KRISHNAV-P DISTRIBUTION AND ITS BIOMEDICAL APPLICATION

Praseeja C B[1], Prasanth C B[2*]

•

[1]Research Scholar, Department of Statistics, Annamalai University, Tamil Nadu, India
prasinikhil@gmail.com
[2]Assistant Professor, Department of Statistics, Sree Keralawarma College, Kerala
prasanthwarriercb@keralavarma.ac.in
[*]Corresponding author, +91 9447337331

Abstract

This paper introduces a novel class of distributions, termed the KrishNav-P distribution (KNPD), developed through the application of a weighted distribution technique to the baseline Fav-Jerry distribution. The statistical properties of the KNPD are rigorously analyzed, and parameter estimation is performed using the maximum likelihood estimation method. To evaluate the practical applicability and performance of the proposed distribution, it is applied to a real-world dataset consisting of weight loss measurements (in kilograms) from 77 secondary school students aged 12 to 15 in the Thrissur District, Kerala. The data, collected between January and June 2021, pertain to students who experienced weight loss following a COVID-19 infection and tested the goodness of fit and the superiority of the distribution over the baseline and other existing distributions are also demonstrated.

Keywords: Weighted distribution, Fav-Jerry distribution, Survival analysis, Entropy, Order statistics.

1. Introduction

Weighted distributions have emerged as a powerful statistical tool, enhancing classical distributions by introducing an additional parameter, commonly referred to as a shape parameter. This parameter allows baseline distributions to become more flexible and reliable when compared to other distributions. Weighted distributions arise when observations are recorded according to a specific stochastic model. If each observation does not have an equal likelihood of being recorded and is instead influenced by a weight function, the resulting distribution deviates from the original.

Since their inception, weighted distributions have been fundamental in probability and statistics, representing a significant advancement in distribution theory. The concept was first proposed by Fisher [6] to address ascertainment bias, and later generalized by Rao [13] to model statistical data where standard distributions were deemed inadequate. These distributions have proven highly effective in modelling statistical data, offering robust methods for model specification, data interpretation, and fitting models to unknown weight functions using samples from both original and weighted distributions.

Weighted distributions have seen significant contributions from numerous researchers, showcasing their versatility in addressing complex statistical problems. For instance, Ajami and

Jahanshahi [1] explored parameter estimation in the weighted Rayleigh distribution, while Bashir *et al.* [2] proposed the Pareto length-biased exponential distribution and demonstrated its application to flood data. Bashir and Rasul [3] introduced the size-biased Janardan distribution, detailing its properties, and Benchiha *et al.* [4] examined the weighted generalized quasi Lindley distribution, including its estimation and applications. Ganaie and Rajagopalan [7] studied the weighted quasi-gamma distribution and its applications. Murthy *et.al.*, [8] explained the weighted case of the Weibull Models whereas Nanuwong and Bodhisuwan [9] constructed the length-biased Beta-Pareto distribution. Praseeja *et al.* [10] derived the Chandbhas-*P* distribution and demonstrated its applications in medical science, later extending their work [11] to explain the characteristics of the KrishSupra-*P* Distribution. and Its Biomedical Application. Saghir *et al.* [14] developed the weighted exponentiated inverted Weibull distribution, while Rajagopalan *et al.* [12] introduced the weighted three-parameter Pranav distribution. Salama *et al.* [15] contributed by proposing the length-biased weighted exponentiated inverted exponential distribution, offering an in-depth analysis of its properties and estimation methods. Collectively, these studies highlight the extensive utility of weighted distributions across various fields and their potential in solving intricate statistical challenges.

The Fav-Jerry distribution, a one-parameter distribution recently introduced by Ekemezie and Obulezi [5], represents a novel addition to the field of probability and statistics. Several key properties of this distribution have been explored, including its quantile function, moments, distributional shape, mean, variance, skewness, kurtosis, and Shannon entropy. Parameter estimation has been performed using maximum likelihood estimation.

2. Methods

KrishNav-*P* Distribution (KNPD)

The probability density function (PDF) of Fav-Jerry distribution is given by

$$f(x;\theta) = \frac{\theta}{(\theta^2+2)}(2+\theta^3 x)e^{-\theta x}; x>0, \theta>0 \quad (1)$$

and cumulative distribution function (CDF) of Fav-Jerry distribution is given by

$$F(x;\theta) = 1 - \left\{1 + \frac{\theta^3 x}{(\theta^2+2)}\right\} e^{-\theta x}; x>0, \theta>0. \quad (2)$$

Assume the random variable X constitutes non-negative condition with PDF $f(x)$. Consider $w(x)$ be its weight function which is non-negative, then probability density function of weighted random variable X_w is given by

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

Here $w(x)$ be the weight function, then $E(w(x)) = \int w(x)f(x)dx < \infty$.

In this paper, we have to obtain the weighted version of Fav-Jerry distribution precisely by taking the weight function as $w(x) = x^c$ and therefore its PDF is defined as

$$f_w(x;\theta,c) = \frac{x^c f(x;\theta)}{E(x^c)}. \quad (3)$$

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

$$E(x^c) = \int_0^\infty x^c \frac{\theta}{(\theta^2+2)}(2+\theta^3 x)e^{-\theta x}dx.$$

After simplification of above equation, we get

$$E(x^c) = \frac{(2\Gamma(c+1) + \theta^2\Gamma(c+2))}{\theta^c(\theta^2+2)}. \tag{4}$$

Therefore, by applying equation (1) and (4) in equation (3), we have acquired the PDF of KNPD which is

$$f_w(x;\theta,c) = \frac{(x^c\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3x)e^{-\theta x} \tag{5}$$

and CDF of KNPD will be obtained by using the expression

$$F_w(x;\theta,c) = \int_0^x f_w(x;\theta,c)dx$$

$$F_w(x;\theta,c) = \int_0^x \frac{x^c\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3x)e^{-\theta x}dx$$

$$F_w(x;\theta,c) = \frac{1}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\int_0^x x^c\theta^{c+1}(2+\theta^3x)e^{-\theta x}dx$$

$$F_w(x;\theta,c) = \frac{1}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\left(2\theta^{c+1}\int_0^x x^c e^{-\theta x}dx + \theta^{c+4}\int_0^x x^{c+1}e^{-\theta x}dx\right) \tag{6}$$

After simplification of equation (6), we have obtained the cumulative distribution function of KNPD which is

$$F_w(x;\theta,c) = \frac{1}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2\gamma(c+1,\theta x)+\theta^2\gamma(c+2,\theta x)) \tag{7}$$

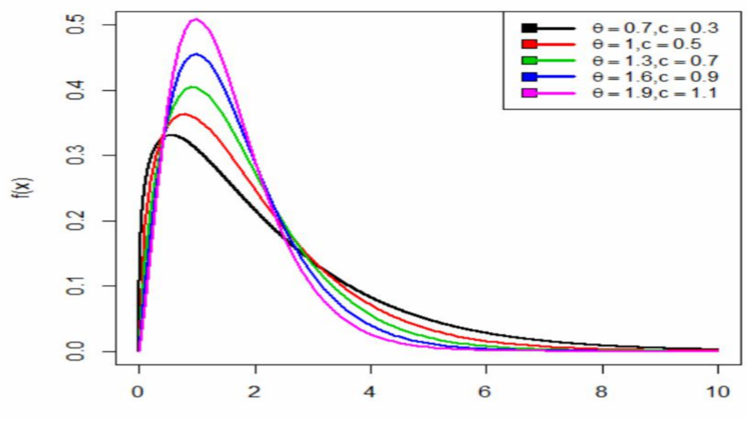

Figure 1: *PDF of KNPD*

The non-symmetric nature of the PDF is clear from the figure. The distribution is positively skewed.

3. Results

Survival Analysis
This section presents specific measures of the KNPD, including the survival function, hazard function, reverse hazard function, and Mills ratio.

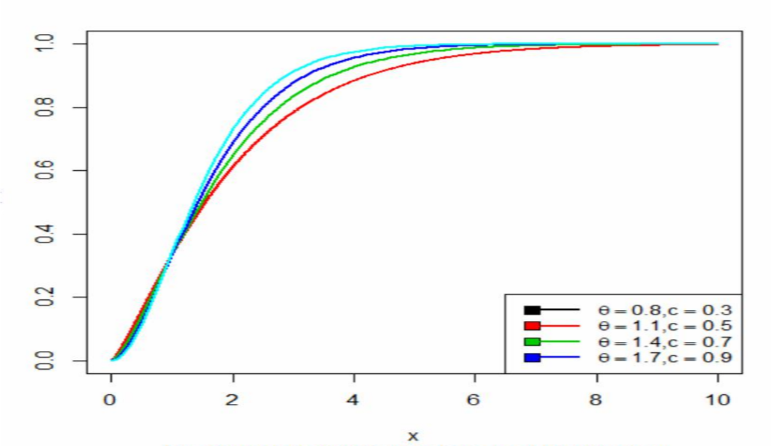

Figure 2: *CDF of KNPD*

Survival function

The survival function, which represents the probability that the random variable exceeds a certain value x, is defined as:

$$S(x) = 1 - F(x).$$

The survival function of KNPD is given by

$$S(x) = 1 - F_w(x;\theta,c)$$

$$S(x) = 1 - \frac{1}{(2\Gamma(c+1) + \theta^2\Gamma(c+2))}(2\gamma(c+1,\theta x) + \theta^2\gamma(c+2,\theta x)).$$

Hazard function

The hazard function, describing the instantaneous rate of failure at time x given survival up to that point, is given by:

$$h(x) = \frac{f(x)}{S(x)}.$$

The hazard function of explored distribution is given by

$$h(x) = \frac{f_w(x;\theta,c)}{1 - F_w(x;\theta,c)}$$

$$h(x) = \frac{x^c\theta^{c+1}(2+\theta^3 x)e^{-\theta x}}{(2\Gamma(c+1) + \theta^2\Gamma(c+2)) - (2\gamma(c+1,\theta x) + \theta^2\gamma(c+2,\theta x))}.$$

Reverse hazard function

The reverse hazard function, representing the rate at which the random variable falls below xxx, is expressed as:

$$r(x) = \frac{f(x)}{F(x)}.$$

The reverse hazard function of proposed distribution is given by

$$h_r(x) = \frac{f_w(x;\theta,c)}{F_w(x;\theta,c)}$$

$$h_r(x) = \frac{x^c\theta^{c+1}(2+\theta^3 x)e^{-\theta x}}{(2\gamma(c+1,\theta x) + \theta^2\gamma(c+2,\theta x))}.$$

Mills Ratio

The Mills ratio, an important measure for analyzing tail behaviour, is defined as:

$$M(x) = \frac{S(x)}{f(x)}$$

which is the ratio of the survival function to the probability density function.

$$M.R = \frac{1}{h_r(x)} = \frac{(2\gamma(c+1,\theta x) + \theta^2\gamma(c+2,\theta x))}{x^c\theta^(c+1)(2+\theta^3 x)e^{-\theta x}}$$

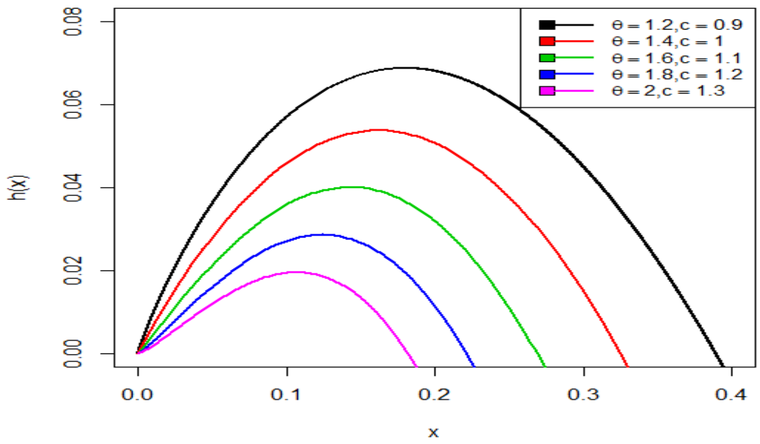

Figure 3: *Reliability of KNPD*

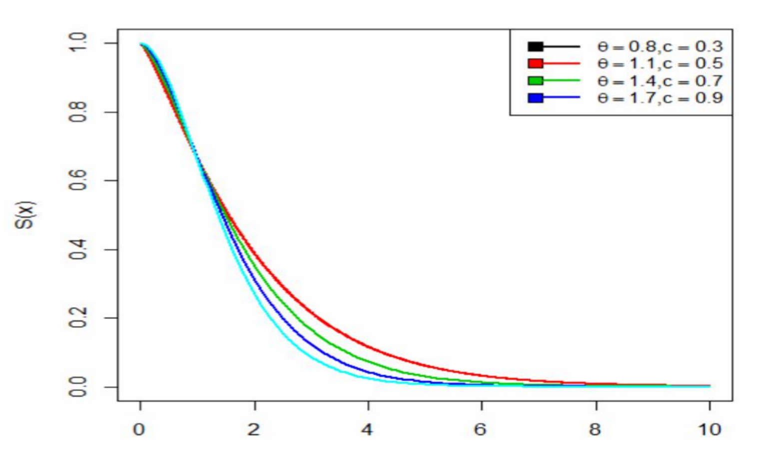

Figure 4: *Hazard function of KNPD*

The nature of the Reliability Function and Hazard Function of KNPD is clear from the figure.

Order Statistics

Order statistics is a useful concept in statistical sciences and it deals with independent and identically distributed random variables used in estimation theory, statistical inference and hypothesis testing in various fields. Consider the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ of a random sample $X_1, X_2 \ldots, X_n$ from a continuous population with probability density function $f_x(x)$ and cumulative distribution function $F_X(x)$, then pdf of r^{th} order statistics $X_{(r)}$ is given by

$$f_{x(r)}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x)(F_X(x))^{r-1}(1-F_X(x))^{n-r}. \quad (8)$$

After applying equation (5) and (7) in equation (8), we will acquire the probability density function of r^{th} order statistics $X_{(r)}$ of KNPD which is given by

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

Now therefore the PDF of higher order statistic $X_{(n)}$ of KNPD is given by

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

and probability density function of first order statistic $X_{(1)}$ of KNPD is given by

$$f_{x(1)}(x) = \frac{nx^c\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3 x)e^{-\theta x}$$
$$\times\left(1-\frac{1}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2\gamma(c+1,\theta x)+\theta^2\gamma(c+2,\theta x))\right)^{n-1}.$$

Entropy

Entropy is a measure of randomness or disorder within a system. In information theory, it quantifies the average uncertainty or the amount of information associated with the potential states or possible outcomes of a variable. It reflects the level of disorganization or unpredictability in the system.

In biostatistics, entropy is used as a measure of uncertainty or disorder in biological systems, often applied to analyze complex data like genomics, proteomics, and fMRI signals. It helps quantify the diversity or complexity of biological processes at various levels, from molecular to organismal. Specific entropy measures like approximate entropy (ApEn) and sample entropy (SampEn) are used to analyze time series data and assess the regularity or predictability of biological signals.

Applications in biostatistics:

i) Genomics and proteomics: Entropy can be used to analyze the diversity of genes, proteins, or metabolites in a biological sample. For instance, comparing the entropy of normal and cancer cell transcriptomes can reveal differences in gene expression complexity.

ii) fMRI data analysis: Entropy, particularly sample entropy, is used to assess the complexity of brain signals from fMRI data. This can help understand brain-based disorders and assess the predictability of brain activity patterns.

iii) Time series analysis: ApEn and SampEn are commonly used to analyze time series data, like physiological signals or movement patterns, to quantify the irregularity or complexity of these signals.

The maximum entropy principle can be used as a basis for statistical inference, providing a way to estimate probabilities or parameters that maximize entropy subject to certain constraints.

Renyi Entropy

Rényi entropy is a fundamental concept in information theory that generalizes various forms of entropy. It serves as the basis for the concept of generalized dimensions, providing a broader perspective on the distribution of probabilities. The Rényi entropy of order α (where $\alpha > 0$ and $\alpha \neq 1$) is defined as:

$$T_R(\alpha) = \frac{1}{1-\alpha}\log\left(\int f_w^{\alpha}(x;\theta,c)dx\right)$$

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

$$T_R(\alpha) = \frac{1}{1-\alpha}\log\left(\left(\frac{\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\right)^{\alpha}\int_0^{\infty}x^{\alpha c}(2+\theta^3x)^{\alpha}e^{-\theta\alpha x}dx\right). \tag{9}$$

After the simplification of equation (9), we get

$$T_R(\alpha) = \frac{1}{1-\alpha}\log\left(\left(\frac{\theta^{\alpha-j-1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))^{\alpha}}\right)\sum_{j=0}^{\infty}\binom{\alpha}{j}2^{\alpha-j}\theta^{3j}\frac{\Gamma(\alpha c+j+1)}{\alpha^{\alpha c+j+1}}\right).$$

The Tsallis entropy of proposed distribution is obtained by using expression as given

$$T_s(\gamma) = \frac{1}{\gamma-1}\left(1-\int_0^{\infty}f_w^{\gamma}(x;\theta,c)dx\right)$$

$$T_s(\gamma) = \frac{1}{\gamma-1}\left(1-\int_0^{\infty}\left(\frac{x^c\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3x)e^{-\theta x}\right)^{\gamma}dx\right)$$

$$T_s(\gamma) = \frac{1}{\gamma-1}\left(1-\left(\frac{\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\right)^{\gamma}\int_0^{\infty}x^{\gamma c}(2+\theta^3x)^{\gamma}e^{-\theta\gamma x}dx\right). \tag{10}$$

After the simplification of equation (10), we obtain

$$T_s(\gamma) = \frac{1}{\gamma-1}\left(1-\left(\frac{\theta^{\gamma-k-1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))^{\gamma}}\right)\sum_{k=0}^{\infty}\binom{\gamma}{k}2^{\gamma-k}\theta^{3k}\frac{\Gamma(\gamma c+k+1)}{\gamma^{\gamma c+k+1}}\right).$$

Statistical Properties

In this section, we describe and derive some key features of the KNPD, including its moments, harmonic mean, and moment-generating function.

Moments

Suppose the random variable X constitutes KNPD with parameter θ and c, then the r^{th} order moment of proposed distribution is given by

$$\mu_r' = E(X^r) = \int_0^{\infty}x^r f_w(x;\theta,c)dx$$

$$\mu_r' = \int_0^{\infty}x^r\frac{x^c\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3x)e^{-\theta x}dx$$

$$\mu_r' = \int_0^{\infty}\frac{x^{c+r}\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3x)e^{-\theta x}dx$$

$$\mu_r' = \frac{\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\int_0^{\infty}x^{c+r}(2+\theta^3x)e^{-\theta x}dx$$

$$\mu_r' = \frac{\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\left(2\int_0^{\infty}x^{(c+r+1)-1}e^{-\theta x}dx+\theta^3\int_0^{\infty}x^{(c+r+2)-1}e^{-\theta x}dx\right). \tag{11}$$

After the simplification of equation (11), we obtain

$$\mu_r' = E(X^r) = \frac{(2\Gamma(c+r+1) + \theta^2\Gamma(c+r+2))}{\theta^r(2\Gamma(c+1) + \theta^2\Gamma(c+2))}. \tag{12}$$

Therefore, by putting $r = 1$ in equation (12), we will obtain the mean of KNPD which is

$$\mu_1' = \frac{(2\Gamma(c+2) + \theta^2\Gamma(c+3))}{\theta(2\Gamma(c+1) + \theta^2\Gamma(c+2))}.$$

Now putting $r = 2, 3$ and 4 in equation (12), we will obtain other three moments of KNPD as

$$\mu_2' = \frac{(2\Gamma(c+3) + \theta^2\Gamma(c+4))}{\theta^2(2\Gamma(c+1) + \theta^2\Gamma(c+2))}$$

$$\mu_3' = \frac{(2\Gamma(c+4) + \theta^2\Gamma(c+5))}{\theta^3(2\Gamma(c+1) + \theta^2\Gamma(c+2))}$$

$$\mu_4' = \frac{(2\Gamma(c+5) + \theta^2\Gamma(c+6))}{\theta^4(2\Gamma(c+1) + \theta^2\Gamma(c+2))}$$

$$\text{Variance} = \frac{(2\Gamma(c+3) + \theta^2\Gamma(c+4))}{\theta^2(2\Gamma(c+1) + \theta^2\Gamma(c+2))} - \left(\frac{(2\Gamma(c+2) + \theta^2\Gamma(c+3))}{\theta(2\Gamma(c+1) + \theta^2\Gamma(c+2))}\right)^2$$

$$S.D(\sigma) = \sqrt{\left(\frac{(2\Gamma(c+3) + \theta^2\Gamma(c+4))}{\theta^2(2\Gamma(c+1) + \theta^2\Gamma(c+2))} - \left(\frac{(2\Gamma(c+2) + \theta^2\Gamma(c+3))}{\theta(2\Gamma(c+1) + \theta^2\Gamma(c+2))}\right)^2\right)}.$$

Harmonic mean

The harmonic mean of explored new distribution is obtained by using the expression as

$$H.M = E\left(\frac{1}{x}\right) = \int_0^\infty \frac{1}{x} f_w(x;\theta,c)dx$$

$$H.M = \int_0^\infty \frac{1}{x} \frac{x^c\theta^{c+1}}{(2\Gamma(c+1) + \theta^2\Gamma(c+2))}(2+\theta^3 x)e^{-\theta x}dx$$

$$H.M = \int_0^\infty \frac{x^{c-1}\theta^{c+1}}{(2\Gamma(c+1) + \theta^2\Gamma(c+2))}(2+\theta^3 x)e^{-\theta x}dx$$

$$H.M = \frac{\theta^{c+1}}{(2\Gamma(c+1) + \theta^2\Gamma(c+2))}\int_0^\infty x^{c-1}(2+\theta^3 x)e^{-\theta x}dx. \tag{13}$$

After simplification of equation (13), we obtain

$$H.M = \frac{\theta(2\Gamma(c) + \theta^2\Gamma(c+1))}{(2\Gamma(c+1) + \theta^2\Gamma(c+2))}.$$

The harmonic mean is a type of average particularly useful in biomedical fields when dealing with rates, ratios, or quantities where smaller values should have more influence. It's calculated by finding the reciprocal of the arithmetic mean of reciprocals of the data. In biomedical contexts, it's used to calculate averages like half-lives, or when dealing with rates of events or processes where the smaller values represent more frequent occurrences.

Uses in Biomedical Areas:

1. Half-life calculations: The harmonic mean can be used to find the average half-life of a substance in the body, as detailed in a study by the National Institutes of Health (NIH).
2. Averaging rates: When calculating the average rate of a biological process, such as the rate of a reaction or the rate of cell growth, the harmonic mean can be useful.

3. Inter-beat intervals: The harmonic mean can be applied to heart inter-beat intervals (time between heartbeats) when transformed to heart rate.

Moment generating function and characteristic function

Suppose X be the random variable constitutes weighted Fav-Jerry distribution, then moment generating function of explored distribution is obtained by applying the expression

$$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} f_w(x;\theta,c)dx.$$

By applying the Taylor's series, we obtain

$$M_X(t) = \int_0^\infty \left(1 + tx + \frac{(tx)^2}{2!} + \dots\right) f_w(x;\theta,c)dx = \int_0^\infty \sum_{k=0}^{\infty} \frac{t^k}{k!} x^k f_w(x;\theta,c)dx$$

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mu'_k = \sum_{k=0}^{\infty} \frac{t^k}{k!}\left(\frac{(2\Gamma(c+k+1)+\theta^2\Gamma(c+k+2))}{\theta^k(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\right)$$

$$M_X(t) = \frac{1}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\sum_{k=0}^{\infty} \frac{t^k}{k!\theta^k}(2\Gamma(c+k+1)+\theta^2\Gamma(c+k+2)).$$

The characteristic function of KNPD is,

$$M_X(it) = \frac{1}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\sum_{k=0}^{\infty} \frac{it^k}{k!\theta^k}(2\Gamma(c+k+1)+\theta^2\Gamma(c+k+2)).$$

Maximum Likelihood Estimation and Fisher's Information Matrix

In this section, we estimate the parameters of KNPD by using the maximum likelihood estimation. Assume the random sample $X_1, X_2 \dots, X_n$ from weighted Fav-Jerry distribution, then likelihood function should be defined as

$$L(x) = \prod_{i=1}^{n} f_w(x;\theta,c)$$

$$L(x) = \prod_{i=1}^{n}\left(\frac{x_i^c\theta^{c+1}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}(2+\theta^3 x_i)e^{-\theta x_i}\right)$$

$$L(x) = \frac{\theta^{n(c+1)}}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))^n}\prod_{i=1}^{n}\left(x_i^c(2+\theta^3 x_i)e^{-\theta x_i}\right).$$

The log likelihood function is given by

$$\log L = n(c+1)\log\theta - n\log(2\Gamma(c+1)+\theta^2\Gamma(c+2)) + c\sum_{i=1}^{n}\log x_i + \sum_{i=1}^{n}\log(2+\theta^3 x_i) - \theta\sum_{i=1}^{n} x_i. \quad (14)$$

Therefore, now differentiating the log likelihood equation (14) with respect to θ and c. The following normal equations must be satisfied

$$\frac{\partial \log L}{\partial\theta} = \frac{n(c+1)}{\theta} - n\left(\frac{(2\theta\Gamma(c+2))}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\right) + \sum_{(i=1)}^{n}\left(\frac{3\theta^2 x_i}{(2+\theta^3 x_i)}\right) - \sum_{(i=1)}^{n} x_i = 0$$

$$\frac{\partial \log L}{\partial c} = n\log\theta - n\psi(2\Gamma(c+1)+\theta^2\Gamma(c+2)) + \sum_{(i=1)}^{n}\log x_i = 0$$

Here $\psi(.)$ is the digamma function.

The likelihood equations given above are too complicated to solve through algebraically. Apply R program for determining the parameters of developed distribution. Also apply the asymptotic normality results for getting the confidence interval. We state that were $\hat{\beta} = (\hat{theta}, \hat{c})$ which represents the maximum likelihood estimate of $\beta = (\theta, c)$. The result can be interpreted as $\sqrt{n}(\hat{\beta} - \beta) \to N_2(0, I^{-1}(\beta))$ here $I^{-1}(\beta)$ is Fisher Transformation Matrix

$$I(\beta) = -\frac{1}{n}\begin{pmatrix} E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) & E\left(\frac{\partial^2 \log L}{\partial \theta \partial c}\right) \\ E\left(\frac{\partial^2 \log L}{\partial c \partial \theta}\right) & E\left(\frac{\partial^2 \log L}{\partial c^2}\right) \end{pmatrix}$$

Here, we can state that

$$E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) = -\frac{n(c+1)}{\theta^2} - n\left(\frac{(2\Gamma(c+1)+\theta^2\Gamma(c+2))(2\Gamma(c+2)) - (2\theta\Gamma(c+2))^2}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))^2}\right) + \sum_{i=1}^{n}\left(\frac{(2+\theta^3 x_i)(6\theta x_i) - (3\theta^2 x_i)^2}{(2+\theta^3 x_i)^2}\right)$$

$$E\left(\frac{\partial^2 \log L}{\partial c^2}\right) = -n\psi'(2\Gamma(c+1)+\theta^2\Gamma(c+2))$$

$$E\left(\frac{\partial^2 \log L}{\partial \theta \partial c}\right) = \frac{n}{\theta} - n\psi\left(\frac{2\theta\Gamma(c+2)}{(2\Gamma(c+1)+\theta^2\Gamma(c+2))}\right).$$

Hence β is not known so $I^{-1}(\beta)$ should be estimated by $I^{-1}(\hat{\beta})$ and it should be applied to attain asymptotic confidence interval for θ and c.

Fisher's Information Matrix (FIM) is a statistical tool that quantifies the amount of information a dataset provides about a parameter. In biomedical applications, FIM helps optimize experimental designs, analyze data from imaging techniques like single molecule microscopy, and improve parameter estimation in models. It's essentially a measure of how well a parameter can be estimated from data.

4. Discussion

Simulation Analysis

Simulation analysis uses statistical methods to study processes and systems by creating artificial models that mimic real-world scenarios. It involves repeated random sampling to approximate complex systems, allowing for the analysis of statistical estimates and model validation.

Key aspects of simulation analysis in statistics include:

Model Creation: Building a model that represents the system or process being analyzed.

Data Generation: Using random sampling to generate data that mimics the real-world conditions.

Statistical Analysis: Analyzing the simulated data to understand the system's behaviour, identify potential outcomes, and validate the model.

Output Analysis: Summarizing and interpreting the simulation results to gain insights and inform decision-making.

Model Validation: Ensuring the simulation model accurately reflects the real-world system it's intended to represent.

Simulation analysis is a powerful statistical tool that enables the study of complex systems, the evaluation of statistical methods, and the support of informed decision-making by creating artificial models and analyzing their simulated outputs.

By considering 5000 simulated random values from the KNPD, we can observe the non-symmetric nature of the distribution.

Table 1: *The Descriptive Statistics of the Simulated data.*

Mean	2.700	Range	5.750
Standard Error	0.105	Minimum	0.090
Median	1.790	Maximum	5.660
Mode	1.030	Sum	13500
Standard Deviation	1.405	Count	5000
Sample Variance	1.972	Largest (1)	5.750
Kurtosis	0.73	Smallest (1)	0.090
Skewness	1.1886	Confidence Level (95.0%)	0.2070

Karl Pearson's Skewness = (Mean-Mode)/SD. The value of this coefficient will be zero for a symmetrical distribution. If mean > mode, the coefficient of skewness is positive else negative. For considerably skewed destitution this coefficient is lies between -1and 1. If it is greater than 1 its highly positive skewed. Generally, if the distribution of data is skewed to the left, then mean < median< mode. If the distribution of data is skewed to the right, then mode < median< mean. Here from Figure 1 & Table 1, both cases imply the distribution is highly positive skewed.

Application

In this section, we have fitted and analyzed the goodness of fit of KNPD by examining and applying a real lifetime data set and observed that KNPD gives a better fit while compared over Fav-Jerry, KrishSupra-*P* Distribution (KSPD), Exponential and Lindley distributions.

The following data, Table 2, (Praseeja *et al.* [10]) represents the weight loss (Kilograms) of 77 randomly selected patients (of age group 12 to 15 - Secondary School Students from Thrissur District of Kerala" (January to June (2021)) after being infected with COVID-19.

Table 2: *weight loss (Kilograms) - patients (of age group 12 to 15) - from Thrissur District.*

1.39	1.44	1.46	1.53	1.59	1.6	1.63	1.68	1.71	1.72
4.02	4.32	4.58	5.55	2.54	0.77	2.93	3.27	3.42	3.47
1.83	1.95	1.96	1.97	2.02	2.13	2.15	2.16	2.22	2.30
1.00	0.99	1.02	1.05	1.07	1.07	1.08	1.08	1.08	1.09
0.10	0.33	0.44	0.56	0.59	0.59	0.72	0.74	0.92	0.93
1.13	1.15	1.16	1.2	1.21	1.22	1.22	1.24	1.30	1.34
2.4	2.45	2.51	2.53	2.54	2.78	1.83	1.95	1.96	1.97
1.76	3.61	2.31	1.12	0.96	1.36	2.02			

Table 3: *Descriptive Statistics of the Data.*

Mean	2.2400	Range	5.4500
Standard Error	0.1047	Minimum	0.1000
Median	1.9600	Maximum	5.5500
Mode	1.0800	Sum	403.4800
Standard Deviation	1.4041	Count	180.0000
Sample Variance	1.9716	Largest (1)	5.5500
Kurtosis	0.51	Smallest (1)	0.1000
Skewness	1.1149	Confidence Level (95.0%)	0.2065

Karl Pearson's Skewness = (Mean-Mode)/SD = 1. 1149.The distribution of data is skewed to the right, then mode < median< mean. Here from Table 3, both cases imply the distribution is highly positive skewed.

In order to compute the comparison criterion values besides with unknown parameters are estimated thoroughly by applying R software. To evaluate and highlight the performance of KNPD in comparison over Fav-Jerry, KSPD, Exponential and Lindley distributions, the considered criterions such as AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), AICC (Akaike Information Criterion Corrected) and $-2\log L$ have been applied. The distribution performs quite better if it has the smaller criterion values of $AIC, BIC, AICC$ and $-2\log L$ as compared over other distributions. The formulas for determining the criterions are,

$$AIC = 2k - 2\log L, \quad BIC = k\log n - 2\log L \quad \text{and} \quad AICC = AIC + \frac{2k(k+1)}{n-k-1}.$$

Here $-2\log L$ is the maximized value of log-likelihood function$(L), n$ is the sample size and k is the number of parameters in statistical model.

It is clearly observed from results in Table 3, that KNPD has smaller criterions of $AIC, BIC, AICC$ and $-2\log L$ in comparison with the Fav-Jerry, Exponential, and Lindley distributions demonstrates that the KNPD offers a superior fit. Additionally, its performance is competitive with the KSPD, further underscoring its effectiveness in modelling the given data.

Table 4: *Shows Analysis and Comparison of Performed Distributions*

Distributions	MLE	S.E	$-2\log L$	AIC	BIC	AICC
KPND	$\hat{\theta} = 1.1054$	$\hat{\theta} = 0.1803$				
	$\hat{c} = 0.7619$	$\hat{c} = 0.2324$	203.939	207.939	212.225	208.139
Fav-Jerry	$\hat{\theta} = 0.5646$	$\hat{\theta} = 0.0955$	217.879	219.879	222.022	219.945
KSPD	$\hat{\theta} = 2.6090$	$\hat{\theta} = 0.3019$				
	$\hat{c} = 1.8501$	$\hat{c} = 0.50912$	191.4293	195.3943	199.9900	195.5960
Exponential	$\hat{\theta} = 0.5697$	$\hat{\theta} = 0.0596$	224.8929	226.8929	229.1696	226.9500
Lindley	$\hat{\theta} = 0.8697$	$\hat{\theta} = 0.0697$	213.051	215.051	217.3277	215.1081

Maximum Likelihood Estimate (MLE), Standard Error (S.E)

5. Conclusion

This paper introduces a novel study on the Fav-Jerry distribution, referred to as KNPD, developed through the application of a weighted technique to its classical counterpart, the Fav-Jerry distribution. The proposed distribution has been extensively analyzed, covering various statistical properties such as moments, the behaviour of its probability density function (PDF) and cumulative distribution function (CDF), mean and variance, survival function, hazard function, reverse hazard rate function, moment-generating function, and harmonic mean. Additionally, its order statistics and Renyi entropy measure have been discussed. The parameters of the new distribution are estimated using the maximum likelihood estimation method. Moreover, the superiority and versatility of the proposed distribution have been demonstrated through a real dataset, revealing that KNPD provides a better fit compared to the Fav-Jerry, Exponential, and Lindley distributions.

References

[1] Ajami, M. and Jahanshahi, S. M. A. (2017). Parameter estimation in weighted Rayleigh distribution. *Journal of Modern Applied Statistical Methods*, 16(2), 256–276.

[2] Bashir, S., Masood, B., Rasul, M. and Waseem, N. (2023). Pareto length-biased exponential distribution: Properties and Applications on the flood data. *International Journal of Contemporary Issues in Social Sciences*, 2(3), 838–850.

[3] Bashir, S. and Rasul, M. (2016). Some properties of the size-biased Janardan distribution. *American Journal of Theoretical and Applied Statistics*, 5(5), 305–310.

[4] Benchiha, S. A., Al-Omari, A. I., Alotaibi, N. and Shrahili, M. (2021). Weighted generalized quasi Lindley distribution: Different methods of estimation, applications for covid-19 and engineering data. *AIMS Mathematics*, 6(11), 11850–11878.

[5] Ekemezie, D. F. N. and Obulezi, O. J. (2024). The Fav-Jerry distribution: Another member in the Lindley class with applications. *Earthline Journal of Mathematical Sciences*, 14(4), 793–816.

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

[7] Ganaie, R. A. and Rajagopalan, V. (2020). A new generalization of quasi gamma distribution with properties and applications. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(3), 3855–3861.

[8] Murthy, D. P., Xie, M. and Jiang, R. (2004). Weibull Models, volume 505. John Wiley & Sons.

[9] Nanuwong, N. and Bodhisuwan, W. (2014). Length biased beta-Pareto distribution and its structural properties with application. *Journal of Mathematics and Statistics*, 10(1), 49–57.

[10] Praseeja C. B., Prasanth C. B., Subramanian C. and Unnikrishnan T. (2023). Characteristics of Chandbhas-*P* Distribution and Its Applications in Medical Science. *RT&A*, 4 (76), 314–324, https://gnedenko.net/Journal/2023/042023/RTA_4_2023-25.pdf

[11] Praseeja, C. B., Prasanth, C. B., Subramanian, C., Unnikrishnan, T. (2024). Characterisation of KrishSupra-*P* Distribution and Its Biomedical Application. *Communications in Mathematics and Applications*, 15(3), 1077–1088. https://doi.org/10.26713/cma.v15i3.2833

[12] Rajagopalan, V., Ganaie, R. A. and Nair, S. R. (2023). A new extension of two parameter Pranav distribution with applications in Industrial and Medical sciences. *Journal of Statistics Applications & Probability*, 12(1), 231–246.

[13] Rao, C. R. (1965). On discrete distributions arising out of method of ascertainment, in classical and contagious discrete distributions. Pergamon press and statistical publishing society, Calcutta. 320–332.

[14] Saghir, A., Tazeem, S. and Ahmad, I. (2017). The weighted exponentiated inverted Weibull distribution: Properties and Application. *Journal of Informatics and Mathematical Sciences*, 9(1), 137–151.

[15] Salama, M. M., Elsherpieny, E. A. and Abd-elaziz, A. A. (2023). The length-biased weighted exponentiated inverted exponential distribution: properties and estimation. *Computational Journal of Mathematical and Statistical Sciences*, 2(2), 181–196.