation (14) and equation (5) in equation (20), we get

$$\begin{aligned} MRL =& \frac{\left(\sum_{m=0}^{\infty}\eta(j)\frac{(2\theta^2)^{\frac{1}{2}}\Gamma\left(\frac{3}{2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}\right) - \left(\sum_{m=0}^{\infty}\eta(j)\frac{(2\theta^2)^{\frac{1}{2}}\gamma\left(\frac{3}{2},\frac{(m+1)t^2}{2\theta^2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}\right)}{1-\left((1+\lambda)\left[1-e^{-\frac{t^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{t^2}{2\theta^2}}\right]^{\alpha}\right)} - t, \\ =& \frac{\left(\sum_{m=0}^{\infty}\eta(j)\frac{(2\theta^2)^{\frac{1}{2}}\Gamma\left(\frac{3}{2},\frac{(m+1)t^2}{2\theta^2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}\right)}{1-\left((1+\lambda)\left[1-e^{-\frac{t^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{t^2}{2\theta^2}}\right]^{\alpha}\right)} - t, \end{aligned}$$

where $\Gamma(a,b) = \int_0^b t^{a-1}e^{-t}dt$.

Similarly, we can obtain MWT for NTRD using the formula given as

$$\begin{aligned} MWT =& t - \frac{\psi_1(t)}{M(t)}, \\ =& t - \frac{\left(\sum_{m=0}^{\infty}\eta(j)\frac{(2\theta^2)^{\frac{1}{2}}\gamma\left(\frac{3}{2},\frac{(m+1)t^2}{2\theta^2}\right)}{(m+1)^{\left(\frac{3}{2}\right)}}\right)}{\left((1+\lambda)\left[1-e^{-\frac{t^2}{2\theta^2}}\right]^{\delta} - \lambda\left[1-e^{-\frac{t^2}{2\theta^2}}\right]^{\alpha}\right)}. \end{aligned}$$

4.5. Stress Strength Reliability(SSR) of NTRD

Let $X_1 \sim NTRD(\alpha, \lambda, \delta, \theta_1)$ be the strength of a system and $X_2 \sim NTRD(\alpha, \lambda, \delta, \theta_2)$ denote the stress that the system is subjected to. Then, the SSR can be computed as

$$\begin{aligned} R &= \int_0^\infty M_1(x) m_2(x) dx, \\ &= \sum_{m=0}^{\infty} \sum_{s=0}^{\infty} \eta^{'}(p)\eta(j) \int_0^\infty \frac{x}{\theta_1^2} e^{-(m+1)\frac{x^2}{2\theta_1^2} - s\frac{x^2}{2\theta_2^2}} dx, \\ &= \sum_{m=0}^{\infty} \sum_{s=0}^{\infty} \frac{\eta^{'}(p)\eta(j)}{\theta_1^2 \Gamma\left(\frac{m+1}{\theta_1^2} + \frac{s}{\theta_2^2}\right)}. \end{aligned}$$

5. Estimation of Parameters

The technique of MLE is employed to estimate the parameters are not known. For a sample $X_1, X_2, ..., X_n$ taken from NTRD randomly, the log-likelihood function denoted by l can be obtained as

$$\begin{aligned} l = \sum_{i=1}^{n} log \left(\delta(1+\lambda) \left[1 - e^{-\frac{x_i^2}{2\theta^2}}\right]^{\delta-1} - \alpha\lambda \left[1 - e^{-\frac{x_i^2}{2\theta^2}}\right]^{\alpha-1} \right) + \\ \sum_{i=1}^{n} log x_i - 2n log\theta - \sum_{n=1}^{n} \frac{x_i}{2\theta^2}. \end{aligned} \tag{21}$$

Equation (21) is differentiated w.r.t α, λ, δ and θ respectively and equated to zero to obtain the normal equations. These normal equations will not be in closed form. Therefore, a number of iterative procedures are available in literature which can be followed to solve such equations and obtain the parameter estimates.

6. Application

The applicability of NTRD from practical standpoint is assessed through a real-life data set which consists of the waiting period (in minutes) of 100 bank clients prior service which has been extracted from Ghitany et.al. [3]. The conventional models that are being used to compare the fits are

- TRD with cdf given by equation (7).
- ERD with cdf given by equation (9).
- Exponential Distribution (ED) with pdf given as

$$m(x) = \frac{e^{-\frac{x}{\theta}}}{\theta} \qquad ; \qquad x, \theta > 0.$$

- RD with pdf given in equation (4).

To choose the best model among the compared models, performance comparing tools such as AIC, AICc, SIC, and HQIC are exploited. These Criterion choose the superior distribution as the one which gives the minimal value. Furthermore, the Kolmogorov-Smirnov (KS)-Statistic (which is used to test if the sample comes from a specific population) and associated p- value (for decision making regarding the hypothesis) is obtained to assess the goodness of fit. All the

calculations are performed using R software. The formulas for computing these criterion are given as:

$$AIC = -2\hat{l} + 2w, \qquad AICc = AIC + \frac{2w(w+1)}{n-w-1},$$
$$SIC = -2\hat{l} + logn, \qquad HQIC = -2\hat{l} + 2wlog(logn),$$
$$KS - statistic = Maximum|M_0(x) - M_n(x)|.$$

where c represents number of parameters, $M_0(x)$ is the observed cdf and $M_n(x)$ is the theoretical cdf.

Table 1 displays the MLE's and corresponding Standard Error (SE) and Table 2 presents comparison of performance of NTRD and compared distributions for waiting time data.

Table 1: *MLE's of NTRD and compared distributions with corresponding standard error (given in parenthesis) for waiting time data*

Model	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\delta}$
NTRD	12.089	1.000	1.269	0.754
	(1.675)	(0.990)	(0.340)	(0.119)
TRD	10.127	0.647		
	(0.778)	(0.172)		
ERD	8.185			0.629
	(0.774)			(0.0777)
RD	8.643			
	(0.432)			
ED	9.877			
	(0.987)			

Table 2: *Comparison of NTRD and compared distributions for waiting time data*

Model	-$\hat{l}$	AIC	SIC	AICc	HQIC	KS-statistic	p-value
NTRD	318.71	645.43	655.85	645.86	649.65	0.0655	0.7831
TRD	323.83	651.66	656.87	651.79	653.77	0.1317	0.0621
ERD	321.51	647.03	658.24	647.45	658.25	0.0945	0.3337
RD	329.24	660.48	663.08	660.52	661.53	0.1736	0.0048
ED	329.02	660.04	662.64	660.08	661.09	0.1730	0.0050

The values reported in Table 2 reveal that NTRD has a minimum value of SIC, AIC, AICc, and HQIC, consequently outperforms the base model and some well-known models for the given data set. Figure 3 further justifies the claim. Also, the QQ-plots are plotted for the proposed and compared distributions for the waiting data in Figure 4.

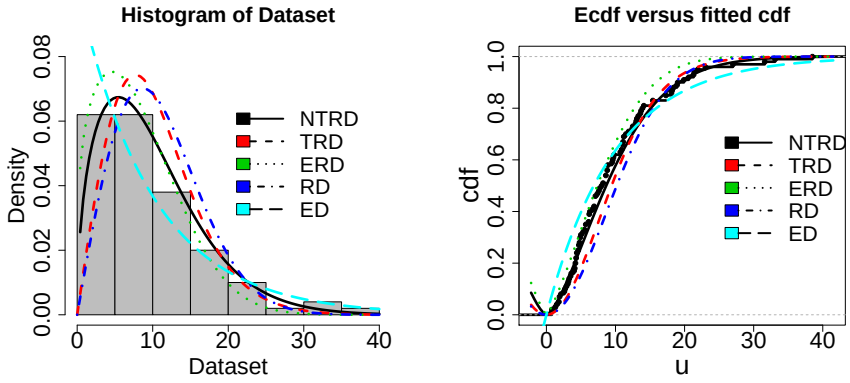
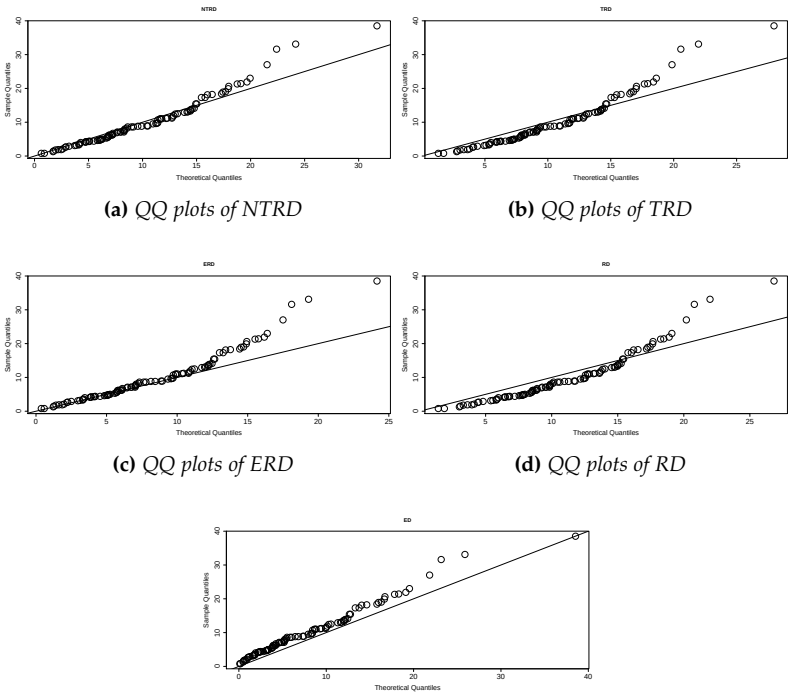


Figure 3: *Fitted densities*



(a) *QQ plots of NTRD* **(b)** *QQ plots of TRD*

(c) *QQ plots of ERD* **(d)** *QQ plots of RD*

(e) *QQ plots of ED*

Figure 4: *QQ-plot of NTRD and competitive models for given dataset*

7. Conclusion

In this manuscript, a flexible generalization of RD is proposed which can acts as a potential substitute for the base model in various situations. Some of its important properties are expounded and parameters are estimated utilizing a very powerful estimation procedure. A data set is incorporated for the illustration of the applicability of NTRD from a practical standpoint and the results prove the superiority of NTRD over various standard Distributions namely TRD, ERD, RD and ED. Additionally, we present graphs to visually illustrate the results.

References

[1] N. Balakrishnan. Approximate mle of the scale parameter of the rayleigh distribution with censoring. *IEEE Transactions on Reliability*, 38(3):355–357, 1989.

[2] A. Bhat and S. P. Ahmad. A new generalization of rayleigh distribution: Properties and applications. *Pakistan journal of statistics*, 36(3), 2020.

[3] M. E. Ghitany, B. Atieh, and S. Nadarajah. Lindley distribution and its application. *Mathematics and computers in simulation*, 78(4):493–506, 2008.

[4] A. S. Malik and S. Ahmad. Alpha power rayleigh distribution and its application to life time data. *International Journal of Enhanced Research in Management & Computer Applications*, 6(11):212–219, 2017.

[5] M. M. Mansour, M. Enayat, S. Hamed, and M. Mohamed. A new transmuted additive weibull distribution based on a new method for adding a parameter to a family of distributions. *Int. J. Appl. Math. Sci*, 8:31–51, 2015.

[6] F. Merovci. Transmuted rayleigh distribution. *Austrian Journal of statistics*, 42(1):21–31, 2013.

[7] L. Rayleigh. Xii. on the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(60):73–78, 1880.

[8] J. Reshi, A. Ahmed, and K. Mir. Characterizations and estimations of size biased generalized rayleigh distribution. *Mathematical Theory and Modeling*, 4(6):87–101, 2014.

[9] J. Salo, H. M. El-Sallabi, and P. Vainikainen. The distribution of the product of independent rayleigh random variables. *IEEE transactions on Antennas and Propagation*, 54(2):639–643, 2006.

[10] A. Sarti, C. Corsi, E. Mazzini, and C. Lamberti. Maximum likelihood segmentation of ultrasound images with rayleigh distribution. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 52(6):947–960, 2005.

[11] W. T. Shaw and I. R. Buckley. The alchemy of probability distributions: beyond gram-charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv preprint arXiv:0901.0434*, 2009.

MEASUREMENT AND DETERMINATION OF STRENGTH OF LOAD-BEARING STRUCTURES MATERIALS BY SHEAR TEST METHOD

Alena Rotaru

All-Russian Scientific Research Institute for Civil Defence and Emergencies of the EMERCOM of Russia
alenarotaru@mail.ru

Abstract

The strength of load-bearing and enclosing structures largely depends on the parameters of their materials. Complex shear testing of concrete is a non-destructive method used to determine the parameters and quality of the mixtures used with high accuracy. This concrete testing method has become widespread due to its versatility and convenience. The material strength is tested by directly impacting the concrete of the structure and causing its partial shearing. During the test, the force needed to tear off a fragment of the structure using a leafed anchor embedded in the bore hole is determined. This method can provide more accurate data on the concrete strength to make a decision on the need for further operation of the building. The concrete test to be shear tested must be located at a sufficient distance from pre-stressed rods. In addition, the test area should not be subjected to heavy operational loads.

Keywords: concrete strength, shear test method, anchor, calibration curve, non-destructive testing, device.

I. Introduction

The shear test method is used to test the strength of concrete. The shear test method has advantages over non-destructive methods for determining the strength of concrete. Modern non-destructive methods use various indirect characteristics (ultrasound propagation rate, diameter of the imprint left on concrete upon impact, etc.) to determine the strength of concrete. The method is based on an empirical proportional relationship between the concrete strength and the pullout force of a special anchor with an expanding cone. The lack of a concrete destruction physical model during the shear test is challenging to find ways to improve the accuracy and reliability of the results. The purpose of the study is to develop a physical model of concrete destruction to determine the design strength by shear testing.

II. Methods

During shear testing, an anchor usually breaks out concrete in the form of a truncated cone, indicating a cleavage failure. This means that the concrete tensile strength is less than its shear strength. Otherwise, it would be a shear failure.

Leafed anchors, which can be of different sizes, are used to apply a pullout force. The anchors are inserted into drilled holes in the measurement area. As in the previous case, a device measures the fracture force. The compressive strength is calculated based on the relationship expressed by the formula

$$R=m1*m2*P$$

where:
m1 is the maximum size factor of coarse filler,
m2 is the conversion factor for compressive strength. It depends on the concrete type and strength gain conditions.
P is the fracture force obtained by testing. This method is one of the most popular ones as it is quite versatile. It ensures to test any part of the structure, since a flat surface is not a must. In addition, it is easy to fix a leafed anchor within the concrete mass manually. However, there are some limitations, such as:
• Dense structure reinforcement, in this case the measurements will be inaccurate.
• Structure thickness, it should be twice the length of the anchor. The ONIX-1.OS device is used for concrete shear testing, as shown in Pictures 1-2.

Picture 1: *Shear test method using ONIX-1.OS*

Picture 2: *Shear test method (crack formation)*

During shear testing, the anchor breaks out concrete, indicating a cleavage failure as shown in Pictures 3–4.

Picture 3: *Breaking out of concrete during a shear test*

Picture 4: *Breaking out of concrete during a shear test*

III. Results

Using the data recorded during the study, we can evaluate the strength of the above material based on the applied load at which shear occurs. The force breaking off a concrete fragment as a result of shearing is multiplied by a correction factor. It is calculated using the following formula:

$$\gamma=h^2/(h-\Delta h)^2$$

where h is the anchor depth,
Δh is the slip value.
If the maximum length of the material fragment broken off during the test is more than twice the minimum length, the result is deemed indicative. The same applies, if the hole depth exceeds the anchor slip value by 5% or more. Indicative values cannot be used to determine the strength class of a material. The tests are deemed invalid, if the tear depth differs from the anchor length by 10% or any reinforcement rod is found at a distance within the hole depth. During the works, the measurement results are recorded in a table, as shown in Table 1.

Table 1: *Measuring concrete strength by shear testing*

Structure	Test site No.	Shear force, MPa	Average strength, MPa
Reinforced concrete column No. 1	1	53.6	51.24
	2	52.1	
	3	50.6	
	4	48.2	
	5	48.9	
	6	49.9	
	7	51.1	
	8	52.9	
	9	53.9	
Reinforced concrete column No. 2	1	44.9	45.25
	2	43.4	
	3	45.6	
	4	46.7	
	5	44.8	
	6	46.7	

Structure	Test site No.	Shear force, MPa	Average strength, MPa
	7	45.8	
	8	44.4	
	9	45.0	

Strength testing of reinforced concrete columns in 9 sections yielded values from 45.25 MPa to 80 MPa. The strength of concrete becomes constant after achieving the concrete strength of about 50 MPa, as reported in various regulations.

IV. Conclusions

It was established that the empirical dependence between concrete compression resistance and anchor pullout force assumed for shear testing of concrete is only possible with a linear dependence between the concrete compressive and tensile strength. However, the actual relationship between the concrete compressive and tensile strength is not linear, so for relatively weak concretes, the potential compressive strength overestimation due to the empirical relationship is neutralized by a reduction factor. The more accurately tensile strength is converted to compressive strength, the more accurate the calculation results using the developed model and empirical pullout force values. One of the most accurate methods for testing the strength of structures. The shear test method allows to determine the forces needed for local concrete failure as an anchor element is pulled out from it. This method has disadvantages, such as high labor intensity and unsuitability for areas with dense reinforcement, as well as partially damaged structures.

References

[1] Kralik J., Kralik J., Probability Assessment of Analysis of High-Rise Buildings Seismic Resistance, Advances in manufacturing science and engineering, Advanced materials Research, 2013.

[2] CSN EN 1992-2 Eurocode 2 – Design of concrete structures, Concrete bridges, Design and detailing rules, 2005.

[3] Non-destructive testing: reference book: in 8 volumes. Vol. 3: Ultrasonic testing – Ed. 2, revised and amended. – Moscow, Mashinostroenie, 2008, 864.

[4] Szonova S.A., Nikolenko S.D., Zyazina T.V., Chernyshova E.V., and Kazbanova I.M., Srength test of the industrial building's load-bearing structures, Journal if Physics: Conderence Series, 2373, 2022.

[5] Litvinov A.E., Novikov V.V., Solod S.A. and Chukarin A.N., Russian Engineering Reseach, 39 (2), 2019, 158-9.

[6] MSL Engineering Limited, Doc ref. C237R001 rev. 1 Nov. 2004, Assessment of repair techniques for ageing or damaged structures.

[7] N-001 Structural Design, NORSOK, Rev. 2004.

[8] Kaloop, M.R., Roy,B., Chaurasia, K., Kim, S.-M., Jang, H.-M., Hu, J.-W., Abdelwahed, B.S., Shear Strength Estimation of Rein- force Concrete Deep Beams Using a Novel Hybrid Metaheuristic Optimized SVR Models. Sustainability 2022, 14, 5238.

[9] Kachouh, N., El-Maaddawy, T., El-Hassan, H., El-Ariss, B., Shear Behavior of Steel-Fiber-Reinforced Recycled Aggregate Concrete Deep Beams. Buildings 2021, 11, 423.

[10] Tariq, M., Khan, A., Ullah, A., Shayanfar, J., Niaz, M., Improved Shear Strength Prediction Model of Steel Fiber Reinforced Concrete Beams by Adopting Gene Expression Programming. Materials, 2022, 15, 3758.

[11] ACI 318-19: Building code requirements for structural concrete, American Concrete Institute, Farmington Hills, America, 2019, 436-449.

[12] Deifalla, A., Awad, A., Seleem, H., Abdelrahman, A., Investigating the behavior of lightweight foamed concrete T-beams under torsion, shear, and flexure. Eng. Struct. 2020, 219, 110741.

[13] Rotaru A., Measurements of bridge structures using non-destructive testing methods and their stability in wind gusts, Reliability: Theory &Applications, 2023, No. 4 (76), 1056-1066.

[14] Jansson A., Lofgren I., Lundgren K., Gylltoft K., Bond of reinforcement in self-compacting steel-fibre-reinforced concrete. Magazine of concrete research, 2012, Volume 64 (7), 617-630.

[15] Kouche, El., Hassanein, H.S., Ultrasonic non-destructive testing using wireless sensor networks. Procedia Computer Science, 10, 136-143.

[16] Sanayei, M., Phelps, J.E., Sipple, J.D., Bell, E.S., Brenner, B.R., Instrumentation, nondestructive testing, and finite-element model updating for bridge evaluation using strain measurements, Journal of bridge engineering, 17(1), 2011, 130-138.

[17] Teplý, B., Keršner, Z., Rovnaník, P. and Chromá, M., Durability vs. Reliability of RC structures. Durability of Building Materials and Components, 2005, 17-20.

[18] Balaji Rao, K., Anoop, M.B., Lakshmanan, N., Gopalakrishnan, S. and Appa Rao, T.V.S.R., Risk-based remaining life assessment of corrosion affected reinforced concrete structural members, 2004.

[19] Odriozola, M.A.B. and Gutiérrez, P.A., Comparative study of different test methods for reinforced concrete durability assessment in marine environment. Materials and Structures, 41(3), 2008, 527-541.

[20] Mehta, P.K. and Gerwick, B.C., Concrete in service of modern world. Dundee, Scotland, UK: Univ. of Dundee, 1996, 1-28.

[21] Lang, Ch., Willmes, M., Non-destructive testing of reinforced concrete structures, Germany, 2018.

[22] Petrakov A.N., Bukin A.V., (2010), Determination of concrete strength by destructive and non-destructive testing methods, No. 1, 89-94.

[23] Rotaru A., Determination of strength characteristics of materials of vertical load-bearing structures, in the collection: Potential of sustainable innovative development: concepts, models and practical application. Collection of articles on the results of the International Scientific and Practical Conference, 2024, 112-114.

[24] Rotaru A., Dynamic load assessment of building structures, E3S Wed of Conferences, 2023, T. 402, 07015.

METHODS AND TOOLS OF INTELLIGENT SUPPORT FOR FORECASTING THE TECHNICAL CONDITION OF CRITICAL SYSTEMS

Oleg Abramov, Dmitry Nazarov

•

Institute of Automation and Control Processes,
Far Eastern Branch of Russian Academy of Sciences, Russia
abramov@iacp.dvo.ru, nazardim@iacp.dvo.ru

Abstract

A variant of an expert-statistical approach to solving the problem of forecasting parametric deviations of critical systems condition is proposed. Issues of development of specialized software (case-based reasonong approach) with the necessary problem orientation (forecasting degradation of technical condition) and allowing to improve the quality of forecast are discussed. An approach to case describing using an ontological model of degradation processes with allowing to take into account both external influences, and internal processes characteristic of specific types of element, is proposed.

Keywords: technical condition, critical system, forecasting, reliability, decision support intelligent system, case-based reasoning

1. Introduction

Forecasting the state of technical devices and systems plays a crucial role in planning their operation and managing technological risks. The ability to predict possible failure moments is especially important for mission-critical objects, the failure of which is associated with significant material losses or catastrophic consequences. In most cases, these are complex systems manufactured in small numbers, operated in varying conditions, and implementing extreme technologies. The operation strategy for such systems should be individualized and preventive (anticipating failures) in nature.

Individualized operational planning is possible provided that current information about the actual condition of each object is obtained. Implementing an individualized operation strategy requires continuous or discrete monitoring and analysis of the technical condition of the object. It is assumed that the real technical condition of the object can be assessed based on the results of monitoring (measurements) of its parameters, and forecasting their changes allows the object to be operated until signs of a dangerous reliability decrease appear, thereby avoiding premature dismantling of units and assemblies, as well as performing other labor-intensive tasks, often of questionable usefulness for operational reliability.

The main difficulties in solving the forecasting problem for synthesis of operational strategies based on the condition are associated with the need to forecast for each object individually, with small volumes of initial information (based on a small set of monitoring results), and in the presence of noise (measurement errors), the statistical properties of which are not reliably known. In these conditions, classical methods of mathematical statistics and the theory of random processes lose their attractive properties, and their use for forecasting leads to significant errors and low forecast reliability.

There are some approaches to solving the problem of individual forecasting and operational planning with a deficit and incomplete reliability of initial information, allowing for sufficiently reliable results in these conditions. Among them is the method of individual guaranteed forecasting [1, 2]. Its main idea is that from the set of possible realizations of the degradation process of the properties (states) of the studied technical object, consistent with the monitoring results (not contradicting them), the "worst" ones are selected. By "worst" realizations, we mean those that can go beyond the operability range before others. Such realizations can be called extreme [2].

Unfortunately, to use the guaranteed forecasting method, it is necessary to have information about the form (model) of the process of changing the parameters of the technical condition of the studied object. Usually, a specific model of the parameter change process is assumed (as a hypothesis), based on which the forecasting problem is solved. In many cases, there is insufficient prior information to substantiate the choice of the preferred model hypothesis, and the sample of posterior data (monitoring results) is too small for any statistical inferences.

If the observation sample is limited and contains too little information for reliable estimation and forecasting, it is advisable to combine all available information, both objective (statistics, measurement results) and subjective (expert), in other words, to use an expert-statistical approach. Within the expert-statistical approach to forecasting, it is natural to use the same action scheme to refine forecasts as what is usually called actions by analogy and based on the cases that constitute the content of the accumulated practical experience of experts.

The essence of the approach to forecasting based on the method of analogs is to combine the capabilities of modern information theory with the apparatus of methods of system theory with artificial intelligence (which in this case are various computer implementations of action schemes by analogy) [3, 4].

The method of analogs is based on the assumption that in forecasting in a number of subject areas, experts try to make forecasts based on their representations of objects or processes, the history of which is well known to them, experts.

It should be noted that the method of analogs is not universal, and the possibility of its application is associated with certain limitations. For the method of analogs to be applicable, the corresponding subject area must have the property that there is always a large number of objects represented by a substantial volume of statistical and expert information describing them. The first step in the forecasting procedure is to choose a class of models describing the phenomenon being forecast. The next step is the process of collecting expert judgments. It is assumed that subject experts can provide estimates of the upper and lower bounds of possible values of the forecasted process (time series), express judgments about the monotonic decrease or increase in the values of the series over time, extract preliminary estimates of the moments of reaching minimum or maximum values of the series from their practical experience, and so on.

As a result, the forecasting problem is reformulated as follows: to find a sequence of future values most consistent with the monitoring results of the forecasted process and expert opinions. Since the length of the forecast base period is insufficient for reliable statistical conclusions, solving the corresponding forecasting problem imposes an additional – but now prioritized – requirement for the model to match fixed expert opinions and only then considering these opinions – observation results.

2. Means of Intelligent Support for Forecasting Methods

The use of the expert-statistical approach in solving forecasting tasks necessitates the application of information technologies for computer support of decision-makers. One way to provide such support is the creation of specialized software with a specific problem orientation (forecasting the degradation of technical condition) that allows improving the quality of the obtained results. Such software can be developed using an approach based on modeling of case-based reasoning (CBR) [4].

Choosing this approach is justified if, by the time the forecasting degradation problem arises,

a certain experience of solving similar problems has already been accumulated, which occurred earlier on similar technical objects (analogues). Representing this experience in the form of precedents and its automated processing using specialized software systems can significantly improve the effectiveness of the results.

Solving the forecasting problem based on precedents is based on recognizing the current problem situation, information about which is presented in the form of a certain pattern (case), and searching for similar pattern contained in the repository of patterns (cases base), followed by their adaptation and use for solving the forecasting task.

Let's highlight the main functions of the intelligent support system that ensure solving the forecasting task using plausible inference based on precedents.

- Formation of a case (analogue).
- Formation of a cases repository based on the forulated model.
- Finding a solution (plausible inference) based on cases.

To implement these functions, the following main modules of the intelligent support system are necessary:

- Module for cases modeling, which provides creation and modification of case models, formulation and updating of cases repositories based on existing models;
- Internal memory, which should provide storage of case models;
- Case inference engine, performing the search for precedents based on a given description;
- Management module, providing interaction between modules and providing an interface for communication with the external environment.

3. Case Structure

Automated systems implementing the case-based approach allow finding solutions to current problems based on past experience of solving similar problems. The success of implementing the case-based approach consists in preliminary analysis of the specificity of the problem being solved and largely depends on the choice of the way to describe the case.

Modeling of reasoning for forecasting the technical condition of a system based on the analysis of parameter drift over time is reduced to analyzing the degradation processes of its elements' properties. Every individual case consists a model of the parameter drift of the element. Each such model takes into account not only fatigue and wear processes, external environmental factors (which can also affect wear), but also more complex processes of mutual influence between elements (electromagnetic fields, heating from neighboring elements, vibrations), also taking into account their spatial arrangement relative to each other. However, in practice, known linear or exponential drift models are usually used [5]. Taking into account various external factors when modeling the drift process of each parameter requires a more complete description of the structure of the studied system.

The ontological approach offers broad possibilities for describing the structure of a technical object, taking into account the interaction of its components when solving tasks related to forecasting the technical condition [6, 7, 8, 10]. An ontology is a conceptual model of the subject area that describes the studied objects, processes, and phenomena in the form of a graph, where the vertices are concepts of the subject area, and the edges are all possible relations between them. For practical tasks, there are models of ontologies of various complexity, but in most cases and within the framework of the considered problem, it is sufficient to use a model in the form of a triplet $O = (X, R, \Phi)$, where X is the set of concepts of the subject area, R is the set of relations between these concepts, and Φ is the set of interpretation functions [6].

The use of ontologies allows describing the structure of a system of any complexity and level of detail with any known types of relationships between its elements and external influences. Such an approach to representing the structure of a technical object, taking into account the interconnections of its elements and factors causing failures, is used in software systems for Prognostic Health Management (PHM). In such systems, the hierarchy of elements with possible causes of their failure is usually described in detail, with quantitative characteristics specified

as intervals of admission and their current values [6], or probabilistic properties such as failure distribution density, mean time to failure, etc. [7]. On the basis of this description, it is possible to build a fault tree, identify the causes of failures, and build reasoning about the system's state in the form of productions based on predicate logic expressing the state of individual nodes in the hierarchy of elements, for example, the OntoProg model [6].

The ontological approach can also be used to build a degradation model of a complex system. When describing its structure, experts specify the factors affecting the performance and wear of each element. For example, Figure 1 illustrates an example of ontological representation of the hierarchy of system elements with a description of factors influencing the values of their operating parameters. The structure of the system is formed by set-theoretic relations "a-part-of" (depicted by solid arrows), describing the hierarchy of elements, and their influence on each other is described by relations based on changes in the physical parameters of the element's degradation model (shown by dashed arrows on the figure). Thus, over time, a rotating shaft due to friction can increase the temperature of surrounding elements, which can be accounted for in the used model of resistor and capacitor drift in the electronic circuit. Over time, a rotating shaft can become a source of vibrations, which negatively affects the quality of contacts of electronic components, accelerates the destruction of solder joints, which can be reflected in the corresponding failure model of the electronic board.

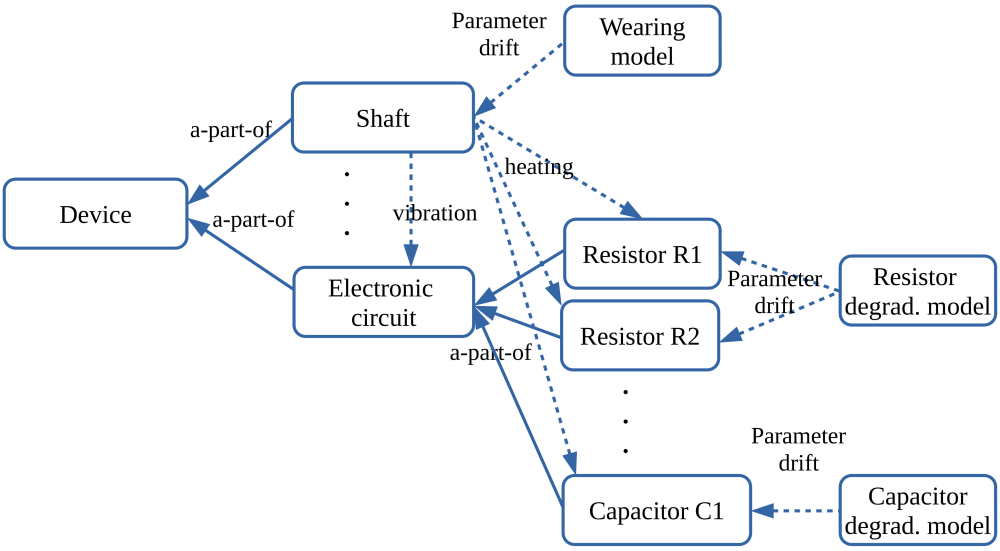


Figure 1: *Example of ontological representation of a technical device including parametric drift models*

Description of a case, considering the mutual influence of system elements and environmental factors, is based on the use of an ontological model and typically is a graph model. Processing graph structures during their storage, loading, and application in specific tasks requires more sophisticated procedures than linear structures of parametric models. A detailed account of all interrelationships influencing the properties of system elements is very labor-intensive and justified when designing expensive or mission-critical systems. On the other hand, overly detailed description of factors affecting parameter degradation may not only hinder expert support but also the reuse of the entire structure, especially for unique systems. Additionally, in similar systems, some components with similar functions may have different implementations (e.g., electrolytic and film capacitors [11]), and their parameter behavior over time and under the influence of the same factors may differ.

Despite the advantages of considering environmental parameters and factors of mutual influence, it is often considered sufficient in practice to use known drift models with a limited set of parameters, based on linear or exponential trends [9]. Thus, the case will include a parameterized drift model for a specific element:

$$C = (x_1, x_2, \ldots, x_{N_p}) \tag{1}$$

where $x_i, i = 1, 2, \ldots, N_p$ are the parameters of the drift model being used, N_p is the number of model parameters. In addition to the parameters of the drift model 1, for its identification (linear, exponential, and polynomial) and correct consideration of its parameters during reuse, the case should contain its conditional name:

$$CASE = (ID, C_p, Descr) \tag{2}$$

where ID is the unique identifier of the drift model type from the known models database, C_p are the parameters of the model 1, $Descr$ is a verbal description, and recommendations for application to the decision-maker. In the case of specifying a unique model not present in the known drift models database, the parameters C_p represent a symbolic description of the model in a specific format, taking into account the rules of the used mathematical expression parser [12].

4. The Module of Case Formation

The formation of a case using a degradation model of parameters is carried out at the stage of designing the technical object by an expert based on their experience and known regularities. The software module for case formation provides unified rules for describing cases based on the parametric model 2 of the case structure, as well as allows interaction with the ontological description of the system structure, in which the parameters of the external environment and the factors influencing other system components are specified.

The case is formed by the expert on the basis of processing and adaptation of the existing structure in the cases repository, or as a completely new one. In different tasks, the same components may or may not use external environmental factors. If a case takes into account any external influencing factor on the degradation process, but its use is not provided in the current task, then such a parameter is usually set to zero or NaN (not a number) when zero values are significant for a certain task.

In the interaction with the user, this software module represents an interactive guide, offering the expert a suitable case as the basis for processing for the current task, or contextual prompts in the form of suitable models for this system and external factors.

The models created by the expert are saved in a single case repository in parametric form. If necessary, ontological information about the system and the relationships between its elements, to which references from the case structure exist, can be saved separately.

5. Case-Based Plausible Inference Module

The search for a suitable case is based on its representation method. The parametric representation of its structure allows the application of cluster search methods, such as the Nearest Neighbor (NN) method [9]. The case inference software module analyzes the current description of the designed system by examining the relationships between its components, considering possible environmental factors, and searches the case database for a suitable model of the degradation process.

The tuning parameters of the case search include fundamental properties such as the metric for assessing the proximity of cases and the number of structures used in NN method. However, in the task of investigating unique systems, it is important to consider the uniqueness factor of their individual elements, which may make the use of a case-based approach not entirely appropriate.

The result of this module's operation is a specific degradation model for a system element, which is integrated into the structure of its representation to assess reliability characteristics, operational reserve, and failure risk [13]. This module provides means of explaining the selection made and proposing alternatives in case of identifying disputed decisions. The final decision on

the use of the proposed model, its adaptation for correct implementation into the model of the studied system, is made by an expert or the decision-making authority.

6. Conclusion

An expert-statistical approach to forecasting the processes of parametric deviations of the technical condition of critical systems based on the method of analogs has been considered. The essence of the method lies in experts' involvement in solving the forecasting problem, based on their identification of analogs of the predicted process from among previously observed processes.

The development of specialized software (tools for intelligent support implementing a case-based reasonong) with the necessary problem orientation (forecasting the degradation of the technical condition) and capable of improving the quality of the obtained results is discussed. The proposed approach to describing a case is based on an ontological model of knowledge about the degradation processes of system components. The method of using domain-specific ontologies for constructing degradation models taking into account all possible factors of both external influences and internal processes characteristic of specific types of components is considered. The use of such models is most typical for tasks within the framework of the functional-parametric direction in reliability theory.

The research was carried out within the state assignment of IACP FEB RAS (Theme FWFW-2021-0003).

References

[1] Abramov, O., Dimitrov, B. (2017). Reliability design in gradual failures: a functional-parametric approach. *Reliability: TheoryApplication*, Vol.12, 4(47):39–48.

[2] Abramov, O., Nazarov, D. (2016). Condition-based maintenance by minimax criteria. *Applied Mathematics in Engineering and Reliability. Proceedings of the 1st International Conference on Applied Mathematics in Engineering and Reliability (Ho Chi Minh City, Vietnam, 4-6 May 2016). CRC Press*, 91–94. DOI: 10.1201/b21348-16.

[3] Mandel, A.S. (2004). Method of analogs in prediction of short time series: An expert-statistical approach. *Automation and Remote Control*, Vol. 65, 4:634–641.

[4] Aamodt, A., Plaza, E. (1994). Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, Vol. 7, 1:39-59.

[5] Jones, J.A. (1999). A toolkit for parametric drift modelling of electronic components. *Reliability Engineering System Safety*, Vol. 63, 1:99–106.

[6] Nunez, D. L., Borsato, M. (2018). OntoProg: an ontology-based model for implementing prognostics health management in mechanical machines. *Advanced Engineering Informatics*, 38:746–759. DOI: 10.1016/j.aei.2018.10.006.

[7] Venceslau, A., Lima, R., Guedes, L.A. and Silva, I. (2014). Ontology for computer-aided fault tree synthesis. *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*:1–4.

[8] Ebrahimipour, V., Rezaie, K., Shokravi, S. (2009). An ontology approach to support FMEA studies. *Annual Reliability and Maintainability Symposium, Fort Worth, TX, USA, 2009*:407–411. DOI: 10.1109/RAMS.2009.4914711.

[9] Weber, R., Schek, H.-J., Blott, S. (1998). A quantitative analysis and performance study for similarity search methods in high dimensional spaces. *VLDB 98 Proceedings of the 24rd International Conference on Very Large Data Bases*:194–205.

[10] Eremeev, A., Varshavsky, P. (2007). Methods and Tools for Reasoning by Analogy in Intelligent Decision Support Systems. *Proceedings of the International Conference on Dependability of Computer Systems. Szklarska Poreba, Poland, 14-16 June 2007, IEEE*:1–4.

[11] Gupta, A., Yadav, O.P., DeVoto, D., Major, J. (2018). A Review of Degradation Behavior and Modeling of Capacitors. *Proceedings of the ASME 2018 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems. ASME 2018*

International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems. San Francisco, California, USA:1–10.

[12] Lange, C. (2013). Ontologies and languages for representing mathematical knowledge on the Semantic Web. *Semantic Web*, 4, 2:119–158.

[13] Abramov, O.V., Nazarov, D.A. (2022). Functional-parametric direction of risk theory. *Proc. of the 13th Asian Control Conference (ASCC 2022), Korea*:1911–1913. DOI: 10.23919/ASCC56756.2022.9828266.

RT&A, No 4(80)
Volume 19, December, 2024

A NEW ALGORITHM FOR MODELING ASYMMETRICAL DATA – AN EMPIRICAL STUDY

K.M. Sakthivel and Vidhya G

•

Department of Statistics, Bharathiar University, Coimbatore 641046, Tamil Nadu, India
sakthithebest@buc.edu.in,vidhyastatistic96@gmail.com

Abstract

In the current era, it is quite challenging to find symmetric data, as the form of most real-world data is asymmetric, meaning it tends to slant towards one side or another. These types of data emerge from various fields, including finance, economics, medicine, and reliability. Traditional statistical models often fail to handle such type of data as most of the statistical procedures are developed under normality assumptions. Therefore, the usual way of modeling these data results in incorrect predictions or leads to wrong decisions. There is no familiar methodology available in the research for modeling asymmetric data. Hence, there is a need to address this research gap as an emerging area of research in statistical modeling. In this paper, we propose a new systematic approach called the Model Selection Algorithm for modeling asymmetric data. In this algorithm, we incorporate various statistical tools and provide a guideline for a step-by-step procedure. Further, we have applied maximum likelihood estimation for parameter estimation, and model selection criteria such as Cramer Von Mises, Anderson Darling, and Kolmogorov Smirnov tests. We used real-time data to demonstrate the effectiveness of the algorithm.

Keywords: Lifetime distributions, Estimation, Information Criteria, Goodness of fit, Model selection.

1. Introduction

In this information era, data plays a significant role in policy and decision-making in various fields. With the advancement of technology, data generation, and its utilization have been increasing exponentially. However, it is important to note that the behavior of data is dynamic and depends on several factors. Therefore, it is crucial to understand the intricacies of data and its behavior to utilize it effectively. As part of our data analysis process, we use statistical models to gain a better understanding of the patterns and trends present in the data. These models help us to delve deeper into the underlying structure of the data and make informed decisions based on the findings from the statistical inference.

The foundation of the probability models is to capture the dynamics and variability of data. A vast amount of literature has been written about probability models, and new results are being produced daily. Choosing the right model for a particular asymmetrical dataset may be difficult for even statistics experts. To facilitate this process, we have developed a framework that considers choosing the best model for asymmetric data under study. Both statisticians and data analysts may benefit from this approach to make sensible Inference.

Data that has an uneven pattern due to an unequal distribution of data points' frequencies is referred to as asymmetrical data or skewed data. This kind of data is not symmetrical since the mean, median, and mode are not equal and also not at the same location. Consequently, the distribution takes on an extended form on one side and a longer or fatter tail on the other. We

will encounter many fields including Finance, Economics, Medicine, etc.

To provide a more sophisticated depiction of asymmetrical data, a combination of probability models may be employed. This technique proves especially valuable when handling data from a range of sources or with varying patterns. Blending different distributions into one mixture distribution enables the model to more effectively capture the distinctive features of the data. This proves particularly advantageous when the true underlying distribution remains unknown. By embracing a mixture model, analysts can more precisely assess the probability of different outcomes and make informed decisions based on the data at hand.

Karl Pearson [25] a prominent biometrician, proposed one of the earliest mixture models by fitting a proportionate mixture of two normal density functions. Subsequently, several writers used the finite mixture model to create bimodal distributions. This approach was limited to bimodal or multimodal datasets until Lindley [22] used the finite mixture model to generate a single parameter distribution for unimodal data. Furthermore, a mixed distribution was created by utilizing different proportions of gamma and exponential distributions, resulting in an improved outcome.

For the past few years, the authors of the study initially developed a model, which was subsequently applied to real-time data. However, the current situation presents a challenge, as there is no clarity on the most appropriate model to use for the available data. The data at hand are asymmetric, which makes it difficult to determine the most suitable distribution. To address this issue, a framework was established to identify the model that best aligns with the data. This involved a meticulous examination of the data properties to select a model that accurately captures the data features. Consequently, a model was developed that can provide reliable predictions through this approach. The study aims to provide a concise and comprehensible process for selecting an appropriate model for asymmetric real-time data. Considerable efforts were made to ensure that the method is straightforward and easy to understand, with the ultimate goal of facilitating the decision-making process for businesses and academics alike.

In our paper, we begin by presenting the framework that we followed for fitting skewed data in section 2. In Section 3, we also provide additional insight into the assumptions and limitations of our approach. We discuss a real-time application to demonstrate the effectiveness of our framework in finding the most suitable mixture model by our proposed methodology. Finally, in Section 4, we discuss the basic properties of the proposed model and also discuss the simulation work that was done for the proposed model.

2. Model Selection Algorithm

This algorithm outlines a comprehensive and systematic approach to analyzing asymmetrical data. It involves a model selection process that is designed to provide a detailed understanding of the data. The first step in this process is to gather the data and partition it using clustering techniques. This allows for the identification of distinct groups within the data, which can then be analyzed separately. After the data has been partitioned, the next step is to fit a probability distribution to each partition. This is done to determine the best-fit model for each group of data. The selection of the best-fit model is based on a variety of factors, including the goodness of fit, the complexity of the model, and the interpretability of the results. Finally, the best-fit models from each partition are combined to propose a comprehensive hybrid model. This model provides a detailed understanding of the data and allows for the identification of patterns and trends that may not be apparent when analyzing the data as a whole. Sakthivel and Vidhya ([26]-[27]) have discussed the algorithm and given different applications.

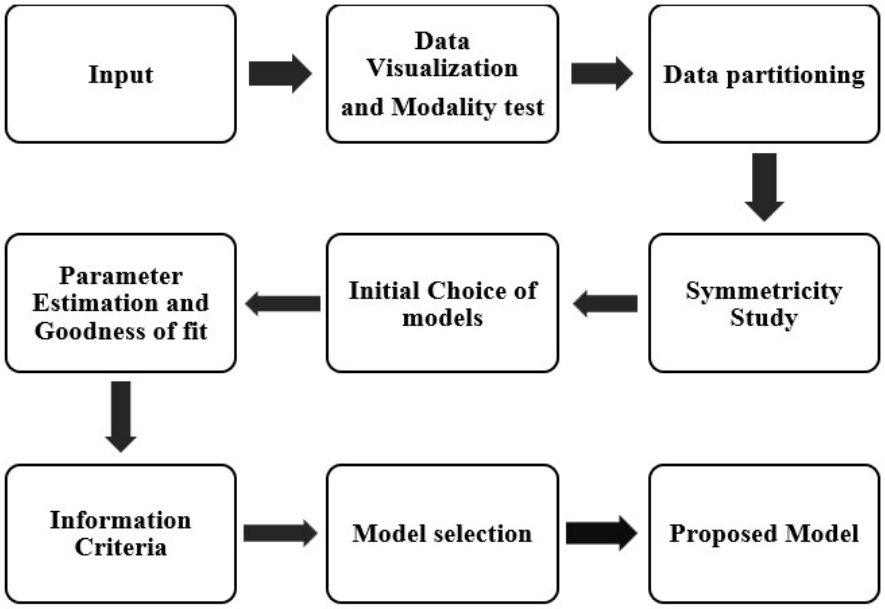


Figure 1: *Framework for selecting a better model*

The steps involved in this algorithm are as follows:

Step 1: Consider the asymmetrical data.

Step 2: Visualize the data to identify the distribution of the random variable. And test the modality of the data.

Step 3: Divide the data into two parts using the clustering technique.

Step 4: Calculate the skewness for both parts of the data to capture the asymmetry.

Step 5: Consider the basic distributions for modeling based on the data characteristics.

Step 6: For the first part of the data, estimate the value of parameters using MLE for suitable probability distributions and determines the model's adequacy by computing the goodness of fit and information metrics.

Step 7: Repeat the process for the second part of the data.

Step 8: Choose a better model from the considered distributions in steps 6 & 7, based on minimized -2LL, AIC, BIC, and AICc values.

Step 9: Propose a new model by combining the selected models from Step 8.

An advanced framework has been developed to effectively analyze asymmetrical data using a suite of analytical tools. In our framework, we incorporate the fundamental models to choose a better model. The distributions incorporated in this framework such as symmetric, heavy-tailed, light-tailed, positively skewed, and negatively skewed models. The framework employs a range of statistical techniques, including K-mean Clustering for initial data partitioning, maximum likelihood estimation for parameter estimation, and statistical tests such as Cramer Von Mises, Anderson Darling, and Kolmogorov Smirnov for rigorous model evaluation. The model selection process is further refined through the application of information criteria, which includes Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Corrected Akaike Information Criteria (AICc). These criteria help to identify the most suitable model for analyzing the data,

ensuring that the insights gained are both accurate and reliable.

This comprehensive approach ensures an efficient methodology for gaining valuable insights from asymmetrical datasets and can be used in a variety of applications, including Engineering, Finance, Marketing, and Healthcare. By utilizing a range of analytical tools and statistical techniques, the framework can effectively analyze complex data sets, providing valuable insights that can be used to inform decision-making processes.

Before using the framework, it is important to consider certain assumptions. Firstly, the data used for analysis should be asymmetrical with dual peaks and should not be perfectly bimodal. Additionally, the framework is specifically designed for situations where the distribution of the data significantly deviates from a normal distribution. To be suitable for analysis using this framework, the data must exhibit some degree of skewness, either positive or negative. Furthermore, the framework is best suited for data that exhibits heavy tails and allows for the existence of outliers in the dataset. It is important to note that this framework is designed for univariate data analysis only.

It is important to keep in mind that there are certain limitations to this procedure. Especially, this framework is designed to handle asymmetrical data and requires dividing the data into exactly two distinct groups. This method has been found to produce superior results for smaller datasets. We use the limited models to choose a better model. Manual processing of this framework with significant amounts of data can present numerous challenges and be time-consuming. In future research, we aim to automate this process through statistical software and produce results using simple codes. This approach will help streamline the process and make it more efficient for larger datasets.

3. Model Selection and Validation

In this section, we would like to apply our framework to the datasets on the glass strength of aircraft windows, as originally reported by Fuller et al. (1994). This dataset is widely recognized and has been utilized by various authors in the literature. The strength data involved can potentially originate from different underlying probability distributions, each representing varying conditions or modes.

It is important to note that the strength of aircraft windows is a crucial factor in ensuring the safety and reliability of air travel. Hence, it is essential to accurately represent the underlying distribution of the strength data to better understand and predict its behavior. By utilizing a mixture of probability distribution, we can more effectively capture the inherent variability in the data and simply fit standard probability distribution. The proposed approach will allow us to identify the most suitable model for the actual strength data to ensure the safety and reliability of aircraft windows.

Further, a mixture of probability distributions is a statistical model that combines two or more probability distributions into a single probability distribution. This concept of mixture enables to modeling of multimodal data, which present different types of defects or variations in the manufacturing process. By incorporating multiple distributions, we can capture the full range of variability in the data, which is crucial for modeling the strength of aircraft windows under different conditions.

The summary statistic of the data is given in Table 1.

Table 1: *Summary of data*

Minimum	First Quartile (Q1)	Median	Mean	Third Quartile (Q3)	Maximum	Skewness
18.83	25.51	29.90	30.81	35.83	45.38	0.4263

It is observed that the mean and median values are different. This indicates that the shape of the density plot is likely to be skewed and not symmetric. These findings can be ensured with the support of other statistical plots given below.

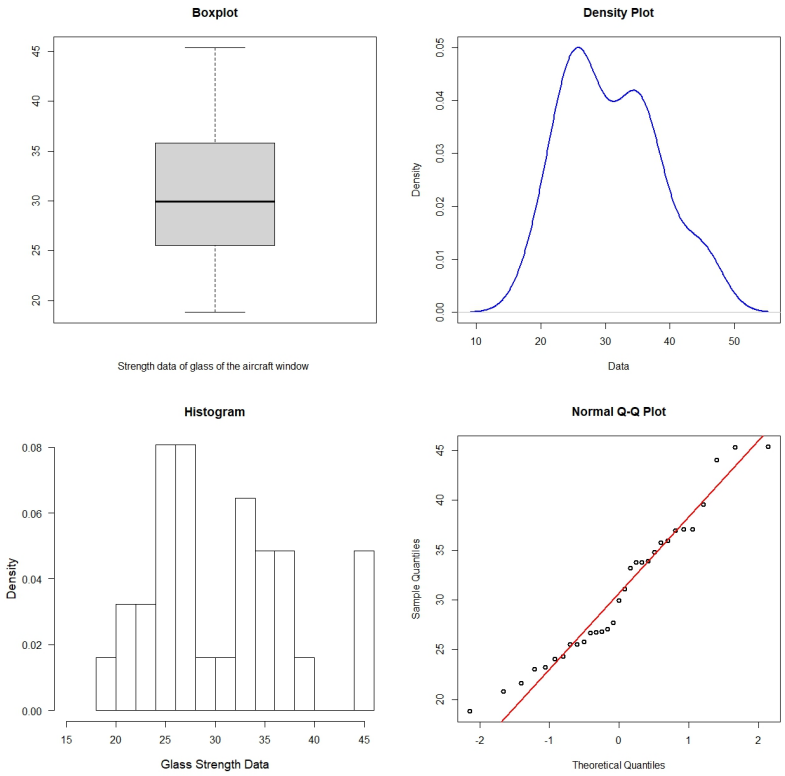


Figure 2: *Graphical representation of the data*

The analysis of the dataset through graphical representation has revealed the presence of two modes, which are depicted in Figure 2. Additionally, we check the modality of the data by using the Hartigan dip test (1985). Dip statistic=0.081364 ($p - value$ =0.09169). The Hartigan dip test confirms the data's unimodality. The Q-Q plot, which shows that the data is not symmetric, indicates that the data is asymmetric. In light of these findings, we have proceeded to step 3 of our analytical framework, which involves the division of the data into two parts using K-mean clustering. Figure 3 presents the clustered data along with the cut line that was used to divide the data.

After partitioning the data, the next process will evaluate the skewness of both partitions of the datasets. We can select the model accordingly if the data is positively skewed or negatively skewed.

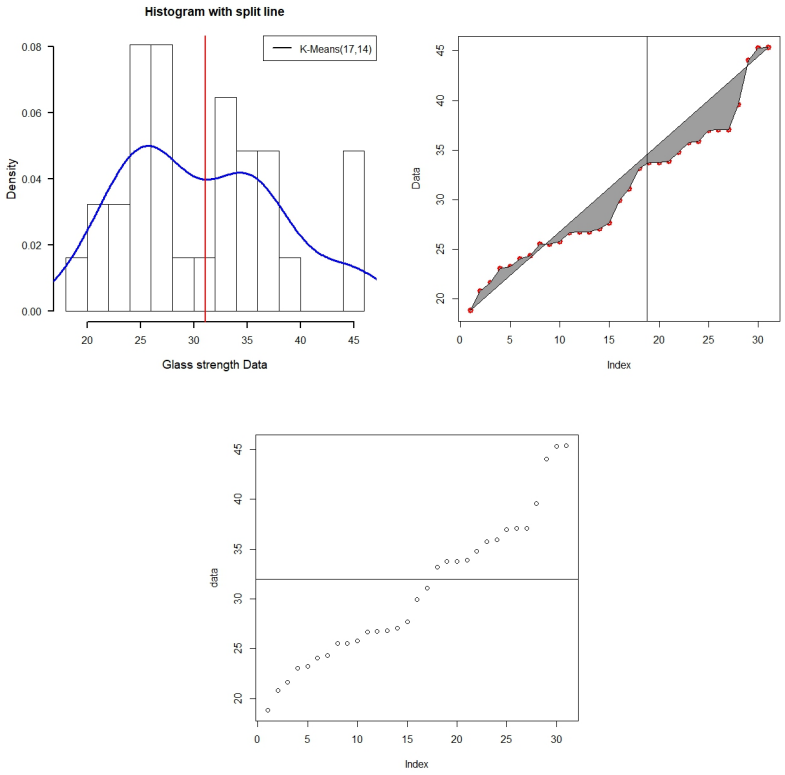


Figure 3: *Data partitioning for data*

Table 2: *Summary of the first part of the data*

Minimum	Q1	Median	Mean	Q3	Maximum	Skewness
18.83	23.23	25.52	25.22	26.78	31.11	-0.1538

Since the initial portion of the data is unimodal and negatively skewed, we can consider a probability distribution that also portrays this particular characteristic. The low value of skewness indicates that the distribution may be nearly symmetric. Therefore, we considered the normal distribution as one of the choices. The present section deals with the computation of the tools mentioned in Section 2 for diverse distributions. These computations are crucial to accomplishing step 6 of the analysis. All statistical findings and calculations have been obtained through the R programming language, and the resulting outcomes have been documented in Table 3.

Lower values for KS, CVM, and AD suggest a better fit between the model and data. Higher values imply a poor fit. We can compare p-values to determine the goodness of fit. A p-value closer to zero denotes a weaker fit, while a p-value closer to unity denotes a better fit.

From Table 3, we are unable to identify which model best matches the data. On the other hand, we might go on to the next process if we filter the model using Table 3. After calculating the information criteria, the data's best fit is selected.

Table 3: *Estimated parameters value and Goodness of fit for the first part of the dataset.*

Model	Estimated Parameter	-2logL	CVM (p values)	AD (p-values)	KS (p-values)
Gamma	$\hat{\alpha}$=66.3065 $\hat{\beta}$=2.6293	86.4328	0.0419 (0.9287)	0.2466 (0.972)	0.1406 (0.8451)
Lognormal	$\hat{\mu}$=3.2200 $\hat{\sigma}$=0.1239	86.7240	0.0468 (0.9014)	0.2746 (0.9553)	0.1480 (0.7992)
Weibull	$\hat{k}$=9.2248 $\hat{\lambda}$=26.5551	86.4999	0.0362 (0.9561)	0.2449 (0.9729)	0.1291 (0.9058)
Cauchy	$\hat{a}$=25.6887 $\hat{b}$=1.7157	91.4211	0.0546 (0.8542)	0.3931 (0.8537)	0.1259 (0.9195)
Logistic	$\hat{a}$=25.2943 $\hat{b}$ =1.73282	86.4194	0.0316 (0.9742)	0.1950 (0.9919)	0.1179 (0.9500)
Normal	$\hat{\mu}$=25.2181 $\hat{\sigma}$=3.0433	86.0835	0.0344 (0.9636)	0.2100 (0.9876)	0.1251 (0.9231)
Gompertz	$\hat{k}$=0.1834 $\hat{\lambda}$=0.0016	96.5179	0.3806 (0.0803)	1.9689 (0.0962)	0.2722 (0.1328)
Gumbel	$\hat{k}$=23.6725 $\hat{\lambda}$=3.0430	89.1066	0.0793 (0.7019)	0.4724 (0.7727)	0.1660 (0.6767)
Laplace	$\hat{\lambda}$=25.5200 $\hat{\beta}$=2.3737	86.9575	0.0408 (0.9343)	0.2566 (0.9666)	0.1100 (0.9717)

Table 4: *Model selection criteria for the first part of the dataset*

Model	AIC	BIC	AICc
Gamma	90.4328	92.0993	91.2899
Lognormal	90.7240	92.3905	91.5812
Weibull	90.4999	92.1664	91.3571
Cauchy	95.4211	97.0875	96.2782
Logistic	90.4194	92.0858	91.2766
Normal	90.0835	91.7499	90.9407
Gompertz	100.5179	102.1843	101.3750
Gumbel	93.1066	94.7730	93.9637
Laplace	90.9575	92.6239	91.8147

From Table 4, selecting a model is a simpler process. It was determined that the distribution with the lowest values of AIC, BIC, and AICc provided the best fit. Hence, the normal distribution is the most suitable fit for the first segment of the data. This process is then repeated for the second part of the data.

Table 5: *Summary of the second part of the dataset*

Minimum	Q1	Median	Mean	Q3	Maximum	Skewness
33.20	34.11	36.45	37.60	38.96	45.38	0.90099

Since the data's second component is skewed towards positive values, we should choose a probability distribution that also displays positive skewness. This will help us accurately represent the shape of the data and ensure more representative statistical analysis and modeling.

Table 6: *Estimated parameters value and Goodness of fit for the second part of the dataset.*

Model	Estimated Parameter	-2logL	CVM (p-values)	AD (p-values)	KS (p-values)
Rayleigh	$\hat{\sigma}$=26.753	110.6552	0.8793 (0.0039)	4.1262 (0.0078)	0.5370 (0.0003)
Lindley	$\hat{\theta}$=0.0519	119.5955	0.9050 (0.0034)	4.3004 (0.0065)	0.5289 (0.0004)
Exponential	$\hat{\theta}$=0.0266	129.5586	1.1305 (0.0009)	5.2124 (0.0024)	0.5866 (0.0000)
Gamma	$\hat{\alpha}$=85.9100 $\hat{\beta}$=2.2846	78.7922	0.1428 (0.4163)	0.9128 (0.4045)	0.2506 (0.2914)
Weibull	$\hat{k}$=8.8277 $\hat{\lambda}$=39.5728	82.47142	0.1983 (0.2725)	1.1507 (0.2861)	0.2829 (0.1745)
Pareto	$\hat{\alpha}$=8.4226	69.7299	0.0427 (0.9254)	0.2184 (0.9324)	0.1266 (0.9573)
Lomax	$\hat{\alpha}$=185.6654 $\hat{\lambda}$=5.3694	132.0715	1.1395 (0.0008)	5.2563 (0.0023)	0.5866 (0.0000)
Lognormal	$\hat{\mu}$=3.6213 $\hat{\sigma}$=0.1064	78.3804	0.1344 (0.4453)	0.8709 (0.4305)	0.2440 (0.3208)
Cauchy	$\hat{a}$=35.8128 $\hat{b}$=1.9214	81.5577	0.1156 (0.5192)	0.8219 (0.4631)	0.2018 (0.5517)
Logistic	$\hat{a}$=36.9773 $\hat{b}$=2.3644	80.0735	0.1038 (0.5729)	0.8596 (0.4378)	0.2024 (0.5483)
Normal	$\hat{\mu}$=37.6033 $\hat{\sigma}$=4.1682	79.7004	0.1608 (0.3614)	1.0064 (0.3524)	0.2633 (0.2402)
Gompertz	$\hat{k}$=0.1200 $\hat{\lambda}$=0.0014	91.5035	0.5587 (0.0269)	2.6816 (0.0407)	0.4545 (0.0037)
Gumbel	$\hat{k}$=35.7494 $\hat{\lambda}$=2.8798	75.6418	0.0814 (0.6912)	0.6522 (0.5971)	0.1805 (0.6872)
Laplace	$\hat{\mu}$=36.6535 $\hat{\beta}$=3.1747	79.7542	0.0815 (0.6908)	0.7508 (0.5152)	0.1684 (0.7628)

Table 7: *Model selection criteria for the second part of dataset*

Model	AIC	BIC	AICc
Gamma	82.7922	84.0703	83.8831
Weibull	86.4714	87.7495	87.5623
Pareto	71.7299	72.3689	72.0633
Lognormal	82.3804	82.3804	83.6585
Cauchy	85.5577	86.8358	86.6486
Logistic	84.0735	85.3516	85.1644
Normal	83.7004	84.9785	84.7913
Gumbel	79.6418	80.9199	80.7328
Laplace	83.7542	85.0323	84.8451

Applying the same methodology to the second half of the data, we can conclude from Tables 6 and 7 that the Pareto distribution is the best fit. We may go on to the next stage once we have successfully obtained the two components required for the two sections of our data. We may infer from step 7 that a Pareto distribution would be appropriate for the second half of the data, and a normal distribution for the first part. Step 9 involves creating a new distribution called the Normal-Pareto distribution (NPD) by combining these two distributions using the finite mixture

model. Section 4 describes the function of NPD.

We must test our suggested model to ensure that it functions accurately and efficiently with data. We will be able to demonstrate that our framework performs better than the current model and yields better outcomes through this approach. Thus, we carry out step 6 again using the suggested model, and Table 8 presents the outcomes.

Table 8: *Estimated parameters value, Goodness of fit, and Model selection criteria for the dataset*

Model	Estimated Parameter	-2LL	CVM	AD	KS	AIC	BIC	AICc
Rayleigh	$\hat{\sigma}$=22.36	236.4	0.8388 (0.0055)	4.3918 (0.0057)	0.3189 (0.0027)	238.4	239.8	238.5
Lindley	$\hat{\theta}$=0.063	253.9	1.1389 (0.0010)	5.8718 (0.0011)	0.3655 (0.0003)	255.9	257.4	256.1
Exponential	$\hat{\theta}$=0.032	274.5	1.7891 (0.000)	8.5303 (0.000)	0.4587 (0.0000)	276.5	277.9	276.6
Gamma	$\hat{\alpha}$=18.93 $\hat{\beta}$=0.614	208.2	0.0816 (0.6863)	0.4387 (0.8085)	0.1349 (0.5785)	212.2	215.1	212.6
Lognormal	$\hat{\mu}$=3.401 $\hat{\sigma}$=0.231	208.0	0.0791 (0.7007)	0.4136 (0.8341)	0.1246 (0.6759)	212.0	214.8	212.4
Weibull	$\hat{k}$=4.635 $\hat{\lambda}$=33.674	210.9	0.0908 (0.6353)	0.5973 (0.6492)	0.1526 (0.4238)	214.9	217.8	215.4
Pareto	$\hat{\alpha}$=2.146	225.5	0.6215 (0.0192)	3.2534 (0.000)	0.25419 (0.0298)	227.5	228.9	227.6
Lomax	$\hat{\alpha}$=113.53 $\hat{\lambda}$=4.392	281.5	2.0613 (0.0000)	9.6293 (0.000)	0.4903 (0.000)	285.5	288.3	285.9
Cauchy	$\hat{a}$=29.258 $\hat{b}$=5.093	225.6	0.1696 (0.3363)	1.2009 (0.2669)	0.1612 (0.3573)	229.6	232.4	230.0
Logistic	$\hat{a}$=30.44 $\hat{b}$=4.224	211.6	0.0966 (0.6051)	0.5729 (0.6727)	0.1425 (0.5094)	215.6	218.5	216.0
Normal	$\hat{\mu}$=30.81 $\hat{\sigma}$=7.135	209.8	0.0936 (0.6203)	0.5559 (0.6892)	0.154 (0.4125)	213.8	216.6	214.2
Gompertz	$\hat{k}$=0.117 $\hat{\lambda}$=0.002	216.7	0.1332 (0.4473)	1.0257 (0.3434)	0.1549 (0.4053)	220.7	223.6	221.1
Gumbel	$\hat{k}$=27.399 $\hat{\lambda}$=5.986	208.2	0.0757 (0.7204)	0.3980 (0.8497)	0.1358 (0.5704)	212.2	215.1	212.7
Laplace	$\hat{\mu}$=29.900 $\hat{\beta}$=6.124	217.3	0.1477 (0.3984)	0.8691 (0.4328)	0.1599 (0.367)	221.3	224.2	221.7
Normal-Pareto (NPD)	$\hat{\theta}$=0.973 $\hat{\mu}$=29.690 $\hat{\sigma}$=8.117 $\hat{\alpha}$=152.23	202.3	0.062 (0.8019)	0.551 (0.6590)	0.122 (0.6984)	210.3	216.0	211.8

By comparing the newly created probability distribution against conventional distributions using every criterion that was used to choose the model, its goodness of fit was assessed. Table 8 and Figure 4 make it clear that our suggested mixed probability model yields the best results, and our methodology helps select a more appropriate model for the skewed data.

However, comparing the traditional distribution alone is not enough to prove that our proposed model is a better fit for the data so we extended the study and collected various distributions using various techniques, and compared it with our proposed model. The results are given in Table 9.

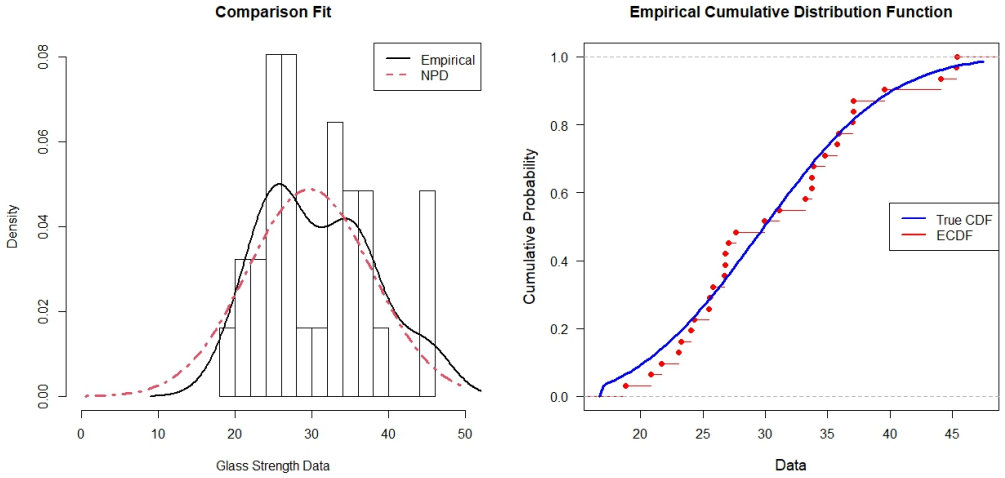


Figure 4: *Comparison fit for the dataset*

Other than the model listed in Table 9, many distributions are taken into account for comparison. The distributions are Shanker, Akash, Rama, Suja, Sujatha, Amarendra, Devya, Shambhu, Aradhana, Akshya, Inverse Rayleigh, Inverse Exponential, Kpenadidum, Iwok-Nwi, Two parameter Pranav, Two parameter Sujatha, Weibull Extended Pranav, Extended Pranav, Weibull-Lindley, Weibull- Pranav, Exponentiated Exponential, Exponentiated Hypoexponential, Generalized Inverted Exponential, Inverted Exponential, Beta Generalized Inverted Exponential, Exponentiated Exponential, Exponentiated Lindley, Exponentiated Akash, Gold, Power Size Biased Two Parameter Akash distributions.

In summary, the Normal-Pareto distribution provides the best fit to the data when compared to other distributions.

4. Key Properties of Normal-Pareto Distribution

From the above statistical analysis, we have examined our proposed model key features in depth. The model satisfy the essential condition for the distribution function, this shows that they can be used for more research and application in pertinent areas.

4.1. Normal-Pareto Distribution (NPD)

The model can be obtained using a finite mixture model.

$$f(x) = w_1 g_1(x) + w_2 g_2(x) \tag{1}$$

Equation (1) is used for developing the Normal-Pareto model. Where $g_1(x) \sim Normal(\mu, \sigma)$, $g_2(x) \sim Pareto(x_m, \alpha)$ and $w_1 = \theta\,; w_2 = 1 - w_1 = 1 - \theta$ and x_m is the minimum of x. To obtain a perfect density function, we utilized a normalizing constant.

Let $X \sim NPD(\theta, \mu, \sigma, \alpha)$ then the probability density function (pdf) and cumulative distribution function (cdf) for the NPD are

Table 9: *Estimated value of parameters, and model selection criteria for data for various distributions*

Model	Estimated Parameter	-LL	AIC	BIC	AICc
A Distribution (A)	$\hat{\alpha}$=125.662	107.950	217.901	219.335	218.039
Inverse Gompertz (IG)	$\hat{\alpha}$=1.249, $\hat{\beta}$=119.762	107.884	219.768	222.636	220.196
Kumaraswamy Inverse Gompertz (KuIG)	$\hat{\alpha}$=79.042, $\hat{\beta}$=18.694 $\hat{\gamma}$=26.554	103.988	213.976	218.278	214.865
Exponentiated Aradhana	$\hat{\alpha}$=19.1870, $\hat{\theta}$=0.2200	104.083	212.165	215.033	212.594
Inverse Weibull (IW)	$\hat{\alpha}$=446.1827, $\hat{\beta}$=4.655	105.323	214.647	217.515	215.075
Lomax Gumbel Type-Two (LGTT)	$\hat{\alpha}$=31.7086, $\hat{\beta}$=0.4549 $\hat{\theta}$=89.5227, $\hat{k}$=0.8379	104.818	217.636	223.372	219.174
Lomax-Gompertz (LomGo)	$\hat{\alpha}$=0.2952, $\hat{\beta}$=3.7704 $\hat{\theta}$=0.0005, $\hat{k}$=0.2523	105.729	219.457	225.193	220.996
Weighted Quasi Akash Distribution	$\hat{\theta}$=0.6152, $\hat{\alpha}$=3.7439 $\hat{\beta}$=16.9691	104.117	214.234	217.528	215.123
Three-parameter Weighted Lindley distribution (TWLD)	$\hat{\theta}$=0.6198, $\hat{\alpha}$=18.300 $\hat{\beta}$=16.9691	104.119	214.238	217.558	215.127
Weighted New quasi-Lindley	$\hat{\alpha}$=4.7687, $\hat{\theta}$=0.6146 $\hat{c}$=-16.9412	104.116	214.232	218.534	215.120
Harris Extended Generalized Exponential Distribution (HEGED)	$\hat{c}$=-0.1121, $\hat{\varphi}$=0.0761 $\hat{\lambda}$=0.1566, $\hat{\theta}$=5.0976	104.093	216.186	221.922	217.724
Marshall Olkin Extended Generalized Exponential Distribution (MOEGE)	$\hat{\varphi}$=0.0761, $\hat{\lambda}$=0.1566 $\hat{\theta}$=5.0976	105.776	217.552	221.854	218.441
Inverse Flexible Weibull (IFW)	$\hat{\alpha}$=61.167, $\hat{\beta}$=0.0859	104.963	213.927	216.795	214.355
Exponentiated Inverse Flexible Weibull (EIFW)	$\hat{\alpha}$=2.376,$\hat{\beta}$=0.164 $\hat{\gamma}$=81.51	104.141	214.282	218.584	215.171
Normal-Pareto (NPD)	$\hat{\theta}$=0.973, $\hat{\mu}$=29.6904 $\hat{\sigma}$=8.1176, $\hat{\alpha}$=152.232	101.167	210.334	216.069	211.872

$$f(x) = \frac{2\left(\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + (1-\theta)\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}\right)}{\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)\theta+2} \tag{2}$$

$$F(x) = \frac{x^{\alpha}\left(\theta\left(erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)+erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-2\right)+2\right)+2x_m^{\alpha}(\theta-1)}{\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)x^{\alpha}} \tag{3}$$

For $x_m \leq x, \alpha > 0, \mu \in \mathbb{R}, \sigma > 0, 0 \leq \theta \leq 1$. Where, x_m is the minimum of x.

The survival function of X is

$$S(x) = \frac{1}{\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)x^{\alpha}} \times x^{\alpha}\left(\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)-\left(\theta\left(erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)+erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-2\right)+2\right)\right) - 2x_m^{\alpha}(\theta-1) \tag{4}$$

The hazard function of X is

$$h(x) = \frac{2x^{\alpha}\left(\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}\right)}{x^{\alpha}\left(\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)-\left(\theta\left(erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)+erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-2\right)+2\right)\right)-2x_m^{\alpha}(\theta-1)} \tag{5}$$

Figure 5 displays the possible shapes of the pdf, cdf, sf, and HF of the Normal-Pareto distribution for the various parameter values.

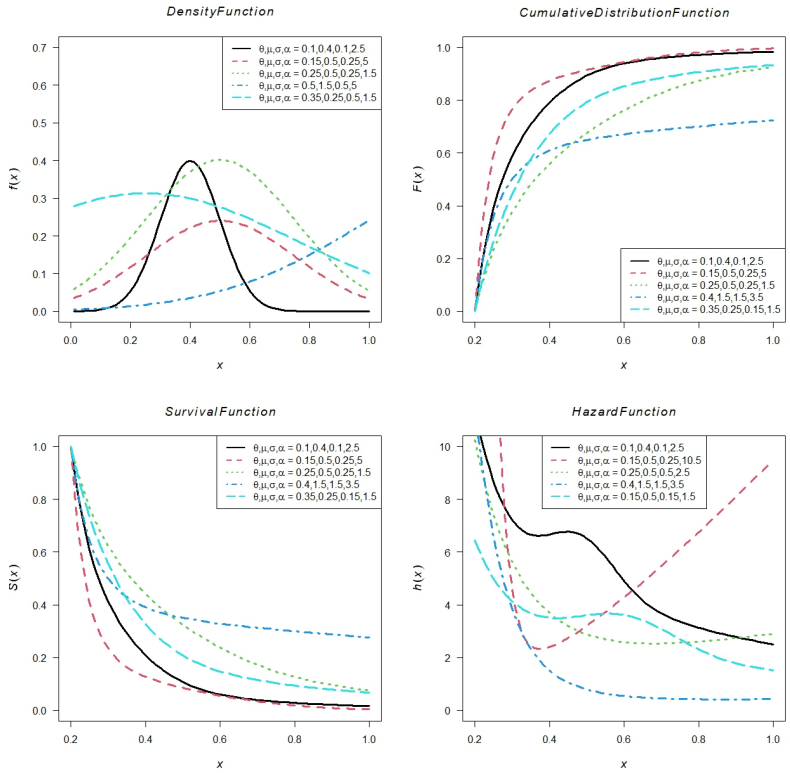


Figure 5: *The shapes of the pdf, cdf, sf, and HF of NPD for different values of the parameters*

The mean value can be calculated using the following equation.

$$Mean = \frac{2^{\frac{3}{2}}\sqrt{\pi}(\alpha\mu\theta-\mu\theta-\alpha x_m\theta+\alpha x_m)+(\alpha-1)\left(2\Gamma\left(1,\frac{(\mu-x_m)^2}{2\sigma^2}\right)\sigma-\sqrt{2}\Gamma\left(\frac{1}{2},\frac{(\mu-x_m)^2}{2\sigma^2}\right)\mu\right)\theta}{\sqrt{2}\sqrt{\pi}(\alpha-1)\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)} \tag{6}$$

The n^{th}-order statistic is given as

$$f_{X_{(n)}}(x) = n\left(\frac{2\left(\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\,\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}\right)}{\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)\theta+2}\right)\left[\frac{x^{\alpha}\left(\theta\left(erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)+erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-2\right)+2\right)+2x_m^{\alpha}(\theta-1)}{\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)x^{\alpha}}\right]^{(n-1)} \tag{7}$$

The first-order statistic is obtained as

$$f_{X_{(1)}}(x) = n\frac{2\left(\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\,\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}\right)}{\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)\theta+2}\times \left[\frac{x^{\alpha}\left(\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)-\left(\theta\left(erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)+erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-2\right)+2\right)\right)+2x_m^{\alpha}(\theta-1)}{\left(\left(erf\left(\frac{\mu-x_m}{\sqrt{2}\sigma}\right)-1\right)\theta+2\right)x^{\alpha}}\right]^{(n-1)} \tag{8}$$

We obtained the maximum likelihood estimator of parameters $(\theta,\mu,\sigma,\alpha)$ of the NPD. Consider the following log-likelihood function l of a random sample $X_1, X_2, \ldots, X_n$ from the density of NPD $(\theta,\mu,\sigma,\alpha)$ given in Equation (9).

$$l = n\ log2 - nlog\left(\theta\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)+2\right) + log\sum\left(\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\,\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}\right) \tag{9}$$

On differentiating Equation (9) with respect to the parameters θ, μ, σ, and α and equating to zero, we obtain the following likelihood equations.

$$\frac{\partial l}{\partial\theta} = \sum\left(\frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}-\sqrt{2\pi}\sigma\alpha x_m^{\alpha}x^{-\alpha-1}}{\sqrt{2\pi}\sigma\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\,\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}}\right) - \frac{n\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)}{\left(\theta\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)+2\right)} = 0 \tag{10}$$

$$\frac{\partial l}{\partial\mu} = \sum\left(\frac{\theta(x-\mu)e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2}\sqrt{\pi}\sigma^3\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\,\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}}\right) - \frac{\sqrt{2}n\theta e^{-\frac{(\mu-x_m)^2}{2\sigma^2}}}{\left(\theta\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)+2\right)\sqrt{\pi}\sigma} = 0 \tag{11}$$

$$\frac{\partial l}{\partial\sigma} = \sum\left(\frac{\theta(x-\mu)^2e^{\frac{-(x-\mu)^2}{2\sigma^2}}-\theta\sigma^2e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2}\sqrt{\pi}\sigma^4\frac{\theta}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}+(1-\theta)\,\frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}}}\right) + \frac{\left(\sqrt{2}\mu-\sqrt{2}x_m\right)n\theta e^{-\frac{(\mu-x_m)^2}{4\sigma^2}}}{\left(\theta\left(erf\left(\frac{\sqrt{2}\mu-\sqrt{2}x_m}{2\sigma}\right)-1\right)+2\right)\sqrt{\pi}\sigma^2} = 0 \tag{12}$$

and

$$\frac{\partial l}{\partial \alpha} = \sum \left(\frac{x_m^{\alpha} \ln(x_m)(1-\theta) x^{-\alpha-1}\alpha - x_m^{\alpha} \ln(x)(1-\theta) x^{-\alpha-1}\alpha + x_m^{\alpha}(1-\theta) x^{-\alpha-1}}{\frac{\theta}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + (1-\theta)\frac{\alpha x_m{}^{\alpha}}{x^{\alpha+1}}} \right) = 0 \qquad (13)$$

Now the MLEs $\hat{\theta}, \hat{\mu}, \hat{\sigma}$,and $\hat{\alpha}$ of the parameters θ, μ, σ,and α of NPD can be obtained by solving the above four likelihood equations with the help of statistical software R.

5. Simulation Study

In this section, we evaluate the performance of ML estimates using a simulation study. For this purpose, we carry out a replication of 1000 times with various sample sizes ranging from 25 to 250 for the Normal-Pareto Distribution (NPD) parameters. We created a random sample of NPD using the Monte Carlo simulation method to generate the samples with the help of R programming. For each sample, we compute the mean value, average bias, and root-mean-square error (RMSE) to assess the performance of the MLEs, and these values are presented in Table 10.

From Table 10, it is observed that the sample size of n increases, and the bias and RMSE tend to decrease. Therefore, a larger sample size indicates more accurate results.

Table 10: *Simulation analysis: Mean, Bias, and RMSE values of NPD for various sample sizes*

		Case (i): θ=0.1, μ =0.5, σ =0.7, α=0.5			**Case (ii): θ=0.5, μ =1.5, σ =0.5, α=0.1**		
n	**Parameters**	**Mean**	**Average Bias**	**RMSE**	**Mean**	**Average Bias**	**RMSE**
25	θ	0.1355	0.0645	0.0907	0.4968	0.0732	0.1079
	μ	26.5063	0.1122	2.6756	6.3569	1.0943	2.3405
	σ	0.1943	0.1941	3.1157	0.0308	0.0303	1.8590
	α	1.0902	0.2871	0.5930	1.1431	0.1124	1.6121
50	θ	0.1062	0.0619	0.0396	0.4959	0.0610	0.0931
	μ	26.6213	0.0046	0.0021	6.2519	0.0575	0.0018
	σ	0.0002	$6.25e^{-06}$	$8.77e^{-05}$	0.0004	0.0001	0.0003
	α	1.0292	0.1007	0.1891	1.0482	0.1078	0.2333
75	θ	0.0997	0.0197	0.0378	0.4908	0.0608	0.0625
	μ	26.6190	0.0015	0.0018	6.2507	0.0011	0.0015
	σ	0.0002	$3.13e^{-06}$	$7.3e^{-05}$	0.0003	0.0003	0.0002
	α	1.0172	0.0374	0.1299	1.0273	0.1011	0.1993
100	θ	0.0985	0.0185	0.0341	0.4900	0.0151	0.0515
	μ	26.6192	0.0012	0.0011	6.2506	0.0009	0.0009
	σ	0.0002	$3.12e^{-06}$	$6.85e^{-05}$	0.0003	$9.5e^{-05}$	0.0001
	α	1.0166	0.0368	0.1246	1.0229	0.0663	0.1604
250	θ	0.0900	0.0004	0.0191	0.4902	0.0098	0.0325
	μ	26.6108	0.0006	0.0007	6.2506	0.0003	0.0007
	σ	0.0002	$6.15e^{-07}$	$5.84e^{-05}$	0.0003	4.92 e^{-05}	7.8 e^{-05}
	α	1.0084	0.0285	0.0748	1.0031	0.0248	0.0932

6. Conclusion

We developed an algorithm that is a comprehensive and specifically designed framework to select the most appropriate model for asymmetric data. This framework is based on a unique combination of probability distributions, which allows us to determine the best possible mixture of probability models. To ensure that our mixture model is better than existing models, we

have used various goodness of fit tests and information criteria. Our approach utilizes a finite mixture model, which combines multiple probability models. We used the maximum likelihood estimation method to estimate the parameters of NPD. To demonstrate the effectiveness of the algorithm, we conducted an experiment where we utilized real-time data and ran it through the algorithm. This enabled us to analyze the data and come up with a new appropriate model based on the finite mixture. To further test the efficiency of the proposed model, we compared it with other models used for the same data sets available in the literature. The algorithm proposed NPD model is found most suitable in this comparison study. Finally, we have obtained the statistical characteristics of our NPD model.

References

[1] Adeyemi, A. O., Adeleke, I. A., and Akarawak, E. E. (2022). Lomax Gumble types two distributions with applications to lifetime data. *International Journal of Statistics and applied mathematics*, 7(1): 36-45.

[2] Adewara, J. A., Adeyeye, J. S., Khaleel, M. A., and Aako, O. L. (2021). Exponentiated Gompertz Exponential (EGoE) distribution: Derivation, properties, and applications. *ISTATISTIK: Journal of the Turkish Statistical Association* , 13(1), 12-28.

[3] Adeyemi, A. O., Adeleke, I. A., and Akarawak, E. E. (2022). Lomax Gumbel types two distributions with applications to lifetime data. *International Journal of Statistics and Applied Mathematics*, 7(1), 36-45.

[4] Ahmad, Z., Hamedani, G. G., and Butt, N. S. (2019). Recent developments in distribution theory: A brief survey and some new generalized classes of distributions. *Pakistan Journal of Statistics and Operation Research*, 15(1), 87-110.

[5] Akaike, H. (1973). Information theory and an extension of the Maximum Likelihood Principle. *On Information Theory (Akademia Kiado, Budapest),* 267 -281.

[6] Alhyasat, K., Kamarulzaman, I., Al-Omari, A. I., and Abu Bakar, M. A. (2020). Power size biased distribution. *Statistics in Transition new series,* 21(3), 73-91.

[7] Alshenawy, R. (2020). A new one-parameter distribution: Properties and estimation with applications to complete and type II censored data. *Journal of Taibah University of Science,* 14(1), 11-18.

[8] Al-Talib, M., Al-Nasser, A., and Ciavolino, E. (2023). Gold distribution is another look at the generalization of Lindley distribution. *Pakistan Journal of Statistics and Operation Research,* 19(2), 241-256.

[9] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron,* 7(11), 63-79.

[10] Badmus, N. I., Amusa, S. O., and Ajiboye, Y. O. (2021). Weibull - Extended Pranav distribution: application to lifetime data sets. *Unilag Journal of Mathematics and Applications,* 1(1), 104-120.

[11] Bakoban. R.A. and Hanaa H. Abu-Zinadah. (2017).. The beta generalized inverted exponential distribution with real data applications. *REVSTAT- Statistical Journal,* 15(1), 65-88.

[12] Chesneau, C. (2017). A new family of distributions based on the hypo-exponential distribution with fitting reliability data. *HAL Archives.* HAL-01519350v5

[13] El-Gohary, A., El-Bassiouny, A.H. and El-Morshedy M. (2015). Inverse Flexible Weibull Extension Distribution. *International Journal of Computer Applications,* 115(2), 46-51.

[14] El-Morshedy, M., El-Bassiouny, A. H., and El-Gohary, A. (2017). Exponentiated Inverse Flexible Weibull Extension Distribution. *Journal of Statistics Applications & Probability- An International Journal,* 6(1), 169-183.

[15] El-Morshedy, M., El-Faheem, A. A., and El-Dawoody, M. (2020). Kumaraswamy inverse Gompertz distribution: Properties and engineering applications to complete, type II right censored and upper record data. *PLoS ONE,* 15(12).

[16] El-Morshedy, M., El-Faheem, A. A., Al-Bossly, A., and El-Dawoody, M. (2021). Exponentiated generalized inverted Gompertz distribution: Properties and estimation methods with applications to symmetric and asymmetric data. *Symmetry,* 13(10), 1868.

[17] Eyob, T., Shanker, R., Shukla, K. K., and Leonida, T. A. (2019). Weighted quasi-Akash distribution: Properties and applications. *American Journal of Mathematics and Statistics,* 9(1), 30-43.

[18] Fuller, E. R., Jr., Freiman, S. W., Quinn, J. B., Quinn, G. D., and Carter, W. C. (1994). *Fracture mechanics approach to the design of glass aircraft windows: A case study.* Proceedings of SPIE 2286, Window and Dome Technologies and Materials IV.

[19] Ganaie, R. A., and Rajagopalan, V. (2022). Exponentiated Aradhana distribution with properties and applications in engineering sciences. *Journal of Scientific Research,* 6(1), 316-325.

[20] Ganaie, R. A., and Rajagopalan, V. (2020). Weighted new quasi-Lindley distribution with properties and applications. *Journal of Xi'an University of Architecture & Technology,* 12(2), 561-575.

[21] Iwok, I. A., and Nwikpe, B. J. (2021). The Iwok-Nwikpe distribution: Statistical properties and application. *European Journal of Statistics and Probability,* 10(1), 33-48.

[22] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B,* 20, 102-107.

[23] Okereke, E. W., and Uwaeme, O. R. (2018). Exponentiated Akash distribution and its applications. *Journal of the Nigerian Statistical Association,* 30.

[24] Omale, A., Asiribo, O. E., and Yahaya, A. (2019). On properties and applications of Lomax-Gompertz distribution. *Asian Journal of Probability and Statistics,* 3(2), 1-17.

[25] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society Series A,* 185, 71-110.

[26] Sakthivel, K.M. and Vidhya, G. (2024). A Systematic Procedure for Modeling Asymmetrical Data. *Advances and Applications in Statistics,* 91(10), 1241-1260.

[27] Sakthivel, K.M. and Vidhya, G. (2024). Statistical Framework for Modeling Asymmetrical data with Dual Peaks. *Indian Journal of Science and Technology,* 17(27), 2829-2840.

[28] Shanker, R., Shukla, K.K., and Mishra, A. (2017). A Three-Parameter weighted Lindley distribution and its applications to model survival time data. *Statistics in Transition- New Series,* 18(2), 291-300.

[29] Sharma, V., Shanker, R., and Shanker, R. (2019). On some one parameter lifetime distributions and their applications. *Annals of Biostatistics & Biometric Applications,* 3(2).

A FUZZY LOGIC APPROACH TO DESIGNING A DOUBLE SAMPLING PLANS FOR ZERO INFLATED POISSON DISTRIBUTION USING IN PYTHON

Kavithanjali S [1] and Sheik Abdullah A [*]

•

[1] Research Scholar, Department of Statistics, Salem Sowdeswari College, Salem-10, India
[*] Assistant Professor , Department of Statistics, Salem Sowdeswari College, Salem-10, India
kavithanjalis2018@.com, sheik.stat@gmail.com

Abstract

Acceptance sampling plan by attributes is a statistical measure used in quality control in various production process. It is mainly determined for identifying whether the lot or the batch of the product is accepted or rejected based on the number of defective items in the sample. Appropriate sampling plan provides defect-free lot. There are several sampling plans are available for determine the sample size. Among the sampling plan, double sampling plan is more effective because it is always giving best result in lot selection compared with other sampling plan. In most of the practical situation, it is very hard to found the product as strictly defective or non-defective. In some situation, quality of the product can be classified several types which are expressed as good, almost good, bad not so bad and so on. This causes ambiguity deficiency in proportion value of lot or process. In mathematical tools, fuzzy set or fuzzy logic is one of the powerful modeling, which has incomplete and imprecise information. The fuzzy set theory is adopted to cope the vagueness in these linguistic expressions for the accepting sampling. In this article double sampling plans, are determined when non-conformities are fuzzy number and being modeled based on Zero-Inflated Poisson (ZIP) distribution. The Operating Characteristic (OC) function and Average Sample Number (ASN) function are evaluated both numerically and graphically in fuzzy and crisp environments.

Keywords: Acceptance double sampling plan, Fuzzy OC curve, Fuzzy average sample number, ZIP distribution, Fuzzy parameter.

1. Introduction

The present research deals with quality control issues raised by the ZIP distribution, an elaborate statistical model used in industrial operations. Traditional sample tactics, although effective for well-behaved distributions, may struggle with the zero-inflated character of certain datasets. To cope with this, we present a unique technique which integrates fuzzy logic into the construction of a double sampling plan optimised for ZIP distribution. Fuzzy logic, popular because of its adaptability in coping with imprecision and uncertainty, strengthens the plan by simulating the unclear boundary between zero and non-zero occurrences in the ZIP distribution. This study expands on the fundamental work of Zadeh and Kosko (1965) in fuzzy logic and is consistent with recent research in quality control systems. This study has a substantial impact on industry, helping to develop quality control techniques. Our objective is to provide a robust double sampling strategy that combines fuzzy logic with ZIP distribution complexities in order to enhance decision-making and consistently detect deviations from quality requirements in an industrial context where technological breakthroughs may result in occasional zero defects. ZIP

distribution used in many of fields such as agriculture, epidemiology, econometrics, public health, process control, medicine, and manufacturing.

Numerous scenarios, such as Bohning, Dietz, and Schlattmann [1] use the ZIP model on dental epidemiology data to examine excess zeros in intervention effects by ZIP regression. Fuzzy probability is examined by Buckley [2], who introduced new methods of probability theory in the environment of uncertainty. Chakraborty [3] introduces a fuzzy optimization technique for single sampling plans that minimizes inspection while managing customer risk by utilizing a Poisson distribution. Duncan AJ [4] Quality Control and Industrial Statistics is a most impact guide that combines statistical methods with practical approaches to quality control in industrial circumstances.Ezzatallah and Gildeh [5] suggest a fuzzy Poisson-based acceptance double sampling strategy for the management of unpredictable defective proportions.Janani K, Vignesh A and et al.,[6] . introduce a novel fuzzy set-based tactics for identify and conserve endemic plant species in the Nilgiris Biosphere Reserve. The work effectively addresses multi-criteria decision-making by leveraging advanced fuzzy operators, which provide improved representation of ambiguity compared to previous study's. Kavithanjali, Sheik Abdullah, and Kamalanathan [7] review SQC methodologies in single and double-sampling plans, pointing at possible effects on quality.Kavithanjali and Sheik Abdullah [8] present a inovative research by integrates of fuzzy logic ZIP distribution for SSPs, managing quality control and risk in uncertainity distribution plan. Kaviyarasu and Asif T Thottathil [9] deals the application of Zero-Inflated Poisson distribution in designing optimal acceptance sampling plans for quality control in manufacturing with a focus on special type double sampling plans.LA Zadeh [10] presents fuzzy sets and the degree of membership, which lays the groundwork for the employment of the conventional theory of sets in fuzzy control.Lambert [11] shows that ZIP regression can be employed for better data analysis in manufacturing by handling excess zeros in count data. Malathi and Muthulakshmi [12] initiate an inquiry into fuzzy logic in double-sampling plans to deal with ambiguity in quality assessments. McLachlan and Peel [13] provide a detailed reference on finite mixture models, which is vital for the analysis of complex data and heterogeneous populations. Naya, Urioste, and Chang [14] employ ZIP models, demonstrating that age is a significant factor in the occurrence of black patches. Ridout, Demetrio, and Hinde [15] provide practical horticulture examples to evaluate models for excess zeros in count data.Schilling and Neubauer [16] provide a comprehensive and authoritative guide on acceptance sampling plan, offering useful insights for quality control in numerous industries. In the context of statistical process control, Xie, He, and Goh [17] establish the ZIP distribution's superiority over the Poisson distribution for over-dispersed data.

In the following sections, we will explain the methods we used, show our results, and talk about why these findings are important for experts working in quality control. We did our analysis using Python and powerful libraries like NumPy, Pandas, SciPy, and Matplotlib to help with statistics and data visualization. We believe this research will help improve methods used to solve problems caused by difficult distributions in industrial environments.

2. Methodology

2.1. Fundamental Definitions

2.1.1 Fuzzy Number: A fuzzy number $(\tilde{N})$ is a fuzzy set on the real line R, characterized by a membership function $\mu_N : R \to [0,1]$, that satisfies the following conditions:

- $(\tilde{N})$ is normal, meaning there exists some x such that $\mu_N(x) = 1$.
- $(\tilde{N})$ is convex, meaning for any $x_1, x_2 \in R$ and $\lambda \in [0,1]$, $\mu_N(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_N(x_1), \mu_N(x_2))$.
- The membership function μ_N is upper semi-continuous, meaning the set $\{x \in R \mid \mu_N(x) \geq \alpha\}$ is closed for every $\alpha \in (0,1]$.
- The support of $(\tilde{N})$, defined as $\text{Supp}(\tilde{N}) = \{x \in R \mid \mu_N(x) > 0\}$, is bounded.

2.1.2 Triangular Fuzzy Number: A triangular fuzzy number $\tilde{N}$ is defined by a triplet (a, b, c), where $a < b < c$. The membership function $\mu_{\tilde{N}}(x)$ is given by:

$$\mu_{\tilde{N}}(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \leq x \leq b, \\ \frac{c-x}{c-b} & \text{if } b < x \leq c, \\ 0 & \text{otherwise.} \end{cases}$$

This function forms a triangular shape with $[a, c]$ as the base and the peak at $x = b$.
2.1.3 α-Cut of Fuzzy t: The α-cut of a fuzzy set $\tilde{N}$ is a crisp set of values where the membership function is at least α. It is defined as:

$$N[\alpha] = \{x \in R \mid \mu_N(x) \geq \alpha\}.$$

The fuzzy number $\tilde{N}[\alpha]$ can be represented by its lower and upper bounds as $N^L[\alpha]$ and $N^U[\alpha]$, where:

$$N^L[\alpha] = \inf\{x \in R \mid \mu_N(x) \geq \alpha\},$$
$$N^U[\alpha] = \sup\{x \in R \mid \mu_N(x) \geq \alpha\}.$$

2.1.4 ZIP Distribution: The Zero-Inflated Poisson (ZIP) distribution, define as ZIP(φ, λ), is used when there is an more number of zero counts. The probability mass function (p.m.f.) is found in Lambert [11] and Mclachlan [13] :

$$P(D = d \mid \varphi, \lambda) = \begin{cases} \varphi + (1-\varphi)e^{-\lambda} & \text{if } d = 0, \\ (1-\varphi)\frac{e^{-\lambda}\lambda^d}{d!} & \text{if } d = 1, 2, \ldots \end{cases}$$

In this distribution:

- φ represents the probability of extra zeros.
- λ is the mean of the underlying Poisson distribution.

The ZIP distribution mean is $(1-\varphi)\lambda$, and the variance is $\lambda(1-\varphi)(1+\varphi\lambda)$.
To extend the ZIP distribution to a fuzzy setting, we replace λ with a fuzzy number $\tilde{\lambda} > 0$. The fuzzy p.m.f. can be represented as:

$$\tilde{P}(d \mid \alpha) = \begin{cases} \varphi + (1-\varphi)e^{-\lambda} & \text{if } d = 0, \\ (1-\varphi)\frac{e^{-\lambda}\lambda^d}{d!} & \text{if } d = 1, 2, \ldots \end{cases}$$

where λ belongs to the α-cut of $\tilde{\lambda}$.

2.2. Python Programming

In this present study on statistical quality control, we used Python programming to create the upper and lower bounds of the Fuzzy Operating Characteristic (OC) Band and the Fuzzy Average Sample Number (ASN) tables. Moreover, we created graphs to visualize the Fuzzy OC, the fuzzy probability of acceptance, and the average sample number curves. Python's extensive analytical abilities and versatile libraries make it straightforward to implement these statistical methods within our study methodology.

3. Operating procedure for DSP under ZIP distribution conditions

Let us consider a circumstance where we analyse the N- lot size for defects with Zero-Inflated Poisson (ZIP) distribution. These are general steps of the typical double sampling plan.
Step 1:

- Take a random sample of size n_1 and count the number of defective items (D_1).
- c_1 is the acceptance number for the first sample.
- c_2 is the acceptance number for both combined samples.

Step 2:

- Accept the lot if $D_1 \leq c_1$.
- Reject the lot if $D_1 > c_2$.
- If $c_1 < D_1 \leq c_2$, proceed to Step 3.

Step 3:

- Take a random sample from second sample n_2 and count the number of defective items (D_2).
- Add D_1 and D_2 together.
- Accept the lot if $D_1 + D_2 \leq c_2$, otherwise reject it.

Step 4:

- The random variables D_1 and D_2 follows the ZIP distribution with parameter $\lambda_1 = n_1 p$ and $\lambda_2 = n_2 p$, given a large sample size and a small probability p.
- Let P_a stand for the acceptance probability of the lot based onto the combined samples.
- $\tilde{P}_a^I$ is for the acceptance probability after the first sample and $\tilde{P}_a^{II}$ for the second sample.

Thus, the overall probability of acceptance is:

$$\tilde{P}_a = \tilde{P}_a^I \cdot \tilde{P}_a^{II}$$

Using the ZIP distribution pmf, the number of nonconforming items in the lot is given by

$$\tilde{\mathrm{P}}\left(\mathrm{D} = \mathrm{d} \mid \varphi, \lambda\right) = \tilde{\mathrm{P}}\left(\mathrm{d}\right) = \begin{cases} \varphi + (1-\varphi)\, e^{-\lambda}, & \text{When d} = 0 \\ (1-\varphi)\, \frac{e^{-\lambda}\lambda^d}{d!}, & \text{When d } = 1, 2, \ldots,\ 0 < \varphi < 1,\ \lambda > 0 \end{cases}$$

Given a sample size of n_1, the probability of finding no deficiencies will be

$$\tilde{P}\left(D = 0\right) = \tilde{P}_a^I\left(\alpha\right) = \varphi + (1-\varphi)\, e^{-n_1 p} \tag{1}$$

Given a sample size of n_2, the probability of finding one deficiencies will be

$$\tilde{P}\left(D = 1, D_1 + D_2 \leq 1\right) = \tilde{P}_a^{II}\left(\alpha\right) = (1-\varphi)\, \mathrm{e}^{-np}\, n_2 p \tag{2}$$

From a sample of size n_1 the probability of finding one or less defects will be

$$\tilde{P}\left(D \leq 1\right) = \varphi + (1-\varphi)\, e^{-n_1 p}\left(1 + n_1 p\right) \tag{3}$$

From a sample of size n_2 the probability of finding one or more defects will be

$$\tilde{P}_a^{II}\left(\alpha\right) = (1-\varphi)\, e^{-np}\, (0.5)(n_1 p)^2 \tag{4}$$

A DSP only accepts a lot if a sample of size n_1 has no faults and a sample of size n_2 has one defect or less. Thus, DSP's $\tilde{P}_a\left(\alpha\right)$ will be provided by

$$\tilde{P}_a\left(\alpha\right) = P_a^I\left(\alpha\right) + P_a^{II}\left(\alpha\right) \tag{5}$$

3.1. Numerical illustration 1

Consider that $\tilde{P} = (0.01,\ 0.02,\ 0.03)$, N=200, n_1=10, n_2=10, c_1= 0 ,c_2=1, $\tilde{\lambda} = n\tilde{p}$, $\varphi = 0.0001$,n=$n_1 + n_2$.

$$\tilde{\mathrm{P}}\,[\alpha] = [0.01 + 0.01\alpha\,,\ 0.03 - 0.01\alpha]$$

From equation(1) the fuzzy probability of a sample of size n_1 having no faults is thus as follows:

$$\tilde{P}\,(D=0) = \tilde{P}_a^I\,(\alpha) = \varphi + (1-\varphi)\ \mathrm{e}^{-10p}$$

$$\tilde{\mathrm{P}}_a^I\,[\alpha] = \left\{\varphi + (1-\varphi)\left(e^{-(0.3-0.1\alpha)}\right),\ \varphi + (1-\varphi)\left(e^{-(0.1+0.1\alpha)}\right)\right\}$$

From equation (2) the fuzzy probability of a sample of size n_2 having one fault is thus as follows:

$$\tilde{P}(D=1, D_1+D_2\leq 1) = \tilde{P}_a^{II}(\alpha) = (1-\varphi)e^{-20p}\cdot 10p$$

$$\tilde{\mathrm{P}}_a^{\mathrm{II}}[\alpha] = \left\{(1-\varphi)\,e^{-(0.2+0.2\alpha)}\,(0.1+0.1\alpha)\,,(1-\varphi)\,e^{-(0.6-0.2\alpha)}\,(0.3-0.1\alpha)\right\}$$

From equation (5) a DSP only accepts a lot if a sample of size n_1 has no faults and a sample of n_2 has one defect or less. Thus, DSP's $\tilde{P}_a\,(\alpha)$ will be provided by

$$\begin{aligned}\tilde{\mathrm{P}}_a\,[\alpha] &= \tilde{P}_a^I\,(\alpha) + \tilde{P}_a^{II}\,(\alpha)\\ &= \Big\{\varphi + (1-\varphi)\,e^{-(0.3-0.1\alpha)} + (1-\varphi)\,e^{-(0.6-0.2\alpha)}\,(0.3-0.1\alpha)\,,\\ &\quad \varphi + (1-\varphi)\,e^{-(0.1+0.1\alpha)} + (1-\varphi)\,e^{-(0.2+0.2\alpha)}\,(0.1+0.1\alpha)\ \}\end{aligned}$$

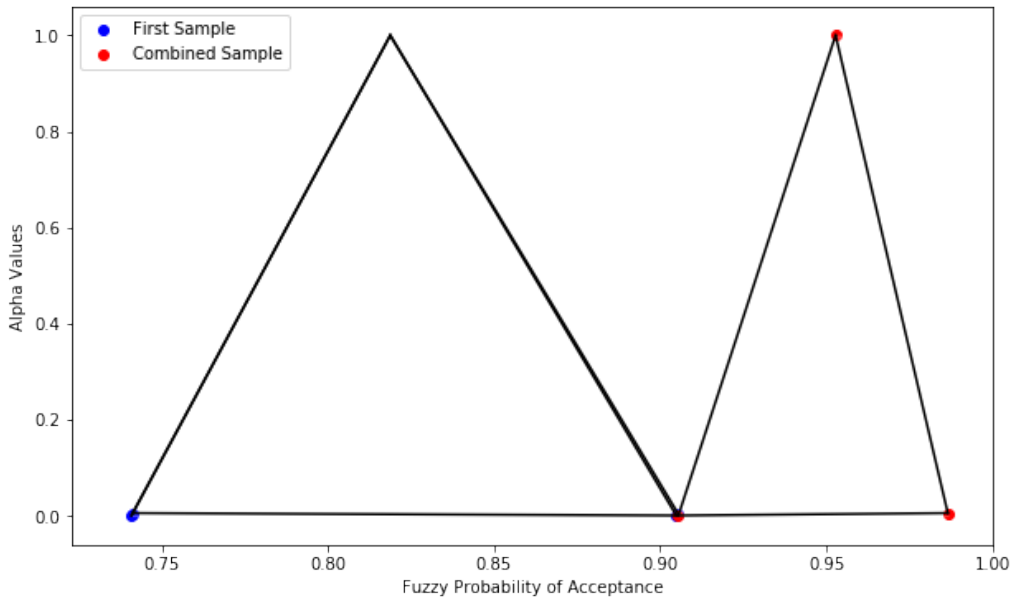


Figure 1: *Fuzzy probability of acceptance with DSPs under ZIP distribution*

Table 1: *Fuzzy probability of acceptance with DSP* $\varphi = 0.0001, \alpha$= *0,0.005,1*

α	$\tilde{P}_a^I$	$\tilde{P}_a^{II}$	$\tilde{P}_a[\alpha]$
0.000	[0.740844, 0.904847]	[0.081865, 0.164627]	[0.905471, 0.986712]
0.005	[0.741215, 0.904395]	[0.082192, 0.164517]	[0.905732, 0.986587]
1.000	[0.818749, 0.818749]	[0.134051, 0.134051]	[0.952799, 0.952799]

As in Figure 1 and Table 1, it shows that 98 to 100 out of every 100 lots accepted in this process.

4. Fuzzy OC curve for DSPs under ZIP distribution

The fuzzy parameter is applied to construct the operating characteristic (OC) curve for the DSP. This curve demostrate the relationship between the fraction of nonconforming items, denoted as p, and the acceptance probability is $P_a(p)$. The OC curve helps differentiate between good and bad lots in the sampling plan. Producer's risk occurs when a customer rejects a product that actually meets the quality standards (i.e., the product is of good quality). On the other hand, consumer's risk happens when a customer accepts a product that fails to meet the standards (i.e., the product quality is poor). A fuzzy parameter with upper and lower bounds can be used to estimate the defective fraction. If these bounds are equal, the process is considered to be in an optimal state.

4.1. Numerical illustration 2

Consider that, $\tilde{P} = (0.01,\ 0.02,\ 0.03)$, $\varphi = 0.0001$, n_1=10, n_2=5, c_1=1, c_2=2,$n = n_1 + n_2$, b_2=0.01, b_3=0.02 following that we have, $\tilde{P}\,[\alpha] = [k + 0.01\alpha,\ 0.02 + k - 0.01\alpha]$, $0 \le k \le 0.98$.
Accordingly, the first sample fuzzy probability of acceptance using equation (3) is

$$\tilde{P}_a^I(\alpha) = \tilde{P}[0,1][\alpha] = [\varphi + (1-\varphi)e^{-10p}(1+10p)], f_1(p) = [\varphi + (1-\varphi)e^{-10p}(1+10p)]$$

Since is decreasing, as follows:

$$\tilde{P}_a^I[0] = \left[\varphi + (1-\varphi)\, e^{-(10k+0.2)}\ \ (1 + (10k+0.2))\,,\ \varphi + (1-\varphi)\, e^{-(10k)}\ \ (1+10k)\right]$$

and accordingly the second sample's fuzzy probability of acceptance from equation (4) is

$$\tilde{P}_a^{II}(\alpha) = \tilde{P}(D_1 = 2,\ D_2 = 0) = \left[(1-\varphi)\, e^{-15p}\ 50p^2\right], f_2\,(p) = \left[(1-\varphi)\, e^{-15p}\ 50p^2\right]$$

$$\tilde{P}_a^{II}\,(0) = \left[(1-\varphi)\, e^{-15k}\ 50\ (k^2),\ (1-\varphi)\, e^{-15(k+0.02)}\ 50(k+0.02)^2\ \right]\ \ 0 \le k \le 0.98$$

Next, we will obtain the α-cut by examining the $f_2(p)$ function in the following manner:

$$\tilde{P}_a^{II}\,(\alpha) = \begin{cases} (1-\varphi)\, e^{-15k}\ 50\ k^2,\ (1-\varphi)\, e^{-15(k+0.02)}\ 50(k+0.02)^2 & , 0 \le k < \frac{1.7}{15} \\ (1-\varphi)\, e^{-15k}\ 50\ k^2\ ,\ 0.12 & , \frac{1.7}{15} \le k < 0.12 \\ (1-\varphi)\, e^{-15(k+0.02)}\ 50(k+0.02)^2,\ 0.12 & , 0.12 \le k < \frac{2}{5} \\ (1-\varphi)\, e^{-15(k+0.02)}\ 50(k+0.02)^2\,, (1-\varphi)\, e^{-15k}\ 50\ k^2 & , \frac{2}{5} \le k \le 0.98 \end{cases}$$

$$\tilde{P}_a[\alpha] = \tilde{P}_a^I\,(0) + \tilde{P}_a^{II}\,(0)$$

Table 2: *Sample values at different k levels with acceptance DSP using ZIP distribution*

k	$\tilde{P}_a^I$	$\tilde{P}_a^{II}$	$\tilde{P}_a(\alpha)$
0.00	[0.982479, 1.000000]	[0.000000, 0.014815]	[0.997294, 1.000000]
0.01	[0.963067, 0.995322]	[0.004303, 0.028690]	[0.991758, 0.999625]
0.02	[0.938454, 0.982479]	[0.014815, 0.043901]	[0.982355, 0.997294]
0.03	[0.909805, 0.963067]	[0.028690, 0.059040]	[0.968845, 0.991758]
0.04	[0.878111, 0.938454]	[0.043901, 0.073175]	[0.951286, 0.982355]
0.05	[0.844211, 0.909805]	[0.059040, 0.085726]	[0.929937, 0.968845]
0.06	[0.808811, 0.878111]	[0.073175, 0.096373]	[0.905184, 0.951286]
0.07	[0.772505, 0.844211]	[0.085726, 0.104982]	[0.877487, 0.929937]

k	$\tilde{P}_a^I$	$\tilde{P}_a^{II}$	$\tilde{P}_a(\alpha)$
0.08	[0.735785, 0.808811]	[0.096373, 0.111554]	[0.847339, 0.905184]
0.09	[0.699059, 0.772505]	[0.104982, 0.116179]	[0.815238, 0.877487]
0.10	[0.662647, 0.735785]	[0.111554, 0.118965]	[0.781683, 0.847339]
0.11	[0.626792, 0.699059]	[0.116179, 0.120058]	[0.747170, 0.815238]
0.12	[0.591660, 0.662647]	[0.118965, 0.119622]	[0.712177, 0.781683]
0.13	[0.557353, 0.626792]	[0.120058, 0.117834]	[0.677139, 0.747170]
0.14	[0.523917, 0.591660]	[0.119622, 0.114881]	[0.642457, 0.712177]
0.15	[0.491356, 0.557353]	[0.117834, 0.110947]	[0.608493, 0.677139]
0.16	[0.462891, 0.524978]	[0.116107, 0.108862]	[0.571753, 0.641086]
0.17	[0.433806, 0.493296]	[0.112817, 0.104399]	[0.538204, 0.606113]
0.18	[0.406065, 0.462891]	[0.108862, 0.099564]	[0.505629, 0.571753]
0.19	[0.379677, 0.433806]	[0.104399, 0.094479]	[0.474156, 0.538204]
0.20	[0.354635, 0.406065]	[0.099564, 0.089248]	[0.443883, 0.505629]
0.21	[0.330921, 0.379677]	[0.094479, 0.083959]	[0.414880, 0.474156]
0.22	[0.308510, 0.354635]	[0.089248, 0.078684]	[0.387195, 0.443883]
0.23	[0.287369, 0.330921]	[0.083959, 0.073486]	[0.360854, 0.414880]
0.24	[0.267458, 0.308510]	[0.078684, 0.068411]	[0.335869, 0.387195]
0.25	[0.248736, 0.287369]	[0.073486, 0.063498]	[0.312234, 0.360854]
0.26	[0.231155, 0.267458]	[0.068411, 0.058777]	[0.289932, 0.335869]
0.27	[0.214669, 0.248736]	[0.063498, 0.054268]	[0.268937, 0.312234]
0.28	[0.199228, 0.231155]	[0.058777, 0.049985]	[0.249214, 0.289932]
0.29	[0.184783, 0.214669]	[0.054268, 0.045939]	[0.230722, 0.268937]
0.30	[0.171284, 0.199228]	[0.049985, 0.042132]	[0.213416, 0.249214]
0.31	[0.158682, 0.184783]	[0.045939, 0.038565]	[0.197247, 0.230722]
0.32	[0.146928, 0.171284]	[0.042132, 0.035236]	[0.182163, 0.213416]
0.33	[0.135975, 0.158682]	[0.038565, 0.032138]	[0.168112, 0.197247]
0.34	[0.125777, 0.146928]	[0.035236, 0.029265]	[0.155041, 0.182163]
0.35	[0.116289, 0.135975]	[0.032138, 0.026607]	[0.142896, 0.168112]
0.36	[0.107469, 0.125777]	[0.029265, 0.024155]	[0.131624, 0.155041]
0.37	[0.099275, 0.116289]	[0.026607, 0.021899]	[0.121175, 0.142896]
0.38	[0.091669, 0.107469]	[0.024155, 0.019828]	[0.111497, 0.131624]
0.39	[0.084612, 0.099275]	[0.021899, 0.017933]	[0.102542, 0.121175]
0.40	[0.078069, 0.091669]	[0.019828, 0.016195]	[0.094264, 0.111497]
0.41	[0.072006, 0.084612]	[0.017930, 0.014610]	[0.086617, 0.102542]
0.42	[0.066391, 0.078069]	[0.016195,0.013167]	[0.079558, 0.094264]

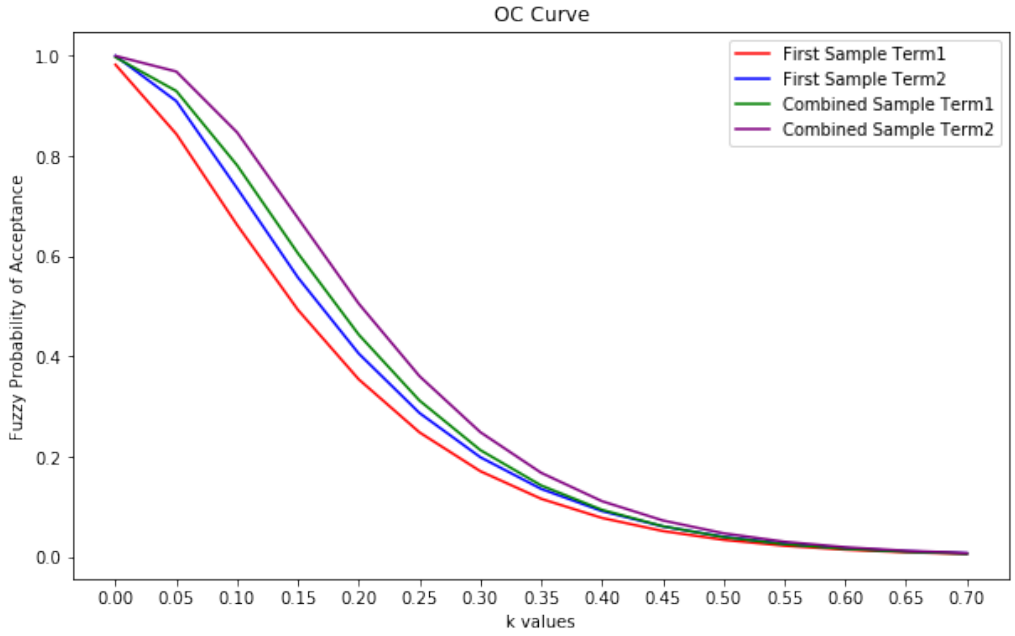


Figure 2: *FOC band with DSP using ZIP distribution* $n_1 = 10, c_1 = 1, n_2 = 5, and c_2 = 2$

The instance shown in Figure 2 is what happens when process quality drops from a perfect condition to a modest state, at which point the FOC band widens. Table 2 will be computed by simplifying and using fuzzy arithmetic.

5. Fuzzy Average Sample Number (FASN) with DSPs using ZIP distribution

The first benefit of a DSP is that it requires a small average sample size to make a sensible choice. In the context of this technique, the sample number is determined by either n_1 or n_2, which must add up to n_1+n_2. When deciding on the first sample (with fuzzy probability $\tilde{P}$ ($D_1 \leq c_1$ or $D_1 > c_2$,)) the $\tilde{P}^I$describes the fuzzy probability of drawing the first sample. If the first sample gives an uncertain result $\tilde{P}(c_1 < D_1 \leq c_2)$, a second sample with a total population of n_1+n_2 is required. This is represented by the fuzzy probability and will be referred to as $\tilde{P}^{II}$. The fuzzy mean formula is used to determine the FASN: $FASN = \tilde{\mu}_{SN}(\alpha) = \{n_1 p_1 + (n_1 + n_2) p_{II}\}$. $P_i \in \tilde{P}_i(\alpha), i = I, II, P_I + P_{II} = 1$ Thus, we obtain $FASN = \{n_1 + n_2 p_{II}\}$

The following illustration is based on FASN using illustration I
$\tilde{P} = (0.01,\ 0.02,\ 0.03)$,N=200, n_1=10, n_2=10 ,c_1= 0 ,c_2=1, $\tilde{\lambda} = np$, $\varphi = 0.0001$, $n = n_1 + n_2$

$$\tilde{P}(\alpha) = [0.01 + 0.01\alpha,\ 0.03 - 0.01\alpha]$$

$$FASN = \left\{10 + 10\left[(1-\varphi)e^{-10p}\, 10p\right]\right\}$$

$$FASN = \left\{10 + (1-\varphi)\ e^{-10p}\, 100p\right\}$$

$$FASN(\alpha) = \left\{10 + (1-\varphi)\ e^{-(0.1+0.1\alpha)}\ (1+\alpha)\ ,10 + (1-\varphi)\ e^{-(0.3-0.1\alpha)}\ (3-\alpha)\right\}$$

Under$\alpha = 0$ we gain $FASN(0) = 10.90, 12.22$

The figure 3 shows triangular fuzzy graph illustrates how the FASN adapts to varying degrees of ambiguity, illustrating the adaptability nature of the sampling plan under different fuzzy probability conditions.The Average Sample Number (ASN) curve in double sampling plans is

used to assess sampling efficiency and resource optimisation. It helps to compare different plans and get insights into sampling performance under varied lot characteristics. ASN curves help in decision-making by indicating the predicted number of samples needed for acceptance or rejection at various quality standards. This ensures balanced inspection costs and quality control. and the sample size relies on whether or not a second sample is needed. We may depict the FASN band in terms of $\tilde{P}$, the fuzzy proportion faulty in an entering lot, using the $\tilde{P}$ structure that was developed.

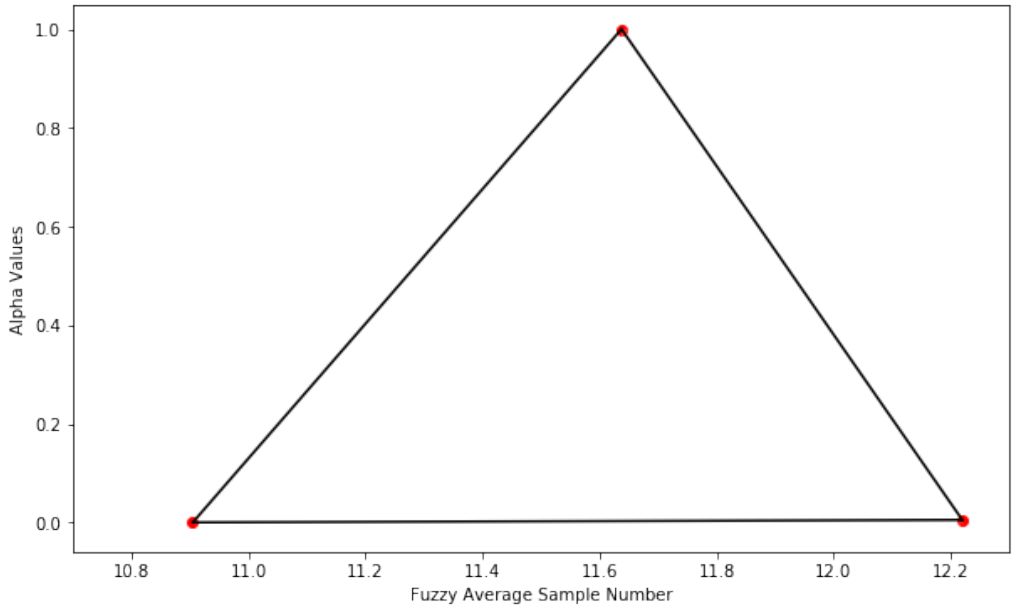


Figure 3: *FASN for a DSP with using ZIP distribution $c_1 = 0$, $c_2 = 1$, $n_1 = n_2 = 10$*

5.1. Numerical illustration 3

Let that $\varphi = 0.0001$, $\mathbf{c}_1$= 0, $\mathbf{c}_2$ = 1, N=200, n_1 = n_2 =10 and b2 = 0.01, b3 = 0.02. Then FASN is obtained as follows

$$\tilde{P}(\alpha) = [k + 0.01\alpha\,,\, k + 0.02 - 0.01\alpha]$$

$$FASN = \left\{10 + (1-\varphi)\, e^{-10p}\, 100p\right\}$$

And α-cut of FASN is:

$$FASN(\alpha) = \begin{cases} FASN^{*},\, FASN^{**} & ,0 \le k < 0.08 \\ FASN^{*},\, 13.68^{**} & ,0.08 \le k < 0.09 \\ FASN^{**},\, 13.68 & ,0.09 \le k < 0.1 \\ FASN^{**},\, FASN^{*} & ,0.1 \le k < 0.98 \end{cases}$$

$$FASN^{*} = \left\{10 + (1-\varphi)\; e^{-(10k+0.1\alpha)} 100\,(k + 0.01\alpha)\;\right\} \text{ and}$$

$$FASN^{**} = \left\{10 + (1-\varphi)\; e^{-(10k+0.2-0.1\alpha)} 100\,(k + 0.02 - 0.01\alpha)\;\right\}$$

Figure 4 and table 3 shows the FASN band, the first sample will determine whether the batch is accepting or reject, depending on how good the process is executed. This implies that the sample number will be less, and if the process quality is average, most of the time, while selecting whether to accept or reject the lot, a second sample should be picked, increasing the sample size.

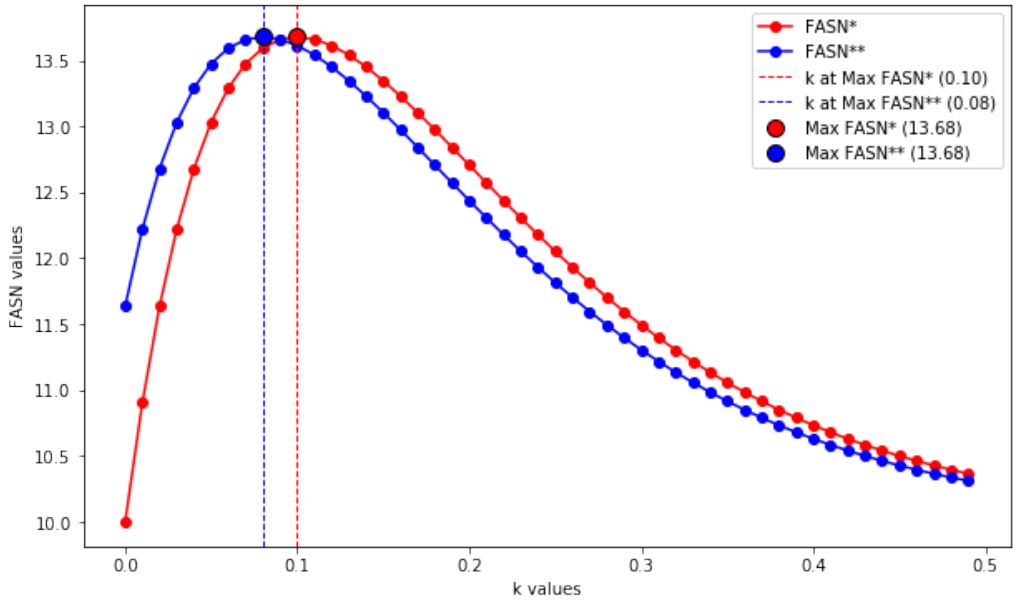


Figure 4: *FASN band for a DSP with using ZIP distribution of $c_1 = 0$, $c_2 = 1$, $n_1 = n_2 = 10$*

Table 3: *FASN for a DSP with using ZIP distribution* $c_1 = 0, c_2 = 1, n_1 = n_2 = 10$

k	FASN*	FASN**	k	FASN*	FASN**
0.00	10.0000	11.6373	0.21	12.5713	12.3057
0.01	10.9047	12.2222	0.22	12.4374	12.1770
0.02	11.6373	12.6810	0.23	12.3057	12.0519
0.03	12.2222	13.0324	0.24	12.1770	11.9309
0.04	12.6810	13.2925	0.25	12.0519	11.8144
0.05	13.0324	13.4757	0.26	11.9309	11.7025
0.06	13.2925	13.5943	0.27	11.8144	11.5955
0.07	13.4757	13.6588	0.28	11.7025	11.4935
0.08	13.5943	13.6784	0.29	11.5955	11.3964
0.09	13.6588	13.6612	0.30	11.4935	11.3043
0.10	13.6784	13.6140	0.31	11.3964	11.2170
0.11	13.6612	13.5426	0.32	11.3043	11.1346
0.12	13.6140	13.4520	0.33	11.2170	11.0568
0.13	13.5426	13.3466	0.34	11.1346	10.9836
0.14	13.4520	13.2300	0.35	11.0568	10.9147
0.15	13.3466	13.1053	0.36	10.9836	10.8500
0.16	13.2300	12.9751	0.37	10.9147	10.7894
0.17	13.1053	12.8415	0.38	10.8500	10.7326
0.18	12.9751	12.7064	0.39	10.7894	10.6794
0.19	12.8415	12.5713	0.40	10.7326	10.6298
0.20	12.7064	12.4374	0.41	10.6794	10.5834

6. Conclusion

In this study, we proposed a fuzzy ZIP distribution-based technique for developing acceptance double sampling plans (DSP) using fuzzy features. These plans are clearly defined since the results are consistent with classical plans when the proportion of damaged items is sharp. The primary parameters of the DSP, proportion defective, and sample size, are considered a triangle fuzzy number. With these parameters, the fuzzy operating characteristic and average sample number curves are generated and given in this paper. It was demonstrated that the plans OC and ASN curves resemble a band lower boundaries. Under this approach, FASN will have a lesser value depending on how perfect or inadequate the process quality.

References

[1] Bohning, D., Dietz, E., and Schlattmann, P. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society*, 162(2):195–209.

[2] Buckley, J. J. (2003). Fuzzy probability: new approach and application. Heidelberg, Germany.

[3] Chakraborty, T. K. (1994). A class of single sampling plan based on fuzzy optimization. *Opsearch*, 63(1):35–43.

[4] Duncan, A. J. (1986). *Quality Control and Industrial Statistics*. Richard D. Irwin, Inc.

[5] Ezzatallah, B., and Bahram, S. G. (2012). Acceptance double sampling plan using fuzzy Poisson distribution. *World Applied Sciences Journal*, 16(11):1578–1588.

[6] Janani, K., Vignesh, A., Veerakumari, P., Vasanth, K., and Rakkiyappan, R. (2023). Identifying native endemic plant species in Nilgiris using the interval type 2 q-rung orthopair fuzzy Bonferroni mean operator. *Computational and Applied Mathematics*, 42:55.

[7] Kavithanjali, S., Sheik Abdullah, A., and Kamalanathan, R. (2024a). A review study on various sampling plans. *Journal of Inventive and Scientific Research Studies*, 1(2):15–28.

[8] Kavithanjali, S., and Sheik Abdullah, A. (2024b). Python implementation of fuzzy logic for zero-inflated Poisson single sampling plans. *Reliability: Theory & Applications*, 3(79):275–281.

[9] Kaviyarasu, V., and Thottathil, A. T. (2018). Designing STDS plan for zero-inflated Poisson distribution through various quality level. *International Journal of Statistics and Applied Mathematics*, 3(4):44–53.

[10] Lotfi A Zadeh. (1965). Fuzzy sets. *Information and Control*, 38(4):338–353.

[11] Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

[12] Malathi, and Muthulakshmi. (2012). Special double sampling plan with fuzzy parameter. *Indian Journal of Applied Research*, 2(1):117–120.

[13] McLachlan, G., and Peel, D. *Finite Mixture Models*. John Wiley and Sons.

[14] Naya, H., Urioste, J. I., Chang, Y. M., Motta, M. R., Kremer, R., and Gianola, D. (2008). A comparison between Poisson and zero-inflated Poisson regression models with an application to the number of black spots in Corriedale sheep. *Genetics Selection Evolution*, 40(1):379–394.

[15] Ridout, M., Demtrio, C. G. B., and Hinde, J. (1998). Models for count data with many zeros. In *International Biometric Conference*, Cape Town.

[16] Schilling, E. G., and Neubauer, D. V. (2009). *Acceptance Sampling in Quality Control*. Boca Raton: CRC Press.

[17] Xie, M., He, B., and Goh, T. N. (2001). Zero-inflated Poisson model in statistical process control. *Computational Statistics & Data Analysis*, 38(2):191–201.

CHARACTERISTIC FEATURES OF CONTROL METHODS IN ELECTROMECHANICAL DEVICES

G.V.Mamedova
•
Azerbaijan State Oil and Industry University, Baku, Azerbaijan
gulschen98@mail.ru

Abstract

Modern electromechanical devices for continuous process control are used in a variety of industrial applications. A control system or electromechanical device used to control the speed and torque of AC motors by varying the frequency and supply voltage converts alternating current of one frequency to alternating current of another frequency. The power section and the control device are the main elements of the control system. The main elements of a control system or electromechanical device are the power part (electrical energy converter) and the control device (controller). Modern frequency converters have a modular architecture, which expands the capabilities of the device, and also, in most cases, allows the installation of additional interface modules for input-output channels. The control device (microcontroller) is controlled by software and is controlled by the main parameters (speed or torque).

Key words: electromechanical device, control system, frequency converter, control device, power part, modulation method, output voltage, three-phase autonomous voltage inverter, semiconductor device, analog-to-digital converter, controller, comporator.

I. Introduction

Research and development of three-phase autonomous voltage inverters (AVI) are highly relevant. Designing a control system is one of the main stages in the development of AVI. Any control of the power converter ultimately boils down to regulating the time of the open state of the power transistor in relation to the period of its operation. This control method is called pulse width modulation (PWM). Given the tendency to build digital (digital-analog) control systems for power converters, a large number of control methods have been developed for the class of three-phase inverters. When implementing them to solve practical problems that require the formation of an output voltage with increased frequencies (1-2 kHz), the control system must be developed on its own due to the lack of ready-made integrated solutions in the form of three-phase PWM control.

By changing the supply voltage and frequency of electrical equipment to control the speed and torque of AC motors, the applied electromechanical devices or control system provide alternating current of one frequency to another alternating current of another frequency.

As an example of such a device, a frequency converter can be shown, that is, an electromechanical device for ensuring continuous control of the process. Typically, this device relates to the control of speed and torque of equipment (asynchronous or synchronous motors). Such devices are widely used in various fields of industry and transport [1-4,9-13]. Table 1 presents the main control methods used in frequency converters.

Table 1. Basic control methods in frequency converters

Control method	Speed range	Speed error	Period of exceeding mocop	Start howl moment	Meaning	Standard application
Skalyar	1:10	5-10	Inaccessib le ny	small	very small	little used pumps, fans, compressors
linear vector	>1:200[2]	0	<1-2	big	big	frequently used taps,elevators, transport, etc.
direct control tion with vector field modulation	>1:200[2]	0	<1-2	big	big	-------
direct control with nonlinear torque	>1:200[2]	0	<1	big	big	--------
direct self-govern ment	>1:200[2]	0	<1-2	big	big	high-performan ce facilities

II. Formulation of the problem

The efficiency of the control system and power section is directly affected by the modulation method. The growth of research in the field of modulation is characterized by the widespread development of power electrical converters, determining economic efficiency and productivity. The main goal of modulation is to achieve the best shape of signals (voltages and currents) with minimal losses. Other additional control tasks such as rectifying DC voltage, reducing input current ripple, reducing overvoltage speed can be solved by using the right modulation method. Modulation methods are divided into IV main groups:

1. pulse width modulation;
2. vector field modulation;
3. harmonic modulation;
4. methods for changing variable frequency circuits.

To generate control pulses using modulation methods, it is possible to synthesize the output voltage with the necessary parameters (shape, frequency, amplitude). Due to the presence of higher harmonics in the output signal, it is necessary to filter the output signal to generate sinusoidal currents.In such devices, the presence of an inductive load (electric motors), if necessary, additional filters are used. The maximum output voltage is determined by the DC link voltage. To effectively apply a heavy load, a high constant DC link voltage is required. However, in practice this voltage is limited by the operating voltage of the semiconductors. For example, low-voltage transistors provide output voltages up to 690 V. In terms of voltage, multi-position converter circuits have been developed to overcome this limitation. Such control modulation devices are complex, but have good power, size, reliability, efficiency and performance. For example, in a three-position converter with a neutral point, the DC voltage is divided by two capacitors, so the phase can be connected to the positive voltage line (by turning on the two upper switches), to the middle point (by turning on the two central switches) or to the negative voltage line (by turning on the two lower switches). Each switch can block only half of the DC link voltage, so the use of semiconductor switches allows you to increase the power of the device. Typically, such devices use high-voltage transistors and thyristors (figure 1).

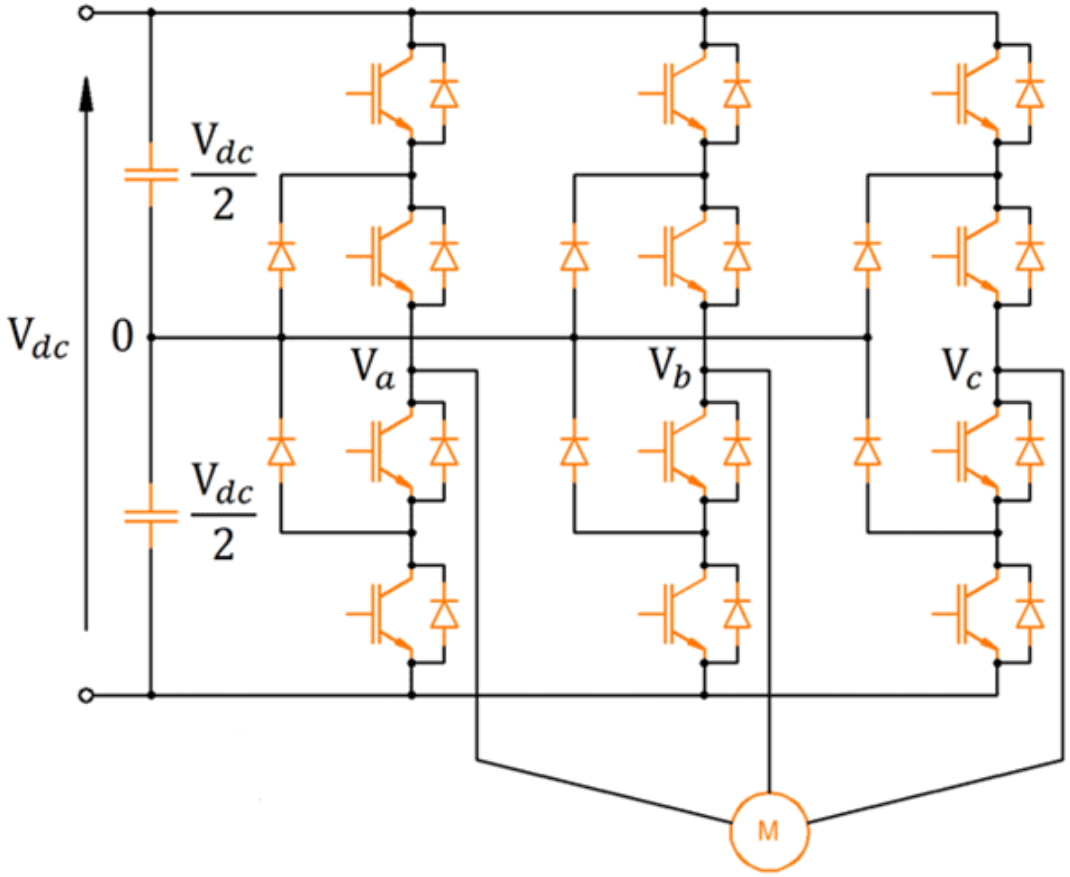


Figure 1. Circuit diagram of a three-position converter with a neutral point

Disadvantages of such devices: imbalance of capacitors created by the asymmetry of the device - this problem is solved by changing the modulation method; uneven distribution of losses - depending on the operating mode, losses from replacing the circuits of central and external keys differ.

The presented converter, in order to achieve an output signal beyond the three-level limit, can be scaled so that by means of capacitors it is divided into more than two values of the DC link voltage. Each of the separated voltages can, using different switches and limiting diodes, be connected to the load. With the increase in power, the advantages of multi-position converters are: power quality, low voltage overvoltage rate and associated electromagnetic interference. Figure 2 shows the voltage dependence of the converter phase [4-6].

Any control of the power converter, ultimately, relative to the period of the operating mode of the power transistor, is driven by the regulation of the period in the open position. As you know, this control method is a method of pulse width modulation. The construction of control systems for power circuits is easily covered by a wide class of industrially produced integrated controllers. In this case, taking into account the trend in the construction of power converters with digital control systems (digital-analog), numerous control methods have been developed for the class of three-phase inverters. However, when solving practical problems, for implementation, which at increased frequencies (1-2 Hz.) requires the formation of an output voltage, the control system is used as a three-phase controller.

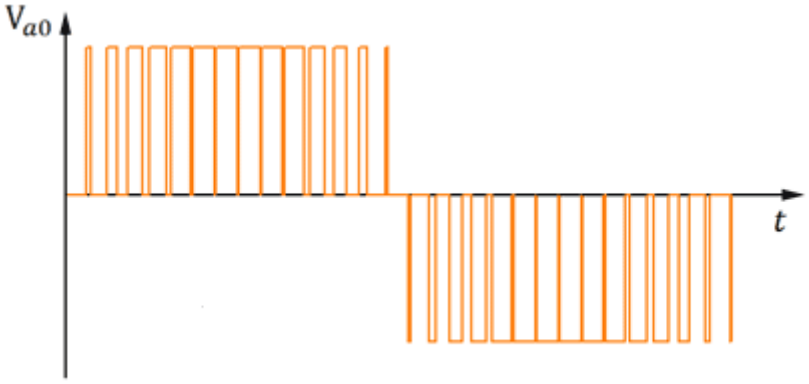


Figure 2. Phase voltage of a three-position converter with a neutral point.

III. Problem solution

Research and development of three-phase autonomous voltage inverters (AVİ) is one of the urgent issues. One of the main stages of the development of AVİ is the design of management systems. Different management of the power converter is adjusted according to the ratio of the time of the power transistor in the open state to its working period. Such a control method is called circular-pulse modulation [1-4]. Numerous control methods have been developed for the classification of three-phase inverters. However, in order to solve practical issues in their implementation, the management system should be developed with special forces.As is known, AVİ is a static converter that converts a constant voltage E_d to an alternating voltage through semiconductor switches (S1-S6). Either field MOSFET-transistors or IGBTs are used as semiconductor switches. Bridge transistors are controlled by the control system . The control of this system is carried out on the basis of existing algorithms in order to provide the load (Z_{load}) with a stabilized alternating three-phase voltage.

Currently, there are three main classifications of IS: analog, digital, mixed (digital-analog). The vector of development in power converters is directed to the application of digital IS [2]. However, the operating speed of modern microcontrollers and analog-digital converters is not so high, 50-100 kHs. and the power converter working at multiple frequencies should ensure the necessary speed of the digital .The most circular-pulse control is characterized by a modulation period, during which the stator winding of the electric motor is sequentially connected to the positive and negative poles of the rectifier. The duration of these states within the impuls modulation period is modulated according to the sinusoidal law. At high clock frequencies of impuls modulation (usually 2.....15kHz.), sinusoidal currents flow in the windings of the motor (due to filter properties). Thus, the shape of the output voltage curve presents a high-frequency bipolar sequence of rectangular pulses. The frequency of the pulses is determined by the frequency of the DIM, the duration (width) of the pulses is modulated according to the sinusoidal law during the period of the output frequency of the AVİ. The shape of the output current (in the windings of an asynchronous electric motor) is practically sinusoidal. The output voltage of the inverter can be adjusted in two ways: the first - by the amplitude method, by changing the input voltage; the second - with impuls modulation, changing the valve rotation program. Due to the development of the modern element base in modern frequency converters (microprocessor, IBGT-transistors), the second method has found wider application.). Such control allows the converter to obtain a high efficiency, equivalent to analog control through frequency and voltage amplitude. Modern inverters are based on fully controlled power semiconductor devices [2].

In the mixed control system, the proportional-integrative-differentiating control operation was performed by means of amplifiers. The linear voltage from the output of the inverter is measured by a voltage transmitter and U_{steady} as a negative feedback signal. the regulation is summed up by the signal of the price. The signal at the output of the impuls modulation regulator is scaled and transferred to a digital series through an analog-to-digital converter, then transferred to a digital pulse generator, which in turn, according to the algorithm, supplies the three-phase AVİ with control pulses using the driver block, is for matching low power control signals with low input resistance of power transistors. According to the structural scheme, the digital pulse shaper is like a prototype of the analog impuls modulation -controller, it forms the control pulses of the function from the sum of the regulation value and feedback signals.

With the development of microprocessor technology, vector impuls modulation algorithms have found wide application. For each phase of the output voltage, there is a period equal to $(\pi/6)$ in the dual period of the output frequency, and the voltage of this phase is maximal according to the module (figure 4). According to the vector impuls modulation algorithm, the switch corresponding to this interval duration (S1-S6) remains open regardless of the modulation factor K_M. This coefficient is a signal of the adjustment price in relative units U_{steady} and varies in the range of (0-1) (figure 3). According to the vector impuls modulation method, the operating period of each phase of the transistor bridge (2π) is divided into 6 equal intervals $(\pi/3)$. Using this method, each $(\pi/3)$ interval is divided into 8 impuls modulation intervals $(\pi/24)$. Thus, the modulation number of the impuls modulation -transformation is M=48.

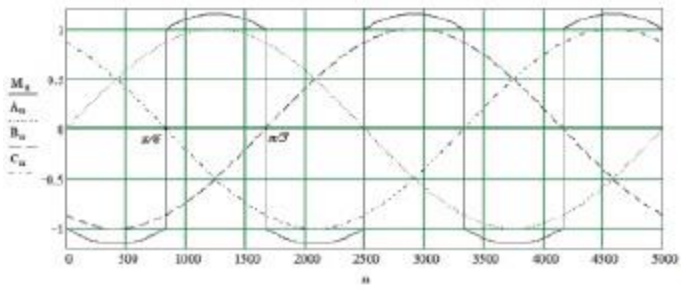
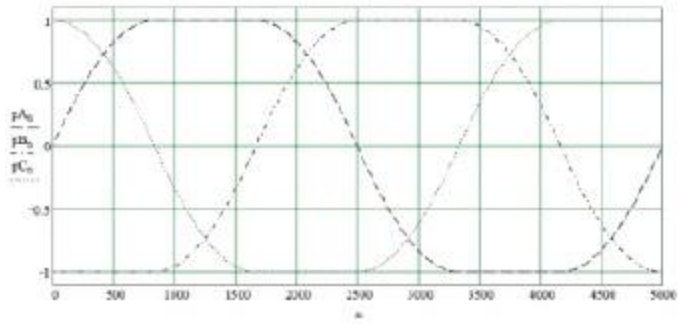
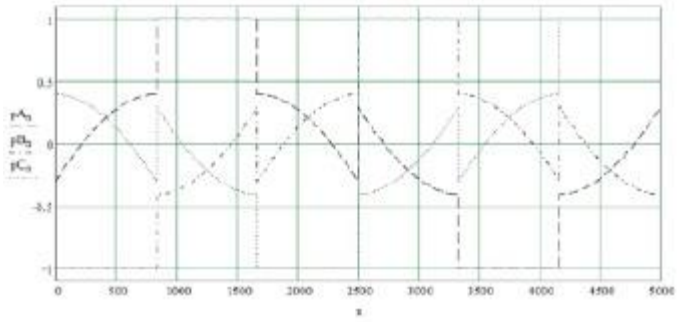


Figure 3. An, Bn, Cn – phase-regulating voltage signals:
Mn – signal to modulation of 3rd harmonics.

If the frequency of the AVİ output voltage (f_{output}) is equal to 1 kHz. Then the switching frequency of power switches: $f_k = M \cdot f_{out.} = 48$ kHz. In this case, impuls modulation period duration: $1/48 \cdot 103 = 20.833$ mks. Practical work experience with modern semiconductor devices shows that the operation at the given frequency in the hard switching mode of the switches is close to the maximum (taking into account the losses of conversions in transistors). That is, to increase the frequency of the output voltage of the AVİ, it is necessary to use transistors with minimal dynamic losses [1].



a)



b)

Figure 4. Control signals: a) modulation signal KM =1;
b) modulation signal KM =0.7.

As is known, the impuls modulation signal is formed by comparing the step voltage with the control voltage through a comparison device. For example, in analog systems - this is an ordinary analog comparator. Due to the fully digital design of the pulse shaping system, the comparison is carried out in digital comparators, a triangular digital opening is used as a two-way impuls modulation opening at the π/24 interval, and this is provided by a reverse counter. In the implementation of this type of vector impuls modulation, they passivate each phase control twice in the π-difference period of the output frequency during (π/3) intervals, that is, switching of power switches with the frequency of impuls modulation does not occur. In this case, according to the control algorithm, either the upper or lower phase switch is open. The other two phases are controlled by extending the duration of the pulses according to the sinusoidal law through impuls modulation . Thus, the average switching frequency of each power switch is 1.5 times less than that of the classical impuls modulation, which reduces the switching losses accordingly.

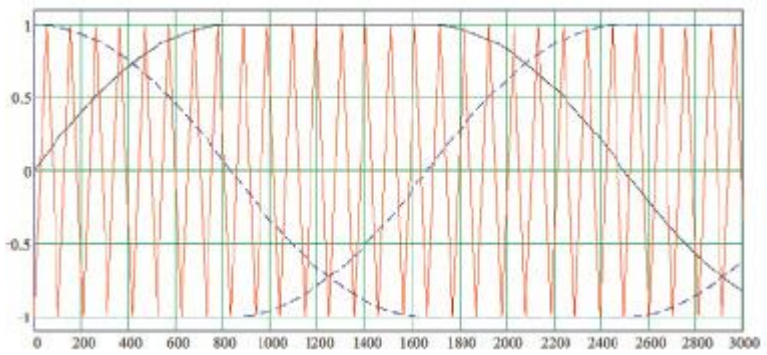


Figure 5. Reference sine functions and signal and shutdowns

Looking at one of the six (π/3) intervals, one switch is fully open and the other two are at 48kHz. they take part in the cycle with the frequency, the width of the pulses in this case changes according to the harmonic law over time. If we take the N/2 opening average of the digital cutting as the conventional level of the zero line of the phase opening (figure 5), then the sinusoidal voltage of the phase opening:

$$U_f = \frac{N}{22} + \left[\frac{N \cdot K_M}{\cos\frac{\pi}{6}} \right] \sin(\theta),$$

here cos(π/6)- fundamental harmonic amplitude increase factor; θ -initial phase.
And the voltage for this interval:

$$U_f = N - \frac{N}{22} + \left[\frac{N \cdot K_M}{\cos\frac{\pi}{6}} \right] \sin(\theta)$$

The expressions obtained for the second and third phases are identical, only the first phase is shifted by θ(2π/3). After several changes, we get the formulas for calculating the coefficients, which represent the width of the pulses varying with time:

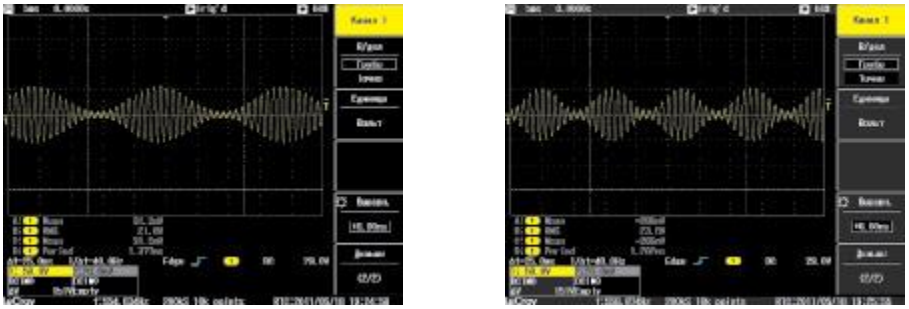
$$A_{1M} = N\left(1 - K \cdot \cos(\theta)\right)$$

$$A_{2M} = \frac{N}{2}\left[2 - K\left(\cos(\theta) + \sqrt{3}\sin(\theta)\right)\right]$$

The calculation of the received coefficients is the same for each (π/3) interval, and their values vary as a function of the modulation coefficient K_M, whose value is determined by the voltage transmitter and the feedback signal received from the digitized analog-to-digital converter. There are two methods of processing received information: calculation of interval coefficients A_1, A_2 (during system operation) and filling of the cells of the coefficients matrix (selection of the numbered line corresponding to the K_M value in the range 0-1), which is divided into a number of values (according to the accuracy of the stabilized voltage). The second option is considered more reliable, because the digital pulse generator consists of a microcontroller and a microcircuit.

In this case, the microcontroller performs the function of initial calculation of the coefficients and filling of the matrix, and the received matrix is used for the selection of subsequent values, in order to transfer them to the inputs of digital comparators. The values of coefficients A_1 and A_2 are calculated according to the formulas and written in the blocks of the matrix. Each block consists of 256 lines. The value of the coefficient K_M =1 corresponds to the maximum value at the output of the analog-to-digital converter. Thus, 16 matrices with 256 rows, 8 matrices for each coefficient, corresponding to the (π/3) sector divided into 8 intervals, need to be created.

The values selected from the corresponding matrix and the digital signal of the cascade opening are given to the digital comparators, at the output of each comparator a rectangular pulse is simultaneously formed. Timed (π/3) sequences consisting of 8 pulses are formed through registers and logic devices. Thus, the received periodic sequences are transmitted to the pulse distribution device. This device, in turn, transmits the necessary control signal to the corresponding transistor according to the logic of the operation of the bridge switches [8-13].



a) b)

Figure 6. Voltage oscillograms at the inverter output (for a given alternating signal):
a-set signal-50 Hz.; b-set signal – 100 Hz.

IV. Conclusions

Autonomous voltage inverter (AVI) using semiconductor switches is a static converter of direct voltage into alternating voltage. MOSFET or İGBT field-effect transistors are used as semiconductor switches. The transistors are controlled by the control system according to a given algorithm, providing a stabilizing alternating three-phase load voltage (Z_{load}). Currently, there are three classes of control systems: analog, digital, mixed (digital-analog). Analogue control systems have an advantage over mixed ones; the vector of development in power converters is aimed at the use of digital control systems. The three-phase autonomous voltage inverter with the vector modulation method has a digital-analog control system with a frequency (1-2kHz), high-speed response for the necessary stabilization and formation of the output voltage, with improved weight and dimensions, it can be used for various three-phase voltage inverters with symmetrical loads. The practical implementation of the developed method is carried out by a device consisting of an analog-to-digital converter controller and a microcircuit. A digital pulse shaper is directly designed in this device.

In order to check the performance of the system, a sinusoidal signal of various frequencies is transmitted to the input of the converter (figure 6) and we have oscillograms at the output of the inverter. As can be seen from the presented oscillograms, the control system has stability and good dynamic feature.The power part (electricity converter) and the control device (controller) are the main elements of a control system or electromechanical device. An example of such devices is a frequency converter, that is, an electromechanical device that provides the control process. The control device of the control system (microcontroller) runs software and controls the main parameters.

References

[1] G.S.Kerimzade, L.Ya.Qurbanov.“ Control and modulation methods in electromechanical devices triplets”,Baku. 2019. “Energy Problems”, №1. pp.94-98.

[2] G.S.Kerimzade, L.Ya.Qurbanov.“Three-phase autonomous inverter control system with vector pulse width modulation”,Baku. 2019.“ScientificWorks”, №2. pp.167-172.

[3] G.S.Kerimzade ”Calculation of parameters of control induction support”, PRZEGLAD Elektrotechnich niczny Publishing house of magazines and technical literature SIGMA-NOT.ISSN 0033-2097,R.100.NR.05/2024. Warszawa.. pp.219-222.

[4] G.S.Kerimzade. ”Method for determining the overall dimensions of an induction control support”, PRZEGLAD Elektrotechniczny Publishing house of magazines and technical literature SIGMA-NOT.ISSN 0033-2097, R.100.NR.06/2024. Warszawa.. pp.235-237.

[5] G.S.Kerimzade. ”Structure of the monitoring and tracking electromechanical control system”,PRZEG LAD Elektrotechniczny Publishing house of magazines and technical literature SIGMA-NOT.ISSN 0033-2097, R.100.NR.07/2024. Warszawa.. pp.295-297.

[6] G.S.Kerimzade, G.V.Mamedova. ” Research of electromechanical devices with levitation elements in control systems”, Journal “Realibility: Theory & Application”. ISSN 1932-2321. Volume 19.№ 2(78). Yune 2024.pp.85-90.

[7] G.V.Mamedova, G.S. Kerimzade.”Design parameters for electromechanical devices with a levitation element”, PRZEGLAD Elektrotechniczny Publishing house of magazines and technical literature SIGMA-NOT.ISSN 0033-2097,R.100. NR 09 / 2024.Warszawa. pp.175.

[8] A.Q.Yudinsev . Three-phase inverter control systems with vector pulse width modulation. Basic Research // - 2015.

[9] M.P.Belov . Technical means of control and automation. Study guide.2009.

[10] A.D.Pozdnev. Electromagnetic and electromechanical processes in frequency-controlled asynchronous electric drive. Cheboksary. Chuvash University.2000.

[11] R.T.Chreyner. Mathematical modeling of an AC electric drive with a semiconductor frequency converter.- Ekaterinburg . 2000.

[12] Rahul Dixit, Bindeshwar Singh, Nupur Mittal. Adjustable speeds drives. Review on different Invertertopologies.-// Sultanpur, India. International Journal of Reviews in Computing, 2012.

[13] Marian P. Kazmierkowski, Leopoldo G.Franquelo, Jose Rodriquez, Marcelo A.Perez, Jose I.Leon “High-Peromance Motor Drives”. // IEEE Industrial Electronicsd, vol.5

A STUDY ON CONVENTIONAL BULK QUEUES IN QUEUEING MODEL

Keerthiga S[1] and Indhira K*

[1,*] Department of Mathematics, School of Advanced Sciences,
Vellore Institute of Technology, Vellore - 632 014, Tamil Nadu, India.
keerthiga.2020@vitstudent.ac.in, kindhira@vit.ac.in.

Abstract

The study on bulk arrival and batch service queueing models is discussed in this article. The mathematical logic of queueing models is crucial in many industries, especially in production lines, to minimize congestion issues. This survey seeks to review and model different occurrences in the area of bulk queues with vacations, breakdowns, and repairs. This research goals to provide enough information to analysts, researchers, and industry professionals to simulate congestion problems and create various performance measures to improve the queueing model.

Keywords: Bulk Arrivals and Batch service,Vacations,Breakdown and Repairs.

1. Introduction

“Queuing theory" is an application of stochastic processes in Operations Research. Predictions of waiting times and the duration of the server(s) are busy are heavily reliant on the concept of “stochastic processes". The inventor of queueing theory was Agner Krarup Erlang (1909-1917), a Copenhagen telephone exchange engineer. The idea was not recognised as having a variety of uses until after World War II. Erlang was the first to address congestion issues in the early twentieth century. His work motivated engineers and mathematicians to use probabilistic methods to solve queueing problems. Many of these findings have been applied to fields such as operations research, computer science, telecommunications, traffic engineering, and reliability theory.

The idea of bulk queues has exploded its popularity in recent years, because of the congestion concerns that have occurred in communication/computer networks and industrial/manufacturing systems. Consumers arriving in large groups or in quantity at a post office, ships arriving in a port in a column, or a few ideas include people going to a restaurant. In many cases, such as an unreliable server,a vacation, and so on,bulk queues can be used.
In this review, a study on conventional bulk queues has been done. We assumed that the bulk queues are supervised by an arbitrary distributions. The structure of the paper is as follows: batch arrival and bulk service are analysed in section 2, Bulk queues with different vacations are investigated in section 3. Bulk queues with breakdown and repairs are presented in section 4. Finally, section 5 discusses the final notes and the summary of the survey.

2. Bulk Arrival and Batch Service

In many real-world queueing situations, the arrival of more than one customer at the same time is called a “bulk queue". Bulk queues can be employed in a number of different production

scenarios and manufacturing processes. Mathematical modelling of bulk arrival and service queues yields various performance measurements for analysing assembly line systems. The optimization method can be used to set the server's minimum and maximum capacities. The group of arrivals can be determined by using a random variable or by a predetermined quantity. This is a challenging problem to solve on a large scale, both mathematically and practically.

Neuts [1] classified bulk queues with Poisson input. The output process has been described as the length of the queue in discrete and continuous time. Customers enter at a counter to a homogeneous Poisson process rate and subsequent service times are conditionally independent. The embedded semi-Markov Process approach has been the subject of the investigation. A single server queue(SSQ) with service interruptions was described by Takine and Sengupta [5]. They have measured the length of a SSQ and the distribution of waiting times during service problems. They have also considered arrivals according to the Markovian Arrival Process (MAP). "bulk service queue with accessible and non-accessible batches at discrete times" was obtained by Goswami et al. [9]. They evaluated steady state probability for both finite and infinite buffers in this model. Numerical results were used to demonstrate the utility of the model variables. Al-Khedhairi [10] studied the Bernoulli schedule which was used to choose and re-service services in a bulk service queue (BSQ). Results were obtained in both discrete and continuous time for the proposed queueing model. In addition, the real-time scenarios includes an explicit explanation of service and re-service types. Sanjeet Singh and Naveen Kapil [21] investigated the system had been in an idle state before the appearance of a client, and once it's in a working state, it may breakdown, and the system may progress on vacation during that time.

Pradhan et al. [24] was completely researched "An infinite-buffer batch-size-dependent batch-service queue with bunch of arrivals and arbitrary serving capacity". It was demonstrated that the Probability Generating Function(PGF) for line content at the service completion time may be used to obtain the probability distributions for departure periods and related performance characteristics. Schwarz and Martin [25] introduced a new discrete-time Markov chain-based approach. The Supplementary Variable Technique(SVT) was used to investigate extensively spread inter-arrival and service times. The suggested technique provides detailed probability distributions for crucial performance indicators such as length of the queue, waiting time, departing batch size, and non-empty batch inter-departure period. Rashmita Sharma [27] examined "$G^X/G/m$ queueing model with finite capacity using the diffusion approximation approach". The two boundaries connected to the fundamental return procedure have been placed at 0 and N for the aim of the result. In this paper, equations for the steady-state distribution, the No. of clients, delay opportunity, and suggest line period have been obtained.

Madhu Jain [31] explained "priority queue with batch arrival, balking, threshold recovery, unreliable server, and optimal service". They explored a SSQ system with two situations of priority for clients: high priority and low priority. Priority clients are admitted with preemptive priority over low-priority clients. Under the priority rule, customers of various classes are suitable to gain varying levels of quality of service. In the event that the server fails while delivering service to priority customers, it is instantaneously transferred for repair. The queue size distribution and other performance indicators were calculated using the matrix geometric method (MGM). Bruneel et al. [33] analysed a discrete-time double-class SSQ system with bulk service at arbitrary slot borders, calculated the steady-state PGF of the No. of clients in the station, and visualised the effect of different arrival process parameters on the average system occupancy. Charan Jeet et al. [34] investigated the waiting time, under the ideas of a Bernoulli feedback (BF) schedule and the maximum entropy principle was used to analyze a bulk arrival unreliable queue with necessary (optional) services. Sushil Ghimire et al. [35] investigated the batch QM with multiple servers. Customers are served individually if there are fewer than the batch size. They also evaluated the probability distribution and performance characteristic of the size of the queue and the waiting time spent for both scenarios. Krishna moorthy et al. [36] examined the client has the option of joining a service at any time within the stages. The service rate at each stage is determined by both the stage and the number of clients serviced in a batch. An analysis based on performance characteristic had been developed, and it was presented to obtain the effect of service rates on

this cost function. Gupta and Pradhan [37] analyzed a batch-service line with unlimited buffers and a Markovian arrival procedure, extensively distributed service time at the batch's departure time, and they were suitable for developing a bi variate vector generating function for system size and server content distribution.

Geetha et al. [41] described the queue parameter was estimated and predictor analysis was performed on non-parametric bulk arrival fuzzy queueing systems. The inverse participation function of the k-stage fuzzy queueing model was initially created, followed by a proposal for a system performance algorithm. The optimality level of the interval was investigated as well as the system's level of uncertainty. Shavej Ali Siddiqui [46] focused on an interdependent system with a BSQ model with distinct capacity. Queueing models with arrival and service procedures that are both Poisson which follow a bi variate Poisson process were studied. These models system characteristics are investigated. Santhi and Saravanan [49] proposed a cloud computing architecture model for examining performance metrics like waiting time for different i-classes of public cloud customers or units accessing cloud databases in size K batches. The overall waiting time and the No. of consumers in the system showed a substantial difference in simulation results. Sinu Lal et al. [51] examined the batch size, which has been decided by the No. of emergency clients waiting in line. This model was shown in a courier delivery system with two different kinds of arrivals, created by a matrix analytic method (MAM). Hemanth Kumar et al. [53] studied the SSQ model with non-homogeneous arrival rates and batch size distribution factors reached system performance measures, which were determined to be significant. Djamila Zirem et al. [56] analysed a bunch of arrivals in queue that included general retrial time, breaks, repairs, and reserved time. In this situation, customers are expected to arrive via compound Poisson process(PP). Cognitive radio networks and manufacturing systems, for example, could benefit from this paradigm. The average queue length in orbit has been determined, as well as system performance metrics.

Yahya Mohamed and Karthikeyan [60] provided an analysis of group arrivals in QM. Bulk queueing models are very helpful in resolving the problem of congestion. A review of the work completed on bulk queues was attempted with this survey and modelling of different phenomena. Kunpeng Yang and Hangguan Shan [61] explored "A tandem open-queuing network with multiple Geo/Geo/1 customers and a server-based bulk service queue". The primary goal of this project is to examine the usual delay as well as the likelihood of overflow for each customer's message going through the system. Alexander Zeifman et al. [63] provided a technique for estimating "convergence bounds for four kinds of multi server queueing systems with an non homogeneous $M/M/S$ queueing system" with probable state-dependent arrival and service strength, as well as possible to arrival of groups and bulk service. Cruz et al. [64] derived various performance metrics for "General bulk-arrival Markovian multi-server finite queues" and investigated their operation in limited-sample conditions.

Jitendra Kumar and Vikas Shinde [67] investigated the performance of bulk queues with several servers using a queue model. In this research, they have presented explicit mathematical formulae for real-world problems such as customer transmit procedure for bulk queues with multi-servers, and this model can be investigated with time-dependent arrival and service rates. MATLAB-9 was used to calculate the numerical results. Kirupa and Udaya Chandrika [68] presented the steady state analysis(SSA) for "an unreliable batch arrival retrial G-queue with M service modes, priorities, active breakdown, delayed repair, and orbital search". PGF have been used to determine the No.of consumers in orbit and in the system. Barbhuiya and Gupta [69] derived an entirely new method for testing the $GI^X/M^Y/1$ queue used to the SVT and the difference equation method to carry through the entire analysis and obtain explicit expressions of system content distributions before the arrival of any time epoch. Atina Ahdika [73] developed a queuing model for a tiny system that integrates dual servers and multi phase processing into a single system. Customers arrive in bunches, and each server has a separate service time. This model was created on the basis of the birth-death process to address the problem of system performance measures. It was calculated by using the global balance condition. Nithya and Haridass [74] studied the purpose of modelling and simulating a bulk queueing system. They

investigated the performance metrics and provided simulation modelling for the textile industry in a group of arrival and service queueing systems. Gupta and Banerjee [75] analysed a bulk arrival and service on a finite buffer queue with line length and batch size dependent service. In a steady state, a SS offers a service that meets the fundamental bulk service criteria. Depending on queue length, the service time distribution followed any generic distribution and represented batch size. The SVT and embedded Markov chain methodology were used to enumerate the joint probability of queue length.

Gupta et al. [76] investigated bulk service queues that has a finite capacity and two vacation rules: Single Vacation(SV) and Multiple Vacation (MV). The service time was supposed to come after a generic form distribution, which has been dependent on the batch size being served. Vacation time was determined by the length of the queue at the moment of vacation inception. Andrew Daw and Jamol Pender [77] discussed the impact of bulk arrivals on countless server queues. A Poisson process was considered to determine the arrival epochs with both stationary and non-stationary arrival rates and they examined fixed and arbitrary arrival batch sizes as well as exponentially and generally distributed service duration's. "A multi-class batch-service QM with restricted flexible service capacity and system occupancy" was presented by Jens Baetens et al. [78]. A single batch server creates a queue for all clients of the same class, up to a class-reliant greatest service capacity. They investigated a BSQ with two discrete-time customer classes, focusing on system occupancy during service initiation chances to compute comparability for slow and heavy traffic in order to decrease the numerical involvement generated by highest service capacity.

Madhu Jain and Sandeep Kaur [79] explored a QM with a (p,N)-policy to estimate the validity and possibility of utilizing the SVT and maximum entropy principle (MEP) to produce different performance measure outcomes. Sree Parimala [82] examined the $M/M(a,b)/(2,1)$ queueing model with heterogeneous servers and batch-based delayed vacation for slow down servers. The two fundamental aims of queueing theory as a whole have been simplified to client wait times and queue length. The quick server has always been existing in the system, but the slow server can only take a postponed vacation. Singh and Srivastava [83] explored the Markovian queueing model with group of arrival and three categories of bulk services,the 1^{st} is a mandatory service, the 2^{nd} and 3^{rd} are optional services depending on the customers preferences. This effort will aid in reducing client wait times at service stations,when they will be dealing with one or more servicing facilities. Bank and Samanta [85] derived the steady-state queue -length distributions of the BMAP/BMSP/1 queue by using analytical results at random, pre-arrival, and post-departure epochs. They have also done a detailed investigation of the system-length distribution at arbitrary time by applying the MGM.

Bharathi Ramesh Kumar et al. [86] discussed that the estimation theory was used to examine the system performance in $FM/FM^{(k)}/\infty$. Initially,they have applied Alpha-cut approach to convert an Erlang queueing system into a fuzzy Erlang system. Finally,the defuzzification values can be measured by using the Robust ranking approach. Deena Merit and Haridass [88] analysed flexible general batch service (FGBS) queueing model to constrict the resting time of consumers. The server starts serving when the required number of consumers is available. Customers that have already arrived must stay for the duration of this period, regardless of when they arrived. This research developed the FGBS rule to enable batching flexibility in production.

3. Bulk Queues with Different vacations

Initially Doshi [2] investigated queueing systems in combination with vacations. They concentrated on single-server queues. There has been quite a little work accomplished on multi-server queues with vacations. A few generic decomposition outcomes and the technique utilised to attain these conclusions are explained for two vacation models. Only in single-server vacation(SSV) models with general arrival procedures is the simple deconstruction known when the vacation order is unaffected by the arrival and service operations. Lee et al. [3] investigated "M^X/G/1 queueing system with MV and N-policy". When the system has been emptied, the server is

shut off. When the queue length attain or surpass a predefined value N (threshold),the server is activated and proceeds servicing consumers.

The single vacation model with probabilistic measures and thresholds was the main focus of Lee et al. [4]. They devised a method for reducing the system's long-term average cost. Arumuganathan and Jeyakumar [7] established "steady-state conditions for batch queueing models deals with N-policy MV, set-up and close-down times". They analysed this model uses to the SVT deals with the appropriate numerical illustration. Many performance indicators such as predicted queue length,anticipated length of busy and idle periods were obtained. The probabilities of the server getting overloaded and of the server which takes a break from the queueing process were also calculated. Chang and Choi [8] explored "a single-server bulk arrival and batch service" in which clients were serviced in arbitrary-size groups and the server took MV when the line was unoccupied. "Discrete time $GI/Geo/1$ queue with working vacation and vacation interruption" was analysed by Tain and Li [12]. The authors of this paper introduced vacation interruption. This is the concept that if there are no consumers,a server goes on vacation. If customers continue in the system after a service has been performed throughout the vacation time, the server will return to normal operations. Otherwise it continues on its vacation.

Samanta et al. [13] proposed "discrete-time $Geo^X/G^{(a,b)}/1/\mathrm{N}$ queues with group of arrival and service under single vacation and multiple vacation policies". The ranges of batch service times and server vacation times has been an integral several's of slot duration. They derived queue size distribution functions at service completion, vacation termination, random, and pre-arrival epochs. Li et al. [14] analysed a single server vacation queue with a general arrival process. The "two policies of working vacations and vacation interruption" have been linked in this model to certain practical problems. The MAM was used to examine the "$GI/M/1$ queue" with these two strategies in order to achieve different performance indicators such as mean length of the queue and pausing time. Li et al. [16] presented "steady state conditions for discrete time batch arrivals and a single service queueing model with working vacations". They modelled the system as an embedded Markov chain(EMC) with a PGF of $M/G/1$ type stationary queue length. In the above scenario, stochastic decomposition was also used to produce an equivalent queue PGF. Zhang and Hou [17] analysed "working vacations and vacation interruptions in an $M/G/1$ queue". They determined the distribution of length of the queue and service status at an random epoch under SSC using the process of a SVT and the MAM. Jeyakumar and Arumuganathan [18] studied "$M^X/G^{(a,b)}/1$ queueing model with multiple vacations, control policy on request for re-service". Re-service has only been offered in this model. if the size of the system is less than the server's minimum threshold value. They were able to generate steady-state conditions along with many performance measures in this investigation by using the SVT and it also includes a cost analysis.

A theoretical replica of a QM that deals with the restricted admissibility policy of arriving batches with multiple vacations has been explained by Haridass and Arumuganathan [20]. The outcome of this research were verified by numerical examples.The above queueing system is also complemented by an optimum cost analysis, which was used to calculate the minimum and maximum capacity of the server. Suganya [22] examined "$M^X/G(a,b)/1$ model with vacation interrupted, optional re service and balking" using the SVT and derived a steady conditions for a batch arrival queueing system. It is feasible to obtain the PGF of system length during any period epoch. Mishra and Pandey [23] considered a bulk queueing model $M/G^k/2$ for two non-identical servers including batch service and a grand vacation process. A few significant performance indicators are incorporated in this model, including a stationary and leaving point system length distribution and the joint distribution of system length and vacation duration. Maragathasundari and Karthikeyan [28] examined a server which may suffer a failure followed by a recover process in the bulk QM. They suggested a two-phase service queueing system, with the 2 phase potentially optional. The server has been allowed to take a long or short vacation when the service is completed, depending on the demands. Kalyanaraman and Nagarajan [29] investigated the arrival of random size X with a "single server, batch service (fixed) with compulsory vacation, and an unreliable server". This representation can be developed by considering the failure interval as a random variable. They have presented several numerical examples to show the model's

analytical compatibility. Jeyakumar and Senthilnathan [32] examined a SSA of a QM with a widely distributed variable batch size service and varied service rates. The PGF of queue size was calculated for any time epoch and for various completion epochs. It has been possible to obtain the models of busy and idle time distributions. Ayyappan and Deepa [38] analysed the steady-state case. They calculated the PGF of the system size for a QM with MV, closedowns, and state-dependent arrival rates and also collected a variety of performance indicators and tested numerically.

Thangaraj and Rajendran [39] designed and investigated "bulk arrival queueing systems with two different cases of service patterns and vacations". The batch service can be distributed in this model and Service will not be disrupted under the suggested technique, if the system length is greater than or equal to the lowest batch size, except during vacation periods. Rajendran and Thangaraj [40] proposed a bulk QM with a single service and vacation, and with the aid of the PGF technique, they were able to identify the suggested model's system size distribution and the performance metrics of the provided queueing system. This investigates group of arrival in QM with dual cases of service patterns on a SV. Vignesh et al. [42] investigated bulk arrival queueing model with multistage heterogeneous service provided by a SS with distinct service time distributions, feedback with BV, and optional server vacation, in addition to restricted admissibility. This study analysed the model's transient solutions, steady-state results, and many performance metrics. Niranjan et al. [47] studied "bulk arrival and service retrial queueing systems with server failures, thresholds, and multiple vacations". Different performance measurements as well as specific cases were examined.The orbit size of PGF at any time epoch was calculated by using the SVT. Niranjan et al. [50] explored "bulk arrival queueing system with server loss and vacation break-off". where the queue length was determined by the service. An ideal cost model was also established for the system, allowing management to decide on the threshold value at which service can start. For the suggested queueing model, the PGF of queue size at all time was computed.This paper also includes other performance measures.

Anitha et al. [52] investigated a queuing system for a consumers are served in three phases, with the first two phases being required and the third optional. After the service has been done, the server can take a short or extended vacation.The PGF of the cloud computing user queue size was determined by using the SVT and a generating function approach. Little's law was used to calculate the models of other performance measures. "State-dependent arrival in a bulk retrial queueing system with active Bernoulli feedback, multiple vacations and threshold" was investigated by Niranjan et al. [54]. The PGF describing the orbit size was examined by using the SVT, and many numerical examples were derived to complement the various performance characteristics. Niranjan et al. [55] examined a bulk service QM with a vacation. Finally, in order to renew their service station, they established a service threshold for the $M^X/G(a,b)/1$queuing system. Niranjan et al. [59] analysed a service QM with WV that was based on bunch of arrivals and batch sizes. Depending on the system size, the server gives a single or fixed-batch service. During WV, the server can provide service in two different modes. The PGF of the queue size at any time was determined. Ayyappan and Supraja [62] investigated "an $M^X/G(a,b)/1$ queueing model with a 2^{nd} optional service that was subjected to server failure and two different vacation and unrestricted admissibility policy". They calculated the PGF of the No. of consumers in the line at a arbitrary interval along with the system size distribution at a departing period in both transient and SSC.

Apoorv Saxena et al. [65] studied "data backup process Quality of Service (QoS) measures" such as the duration of period the supporting server is busy and the repetition of new connections. This approach enables us to easily evaluate the QoS dependent on model parameters and also compute backup parameters. Ayyappan and Karpagam [70] analysed "Bulk queue with an unreliable server, immediate feedback, N-Policy, Bernoulli Schedule MV, and a standby server". The stand-by server has only been used while the main server is being repaired. The main server may be allowed to take a short break when each service is done. The queue size PGF was estimated along with few significant performance measures. Kalyanaraman and Nagarajan [72] studied BV in the $M^{[X]}/G^K/1$ queue. Moreover,the server may suffer a breakdown, causing the

server to wait an unknown amount of time for repairs. This model was examined by using the SVT. “A bulk arrival retrial queue with starting failures and exponentially distributed multiple working vacation" was presented by Pazhani Bala Murugan and Vijaykrishnaraj [80]. Customers utilizing FCFS can connect to the server from orbit. When a server starts to fail, it is immediately dispatched to be repaired. The repair time is completely random. The SVT was used to calculate the PGF for the No. of consumers in orbit.

Karpagam et al. [81] proposed rework and repair processes were essential components of the manufacturing process. When a problem arises and a manufacturing issue needs to be addressed prior to delivery to the customer, these processes must pass the same level of inspection and repair as normal production. Tamrakar and Banerjee [84] discussed an unlimited capacity BSQ with SV and MV by using the SVT and the bi variate generating function method. Steady-state joint probabilities have been established at different epochs. Sadhna Singh and Srivastava [87] investigated the SS Markovian queueing system that has an idle and busy server, a vacation,a failure and recover states. The Markovian queueing model allows for repeated service completion attempts as well as customer impatience. They have compared it to the FCFS and bulk service strategies by calculating the total probability of these scenarios. Ayyappan and Nirmala [89] investigated “customer impatience in a non-Markovian multiple vacations queueing system using an inconsistent bulk queueing model with two types of vacation" on a single server. The server was permitted to take binary sorts of vacations based on queue length, and the PGF of the system size distribution in random and departure time was also examined. Bouchentouf and Guendouzi [90] discussed a single server batch arrival Bernoulli feedback queueing system with waiting server, K-variant vacations and impatient customers. Pradhan and Karan [91] explored of an infinite-buffer batch-size-dependent bulk service queue with server breakdown and multiple vacation.

4. Bulk Queues with Breakdown and Repairs

Madan et al. [6] investigated the steady-state behaviour of two $M^X/M^{(a,b)}/1$ queue model with random breakdowns. They assumed that the recover period for 1^{st} model was exponential, whereas for next it was predictable. In the BSQ model, all writers deal with server breakdown (SB) and a server that can only serve one client at a time, with the exception of Madan et al. [6]. In this research on queueing systems with SB, suspending the server before it has completed its batch of services is practically impossible. Ke [11] studied the operation of batch arrival queues with SB and start-up/close-down times under vacation policies. single vacation and multiple vacation policies are the two cases of vacation policies. when a consumers arrives through a shutdown interval, the server will start instantly and without start-up time.

Wang et al. [15] derived on characterizing the different system performance for the “T-policy of the M/G/1 queue with server breakdown and general start-up times". At least one person must join the queue after the server has been reopened after a specified length of time. The server needs start-up time before starting the service. Jeyakumar and Senthilnathan [19] investigated “$M^X/G^{(a,b)}/1$ QM with MV and close down period". The behaviour of a server failure without interruption. They developed a PGF for the service, vacation, and renovation completion epochs with a close down period. Sumitha and Udaya Chandrika [26] analysed a “batch arrival retrial QM with delay time and multi-stage repair" using for SVT. In a steady state, the PGF of the server state and performance measures has been constructed with numerical results. The model can be discussed in a fuzzy environment. Nithya and Haridass [30] explored a bulk QM with breakdown, batch control, and MV. The PGF for queue size at any time epoch was determined. In particular, a amount model has been developed to produce managerial decisions about how to reduce total costs.

Bharathidass et al. [43] examined the ideas like vacation, break down, and repair on a single server in an Erlang bulk service queue. The whole arrival at the service point through the PP and differing arrival rates depending on the servers position. The units were provided in bunches according to the usual bulk service rules during the k-service phase. Sasikala et al.

[44] examined "The behaviour of the $M^X/G(a,b)/1$ queue in the presence of server breakdown without interruption, multiple vacations, setup time, and N-policy". The proposed queueing systems steady state equations were derived and also developed, along with queue size PGF. Jeyakumar and Senthilnathan [45] studied a variable bulk service QM with various WV and server breakdowns. The queue length has been calculated at different arrival rates, service rates during WV and regular periods of service over an interval of time and finally, a PGF was also obtained on system length at any random period and various completion epochs.

Singh et al. [48] developed a single-repairable server QM with batch input and state-dependent rates. Throughout the repair, the server has to go through a No. of different required steps. The PGF of the queue size distributions are created by using the SVT. Ayyappan and Deepa [57] investigated the steady-state case of a "$M^{[X]}/G(a,b)/1$ queueing model for bulk queues with multiple vacation,close down,essential and optional repair". They determined the PGF of the system size and obtained different performance measures that have been numerically verified. "Analysis of Batch Arrival Bulk Service Queueing System with Breakdown, Different Vacation Policies and Multi phase Repair" was developed by Thangaraj and Rajendran [58]. The queue size distribution and performance indicators of a wide queueing system were constructed, and particular examples of the suggested QM were evaluated.

Ebenesar Anna Bagyam and Udaya Chandrika [66] examined in a proper sequence, the server gives M-phases of heterogeneous service. In a SS retrial queueing model, customers arrive in bunches through a PP. After the 1^{st} stage of service has been completed, the clients have the option of continuing on to the third stage or departing from the server. Deepa and Azhagappan [71] analysed a single group of arrival and bulk service queue with MV close downs and repairs. The server begins a close down and then goes on a random-length MV, if the system is unoccupied or the server is ready to serve after the repair but no one is waiting. The steady-state PGF of the line size at any time was derived by utilizing the SVT. Niranjan and devi Latha [92] investigated two-phase heterogeneous and batch service queuing system with breakdown in two-Phases, feedback, and vacation.

5. Conclusion

This survey aims to assess the work done on bulk queues, which are used to estimate the system in advance and avoid loss by utilising these models to simulate different phenomena. It can aid researchers, operations analysts, engineers, and statisticians in the use of these models. Many workers have combined the concepts of group of arrival and batch service queues with vacations, breakdowns, and repairs. These models can be used to help reduce traffic congestion and act as a source of inspiration for researchers in the field of queueing theory. A wide range of literature has been reviewed, and appropriate citations have been produced. This review provides a comprehensive overview of the incorporation of bulk queues in different scenarios. Batch arrival and bulk service queueing models play a crucial role in forecasting queue lengths, waiting periods, and other performance metrics in queue systems.

References

[1] M.F. Neuts, *A general class of bulk queues with Poisson input*,Ann. Math. Statis.,38(1967),759-770.
[2] B.T. Doshi, *Single server queues with vacations: a survey*,Queueing Systems-1 (1986),29-66, https://doi.org/10.1007/BF01149327.
[3] H.W.Lee, S.S.Lee and K.C.Chae, *Operating characteristics of M^X/G/1 queue with N-policy*,Queueing Systems., 15 (1994) 387-399, https:// https://doi.org/10.1007/BF01189247.
[4] Soon Seok Lee,Ho Woo Lee,Seung Hyun Yoon and K.C.Chae, *Batch Arrival Queue With N-Policy and Single Vacation*,Computers Operations Research., 22(1995) 173-189,https://doi.org/10.1016/0305-0548(94)E0015-Y.
[5] T.Takine and B.Sengupta,*A single server queue with service interruptions*,Queueing Systems., 26 (1997)285-300, https://doi.org/10.1023/A:1019189326131 .

[6] K.C.Madan, W.Abu-Dayyeh and M.Gharaibeh,*Steady state analysis of two $M^X/M^{(a,b)}/1$ queue models with random breakdowns*, Queueing Systems., 26 (1997)285-300.

[7] R.Arumuganathan and S.Jeyakumar, *Steady state analysis of a bulk queue with multiple Vacations, setup times with N-policy and closedown times*,Applied Mathematical Modelling, 29 (2005) 972-986,https:// doi:10.1016/j.apm.2005.02.013.

[8] S.H.Chang and W.D.Choi,*Performance analysis of a finite buffer discrete time queue with bulk arrival, bulk service and vacations*,Comput.Oper.Res., 32 (2005) 2213-2234,https://doi:10.1016/j.cor.2004.01.004.

[9] V.Goswami,J.R.Mohanty and S.K.Samanta,*Discrete time bulk service queues with accessible and non accessible batches*,Applied Mathematics and Computation,182 (2006) 898-906,https://doi.org/10.1016/j.amc.2006.04.047 .

[10] A.Alkhedhairi,*A Bulk Service Queue with a Choice of Service and Reservice under Bernoulli Schedule*,Int. J. Contemp. Math. Sciences, 2 (2007), 1107-1120.

[11] J.C.Ke,*Batch arrival queues under vacation policies with server breakdowns and startup/closedown times*,Applied Mathematical Modelling,31 (2007) 1282-1292,.https:// doi:10.1016/j.apm.2006.02.010.

[12] J.Li and N.Tian,*The discrete-time $GI/Geo/1$ queue with working vacations and vacation interruption*, J. Appl. Math. Comput., 185 (2007) 1-10,https: doi:10.1016/j.amc.2006.07.008.

[13] S.K.Samanta, M.L.Chaudhry and U.C.Gupta,*Discrete-time $Geo^X/Geo^{(a,b)}/1/N$ queues with single and multiple vacations*, Math. Comput. Model, 45 (2007) 93-108, https://doi.org/10.1016/j.mcm.2006.04.008.

[14] Ji-Hong Li, Nai-Shuo Tian and Zhan-You Ma,*Performance analysis of $GI/M/1$ queue with working vacations and vacation interruption*,Applied Mathematical Modelling 32 (2008) 2715-2730,.https://doi.org/10.1016/j.apm.2007.09.017.

[15] T.Y.Wang, K.H Wang and W.L.Pearn,*Optimization of the T policy $M/G/1$ queue with server breakdowns and general startup times*,Journal of Computational and Applied Mathematics,228 (2009) 270-278, https://doi.org/10.1016/j.cam.2008.09.021.

[16] Ji-hong Li,Wei-qi Liu and Nai-shuo Tian,*Steady-state analysis of a discrete-time batch arrival queue with working vacations*,Elsevier , 67 (2010), 897-912,https://doi.org/10.1016/j.peva.2010.03.001.

[17] Mian Zhang and Zhengting Hou,*Performance analysis of $M/G/1$ queue with working vacations and vacation interruption*,Journal of Computational and Applied Mathematics,234 (2010) 2977-2985, https://doi.org/10.1016/j.cam.2010.04.010.

[18] S.Jeyakumar and R.Arumuganathan,*A Non-Markovian Bulk Queue with Multiple Vacations and Control Policy on Request for Re-Service*,Quality Technology and Quantitative Management,8 (2011) 253-269,http://dx.doi.org/10.1080/16843703.2011.11673258.

[19] S.Jeyakumar and B.Senthilnathan,*A study on the behaviour of the server breakdown without interruption in a $M^X/G^{(a,b)}/1$ queueing system with multiple vacations and closedown time*,Applied Mathematics and Computation, 219 (2012), 2618-2633,https://doi.org/10.1016/j.amc.2012.08.096

[20] M.Haridass and R.Arumuganathan,*Optimum Cost Analysis of a Bulk Queueing System with Multiple Vacations and Restricted Admissibility of Arriving Batches*,International Journal of Operations Research,9 (2012) 27-43.

[21] Sanjeet singh and Naveen,*Analysis of a Bulk Arrival Bulk Service Queueing Model for Non Reliable Server*,International Journal Computer Engineering and Management, 17 (2014), 22307893.

[22] S.Suganya,*$M^X/G(a,b)/1$ with vacation interrupted, optional reservice and balking*,International Journal of Scientific and Research Publications,4 (2014), 1-9.

[23] S.S.Mishra and N.K.Pandey ,*Analysis of bulk queueing model $M/G^k/2$ for non identical servers with a grand vacation policy*,docs.vijnanaparishadofindia.org,Vol :31/32,(2002).

[24] S.Pradhan,U.C.Gupta and S.K.Samanta,*Analyzing an infinite-buffer batch arrival and batch service queue under batch-size-dependent service policy*,Journal of the Korean Statistical Society, (2015).1226-3192, https://doi.org/10.1016/j.jkss.2015.08.004.

[25] Justus Arne Schwarz and Martin Epp,*Performance evaluation of a transportation-type bulk queue with generally distributed inter-arrival times*, International Journal of Production Research, (2015),http://dx.doi.org/10.1080/00207543.2015.1092613.

[26] D.Sumitha and K.Udaya Chandrika,*Batch arrival retrial queue with delay time and additional multi-optional repair*,International Journal of Advanced Computer Research,Vol- 7(28) (2016) ISSN (Print): 2249-7277, http://dx.doi.org/10.19101/IJACR.2017.728006.

[27] Rashmita Sharma,*A Batch Arrival Multi-Server Queueing Model With Finite Capacity Using Diffusion Approximation*,International Journal of Innovative Research and Advanced Studies (IJIRAS) Volume 3,(2016) ISSN: 2394-4404.

[28] S.Maragathasundari and K.Karthikeyan,*A Bulk Queuing Model of Optional Second Phase Service with Short and Long Vacations*,International Journal of scientific research in science and Technology, Volume 2 Issue 5(2016) Print ISSN: 2395-6011.

[29] R.Kalyanaraman and P.Nagarajan,*Bulk Arrival, Fixed Batch Service Queue with Unreliable Server and with Compulsory Vacation*,Indian Journal of Science and Technology, Vol 9(38),(2016) ISSN (Print) : 0974-6846,https:// DOI: 10.17485/ijst/2016/v9i38/101793

[30] R.P.Nithya and M.Haridass,*optimum cost analysis of a bulk queueing system with breakdown, controlled arrival and multiple vacations*,International Journal of Pure and Applied Mathematics, Volume 109 No. 9 (2016), 81-89.

[31] Madhu Jain,*priority queue with batch arrival, balking, threshold recovery, unreliable server and optimal service*,RAIRO Operations Research,51(2017) 417-432,https://DOI: 10.1051/ro/2016032.

[32] S.Jeyakumar and B.Senthilnathan,*Steady state analysis of bulk arrival and bulk service queueing model with multiple working vacations*,International Journal of Mathematics in Operational Research,Vol. 9, No.3(2016),pp.375-394.

[33] Jens Baetens, Bart Steyaert, Dieter Claeys and Herwig Bruneel,*System Occupancy of a Two-Class Batch-Service Queue with Class-Dependent Variable Server Capacity*, International conference of Analytical and stochastic Modelling and Applications,(2016) pp.32-44,https:// DOI: 10.1007/978-3-319-43904-4-3.

[34] Charan Jeet Singh,Sandeep Kaur and Madhu Jain,*Waiting Time of Bulk Arrival Unreliable Queue with Balking and Bernoulli Feedback using Maximum Entropy Principle*,Journal of Statistical Theory and Practice (2016) 1559-8608, http://dx.doi.org/10.1080/15598608.2016.1251365

[35] Sushil Ghimire,Ram prasad Ghimire, Gyan bahadur Thapa and Sara Fernande,*Multi-server batch service queueing model with variable service rates*,International Journal of Applied Mathematics and Statistical Sciences (IJAMSS) (2017) ISSN(P): 2319-3972.

[36] Achyutha Krishnamoorthy, Anu Nuthan Joshua and Vladimir Vishnevsky,*Analysis of a k-Stage Bulk Service Queuing System with Accessible Batches for Service*,Journal of Statistical Theory and Practice,ISSN: 1559-8608 (Print) 1559-8616,https://DOI:10.3390/math9050559.

[37] U.C.Gupta and S.Pradhan,*Analysis of an infinite buffer batch size dependent service queue with Markovian arrival process*,Annals of Operation Research,(2019),https:// DOI 10.1007/s10479-017-2476-5 .

[38] G.Ayyappan and T.Deepa,*Analysis of Batch Arrival Bulk Service Queue with Two Fluctuating Modes of Service Closedown Multiple Vacation and State Dependent Arrival Rate*,International Journal of Scientific and Innovative Mathematical Research,Volume 5, Issue 11, (2017), PP 6-19, http://dx.doi.org/10.20431/23473142.0511002

[39] M.Thangaraj and P.Rajendran,*Analysis of Batch Arrival Queueing System with Two Types of Service and Two Types of Vacation*,International Journal of Pure and Applied Mathematics Volume 117 No. 11 (2017), 263-272.

[40] M.Thangaraj and P. Rajendran,*Analysis of bulk queueing system with single service and single vacation*,IOP Conf. Series: Materials Science and Engineering 263 (2017),https://doi:10.1088/1757-899X/263/4/042106.

[41] Sivaraman Geetha,Bharathi Ramesh Kumar and Sankar Murugesan, *Predictor Analysis on Non-parametric Bulk Arrival Fuzzy Queueing System*,Global Journal of Technology and Optimization (2017),https:// doi: 10.4172/2229-8711.1000212 .

[42] P.Vignesh, S.Srinivasan and S.Maragathasundari, *Analysis of non- Markovian batch arrival queueing model with multi stages of service of restricted admissibility,feedback service and three optional vacations in production and manufacturing*, International Journal of Mathematics in operation Research, Vol. 11, No.3,(2017).

[43] S Bharathidass,.V Arivukkarasu and V Ganesan, *Bulk service queue with server breakdown and repairs*,International Journal of Statistics and Applied Mathematics (2018) 3(1): 136-142.

[44] S.Sasikala, K.Indhira and V.M.Chandrasekaran, *General bulk service queueing system with N-policy, multiple vacations, setup time and server breakdown without interruption*,IOP Conf. Series: Materials Science and Engineering 263 (2017) 042154,https:// doi:10.1088/1757-899X/263/4/042154.

[45] S.Jeyakumar and B.Senthilnathan,*Modelling and Analysis of a bulk service queueing model with multiple working vacations and server breakdown*,RAIRO Operations Research , 51 (2017) 485-508,https:// DOI: 10.1051/ro/2016037.

[46] Shavej Ali Siddiqui, *Optimization of interdependent queueing model with bulk service having variable Capacity*,International Release of Mathematics and Mathematical Sciences, Vol.1, No.1, pp.16-25, (2017).

[47] S.P Niranjan ,V.M Chandrasekaran and K Indhira, *Optimum cost analysis of batch service retrial queueing system with server failure, threshold and multiple vacations*,Songklanakarin Journal of Science and Technology,(2017)-0147.R2.

[48] Charan Jeet singh,Sandeep Kaur and Madhu Jain, *Performance analysis of bulk arrival queue with balking,optional service,delayed repair and multi-phase repair*,Ains Shams Engineering Journal,(2017), https://dx.doi.org/10.1016/j.asej.2016.08.025 .

[49] K.Santhi and R.Saravanan, *Performance Analysis of Cloud Computing Bulk Service Using Queueing Models*,International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6487-6492.

[50] S.P.Niraijan,V.M.Chandrasekaran and K.Indhira, *Queue size dependent service in bulk arrival queueing system with server loss and vacation break-off* ,Int. J. Knowledge Management in Tourism and Hospitality, Vol. 1, No. 2, 2017.

[51] T.S. Sinu Lal , A.Krishnamoorthy, and V.C.Joshua, *Queue with Batch Service, Batch Size Determined by Emergency Customers*,ceur-ws.org.(2017).

[52] K.Anitha, S.Maragathasundari and M.Bala, *Queuing System of Bulk Arrival Model With Optional Services in Third Stage and Two Different Vacation Policies*, International Journal of Mathematics And its Applications,Volume 5, Issue 4(2017), 711-721.

[53] M. Hemanth Kumar,M. Venkateswaran,T. Ganesh , P.R.S. Reddy, *Single Server Queueing Model with Non Homogeneous Compound Poisson Bulk Arrivals with Intervened Poisson Distribution*,Research and Reviews: Journal of Statistics,(2017) Volume 6, Issue 3, ISSN: 2348-7909.

[54] S.P. Niranjan,V.M. Chandrasekaran and K. Indhira, *State dependent arrival in bulk retrial queueing system with immediate Bernoulli feedback, multiple vacations and threshold*,IOP Conference Series: Materials Science and Engineering, 263 (2017) 042144,.https:// doi:10.1088/1757-899X/263/4/042144.

[55] S.P Niranjan , K.Indhira and V.M. Chandrasekaran , *Twofold Control Strategy of a Batch Service Queueing System With Vacation and Restriction on Number of Services for Renewal of Service Station*,International Journal of Pure and Applied Mathematics Volume 117 No. 13 (2017), 301-311.

[56] Djamila Zirem, Mohamed Boualem, Karima Adel-Aissanou and Djamil Aissani, *Analysis of a single server batch arrival unreliable queue with balking and general retrial time*,Quality Technology and Quantitative Management,(2018) ISSN: (Print) 1684-3703,https://doi.org/10.1080/16843703.2018.1510359.

[57] G.Ayyappan and T.Deepa, *Analysis of batch arrival bulk service queue with multiple vacation closedown essential and optional repair*,An International Journal of Applications and Applied Mathematics, Vol. 13, Issue 2 (2018), pp. 578 - 598.

[58] M.Thangaraj and P. Rajendran, *Analysis of Batch Arrival Bulk Service Queueing System with Breakdown, Different Vacation Policies and Multiphase Repair*,Advances in Algebra and Analysis (2018),https:// DOI: 10.1007/978-3-030-01120-8-30.

[59] S.P Niranjan, K.Indhira, and V.M.Chandrasekaran, *Analysis of bulk arrival queueing system with batch size dependent service and working vacation*,AIP Conference Proceedings 1952, 020061 (2018),https://doi.org/10.1063/1.5032023.

[60] S.Yahya Mohamed,N.Karthikeyan, *Analysis of bulk arrivals in queueing models*,International Research Journal Of Multidisciplinary Studies,(2018)Vol. 4, Issue 4, 2454-849.

[61] Hangguan Shan and Kunpeng Yang, *Analysis of Queueing Networks with Geo/Geo/1 Clients and Batch Service Queue*,International Conference on Wireless Communications and Signal Processing, (2018),https:// DOI: 10.1109/WCSP.2018.8555930.

[62] G.Ayyappan and R.Supraja, *Batch arrival bulk service queue with unreliable server, second optional service, two different vacations and restricted admissibility policy*,An International Journal of Applications and Applied Mathematics, Vol. 13, Issue 2 (2018),pp. 600 - 627.

[63] Alexander Zeifman,Rostislav Razumchik,Yacov Satin,Ksenia Kiseleva,Anna Korotysheva,Victor Korolev, *Bounds on the rate of convergence for one class of inhomogeneous Markovian queueing models with possible batch arrivals and services*,International Journal of Applied Mathematics and Computer Science,(2018).Vol. 28, No.1, 141-154,https://DOI: 10.2478/amcs-2018-0011.

[64] F.R.B.Cruz, M.A.C.Santos, F.L.P.Oliveira and R.C.Quinino,*Estimation in a general bulk arrival Markovian multi server finite queue*,Operational Research,(2018),https://doi.org/10.1007/s12351-018-0433y.

[65] Apoorv Saxena,Dieter Claeys,Herwig Bruneel,Bo Zhang,Joris Walraevens, *Modeling Data backups as a batch-service queue with vacations and exhaustive policy*,Computer Communications,(2018) 46-59, https://doi.org/10.1016/j.comcom.2018.07.014

[66] J.Ebenesar Anna Bagyam and K.Udaya Chandrika, *Multi-stage bulk arrival retrial queue with active breakdown and delayed repair*,International Journal of Pure and Applied Mathematics, Volume 119 No.12 2018, 16045-16058.

[67] Jitendra Kumar and Vikas Shinde, *Performance Evaluation Bulk Arrival and Bulk Service with Multi Server using Queue Model*,International Journal of Research in Advent Technology, Vol.6, No.11,(2018) E-ISSN: 2321-9637.

[68] K.Kirupa and K.Udaya Chandrika,*Unreliable Batch Arrival Retrial G-queue with Fluctuating Modes of Service, Preemptive Priority and Orbital Search*,International Journal of Mathematics Trends and Technology (IJMTT)(2018).

[69] F.P.Barbhuiya and U.C.Gupta,*A difference equation approach for analysing a batch service queue with the batch renewal arrival process*,Journal of Difference Equations and Applications(2019),ISSN: 1023-6198,https://doi.org/10.1080/10236198.2019.1567723.

[70] G.Ayyappan and S.Karpagam,*Analysis of a bulk queue with unreliable server,immediate feedback,N-Policy,Bernoulli Schedule multiple vacation and stand by server*,Ains Shams Engineering Journal (2019) 873-880,https://doi.org/10.1016/j.asej.2019.03.008.

[71] T.Deepa and A.Azhagappan,*Analysis of Batch Arrival Single and Bulk Service Queue with Multiple Vacation Closedown and Repair*,An International Journal of Applications and Applied Mathematics, Vol.14,Issue2 (2019),pp.617-639.

[72] R.Kalyanaraman and P.Nagarajan , *Bulk Arrival, Fixed Batch Service Queue With Unreliable Server, Bernoulli Vacation And With Delay Time*,AIP Conference Proceedings 2177, 020034 (2019),https://doi.org/10.1063/1.5135209.

[73] Atina Ahdika,*Mathematical model and numerical simulation of dual-channel multiphase bulk arrival queuing system with heterogeneous service time*,International Conference on Mathematics and Natural Sciences (2019), https://doi:10.1088/1742-6596/1127/1/012070.

[74] R.P Nithya and M.Haridass,*Modelling and simulation analysis of a bulk queueing system*,The International Journal of Systems and Cybernetics,Volume 50,(2019),pp.263-283(21),https://doi.org/10.1108/K-07-2018-0414.

[75] G.K.Gupta and A.Banerjee,*On Finite Buffer Bulk Arrival Bulk Service Queue with Queue Length and Batch Size Dependent Service*,International Journal of Applied and Computational Mathematics (2019),https://doi.org/10.1007/s40819-019-0617-z.

[76] G.K.Gupta, A.Banerjee and U.C.Gupta ,*On finite-buffer batch-size dependent bulk service queue with queue-length dependent vacation*,Quality Technology and Quantitative Management,(2019),https:// DOI: 10.1080/16843703.2019.1675568.

[77] A.Daw and J.Pender,*On the distributions of infinite server queues with batch arrivals*,Queueing Systems 91,367-401 (2019),https://doi.org/10.1007/s11134-019-09603-4.

[78] J. Baetens, B. Steyaert and D. Claeys, *System occupancy in a multiclass batch-service queueing system with limited variable service capacity*,Annals of Operations Research 293, 3-26 (2020), https://doi.org/10.1007/s10479-019-03470-1.

[79] M. Jain and S. Kaur, *(p,N)-Policy for Unreliable Server Bulk Queue with Bernoulli Feedback*,International Journal of Applied and Computational Mathematics 6,170(2020),https://doi.org/10.1007/s40819-020-00912-4.

[80] S.Pazhani Bala Murugan and R.Vijaykrishnaraj,*A Bulk Arrival Retrial Queue with Starting Failures and Exponentially Distributed Multiple Working Vacation*,, Journal of Xi'an University of Architecture and Technology,Volume XII,Issue II,(2020) ISSN No : 1006-7930.

[81] S.Karpagam,G.Ayyappan and B.Somasundaram,*A Bulk Queueing System with Rework in Manufacturing Industry with Starting Failure and Single Vacation*,International Journal of Applied and Computational Mathematics (2020),https://doi.org/10.1007/s40819-020-00922-2.

[82] R.Sree Parimala,*A New Investigation on Heterogeneous Bulk Service Queueing Model*,International Journal of Mathematics and Statistics Invention (2020) PP-47-51.

[83] S.Singh and R.K.Srivastava,*Markovian Queueing System with Bulk Arrival and Three Types of Essential and Optional Services*,International Journal of Mathematics And its Applications, 8(4)(2020), 103-112.ISSN: 2347-1557.

[84] G.K.Tamrakar and A.Banerjee,*Study on Infinite Buffer Batch Size Dependent Bulk Service Queue with Queue Length Dependent Vacation*,International Journal of Applied and Computational Mathematics (2021)7:252,https://doi.org/10.1007/s40819-021-01194-0.

[85] B.Bank and S.K.Samanta,*Performance Analysis of a Versatile Correlated Batch-arrival and Batch-service Queue*,Queueing Models and Service Management Vol. 4, No. 1, page 1-30, (2021).

[86] Bharathi Ramesh Kumar,Pothiraj Narayansamy and Subramanian Bhuvaneswari, *Optimal Control Analyze the Bulk arrival Queueing system using Estimation Theory*,Journal of Physics: Conference Series 1850 (2021) 012072, https://doi:10.1088/1742-6596/1850/1/012072.

[87] Sadhna Singh and R.K.Srivastava,*Markovian Queueing System for Bulk Arrival and Retrial Attempts with Multiple Vacation Policy*,International Journal of Mathematics Trends and Technology,Volume-67,Issue-1,21-28(2021),https:
DOI: 10.14445/22315373/IJMTT-V67I1P504.

[88] C.K.Deena Merit and M.Haridass,*Analysis of Flexible Batch Service Queueing System to Constrict Waiting Time of Customers*,Mathematical Problems in Engineering Volume (2021), Article ID-3601085,16, https://doi.org/10.1155/2021/3601085.

[89] G.Ayyappan and M.Nirmala,*Analysis of customers impatience on bulk service queueing system with unreliable server,setup time and two types of multiple vacations*, Int. J. Industrial and Systems Engineering, Vol.38, No.2,(2021),https:
DOI: 10.1504/IJISE.2021.115321.

[90] A. A. Bouchentouf, & A.Guendouzi ,*Single server batch arrival Bernoulli feedback queueing system with waiting server, K-variant vacations and impatient customers*,Operations Research Forum (Vol. 2, pp. 1-23)(2021).

[91] S.Pradhan and P.Karan,*Performance analysis of an infinite-buffer batch-size-dependent bulk service queue with server breakdown and multiple vacation*,Journal of Industrial & Management Optimization,vol.19,No.6,(2023).

[92] S. P.Niranjan,& S.D. Latha, *Analyzing the Two-Phase Heterogeneous and Batch Service Queuing System with Breakdown in Two-Phases, Feedback, and Vacation*,Baghdad Science Journal,(2024). https://doi.org/10.21123/bsj.2024.9126.

PROFIT ANALYSIS OF REPAIRABLE WARM STANDBY SYSTEM

[1] Shiv Kant, [2]Shashi Kant, [3*]Mohit Yadav, [4]Arunita Chaukiyal, [5]Bindu Jamwal

[1]Department of Computer Science and Engineering (AI & DS), GNIOT, Greater Noida
[2]Department of Computer Science and Engineering, Sharda University, Greater Noida
[3*]Department of Mathematics, University Institute of Sciences,
Chandigarh University, Mohali, Chandigarh
[4]Department of Computer Science, Acharya Narendra Dev College, University of Delhi
[5]School of Engineering and Technology, Raffles University, Neemrana
kaushikshivkant@gmail.com, shashi.kant1@sharda.ac.in, mohit.e15793@cumail.in,
arunita@yahoo.com, bindujamwal1810@gmail.com
*Corresponding Author

Abstract

In the generation of science and technology, every company wants to increase the reliability of their products. So, they used the concept of warm standby redundancy, timely repair of the failed unit. This paper aims to explore the system of two non identical units where the primary unit is operative and the secondary unit is in warm standby mode. When the primary unit fails due to any fault then secondary unit starts working immediately. Here, times of failure of unit and times of repair of unit follow general distributions. Such types of systems are used in companies to prevent losses. The system's behaviour is calculated by using concepts of mean time to system failure, availability, busy period of the server, expected number of visits made by the server and profit values using the semi Markov process and regenerative point technique. Tables are used to explore the performance of the system.

Keywords: Warm standby, non identical units, regenerative point, semi Markov process

I. Introduction

Reliability and maintainability are the essential parameters of items and products that satisfy customers' requirements. In today's era, several approaches for performance improvement of maintainable systems have been adopted by scientists and engineers during designing them. A large amount of research work has been done on repairable systems such that Jack and Murthy [4] discovered the role of limited warranty and extended warranty for the product. Wang and Zhang [8] examined the repairable system of two non identical components under repair facility using geometric distributions. Deswal and Malik [3] evaluate the non identical units system under different working conditions by using the semi Markov process. Malik and Rathee [7] threw light on the two parallel units system under preventive maintenance and maximum operation time. Levitin et al. [6] explored the results of optimal preventive replacement of failed units in a cold standby system by using the poisson process.

Agarwal et al. [1] described the reliability and availability of water reservoir system under repair facility. Chaudhary and Sharma [2] explored the parallel non identical units system that gives priority to repair over preventive maintenance. Jia et al. [5] explored the two unit system under demand and energy storage techniques.

II. System Assumptions

There are following system assumptions:

- Initially, the system has two non identical units such that one is operative and the other is warm standby.
- When the operative unit fails then the warm standby unit starts working.
- An expert repairman is always available to repair the failed unit.
- The failed unit behaves like a new one after repair.
- Repair times and failure time follow general distribution.

III. System Notations

There are following system notations:

R	Collection of regenerative states S_i $(i = 0,1,3)$
O / Ds	Operative unit / warm standby unit of the system
λ / w	Failure rate of the unit/ rate by which the server needs refreshment
$q_{r,s}(t)/Q_{r,s}(t)$	PDF/ CDF of first passage time from rth to sth regenerative state or sth failed state without halting in any other $S_i \in R$ in (0,1]
$M_r(t)$	Represents the probability of the system that it initially works $S_r \in R$ at a time (t) without moving through another state $S_i \in R$
$W_r(t)$	Probability that up to time (t) the server is busy at the state S_r without transit to another state $S_i \in R$ or before return to the same state through one or more non regenerative states
$\oplus/\otimes$	Laplace convolution / Laplace Stieltjes Convolution
$*/**/'$	Laplace Transform/ Laplace Stieltjes Transform/ Function's derivative
○ / ● / ▭	Upstate/ regenerative state/ failed state

IV. State Descriptions

The individual state description is given by the table 1:

Table 1: *State Descriptions*

States	Descriptions
S_0	It is a regenerative up state with two units - A_0 and B_s.
S_1	This regenerative up state has two units such that (A_{Fur}) and (B_0).
S_2	It is a down state where one unit is failed under repair continuously from previous state (A_{FUR}) and the other is failed under repair (B_{Fur}).
S_3	It is a regenerative up state with two units such that one is operative (A_0) and the other is failed under repair (B_{Fur}).
S_4	It is a down state where one unit fails under repair continuously from previous state (A_{FUR}) and the other unit is failed under repair (B_{Fur}).

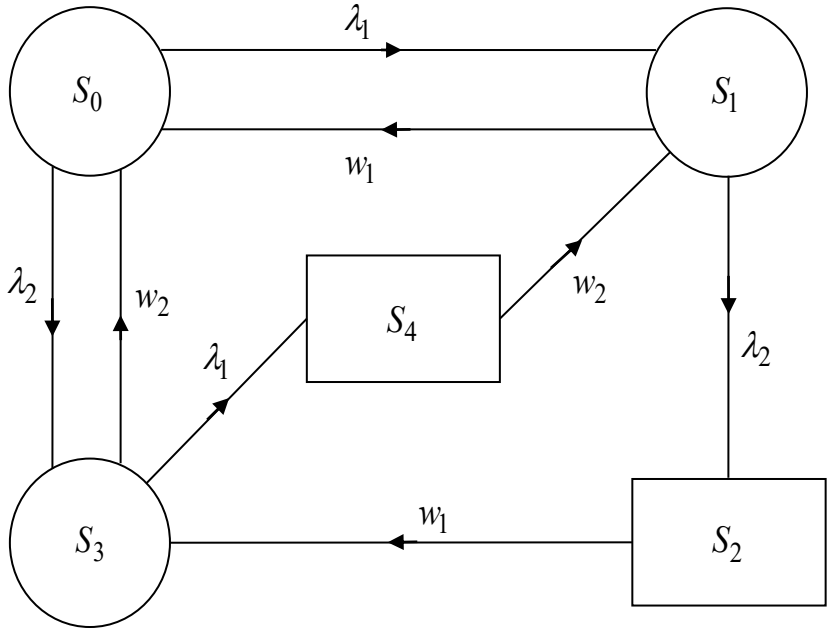


Figure 1: *State Transition Diagram*

Where, $S_0 = (A_o, B_s)$, $S_1 = (A_{Fur}, B_O)$, $S_2 = (A_{FUR}, B_{Fur})$
$S_3 = (A_O, B_{Fur})$, $S_4 = (A_{Fur}, B_{FUR})$

V. Transition Probabilities

The transition probabilities are calculated

$$p_{0,1} = \frac{\lambda_1}{\lambda_1 + \lambda_2},\ p_{0,3} = \frac{\lambda_2}{\lambda_1 + \lambda_2},\ p_{1,0} = \frac{w_1}{w_1 + \lambda_2},\ p_{1,2} = \frac{\lambda_2}{w_1 + \lambda_2}$$

$$p_{3,0} = \frac{w_2}{\lambda_1 + w_2},\ p_{3,4} = \frac{\lambda_1}{\lambda_1 + w_2},\ p_{2,3} = p_{4,1} = 1 \quad (1)$$

It has been conclusively established that

$$p_{0,1} + p_{0,3} = 1,\ p_{1,0} + p_{1,2} = 1,\ p_{3,0} + p_{3,4} = 1 \quad (2)$$

VI. Mean Sojourn Time

Let μ_i represents the mean sojourn time. Mathematically, time consumed by a system in a particular state is, $\mu_i = \sum_j m_{i,j} = \int_0^\infty P(T > t)\,dt$. Then

$$\mu_0 = m_{0,1} + m_{0,3} = \frac{1}{\lambda_1 + \lambda_2},\ \mu_1 = m_{1,0} + m_{1,2} = \frac{1}{w_1 + \lambda_2},\ \mu_2 = m_{2,3} = \frac{1}{w_1}$$

$$\mu_3 = m_{3,0} + m_{3,4} = \frac{1}{\lambda_1 + w_2},\ \mu_4 = m_{4,1} = \frac{1}{w_2}$$

$$\mu'_1 = m_{1,0} + m_{1,3.2} = \frac{1}{w_1 + \lambda_2},\ \mu'_3 = m_{3,0} + m_{3,1.4} = \frac{1}{\lambda_1 + w_2} \quad (3)$$

VII. Reliability Measures Evaluations

I. Mean Time to System Failure (MTSF)

Let the cumulative distribution function of the first elapsed time be $\varphi_i(t)$ from the regenerative state S_i to the failed state of the system. Treating the failed states as an absorbing state then the repetitive interface for $\varphi_i(t)$ being

$$\varphi_0(t) = Q_{0,1}(t) \otimes \varphi_1(t) + Q_{0,3}(t) \otimes \varphi_3(t)$$
$$\varphi_1(t) = Q_{1,0}(t) \otimes \varphi_0(t) + Q_{1,2}(t)$$
$$\varphi_3(t) = Q_{3,0}(t) \otimes \varphi_0(t) + Q_{3,4}(t) \quad (4)$$

Taking LST of the relation (4) and solving for $\varphi_0^{**}(s)$ then

$$MTSF = \lim_{s\to 0} \frac{1-\varphi_0^{**}(s)}{s} \quad (5)$$

System reliability can be obtained by using the inverse LT of equation (5). We have

$$MTSF = \frac{\mu_0 + p_{0,1}\mu_1 + p_{0,3}\mu_3}{1 - p_{0,1}p_{1,0} - p_{0,3}p_{3,0}} \quad (6)$$

II. Availability of the system

From the transition diagram, the system is available at the regenerative up states S_0, S_1 and S_2. Let $A_i(t)$ is the probability that the system is in upstate at time (t) specified that the system arrives at the regenerative state S_i at t = 0. Then the repetitive interface for $A_i(t)$ is

$$A_0(t) = M_0(t) + q_{0,1}(t) \oplus A_1(t) + q_{0,3}(t) \oplus A_3(t)$$
$$A_1(t) = M_1(t) + q_{1,0}(t) \oplus A_0(t) + q_{13.2}(t) \oplus A_3(t)$$
$$A_3(t) = M_3(t) + q_{3,0}(t) \oplus A_0(t) + q_{3,1.4}(t) \oplus A_1(t) \quad (7)$$

where, $M_0(t) = e^{-(\lambda_1+\lambda_2)t}$, $M_1(t) = e^{-(w_1+\lambda_2)t}$, $M_3(t) = e^{-(\lambda_1+w_2)t}$ (8)

Using LT of the above relation (7), there exist

$$\therefore A_0 = \lim_{s\to 0} \frac{N_A}{D_1'} = \frac{\mu_0[1 - p_{1,3}p_{3,1}] + \mu_1[p_{0,1} + p_{0,3}p_{3,1}] + \mu_3[p_{0,3} + p_{0,1}p_{1,3}]}{\mu_0[1 - p_{1,3}p_{3,1}] + \mu'_1[p_{0,1} + p_{0,3}p_{3,1}] + \mu'_3[p_{0,3} + p_{0,1}p_{1,3}]} \quad (9)$$

III. Busy Period of the Server

Let $B_i(t)$ is the probability that the repairman is busy due to the repair of the failed unit at time 't' specified that the system arrives at the regenerative state S_i at t = 0. Then the repetitive interface for $B_i(t)$ is

$$B_0(t) = q_{0,1}(t) \oplus B_1(t) + q_{0,3}(t) \oplus B_3(t)$$
$$B_1(t) = W_1(t) + q_{1,0}(t) \oplus B_0(t) + q_{1,3.2}(t) \oplus B_3(t)$$
$$B_3(t) = W_3(t) + q_{3,0}(t) \oplus B_0(t) + q_{31.4}(t) \oplus B_1(t) \quad (10)$$

where, $W_1(t) = w_1 e^{-(w_1+\lambda_2)t} + ...$

$$W_3(t) = w_2 e^{-(\lambda_1+w_2)t} + ... \quad (11)$$

Using LT on relations (10) then we get

$$B_0 = \lim_{s\to 0} \frac{N_B}{D_1'} = \frac{\mu'_1[p_{0,1} + p_{0,3}p_{3,1}] + \mu'_3[p_{0,3} + p_{0,1}p_{1,3}]}{\mu_0[1 - p_{1,3}p_{3,1}] + \mu'_1[p_{0,1} + p_{0,3}p_{3,1}] + \mu'_3[p_{0,3} + p_{0,1}p_{1,3}]} \quad (12)$$

IV. Estimated number of visits made by the server

Let $N_i(t)$ is the estimated number of visits made by the repairman for repair in (0, t] specified that the system arrives at the regenerative state S_i at t = 0. Then the repetitive interface for $N_i(t)$ is

$$N_0(t) = Q_{0,1}(t) \otimes [1 + N_1(t)] + Q_{0,3}(t) \otimes [1 + N_3(t)]$$
$$N_1(t) = Q_{1,0}(t) \otimes N_0(t) + Q_{1,3.2}(t) \otimes N_3(t)$$
$$N_3(t) = Q_{30}(t) \otimes N_0(t) + Q_{31.4}(t) \otimes N_1(t) \quad (13)$$

Using LST of the above relations (13) then we get

$$N_0 = \lim_{s\to 0} \frac{N_v}{D_1'} = \frac{1 + p_{1,3}p_{3,1}}{\left[\begin{matrix}\mu_0[1 - p_{1,3}p_{3,1}] + \mu'_1[p_{0,1} + p_{0,3}p_{3,1}] \\ + \mu'_3[p_{0,3} + p_{0,1}p_{1,3}]\end{matrix}\right]} \quad (14)$$

V. Profit Analysis

The profit function of the system is defined by

$$P = T_0 A_0 - T_1 B_0 - T_2 N_0 \quad (15)$$

where, $T_0 = 5000$ (Revenue per unit up-time)

$T_1 = 600$ (Charge per unit for server busy period)

$T_2 = 100$ (Charge per visit made by the server)

VIII. Discussion

Let $\lambda_1 = \lambda_2 = \lambda$ and $w_1 = w_2 = w$ then the reliability measures like MTSF, availability of the system, busy period of the server, expected number of visits made by the server and profit

Table 2: *MTSF vs. Repair Rate*

w ↓	λ=0.55	λ=0.6,	λ=0.65
0.55	4.76082	4.413408	4.192308
0.6	4.929006	4.585448	4.290541
0.65	5.063985	4.728682	4.36747
0.7	5.174709	4.849785	4.429348
0.75	5.267176	4.953519	4.480198
0.8	5.345557	5.043371	4.522727
0.85	5.412844	5.121951	4.558824
0.9	5.471236	5.191257	4.589844
0.95	5.522388	5.252838	4.616788
0.1	5.567568	5.307918	4.640411

Table 3: *Availability vs. Repair Rate*

w ↓	λ=0.55	λ=0.65,	λ=0.6
0.55	0.686403	0.670466	0.657986
0.6	0.69355	0.678566	0.662905
0.65	0.699041	0.684997	0.666652
0.7	0.703393	0.690227	0.669601
0.75	0.706926	0.694563	0.671982
0.8	0.709852	0.698216	0.673945
0.85	0.712315	0.701336	0.675592
0.9	0.714416	0.704032	0.676992
0.95	0.716231	0.706384	0.678198
0.1	0.717813	0.708454	0.679247

Table 4: *Profit vs. Repair Rate*

w ↓	λ=0.55	λ=0.65,	λ=0.6
0.55	2722.902	2572.193	2533.318
0.6	2788.908	2650.579	2583.419
0.65	2838.719	2711.387	2620.861
0.7	2877.66	2759.963	2649.913
0.75	2908.947	2799.676	2673.117
0.8	2934.638	2832.757	2692.08
0.85	2956.115	2860.746	2707.868
0.9	2974.335	2884.737	2721.218
0.95	2989.99	2905.531	2732.655
0.1	3003.585	2923.73	2742.563

values are calculated. It can be seen from the tables 2, 3 and 4 that the tendency of MTSF, availability of system and profit values increase smoothly with respect to increments in repair rate(θ) whereas these values declines corresponding to increment in failure rate.

IX. Conclusion

It is calculated that the MTSF, availability and profit values of the two non identical unit system increase with respect to increments in repair rate but these reliability values decline when failure rate of unit is enhanced. The idea of repair facility is used by corporate sectors, industries, cybercafés, education, university systems, etc.

References

[1] Agrawal, A., Garg, D., Kumar, A. and Kumar, R. (2021). Performance analysis of the water treatment reverses osmosis plant. *Reliability Theory and Applications*, *16*(3): 16-25.

[2] Chaudhary, P. and Sharma, A. (2022). A two non-identical unit parallel system with priority in repair. *Reliability Theory and Applications*, *18*(3 (74)), 308-315.

[3] Deswal, S. and Malik, S. C. (2015). Reliability measures of a system of two non-identical units with priority subject to weather conditions. *Journal of Reliability and Statistical Studies*, 181-190.

[4] Jack, N., and Murthy, D. P. (2007). A flexible extended warranty and related optimal strategies. *Journal of the Operational Research Society*, 58(12), 1612–1620.

[5] Jia, H., Peng, R., Yang, L., Wu, T., Liu, D., and Li, Y. (2022). Reliability evaluation of demand based warm standby system with capacity storage.

[6] Levitin, G., Finkelstein, M., and Dai, Y. (2020). Optimal preventive replacement policy for homogeneous cold standby systems with reusable elements. *Reliability Engineering and System Safety*, *204*, 107135.

[7] Malik, S. C. and Rathee, R. (2016). Reliability modelling of a parallel system with maximum operation and repair times. *International Journal of Operational Research*, 25(1), 131-142.

[8] Wang, G. J. and Zhang, Y. L. (2007). An optimal replacement policy for repairable cold standby system with priority in use. *International Journal of Systems Science*, *38*(12), 1021-1027.