ISSN 1973-2201 STATISTICA, anno LXXXIV, n. 3, 2024
DOI: https://doi.org/10.60923/issn.1973-2201/18579

A POSITIVE INFLATED DISCRETE DISTRIBUTION: PROPERTIES AND APPLICATIONS

Bilal Ahmad Peer [1]
Department of Mathematical Sciences, Islamic University of Science & Technology, India
Zehra Skinder
Department of Mathematical Sciences, Islamic University of Science & Technology, India

Summary

We consider data modelling under one inflation for zero-truncated count data, as they typically arise in capture-recapture modelling. One-inflation in zero-truncated count data has recently found considerable attention. In this regard, zero-truncated New Discrete distribution and a distribution to a point mass at one are used to create a one-inflated model namely one-inflated zero-truncated New Discrete distribution. Its reliability characteristics, generating functions, and distributional properties are investigated in some detail. which includes survival function, hazard rate function, probability generating function, characteristic function, variance, skewness, and kurtosis. Monte Carlo Simulation have been undertaken to evaluate the effectiveness of the maximum likelihood estimators. To test the compatibility of our proposed model, the baseline model and the proposed model are distinguished by using the two different test procedures. The adaptability of the suggested model is demonstrated using two real-life datasets from separate domains by taking various performance measures into consideration.

Keywords: Goodness of fit; Hypothesis testing; New discrete distribution; One-inflation; Simulation; Zero-truncation.

1. Introduction

In every area of research, including biology, public health, engineering, sociology, epidemiology, and insurance, the statistical analysis and modelling of count data are fundamental. To improve decision-making while working with count data, we fit a valid probability model to the count data.

Truncation of probability distributions is an essential statistical feature that has several applications in different areas. When a specific range of values for the variables is

[1] Corresponding Author. E-mail: bilalahmadpz@gmail.com

ignored or cannot be seen, the resulting model is said to be truncated. It is preferable to use a zero-truncated probability distribution instead of a discrete distribution when the data is to be represented or produced without zeros. Many datasets exclude zero counts, such as the number of siblings in a family, the number of passengers in a car including the driver, the number of articles published in different journals from various disciplines, the number of disturbing events reported by patients, the number of flowers bloomed, and the number of times a voter has cast a ballot in a general election, etc. Zero-truncated probability models behave well when modelling such type of situations and the results drawn from them seem quite sound. Positive count data modelling can be traced back to the mid-twentieth century, when the first truncated model, known as the zero-truncated Poisson distribution (ZTPD), was put forward by David and Johnson (1952) to model such types of data. Later on, several zero-truncated models like zero-truncated Negative Binomial distribution (ZTNBD) proposed by Sampford (1955), zero-truncated Poisson–Lindley distribution proposed by Ghitany *et al.* (2008), Lagrangian zero-truncated Katz distribution introduced by Shibu *et al.* (2023), zero-truncated discrete Lindly distribution introduced by Kiani (2020) to model such type of data. Irshad *et al.* (2022) presented a generalization of the zero-truncated Poisson distribution (ZTPD), namely Lagrangian zero-truncated Poisson distribution (LZTPD), to model both over-dispersed and under-dispersed count datasets. The baseline model, that we have used in this article introduced by Elah *et al.* (2023) called zero-truncated New discrete distribution (ZTNDD) to analyze different real-life applications in various fields. Let a random variable y follow a ZTNDD with probability mass function (pmf) given in Eq. (1)

$$P(y;\zeta)=\frac{\zeta-1}{\zeta^{y}}; \quad y=1,2,3,\dots \quad \zeta>1, \tag{1}$$

with Mean$=\frac{\zeta}{\zeta-1}$ and Variance$=\frac{\zeta}{\zeta^2+1-2\zeta}$. The motivation behind choosing this model as the base model is that the model has only one parameter and is unimodal. Besides that, the model is equi-dispersed, over-dispersed, and under-dispersed as well. This means that the model is suitable for all three types of data sets.

Besides, inflated models based on the zero-truncated distribution have been examined to explain the large number of ones in the dataset. When the observed data contains a larger number of ones than expected, it is referred to as one-inflation in zero-truncation. Several one-inflated models, truncated at zero have been recently attracted by several researchers like one-inflated positive Poisson model introduced by Godwin and Böhning (2017) to deal with phenomena of excess 1's. Godwin (2017, 2019) proposed a one-inflated zero-truncated Negative Binomial (OIZTNB) model and a positive Poisson mixture model, respectively, and used them as truncated distributions in the Horvitz–Thompson estimation of unknown population size. They also analyzed the various applications in different fields to check the model's adaptability. Böhning and van der Heijden (2019) motivated process prompted by British police data on drunk driving revealed high one-inflation and led them to analyse data under one-inflation. Tajuddin *et al.* (2021) investigated the parameter estimation techniques of one-inflated positive

Poisson distribution and compared different estimation methods in terms of unbiasedness, consistency, efficiency, and deficiency and found that all the estimators are consistent and asymptotically normal. One-inflation index was also developed to analyse the presence of excess ones in the dataset. Some of the recent works regarding the inflation aspect in count data are by Skinder *et al.* (2023) and Wani and Ahmad (2023). One of the count models introduced by Tajuddin *et al.* (2022) namely one-inflated positive Poisson Lindly distribution to estimate the population size of criminals. Kaskasamkul and Bohning (2017) proposed the inflated count model based on the geometric distribution for the estimation of population size.

For positive count data, most of the data comes from the count "1" because "0" counts remain unobserved and it is believed that excessive number of "1" counts can contribute to the dispersion (over or under) in the data. Although statistical modelling in this subject has come a long way, new models are still required from time to time. These new models were inspired by the emerging trends that frequently occur in our count data. As a result, we have developed a simple and flexible two-parametric probability model that can handle the statistical dispersion (over and under-dispersion) and the excess of "1" counts in the data. The model developed may be used as an alternative to some other distributions mentioned in the application Section as it provides better fitting as compared to them.

The remainder of the paper is organized as follows. In Section 2, the proposed model along with its probability mass function (pmf) and cumulative distribution function (cdf) are introduced. Further, reliability characteristics along with generating functions, and distributional properties are also presented in this Section. In Section 3, we discuss the estimation of the parameters of the proposed model by the maximum likelihood method. A rigorous simulation study is discussed in Section 4. In Section 5, different test procedures are applied for examination to check the significance of the inflation parameter. Specific real-life applications are considered in Section 6 from various domains to highlight the functionality of the model. The conclusion is discussed in Section 7 itself.

2. Methodology

By combining the ZTNDD with a point mass δ, a distribution is obtained that accounts for the inflated frequency at one. Take an experiment into consideration that led to the following two responses:

- the first response generates only one count with probability δ, $0 < \delta < 1$;
- the second response is governed by ZTNND with probability $(1-\delta)$.

Also, suppose that the experiment is repeated several times independently. Assume that the first response occurs with probability δ and the second response occurs with probability $(1-\delta)$. Now, take a count variable, say Z, into consideration that takes some distribution which allows for frequent one-valued observations. When the first response occurs, Z is set at $Z=1$ and when the second response occurs, i.e., the counts are generated according to ZTNDD random variable. Thus, for $Z=1$, which could be from the occurrence of either response first with probability δ, or response second with probability $(1-\delta)$. We could have

$$P(Z=1)=\delta+(1-\delta)\frac{(\zeta-1)}{\zeta}; \quad z=1.$$

For $Z>1$, the pmf of Z follows the ZTNDD written as

$$P(Z=z)=(1-\delta)\frac{(\zeta-1)}{\zeta^z}; \qquad z=2,3,4,\ldots .$$

Hence, the pmf of count variable Z is obtained by combining the above two equations and is given as

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

where $\delta\in[0,1]$ is the parameter of the mixture distribution, $\zeta>1$. Eq. (2) is known as one-inflated zero-truncated New Discrete distribution (OIZTNDD).

In order to prove that the above function $P(Z=z)$ is a proper pmf, consider

$$\begin{aligned}\sum_{z=1}^{\infty}P(Z=z)&=\delta+(1-\delta)\frac{(\zeta-1)}{\zeta}+(1-\delta)\sum_{z=2}^{\infty}\frac{(\zeta-1)}{\zeta^z}\\ &=\delta+(1-\delta)\sum_{z=1}^{\infty}\frac{(\zeta-1)}{\zeta^z}\\ &=\delta+(1-\delta)\sum_{z=1}^{\infty}g(z),\end{aligned}$$

where $g(z)=\frac{(\zeta-1)}{\zeta^z}$ is the pmf of the ZTNDD given in Eq.(1). Thus, $\sum_{z=1}^{\infty}g(z)=1$.

Clearly, when $\delta=0$, the distribution in Eq. (2) reduces to ZTNDD with pmf given in Eq. (1).

For $z\geq 2$, the OIZTNDD is unimodal, which can be obtained based on the decreasing function from the following ratio

$$\begin{aligned}\frac{P(Z=z+1)}{P(Z=z)} &= \frac{(1-\delta)\frac{(\zeta-1)}{(\zeta^{z+1})}}{(1-\delta)\frac{(\zeta-1)}{\zeta^{z}}}\\ &= \frac{1}{\zeta}.\end{aligned}$$

The pmf plots given in Figure 1 for different combinations of parametric values indicates that the OIZTNDD in Eq. (2) is unimodal. Further, the mode is at one for different combinations of parameters. Moreover, the tail shows a rapid decrease as the value increases for different combinations of parameters.

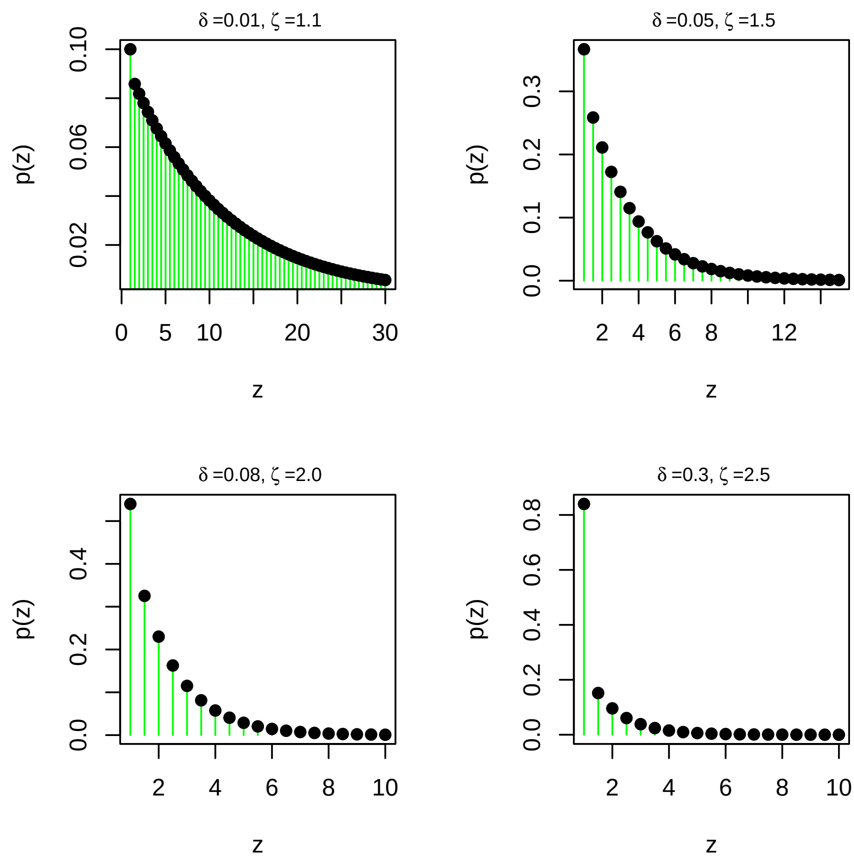

Figure 1 – The pmf plots of OIZTNDD.

2.1. *Cumulative distribution function (cdf)*

THEOREM 1. *If Z ∼ OIZTNDD* (δ, ζ), *then its cdf is given as*

$$F(Z=z)=\delta-\frac{(1-\delta)(1-\zeta^{z})}{\zeta^{z}}. \tag{3}$$

PROOF. If Z ∼ OIZTNDD (δ, ζ), then its cdf is as follows

$$\begin{aligned}
F(Z) &= P(Z\leq z)\\
&= \sum_{y=1}^{z} P(Z=y)\\
&= \delta+(1-\delta)\sum_{y=1}^{z}\frac{(\zeta-1)}{\zeta^{y}}\\
&= \delta+\frac{(1-\delta)(\zeta-1)}{\zeta}\left[1+\frac{1}{\zeta}+\frac{1}{\zeta^{2}}+\frac{1}{\zeta^{3}}+\ldots+\frac{1}{\zeta^{z-1}}\right]\\
&= \delta+(1-\delta)\left[\frac{(\zeta-1)(1-\zeta^{z})}{\zeta^{z}}\right]\\
&= \delta-\frac{(1-\delta)(1-\zeta^{z})}{\zeta^{z}}.
\end{aligned}$$

Hence proved. □

The cdf plots of OIZTNDD in Eq. (2) with different combinations of parameters δ and ζ are provided in Figure 2.

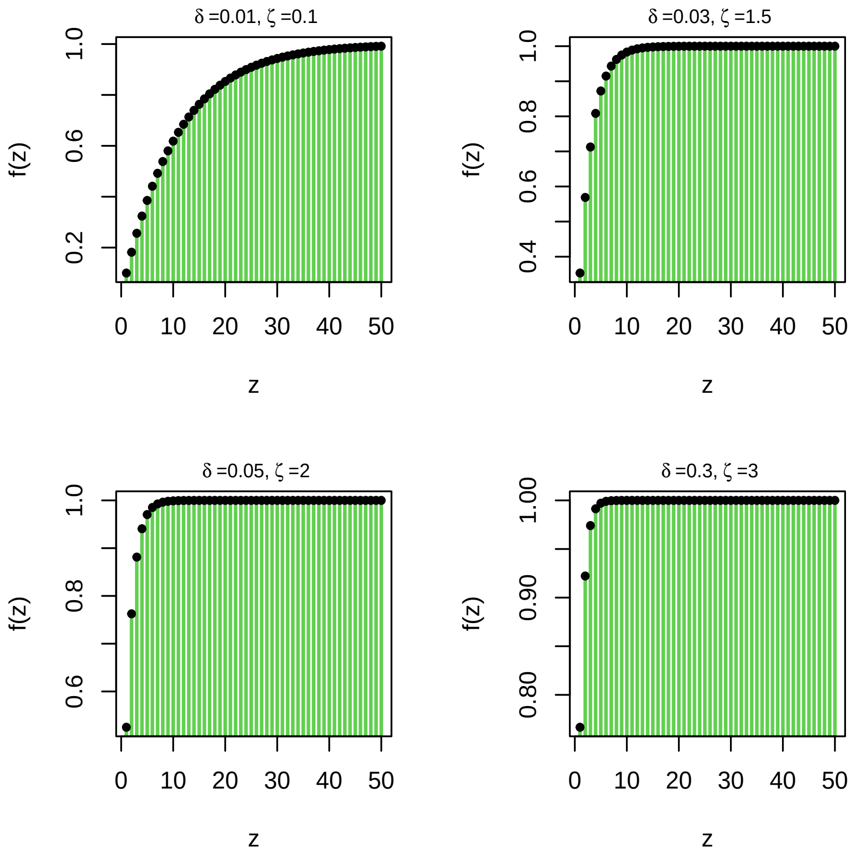

Figure 2 – The cdf plots of OIZTNDD.

2.2. *Reliability characteristics along with generating functions*

In this part, various reliability characteristics like survival analysis, hazard function, and reverse hazard function are discussed along with generating functions. Further, the distributional properties are also discussed.

2.2.1. *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 OIZTNDD (δ,ζ) is as follows

$$\begin{aligned} S(Z) &= 1-F(Z) \\ &= 1-\left[\delta-\frac{(1-\delta)(1-\zeta^z)}{\zeta^z}\right] \\ &= \frac{1-\delta}{\zeta^z}. \end{aligned} \tag{4}$$

2.2.2. *Hazard Rate Function (HRF)*

Let z_1, z_2, z_3,..., z_n be a random sample from OIZTNDD (δ,ζ) as given by Eq.(2). Let Y be the number of z_i's taking the value one. Then Eq. (2) can be written as follows

$$P(Z=z_i)=\left[\delta+(1-\delta)\frac{(\zeta-1)}{\zeta}\right]^Y\left[(1-\delta)\frac{(\zeta-1)}{\zeta^{z_i}}\right]^{1-Y}.$$

Now, using $S(Z)$ from Eq. (4), the Hazard Rate of OIZTNDD (δ,ζ) is given as

$$\begin{aligned} H(Z) &= \frac{P(Z)}{S(Z)} \\ &= \frac{\zeta^z\left[\delta+(1-\delta)\frac{(\zeta-1)}{\zeta}\right]^Y\left[(1-\delta)\frac{(\zeta-1)}{\zeta^{z_i}}\right]^{1-Y}}{(1-\delta)}. \end{aligned}$$

2.2.3. *Moment Generating Function (MGF)*

THEOREM 2. *If $Z\sim$ OIZTNDD (δ,ζ), then its MGF $M_z(t)$ is given as*

$$M_z(t)=\delta e^t+(1-\delta)(\zeta-1)\left[\frac{e^t}{\zeta-e^t}\right]. \tag{5}$$

PROOF. If $Z\sim$ OIZTNDD (δ,ζ), then its MGF is follows as

$$\begin{aligned} M_z(t) &= E(e^{tz}) \\ &= \sum_{z=1}^{\infty} e^{tz}P(Z=z) \\ &= e^t\left[\delta+\frac{(1-\delta)(\zeta-1)}{\zeta}\right]+\sum_{z=2}^{\infty} e^{tz}\frac{(1-\delta)(\zeta-1)}{\zeta^z} \\ &= \delta e^t+\frac{(1-\delta)(\zeta-1)}{\zeta}\left[e^t+\frac{e^{2t}}{(\zeta-e^t)}\right] \\ &= \delta e^t+(1-\delta)(\zeta-1)\left[\frac{e^t}{(\zeta-e^t)}\right]. \end{aligned}$$

Hence proved. □

REMARK 3. *Putting $e^t = S$ in Eq. (5), the Probability Generating Function, of $P_z(S)$ OIZTNDD (δ,ζ) is defined as*

$$P_z(S) = \delta S + (1-\delta)(\zeta-1)\left[\frac{S}{(\zeta-S)}\right]. \tag{6}$$

REMARK 4. *Putting $e^t = e^{it}$ in Eq. (5), the Characteristic Function, $\phi_z(t)$ of OIZTNDD (δ,ζ) is defined as*

$$\phi_z(t) = \delta e^{it} + (1-\delta)(\zeta-1)\left[\frac{e^{it}}{(\zeta-e^{it})}\right]. \tag{7}$$

Through MGF, we have derived the first four raw moments of the proposed distribution by differentiating Eq. (5) at $t=0$. The first four raw moments of the proposed distribution are as fallows

$$\mu_1' = \frac{\zeta-\delta}{\zeta-1}, \tag{8}$$

$$\mu_2' = \frac{(\zeta^2+\zeta+\delta-3\zeta\delta)}{(\zeta-1)^2}, \tag{9}$$

$$\mu_3' = \delta+(1-\delta)\frac{\zeta^3+4\zeta^2+\zeta}{(\zeta-1)^3}, \tag{10}$$

$$\mu_4' = \delta+(1-\delta)\frac{\zeta^4+11\zeta^3+11\zeta^2+\zeta}{(\zeta-1)^4}. \tag{11}$$

Therefore, the mean and variance of the proposed model are respectively given as

$$\begin{aligned} \text{Mean} &= \frac{\zeta-\delta}{\zeta-1}, \\ \text{Variance} &= [\mu_2' - \mu_1'^2] = \frac{\zeta+\delta-\zeta\delta-\delta^2}{(\zeta-1)^2}. \end{aligned}$$

2.2.4. *Index of Dispersion*

If Z ∼ OIZTNDD (δ,ζ), then the Index of Dispersion (ID) is as follows

$$\begin{aligned} \text{ID} &= \frac{\text{Variance}}{\text{Mean}} \\ &= \frac{(\zeta+\delta-\zeta\delta-\delta^2)}{(\zeta-\delta)(\zeta-1)}. \end{aligned} \tag{12}$$

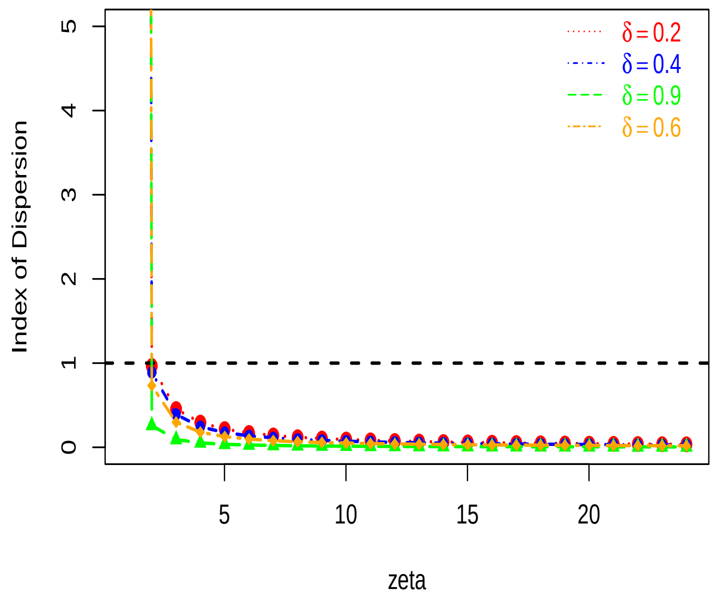

Figure 3 – The ID plot of OIZTNDD.

As can be seen from Figure 3, the proposed model is both over-dispersed as well as under-dispersed. At lower values of δ and ζ, the model is over-dispersed. It may be equi-dispersed at some value of parameters δ and ζ. Hence, with an increase in value of δ and ζ, dispersion index decreases. So, we can say that model shows over-dispersion at lower values of δ and ζ and the model shows under-dispersion, as the parametric value increases for both δ and ζ.

2.2.5. *Coefficient of Skewness*

If Z $\sim$ OIZTNDD (δ,ζ), then the Pearson's β_1 coefficient is as follows

$$\begin{aligned}\beta_1 &= \frac{\mu_3^2}{\mu_2^3} = \frac{[\mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3]^2}{[\mu_2' - \mu_1'^2]^3} \\ &= \frac{[\zeta^2 + \zeta + 2\zeta\delta + 3\delta^2 - \delta(\zeta^2 + 3\zeta\delta + 2\delta^2 + 1)]^2}{[\zeta + \delta - \zeta\delta - \delta^2]^3}. \qquad (13)\end{aligned}$$

From Figure 4, it can be observed that frequencies are lowest at the lower values of parameters δ and ζ and they rapidly increase as the parametric value increases, which means that the model is positively skewed.

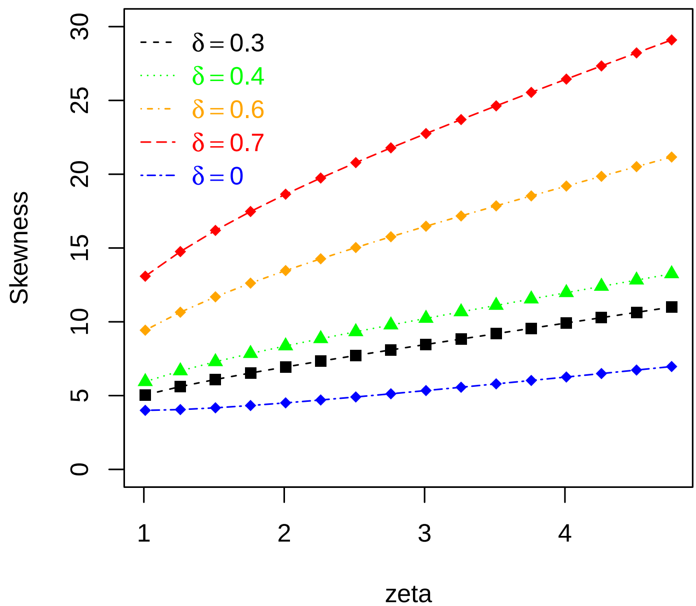

Figure 4 – The Skewness Plot of OIZTNDD.

2.2.6. Coefficient of Kurtosis

If Z $\sim$ ZTOINDD (δ, ζ), then the Pearson's β_2 coefficient is as follows

$$\begin{aligned}
\beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{[\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4]}{[\mu_2' - {\mu_1'}^2]^2} \\
&= \frac{\left[\begin{array}{c} \zeta^3 + 7\zeta^2 + \zeta + 23\zeta^3\delta + 15\zeta\delta + 2\zeta\delta^2 + 12\zeta\delta^3 \\ +6\delta^3 + \delta - \delta(21\zeta^2 + 34\zeta^2\delta + 4\delta + 3\delta^3)] \end{array}\right]}{[3\zeta\delta + \zeta + \delta - \delta^2]^2}. \qquad (14)
\end{aligned}$$

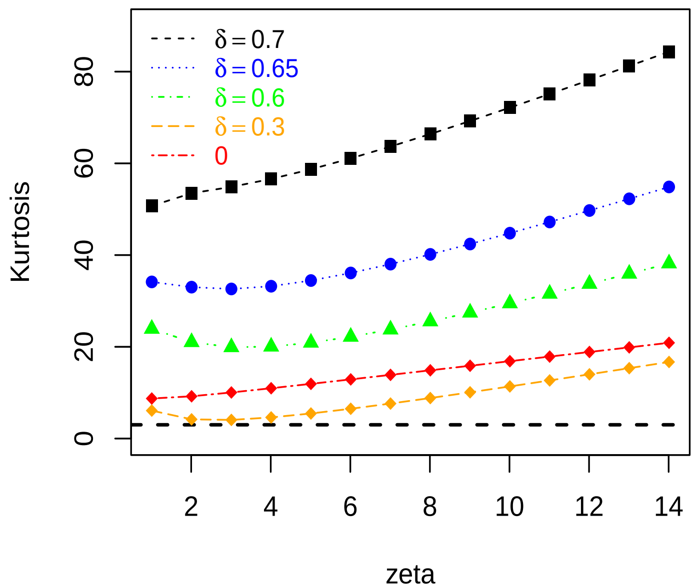

Figure 5 – The Kurtosis Plot of OIZTNDD.

Furthermore, from Figure 5, it can be observed that OIZTNDD in Eq. (2) is leptokurtic as the value of kurtosis increases with increase in parametric values for different combinations of δ and ζ.

3. Parameter Estimation

The parameters δ and ζ of the proposed model can be obtained by using the maximum likelihood method of estimation. The loglikelihood function L for a random variable Z, that follows the ZTOIND distribution is given as

$$\begin{aligned} \ln L &= n_0 \ln\left[\delta + (1-\delta)\frac{(\zeta-1)}{\zeta}\right] + (n-n_0)\ln(1-\delta) \\ &+ (n-n_0)\ln(\zeta-1) - \sum_{z=2}^{\infty} n_z z \ln(\zeta), \end{aligned} \tag{15}$$

where n_z refers to the number of z-valued observations and $n = \sum_{z=1}^{\infty} n_z$. The partial derivatives of the parameters δ and ζ, are obtained in the following equations:

$$\frac{\partial \ln L}{\partial \delta} = \frac{n_0}{[\delta+\zeta-1]} - \frac{(n-n_0)}{(1-\delta)}, \tag{16}$$

$$\frac{\partial \ln L}{\partial \zeta} = \frac{n_0(1-\delta)}{\zeta[\delta\zeta+(1-\delta)(\zeta-1)]} + \frac{(n-n_0)}{(\zeta-1)} - \frac{\sum_{y=2}^{\infty} n_z z}{\zeta}. \tag{17}$$

Now, letting $\frac{\partial \ln L}{\partial \delta}$=0, Eq. (16) reduces to

$$\frac{\partial \ln L}{\partial \delta} = \frac{n_0}{[\delta+\zeta-1]} - \frac{(n-n_0)}{(1-\delta)} = 0$$

$$\Rightarrow \hat{\delta} = 1 - \frac{\zeta(n-n_0)}{n}. \tag{18}$$

Now, letting $\frac{\partial \ln L}{\partial \zeta}$=0, Eq. (17) reduces to

$$\frac{\partial \ln L}{\partial \delta} = \frac{n_0(1-\delta)}{\zeta[\delta\zeta+(1-\delta)(\zeta-1)]} + \frac{(n-n_0)}{(\zeta-1)} - \frac{\sum_{z=2}^{\infty} n_z z}{\zeta} = 0. \tag{19}$$

Since, the differential equation for parameter ζ in Eq. (19) is not in closed form, so we use `fitdistrplus` function in R-software (Delignette-Muller and Dutang, 2015) for obtaining the ML estimator of the proposed model.

The log-likelihood function is continuously differential for all valid parameter values for $\delta < 1$ and $\zeta > 1$, all the terms are ratios of continuous functions and the sum is finite for any finite sample.

The second order differentiations of the log-likelihood function are given here

$$\frac{\partial^2 \ln L}{\partial \delta^2} = -\frac{n_0}{[\delta+\zeta-1]^2} - \frac{(n-n_0)}{(1-\delta)^2},$$

$$\frac{\partial^2 \ln L}{\partial \zeta^2} = -\frac{n_0(1-\delta)(\delta+2\zeta-1)}{\zeta[\delta+\zeta-1]} - \frac{(n-n_0)}{(\zeta-1)^2} + \frac{\sum_{z=2}^{\infty} n_z z}{\zeta^2},$$

and

$$\frac{\partial^2 \ln L}{\partial \delta \partial \zeta} = -\frac{n_0}{[\delta+\zeta-1]^2}.$$

All second-order partial derivatives exist and are continuous for parameters δ and ζ.

THEOREM 5. *The MLE of $\hat{\zeta}$ of ζ is consistent and asymptotically normal, such that*

$$\sqrt{n}(\hat{\zeta}-\zeta) \xrightarrow{d} N(0, I^{-1}(\zeta)),$$

where

$$\begin{aligned} I(\zeta) &= \left[\delta+\frac{(1-\delta)(\zeta-1)}{\zeta}\right]*\frac{\zeta^2-[\zeta\delta+(1-\delta)(\zeta-1)]^2}{\zeta^2[\zeta\delta+(1-\delta)(\zeta-1)]^2} \\ &\quad +\frac{(1-\delta)[\zeta^2-(\zeta-1)^2(2\zeta-1)]}{\zeta^3(\zeta-1)^2}. \end{aligned}$$

PROOF. The regularity conditions under which the MLE $\hat{\zeta}$ is consistently and asymptotically normal is satisfied by the OIZTND distribution (Hogg *et al.*, 2005), therefore

$$\begin{aligned} I(\zeta) &= -E\left(\frac{\partial^2 \ln f(z)}{\partial\zeta^2}\right) \\ &= -E\left[\frac{\partial^2}{\partial\zeta^2}\left\{\begin{array}{c} I_{(z=1)}\ln\left(\delta+\frac{(1-\delta)(\zeta-1)}{\zeta}\right) \\ +I_{(z>1)}(\ln(1-\delta)+\ln(\zeta-1)-z\ln(\zeta)) \end{array}\right\}\right] \\ &= -E\left[I_{(z=1)}U+I_{(z>1)}V\right] \\ &= -\left[Uf(1)+\sum_{z=2}^{\infty}Vf(z)\right], \end{aligned}$$

where, $f(z)=P(Z=z)$, and U and V are respectively given as

$$U=\frac{1}{\zeta^2}-\frac{1}{[\zeta\delta+(1-\delta)(\zeta-1)]^2}$$

and

$$V=\frac{z}{\zeta^2}-\frac{1}{(\zeta-1)^2}.$$

Solving the summation in $I(z)$, gives

$$\begin{aligned} I(\zeta) &= \left[\delta+\frac{(1-\delta)(\zeta-1)}{\zeta}\right]*\frac{\zeta^2-[\zeta\delta+(1-\delta)(\zeta-1)]^2}{\zeta^2[\zeta\delta+(1-\delta)(\zeta-1)]^2} \\ &\quad +\frac{(1-\delta)[\zeta^2-(\zeta-1)^2(2\zeta-1)]}{\zeta^3(\zeta-1)^2}. \end{aligned}$$

Therefore, from Theorem 5, the asymptotic $100(1-\alpha)$ confidence interval of ζ is given as

$$\hat{\zeta}\mp z_{\frac{\alpha}{2}}\frac{I^{-1/2}(\zeta)}{\sqrt{n}}.$$

□

THEOREM 6. *The MLE of $\hat{\delta}$ of δ is consistent and asymptotically normal, such that*

$$\sqrt{n}(\hat{\delta}-\delta)\xrightarrow{d} N(0, I^{-1}(\delta)),$$

where

$$I(\delta)=\frac{1}{\zeta[\delta\zeta+(1-\delta)(\zeta-1)]}-\frac{[\zeta\delta(1-\delta)(\zeta-1)-\zeta]}{\zeta(1-\delta)^2}.$$

PROOF. The regularity conditions under which the MLE $\hat{\zeta}$ is consistently and asymptotically normal is satisfied by the OIZTND distribution (Hogg *et al.*, 2005), therefore

$$\begin{aligned}
I(\delta) &= -E\left(\frac{\partial^2 \ln f(z)}{\partial\delta^2}\right)\\
&= -E\left[\frac{\partial^2}{\partial\delta^2}\left\{\begin{array}{c} I_{(z=1)}\ln\left(\delta+\frac{(1-\delta)(\zeta-1)}{\zeta}\right)\\ +I_{(z>1)}(\ln(1-\delta)+\ln(\zeta-1)-z\ln(\zeta))\end{array}\right\}\right]\\
&= -E\left[I_{(z=1)}W+I_{(z>1)}X\right]\\
&= -[Wf(1)+X(1-f(z))],
\end{aligned}$$

where, $f(z)=P(Z=z)$, and

$$W=-\frac{1}{[\delta\zeta+(1-\delta)(\zeta-1)]^2}.$$

Substituting the value of W in the above equation, we get

$$I(\delta)=\frac{1}{\zeta[\delta\zeta+(1-\delta)(\zeta-1)]}-\frac{[\zeta\delta(1-\delta)(\zeta-1)-\zeta]}{\zeta(1-\delta)^2}.$$

Therefore, from Theorem 6, the asymptotic $100(1-\alpha)$ confidence interval of δ is given as

$$\hat{\delta}\mp z_{\frac{\alpha}{2}}\frac{I^{-1/2}(\delta)}{\sqrt{n}}.$$

□

REMARK 7. *From Theorem 5 and 6, it is obvious that the regularity condition for parameters δ & ζ are fulfilled (see Appendix A).*

4. Simulation

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 through the use of discrete variant of the inverse CDF technique in R-software (R Core Team, 2023).

The procedure was repeated 1000 times for calculation of Bias, Variance, Mean Square Error (MSE), and Mean Relative Estimate (MRE) and the results are given in Table 1.

$$\text{Bias} = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\hat{\zeta} - \zeta \qquad \text{Variance} = \frac{1}{N}\sum_{i=1}^{N}\hat{\zeta}^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{\zeta}\right)^2$$

$$\text{MSE} = \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\hat{\zeta}-\zeta)^2 \quad \text{MRE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\zeta}}{\zeta}.$$

Here, $\hat{\zeta}$ is the estimate of ζ and N=1000, is the number of replications. It can be seen from the Table, that as the sample size increases, the Bias, Variance and MSE decreases and are close to zero for large sample sizes. Also, MRE tends to be 1 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.

5. Hypothesis Testing

This part considers how to test whether the OIZTNDD is preferred over ZTNDD. The latter model is nested in the former. Specifically, when $\delta = 0$ the OIZTNDD collapses to the ZTNDD. For that, we have checked the significance of the inflation parameter (δ) by likelihood ratio test and wald's test.

5.1. *Likelihood Ratio Test*

In order to test the significance of the inflation parameter δ of the OIZTNDD, the likelihood ratio test is carried out to distinguish between ZTNDD (ζ) and OIZTNDD (δ, ζ). Here the null hypothesis is

$$H_0 : \delta = 0$$

and the alternative hypothesis is

$$H_1 : \delta \neq 0.$$

In case of likelihood ratio test, (Skinder *et al.*, 2023), the test statistic is given by

$$-2\ln\xi = 2(l_1 - l_2), \tag{20}$$

where, $l_1 = \ln L(\hat{\omega}; z)$, $\hat{\omega}$ is the maximum likelihood estimator for $\omega = (\delta, \zeta)$ without limitation, and $l_2 = lnL(\hat{\omega}^*, z)$, in which $\hat{\omega}^*$ is the maximum likelihood estimator for

TABLE 1
Simulation table of MLE's for the proposed model.

Sample Size(n)	Parameter	$\delta=0.1,\zeta=1.5$ Bias	Variance	MSE	MRE	$\delta=0.6,\zeta=1.3$ Bias	Variance	MSE	MRE
25	$\hat{\delta}$	0.019	0.015	0.015	1.194	-0.034	0.013	0.014	0.942
	$\hat{\zeta}$	-0.010	0.016	0.0157	0.993	0.114	0.082	0.096	1.088
75	$\hat{\delta}$	-0.008	0.008	0.007	0.919	-0.008	0.005	0.005	0.987
	$\hat{\zeta}$	0.019	0.008	0.008	1.012	0.025	0.013	0.014	1.019
100	$\hat{\delta}$	-0.012	0.004	0.005	0.879	-0.008	0.004	0.004	0.986
	$\hat{\zeta}$	0.015	0.005	0.006	1.009	0.017	0.004	0.004	1.013
300	$\hat{\delta}$	0.007	0.002	0.002	1.074	-0.001	0.001	0.001	0.998
	$\hat{\zeta}$	-0.007	0.001	0.002	0.995	0.003	0.002	0.002	1.002
600	$\hat{\delta}$	0.005	0.001	0.001	1.050	0.001	0.001	0.001	1.001
	$\hat{\zeta}$	0.001	0.001	0.001	1.002	-0.001	0.001	0.001	0.999

sample Size(n)	Parameter	$\delta=0.2,\zeta=2$ Bias	Variance	MSE	MRE	$\delta=0.4,\zeta=1.7$ Bias	Variance	MSE	MRE
25	$\hat{\delta}$	-0.008	0.035	0.035	0.962	-0.064	0.047	0.051	0.839
	$\hat{\zeta}$	0.119	0.254	0.268	1.059	0.259	0.286	0.354	1.153
75	$\hat{\delta}$	-0.003	0.015	0.015	0.983	-0.002	0.010	0.010	0.995
	$\hat{\zeta}$	-0.006	0.041	0.041	0.997	0.015	0.035	0.035	1.009
100	$\hat{\delta}$	-0.014	0.013	0.014	0.928	-0.007	0.012	0.012	0.983
	$\hat{\zeta}$	0.043	0.052	0.054	1.022	0.028	0.037	0.038	1.017
300	$\hat{\delta}$	-0.001	0.005	0.005	0.993	0.012	0.008	0.004	1.031
	$\hat{\zeta}$	0.024	0.019	0.019	1.018	-0.009	0.006	0.006	0.995
600	$\hat{\delta}$	0.009	0.002	0.002	1.044	0.006	0.002	0.002	1.015
	$\hat{\zeta}$	0.001	0.007	0.007	1.001	-0.001	0.005	0.005	0.999

Sample Size(n)	Parameter	$\delta=0.2,\zeta=3$ Bias	Variance	MSE	MRE	$\delta=0.5,\zeta=1.5$ Bias	Variance	MSE	MRE
25	$\hat{\delta}$	0.016	0.054	0.055	1.081	-0.039	0.028	0.029	0.921
	$\hat{\zeta}$	0.581	4.918	5.268	1.194	0.115	0.087	0.100	1.076
75	$\hat{\delta}$	-0.021	0.037	0.037	0.893	-0.008	0.007	0.007	0.983
	$\hat{\zeta}$	0.277	1.167	1.243	1.092	0.018	0.011	0.012	1.013
100	$\hat{\delta}$	-0.007	0.032	0.032	0.963	-0.004	0.006	0.006	0.992
	$\hat{\zeta}$	0.106	0.329	0.341	1.035	0.034	0.015	0.016	1.023
300	$\hat{\delta}$	-0.029	0.012	0.013	0.856	0.008	0.002	0.002	1.015
	$\hat{\zeta}$	0.098	0.113	0.123	1.033	-0.006	0.004	0.004	0.996
600	$\hat{\delta}$	-0.019	0.009	0.009	0.905	-0.009	0.001	0.001	0.982
	$\hat{\zeta}$	-0.005	0.077	0.077	0.998	0.002	0.002	0.002	1.001

ω under the null hypothesis H_0. The test statistic described in Eq. (20) is asymptotically distributed as χ^2 with one degree of freedom.

5.2. Wald's test

Here for testing the significance of the inflation parameter δ of OIZTNDD, we assess wald's test. To test the null hypothesis

$$H_0 : \delta = 0$$

and the alternative hypothesis is

$$H_1 : \delta \neq 0.$$

In case of wald's test, the test statistic (Jansakul and Hinde, 2008) is given by

$$W_\delta = \frac{\hat{\delta}^2}{\mathrm{Var}(\hat{\delta})}, \tag{21}$$

where $\mathrm{Var}(\hat{\delta})$ represents the diagonal element of Fisher's information matrix at $\delta = \hat{\delta}$ and $\zeta = \hat{\zeta}$. The test statistic given in Eq. (21) is asymptotically distributed as χ^2 with one degree of freedom.

6. Applications

In this part, we study the practical significance of one-inflated zero-truncated New Discrete distribution. Two real-life datasets are taken to compare OIZTNDD with few other distributions like zero-truncated New Discrete distribution (ZTNDD), zero-truncated Negative Binomial distribution (ZTNBD), zero-truncated Poisson Lindly distribution (ZTPLD), zero-truncated Discrete Lindly distribution (ZTDLD), zero-truncated two parameter discrete Lindly distribution (ZTTPDLD) and zero-truncated Poisson distribution (ZTPD) to check the performance measures of the proposed model.

Here, we consider two datasets, the first dataset shown in Table 2 has been taken from Umar *et al.* (2019) The data is related to the number of people injured from a major road accident that occurred in Bangkok, Thailand, information gathered from the Department of Highways, Ministry of Transport, Thailand.

TABLE 2
Dataset I: number of people injured from a major road accident. Umar et al. (2019).

Claims	1	2	3	4	5	6
Observed Count	300	71	18	9	4	3

TABLE 3
Dataset II: number of Publications in the Review of Applied Entomology Williams (1943).

Claims	1	2	3	4	5	6	7	8	9	10
Observed Count	285	70	32	10	4	3	3	1	2	1

The second dataset shown in Table 3 has been taken from Williams (1943) and has been recently used by Hassan and Ahmad (2009). The dataset is related to the number of publications in the review of applied entomology.

These two datasets are used to demonstrate and analyze the performance of various metrics. We have fitted OIZTNDD, ZTDD, ZTNBD, ZTPLD, ZTDLD, ZTTPDLD, and ZTPD to both datasets for comparison. To check the adaptability of the model, we have computed expected frequencies, the values of maximized log-likelihood, χ^2 statistic along with associated p-value, Akaike's Information Criteria (AIC), and Bayesian Information Criteria (BIC) through *R* software R Core Team (2023) shown in Tables 4 and 5.

From Tables 4 and 5, it has been revealed that OIZTNDD provides the best fit compared to existing models – ZTNDD, ZTNBD, ZTPLD, ZTDLD, ZTTPDLD, and ZTPD as our proposed model has lowest AIC and BIC, and also has highest p-value in both datasets. Also, the main advantage of the proposed model is in its versatility to handle both over-dispersed as well as under-dispersed count data.

Here, we have also adopted the likelihood ratio test and the wald's test for testing the significance of the inflation parameter δ. In case of likelihood ratio test, the computed value for both datasets shown in Table 6 are 6.06 and 20.88. In case of wald's test, the computed value for both datasets shown in Table 7 are 9.18 and 37.64. The null hypothesis is rejected in all cases of likelihood ratio test and wald's test as the value for both the datasets is greater than 3.84 for one degree of freedom. Hence, we conclude that δ, the additional parameter in the model is significant, as discussed in Section 5.
Here, we have also plotted the estimated pmfs for both the datasets. From Figures 6 and 7, it can be seen that OIZTNDD yields better fit to both the datasets compared to all the existing models considered in the paper. This also supports the suitability of the proposed model to the given datasets.

TABLE 4
Expected frequencies and χ^2 values for fitted models.

Claims	Observed Count	**OIZTNDD**	ZTNDD	ZTNBD	ZTPLD	ZTDLD	ZTTPDLD	ZTPD
1	300	300	288	287	287	284	288	275
2	71	67	83	84	84	88	83	100
3	18	24	24	24	24	24	24	24
4	9	9	7	7	7	6	7	4
5	4	3	2	2	2	2	2	1
6	3	1	1	1	1	0	1	0
Degrees of Freedom		**2**	2	2	1	2	2	1
ML Estimates		$\hat{\delta}$=**0.287** $\hat{\zeta}$=**2.758**	$\hat{\zeta}$=3.455	$\hat{p}$=0.277 $\hat{r}$=1.111	$\hat{\theta}$=2.983	$\hat{\theta}$=1.573	$\hat{p}$=0.289 $\hat{\beta}$=0.001	$\hat{\theta}$=0.727
χ^2-value		**2.53**	7.33	7.70	7.70	11.12	7.33	28.84
p-value		**0.281**	0.025	0.005	0.021	0.003	0.006	<0.001
$-\log\hat{l}$		**339.92**	342.95	343.32	343.51	345.83	342.95	357.65
AIC		**683.85**	687.91	690.65	689.02	693.66	689.91	717.31
BIC		**691.86**	692.81	698.66	693.02	697.66	697.92	721.32

TABLE 5
Expected frequencies and χ^2 values for fitted models.

Claims	Observed Count	**OIZTNDD**	ZTNDD	ZTNBD	ZTPLD	ZTDLD	ZTTPDLD	ZTPD
1	285	285	258	257	256	251	258	236
2	70	65	96	97	98	103	96	121
3	32	31	36	36	36	38	36	41
4	10	15	13	13	13	13	13	10
5	4	7	5	5	5	4	5	2
6	3	4	2	2	2	1	2	0
7	3	2	1	1	1	0	1	0
8	1	1	0	0	0	0	0	0
9	2	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0
Degrees of Freedom		**3**	3	2	2	3	3	2
ML Estimates		$\hat{\delta}$**=0.369** $\hat{\zeta}$**=0.514**	$\hat{\zeta}=0.627$	$\hat{p}=0.359$ $\hat{r}=1.111$	$\hat{\theta}=2.133$	$\hat{\theta}=1.295$	$\hat{p}=0.373$ $\hat{\beta}=0.001$	$\hat{\theta}=1.021$
χ^2-value		**4.65**	15.50	17.13	16.92	33.01	11.00	54.11
p-value		**0.198**	0.001	<0.001	<0.001	<0.001	0.004	<0.001
$-\log \hat{l}$		**423.03**	433.47	434.50	435.71	441.36	433.47	475.83
AIC		**850.06**	868.94	873.00	873.42	884.73	870.94	953.67
BIC		**858.10**	872.96	881.04	877.44	884.73	878.98	957.69

TABLE 6
Calculated value of test statistic in case of likelihood ratio test.

	Test statistic	Degrees of freedom	Critical value
Dataset I	6.06	2	3.841
Dataset II	20.88	3	3.841

TABLE 7
Calculated value of test statistic in case of Wald's test.

	Test statistic	Degrees of freedom	Critical value
Dataset I	9.18	2	3.841
Dataset II	37.64	3	3.841

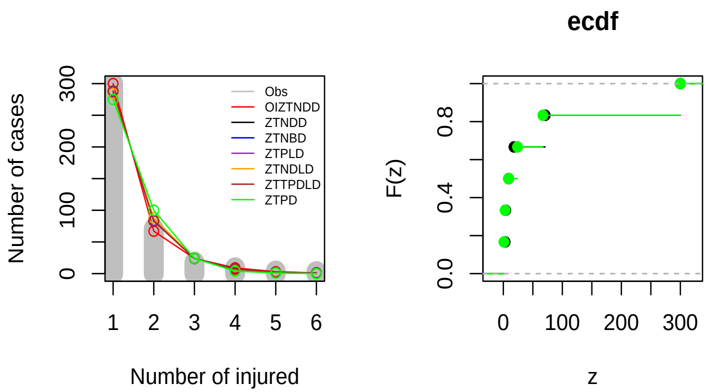

Figure 6 – The observed and expected pmf under OIZTNDD, ZTNDD, ZTNBD, ZTPLD, ZTNDLD, ZTTPDLD, and ZTPD, and Empirical cumulative distribution function (ecdf) of Dataset I.

7. Conclusion

A new one-inflated version of truncated distribution is proposed in this paper namely one-inflated zero-truncated New Discrete distribution (OIZTNDD). The main advantage of the proposed model is in its versatility to handle both over-dispersed as well as under-dispersed count data. Key statistical properties of the distribution including generating functions, reliability characteristics, and moments have been derived.For parametric estimation purpose, the maximum likelihood method of estimation has been used. A simulation study has been done to evaluate the proficiency of the estimation measures considered in this paper. Further, the procedure of the likelihood ratio test and wald's test are carried out to test the significance of the inflation parameter. Two real-life datasets are reviewed to demonstrate the practicality of the proposed model juxtaposed to the existent models ZTNDD, ZTNBD, ZTPLD, ZTDLD, ZTTPDLD, and ZTPD. We can see that OIZTNDD in terms of Chi-square value and p-value gives the best fit as the existent models do not show the best fit. The information measures like Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) have the lowest values among all other competing distributions in terms of numerical value, revealing that OIZTNDD can be considered as a suitable model in comparison to other existing models as discussed in this paper.

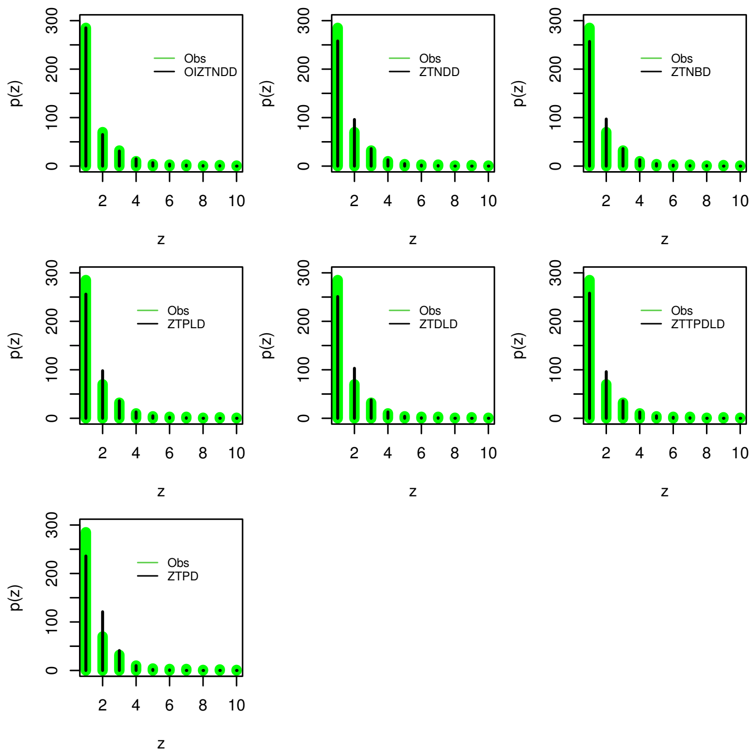

Figure 7 – The frequency plots of dataset II.

APPENDIX

A. CONDITION 1: COMPACTNESS OF PARAMETER SPACE

THEOREM 8 (PARAMETER SPACE COMPACTNESS). *The parameter space* $\Theta=(\delta,\zeta)$ *can be restricted to a compact subset* $\Theta_{\epsilon,M}$ *while maintaining consistency of the MLE.*

PROOF. Let $\epsilon>0$ be small and $M<\infty$ be large. Define

$$\Theta_{\epsilon,M}=[\epsilon,1-\epsilon]\times[1+\epsilon,M].$$

1. Verify closedness:
   - the intervals $[\epsilon,1-\epsilon]$ and $[1+\epsilon,M]$ are closed by construction,
   - their product is closed in $\mathbb{R}^2$.
2. Verify boundedness:

- δ is bounded by ϵ and $1-\epsilon$,
- ζ is bounded by $1+\epsilon$ and M.

3. Apply Heine-Borel Theorem:

- $\Theta_{\epsilon,M}$ is closed and bounded in $\mathbb{R}^2$,
- therefore, $\Theta_{\epsilon,M}$ is compact.

□

B. Condition 2: Continuous Differentiability

Theorem 9 (First-Order Differentiability). *The log-likelihood function is continuously differentiable in both parameters.*

Proof.

1. Differentiate Eq. (15) with respect to δ and let $A = \delta + (1-\delta)\frac{(\zeta-1)}{\zeta}$

$$\begin{aligned}\frac{\partial \log L}{\partial \delta} &= n_0 \cdot \frac{1}{A} \cdot \frac{\partial A}{\partial \delta} + \frac{-1}{1-\delta}(n-n_0) \\ &= n_0 \cdot \frac{1}{A} \cdot \left(1 - \frac{\zeta-1}{\zeta}\right) - \frac{n-n_0}{1-\delta}.\end{aligned}$$

This derivative is continuous because:

- $A > 0$ for all valid parameter values,
- $\delta < 1$ for all valid parameter values,
- all terms are ratios of continuous functions.

2. Differentiate Eq. (15) with respect to ζ

$$\begin{aligned}\frac{\partial \log L}{\partial \zeta} &= n_0 \cdot \frac{1}{A} \cdot \frac{\partial A}{\partial \zeta} + \frac{n-n_0}{\zeta-1} - \sum_{z=2}^{\infty} \frac{n_z z}{\zeta} \\ &= n_0 \cdot \frac{(1-\delta)}{\zeta^2 A} + \frac{n-n_0}{\zeta-1} - \sum_{z=2}^{\infty} \frac{n_z z}{\zeta}.\end{aligned}$$

This derivative is continuous because:

- $\zeta > 1$ for all valid parameter values,
- the sum is finite for any finite sample,

- all terms are ratios of continuous functions.

□

THEOREM 10 (SECOND-ORDER DIFFERENTIABILITY). *All second-order partial derivatives exist and are continuous.*

PROOF.

1. Pure second derivative with respect to δ

$$\frac{\partial^2 \log L}{\partial \delta^2} = -n_0 \cdot \frac{\left(1 - \frac{\zeta-1}{\zeta}\right)^2}{A^2} - \frac{n - n_0}{(1-\delta)^2}.$$

2. Pure second derivative with respect to ζ

$$\frac{\partial^2 \log L}{\partial \zeta^2} = -n_0 \cdot \frac{2(1-\delta)}{\zeta^3 A} - \frac{n - n_0}{(\zeta-1)^2} + \sum_{z=2}^{\infty} \frac{n_z z}{\zeta^2}.$$

3. Mixed partial derivatives

$$\frac{\partial^2 \log L}{\partial \delta \partial \zeta} = \frac{\partial^2 \log L}{\partial \zeta \partial \delta} = -n_0 \cdot \frac{1}{\zeta^2 A^2}.$$

Continuity follows from:

- all denominators are strictly positive in $\Theta_{\epsilon,M}$,
- all components are continuous functions,
- the parameter space is compact.

□

C. CONDITION 3: FISHER INFORMATION MATRIX

THEOREM 11 (FISHER INFORMATION PROPERTIES). *The Fisher Information matrix $I(\theta)$ exists, is continuous and is positive definite.*

PROOF.

1. The Fisher Information matrix is:

$$I(\theta) = -E \begin{bmatrix} \frac{\partial^2 \log L}{\partial \delta^2} & \frac{\partial^2 \log L}{\partial \delta \partial \zeta} \\ \frac{\partial^2 \log L}{\partial \zeta \partial \delta} & \frac{\partial^2 \log L}{\partial \zeta^2} \end{bmatrix}.$$

2. Existence follows from:
   - all second derivatives are continuous,
   - the expectation exists for all components,
   - the parameter space is compact.
3. Positive Definiteness: for any non-zero vector $v = (v_1, v_2)$:

$$v^T I(\theta) v > 0. \tag{22}$$

   This follows from:
   - the log-likelihood function is strictly concave,
   - the second derivatives are negative definite,
   - the expectation preserves positive definiteness.

□

D. CONDITION 4: BOUNDED THIRD DERIVATIVES

THEOREM 12 (BOUNDEDNESS OF THIRD DERIVATIVES). *All third derivatives exist and are bounded in* $\Theta_{\epsilon,M}$.

PROOF.

1. Existence: all third derivatives exist because:
   - the log-likelihood function is infinitely differentiable in $\Theta_{\epsilon,M}$,
   - all denominators remain strictly positive.
2. Boundedness: for any third derivative $\frac{\partial^3 \log L}{\partial \theta_i \partial \theta_j \partial \theta_k}$:
   - the derivative is continuous on $\Theta_{\epsilon,M}$,

- $\Theta_{\epsilon,M}$ is compact,
- by the extreme value Theorem, the derivative is bounded.

□

E. Main Result

Theorem 13 (Asymptotic Normality of MLE). *Under the verified conditions, for the Maximum Likelihood Estimator $\hat{\theta}_n$:*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

where θ_0 is the true parameter value.

Proof.

1. We have shown:
   - compactness of parameter space,
   - twice continuous differentiability,
   - existence and positive definiteness of Fisher Information,
   - boundedness of third derivatives.

   Hence the Theorem.
2. Apply Taylor expansion around the true parameter θ_0:
   $$\begin{aligned} 0 &= \frac{\partial \log L(\hat{\theta}_n)}{\partial \theta} \\ &= \frac{\partial \log L(\theta_0)}{\partial \theta} + (\hat{\theta}_n - \theta_0) \frac{\partial^2 \log L(\theta^*)}{\partial \theta^2}, \end{aligned}$$
   where θ^* lies between $\hat{\theta}_n$ and θ_0.
3. By the Central Limit Theorem:
   $$\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, I(\theta_0)).$$
4. By combining the above:
   $$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

□

F. Implications

The asymptotic variance-covariance matrix of the MLE is:

$$\mathrm{Var}(\hat{\theta}_n) \approx \frac{I(\theta_0)^{-1}}{n},$$

for large sample sizes.

Remark 14. *This result allows for:*

- *construction of confidence intervals,*
- *hypothesis testing,*
- *efficiency comparisons with other estimators.*

References

D. Böhning, P. G. van der Heijden (2019). *The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in britain*. Annals of Applied Statistics, 13, no. 2, pp. 1198–1211.

F. David, N. Johnson (1952). *The truncated Poisson distribition*. Biometrics, 8, no. 4, pp. 275–285.

M. L. Delignette-Muller, C. Dutang (2015). `fitdistrplus`*: An R package for fitting distributions*. Journal of Statistical Software, 64, no. 1, pp. 1–34.

N. Elah, P. B. Ahmad, M. A. Wani (2023). *A new zero-truncated distribution and its applications to count data*. Reliability: Theory & Applications, 18, no. 2 (73), pp. 327–339.

M. Ghitany, D. K. Al-Mutairi, S. Nadarajah (2008). *Zero-truncated Poisson–Lindley distribution and its application*. Mathematics and Computers in Simulation, 79, no. 3, pp. 279–287.

R. T. Godwin (2017). *One-inflation and unobserved heterogeneity in population size estimation*. Biometrical Journal, 59, no. 1, pp. 79–93.

R. T. Godwin (2019). *The one-inflated positive Poisson mixture model for use in population size estimation*. Biometrical Journal, 61, no. 6, pp. 1541–1556.

R. T. Godwin, D. Böhning (2017). *Estimation of the population size by using the one-inflated positive Poisson model*. Journal of the Royal Statistical Society Series C: Applied Statistics, 66, no. 2, pp. 425–448.

A. HASSAN, P. B. AHMAD (2009). *Misclassification in size-biased modified power series distribution and its applications*. Journal of the Korean Society for Industrial and Applied Mathematics, 13, no. 1, pp. 55–72.

R. HOGG, J. MCKEAN, A. CRAIG (2005). *Maximum Likelihood Estimation*. Pearson Prentice Hall, New Jersey.

M. R. IRSHAD, C. CHESNEAU, D. S. SHIBU, M. MONISHA, R. MAYA (2022). *Lagrangian zero truncated Poisson distribution: properties regression model and applications*. Symmetry, 14, no. 9, p. 1775.

N. JANSAKUL, J. P. HINDE (2008). *Score tests for extra-zero models in zero-inflated negative binomial models*. Communications in Statistics - Simulation and Computation, 38, no. 1, pp. 92–108.

P. KASKASAMKUL, D. BOHNING (2017). *Population size estimation for one-inflated count data based upon the geometric distribution*. In P. V. D. H. D. BOHNING, J. BUNGEAND (eds.), *Capture-recapture methods for the social and medical sciences*, Chapman and Hall/CRC, New York, pp. 191–209.

T. H. KIANI (2020). *A zero truncated discrete distribution: theory and applications to count data*. Pakistan Journal of Statistics and Operation Research, pp. 167–190.

R CORE TEAM (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

M. SAMPFORD (1955). *The truncated negative binomial distribution*. Biometrika, 42, no. 1/2, pp. 58–69.

D. S. SHIBU, C. CHESNEAU, M. MONISHA, R. MAYA, M. R. IRSHAD (2023). *A novel zero-truncated Katz distribution by the lagrange expansion of the second kind with associated inferences*. Analytics, 2, no. 2, pp. 463–484.

Z. SKINDER, P. B. AHMAD, N. ELAH (2023). *A new zero-inflated count model with applications in medical sciences*. Reliability: Theory & Applications, 18, no. 3 (74), pp. 841–855.

R. R. M. TAJUDDIN, N. ISMAIL, K. IBRAHIM (2021). *Comparison of estimation methods for one-inflated positive Poisson distribution*. Journal of Taibah University for Science, 15, no. 1, pp. 869–881.

R. R. M. TAJUDDIN, N. ISMAIL, K. IBRAHIM (2022). *Estimating population size of criminals: A new horvitz–thompson estimator under one-inflated positive Poisson–Lindley model*. Crime & Delinquency, 68, no. 6-7, pp. 1004–1034.

M. A. UMAR, K. JIMOH, W. B. YAHYA (2019). *A note on the applications of zero truncated distributions*. Professional Statisticians Society of Nigeria Edited Proceedings of 3rd International Conference, 3, pp. 499–505.

M. K. WANI, P. B. AHMAD (2023). *Zero-inflated Poisson-Akash distribution for count data with excessive zeros*. Journal of the Korean Statistical Society, 52, no. 3, pp. 647–675.

C. WILLIAMS (1943). *The numbers of publications written by biologists*. Annals of Eugenics, 12, no. 1, pp. 143–146.