A NEW ZERO-INFLATED COUNT MODEL WITH APPLICATIONS IN MEDICAL SCIENCES

Zehra Skinder[1]   Peer Bilal Ahmad*[2]   Na Elah[3]

•

Department of Mathematical Sciences,
Islamic University of Sciene & Technology-192122, Kashmir[1,2,3]
zehraskinder7@gmail.com
bilalahmadpz@gmail.com*
naaelashah@gmail.com

Abstract

Inflated models are used whenever there are too many frequencies at a given count. In this regard, Poisson moment exponential distribution and a distribution to a point mass at zero are used to create a zero-inflated model namely Zero-Inflated Poisson Moment Exponential Distribution. Its distributional and reliability characteristics are investigated in some detail. A simulation exercise is undertaken to evaluate the effectiveness of the maximum likelihood estimators. The adaptability of the suggested distribution is demonstrated using three real datasets from various domains (e.g., vaccine adverse events, medical science data, epileptic seizure counts). The suggested distribution and the Poisson moment exponential distribution are distinguished by using the two different test procedures.

Keywords: goodness of fit, poisson Moment exponential distribution, likelihood ratio test, zero-inflation, wald test

1. Introduction

Statistical modelling of count observations is an essential part in several areas of scientific research. Frequent zeros in count observations are so common in areas like ecology, epidemiology, public health, engineering, etc. Examples of such data includes the number of foetal movements count per 5 seconds by Leroux *et.al* [16], number of HIV infected patients count by Van den Broek [32] in 1995, number of migrants count at household level by Shukla *et.al*[28] in 2006, number of accidents count due to heavy vehicular traffic for the year 2010 by Sharma *et.al* [27] in 2013, number of suicide cases count due to COVID-19 in India by Rahman *et.al* [23] in 2022, number of antenatal care service visits count by Bekalo *et.al* [4] in 2021. In order to model count observations with frequent zeros, number of Zero-Inflated models have been studied in the literature. The idea of zero-inflation was first given by Neyman [22] in 1939 and feller [9] in 1943 to overcome the situation of more zeros. Zero-inflated Poisson distribution (ZIPD) introduced by Mullahy [20] in 1986 as a mixture between Poisson distribution and a distribution at point mass zero with mixing probability (δ). The probability mass function of the distribution X (δ, ζ) is as follows.

$$P(x,\delta,\zeta) = \begin{cases} \delta + (1-\delta)e^{-\zeta}; x=0 \\ (1-\delta)\frac{e^{-\zeta}\zeta^x}{x!}; x>0 \end{cases}$$

where δ is a zero-inflation parameter (0<δ<1), $\zeta \geq 0$ and if δ=0, the distribution reduces to Poisson distribution. Several authors investigated ZIPD such as Singh [30] in 1962, Martin and Katti [17] in 1965, Goralski [10] in 1977, Lambert [14] in 1992, Bohning [5] in 1998 and

sim *et.al* [29] in 2018. Gupta [11] developed a generalized version of Zero-inflated Poisson model called as zero adjusted generalized Poisson model. The parameters of the zero-inflated Poisson model were estimated by Nanjundan and Naika [21] in 2012 by using moment method of estimation and compared with maximum likelihood estimators. Beckett *et.al* [3] used some natural calamities data to study the zero-inflated Poisson model and compared moment method and maximum likelihood method of estimation. Hall [12] in 2000 introduced the zero-inflated binomial distribution. Zero-inflated negative binomial distribution (ZINBD) was distinguished by Suresh *et.al* [31] in 2015. Zero-inflated negative binomial distribution studied by Mwalili *et.al* [19] in 2008 to accommodate extravagant zeros. Ahmad *et.al* [1] in 2014 studied the zero-inflated generalized power series distribution using Bayes estimators of functions of parameters under varied loss functions. Sandhyaa *et.al* [25] in introduced a model called Inflated-parameter Harris Distribution. Several structural properties were explored and characterization on the basis of probability generating function was also given. To check the applicability of the model, real life-data was also considered. Junnumtuam *et.al* [13] introduced a new discrete distribution called the Zero-Inflated Cosine Geometric (ZICG) Distribution for modelling over dispersed data with excessive zeros. Various structural properties like moment generating function, mean and variance were also obtained. Furthermore, confidence interval was also constructed by using the Wald method. The Bayesian method with highest posterior density method was also used to estimate the true confidence intervals. Dara and Ahmad [7] in 2012 introduced the Moment exponential distribution (ME) by weighting the exponential distribution in conformity with Fishers theory (1934). Scollnik [26] in 2022 obtained Bayesian analyses of an exponential-Poisson and related zero augmented type models. Maya *et.al* [18] in 2023 analysed the applications of Poisson moment exponential distribution in the contexts of time series analysis and regression analysis for real world phenomenon. Ahsan-ul-Haq [2] in 2022 introduced the Poisson moment exponential distribution (PMExD) by combining the Moment exponential and Poisson distribution by compounding technique and showed that the model is over dispersed and flexible for statistical data analysis. The probability mass function (p.m.f) of the PMExD is as follows.

$$P(Y = y) = \frac{\zeta^y (1 + y)}{(1 + \zeta)^{2+y}},$$

Where y=0, 1, 2, 3, ..., and $\zeta > 0$. The PMExD has been found in immense applications in various fields of medical sciences, engineering, entomology and education.

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

This paper is organized as follows. In section 2, we show the derivation of the ZIPMExD, cumulative distribution function. Also, the shapes of probability mass function (p.m.f) and cumulative distribution function (c.d.f) are presented in this section. In section 3, we have obtained the various structural properties along with reliability characteristics and generating functions. In section 4, we discuss the estimation of the parameters of the ZIPMExD by two different methods. A rigorous simulation study is also discussed in this section. In section 5, to check the significance of the inflation parameter, different test procedures are applied for examination. Certain real life data applications are considered in section 6 for highlighting the functionality of the model. Also, it is important to highlight that zero-inflated version of PMExD is not studied yet in the literature.

2. Zero-Inflated Poisson Moment Exponential Distribution

In this part, we present Zero-Inflated Poisson Moment Exponential distribution (ZIPMExD). We have derived the propability mass function of the proposed model along with cumulative distribution function.

Theorem 1. Let Y $\sim$ ZIPMExD (π, δ). Then the probability mass function (p.m.f) of Y is given as

$$P(Y=y) = \begin{cases} \pi + (1-\pi)\frac{1}{(1+\delta)^2} & ; y = 0 \\ \\ (1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} & ; y = 1,2,3,..., \end{cases}$$

Proof: If Y is a random variable of Poisson Moment Exponential distribution, then the probability mass function (p.m.f) of Y can be defined as

$$h(y) = \frac{\delta^y(1+y)}{(1+\delta)^{y+2}} \qquad ; y = 0,1,2,...,; \delta > 0$$

The Zero-inflated distribution is an extra proportion added to the proportion of zero, then the probability mass function of Zero-inflated distribution is given as

$$p(Y=y) = \begin{cases} \pi + (1-\pi)h(y=0) & ; y = 0 \\ (1-\pi)h(y) & ; y = 1,2,... \end{cases}$$

where $0 < \pi < 1$.

Then, the p.m.f of the ZIPMExD(π, δ) is obtained by substituting the probability mass function of the Poisson moment exponential random variable into Zero-inflated model. Therefore, it can be written as

$$\begin{cases} \pi + (1-\pi)\frac{1}{(1+\delta)^2} & ; y = 0 \\ \\ (1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} & ; y = 1,2,3,..., \end{cases} \tag{1}$$

where $0 < \pi < 1$ and $\delta > 0$

The Cumulative Distribution Function of ZIPMExD (π, δ) can be given as

$$F(Y) = P(Y \le y)$$

$$= \sum_{z=0}^{y} P(Y=z)$$

$$= \pi + \frac{(1-\pi)}{(1+\delta)^2} \sum_{z=0}^{y} \frac{\delta^z(1+z)}{(1+\delta)^z}$$

$$= [1 - (1-\pi)(y+\delta+2)\delta^{y+1}(1+\delta)^{-(y+2)}] \tag{2}$$
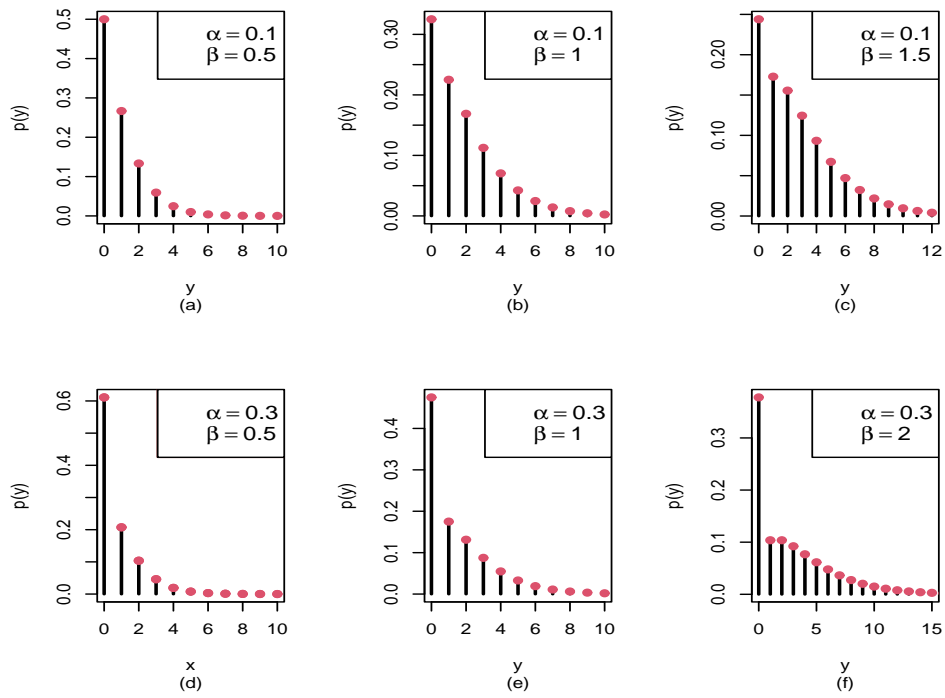
Figure 1: *The Pmf plots of ZIPMExD*
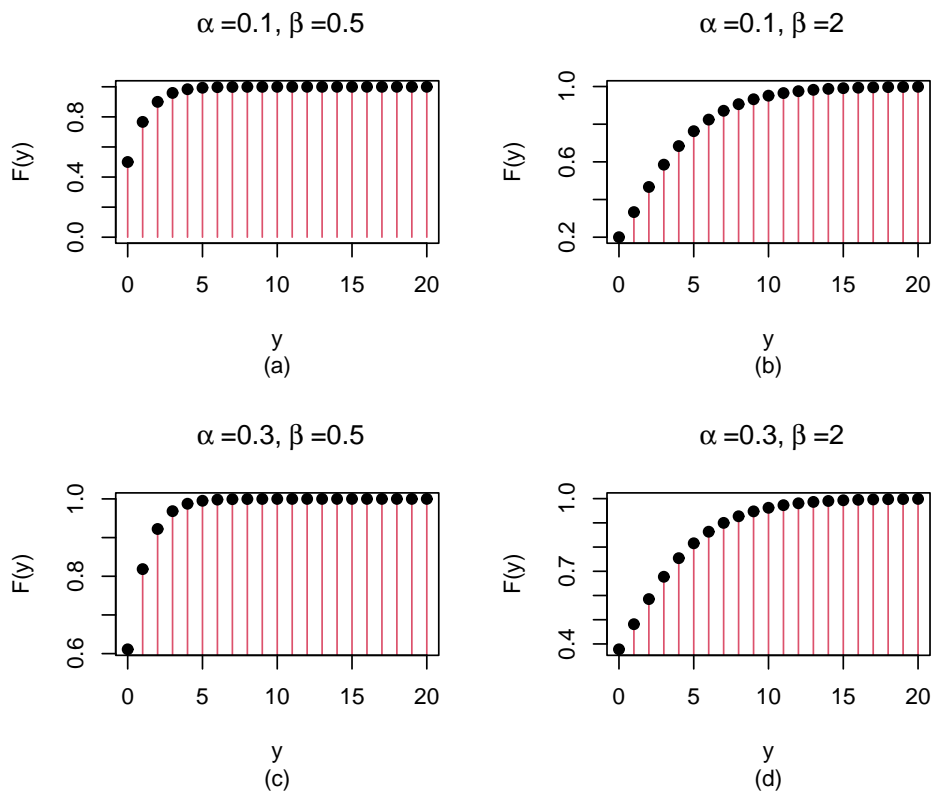
Figure 2: *The Cdf plots of ZIPMExD*

3. Structural properties along with reliability characteristics and generating functions.

In this part, we have obtained the survival function, hazard rate function, generating functions along with associated measures like index of dispersion (ID), skewness (β_1) and kurtosis (β_2) of the ZIPMExD (π, δ).

3.1. Survival Function (SF):

The Survival Function of ZIPMExD (π, δ) is as follows.

$$S(Y) = 1 - F(Y)$$

$$= (1 - \pi) \left[\frac{(y + \delta + 2)\delta^{y+1}}{(1 + \delta)^{y+2}} \right] \tag{3}$$

3.2. Hazard Rate (HR):

Let $y_1, y_2, y_3, ..., y_n$ be a random sample from ZIPMEx (π, δ) distribution as given by equation (1) Define Z be the number of y_i' s taking the value zero. Then equation (1) can be written as follows

$$P(Y = y_i) = \left[\pi + (1 - \pi)\frac{1}{(1 + \delta)^2} \right]^Z \left[(1 - \pi)\frac{\delta^y(1 + y)}{(1 + \delta)^{y+2}} \right]^{1-Z}$$

Now, using S(Y) from equation (3). The Hazard Rate of ZIPMExD (π, δ) is given as

$$H(y) = \frac{P(y)}{S(y)} = \frac{\left[\pi + (1 - \pi)\frac{1}{(1+\delta)^2} \right]^Z \left[(1 - \pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} \right]^{1-Z}}{\left[(1 - \pi)\frac{(y+\delta+2)\delta^{y+1}}{(1+\delta)^{y+2}} \right]}$$

3.3. Reverse Hazard Rate (RHR):

$$R(y) = \frac{P(y)}{F(y)} = \frac{\left[\pi + (1 - \pi)\frac{1}{(1+\delta)^2} \right]^Z \left[(1 - \pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} \right]^{1-Z}}{[1 - (1 - \pi)(y + \delta + 2)\delta^{y+1}(1 + \delta)^{-(y+2)}]}$$

3.4. Moments and associated measures

3.4.1 Moment Generating Function:

The Moment Generating Function, $M_y(t)$ of ZIPMExD (π, δ) distribution is given as

$$M_y(t) = E(e^{tx}) = \sum_{y=0}^{\infty} e^{ty}P(Y = y)$$

$$= \pi + \frac{(1 - \pi)}{(1 + \delta)^2} \sum_{y=0}^{\infty} e^{ty} \left[\frac{\delta^y(1 + y)}{(1 + \delta)^y} \right]$$

$$= \pi + \frac{(1 - \pi)}{(1 + \delta)^2} \left[\sum_{y=0}^{\infty} \frac{(e^t\delta)^y}{(1 + \delta)^y} + \sum_{y=0}^{\infty} y\frac{(e^t\delta)^y}{(1 + \delta)^y} \right]$$

$$= \pi + \frac{(1-\pi)}{(1+\delta)^2}\left[\frac{(1+\delta)}{(1+\delta-e^t\delta)^2}\right] \tag{4}$$

Putting $e^t = e^{it}$ in equation (4), the Characteristic Function, $\phi_y(t)$ of ZIPMExD (π, δ) is defined as

$$\phi_y(t) = \pi + \frac{(1-\pi)}{(1+\delta)^2}\left[\frac{(1+\delta)}{(1+\delta-e^{it}\delta)^2}\right] \tag{5}$$

Through MGF, we have derived the first four raw moments of the proposed distribution by differentiating equation (4) at t=0. The first four raw moments of the proposed distribution are as fallows.

$$\mu'_1 = (1-\pi)2\delta \tag{6}$$

$$\mu'_2 = (1-\pi)2\delta(1+3\delta) \tag{7}$$

$$\mu'_3 = (1-\pi)[2\delta(1+9\delta+12\delta^2)] \tag{8}$$

$$\mu'_4 = (1-\pi)[2\delta(1+21\delta+72\delta^2+60\delta^3)] \tag{9}$$

3.4.2 Central Moments:

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

$$\mu_2 = (1-\pi)2\delta[1+\delta+2\pi\delta] \tag{10}$$

$$\mu_3 = (1-\pi)2\delta[1+3\delta+2\delta^2+6\pi\delta+34\pi\delta^2+8\pi^2\delta^2] \tag{11}$$

$$\mu_4 = (1-\pi)2\delta[1+13\delta+24\delta^2+12\delta^3+24\pi\delta^2+8\pi\delta+24\pi\delta+24\pi\delta^3+24\pi^2\delta^2+24\pi^3\delta^3] \tag{12}$$

Remark 4.1: The ZIPMExD is over dispersed for any $\delta > 0$ and $\pi = [0,1]$.

proof: Suppose that the ZIPMExD is under dispersed. Then clearly
Mean>Variance, which implies that

$$(1-\pi)2\delta > (1-\pi)2\delta[1+\delta+2\pi\delta]$$

$$\Rightarrow 1 > [1+\delta+2\pi\delta]$$

which shows that $[1+\delta+2\pi\delta] < -1$, which is impossible for any $\delta > 0$ and $\pi = [0,1]$. Hence the proof. The dispersion index (DI) of the proposed distribution is

$$Dispersion Index = \frac{var(y)}{mean(y)} = \frac{(1-\pi)2\delta[1+\delta+2\pi\delta]}{(1-\pi)2\delta}$$

$$= [1+\delta+2\pi\delta] > 1 \tag{13}$$

Further, Coefficient of variation(CV), Skewness and Kurtosis of the proposed model are given as follows:

$$CV = \frac{SD(y)}{Mean(y)} = \frac{\sqrt{(1-\pi)2\delta[1+\delta+2\pi\delta]}}{(1-\pi)2\delta} \tag{14}$$

$$Skewness(\sqrt{\beta_1}) = \frac{[1 + 3\delta + 2\delta^2 + 6\pi\delta + 8\pi^2\delta^2 + 34\pi\delta^2]}{\sqrt{(1-\pi)(2\delta)(1+\delta+2\pi\delta)^3}} \tag{15}$$

$$Kurtosis(\beta_2) = \frac{[1 + 13\delta + 24\delta^2 + 12\delta^3 + 8\pi\delta + 24\pi\delta^2 + 24\pi\delta^3 + 24\pi^2\delta^2 + 24\pi^3\delta^3]}{(1-\pi)2\delta[1+\delta+2\pi\delta]^2} \tag{16}$$

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

	π=0.1					π=0.3					π=0.6				
$\delta\rightarrow$	0.5	1	1.5	2	2.5	0.5	1	1.5	2	2.5	0.5	1	1.5	2	2.5
Mean	0.900	1.800	2.700	3.600	4.500	0.700	1.400	2.100	2.800	3.500	0.400	0.800	1.200	1.600	2.000
Variance	1.440	3.960	7.560	12.240	18.000	1.260	3.640	7.140	11.760	17.5	0.840	2.560	5.160	8.640	13.000
DI	1.600	2.200	2.800	3.400	4.000	1.800	2.600	3.400	4.200	5.000	2.100	3.200	4.300	5.400	6.500
CV	1.300	1.105	1.018	0.971	0.942	1.603	1.362	1.272	1.224	1.195	2.292	2.000	1.892	1.837	1.802
β_1	2.171	2.302	2.432	2.532	2.607	3.281	3.777	4.102	4.324	4.482	5.517	6.421	6.939	7.260	7.488
β_2	6.754	6.136	5.929	5.828	5.769	7.548	6.594	6.250	6.074	5.967	11.795	10.134	9.524	9.204	9.005

From the above table, it can be seen that for different combinations of parameters, the value of dispersion index is greater than one. So, the proposed model is over dispersed. For skewness, it can be seen that model is rightly skewed as the value of skewness increases for different combinations of parameters. Furthermore, from the table, it can be observed that the ZIPMExD is leptokurtic as the value of kurtosis is greater than three for different combinations of parameters.

4. Parametric Estimation

In this part, we have discussed the parametric estimation of the ZIPMExD (π, δ) by moment method of estimation and maximum likelihood method of estimation.

4.1. Moment Method of Estimation (MME)

The parameters π and δ of the proposed model can be obtained using this method as follows: Considering the first two raw moments from equation number (6) and (7)

$$\hat{\delta} = \frac{\mu_1'}{2 - 2\pi} \tag{17}$$

Now, from equation number (7), we have

$$\mu_2' = (1-\pi)2\delta(1+3\delta) \tag{18}$$

Putting the value of δ from (17) to (18), we get

$$\mu_2' = 2\left[\frac{\mu_1'}{2-2\pi}\right] + 6\left[\frac{\mu_1'}{2-2\pi}\right]^2 - 2\pi\left[\frac{\mu_1'}{2-2\pi}\right] - 6\pi\left[\frac{\mu_1'}{2-2\pi}\right]^2$$

$$2(\mu_2' - \mu_1')\pi^2 - (4\mu_2' - 4\mu_1' - 3\mu_1'^2)\pi + (2\mu_2' - 2\mu_1' - 3\mu_1'^2) = 0 \tag{19}$$

we can get estimated value of π on solving the above quadratic equation and that value of π has been used to estimate the value of δ in equation (17).

4.2. Maximum Likelihood Estimation Method (MLE)

The parameters π and δ of equation (1) can be obtained using this method as follows:
Let $y_1, y_2, y_3, \dots, y_n$ be a random sample from ZIPMExD (π, δ) as given by equation (1) and let for i=1, 2, 3, … , n

$$b_i = \begin{cases} 1; if \quad y_i = 0 \\ 0; \quad otherwise \end{cases}$$

then, for i=1, 2, 3, …, n equation (1) can be expressed as follows

$$P(Y = y_i) = \left[\pi + \frac{(1-\pi)}{(1+\delta)^2}\right]^{b_i} \left[\frac{(1-\pi)(1+y_i)\delta^{y_i}}{(1+\delta)^{2+y_i}}\right]^{1-b_i}$$

Hence the likelihood function; L=$L(\pi, \delta; y_1, y_2, y_3, \dots, y_n)$ will be

$$L = \prod_{i=1}^{n} \left[\pi + \frac{(1-\pi)}{(1+\delta)^2}\right]^{b_i} \left[\frac{(1-\pi)(1+y_i)\delta^{y_i}}{(1+\delta)^{2+y_i}}\right]^{1-b_i}$$

$$= \left[\pi + \frac{(1-\pi)}{(1+\delta)^2}\right]^{n_0} \prod_{i=1}^{n} \left[\frac{(1-\pi)(1+y_i)\delta^{y_i}}{(1+\delta)^{2+y_i}}\right]^{d_i}$$

Where d_i=1-b_i, $n_0 = \sum_{i=1}^{n} b_i$. Note that the number of zeros in the sample are represented by n_0. Therefore,

$$\log L = n_0 \log \left[\pi + \frac{(1+\pi)}{(1+\delta)^2}\right] + (n - n_0) \log(1 - \pi) + \log \delta \sum_{i=1}^{n} d_i y_i + \log(1 + \delta) \sum_{i=1}^{n} d_i(y_i + 2)$$

$$\frac{\partial \log L}{\partial \pi} = \frac{n_0[(1+\delta)^2 - 1]}{(1+\delta)^2 \pi + (1-\pi)} - \frac{(n - n_0)}{(1-\pi)} \tag{20}$$

$$\frac{\partial \log L}{\partial \delta} = -\frac{2n_0(1-\pi)}{\pi(1+\delta)^3 + (1-\pi)(1+\delta)} + \frac{1}{\delta} \sum_{i=1}^{n} d_i y_i - \frac{1}{(1+\delta)} \sum_{i=1}^{n} d_i(y_i + 2) \tag{21}$$

$$Let, p = \pi + \frac{(1-\pi)}{(1+\delta)^2} \tag{22}$$

Now, let $\frac{\partial \log L}{\partial \pi}$=0, then from equation (20) and using equation (22),

$$1 - \pi = \frac{p(n - n_0)(1+\delta)^2}{n_0[(1+\delta)^2 - 1]} \tag{23}$$

Now, letting $\frac{\partial \log L}{\partial \delta}$=0, using equation (22), equation (21) reduces

$$-\frac{n_0[2(1-\pi)]}{p(1+\delta)^3} + \sum_{i=1}^{n} \frac{d_i y_i}{\delta} - \sum_{i=1}^{n} \frac{d_i(y_i + 2)}{(1+\delta)} = 0 \tag{24}$$

Now, if we replace p by their sample relative frequencies, i.e., by their sample estimates, the proportion of zeros in the sample, i.e., $\hat{p}$=n_0/n and then Equation (23) reduces to

$$1 - \pi = \frac{(n - n_0)(1+\delta)^2}{n[(1+\delta)^2 - 1]} \tag{25}$$

Now using equation (23), equation (24) can be written as

$$-\frac{[2(n-n_0)(1+\delta)^2]}{[(1+\delta)^2-1]} + \sum_{i=1}^{n}\frac{d_i y_i}{\delta} - \sum_{i=1}^{n}\frac{d_i(y_i+2)}{(1+\delta)} = 0$$

$$\Rightarrow M(\delta) = 0 \qquad (26)$$

Where, $M(\delta) = -\frac{[2(n-n_0)(1+\delta)^2]}{[(1+\delta)^2-1]} + \sum_{i=1}^{n}\frac{d_i y_i}{\delta} - \sum_{i=1}^{n}\frac{d_i(y_i+2)}{(1+\delta)}$

Hence by any numerical means, say Newton Rapson method. Equation (26) can be solved to obtain $\hat{\delta}$ numerically. i.e., M ($\hat{\delta}$)=0
similarly, using equation (22), π can be estimated

$$\hat{\pi} = \frac{1}{n}\left[n_0 - \frac{(n-n_0)}{[(1+\delta)^2-1]}\right] \qquad (27)$$

Therefore, the maximum likelihood estimates (MLE) of the parameters δ and π are given by solving equation (26) numerically to find $\hat{\delta}$ and $\hat{\pi}$ given by equation (27) respectively.
In order to calculate the asymptotic variance-covariance matrix of the estimates the second order differentiations of the log-likelihood function are given here

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

$$\frac{\partial \log L}{\partial \delta^2} = \frac{2n_o(1-\pi)[2(1+\delta)^2\pi+(1-\pi)]}{[\pi(1+\delta)^3+(1-\pi)(1+\delta)]^2} - \sum_{i=1}^{n}\frac{d_i y_i}{\delta^2} + \sum_{i=1}^{n}\frac{d_i(y_i+2)}{(1+\delta)^2}$$

$$\frac{\partial \log L}{\partial \delta \partial \pi} = \left[\frac{2n_0[\pi(1+\delta)^2+(1-\pi)] - 2n_0(1-\pi)[(1+\delta)^2-1]}{[\pi(1+\delta)^2+(1-\pi)]^2}\right]$$

By inverting the Fisher's information matrix (I), the asymptotic variance-covariance matrix of the maximum likelihood estimates of δ and π for ZIPMExD (δ, π) can be obtained as

$$I = \begin{bmatrix} E\left(-\frac{\partial \log L}{\partial \pi^2}\right) & E\left(-\frac{\partial \log L}{\partial \delta \partial \pi}\right) \\ E\left(-\frac{\partial \log L}{\partial \pi \partial \delta}\right) & E\left(-\frac{\partial \log L}{\partial \delta^2}\right) \end{bmatrix}$$

The ingredients of the above Fisher's information matrix can be obtained as

$$E\left(-\frac{\partial \log L}{\partial \pi^2}\right) = \left(-\frac{\partial \log L}{\partial \pi^2}\right)\Bigg|_{\pi=\hat{\pi},\delta=\hat{\delta}} \qquad (28)$$

The asymptotic distribution of the maximum likelihood estimator $(\hat{\delta}, \hat{\pi})$ is given by

$$\sqrt{n}\begin{pmatrix}\hat{\delta}\\\hat{\pi}\end{pmatrix}_{MLE} \xrightarrow{L} AN\left(\begin{pmatrix}\hat{\delta}\\\hat{\pi}\end{pmatrix}, I^{-1}\right), as\, n \to \infty$$

5. Testing

In this part, we have checked the significance of inflation parameter (π) by likelihood ratio test and score test.

5.1. Likelihood Ratio Test

In order to test the significance of the inflation parameter π of the ZIPMExD, The Likelihood Ratio Test (LRT) is carried out to distinguish between PMExD (ζ) and ZIPMExD (π, δ). Here the null hypothesis is

$$H_o : \pi = 0 \, Vs \, H_1 \pi \neq 0$$

In case of LRT, test statistic is given by

$$-2ln\xi = 2(l_1 - l_2), \tag{29}$$

where, $l_1 = lnL(\hat{\theta}; y)$, Where $\hat{\theta}$ is the maximum likelihood estimator for $\theta = (\pi, \delta)$ without limitation, and $l_2 = lnL(\hat{\theta}^*, y)$, in which $\hat{\theta}^*$ is the maximum likelihood estimator for θ under the null hypothesis H_o. The test statistic described in equation (29) is asymptotically distributed as χ^2 with one degree of freedom.

Table 2: *Calculated value of test statistic in case of Likelihood Ratio Test.*

	$lnL(\hat{\theta}^*; y)$	$lnL(\hat{\theta}; y)$	Test statistic
Dataset 1	-6752.66	-6736.52	32.28
Dataset 2	-476.68	-430.85	91.66
Dataset 3	-595.86	-593.64	4.44

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

5.2. Wald test

Here for testing the significance of inflation parameter π of ZIPMExD we assess the Wald test. To test the null hypothesis

$$H_0 : \pi = 0 \, Vs \, H_1 : \pi \neq 0$$

In case of Wald test, test statistic is given by

$$W_\pi = \frac{\hat{\pi}^2}{Var(\hat{\pi})}, \tag{30}$$

Where $Var(\hat{\pi})$ represents the diagonal element of Fisher information matrix at $\pi = \hat{\pi}$ and $\delta = \hat{\delta}$. The test statistic given in equation (30) is asymptotically distributed as chi^2 with one degree of freedom.

Table 3: *Calculated value of test statistic in case of Wald Test.*

	Test statistic
Dataset 1	39.73
Dataset 2	283.96
Dataset 3	4.72

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

5.3. Simulation

In this section, we carry 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. The procedure was repeated 1000 times for calculation of Bias, Variance, Mean Square Error (MSE) and Coverage Probability and the results are given in Table2. It can be seen from the table, that as the sample size increases, the bias, variance and mean square error decreases and are close to zero for large sample sizes. Also, the coverage probability tends to 0.95 as the sample size increases. These results suggest that maximum likelihood estimates are consistent and therefore can be used in estimating the unknown parameters of the proposed model.

Table 4: *Simulation table of MLE's for proposed model*

Sample Size(n)	Parameter	$\pi = 0.3, \delta = 0.9$				$\pi = 0.5, \delta = 2$			
		Bias	Variance	MSE	Coverage probability (95%)	Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\pi}$	-0.06354	0.02335	0.02739	0.98	-0.03026	0.00832	0.00923	1.00
	$\hat{\delta}$	-0.08737	0.06415	0.07178	0.90	-0.00590	0.24814	0.24818	1.00
75	$\hat{\pi}$	-0.01913	0.00818	0.00854	0.98	-0.01438	0.00422	0.00442	0.96
	$\hat{\delta}$	-0.03034	0.02751	0.02843	0.92	0.00707	0.06000	0.06005	0.98
100	$\hat{\pi}$	0.00948	0.00646	0.00655	0.96	-0.00521	0.00188	0.00191	1.00
	$\hat{\delta}$	0.03178	0.02177	0.02278	0.94	0.00771	0.04532	0.04538	1.00
300	$\hat{\pi}$	-0.01206	0.00241	0.00256	0.92	-0.00828	0.00088	0.00095	1.00
	$\hat{\delta}$	-0.00248	0.00565	0.00565	0.98	-0.03233	0.01900	0.02005	0.96
600	$\hat{\pi}$	-0.01220	0.00065	0.00080	0.98	0.00144	0.00064	0.00064	0.94
	$\hat{\delta}$	-0.00941	0.00263	0.00272	0.96	0.02842	0.01513	0.01594	0.90
Sample Size(n)	Parameter	$\pi = 0.5, \delta = 2.5$				$\pi = 0.5, \delta = 0.85$			
		Bias	Variance	MSE	Coverage probability (95%)	Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\pi}$	-0.00174	0.01119	0.01120	0.94	-0.05718	0.02309	0.02636	0.98
	$\hat{\delta}$	-0.01929	0.62737	0.62775	0.90	-0.00713	0.11008	0.11013	0.94
75	$\hat{\pi}$	-0.01586	0.00481	0.00506	0.96	0.02007	0.00812	0.00853	0.94
	$\hat{\delta}$	-0.03521	0.09355	0.09479	0.94	0.01183	0.03575	0.03589	0.92
100	$\hat{\pi}$	-0.00893	0.00341	0.00349	0.94	0.00874	0.00646	0.00654	0.94
	$\hat{\delta}$	-0.05368	0.09101	0.09389	0.98	0.00454	0.02240	0.02242	0.96
300	$\hat{\pi}$	0.00236	0.00110	0.00111	0.98	-0.00032	0.00243	0.00243	0.92
	$\hat{\delta}$	0.01784	0.03013	0.03045	0.98	0.00635	0.00626	0.00630	0.98
600	$\hat{\pi}$	-0.00198	0.00034	0.00035	1.00	0.00133	0.00097	0.00097	0.94
	$\hat{\delta}$	0.02452	0.01711	0.01771	0.96	-0.00472	0.00378	0.00381	0.96
Sample Size(n)	Parameter	$\pi = 0.1, \delta = 0.5$				$\pi = 0.5, \delta = 1.6$			
		Bias	Variance	MSE	Coverage probability (95%)	Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\alpha}$	-0.01360	0.00971	0.00989	0.98	-0.02185	0.01302	0.01350	0.98
	$\hat{\theta}$	0.00515	0.05491	0.05494	0.96	-0.14024	0.15844	0.17811	0.90
75	$\hat{\alpha}$	0.00575	0.00632	0.00635	0.98	-0.01570	0.00438	0.00463	0.96
	$\hat{\theta}$	0.01937	0.01274	0.01312	0.98	-0.03387	0.09477	0.09592	0.86
100	$\hat{\alpha}$	0.00114	0.00623	0.00624	0.94	-0.00752	0.00470	0.00476	0.94
	$\hat{\theta}$	0.01932	0.01359	0.01396	1.00	-0.04282	0.05458	0.05642	0.94
300	$\hat{\alpha}$	-0.00765	0.00200	0.00206	1.00	0.00347	0.00127	0.00128	0.96
	$\hat{\theta}$	-0.01014	0.00706	0.00717	0.90	0.02095	0.01841	0.01885	0.96
600	$\hat{\alpha}$	-0.00306	0.00152	0.00153	0.92	-0.00371	0.00066	0.00067	0.96
	$\hat{\theta}$	-0.00216	0.00262	0.00263	0.94	-0.02580	0.01039	0.01105	0.88

6. APPLICATIONS

In this part, we study the practical significance of Zero-Inflated Poisson Moment Exponential Distribution(ZIPMExD). Three real life data sets are taken to compare Zero-Inflated Poisson Moment Exponential Distribution (ZIPMED) with few other distributions.

6.1. Data set 1

The dataset from Table 5 consists of frequencies regarding the number of vaccine adverse events originally given by Rose *et.al* [24] in 2006. Total number of events were recorded after each of the four injections for the 1005 study participants, which results in 4020 observations. Dar*et.al* [8] recently used Poisson weighted exponential distribution to fit the number of vaccine adverse events data. After analysing data through R software, we can see that our model performs better than other competing models because of highest p-value i.e., (0.928) among all other competing distributions and we can also see that our model favours the criteria i.e., Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) among all other competing models because of lowest values. The other competing models we use in this paper are Poisson Moment distribution (PMD), Zero-inflated Poisson Distribution (ZIPD), Poisson Distribution (PD), Negative Binomial Distribution (NBD), Discrete Weibull Distribution (DWD) and Zero-inflated Negative Binomial Distribution (ZINBD).

Table 5: *Expected frequencies and χ^2 values for fitted models*

Claims	observed count	ZIPMExD	PMExD	ZIPD	PD	NBD	DWD	ZINBD
0	1437	1437	1307	1437	891	1119	1411	1437
1	1010	1009	1124	787	1342	1225	1066	958
2	660	681	724	803	1011	838	668	708
3	428	408	415	546	508	459	393	436
4	236	230	223	279	191	220	223	241
5	122	124	115	114	58	96	123	125
6	62	65	58	39	14	39	66	61
7	34	33	28	11	3	15	35	29
8	14	17	14	3	1	6	18	13
9	8	8	7	1	0	2	9	6
10	4	4	3	0	0	1	5	3
11	4	2	1	0	0	0	2	1
12	1	1	1	0	0	0	1	1
Degrees of Freedom		8	8	5	6	9	8	7
ML Estimates		$\hat{\pi}$=0.0836 $\hat{\delta}$=0.7534	$\hat{\lambda}$=2.0405	$\hat{\lambda}$=1.5069	$\hat{p}$=0.5032	$\hat{q}$=0.6491	$\hat{p}$=0.6020	
		$\hat{\delta}$=0.8324		$\hat{\pi}$=0.2614		$\hat{r}$=1.5267	$\hat{\beta}$=1.1469	$\hat{r}$=2.6000
								$\hat{a}$=0.1229
χ^2-value		3.08	36.64	301.57	1516.9	10.12	8.79	7.23
$p-value$		0.928	< 0.001	< 0.001	< 0.001	0.320	0.359	0.404
$-\hat{log}$		6736.52	6752.66	6868.79	7231.13	6740.60	6739.67	6737.84
AIC		13477.04	13507.32	13741.58	14464.26	13485.21	13483.35	13481.68
BIC		13489.64	13513.62	13754.18	14470.56	13497.80	13495.95	13500.58

6.2. Dataset 2

The dataset from Table 6 has been taken from [15]. The dataset is related to the HIV exposed infant data. The data has been taken from three concerned regions, Nairobi, Kisumu and Mombasa and the data reveals zero-inflation because of the measures that have been put in place to reduce the rate of Mother to Child Transmission (MTCT). A total of 494 samples were collected from 60 health centres in Kenya from these three regions for analysis. From the table, we can see that our model outperforms other competing models because of highest p-value among all other competing distributions and we can also see that our model has lowest criteria i.e., Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) among all other competing models.

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

Claims	Observed count	ZIPMExD	PMExD	ZIPD	PD	NBD	DWD	ZINBD
0	378	378	323	378	308	354	336	378
1	59	54	123	47	145	86	108	57
2	26	31	35	38	34	31	35	30
3	13	16	9	20	5	13	11	15
4	7	8	2	8	1	5	4	8
5	11	4	0	3	0	2	1	6
Degrees of Freedom		3	2	2	5	2	2	2
ML Estimates		$\hat{\pi}$=0.6215	$\hat{\delta}$=0.2358	$\hat{\lambda}$=1.6051	$\hat{\lambda}$=0.4716	$\hat{p}$=0.5145	$\hat{q}$=0.3207	$\hat{p}$=0.6639
		$\hat{\delta}$=0.6230		$\hat{\pi}$=0.7061		$\hat{r}$=0.5000	$\hat{\beta}$=1.0010	$\hat{r}$=2.6000
								$\hat{a}$=0.6416
χ^2		2.04	69.90	13.76	172.97	63.53	63.96	3.15
$p-value$		0.564	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.207
$-l\hat{o}g$		430.85	476.68	435.39	524.23	438.52	456.12	431.10
AIC		865.70	955.36	874.79	1050.64	881.04	916.25	868.20
BIC		874.10	959.56	883.20	1054.84	889.45	924.66	880.81

6.3. Dataset 3

The dataset from Table7 represents the frequencies of epileptic seizure counts reported in [6]. The measures of goodness-of-fit for all competing distributions are presented and it is evident that the proposed distribution fits well, as it has the highest p-value and lowest AIC and BIC criteria. So, in this regard we see that our model fits better than other fitted models on the given data set.

Table 7: *Expected frequencies and χ^2 values for fitted models*

Claims	Observed count	ZPMExD	PMExD	ZIPD	PD	NBD	DWD	ZINB
0	126	126	112	126	75	120	120	126
1	80	85	97	65	116	93	93	79
2	59	59	64	69	89	59	59	62
3	42	36	37	49	46	35	35	39
4	24	21	20	26	18	20	20	22
5	8	11	11	11	5	11	11	12
6	5	6	5	4	1	6	6	6
7	4	3	3	1	0	3	3	3
8	3	2	1	0	0	2	2	1
Degrees of Freedom		5	6	4	2	5	5	4
ML Estimates		$\hat{\pi}$=0.0959	$\hat{\delta}$=0.7720	$\hat{\lambda}$=2.1196	$\hat{\lambda}$=1.5441	$\hat{p}$=0.5009	$\hat{q}$=0.6577	$\hat{p}$=0.6455
		$\hat{\delta}$=0.8540		$\hat{\pi}$=0.2715		$\hat{r}$=1.5500	$\hat{\beta}$=1.1560	$\hat{r}$=3.3845
								$\hat{a}$=0.1710
χ^2-value		3.5	9.66	16.68	82.45	6.10	6.10	2.87
$p-value$		0.622	0.139	0.002	<0.001	0.296	0.296	0.579
$-l\hat{o}g$		575.70	577.92	599.63	636.04	594.94	594.74	576.00
AIC		1155.41	1157.84	1203.27	1274.09	1193.88	1193.49	1156.41
BIC		1162.09	1163.07	1210.99	1277.95	1201.60	1201.22	1167.93

7. Conclusion

A new Zero-inflated version of Poisson moment exponential distribution is introduced in this paper namely Zero-inflated Poisson Moment Exponential Distribution (ZIPMExD). Key statistical properties of the distribution including generating functions, reliability characteristics and moments have been derived. For parametric estimation purpose, two different methods i.e., moment method and maximum likelihood method of estimation have been used. Simulation study has been done for evaluating the proficiency of the estimation measures considered in this paper. Further, the procedure of Log Likelihood ratio test and Wald test are designed for testing the significance of inflation parameter. Three real life data sets are reviewed for demonstrating the practicality of the introduced model juxtapose to the existent models being PMExD, PD, ZIPD, DWD, NBD and ZINBD. We can see that ZIPMExD in terms of Chi-square value and p-value gives best fit as the existent models do not show best fit. The information measures like Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in terms of numerical value reveals that ZIPMExD can be considered as a suitable model in comparison to other models as discussed in this paper.

References

[1] Ahmad, P. B. (2019). Bayesian Analysis of Zero-Inflated Generalized Power Series Distributions Under Different Loss Functions. Bayesian Analysis and Reliability Estimation of Generalized Probability Distributions, 1.

[2] Ahsan-ul-Haq, M. (2022). On Poisson moment exponential distribution with applications. Annals of Data Science, 1-22.

[3] Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., & Pal, N. (2014). Zero-inflated Poisson (ZIP) distribution: Parameter estimation and applications to model data from natural calamities. Involve, a Journal of Mathematics, 7(6), 751-767.

[4] Bekalo, D. B., & Kebede, D. T. (2021). Zero-inflated models for count data: an application to number of antenatal care service visits. Annals of Data Science, 8, 683-708.

[5] B?hning, D. (1998). Zero-inflated Poisson models and CA MAN: A tutorial collection of evidence. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 40(7), 833-843.

[6] Chakraborty, S. (2010). On some distributional properties of the family of weighted generalized Poisson distribution. Communications in Statistics—Theory and Methods, 39(15), 2767-2788.

[7] Dara ST, Ahmad M (2012) Recent advances in moment distribution and their hazard rates. LAP LAMBERT Academic Publishing, Chisinau

[8] Dar, S. A., Hassan, A., Ahmad, P. B., & Wani, S. A. (2021). A new count data model applied in the analysis of vaccine adverse events and insurance claims. Statistics in Transition new series, 22(3), 157-174.

[9] Feller, W. (1943). On a general class of" contagious" distributions. The Annals of mathematical statistics, 14(4), 389-400.

[10] GORALSKI, A. (1977). DISTRIBUTION Z-POISSON.

[11] Gupta, P. L., Gupta, R. C., & Tripathi, R. C. (1996). Analysis of zero-adjusted count data. Computational Statistics & Data Analysis, 23(2), 207-218.

[12] Hall, D. B. (2000). Zero?inflated Poisson and binomial regression with random effects: a case study. Biometrics, 56(4), 1030-1039.

[13] Junnumtuam, S., Niwitpong, S. A., & Niwitpong, S. (2023). Bayesian Computation for the Parameters of a Zero-inflated Cosine Geometric Distribution with Application to COVID-19 Pandemic Data. CMES-COMPUTER MODELING IN ENGINEERING & SCIENCES, 135(2), 1229-1254.

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

[15] Kibika, S. A. (2020). The Zero Inflated Negative Binomial-Shanker distribution and its application to HIV exposed infant data (Doctoral dissertation, Strathmore University).

[16] Leroux, B. G., & Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. Biometrics, 545-558.

[17] Martin, D. C., & Katti, S. K. (1965). Fitting of certain contagious distributions to some available data by the maximum likelihood method. Biometrics, 21(1), 34-48.

[18] Maya, R., Huang, J., Irshad, M. R., & Zhu, F. (2023). On Poisson Moment Exponential Distribution with Associated Regression and INAR (1) Process. Annals of Data Science, 1-19.

[19] Mwalili, S. M., Lesaffre, E., & Declerck, D. (2008). The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. Statistical methods in medical research, 17(2), 123-139.

[20] Mullahy, J. (1986). Specification and testing of some modified count data models. Journal of econometrics, 33(3), 341-365.

[21] Nanjundan, G., & Naika, T. R. (2012). Asymptotic comparison of method of moments estimators and maximum likelihood estimators of parameters in zero-inflated poisson model.

[22] Neyman, J. (1939). On a new class of" contagious" distributions, applicable in entomology and bacteriology. The Annals of Mathematical Statistics, 10(1), 35-57.

[23] Rahman, T., Hazarika, P. J., Ali, M. M., & Barman, M. P. (2022). Three-Inflated Poisson Distribution and its Application in Suicide Cases of India During Covid-19 Pandemic. Annals of Data Science, 9(5), 1103-1127.

[24] Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of biopharmaceutical statistics, 16(4), 463-481.

[25] Sandhyaa, E., & Abrahamb, T. L. (2016). Inflated-parameter Harris distribution. JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE-JMCS, 16(1), 33-49.

[26] Scollnik, D. P. (2022). Bayesian analyses of an exponential-Poisson and related zero augmented type models. Journal of Applied Statistics, 49(4), 949-967.

[27] Sharma, A. K., & Landge, V. S. (2013). Zero inflated negative binomial for modeling heavy vehicle crash rate on Indian rural highway. International Journal of Advances in Engineering & Technology, 5(2), 292.

[28] Shukla, K. K., & Yadava, K. N. S. (2006). The Distribution of the Number of Migrants at the Household Level. Journal of Population and Social Studies [JPSS], 14(2), 153-166.

[29] Sim, S. Z., Gupta, R. C., & Ong, S. H. (2018). Zero-inflated Conway-Maxwell Poisson distribution to analyze discrete data. The international journal of biostatistics, 14(1).

[30] Singh, S. N. (1962, January). Note on inflated Poisson-distribution. In ANNALS OF MATHEMATICAL STATISTICS (Vol. 33, No. 3, p. 1210). IMS BUSINESS OFFICE-SUITE 7, 3401 INVESTMENT BLVD, HAYWARD, CA 94545: INST MATHEMATICAL STATISTICS.

[31] Suresh, R., Nanjundan, G., Nagesh, S., & Pasha, S. (2015). On a characterization of Zero-inflated negative binomial distribution. Open Journal of Statistics, 5(06), 511-513.

[32] Van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. Biometrics, 738-743.