**JIRSS (2023)**
**Vol. 22, No. 01, pp 137-160**
**DOI: 10.22034/jirss.2024.707642**

# On Zero-inflated Extended Alternative Hyper Poisson Distribution and its Applications

**Chandrasekharannair Satheesh Kumar** [1], **Rakhi Ramachandran** [1]

[1] Department of Statistics, University of Kerala, Thiruvananthapuram - 695 581, India.

Received: 28/03/2023, Accepted: 15/09/2023, Published online: 15/06/2024

**Abstract.** In this paper we propose a zero-inflated version of the extended alternative hyper-Poisson distribution of Kumar and Nair (2013b) and investigate some of its important properties and applications. We derive expressions for its probability generating function, mean, variance, etc. along with recursion formulae for probabilities, raw moments and factorial 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, generalized likelihood ratio test procedure is 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.** Count Data Modeling, Generalized Hermite Distribution, Generalized Likelihood Ratio Test, Simulation, Zero-inflated Hermite Distribution.

**MSC:** 60E05, 62E15.

## 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 psychosocial researchs. Poisson distribution models are widely used to study 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

Corresponding Author : Chandrasekharannair Satheesh Kumar (drcsatheeshkumar@gmail.com).
Rakhi Ramachandran (rakhir3@gmail.com).

often creates extra variability and consequently the variance will be larger than the mean. This renders the Poisson distribution inappropriate for modeling count data in such instances.

To model count data with excess zeros, Lambert (1992) considered the zero-inflated Poisson (ZIP) distribution, whose probability mass function (pmf) is the following.

$$g_1(x) = P(X = x) = \begin{cases} \pi + (1-\pi)e^{-\lambda}, & x = 0 \\ (1-\pi)\frac{e^{-\lambda}\lambda^x}{x!}, & x = 1, 2, ..., \end{cases} \tag{1.1}$$

where $\pi \in [0, 1]$ 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 Singh (1962), Cohen (1963), Martin and Katti (1965), Goraski (1977), Bohning (1998), Kemp (2002) Johnson et al. (2005) and references therein.

Kumar and Nair (2013b) developed an extended version of the alternative hyper-Poisson distribution namely the extended alternative hyper-Poisson distribution (EAHPD) through the pmf

$$\begin{aligned} g_2(x) &= P(X = x) \qquad (1.2) \\ &= \sum_{k=0}^{[\frac{x}{m}]} \left\{ \frac{[x-(m-1)k]!}{(\lambda)_{x-(m-1)k}} \phi[1 + x - (m-1)k; \lambda + x - (m-1)k; -(\theta_1 + \theta_2)] \right. \\ &\times \left. \frac{\theta_1^{x-mk}\theta_2^k}{(x-mk)!k!} \right\}, \end{aligned}$$

for $x = 0, 1, 2, ...$, $\lambda > 0$, $\theta_1 > 0$, $\theta_2 \geq 0$, $[k]$ denote the integer part of k, m is fixed but positive integer and $\phi(a; b; x)$ is the confluent hypergeometric function (also called Kummer M function). For more details on confluent hypergeometric series see Mathai and Haubold (2008) or Slater (1966). Clearly, the EAHPD with $m = 1$ and/or $\theta_2 = 0$ is the alternative hyper-Poisson distribution (AHPD) of Kumar and Nair (2012) whereas when $m = 2$, the EAHPD reduces to the modified alternative hyper-Poisson distribution (MAHPD) of Kumar and Nair (2013a). When $\lambda$ is a positive integer, it reduces to the generalized version of the displaced Poisson distribution of Staff (1964) and when $\lambda = 1$, the EAHPD reduces to the generalized Hermite distribution of Gupta and Jain (1974). In addition to that, when $\lambda = 1$ and $m = 2$, it reduces to the Hermite distribution of Kemp and Kemp (1965).

In this paper, we try to construct a zero-inflated version of the extended alternative hyper-Poisson distribution through the name "the zero-inflated extended alternative hyper-Poisson distribution (ZIEAHPD)". The rest of the paper is organized as follows. In section 2, we present the definition of the ZIEAHPD and derive its probability generating function, mean, variance and recursion formulae for probabilities, factorial moments and raw moments. The identifiability condition of the model is also derived. The estimation of the parameters of the ZIEAHPD is discussed in section 3 and a generalized likelihood ratio test procedure is suggested in section 4. In section 5, a brief simulation study is carried out for assessing the performance of the maximum likelihood estimators of the parameters of the distribution. Finally, certain real life data applications are considered in section 6 for highlighting the usefulness of the distribution.

## 2 Definition and Properties

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

**Definition 2.1.** Let $\xi$ be a degenerate random variable at the point zero and let X follows EAHPD with pmf (1.2) Assume that $\xi$ and X are statistically independent. Then a discrete random variable Y is said to follow "the zero-inflated extended alternative hyper-Poisson distribution" or in short "the ZIEAHPD" if its pmf has the following form:

$$\begin{aligned} f(y) &= P(Y=y) \\ &= \pi P(\xi=y)+(1-\pi)P(X=y) \\ &= \begin{cases} \pi+(1-\pi)\,\phi[1;\lambda;-(\theta_1+\theta_2)], & y=0 \\ (1-\pi)\sum\limits_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\lambda)_\delta}\phi[1+\delta;\lambda+\delta;-(\theta_1+\theta_2)]\frac{\theta_1^{y-mk}\theta_2^k}{(y-mk)!k!}, & y=1,2,\ldots \\ 0, \ otherwise, \end{cases} \end{aligned} \tag{2.1}$$

in which $\delta=y-(m-1)k$, $[\frac{y}{m}]$ denotes the integer part of $\frac{y}{m}$, $\pi\in[0,1]$, $\lambda>0$, $\theta_1>0$, $\theta_2\geq 0$ and $m$ is a fixed but arbitrary positive integer.

In order to prove that the function $f(y)$ given in (2.1) is a proper pmf, consider

$$\begin{aligned} \sum_{y=0}^{\infty} f(y) &= \pi+(1-\pi)\phi[1;\lambda;-(\theta_1+\theta_2)]+(1-\pi)\sum_{y=1}^{\infty}\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\lambda)_\delta} \\ &\quad\times\phi[1+\delta;\lambda+\delta;-(\theta_1+\theta_2)]\frac{\theta_1^{y-mk}\theta_2^k}{(y-mk)!k!} \\ &= \pi+(1-\pi)\sum_{y=0}^{\infty}\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\lambda)_\delta}\phi[1+\delta;\lambda+\delta;-(\theta_1+\theta_2)]\frac{\theta_1^{y-mk}\theta_2^k}{(y-mk)!k!} \\ &= \pi+(1-\pi)\sum_{y=0}^{\infty} f_1(y), \end{aligned}$$

where $f_1(y) = \sum_{k=0}^{[\frac{y}{m}]} \frac{\delta!}{(\lambda)_\delta}\phi[1+\delta;\lambda+\delta;-(\theta_1+\theta_2)]\frac{\theta_1^{y-mk}\theta_2^k}{(y-mk)!k!}$ is the pmf of the EAHPD given in (1.2). Thus $\sum_{y=0}^{\infty} f(y) = 1$.

Clearly,

1. When $\pi = 0$, ZIEAHPD $\rightarrow$ EAHPD of Kumar and Nair (2013b).

2. When $m = 1$ or $\theta_2 = 0$, ZIEAHPD $\rightarrow$ zero-inflated alternative hyper-Poisson distribution (ZIAHPD) of Kumar and Ramachandran (2021).

3. When $m = 1$ or $\theta_2 = 0$ and $\pi = 0$, ZIEAHPD $\rightarrow$ alternative hyper-Poisson distribution (AHPD) of Kumar and Nair (2012).

4. When $m = 1$ and $\lambda = 1$, ZIEAHPD $\rightarrow$ ZIPD ($\pi$, $\theta_1 + \theta_2$) of Lambert (1992).

5. When $m = 2$ and $\pi = 0$, ZIEAHPD $\rightarrow$ modified alternative hyper-Poisson distribution (MAHPD) of Kumar and Nair (2013a).

6. When $\lambda = 1$ and $\theta_2 = 0$, ZIEAHPD $\rightarrow$ ZIPD of Lambert (1992).

7. When $m = 2$ and $\lambda = 1$, ZIEAHPD $\rightarrow$ zero-inflated Hermite distribution (ZIHD) of Kumar and Ramachandran (2020).

8. When $m = 2$, $\lambda = 1$ and $\pi = 0$ ZIEAHPD $\rightarrow$ Hermite distribution (HD) of Kemp and Kemp (1965).

Probability mass plots of ZIEAHPD for particular values of $\pi, \lambda, \theta_1$ and $\theta_2$, for $m = 1, 2, 3$ and 4 are given in Figure 1.

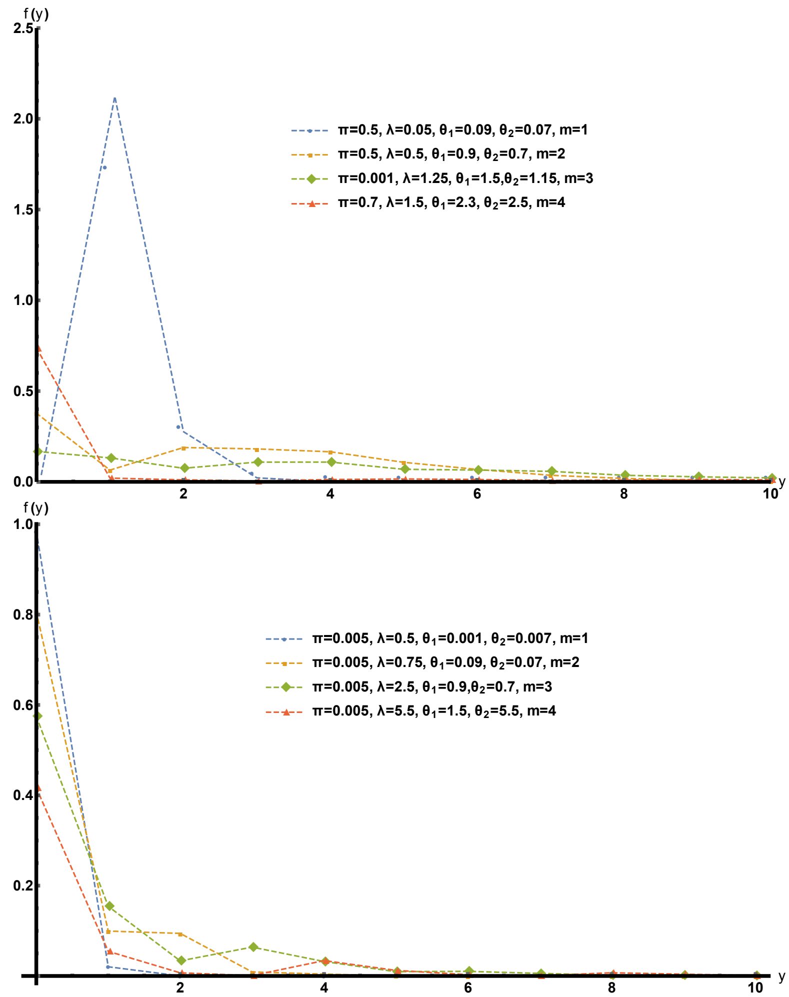


Figure 1: Probability plots of ZIEAHPD for different values of its parameters.

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

**Result 2.1.** *The probability generating function (pgf) $G(t)$ of the ZIEAHPD with pmf (2.1) is the following.*

$$G(t) = \pi + (1 - \pi)\phi[1; \lambda; \theta_1(t - 1) + \theta_2(t^m - 1)]. \tag{2.2}$$

*Proof.* By definition, the pgf of the ZIEAHPD is

$$
\begin{aligned}
G(t) = E(t^Y) &= \sum_{y=0}^{\infty} t^y \, P(Y = y) \\
&= \pi + (1-\pi)\phi[1; \lambda; -(\theta_1 + \theta_2)] + (1-\pi) \sum_{y=1}^{\infty} \sum_{k0}^{[\frac{y}{m}]} \frac{[y-(m-1)k]!}{(\lambda)_{y-(m-1)k}} \\
&\times \frac{(\theta_1 t)^{y-mk} (\theta_2 t^m)^k}{(y-mk)!k!} \phi[1 + y - (m-1)k; \lambda + y - (m-1)k; -(\theta_1 + \theta_2)] \\
&= \pi + (1-\pi) \sum_{y=0}^{\infty} \sum_{k=0}^{[\frac{y}{m}]} \frac{[y-(m-1)k]!}{(\lambda)_{y-(m-1)k}} \frac{(\theta_1 t)^{y-mk} (\theta_2 t^m)^k}{(y-mk)!k!} \\
&\times \phi[1 + y - (m-1)k; \lambda + y - (m-1)k; -(\theta_1 + \theta_2)].
\end{aligned}
$$

Now on expanding the confluent hypergeometric function, in the light of the following identity:

$$
\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} A(s, r) = \sum_{r=0}^{\infty} \sum_{s=0}^{[\frac{r}{m}]} A(s, r - ms), \tag{2.3}
$$

we get,

$$
G(t) = \pi + (1-\pi) \sum_{y=0}^{\infty} \sum_{k=0}^{\infty} \sum_{r=0}^{\infty} \frac{(y+k)!}{(\lambda)_{y+k}} \frac{(\theta_1 t)^y (\theta_2 t^m)^k}{y!k!} \frac{(1+y+k)_r \, [-(\theta_1 + \theta_2)]^r}{r! \, (\lambda + y + k)_r}.
$$

Next, apply the following identity:

$$
\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} A(s, r) = \sum_{r=0}^{\infty} \sum_{s=0}^{r} A(s, r - s), \tag{2.4}
$$

results in:

$$\begin{aligned}
G(t) &= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{\infty}\sum_{k=0}^{y}\frac{y!}{(\lambda)_y}\frac{(\theta_1 t)^{y-k}(\theta_2 t^m)^k}{(y-k)!k!}\frac{(1+y)_r\,[-(\theta_1+\theta_2)]^r}{r!\,(\lambda+y)_r} \\
&= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{\infty}\left\{\sum_{k=0}^{y}\binom{y}{k}(\theta_1 t)^{y-k}(\theta_2 t^m)^k\right\}\frac{(1+y)_r\,[-(\theta_1+\theta_2)]^r}{r!\,(\lambda+y)_r\,(\lambda)_y} \\
&= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{\infty}(\theta_1 t+\theta_2 t^m)^y\frac{(1+y)_r\,[-(\theta_1+\theta_2)]^r}{r!\,(\lambda+y)_r\,(\lambda)_y} \\
&= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{\infty}(\theta_1 t+\theta_2 t^m)^y\frac{(1+y)_r\,[-(\theta_1+\theta_2)]^r}{r!\,(\lambda)_{y+r}} \\
&= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{\infty}t^y(\theta_1+\theta_2 t^m)^y\frac{(y+r)!\,[-(\theta_1+\theta_2)]^r}{y!r!\,(\lambda)_{y+r}}.
\end{aligned}$$

On applying (2.4), we have

$$\begin{aligned}
G(t) &= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{r}(\theta_1 t+\theta_2 t^m t)^y\frac{r!\,[-(\theta_1+\theta_2)]^{r-y}}{y!(r-y)!\,(\lambda)_r} \qquad (2.5) \\
&= \pi + (1-\pi)\sum_{r=0}^{\infty}\sum_{y=0}^{r}\binom{r}{y}(\theta_1 t+\theta_2 t^m)^y\frac{[-(\theta_1+\theta_2)]^{r-y}}{(\lambda)_r} \\
&= \pi + (1-\pi)\sum_{r=0}^{\infty}\frac{[(\theta_1 t+\theta_2 t^m)+-(\theta_1+\theta_2)]^r}{(\lambda)_r},
\end{aligned}$$

which leads to (2.2). □

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

*Remark* 1. Mean and Variance of the ZIEAHPD with pgf (2.2) are, respectively,

$$Mean = G'(1) = \frac{(1-\pi)(\theta_1+m\theta_2)}{\lambda} = \eta, \qquad (2.6)$$

and

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

Proof is straight forward, hence we omit it.

*Remark* 2. The third and fourth factorial moments of the ZIEAHPD are, respectively,

$$\mu_{[3]} = E[Y(Y-1)(Y-2)] = \frac{(1-\pi)\theta_2 m(m-1)}{\lambda}\left\{1+\frac{2(\theta_1+m\theta_2)}{\lambda+1} + \frac{12(\theta_1+m\theta_2)^2}{(\lambda+1)(\lambda+2)}\right\}, \tag{2.7}$$

and

$$\mu_{[4]} = E[Y(Y-1)(Y-2)(Y-3)] = \frac{(1-\pi)\theta_2 m(m-1)}{\lambda}\left\{(m-2)\frac{(\theta_1+m\theta_2)}{(\lambda+1)} + 18\frac{(\theta_1+m\theta_2)}{(\lambda+2)^2} + 6\frac{(\theta_1+m\theta_2)}{(\lambda+2)} + (m-2) \frac{2}{(\lambda+1)} m(m-1)\theta_2 + 12m(m-1) \theta_2\frac{(\theta_1+m\theta_2)^2}{(\lambda+2)}\right\}. \tag{2.8}$$

By using Remark 1 and Remark 2, we have computed measures of skewness and kurtosis with the help of *Mathematica* software and plotted the values in Figures 2 and 3. From the figures, it can be seen that the distribution enjoys positively and negatively skewed nature and both platykurtic and leptokurtic behaviour.

*Remark* 3. The ZIEAHPD is over dispersed when $\lambda > 1$ and under dispersed when $\lambda < 1$ provided

$$\left[\frac{1-\lambda}{1+\lambda} + \pi\right]\frac{(\theta_1+m\theta_2)^2}{\lambda} < m(m-1)\theta_2,$$

for $m > 1$ and $\theta_2 \geq 0$.

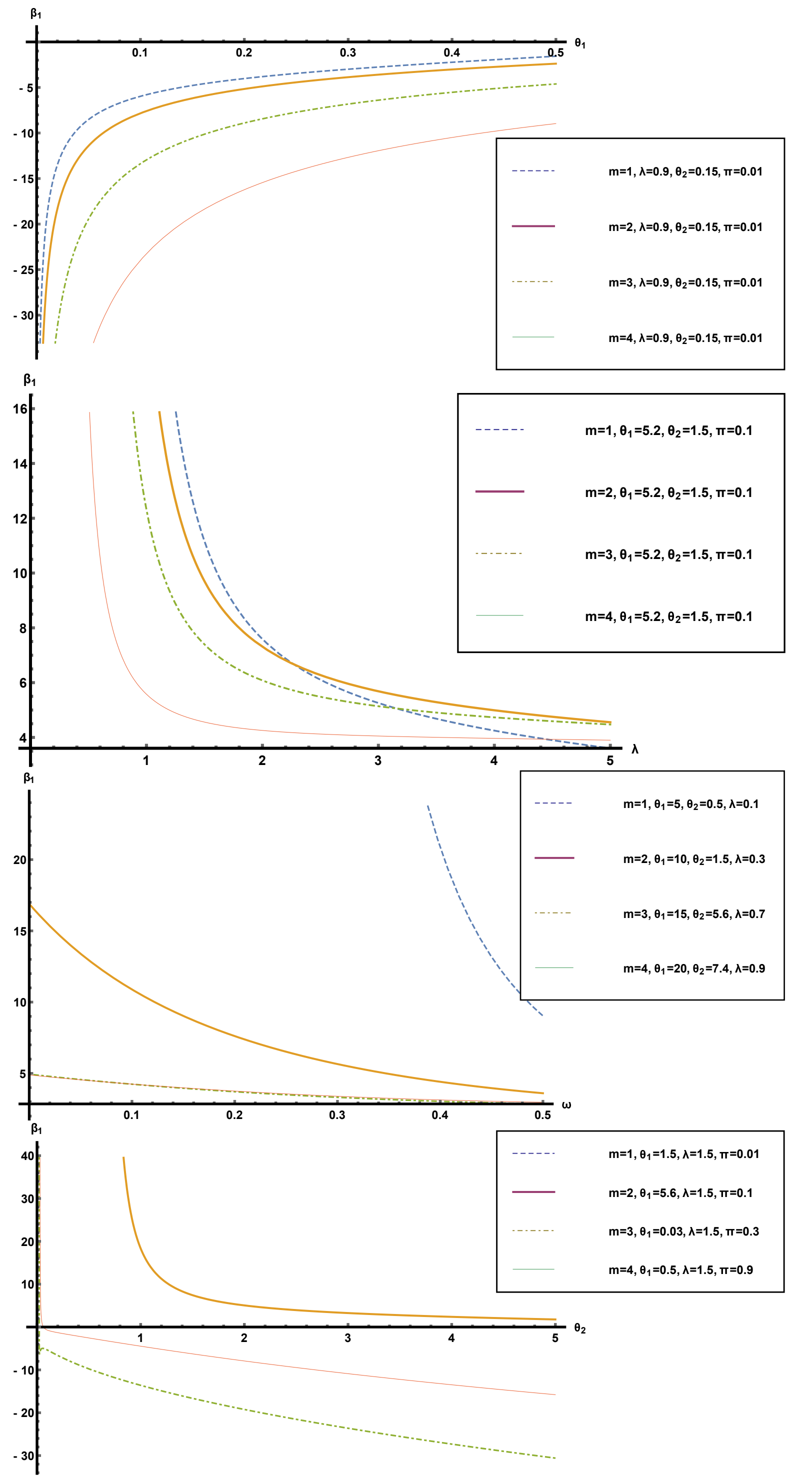


Figure 2: Plots of skewness of ZIEAHPD for particular values of its parameters.

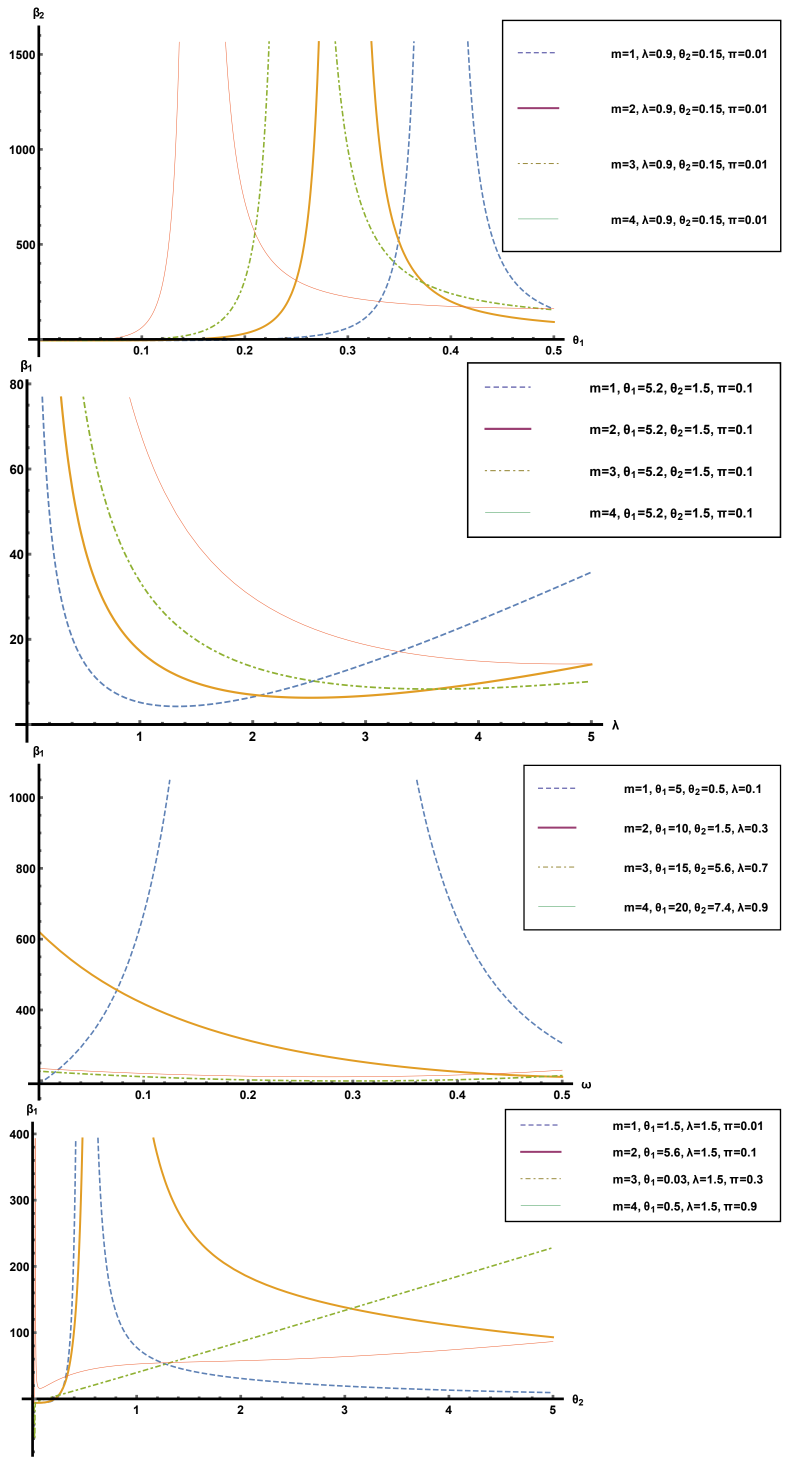


Figure 3: Plots of kurtosis of ZIEAHPD for particular values of its parameters.

**Result 2.2.** *The following is a simple recursion formula for probabilities $f_y(\lambda^*)$ of the ZIEAHPD with pgf (2.2)*

$$f_1(\lambda^*) = \frac{\theta_1}{\lambda}\left\{f_0(\lambda^*) - \pi\right\}, \text{ for } y = 0, 1, ..., m-1, \tag{2.9}$$

and

$$(y+1)f_{y+1}(\lambda^*) = \frac{1}{\lambda}\left(\theta_1 f_y(\lambda^*+1) + m\theta_2 f_{y-m+1}(\lambda^*+1)\right), \text{ for } y \geq m. \tag{2.10}$$

*Proof.* From (2.2), we have

$$\begin{aligned} G(t) &= \pi + (1-\pi)\phi(1;\lambda;\theta_1(t-1) + \theta_2(t^m-1)) &(2.11)\\ &= \sum_{y=0}^{\infty} t^y f_y(\lambda^*). &(2.12)\end{aligned}$$

On differentiating (2.11) and (2.12) with respect to $t$, we obtain the following:

$$\frac{(1-\pi)}{\lambda}\phi[2;\lambda+1;\theta_1(t-1)+\theta_2(t^m-1)](\theta_1 + m\theta_2 t^{m-1}) = \sum_{y=0}^{\infty}(y+1)f_{y+1}(\lambda^*)t^y. \tag{2.13}$$

Replacing $\lambda$ by $\lambda+1$ in (2.11) and (2.12) results in the following equality:

$$\pi + (1-\pi)\phi[2;\lambda+1;\theta_1(t-1)+\theta_2(t^m-1)] = \sum_{y=0}^{\infty} f_y(\lambda^*+1)t^y. \tag{2.14}$$

Combining (2.13) and (2.14) together leads to:

$$\frac{1}{\lambda}\left(\sum_{y=0}^{\infty} t^y f_y(\lambda^*+1) - \pi\right)(\theta_1 + m\theta_2 t^{m-1}) = \sum_{y=0}^{\infty}(y+1)f_{y+1}(\lambda^*)t^y, \tag{2.15}$$

that is,

$$\sum_{y=0}^{\infty}(y+1)f_{y+1}(\lambda^*)t^y = \frac{(\theta_1 + m\theta_2 t^{m-1})}{\lambda}\left(\sum_{y=0}^{\infty} t^y f_y(\lambda^*+1) - \pi\right). \tag{2.16}$$

Now, on equating the coefficients of $t^0$ on both sides of (2.16), we get (2.9), and on equating the coeffients of $t^y$ on both sides of (2.16), we get (2.10). □

**Result 2.3.** *A simple reccursion formula for factorial moments $\mu_{[r]}(\lambda^*)$ of the ZIEAHPD, for $r \geq 1$, is the following:*

$$\mu_{[r+1]}(\lambda^*) = \frac{1}{\lambda}\left[\theta_1\mu_{[r]}(\lambda^*+1) + m\theta_2\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{r!}{(r-k)!}\mu_{[r-k]}(\lambda^*+1)\right], \tag{2.17}$$

*in which $\mu_{[0]}(\lambda^*) = 1$ and $\mu_{[1]} = \frac{(1-\pi)(\theta_1+m\theta_2)}{\lambda}$.*

*Proof.* The factorial moment generating function $F(t)$ of the ZIEAHPD with pgf (2.2) has the following series representation:

$$
\begin{aligned}
F(t) &= G(1+t) \\
&= \pi + (1-\pi)\phi[1;\lambda;\theta_1 t + \theta_2((1+t)^m - 1)] \qquad (2.18)\\
&= \sum_{r=0}^{\infty} \mu_{[r]}(\lambda^*)\frac{t^r}{r!}. \qquad (2.19)
\end{aligned}
$$

Differentiate (2.18) and (2.19) with respect to $t$ to obtain

$$
\frac{(1-\pi)}{\lambda}\phi[2;\lambda+1;\theta_1 t + \theta_2((1+t)^m - 1)](\theta_1 + m\theta_2(1+t)^{m-1}) = \sum_{r=0}^{\infty} \mu_{[r+1]}(\lambda^*)\frac{t^r}{r!}. \qquad (2.20)
$$

By replacing $\lambda$ by $\lambda + 1$ in (2.18) and (2.19), we get the following equality from (2.20):

$$
\pi + (1-\pi)\phi[2;\lambda+1;\theta_1 t + \theta_2((1+t)^m - 1)] = \sum_{r=0}^{\infty} \mu_{[r]}(\lambda^* + 1)\frac{t^r}{r!}. \qquad (2.21)
$$

Combining (2.20) and (2.21) together we get:

$$
\begin{aligned}
\sum_{r=0}^{\infty} \mu_{[r+1]}(\lambda^*)\frac{t^r}{r!} &= \frac{1}{\lambda}\theta_1\left(\sum_{r=0}^{\infty}\mu_{[r]}(\lambda^*+1)\frac{t^r}{r!} - \pi\right) + \frac{m\theta_2}{\lambda}\left(\sum_{r=0}^{\infty}\sum_{k=0}^{m-1}\binom{m-1}{k}\right. \\
&\qquad \left. \mu_{[r]}(\lambda^*+1)\frac{t^{r+k}}{(r+k)!} - \pi\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{t^k}{k!}\right).
\end{aligned}
\qquad (2.22)
$$

Now, on equating the coefficients of $\frac{t^r}{r!}$ on both sides of (2.22), we get (2.17). □

**Result 2.4.** *The following is a simple recursion formula for the raw moments $\mu_r(\lambda^*)$ of the ZIEAHPD, for $r \geq 0$:*

$$
\mu_{r+1}(\lambda^*) = \frac{1}{\lambda}\sum_{j=0}^{r}(\theta_1 + m^{j+1}\theta_2)\left[\binom{r}{j}\mu_{r-j}(\lambda^*+1) - \pi\right]. \qquad (2.23)
$$

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

$$
\begin{aligned}
\phi(t) &= G(e^{it}) \\
&= \sum_{r=0}^{\infty}\mu_r(\lambda^*)\frac{(it)^r}{r!} \qquad (2.24)\\
&= \pi + (1-\pi)\phi[1;\lambda;\theta_1(e^{it}-1) + \theta_2(e^{mit}-1)]. \qquad (2.25)
\end{aligned}
$$

On differentiating (2.24) and (2.25) with respect to $t$, we get the following:

$$\frac{(1-\pi)}{\lambda}\phi[2;\lambda+1;\theta_1(e^{it}-1)+\theta_2(e^{mit}-1)]\left(\theta_1e^{it}+m\theta_2e^{mit}\right)=\sum_{r=0}^{\infty}\mu_{r+1}(\lambda^*)\frac{(it)^r}{r!}. \quad (2.26)$$

By replacing $\lambda$ by $\lambda+1$, we get the following from (2.26):

$$\pi+(1-\pi)\phi[2;\lambda+1;\theta_1(e^{it}-1)+\theta_2(e^{mit}-1)]=\sum_{r=0}^{\infty}\mu_r(\lambda^*+1)\frac{(it)^r}{r!}. \quad (2.27)$$

Combining (2.26) and (2.27), we get

$$\sum_{r=0}^{\infty}\mu_{r+1}(\lambda^*)\frac{(it)^r}{r!}=\frac{1}{\lambda}\left(\theta_1e^{it}+m\,\theta_2e^{mit}\right)\left[\sum_{r=0}^{\infty}\mu_r(\lambda^*+1)\frac{(it)^r}{r!}-\pi\right]. \quad (2.28)$$

Expanding the exponential terms and on rearranging the terms of (2.28), and on equating the coefficients of $\frac{(it)^r}{r!}$, we get (2.23). □

## 2.1 Model Identifiability

Let $Y$ be a discrete random varible with pmf $p(y)=P(Y=y)$ of the form $p(y)=\pi_1p_1(y)+\pi_2p_2(y)+...+\pi_gp_g(y)$, where for each $j=1,2,\dots,g$; $\pi_j>0$ such that $\sum_{j=1}^{g}\pi_j=1$, $p_j(y)\geq 0$ and $\sum_{j=1}^{g}p_j(y)=1$. Then, we say that $Y$ has a mixture distribution and $p(y)$ is a finite mixture of distributions. The parameters $\pi_1,\pi_2,\dots,\pi_g$ are known as mixing weights and $p_1,p_2,\dots,p_g$ are the components of the mixture. We denote $\Psi$ as the collection of all distinct parameters occuring in the components and $\Theta$ as the complete collection of all distinct parameters occuring in the mixture model.

A parametric family of densities $f(y_j;\Psi)$ is identifiable if distinct values of the parameter $\Psi$ determine distinct members of the family of densities $\left\{f(y_j;\psi):\psi\in\Omega\right\}$, where $\Omega$ is a specified parameter space; that is

$$f(y_j;\psi)=f(y_j;\psi^*), \quad (2.29)$$

if and only if

$$\psi=\psi^*. \quad (2.30)$$

Identifiability for mixture distribution is slightly different. Suppose that $f(y_j;\psi)$ has two component densities, say, $f_i(y_j;\beta_j)$ and $f_h(y_j;\beta_j)$, that belongs to the same parametric family.

Let

$$f(y_j;\psi)=\sum_{i=1}^{g}\pi_if_i(y_j;\beta_i),$$

and

$$f(y_j;\psi*) = \sum_{i=1}^{g} \pi_i f_i(y_j;\beta_i^*),$$

be any two members of a parametric family of mixture densities. The class of finite mixtures is said to be identifiable for $\psi \in \Omega$ if $f(y_j;\psi) \equiv f(y_j;\psi^*)$ if and only if $g = g*$ and we can permute the component labels so that $\pi_i = \pi_i^*$ and $f_i(y_j;\beta_i) = f_i(y_j;\beta_i^*)$ for $i = 1, 2, ..., g$.

Now, we use the following lemma which we need in the sequel for establishing the identifiability condition of the model considered in this paper.

**Lemma 2.1.** *(Titterington et al., 1985) A necessary and sufficient conditionthat a model to be identifiable is that the distribution function of convex combination of mixture densities is linearly independent over the field of real numbers.*

**Definition 2.2.** A random variable $Y$ is said to have g component mixture model of zero-inflated extended alternative hyper-Poisson distribution if it has the following pmf $p(y)$, in which $0 \leq \pi_j \leq 1$, for $j = 1, 2, ..., k$, $\sum_{j=1}^{k} \pi_j = 1$ and $y = 0, 1, 2, ...$

$$f(y) = \sum_{j=1}^{k} \pi_j q_j(y), \tag{2.31}$$

where

$$\begin{aligned} q_j(y) &= \sum_{k=0}^{[\frac{x}{m}]} \left\{ \frac{[x-(m-1)k]!}{(\lambda_j)_{x-(m-1)k}} \phi[1+x-(m-1)k; \lambda_j + x-(m-1)k; -(\theta_{1j}+\theta_{2j})] \right. \\ &\times \left. \frac{\theta_{1j}^{x-mk}\theta_{2j}^{k}}{(x-mk)!k!} \right\}, \end{aligned} \tag{2.32}$$

with $\lambda_j > 0$, $\theta_{1j} > 0$ and $\theta_{2j} \geq 0$ for each $j = 1, 2, \ldots, k$.

**Result 2.5.** *The ZIEAHPD with p.m.f $f(y)$ given in (2.1) is identifiable with respect to the parameters $\lambda$, $\theta_1$ and $\theta_2$.*

*Proof.* Consider the equation

$$a_1 G_1(y) + a_2 G_2(y) = 0, \tag{2.33}$$

where $a_1$ and $a_2$ are any two arbitrary real numbers and $G_1(y) = \sum_{j=1}^{y} f(j)$ and $G_2(y) = \sum_{j=1}^{y} h(j)$, for $y = 0, 1, ...,$ where $h_j$ obtained from $f_j$ by replacing $\lambda_j$ by $\sigma_j$, $\theta_{1j}$ by $\rho_j$ and $\theta_{2j}$ by $\tau_j$. Assume that for each $j = 1, 2$, $\lambda_j \neq \sigma_j$, $\theta_{1j} \neq \rho_j$ and $\theta_{2j} \neq \tau_j$. Thus,

$$\begin{aligned} G_1(y) &= \pi + (1-\pi)\phi[1+j;\lambda+j;-(\theta_1+\theta_2)] \\ &+ \sum_{j=1}^{y}(1-\pi)\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\lambda_j)_\delta}\phi[1+\delta;\lambda+\delta;-(\theta_1+\theta_2)]\frac{\theta_{1j}^{y-mk}\theta_{2j}^{k}}{(y-mk)!k!}, \end{aligned} \tag{2.34}$$

and

$$\begin{aligned} G_2(y) &= \pi + (1-\pi)\phi[1+j;\sigma+j;-(\rho+\tau)] \\ &+ \sum_{j=1}^{y}(1-\pi)\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\sigma_j)_\delta}\phi[1+\delta;\sigma+\delta;-(\rho+\tau)]\frac{\rho_j^{y-mk}\tau_j^{k}}{(y-mk)!k!}. \end{aligned} \tag{2.35}$$

Combining (2.33), (2.34) and (2.35), we have

$$\begin{aligned} &a_1\sum_{j=0}^{y}\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\lambda_1)_\delta}\phi[1+\delta;\lambda_1+\delta;-(\theta_{11}+\theta_{21})]\frac{\theta_{11}^{y-mk}\theta_{21}^{k}}{(y-mk)!k!} = a_2\sum_{j=0}^{y}\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\sigma_1)_\delta} \\ &\times\phi[1+\delta;\sigma_1+\delta;-(\rho_1+\tau_1)]\frac{\rho_1^{y-mk}\tau_1^{k}}{(y-mk)!k!}, \end{aligned} \tag{2.36}$$

and

$$\begin{aligned} &a_1\sum_{j=0}^{y}\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\lambda_2)_\delta}\phi[1+\delta;\lambda_2+\delta;-(\theta_{12}+\theta_{22})]\frac{\theta_{12}^{y-mk}\theta_{22}^{k}}{(y-mk)!k!} = a_2\sum_{j=0}^{y}\sum_{k=0}^{[\frac{y}{m}]}\frac{\delta!}{(\sigma_2)_\delta} \\ &\times\phi[1+\delta;\sigma_2+\delta;-(\rho_2+\tau_2)]\frac{\rho_2^{y-mk}\tau_2^{k}}{(y-mk)!k!}. \end{aligned} \tag{2.37}$$

Eliminate $a_1$ using (2.31) and (2.32), we have $a_2 = 0$. From (2.31), we obtain $a_1 = 0$ shows that $G_1$ and $G_2$ are linearly independent. □

# 3 Maximum Likelihood Estimation

Here we consider the estimation of the parameters $\pi$, $\lambda$, $\theta_1$ and $\theta_2$ of the ZIEAHPD by the method of maximum likelihood. For any $y = 0, 1, 2, ...$, let $A(y)$ be the observed frequency of $y$ events and let $z$ be the highest value of $y$ 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 pmf of the ZIEAHPD given in (2.1).
Now, $L(\Theta; y)$ can be written as

$$L(\Theta; y) = (f_0)^s \prod_{y=1}^{z} (f_1(y))^{A(y)},$$

where $s = A(0)$, $f_0$ is the pmf of the ZIEAHPD when $y = 0$ and $f_1(.)$ is the pmf of the distribution when $y = 1, 2, \ldots$ .

Then the log-likelihood function can be written as

$$\begin{aligned}
\ln L(\theta; y) &= s \ln\Big(\pi + (1-\pi)\phi[1; \lambda; -(\theta_1 + \theta_2)]\Big) + \sum_{y=1}^{z} A(y)\, ln(1-\pi) \\
&+ \sum_{y=1}^{z} A(y) \ln \left\{ \sum_{k=0}^{[\frac{y}{m}]} \frac{(y-(m-1)k)!}{(\lambda)_{y-(m-1)k}} \right. \\
&\times \left. \phi[1 + y - (m-1)k; \lambda + y - (m-1)k; -(\theta_1 + \theta_2)] \right\} \\
&+ \sum_{y=1}^{z} A(y)\, ln \left\{ \sum_{k=0}^{[\frac{y}{m}]} \frac{\theta_1^{y-mk} \theta_2^k}{(y-mk)!k!} \right\}. \qquad (3.1)
\end{aligned}$$

Assume that $\hat{\pi}$, $\hat{\lambda}$, $\hat{\theta_1}$ and $\hat{\theta_2}$ be the maximum likelihood estimators of the parameters $\pi$, $\lambda$, $\theta_1$ and $\theta_2$ of the ZIEAHPD. Now, on differentiating the log-likelihood function (3.1) with respect to $\pi$, $\lambda$, $\theta_1$ and $\theta_2$ and equating to zero, we obtain the following likelihood equations:

$$\frac{\partial \log L}{\partial \pi} = 0,$$

$$\frac{s\Big(1 - \phi[1; \lambda; -(\theta_1 + \theta_2)]\Big)}{\pi + (1-\pi)\phi[1; \lambda; -(\theta_1 + \theta_2)} - \sum_{y=1}^{z} A(y) \frac{1}{(1-\pi)} = 0, \qquad (3.2)$$

$$\begin{aligned}
\frac{\partial \log L}{\partial \lambda} = 0 \qquad &\Longrightarrow \tfrac{s\ (1-\pi)}{\pi+(1-\pi)\phi[1;\lambda;-(\theta_1+\theta_2)]} \textstyle\sum_{r=0}^{\infty} \tfrac{[-(\theta_1+\theta_2)]^r}{(\lambda)_r} [\psi(\lambda) - \psi(\lambda + r)] + \\
&\textstyle\sum_{y=1}^{z} A(y) \sum_{k=0}^{[\frac{y}{m}]} \dfrac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^k}{(y-mk)!k![(\lambda)_{y-(m-1)k}]^2} \Bigg\{ \sum_{r=0}^{\infty} \tfrac{(1+y-(m-1)k)_r}{(\lambda+y-(m-1)k)_r} \\
&\tfrac{[-(\theta_1+\theta_2)^r]}{r!} \Big[ \psi(\lambda + y - (m-1)k) - \psi(\lambda + y - (m-1)k + r)] \\
&- \phi[1; \lambda; -(\theta_1 + \theta_2)] \tfrac{1}{(\lambda)}_{y-(m-1)k} \\
&\Big[\psi(\lambda + y - (m-1)k) - \psi(\lambda + y - (m-1)k + r)] \Bigg\} = 0, \qquad (3.3)
\end{aligned}$$

$$\frac{\partial \log L}{\partial \theta_1} = 0 \qquad \Longrightarrow \frac{-s\frac{(1-\pi)}{\lambda}\phi[2;\lambda+1;-(\theta_1+\theta_2)]}{\pi+(1-\pi)\phi[1;\lambda;-(\theta_1+\theta_2)]} + \sum_{y=1}^{z} A(y)\frac{1}{\xi}$$
$$\Bigg\{\sum_{k=0}^{[\frac{y}{m}]} \frac{(y-(m-1)k)!\theta_1^{y-(m-1)k-1}\theta_2^{k}}{(\lambda)_{y-(m-1)k}(y-(m-1)k-1)!k!}$$
$$\phi[1+y-(m-1)k;\lambda+y-(m-1)k;-(\theta_1+\theta_2)]$$
$$-\sum_{k=0}^{[\frac{y}{m}]} \frac{(y-(m-1)k)!\theta_1^{y-(m-1)k}\theta_2^{k}}{(\lambda)_{y-(m-1)k}(y-mk)!k!}\frac{(1+y-(m-1)k)}{(\lambda+y-(m-1)k)}$$
$$\phi[2+y-(m-1)k;\lambda+1+y-(m-1)k;-(\theta_1+\theta_2)]\Bigg\} = 0. \tag{3.4}$$

and

$$\frac{\partial \log L}{\partial \theta_2} = 0 \qquad \Longrightarrow \frac{-s\frac{(1-\pi)}{\lambda}\phi[2;\lambda+1;-(\theta_1+\theta_2)]}{\pi+(1-\pi)\phi[1;\lambda;-(\theta_1+\theta_2)]} + \sum_{y=1}^{z} A(y)\frac{1}{\xi}$$
$$\Bigg\{\sum_{k=0}^{[\frac{y}{m}]} \frac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^{k-1}}{(\lambda)_{y-(m-1)k}(y-mk)!(k-1)!}$$
$$\phi[1+y-(m-1)k;\lambda+y-(m-1)k;-(\theta_1+\theta_2)]$$
$$-\sum_{k=0}^{[\frac{y}{m}]} \frac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^{k}(1+y-(m-1)_k)}{(\lambda)_{y-(m-1)k}(y-mk)!k!(\lambda+y-(m-1)_k)}$$
$$\phi[2+y-(m-1)k;\lambda+1+y-(m-1)k;-(\theta_1+\theta_2)]\Bigg\} = 0. \tag{3.5}$$

in which $\psi(\lambda) = \frac{\partial}{\partial\lambda} log\ \lambda(\lambda)$ and
$\xi = \sum_{k=0}^{[\frac{y}{m}]} \frac{(y-(m-1)k)!\theta_1^{y-mk}\theta_2^{k}}{(\lambda)_{y-(m-1)k}(y-mk)!k!}\phi[1+y-(m-1)k;\lambda+y-(m-1)k;(\theta_1+\theta_2)].$

On solving the likelihood equations (3.2), (3.3), (3.4), and (3.5) with the help of Mathematical software, say *Mathematica*, one can obtain the maximum likelihood estimators of the parameters of the proposed distribution.

## 4 Generalized Likelihood Ratio Test

Now we consider the generalized likelihood ratio test (GLRT) procedure for testing the significance of the additional parameters of the model. The null hypotheses considered here are

1. $H_0^{(1)} : \theta_2 = 0$ against the alternative hypothesis $H_1^{(1)} : \theta_2 \neq 0$
2. $H_0^{(2)} : \pi = 0$ against the alternative hypothesis $H_1^{(2)} : \pi \neq 0$

3. $H_0^{(3)} : \theta_2 = 0, \lambda = 1$ against the alternative hypothesis $H_1^{(3)} : \theta_2 \neq 0, \lambda \neq 1$.

The test statistic is

$$-2 \ln \psi = 2 \left(\iota_1 - \iota_2\right), \tag{4.1}$$

where, $\iota_1 = \ln L(\hat{\Theta}; y)$, $\hat{\Theta}$ is the maximum likelihood estimator for $\Theta = (\pi, \lambda, \theta_1, \theta_2)$ with no restrictions, $\iota_2 = lnL(\hat{\Theta}^*; y)$, and $\hat{\Theta}^*$ is the maximum likelihood estimator for $\Theta$ under the null hypothesis $H_0$. The test statistic given in (4.1) follows chi-square distribution with one degree of freedom for those hypotheses having one parameter restriction and two degrees of freedom for those hypotheses having two parameter restrictions.

# 5 Simulation

It is quite difficult to compare the theoretical performances of the estimators of different parameters of the ZIEAHPD 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. Since we cannot apply inverse transformation method to simulate ZIEAHPD random samples, we have generated the samples as per the following steps:

1. Specification of the values of the parameter sets as ($\pi$, $\lambda$, $\theta_1$, $\theta_2$) = (0.7, 0.3, 0.25, 0.6) and (0.2, 0.3, 0.1, 0.7) in case of over-dispersion and under-dispersion, respectively.

2. Specification of the sample size. Here we considered samples of sizes 100, 200 and 500.

3. Generation of pseudo-random sample of ZIEAHPD utilizing Metropolis-Hastings algorithm for computation of the maximum likelihood estimates of the parameters using maximum likelihood estimation.

4. Generation of 1000 samples.

5. Computation of Bias and standard errors.

The numerical results thus obtained are summarised in Table 1. We have used the R Software (R Core Team, 2019) to find the estimates and sample generation. Probability plots corresponding to the simulated data sets in case of over dispersion and under dispersion are as given in Figures 4, 5 and 6. From Table 1 and Figures 4, 5 and 6, 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.

Table 1: Absolute bias and standard errors (given in paranthesis) corresponding to the estimates obtained via MLE for simulated samples in case of both parameter sets for $m = 1,\ 2$ and 3.

| | | | MLE | | | |
|---|---|---|---|---|---|---|
| m | Parameter set | Sample size | $\hat{\pi}$ | $\hat{\lambda}$ | $\hat{\theta_1}$ | $\hat{\theta_2}$ |
| 1 | (1) | 100 | 0.36 | 0.06 | 0.24 | 0.86 |
| | | | (0.09) | (0.032) | (0.151) | (0.12) |
| | | 200 | 0.05 | 0.006 | 0.098 | 0.078 |
| | | | (0.01) | (0.0026) | (0.064) | (0.098) |
| | | 500 | 0.0015 | 0.0015 | 0.006 | 0.0087 |
| | | | (0.0025) | (0.00059) | (0.00124) | (0.00066) |
| | (2) | 100 | 0.32 | 0.04 | 0.24 | 0.15 |
| | | | (0.14) | (0.0222) | (0.024) | (0.089) |
| | | 200 | 0.11 | 0.0026 | 0.11 | 0.06 |
| | | | (0.068) | (0.00719) | (0.002) | (0.0054) |
| | | 500 | 0.079 | 0.00078 | 0.006 | 0.0078 |
| | | | (0.0012) | (0.00067) | (0.0006) | (0.00099) |
| 2 | (1) | 100 | 0.113 | 0.04 | 0.0987 | 0.088 |
| | | | (0.06) | (0.095) | (0.012) | (0.0025) |
| | | 200 | 0.025 | 0.01 | 0.0102 | 0.0088 |
| | | | (0.0018) | (0.010) | (0.0026) | (0.0010) |
| | | 500 | 0.009 | 0.0025 | 0.0089 | 0.00074 |
| | | | (0.00065) | (0.00113) | (0.0009) | (0.00014) |
| | (2) | 100 | 0.55 | 0.08 | 0.61 | 0.99 |
| | | | (0.012) | (0.025) | (0.105) | (0.015) |
| | | 200 | 0.06 | 0.002 | 0.11 | 0.01 |
| | | | (0.0087) | (0.0016) | (0.025) | (0.0021) |
| | | 500 | 0.0087 | 0.0009 | 0.0025 | 0.0012 |
| | | | (0.00042) | (0.00047) | (0.000235) | (0.0006) |
| 3 | (1) | 100 | 0.26 | 0.01 | 0.32 | 0.24 |
| | | | (0.012) | (0.08) | (0.01) | (0.05) |
| | | 200 | 0.0014 | 0.007 | 0.04 | 0.08 |
| | | | (0.006) | (0.01) | (0.006) | (0.0048) |
| | | 500 | 0.00098 | 0.00085 | 0.002 | 0.0059 |
| | | | (0.001) | (0.0004) | (0.00063) | (0.0003) |
| | (2) | 100 | 0.74 | 0.15 | 0.123 | 0.24 |
| | | | (0.032) | (0.06) | (0.096) | (0.023) |
| | | 200 | 0.12 | 0.08 | 0.012 | 0.0147 |
| | | | (0.008) | (0.009) | (0.0058) | (0.0087) |
| | | 500 | 0.025 | 0.009 | 0.00456 | 0.00485 |
| | | | (0.0004) | (0.000123) | (0.00054) | (0.000156) |

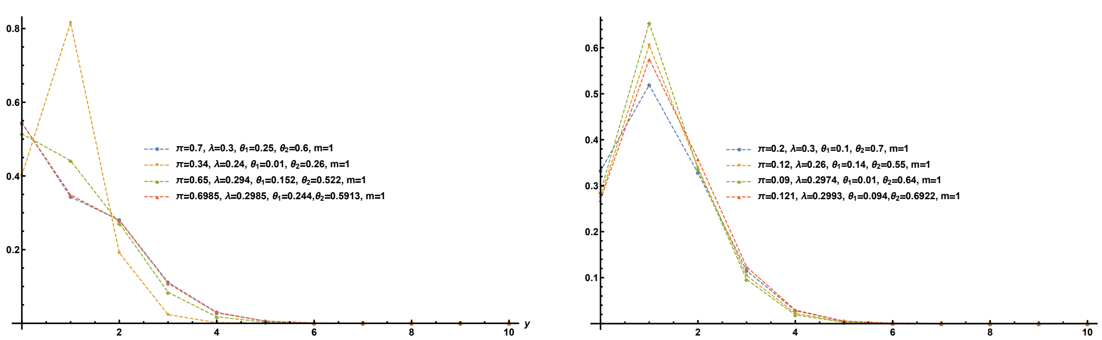


Figure 4: Probability plots corresponding to parameter set-1 and parameter set-2 for $m = 1$

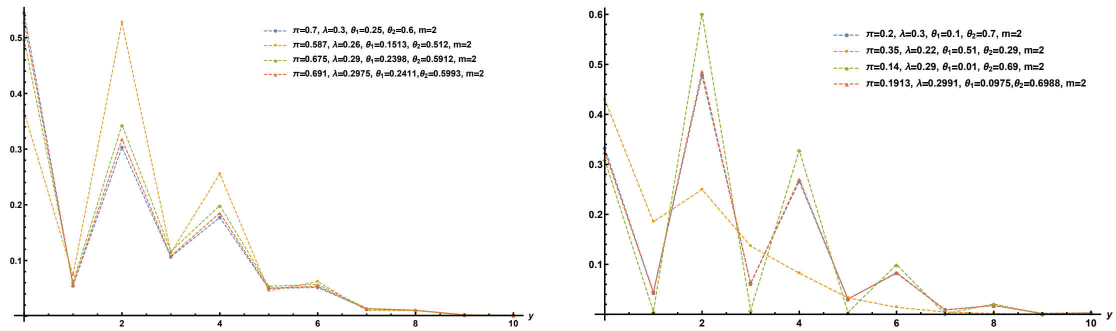


Figure 5: Probability plots corresponding to parameter set-1 and parameter set-2 for $m = 2$

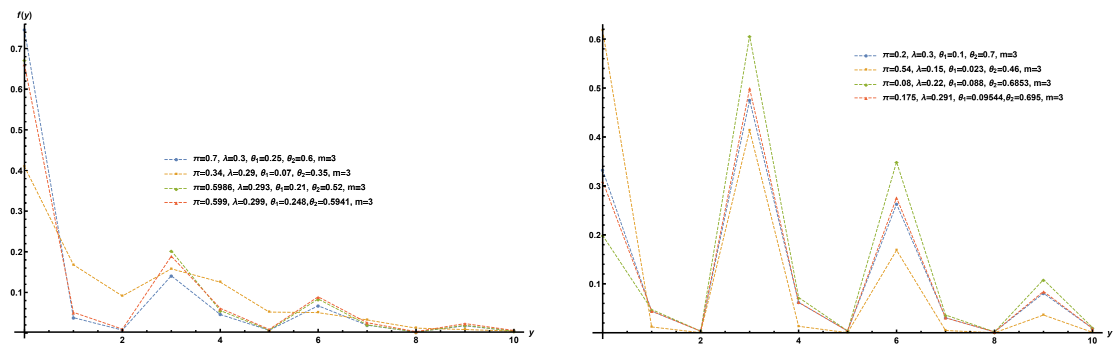


Figure 6: Probability plots corresponding to parameter set-1 and parameter set-2 for $m = 3$

# 6 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 on the distribution of Ribes (Fracker and Brischle, 1944) and the second data is on the distribution of the counts of red mites on apple leaves (Bliss and Fisher, 1953). We have fitted the ZIEAHPD to these two data sets and considered the fitting of the models - zero-inflated alternative hyper-Poisson distribution (ZIAHPD), ZIPD, zero-inflated Hermite distribution (ZIHD) and extended alternative hyper-Poisson distribution (EAHPD) for comparison. For comparing the models, we computed the values of $\chi^2$ and information measures, such as Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and Corrected Akaike's Information Criterion (AICc). The numerical results obtained are presented in Tables 2 and 3. Based on the computed values of $\chi^2$, AIC, BIC and AICc, as presented in Tables 2 and 3, it can be observed that ZIEAHPD for $m = 3$ gives a better fit in the case of the first data set and for $m = 4$ in the case of second data set considered in the paper.

We have also plotted the observed frequency curves of the data sets along with the fitted densities corresponding to the ZIAHPD, ZIPD, ZIHD, EAHPD and the ZIEAHPD. From Tables 2 and 3 and Figure 7, it can be seen that these models do not provide the best fit to the data sets while the ZIEAHPD gives best fit based on the p-value and chi-square value. The values of information measures like AIC, BIC and AICc also support the fact that the ZIEAHPD can be considered as a suitable model compared to the other models discussed in the paper.

Table 2: Distribution of Ribes (Fracker and Brischle, 1944) and the expected frequencies computed using ZIAHPD, $ZIEAHPD_{m=2}$, ZIPD, ZIHD, $EAHPD_m = 3$, $EAHPD_m = 4$ and $ZIEAHPD_{m=3}$.

| *Count* | *Observed frequency* | *ZIAHPD* | $ZIEAHPD_{m=2}$ | *ZIPD* | *ZIHD* | $EAHPD_{m=3}$ | $EAHPD_{m=4}$ | $ZIEAHPD_{m=3}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 43 | 11.23 | 48.12 | 59.55 | 38.28 | 33.85 | 33.3 | 41.72 |
| 1 | 15 | 19.345 | 8.05 | 9.42 | 5.3 | 21.52 | 23.03 | 13.27 |
| 2 | 8 | 29.85 | 7.03 | 5.1 | 8.4 | 14.09 | 13.04 | 6.4 |
| 3 | 6 | 14.27 | 6.5 | 3.5 | 5.7 | 5.38 | 5.39 | 7.5 |
| 4 | 3 | 4.2 | 5.13 | 1.7 | 5.2 | 2.9 | 2.91 | 5.85 |
| 5 | 4 | 0.95 | 3.29 | 0.68 | 6.06 | 1.55 | 1.55 | 4.7 |
| 6 | 0 | 0.16 | 1.15 | 0.041 | 5.66 | 0.68 | 0.75 | 0.06 |
| 7 | 1 | 0.025 | 0.73 | 0.009 | 5.4 | 0.03 | 0.03 | 0.5 |
| *Total* | 80 | 80 | 80 | 80 | 80 | 80 | 80 | 80 |
| *df* | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Estimates* | | $\pi$=0.59<br>$\lambda$=0.20<br>$\theta$=0.90 | $\pi$=0.7<br>$\lambda$=0.9<br>$\theta_1$=0.3<br>$\theta_2$=0.35 | $\pi$=0.85<br>$\lambda$=1.5 | $\pi$=0.49<br>$\lambda$=18.8<br>$\theta$=0.41 | $\lambda$=0.8<br>$\theta_1$=0.51<br>$\theta_2$=0.15 | $\lambda$=0.81<br>$\theta_1$=0.5<br>$\theta_2$=0.15 | $\pi$=0.79<br>$\lambda$=0.20<br>$\theta_1$=0.30<br>$\theta_2$=0.35 |
| $\chi^2$*-value* | | 155.99 | 97.48 | 64.40 | 63.49 | 41.13 | 44.45 | 2.12 |
| *P-value* | | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0828 |
| *AIC* | | 375.3 | 343.8 | 354.9 | 1129.7 | 329.6 | 339.7 | 303.34 |
| *BIC* | | 375.6 | 344.2 | 354.99 | 1129.9 | 330.1 | 340.2 | 303.6 |
| *AICc* | | 381.34 | 357.1 | 357.2 | 1135.7 | 335.6 | 345.7 | 316.67 |

Now by adopting the test procedure discussed in section 4, we test $H_0^{(1)} : \theta_2 = 0$ against the alternative hypothesis $H_1^{(1)} : \theta_2 \neq 0$, $H_0^{(2)} : \pi = 0$ against the alternative hypothesis $H_1^{(2)} : \pi \neq 0$, and $H_0^{(3)} : \theta_2 = 0, \lambda = 1$ against the alternative hypothesis $H_1^{(3)} : \theta_2 \neq 0, \lambda \neq 1$. The numerical results obtained are given in Table 4.

Based on the computed values of GLRT and its p-values from Table 4, one can observe that the null hypotheses are rejected for both data sets, which indicate the suitability of the proposed model ZIEAHPD for both data sets considered in this paper.

Table 3: Distribution of the counts of red mites on apple leaves (Bliss and Fisher, 1953) and the expected frequencies computed using ZIAHPD, $ZIEAHPD_{m=2}$, ZIPD, ZIHD, $EAHPD_m$ = 3, $EAHPD_m$ = 4, $ZIEAHPD_{m=3}$ and $ZIEAHPD_{m=4}$.

| *Count* | *Observed frequency* | *ZIAHPD* | *ZIEAHPD$_{m=2}$* | *ZIPD* | *ZIHD* | *EAHPD$_{m=3}$* | *EAHPD$_{m=4}$* | *ZIEAHPD$_{m=3}$* | *ZIEAHPD$_{m=4}$* |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 70 | 38.8 | 99.4 | 87.34 | 100.74 | 62.572 | 63.1 | 79.3 | 68.55 |
| 1 | 38 | 56.6 | 19.2 | 24.5 | 17.5 | 45.2 | 45.17 | 18.5 | 32.16 |
| 2 | 17 | 38.23 | 10.3 | 19.8 | 16.06 | 26.6 | 26.3 | 5.3 | 15.07 |
| 3 | 10 | 11.3 | 7.5 | 3.6 | 8.4 | 10.32 | 10.1 | 18.9 | 12.27 |
| 4 | 9 | 4.51 | 5.6 | 8.39 | 4.6 | 3.8 | 3.567 | 12.5 | 11.3 |
| 5 | 3 | 0.52 | 4.3 | 3.7 | 2.1 | 1.16 | 1.4 | 2.8 | 5.18 |
| 6 | 2 | 0.037 | 1.6 | 1.85 | 0.5 | 0.27 | 0.28 | 7.6 | 3.33 |
| 7 | 1 | 0.0027 | 1.2 | 0.8 | 0.1 | 0.06 | 0.068 | 4.5 | 1.4 |
| 8 | 0 | 0.000312 | 0.9 | 0.02 | 0 | 0.0182 | 0.015 | 0.6 | 0.74 |
| *Total* | 150 | 150 | 150 | 150 | 150 | 150 | 150 | 150 | |
| *df* | | 1 | 1 | 3 | 2 | 1 | 1 | 3 | 2 |
| *Estimates* | | $\pi$=0.58<br>$\lambda$=0.21<br>$\theta$=0.601 | $\pi$=0.8<br>$\lambda$=0.91<br>$\theta_1$=0.3<br>$\theta_2$=0.35 | $\pi$=0.75<br>$\lambda$=1.25 | $\pi$=0.5<br>$\lambda$=0.67<br>$\theta$=0.4 | $\lambda$=0.82<br>$\theta_1$=0.51<br>$\theta$2=0.16 | $\lambda$=0.81<br>$\theta_1$=0.51<br>$\theta_2$=0.15 | $\pi$=0.76<br>$\lambda$=0.2<br>$\theta_1$=0.31<br>$\theta_2$=0.3 | $\pi$=0.58<br>$\lambda$=0.5<br>$\theta_1$=0.9<br>$\theta_2$=0.7 |
| *$\chi^2$-value* | | 62.5 | 34.8 | 22.7 | 38.9 | 22.7 | 22.8 | 60.08 | 4.25 |
| *P-value* | | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.2 | |
| *AIC* | | 709.9 | 699.4 | 698.8 | 814.1 | 634.6 | 659 | 645.4 | 627.5 |
| *BIC* | | 710.5 | 700.1 | 699.2 | 814.8 | 635.1 | 659.9 | 646.18 | 628.3 |
| *AICc* | | 714.7 | 709.4 | 700.8 | 818.9 | 639.4 | 663.8 | 658.7 | 637.5 |

Table 4: Computed values of $L(\hat{\theta}^*; y)$, $L(\hat{\theta}; y)$, generalized likelihood ratio test statistics and p-values of the ZIEAHPD.

| Data | Test | $L(\hat{\theta}^*; y)$ | $L(\hat{\theta}; y)$ | GLRT | d.f | Chi-square value (tabled value) | P-value |
|---|---|---|---|---|---|---|---|
| Dataset 1 | Test 1 | -184.67 | -147.67 | 74 | 1 | 3.84 | 0.0001 |
| | Test 2 | -161.8 | -147.67 | 28.3 | 1 | 3.84 | 0.0001 |
| | Test 3 | -175.4 | -147.67 | 55.4 | 2 | 5.99 | 0.0001 |
| Dataset 2 | Test 1 | -351.9 | -309.8 | 84.8 | 1 | 3.84 | 0.0001 |
| | Test 2 | -326.5 | -309.8 | 33.4 | 1 | 3.84 | 0.0001 |
| | Test 3 | -347.5 | -309.8 | 75.4 | 2 | 5.99 | 0.0001 |

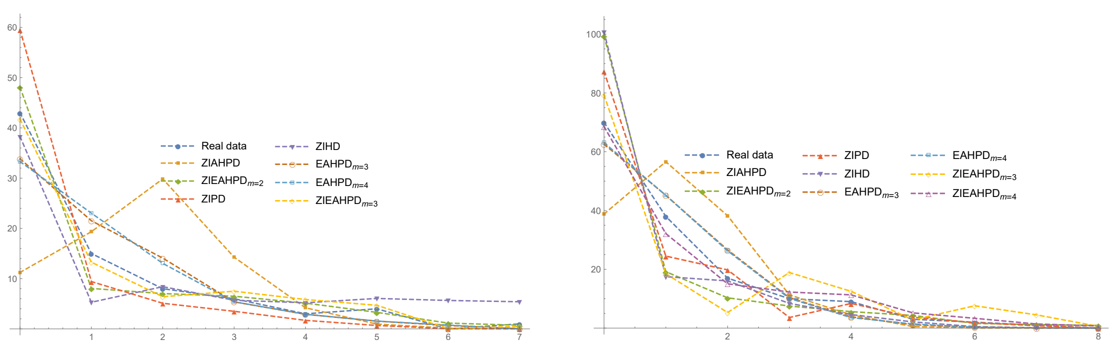


Figure 7: Frequency curves corresponding to various models based on data set 1 and data set 2.

## 7 Conclusion

This article introduces a more flexible class of the zero-inflated Poisson distribution namely "the zero-inflated extended alternative hyper-Poisson distribution (ZIEAHPD)." The statistical properties of the distribution such as moments, generating function, recursion formulae, etc are presented. The estimation of the parameters of the distribution have been obtained using maximum likelihood estimation. The generalized likelihood ratio test procedure is constructed for testing the significance of the additional parameters of the model and a brief simulation study has been carried out for assessing the efficiency of the estimation procedure discussed in the paper. Finally, two real life data applications are considered for illustrating the usefulness of the proposed model compared to the other existing models such as ZIAHPD, ZIPD, ZIHD and EAHPD.

**Acknowledgements**

The authors are highly thankful to the Editor in Chief, the Associate editor and the anonymous Referees for their fruitful comments on an earlier version of the paper that greatly improved the quality and presentation of the paper. The first author is particularly thankful to Department of Science and Technology, New Delhi (Government of India) for financial support (MTR/2017/000942).

## References

Bliss, C. I., and Fisher, R. A.(1953). Fitting the negative binomial distribution to biological data. *Biometrics*, **9**, 176-200.

Bohning, D. (1998). Zero-inflated Poisson models and CA MAN: A tutorial collection of evidence. *Biometrical Journal*, **40**, 833-843.

Cohen, A. C. (1963). Estimation in mixtures of discrete distributions (pp. 373-378). Statistical Pub. Society.

Fracker, S. B., and Brischle, H. A. (1944). Measuring the local distribution of Ribes. *Ecology*, **25**, 283-303.

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

Gupta, R. P., and Jain, G. C. (1974). A generalized Hermite distribution and its properties. *SIAM Journal on Applied Mathematics*, **27**, 359-363.

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

Kemp, C. (2002). q-analogues of the hyper-Poisson distribution. *Journal of Statistical Planning and Inference*, **101**, 179-183.

Kemp, C. D., and Kemp, A. W. (1965). Some properties of the Hermite distribution. *Biometrika*, **52**, 381-394.

Kumar, C. S., and Nair, B. U. (2012). An alternative hyper-Poisson distribution. *Statistica*, **72**, 357-369.

Kumar, C. S., and Nair, B. U. (2013a). Modified alternative hyper-Poisson distribution. *Collection of Recent Statistical Methods and Applications*, Vol-2, No-1, 97-109.

Kumar, C. S., and Nair, B. U. (2013b). On extended alternative hyper-Poisson distribution. *The Aligarh Journal of Statistics*, **33**, 119-128.

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

Kumar, C. S., and Ramachandran, R (2021). On zero-inflated alternative hyper-Poisson distribution. *Statistica*, **81**, 423-446.

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

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.

Mathai, A. M., and Haubold, H. J. (2008). *Special Functions for Applied Scientists*. Springer-Verlag, NewYork.

R Core Team. (2019). R: A language and statistical, statistical computing. Vienna, Austria: R Foundation for Computing. https://www. R-project.org.

Singh, S. N. (1962). Note on inflated Poisson distribution. *Annals of Mathematical Statistics*, **33**, 1210.

Slater, L. J. (1966). *Generalized Hypergeometric Functions*. Cambridge University Press, Cambridge.

Staff, P. J. (1964). The displaced Poisson distribution. *Australian Journal of Statistics*, **6**, 12-20.

Titterington, D. M., A. F. M. Smith, and U. E. Markov. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, New York.