Earthline Journal of Mathematical Sciences
E-ISSN: 2581-8147
Volume 12, Number 2, 2023, Pages 255-270
https://doi.org/10.34198/ejms.12223.255270

On Poisson-Samade Distribution: Its Applications in Modelling Count Data

S. A. Aderoju[1,*], I. Adeniyi[2], J. B. Olaifa[1] and A. Olaosebikan[1]

[1] Department of Mathematics and Statistics, Kwara State University, Malete, P.M.B. 1530, Ilorin, Kwara State, Nigeria
e-mail: samueladeroju77@gmail.com[*]

[2] Department of Statistics, Federal University Lokoja, P.M.B. 1154, Lokoja, Kogi State, Nigeria

Abstract

A new mixed Poisson model is proposed as a better alternative for modelling count data in the presence of overdispersion and/or heavy-tail. The mathematical properties of the model were derived. The maximum likelihood estimation method is employed to estimate the model's parameters and its applications to the three real data sets discussed. The model is used to model sets of frequencies that have been used in different literature on the subject. The results of the new model were compared with Poisson, Negative Binomial and Generalized Poisson-Sujatha distributions (POD, NBD and GPSD, respectively). The parameter estimates expected frequencies and the goodness-of-fit statistics under each model are computed using R software. The results show that the proposed PSD fits better than POD, NBD and GPSD for all the data sets considered. Hence, PSD is a better alternative provided to model count data exhibiting overdispersion property.

1. Introduction

The Poisson model has been referred to as the standard model in modelling count data. The probability mass function (pmf) of the Poisson distribution is:

$$h(x|\mu) = \frac{\mu^x e^{-\mu}}{x!}, \qquad \text{for } x = 0, 1, 2, \ldots \text{ and } \mu > 0,$$

$$\text{with } mean = variance = \mu.$$

Received: March 14, 2023; Revised & Accepted: May 1, 2023; Published: May 7, 2023
2020 Mathematics Subject Classification: 60E05, 62H05.
Keywords and phrases: Samade distribution, Poisson distribution, mixed model, overdispersion, moments.
[*]Corresponding author

Obviously, the mean and variance of the Poisson distribution are equal, which is a rare situation for real life data. More often than none, the variance is often higher than the average. In such a scenario, which reflects the presence of overdispersion in the data. "The Poisson model fits overdispersed count data poorly" [1]. Therefore, alternative distributions such as Negative Binomial distribution (NBD), which have a dispersion parameter that accounts for the overdispersion have been proposed [1-5]. The NBD has *pmf* defined as:

$$P(x|k,p) = \binom{k+x-1}{x}(1-p)^x p^k; \quad \text{for } x = 0,1,2,\ldots \quad 0 \le p \le 1 \tag{1}$$

where $E(X) = \dfrac{k(1-p)}{p}$, $Var(X) = \dfrac{k(1-p)}{p^2}$.

The NBD is suitable for overdispersed count data but may not be appropriate for modelling data exhibiting heavy-tailed.

Several other models have been suggested for handling overdispersion issues in different fields (see [6-10] for examples and more details), one of the most recent ones is a new generalized Poisson-Sujatha distribution [5]. The pmf of the new generalized Poisson-Sujatha distribution (NGPSD) was defined as:

$$Pr(y|\beta,\alpha) = \frac{\theta^3}{\theta^2+\alpha\theta+2\alpha}\left(\frac{\alpha(y^2+\theta+3)+(\theta+4)\alpha y+(\theta^2+2\theta+1)}{(\theta+1)^{y+3}}\right); \quad y = 0,1 \ldots \tag{2}$$

for $\theta > 0$ and $\alpha \ge 0$,

with $E(X) = \dfrac{\theta^2+2\alpha(\theta+3)}{\theta\left(\theta^2+\alpha(\theta+2)\right)}$ and

$$Var(X) = \frac{\theta^4(\theta+1)+\alpha\theta^2(16+12\theta+3\theta^2)+2\alpha^2(6+12\theta+6\theta^2+\theta^3)}{\theta^2(\theta^2+\alpha(2+\theta))^2},$$

where θ and α are the shape and scale parameters, respectively.

The purpose of this study is to propose a new model that is capable of handling overdispersed and/or heavy-tailed count data. The model is to provide a better alternative to NBD and other similar models. In Section 2, we derive the new model. In Section 3, we present the mathematical properties of the model. Estimation of the model's parameters is provided in Section 4 while Section 5 is dedicated to its application to real data sets and concluding remarks are in Section 6.

2. The Poisson-Samade Distribution

Samade distribution (SD) proposed by [11] is a two-parameter lifetime distribution which was defined as:

$$f(x|\alpha,\theta) = \begin{cases} \dfrac{\theta^4}{\theta^4+6\alpha}(\theta+\alpha x^3)e^{-\theta x}, & \text{for } x,\alpha,\theta > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

The distribution is a mixture of exponential (θ) and gamma (4, θ) distributions.

Definition. A random variable X is said to follow a Poisson-Samade distribution (PSD) if it follows

$$X|\mu \sim Po(\mu)$$

$$\text{while } \mu|\alpha,\theta \sim Sa(\alpha,\theta)$$

for $\mu > 0$, $\alpha > 0$ and $\theta > 0$. Hence, we denote the unconditional distribution of PSD by $PS(\alpha,\theta)$.

Theorem. *If* $X \sim PS(\alpha,\theta)$, *then the probability mass function* (*pmf*) *of* X *is*

$$f(X|\alpha,\theta) = \frac{\theta^4}{\theta^4+6\alpha}\left(\frac{\theta(\theta+1)^3+\alpha(x^3+6x^2+11x+6)}{(\theta+1)^{x+4}}\right); \quad x = 0,1\ldots \tag{3}$$

$for\ \theta > 0\ and\ \alpha \geq 0.$

Proof. Suppose $X|\mu \sim Po(\mu)$ and $\mu|\alpha,\theta \sim Sa(\alpha,\theta)$, then the pmf of unconditional random variable X is given as:

$$f(x) = \int_0^\infty h(X = x|\mu)g(\mu|\alpha,\theta)d\mu$$

where

$$h(X = x|\mu) = \begin{cases} \dfrac{\mu^x e^{-\mu}}{x!}, & \text{for } x = 0,1,2\ldots \text{ and } \mu > 0 \\ 0, & \text{elsewhere} \end{cases}$$

and

$$g(\mu|\alpha,\theta) = \begin{cases} \dfrac{\theta^4}{\theta^4+6\alpha}(\theta+\alpha\mu^3)e^{-\theta\mu}, & \text{for } \mu,\alpha,\theta > 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$\therefore \quad f(x) = \int_0^\infty \frac{\mu^x e^{-\mu}}{x!} \frac{\theta^4}{\theta^4 + 6\alpha} (\theta + \alpha\mu^3) e^{-\theta\mu} d\mu$$

$$= \frac{\theta^4}{x!\,(\theta^4 + 6\alpha)} \int_0^\infty \mu^x e^{-\mu(\theta+1)} (\theta + \alpha\mu^3) d\mu$$

$$= \frac{\theta^4}{x!\,(\theta^4 + 6\alpha)} \left[\theta \int_0^\infty \mu^x e^{-\mu(\theta+1)} d\mu + \alpha \int_0^\infty \mu^{x+3} e^{-\mu(\theta+1)} d\mu\right]$$

$$= \frac{\theta^4}{x!\,(\theta^4 + 6\alpha)} \left[\frac{\theta\Gamma(x+1)}{(\theta+1)^{x+1}} + \frac{\alpha\Gamma(x+4)}{(\theta+1)^{x+4}}\right]$$

$$= \frac{\theta^4 \cancel{x!}}{\cancel{x!}(\theta^4 + 6\alpha)} \left[\frac{\theta}{(\theta+1)^{x+1}} + \frac{\alpha(x+3)(x+2)(x+1)}{(\theta+1)^{x+4}}\right]$$

$$= \frac{\theta^4}{(\theta^4 + 6\alpha)} \left[\frac{\theta}{(\theta+1)^{x+1}} + \frac{\alpha(x+3)(x+2)(x+1)}{(\theta+1)^{x+4}}\right]$$

$$\therefore f(x) = \frac{\theta^4}{(\theta^4 + 6\alpha)} \left[\frac{\theta(\theta+1)^3 + \alpha(x^3 + 6x^2 + 11x + 6)}{(\theta+1)^{x+4}}\right]; \quad x = 0,1,2 \ldots$$

where $\alpha > 0$ and $\theta > 0$.

Note that

$$\sum_{x=0}^{\infty} \frac{\theta^4}{(\theta^4 + 6\alpha)} \left[\frac{\theta(\theta+1)^3 + \alpha(x^3 + 6x^2 + 11x + 6)}{(\theta+1)^{x+4}}\right] = 1,$$

which shows $f(x)$ is true *pmf*.

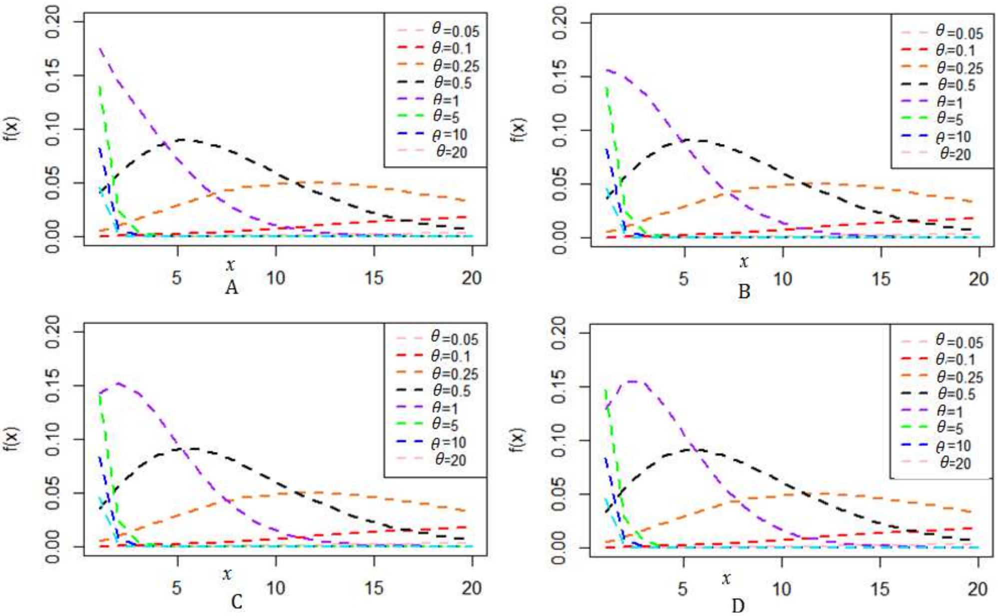

Figure 1: Graph of the pmf of PSD when $\hat{\alpha}$ = 0.25, 0.5, 1.0 and 5.0 (in Figure 1A, B, C, and D, respectively) at various values of $\widehat{\theta}$.

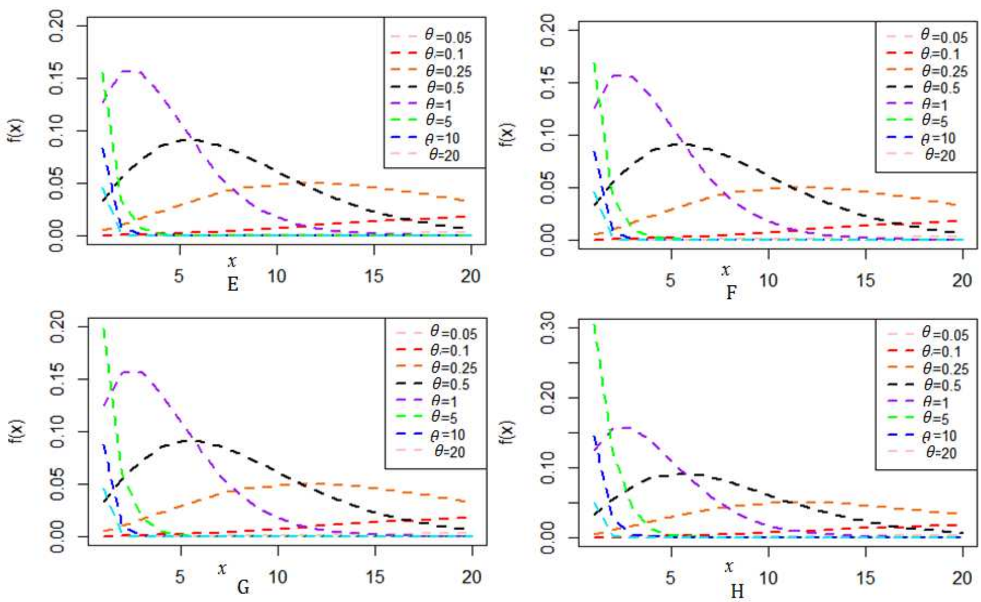

Figure 2: Graph of the pmf of PSD when $\hat{\alpha}$ = 10, 20, 50 and 100 (in Figure 2E, F, G, and H, respectively) at various values of $\hat{\theta}$.

The cumulative distribution function (CDF) is obtained as:

$$F(x) = \sum_{t=0}^{x} f(t)$$

$$F(x) = \sum_{t=0}^{x} \frac{\theta^4}{(\theta^4 + 6\alpha)} \left[\frac{\theta(\theta+1)^3 + \alpha(t^3 + 6t^2 + 11t + 6)}{(\theta+1)^{t+4}} \right]$$

$$\begin{aligned} F(x) = \frac{1}{(\theta^4 + 6\alpha)(\theta+1)^{x+4}} &[\theta^4(\theta+1)^3\big((1+\theta)(\theta+1)^x - 1\big) \\ &+ \alpha(6\theta^4(\theta+1)^x + 6((\theta+1)^x - 1) + 6\theta(4(\theta+1)^x - x - 4) \\ &+ 3\theta^2(12(\theta+1)^x - x^2 - 7x - 12) \\ &+ \theta^3(24(\theta+1)^x - x^3 - 9x^2 - 26x - 24))] \end{aligned}$$

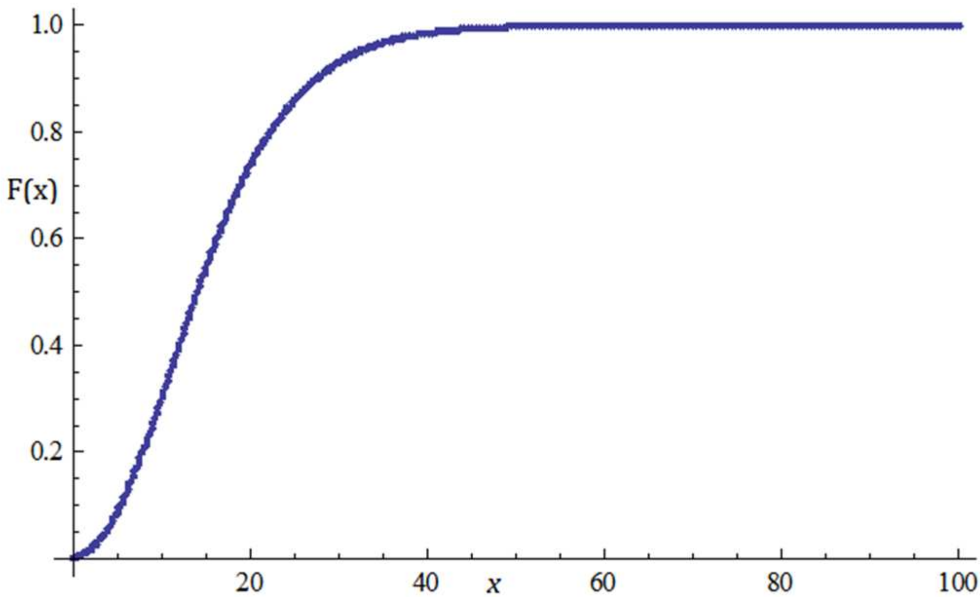

Figure 3: The CDF curve of the PSD.

3. The Mathematical Properties of Poisson-Samade Distribution

We present some of the mathematical characteristics of the PSD in this section. The r^{th} moment is defined as:

$$\mu_r = E(X^r) = \sum_{x=0}^{\infty} x^r f(x)$$

$$= \sum_{x=0}^{\infty} \frac{x^r \theta^4}{(\theta^4 + 6\alpha)} \left(\frac{\theta(\theta+1)^3 + \alpha(x^3 + 6x^2 + 11x + 6)}{(\theta+1)^{x+4}} \right).$$

The first four moments were obtained as:

$$\mu_1 = \frac{24\alpha + \theta^4}{6\alpha\theta + \theta^5},$$

$$\mu_2 = \frac{24\alpha(\theta+5) + \theta^4(2+\theta)}{6\alpha\theta^2 + \theta^6},$$

$$\mu_3 = \frac{24\alpha(30 + 15\theta + \theta^2) + \theta^4(6 + 6\theta + \theta^2)}{6\alpha\theta^3 + \theta^7},$$

$$\mu_4 = \frac{24\alpha(210 + 180\theta + 35\theta^2 + \theta^3) + \theta^4(24 + 36\theta + 14\theta^2 + \theta^3)}{6\alpha\theta^4 + \theta^8}.$$

Recall that $\mu_1$is the meanwhile variance is obtained as $Var(X) = \sigma^2 = \mu_2 - (\mu_1)^2$

$$\sigma^2 = \frac{24\alpha(\theta+5) + \theta^4(2+\theta)}{6\alpha\theta^2 + \theta^6} - \left(\frac{24\alpha + \theta^4}{6\alpha\theta + \theta^5} \right)^2,$$

$$\sigma^2 = \frac{(144\alpha^2 + \theta^8)(\theta+1) + 6\alpha\theta^4(14 + 5\theta)}{(6\alpha\theta + \theta^5)^2}.$$

Coefficient of Variation (CV)

$$CV = \frac{\sqrt{\sigma^2}}{\mu_1},$$

$$CV = \frac{\sqrt{\frac{(144\alpha^2 + \theta^8)(\theta+1) + 6\alpha\theta^4(14+5\theta)}{(6\alpha\theta + \theta^5)^2}}}{\frac{24\alpha + \theta^4}{6\alpha\theta + \theta^5}},$$

$$CV = \frac{(6\alpha\theta + \theta^5)\sqrt{\left((144\alpha^2 + \theta^8)(\theta+1) + 6\alpha\theta^4(14+5\theta)\right)/(6\alpha\theta + \theta^5)^2}}{(24\alpha + \theta^4)}.$$

Index of Dispersion (ID) is obtained as

$$ID = \frac{\sigma^2}{\mu_1} = \frac{144\alpha^2(1+\theta) + \theta^8(1+\theta) + 6\alpha\theta^4(14+5\theta)}{(24\alpha+\theta^4)(6\alpha\theta+\theta^5)}.$$

The corresponding kurtosis (K_s) and skewness (S_k) were obtained as follows:

$$K_s = \frac{\mu_4}{\sigma^4} = \frac{(6\alpha+\theta^4)^3(\theta^4(24+36\theta+14\theta^2+\theta^3)+24\alpha(210+180\theta+35\theta^2+\theta^3))}{(144\alpha^2(1+\theta)+\theta^8(1+\theta)+6\alpha\theta^4(14+5\theta))^2},$$

$$S_k = \frac{\mu_3}{\sigma^3} = \frac{\theta^4(6+6\theta+\theta^2)+24\alpha(30+15\theta+\theta^2)}{(6\alpha\theta^3+\theta^7)\left(\frac{144\alpha^2(1+\theta)+\theta^8(1+\theta)+6\alpha\theta^4(14+5\theta)}{(6\alpha\theta+\theta^5)^2}\right)^{3/2}}.$$

Table 1 shows the nature of the Mean (μ), variance (σ^2), Index of Dispersion (ID), coefficient of skewness (S_k), and coefficient of kurtosis (K_s) of the PSD for varying values of the parameters.

Table 1: Summary statistics of the moments of PSD.

θ	α	μ	σ^2	ID	S_k	K_s
	0.5	7.8775	24.3523	3.0913	9.8594	32.6783
0.5	1.0	7.9381	24.1817	3.0462	10.0603	33.4698
	5.0	7.9875	24.0372	3.0093	10.2303	34.1446
	0.5	3.2500	8.1875	2.5192	6.0292	19.3443
1.0	1.0	3.5714	8.2448	2.3085	6.7402	21.6434
	5.0	3.9032	8.0874	2.0719	7.7604	25.2491
	0.5	0.2028	0.2451	1.2084	4.1649	19.0638
5.0	1.0	0.2057	0.2502	1.2164	4.1773	19.2351
	5.0	0.2274	0.2887	1.2691	4.2299	19.8568

From Table 1, we observed that for varying values of the parameters, the mean (μ) is either less than or closer to the variance (σ^2) which makes the distribution flexible for analysing over-dispersed ($\mu < \sigma^2$) and equal-dispersed ($\mu \cong \sigma^2$) data sets. The PSD also exhibits positively skewed ($S_k > 0$) which conform with the plots of the pmf of the distribution in Figure 1.

4. Estimation of the Parameters of Poisson-Samade Distribution

The likelihood function of the PSD is given as:

$$L(\theta,\alpha|x_i)=\prod_{i=1}^{n}\frac{\theta^4}{(\theta^4+6\alpha)}\left(\frac{\theta(\theta+1)^3+\alpha({x_i}^3+6{x_i}^2+11x_i+6)}{(\theta+1)^{x_i+4}}\right). \tag{4}$$

The log-likelihood function is:

$$\ell=nlog(\theta^4)-nlog(\theta^4+6\alpha)+\sum_{i=1}^{n}\log\left(\theta(\theta+1)^3+\alpha({x_i}^3+6{x_i}^2+11x_i+6)\right)$$
$$-\sum_{i=1}^{n}(x_i+4)\log(\theta+1). \tag{5}$$

Therefore, the MLE of the parameters can be obtained numerically by solving the following partial differential equations

$$\frac{\partial\ell}{\partial\theta}=\frac{4n\theta^3}{\theta^4+6\alpha}+\frac{4n}{\theta}-\sum_{i=1}^{n}\frac{x_i+4}{\theta+1}+\sum_{i=1}^{n}\frac{3\theta(\theta+1)^2+\theta(\theta+1)^3}{\alpha({x_i}^3+6{x_i}^2+11x_i+6)+\theta(\theta+1)^3},$$

$$\frac{\partial^2\ell}{\partial\theta^2}=\frac{16n\theta^6}{(\theta^4+6\alpha)^2}-\frac{12n\theta^2}{\theta^4+6\alpha}-\frac{4n}{\theta^2}+\sum_{i=1}^{n}\frac{x_i+4}{(\theta+1)^2}$$
$$+(\theta+1)\sum_{i=1}^{n}\frac{12\theta-\frac{(\theta+1)^3(4\theta+1)^2}{\alpha({x_i}^3+6{x_i}^2+11x_i+6)+\theta(\theta+1)^3}+6}{\alpha({x_i}^3+6{x_i}^2+11x_i+6)+\theta(\theta+1)^3},$$

$$\frac{\partial\ell}{\partial\alpha}=-\frac{6n}{\theta^4+6\alpha}+\sum_{i=1}^{n}\frac{{x_i}^3+6{x_i}^2+11x_i+6}{\alpha({x_i}^3+6{x_i}^2+11x_i+6)+\theta(\theta+1)^3},$$

$$\frac{\partial^2\ell}{\partial\alpha^2}=\frac{36n}{(\theta^4+6\alpha)^2}-\sum_{i=1}^{n}\frac{({x_i}^3+6{x_i}^2+11x_i+6)^2}{(\alpha({x_i}^3+6{x_i}^2+11x_i+6)+\theta(\theta+1)^3)^2}.$$

Thus, maximum likelihood estimates of (5) can be carried out using iterative methods such as Fisher Score Algorithm, Bisection method Regula-Falsi method or Newton-Raphson (NR) iterative method among others.

However, we obtained the MLEs of the parameters by direct maximization of the log-likelihood function using the "***optim***" routine of R software ([12]) with the "***L-BFGS-B***" method. This can as well be done by using **PROC NLMIXED** in SAS.

5. Application

To examine the goodness-of-fit of the PSD, we apply the proposed model to some data sets and compared its performance with models considered in some existing literature.

5.1. Data Set 1

The dataset was obtained from [13] representing epileptic seizure counts (see [14]). The Mean and variance from the data are 1.544 and 2.883, respectively.

Table 2: Observed and expected frequencies of epileptic seizure counts

X	Observed Frequency	Models			
		POD	NBD	NGPSD	PSD
0	126	74.9	120.2	122.3	123.9
1	80	115.7	93.0	89.6	85.4
2	59	89.3	59.2	58.8	59.0
3	42	46.0	35.0	35.8	37.4
4	24	17.8	19.8	20.6	21.8
5	8	5.5	11.0	11.4	11.8
6	5	1.4	6.0	6.1	6.1
7	4	$\left.\begin{matrix}0.3\\0.1\end{matrix}\right\}0.4$	3.2	3.2	3.0
≥ 8	3		3.6	3.2	2.6
Total	351	351	351	351	351
MLE		$\hat{\mu} = 1.5442$	$\hat{r} = 1.550$ $\hat{p} = 0.5009$	$\hat{\alpha} = 1.3155$ $\hat{\theta} = 1.3716$	$\hat{\alpha} = 3.7438$ $\hat{\theta} = 1.3716$
LogLik		-636.0455	-594.9419	-594.0483	-593.2133
$\boldsymbol{X^2}$		80.76	5.6735	4.2010	2.9818
df		6	6	6	6
$\boldsymbol{G^2}$		70.15	5.7061	4.2708	3.0961
P-value		< 0.001	0.4607	0.6495	**0.8111**
AIC		1274.091	1193.884	1192.097	**1190.427**
BIC		1272.091	1194.043	1192,256	**1190.586**
Mean	1.544	1.544	1.5442	1.5448	1.5442
Variance	2.883	1.544	3.0827	3.0433	2.9445
Lawal's rule		0.1620	0.068	0.068	0.068

The underlying distribution for this data is Poisson. However, when the Poisson model is applied to the data in Table 2, it fits very poorly, with $\hat{\mu} = 1.54$ and $\hat{\sigma}^2 = 1.54$. The observed mean, $\mu = 1.544$ but the observed variance is, $\sigma^2 = 2.883$. Obviously, the variance is greater than the expected value in the observed data. As expected, Poisson distribution (POD) failed to fit the data. Therefore, we would need models that will account for the over-dispersion in the data, so we used other models considered by Bhati et al. [13], Aderoju [5] and the PSD. The results show that the proposed PSD fits the data better than other models, using X^2, Deviance (G^2) and p-values. Alternative measures of fit provided by the AIC, Log-likelihood (LogLik) and BIC also strongly support the PSD as the best model compared to the NBD and NGPSD.

Moreover, [15] proposed a rule that the expected value can be as small as $\frac{r}{d^{3/2}}$ (where r is the number of expected values less than 3 and d is the degree of freedom under such a model) without violating the X^2 assumption. We refer to this as "Lawal's rule". Hence, the minimum expected values under PSD can be as small as $\frac{1}{6^{3/2}} = 0.068$, that is, only one expected value is less than 3 and the $d.f = (9-2-1) = 6$. So, we do not need to collapse any cell. The same thing applies to other models except for the POD. For the Poisson model, the minimum expected frequency is 0.1620. Therefore, we collapsed cells $X = 7$ and 8, now the $X^2 = 80.76$ on 6 $d.f$.

5.2. Data Set 2

The second dataset is the number of European corn-borer available in [16]. The observed mean and variance are 0.648 and 0.8479, respectively.

Table 3: Observed and expected number of European corn-borer.

x	**Observed Frequency**	**Models**			
		POD	**NBD**	**NGPSD**	**PSD**
0	188	169.5	185.9	186.8	187.1
1	83	109.8	89.2	87.8	86.6
2	36	35.6	33.0	33.1	34.1
3	14	7.7	11.0	11.2	11.5
4	2	$\left.\begin{matrix}1.2\\0.2\end{matrix}\right\}1.4$	3.5	3.6	3.5
≥5	1		1.4	1.5	1.0
Total	**324**	**324**	**324**	**324**	**324**

MLE		$\hat{\lambda} = 0.6481$	$\hat{r} = 1.8485$ $\hat{p} = 0.7404$	$\hat{\alpha} = 10.5494$ $\hat{\theta} = 3.3169$	$\hat{\alpha} = 173.153$ $\hat{\theta} = 4.6958$
LogLik		-362.2451	-355.955	-355.8283	-355.4974
$\boldsymbol{X^2}$		15.5481	2.1755	1.8928	1.4315
df		3	3	3	3
$\boldsymbol{G^2}$		14.6161	2.3709	2.2079	1.9295
P-value		0.0004	0.5368	0.5949	**0.6947**
AIC		726.4902	715.910	715.6566	**714.9949**
BIC		724.4902	715.1288	714.8755	**714.2138**
Mean	0.6481	0.6481	0.6482	0.6483	0.6481
Variance	0.8479	0.6481	0.8754	0.8852	0.8748
Lawal's rule		0.25	0.1925	0.1925	0.1925

From Table 3, the Poisson model is applied to the data, but it failed to fit, with $\hat{\mu} = \hat{\sigma}^2 = 0.6481$. The observed mean, $\mu = 0.6481$ but $\sigma^2 = 0.8479$. Clearly, the variance is not the same as the expected value in the observed data. Asexpected, POD failed to fit the data, hence, we need a more flexible (Poisson mixture) model developed for this situation. The results (using X^2, Deviance (G^2) and p-values) shown that the proposed PSD fits the data better than NBD and NGPSD, though they equally fit the data. Alternative model selection measures are provided by the AIC, Log-likelihood (LogLik) and BIC; these also show that the PSD provides a better alternative to the existing models used.

By Lawal's rule, the minimum expected value under PSD, NGPSD and NBD is 0.1925 and none of the expected values is less than that. So, we do not need to collapse any cell. For the Poisson model, the minimum expected frequency is 0.25. Hence, we collapsed cells $X = 4$ and 5, now the $X^2 = 15.5481$ on 3 $d.f$.

5.3. Data Set 3

In Table 4 the observed frequencies refer to the number of crimes for every month from 1982 to January 1993 making 145 observations in Greece. "The data show overdispersion with mean =2.2413 and variance = 3.3833, making the assumption of a mixed Poisson distribution plausible" ([17]).

Table 4: Number of monthly crimes for the period 1982-1993 in Greece.

x	**Observed Frequency**	**Models**			
		POD	**NBD**	**NGPSD**	**PSD**
0	21	15.4	23.5	30.3	25.6
1	41	34.6	35.2	34.0	34.3
2	32	38.7	32.2	28.0	30.8
3	16	28.9	23.2	20.0	22.4
4	19	16.2	12.4	13.2	14.3
5	8	7.3	8.2	8.2	8.4
6	4	2.7	4.3	4.9	4.6
7	1	0.9	2.1	2.9	2.4
8	2	$\left.\begin{matrix}0.2\\0.1\end{matrix}\right\}0.3$	1.0	1.6	1.2
≥ 9	1		0.5	1.9	1.0
Total	145	145	145	145	145
MLE		$\hat{\lambda} = 2.2414$	$\hat{r} = 4.4949$ $\hat{p} = 0.6673$	$\hat{\alpha} = 28.60$ $\hat{\theta} = 1.160$	$\hat{\alpha} = 46.50$ $\hat{\theta} = 1.7346$
LogLik		-281.0803	-274.5055	-276.7761	-274.875
$\boldsymbol{X^2}$		35.6264	7.0328	9.7239	7.0026
df		7	7	7	7
$\boldsymbol{G^2}$		20.5506	6.7007	10.6597	7.1835
P-value		0.000009	0.298	0.2048	**0.4286**
AIC		564.1606	**553.011**	557.5521	553.7499
BIC		562.1606	**553.4055**	557.9466	554.1444
Mean	2.241	2.2414	2.2414	2.2490	2.2516
Variance	3.4066	2.2414	3.3591	4.3867	3.6408
Lawal's rule		0.1768	0.1619	0.1619	0.1619

Similarly, for the crimes data in Table 4, the Poisson model is applied but fits poorly, with $\hat{\mu} = \hat{\sigma}^2 = 2.2414$. The observed mean is $\mu = 2.2410$ but the observed variance, $\sigma^2 = 3.4066$. Clearly, there is a presence of overdispersion in the data. The POD fits the data very poorly as anyone would expect. There is a need for a more flexible model that can handle overdispersed count data. Hence, we used the proposed PSD along with NBD and NGPSD. The results show that the proposed PSD fits the data better than NBD and NGPSD, though they equally fit the data fairly (using the X^2 and the p-values).

Similarly, the minimum expected value under PSD, NGPSD and NBD is 0.1619 and none of the expected values is less than that. So, we do not need to collapse any cell in this example as well. For the POD, the minimum expected frequency is 0.1768. Hence, we collapsed cells $X = 8$ and 9, now the $X^2 = 35.6264$ on 7 $d.f.$

6. Conclusion

In this paper, a new Poisson mixed model called Poisson-Samade distribution (PSD) has been developed by compounding Poisson distribution with Samade distribution. The expression for the k^{th} moment has been derived and hence the first four moments and variance are derived. The method of maximum likelihood estimation has been discussed for estimating the parameters of the proposed distribution. The distribution has been fitted using maximum likelihood estimate to three datasets to test its goodness of fit over Poisson distribution (POD), Negative Binomial distribution (NBD) and New generalized Poisson-Sujatha distribution (NGPSD) and found that PSD gives a much closer fit than POD, NBD and NGPSD in the considered datasets.

Conflicts of interest

No conflict of interest

Acknowledgments

We sincerely thank the associate editor and the reviewers for their constructive suggestions and comments.

References

[1] B. H. Lawal, On the negative binomial-generalized exponential distribution and its applications, *Applied Mathematical Sciences* 11(8) (2017), 345-360. https://doi.org/10.12988/ams.2017.612288

[2] R. A. Fisher, The negative binomial distribution, *Ann. Eugen., Lond.* 11 (1941), 182.

[3] Ashenaf A. Yirga, Sileshi F. Melesse, Henry G. Mwambi and Dawit G. Ayele, Negative binomial mixed models for analysing longitudinal CD4 count data, *Scientific Reports* 10 (2020), 16742. https://doi.org/10.1038/s41598-020-73883-7

[4] M. R. Sampford, The truncated negative binomial distribution, *Biometrika* 42(1/2) (1955), 58-69. https://doi.org/10.1093/biomet/42.1-2.58

[5] S. A. Aderoju, A new generalized Poisson-Sujatha distribution and its applications, *Applied Mathematical Sciences* 14(5) (2020), 229-234. https://doi.org/10.12988/ams.2020.914185

[6] Sirinapa Aryuyuen, The negative binomial-new generalized Lindley distribution for count data: Properties and application, *Pakistan Journal of Statistics and Operation Research* 18(1) (2022), 167-177. http://dx.doi.org/10.18187/pjsor.v18i1.2988

[7] Xinyan Zhang, Himel Mallick, Zaixiang Tang, Lei Zhang, Xiangqin Cui, Andrew K. Benson and Nengjun Yi, Negative binomial mixed models for analyzing microbiome count data, *BMC Bioinformatics* 18 (2017), 4. https://doi.org/10.1186/s12859-016-1441-7

[8] F. C. Opone, E. A. Izekor, I. U. Akata and F. E. U. Osagiede, A discrete analogue of the continuous Marshall-Olkin Weibull distribution with application to count data, *Earthline Journal of Mathematical Sciences* 5(2) (2021), 415-428. https://doi.org/10.34198/ejms.5221.415428

[9] S. A. Aderoju and E. T. Jolayemi, On zero-truncated negative binomial with excess ones, *Asian Journal of Probability and Statistics* (2023), to appear.

[10] Ademola Abiodun Adetunji and Shamsul Rijal Muhammad Sabri, An alternative count distribution for modeling dispersed observations, *Pertanika J. Sci. & Technol.* 31(3) (2023), 1587-1603. https://doi.org/10.47836/pjst.31.3.25

[11] S. Aderoju, Samade probability distribution: its properties and application to real lifetime data, *Asian Journal of Probability and Statistics* 14(1) (2021), 1-11. https://doi.org/10.9734/ajpas/2021/v14i130317

[12] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2022. https://www.R-project.org/.

[13] D. Bhati, D. V. S Sastry and PZ Maha Qadri, A new generalized Poisson-Lindley distribution: Applications and properties, *Austrian Journal of Statistics* 44 (2015), 35-51. https://doi.org/10.17713/ajs.v44i4.54

[14] S. Chakraborty, On some distributional properties of the family of weighted generalized Poisson distribution, *Communications in Statistics - Theory and Methods* 39(15) (2010), 2767-2788. https://doi.org/10.1080/03610920903129141

[15] H. B. Lawal, Tables of percentage points of Pearson's goodness-of-fit statistic for use with small expectations, *Applied Statistics* 29 (1980), 292-298. https://doi.org/10.2307/2346904

[16] J. U. McGuire, T. A. Brindley and T. A. Bancroft, The distribution of European corn borer larvae *Pyrausta nubilalis* (Hbn.), in field corn, *Biometrics* 13 (1957), 65-78. https://doi.org/10.2307/3001903

[17] D. Karlis, EM algorithm for mixed Poisson and other discrete distributions, *Astin Bulletin* 35(1) (2005), 3-24. https://doi.org/10.1017/s0515036100014033