Journal of Statistics *ISSN-P: 1684–8403*
Volume 29, 2025, pp. 110–127 *ISSN-E: 2957–8981*

Zero-Inflated Poisson XLindley Distribution for Medical Science Modeling

Muhammad Ahsan-ul-Haq[1*], *Muhammad Nasir Saddam Hussain*[2], *Junaid Talib*[3]
Saadia Tariq[4]

Abstract

This paper introduces and investigates a new one-parameter zero-inflated count distribution. The new model is named the zero-inflated Poisson XLindley (ZIPXL) distribution. The fundamental mathematical characteristics of the ZIPXL model–including survival analysis, hazard function, generating functions, moments (mean and variance), dispersion index, skewness coefficient, kurtosis, and order statistics—are derived. Maximum likelihood is used to estimate the parameters of the ZIPXL distribution. An intensive simulation study is conducted to assess the performance of these estimators. The research demonstrates the practical utility and flexibility of the new distribution in managing excess zero data in real-world applications, using two real-world datasets from the medical field. The research compares the ZIPXL distribution with the zero-inflated Poisson moment exponential distribution and the zero-inflated Poisson distribution. Provides evidence that the ZIPXL distribution performs effectively in examining overdispersed count data.

Keywords: Poisson XLindley distribution, Overdispersion, Zero inflation, Count data.

1. Introduction

Count data modeling is applicable in various domains, including medical sciences, engineering, agronomy, economics, and weather forecasting. Over the past few decades, it has been observed that the Poisson distribution is the most widely used one-parameter probability distribution for analyzing count data across the different fields mentioned above. Count data, on the other hand, frequently demonstrates greater variance than the predicted hypothesized model, a fact explained by the idea of statistical dispersion. As the Poisson distribution holds an important property of mean-variance equality, it cannot be used for over-dispersed data sets. Overdispersion exists for a variety of causes; however, in certain circumstances, the cause of overdispersion is an excessive amount of zeros in the dataset, often

*Corresponding author
[1]College of Statistical Sciences, University of the Punjab, Lahore, Pakistan.
Email: ahsanshani36@gmail.com
[2]Department of Statistics, Govt. Murray Graduate College Sialkot, Pakistan.
Email: iqbalnasir945@gmail.com
[3]School of Statistics, Minhaj University Lahore, Pakistan.
Email: junaidtalib164@gmail.com
[4]School of Statistics, Minhaj University Lahore, Pakistan.
Email: sadiasajad100@gmail.com

known as zero-inflated data. To get rid of this challenge, researchers move towards zero-inflated models, which offer a more suitable alternative to conventional discrete distributions. Illustrative examples of zero-inflated data include the count of fetal movements per five seconds (Leroux & Puterman, 1992), the enumeration of HIV-infected patients (Van den Broek, 1995), household-level migrant counts (Shukla & Yadava, 2006), accidents attributed to heavy vehicular traffic in 2010 (Sharma & Landge, 2013), and suicide cases related to COVID-19 in India (Rahman et al., 2022).

To effectively model the zero-inflated count data sets, various zero-inflated discrete models are developed and studied. For example, Zamani et al. (2023) developed a zero-inflated Poisson quasi-Lindley distribution, investigated its statistical properties, and applied it to data from the US National Medical Expenditure Survey. Aryuyuen et al. (2014) investigated the zero-inflated negative binomial-generalized exponential distribution and utilized it to analyze two datasets: one from the medical industry and the other from consumer goods. Skinder et al. (2023) developed a zero-inflated Poisson moment exponential distribution and applied it to three real datasets from distinct areas, including vaccine adverse event reports, HIV-exposed newborn data, and epileptic seizure counts. Junnumtuam et al. (2022) invented the zero-inflated-Cosine geometric distribution to characterize data with too many zeros. Several structural traits were identified, including the moment-generating function, mean, and variance. Additionally, the Wald approach was utilized to generate confidence intervals. The real confidence interval was estimated using the Bayesian approach with the highest posterior density. Sabri and Adetunji (2023) proposed the zero-inflated Poisson transmuted weighted exponential distribution. They examined the suggested model's probabilistic and reliability features. The parameters of the novel zero-inflated model were determined using the maximum likelihood method. The model's applicability was shown with five real-world datasets. Wani and Ahmad (2023) proposed a zero-inflated Poisson-Akash distribution to represent the dataset with excess zeros. Various statistical features of the aforementioned model were exclusively investigated.

2. The Poisson XLindley distribution

Ahsan-ul-Haq et al. (2022) introduced the Poisson XLindley (PXL) distribution, formulated by compounding the XLindley and Poisson distributions using a compounding technique. The PXL distribution is an important discrete model because it provides a versatile and reliable tool for modeling count data, particularly when dealing with positively skewed and leptokurtic distributions. It adequately handles equi- and over-dispersed phenomena, making it appropriate for a wide range of applications in many domains. The probability mass function (pmf) of the PXL distribution with random variable Z is as follows:

$$P(Z = z, \delta) = \frac{\delta^2 \left(z + \delta^2 + 3(1+\delta)\right)}{(1+\delta)^{4+z}}, \tag{1}$$

where $z = 0, 1, 2, 3, \ldots$ and $\delta > 0$. The mean and variance of the PXL model are

$$\mathrm{E}[Z] = \frac{\delta^2 + 2\delta + 2}{\delta(1+\delta)^2},$$

and

$$\mathrm{Var}(Z) = \frac{\delta^5 + 5\delta^4 + 11\delta^3 + 14\delta^2 + 10\delta + 2}{\delta^2(1+\delta)^2},$$

respectively. Data patterns change over time due to various factors, such as the advancement of new technology, which has resulted in the rise of complicated count data sets, driven by advances in data gathering, storage, and processing. The conventional discrete models discussed earlier may not be best suited for these data sets. Therefore, there arises a need to innovate and build new zero-inflated models to properly address zero-inflated data sets. In this connection, this study aims to introduce a novel zero-inflated statistical model known as the ZIPXL distribution, along with its distributional properties and various other essential aspects. The suggested distribution is compared to other models, including the zero-inflated Poisson moment exponential (Skinder et al., 2023) and the zero-inflated Poisson distributions, using two real-life zero-inflated data sets.

The rest of the study is organized as follows: In Section 2, we describe the functional form of the Poisson XLindley distribution, a mixed model that can handle instances when variance far exceeds mean, making it ideal for modeling over-dispersed count data. Section 3 dives into the origins and analysis of the zero-inflated Poisson XLindley distribution. Section 4 goes into great detail on many statistical and dependability aspects. Section 5 discusses the parameter estimates for the new suggested zero-inflated model. Section 6 presents a carefully planned simulation investigation. Finally, in Section 7, the performance of the novel model was tested using two real-world data sets and compared to well-known zero-inflated models.

3. Zero-Inflated Poisson XLindley distribution

In zero-inflated models, it is hypothesized that there are two distinct elements for zero observations. It is presumed that the first element generates zeros exclusively, denoted as "structural zeros," with a probability ϕ and the second component produces counts from a count model, with a probability of $(1 - \phi)$ for "count observations." To explain this phenomenon, let's assume a random variable Y with a discrete distribution. The pmf of a random variable Z for a zero-inflated distribution can be expressed as follows:

In zero-inflated models, it is hypothesized that there are two distinct elements for zero observations. It is presumed that the first element generates zeros exclusively, denoted as "structural zeros," with a probability ϕ and the second component produces counts from a count model, with a probability of $(1 - \phi)$ for "count observations." To explain this phenomenon, let's assume a random variable Y with a count model. The pmf of a random variable Z for a zero-inflated model can be

expressed as follows:

$$P(Z=z)=\begin{cases}\phi+(1-\phi)p(0), & z=0\\ (1-\phi)p(z), & z=1,2,3,\ldots\end{cases} \tag{2}$$

where $p(z)$ represents the pmf of the random variable Z, and ϕ denotes the zero-inflation parameter $(0<\phi<1)$.

By combining equations (1) and (3), the PMF of the ZIPXL model is given as follows:

$$P(Z=z;\phi,\delta)=\begin{cases}\phi+(1-\phi)\dfrac{\delta^2(\delta^2+3(1+\delta))}{(1+\delta)^4}, & z=0\\ (1-\phi)\dfrac{\delta^2(z+\delta^2+3(1+\delta))}{(1+\delta)^{4+z}}, & z=1,2,3,\ldots\end{cases} \tag{3}$$

where ϕ signifies the probability associated with zero events, constrained within the interval $0<\phi<1$, and $\delta>0$.

Figure 1 displays pmf plots for the ZIPXL distribution, featuring various parameter combinations. It demonstrates that even when the proportion of zero counts (ϕ) is small (e.g., $\phi=0.1$), the probability of encountering zero values remains remarkably high. This validates the distribution's capacity to represent datasets with a significant percentage of zero counts, which is particularly valuable for situations where zeros are excessive. For instance, when $\phi=0.1$ and $\delta=1$, the

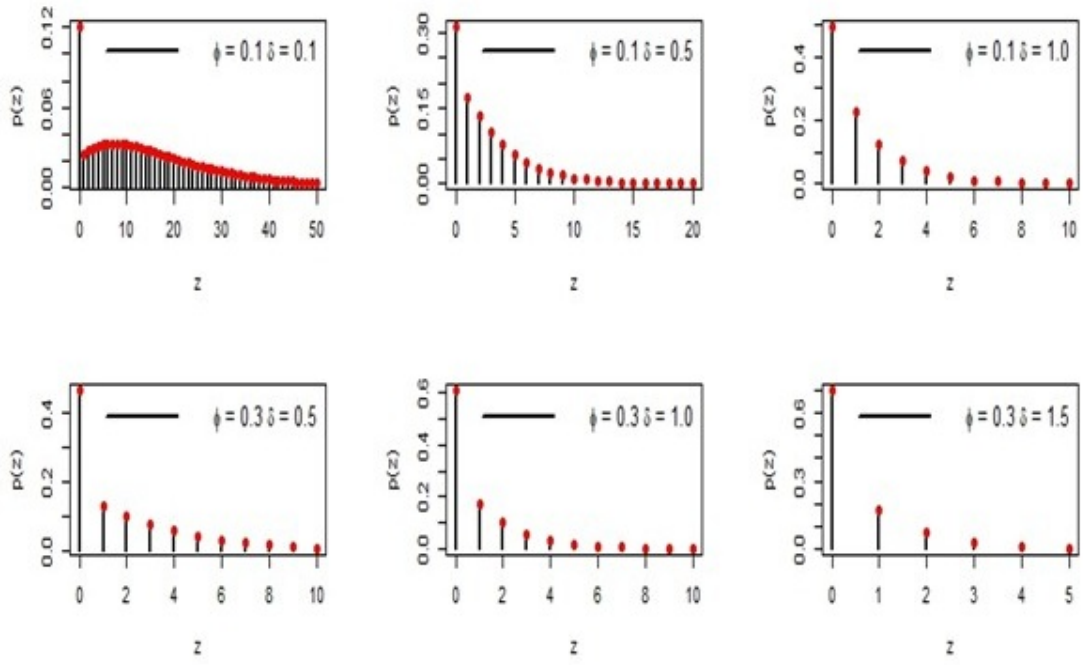

Figure 1: The pmf plots of ZIPXL distribution for different parameter values.

probability of observing $z=0$ surpasses 0.5. These pmf plots also reveal that the ZIPXL distribution consistently exhibits uni-modality, with the mode always

centered at zero. Beyond this, it proves to be a versatile choice for analyzing data characterized by right-skewness and zero inflation. Notably, as the zero-inflation parameter increases, the concentration of probability mass at zero becomes increasingly evident. Furthermore, higher values of the parameter δ result in a swift decline in the distribution's tail.

The cumulative distribution function (CDF) is given by

$$F(z;\delta,\phi) = \phi + \sum_{k=0}^{z}(1-\phi)\cdot\frac{\delta^2(k+\delta^2+3(1+\delta))}{(1+\delta)^{4+k}},$$

or equivalently,

$$F(z;\delta,\phi) = 1-(1-\phi)\cdot\frac{1+\delta(4+z+\delta(3+\delta))}{(1+\delta)^{4+z}}. \tag{4}$$

4. Statistical characteristics

The subsequent section of the study will examine ZIPXL distribution, systematically examining the reliability measures. Furthermore, the study explores various statistical dimensions such as probability-generating function, moment-generating function, moments, and dispersion index among others to obtain a thorough understanding of distribution attributes and dynamics.

4.1. Reliability function

The measure of probability that a system survives beyond a specific time, in the view of a distribution. It can be expressed in terms of a complement to the distribution function of the model. In the case of the ZIPXL distribution, the calculation of the reliability function or survival function unfolds as follows.

$$R(z;\delta,\phi) = P(Z>z) = (1-\phi)\cdot\frac{1+\delta(4+z+3\delta+\delta^2)}{(1+\delta)^{4+z}}.$$

4.2. Hazard rate

Consider a random sample $z_1, z_2, \ldots, z_n$ drawn from the ZIPXL distribution as described by equation (4). Let Y be the number of z_i taking the value zero. Equation (4) can thus be reformulated in the following manner:

$$P(Z=z;\delta,\phi) = \left[\phi+(1-\phi)\cdot\frac{\delta^2(\delta^2+3+3\delta)}{(1+\delta)^4}\right]^Y\left[(1-\phi)\cdot\frac{\delta^2(z+\delta^2+3+3\delta)}{(1+\delta)^{4+z}}\right]^{1-Y}.$$

Now, using $R(z)$ from equation (5), the hazard rate function of the ZIPXL distri-

bution is given by:

$$h(z) = \frac{\left[\phi+(1-\phi)(1+\delta)^{-4}(\delta^4+3\delta^2+3\delta^3)\right]^Y\left[(1-\phi)(1+\delta)^{-(4+z)}(z\delta^2+\delta^4+3\delta^2+3\delta^3)\right]^{1-Y}}{(1-\phi)(1+\delta)^{-(4+z)}(1+4\delta+z\delta+3\delta^2+3\delta^3)}.$$

4.3. Reverse hazard rate

The reverse hazard rate function $r^*(z)$ of the ZIPXL distribution can be expressed as:

$$r^*(z) = \frac{(1+\delta)^{z+4}\left[\phi+(1-\phi)(1+\delta)^{-4}(\delta^4+3\delta^2+3\delta^3)\right]^Y\left[(1-\phi)(1+\delta)^{-(z+4)}(z\delta^2+\delta^4+3\delta^2+3\delta^3)\right]^{1-Y}}{(1+\delta)^{4+z}-(1-\phi)[1+\delta(4+z+3\delta+\delta^2)]}.$$

4.4. Cumulative hazard rate

The cumulative hazard rate function $H(z)$ of the ZIPXL distribution is given by:

$$H(z) = -\ln\left[(1-\phi)\cdot(1+\delta)^{-(z+4)}\left(1+4\delta+z\delta+3\delta^2+3\delta^3\right)\right].$$

4.5. Mills ratio

The Mills ratio of a probability distribution is the ratio of its survival function to its probability mass function. Formally, for a ZIPXL random variable Z, the Mills ratio is defined as:

$$M(z) = \frac{(1-\phi)\left(1+4\delta+z\delta+3\delta^2+3\delta^3\right)(1+\delta)^{-(z+4)}}{[\phi+(1-\phi)(1+\delta)^{-4}(\delta^4+3\delta^2+3\delta^3)]^Y\left[(1-\phi)(1+\delta)^{-(z+4)}(z\delta^2+\delta^4+3\delta^2+3\delta^3)\right]^{1-Y}}$$

4.6. Generating functions and moments

The probability-generating function (PGF) of the ZIPXL distribution can be obtained as:

$$P_Z(t) = \sum_{z=0}^{\infty} t^z P(Z=z),$$

which can be expressed as:

$$\begin{aligned} P_Z(t) &= \phi + \sum_{z=0}^{\infty} t^z(1-\phi)\frac{\delta^2(z+\delta^2+3+3\delta)}{(1+\delta)^{4+z}} \\ &= \phi + \frac{(1-\phi)\delta^2}{(1+\delta)^4}\sum_{z=0}^{\infty} t^z\frac{z+\delta^2+3+3\delta}{(1+\delta)^z} \end{aligned}$$

$$
\begin{aligned}
&= \phi + \frac{(1-\phi)\delta^2}{(1+\delta)^4} \sum_{z=0}^{\infty} \left(\frac{t^z z}{(1+\delta)^z} + \frac{t^z(\delta^2+3+3\delta)}{(1+\delta)^z} \right) \\
&= \phi + \frac{(1-\phi)\delta^2}{(1+\delta)^4} \left(\sum_{z=0}^{\infty} z \left(\frac{t}{1+\delta} \right)^z + (\delta^2+3+3\delta) \sum_{z=0}^{\infty} \left(\frac{t}{1+\delta} \right)^z \right) \\
&= \phi + \frac{(1-\phi)\delta^2}{(1+\delta)^4} \cdot \frac{(1+\delta)^2 \left(3 - t(2+\delta) + 3\delta + \delta^2\right)}{(1-t+\delta)^2}
\end{aligned}
$$

thus simplifying to:

$$
P_Z(t) = \phi + \frac{(1-\phi)\delta^2}{(1+\delta)^2} \frac{3 + 3\delta + \delta^2 - t(2+\delta)}{(1-t+\delta)^2}.
$$

Similarly, the moment generating function (MGF) of the ZIPXL distribution can be expressed as:

$$
M_Z(t) = \phi + \frac{(1-\phi)\delta^2}{(1+\delta)^2} \frac{3 + 3\delta + \delta^2 - e^t(2+\delta)}{(1-e^t+\delta)^2}.
$$

The first four moments about the origin are given by:

$$
m_1' = \frac{(1-\phi)(\delta^2+2\delta+2)}{\delta(1+\delta)^2},
$$

$$
m_2' = \frac{(1-\phi)(\delta^3+4\delta^2+6\delta+6)}{\delta^2(1+\delta)^2},
$$

$$
m_3' = \frac{(1-\phi)(\delta^4+8\delta^3+20\delta^2+30\delta+24)}{\delta^3(1+\delta)^2},
$$

$$
m_4' = \frac{(1-\phi)(\delta^5+16\delta^4+66\delta^3+138\delta^2+192\delta+120)}{\delta^4(1+\delta)^2}.
$$

Using moments about the origin, the central moments can be written as:

$$
\mu_2 = \frac{(1-\phi)\left(\delta^4\phi + 4\delta^3\phi + 8\delta^2\phi + 8\delta\phi + 4\phi + \delta^5 + 5\delta^4 + 11\delta^3 + 14\delta^2 + 10\delta + 2\right)}{\delta^2(1+\delta)^4},
$$

$$
\begin{aligned}
\mu_3 = \frac{(1-\phi)}{\delta^3(1+\delta)^6} \Bigg(& 4 + 30\delta + 98\delta^2 + 152\delta^3 + 141\delta^4 + 87\delta^5 + 36\delta^6 + 9\delta^7 + \delta^8 + 4\phi + 48\delta\phi \\
& + 114\delta^2\phi + 148\delta^3\phi + 120\delta^4\phi + 63\delta^5\phi + 20\delta^6\phi + 3\delta^7\phi + 16\phi^2 + 48\delta\phi^2 \\
& + 72\delta^2\phi^2 + 64\delta^3\phi^2 + 36\delta^4\phi^2 + 12\delta^5\phi^2 + 2\delta^6\phi^2 \Bigg),
\end{aligned}
$$

$$\mu_4 = \frac{(1-\phi)}{\delta^4(1+\delta)^8}\Bigg(24 + 240\delta + 1010\delta^2 + 2430\delta^3 + 3678\delta^4 + 3747\delta^5 + 2692\delta^6 + 1395\delta^7 + 515\delta^8 + 127\delta^9 + 18\delta^{10} + \delta^{11} + 48\phi + 336\delta\phi + 1168\delta^2\phi + 2280\delta^3\phi + 2872\delta^4\phi + 2512\delta^5\phi + 1580\delta^6\phi + 716\delta^7\phi + 225\delta^8\phi + 44\delta^9\phi + 4\delta^{10}\phi + 144\delta\phi^2 + 528\delta^2\phi^2 + 984\delta^3\phi^2 + 1164\delta^4\phi^2 + 948\delta^5\phi^2 + 540\delta^6\phi^2 + 210\delta^7\phi^2 + 51\delta^8\phi^2 + 6\delta^9\phi^2 + 48\phi^3 + 192\delta\phi^3 + 384\delta^2\phi^3 + 480\delta^3\phi^3 + 408\delta^4\phi^3 + 240\delta^5\phi^3 + 96\delta^6\phi^3 + 24\delta^7\phi^3 + 3\delta^8\phi^3\Bigg).$$

The following values for the dispersion index (DI), coefficient of variation (CV), coefficient of skewness (CS), and coefficient of kurtosis (CK) for the ZIPXL distribution are, respectively, derived as:

$$DI = \frac{\delta^4\phi + 4\delta^3\phi + 8\delta^2\phi + 8\delta\phi + 4\phi + \delta^5 + 5\delta^4 + 11\delta^3 + 14\delta^2 + 10\delta + 2}{\delta(1+\delta)^2(\delta^2 + 2\delta + 2)},$$

$$CV = \sqrt{\frac{(1-\phi)(\delta^4\phi + 4\delta^3\phi + 8\delta^2\phi + 8\delta\phi + 4\phi + \delta^5 + 5\delta^4 + 11\delta^3 + 14\delta^2 + 10\delta + 2)}{(\delta^2 + 2\delta + 2)(1-\phi)}},$$

$$CS = \frac{\mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3}{(\mu_2' - (\mu_1')^2)^{3/2}},$$

$$CK = \frac{\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4}{(\mu_2' - (\mu_1')^2)^2}.$$

The above-derived measures including mean-variance, DI, CV, skewness, and kurtosis are computed numerically and presented in Table 1.

Examining the table above reveals that the dispersion index exceeds unity across multiple parameter combinations, indicating that the suggested model is overdispersed. Skewness analysis demonstrates a constant rightward skew, with higher skewness values recorded for various parameter combinations. Furthermore, the ZIPXL model is leptokurtic, with kurtosis greater than three across a wide range of parameter combinations.

4.7. Order statistics

Given a sample of n independently and identically distributed random variables from a distribution with CDF $F(z)$ and PMF $P(Z = z)$, the cumulative distribution function (CDF) of the l^{th} order statistic $Z_{(l)}$ is given by:

$$F_{Z_{(l)}}(z) = \sum_{j=l}^{n} \binom{n}{j} [F(z)]^j [S(z)]^{n-j},$$

Table 1: Numerical values of descriptive measures using different combinations of parameters.

ϕ	δ	Mean	Variance	DI	CV	Skewness	Kurtosis
0.1	0.1	16.4380	223.7503	13.6118	0.9100	1.4227	5.9380
	0.5	2.6000	9.4400	3.6308	1.1817	1.8271	7.6632
	1.0	1.1250	2.5594	2.2750	1.4220	2.1036	9.1793
	1.5	0.6960	1.2676	1.8212	1.6176	2.2760	10.1643
	2.0	0.5000	0.8000	1.6000	1.7889	2.4108	10.9180
	3.0	0.3188	0.4421	1.3871	2.0861	2.6396	12.1883
0.3	0.1	12.7851	220.7307	17.2646	1.1621	1.5808	6.2670
	0.5	2.0222	8.5106	4.2085	1.4426	2.1047	8.9038
	1.0	0.8750	2.2094	2.5250	1.6987	2.4529	11.1333
	1.5	0.5413	1.0696	1.9759	1.9105	2.6680	12.6094
	2.0	0.3889	0.6654	1.7111	2.0976	2.8316	13.7360
	3.0	0.2479	0.3615	1.4580	2.4251	3.0996	15.5798
0.5	0.1	9.1322	191.0238	20.9175	1.5134	1.9877	7.8636
	0.5	1.4444	6.9136	4.7863	1.8203	2.5931	11.7246
	1.0	0.6250	1.7344	2.7750	2.1071	3.0117	15.0297
	1.5	0.3867	0.8238	2.1306	2.3474	3.2749	17.2765
	2.0	0.2778	0.5062	1.8222	2.5612	3.4744	19.0058
	3.0	0.1771	0.2707	1.5288	2.9382	3.7956	21.8073
0.7	0.1	5.4793	134.6297	24.5706	2.1176	2.7964	12.5692
	0.5	0.8667	4.6489	5.3639	2.4877	3.5245	18.9805
	1.0	0.3750	1.1344	3.0251	2.8402	4.0523	24.5978
	1.5	0.2320	0.5302	2.2853	3.1386	4.3926	28.5174
	2.0	0.1667	0.3222	1.9328	3.4051	4.6527	31.5784
	3.0	0.1063	0.1700	1.5992	3.8787	5.0653	36.4993
0.9	0.1	1.8264	51.5484	28.2240	3.9311	5.2960	37.9624
	0.5	0.2889	1.7165	5.9415	4.5350	6.4523	56.7135
	1.0	0.1250	0.4094	3.2752	5.1187	7.3278	73.5441
	1.5	0.0773	0.1887	2.4411	5.6196	7.9078	85.5853
	2.0	0.0556	0.1136	2.0432	6.0620	8.3521	95.0660
	3.0	0.0354	0.0592	1.6723	6.8732	9.0583	110.3840

where $S(z) = 1 - F(z)$ is the survival function. Based on Eq. (4), the CDF of the l^{th} order statistic from the ZIPXL distribution can be expressed as:

$$F_{Z_{(l)}}(z) = \sum_{j=l}^{n} \binom{n}{j} \left\{1 - (1-\phi)(1+\delta)^{-(4+z)} \left(1 + 4\delta + z\delta + \delta^2(3+\delta)\right)\right\}^j \times \left\{(1-\phi)(1+\delta)^{-(4+z)} \left(1 + 4\delta + z\delta + \delta^2(3+\delta)\right)\right\}^{n-j}.$$

The probability mass function (PMF) for the l^{th} order statistic of a discrete model

is given by:

$$f_{Z_{(l)}}(z) = \frac{n!}{(l-1)!(n-l)!}[F(z)]^{l-1}[S(z)]^{n-l}P(Z=z).$$

Utilizing Eqs. (3) and (4), the PMF corresponding to the l^{th} order statistic of the ZIPXL distribution can be derived as:

$$f_{Z_{(l)}}(z) = \frac{n!}{(l-1)!(n-l)!}\left[\left\{1-(1-\phi)(1+\delta)^{-(4+z)}\left(1+4\delta+z\delta+\delta^2(3+\delta)\right)\right\}^{l-1}\right.$$
$$\left.\times\left\{(1-\phi)(1+\delta)^{-(4+z)}\left(1+4\delta+z\delta+\delta^2(3+\delta)\right)\right\}^{n-l}P(Z=z)\right].$$

The minimum and maximum order statistics of the ZIPXL distribution can be obtained by substituting $l=1$ and $l=n$, respectively, into the general expression for the PMF of the l^{th} order statistic.

5. Parameter estimation

In this segment, we explore the parametric estimation of the ZIPXL distribution through the utilization of the maximum likelihood (ML) estimation (MLE) method.

Let $z_1, z_2, \ldots, z_n$ be a random sample drawn from the ZIPXL distribution as defined in equation (2), and this holds for each $i = 1, 2, 3, \ldots, n$. Define the indicator variable:

$$t_i = \begin{cases} 1 & \text{if } z_i = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Accordingly, for $i = 1, 2, \ldots, n$, equation (1) takes the following form:

$$P(Z=z_i) = \left[\phi+(1-\phi)\frac{\delta^4+3\delta^2+3\delta^3}{(1+\delta)^4}\right]^{t_i}\left[(1-\phi)\frac{\delta^2 z_i+\delta^4+3\delta^2+3\delta^3}{(1+\delta)^{4+z_i}}\right]^{1-t_i}.$$

Using the above equation, the likelihood function is:

$$l(\phi,\delta \mid z) = \prod_{i=1}^{n}\left[\phi+(1-\phi)\frac{\delta^4+3\delta^2+3\delta^3}{(1+\delta)^4}\right]^{t_i}\left[(1-\phi)\frac{\delta^2 z_i+\delta^4+3\delta^2+3\delta^3}{(1+\delta)^{4+z_i}}\right]^{1-t_i}.$$

Let L denote the log-likelihood function. It can be expressed as:

$$L = \log\left\{\left[\phi+(1-\phi)\frac{\delta^4+3\delta^2+3\delta^3}{(1+\delta)^4}\right]^{\sum_{i=1}^{n}t_i}\left[(1-\phi)\frac{\delta^2 z_i+\delta^4+3\delta^2+3\delta^3}{(1+\delta)^{4+z_i}}\right]^{n-\sum_{i=1}^{n}t_i}\right\}.$$

Define $n_0 = \sum_{i=1}^{n} t_i$, the number of zeros in the sample. Then the log-likelihood simplifies to:

$$L = n_0 \log\left(\phi + (1-\phi)\frac{\delta^4 + 3\delta^2(1+\delta)}{(1+\delta)^4}\right) + (n - n_0)\log\left((1-\phi)\frac{\delta^2 z_i + \delta^4 + 3\delta^2 + 3\delta^3}{(1+\delta)^{4+z_i}}\right).$$

The partial derivatives of the log-likelihood with respect to the parameters are:

$$\frac{\partial \log L}{\partial \phi} = \frac{n_0\left((1+\delta)^4 - \delta^2(\delta^2 + 3(1+\delta))\right)}{\phi(1+\delta)^4 + \delta^2(1-\phi)(\delta^2 + 3(1+\delta))} - \frac{n - n_0}{1-\phi}, \tag{5}$$

and

$$\begin{aligned}\frac{\partial \log L}{\partial \delta} = {} & \frac{n_0(1+\delta)^4\delta^3(4 + 9\delta(5+\delta))(1-\phi)}{\left(\phi(1+\delta)^4 + \delta^2(1-\phi)(\delta^2 + 3(1+\delta))\right)(1+\delta)^5} \\ & + (n - n_0)\frac{\delta\left[2(z_i+3) - (z_i - 3)\delta^2 + z_i\delta^3 + \delta(3 + z_i(5+z_i))\right]}{\delta^2(1+\delta)(z_i + \delta^2 + 3(1+\delta))}.\end{aligned} \tag{6}$$

Since Eqs. (5) and (6) lack closed-form solutions, they can be solved numerically.

6. Simulation Study

This section presents a comprehensive Monte Carlo simulation study to assess the performance of the maximum likelihood (ML) estimators for the ZIPXL distribution. The simulation procedure involves generating 10,000 independent samples of varying sizes, specifically $n = 25, 50, 100, 200$, and 300.

The simulation results, including the average estimates (Avg.E), absolute biases (AB), and mean square errors (MSE), are presented in Table 2. Additionally, Figure 2 illustrates the MSEs based on all combinations of parameters.

The performance metrics are defined as follows:

$$\text{Avg.E} = \frac{1}{N}\sum_{i=1}^{N}\hat{\omega}_i, \qquad \text{AB} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{\omega}_i - \omega\right|, \qquad \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\omega}_i - \omega\right)^2,$$

where $N = 10{,}000$, $\hat{\omega}_i$ denotes the estimate of the parameter ω in the i-th simulation, and ω is the true value of the parameter.

Table 2 presents the simulation results for the ZIPXLD model, showing the performance of parameter estimates across varying sample sizes and different combinations of parameter values. The table reports the average estimates (Avg.Es), absolute biases (ABsB), and mean squared errors (MSEs) for both parameters δ and ϕ, evaluated at nine distinct parameter settings.

Table 2: Simulation results for different combinations of true parameter values (ϕ, δ).

ϕ	δ	n	$\hat{\delta}$			$\hat{\phi}$		
			Avg.E	ABsB	MSEs	Avg.E	ABsB	MSEs
0.1	0.1	25	0.1030	0.0030	0.0003	0.1009	0.0009	0.0044
		50	0.1010	0.0010	0.0001	0.1010	0.0010	0.0022
		75	0.1007	0.0007	0.0001	0.0996	0.0004	0.0011
		100	0.1004	0.0004	0.0000	0.1000	0.0000	0.0006
		200	0.1003	0.0003	0.0000	0.0998	0.0002	0.0004
	0.5	25	0.5181	0.0181	0.0146	0.1067	0.0067	0.0117
		50	0.5089	0.0089	0.0063	0.0999	0.0001	0.0069
		75	0.5044	0.0044	0.0031	0.0981	0.0019	0.0041
		100	0.5038	0.0038	0.0016	0.0977	0.0023	0.0023
		200	0.5016	0.0016	0.0010	0.0991	0.0009	0.0016
	1.0	25	1.0444	0.0444	0.0976	0.1256	0.0256	0.0227
		50	1.0166	0.0166	0.0458	0.1147	0.0147	0.0140
		75	1.0104	0.0104	0.0214	0.1036	0.0036	0.0086
		100	1.0087	0.0087	0.0115	0.0990	0.0010	0.0054
		200	1.0059	0.0059	0.0080	0.0977	0.0023	0.0039
0.5	0.1	25	0.1057	0.0057	0.0007	0.4985	0.0015	0.0106
		50	0.1026	0.0026	0.0003	0.5000	0.0000	0.0054
		75	0.1013	0.0013	0.0001	0.4997	0.0003	0.0026
		100	0.1004	0.0004	0.0001	0.5008	0.0008	0.0013
		200	0.1004	0.0004	0.0000	0.5001	0.0001	0.0009
	0.5	25	0.5698	0.0698	0.0869	0.4815	0.0185	0.0211
		50	0.5292	0.0292	0.0171	0.4891	0.0109	0.0102
		75	0.5132	0.0132	0.0064	0.4955	0.0045	0.0047
		100	0.5056	0.0056	0.0031	0.4980	0.0020	0.0023
		200	0.5042	0.0042	0.0020	0.4984	0.0016	0.0016
	1.0	25	1.3625	0.3625	1.7993	8.0612	7.5612	3.2491
		50	1.1252	0.1252	0.2301	0.4693	0.0307	0.0259
		75	1.0564	0.0564	0.0688	0.4842	0.0158	0.0117
		100	1.0281	0.0281	0.0272	0.4912	0.0088	0.0054
		200	1.0181	0.0181	0.0171	0.4942	0.0058	0.0034
0.8	0.1	25	0.1385	0.0385	0.5029	1.1795	0.3795	5.3509
		50	0.1070	0.0070	0.0010	0.7983	0.0017	0.0033
		75	0.1032	0.0032	0.0004	0.7994	0.0006	0.0016
		100	0.1014	0.0014	0.0002	0.8004	0.0004	0.0008
		200	0.1010	0.0010	0.0001	0.8003	0.0003	0.0005
	0.5	25	1.1883	0.6883	11.0880	6.6822	6.1822	2.8623
		50	0.6606	0.1606	2.9467	2.9109	2.4109	1.9702
		75	0.5363	0.0363	0.0228	0.7961	0.0039	0.0025

Continued on next page...

Continued on from previous page...

ϕ	δ	n	$\hat{\delta}$			$\hat{\phi}$		
			Avg.E	ABsB	MSEs	Avg.E	ABsB	MSEs
		100	0.5171	0.0171	0.0088	0.7974	0.0026	0.0013
		200	0.5112	0.0112	0.0057	0.7983	0.0017	0.0009
		25	3.5146	2.5146	46.8740	2.8947	2.0947	1.5071
		50	2.1813	1.1813	25.3180	1.5809	0.7809	0.0996
	1.0	75	1.2147	0.2147	0.7105	0.7796	0.0204	0.0082
		100	1.0801	0.0801	0.1015	0.7909	0.0091	0.0028
		200	1.0479	0.0479	0.0522	0.7945	0.0055	0.0017

As observed, the average estimates for both parameters tend to approach their respective true values as the sample size increases, indicating consistency of the estimators. Furthermore, both the absolute bias and MSE decrease with increasing sample size across all parameter configurations, which demonstrates the improved estimation accuracy and efficiency with larger samples.

It is also noteworthy that, when the value of ϕ is fixed and δ increases, both the absolute bias and MSE generally exhibit an increasing trend. This suggests that the accuracy of the estimates is more sensitive to larger values of δ, especially in smaller sample sizes. These findings are also visually supported by Figure 2, where the bias and MSE curves clearly reflect the trends observed in Table 2.

7. Applications

This section carefully investigates the practical application of the Zero-Inflated Poisson XLindley (ZIPXL) distribution. We conducted a comparison study using two real-world datasets to compare our proposed model to the zero-inflated Poisson moment exponential (ZIPMEx) model and the usual zero-inflated Poisson (ZIP) distribution. The model parameters are estimated across all versions using the MLE approach. To find the best-fitting model, numerous goodness-of-fit metrics are applied, including the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Chi-square test, along with related $p-$values.

7.1. Infant HIV data

The first dataset is about the HIV-exposed infant and is taken from (Kibika, 2020). Data was gathered from three significant regions: Nairobi, Kisumu, and Mombasa, indicating a prevalence of zero inflation due to implemented measures aimed at reducing the rate of mother-to-child transmission. A total of 494 samples were collected from 60 healthcare centers across these three regions in Kenya for analysis. The dataset is detailed in Table 3. Observed and fitted frequencies for this dataset are plotted in Figure 3.

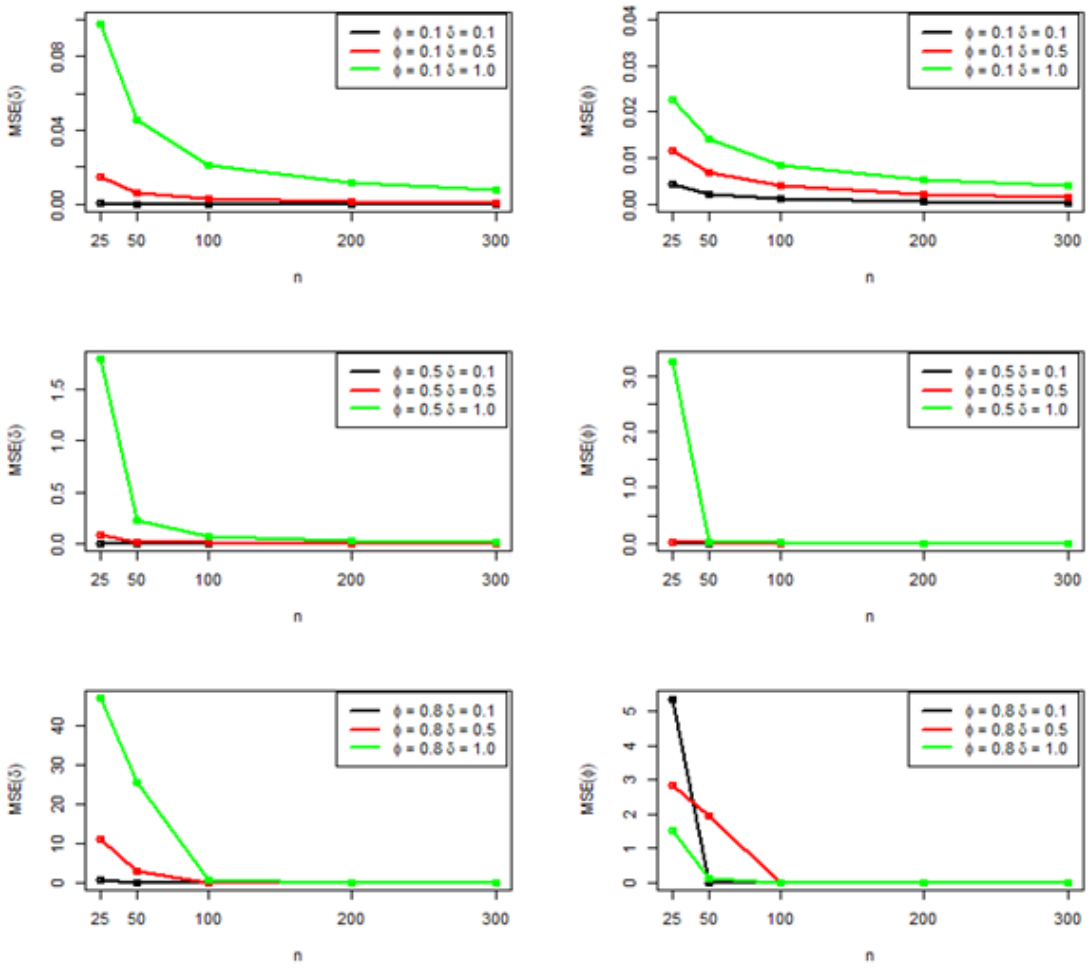

Figure 2: MSEs plots for all combinations of parameters and different sample sizes.

Table 3 demonstrates the observed and fitted frequencies of ZIPXL, ZIPMEx, and Poisson distributions for the infant HIV data. To evaluate the performance of each model, different measures such as chi-square values, $p-$values, log-likelihoods, AIC, and the BIC are utilized. The ZIPXL distribution shows the best fit, with a high $p-$value (0.4629), low Chi-square value (2.5689), highest log-likelihood (-430.66), lowest AIC (865.32), and BIC (873.73), indicating it fits the observed data well. Figure 3 shows the graph of the observed and expected frequencies for the infant HIV data, and it also indicates that the ZIPXL distribution best fits the data amongst all the competing models.

7.2. Criminal sociology data

The second dataset is about criminal sociology with a sample of people with deviant behavior provided by (Diekmann, 1981), and the author discovered that the standard Poisson distribution does not provide a satisfactory fit to the data. This dataset is also fitted by (Böhning, 1998) using ZIP distribution. The MLEs and goodness-of-fit measures are given in Table 4. Observed and fitted frequencies for this dataset are plotted in Figure 4.

Table 3: The MLEs and goodness-of-fit for the infant HIV Data.

Z	Observed	ZIPXL	ZIPMEx	ZIP
0	378	377.9994	377.9955	378.7241
1	59	57.2867	54.4935	46.5060
2	26	29.2519	31.3797	37.3258
3	13	14.7850	16.0620	19.9718
4	7	7.4102	7.7077	8.0147
5	11	7.2667	6.3617	3.4576
Total	494	494	494	494
MLEs				
	$\hat{\delta}$	1.1779	0.6215	0.7062
	$\hat{\phi}$	0.5409	0.6231	1.6052
Goodness-of-fit measures				
	$-\ell$	430.66	430.85	435.40
	AIC	865.32	865.70	874.80
	BIC	873.73	874.11	883.20
	df	3	3	3
	$p-$value	0.4629	0.1495	0.0015
	χ^2	2.5689	5.3255	12.943

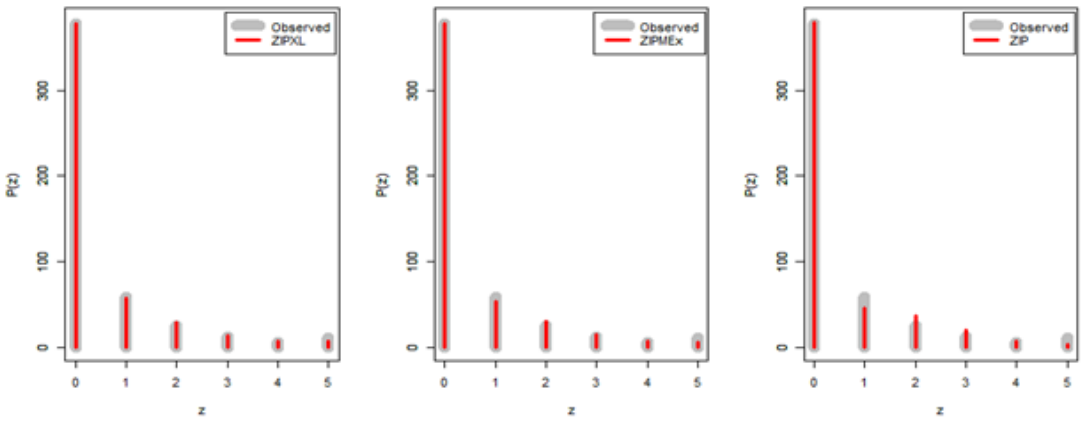

Figure 3: The observed and expected frequencies for the infant HIV data.

It can be observed from Table 4 that the ZIPXL distribution performs better compared to other competing distributions, as evidenced by its highest $p-$value (0.0278) compared to others. Furthermore, our model meets favorable requirements, with the lowest AIC (2331.90) and BIC (2344.64) values among its peers. Figure 4 illustrates a graph of the observed and expected frequencies for the second dataset, indicating that the ZIPXL distribution best matches the data among all competing models. These findings highlight the robustness and usefulness of our suggested model in capturing the fundamental properties of the data, making it a better alternative for analysis.

Table 4: The MLEs and goodness-of-fit for the criminal sociology data.

Z	Observed	ZIPXL	ZIPMEx	ZIP
0	4037	4036.9866	4036.9747	4036.9988
1	219	208.7013	207.3834	204.5400
2	29	43.7349	45.6980	50.1512
3	9	9.1561	8.9509	8.1977
4	5	1.9151	1.6436	1.0050
5	2	0.5058	0.3494	0.1072
Total	4301	4301	4301	4301
Estimates				
	$\hat{\delta}$	0.7073	0.7745	0.8416
	$\hat{\phi}$	3.9250	0.1722	0.4904
Goodness-of-fit measures				
	$-\ell$	1163.95	1165.69	1171.40
	AIC	2331.90	2335.38	2346.80
	BIC	2344.64	2348.12	2359.54
	df	1	1	1
	$p-$value	0.0278	0.0026	0.0001
	χ^2	7.1624	9.0881	14.7503

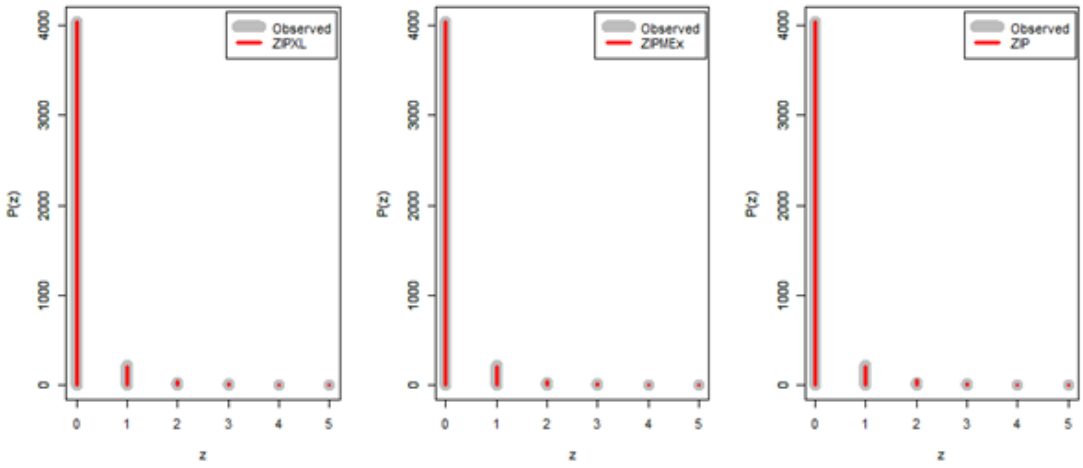

Figure 4: The observed and expected frequencies for the criminal sociology data.

8. Conclusion

When dealing with data having an excess of zeros, accurate count distribution modeling is crucial. Zero-inflated models, first introduced by Lambert (1992), address this issue by accounting for both zero inflation and the underlying counting process, leading to more reliable statistical inferences and predictions across various domains. So, this research paper introduces zero-inflated Poisson XLindley distribution (ZIPXL), a novel extension of the Poisson XLindley distribution. Through meticulous derivation of key statistical properties and application of the maximum likelihood estimation method, we have shown ZIPXL's ability to handle complex data patterns. Our simulation studies give additional proof of the accuracy of

our estimating methods. Real-world dataset analysis verifies practical utility of ZIPXL, proving its superior fit compared to other models using a variety of assessment criteria such as Chi-square values, $p-$values, and information criteria such as AIC and BIC. Overall, ZIPXL appears to be a viable method for analyzing over-dispersed count data with zero inflation, providing a solid and useful tool for both scholars and practitioners.

References

Ahsan-ul-Haq, M., Al-Bossly, A., El-Morshedy, M., & Eliwa, M. S. (2022). Poisson Xlindley distribution for count data: Statistical and reliability properties with estimation techniques and inference. *Computational Intelligence and neuroscience*, *2022*(1), 6503670. https://doi.org/10.1155/2022/6503670

Aryuyuen, S., Bodhisuwan, W., & Supapakorn, T. (2014). Zero-inflated negative binomial-generalized exponential distribution and its applications. *Songklanakarin Journal of Science and Technology*, *36*(4), 483–491.

Böhning, D. (1998). Zero-inflated Poisson models and CA MAN: A tutorial collection of evidence. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *40*(7), 833–843. https://doi.org/10.1002/(SICI)1521-4036(199811)40:7⟨833::AID-BIMJ833⟩3.0.CO;2-O

Diekmann, A. (1981). Forschungsnotiz. ein einfaches stochastisches modell zur analyse von häufigkeitsverteilungen abweichenden verhaltens. *Zeitschrift für Soziologie*, *10*(3), 319–325. https://doi.org/10.1515/zfsoz-1981-0307

Junnumtuam, S., Niwitpong, S.-A., & Niwitpong, S. (2022). A zero-and-one inflated cosine geometric distribution and its application. *Mathematics*, *10*(21), 4012. https://doi.org/10.3390/math10214012

Kibika, S. A. (2020). *The zero inflated negative binomial-shanker distribution and its application to hiv exposed infant data* [Doctoral dissertation, Strathmore University].

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14. https://doi.org/10.2307/1269547

Leroux, B. G., & Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*, 545–558. https://doi.org/10.2307/2532308

Rahman, T., Hazarika, P. J., Ali, M. M., & Barman, M. P. (2022). Three-inflated Poisson distribution and its application in suicide cases of india during COVID-19 pandemic. *Annals of data science*, *9*(5), 1103–1127.

Sabri, S. R. M., & Adetunji, A. A. (2023). Zero-inflated Poisson transmuted weighted exponential distribution: Properties and applications. *Borneo Science— The Journal of Science and Technology*, *44*(2). https://doi.org/10.51200/bsj.v44i2

Sharma, A., & Landge, V. (2013). Zero-inflated negative binomial for modeling heavy vehicle crash rate on indian rural highway. *International Journal of Advances in Engineering & Technology*, *5*(2), 292.

Shukla, K. K., & Yadava, K. (2006). The distribution of the number of migrants at the household level. *Journal of Population and Social Studies*, *14*(2), 153–166.

Skinder, Z., Ahmad, P. B., & Elah, N. (2023). A new zero-inflated count model with applications in medical sciences. *Reliability: Theory & Applications*, *18*(3 (74)), 841–855.

Van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, 738–743. https://doi.org/10.2307/2532959

Wani, M. K., & Ahmad, P. B. (2023). Zero-inflated Poisson-Akash distribution for count data with excessive zeros. *Journal of the Korean Statistical Society*, *52*(3), 647–675. https://doi.org/10.1007/s42952-023-00216-5

Zamani, H., Pakdaman, Z., & Shekari, M. (2023). Zero-inflated Poisson quasi-lindley regression for modeling number of doctor visit data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *9*(1), 1–15. https://doi.org/10.1080/23737484.2023.2164941