The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
19(2): S.I (1), 928-935, 2024 www.thebioscan.com

A New Poisson Mixture Distribution: Characterization and Biomedical Application

Rekha Radhakrishnan[1] , D Venkatesan[2] and Prasanth C.B[3*]

[1]Research Scholar, Department of Statistics, Annamalai University,Tamil Nadu,

[2]Professor, Department of Statistics, Annamalai University,Tamil Nadu,

[3]Assistant Professor, Department of Statistics, Sree Keralavarma College ,Kerala.

[*](Corresponding author) prasanthwarriercb@keralavarma.ac.in

DOI: 10.63001/tbs.2024.v19.i02.S.I(1).pp928-935

KEYWORDS
Poisson distribution,
moments,
maximum-likelihood estimates,
real data,
Suja distribution.

Received on:

22-10-2024

Accepted on:

20-11-2024

Published on:

28-12-2024

ABSTRACT

This study presents a novel Poisson mixture distribution, combining elements of the Poisson and Suja distributions. The structural properties of this distribution are derived, including the formulation of the r-th central moments. Additionally, formulas for the coefficient of variation, skewness, and kurtosis are provided, with their behaviors illustrated through graphical representations. Key statistical properties, such as the hazard rate function and generating functions, are also discussed. Methods for parameter estimation, including maximum likelihood estimation and the method of moments, are explored. A simulation study has been conducted to assess the model. In many practical scenarios, real-world datasets do not fit well with conventional distributions. In this case, a dataset of newborn babies' weights from a hospital in Kerala over a specific period is analyzed, focusing on the number of newborns with critically low birth weight (Extremely Low Birth Weight, or ELBW). This data is fitted using the Poisson-Suja distribution. The goodness of fit of the proposed distribution is demonstrated using the count dataset, showing it outperforms the Poisson, Poisson Lindley (PL), and Poisson Akash (PA) distributions.

INTRODUCTION

Sankaran (1970) introduced the discrete Poisson-Lindley distribution, and Shanker (2015) further explored the Shanker distribution, its applications, as well as the Akash distribution and its uses. Shanker and Hagos (2015) provided insights into the Poisson-Lindley distribution and its relevance to biological sciences. In addition, Shanker and Fesshaye (2016) analyzed the modeling of lifetime data using the Akash, Shanker, Lindley, and exponential distributions. Shanker (2016) presented the discrete Poisson-Shanker distribution, and Shanker et al. (2015) examined lifetime data modeling with the exponential and Lindley distributions. Johnson et al. (2005) offered a comprehensive overview of discrete distributions, including their properties and extensions. Fisher (1922) laid the mathematical foundations for theoretical statistics and distributions, while Teicher (1954) investigated distribution mixtures. Rao and Rubin (1964) focused on characterizing the Poisson distribution, and Simar (1976) discussed maximum likelihood estimation for compound Poisson processes. Patil and Ord (1976) explored size-biased sampling and related form-invariant weighted distributions, and Altham (1978) studied two generalizations of the binomial distribution. Consul and Jain (1973) proposed a generalization of the Poisson distribution, and Willmot (1986) explored mixed compound Poisson distributions. Brix and Diggle (2001) detailed spatiotemporal prediction for log-Gaussian Cox processes, while Lambert (1992) analyzed zero-inflated Poisson regression with applications to manufacturing defects. El-Shaarawi and Zhu (2002) discussed modeling skewed data using mixtures of distributions, and Karlis and Xekalaki (2005) conducted an in-depth study on mixed Poisson distributions. Rao (1997) emphasized the practical applications of statistical models, including mixtures, in real-world situations. Lastly, Aitchison and Ho (1989) examined the multivariate Poisson-log normal distribution.

Shanker (2017) introduced the Suja distribution, providing its probability density function (f(x)) and cumulative distribution function (F(x)) for various fields, including biomedical sciences, engineering, insurance, and finance. Also, Sanker (2017) explained Suja Distribution and application in lifetime data. In this context,

$$f(x,\theta) = \frac{\theta^5}{(\theta^4+24)}(1+x^4)e^{-\theta x} \quad ,\theta > 0,\ \ x > 0 \text{ and} \tag{1}$$

$$F(x) = 1-[1+\frac{\theta^4x^4+4\theta^3x^3+12\theta^2x^2+24\theta x}{(\theta^4+24)}]e^{-\theta x}, x>0,\ \ \theta > 0 \tag{2}$$

The Suja distribution is a mixture of Exponential (θ) and Gamma (5,θ) distributions, with the mixing proportion $\frac{\theta^4}{(\theta^4+24)}$.

We have,

$f(x,\theta)$= $pg_1(x)$+(1-p)$g_2(x)$ where p= $\frac{\theta^4}{(\theta^4+24)}$,

$g_1(x) = \theta e^{-\theta x}$ and $g_2(x)$ = $\frac{\theta^5 x^4 e^{-\theta x}}{24}$.

Shanker (2017) analyzed the characteristics of the Suja distribution, highlighting its advantages. Previous studies have demonstrated that the Suja distribution outperforms the Lindley, Akash, and exponential distributions in accuracy. It is expected that the Poisson mixture of the Suja distribution will provide a superior model for count data compared to the Poisson, Poisson-Lindley, and Poisson-Akash distributions.

The Poisson Suja (PS) distribution, a Poisson mixture of the Suja distribution, is introduced in this study, along with a comprehensive examination of its properties, including its form, moments, coefficient of variation, skewness, and kurtosis. The maximum likelihood estimation method for parameter estimation is also discussed. Additionally, the study evaluates the goodness of fit of the PS distribution compared to the Poisson, Poisson Lindley, and Poisson Akash distributions. The paper concludes with a discussion of its findings and implications.

2 Poisson Suja Distribution

When the parameter λ of the Poisson distribution (PD) coincides with the Suja distribution, the Poisson mixture of the Suja distribution can be derived and hereby termed as Discrete Poisson-Suja Distribution(DPSD) or simply Poisson-Suja Distribution(PSD) .

2.1 structural properties of a probability distribution

The structural properties of a probability distribution describe its mathematical and statistical characteristics that define its behavior and applications. These properties include aspects like moments, skewness, kurtosis, and generating functions.

2.1.1 Probability Mass Function (PMF) of PSD

Let λ follow the Suja distribution with a mixing proportion. The resulting probability mass function (PMF) of the DPSD is given by:

$P(X=x) = \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} . f_{suja}(\lambda)\, d\lambda$ where $f_{suja}(\lambda)$ represents the probability density function of the Suja distribution. Substituting the expression for $f_{Suja}(\lambda)$, the explicit form of the Poisson-Suja distribution can be obtained.

$$= \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} . \frac{\theta^5}{(\theta^4+24)}(1+\lambda^4)e^{-\theta\lambda}\, d\lambda$$

$$P(X=x) = \frac{\theta^5}{(\theta^4+24)} \frac{(\theta+1)^4+x^4+10x^3+35x^2+50x+24}{(\theta+1)^{x+5}} \quad , \quad x=0,1,2,\ldots \theta > 0 \qquad (3)$$

The Poisson Suja distribution, characterized by the parameter θ, is defined as a Poisson mixture of the Suja distribution. Its probability mass function (pmf) demonstrates the distribution's behavior for various values of the parameter θ. A graphical representation of the pmf is provided to illustrate the impact of different parameter values on the distribution's shape and behavior.

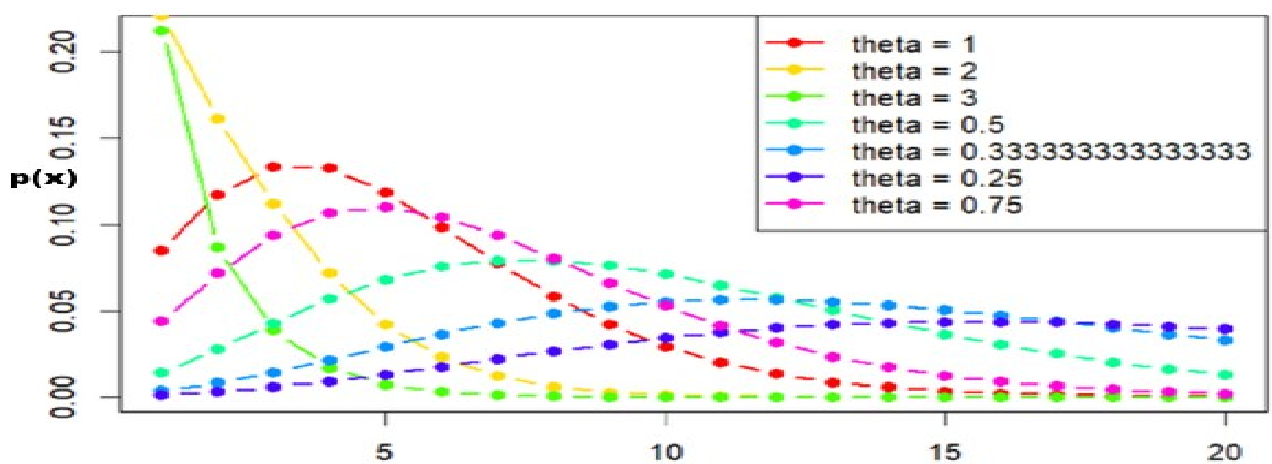

Figure 1. The shape of the probability mass function for different values of theta

2.1.2 Moments

The general form of the r-th factorial moment about the origin for the Poisson Suja distribution can be derived using the gamma integral property and mathematical simplifications. The factorial moment is expressed as:

$$\mu_r' = E[X(X-1)\cdots(X-r+1)]$$

The Poisson Suja distribution's rth factorial moment can be found by

$$\mu'_{(r)} = E\left[E\left(X^{(r)}/\lambda\right)\right] \text{ where } X^{(r)} = x(x-1)(x-2)\ldots..(x-r+1)$$

$$= \frac{\theta^5}{(\theta^4+24)} \int_0^{\infty} \sum_0^{\infty} \left[X^{(r)} \frac{e^{-\lambda}\lambda^x}{x!}\right](1+\lambda^4)e^{-\theta\lambda}\, d\lambda$$

$$= \frac{\theta^5}{(\theta^4+24)} \int_0^{\infty} \sum_0^{\infty} [X^{(r)} \frac{e^{-\lambda}\lambda^x}{(x-r)!}] (1+\lambda^4)e^{-\theta\lambda}\, d\lambda$$

Substituting (x+r) in x within the bracket we obtain,

$$\mu'_{(r)} = \frac{\theta^5}{(\theta^4+24)} \int_0^{\infty} \lambda^r \sum_0^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-r)!} (1+\lambda^4)e^{-\theta\lambda}\, d\lambda$$

$$= \frac{\theta^5}{(\theta^4+24)} \int_0^{\infty} \lambda^r (1+\lambda^4)e^{-\theta\lambda}\, d\lambda$$

For the Poisson Suja distribution, this can be computed by applying the gamma integral property along with simplifications, the explicit form of the r-th factorial moment can be obtained.

$$\mu'_{(r)} = \frac{r!}{\theta^r}\left[\frac{(\theta^4+(r+1)(r+2)(r+3)(r+4))}{(\theta^4+24)}\right] \quad r=1,2,3,\ldots \qquad (4)$$

We obtain the first four factorial moments about origin by substituting r=1,2,3, 4. By applying the connection between factorial moments and moments about origin, the following 4 moments about origin" are obtained:

$$\mu_1' = \frac{\theta^4+120}{\theta(\theta^4+24)} \qquad \mu_2' = \frac{\theta^5+2\theta^4+120\theta+720}{\theta^2(\theta^4+24)}$$

$$\mu_3' = \frac{\theta^6+6\theta^5+6\theta^4+120\theta^2+2160\theta+5040}{\theta^3(\theta^4+24)}$$

$$\mu_4' = \frac{\theta^7+14\theta^6+36\theta^5+24\theta^4+120\theta^3+5040\theta^2+30240\theta+40320}{\theta^4(\theta^4+24)}$$

The graph of the mean for Poisson Suja distribution for various values of θ is given in Figure 2

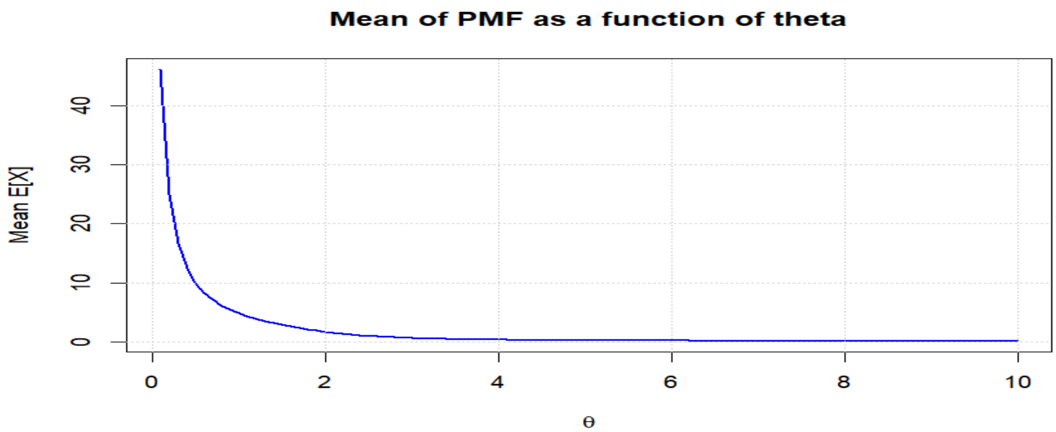

Figure 2. Mean of PS distribution

Using the relationship between moments about the mean (μr) and moments about the origin (Mr), the first four moments about the mean for the Poisson Suja distribution can be expressed as follows:

First Moment about the Mean (μ_1) is equal to zero always. $\mu_1=0$ (By definition of moments about the mean.)

Second Moment about the Mean (μ_2) (Variance): = $M_2-(M_1)^2$

Third Moment about the Mean (μ_3) (Skewness-related): = $M_3-3M_2M_1+2(M1)^3$

Fourth Moment about the Mean (μ4) (Kurtosis-related): = $M_4-4M_3M_1+6M_2(M_1)^2-3(M_1)^4$

Here, M_1,M_2,M_3,M_4 are the first four moments about the origin for the Poisson Suja distribution, which can be derived based on its probability mass function. Substituting these values will yield the explicit forms of μ2,μ3 and μ4.

$$\mu_2 = \frac{\theta^9+\theta^8+144\theta^5+528\theta^4+2880\theta+2880}{\theta^2(\theta^4+24)^2}$$

$$\mu_3 = \frac{\theta^{14}+3\theta^{13}+2\theta^{12}+168\theta^{10}+1656\theta^9+3024\theta^8+6336\theta^6+46656\theta^5+3456\theta^4+132\theta^3+69120\theta^2+207360\theta+138240}{\theta^3(\theta^4+24)^3}$$

$$\mu_4 = (\theta^{19}+10\theta^{18}+18\theta^{17}+9\theta^{16}+192\theta^{15}+4896\theta^{14}+22464\theta^{13}+23904\theta^{12}+10368\theta^{11}+281088\theta^{10}+946944\theta^9+528768\theta^8+221184\theta^7+5584896\theta^6+12939264\theta^5+11114496\theta^4+165888\theta^3+36495360\theta^2+69672960\theta+34836480)\ [\theta^4(\theta^4+24)^4]^{-1}$$

2.1.3 Coefficient of Variation, Coefficient of Skewness and Coefficient of Kurtosis

The coefficient of variation(CV) is a measure of the relative dispersion or variability of a dataset, expressed as a proportion of the mean. It is calculated as: CV = $\frac{\sigma}{\mu}$ ×100, where σ is the standard deviation and μ is the mean. The CV is useful for comparing the variability of datasets with different units or scales. A higher CV indicates greater relative variability. Interpretation: CV = 0: No variability (all values are identical).Higher CV: More dispersion relative to the mean.

The coefficient of skewness (β_1 and β_2) quantifies the asymmetry of a probability distribution. It is defined as:

Skewness = $\beta_2 = \sqrt{\beta_1} = \frac{\mu_3}{\sigma^3}$, where μ_3 is the third central moment and σ is the standard deviation. It helps determine whether a distribution is symmetric or has a longer tail on one side. Interpretation: Skewness = 0: The distribution is symmetric, Skewness > 0: The distribution is positively skewed (longer tail on the right) and Skewness < 0: The distribution is negatively skewed (longer tail on the left).

The coefficient of kurtosis measures the "tailedness" or peakedness of a distribution. It is given by:Kurtosis = $\gamma_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$ where μ_4 the fourth central moment and σ is the standard deviation. Kurtosis indicates how heavy or light the tails of a distribution are compared to a normal distribution. Interpretation: Excess Kurtosis = γ_2 = Kurtosis – 3. If γ_2= 0: Mesokurtic (normal-like tails), If γ_2 > 0: Leptokurtic (heavy tails, sharp peak) and If γ_2 < 0: Platykurtic (light tails, flat peak).

These coefficients are essential for understanding the shape, spread, and variability of a distribution in statistical analysis. The CV, β_2 and γ_2 of the Poisson Suja distribution can be obtained by employing the subsequent formula: CV =$\frac{\sigma}{\mu_1'}$, $\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}}$, $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

The behaviour of CV, skewness and kurtosis of Poisson Suja distribution for different values of θ are given in Figure 3, Figure 4 and Figure 5.

$$CV = \frac{\sigma}{\mu_1'} = \frac{\sqrt{\frac{\theta^9+\theta^8+144\theta^5+528\theta^4+2880\theta+2880}{\theta^2(\theta^4+24)^2}}}{\frac{\theta^4+120}{\theta(\theta^4+24)}} \quad (5)$$

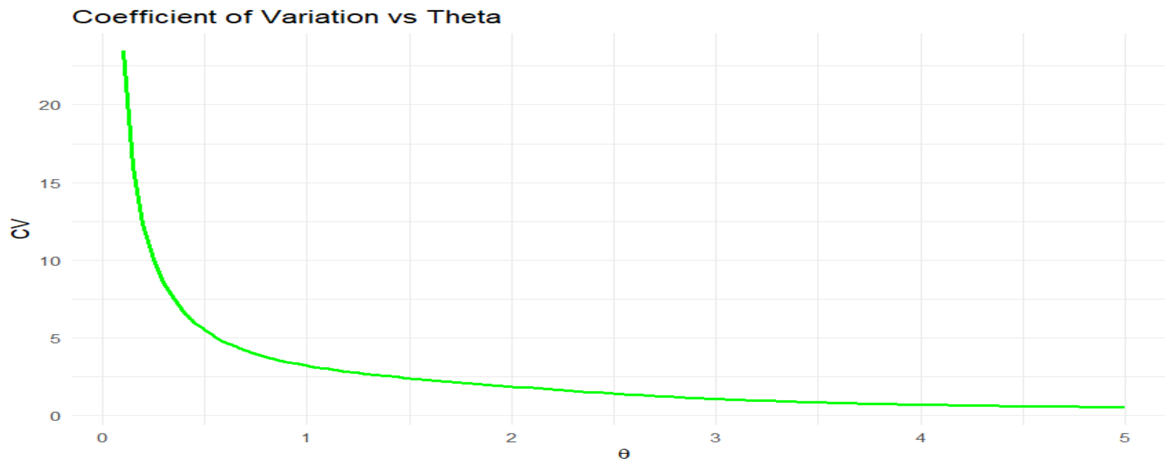

Figure 3. Coefficient of Variation & theta

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}}$$

$$\sqrt{\beta_1} = \frac{(\theta^{14}+3\theta^{13}+2\theta^{12}+168\theta^{10}+1656\theta^{9}+3024\theta^{8}+6336\theta^{6}+46656\theta^{5}+3456\theta^{4}+132\theta^{3}+69120\theta^{2}+207360\theta+138240)}{(\theta^{9}+\theta^{8}+144\theta^{5}+528\theta^{4}+2880\theta+2880)^{3/2}} \quad (6)$$

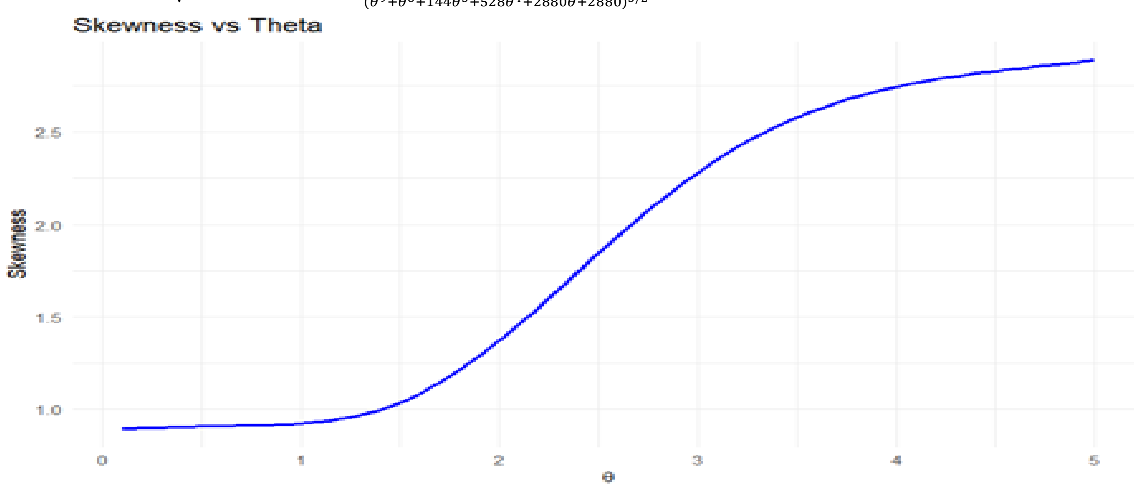

Figure 4. Coefficient of Skewness & theta

The graph of the Poisson Suja distribution reveals the following observations and properties regarding skewness as the parameter θ varies:

Low Values of θ: At very low values of θ, the skewness increases, indicating a distribution with a pronounced right tail. This suggests that the probability mass is concentrated on the lower values, with a long tail extending to the right.

Skewness Peak (θ≈3.5): At around θ≈3.5, the skewness reaches its maximum value. This corresponds to the point where the distribution is most strongly skewed to the right, indicating the highest asymmetry.

Post-Peak Decrease in Skewness: Beyond the peak, as θ increases, the skewness decreases. This implies that the distribution becomes more symmetric, with the right tail diminishing in length relative to the central mass.

Secondary Increase in Skewness: At higher values of θ, the skewness begins to increase gradually again after the initial decrease. This secondary increase indicates that the distribution regains a rightward skew at larger θ values, though not as pronounced as the initial peak.

These observations highlight the dynamic behavior of skewness in the Poisson Suja distribution, reflecting its adaptability across a range of parameter values.

Asymptotic Behaviour: For larger values of θ, the skewness continues to increase but at a slower rate. This suggests that the distribution is still slightly right-skewed as θ increases further.

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\beta_2 = \frac{(\theta^{19}+10\theta^{18}+18\theta^{17}+9\theta^{16}+192\theta^{15}+4896\theta^{14}+22464\theta^{13}+23904\theta^{12}+10368\theta^{11}+281088\theta^{10}+946944\theta^{9}+528768\theta^{8}+221184\theta^{7}+5584896\theta^{6}+12939264\theta^{5}+11114496\theta^{4}+165888\theta^{3}+36495360\theta^{2}+69672960\theta+34836480)}{(\theta^{9}+\theta^{8}+144\theta^{5}+528\theta^{4}+2880\theta+2880)^{2}} \quad (7)$$

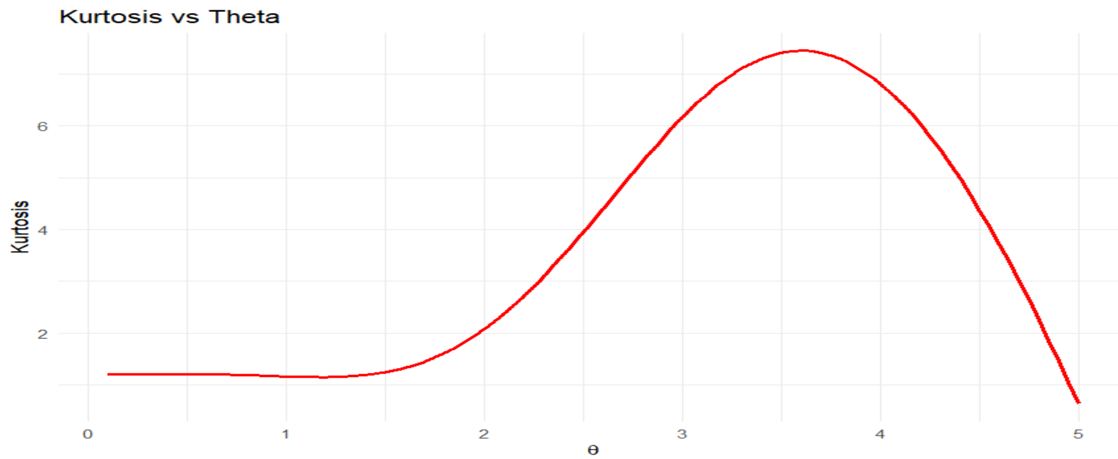

Figure 5: Coefficient of Kurtosis & Theta

The graph of the Poisson Suja distribution illustrates the following observations and properties regarding kurtosis as the parameter θ changes:

Rapid Increase at Low θ: At very low values of θ, the kurtosis rises sharply, indicating a highly peaked and heavy-tailed distribution. This reflects a distribution that is initially leptokurtic.

Peak Kurtosis (θ≈3.5): The kurtosis reaches its maximum value around θ≈3.5, marking the most leptokurtic point of the distribution. At this value, the distribution is at its most peaked and exhibits the heaviest tails.

Gradual Decline After the Peak: Beyond the peak, as θ increases, the kurtosis decreases gradually. This suggests that the distribution becomes less peaked and less heavy-tailed, moving away from leptokurtic behavior.

Stabilization at High θ: For large values of θ, the kurtosis approaches a relatively stable value. This indicates that the distribution transitions to a more platykurtic shape, becoming flatter and more normal-like.

Overall Trend: The kurtosis exhibits an initial rapid rise to a peak, followed by a gradual decline and eventual stabilization. This implies that at low θ, the distribution is highly peaked and heavy-tailed, while at higher θ, it becomes flatter and more symmetric.

These observations suggest that the Poisson Suja distribution is versatile, with its shape and tail behavior strongly influenced by the parameter θ.

3. Estimation of Parameter

3.1 Maximum Likelihood Estimate

The application of maximum likelihood estimation (MLE) for estimating the parameters of the Poisson Suja distribution is detailed in this section.

Let x1,x2,x3,…,xn represent a random sample of size nnn drawn from the Poisson Suja distribution. Let fx denote the observed frequency for X=x (x=1,2,…,k), where k is the largest observed value with a non-zero frequency. The total frequency satisfies: Σfx = n

The likelihood function L for the Poisson Suja distribution is given by:

$$L=\prod P(X=x)^{fx}$$

where P(X=x) is the probability mass function (pmf) of the Poisson Suja distribution. Substituting the expression for P(X=x), the likelihood function becomes:

$$L=\prod \left(\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} . f_{suja}(\lambda)\, d\lambda\right)^{fx}$$

Taking the natural logarithm, the log-likelihood function is:

$$\log(L) = \sum f_x \log\left(\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} . f_{suja}(\lambda)\, d\lambda\right).$$

The parameter θ of the Poisson Suja distribution is estimated by maximizing this log-likelihood function with respect to θ. Numerical methods or optimization techniques are typically used to solve for θ, as the integration and logarithmic terms make an analytical solution challenging.

$$L = \left(\frac{\theta^5}{\theta^4+24}\right)^n \frac{1}{(\theta+1)^{\sum_{x=1}^{k} fx(x+5)}} \prod_{x=1}^{k}\left((\theta+1)^4 + x^4 + 10x^3 + 35x^2 + 50x + 24\right)^{fx}$$

This provides "the log likelihood function as

$$\text{Log } L = n\log\left(\frac{\theta^5}{\theta^4+24}\right)^n - \sum_{x=1}^{k} fx(x+5)\log(\theta+1) + \sum_{x=1}^{k} fx\left((\theta+1)^4 + x^4 + 10x^3 + 35x^2 + 50x + 24\right)$$

The log" probability function's first derivative can be found using

$$\frac{\partial \log L}{\partial \theta} = \frac{5n}{\theta} - \frac{4n\theta^3}{\theta^4+24} - \frac{\sum_{x=1}^{k} fx(x+5)}{\theta+1} + \frac{\sum_{x=1}^{k} fx4(\theta+1)^3}{(\theta+1)^4+x^4+10x^3+35x^2+50x+24} \quad (8)$$

The Maximum likelihood estimate $\hat{\theta}$ of θ for Poisson Suja distribution is the solution of the equation $\frac{\partial \log L}{\partial \theta}=0$ and is given by the non-linear equation.The given non linear equation" can be solved by numerical iteration methods.

3.2 Method of Moment Estimate

Let (x1,x2,x3,…,xn) be a random sample of size n from PSD. Eqating the population mean to the corresponding sample mean $(\bar{x})$.

$$\bar{x} = \frac{\theta^4+120}{\theta(\theta^4+24)}$$

We get the equation $\bar{x}\theta^5 - \theta^4 + 24\theta\bar{x} - 120 = 0$,the solution of the equation gives the moment estimator for the parameter θ.This equation can be solved by Newton Raphson method.

4 Statistical Properties

4.1 Increasing Hazard Rate

The Poisson Suja distribution has an increasing hazard rate because

$$\frac{P(x+1,\theta)}{P(x,\theta)} = \frac{1}{(\theta+1)}\left[1+\frac{4x^3+36x^2+104x+116}{(\theta+1)^4+x^4+10x^3+35x^2+50x+24}\right]$$

The function shows a decreasing trend indicating that it is log concave.Which impliews that the PSD exhibit an increasing hazard rate and follows a unimodal pattern.The relationship between logconcavity, unimodality and increasing hazard rate in descrete distributionhas been extensively studied by Grandell(1997).

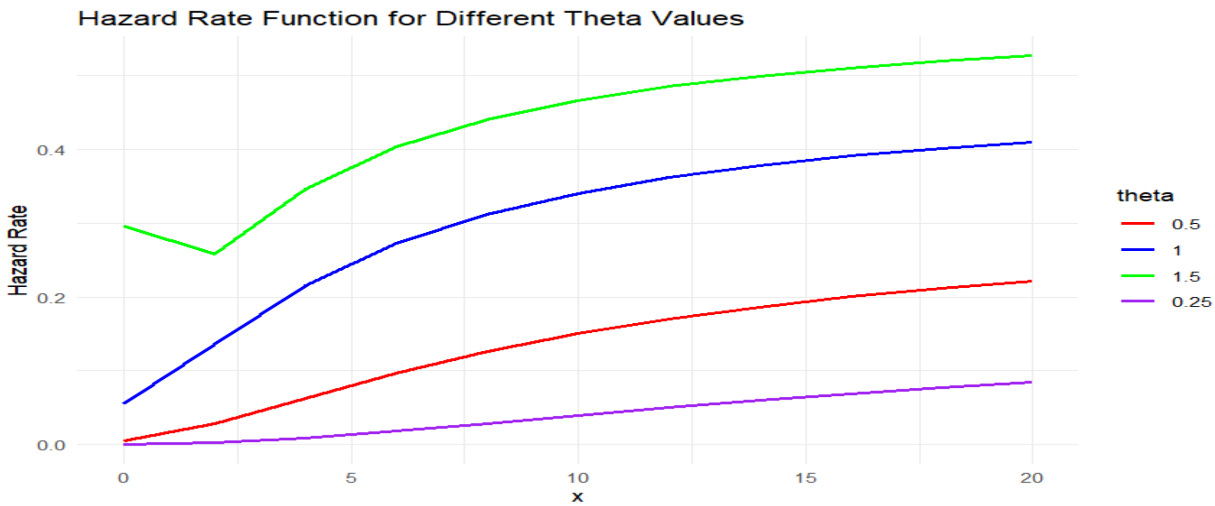

Figure 6: Hazard rate function for different values of theta

4.2 Overdispersion

The Poisson Suja distribution is always over dispersed ($\sigma^2 > \mu$).

We have $\sigma^2 = \frac{\theta^9+\theta^8+144\theta^5+528\theta^4+2880\theta+2880}{\theta^2(\theta^4+24)^2}$

$$= \frac{\theta^4+120}{\theta(\theta^4+24)}\left[1+\frac{\theta^8+528\theta^4+2880}{\theta(\theta^4+24)(\theta^4+120)}\right]$$

$$= \mu\left[1+\frac{\theta^8+528\theta^4+2880}{\theta(\theta^4+24)(\theta^4+120)}\right] > \mu$$

This shows that PS distribution is always over dispersed .Therefore PSD is used for discrete data set which is overdispersed in nature.

4.3 Generating Functions

The probability generating function of PSD is given as

$$P_x^{(t)} = \sum_{x=0}^{\infty} t^x \frac{\theta^5}{(\theta^4+24)}\left[\frac{(\theta+1)^4+x^4+10x^3+35x^2+50x+24}{(\theta+1)^{x+5}}\right]$$

$$= \frac{\theta^5}{(\theta^4+24)}\sum_{x=0}^{\infty} t^x \left[\frac{(\theta+1)^4+x^4+10x^3+35x^2+50x+24}{(\theta+1)^{x+5}}\right]$$

$$= \frac{\theta^5}{(\theta^4+24)(\theta+1)^5}\sum_{x=0}^{\infty} t^x \left[\frac{(\theta+1)^4+x^4+10x^3+35x^2+50x+24}{(\theta+1)^{x}}\right]$$

$$P_x^{(t)} = \frac{\theta^5}{(\theta^4+24)(\theta+1)^4}\left[\frac{(\theta+1)^4+24}{(\theta+1-t)} + \frac{50t}{(\theta+1-t)^2} + \frac{35t(\theta+1+t)}{(\theta+1-t)^3} \frac{10t\left((\theta+1)^2+4t(\theta+1)+t^2\right)}{(\theta+1-t)^4} \frac{(\theta+1)^3t+11t^2(\theta+1)^2+11t^3(\theta+1)+t^4}{(\theta+1-t)^5}\right]$$

The moment generating function of PSD is thus given as

$$M_x(t) = \frac{\theta^5}{(\theta^4+24)(\theta+1)^4}\left[\frac{(\theta+1)^4+24}{(\theta+1-e^t)} + \frac{50e^t}{(\theta+1-e^t)^2} + \frac{35e^t(\theta+1+e^t)}{(\theta+1-e^t)^3} + \frac{10e^t\left((\theta+1)^2+4e^t(\theta+1)+e^{2t}\right)}{(\theta+1-e^t)^4} + \frac{(\theta+1)^3e^t+11e^{2t}(\theta+1)^2+11e^{3t}(\theta+1)+e^{4t}}{(\theta+1-e^t)^5}\right]$$

5. Simulation Analysis

Simulation analysis uses models to represent real-world systems or processes. These models are then used to run simulations, allowing analysts to explore different scenarios and conditions without real-world implementation. Simulation provides data and insights that support better decision-making, leading to improved outcomes. By conducting simulated variable analysis with the pdf of PSD we get the following results,Table 1.

Table 1 The descriptive statistics of simulated data

Simulated sample size n = 5000			
Mean	0.9783	Sample Variance	1.9257
Median	0	Kurtosis	0.9360
Mode	0	Skewness	1.4099
Minimum	0	Largest(1)	5
Maximum	5	Smallest(1)	0

Skewness and kurtosis are two statistical measures used to describe the shape and characteristics of a distribution. They provide insights into the symmetry and the tail behavior of the data. urtosis measures the "tailedness" or the sharpness of the peak of a probability distribution. It provides insight into how outliers or extreme values are distributed. **Skewness** focuses on the *symmetry* of the data, while **kurtosis** focuses on the *tail behavior* (extremes or outliers) of the distribution.Skewness measures the asymmetry of a probability distribution. It indicates whether the data are skewed to the left (negative skewness) or to the right (positive skewness).It is known that the value of Skewness coefficient will be zero for a symmetrical distribution. If mean >mode, the coefficient of skewness is positive else negative. It is lies between −1 and 1 for considerably skewed destitution. Its highly positive skewed if it is greater than 1. That is, here the distribution from simulation analysuis shows a non-symmetric, highly positively skewed distribution.

6. Application

Birth weight, the weight of a baby at birth is an important indicator of a baby's health and development Low Birth Weight (LBW) is a birth weight of less than 2,500 grams. Very Low Birth Weight (VLBW) is a birth weight of less than 1,500 grams. Extremely Low Birth Weight (ELBW)is a birth weight of less than 1,000 grams. Normal Birth Weight: A birth weight between 2,500 and 4,000 grams. High Birth Weight: A birth weight greater than 4,000 grams.

A real data set of weight of newborn babies from a Hospital at Kerala is noted for a particular period. The number of newborn babies with crucial under weight (Extremely Low Birth Weight (ELBW)) is the variable under concern and has been fitted with the Poisson Suja distribution. The Poisson Distribution(PD), Poisson Lindley Distribution (PLD), Poisson -Akash Distribution (PAD), and Poisson Suja Distribution(PSD) that have been fitted for the count data set are shown at Table 2.

Table 2 Distribution of weight of newborn babies with ELBW

No. of of newborn babies with ELBW per six months period	Observed Frequency	Expected Frequency			
		PD	PLD	PAD	PSD
0	35	27.41	33.06	33.5	34.58
1	11	21.47	15.27	14.7	13.3
2	8	8.4	6.74	6.6	6.27
3	4	2.19	2.88	2.9	3.13
4	2	0.43	1.207	1.29	1.52
Total	60	60	60	60	60
ML Estimate		0.7833	1.7434	2.0779	2.737
Chi-square		7.98	2.20	1.40	1.28
P value		0.0047	0.1380	0.4966	0.7337

R software is employed to estimate the unknown parameters and compute model comparison criteria. The criteria considered include the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), corrected Akaike Information Criterion (AICC), and -2logL. The distribution that yields lower values for AIC, BIC, AICC, and -2logL is considered the better fit. The formulas below are used to calculate these criterion values.

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

Here, k represents the number of parameters in the statistical model, n denotes the sample size, and −2logL is the maximized value of the log-likelihood function for the model being analyzed.

Table 3: Comparison and Performance of Fitted distributions for the Real Data Set.

Distributions	MLE	S.E	-2logL	AIC	BIC	AICC
PD	0.7833	0.1142	155.091	157.091	159.185	157.160
PLD	1.743	0.2817	146.70	148.70	150.79	148.771
PAD	2.077	0.1648	146.418	148.418	150.513	148.487
PSD	2.734	0.1841	146.201	148.201	150.295	148.270

The PSD fits data much more closely than the PD, PLD, and PA distributions since PSD has the lesser *AIC, BIC, AICC* and *-2logL* values as compared to the other distributions.

CONCLUSION

In this study, a new Poisson mixture distribution was developed by combining the Poisson distribution with the Suja distribution, and its properties were thoroughly characterized. The probability mass function was presented, along with expressions for both raw and central moments. Formulas for the coefficient of skewness and kurtosis were also derived. Parameter estimation was discussed using both maximum likelihood estimation and the method of moments. The generating functions for the distribution were derived as well. To assess its goodness of fit, the Poisson Suja (PS) distribution was applied to a real data set and compared to the Poisson (PD), Poisson Lindley (PL), and Poisson Akash (PA) distributions. The results indicated that the Poisson Suja distribution offers a superior fit.

REFERENCE

- Aitchison, J., & Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika, 76*(4), 643-653. https://doi.org/10.1093/biomet/76.4.643
- Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 27*(3), 162–167. https://doi.org/10.2307/2346948
- Brix, A., & Diggle, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(4), 823-841. https://doi.org/10.1111/1467-9868.00313
- Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics, 15*(4), 791–799. https://doi.org/10.2307/1267352
- El-Shaarawi, A. H., & Zhu, R. (2002). Modelling skewed data with mixtures of distributions. *Canadian Journal of Statistics, 30*(4), 563-581. https://doi.org/10.2307/3316156
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222*, 309–368. https://doi.org/10.1098/rsta.1922.0009

- Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3rd ed.). John Wiley & Sons.
- Karlis, D., & Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review, 73*(1), 35–58. https://doi.org/10.1111/j.1751-5823.2005.tb00251.x
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34*(1), 1–14. https://doi.org/10.2307/1269547
- Patil, G. P., & Ord, J. K. (1976). On size-biased sampling and related form-invariant weighted distributions. *Biometrika, 63*(1), 145–153. https://doi.org/10.1093/biomet/63.1.145
- Rao, C. R. (1997). *Statistics and truth: Putting chance to work* (2nd ed.). World Scientific.
- Rao, C. R., & Rubin, H. (1964). On a characterization of the Poisson distribution. *Sankhya: The Indian Journal of Statistics, Series A, 26*(4), 295–298.
- Sankaran, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics, 26*(1), 145–149. https://doi.org/10.2307/2529053
- Shanker, R. (2015). Akash distribution and its applications. *International Journal of Probability and Statistics, 4*(3), 65–75. https://doi.org/10.5923/j.ijps.20150403.01
- Shanker, R. (2015). Shanker distribution and its applications. *International Journal of Statistics and Applications, 5*(6), 338–348. https://doi.org/10.5923/j.statistics.20150506.08
- Shanker, R. (2016). The discrete Poisson-Shanker distribution. *Jacobs Journal of Biostatistics, 1*(1), 1–7. https://doi.org/10.15406/bbij.2017.05.00121
- Shanker, R. (2017). Suja distribution and its applications. *International Journal of Probability and Statistics, 6*(2), 11–19. https://doi.org/10.5923/j.ijps.20170602.01
- Shanker, R. (2017). Suja distribution and its applications in modeling lifetime data. *International Journal of Statistics and Reliability Engineering, 4*(1), 47–56.
- Shanker, R., & Fesshaye, T. (2016). On modeling of lifetime data using Akash, Shanker, Lindley, and exponential distributions. *Biometrics & Biostatistics International Journal, 3*(6), 1-9. https://doi.org/10.15406/bbij.2016.03.00084
- Shanker, R., & Hagos, F. (2015). On Poisson-Lindley distribution and its applications to biological sciences. *Biometrics & Biostatistics International Journal, 2*(4), 1–5. https://doi.org/10.15406/bbij.2015.02.00036
- Shanker, R., Hagos, F., & Sujatha, S. (2015). On modeling of lifetime data using exponential and Lindley distributions. *Biometrics & Biostatistics International Journal, 2*(5), 1-9. https://doi.org/10.15406/bbij.2015.02.00042
- Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *The Annals of Statistics, 4*(6), 1200–1209. https://doi.org/10.1214/aos/1176343651
- Teicher, H. (1954). On the mixtures of distributions. *The Annals of Mathematical Statistics, 25*(3), 581–587. https://doi.org/10.1214/aoms/1177728735
- Willmot, G. E. (1986). Mixed compound Poisson distributions. *Biometrika, 73*(2), 317–321. https://doi.org/10.1093/biomet/73.2.317
- Grandell, J. (1997): Mixed Poisson Processes, Chapman& Hall, London.