Asian Journal of Statistical Sciences
Vol. 5, No-2, (2025), pp. 155-168

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

On a Wide Class of Zero-inflated Count Data Models and its Applications

C. Satheesh Kumar[a] and Rakhi Ramachandran[b]

[a] *Department of Statistics, University of Kerala, Thiruvananthapuram, PIN: 695 581, Kerala*
[b]*Department of Statistics, Sree Neelakanta Government Sanskrit College, Pattambi, Palakkad, PIN: 679306, Kerala*

ARTICLE HISTORY
Compiled July 13, 2026

Received: 15 May 2024; Accepted: 01 December 2025

ABSTRACT
Count data arises frequently in every field of human activity, but regularly violate the equidispersion constraint imposed by the most popular distribution for analyzing these data, the Poisson distribution. The hyper-Poisson distribution, a generalization of the Poisson distribution is suitable for modeling count data having under dispersed, over dispersed and equi dispersed data sets. In this paper we consider a wide class of zero-inflated distributions as extended vesions of the zero-inflated Hermite distribution, the zero-inflated hyper-Poisson distribution and the zero-inflated modified hyper-Poisson distribution. We investigate several important properties of the distribution such as expressions for probability generating function, mean, variance, skewness, kurtosis etc. along with recursion formulae for probabilities, factorial moments and raw moments. The estimation of the parameters of the distribution is also attempted and it has been fitted to certain real life data sets for highlighting its practical relevance. Further, certain test procedures are applied for examining the significance of the parameters of the model and a simulation study is conducted for assessing the performance of the maximum likelihood estimators of the parameters of the distribution.

KEYWORDS
confluent hypergeometric series; count data modeling; generalized likelihood ratio test; Rao's efficient score test; simulation

1. Introduction

Count (or frequency) responses such as number of heart attacks, number of days of alcohol drinking, number of suicide attempts, and number of unprotected sexual encounters during a period of time arise quite often in biomedical and psycho social research. Poisson distribution models are widely used to handle such situations. One major limitation of the Poisson model is that the mean is identical to the variance. In practice, heterogeneity in study populations due to data clustering or other factors often creates extra variability, resulting in variance that is larger than the mean. This renders the Poisson distribution inappropriate for modeling count data in several such instances.

CONTACT Author[a]. Email:drcsatheeshkumar@gmail.com

To model count data with excess zeros, [20] considered the zero-inflated Poisson (ZIP) distribution, whose probability mass function (p.m.f) is the following.

$$g_1(x) = P(X = x) = \begin{cases} \omega + (1-\omega)e^{-\lambda}, & x = 0 \\ (1-\omega)\frac{e^{-\lambda}\lambda^x}{x!}, & x = 1, 2, ..., \end{cases} \tag{1}$$

where $\omega \in [0, 1]$ and $\lambda > 0$. Many authors have utilized the ZIP distribution for modeling count data with an excessive number of zeros in practical contexts and considerable quantum of work has been done in this regard. For example, see [24], [3], [21], [6], [2], [8], [7] and references therein.

The "hyper-Poisson distribution (HPD)", a generalized Poisson family of distributions developed by [1] has the following p.m.f., for $x = 0, 1, 2, ...$.

$$g_2(x) = P(X = x) = \frac{\Gamma(\lambda)}{\Gamma(\lambda + x)} \frac{\theta^x}{\phi(1, \lambda, \theta)}, \tag{2}$$

in which λ positive, θ positive and $\phi(1, \lambda, \theta)$ is the confluent hypergeometric function. For more details on confluent hypergeometric function, see [22] or [25]. Throughout in this paper we write HPD(λ, θ) to denote a distribution with p.m.f. (2). The HPD(λ, θ) with $\lambda < 1$ is known in the literature as "sub-Poisson distribution" while that with $\lambda > 1$ is known as "super-Poisson distribution" (for details, see [1]). The hyper-Poisson distribution is a member of the Kemp family of distribution studied by [9]. It has been further studied by [11], [10], [12], [13], [14], [15] etc.

[17] developed an over dispersed zero-inflated model, namely the zero-inflated Hermite distribution (ZIHD) which is heavy tailed compared to zero-inflated generalized Poisson distribution of [5]. The p.m.f. of the ZIHD is given by

$$g_3(x) = \begin{cases} \omega + (1-\omega)e^{-(\lambda_1+\lambda_2)}, x = 0 \\ (1-\omega)\sum\limits_{j=0}^{[\frac{x}{2}]} e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^{x-2j}\lambda_2^j}{(x-2j)!j!}, x = 1, 2, ..., \end{cases} \tag{3}$$

where $[\frac{x}{2}]$ denotes the integer part of $\frac{x}{2}$, with $\omega \in [0, 1]$, $\lambda_1 > 0$ and $\lambda_2 \geq 0$.

As a generalization to the ZIPD, [16] developed the zero-inflated hyper-Poisson distribution (ZIHPD) through the p.m.f.

$$g_4(x) = \begin{cases} \omega + (1-\omega)\frac{1}{\phi(1;\lambda;\theta)}, x = 0 \\ (1-\omega)\frac{\Gamma(\lambda)}{\Gamma(\lambda+x)}\frac{\theta^x}{\phi(1;\lambda;\theta)}, x = 1, 2, ... \end{cases} \tag{4}$$

with $\omega \in [0, 1]$, $\lambda > 0$ and $\theta > 0$. Obviously, the ZIHPD with $\lambda = 1$ is the ZIPD having p.m.f. (1).

[18] further considered a generalized class of zero-inflated models namely the zero-inflated modified hyper-Poisson distribution (ZIMHPD) through the p.m.f.

$$g_5(x) = \begin{cases} \omega + \frac{(1-\omega)}{\phi(1;\lambda;\theta_1+\theta_2)}, x = 0 \\ \frac{(1-\omega)}{\phi(1;\lambda;\theta_1+\theta_2)} \sum\limits_{k=0}^{[\frac{x}{2}]} \frac{(x-k)!\theta_1^{x-2k}\theta_2^k}{(\lambda)_{x-k}(x-2k)!k!}, x = 1, 2, ... \end{cases} \tag{5}$$

with $\omega \in [0, 1]$, $\lambda > 0$, $\theta_1 > 0$ and $\theta_2 \geq 0$.

In this paper we suggest a more wide class of zero-inflated models as an extension of the ZIHD, the ZIHPD and the ZIMHPD, named the "zero-inflated extended hyper-Poisson (ZIEHPD)". The rest of the paper is organized as follows. Section 2 contains definition and important properties of the ZIEHPD. In section 3 we discuss estimation of the parameters of the ZIEHPD. In section 4 certain test procedures are suggested. Certain real life data applications are considered for highlighting the usefulness of the model are discussed in section 5. Finally, in section 6 a brief simulation study is carried out for assessing the performance of the maximum likelihood estimators of the parameters of the distribution.

2. Definition and Properties

Here first we present the definition of the zero-inflated extended hyper-Poisson distribution and discuss some of its properties.

Definition 2.1. Let ξ be a degenerate random variable at the point zero and let Z follows the extended hyper-Poisson distribution $(\lambda, \theta_1, \theta_2)$. Assume that ξ and Z are statistically independent. Then the p.m.f. of the "the zero-inflated extended hyper-Poisson distribution" or in short "the ZIEHPD" has the following form.

$$\begin{aligned} f(y) &= P(Y = y) \\ &= \omega P(\xi = y) + (1-\omega) P(Z = y) \\ &= \begin{cases} \omega + (1-\omega)\,\delta, \quad y = 0 \\ (1-\omega)\delta \sum\limits_{k=0}^{[\frac{y}{m}]} \frac{[y-(m-1)k]!\ \theta_1^{y-mk}\ \theta_2^{k}}{(\lambda)_{y-(m-1)k}\ (y-mk)!\ k!}, \ y = 1, 2, ... \\ 0, \ otherwise, \end{cases} \end{aligned} \tag{6}$$

where $\delta = \frac{1}{\phi(1;\lambda;\theta_1+\theta_2)}$, $[\frac{y}{m}]$ denotes the integer part of $\frac{y}{m}$, $\omega \in [0,1]$, $\lambda > 0$, $\theta_1 > 0$, $\theta_2 \geq 0$ and m is a fixed but arbitrary positive integer.

(1) When $m = 1$ and $\lambda = 1$ or $\lambda = 1$ and $\theta_2 = 0$ ZIEHPD $\rightarrow$ ZIPD $(\omega, \theta_1 + \theta_2)$ of [20].
(2) When $\omega = 0$, ZIEHPD $\rightarrow$ extended hyper-Poisson distribution (EHPD) of [10].
(3) When $\omega = 0$, $m = 2$ and $\lambda \neq 1$ ZIEHPD $\rightarrow$ modified hyper-Poisson distribution (MHPD) of [11].
(4) When $m = 1$ or $\theta_2 = 0$ or both ZIEHPD $\rightarrow$ ZIHPD of [16].
(5) When $m = 2$ and $\lambda = 1$ ZIEHPD $\rightarrow$ ZIHD of [17].

Probability mass plots of ZIEHPD for particular values of ω, λ, θ_1 and θ_2, for $m = 1, 2, 3$ and 4 are given in Figure-1.

Next we present certain properties of the ZIEHPD through the following results.

Proposition 2.2. *The probability generating function (p.g.f.) $G(t)$ of the ZIEHPD with p.m.f (6) is the following.*

$$G(t) = \omega + (1-\omega)\delta\phi(1;\lambda;\theta_1 t + \theta_2 t^m) \tag{7}$$

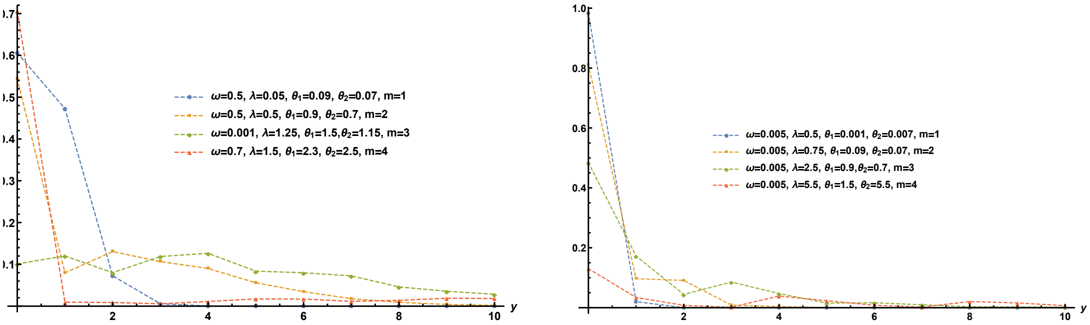

Figure 1. Probability plots of ZIEHPD for different values of its parameters.

Proof. By definition, the p.g.f. of the ZIEHPD is

$$
\begin{aligned}
G(t) = E(t^Y) &= \sum_{y=0}^{\infty} t^y \; P(Y=y) \\
&= \quad \omega + (1-\omega)\delta + (1-\omega)\sum_{y=1}^{\infty}\sum_{k=0}^{[\frac{y}{m}]} t^y \delta \tfrac{[y-(m-1)k]! \; \theta_1^{y-mk}\theta_2^k}{(\lambda)_{y-(m-1)k} \; (y-mk)! \; k!} \\
&= \omega + (1-\omega)\delta + (1-\omega)\delta \left\{ \sum_{y=0}^{\infty}\sum_{k=0}^{[\frac{y}{m}]} \frac{[y-(m-1)k]!(\theta_1 t)^{y-mk}(\theta_2 t^m)^k}{(\lambda)_{y-(m-1)k}(y-mk)!k!} - 1 \right\} \\
&= \omega + (1-\omega)\delta + (1-\omega)\delta[\phi(1;\lambda;\theta_1 t + \theta_2 t^m) - 1]
\end{aligned}
$$

On rearranging the terms, we get (7). $\square$

By using Proposition 2.1, one can easily obtain the following remarks.

Remark 1. Mean and Variance of the ZIEHPD with p.g.f. (7) are

$$
Mean = G'(1) = \frac{(1-\omega)\psi_1(\theta_1 + m\theta_2)}{\lambda} = \eta \tag{8}
$$

and

$$
\begin{aligned}
Variance &= G''(1) + G'(1) - G'(1)^2 \\
&= \frac{(1-\omega)}{\lambda}\left((\theta_1 + m\theta_2)^2 \left\{ \frac{2}{\lambda+1}\psi_2 - \frac{1-\omega}{\lambda}\psi_1^2 \right\} + \psi_1(\theta_1 + m^2\theta_2) \right),
\end{aligned}
$$

where

$$
\psi_j = \delta\phi(j+1;\lambda+j;\theta_1+\theta_2), \; j = 1, 2...., m
$$

Proof is straight forward, hence we omit it.

Result 2.3. *The following is a simple recursion formula for probabilities $f_y(1,\lambda)$ of the ZIEHPD with p.g.f. (7)*

$$f_1(1,\lambda) = \frac{\Psi_1}{\lambda}\left\{f_0(2,\lambda+1)(\theta_1+m\theta_2)-\omega\right\}, for\ y=0,1,...,m-1 \tag{9}$$

and

$$(y+1)f_{y+1}(1,\lambda) = \frac{\Psi_1}{\lambda}\left\{\theta_1 f_y(2,\lambda+1)+m\theta_2\ f_{y-m+1}(2,\lambda+1)\right\},, for\ y \geq m. \tag{10}$$

Proof. From (7), we have

$$\begin{aligned} G(t) &= \omega+(1-\omega)\delta\phi(1;\lambda;\theta_1 t+\theta_2 t^m) &(11)\\ &= \sum_{y=0}^{\infty} t^y f_y(1,\lambda). &(12)\end{aligned}$$

On differentiating (11) and (12) with respect to t, we obtain the following.

$$\frac{(1-\omega)\delta}{\lambda}(\theta_1+m\theta_2 t^{m-1})\phi(2;\lambda+1;\theta_1 t+\theta_2 t^m) = \sum_{y=0}^{\infty}(y+1)t^y f(y+1). \tag{13}$$

Replacing λ by $\lambda+1$ in (11) and (12) to obtain the following.

$$\omega+(1-\omega)\delta^*\phi(2;\lambda+1;\theta_1 t+\theta_2 t^m) = \sum_{y=0}^{\infty} t^y f_y(2,\lambda+1), \tag{14}$$

where $\delta^* = \phi^{-1}(2;\lambda+1;\theta_1+\theta_2)$.
Combining (13) and (14) together lead:

$$\begin{aligned}\sum_{y=0}^{\infty}(y+1)t^y f_{y+1}(1,\lambda) &= \frac{\psi_1}{\lambda}\theta_1\left(\sum_{y=0}^{\infty} t^y f_y(2,\lambda+1)-\omega\right)\\ &+ \frac{\psi_1}{\lambda}m\theta_2\left(\sum_{y=0}^{\infty} t^{y+m-1} f_y(2,\lambda+1)-\omega t^{m-1}\right). &(15)\end{aligned}$$

Now, on equating the coefficients of t^0 on both sides of (15), we get (9), and on equating the coefficients of t^y on both sides of (15), we get (10), □

Result 2.4. *A simple recursion formula for factorial moments $\mu_{[r]}(1,\lambda)$ of the ZIEHPD is the following, for $r \geq 1$, in which $\mu_{[1]} = \frac{(1-\omega)\psi_1(\theta_1+m\theta_2)}{\lambda}$.*

$$\mu_{[r+1]}(1,\lambda) = \frac{\psi_1}{\lambda}(1-\omega)\left[\theta_1\mu_{[r]}(2,\lambda+1)+m\theta_2\sum_{k=0}^{m-1}\binom{m-1}{k}r^{(k)}\mu_{[r-k]}(2,\lambda+1)\right], \tag{16}$$

in which $\mu_{[0]}(1,\lambda)=1$ and $r^{(k)} = r(r-1)(r-2)...(r-k+1)$, for any positive integer k and $r^{(0)}=1$.

Proof. The factorial moment generating function $F(t)$ of the ZIEHPD with p.g.f (7) has the following series representation.

$$\begin{aligned} F(t) &= G(1+t) \\ &= \omega + (1-\omega)\delta\phi(1;\gamma;\theta_1(1+t)+\theta_2(1+t)^m) && (17) \\ &= \sum_{r=0}^{\infty} \mu_{[r]}(1,\lambda)\frac{t^r}{r!}. && (18) \end{aligned}$$

Differentiating (17) and (18) with respect to t to obtain

$$\frac{(1-\omega)\delta}{\lambda}\phi[2;\lambda+1;\theta_1(1+t)+\theta_2(1+t)^m](\theta_1+m\theta_2(1+t)^{m-1}) = \sum_{r=0}^{\infty}\mu_{[r+1]}(1,\lambda)\frac{t^r}{r!}. \quad (19)$$

Replacing λ by $\lambda+1$ in (17) and (18) we get the following from (19).

$$\begin{aligned} \sum_{r=0}^{\infty}\mu_{[r+1]}(1,\lambda)\frac{t^r}{r!} &= \frac{\psi_1}{\lambda}\left[\theta_1\sum_{r=0}^{\infty}\mu_{[r]}(2,\lambda+1)\frac{t^r}{r!}-\omega\right]+m\theta_2 && (20) \\ &\times \sum_{r=0}^{\infty}\sum_{k=0}^{m-1}\binom{m-1}{k}\mu_{[r]}(2,\lambda+1)\frac{t^{r+k}}{r!}-\omega m\theta_2\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{t^k}{k!}. \end{aligned}$$

Now, on equating the coefficients of $\frac{t^r}{r!}$ on both sides of (21), we get (16). $\square$

Result 2.5. *The following is a simple recursion formula for the raw moments $\mu_r(1,\lambda)$ of the ZIEHPD, for $r \geq 0$.*

$$\mu_{r+1}(1,\lambda) = \frac{\psi_1}{\lambda}(1-\omega)\sum_{k=0}^{r}(\theta_1+m^{k+1}\theta_2)\left[\binom{r}{k}\mu_{r-k}(2,\lambda+1)-\omega\right]. \quad (21)$$

Proof. For any $t \in R = (-\infty,\infty)$ and $i = \sqrt{-1}$, the characteristic function of the ZIEHPD is

$$\begin{aligned} \phi(t) &= \sum_{r=0}^{\infty}\mu_r(1,\lambda)\frac{(it)^r}{r!} && (22) \\ &= G(e^{it}) \\ &= \omega+(1-\omega)\delta\phi(1;\lambda;\theta_1 e^{it}+\theta_2 e^{mit}). && (23) \end{aligned}$$

On differentiating (22) and (23) with respect to t, we get the following.

$$(1-\omega)\frac{\delta}{\lambda}\phi[2;\lambda+1;\theta_1 e^{it}+\theta_2 e^{mit}](\theta_1 e^{it}+m\theta_2 e^{mit}) = \sum_{r=0}^{\infty}\mu_{r+1}(1,\lambda)\frac{(it)^r}{r!} \quad (24)$$

Replacing λ by $\lambda+1$, we get the following from (24).

$$\sum_{r=0}^{\infty}\mu_{r+1}(1,\lambda)\frac{(it)^r}{r!}=(1-\omega)\psi_1\frac{(\theta_1e^{it}+m\theta_2e^{mit})}{\lambda}\left[\sum_{r=0}^{\infty}\mu_r(2,\lambda+1)\frac{(it)^r}{r!}-\omega\right] \tag{25}$$

Expanding the exponential terms and on rearranging the terms, and on equating the coefficients of $\frac{(it)^r}{r!}$ we get (21). □

3. Maximum Likelihood Estimation

In this section, we discuss the maximum likelihood estimation procedure for estimating the parameters ω, λ, θ_1 and θ_2 of the ZIEHPD with p.m.f. (6).

For any $y=0,1,2,...$, suppose $a(y)$ be the observed frequency of y events and let z be the highest value of y to be observed. Then the likelihood function of the sample is given by

$$L(\Theta;y)=\prod_{y=0}^{z}[f(y)]^{a(y)},$$

where $f(y)$ is the p.m.f of the ZIEHPD (6). Then the likelihood function can be written as

$$L(\Theta;y)=(f_0)^{a(0)}\prod_{y=1}^{z}[f_1(y)]^{a(y)}.$$

Taking the log-likelihood function, we get

$$\ln L(\Theta;y)=a(0)\ln(\omega+(1-\omega)\delta)+\sum_{y=1}^{\infty}a(y)\ln\left((1-\omega)\delta\sum_{k=0}^{[\frac{y}{m}]}\frac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^k}{(\lambda)_{y-(m-1)k}(y-mk)!k!}\right). \tag{26}$$

The MLEs of $\Theta=(\omega,\ \lambda,\ \theta_1,\ \theta_2)$ can be directly obtained by maximizing the log-likelihood function (26) or alternatively, by finding the solution of the following four likelihood equations using MATHEMATICA software:

$$\frac{\partial\ln L}{\partial\omega}=\frac{a(0)(1-\delta)}{\omega+(1-\omega)}-\sum_{y=1}^{\infty}\frac{a(y)}{(1-\omega)}=0, \tag{27}$$

$$\frac{\partial\ln L}{\partial\theta_1}=\frac{a(0)}{\omega+\frac{(1-\omega)}{\phi(1,\lambda,\theta_1+\theta_2)}}\frac{(1-\omega)}{\lambda}\phi(2,\lambda+1,\theta_1+\theta_2)+\sum_{y=1}^{z}a(y)\left\{\frac{\phi(2,\lambda+1,\theta_1+\theta_2)}{\lambda}\right.$$

$$-\eta \sum_{k=0}^{[y/m]} \frac{(y-(m-1)k)!\theta_1^{y-mk-1}\theta_2^k}{(\lambda)_{y-(m-1)k}(y-mk-1)!k!} = 0, \tag{28}$$

$$\frac{\partial \ln L}{\partial \theta_2} = \frac{a(0)}{\omega + \frac{(1-\omega)}{\phi(1,\lambda,\theta_1+\theta_2)}} \frac{(1-\omega)}{\lambda}\phi(2,\lambda+1,\theta_1+\theta_2) + \sum_{y=1}^{z} a(y)\left\{\frac{\phi(2,\lambda+1,\theta_1+\theta_2)}{\lambda}\right.$$

$$\left. -\eta \sum_{k=0}^{[y/m]} \frac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^{k-1}}{(\lambda)_{y-k}(y-mk)!(k-1)!} = 0 \right. \tag{29}$$

and

$$\frac{\partial logL}{\partial \lambda} = \frac{a(0)(1-\omega)}{\omega+(1-\omega)\delta} \sum_{r=0}^{\infty} (1)_r \frac{(\theta_1+\theta_2)^r}{r!\,(\lambda)_r}[\psi(r)-\psi(\lambda+r)] - \sum_{y=1}^{z} a(y)\times$$

$$\left\{\delta \sum_{r=0}^{\infty} (1)_r \frac{(\theta_1+\theta_2)^r}{r!\,(\lambda)_r}[\psi(\lambda)-\psi(\lambda+r)] + \frac{\eta}{(\lambda)_{y-(m-1)k}}[\psi(\lambda)-\psi(\lambda+y-(m-1)k]\right\} = 0, \tag{30}$$

where

$$\eta = \left[\sum_{k=0}^{[y/m]} \frac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^k}{(\lambda)_{y-(m-1)k}(y-mk)!k!}\right]^{-1}.$$

4. Testing

4.1. Generalized Likelihood Ratio Test

Here we adopt the following generalized likelihood ratio test (GLRT) for testing the significance of the inflation parameter ω.

The null hypothesis H_0 against the alternative hypothesis H_1 is given by

$$H_0 : \omega = 0 \; Vs \; the \; alternative \; hypothesis \; H_1 : \omega \neq 0.$$

The test statistic in the case of GLRT is given by

$$LRT = -2\ln\left[\frac{Sup_{\theta\in\Theta_0} L(\hat{\theta}^*/y)}{Sup_{\theta\in\Theta_1} L(\hat{\theta}/y)}\right] \tag{31}$$

in which $\hat{\theta}$ is the maximum likelihood estimate of $\Theta = (\omega, \lambda, \theta_1, \theta_2)$ with no restrictions and $\hat{\theta^*}$ is the maximum likelihood estimate of θ under H_0. The test statistic given in (31) is asymptotically distributed as chi-square with one degree of freedom.

4.2. Rao's Efficient Score Test

We here consider Rao's efficient score test (REST) for testing the significance of the inflation parameter ω of the ZIEHPD. Thus, to test the null hypothesis

$$H_0 : \omega = 0 \ Vs \ the\ alternative\ hypothesis\ H_1 : \omega \neq 0,$$

the test statistic is

$$S = T' \, \Phi^{-1} \, T, \tag{32}$$

where $T' = \left(\frac{1}{\sqrt{n}} \frac{\partial \log L}{\partial \omega}, \ \frac{1}{\sqrt{n}} \frac{\partial \log L}{\partial \lambda}, \ \frac{1}{\sqrt{n}} \frac{\partial \log L}{\partial \theta_1}, \ \frac{1}{\sqrt{n}} \frac{\partial \log L}{\partial \theta_2} \right)$ and Φ is the Fisher information matrix. The test statistic given in (32) is asymptotically distributed as chi-square with one degree of freedom (df). For more details regarding REST, see [23].

5. Applications

In this section we illustrate all the procedures discussed in sections 3 and 4 with the help of certain real life data sets. The first data set is taken from [19] and the second data is taken from [4]. We have fitted the ZIEHPD to all these data sets and considered the fitting of the models - extended hyper-Poisson distribution (EHPD), zero-inflated Poisson distribution (ZIPD), zero-inflated hyper-Poisson distribution (ZIHPD), zero-inflated Hermite distribution (ZIHD) and the zero-inflated modified hyper-Poisson distribution (ZIMHPD) for comparison. For comparing the models we computed the values of χ^2 statistic, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and second-order Akaike information criterion (AICc). All these numerical results are presented in Tables 1, and 2. Based on the computed values of χ^2 statistic, AIC, BIC and AICc given in Tables 1 and 2 it can be seen that the ZIEHPD for $m = 3$ gives a better fit in the case of first data set and $m = 4$ in the case of second data set considered in the paper.

The computed values of the test statistic given in (31) and (32) are included in Table 3. The null hypothesis is rejected in both the cases, since the critical value for the test at 5% level of significance and one degree of freedom is 3.84. Hence we conclude that the additional parameter ω in the model is significant in both cases of the two data sets.

Table 1. *Data set based on the distribution of corn borers in field experiment arranged in 15 randomized blocks taken from [19] and the expected frequencies computed using $EHPD_{m=3}$, ZIPD, ZIHPD, ZIHD, ZIMHPD, $ZIEHPD_{m=3}$.*

Count	*Observed frequency*	$EHPD_{m=3}$	$EHPD_{m=4}$	*ZIPD*	*ZIHPD*	*ZIHD*	*ZIMHPD*	$ZIEHPD_{m=3}$
0	47	17.72	15.65	46.57	88	52.63	65.9	55.15
1	23	15.74	13.65	9.67	12.23	8.35	20.22	21.27
2	27	30.43	29.57	15.72	6.91	11.76	18.93	23.34
3	9	17.84	17.37	17.63	5.35	13.74	6.8	8.51
4	7	19.62	20.27	13.84	3.61	12.47	3.86	6.23
5	3	9.03	9.86	9.2	2.46	9.42	2.64	2.76
6	1	6.73	8.13	4.87	1.29	6.12	1.22	1.85
7	3	2.89	5.4	2.5	0.15	5.51	0.43	0.89
Total	120	120	120	120	120	120	120	120
df		3	4	4	1	4	1	1
Estimates of the parameters		$\hat{\lambda}$=0.5 $\hat{\theta}_1$=0.5 $\hat{\theta}_2$=0.8	$\hat{\lambda}$=0.49 $\hat{\theta}_1$=0.49 $\hat{\theta}_2$=0.9	$\hat{\omega}$=0.36 $\hat{\lambda}$=3.25	$\hat{\omega}$=0.6 $\hat{\lambda}_1$=12 $\hat{\theta}_2$=7.002	$\hat{\omega}$=0.423 $\hat{\lambda}_1$=3.16 $\hat{\theta}_2$=0.28	$\hat{\omega}$=0.5 $\hat{\lambda}$=0.11 $\hat{\theta}_1$=0.2 $\hat{\theta}_2$=0.20	$\hat{\omega}$=0.5 $\hat{\lambda}$=0.1 $\hat{\theta}_1$=0.5 $\hat{\theta}_2$=0.1
χ^2*-value*		71.92	94.3	40.82	95.09	59.8	14.22	3.89
P-value		0.0001	0.0001	0.0001	0.0001	0.0001	0.002	0.21
AIC		553.6	613.8	424.14	599.02	453.58	4368	421.58
BIC		553.8	614.03	424.31	599.25	453.63	437.2	421.9
AICc		555.98	426.54	605.02	459.6	450.13	424.9	

Table 2. *Data set based on the analysis of thunderstorm events at Cape Kenedy, Florida for June ([4]) and the expected frequencies computed using $EHPD_{m=3}$, $EHPD_{m=4}$, ZIPD, ZIHPD, ZIHD, ZIMHPD, $ZIEHPD_{m=3}$ and $ZIEHPD_{m=4}$.*

Count	*Observed frequency*	$EHPD_{m=3}$	$EHPD_{m=4}$	*ZIPD*	*ZIHPD*	*ZIHD*	*ZIMHPD*	$ZIEHPD_{m=3}$	$ZIEHPD_{m=4}$
0	187	59.20	67.3	185.447	159.05	185.73	240.85	229.67	183.85
1	77	60.19	53.5	33.27	94.2	30.45	33.63	53.34	74.7
2	40	81.44	83.7	40.73	55.133	35.8	30.66	19.70	37.84
3	17	52.94	49	33.94	16.9	30.24	10.4	13.36	20.2
4	6	46.01	51.1	21.41	4	24.58	8.6	8.63	8.21
5	2	20.02	14.9	10.72	0.632	14.8	4.5	3.90	3.1
6	1	10.2	10.5	4.483	0.085	8.4	1.36	1.40	2.1
Total	330	330	330	330	330	330	330	330	330
df		3	3	4	2	3	1	1	1
Estimates of the parameters		$\hat{\lambda}$=0.5 $\hat{\theta}_1$=0.6 $\hat{\theta}_2$=0.7	$\hat{\lambda}$=0.501 $\hat{\theta}_1$=0.5 $\hat{\theta}_2$=0.8	$\hat{\omega}$=0.52 $\hat{\lambda}$=2.51	$\hat{\omega}$=0.4 $\hat{\lambda}$=0.2 $\hat{\theta}$=0.70	$\hat{\omega}$=0.53 $\hat{\lambda}$=1.15 $\hat{\theta}$=1.01	$\hat{\omega}$=0.69 $\hat{\lambda}$=0.12 $\hat{\theta}_1$=0.2 $\hat{\theta}_2$=0.15	$\hat{\omega}$=0.7 $\hat{\lambda}$=0.101 $\hat{\theta}_1$=0.49 $\hat{\theta}_2$=0.1	$\hat{\omega}$=0.5 $\hat{\lambda}$=0.11 $\hat{\theta}_1$=0.505 $\hat{\theta}_2$=0.106
χ^2*-value*		385.38	326.5	86.85	26	109.09	77.18	41.92	2.28
P-value		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.13
AIC		1349.6	1522.8	888.5	1070.1	985.8	861.6	848.24	839.48
BIC		1349.62	1522.68	888.5	1070.02	986.03	861.3	848.024	839.26
AICc		1357.6	1530.8	891.5	1075.02	993.8	881.6	868.2	859.4

We have plotted the observed frequency curves of both the data sets along with the fitted density corresponding to the EHPD, ZIPD, ZIHPD, ZIHD, ZIMHPD and ZIEHPD. From the Tables and figures, it can be seen that the ZIEHPD yields better fit to the data set compared to all the existing models considered in the paper. This also supports the suitability of the proposed model ZIEHPD to both the data set.

Table 3: Calculated values of the test statistic for GLRT and REST for ZIEHPD.

Calculated values of	*Data set* 1	*Data set* 2
GLRT	134.042	685.36
REST	4.25	113.34

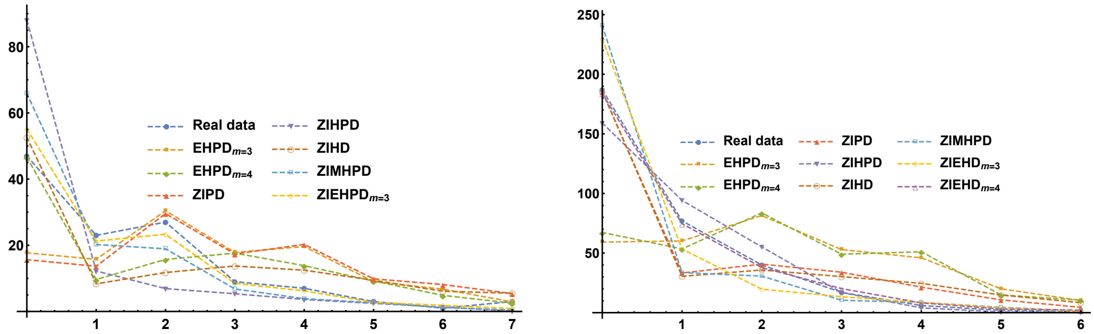

Figure 2. Frequency curves corresponding to various models based on data set 1 and data set 2.

6. Simulation

It is quite difficult for comparing the theoretical performances of estimators of different parameters of the ZIEHPD obtained by the method of maximum likelihood. So, in this section, we have attempted a brief simulation study for comparing the performances of the estimators, and computed their absolute bias and standard errors. The numerical results thus obtained are summarised in Table 4. The initial parameter values used to generate the data sets are:

(1) $\omega = 0.7,\ \lambda = 0.5,\ \theta_1 = 1.2,\ \theta_2 = 0.5$ (over-dispersion)
(2) $\omega = 0.5,\ \lambda = 0.1,\ \theta_1 = 0.5,\ \theta_2 = 0.1$ (under-dispersion)

From Table 4 it can be observed that both the absolute bias and standard errors of the parameters are in the decreasing order as the sample size increases.

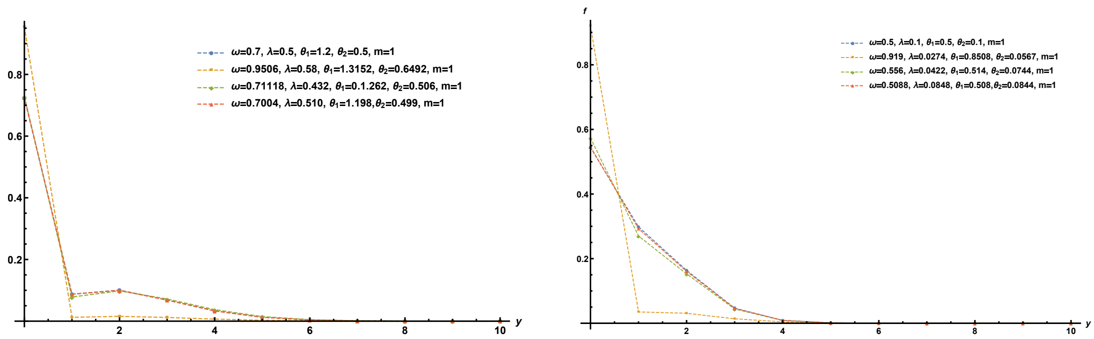

Figure 3. Probability plots corresponding to parameter set-1 and parameter set-2 for $m = 1$

Table 4. *Absolute bias and standard errors (given in parenthesis) corresponding to the estimates obtained via MLE for simulated samples in case of parameter sets: 1, 2 and 3.*

m	Parameter set	Sample size	MLE $\hat{\omega}$	MLE $\hat{\lambda}$	MLE $\hat{\theta}_1$	$\hat{\theta}_2$
1	(1)	100	0.2506	0.08	0.1152	0.1495
			(0.0727)	(0.0584)	(0.1231)	(0.0720)
		200	0.09	0.068	0.062	0.006
			(0.0111)	(0.0064)	(0.0276)	(0.01631)
		500	0.0404	0.0104	0.002	0.004
			(0.0097)	(0.00059)	(0.00127)	(0.000066)
	(2)	100	0.417	0.0726	0.3508	0.1567
			(0.1781)	(0.0222)	(0.1320)	(0.0543)
		200	0.056	0.0578	0.014	0.0256
			(0.0174)	(0.00719)	(0.0094)	(0.0069)
		500	0.0088	0.0152	0.008	0.0156
			(0.00076)	(0.00067)	(0.00138)	(0.00088)
2	(1)	100	0.232	0.06	0.0452	0.06
			(0.082)	(0.0678)	(0.0703)	(0.0478)
		200	0.04	0.0432	0.0277	0.0304
			(0.01778)	(0.0119)	(0.0312)	(0.0115)
		500	0.008	0.015	0.0058	0.004
			(0.0013)	(0.00225)	(0.00145)	(0.0017)
	(2)	100	0.2256	0.0608	0.1828	0.1656
			(0.1362)	(0.0154)	(0.1025)	(0.0558)
		200	0.0488	0.035	0.0186	0.0509
			(0.0232)	(0.00297)	(0.0165)	(0.0034)
		500	0.0172	0.0174	0.004	0.00129
			(0.000755)	(0.000724)	(0.000246)	(0.0000063)
3	(1)	100	0.1406	0.085	0.032	0.1905
			(0.0777)	(0.0409)	(0.0451)	(0.0494)
		200	0.058	0.07	0.004	0.03
			(0.0066)	(0.0064)	(0.0028)	(0.0027)
		500	0.007	0.0036	0.002	0.0174
			(0.0014)	(0.00079)	(0.000874)	(0.00043)
	(2)	100	0.0798	0.1806	0.134	0.304
			(0.0445)	(0.0418)	(0.0963)	(0.1502)
		200	0.052	0.0742	0.0722	0.0762
			(0.0062)	(0.0086)	(0.00655)	(0.007)
		500	0.014	0.0129	0.0156	0.0114
			(0.00068)	(0.000211)	(0.00048)	(0.000185)

Acknowledgement(s)

We are grateful to both the referees for their careful review and constructive comments that substantially improved the article.

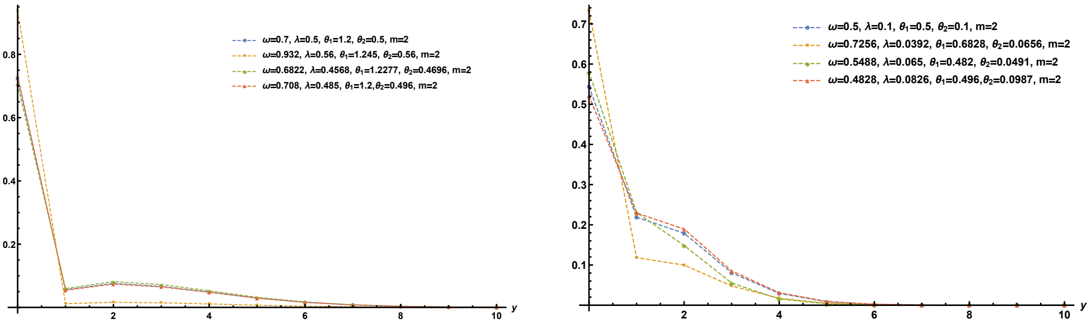

Figure 4. Probability plots corresponding to parameter set-1 and parameter set-2 for $m = 2$

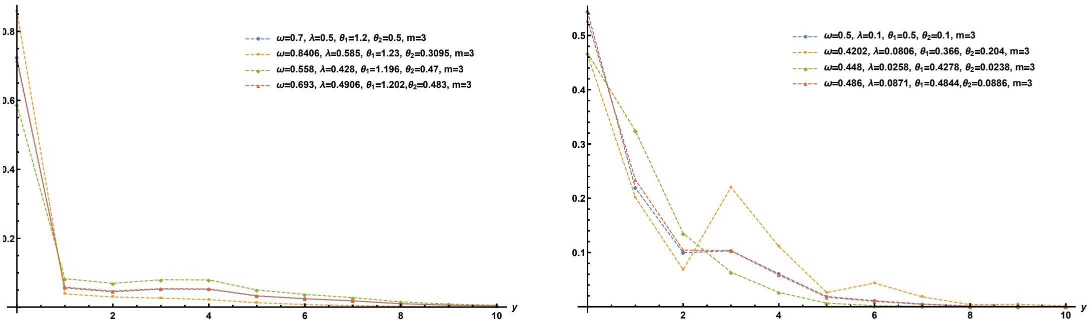

Figure 5. Probability plots corresponding to parameter set-1 and parameter set-2 for $m = 3$

References

[1] Bardwell, G. E. and Crow, E. L. (1964). A two-parameter family of hyper-Poisson distributions. *Journal of the American Statistical Association*, 59(305):133–141.

[2] Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., and Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society: Series A*. **162**, 195–209.

[3] Cohen, A. C. (1963). *Estimation in Mixtures of Discrete Distributions*. Statistical Publishing Society.

[4] Falls, L. W., Williford, W. O., and Carter, M. C. (1971). Probability distributions for thunderstorm activity at cape kennedy, florida. *Journal of Applied Meteorology and Climatology*, 10(1):97–104.

[5] Famoye, F. and Singh, K. P. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*. **4**, 117–130.

[6] Goraski, A. (1977). Distribution z-Poisson. *Institute of Statistical University, Paris*. **12**, 45–53.

[7] Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. John Wiley and Sons, New York.

[8] Kemp, C. (2002). q-analogues of the hyper-Poisson distribution. *Journal of Statistical Planning and Inference*, 101(1-2):179–183.

[9] Kumar, C. S. (2009). Some properties of kemp family of distributions. *Statistica*, 69(4):311–316.

[10] Kumar, C. S. and Nair, B. U. (2011a). An extended version of hyper poisson distribution and some of its applications. *Journal of Applied Statistical Science*, 19(1):81.

[11] Kumar, C. S. and Nair, B. U. (2011b). A modified version of hyper-poisson distribution and its applications. *Journal of Statistics and Applications*, 6(1/2):23.

[12] Kumar, C. S. and Nair, B. U. (2012). An alternative hyper-Poisson distribution. *Statistica*, 72(3):357–369.

[13] Kumar, C. S. and Nair, B. U. (2013a). On stuttering hyper-poisson distribution and its properties. *Sri Lankan Journal of Applied Statistics*, 14(1).

[14] Kumar, C. S. and Nair, B. U. (2013b). Order k version of the alternative hyper-poisson distribution. *Economic Quality Control*, 28(2):97–104.

[15] Kumar, C. S. and Nair, B. U. (2016). On generalized hyper-Poisson distribution of order k and its application. *Communications in Statistics-Theory and Methods*, 45(22):6720–6731.

[16] Kumar, C. S. and Ramachandran, R. (2019). On zero-inflated hyper-Poisson distribution and its applications. *Communications in Statistics-Simulation and Computation*, pages 1–14.

[17] Kumar, C. S. and Ramachandran, R. (2020). On some aspects of a zero-inflated overdispersed model and its applications. *Journal of Applied Statistics*, 47(3):506–523.

[18] Kumar, C. S. and Ramachandran, R. (2022). A generalization to zero-inflated hyper-poisson distribution: Properties and applications. *Communications in Statistics-Theory and Methods*, pages 1–14.

[19] Kumar, C. S. and Riyaz, A. (2016). An order k version of the alternative zero-inflated logarithmic series distribution and its applications. *Journal of Applied Statistics*, 43(14):2681–2695.

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

[21] Martin, D. C. and Katti, S. (1965). Fitting of certain contagious distributions to some available data by the maximum likelihood method. *Biometrics*. **21**, 34–48.

[22] Mathai, A. M. and Haubold, H. J. (2008). *Special functions for applied scientists*, volume 4. Springer, New York.

[23] Rao, C. R. (1973). *Linear statistical inference and its applications*, volume 2. Wiley New York.

[24] Singh, S. (1962). Note on inflated Poisson-distribution. In *Journal of the Indian Statistical Association*. **33**, 140-144.

[25] Slater, L. J. (1966). *Generalized hypergeometric functions*. Cambridge University Press, Cambridge.