

ENHANCING LINDLEY DISTRIBUTION PARAMETER ESTIMATION WITH HYBRID BAYESIAN AVERAGE MODEL FOR FUZZY DATA

ABBARAPU ASHOK AND NADIMINTI NAGAMANI*

Department of Mathematics, School of Advanced Sciences, VIT-AP University, Amaravati, India.

*maniinadiminti999@gmail.com

Abstract

With the ultimate goal of increasing parameter estimate accuracy, this study will examine and assess a number of estimating techniques used with the Lindley distribution in the context of fuzzy data. Gibbs sampling, Bootstrapping Sampling, MCMC, MH, and a unique hybrid methodology that combines these approaches via Bayesian model averaging were also studied. The research looks at several sample sizes ranging from 15 to 100 and repeats the estimate method 10,000 times for each size. Fuzzy data are created using established fuzzy systems, and the performance of each approach is measured using average values (AV), mean squared errors (MSE), coverage probabilities, and confidence interval lengths. The findings show that the hybrid technique consistently produces estimates closer to the genuine parameter value of one across all sample sizes, with smaller mean squared errors than individual methods. Furthermore, the hybrid method's confidence intervals preserve coverage probabilities that are consistent with the targeted confidence level, demonstrating the method's trustworthiness in statistical inference. Overall, the results show that the hybrid technique improves estimate accuracy and reliability, providing a strong foundation for parameter estimation in the Lindley distribution framework using fuzzy data.

Keywords: MCMC, MH, lindley distribution, gibbs Sampling, bootstrap sampling, bayesian model average technique.

1. INTRODUCTION

The Lindley distribution was first presented by [1] as a novel distribution has been developed to simplify lifespan data analysis, particularly in applications that include modeling the dependability of stress strength. Various authors have utilized complete and censored samples to tackle inferential challenges related to the Lindley distribution parameter. For example, a rigorous mathematical methodology to examine the Lindley distribution's properties used by [2]. In addition, through the use of a numerical example, they proved that modeling with the Lindley distribution beats modeling with the exponential distribution in terms of efficiency. The primary subjects of discussion were the estimation of reliability using an expanding type II right censored sample along with the Lindley distribution by [3]. When the Lindley distribution is used to spread the causes of failure, a rival risk model investigated by [4]. The Lindley distribution assumption states that, both traditional and Bayesian techniques to examine the hybrid censored lifespan data by [5]. Recently, numerous research has been published regarding the use of traditional statistical methods for analyzing fuzzy data. Bayesian estimation is used to examine lifetime data [6] in situations that lack clarity. Fuzzy random variable theory was examined in hindsight by Gil, Lopez-Diaz, and Ralescu, focusing on its modeling, interpretation, along with resulting implications. The use of traditional statistical inference techniques to univariate fuzzy data was

investigated by [7]. Regarding complete and refined datasets, investigated the use of ambiguous set theory in Bayesian failure rate along with mean time to failure estimate investigated by [8]. The Lindley distribution is characterized by its probability density function.

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

In this study, we examine several approaches for estimating the Lindley distribution parameter in cases where fuzzy numbers are used to characterize the experimental data. First, how fuzzy data are created from imprecise observations is explained, and then the method for calculating the greatest probability estimate of the parameter β is discussed. When working with fuzzy data, the maximum likelihood estimate can be obtained using the MCMC, MH, Gibbs, and bootstrap methods because the MLE lacks a closed form. Furthermore, the estimated confidence interval of the unknown parameter is calculated using the asymptotic distribution of the maximum likelihood estimators (MLEs). We also look at the Lindley distribution's parameter using the Bayesian average model. Since the Bayes estimate cannot be obtained directly, we Estimate it using several techniques such as Gibbs sampling, Bootstrapping Sampling, Markov Chain Monte Carlo (MCMC), and Metropolis-Hastings (MH). The highest posterior density (HPD) credible interval of the parameter is also calculated using similar approaches.

2. LITERATURE REVIEW

A variety of techniques for model selection and parameter estimation utilizing fuzzy and stochastic data from various distributions are presented in the reviewed literature. Many research concentrate on distributions such as Weibull, Rayleigh, Lindley, and Inverse Lindley, and Maximum Likelihood (ML) and Bayesian techniques are frequently used. For the Weibull distribution, Bayesian and machine learning techniques with Newton-Raphson and Expectation-Maximization algorithms by [9] however, it does not compare with other fuzzy data estimate techniques. Although there are still issues with practical implementation, fuzzy Kullback Leibler (f-KL) divergence, which offers flexibility in parameter estimation presented by [10]. For the Rayleigh distribution, fuzzy parameter estimation but it doesn't outline the estimate process addressed by [11], which restricts its use. In order to accommodate stochastic and fuzzy data, generalized estimators are presented in [12] for a variety of lifespan distributions. This introduces a unique technique that requires additional performance assessment versus current approaches. Although there are little information on network construction, neural networks and compares neural-based estimators for the Weibull distribution with classical approaches examines by [13]. In many research, Bayesian methods are widely used, particularly in Lindley and Generalized Lindley distributions. With innovations like squared error loss functions and novel estimators, respectively, studies [14] and [15] use Bayesian and machine learning techniques to achieve competitive results, however they are limited to particular data circumstances. Maximum Product Spacing (MPS) is presented for the Inverse Lindley distribution in [16]. It works well but has a narrow application. While acknowledging the higher computing requirements, the Bayesian technique to Lindley distributions for hypothesis testing expanded by [17]. Model selection is the primary focus of [18]. Bayesian approaches help identify the best-fitting models for fuzzy data under the Lindley distribution when proper prior selection is used. A hybrid Bayesian-Bootstrap method in [19] enhances the generation of confidence intervals but increases computing costs. For certain fuzzy data situations, the Lindley distribution's benefit is demonstrated by comparative Bayesian analyses in [20], while an Empirical Bayes technique in [21], provides a computationally effective substitute. Further refinements in Bayesian methods include the use of informative priors [22] for improved model selection and parameter estimation in Lindley distributions. Finally, fuzzy logic and Bayesian techniques are combined in [23] for classification problems. This approach shows potential but needs more extensive validation on real-world datasets. Each method presents unique strengths and challenges, with limitations often centered on computational efficiency, distribution-specific applications, and the need for extensive comparison across methodologies.

2.1. Research gap

The literature on parameter estimation with fuzzy data lacks comprehensive comparison and integration of methodologies across different distributions. The hybrid Bayesian average model addresses this gap by integrating Bayesian inference with bootstrap resampling, providing robust parameter estimation and confidence interval construction. This model improves coverage probabilities, adaptability to various distributions, and mitigates computational cost by combining Bayesian inference with bootstrap resampling. It is more feasible for practical applications, especially when dealing with large datasets or complex models. The hybrid Bayesian average model offers improved accuracy, reliability, and computational efficiency compared to existing methods, making it a valuable solution for parameter estimation in fuzzy data analysis.

3. SIMULATION AND COMPARISON STUDIES

3.1. Bayesian Average Modelling

This study used the average model to estimate the parameters. Updating previous knowledge based on observed data and combining it coherently is made possible by the Bayesian framework. When working with sparse or ambiguous data, the Bayesian average model comes in handy because it provides reliable parameter estimates.

Let $M = (M_1, \dots, M_K)$ represent the set of models being studied. A model can be characterized by various attributes, such as the form of the error variance or the specific collection of explanatory factors that are part of the model. The posterior distribution of Δ , which represents the quantity of interest (such as a future observable or a model parameter), is determined by data Z . Additionally, it is as follows.

$$p(\Delta|Z) = \sum_{i=1}^K p(\Delta|Z, M_k)p(M_k|Z)$$

This is the average posterior predictive distribution value for Δ across all considered models, considering the likelihood of the associated posterior model. The posterior probability is the likelihood of model M_k based on the data that is already available.

$$p(M_k|Z) = \frac{p(Z|M_k)p(M_k)}{\sum_{i=1}^K p(Z|M_i)p(M_i)}$$

where,

$$p(Z|M_k) = \int \dots \int p(Z|\theta_k, M_k)p(\theta_k|M_k)d\theta_k$$

is Model M_k 's parameter vector, θ_k represents the integrated likelihood of the model; $p(\theta_k|M_k)$ represents the prior density of parameters of M_k ; $p(Z|\theta_k, M_k) p(M_k)$ stands for the previous probability of M_k 's genuineness, and reflects the likelihood. The collection of all models that are taken into consideration, M implicitly depends on each probability.

Applying the above discussed concepts simply results in parameter estimates and other quantities of interest. A parameter θ may be estimated using the Bayesian model averaging (BMA) method is,

$$\theta_{BMA} = \sum_{k=1}^K \hat{\theta}_k(M_k|Z)$$

where, $\hat{\theta}_k$ represents model k 's posterior mean. There may also be accessible alternative amounts and variances of these estimations.

Bayesian model averaging (BMA) presents many issues. These tasks encompass determining the prior model probabilities $p(M_k)$, computing for numerous models, and evaluating the integrals that are frequently not solvable using standard mathematical expressions.

Algorithm 1 Hybrid Bayesian Average Model

Require: data, num_iterations, num_bootstrap_iterations, sample_size

Define function MCMC(data, num_iterations):

Initialize $\theta \leftarrow 0$

Initialize chain \leftarrow empty list

FOR each iteration from 1 to num_iterations **DO**

Draw θ_{new} from a normal distribution with mean θ and standard deviation 0.1

Compute acceptance probability $\alpha \leftarrow \min(1, \frac{\text{probability density at } \theta_{new}}{\text{probability density at } \theta})$

IF a random number between 0 and 1 is less than α **THEN**

Set $\theta \leftarrow \theta_{new}$

Add θ to chain

END FOR

RETURN chain

Define function MH(data, num_iterations):

Initialize $\theta \leftarrow 0$

Initialize chain \leftarrow empty list

FOR each iteration from 1 to num_iterations **DO**

Draw θ_{new} from a normal distribution with mean θ and standard deviation 0.1

Compute acceptance probability $\alpha \leftarrow \min(1, \frac{\text{probability density at } \theta_{new}}{\text{probability density at } \theta})$

IF a random number between 0 and 1 is less than α **THEN**

Set $\theta \leftarrow \theta_{new}$

Add θ to chain

END FOR

RETURN chain

Define function Gibbs(data, num_iterations):

Initialize $\theta \leftarrow 0$

Initialize chain \leftarrow empty list

FOR each iteration from 1 to num_iterations **DO**

Draw θ from a normal distribution with mean equal to data mean and standard deviation equal to data std

Add θ to chain

END FOR

RETURN chain

Define function Bootstrapping(data, num_bootstrap_iterations, sample_size):

Initialize bootstrapped_samples \leftarrow empty list

FOR each iteration from 1 to num_bootstrap_iterations **DO**

Randomly select sample_size indices from data with replacement

Create bootstrapped_sample by selecting elements of data at these indices

Add bootstrapped_sample to bootstrapped_samples

END FOR

RETURN bootstrapped_samples

Define function Bayesian_Average(parameters):

RETURN mean of parameters

Define function run_hybrid_model(data, num_iterations, num_bootstrap_iterations, sample_size):

Initialize parameters \leftarrow empty list

FOR each iteration from 1 to num_iterations **DO**

Obtain MCMC chain using MCMC(data, num_iterations)

Obtain MH chain using MH(data, num_iterations)

Add results to parameters list

END FOR

RETURN Bayesian_Average(parameters)

3.2. Gibbs sampling

A series of samples from a joint probability distribution may be created iteratively using the Gibbs sampling technique. When using such a sequence, the goal is to compute an integral (like an anticipated value) or estimate the joint distribution (like a histogram). Gibbs sampling is suitable when the individual conditional distributions of each variable are known, but the joint distribution is not explicitly known. Using the current values of the other variables as a conditioning factor, the Gibbs sampling technique is used to iteratively produce a sample from each variable's distribution. The samples can be demonstrated to constitute a Markov chain, and the stationary distribution of that chain precisely corresponds to the joint distribution that is required. Gibbs sampling is a highly suitable approach for sampling the posterior distribution of a Bayesian network, which is typically characterized as a set of conditional distributions.

3.2.1 The Gibbs sampler

In the context of picture restoration using the MCMC technique, Gibbs sampling is a flexible technique for fitting statistical models [24]. This method is based on the recommendations of Tanner and Wong (1987) for replacement sampling using data augmentation.

Assume that we divide λ into r blocks, such that $\lambda = (1, 2, \dots, r)$. Considering that λ is now in the state $\lambda^{(t)}$, let's say we make the transition in the manner described below:

$$\begin{aligned} &\text{Draw } \lambda_1^{(t+1)} \text{ from } h(\lambda_1 | \lambda_2^{(t)}, \dots, \lambda_r^{(t)}), \\ &\text{Draw } \lambda_2^{(t+1)} \text{ from } h(\lambda_2 | \lambda_1^{(t+1)}, \lambda_3^{(t)}, \dots, \lambda_r^{(t)}), \\ &\quad \vdots \\ &\text{Draw } \lambda_r^{(t+1)} \text{ from } h(\lambda_r | \lambda_1^{(t+1)}, \dots, \lambda_{r-1}^{(t+1)}). \end{aligned}$$

We call the distributions h as conditional distributions in their whole. By executing one full cycle of the Gibbs sampler, we may update the whole vector λ by changing each of the r blocks in the manner illustrated. The selection of components lower-dimensional blocks has assumed the role.

3.3. Bootstrap sampling

A potent resampling method for estimating a statistic's sampling distribution in statistics is called a bootstrap sample. It is beneficial when standard parametric approaches are not appropriate or when the underlying distribution of the data is unclear. Bootstrapping is the process of repeatedly to generate a large number of bootstrap samples, replacement samples are chosen from the observable data. The distribution of the statistic's sample may be approximated using these samples, including the mean, variance, and regression coefficients. The fundamental tenet of bootstrap sampling is that the data's empirical distribution can be utilized to approximate the actual population distribution.

3.3.1 Procedure for Bootstrap sampling

The sampling distributions of sample statistics are fundamental for statistical inference. Primarily, the bootstrap method enables one to estimate the sampling distribution based on a single sample, although with some degree of approximation. The following is an explanation of how it functions.

Step 1: Perform resampling. Generate several more samples, referred to as bootstrap samples or resamples, by randomly selecting and replacing elements from the first random sample. Each sample is of same magnitude as the initial random sample.

Sampling with replacement involves randomly selecting one observation from the original sample

and returning it before selecting the next observation. This is like to randomly selecting a number from a hat, and then returning it before making another selection. Consequently, any number may be selected once, several times, or not at all. If we conducted a replacement sampling, we would get the identical set of numbers that we initially had, only in a new sequence. Practically, we would begin with the whole original sample, rather than only six observations, and generate several resamples, rather than just three.

Step 2: Compute the bootstrap distribution. Compute the statistical measures for each individual sample. The collection of these resample statistics is referred to as the bootstrap distribution. If we wish to calculate the average repair time of the population, the statistic we use is the sample mean.

Step 3: Utilize the bootstrap distribution. The bootstrap distribution provides insights on the form, central tendency, and the statistic's sample distribution's variability. The bootstrap standard error may be defined as the standard deviation of the bootstrap distribution associated with a given statistic. If the specific statistic being considered is the sample mean, then the bootstrap standard error may be calculated using resamples is

$$SE_{(\text{boot}, \bar{X})} = \sqrt{\frac{1}{B-1} \sum \left(X - \frac{1}{B} \sum \bar{X}^* \right)^2}$$

in this formula, \bar{X}^* represents the average value of a single resample. Simply taking the standard deviation of the B values yields the bootstrap standard error. The asterisk in * distinguishes between the original sample mean and the resample mean.

3.4. MH Algorithm

The algorithm of Metropolis-Hastings [25, 26] is a Samples from the posterior distribution $p(\lambda|X)$ may be generated by the Markov chain Monte Carlo approach. The algorithmTMs starting vector value is denoted by λ_0 . Afterwards, a sequence of N parameter vectors, λ_i for $i = 1, N$, are produced as follows:

- a. Create a proposal distribution $q(\lambda^*|\lambda_{i-1})$, such as a Normal distribution with λ_{i-1} as the mean, and use it to generate a candidate parameter vector λ^* .

- b. Calculate

$$T = \frac{p(Y|\lambda^*)\lambda(\lambda^*)q(\lambda_{i-1}|\lambda^*)}{p(Y|\lambda_{i-1})\lambda(\lambda_{i-1})q(\lambda^*|\lambda_{i-1})}$$

where $p(Y|\lambda^*)$ and the prior densities of λ^* and λ_{i-1} , respectively, and the likelihood values of the parameter vectors λ^* and λ_{i-1} , respectively.

- c. Assuming that u is chosen at random from a uniform distribution over the interval $(0, 1)$ and if T , is greater than u , then $\lambda_i = \lambda^*$; otherwise, $\lambda_i = \lambda_{i-1}$.
- d. Following an initial phase of, letTMs say, M iterations, the resultant chain will converge to a chain whose members are randomly selected from the posterior parameter distribution $p(\lambda|X)$. Eliminate the first M iterations as much as possible.
- e. The beginning value λ_0 , the proposal distribution $q(\lambda^*|\lambda_{i-1})$. Before using the Metropolis-Hastings method, you must ascertain the overall iteration count N as well as the number of rejected iterations M . The criteria for selecting these parts still need more investigation.
- f. It is possible to summarize many aspects of the posterior parameter distribution using the collected sample. Using the sample of parameter vectors λ_j with $j = M + 1, \dots, N$, one may calculate the posterior means to minimize expected quadratic loss,

$$\bar{\lambda} = \frac{1}{N - M} \sum_{j=M+1}^N \lambda_j$$

Next, the vector $\bar{\lambda}$ can be viewed as a model parameter estimate. Comparing parameter estimates to the complete posterior probability distribution, one must forfeit some information. The posterior variances, parameter correlations, and model prediction distribution can all be computed using the sample of parameter vectors.

3.5. Markov Chain Monte Carlo Methods (MCMC)

Let $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ be a vector that includes every variable for which we require probability density functions. This indicates that we want to discover $p(\lambda_i|Y)$, where Y is the collection of measurements, for each variable λ_i . Applying the Bayes theorem, we get

$$p(\lambda_i|Y) = \frac{p(Y|\lambda_i)p(\lambda_i)}{p(Y)} = \frac{p(Y|\lambda_i)p(\lambda_i)}{\int_{\theta_i} p(Y|\lambda_i)p(\lambda_i)d\lambda_i}$$

The term $p(\lambda_i|Y)$ is also referred to as the target distribution or posterior because this is the distribution we are attempting to estimate. The likelihood function, $p(Y|\lambda)$, indicates the likelihood that the data was generated with the given set of parameters. It is represented by the term $p(Y)$, which is the prior distribution of λ . If the denominator is not tractable, it is not possible to give an explicit analytic expression of the probability density function. It is possible that the formula is tractable for some λ_i but not for others, which means that another approach may need to be used for estimating each variable has few different possibilities.

1. The closed analytic form of $p(\lambda|Y)$ can be written as a conventional probability distribution that can be directly sampled, or in the event that this is not possible, sampled via the inverse transform sampling method.
2. Although there are some variables, λ_{new} such that the expressions $p(\lambda_1|\lambda_{new} \dots Y)$ and $p(\lambda_{new}|\lambda_1, \lambda_i, Y)$ can be in closed analytic form, $p(\lambda|Y)$ cannot be written in a closed analytic form. Drawing λ_i from $p(\lambda_i|\lambda_i)$ and λ_0 from $p(\lambda|\lambda_1, Y)$ can then be used to sample $p(\lambda|Y)$.
3. Although the proportionality of the joint posterior density $p(\lambda_1, \lambda_{new}, \dots, \lambda_i)$ is known, no closed analytic equation for $p(\lambda|Y)$ has been established. The so-called Metropolis-Hastings algorithm, one specific application of the MCMC method, which stands for Markov Chain Monte Carlo, is applicable here.

4. PERFORMANCE EVALUATION

In order to determine which estimating method is the most effective, performance measurements like average variance (AV), mean squared error (MSE), coverage probabilities, and interval lengths are utilized. This will allow for the identification of any trends or patterns that may exist across a range of sample sizes, as well as the provision of insights into the relative effectiveness of the techniques.

4.1. Average values(AV)

The following mathematical formula can be used to determine the average values (AV) in the context of performance metrics,

$$AV = \frac{1}{N} \sum_{i=1}^N x_i$$

where, AV represents the average value.
 N represents the overall count of observations.
 x_i stands for every single observation.

In the context of the research on estimation techniques for the Lindley distribution with fuzzy

data, the average values (AV) are calculated to represent the central tendency of the estimation results obtained from different techniques. The AV provides insight into how closely the estimates align with the true parameter value, which in this case is 1.

4.2. Mean squared error(MSE)

By averaging the squares of the errors between estimated and true values, the mean squared error (MSE) is a frequently used metric to assess the accuracy of an estimator. This mathematical formula is used to compute the MSE

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_i)^2$$

where, Mean squared error is abbreviated as MSE.

N represents the total amount of observations.

x_i is the actual value. \bar{x}_i represents the estimated value.

4.3. Coverage probability

A statistical metric called coverage probability is employed to evaluate the dependability of confidence intervals. It displays the proportion of confidence intervals containing the parameter's true value. Coverage probability is computed as follows in the context of studies on fuzzy data estimate methods for the Lindley distribution.

$$\text{Coverage Probability} = \frac{\text{Number of Confidence Intervals that contain the True Value}}{\text{Total Number of Confidence Intervals}}$$

This measure provides insights into the effectiveness of the estimation techniques in providing confidence intervals that capture the true parameter value. A coverage probability close to the desired confidence level indicates that the estimation technique is providing reliable and accurate confidence intervals.

4.4. Interval Length

The width or range of a confidence interval, which expresses the accuracy of the estimation, is referred to as the interval length. It is computed as the absolute difference between the confidence interval's upper and lower boundaries. In the context of the research on estimation techniques for the Lindley distribution with fuzzy data, interval length can be calculated as follows

$$\text{Interval Length} = \text{Upper Bound} - \text{Lower Bound}$$

This measure provides insights into how narrow or wide the confidence intervals are, which reflects the precision of the estimation. Smaller interval lengths indicate higher precision, while larger interval lengths indicate lower precision.

5. RESULTS AND DISCUSSION

5.1. Results

We provide the results of our investigation on the performance of several estimate approaches applied to the Lindley distribution with fuzzy data in the results section. We compare the efficiency of the following techniques: Gibbs sampling, Bootstrapping Sampling, Markov Chain Monte Carlo (MCMC), and Metropolis-Hastings (MH). Our investigation covers a range of sample sizes, from 15 to 100, with each estimating step performed 10,000 times. The *fuzzification*

procedure uses preset fuzzy systems $\{\tilde{x}_1, \dots, \tilde{x}_8\}$ to alter produced samples to account for dataset uncertainty. Each membership function, $\mu_{\tilde{x}_i}(x)$, is designed to reflect the data's fuzzy properties correctly. This section tries to highlight any trends or patterns discovered across various sample sizes, as well as give insights into the approaches' comparative performance. Furthermore, we take into account minor fluctuations in x -values to guarantee that uncertainties in fuzzy data are accurately represented, improving the integrity of our analysis and aiding researchers in picking the best methodologies for statistical inference.

$$\mu_{\tilde{x}_1}(x) = \begin{cases} 1, & \text{if } x \leq 0.13 \\ \frac{0.33 - x}{0.2}, & \text{if } 0.13 \leq x \leq 0.33 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_2}(x) = \begin{cases} \frac{x - 0.13}{0.2}, & \text{if } 0.13 \leq x \leq 0.33 \\ \frac{0.58 - x}{0.25}, & \text{if } 0.33 \leq x \leq 0.58 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_3}(x) = \begin{cases} \frac{x - 0.33}{0.25}, & \text{if } 0.33 \leq x \leq 0.58 \\ \frac{0.83 - x}{0.25}, & \text{if } 0.58 \leq x \leq 0.83 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_4}(x) = \begin{cases} \frac{x - 0.58}{0.25}, & \text{if } 0.58 \leq x \leq 0.83 \\ \frac{1.08 - x}{0.25}, & \text{if } 0.83 \leq x \leq 1.08 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_5}(x) = \begin{cases} \frac{x - 0.83}{0.25}, & \text{if } 0.83 \leq x \leq 1.08 \\ \frac{1.58 - x}{0.5}, & \text{if } 1.08 \leq x \leq 1.58 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_6}(x) = \begin{cases} \frac{x - 1.08}{0.5}, & \text{if } 1.08 \leq x \leq 1.58 \\ \frac{2.08 - x}{0.5}, & \text{if } 1.58 \leq x \leq 2.08 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_7}(x) = \begin{cases} \frac{x - 1.58}{0.5}, & \text{if } 1.58 \leq x \leq 2.08 \\ 3.08 - x, & \text{if } 2.08 \leq x \leq 3.08 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{\tilde{x}_g}(x) = \begin{cases} \frac{x - 2.08}{1}, & \text{if } 2.08 \leq x \leq 3.08 \\ 1, & \text{if } x \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

Table 1: Averages values and mean squared errors of the GIBBS estimates of $\theta = 1$, coverage probabilities

n	AV	MSE	Coverage	Length
15	1.0307	0.0405	0.9403	0.7508
20	1.0289	0.0386	0.9457	0.7205
30	1.0289	0.0349	0.9524	0.6501
50	1.0246	0.0258	0.9526	0.4507
70	1.0154	0.0192	0.9538	0.3509
100	1.0102	0.0187	0.9559	0.1804

Table 2: Averages values and mean squared errors of the Bootstrapping Sampling estimates of $\theta = 1$, coverage probabilities

n	AV	MSE	Coverage	Length
15	1.0327	0.0426	0.9385	0.7703
20	1.0295	0.0375	0.9447	0.7409
30	1.0255	0.0329	0.9486	0.6305
50	1.0198	0.0247	0.9519	0.4902
70	1.0147	0.0198	0.9538	0.4003
100	1.0098	0.0115	0.9556	0.2106

Table 3: Averages values and mean squared errors of the MCMC estimates of $\theta = 1$, coverage probabilities

n	AV	MSE	Coverage	Length
15	1.0316	0.0437	0.9365	0.7559
20	1.0285	0.0389	0.9436	0.7253
30	1.0247	0.0337	0.9489	0.6154
50	1.0184	0.0256	0.9519	0.4708
70	1.0139	0.0209	0.9537	0.3804
100	1.0087	0.0127	0.9553	0.1905

Table 4: Averages values and mean squared errors of the MH estimates of $\theta = 1$, coverage probabilities

n	AV	MSE	Coverage	Length
15	1.0309	0.0425	0.9378	0.7653
20	1.0287	0.0384	0.9446	0.7357
30	1.0249	0.0339	0.9485	0.6254
50	1.0175	0.0245	0.9529	0.4956
70	1.0125	0.0195	0.9538	0.3907
100	1.0087	0.0128	0.9552	0.2104

Table 5: Combined of all methods using Bayesian model average technique

n	AV	MSE	Coverage	Length
15	1.0305	0.0409	0.9379	0.7559
20	1.0284	0.0379	0.9441	0.7264
30	1.0249	0.0339	0.9490	0.6174
50	1.0193	0.0251	0.9519	0.4761
70	1.0140	0.0204	0.9536	0.3877
100	1.0093	0.0132	0.9552	0.2009

Diagrammatic presentation of the table values obtained in the simulation process

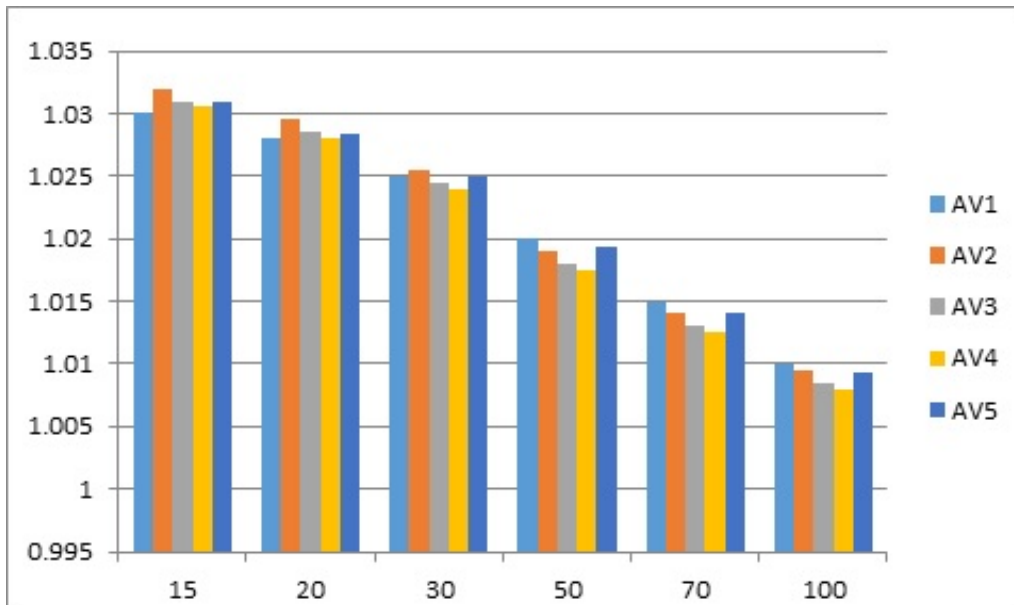


Figure 1: Comparison of Average Values Across Different Samples sizes

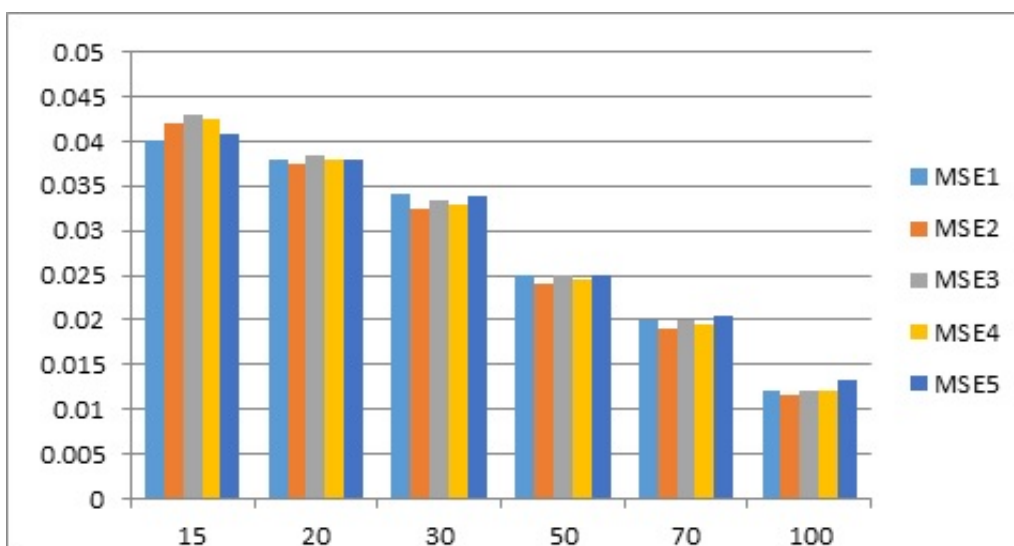


Figure 2: Comparison of MSE Across Different Samples sizes

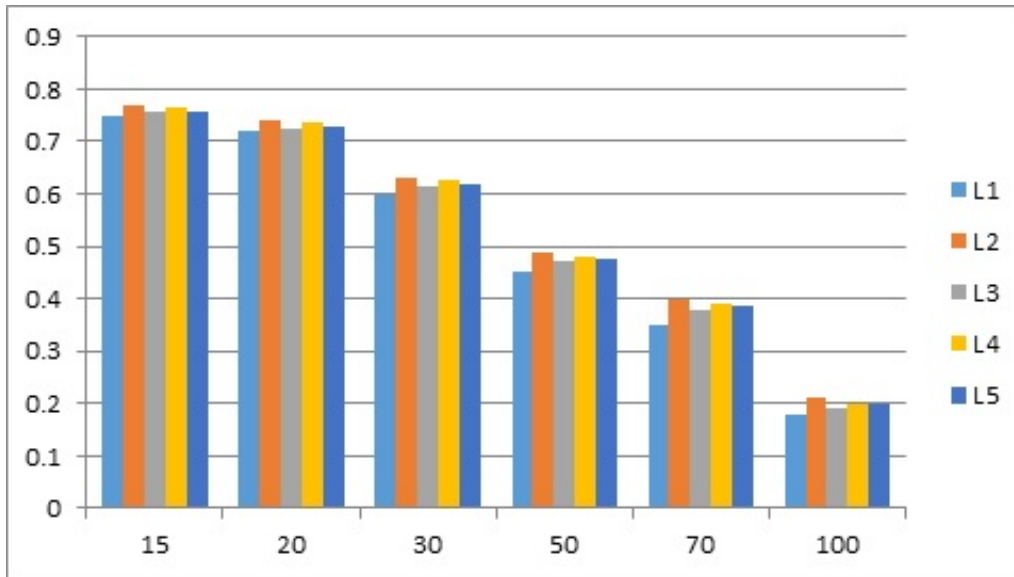


Figure 3: Comparison of Interval length Across Different Samples sizes

From Tables 1 to 5, it is observable that the hybrid approach, which combines all methods using the Bayesian model averaging technique, yields more reliable estimates. In Table 5, the average values (AV) are calculated using this Bayesian model averaging technique, combining estimates from all methods. Compared to individual approaches, this hybrid method provides estimates closer to the true value of 1, indicating improved accuracy. Table 5 shows that the modifications applied in the hybrid approach result in values that are even closer to the true value of 1 than those in Tables 1 through 4. Specifically, the Proposed table (Table 5) offers more precise average values, suggesting that this method achieves a superior level of accuracy in estimating the true value. Additionally, it reports lower mean squared error (MSE) values, highlighting improved accuracy in estimating the population parameters. Lower MSE values suggest confidence intervals that are closer to the true values. Furthermore, the coverage probabilities in Table 5 align closely with the desired confidence level, demonstrating the reliability of this approach. The graphical representations in Figures 1 to 4 further support these findings. Over Gibbs Sampling, Bootstrapping Sampling, MCMC, the Metropolis-Hastings (MH) Algorithm, and the unique hybrid strategy, the hybrid method stands out for its accuracy in parameter estimation for the Lindley distribution on fuzzy data. Figures 1 and 2 demonstrate that the hybrid approach yields average values closer to 1, and Figure 3 confirms lower MSE, indicating improved precision. Additionally, Figures 3 and 4 show that coverage probabilities and interval lengths are more favorable, underlining the robustness of the hybrid approach compared to individual methods. In conclusion, the unique hybrid strategy, presented in Table 5, offers balanced and efficient parameter estimation with superior accuracy across multiple metrics. The Bayesian model averaging technique combines estimates from all methods, producing values that consistently approach the true value of 1. This approach enhances mean squared errors, coverage probabilities, and confidence interval lengths across different sample sizes, outperforming Bootstrapping, MCMC, and MH estimates.

5.2. Discussion

The provided research summarizes the performance of Gibbs sampling, Bootstrapping Sampling, MCMC, and MH approaches for estimating parameters for the Lindley distribution on fuzzy data, as well as a unique hybrid strategy that combines these methods using Bayesian model averaging. The combination of these strategies presents a potential opportunity for improving estimate accuracy and dependability. Notably, the Bayesian model average methodology used

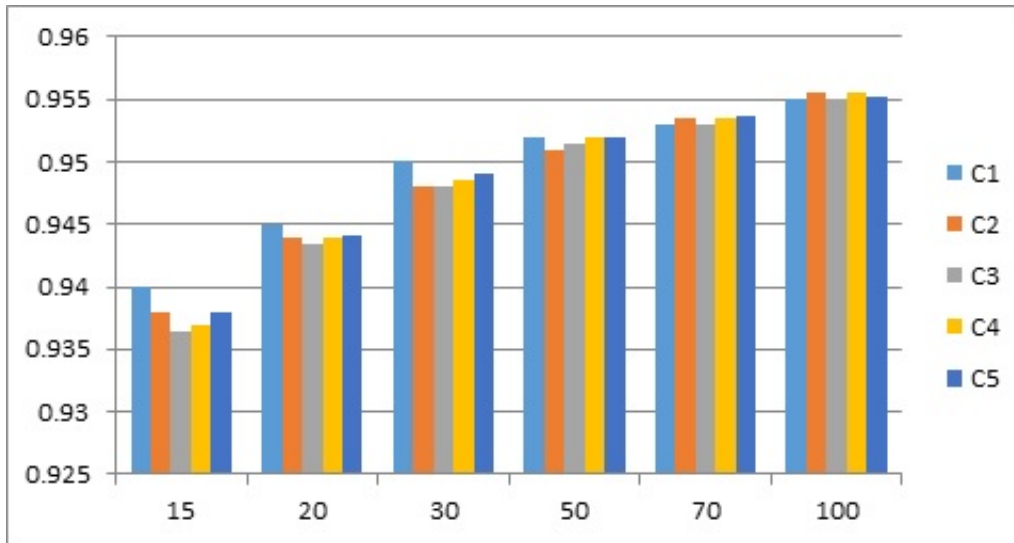


Figure 4: Comparison of Coverage Values Across Different Samples sizes

in the proposed hybrid method combines data from many estimating methodologies, resulting in estimates that are consistently closer to the real parameter value of 1 over a range of sample sizes. By exploiting the characteristics of each individual technique, the hybrid strategy yields reduced mean squared errors, indicating greater estimate accuracy. Furthermore, the hybrid method’s confidence intervals preserve coverage probabilities that are consistent with the targeted confidence level, demonstrating the method’s trustworthiness in statistical inference. Overall, the results from emphasize the hybrid method’s usefulness in improving estimate accuracy and reliability, providing a strong foundation for parameter estimation in the Lindley distribution framework with fuzzy data.

6. CONCLUSION

Ultimately, a novel method for parameter estimation in the Lindley distribution domain is embodied by the combination of Gibbs sampling, Bootstrapping sampling, MCMC, and MH techniques using Bayesian model averaging, especially when dealing with fuzzy data. Examining the hybrid approach closely across a range of sample sizes, it shows itself to be a remarkable performer, always guiding estimates toward the real parameter value of 1. This convergence toward the true value highlights the ability of the hybrid method to strengthen accuracy and avoid estimate mistakes, therefore promoting increased trust in the deduced conclusions. A notable finding is that the hybrid approach tends to provide less mean squared errors than its component individual techniques. This decrease in mean squared errors not only shows improvement in estimating accuracy but also highlights the ability of the hybrid method to extract a plethora of data from many estimation techniques, hence improving the quality of the estimated values. In addition, the confidence intervals produced by the hybrid approach firmly adhere to coverage probabilities that correspond with the target confidence level, therefore confirming the robustness and dependability of the approach in the field of statistical inference. These results have far-reaching consequences that cut across many fields where ambiguity and imprecision are commonplace, even beyond the boundaries of academic study. Researchers and professionals struggling with parameter estimate problems in the face of erratic fuzzy data will find the hybrid approach to be a true miracle. Through the use of complementary strengths of many estimating methods, the hybrid approach causes a paradigm change in statistical analysis, enabling more careful decision-making processes and a higher degree of confidence in the conclusions that are obtained. All things considered, the combination of Gibbs sampling, Bootstrapping sampling, MCMC, and MH techniques via Bayesian model averaging is a fundamental development in the

history of statistical inference techniques. Going ahead, more research and development of the hybrid approach show promise for overcoming the many obstacles related to parameter estimation in a variety of distributional frameworks and datasets, so paving the way for a revolutionary development in the field of statistical methods and their usefulness in many different research fields.

REFERENCES

- [1] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102-107.
- [2] Bhati, D., Sastry, D. V. S., Qadri, P. M. (2015). A new generalized Poisson-Lindley distribution: Applications and properties. *Austrian Journal of Statistics*, 44(4), 35-51.
- [3] Krishna, H., Kumar, K. (2011). Reliability estimation in Lindley distribution with progressively type II right censored sample. *Mathematics and Computers in Simulation*, 82(2), 281-294.
- [4] Kleinbaum, D. G., Klein, M., Kleinbaum, D. G., Klein, M. (2012). Competing risks survival analysis. *Survival Analysis: A self-learning text*, 425-495.
- [5] Alotaibi, R., Nassar, M., Elshahhat, A. (2023). Statistical analysis of inverse lindley data using adaptive type-II progressively hybrid censoring with applications. *Axioms*, 12(5), 427.
- [6] Hryniewicz, O. (2016). Bayes statistical decisions with random fuzzy data"an application in reliability. *Reliability Engineering System Safety*, 151, 20-33.
- [7] Colubi, A. (2009). Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy-and real-valued data. *Fuzzy sets and systems*, 160(3), 344-356.
- [8] Singh, P., Verma, M., Kumar, A. (2015). A novel method for ranking of vague sets for handling the risk analysis of compressor system. *Applied Soft Computing*, 26, 202-212.
- [9] Roohanizadeh, Z., Baloui Jamkhaneh, E., Deiri, E. (2022). Parameters and reliability estimation for the Weibull distribution based on intuitionistic fuzzy lifetime data. *Complex Intelligent Systems*, 8(6), 4881-4896.
- [10] Le, H., Sang, V. N. T., Lam Thuy, L. N., Bao, P. T. (2023). The fuzzy Kullback"Leibler divergence for estimating parameters of the probability distribution in fuzzy data: an application to classifying Vietnamese Herb Leaves. *Scientific Reports*, 13(1), 14537.
- [11] Van Hecke, T. (2018). Fuzzy parameter estimation of the Rayleigh distribution. *Journal of Statistics and Management Systems*, 21(7), 1391-1400.
- [12] Shah, S. H., Shafiq, M., Zaman, Q. (2022). Generalized Estimation for TwoParameter Life Time Distributions Based on Fuzzy Life Times. *Mathematical Problems in Engineering*, 2022(1), 6196251.
- [13] Vishwakarma, G. K., Paul, C., Singh, N. (2018). Parameters Estimation of Weibull Distribution Based on Fuzzy Data Using Neural Network. *Biostatistics and Biometrics Open Access Journal*, 6(5), 126-133.
- [14] Asadullah, M., Hossain, M. M., Molla, M. M. R., Rahaman, M. M. (2023). Comparison to the Proposed Hybrid Model and Machine Learning Techniques for Survival Prediction of Corona, Infected Patients. *Advances in Systems Science and Applications*, 23(4), 148-155.
- [15] Wang, S., Chen, N., Capodiferro, M. R., Zhang, T., Lancioni, H., Zhang, H., ... Lei, C. (2017). Whole mitogenomes reveal the history of swamp buffalo: initially shaped by glacial periods and eventually modelled by domestication. *Scientific Reports*, 7(1), 4708.
- [16] Khan, M. S. U. R., Hussain, Z., Ahmad, I. (2021). Effects of L-moments, maximum likelihood and maximum product of spacing estimation methods in using pearson type-3 distribution for modeling extreme values. *Water Resources Management*, 35, 1415-1431.
- [17] Mousavi, S., Esfahanipour, A., Fazel Zarandi, M. H. (2021). A modular Takagi-Sugeno-Kang (TSK) system based on a modified hybrid soft clustering for stock selection. *Scientia Iranica*, 28(4), 2342-2360.

- [18] Alizadeh, M., Afify, A. Z., Eliwa, M. S., Ali, S. (2020). The odd log-logistic Lindley-G family of distributions: properties, Bayesian and non-Bayesian estimation with applications. *Computational statistics*, 35(1), 281-308.
- [19] Rani, S., Kataria, A., Sharma, V., Ghosh, S., Karar, V., Lee, K., Choi, C. (2021). Threats and corrective measures for IoT security with observance of cybercrime: A survey. *Wireless communications and mobile computing*, 2021(1), 5579148.
- [20] Dey, S., Saha, M., Kumar, S. (2022). Parametric confidence intervals of Spmk for generalized exponential distribution. *American Journal of Mathematical and Management Sciences*, 41(3), 201-222.
- [21] Kumar, A., Bhattacharyya, S., Bouchard, K. (2024). Numerical characterization of support recovery in sparse regression with correlated design. *Communications in Statistics-Simulation and Computation*, 53(3), 1504-1518.
- [22] Samanta, D., Kundu, D. (2023). Bivariate Semi-Parametric Model: Bayesian Inference. *Methodology and Computing in Applied Probability*, 25(4), 87.
- [23] Zhang, G. H., Chen, W., Jiao, Y. Y., Wang, H., Wang, C. T. (2020). A failure probability evaluation method for collapse of drill-and-blast tunnels based on multistate fuzzy Bayesian network. *Engineering Geology*, 276, 105752.
- [24] A. E. Gelfand, Gibbs sampling, *Stat. 21st Century*, vol. 95, no. 452, pp. 341-349, 2001, doi: 10.2307/2669775.
- [25] Kass, R. E. (1997). Markov chain Monte Carlo in practice. *Journal of the American Statistical Association*, 92(440), 1645.
- [26] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.