Research Article
Open Access
CrossMark

Statistical properties, entropy measures and applications of poisson-shanker distribution in automobile insurance claim frequency data

Abstract

The Poisson-Shanker distribution was introduced by Shanker to model over-dispersed count data from biological sciences. The literature review reveals that some useful statistical properties, reliability properties and entropy measures of Poisson-Shanker distribution have not been studied. The main purpose of this paper is to study statistical properties, reliability properties, entropy measures which were not studied earlier. The Poisson-Shanker distribution has been shown to be log-concave and the two-component mixture of negative binomial distributions. The cumulative distribution function, survival function, hazard function and mean residual life function of the Poisson-Shanker distribution have been derived and their natures for varying values of parameter have been discussed. The expressions for other reliability properties including reverse hazard function, the second rate of failure, the cumulative hazard function and the Mills ratio of the Poisson-Shanker distribution have also been derived. Various entropy measures including the Shannon, Rényi and Tsallis entropy measures are obtained in a form suitable for numerical evaluation to quantify the degree of uncertainty, randomness or information content associated with the distribution. The simulation study has been presented to show the consistency of maximum likelihood estimator of the parameter of the distribution. The applications of the Poisson-Shanker distribution in automobile insurance claim frequency data have been discussed and the goodness of fit has been compared with the goodness of fit given by Poisson distribution, Poisson-Lindley distribution and Poisson-Garima distribution.

Keywords: Poisson-Shanker distribution, reliability properties, entropy measures, simulation, automobile insurance claim frequency data

Volume 15 Issue 1 - 2026

Rama Shanker,[1] Riki Tabassum,[1] Ipshita Sarma,[1] Sujatha Selvaraj[2]

[1]Department of Statistics, Assam University, India
[2]Department of Quantitative Economics, University West Canada, Canada

Correspondence: Ipshita Sarma, Department of Statistics, Assam University, Silchar, Assam, India

Received: May 27, 2026 | **Published:** June 10, 2026

Introduction

The Poisson distribution (PD), being the limiting distribution of binomial distribution, is the first discrete distribution for count data of equi-dispersed (mean equal to variance) nature. Count data arises in insurance, agriculture, environmental science, biological sciences, clinical trials, engineering etc. The count data, in general, are either over-dispersed (variance greater than mean) or under-dispersed (variance less than mean) and in these situations the PD fails to capture the dispersion in the dataset. During recent decades several count distributions have been introduced in Statistics literature including negative binomial distribution which is the Poisson compound of gamma distribution, Poisson-Lindley distribution (PLD) of Sankaran[1] which is the Poisson compound of Lindley distribution (LD) of Lindley, Poisson-Shanker distribution (PSD) of Shanker[2] which is the Poisson compound of Shanker distribution (SD) of Shanker,[3] Poisson-Garima distribution (PGD) of Shanker[4] which is the Poisson compound of Garima distribution (GD) of Shanker.[5] Shanker et al.,[6] have detailed study on statistical and reliability properties along with some new applications of PGD.

The SD introduced by Shanker[3] is defined by its probability density function (pdf)

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

Shanker and Shukla[4,7] proposed weighted Shanker distribution and the power Shanker distribution to model real lifetime data from biological sciences and engineering. Helal et al.,[8] proposed another weighted Shanker distribution by taking the weight function the pdf of exponential distribution and discussed some of its properties. Ray and Shanker[9] derived exponential-Shanker distribution (E-SD) by compounding exponential distribution with the SD. Ray and Shanker[10] also proposed gamma-Shanker distribution (G-SD) by compounding gamma distribution with the SD which includes E-SD as a particular case.

Shanker[2] derived PSD by compounding PD with SD. The genesis of PSD is that a random variable X is said to be PSD if it follows the following stochastic representation $X\mid\lambda\sim\text{Poisson}(\lambda)$ distribution and $\lambda\mid\beta\sim\text{Shanker}(\beta)$ distribution for $\lambda>0,\beta>0$ and the unconditional distribution of this stochastic representation is the PSD (β). That is, the probability mass function (pmf) of PSD obtained by Shanker[2] is

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

$$=\int_0^\infty\frac{e^{-\lambda}\lambda^x}{x!}\frac{\beta^2}{\beta^2+1}(\beta+\lambda)e^{-\beta\lambda}d\lambda$$

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

The mean of the PSD is given by $\mu_1' = \dfrac{\beta^2+2}{\beta(\beta^2+1)}$.

The detailed discussion about its properties, estimation of parameter, and applications has been discussed by Shanker.[2] Shanker et al.,[11] have detailed study on the applications of PSD in different fields of knowledge. Shanker[12,13] derived size-biased Poisson –Shanker distribution (SBPSD) and the zero-truncated Poisson-Shanker distribution (ZTPSD) for modeling count data which structurally excludes zero counts. Wararit[14] discussed the Bootstrap methods for estimating the confidence interval for the index of dispersion of ZTPSD. The literature review on PSD reveals that some useful statistical properties, reliability properties, the entropy measures, simulation study and applications of PSD for modeling automobile insurance claim frequency data have not been discussed.

The main purpose of this paper is to derive some interesting statistical properties, reliability properties, entropy measures and applications of PSD to model automobile insurance claim frequency data, which has not been studied earlier. It has been shown that PSD is log-concave and hence unimodal and is also a two-component mixture of negative binomial distributions. Some entropy measures including the Shannon, Rényi and Tsallis entropy measures are obtained in a form suitable for numerical evaluation to quantify the degree of uncertainty, randomness or information content for PSD have been derived. The natures of cumulative distribution function, survival function, hazard function, reverse hazard function and mean residual life function of PSD has been presented graphically. The consistency of the maximum likelihood estimator of PSD has been studied using simulation. The goodness of fit of the PSD has been established with automobile insurance claim frequency data and observed that it provides best fit over other competing one parameter over-dispersed distributions namely PLD and PGD.

Statistical properties

In this section two important results of PSD, namely the log-concavity and unimodality and the two-component mixture of negative binomial distributions, have been discussed in terms of theorems

Theorem 1: The PSD is log-concave and unimodal.

Proof: From the results of Gupta et al.,[15] Steutel[16] and Bagnoli and Bergstrom,[17] any discrete distribution with pmf $P(x;\beta)$ is log-concave if

$$\left[P(x+1;\beta)\right]^2 > P(x;\beta)\cdot P(x+2;\beta), \forall x\in\mathbb{N}.$$

Now, for $\beta>0$, it can be easily shown that the pmf of PSD satisfy the above inequality. That is,

$$\left[P(x+1;\beta)\right]^2 - P(x;\beta)\cdot P(x+2;\beta) > 0, \forall x\in\mathbb{N}$$

.

Again, using the result of Keilson and Gerber[18] which states that any log-concave pmf is always strongly unimodal, the PSD is unimodal.

Theorem 2: The PSD is the two-component mixture of negative binomial distributions and can be expressed as

$$P(x;\beta) = w_1 P_1(x;\beta) + w_2 P_2(x;\beta),$$

where $P_i(x;\beta)$ is the pmf of the negative binomial distribution (NBD) with parameters, the number of successes i and $p_1 = \dfrac{\beta^2}{\beta^2+1}$ with $P_1(x;\beta) = \dfrac{\beta}{(\beta+1)^{x+1}}$ as $NBD\left(1, \dfrac{\beta}{\beta+1}\right)$,

$p_2 = \dfrac{1}{\beta^2+1}$ with $P_2(x;\beta) = \dfrac{(x+1)\beta^2}{(\beta+1)^{x+2}}$ as the $NBD\left(2, \dfrac{\beta}{\beta+1}\right)$.respectively.

Proof: We have

$$P(x;\beta) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} \frac{\beta^2}{\beta^2+1}(\beta+x)e^{-\beta x} d\lambda$$

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

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

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

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

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

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

Since PSD is the two-component mixture of NBD and the NBD is a suitable two- parameter probability model for over-dispersed count data, PSD would be suitable for over-dispersed count data.

Reliability properties

In this section, several important reliability characteristics of the PSD are discussed.

Cumulative distribution function (CDF) and Survival function

The CDF $F(x)$ of the PSD can be obtained as

$$F(x) = P(X \le x) = \sum_{k=0}^{x} p(k) = \frac{\beta^2}{\beta^2+1}\sum_{k=0}^{x}\frac{k+\beta^2+\beta+1}{(\beta+1)^{k+2}} = 1 - \frac{\beta^3+\beta^2+\beta x+2\beta+1}{(\beta^2+1)(\beta+1)^{x+2}}$$

The survival function $S(x)$ of the PSD is given by

$$S(x) = P(X > x) = 1 - F(x) = \frac{\beta^3+\beta^2+\beta x+2\beta+1}{(\beta^2+1)(\beta+1)^{x+2}}$$

The graph of the $F(x)$ and $S(x)$ of the PSD for different values of the parameter β are shown in Figures 1& 2, respectively.

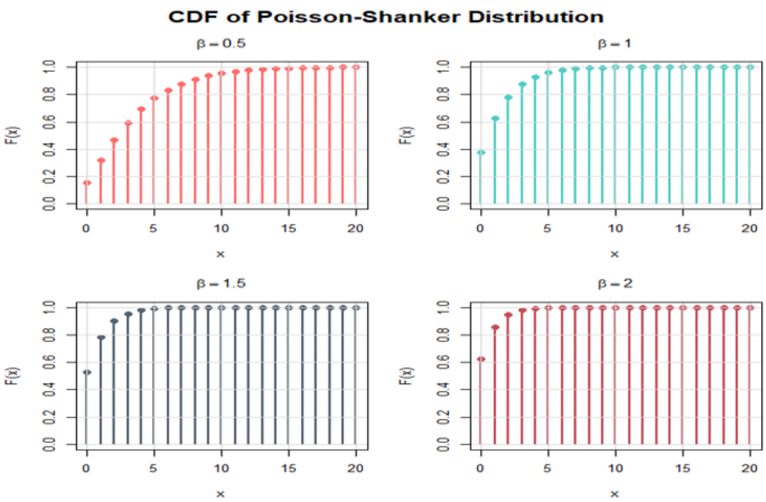

Figure 1 The $F(x)$ of PSD.

Citation: Shanker R, Tabassum R, Sarma I, et al. Statistical properties, entropy measures and applications of poisson-shanker distribution in automobile insurance claim frequency data. *Biom Biostat Int J.* 2026;15(1):51–60. DOI: 10.15406/bbij.2026.15.00448

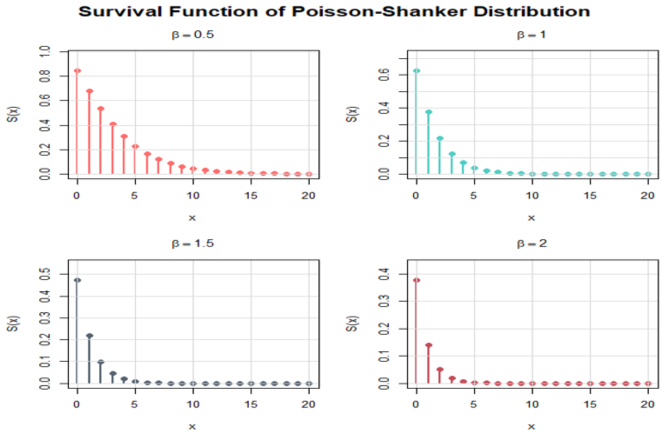

Figure 2 The $S(x)$ of PSD.

Hazard function

Using the result, $S(x-1)=\dfrac{\beta^3+\beta^2+\beta x+\beta+1}{(\beta^2+1)(\beta+1)^{x+1}}$, the hazard function, $h(x)$ of the PSD is obtained as

$$h(x)=\frac{P(x)}{S(x-1)}=\frac{\beta^2\left(x+\beta^2+\beta+1\right)}{(\beta+1)\left(\beta^3+\beta^2+\beta x+\beta+1\right)}.$$

The natures of $h(x)$ the PSD for different values of the parameter β are shown in Figure 3.

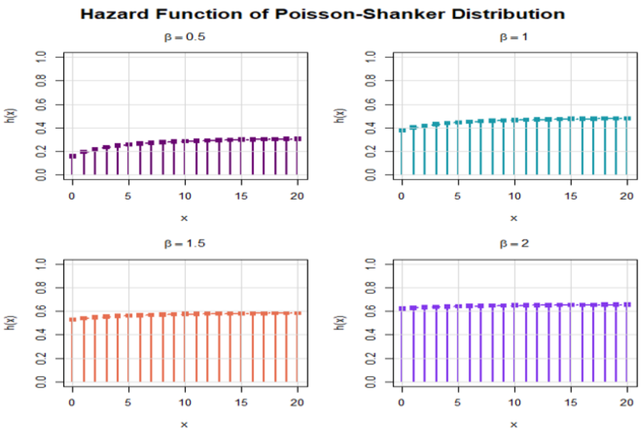

Figure 3 The $h(x)$ of PSD.

Mills ratio

The Mills ratio $M(x)$ of the PSD can be obtained as

$$M(x)=\frac{P(x)}{S(x)}=\frac{\beta^2\left(x+\beta^2+\beta+1\right)}{\beta^3+\beta^2+\beta x+2\beta+1}$$

The natures of $M(x)$ the PSD for different values of the parameter β are shown in Figure 4.

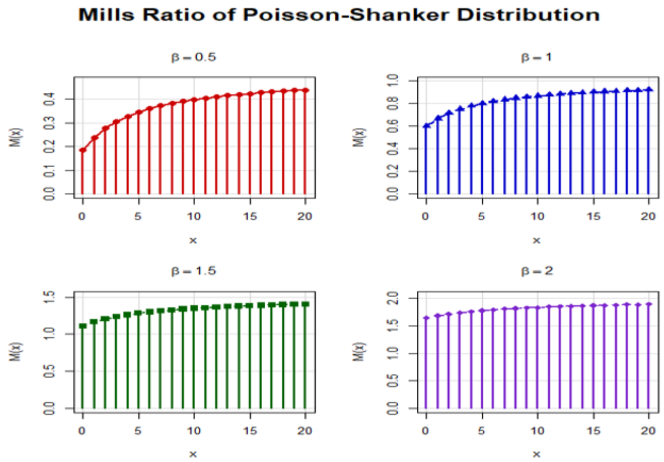

Figure 4 $M(x)$ of PSD.

Mean residual life function

The mean residual life function (MRL) $m(x)$ is defined as

$$m(x)=E\left(X-x \mid X>x\right)$$

For a discrete random variable, $m(x)$ is defined as

$$m(x)=\frac{1}{S(x)}\sum_{k=x}^{\infty}S(k)$$

Substituting the survival function and simplifying, the $m(x)$ of the PSD can be expressed as

$$m(x)=\frac{(\beta+1)\left(\beta^3+\beta^2+\beta x+2\beta+2\right)}{\beta\left(\beta^3+\beta^2+\beta x+2\beta+1\right)}.$$

The natures of $m(x)$ of the PSD for different values of the parameter β are shown in Figure 5.

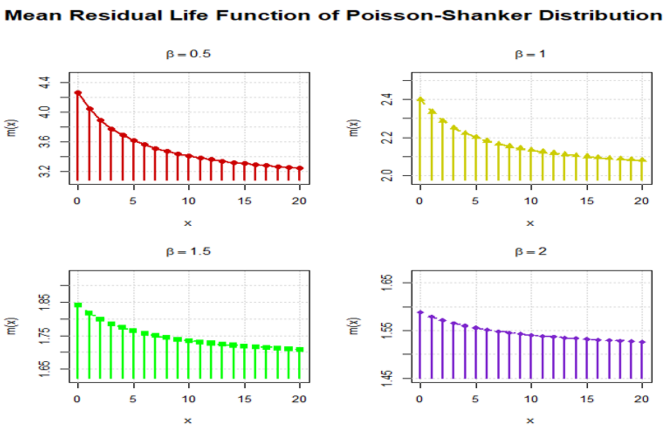

Figure 5 The $m(x)$ of PSD.

Entropy measures

Entropy measures are useful in information theory and statistical distribution theory because they quantify the degree of uncertainty, randomness or information content associated with a probability model. In this section, the Shannon, Rényi and Tsallis entropy measures of the PSD are obtained in a form suitable for numerical evaluation.

Let X be a Poisson-Shanker random variable with probability mass function

$$p(x;\beta)=\frac{\beta^2}{\beta^2+1}\frac{x+\beta^2+\beta+1}{(\beta+1)^{x+2}},\quad x=0,1,2,...\quad \beta>0$$

Since X is a discrete random variable, all entropy measures are expressed in terms of summations over $x=0,1,2,\ldots$.

Shannon entropy

The Shannon entropy introduced by Shannon[19] is one of the most commonly used measures of uncertainty associated with a probability distribution. For a discrete random variable X with pmf $p(x)$, it is defined as

$$H(X)=-\sum_{x=0}^{\infty}p(x)\log p(x),$$

where $\log$ denotes the natural logarithm. Substituting the pmf of the PSD, we get

$$H(X)=-\sum_{x=0}^{\infty}\left[\frac{\beta^2}{\beta^2+1}\frac{x+\beta^2+\beta+1}{(\beta+1)^{x+2}}\right]\log\left[\frac{\beta^2}{\beta^2+1}\frac{x+\beta^2+\beta+1}{(\beta+1)^{x+2}}\right].$$

Using the logarithmic identity $\log(ab/c)=\log a+\log b-\log c$, the logarithmic part can be written as

$$\log p(x)=\log\left(\frac{\beta^2}{\beta^2+1}\right)+\log\left(x+\beta^2+\beta+1\right)-(x+2)\log(\beta+1).$$

Therefore,

$$H(X)=-\sum_{x=0}^{\infty}p(x)\left[\log\left(\frac{\beta^2}{\beta^2+1}\right)+\log\left(x+\beta^2+\beta+1\right)-(x+2)\log(\beta+1)\right]$$

Separating the summation terms gives

$$H(X)=-\log\left(\frac{\beta^2}{\beta^2+1}\right)\sum_{x=0}^{\infty}p(x)-\sum_{x=0}^{\infty}p(x)\log\left(x+\beta^2+\beta+1\right)+\log(\beta+1)\sum_{x=0}^{\infty}(x+2)p(x)$$

Since $\sum_{x=0}^{\infty}p(x)=1$ and $\sum_{x=0}^{\infty}(x+2)p(x)=E(X)+2$, the Shannon entropy of the PSD is obtained as

$$H(X)=-\log\left(\frac{\beta^2}{\beta^2+1}\right)-\sum_{x=0}^{\infty}p(x)\log\left(x+\beta^2+\beta+1\right)+\left(E(X)+2\right)\log(\beta+1).$$

Thus, the Shannon entropy is expressed in a computable infinite-series form. The presence of the term $\log\left(x+\beta^2+\beta+1\right)$ prevents further elementary simplification in closed form.

Rényi entropy

The Rényi entropy introduced by Rényi[20] generalizes Shannon entropy through an order parameter. For $\delta > 0$ and $\delta \neq 1$, the Rényi entropy of a discrete random variable X is defined as

$$H_\delta(X) = \frac{1}{1-\delta} log\left(\sum_{x=0}^{\infty} p(x)^\delta\right)$$

For the PSD, we have

$$p(x)^\delta = \left[\frac{\beta^2}{\beta^2+1}\frac{x+\beta^2+\beta+1}{(\beta+1)^{x+2}}\right]^\delta$$

This gives

$$\sum_{x=0}^{\infty} p(x)^\delta = \sum_{x=0}^{\infty}\left[\frac{\beta^2}{\beta^2+1}\frac{x+\beta^2+\beta+1}{(\beta+1)^{x+2}}\right]^\delta$$

Taking the terms that are independent of x outside the summation, we get

$$\sum_{x=0}^{\infty} p(x)^\delta = \left(\frac{\beta^2}{\beta^2+1}\right)^\delta \frac{1}{(\beta+1)^{2\delta}}\sum_{x=0}^{\infty}\frac{\left(x+\beta^2+\beta+1\right)^\delta}{(\beta+1)^{x\delta}}.$$

Hence, the Rényi entropy of the PSD is given by

$$H_\delta(X) = \frac{1}{1-\delta} log\left[\left(\frac{\beta^2}{\beta^2+1}\right)^\delta \frac{1}{(\beta+1)^{2\delta}}\sum_{x=0}^{\infty}\frac{\left(x+\beta^2+\beta+1\right)^\delta}{(\beta+1)^{x\delta}}\right].$$

This expression is valid for $\delta > 0$, $\delta \neq 1$ and $\beta > 0$.

Tsallis entropy

The Tsallis entropy proposed by Tsallis[21] is another generalized entropy measure frequently used in statistical mechanics and information theory. For $\delta > 0$ and $\delta \neq 1$, it is defined as

$$T_\delta(X) = \frac{1}{\delta-1}\left[1-\sum_{x=0}^{\infty} p(x)^\delta\right]$$

Using the expression already obtained for $\sum_{x=0}^{\infty} p(x)^\delta$, we have

$$\sum_{x=0}^{\infty} p(x)^\delta = \left(\frac{\beta^2}{\beta^2+1}\right)^\delta \frac{1}{(\beta+1)^{2\delta}}\sum_{x=0}^{\infty}\frac{\left(x+\beta^2+\beta+1\right)^\delta}{(\beta+1)^{x\delta}}.$$

Therefore, the Tsallis entropy of the PSD is obtained as

$$T_{\delta}(X)=\frac{1}{\delta-1}\left[1-\left(\frac{\beta^2}{\beta^2+1}\right)^{\delta}\frac{1}{(\beta+1)^{2\delta}}\sum_{x=0}^{\infty}\frac{\left(x+\beta^2+\beta+1\right)^{\delta}}{(\beta+1)^{x\delta}}\right].$$

Thus, the Shannon, Rényi and Tsallis entropies of the PSD are obtained in tractable series forms. These expressions can be evaluated numerically for different values of the parameter β and, in the cases of Rényi and Tsallis entropies, for different values of the entropy order δ

Parameter estimation

Method of moment estimation (MOME): Equating the population mean to the corresponding sample mean $\bar{x}$, the method of moment estimator is the solution of the following cubic equation in β

$$\bar{x}\,\beta^3-\beta^2+\beta\,\bar{x}-2=0.$$

This cubic equation in β can be solved using Newton-Raphson method for the method of moment estimate of the parameter β.

Method of maximum likelihood estimation: Assuming $(x_1,x_2,x_3,....x_n)$ from PSD(β) and f_x 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 log-likelihood equation for estimating the parameter β of PSD is given by

$$\frac{2n}{\beta}-\frac{2n\beta}{\beta^2+1}-\sum_{x=1}^{k}\frac{(x+2)f_x}{\beta+1}+\sum_{x=1}^{k}\frac{(2\beta+1)f_x}{x+\beta^2+\beta+1}=0$$

.

This non-linear log-likelihood equation can be solved using Newton-Raphson method for the maximum likelihood estimate of the parameter β.

Simulation

To study the performance of the maximum likelihood estimator (MLE) for the parameter of the PSD, a simulation study was carried out. Random samples were generated using the inverse transformation method. The study was conducted for fixed parameter values $\beta=5$ and $\beta=12$ with different sample sizes (n = 50, 100, 200,300). For each case, 5,000 random samples were generated, and the MLEs were obtained by minimizing the negative log-likelihood function using the L-BFGS-B optimization method.

For every sample size, the mean estimate, variance, bias, and mean squared error (MSE) were calculated. The results show that the MLE performs well for the PSD. As the sample size increases, the bias, variance, and MSE decrease gradually, showing that the estimator becomes more accurate and stable. Also, the MSE values are very close to the variances, indicating that the estimator is nearly unbiased. Hence, the simulation results support the consistency and efficiency of the MLE. The formulas for bias and mean square error (MSE) are as follows (Tables 1&2):

$$Bias(\hat{\beta})=E(\hat{\beta})-\beta=\frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i-\beta$$

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

Table 1 Simulation table of PSD for $\beta=5$

n	$E(\hat{\beta})$	$Var(\hat{\beta})$	BIAS	MSE
50	5.5339	3.3883	0.5339	3.6727
100	5.2964	1.7030	0.2964	1.7906
200	5.1681	0.7702	0.1681	0.7983
300	5.1024	0.4810	0.1024	0.4914
500	5.0487	0.2677	0.0487	0.2700

Table 2 Simulation table of PSD for $\beta=12$

n	$E(\hat{\beta})$	$Var(\hat{\beta})$	BIAS	MSE
50	12.3037	0.4726	0.3037	0.5648
100	12.2460	0.5026	0.2460	0.5631
200	12.2519	0.4799	0.2519	0.5432
300	12.2267	0.4746	0.2266	0.5259
500	12.1848	0.4477	0.1848	0.4818

Applications to automobile insurance claim frequency data

The most important issues in statistical research are the modelling of count data for automobile insurance claim frequency data. The goodness of fit (GOF) test of PSD over, PD, PLD, and PGD for automobile insurance claim frequency have been discussed in this section.

Automobile insurance plays a critical role in protecting individuals and organizations from financial losses arising from road accidents and related risks. One of the central challenges faced by insurers is accurately predicting the occurrence of insurance claims. Claim frequency, defined as the number of claims reported by policyholders during a specified exposure period, is a key component of actuarial risk assessment and premium determination. Accurate modeling of claim frequency enables insurers to establish fair and adequate premiums, improve underwriting decisions, and maintain financial stability.

Automobile claim frequency data are typically characterized as count data, where the number of claims assumes non-negative integer values. These data often exhibit important statistical features such as over-dispersion and heterogeneity due to differences in policyholder characteristics, vehicle attributes, driving behaviour, and environmental conditions. Factors such as driver's age, gender, driving experience, vehicle type, geographic location, and exposure duration significantly influence the likelihood of claim occurrence. Understanding the relationship between these risk factors and claim frequency is essential for effective risk classification and pricing strategies.

Traditional statistical approaches, particularly generalized linear models (GLMs) such as the Poisson regression model, have been widely used to model claim frequency due to their suitability for count data. However, the assumption of equality between the mean and variance in the Poisson model is often violated in real-world insurance data, resulting in over-dispersion. To address this limitation, alternative models such as the negative binomial regression model have been proposed, offering greater flexibility and improved model performance.

Modeling of automobile claim frequency is essential for ensuring actuarially sound decision-making and promoting the long-term sustainability of insurance operations. During recent decades several researchers have tried to provide a suitable distribution which can capture the over-dispersion in the automobile insurance claim data. Before using the probability distribution to model automobile insurance frequency claim data, we have to understand the structure of collective risk model. In collective risk model, the three important ingredients are aggregate claim, number of claims and the amount of each claim and all these ingredients are random variable. In the collective risk theory, the random variable of interest is the aggregate claim given by $X=\sum_{i=1}^{N} X_i$, where N is the random variable denoting the number of claims and $X_i\ (i=1,2,3,..)$ is the random variable denoting the size or amount of the ith claim. Assuming that $X_1, X_2, X_3, ...$ are independent and identically distributed random variables which are also independent of the random number of claims N, the probability distribution of the sum $X=\sum_{i=1}^{N} X_i$ will be a compound distribution. The detailed study about the use of compound distributions to model the insurance claim data using compound distributions are available in Nadarajah and Kotz[22,23] and Rolski et al.,[24] Klugman et al.[25] The considered automobile insurance claim frequency data are presented in the following Table 3. The GOF measures of datasets 1 to 8 in table 3 are given in Tables 4-11 along with the maximum likelihood estimate of the parameter and their standard error (SE). Based on the values of AIC (Akaike information criterion) and BIC (Bayesian information criterion), it is quite obvious that except for the table 11, in all datasets the PSD provides the best fit and it can be concluded that the PSD is a suitable model to model automobile insurance claim frequency data.

Table 3 Data sets of automobile insurance claim frequency available in Nadarajah and Kotz[22,23] and Rolski et al.,[24] Klugman et al.[25]

Data set														
1	No. of claims	0	1	2	3	4	5	6	7					
	Observed frequency	7840	1317	239	42	14	4	4	1					
2	No. of claims	0		1	2	3	4	5						
	Observed frequency	3719		232	38	7	3	1						
3	No. of claims	0		1	2	3	4							
	Observed frequency	96978		9240	704	43	9							
4	No. of claims	0		1	2	3	4	5	6					
	Observed frequency	20592		2651	297	41	7	0	1					
5	No. of claims	0		1	2	3	4	5	6	7	8	9	10	11
	Observed frequency	71087		6744	2067	690	248	95	34	17	4	3	3	2
6	No. of claims	0		1	2	3	4							
	Observed frequency	530642		334495	2585	211	25							
7	No. of claims	0		1	2	3	4	5	6					
	Observed frequency	103704		14075	1766	255	45	6	2					
8	No. of claims	0		1	2	3	4	5						
	Observed frequency	103704		46545	3935	317	28	3						

Table 4 GOF table for data set 1

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.2144	0.0048	10981.56	10983.56	10990.72
PLD	5.3998	0.1182	10712.56	10714.56	10721.71
PGD	5.3074	0.1191	10711.89	10713.89	10721.05
PSD	4.8562	0.1105	10709.63	10711.63	10718.79

Table 5 GOF table for data set 2

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.0865	0.0046	2492.15	2494.15	2500.45
PLD	12.4344	0.6544	2415.30	2417.30	2423.60
PGD	12.3757	0.6564	2415.25	2417.25	2423.54
PSD	11.6457	0.6433	2414.85	2416.85	2423.15

Table 6 GOF table for data set 3

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.1011	0.0010	72376.51	72378.51	72388.09
PLD	10.7345	0.1010	72245.06	72247.06	72256.64
PGD	10.6725	0.1014	72245.34	72247.34	72256.92
PSD	9.9912	0.0989	72247.14	72249.14	72258.72

Table 7 GOF table for data set 4

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.1442	0.0025	20595.69	20597.69	20605.75
PLD	7.7279	0.1297	20447.76	20449.76	20457.82
PGD	7.6520	0.1305	20447.79	20449.79	20457.86
PSD	7.0698	0.1249	20447.99	20449.99	20458.06

Table 8 GOF table for data set 5

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.1833	0.0015	88962.52	88964.52	88973.82
PLD	6.2385	0.0503	82754.82	82756.82	82766.13
PGD	6.1456	0.0506	82734.74	82736.74	82746.04
PSD	5.6262	0.0474	82655.13	82657.13	82666.43

Table 9 GOF table for data set 6

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.0695	0.0004	293409.7	293411.7	293422.9
PLD	15.2759	0.0755	291591.3	291593.3	291604.5
PGD	15.2269	0.0756	291590.9	291592.9	291604.1
PSD	14.4591	0.0746	291587.6	291589.6	291600.8

Table 10 GOF table for data set 7

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.1551	0.0011	110216.9	110218.6	110228.6
PLD	7.2292	0.0519	109231.4	109233.4	109243.1
PGD	7.1503	0.0523	109231.3	109233.3	109243.0
PSD	6.5908	0.0498	109231.2	109233.2	109242.9

Table 11 GOF table for data set 8

Model	$\hat{\beta}$	SE	$-2\log L$	AIC	BIC
PD	0.1317	0.0006	342746.3	342748.3	342759.3
PLD	8.3951	0.0349	342924.9	342926.9	342937.9
PGD	8.3234	0.0351	342930.1	342932.1	342943.1
PSD	7.7160	0.0338	342956.8	342958.8	342969.7

Conclusion and future works

The Poisson-Shanker distribution (PSD) is log-concave and unimodal, and is a two-component mixture of negative binomial distributions. The cumulative distribution functions, survival function, hazard function, mean residual life functions of the PSD have been derived and their behaviours' for different values of parameter have been presented graphically. The expressions for reverse hazard function, the second rate of failure; the cumulative hazard function and the Mills ratio of the PSD have also been derived. The Shannon, Rényi and Tsallis entropy measures of the PSD are obtained in a form suitable for numerical evaluation to quantify the degree of uncertainty, randomness or information content associated with the PSD. The consistency of maximum likelihood estimator of the parameter of the PSD has been shown through simulation study. The PSD has been applied to eight real automobile insurance claim frequency datasets and found to provide quite satisfactory fit over other competing one parameter over-dispersed count distributions namely Poisson-Lindley distribution and Poisson-Garima distribution. Therefore, for automobile insurance claim frequency data, PSD can be considered as an important discrete probability model in insurance.

Future research on the **PSD** in finance and insurance and in biomedical sciences focuses on addressing its limitations in handling complex, heavy-tailed, or multi-modal data. Key areas of investigation will involve expanding to multivariate environments, implementing robust Bayesian estimations, and applying the distribution to integer-valued time series. Following are the important core areas of future research on PSD:

i Multivariate and Bivariate Modeling: Since the present PSD is uni-dimensional, the future research will focus on extending PSD into multivariate and bivariate frameworks using copula functions. This will be helpful to actuaries to model the joint frequencies of different risks occurring simultaneously, such as natural disasters and resultant supply chain disruptions.

ii Bayesian Estimation and Credibility Theory: Since the parameter of PSD are estimated using method of moment or the method of maximum likelihood, the future research will integrate Bayesian methodologies especially Markov Chain Monte Carlo (MCMC) to estimate risk parameter. This will help the insurance companies to seamlessly update claim probability forecasts when incorporating historical records leading to more accurate Bonus-Malus Systems (BMS) in motor or health insurance.

iii Time Series of Counts (or INAR Models): Since insurance claims and stock market transaction counts often exhibit serial correlation over time, the future research will focus on integrating PSD into integer –valued autoregressive (INAR) time series models. This will help to model claim arrivals dynamically over consecutive quarters and hence capturing seasonal spikes in liability or property damage claims.

iv Zero-inflated and Heavy-Tailed Extensions: Since financial markets and insurance portfolios frequently see an exceptionally high volume of zero trades or zero claims, conflating PSD would be highly beneficial. Therefore the future work will heavily focus on zero-inflated Poisson-Shanker distribution (ZI-PSD) which would be useful in modeling of loss severities where claim amounts have a minimum threshold but display significant tail-variance.

v Aggregate Claim Modeling and Ruin Theory: The compound versions of the PSD will have applications in the computation of aggregate claim distributions within the Cramer-Lundberg risk model. Therefore, by deriving analytical expressions for ruin probability, researchers can determine appropriate capital reserves and ruin probability for insurance firms dealing with over-dispersed and heavy-tailed data.

vi Actuarial and Insurance Studies: The future research on PSD will focus to model the severity and frequency of insurance claims.

vii Biostatistics and Epidemiology: The future research on PSD will focus to track the spread of diseases, analyze cytogenetic dosimetry lesions or model patient hospitalization rates.

Acknowledgments

None.

Conflicts of interest

The authors declares that there are no conlficts of interest.

References

1. Sankaran M. The discrete Poisson-Lindley distribution. *Biometrics*. 1970;26(1):145–149.
2. Shanker R. The discrete Poisson-Shanker distribution. *Jacobs J Biostat*. 2016;1(1):1–7.
3. Shanker R. Shanker distribution and its applications. *Int J Stat Appl*. 2015;5(6):338–348.
4. Shanker R. The discrete Poisson-Garima distribution. *Biom Biostat Int J*. 2017;5(2):48–53.
5. Shanker R. Garima distribution and its application to model behavioural science data. *Biom Biostat Int J*. 2016;4(7):275–281.

6. Shanker R, Baishya J, Tabassum R. Statistical properties and applications of Poisson-Garima distribution. *Biom Biostat Int J*. 2025;14(2):45–51.

7. Shanker R. A size-biased Poisson-Shanker distribution and its applications. *Int J Probab Stat*. 2017;6(3):33–44.

8. Helal TS, Elsehetry AM, Elshaarawy RS. Statistical properties of weighted Shanker distribution. *J Bus Environ Sci*. 2022;1(1):141–153.

9. Ray M, Shanker R. A compound of exponential and Shanker distribution. *J Sci Res Banaras Hindu Univ*. 2023;67(4):39–46.

10. Ray M, Shanker R. A compound of gamma and Shanker distribution. *Reliab Theory Appl*. 2023;18(2):87–99.

11. Shanker R, Hagos F, Tekie AL, et al. On discrete Poisson-Shanker distribution and its applications. *Biom Biostat Int J*. 2017;5(1):6–14.

12. Shanker R, Shukla KK. Weighted Shanker distribution and its applications to model lifetime data. *J Appl Quant Methods*. 2017;12(2):1–17.

13. Shanker R, Shukla KK. Power Shanker distribution and its application. *Turkiye Klinikleri J Biostat*. 2017;9(3):175–187.

14. Wararit P. Bootstrap methods for estimating the confidence interval for the index of dispersion of the zero-truncated Poisson-Shanker distribution. *Prog Appl Sci Technol*. 2023;13(2):17–25.

15. Gupta PL, Gupta RC, Tripathi RC. On the monotonic properties of discrete failure rates. *J Stat Plann Inference*. 1997;65(2):255–268.

16. Steutel FW. Log-concave and log-convex distributions. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B, eds. *Encyclopedia of Statistical Sciences*. New York, NY: Wiley; 2004.

17. Bagnoli M, Bergstrom TC. Log-concave probability and its applications in rationality and equilibrium. In: Colomer JE, ed. *Rationality and Equilibrium: A Symposium in Honour of Marcel K. Richter*. Berlin, Germany: Springer; 2006:217–241.

18. Keilson J, Gerber H. Some results for discrete unimodality. *J Am Stat Assoc*. 1971;66(334):386–389.

19. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27(3):379–423.

20. Rényi A. On measures of entropy and information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Vol 4. Berkeley, CA; 1961:547-562.

21. Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. *J Stat Phys*. 1988;52(1-2):479–487.

22. Nadarajah S, Kotz S. Compound mixed Poisson distribution I. *Scand Actuar J*. 2006b;3(1):141–162.

23. Nadarajah S, Kotz S. Compound mixed Poisson distribution II. *Scand Actuar J*. 2006a;3(1):163–181.

24. Rolski T, Schmidli H, Teugels JL. *Stochastic Processes for Insurance and Finance*. New York, NY: John Wiley & Sons; 1999.

25. Klugman SA, Panjer HH, Willmot GE. *Loss Models: From Data to Decisions*. 3rd ed. New York, NY: Wiley; 2012.