

# NEW DISCRETE DISTRIBUTION FOR ZERO-INFLATED COUNT DATA

Peer Bilal Ahmad<sup>1</sup> Mohammad Kafeel Wani<sup>2</sup>



<sup>1,2</sup>Department of Mathematical Sciences, Islamic University of Science & Technology, Kashmir,  
INDIA, 192122

<sup>1,\*</sup>[bilalahmadpz@gmail.com](mailto:bilalahmadpz@gmail.com) , <sup>2</sup>[wanimk5@gmail.com](mailto:wanimk5@gmail.com)

## Abstract

*Over-dispersed models are commonly utilized when the variation is more than what the model actually predicts. Since one of the reasons for over-dispersion is the large number of zeros, we employ zero-inflated models instead of more traditional ones to handle this observed occurrence. We present a zero-inflated version of a discrete distribution that was developed in 2021 in our research. Significant statistical characteristics of the suggested model have been identified, such as moments, the over-dispersion feature, generating functions, and related measures, among others. We have carried the parametric estimation using the maximum likelihood estimate. Maximum likelihood estimates are checked for usefulness in a simulation exercise. We evaluated the applicability of our developed model using three real-world data sets,*

**Keywords:** Over-dispersion, Zero-inflation, Discrete distribution, Simulation, Goodness-of-fit, Testing of hypothesis.

## I. Introduction

To perform statistical analysis, statisticians use one of several methods, and these methods are the building blocks of statistical models. Mathematical representations of observable data are provided by statistical models. We choose statistical modeling of data for the purpose of understanding a wide range of random events across disciplines. Its applications are not limited to mathematical and statistical studies; rather, they permeate a wide variety of fields of study. Count data plays an important role in almost every scientific study, no matter how big or small. This data is used to draw inferences in relation to the population from which it is collected but, typically this data exhibits more variation than what is predicted by our hypothesized model. More precisely, this observable fact is called as over-dispersion (variance goes beyond mean). One cause of over-dispersion in count data is the presence of many more zeros than predicted by a statistical model. This phenomenon of finding excessive number of zeros is referred to as zero-inflation, and in order to model such dilemma, we use zero-inflated models rather than the more often used standard models.

Over-dispersion in count data due to zero-inflation is common, thus researchers are always developing new ideas and methods to shed light on this phenomenon. In order to deal with an excessive amount of zeros in count data, Lambert developed a new model called as zero-inflated Poisson (ZIP) regression model [7]. She used this model to investigate manufacturing flaws and

found that ZIP regression model is both simple and effective. Böhning argued that ZIP distribution is usually capable of easily dealing with the situation where there is an excessive quantity of zero counts [3]. To deal with count data containing an excessive number of zeros and ones, Melkersson & Olsson offered an improved version of the ZIP distribution, which he named as the zero-one inflated Poisson (ZOIP) distribution [9]. Yau *et al.* presented a mixed regression model of zero-inflated Negative Binomial (ZINB) to examine pancreatic disorder Length of Stay (LOS) times that account for same-day discharges [17]. Gilthroe *et al.* took into account biological count data with an excessive quantity of zeros, and he sought to address variety of factors [5]. An overview of the field of statistical modeling of over-dispersed data was provided by discussing its antecedents, motivations, pioneering contributions, major milestones, and practical uses [16]. Zhang *et al.* made an effort to investigate the characteristics and patterns of ZOIP distribution [19]. In order to evaluate the capacity to incorporate zero-inflation and over-dispersion in count data, Pittman *et al.* evaluated a number of methods, including ZIP, ZINB, and Hurdle Poisson (HUP) regression model [10]. Tüzen *et al.* analyzed the implementation of count data models using simulated data, which allowed for a wide range of outliers and zero-inflation scenarios [14]. They considered Poisson, Negative-Binomial, ZIP, ZINB, HUP and Negative-Binomial Hurdle models to check the compatibility of these models in presence of outliers and excess zeros. Bodhisuwan & Kehler proposed a new distribution called the zero-inflated Negative-Binomial-Exponential (ZINB-E) distribution [2]. To address the issue of too many zeros in count data, Rivas & Campos introduced the zero-inflated Waring (ZIW) distribution [11]. If the Waring distribution can't sufficiently characterize the behavior of the data, as is often the case when there is a large frequency of observed zeros, then the ZIW distribution is thought to be a better fit. Young, Roemmele & Shi evaluated a study that provided a snapshot of the current level of knowledge in the field of zero-inflation [18]. Ahmad & Wani introduced a compound model for handling over-dispersed count data. They used four different data sets and compared the fit with several potential models of interest. The fitting results showed the flexibility of the devised model in handling over-dispersed count data [1]. One of the recent works in zero-inflation aspect of the count data is by Wani & Ahmad [15]. They put forward the zero-inflated version of Poisson-Akash distribution. Much advancement has been made in this field of statistical modeling, yet there is still a consistent need for new models to be created. These new models are driven by the regular emergence of unexpected patterns in count data. We intend to extend this contribution by a devising a new zero-inflated model with a very clear-cut Probability function.

## 2. Zero-Inflated Discrete Distribution

A discrete distribution (DD) was proposed by Jain *et al.* in 2021 by discretization of a continuous distribution [6]. If  $Z$  follows the DD, then the probability mass function (PMF) of the Discrete Distribution is given as follows

$$P(Z = z) = \frac{\theta - 1}{\theta^{z+1}} \quad ; \theta > 1, z = 0, 1, 2, 3, \dots \quad (1)$$

The Discrete distribution (1) is itself an over-dispersed model but it also suffers at times to handle the excessive number of zeros in count data. We have thus made an effort to put forward the zero-inflated version of Discrete distribution. If  $X$  is a random variable following the Discrete distribution with parameter  $\theta > 1$  and  $\alpha$  ( $0 < \alpha < 1$ ) is the extra amount added to the proportion of point zero (zero-inflated distribution), then the probability mass function (PMF) of zero-inflated Discrete distribution (ZIDD) can be written as follows

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

The Cumulative distribution function (CDF) of ZIDD can be expressed as

$$F_x(X) = 1 - (1 - \alpha) \frac{1}{\theta^{x+1}} \quad (3)$$

It can be seen from the plots of PMF given in Fig. 1 that the model has mode at point zero. Furthermore, it is positively skewed and the tail shows a rapid decrease as parameters take higher values.

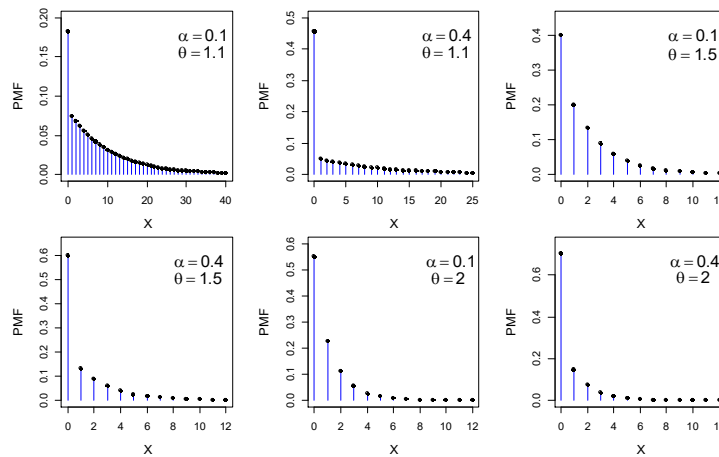


Fig. 1: Plots of Probability mass function of ZIDD for different choices of parameter values

### 3. Statistical Properties

In this section, we have derived some vital statistical characteristics of the newly developed model.

#### 3.1 Generating Functions

When dealing with discrete random variables, the Probability Generating Function (PGF) is an important tool. Its major benefit is that it makes it straightforward for us to explain the distribution of  $X+Y$  when they are independent. The PGF of ZIDD can be obtained as

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

$$P_x(t) = t^0 \left( \alpha + (1 - \alpha) \frac{\theta - 1}{\theta} \right) + (1 - \alpha) \left( \frac{\theta - 1}{\theta} \right) \sum_{x=1}^{\infty} \left( \frac{t}{\theta} \right)^x$$

$$P_x(t) = \alpha + (1 - \alpha) \left( \frac{\theta - 1}{\theta} \right) + (1 - \alpha) \left( \frac{\theta - 1}{\theta} \right) \left( \frac{t}{\theta} \right) \left( \frac{\theta}{\theta - t} \right)$$

$$P_x(t) = \alpha + (1 - \alpha) \left( \frac{\theta - 1}{\theta - t} \right) \tag{4}$$

In equation (4), take  $t = e^t$ , that will yield the Moment Generating Function of ZIDD as follows

$$M_x(t) = \alpha + (1 - \alpha) \left( \frac{\theta - 1}{\theta - e^t} \right) \tag{5}$$

### 3.2 Moments and Related Characteristics

The  $r$ th moment about origin (Raw Moment) of ZIDD is obtained by employing its PMF (2). It follows that

$$\begin{aligned} E(X^r) &= \sum_{\forall x} x^r P(X = x) \\ E(X^r) &= 0^r [\alpha + (1 - \alpha)P_{DD}(X = 0)] + 1^r [(1 - \alpha)P_{DD}(X = 1)] + 2^r [(1 - \alpha)P_{DD}(X = 2)] + \dots \\ E(X^r) &= (1 - \alpha) \sum_{x=0}^{\infty} X^r [P_{DD}(X = x)] \\ E(X^r) &= (1 - \alpha)E_{DD}(X^r) \end{aligned}$$

With  $P_{DD}(X=x)$  and  $E_{DD}(X^r)$  representing the PMF and the  $r$ th Raw Moment of the baseline model respectively.

As a result, the mean and variance of ZIDD comes out to be as follows

$$\text{Mean} = (1 - \alpha) \frac{1}{\theta - 1} \qquad \text{Variance} = (1 - \alpha) \frac{\alpha + \theta}{(\theta - 1)^2}$$

Some of the statistical properties of our proposed model can be expressed by means of Raw and Central moments. These properties include Index of Dispersion (IOD), Coefficient of Variation (C.V), Coefficient of Skewness and Coefficient of Kurtosis.

$$\begin{aligned} IOD &= \frac{\sigma^2}{\mu_1'} = \frac{\alpha + \theta}{\theta - 1} & C.V &= \frac{\sigma}{\mu_1'} = \left( \frac{\alpha + \theta}{1 - \alpha} \right)^{1/2} \\ Skewness &= \frac{\mu_3}{\mu_2^{3/2}} = \frac{\theta^2 + \theta + 3\alpha\theta + 2\alpha^2 - \alpha}{\sqrt{1 - \alpha}(\alpha + \theta)^{3/2}} \\ Kurtosis &= \frac{\mu_4}{\mu_2^2} = \frac{\theta^3 + 7\theta^2 + \theta + 2 + 4\alpha\theta^2 + 6\alpha^2\theta + 4\alpha\theta + \alpha^3 + 3\alpha^2 - 5\alpha}{(1 - \alpha)(\alpha + \theta)^2} \end{aligned}$$

From the plots of Coefficient of Skewness and Kurtosis given in Fig. 3, it can be noted that both of these increase monotonically for greater values of the parameter. Moreover, our proposed model possesses positive skewness and a leptokurtic shape.

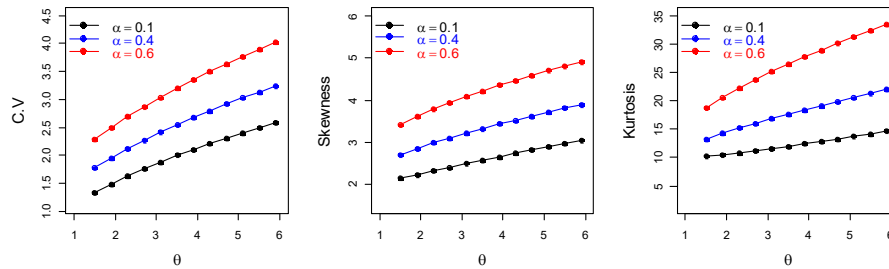


Fig. 2: Plots of Coefficient of Variation, Skewness and Kurtosis for some values of parameters

A significant characteristic enjoyed by our proposed model lies in the fact that it is always over-dispersed i.e., variance is always going to surpass the mean. We have

$$Varinace = (1-\alpha) \frac{\alpha + \theta}{(\theta - 1)^2} = Mean + (1-\alpha) \frac{(1+\alpha)}{(\theta - 1)^2}$$

The second term is obviously positive as  $\theta > 1$  and  $\alpha$  ( $0 < \alpha < 1$ ). This proves the over-dispersion property of ZIDD. The Index of Dispersion has also been plotted (see Fig. 3) for a choice of parameter values, which graphically demonstrates the over-dispersion of the model.

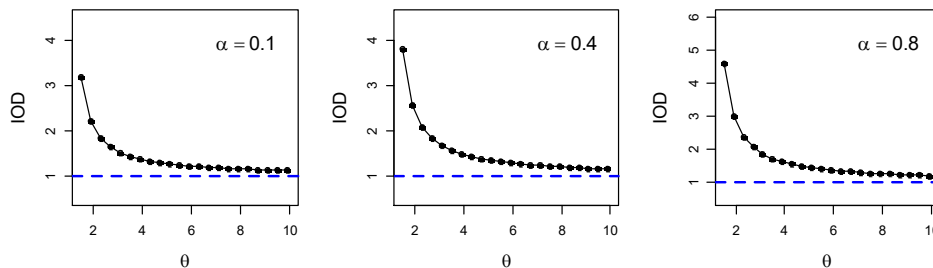


Fig. 3: Index of Dispersion plots for various values of parameters

#### 4. Parametric Inference

The foundation for estimation is in actual fact clear-cut. Knowing the parameter put forwards information concerning the entire population when sampling is done from a population that is represented by a specific distribution. So, it makes sense to carry out the estimation of parameters. The maximum likelihood estimation (MLE) is more often used for the reason that it enjoys greater efficiency and improved numerical stability. MLE is a statistical technique for estimating the parameters of a probability distribution that is assumed given some observed data.

The likelihood function of ZIDD can be defined as follows if  $x_1, x_2, \dots, x_n$  is the random sample of size  $n$  from ZIDD and  $Y$  is the number of  $x_i$ 's having value zero

$$L = \left( \alpha + (1-\alpha) \frac{\theta - 1}{\theta} \right)^Y \prod_{\substack{i=1 \\ x_i \neq 0}}^n (1-\alpha) \frac{\theta - 1}{\theta^{x_i + 1}}$$

The log-likelihood can thus be written as

$$\log L = Y \log(\alpha + \theta - 1) - (n + Y + \sum_{\substack{i=1 \\ x_i \neq 0}}^n x_i) \log \theta + (n - Y) \log(1 - \alpha) + (n - Y) \log(\theta - 1)$$

In this study we have used *firdistrplus* in R-software to obtain the ML estimates [4].

### 5. Simulation Study

A simulation study has been undertaken in this part to evaluate the finite sample performance of the ML estimates of ZIDD. We attempt the Monte Carlo Simulation study by employing the *discrete inverse transform method*. In order to calculate Average Values (AVs), Average Biases (ABs), Mean Square Errors (MSEs), Mean Relative Estimates (MRESs), Mean Relative Errors (MRERs), and Average Dispersion Indices (AVDIs), we considered four different values for parameter and repeated the course of action N=1000 times starting from a small sample to large sample (n=25, 75, 100, 300, 600). The results are provided in Table 1. As it can be seen from Table 1, the ML estimates are asymptotically unbiased and consistent.

**Table 1:** Simulation results for maximum likelihood estimates of parameters of proposed model

	$\hat{\theta}$					$\hat{\alpha}$				
	AVs	ABs	MSEs	MRESs	MRERs	AVs	ABs	MSEs	MRESs	MRERs
n	Parameter Set 1: $\theta=1.5$ $\alpha=0.4$									
25	1.614	0.114	0.175	1.076	0.163	0.387	0.012	0.035	0.968	0.3903
75	1.508	0.008	0.011	1.005	0.055	0.401	0.001	0.010	1.004	0.2049
100	1.529	0.029	0.012	1.019	0.056	0.385	0.014	0.007	0.984	0.1713
300	1.503	0.003	0.002	1.002	0.024	0.398	0.001	0.002	0.996	0.0876
600	1.504	0.004	0.001	1.002	0.021	0.395	0.004	0.000	0.998	0.0213
n	Parameter Set 2: $\theta=1.8$ $\alpha=0.4$									
25	1.959	0.159	0.527	1.088	0.214	0.380	0.019	0.049	0.950	0.462
75	1.856	0.056	0.046	1.031	0.094	0.396	0.003	0.018	0.992	0.263
100	1.793	0.006	0.036	0.996	0.087	0.387	0.012	0.011	0.967	0.221
300	1.809	0.009	0.009	1.005	0.039	0.396	0.003	0.003	0.991	0.114
600	1.809	0.009	0.007	1.005	0.039	0.403	0.003	0.001	1.007	0.089
n	Parameter Set 3: $\theta=2.0$ $\alpha=0.2$									
25	2.238	0.238	0.574	1.110	0.235	0.172	0.027	0.034	0.862	0.819
75	2.028	0.028	0.066	1.014	0.089	0.186	0.013	0.017	0.933	0.532
100	2.022	0.022	0.064	1.011	0.090	0.205	0.005	0.015	1.027	0.528
300	2.017	0.017	0.024	1.008	0.0621	0.188	0.011	0.007	0.943	0.329
600	1.994	0.005	0.006	0.997	0.0333	0.209	0.009	0.002	1.049	0.186
n	Parameter Set 4: $\theta=2.5$ $\alpha=0.3$									
25	2.674	0.174	1.501	1.069	0.260	0.288	0.011	0.054	0.960	0.667
75	2.495	0.044	0.204	0.998	0.157	0.306	0.006	0.038	1.023	0.565
100	2.542	0.042	0.171	1.016	0.130	0.306	0.006	0.017	1.021	0.354
300	2.565	0.035	0.083	1.026	0.095	0.296	0.003	0.012	0.988	0.317
600	2.511	0.011	0.027	1.004	0.055	0.299	0.001	0.005	0.997	0.190

### 6. Data Fitting

Real life datasets from different fields have been employed to test the compatibility of our proposed model in presence of over-dispersion caused by zero-inflation. In addition to this, we compared the fitting results from our proposed model with other statistical models of competing interest. The models with which we have compared our devised models include Poisson distribution (PD), zero-inflated Poisson distribution (ZIPD), zero-inflated Negative-Binomial

distribution (ZINBD), Discrete Weibull distribution (DWD) and Discrete distribution (DD). In order to estimate the parameters of each distribution, we used maximum likelihood estimation method.

### 6.1 Data set 1

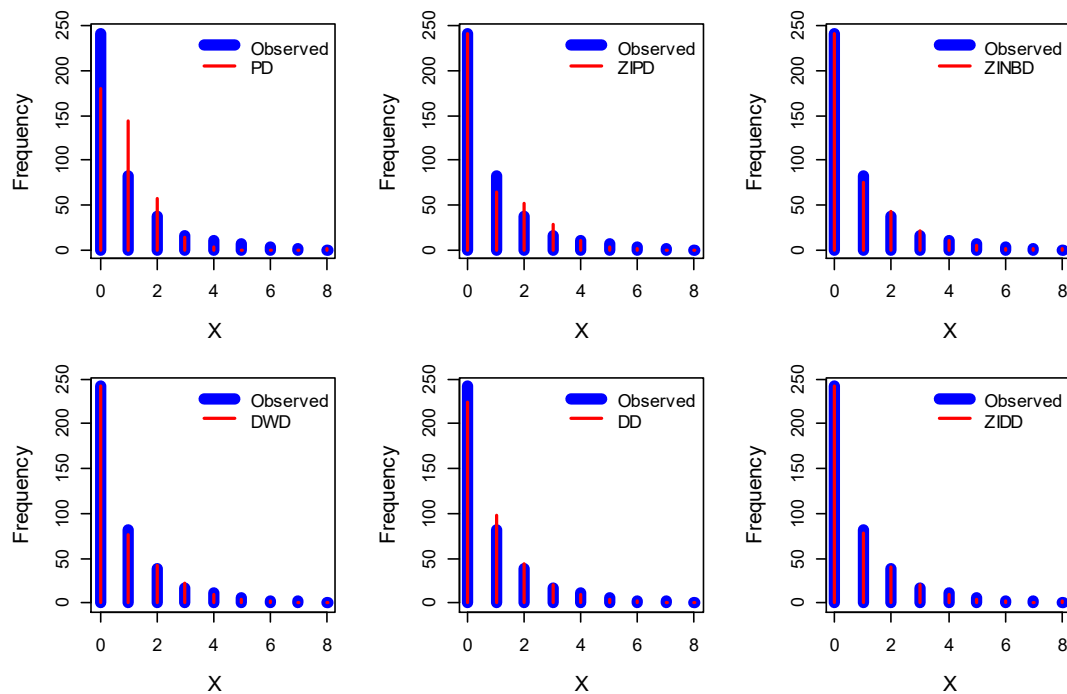
This dataset stands for the observed number of households according to total number of migrants [13]. The data is expressed in Table 2 and the performance of fitting this data is summarized in Table 3. From the fitting results, it is obvious that our model performs better than other competing models of interest. The plots for observed an expected frequencies under different models given in Fig. 4 provides a clearer view of the fitting results.

**Table 2: Data set 1**

Number of Households	0	1	2	3	4	5	6	7	8
Observed Frequency	242	82	38	17	11	7	3	2	0

**Table 3: Fitting results of Data set 1**

Model	$\chi^2$	d.f	p-value	L	AIC	BIC
PD	118.64	3	0.0001	-555.060	1112.121	1116.117
ZIPD	16.80	2	0.0002	-501.796	1007.592	1015.585
ZINBD	3.29	2	0.1930	-492.597	991.195	1003.185
DWD	1.57	4	0.8141	-492.362	988.724	996.717
DD	9.41	4	0.0516	-495.785	993.571	997.567
ZIDD	1.51	4	0.8248	-492.030	988.060	996.053



**Fig. 4: Observed and expected frequencies plots for PD, ZIPD, ZINBD, DWD, DD, and ZIDD for Data set 1**

### 6.2 Data set 2

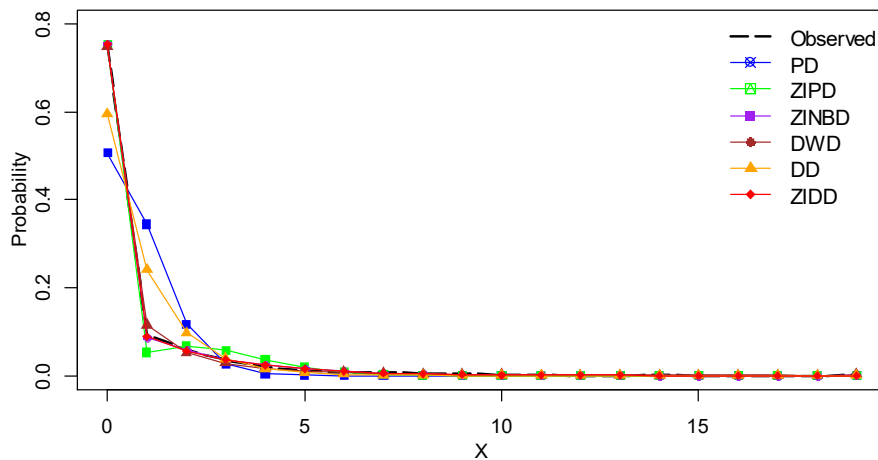
This dataset presents the number of spots in southern beetle [8]. The data is presented in Table 4 and the fitting results are given in Table 5. The performance measures indicate that our model suffers minimum loss compared to other models and the value of Chi-square is comparatively smaller. Moreover, the plots for observed and expected PMFs are given in Fig. 5.

**Table 4:** Data set 2

Number of Spots	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Observed Frequency	1169	144	92	54	29	18	10	12	6	9	3	2	0	0	1	0	0	0	0	1

**Table 5:** Fitting results of Data set 2

Model	$\chi^2$	d.f	p-value	L	AIC	BIC
PD	1361.80	4	0.0001	-2291.139	4584.277	4589.623
ZIPD	184.65	5	0.0001	-1648.989	3301.978	3312.670
ZINBD	9.94	6	0.1272	-1554.612	3115.223	3131.261
DWD	14.30	7	0.0460	-1560.532	3125.064	3135.756
DD	432.62	5	0.0001	-1757.427	3516.855	3522.201
ZIDD	7.96	7	0.3361	-1554.001	3112.002	3122.694



**Fig. 5:** Plots of observed and expected PMFs under PD, ZIPD, ZINBD, DWD, DD, and ZIDD for Data set 2

### 6.3 Data set 3

This dataset represents the number of units of Brand K of Chatfield bought by numbers of consumers over a number of weeks [12]. The data set is given in Table 6 and the Table 7 presents



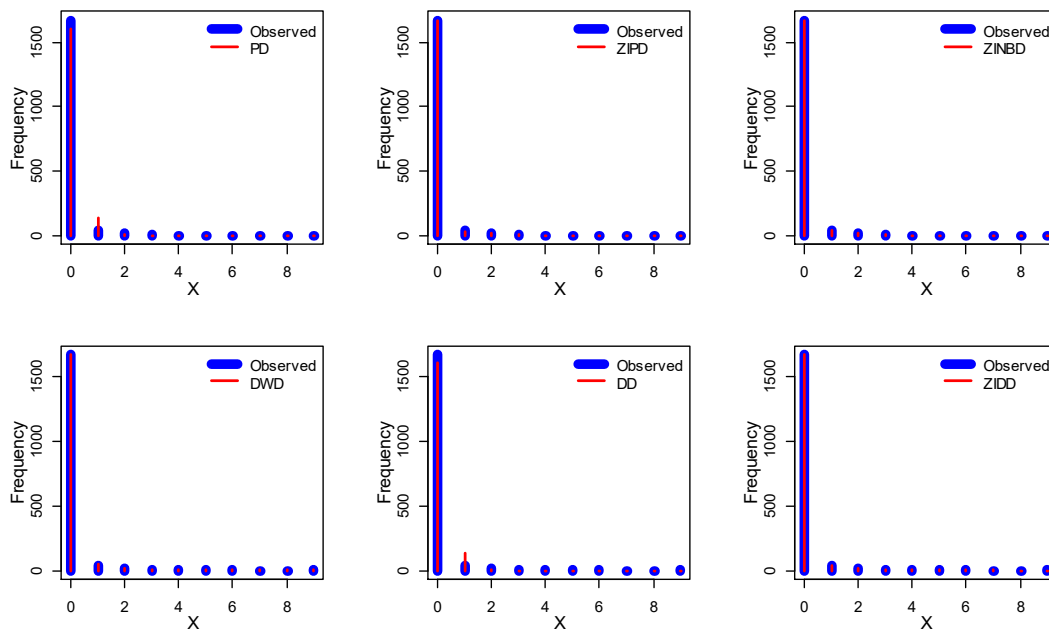
the fitting results of this data set. The results from the fitting table prove that the devised model shows better fitting compared to other competing models. In addition to this, the expected frequencies are closer to observed frequencies in case of our model (see Fig. 6).

**Table 6:** Data set 3

Brand K	0	1	2	3	4	5	6	7	8	9
Number of Consumers	1671	43	19	9	2	3	1	0	0	2

**Table 7:** Fitting results of data set 3

Model	$\chi^2$	d.f	p-value	L	AIC	BIC
PD	221.99	1	0.0001	-612.908	1227.8171	1233.2840
ZIPD	6.28	2	0.0433	-439.794	883.5894	894.5241
ZINBD	1.40	1	0.0942	-429.660	865.3218	881.7239
DWD	2.48	2	0.2893	-429.407	862.8156	873.7504
DD	118.40	1	0.0001	-537.381	1076.763	1082.230
ZIDD	0.48	2	0.7866	-429.334	862.6682	873.6030



**Fig. 6:** Observed and expected frequency plots for Data set 3 under PD, ZIPD, ZINBD, DWD, DD, and ZIDD

## 7. Testing of Hypothesis

In order to test the significance of the zero-inflation parameter of our proposed model, we take on different test statistics to test the null hypothesis given as follows

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

### 7.1 Likelihood Ratio test

The Likelihood ratio test (LRT) evaluates the ratio of two log-likelihood functions in order to test

the null hypothesis  $H_0$  against an alternative hypothesis  $H_1$ . In case of LRT, the test statistic is

$$LRT_{\alpha} = -2 \times [L(\hat{\theta}) - L(\hat{\alpha}, \hat{\theta})],$$

where  $L(\hat{\theta})$  and  $L(\hat{\alpha}, \hat{\theta})$  are the maximum log-likelihood under DD and ZIDD respectively. The LRT test statistic is asymptotically distributed as Chi-square with one degree of freedom. The LRT for all the three datasets is respectively given as

$$LRT_1=7.51, \quad LRT_2=406.852, \quad LRT_3=216.094.$$

## 7.2 Wald test

This test is used to determine the presence or absence of an effect. In this section, we will construct a Wald test for the effect of zero-inflation parameter in our proposed model. The test statistic under Wald test is given by

$$Wald_{\alpha} = \frac{\hat{\alpha}^2}{Var(\hat{\alpha})},$$

where  $Var(\hat{\alpha})$  represents the pertinent diagonal component of the Fisher information matrix calculated at  $\alpha = \hat{\alpha}$  and  $\theta = \hat{\theta}$ . The Wald test statistic is asymptotically distributed as Chi-square with one degree of freedom. The Wald test statistic value for all the three datasets can be correspondingly given as

$$Wald_1=9.53, \quad Wald_2=1014.12, \quad Wald_3=517.44.$$

On comparing the LRT and Wald test values from all data sets with the critical value (3.84), we reject the null hypothesis in case of all the three tests and draw the conclusion that the zero-inflation parameter in our proposed model is of significant importance.

## 8. Conclusion

In this research we have made an effort to present a new zero-inflated count data model. It was investigated how the probability mass function behaves for varied values of parameters. We discussed some important statistical properties of our proposed model. Simulation study was carried out to test the performance of maximum likelihood estimates and the results were pretty much significant. For the testing the compatibility of our proposed model, we tested the proposed distribution on real datasets using different performance metrics like Chi-square Goodness-of-fit, AIC, BIC etc. Moreover, we compared the fitting results of our devised model with other competing models. The results verified that our proposed model is adaptable and can be considered for handling over-dispersion in count data caused by zero-inflation. Finally, we carried out the Likelihood Ratio test and the Wald test on all datasets to see the significance of zero-inflation parameter and the results were significant.

## Acknowledgements

We are highly thankful to the editor-in-chief and the referees. The second author is particularly thankful to the Department of Science and Technology (Government of India) for INSPIRE fellowship (DST/INSPIRE/03/2022/002460).

## Conflict of Interest

The authors report no conflict of interest.

## References

- [1] Ahmad, P. B., & Wani, M. K. (2023). A New Compound Distribution and Its Applications in Over-dispersed Count Data. *Annals of Data Science*. <https://doi.org/10.1007/s40745-023-00478-0>
- [2] Bodhisuwan, R., & Kehler, A. (2021). The Zero-inflated Negative Binomial-Exponential Distribution and Its Application. *Lobachevskii Journal of Mathematics*, 42(2), 300–307. <https://doi.org/10.1134/s1995080221020062>
- [3] Böhning, D. (1998). Zero-Inflated Poisson Models and C.A.MAN: A Tutorial Collection of Evidence. *Biometrical Journal*, 40(7), 833–843. [https://doi.org/10.1002/\(sici\)1521-4036\(199811\)40:7<833::aid-bimj833>3.0.co;2-o](https://doi.org/10.1002/(sici)1521-4036(199811)40:7<833::aid-bimj833>3.0.co;2-o)
- [4] Delignette-Muller, M. L., & Dutang, C. (2015). Fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4). <https://doi.org/10.18637/jss.v064.i04>
- [5] Gilthorpe, M. S., Frydenberg, M., Cheng, Y., & Baelum, V. (2009). Modelling count data with excessive zeros: The need for class prediction in zero-inflated models and the issue of data generation in choosing between zero-inflated and generic mixture models for dental caries data. *Statistics in Medicine*, 28(28), 3539–3553. Portico. <https://doi.org/10.1002/sim.3699>
- [6] Jain, S., Siddiqui, S. A., Dwivedi, S., Siddiqui, I., & Kamal, M. (2021). A NEW DISCRETE DISTRIBUTION WITH ITS MATHEMATICAL PROPERTIES. *Int. J. Agricult. Stat. Sci. Vol*, 17(2), 693–697.
- [7] Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1), 1. <https://doi.org/10.2307/1269547>
- [8] Lin, S. K. (1985). Characterization of lightning as a disturbance to the forest ecosystem in East Texas (Doctoral dissertation, Texas A & M University).
- [9] Melkersson, M., & Olsson, C. (1999). Is visiting the dentist a good habit?: Analyzing count data with excess zeros and excess ones. *Univeristy of Umeå*.
- [10] Pittman, B., Buta, E., Krishnan-Sarin, S., O'Malley, S. S., Liss, T., & Gueorguieva, R. (2018). Models for Analyzing Zero-Inflated and Overdispersed Count Data: An Application to Cigarette and Marijuana Use. *Nicotine & Tobacco Research*, 22(8), 1390–1398. <https://doi.org/10.1093/ntr/nty072>
- [11] Rivas, L., & Campos, F. (2021). Zero inflated Waring distribution. *Communications in Statistics – Simulation and Computation*, 1–16. <https://doi.org/10.1080/03610918.2021.1944638>
- [12] Shoukri, M. M., & Consul, P. C. (1987). Some Chance Mechanisms Generating the Generalized Poisson Probability Models. *Biostatistics*, 259–268. [https://doi.org/10.1007/978-94-009-4794-8\\_15](https://doi.org/10.1007/978-94-009-4794-8_15)
- [13] Shukla, K. K., Shanker, R., & Tiwari, M. K. (2021). A new one parameter discrete distribution and its applications. *Journal of Statistics and Management Systems*, 25(1), 269–283. <https://doi.org/10.1080/09720510.2021.1893475>
- [14] Tüzen, F., Erbaş, S., & Olmuş, H. (2018). A simulation study for count data models under varying degrees of outliers and zeros. *Communications in Statistics – Simulation and Computation*, 49(4), 1078–1088. <https://doi.org/10.1080/03610918.2018.1498886>
- [15] Wani, M. K., & Ahmad, P. B. (2023). Zero-inflated Poisson-Akash distribution for count data with excessive zeros. *Journal of the Korean Statistical Society*. <https://doi.org/10.1007/s42952-023-00216-5>
- [16] Kekalaki, E. (2014). On the distribution theory of over-dispersion. *Journal of Statistical Distributions and Applications*, 1(1). <https://doi.org/10.1186/s40488-014-0019-z>
- [17] Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. *Biometrical Journal*, 45(4), 437–452. <https://doi.org/10.1002/bimj.200390024>

[18] Young, D. S., Roemmele, E. S., & Shi, X. (2022). Zero-inflated modeling part II: Zero-inflated models for complex data structures. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(2), e1540. <https://doi.org/10.1002/wics.1540>

[19] Zhang, C., Tian, G.-L., & Ng, K.-W. (2016). Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods. *Statistics and Its Interface*, 9(1), 11–32. <https://doi.org/10.4310/sii.2016.v9.n1.a2>