STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–12.
Published online in International Academic Press (www.IAPress.org)

One-inflated zero-truncated Binomial distribution: Distributional Properties, Estimation and Applications

Kalpasree Sharma [1], Partha Jyoti Hazarika[1], Morad Alizadeh [2,*]

[1]*Department of Statistics, Dibrugarh University, Assam, India*
[2]*Department of Statistics, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran*

Abstract The analysis of zero-truncated count data plays a vital role in various fields, such as public health, ecology, manufacturing and social sciences.This work introduces a one-inflated extension of the zero-truncated binomial distribution, framed as a two-component model. Some of the key statistical properties of the distribution, including moments, generating functions, skewness, kurtosis and index of dispersion are derived. Also characterization of the proposed distribution based on conditional distribution is presented. The maximum likelihood estimation strategy is discussed and a simulation study is conducted to assess the performance of the maximum likelihood estimators. Finally to test the compatibility of our model we test two real-life datasets.

Keywords Count data, one-inflation; zero-truncation; Overdispersion; Binomial distribution; Estimation methods; Simulation.

AMS 2010 subject classifications 62E15, 62E99

DOI: 10.19139/soic-2310-5070-2888

1. Introduction

Count data are ubiquitous in various scientific fields, ranging from public health and ecology to actuarial science and industrial reliability. The Poisson distribution serves as a fundamental tool for analyzing such data, however, the applicability of the Poisson model is limited when its assumptions are violated in practice.

One important limitation arises when datasets contain more zeros than expected under a Poisson distribution—a phenomenon known as zero-inflation. This occurs, for example, in epidemiological studies where a large portion of the population might not be exposed to the risk factor of interest, resulting in many true zero counts. Zero-inflated models, such as the zero-inflated Poisson (ZIP) ([1]), zero-inflated negative binomial (ZINB) ([2]), zero-inflated COM-Poisson (ZICOM) ([3]) and zero-inflated discrete Lindley (ZIDL) ([4]) address this issue by combining a count model with a separate process that models the excess zeros. These models are particularly useful when the observed zeros come from a mixture of structural zeros (those that must occur) and sampling zeros (those that might occur by chance).

Another common scenario involves the absence of zeros due to the nature of the data collection process. For instance, in insurance datasets, we only see data from policyholders who've made at least one claim; people with zero claims aren't included. Similarly, in medical studies, patients who never visited a clinic or hospital may be excluded by design. In such cases, standard count models like Poisson no longer apply directly, since they assign positive probability to zero counts. Zero-truncated distributions have been extensively studied to address

*Correspondence to: Morad Alizadeh (Email: moradalizadeh78@gmail.com) Department of Statistics, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University,Bushehr 75169, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

such data structures. These distributions are constructed by conditioning the standard count distributions on the event that the count is strictly positive. Some of the significant works on zero-truncated models are the zero-truncated Poisson (ZTP) ([5]) distribution, the zero-truncated Negative-Binomial (ZTNB) ([6]), zero-truncated Poisson-Lindley (ZTPL) ([7]) distribution, just to cite a few.

While zero-truncated models are effective in cases where zeros are structurally absent or unrecorded, they often fail to adequately model datasets where the frequency of ones is disproportionately high. This phenomenon, referred to as one-inflation, results in underestimation of the lower tail of the distribution when using standard zero-truncated models.

In recent times there has been growing interest in addressing the dual challenge of zero-truncation and one-inflation in count data, and this had led to the development of notable one-inflated zero-truncated models in the literature. And some of the key contributions involve the one-inflated positive Poisson (OIPP) distribution by [8] and one-inflated zero-truncated Negative-Binomial by [9] and used these as the truncated count distribution in Horvitz–Thompson estimation of the population size.

In this study, we focus on the relatively less explored binomial distribution, particularly due to its inherently underdispersed nature. Although several extensions, such as the zero-truncated binomial ([10]) and the one-inflated binomial ([11]) distributions, have been discussed in the literature, to the best of our knowledge no attempt has been made so far to develop a one-inflated zero-truncated binomial distribution. The proposed distribution is structurally straightforward yet exhibits considerable flexibility in form. It shows positive skewness, with its shape accommodating varying degrees of kurtosis, thereby capturing a broad range of data behaviors. A particularly valuable feature is its ability to handle both underdispersion and overdispersion effectively, addressing a limitation commonly found in many conventional models. Furthermore, the availability of closed-form expressions for key statistical measures enhances its practical utility, especially in applications involving the modeling of rare events and extreme observations.

The rest of the paper is structured as follows. Section 2 introduces the formulation of the OIZTB distribution and includes a graphical representation of its probability mass function (pmf). In Section 3, we delve into key statistical characteristics of the distribution, such as generating functions, moments and related attributes, the index of dispersion, and their respective graphical analyses. Section 4 focuses on the characterization of the proposed distribution based on conditional structure. Section 5 outlines the maximum likelihood estimation (MLE) method. A simulation study is presented in Section 6 to assess the performance and reliability of the MLE approach. Section 7 demonstrates the real-world applicability of the proposed model by analyzing two empirical datasets. Finally, Section 8 concludes the paper with a summary of findings and observations.

2. Formulation of the One-Inflated Zero-Truncated Binomial (OIZTB) Distribution

Suppose Y follows a zero-truncated binomial (ZTB) distribution with parameters n and p, then its pmf can be obtained as

$$\begin{aligned} P(Y=y) &= \frac{f(y)}{1-f(0)} \\ &= \frac{\binom{n}{y}\, p^y\, q^{n-y}}{1-q^n} \qquad y=1,2,3,..\,;\;\; q=1-p \end{aligned} \tag{1}$$

Now, let us take a r.v X such that $X \sim \text{ZTB}(n,p)$ and to address the phenomenon of one-inflation we introduce an inflation parameter π $(0<\pi<1)$ to it. Thus, the pmf of one-inflated zero-truncated Binomial (OIZTB) distribution is given as

$$P(x;n,p,\pi) = \begin{cases} \pi + \dfrac{(1-\pi)\, n\, p\, (1-p)^{n-1}}{1-(1-p)^n} & x=1, \\ (1-\pi)\dfrac{\binom{n}{x}\, p^x\, (1-p)^{n-x}}{1-(1-p)^n} & x=2,3,... \end{cases} \tag{2}$$

The pmf plots of the OIZTB distribution for various parameter settings are shown in Figure 1. These plots indicate that as both the sample size n and the inflation parameter, π increase, the probability mass becomes increasingly concentrated at a single point. Conversely, when the value of n increases and π decreases, the distribution tends to exhibit a bell-shaped pattern.

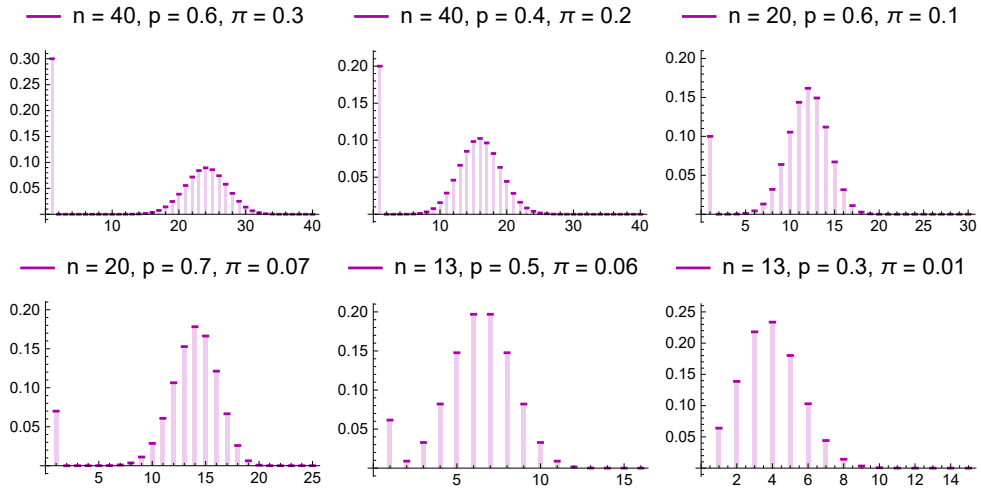

Figure 1. Plot of PMF of OIZTB(n, p, π) for different values of n, p and π.

The cumulative distribution function (cdf) is given simultaneously as

$$F(X \leq k) = \begin{cases} \pi + \dfrac{(1-\pi)\, n\, p\, (1-p)^{n-1}}{1-(1-p)^n} & x = 1, \\ \dfrac{(1-\pi)}{1-(1-p)^n} \sum_{x=2}^{k} \binom{n}{x} p^x (1-p)^{n-x} & x = 2, 3, \ldots \end{cases} \tag{3}$$

Particular cases:

(1) When $\pi \to 0$, (2) reduces to ZTB with parameters n and p as given in (1).
(2) When $\pi \to 1$, (2) degenerates to a point mass at 1.
(3) When $\pi \to 0$, and $n \to \infty$, (2) reduces to a normal distribution.

3. Statistical Properties

3.1. Probability Generating Function (pgf)

If $X \sim$ OIZTB (n, p, π), then the probability generating function can be derived as

$$\begin{aligned} G_X(s) &= E(s^x) \\ &= \sum_{x=1}^{\infty} s^x Pr(X = x) \\ &= p(1) + \sum_{x=2}^{\infty} s^x p(x) \\ &= \pi + \frac{(1-\pi)(1-p)^{n-1}}{1-(1-p)^n} \left(p - nps + (1-p)(1 + ps(1-p)^{-1})^n\right) \end{aligned}$$

Remark 1. The moment generating function (mgf) and the characteristic function (cf) are, respectively, given as

$$M_X(t) = \pi + \frac{(1-\pi)(1-p)^{n-1}}{1-(1-p)^n}\big(p - npe^t + (1-p)(1+pe^t(1-p)^{-1})^n\big), \quad \text{and} \tag{4}$$

$$\Phi_X(t) = \pi + \frac{(1-\pi)(1-p)^{n-1}}{1-(1-p)^n}\big(p - npe^{it} + (1-p)(1+pe^{it}(1-p)^{-1})^n\big). \tag{5}$$

3.2. Moments and Related Measures

If $X \sim$ OIZTB (n, p, π), then the r^{th} order moment about zero is

$$\begin{aligned}\mu_r^{'} &= \sum_{x=1}^{\infty} x^r P(X=x) \\ &= \pi + \frac{(1-\pi)\ n\ p\ (1-p)^{n-1}}{1-(1-p)^n} + \sum_{x=2}^{\infty} x^r (1-\pi)\frac{\binom{n}{x}p^x(1-p)^{n-x}}{1-(1-p)^n}\end{aligned} \tag{6}$$

The first four moments can be obtained by substituting r=1,2,3,4 in Eq (6) as

$$\begin{aligned}\mu_1^{'} &= \pi + \frac{n(1-\pi)}{1-(1-p)^n} \\ \mu_2^{'} &= \pi + \frac{(1-\pi)np(1+(n-1)p)}{1-(1-p)^n} \\ \mu_3^{'} &= \pi + \frac{(1-\pi)np(1-p)^{n-1}}{1-(1-p)^n}\big((1-p)^{1-n}(1+(n-1)p(3+(n-2)p)\big) \\ \mu_4^{'} &= \pi + \frac{(1-\pi)np(1-p)^{n-1}}{1-(1-p)^n}\xi(p)\end{aligned}$$

where $\xi(p) = {}_4F_3\left(\begin{matrix}2,\ 2,\ 2,\ 1-n \\ 1,\ 1,\ 1\end{matrix};\ \frac{p}{p-1}\right)$ is the generalized hypergeometric function ([12]).

Thus, the mean and variance of the OIZTB distribution are, respectively obtained as

$$E(X) = \pi + \frac{n(1-\pi)}{1-(1-p)^n}, \quad \text{and} \tag{7}$$

$$V(X) = \pi + \frac{(1-\pi)np(1+(n-1)p)}{1-(1-p)^n} - \left(\pi + \frac{n(1-\pi)}{1-(1-p)^n}\right)^2 \tag{8}$$

Subsequently, the central moments can be obtained by using the raw moments, but due to lengthy and complex structure, we avoid presenting them here.

Since the coefficients of skewness and kurtosis do not have concise closed-form expressions for our distribution, we refrain from presenting them analytically. Instead, we illustrate their behavior through the plots shown in Figures 2 and 3. From these plots, it is evident that the distribution exhibits positive skewness and a wide range of kurtosis—ranging from platykurtic to mesokurtic, and even leptokurtic shapes. The plots of coefficients of skewness and kurtosis are presented respectively as

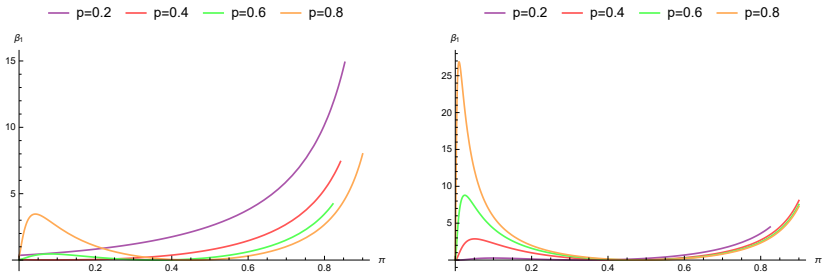

Figure 2. Plots showing coefficient of skewness as a function of p and π.

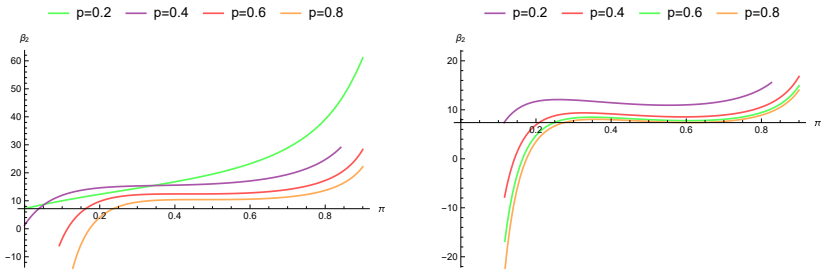

Figure 3. Plots showing coefficient of kurtosis as a function of p and π.

3.3. Index of Dispersion

The index of dispersion (I_x) is a ratio of variance to mean that measures how dispersed count data are. When the index value is equal to 1, the data are equidispersed as in a Poisson model. Values greater than 1 suggest overdispersion, while values less than 1 imply underdispersion. Thus, it is a very useful tool for diagnosing the fit of count data models. The index of dispersion for our proposed model is obtained as

$$I_x = 1 + \frac{(1-\pi)n(n-1)p^2}{\pi\big(1-(1-p)^n\big) + (1-\pi)np} - \frac{np(1-\pi)}{1-(1-p)^n}$$

Figure 4 displays the plots of the index of dispersion for the OIZTB distribution. From these plots, it is clear that the distribution is flexible enough to capture equidispersion, overdispersion, and underdispersion, demonstrating its adaptability to a wide range of count data behaviors.

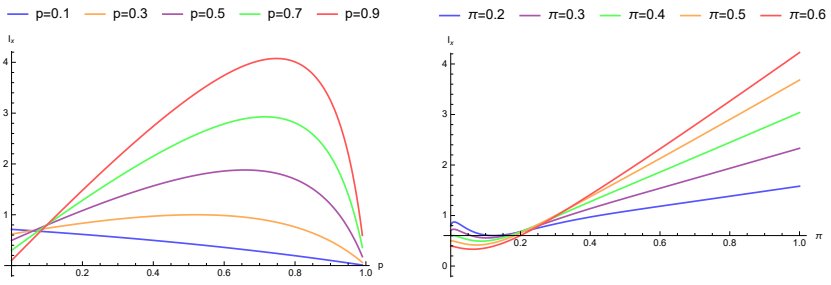

Figure 4. Plots of index of dispersion for n=10, across different possible values of p and π.

4. Characterization of OIZTB Distribution

4.1. Characterization Based on Conditional Distribution

Here, we present the characterization of OIZTB based on conditional distribution of a certain function of the random variable.
First let us rewrite the pmf and cdf of OIZTB in the following manner.

$$f(x) = \begin{cases} \pi + (1-\pi)np(1-p)^{n-1}, & x = 1 \\ Cp^x P(x), & x \in \mathbb{N}, \end{cases} \tag{9}$$

and

$$F(x) = \begin{cases} \pi + (1-\pi)np(1-p)^{n-1}, & x = 1 \\ C\sum_{x=2}^{k} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \mathbb{N} \setminus \{1\}, \end{cases} \tag{10}$$

where $P(x) = \binom{n}{x}(1-p)^{n-x}$, and $C = \dfrac{(1-\pi)}{1-(1-p)^n}$.

Theorem. Let $X : \Omega \to \mathbb{N}$ be a $r.v$. The pmf of X is (9) if and only if

$$E\left\{ [P(x)]^{-1} \mid X \leq k \right\} = \frac{p^2(1-p^{k-1})(1-p)^{-1}}{\sum_{x=2}^{k} p^x P(x)}. \tag{11}$$

Proof. If X has pmf (9) , then for $k \in \mathbb{N} \setminus \{1\}$, the left-hand side of (11) using finite geometric sum formula, will be

$$(F(k))^{-1} \sum_{x=2}^{k} C\, p^x = \frac{p^2\,(1-p^{k-1})\,(1-p)^{-1}}{\sum_{x=2}^{k} p^x P(x)}.$$

On the other hand, if (11) holds, then

$$\sum_{x=2}^{k} \left\{ [P(x)]^{-1} f(x) \right\} = F(k)\, \frac{p^2\,(1-p^{k-1})\,(1-p)^{-1}}{\sum_{x=2}^{k} p^x P(x)}. \tag{12}$$

From equation (12) we can also derive,

$$\begin{aligned} \sum_{x=2}^{k-1} \left\{ [P(x)]^{-1} f(x) \right\} &= F(k-1)\, \frac{p^2\,(1-p^{k-2})\,(1-p)^{-1}}{\sum_{x=2}^{k-1} p^x P(x)} \\ &= (F(k) - f(k)) \times \frac{p^2\,(1-p^{k-1})\,(1-p)^{-1}}{\sum_{x=2}^{k-1} p^x P(x)}. \end{aligned} \tag{13}$$

Now, subtracting Equation (13) from Equation (12), we get

$$\begin{aligned} & f(k)\left\{ [P(k)]^{-1} - \frac{p^2\,(1-p^{k-2})\,(1-p)^{-1}}{\sum_{x=2}^{k} p^x P(x)} \right\} \\ & = F(k)\left\{ \frac{p^2\,(1-p^{k-1})\,(1-p)^{-1}}{\sum_{x=2}^{k} p^x P(x)} - \frac{p^2\,(1-p^{k-2})\,(1-p)^{-1}}{\sum_{x=2}^{k-1} p^x P(x)} \right\} \end{aligned} \tag{14}$$

From the above equality, we have

$$\frac{f(k)}{F(k)} = \frac{\dfrac{p^2\,(1-p^{k-1})\,(1-p)^{-1}}{\sum_{x=2}^{k}\,p^x P(x)} - \dfrac{p^2\,(1-p^{k-2})\,(1-p)^{-1}}{\sum_{x=2}^{k-1}\,p^x P(x)}}{[P(k)]^{-1} - \dfrac{p^2\,(1-p^{k-2})\,(1-p)^{-1}}{\sum_{x=2}^{k}\,p^x P(x)}}, \tag{15}$$

and the RHS of this equation, after some further simplification, gives the reverse hazard rate function corresponding to the pmf (10), hence confirming that X indeed follows this distribution.

5. Method of Estimation (Maximum Likelihood Estimation Method)

Let $X \sim$ OIZTB (n, p, π) and let us take a random sample of size n, say, $x_1, x_2, x_3, ..., x_n$ from this distribution . Let,

$$U_i = \begin{cases} 1, & x_i = 1, \\ 0, & otherwise \end{cases} \tag{16}$$

Now, the pmf of OIZTB (n, p, π) can be written as

$$Pr(X = x_i) = \left[\pi + \frac{(1-\pi)\,np(1-p)^{n-1}}{1-(1-p)^n}\right]^{U_i} \left[\frac{(1-\pi)\binom{n}{x_i}p^{x_i}(1-p)^{n-x_i}}{1-(1-p)^n}\right]^{1-U_i} \tag{17}$$

The likelihood function, $L = L(n, p, \pi; x_1, x_2, ...x_n)$ will be

$$L = \left[\pi + \frac{(1-\pi)\,np(1-p)^{n-1}}{1-(1-p)^n}\right]^{n_0} \prod_{i=1}^{n}\left[\frac{(1-\pi)\binom{n}{x_i}p^{x_i}(1-p)^{n-x_i}}{1-(1-p)^n}\right]^{u_i}, \tag{18}$$

where $n_0 = \sum_{i=1}^{n} U_i$, is the number of ones in the given sample and $(1 - U_i) = u_i$. Thus, we can write the log-likelihood function as

$$\log L = n_0 \log\left[\pi + \frac{(1-\pi)\,np(1-p)^{n-1}}{1-(1-p)^n}\right] + \log\prod_{i=1}^{n}\left[\frac{(1-\pi)\binom{n}{x_i}p^{x_i}(1-p)^{n-x_i}}{1-(1-p)^n}\right]^{u_i}, \tag{19}$$

Differentiating Equation (19) with respect to p and π, we obtain the score functions as follows

$$\begin{aligned}\frac{\delta}{\delta p}\log L = {} & \frac{n_0\Big((1-p)^{n-1} - (n-1)(1-p)^{n-2}p - \dfrac{np(1-p)^{2n-2}}{1-(1-p)^n}\Big)}{\pi\big(1-(1-p)^n\big) + (1-\pi)np(1-p)^{n-1}} + \frac{n(n-n_0)}{p-1} \\ & - \frac{\sum_{i=1}^{n} u_i x_i}{p(p-1)} - \frac{n(n-n_0)(1-p)^{n-1}}{1-(1-p)^n}\end{aligned} \tag{20}$$

$$\frac{\delta}{\delta\pi}\log L = \frac{n_0(1-(1-p)^n)\Big(1 - \dfrac{np(1-p)^{n-1}}{1-(1-p)^n}\Big)}{\pi(1-(1-p)^n) + (1-\pi)np(1-p)^{n-1}} - \frac{n-n_0}{1-\pi} \tag{21}$$

Here we observe that because of their structural complexity, the maximum likelihood estimators (MLEs) do not have a closed-form structure. Therefore, the log-likelihood function is solved directly using a suitable optimization technique.

6. Simulation

To evaluate the performance of the MLEs for the parameters (p and π) of the OIZTB distribution, we conducted a Monte Carlo simulation study. Parameter n is assumed to be known. For each replication, we generated synthetic data by simulating from the one-inflated zero-truncated binomial model with specified parameter values. The MLEs were then computed for each simulated dataset, and the bias and mean squared error (MSE) were calculated across all replications. Simultaneously, the standard error (S.E) is also computed .This process was repeated N=10,000 times to assess the consistency of the maximum likelihood estimators. Simulations were carried out for various sample sizes (50, 100, 200, 500 and 1000), across different parameter values that included different number of binomial trials($n = 5$ and $n = 10$), to observe whether the estimators ($\hat{\pi}$ and $\hat{p}$) converge to the true values as the sample size increases.

Table 1. Simulation results across different sample sizes, assuming $n = 5$.

ine **Sample size**	π	p	n	Bias($\hat{\pi}$)	MSE($\hat{\pi}$)	Bias($\hat{p}$)	MSE($\hat{p}$)	S.E.($\hat{\pi}$)	S.E.($\hat{p}$)
ine 50	0.1	0.2	5	0.0482	0.0343	0.0156	0.0034	0.1788	0.0562
100	0.1	0.2	5	0.0336	0.0226	0.0105	0.0018	0.1465	0.0411
200	0.1	0.2	5	0.0170	0.0145	0.0058	0.0010	0.1192	0.0311
500	0.1	0.2	5	0.0055	0.0080	0.0023	0.0004	0.0893	0.0198
1000	0.1	0.2	5	-0.0006	0.0051	0.0009	0.0002	0.0714	0.0141
ine 50	0.3	0.5	5	-0.0101	0.0095	-0.0033	0.0031	0.0969	0.0556
100	0.3	0.5	5	-0.0061	0.0047	-0.0017	0.0015	0.0683	0.0387
200	0.3	0.5	5	-0.0022	0.0023	-0.0009	0.0007	0.0479	0.0264
500	0.3	0.5	5	-0.0012	0.0009	-0.0004	0.0003	0.0299	0.0173
1000	0.3	0.5	5	-0.0004	0.0004	-0.0004	0.0001	0.0199	0.0099
ine 50	0.8	0.7	5	-0.0024	0.0034	-0.0054	0.0065	0.0583	0.0804
100	0.8	0.7	5	-0.0018	0.0017	-0.0027	0.0028	0.0412	0.0528
200	0.8	0.7	5	-0.0004	0.0009	-0.0012	0.0013	0.0299	0.0360
500	0.8	0.7	5	-0.0004	0.0003	-0.0003	0.0005	0.0173	0.0224
1000	0.8	0.7	5	-0.0001	0.0002	0.0002	0.0003	0.0141	0.0173
ine									

Table 1 and 2 summarize the simulation results showing the bias, mean squared error (MSE) and standard error (S.E) for the estimators of π and p across different sample sizes. In all three scenarios, the estimators demonstrate consistent behavior: as the number of samples increases, both the bias and MSE decrease steadily. This indicates that the estimators are likely consistent and asymptotically unbiased.

For smaller sample sizes (e.g., $n = 50$), the bias and MSE are relatively higher, however, with larger sample sizes ($n = 500$ or 1000), both bias and MSE values approach zero, suggesting improved estimator accuracy and precision. Small fluctuations in bias at larger sample sizes are minor and can be attributed to simulation noise rather than a lack of consistency. Also, we can see that as the sample size increases, the S.E. tends to decrease.

Overall, the results support the conclusion that the estimators for π and p perform well as the sample size increases, both in terms of bias reduction and MSE minimization.

7. Data Analysis

To evaluate the practical applicability of our proposed distribution, we analyze two real-world datasets in this section. For comparison, we consider several established alternatives, including the Zero-Truncated Binomial (ZTB), One-Inflated Binomial (OIB), and Conway-Maxwell Binomial (CMB) ([13]) distributions.

Table 2. Simulation results across different sample sizes, assuming $n = 10$.

ine **Sample size**	π	p	n	Bias($\hat{\pi}$)	MSE($\hat{\pi}$)	Bias($\hat{p}$)	MSE($\hat{p}$)	S.E.($\hat{\pi}$)	S.E.($\hat{p}$)
ine 50	0.1	0.2	10	0.0048	0.0101	0.0021	0.0008	0.1004	0.0282
100	0.1	0.2	10	0.0009	0.0063	0.0010	0.0004	0.0794	0.0199
200	0.1	0.2	10	-0.0021	0.0038	0.0001	0.0002	0.0616	0.0141
500	0.1	0.2	10	-0.0023	0.0081	-0.0001	0.0001	0.0424	0.0099
1000	0.1	0.2	10	-0.0006	0.0009	0.0000	0.0000	0.0299	0.0000
ine 50	0.3	0.5	10	0.0004	0.0043	-0.0007	0.0008	0.0656	0.0283
100	0.3	0.5	10	-0.0009	0.0022	-0.0004	0.0004	0.0469	0.0199
200	0.3	0.5	10	-0.0002	0.0011	0.0000	0.0002	0.0332	0.0141
500	0.3	0.5	10	-0.0001	0.0004	-0.0001	0.0001	0.0199	0.0099
1000	0.3	0.5	10	0.0000	0.0002	0.0000	0.0000	0.0141	0.0000
ine 50	0.8	0.7	10	0.0002	0.0032	0.0004	0.0023	0.0566	0.0479
100	0.8	0.7	10	0.0003	0.0016	-0.0001	0.0011	0.0399	0.0332
200	0.8	0.7	10	0.0002	0.0008	-0.0001	0.0006	0.0283	0.0245
500	0.8	0.7	10	-0.0001	0.0003	0.0000	0.0002	0.0173	0.0141
1000	0.8	0.7	10	0.0001	0.0002	0.0000	0.0001	0.0141	0.0100
ine									

For each model, expected frequencies are computed based on the observed frequencies (OF) in the data. Model performance is then assessed using multiple criteria: the negative log-likelihood (-L), the corrected Akaike Information Criterion (AICc), and the Chi-square statistic (χ^2) along with its corresponding p-value. These metrics provide a basis for evaluating the goodness-of-fit across competing models.

7.1. Illustration I

The first dataset is obtained from the work of [14], who reports a study originally conducted by Edwards and Eberhardt in 1967 involving a live-trapping experiment on a confined population of 135 wild cottontail rabbits. The rabbits were kept in a 40-acre rabbit-proof enclosure, and live-trapping was carried out over 18 consecutive nights. The observed capture frequencies were as follows: 43 rabbits were captured once, 16 were captured twice, 8 three times, 6 four times, none five times, 2 six times, and 1 rabbit was captured seven times, resulting in a total of 76 captured individuals. Since the remaining 59 rabbits were never caught during the trapping period, the data naturally aligns with the structure of a one-inflated zero-truncated distribution, where zero captures are unobserved and the frequency of one capture is notably inflated (for detail see [14]). The characteristics of the data is represented by the non-parametric plot given in Figure 5.

Table 3 presents the goodness-of-fit results for the OIZTB model alongside other competing models. At the 0.05 significance level, the OIZTB model demonstrates the best overall fit to the dataset. Specifically, it yields the lowest AICc value and a chi-square statistic with a high p-value, indicating a good fit between the observed and expected frequencies. Also from Figure 6 we can say that our distribution provides the best fit.

7.2. Illustration II

As our second dataset, we use manuscript count data from [15] (cited in [16]), which records the survival of English medieval literary works. The dataset shows a notable excess of singletons, with 42 works surviving in only one manuscript, making it well-suited for modeling underrepresented cultural data (for detail see [15]). The characteristics of the data is represented by the non-parametric plot given in Figure 7.

Table 4 presents the goodness-of-fit results for the OIZTB model alongside other competing models. At the 0.05 significance level, the OIZTB model demonstrates the best overall fit to the dataset II. It yields the lowest

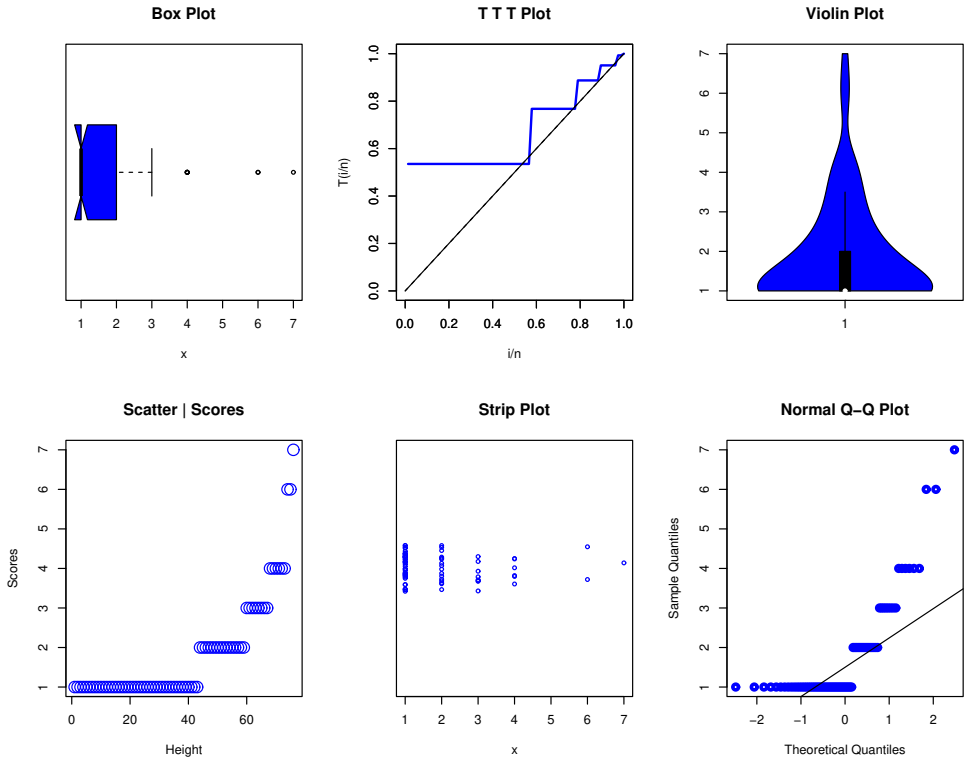

Figure 5. Non-parametric visualization plots for dataset I.

Table 3. The goodness-of-fit test for dataset I.

ineine x	OF	ZTB	OIB	CMB	OIZTB
ineine 1	43	31.97	46.73	21.43	44.97
2	16	27.06	11.53	22.12	13.42
3	8	12.72	9.33	14.13	10.46
4	6	3.59	4.53	5.95	4.89
5	0	0.61	1.32	1.65	1.37
6	2	0.06	0.21	0.28	0.21
7	1	0.01	2.33	10.42	0.67
ineine Total	76	76	76	76	76
ineine $-L$		111.17	102.48	122.14	100.92
ine MLEs		$\hat{q}$=0.78	$\hat{\pi}$=0.51	$\hat{p}$=0.29	$\hat{\pi}$=0.48
			$\hat{p}$=0.33	$\hat{\nu}$=0.82	$\hat{p}$=0.32
ine d.f		1	1	2	1
AICc		228.34	220.96	254.28	217.86
ine χ^2		8.33	2.26	33.14	1.64
p.value		0.003	0.133	¡0.0001	0.200
ineine					

AICc value and a chi-square statistic with a high p-value, indicating a good fit between the observed and expected frequencies. This can be also inferred from Figure 8 that our distribution provides the best fit.

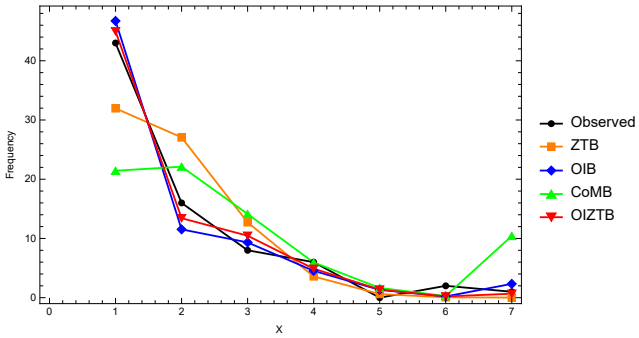

Figure 6. Plot of observed and expected frequency of ZTB, OIB, CoMB and OIZTB for dataset I.

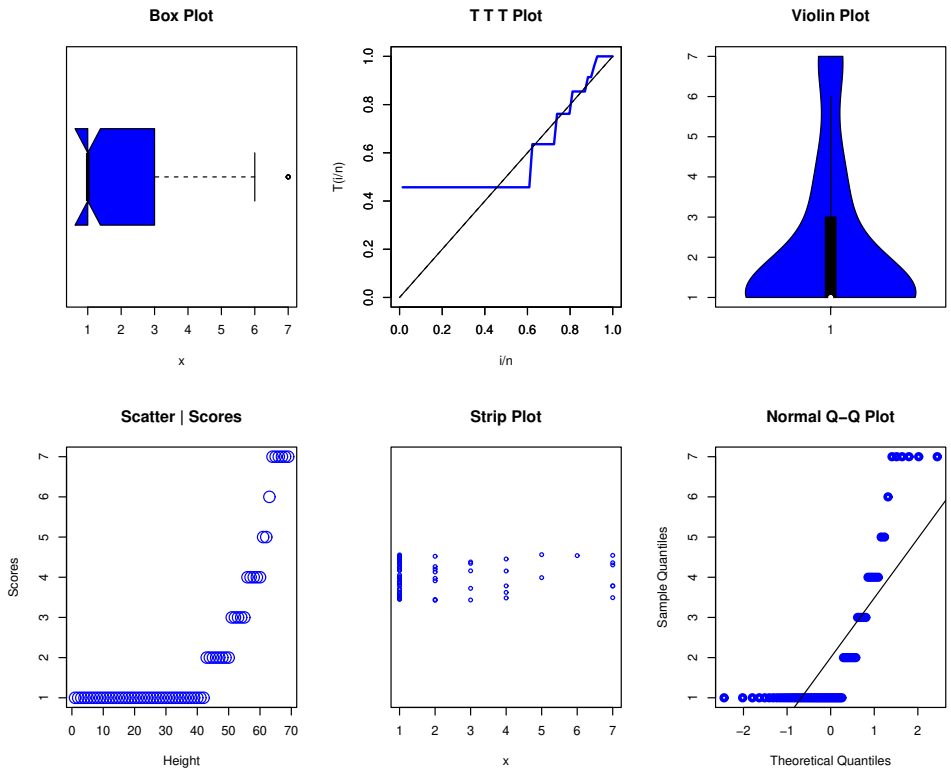

Figure 7. Non-parametric visualization plots for dataset II.

8. Conclusion

In this paper, we introduced a new discrete distribution specifically designed for count data exhibiting an excess of low-frequency counts, particularly singletons. The proposed distribution is flexible and capable of capturing asymmetry and varying degrees of kurtosis in the data. A key strength of the model is its ability to accommodate all three forms of dispersion—overdispersion, equidispersion, and underdispersion within a unified framework.

Table 4. The goodness-of-fit test for dataset II.

ineine x	OF	ZTB	OIB	CMB	OIZTB
ineine 1	42	20.72	41.99	14.85	42.00
2	8	24.39	2.89	12.86	2.95
3	5	15.95	6.27	10.75	6.34
4	5	6.26	8.14	7.24	8.17
5	2	1.47	6.35	4.77	6.32
6	1	0.19	2.75	2.85	2.72
7	6	0.01	10.51	16.35	0.50
ineine Total	69	69	69	69	69
ineine $-L$		144.31	106.33	130	106.21
ine MLEs		$\hat{q}$=0.72	$\hat{\pi}$=0.59	$\hat{p}$=0.41	$\hat{\pi}$=0.59
			$\hat{p}$=0.56	$\hat{\nu}$=0.17	$\hat{p}$=0.56
ine d.f		2	2	3	2
AICc		292.62	222.66	268	222.50
ine χ^2		45.20	10.52	64.59	10.18
p.value		¡0.0001	0.005	¡0.0001	0.006
ineine					

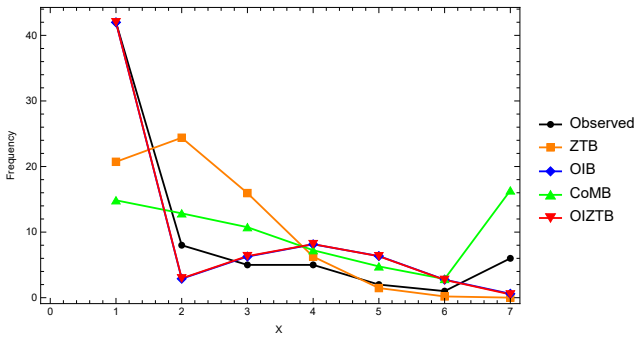

Figure 8. Plot of observed and expected frequency of ZTB, OIB, CoMB and OIZTB for dataset II.

We studied the MLE method and conducted a simulation study to evaluate the consistency and performance of the maximum likelihood estimators. To further validate the practical utility of the model, we applied it to two real-world datasets. In both cases, the proposed distribution outperformed competing models, demonstrating superior fit and adaptability.

However, one of the key limitations of this study is that we have assumed the number of trials (n) to be known. The data analysis is also conducted by taking the value of n to be known and here we have estimated only the parameters p and π. A simultaneous study can also be carried out taking n to be unknown and thus a need to estimate it. However, we have refrained ourselves from taking up this course as this area needs some more research. Future research directions may include extending the model to regression and hierarchical frameworks, enabling the incorporation of covariates and group-level effects. Additionally, a Bayesian formulation of the distribution could enhance inferential robustness and provide greater flexibility in modeling uncertainty.

Data and Code Availability

The datasets used in this research paper are accessible publicly and we have also cited the sources. The R code for analysis section is provided in the link https://github.com/skalpasree-debug/OIZTB.git.

REFERENCES

1. D. Lambert, *Zero-inflated poisson regression, with an application to defects in manufacturing*, Technometrics, vol. 34, no. 1, pp. 1–14, 1992.
2. D. C. Heilbron, *Zero-altered and other regression models for count data with added zeros*, Biometrical Journal, vol. 36, no. 5, pp. 531–547, 1994.
3. G. D. Barriga and F. Louzada, *The zero-inflated conway–maxwell–poisson distribution: Bayesian inference, regression modeling and influence diagnostic*, Statistical Methodology, vol. 21, pp. 23–34, 2014.
4. W. W. Mohammed, K. Sharma, P. J. Hazarika, G. Hamedani, M. S. Eliwa, and M. El-Morshedy, *Zero-inflated discrete lindley distribution: Statistical and reliability properties, estimation techniques, and goodness-of-fit analysis*, AIMS MATHEMATICS, vol. 10, no. 5, pp. 11382–11410, 2025.
5. F. David, and N. Johnson, *The truncated poisson*, Biometrics, vol. 8, no. 4, pp. 275–285, 1952.
6. M. Sampford, *The truncated negative binomial distribution*, Biometrika, vol. 42, no. 1/2, pp. 58–69, 1955.
7. M. E. Ghitany, D. K. Al-Mutairi, and S. Nadarajah, *Zero-truncated poisson–lindley distribution and its application*, Mathematics and Computers in Simulation, vol. 79, no. 3, pp. 279–287, 2008.
8. R. T. Godwin, and D. Böhning, *Estimation of the population size by using the one-inflated positive poisson model*, Journal of the Royal Statistical Society Series C: Applied Statistics, vol. 66, no. 2, pp. 425–448, 2017.
9. R. T. Godwin, *One-inflation and unobserved heterogeneity in population size estimation*, Biometrical Journal, vol. 59, no. 1, pp. 79–93, 2017.
10. Z. Zapata, S. A. Sedory, and S. Singh, *Zero-truncated binomial distribution as a randomization device*, Sociological Methods & Research, vol. 51, no. 2, pp. 800–815, 2022.
11. T. Rahman, P. Hazarika, and M. Barman, *One inflated binomial distribution and its real-life applications*, Indian Journal of Science and Technology, vol. 14, no. 22, pp. 1839–1854, 2021.
12. Gradshteyn, Izrail Solomonovich and Ryzhik, Iosif Moiseevich, *Table of integrals, series, and products*, Academic press, London, 2014
13. J. B. Kadane, *Sums of possibly associated bernoulli variables: The conway–maxwell-binomial distribution*, Bayesian Analysis, vol. 11, no. 2, pp. 403–420, 2016.
14. A. Chao, *Estimating the population size for capture-recapture data with unequal catchability*, Biometrics, vol. 43, no. 4, pp. 783–791, 1987.
15. M. Kestemont, F. Karsdorp, E. de Bruijn, M. Driscoll, K. A. Kapitan, P. Ó Macháin, D. Sawyer, R. Sleiderink, and A. Chao, *Forgotten books: The application of unseen species models to the survival of culture*, Science, vol. 375, no. 6582, pp. 765–769, 2022.
16. D. Böhning, and H. Friedl, *One-inflation and zero-truncation count data modelling revisited with a view on horvitz–thompson estimation of population size*, International Statistical Review, vol. 92, no. 3, pp. 406–430, 2024.