Thailand Statistician
January 2026; 24(1): 137-153
http://statassoc.or.th
Contributed paper

On Zero-Inflated Poisson Garima Distribution and its Applications to Count Data

Peer Bilal Ahmad* and Zehra Skinder
Department of Mathematical Sciences, Islamic University of Science & Technology, Awantipura, Jammu and Kashmir, India
*Corresponding author; Email: bilalahmadpz@gmail.com

Received: 20 September 2023
Revised: 11 July 2024
Accepted: 4 August 2025

Abstract

Excessive zero counts are one of the causes of over-dispersion in count data that is often observed in different fields. In this paper, we propose a new zero-inflated model namely 'the zero-inflated Poisson Garima distribution' to handle excessive zero counts. Various structural properties including reliability characteristics, generating functions, moments, etc. are obtained. Also, the parametric estimation of the proposed model is obtained using maximum likelihood method of estimation. Furthermore, a simulation study is carried out to check the behaviour of maximum likelihood estimators. Moreover, the proposed model and the baseline model are distinguished using two different test procedures. Finally, two real-life data sets taken from different domains are considered to validate the empirical applications of the proposed model.

Keywords: Poisson Garima distribution, zero-inflated distribution, maximum likelihood estimation, goodness of fit, testing of hypothesis.

1. Introduction

We require a robust class of discrete probability distributions in addition to the traditional discrete distributions in order to model real-world data. Recent years have seen the introduction of numerous generalized discrete distributions in this regard. Zero-inflated discrete distributions are effective at capturing situations where there is an excess of zeros and over-dispersion. The problem of zero inflation becomes prominent in count data modelling due to the large occurrence of zeros than that admitted by the conventional Poisson distribution. Zero-inflated count models provide an important framework to model this type of situation. For instance, Lambert (1992) used zero-inflated Poisson (ZIP) distribution to handle count data with excess of zeros. Neyman (1939), Cohen (1960), and Singh (1962) used zero-inflated Poisson distribution to analyse various types of zero-inflated count data sets. Gupta et al. (1996) developed a generalized version of zero-inflated Poisson model called as zero adjusted generalized Poisson model. Also, some useful models were employed by Ridout et al. (1998) for fitting discrete data with excess zeros. da Silva et al. (2018) introduced zero-modified Poisson-Sujata distribution for explaining count data exhibiting inflation or deflation of zeros. They took into consideration data sets from the biological sciences to explain the real-life applicability of the devised model. Johnson et al. (2015) proposed a Chi-Square statistic for comparing proportions of zeros among zero-inflated distributions. The proposed test-statistic can be used for both discrete as well as continuous type of data. Bar-Lev and Ridder (2023) examined two groups of exponential dispersion models exhibiting zero-inflation and over-dispersion for fitting data sets. To calculate the actual associated probability distribution, they developed a numerical procedure. Sellers and Raim (2016) proposed the zero-inflated variant of the Conway Maxwell-Poisson regression model for count data. They also put forward various applications of devised model across different disciplines such as health, psychology, engineering, and business etc. Rahman et al. (2021) introduced one-inflated Binomial model having inflated frequency at point one. Real-life examples were discussed by them to portray the practical importance of the suggested model. The three-inflated version of Poisson distribution introduced by Rahman et al. (2022) to model number of a suicide cases in India during the Covid-19 pandemic. Sadeghkhani and Ahmed (2020) introduced a multiple-inflated Poisson distribution having inflation at multiple points. For future observation, the Bayes predictive model was studied under a loss function called the Kullback Leibler loss function, and a class of shrinkage priors was also taken into account. In order to explain the performance of presented pmf estimator's work, plug-in type pmf estimators were also explored. Sandhyaa and Abrahamb (2016) introduced a model called inflated-parameter Harris distribution. Several structural properties were explored and characterization on the basis of probability generating function was also given. To check the applicability of the model, real life-data was also considered. One of the count models proposed by Altun (2018) is called the zero-inflated Poisson-Lindley regression model. The model is suitable for over-dispersed and zero-inflated data sets. Rivas and Campos (2023) proposed a new distribution called zero-inflated Waring distribution. The model illustrates data behaviour correctly when data shows a large incidence of observed zeros. Some of the zero-inflated models by Tawiah et al. (2021) were used to explore the features of death trends during COVID-19 in Ghana on a daily basis. Junnumtuam et al. (2023) introduced a new discrete distribution called the zero-inflated Cosine Geometric (ZICG) distribution for modelling over-dispersed data with excessive zeros. Various structural properties like moment generating function, mean and variance were also presented. Furthermore, the confidence interval was also constructed by using the Wald's method. The Bayesian method with highest posterior density method was also used to estimate the true confidence intervals. Bekalo and Kebede (2021) proposed a zero-inflated model for count data with applications to the number of total visits regarding antenatal care service. Simmachan et al. (2022) studied road accident fatalities exhibiting under-dispersion and zero inflation by investigating the Conway Maxwell Poisson regression and its zero-inflated version. One of the recent works in the field of zero inflation is the zero-inflated version of a Poisson mixture model given by Wani & Ahmad (2023). They studied the model at length and also considered real-world data for testing the compatibility of the devised model. The baseline distribution, that we have used in this paper is by Shanker (2017) called the Poisson Garima distribution (PGD), obtained by combining the Garima distribution and the Poisson distribution by compounding technique and showed that the model is over-dispersed and flexible for statistical data analysis and several statistical properties were explored. The probability mass function (pmf) of the PGD is as follows.

$$p(X=x)=\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\frac{\theta x+(\theta^2+3\theta+1)}{(\theta+1)^x}, \tag{1}$$

where $x=0,1,2,3,...$, and $\theta>0$. The PGD has found immense applications in different fields.

Since in many practical situations, the different models like Poisson distribution, zero-inflated Poisson distribution, zero-inflated negative binomial distribution, etc. are not preferable. In such cases, a zero-inflated version of the PGD provides a better fit. For example, in the application section, different real-life datasets are considered. Only the zero-inflated version of PGD gives the best fit in comparison to existing models. So, in this paper, we introduce zero-inflated Poisson Garima distribution (ZIPGD) along with distributional properties and other important aspects.

This paper is organized as follows. In Section 2, we show the derivation of the ZIPGD, cumulative distribution function. Also, the shapes of the pmf and cumulative distribution function (cdf) are presented in this Section. In Section 3, we have obtained the various structural properties along with reliability characteristics and generating functions. In Section 4, the estimation of the parameters of the ZIPGD is discussed. A rigorous simulation study is also discussed in this Section. In Section 5, the likelihood ratio test and wald's test is carried out to check the significance of the inflation parameter. Certain real-life data applications are considered in Section 6 to highlight the usefulness of the model. Also, the zero-inflated version of PGD is not studied yet in the literature.

2. Zero-Inflated Poisson Garima Distribution (ZIPGD)

In this part, we introduce the novel zero-inflated version of PGD namely ZIPGD through the following theorem.

Theorem 1 *Let a random variable X∼ZIPGD (α, θ). Then the pmf of X is given as*

$$P(X=x)=\begin{cases}\alpha+(1-\alpha)\frac{\theta(\theta^2+3\theta+1)}{(\theta^3+4\theta^2+5\theta+2)} & ;x=0\\ \\ (1-\alpha)\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\frac{\theta x+(\theta^2+3\theta+1)}{(\theta+1)^x} & ;x=1,2,...\end{cases} \tag{2}$$

where $\alpha\in$*[0,1] and* $\theta>0$.

Proof: If X is a random variable of PGD, then the pmf of X can be defined as

$$p(X=x)=\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\frac{\theta x+(\theta^2+3\theta+1)}{(\theta+1)^x} \qquad ;x=0,1,2,3,...;\theta>0$$

The zero-inflated distribution is an extra proportion added to the proportion of zero, then the pmf of zero-inflated distribution is given as

$$P(X=x)=\begin{cases}\alpha+(1-\alpha)p(X=0) & ;x=0\\ (1-\alpha)p(X=x) & ;x=1,2,...\end{cases}$$

where $\alpha\in$ [0,1].

Then, the pmf of the ZIPGD (α,θ) is obtained by substituting the pmf of the Poisson Garima random variable into zero-inflated model. Therefore, it can be written as

$$P(X=x)=\begin{cases}\alpha+(1-\alpha)\frac{\theta(\theta^2+3\theta+1)}{(\theta^3+4\theta^2+5\theta+2)} & ;x=0\\ \\ (1-\alpha)\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\frac{\theta x+(\theta^2+3\theta+1)}{(\theta+1)^x} & ;x=1,2,...\end{cases}$$

where $\alpha\in$[0,1] and $\theta>0$.
Hence proved.

Cumulative Distribution Function (cdf)

Theorem 2 *If X* $\sim$ *ZIPGD (α, θ), then its cdf is given as*

$$F(X=x)=1-\frac{(1-\alpha)(\theta^2+4\theta+x\theta+2)}{(\theta^3+4\theta^2+5\theta+2)(\theta+1)^x}. \tag{3}$$

Proof: If $X\sim$ ZIPGD (α,θ), then its cdf is as follows

$$F(x)=P(X\leq x)$$

$$= \sum_{z=0}^{x} P(X = z)$$

$$= \alpha + \frac{(1-\alpha)\theta(\theta^2+3\theta+1)}{(\theta+2)(\theta+1)} + \sum_{z=1}^{x} \frac{(1-\alpha)\theta(\theta z+\theta^2+3\theta+1)}{(1+\theta)^{z+2}(\theta+2)}$$

$$= \alpha + (1-\alpha) - \frac{(1-\alpha)(\theta^2+4\theta+x\theta+2)}{(\theta+1)^x(\theta^3+4\theta^2+5\theta+2)}$$

$$= 1 - \frac{(1-\alpha)(\theta^2+4\theta+x\theta+2)}{(\theta+1)^x(\theta^3+4\theta^2+5\theta+2)}$$

Hence proved.

The pmf plots given in Figure 1 for different combinations of parametric values indicate that the ZIPGD (2) is uni-modal. The mode is at zero for different combinations of parameters. Moreover, the tail shows a rapid decrease as the value increases for different combinations of parameters. Also, the proposed model is developed to deal with the data that shows inflation of zero observations and also can be highly suitable for right tailed data with little inflation at zero. Also, the cdf plots for different combinations of parametric values are given in Figure 2.

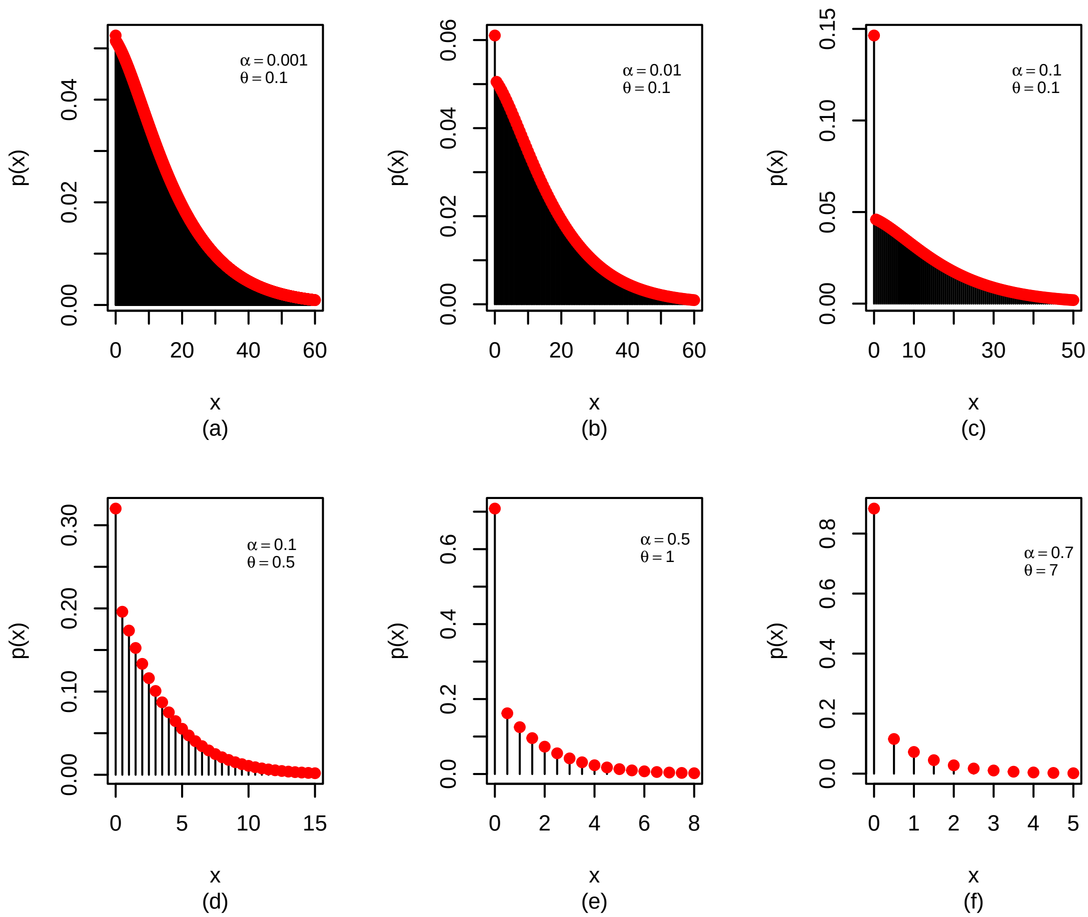

Figure 1 pmf plots of ZIPG Distribution of (α,θ) for (a) (α=0.001,θ=0.1), (b) (α=0.01,θ=0.1), (c) (α=0.1,θ=0.1), (d) (α=0.1,θ=0.5), (e) (α=0.5,θ=1), (f) (α=0.7,θ=7)

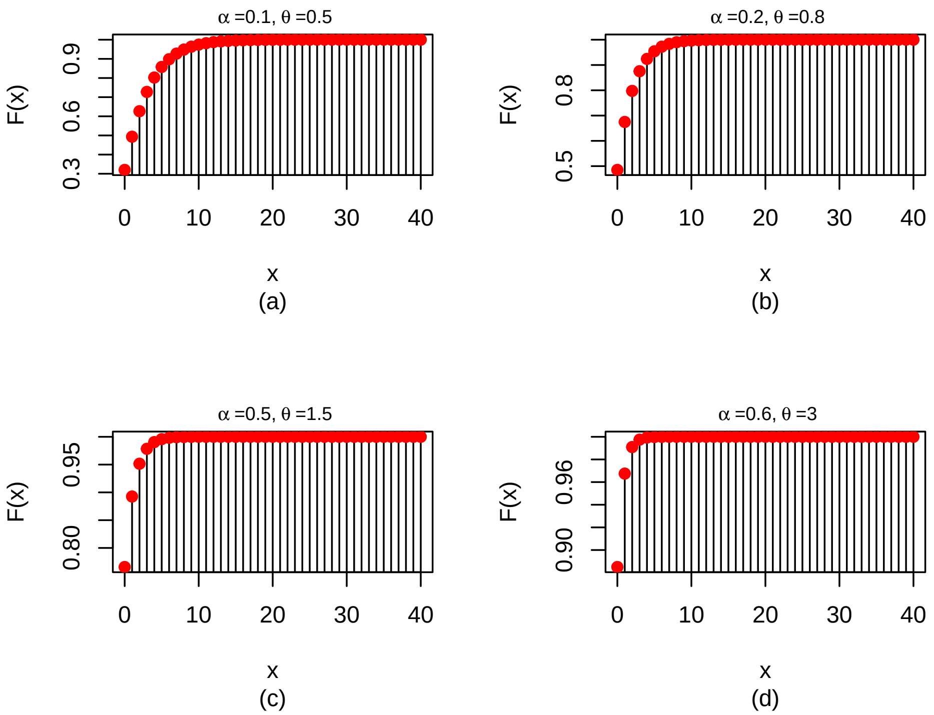

Figure 2 cdf plots of ZIPG Distribution of (α,θ) for (a) (α=0.1,θ=0.5), (b) (α=0.2,θ=0.8), (c) (α=0.5,θ=1.5), (d) (α=0.6,θ=3)

3. Structural Properties Along with Reliability Characteristics and Generating Functions

In this section, we have obtained several reliability characteristics like survival function, hazard rate function, and generating functions along with associated measures like index of dispersion (ID), skewness ($\sqrt{\beta_1}$), and kurtosis (β_2) of the ZIPGD (α, θ).

3.1. Survival function (SF)

The probability that a system will survive beyond a certain time period is called survival function. It is also called as reliability function or survivor function and is denoted by S". The Survival Function of ZIPGD (α, θ) is as follows

$$\begin{aligned} S(x) = 1 - F(x) &= 1 - \left[1 - \frac{(1-\alpha)(\theta^2+4\theta+\theta x+2)}{(\theta^3+4\theta^2+5\theta+2)(\theta+1)^x}\right] \\ &= \frac{(1-\alpha)(\theta^2+4\theta+\theta x+2)}{(\theta^3+4\theta^2+5\theta+2)(\theta+1)^x}. \end{aligned} \tag{4}$$

3.2. Hazard rate function (HRF)

Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from ZIPGD (α, θ) as given by Equation (2).

Define Y be the number of x_i's taking the value zero. Then Equation (2) can be written as follows

$$P(X = x_i) = \left[\alpha + (1-\alpha)\frac{\theta(\theta^2+3\theta+1)}{(\theta^3+4\theta^2+5\theta+2)}\right]^Y \left[(1-\alpha)\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\frac{\theta x_i+(\theta^2+3\theta+1)}{(1+\theta)^{x_i}}\right]^{1-Y}.$$

Now, using $S(x)$ from Equation (4). The Hazard Rate Function of ZIPGD (α, θ) is given as

$$R(x) = \frac{P(x)}{S(x)}$$

$$= \frac{(\theta^3+4\theta^2+5\theta+2)(1+\theta)^x \left[\alpha+(1-\alpha)\frac{\theta(\theta^2+3\theta+1)}{(\theta^3+4\theta^2+5\theta+2)}\right]^Y \left[(1-\alpha)\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\frac{\theta x_i+(\theta^2+3\theta+1)}{(1+\theta)^{x_i}}\right]^{1-Y}}{(1-\alpha)(\theta^2+4\theta+\theta x+2)}.$$

3.3. Moments and associated measures

3.3.1 Probability generating function

The probability generating function, $P_x(t)$ of ZIPGD (α, θ) is given as

$$P_x(t) = E(t^x)$$

$$P_x(t) = \sum_{x=0}^{\infty} t^x P(X=x)$$

$$P_x(t) = \alpha + (1-\alpha)\frac{\theta}{(\theta^3+4\theta^2+5\theta+2)}\sum_{x=0}^{\infty} t^x \frac{(\theta x + (\theta^2+1+3\theta))}{(1+\theta)^x}$$

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

$$P_x(t) = \alpha + (1-\alpha)\frac{\theta^3+\theta^2(4-t)+2\theta(2-t)+(1-t)}{(\theta^2+3\theta+2)(\theta-t+1)^2} \tag{5}$$

Putting $t=e^t$ in Equation (5), the moment generating function, $M_x(t)$ of ZIPGD (α, θ) is defined as

$$M_x(t) = \alpha + (1-\alpha)\frac{\theta^3+\theta^2(4-e^t)+2\theta(2-e^t)+(1-e^t)}{(\theta^2+3\theta+2)(\theta-e^t+1)^2}$$

Putting $t=e^{it}$ in Equation (5), the characteristic function, $\phi_x(t)$ of ZIPGD (α, θ) is defined as

$$\phi_x(t) = \alpha + (1-\alpha)\frac{\theta^3+\theta^2(4-e^{it})+2\theta(2-e^{it})+(1-e^{it})}{(\theta^2+3\theta+2)(\theta-e^{it}+1)^2}$$

By using Equation (5) one can easily obtain the first four raw moments of the proposed distribution.

$$\mu_1' = \frac{(1-\alpha)(\theta+3)}{\theta(\theta+2)}$$

$$\mu_2' = \frac{(1-\alpha)(\theta^2+5\theta+8)}{\theta^2(\theta+2)}$$

$$\mu_3' = \frac{(1-\alpha)(\theta^3+9\theta^2+30\theta+30)}{\theta^3(\theta+2)}$$

$$\mu_4' = \frac{(1-\alpha)(\theta^4+17\theta^3+92\theta^2+204\theta+144)}{\theta^4(\theta+2)}$$

3.3.2 Central moments

The central moments of the proposed distribution are obtained by using the relationship between raw moments and central moments. These are as follows

$$\mu_2 = \frac{(1-\alpha)(\theta^3+6\theta^2+12\theta+7+\alpha(\theta^2+6\theta+9))}{\theta^2(\theta+2)^2}$$

$$\mu_3 = \frac{(1-\alpha)}{\theta^3(\theta+2)^3}\left[\theta^5 + \theta^4(10+3\alpha) + \theta^3(42+28\alpha+2\alpha^2) + \theta^2(87+99\alpha+18\alpha^2)\right.$$
$$\left. + \theta(84+156\alpha+54\alpha^2) + (30+90\alpha+54\alpha^2)\right]$$
$$\mu_4 = \frac{(1-\alpha)}{\theta^4(\theta+2)^4}\left[\theta^7 + \theta^6(17+4\alpha) + \theta^5(148+58\alpha+6\alpha^2) + \theta^4(607+361\alpha+75\alpha^2+3\alpha^3)\right.$$
$$+\theta^3(1402+1206\alpha+378\alpha^2+36\alpha^3) + \theta^2(1816+2232\alpha+960\alpha^2+162\alpha^3)$$
$$\left. +\theta(1224+2036\alpha+1224\alpha^2+324\alpha^3) + (519\alpha+621\alpha^2+243\alpha^3)\right]$$

Mean (μ_1'), variance (σ^2), and coefficient of variation (CV) of the proposed model are given as respectively.

$$\mu_1' = \frac{(1-\alpha)(\theta+3)}{\theta(\theta+2)}$$
$$\sigma^2 = \frac{(1-\alpha)\left[\theta^3+6\theta^2+12\theta+7+\alpha(\theta^2+6\theta+9)\right]}{\theta^2(\theta+2)^2}$$
$$CV = \frac{\sqrt{(1-\alpha)(\theta^3+6\theta^2+12\theta+7)+\alpha(\theta^2+6\theta+9)}}{(1-\alpha)(\theta+3)}.$$

Remark

The ZIPGD (α, θ) is over-dispersed for $\theta > 0$ and $\alpha \in$ [0,1].
Proof: Suppose that the ZIPGD (α, θ) is under-dispersed. Then clearly Mean $>$ Variance, which implies that

$$\theta(\theta+3)(\theta+2) > (\theta^3+6\theta^2+12\theta+7+\alpha(\theta^2+6\theta+9))$$
$$\Rightarrow -(\theta^2+6\theta+7) > \alpha(\theta^2+6\theta+9)$$

which shows that $\alpha(\theta^2+6\theta+9) < -(\theta^2+6\theta+7)$, which is impossible for any $\theta > 0$ and $\alpha \in$ [0,1].
Hence the proof.

Index of Dispersion (ID)**:** It is a ratio used in probability theory that expresses how dispersed a given distribution is. The index of dispersion, sometimes referred to as the Variance to Mean Ratio or the dispersion index, is a normalised statistic as opposed to the standard deviation. When data are evenly distributed, the ID is zero; when data are under-dispersed, the ID is less than one; and is greater than 1 in the case of over-dispersed data.

The ID of the ZIPGD (α, θ) is given by

$$ID = \frac{[(\theta^3+6\theta^2+12\theta+7)+\alpha(\theta^2+6\theta+9)]}{\theta(\theta+2)(\theta+3)}$$
$$Skewness(\sqrt{\beta}_1) = \frac{1}{(1-\alpha)^{1/2}[(\theta^3+6\theta^2+12\theta+7)+\alpha(\theta^2+9\theta+9)]^{3/2}}\left[\theta^5+\theta^4(10+3\alpha)\right.$$
$$+\theta^3(42+28\alpha+2\alpha^2)+\theta^2(87+99\alpha+18\alpha^2)$$
$$\left. +\theta(84+156\alpha+54\alpha^2)+(30+90\alpha+54\alpha^2)\right].$$

From Table 1, we can see that for different combinations of parameters of α and θ. ZIPGD is over dispersed because the value of dispersion index is greater than one. Also, the skewness coefficient increases when the parameter value of α and θ increases. So, the distribution is positively skewed and has a long right tail. Further, it can be observed from the table, that the distribution has positive

Table 1 Behaviour of the model's descriptive statistics for various parameter values

	α=0.1					α=0.3					α=0.6				
$\theta\rightarrow$	**0.5**	**1**	**1.5**	**2**	**2.5**	**0.5**	**1**	**1.5**	**2**	**2.5**	**0.5**	**1**	**1.5**	**2**	**2.5**
μ_1'	2.520	1.200	0.770	0.560	0.440	1.960	0.930	0.680	0.430	0.340	1.400	0.600	0.4200	0.310	0.240
σ^2	9.920	2.960	1.470	0.890	0.600	8.810	2.550	1.240	0.740	0.500	7.080	2.000	0.960	0.570	0.380
ID	3.620	2.300	1.850	1.630	1.500	4.180	2.560	2.020	1.760	1.600	4.740	2.830	2.200	1.880	1.700
CV	1.190	1.380	1.550	1.700	1.840	1.460	1.650	1.830	2.000	2.160	1.840	2.060	2.260	2.450	2.630
$\sqrt{\beta_1}$	2.13	1.870	1.610	1.400	1.250	2.790	2.500	2.170	1.880	1.660	3.690	3.380	2.950	2.560	2.250
β_2	5.910	6.360	6.610	7.140	7.870	8.100	8.540	8.770	9.370	10.260	12.140	12.580	12.750	13.480	14.650

kurtosis. So, we can say that distribution is Leptokurtic i.e., thick tailed distribution.

$$Kurtosis(\beta_2) = \frac{1}{(1-\alpha)[(\theta^3+6\theta^2+12\theta+7)+\alpha(\theta^2+6\theta+9)]^2}\big[\theta^7+\theta^6(17+4\alpha)$$
$$+\theta^5(148+58\alpha+6\alpha^2)+\theta^4(607+361\alpha+75\alpha^2+3\alpha^3)$$
$$+\theta^3(1402+1206\alpha+378\alpha^2+36\alpha^3)+\theta^2(1816+2232\alpha+960\alpha^2+162\alpha^3)$$
$$+\theta(1224+2036\alpha+1224\alpha^2+324\alpha^3)+(519\alpha+621\alpha^2+243\alpha^3)\big]$$

4. Parametric Estimation

In this part, we discuss the parametric estimation of the ZIPGD (α, θ) by maximum likelihood method of estimation.

Maximum likelihood Estimation (MLE)

Let x_1, x_2, x_3, ..., x_n be a random sample of size n from ZIPGD(α, θ) and Y be the number of x_i's taking value zero, then the likelihood function of ZIPGD (α, θ) is given as:

$$L(\alpha,\theta/x) = \left[\alpha+(1-\alpha)\frac{\theta(\theta^2+3\theta+1)}{(\theta^3+4\theta^2+5\theta+2)}\right]^Y \prod_{\substack{x_i=1\\x_i\neq 0}}^{n}\left[(1-\alpha)\frac{\theta(\theta x_i+\theta^2+3\theta+1)}{(\theta+1)^{x_i}(\theta^3+4\theta^2+5\theta+2)}\right]$$

$$L(\alpha,\theta/x) = \left[\frac{[\alpha(\theta^3+4\theta^2+5\theta+2)+(1-\alpha)(\theta^3+3\theta^2+\theta)]^Y(1-\alpha)^{n-Y}\theta^{n-Y}}{(\theta^3+4\theta^2+5\theta+2)^n(1+\theta)^{\sum_{i=1}^n x_i}}\right]$$
$$\prod_{\substack{x_i=1\\x_i\neq 0}}^{n}(\theta x_i+\theta^2+3\theta+1)$$

$$\begin{aligned} lnL(\alpha,\theta/x) =& Y\ln[\alpha(\theta^3+4\theta^2+5\theta+2)+(1-\alpha)(\theta^3+3\theta^2+\theta)]+(n-Y)\ln(1-\alpha) \\ &+(n-Y)\ln\theta - n\ln(\theta^3+4\theta^2+5\theta+2)-\sum_{i=1}^{n}x_i\ln(\theta+1) \\ &+\sum_{\substack{x_i=1\\x_i\neq 0}}^{n} ln(\theta x_i+\theta^2+3\theta+1). \end{aligned} \tag{6}$$

Differentiating Equation (6) with respect to α and θ respectively, and equating to zero, we get

$$\frac{\partial}{\partial\alpha}\ln L(\alpha,\theta/x)=\frac{Y(\theta^2+4\theta+2)}{[\alpha(\theta^3+4\theta^2+5\theta+2)+(1-\alpha)(\theta^3+3\theta^2+\theta)]}-\frac{(n-Y)}{(1-\alpha)}=0. \tag{7}$$

On solving Equation (7), we get the estimate of parameter α.

Therefore,

$$\hat{\alpha}=\frac{Y(\hat{\theta}^3+4\hat{\theta}^2+5\hat{\theta}+2)-n(\hat{\theta}^3+3\hat{\theta}^2+\hat{\theta})}{n(\hat{\theta}^2+4\hat{\theta}+2)}. \tag{8}$$

Similarly,

$$\begin{aligned} \frac{\partial}{\partial\theta}\ln L(\alpha,\theta/x) =& \frac{Y[\alpha(2\theta+4)+(3\theta^2+6\theta+1)]}{[\alpha(\theta^2+4\theta+2)+(\theta^3+3\theta^2+\theta)]}+\frac{(n-Y)}{\theta}-\frac{n(3\theta^2+8\theta+5)}{(\theta^3+4\theta^2+5\theta+2)} \\ &-\frac{\sum_{i=1}^{n}x_i}{(1+\theta)}+\sum_{\substack{x_i=1\\x_i\neq 0}}^{n}\left[\frac{(x_i+2\theta+3)}{(\theta x_i+\theta^2+3\theta+1)}\right]=0. \end{aligned}$$

Since, the differential equation for parameter θ is not in closed form, various numerical methods like the Newton-Raphson method, Bisection method, regular Falsi Method, etc., can be employed to find the solution of the above differential equation. We have used R software R Core Team (2023) to obtain the ML estimators through the Newton-Raphson method. The Newton-Raphson method is one of the most common techniques used to find the roots of given equations. Generally speaking, it can be applied to find solutions for the system of equations. It is an algorithm for finding the roots of the function, which produces successively better approximations to the roots of the function with the given initial guess say x_0. The formula for finding the sufficiently precise value of the function $f(x)$ is

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}.$$

5. Simulation Study

In this section, we carry out a simulation study to investigate the finite sample behaviour of the maximum likelihood estimators for different sample sizes ($n = 25, 75, 100, 300, 600$) on various parameter settings including (($\alpha = 0.1, \theta = 0.5$), ($\alpha = 0.1, \theta = 1$), ($\alpha = 0.1, \theta = 2$), ($\alpha = 0.25, \theta = 0.5$), ($\alpha = 0.5, \theta = 1$), ($\alpha = 0.5, \theta = 2$)). The procedure was repeated 1000 times for the calculation of bias, variance, mean square error (MSE), and coverage probability and the results are given in Table 2. It can be seen from the table, that as the sample size increases, the bias, variance, and mean square error decreases and are close to zero for large sample sizes. Also, the coverage probability tends to be 0.95 as the sample size increases. These results suggest that maximum likelihood estimates are consistent and therefore can be used in estimating the unknown parameters of the proposed model.

Table 2 Simulation table of MLE's for the proposed model

Sample Size(n)	Parameter	$\alpha=0.1, \theta=0.5$ Bias	Variance	MSE	Coverage probability (95%)	$\alpha=0.1, \theta=1$ Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\alpha}$	0.02649	0.01124	0.01194	1.00	0.01629	0.01759	0.01786	0.98
	$\hat{\theta}$	0.01919	0.02680	0.02717	0.94	0.03157	0.10506	0.10606	0.94
75	$\hat{\alpha}$	-0.00630	0.00437	0.00441	0.98	0.00872	0.00981	0.00989	0.94
	$\hat{\theta}$	0.00059	0.00513	0.00513	0.94	0.02255	0.02902	0.02953	0.96
100	$\hat{\alpha}$	0.00298	0.00299	0.00299	1.00	0.01413	0.00832	0.00852	0.98
	$\hat{\theta}$	-0.00322	0.00397	0.00398	0.94	0.00366	0.02556	0.02557	0.98
300	$\hat{\alpha}$	0.00149	0.00119	0.00119	0.96	0.00877	0.00278	0.00285	1.00
	$\hat{\theta}$	0.00254	0.00096	0.00097	0.98	0.00360	0.00787	0.00789	0.98
600	$\hat{\alpha}$	-0.00388	0.00068	0.00069	0.98	0.00222	0.00140	0.00141	0.98
	$\hat{\theta}$	0.00542	0.00067	0.00070	0.96	0.00855	0.00389	0.00396	0.96

Sample Size(n)	Parameter	$\alpha=0.1, \theta=2$ Bias	Variance	MSE	Coverage probability (95%)	$\alpha=0.25, \theta=0.5$ Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\alpha}$	0.03098	0.03783	0.03879	0.96	-0.02011	0.01635	0.01676	1.00
	$\hat{\theta}$	0.06792	0.85409	0.85870	0.94	0.06447	0.03633	0.04049	0.96
75	$\hat{\alpha}$	0.02798	0.01630	0.01709	0.98	0.00595	0.00504	0.00507	0.96
	$\hat{\theta}$	0.02012	0.32090	0.32130	0.98	0.01586	0.00672	0.00697	0.98
100	$\hat{\alpha}$	0.06063	0.01311	0.01679	0.96	-0.00410	0.00520	0.00522	0.94
	$\hat{\theta}$	0.02525	0.00547	0.00611	0.94	-0.03173	0.18858	0.18959	1.00
300	$\hat{\alpha}$	-0.01070	0.00703	0.00715	0.98	0.00076	0.00137	0.00138	0.98
	$\hat{\theta}$	0.02150	0.04878	0.04924	0.98	0.00652	0.00140	0.00144	0.98
600	$\hat{\alpha}$	-0.00123	0.00451	0.00451	0.98	-0.00272	0.00066	0.00067	0.94
	$\hat{\theta}$	0.02591	0.02718	0.02785	0.98	0.00916	0.00087	0.00096	0.94

Sample Size(n)	Parameter	$\alpha=0.5, \theta=1$ Bias	Variance	MSE	Coverage probability (95%)	$\alpha=0.5, \theta=2$ Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\alpha}$	-0.04526	0.05407	0.05612	0.92	-0.19750	0.08476	0.12376	0.98
	$\hat{\theta}$	0.31897	0.52603	0.62778	0.94	2.17655	20.29304	25.03040	0.92
75	$\hat{\alpha}$	-0.01914	0.01336	0.01373	0.94	-0.06464	0.04439	0.04857	0.90
	$\hat{\theta}$	0.14155	0.08819	0.10822	0.96	0.38548	1.60616	1.75476	0.92
100	$\hat{\alpha}$	-0.01793	0.00891	0.00923	1.00	-0.06439	0.02372	0.02787	0.96
	$\hat{\theta}$	0.10182	0.08047	0.09084	0.94	0.28519	0.54369	0.62503	0.96
300	$\hat{\alpha}$	-0.00456	0.00364	0.00366	0.86	-0.01743	0.01285	0.01315	0.96
	$\hat{\theta}$	0.02943	0.02248	0.02334	0.94	0.08660	0.23741	0.24491	0.96
600	$\hat{\alpha}$	0.00253	0.00191	0.00192	0.90	0.00136	0.00355	0.00355	0.94
	$\hat{\theta}$	0.01509	0.01259	0.01282	0.88	0.01065	0.07085	0.07096	0.98

6. Testing of Hypothesis

In this part, we have checked the significance of the inflation parameter (α) by the likelihood ratio test and the wald's test.

6.1. Likelihood ratio test (LRT)

In order to test the significance of inflation parameter α of the ZIPGD (α, θ). The LRT is carried out to distinguish between PGD (θ) and ZIPGD (α, θ). Here the null hypothesis is

$$H_o : \alpha = 0 \text{ vs the alternative hypothesis } H_1 : \alpha \neq 0$$

In case of LRT, the test statistic is given by

$$-2ln\lambda = 2(l_1 - l_2), \tag{9}$$

where $l_1 = lnL(\hat{\lambda}; y)$, with $\hat{\lambda}$ is the maximum likelihood estimator for $\lambda = (\alpha, \theta)$ without limitation, and $l_2 = lnL(\hat{\lambda}^*, x)$, in which $\hat{\lambda}^*$ is the maximum likelihood estimator for λ under the null hypothesis H_o. The test statistic described in Equation (9) is asymptotically distributed as χ^2 with one degree of freedom.

Table 3 Calculated value of test statistic in case of LRT

	$lnL(\hat{\lambda}^*; y)$	$lnL(\hat{\lambda}; y)$	Test statistic
Dataset 1	-341.30	-332.75	8.55
Dataset 2	-501.80	-492.05	9.75

Since the critical value at 5% level of significance is 3.84 at one degree of freedom. It can be seen from Table 3, that the null hypothesis is rejected in all the two data sets. Hence, we can say that the additional parameter in the model is significant.

6.2. Wald's test

Here for testing the significance of inflation parameter α of the ZIPGD (α, θ), we assess the Wald's test. To test the null hypothesis

$$H_0 : \alpha = 0 \text{ vs the alternative hypothesis } H_1 : \alpha \neq 0.$$

In case of Wald's test, the test statistic is given by

$$W_\alpha = \frac{\hat{\alpha}^2}{Var(\hat{\alpha})}, \tag{10}$$

where, $Var(\hat{\alpha})$ represents the diagonal element of Fisher information matrix at $\alpha = \hat{\alpha}$ and $\theta = \hat{\theta}$. The test statistic given in Equation (10) is asymptotically distributed as χ^2 with one degree of freedom.

Table 4 Calculated value of test statistic in case of Wald's Test

	Test statistic
Dataset 1	124.81
Dataset 2	11.87

Since at one degree of freedom, the critical value at 5% level of significance is 3.84. It can be seen from Table 4, that the null hypothesis is rejected in all the two data sets. Hence, we can say that the additional parameter in the model is significant.

7. Applications

In this section, the practical significance of ZIPGD is validated by fitting the proposed model to two real-life data sets. Along with the proposed model, other prominent distributions such as zero-inflated Poisson distribution (ZIPD), PGD, zero-inflated Negative Binomial distribution (ZINBD), and Poisson distribution (PD) are also being fitted and the fitted models are compared on the basis of well-known criteria like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Chi-Square values.

7.1. Data Set I

The first data set in Table 5 consists of frequencies regarding stillbirths in 402 litters of New Zealand white rabbits, originally used by Morgan et al. (2007) when discussing score test statistic. We have calculated the expected frequencies for the Zero-Inflated Poisson Garima model and for other competing models. Also, in order to check the goodness of fit, we have computed Pearson's chi-square statistic and corresponding p-values. The results are provided in Table 6. On the basis of p-values, we perceive that our proposed model has the highest p-value and hence gives a better fit than other competing models. Further, AIC and BIC were also calculated for all the competing models for model comparison (see Table 7). Since our proposed model has the least values of AIC and BIC, implying the outperformance of our proposed model over other competing models on this data set. Thus, these results suggest that our proposed model fits well on this type of data.

Table 5 Number of stillbirths

No. of Stillbirths	0	1	2	3	4	5	6
Observed Counts	313	48	20	7	5	2	6

Table 6 Expected frequencies and χ^2 values for fitted models

Number of stillbirths(X)	Observed Counts	$ZIPGD$	$ZIPD$	PGD	$ZINBD$	PD
0	314	314	314	280	314	260
1	48	44	36	86	41	113
2	20	22	28	26	25	25
3	7	11	15	8	13	4
4	5	6	6	2	6	0
5	2	3	2	1	2	0
6	6	1	0	0	1	0
Degrees of Freedom		2	2	2	1	1
ML Estimates		$\hat{\alpha}$=0.5745	$\hat{\lambda}$=1.5783	$\hat{\zeta}$=2.79625	$\hat{p}$=0.6936	$\hat{\lambda}$=0.4353
		$\hat{\theta}$=0.1.2758	$\hat{\phi}$=0.7241		$\hat{r}$=3.0052	
					$\hat{a}$=0.6714	
χ^2		2.90	13.68	29.67	6.74	52.78
$p-value$		0.2345	0.0010	$<$ 0.0001	0.0094	$<$ 0.0001

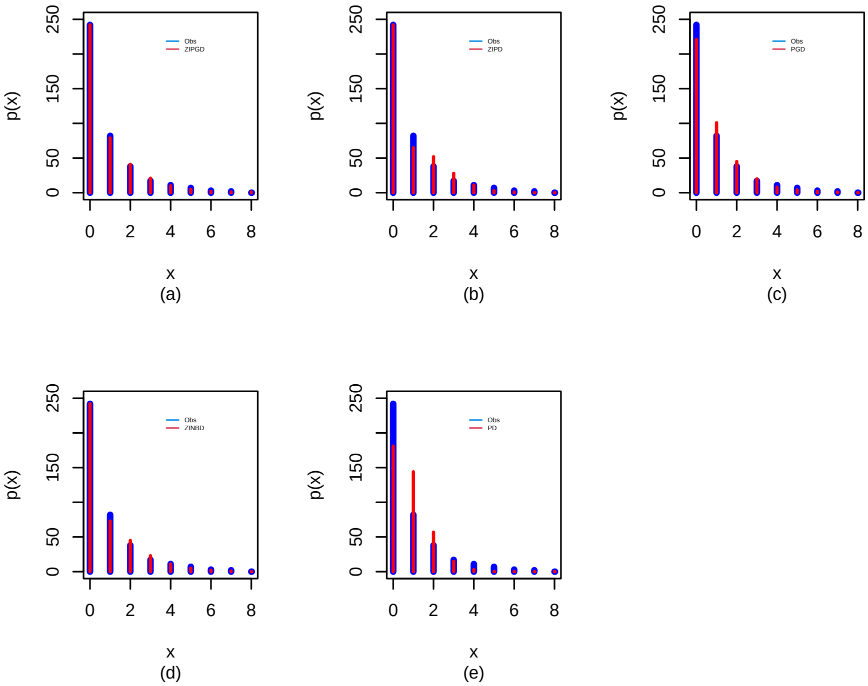

Figure 3 Frequency plots of fitted distributions with respect to ZIPGD(a), ZIPD(b), PGD(c), ZINBD(d), PD(e) for data set I

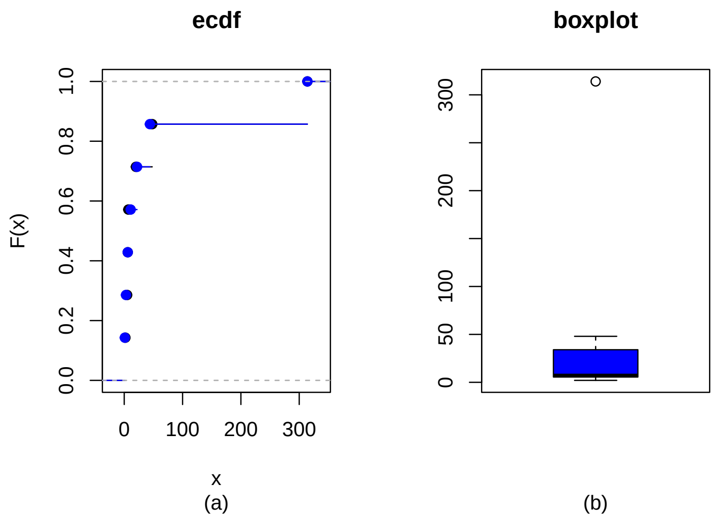

Figure 4 Empirical cumulative distribution function plot (ecdf) (a) and boxplot (b) for data set I

Table 7 AIC and BIC values for fitted distributions

Criterion	$ZIPGD$	$ZIPD$	PGD	$ZINBD$	PD
$\log l$	-332.75	-341.30	-355.65	-334.81	-411.89
AIC	669.50	686.60	713.31	675.62	825.78
BIC	677.49	694.59	717.30	687.61	829.77

7.2. Data Set II

The second data set is regarding the distribution of number of households on the basis of number of migrants (for details see Shukla and Yadava (2006)). The data set is given in Table 8. For all the fitted models, we obtained the expected frequencies and the chi-square statistic along with the respective p-values. The results are shown in Table 9. From Table 9, it is apparent that our proposed model has the highest p-value among all the competing models, suggesting that our proposed model fits better than other fitted models on the given data set. Also, for the model comparison, we computed AIC and BIC for all the fitted models and the results are presented in Table 10. Since the model that has the least values of AIC and BIC is considered to fit best among the group of competing models on the given data set. It can be observed that our proposed model has the least values of AIC and BIC, proving the claim that our proposed model performs better than other competing models on this data set.

Table 8 Number of households

No. of Households	0	1	2	3	4	5	6	7	8
Observed Counts	242	82	38	17	11	7	3	2	0

Table 9 Expected frequencies and χ^2 values of the fitted models

Number of Households(X)	Observed Counts	$ZIPGD$	$ZIPD$	PGD	$ZINBD$	PD
0	242	242	242	221	242	181
1	82	79	65	101	73	144
2	38	41	52	45	45	57
3	17	21	28	20	23	15
4	11	10	11	8	11	3
5	7	5	3	4	5	0
6	3	2	1	2	2	0
7	2	1	0	1	1	0
8	0	1	0	0	0	0
Degrees of Freedom		2	2	2	1	1
ML Estimates		$\hat{\alpha}$=0.2312	$\hat{\lambda}$=1.5936	$\hat{\zeta}$=1.6074	$\hat{p}$=0.6919	$\hat{\lambda}$=0.7960
		$\hat{\theta}$=1.2619	$\hat{\phi}$=0.5004		$\hat{r}$=3.0000	
					$\hat{a}$=0.4054	
χ^2		2.20	16.80	11.81	5.76	80.47
p-value		0.532	< 0.0001	0.018	0.056	0.00

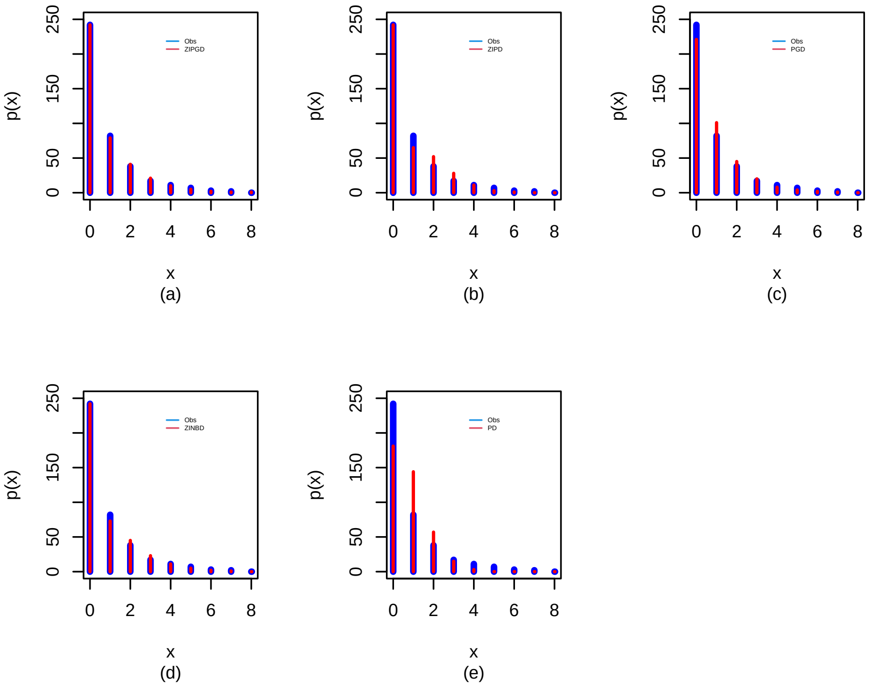

Figure 5 Frequency plots of fitted distributions with respect to ZIPGD(a), ZIPD(b), PGD(c), ZINBD(d), PD(e) for data set II.

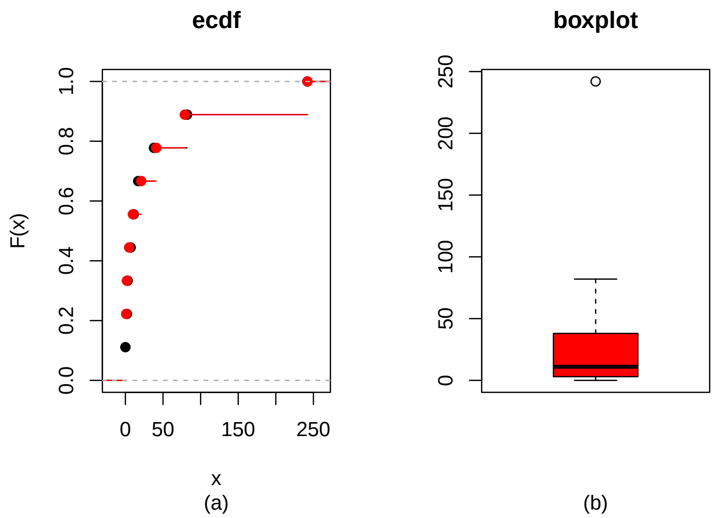

Figure 6 Empirical cumulative distribution function(ecdf) plot (a) and boxplot (b) for data set II

Table 10 AIC and BIC values for fitted distributions

Criterion	$ZIPGD$	$ZIPD$	PGD	$ZINBD$	PD
$\log l$	-492.05	-501.80	-497.02	-493.81	-555.06
AIC	988.10	1007.59	996.04	993.62	1112.12
BIC	996.09	1015.59	1000.04	1005.61	1116.12

8. Conclusion

In this study, a zero-inflated version of the Poisson Garima distribution is introduced, and various distributional properties of the newly developed model are discussed in detail. For parametric estimation, the maximum likelihood estimation is used and an extensive simulation study was carried out to study the finite sample behaviour of the maximum likelihood estimators. It was revealed by the simulation study that the maximum likelihood estimates for the parameters of proposed model are consistent and therefore can be used for fitting purposes. The practical applicability of the proposed model was validated by fitting the proposed model along with other famous probability distributions to two real-life data sets and the results suggested that the proposed model outperforms other competing models. Therefore, when dealing with the data that exhibits over proportion of zeros, the proposed model namely zero-inflated Poisson Garima distribution may be an important alternative to the already existing models and may provide better fitting results.

References

Altun E. A new zero-inflated regression model with application. statistikiler Dergisi: statistik ve Akterya. 2018; 11(2): 73-80.

Bar-Lev SK, Ridder A. Exponential dispersion models for overdispersed zero-inflated count data. Commun Stat Simul Comput. 2023; 52(7): 3286-3304.

Bekalo DB, Kebede DT. Zero-inflated models for count data: an application to number of antenatal care service visits. Ann Data Sci. 2021; 8(4): 683-708.

Cohen Jr AC. Estimating the parameters of a modified Poisson distribution. J Am Stat Assoc. 1960; 55(289): 139-143.

da Silva WB, Ribeiro AM, Conceio KS, Andrade MG, Neto FL. On zero-modified Poisson-Sujatha distribution to model overdispersed count data. Aust J Stat. 2018; 47(3): 1-19.

Gupta PL, Gupta RC, Tripathi RC. Analysis of zero-adjusted count data. Comput Stat Data Anal. 1996; 23(2): 207-218.

Johnson WD, Burton JH, Beyl RA, Romer JE. A simple chi-square statistic for testing homogeneity of zero-inflated distributions. Open J Stat. 2015; 5(6): 483-493.

Junnumtuam S, Niwitpong SA, Niwitpong S. Bayesian computation for the parameters of a zero-Inflated cosine geometric distribution with application to COVID-19 pandemic data. Comp Modeling Eng Sci. 2023; 135(2): 1229-1254

Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992; 34(1): 1-14.

Morgan BT, Palmer KJ, Ridout MS. Negative score test statistic. Am Stat. 2007; 61(4): 285-288.

Neyman J. On a new class of ”contagious” distributions, applicable in entomology and bacteriology. Ann Math Stat. 1939; 10(1): 35-57.

R Core Team. R: A language and environment for statistical computing. R Foundation for stat Comput, Vienna, Austria. URL https: //www. R-project. org/. [Accessed. 2023 Jun; 12].

Rahman T, Hazarika PJ, Barman MP. One inflated binomial distribution and its real-life applications. Indian J Sci Technol. 2021; 14(22): 1839-1854.

Rahman T, Hazarika PJ, Ali MM, Barman MP. Three-Inflated Poisson distribution and its application in suicide cases of India during Covid-19 pandemic. Ann Data Sci. 2022; 9(5): 1103-1127.

Ridout M, Demtrio CG, Hinde J. Models for count data with many zeros. In Proce XIXth int biom conf 1998 Dec 14; 19: 179-192. Proceedings of the XIXth international biometric conference. Cape Town, South Africa: International Biometric Society Invited Papers.

Rivas L, Campos F. Zero inflated waring distribution. Commun Stat Simul Comput. 2023; 52(8): 3676-3691.

Sadeghkhani A, Ahmed SE. The application of predictive distribution estimation in multiple-inflated Poisson models to ice hockey data. Model Assist Stat Appl. 2020; 15(2): 127-137.

Sandhyaa E, Abrahamb TL. Inflated-parameter Harris distribution. J Math Comp Sci. 2016; 16(1): 33-49.

Sellers KF, Raim A. A flexible zero-inflated model to address data dispersion. Comput Stat Data Anal. 2016; 99: 68-80.

Simmachan T, Wongsai N, Wongsai S, Lerdsuwansri R. Modeling road accident fatalities with underdispersion and zero-inflated counts. PLoS one. 2022; 17(11): 1-23.

Shanker R. The discrete Poisson-Garima distribution. Biom Biostat Inter J. 2017; 5(2): 48-53.

Shukla KK, Yadava KN. The distribution of the number of migrants at the household level. J Popul Soc Stud. 2006; 14(2): 153-166.

Singh S. Note on inflated Poisson-distribution. Ann Math Stat 1962; 33(3):1203-1210.

Tawiah K, Iddrisu WA, Asampana Asosega K. Zero-inflated time series modelling of COVID-19 deaths in Ghana. J Environ Public Health. 2021; https://doi.org/10.1155/2021/5543977.

Wani MK, Ahmad PB. Zero-inflated Poisson-Akash distribution for count data with excessive zeros. J Korean Stat Soc. 2023;52(3):647-675.