Riki Tabassum – Rama Shanker

On statistical properties and application of Poisson-Sujatha distribution

Riki Tabassum, Department of Statistics, Assam University, Silchar, Assam, India
Email:rikitabassum18@gmail.com
Rama Shanker, Department of Statistics, Assam University, Silchar, Assam, India
Email:shankerrama2009@gmail.com

The Poisson-Sujatha distribution, the Poisson compound of Sujatha distribution, was introduced to model count data of over-dispersed nature. The proposed distribution has been shown to be log-concave and is a two-component mixture of negative binomial distributions. The hazard function of the distribution has been derived and is increasing and thus the distribution is unimodal. The simulation study has been presented to show the consistency of maximum likelihood estimator of the parameter of Poisson-Sujatha distribution. The Poisson-Sujatha distribution has been applied to three real datasets one from biological science and two from thunderstorm events and found to provide quite satisfactory fit over other competing one parameter over-dispersed count distributions including Poisson-Lindley distribution, Poisson-Garima distribution, Poisson-Komal distribution and Poisson-Akash distribution.

Keywords: Sujatha distribution, Poisson-Sujatha distribution, statistical properties, maximum likelihood estimation, simulation, goodness of fit

The Poisson distribution is one of the earliest and most widely known discrete probability models used for modeling count data, especially when the data exhibit equi-dispersion, i.e., when the mean equals the variance. As a limiting case of the binomial distribution, it has been widely applied across diverse fields such as insurance, agriculture, environmental studies, biological sciences, clinical trials, and engineering. However, real-world count data often display dispersion patterns that deviate from the equi-dispersion assumption. In many practical situations, the data tend to be either over-dispersed (variance greater than the mean) or under-dispersed (variance less than the mean), which limits the applicability of the Poisson distribution. To address this, a variety of generalized count distributions have been proposed over the past few decades, particularly to model over-dispersion. Notable among these is the negative Binomial distribution, which arises as a Poisson-gamma mixture and is commonly used when data exhibit high variability. Another important development is the Poisson-Lindley distribution (PLD), introduced by *Sankaran (1970)*, which results from compounding the Poisson distribution with the Lindley distribution proposed by *Lindley (1958)*. Building on this line of research, *Shanker (2016a)* introduced the Sujatha distribution, and later, *Shanker (2016b)* proposed the Poisson-Sujatha distribution (PSD) as the Poisson compound of Poisson distribution with Sujatha distribution. Subsequent research includes the Discrete Poisson-Akash distribution (PAD) by *Shanker (2017a)* and the Poisson-Garima distribution (PGD) developed by *Shanker (2017b)*. More recently, *Shanker et al. (2025)* introduced the Poisson-Komal distribution (PKD). These models offer greater flexibility in capturing the distributional characteristics of real-world count data and continue to enhance the statistical toolbox for modeling dispersion.

The Sujatha distribution introduced by *Shanker (2016a)* is defined by its probability density function (pdf)

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

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

In recent times, researchers have shown growing interest in extending the Sujatha distribution and Poisson-Sujatha distributions to better suit different types of count data. To handle datasets where zero counts are not present, *Shanker and Hagos (2015a, 2016)* proposed the zero-truncated and size-biased versions of the PSD. Recognizing the need to account for excess zeros in data, *Da Silva et al. (2018)* introduced a zero-modified Poisson-Sujatha model. Around the same time, *Subramanian and Shenbagaraja (2020)* developed the length-biased quasi-Sujatha distribution, particularly useful for length-biased sampling situations. Further contributions came from *Aderoju and Adeniyi (2022),* who introduced a flexible model known as the zero-inflated generalized Poisson-Sujatha distribution (ZIGPSD). They showed how reparametrizing the probability function could convert it into a hurdle model, making it more effective in handling zero-modified data. Their work also included theoretical insights and practical examples to support the model's usefulness. Other developments include the Alpha Power Transformed Sujatha distribution proposed by *Mohiuddin and Kannan (2022)*, which explored how transformation techniques could enhance flexibility and fit. More recently, *Shanker and Prodhani (2024)* presented a size-biased Sujatha distribution and demonstrated its application in modeling flood data. Further, *Prodhani and Shanker (2024)* introduced a three-parameter power Sujatha distribution, expanding the family further with detailed properties and real-life applications.

A random variable X is said to be Poisson-Sujatha distribution (PSD) if it follows the stochastic representation $X \mid \lambda \sim$ Poisson(λ) and $\lambda \mid \theta \sim$ Sujatha (θ) for $\lambda > 0$, $\theta > 0$ and the unconditional distribution of the stochastic representation as PSD (θ). That is the probability mass function of PSD can be obtained as

$$P(X=x)=\int_0^\infty P(X=x\mid\lambda)\,f(\lambda\mid\theta)\,d\lambda$$

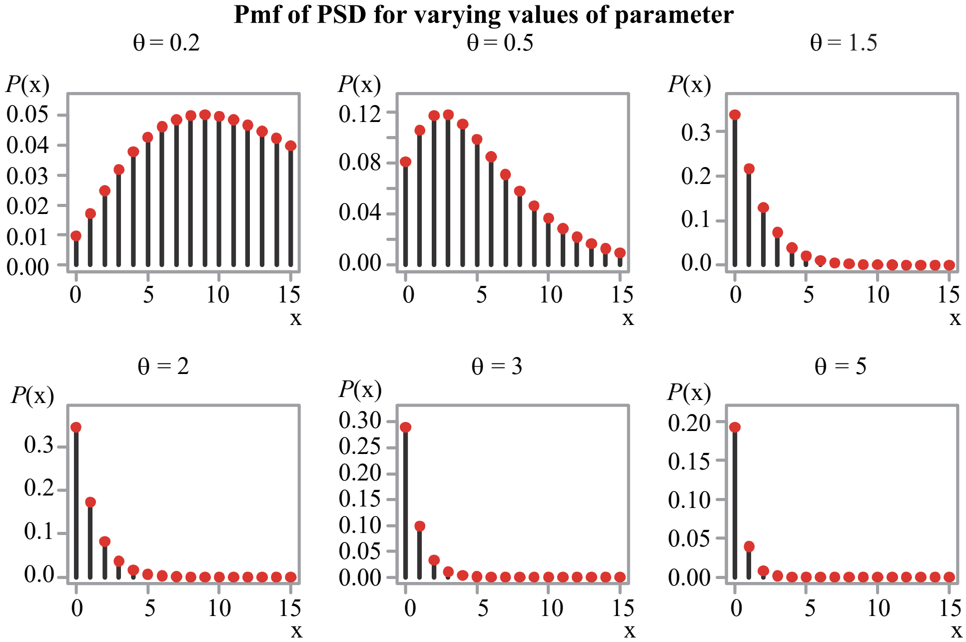
$$=\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\cdot\frac{\theta^3}{\theta^2+\theta+2}\left(1+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

$$=\frac{\theta^3}{\theta^2+\theta+2}\cdot\frac{x^2+(\theta+4)x+\left(\theta^2+3\theta+4\right)}{(\theta+1)^{x+3}}\;;x=0,1,2,...,\theta>0$$

The nature of the pmf (probability mass function) of PSD for varying values of parameter are shown in figure 1 and it is clear that PSD highly positively skewed for increasing values of parameter.

Figure 1

Pmf of PSD for varying values of parameter



The first four moments about origin and variance of PSD obtained by *Shanker (2016b)* is given by

$$\mu_1' = \frac{\theta^3 + 2\theta + 6}{\theta\left(\theta^3 + \theta + 2\right)}$$

$$\mu_2' = \frac{\theta^4 + 2\theta^3 + 2\theta^2 + 12\theta + 24}{\theta^2\left(\theta^3 + \theta + 2\right)}$$

$$\mu_3' = \frac{\theta^5 + 6\theta^4 + 8\theta^3 + 24\theta^2 + 96\theta + 120}{\theta^3\left(\theta^3 + \theta + 2\right)}$$

$$\mu_4' = \frac{\theta^6 + 14\theta^5 + 38\theta^4 + 72\theta^3 + 312\theta^2 + 840\theta + 720}{\theta^4\left(\theta^3 + \theta + 2\right)}$$

The detailed discussion about its properties, estimation of parameter, and applications has been discussed by *Shanker(2016c)* and it has been shown that it is better than Poisson distribution and Poisson-Lindley distribution by *Shanker and Hagos (2015b)* for modeling count data in various fields of knowledge. Although, *Shanker (2016b)* discussed several statistical properties of PSD butit seems that some interesting and useful statistical and reliability properties of PSD have not been discussed by *Shanker (2016b)*.

The main purpose of this paper is to derive some interesting and useful statistical and reliability properties of PSD, which has not been studied earlier by *Shanker (2016b)*. We have shown that the proposed distribution is log-concave and is a three-component mixture of negative binomial distributions. The hazard function of the distribution has been shown to be increasing and thus the distribution is unimodal. For showing the consistency of the maximum likelihood estimator of the distribution, simulation study has been presented. The goodness of fit of the PSD has been established with three real datasets and observed that it provides best fit over other competing one parameter over-dispersed distributions.

1. Statistical properties

In this section three important statistical results of PSD has been made to proved. The PSD is positively skewed, unimodal and decreasing which is shown in theorems 1 and 2. In theorem 3, it has been proved that PSD is also a three-component mixture of negative binomial distributions with different parameter (number of successes) and for the same probability of success. It can be easily shown that PSD has increasing hazard rate (IHR) and is unimodal.

Since $Q(x,\theta)=\dfrac{P(x+1,\theta)}{P(x,\theta)}=\dfrac{1}{\theta+1}\left[1+\dfrac{2x+\theta+5}{x^2+(\theta+4)x+(\theta^2+3\theta+4)}\right]$ is a decreasing function of x for a given $\theta, P(x,\theta)$ is log-concave. This implies that PSD has an increasing hazard rate and is unimodal. *Grandell (1997)* has detailed discussion about relationship between log-concavity, IHR and Unimodality of discrete distributions.

Theorem 1: The $Q(x,\theta)$ is decreasing function of x for given θ.

Proof: We have

$$Q(x,\theta)=\frac{P(x+1,\theta)}{P(x,\theta)}=\frac{1}{\theta+1}\left[1+\frac{2x+\theta+5}{x^2+(\theta+4)x+(\theta^2+3\theta+4)}\right]$$

This gives $Q'(x,\theta)=\dfrac{-\left(2x^2+2\theta x+10x-\theta^2+3\theta+12\right)}{(\theta+1)\left[x^2+(\theta+4)x+(\theta^2+3\theta+4)\right]^2}$.

Since $Q'(x,\theta)<0$, $Q(x,\theta)$ is decreasing function of x for given θ.

Theorem 2: The pmf $P(x,\theta)$ of PSD is log-concave.

Proof: Given that

$$P(x,\theta)==\frac{\theta^3}{\theta^2+\theta+2}\cdot\frac{x^2+(\theta+4)x+(\theta^2+3\theta+4)}{(\theta+1)^{x+3}}$$

We have

$$\log P(x,\theta)=3\log\theta+\log\left[x^2+(\theta+4)x+(\theta^2+3\theta+4)\right]-\log\left(\theta^2+\theta+2\right)-(x+3)\log(\theta+1)$$

Assuming $g(x,\theta)=\log P(x,\theta)$, we have

$$g'(x,\theta)=\frac{2x+\theta+4}{x^2+(\theta+4)x+(\theta^2+3\theta+4)}-\log(\theta+1)$$

and $g''(x,\theta)=\dfrac{-\left(2x^2+2\theta x+10x-\theta^2+2\theta+8\right)}{\left[x^2+(\theta+4)x+(\theta^3+3\theta+4)\right]^2}<0$.

This means that the pmf of PSD is log-concave.

Theorem 3: The PSD is a three-component mixture of negative binomial distributions and can be expressed as

$$P(x;\theta)=p_1 P_1(x;\theta)+p_2 P_2(x;\theta)+p_3 P_3(x;\theta) \quad ; p_1+p_2+p_3=1,$$

where $P_i(x;\theta)$ is the pmf of the negative binomial distribution(NBD) with parameter the number of successes i and $p_1=\dfrac{\theta^2}{\theta^2+\theta+2}$, $p_2=\dfrac{\theta}{\theta^2+\theta+2}$, $p_3=\dfrac{2}{\theta^2+\theta+2}$ with $P_1(x;\theta)=\dfrac{\theta}{(\theta+1)^{x+1}}$ as $NBD\left(1,\dfrac{\theta}{\theta+1}\right)$ and $P_2(x;\theta)=\dfrac{(x+1)\theta^2}{(\theta+1)^{x+2}}$ as the $NBD\left(2,\dfrac{\theta}{\theta+1}\right)$ and $P_3(x;\theta)=\dfrac{(x+1)(x+2)\theta^3}{2(\theta+1)^{x+3}}$ as the $NBD\left(3,\dfrac{\theta}{\theta+1}\right)$,respectively.

Proof: We have

$$P(x;\theta)=\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\frac{\theta^3}{\theta^2+\theta+2}\left(\theta+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

$$=\int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\left\{\frac{\theta^2}{\theta^2+\theta+2}\left(\theta e^{-\theta\lambda}\right)+\frac{\theta}{\theta^2+\theta+2}\left(\frac{\theta^2}{\Gamma(2)}e^{-\theta\lambda}\lambda\right)+\frac{2}{\theta^2+\theta+2}\left(\frac{\theta^3}{\Gamma(3)}e^{-\theta\lambda}\lambda^2\right)\right\}d\lambda$$

$$=\frac{\theta^2}{\theta^2+\theta+2}\left[\frac{\theta}{x!}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^x d\lambda\right]+\frac{\theta}{\theta^2+\theta+2}\left[\frac{\theta^2}{x!\Gamma(2)}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^{x+1}d\lambda\right]$$

$$+\frac{2}{\theta^2+\theta+2}\left[\frac{\theta^2}{x!\Gamma(3)}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^{x+2}d\lambda\right]$$

$$=\frac{\theta^2}{\theta^2+\theta+2}\left[\frac{\theta}{x!}\frac{\Gamma(x+1)}{(\theta+1)^{x+1}}\right]+\frac{\theta}{\theta^2+\theta+2}\left[\frac{\theta^2}{x!\Gamma(2)}\frac{\Gamma(x+2)}{(\theta+1)^{x+2}}\right]+\frac{2}{\theta^2+\theta+2}\left[\frac{\theta^3}{x!\Gamma(3)}\frac{\Gamma(x+3)}{(\theta+1)^{x+3}}\right]$$

$$=\frac{\theta^2}{\theta^2+\theta+2}\left[\frac{\theta}{(\theta+1)^{x+1}}\right]+\frac{\theta}{\theta^2+\theta+2}\left[\frac{(x+1)\theta^2}{(\theta+1)^{x+2}}\right]+\frac{2}{\theta^2+\theta+2}\left[\frac{(x+1)(x+2)\theta^3}{2(\theta+1)^{x+3}}\right]$$

$$=\frac{\theta^2}{\theta^2+\theta+2}\left[\binom{x+1-1}{x}\left(\frac{\theta}{\theta+1}\right)^1\left(\frac{1}{\theta+1}\right)^x\right]+\frac{\theta}{\theta^2+\theta+2}\left[\binom{x+2-1}{x}\left(\frac{\theta}{\theta+1}\right)^2\left(\frac{1}{\theta+1}\right)^x\right]$$

$$+\frac{2}{\theta^2+\theta+2}\left[\binom{x+3-1}{x}\left(\frac{\theta}{\theta+1}\right)^3\left(\frac{1}{\theta+1}\right)^x\right]$$

$$=\frac{\theta^2}{\theta^2+\theta+2}\left[NBD\left(1,\frac{\theta}{\theta+1}\right)\right]+\frac{\theta}{\theta^2+\theta+2}\left[NBD\left(2,\frac{\theta}{\theta+1}\right)\right]+\frac{2}{\theta^2+\theta+2}\left[NBD\left(3,\frac{\theta}{\theta+1}\right)\right]$$

This completes the proof.

According to this result of PSD, although the PSD is a three-component mixture of negative binomial distribution, the existence of the three modes is not visible in any of the figure of the pmf of the PSD for the selected values of the parameter θ. This is the indication that the three modes which come from the three sub-populations must be located very close to each other. *Tajuddin et al. (2022)* showed that if the modes of the sub-populations are located very close to each other, the population will have single mode. Further, if the existence of the modes of the sub-populations can be certain, then the true distribution may be considered as one of the important distributions for over-dispersed count data from any field of knowledge.

2. Reliability properties

The reliability properties including reverse hazard rate function, second rate of failure, cumulative hazard function and Mills ratio of a distribution depends on cumulative distribution function, survival function and hazard function of the distribution. The following theorem 4 deals with the cumulative distribution function (cdf), survival function and the hazard function of PSD. The expression for reverse hazard rate function, second rate of failure, cumulative hazard function and Mills ratio of PSD have also been obtained.

Theorem 4: The cumulative distribution function (cdf), survival function and the hazard function of PSD are given by

$$F(x;\theta)=1-\frac{x^2\theta^2+x\left(\theta^3+6\theta^2+2\theta\right)+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}{\left(\theta^2+\theta+2\right)\left(\theta+1\right)^{x+3}}$$

$$S(x;\theta)=\frac{x^2\theta^2+x\left(\theta^3+6\theta^2+2\theta\right)+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}{\left(\theta^2+\theta+2\right)\left(\theta+1\right)^{x+3}}, \text{ and}$$

$$h(x;\theta)=\frac{\theta^3\left\{x^2+\left(\theta+4\right)x+\left(\theta^2+3\theta+4\right)\right\}}{x^2\theta^2+x\left(\theta^3+6\theta^2+2\theta\right)+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}$$

Proof: We have

$$F(x)=F(x,\theta)=P(X\le x)=1-P(X\ge x+1)$$

$$=1-\sum_{t=x+1}^{\infty}\int_0^{\infty}P(X=t\,|\,\lambda)f(\lambda)d\lambda$$

$$=1-\sum_{t=x+1}^{\infty}\int_0^{\infty}\frac{e^{-\lambda}\lambda^t}{t!}\frac{\theta^3}{\theta^2+\theta+2}\left(1+\lambda+\lambda^2\right)e^{-\theta\lambda}d\lambda$$

$$=1-\sum_{t=x+1}^{\infty}\frac{\theta^3}{\left(\theta^2+\theta+2\right)}\frac{\left\{t^2+(\theta+4)t+\left(\theta^2+3\theta+4\right)\right\}}{(\theta+1)^{t+3}}$$

$$=1-\frac{\theta^3}{\left(\theta^2+\theta+2\right)(\theta+1)^3}\sum_{t=x+1}^{\infty}\frac{t^2+(\theta+4)t+\left(\theta^3+3\theta+4\right)}{(\theta+1)^t}$$

$$=1-\frac{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}{\left(\theta^2+\theta+2\right)(\theta+1)^{x+3}}.$$

The survival function of PSD can be obtained as

$$S(x)=S(x,\theta)=1-F(x,\theta)=\frac{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}{\left(\theta^2+\theta+2\right)(\theta+1)^{x+3}}.$$

The hazard function of PSD can be expressed as

$$h(x)=h(x,\theta)=\frac{P(x,\theta)}{S(x,\theta)}=\frac{\theta^3\left\{x^2+(\theta+4)x+\left(\theta^2+3\theta+4\right)\right\}}{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}$$

The natures of cdf, survival function and hazard function of PSD for varying values of parameter are shown in the following figure 2 and it is obvious from the figure that the PGD has a valid cdf since $F(x)\to 1$ as $x\to\infty$. Further, the hazard rate function shows an increasing pattern with a limiting value of θ, which means that $\lim_{x\to\infty}h(x)=\theta$.

Figure 2

CDF of PSD for varying values of parameter

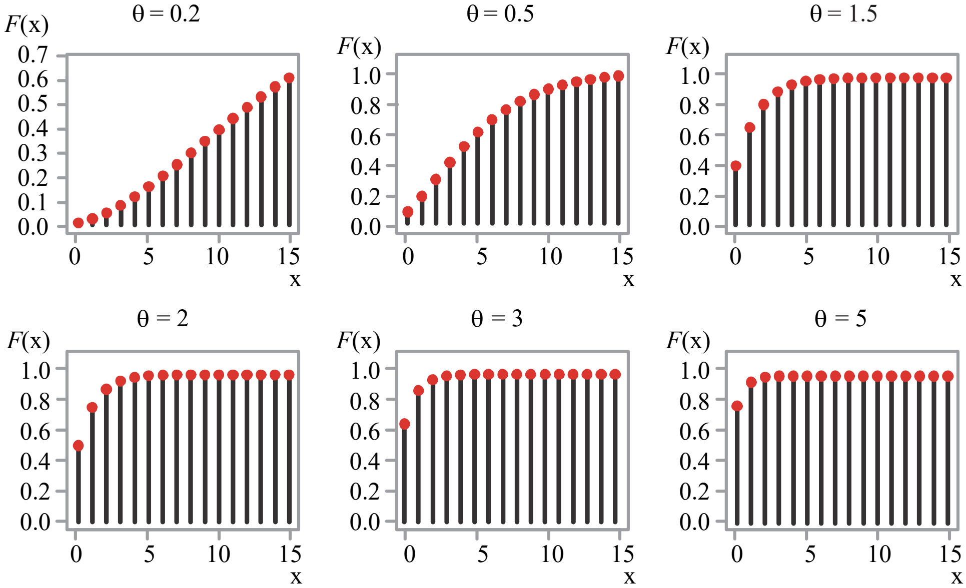


Figure 3

Survival function of PSD for varying values of parameter

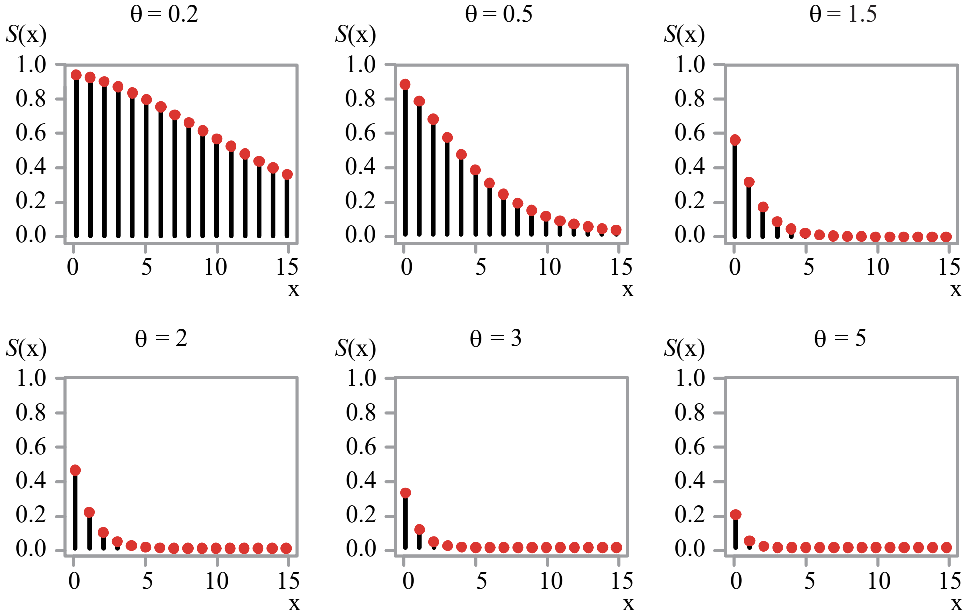
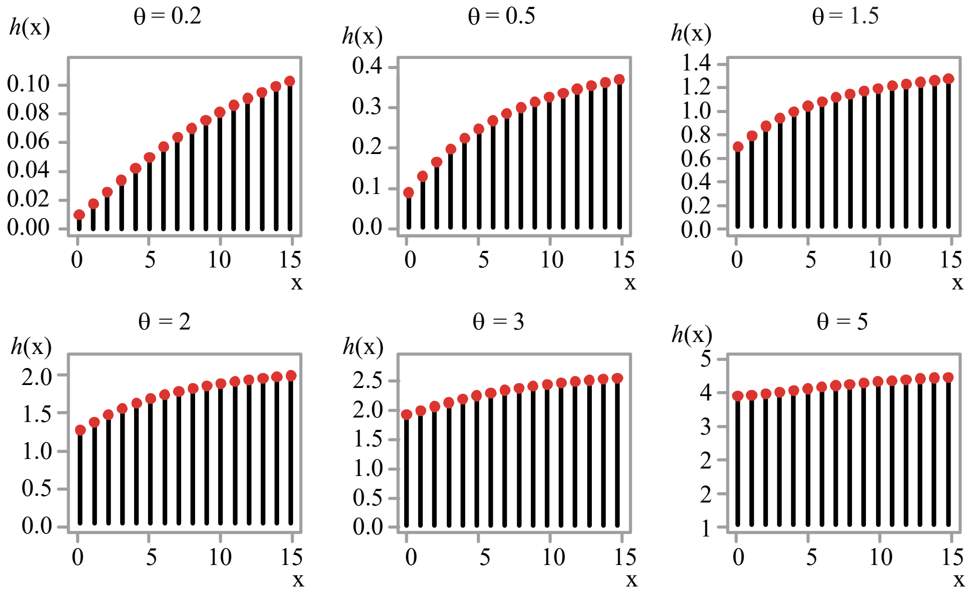


Figure 4

Hazard function of PSD for varying values of parameter

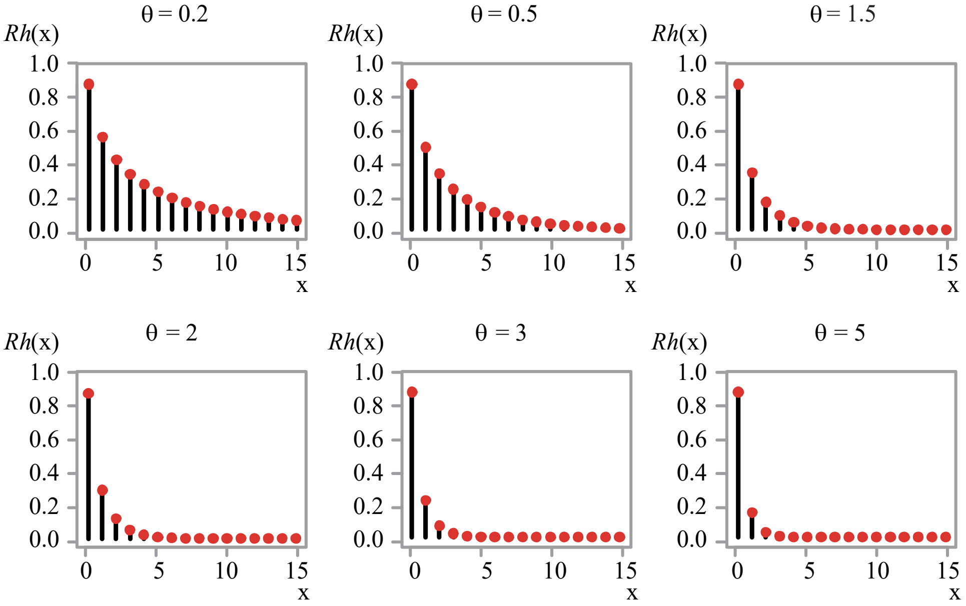


–The reverse hazard rate function $Rh(x;\theta)$ and the second rate of failure $SRF(x;\theta)$ of the PSD can be obtained as

$$Rh(x;\theta)=\frac{P(x;\theta)}{F(x;\theta)}=\frac{\theta^3\left\{x^2+(\theta+4)x+\left(\theta^2+3\theta+4\right)\right\}}{\left[\left(\theta^2+\theta+2\right)(\theta+1)^{x+3}-\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)\right]}$$

Figure 5

Reverse hazard function of PSD for varying values of parameter



$$SRF(x;\theta)=\ln\left[\frac{S(x;\theta)}{S(x+1;\theta)}\right]=\ln\left[\frac{(\theta+1)\left\{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)\right\}}{\theta^2(x+1)^2+\left(\theta^3+6\theta^2+2\theta\right)(x+1)+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}\right]$$

The cumulative hazard function $H(x;\theta)$ and Mills ratio $M(x;\theta)$ of PSD are given by

$$H(x;\theta)=-\ln S(x;\theta)=-\ln\left[\frac{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}{\left(\theta^2+\theta+2\right)(\theta+1)^{x+3}}\right],$$

and

$$M(x;\theta)=\frac{S(x;\theta)}{P(x;\theta)}=\frac{\theta^2x^2+\left(\theta^3+6\theta^2+2\theta\right)x+\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)\theta^3\left\{x^2+(\theta+4)x+\left(\theta^2+3\theta+4\right)\right\}}{\theta^3\left\{x^2+(\theta+4)x+\left(\theta^2+3\theta+4\right)\right\}}$$

3. Parameter estimation

3.1 Method of maximum likelihood estimation

Let $(x_1, x_2, \dots, x_n)$ be a random sample of size *n* from the PSD. Let f_x be the observed frequency in the sample corresponding to $X = x$ ($x = 1, 2, 3, ..., k$) such that $\sum_{x=1}^{k} f_x = n$, where k is the largest observed value having non-zero frequency. The likelihood function, L, of the PSD is given by

$$L = \left(\frac{\theta^3}{\theta^2 + \theta + 2} \right)^n \frac{1}{(\theta+1)^{\sum_{x=1}^{k} f_x (x+3)}} \prod_{x=1}^{k} \left[x^2 + (\theta+4)x + (\theta^2 + 3\theta + 4) \right]^{f_x}$$

The log-likelihood function is given by

$$\log L = n \log\left(\frac{\theta^3}{\theta^2 + \theta + 2} \right) - \sum_{x=1}^{k} f_x (x+3) \log(\theta+1) + \sum_{x=1}^{k} f_x \log\left[x^2 + (\theta+4)x + (\theta^2 + 3\theta + 4) \right]$$

The first derivative of the log likelihood function is given by

$$\frac{d \log L}{d\theta} = \frac{n(\theta^2 + 2\theta + 6)}{\theta(\theta^2 + \theta + 2)} - \frac{n(\bar{x}+3)}{\theta+1} + \sum_{x=1}^{k} \frac{\left[x + (2\theta+3) \right] f_x}{\left[x^2 + (\theta+4)x + (\theta^2 + 3\theta + 4) \right]}$$

The maximum likelihood estimate, $\hat{\theta}$ of θ is the solution of the equation $\frac{\partial \log L}{\partial \theta} = 0$ and is given by solution of the following non-linear equation

$$\frac{n(\theta^2 + 2\theta + 6)}{\theta(\theta^2 + \theta + 2)} - \frac{n(\bar{x}+3)}{\theta+1} + \sum_{x=1}^{k} \frac{\left[x + (2\theta+3) \right] f_x}{\left[x^2 + (\theta+4)x + (\theta^2 + 3\theta + 4) \right]} = 0$$

This non-linear equation can be solved by any numerical iteration methods such as Newton- Raphson, Bisection method or Regula falsi method etc.

4. Simulation study

To see the performance of the maximum likelihood estimator (MLE) for the PSD, we conducted an extensive simulation study. Using the inverse transform sampling method we generated random samples from the distribution. For each chosen sample size (50, 100, 200, 300, 400, and 500), we repeated the sampling process

10,000 times to ensure that the results were stable and reliable. For every set of simulation, we computed the bias and mean squared error (MSE) of the parameter estimate to evaluate the accuracy and consistency of the estimator. The findings, summarized in table 1 and table 2, clearly indicate that both bias and MSE decrease as the sample size increases. We tested performance of the estimator under two different true parameter values (2 and 2.5), and in both cases, the MLE showed satisfactory behavior. The formula for bias and MSE is

$$bias(\hat{\theta}) = E\left[\hat{\theta}\right] - \theta = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i - \theta \qquad MSE\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta\right)^2$$

Table1

Simulation result of PSD for $\theta = 2$

N	Mean	BIAS	MSE	Variance
50	2.0478	0.0479	0.0950	0.0927
100	2.0266	0.0266	0.0436	0.0429
200	2.0117	0.0117	0.0207	0.0206
300	2.0070	0.0070	0.0138	0.0137
400	2.0068	0.0068	0.1025	0.0102
500	2.0051	0.0051	0.0085	0.0085

Table2

Simulation result of PSD for $\theta = 2.5$

N	Mean	BIAS	MSE	Variance
50	2.5798	0.0799	0.2060	0.1996
100	2.5456	0.0456	0.0900	0.0879
200	2.5194	0.0194	0.0413	0.0409
300	2.5109	0.0109	0.0275	0.0270
400	2.5117	0.0117	0.0207	0.0205
500	2.5081	0.0081	0.0171	0.0170

5. Applications

In this section, the goodness of fit of PSD has been presented with three example of observed count data set and the goodness of fit of the Poisson distribution (PD), PLD, PGD, PKD and PAD for comparison. The first data set related to the number of corn borers available in *McGuire et al. (1957)*. The second and third data set in table 4 and 5 are related to number of days that experienced the number of thunderstorms events at Cape Kennedy, Florida for the month of June and July, 11-year period of record January 1957 to December 1967 available in *Falls et al.*

(1971).The mean and the variance of dataset in Table 3,4 and 5 are (0.6481, 0.8479), (0.7515,1.1721) and (0.8739, 1.2811) respectively and it is quite obvious that the data sets are over-dispersed. The Goodness of fit of the considered distributions are presented in Table 3,4 and 5 respectively. The goodness of fit measures in Table 3, 4 and 5 shows that PSD provides much better fit over PLD, PGD, PKD and PAD thus PSD can be considered as one of the important distributions for discrete data where events are dependent. The violine plot of the three data set in Table 3,4 and 5 are presented in Figure 6.

Table 3

Observed and expected number of European corn-borer available in *McGuire et al. (1957)*

Number of corn-borer per plant	Observed frequency	PD	PLD	PGD	PKD	PAD	PSD
0	188	169.5	194.0	195.0	195.6	196.3	193.7
1	83	109.8	79.5	78.7	78.0	76.5	79.6
2	36	35.6	31.3	31.0	30.8	30.8	31.6
3	14	7.7	12.0	12.0	12.0	12.4	12.1
4	2	1.2	4.5	4.6	4.7	4.9	4.5
5	1	0.2	2.7	2.7	2.9	3.1	2.5
Total	324	324	324	324	324	324	324
$\hat{\theta}(SE)$		0.6481 (0.0447)	2.04324 (0.1513)	1.9325 (0.15301)	1.79375 (0.13197)	2.345109 (0.1372)	2.4717 (0.1537)
–2logL		724.4902	714.0992	714.6243	715.195	715.6964	713.837
χ^2		15.423	1.2975	1.5668	1.8386	2.3478	1.1565
d.f		2	2	2	2	2	2
P value		0.0015	0.7297	0.6669	0.6066	0.5034	0.7635

Table 4

Number of thunderstorm events in the month of June

x	Observed frequency	PD	PLD	PGD	PKD	PAD	PSD
0	187	155.6	185.3	186.5	187.2	187.9	184.8
1	77	117.0	83.6	82.4	81.7	80.3	83.6
2	40	44.0	36.0	35.5	35.2	35.3	36.3
3	17	11.0	15.0	15.0	15.0	15.4	15.2
4	6	2.1	6.1	6.3	6.3	6.6	6.1
5	2	0.3	2.5	2.6	2.7	2.7	2.4
6	1	0.0	1.6	1.7	1.9	1.8	1.6
Total	330	330	330	330	330	330	330
$\hat{\theta}(SE)$		0.7515 (0.0477)	1.8043 (0.1257)	1.6906 (0.1268)	1.5886 (0.1088)	2.1338 (0.1159)	2.2299 (0.1295)
–2logL		824.506	788.8836	789.0402	789.2348	788.8993	788.6261
χ^2		32.233	1.3893	1.4338	1.4609	1.3292	1.2573
d.f		2	3	3	3	3	3
P value		0.0000	0.8460	0.8383	0.8335	0.8564	0.8686

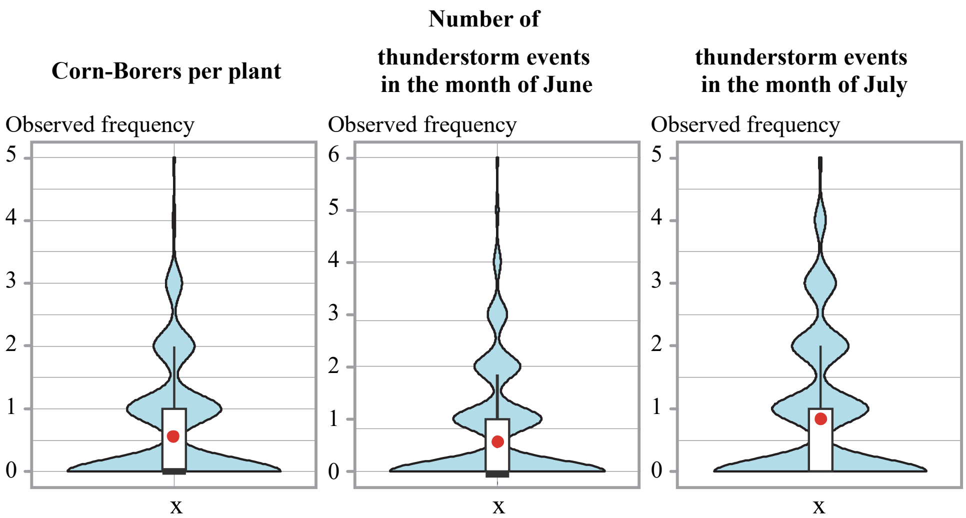
Table 5

Number of thunderstorm events in the month of July

x	Observed frequency	PD	PLD	PGD	PKD	PAD	PSD
0	177	142.3	177.7	179.4	180.0	180.0	176.5
1	80	124.4	88.0	86.7	86.0	84.8	88.4
2	47	54.3	41.5	40.7	40.4	40.9	42.2
3	26	15.8	18.9	18.8	18.8	19.4	19.2
4	9	3.5	8.4	8.5	8.6	8.9	8.5
5	2	0.6	6.5	6.9	7.2	7.0	6.2
Total	341	341	341	341	341	341	341
$\hat{\theta}(SE)$		0.8739 (0.506)	1.5835 (0.1032)	1.4726 (0.1040)	1.4011 (0.0890)	1.9274 (0.0969)	1.9958 (0.1078)
–2logL		911.0165	880.5501	881.2906	881.7636	880.7007	879.6258
χ^2		39.972	5.1469	5.3907	5.7625	4.9869	4.6852
d.f		2	3	3	3	3	3
P value		0.0000	0.2725	0.2495	0.2176	0.2886	0.3211

Figure 6

Violin plot for the three data sets



6. Conclusion

In this paper some interesting and useful statistical and reliability properties of the Poisson-Sujatha distribution, the Poisson compound of Sujatha distribution, has been derived and discussed. The proposed distribution has been observed to be log-concave and is a two-component mixture of negative binomial distributions. The hazard function of the distribution is increasing and thus the distribution is unimodal. The consistency of maximum likelihood estimator of the parameter of Poisson-Garima distribution has been shown using simulation study. The applications and the goodness of fit of the Poisson-Sujatha distribution has been established with three real datasets and found to compete well with other one parameter over-dispersed count distributions namely Poisson-Lindley distribution (PLD), Poisson-Garima distribution (PGD), Poisson-Komal distribution (PKD), Poisson-Akash distribution (PAD).

References

Aderoju, S. A. – Adeniyi, I. (2022): On zero inflated generalized Poisson-Sujatha distribution and its applications. *StatisticaApplicata – Italian Journal of Applied Statistics*, 34(1), 7–22. https://doi.org/10.26398/IJAS.0034-001

Da Silva,W. B. – Ribeiro, A. M. T. – Conceição, K. S. – Andrade, M. G. – Neto F. L. (2018):On Zero-Modified Poisson-Sujatha Distribution to Model Over dispersed Count Data. *Austrian Journal of Statistics*, 47(3), 1–19. https://doi.org/10.17713/ajs.v47i3.590

Falls, L.W. – Wilford, W. O. – Carter, M. C. (1971): Probability distribution for thunderstorm activity at Cape Kennedy, Florida. *Journal of AppliedMeteorology*,10(1), 97–104. https://doi.org/10.1175/1520-0450(1971)010<0097:PDFTAA>2.0.CO;2

Grandell, J. (1997): *Mixed Poisson Processes*. Chapman & Hall, London. https://doi.org/10.1201/9781003059950

Lindley, D. V. (1958): Fiducial distributions and Bayes theorem. *Journal of Royal Statistical Society*, 20(1), 102–107. https://doi.org/10.1111/j.2517-6161.1958.tb00278.x

McGuire, J. U. – Brindley, T. A. – Bancroft, T. A. (1957): The distribution of European corn-borer larvae pyrausta in field corn. *Biometrics*, 13, 65–78. https://doi.org/10.2307/3001903

Mohiuddin, M. – Kannan, R. (2022): Characterization and Estimation of Alpha Power Sujatha Distribution with Applications to Engineering Data. *International Journal of Research in Engineering and Technology,*9(3), 1421–1432.

Prodhani, H. R. – Shanker, R. (2024): A three-parameter power Sujatha distribution with properties and application. *International Journal of Statistics and Reliability Engineering*, 11(1), 70–78.

Sankaran, M. (1970): The discrete Poisson-Lindley distribution. *Biometrics*, 26(1), 145–149. https://doi.org/10.2307/2529053

Shanker, R. (2016a): Sujatha distribution and its application. *Statistics in Transition New Series*, 17(3), 391–410. https://doi.org/10.21307/stattrans-2016-029

Shanker, R. (2016b): The discrete Poisson-Sujatha distribution. *International Journal of Probability and Statistics*, 5(1), 1–9. https://doi.org/10.5923/j.ijps.20160501.01

Shanker, R. (2016c): The discrete Poisson-Sujatha distribution. *International journal of probability and Statistics,*5(1),2168–4871. https://doi:10.5923/j.ijps.20160501.01

Shanker, R. (2017a): The discrete Poisson-Akash distribution. *International journal of probability and statistics*, 6(1), 1–10. https://doi.org/10.5923/j.ijps.20170601.01

Shanker, R. (2017b): The discrete Poisson-Garima distribution. *Biometrics & Biostatistics International Journal*, 5(2), 48–53. https://doi.org/10.15406/bbij.2017.05.00127

Shanker, R. – Hagos, F. (2015a): Zero-truncated Poisson-Sujatha Distribution with applications. *American Journal of Mathematics and Statistics,*24, 55–62.

Shanker, R. – Hagos, F. (2015b): On Poisson-Lindley distribution and its Applications to Biological Sciences.*Biometrics & Biostatistics International Journal,* 2(4), 103–107. https://doi.org/10.15406/bbij.2015.02.00036

Shanker, R. – Hagos, F. (2016): Size-biased Poisson-Sujatha Distribution with applications. *American Journal of Mathematics and Statistics,* 6(4), 145–154. https://doi.org/10.5923/j.ajms.20160604.02

Shanker, R. – Prodhani, H. R. (2024): Size-Biased Sujatha distribution with properties and application to model flood data. *Research and Review: Journal of Statistics,*13(1), 31–47.

Shanker, R. – Tabassum, R. – Sarma, I. – Saikia, P. (2025): The Poisson compound of Komal distribution with statistical properties and applications. *Biomatrices & Biostatistics International Journal*, 14(1), 35–43. https://doi.org/10.15406/bbij.2025.14.00432

Subramanian, C. – Shenbagaraja, R. (2020): Length Biased Quasi-Sujatha Distribution with Properties and Applications to Bladder Cancer Data. *Journal of Information and Computational Science*, 10(1), 493–506.

Tajuddin, R. R. M. –Ismail, N. – Ibrahim, K. (2022): Several two-component mixture distributions for count data.*Communication in Statistics-Simulation and Computation*, 51, 3760–3771. https://doi.org/10.1080/03610918.2020.1722834