STATISTICA, anno LXXXI, n. 4, 2021

ON ZERO-INFLATED ALTERNATIVE HYPER-POISSON DISTRIBUTION

C. Satheesh Kumar [1]
Department of Statistics, University of Kerala, Thiruvananthapuram, India
Rakhi Ramachandran
Department of Statistics, University of Kerala, Thiruvananthapuram, India

1. INTRODUCTION

Count data with excess zeros are so common in many disciplines like agriculture, biology, ecology, engineering, epidemiology, psychology, public health, sociology, etc. Examples of such data include the number of family members with cholera in a village in India (M'Kendrick, 1925), the number of over 80-year-old female deaths per day (Hasselblad, 1969), the number of feotus movements per 5 seconds (Leroux and Puterman, 1992), the number of HIV infected patients (Van den Broek, 1995) and the number of ambulance requests for heat-related illnesses (Bassil *et al.*, 2010).

To model count data with excess zeros, several zero-inflated models have been studied in the literature, among them, zero-inflated Poisson distribution (ZIPD) is of particular interest. The probability mass function (p.m.f) of the distribution is

$$f(x) = \begin{cases} \pi + (1-\pi)e^{-\lambda}, & for\ x = 0 \\ (1-\pi)\frac{e^{-\lambda}\lambda^x}{x!}, & for\ x = 1, 2, \ldots. \end{cases} \tag{1}$$

The ZIPD has been further investigated by several authors such as Singh (1962), Cohen (1963), Martin and Katti (1965), Goraski (1977), Lambert (1992), Bohning (1998), Kemp (2002), Barriga and Louzada (2014), and Sim *et al.* (2018). Zero-inflated versions of the generalized Poisson distribution (ZIGPD) and negative binomial distribution (ZINBD) were also studied in the literature. Kumar and Ramachandran (2019) considered zero-inflated hyper-Poisson distribution (ZIHPD) as a relatively better model compared to both ZIGPD and ZINBD in certain situations, and shown that it is suitable for both over dispersed and under dispersed data sets, while the ZIGPD overestimates the dispersion and ZINBD underestimate the dispersion. Saez-Castillo and Conde-Sanchez (2013) proposed a regression model for the hyper-Poisson distribution (HPD).

[1] Corresponding Author. E-mail: drcsatheesh@gmail.com

Kumar and Nair (2012) considered an alternative form of the hyper-Poisson distribution (AHPD). The p.m.f of the AHPD as follows, for $m = 0, 1, 2, ...$

$$P(M = m) = \frac{\nu^m}{(\rho)_m} \phi(1 + m; \rho + m; -\nu), \tag{2}$$

in which $\rho > 0$, $\nu > 0$ and

$$\phi(a; b; \nu) = \sum_{p=0}^{\infty} \frac{(a)_p \, \nu^p}{(b)_p \, p!}$$

is the confluent hypergeometric function. The AHPD contains a symbol called Pochhammer symbol defined as follows

$$(a)_p = a(a+1)(a+2)...(a+p-1) = \frac{\Gamma(a+p)}{\Gamma(a)}, \; for \; p = 1, 2, ... \; and \; (a)_0 = 1.$$

A distribution with p.m.f (2), they termed as the "alternative hyper-Poisson distribution (AHPD)" and hereafter in this paper we use $AHPD$ to denote the distribution. The Poisson distribution is the special case of $AHPD$ when $\rho = 1$. Moreover, over-dispersion and under-dispersion in cases of $\rho > 1$ and $\rho < 1$ is also one of the important characteristic of the $AHPD$.

Since in certain practical situations the zero-inflated models such as ZIPD, ZIGPD, ZINBD, ZIHPD etc. are not suitable, while a zero-inflated version of the AHPD gives better fit. For example in the case of the real life data sets considered in the application section of this paper only the zero-inflated version of the AHPD provides best fit whereas all other existing models are not suitable. So through this paper, we propose a zero-inflated alternative hyper-Poisson distribution (ZIAHPD) and study its important properties and other important aspects. Also it is important to note that zero-inflated version of the AHPD is not studied yet in the literature.

This paper is organized as follows. In Section 2, we present the definition of the ZIAHPD and obtain its probability generating function, expression for its mean, variance, factorial moments and recursion formulae for probabilities, factorial moments and raw moments. The identifiability condition of the model is also derived. Further, we discuss the esimation of the parameters of the model along with its relevence with the help of two real life data sets. A test procedure called the generalized likelihood ratio test (GLRT) is applied for examining the significance of the inflation parameter. In addition, to identify the performance of the maximum likelihood estimator of the parameters, we have conducted a simulation study.

Consider the following series representations, those we need in the sequel.

$$\sum_{x=0}^{\infty} \sum_{p=0}^{\infty} A(p, x) = \sum_{x=0}^{\infty} \sum_{p=0}^{x} A(p, x - p) \tag{3}$$

$$\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{(l)_{j+k}(m)_k\, x^j y^k}{(n)_{j+k}\, j!\, k!} = \xi_1(l,m;n;x,y) \tag{4}$$

where $\xi_1(l,m;n;x,y)$ is the Horn-Appel function, which is a generalization of the hypergeometric function in two variables.

2. Definition and Properties

In this Section, we introduce the ZIAHPD through deriving its important statistical properties.

Definition 1. *Let Λ be a random variable degenerated at the point zero and let X follows AHPD(ρ,ν). Assume that Λ and X are independent. Then the random variable W is said to follow "the zero-inflated alternative hyper-Poisson distribution" (or in short "the ZIAHPD") if its p.m.f is of the form*

$$\begin{aligned} g(w) = P(W=w) &= \pi P(\Lambda = w) + (1-\pi)P(X=w) \\ &= \begin{cases} \pi + (1-\pi)\phi(1;\rho;-\nu), & w=0 \\ (1-\pi)\frac{\nu^w}{(\rho)_w}\phi(1+w;\rho+w;-\nu), & w=1,2,\ldots \\ 0, \ otherwise, \end{cases} \end{aligned} \tag{5}$$

in which $\pi \in [0,1]$, $\rho > 0$ and $\nu > 0$.

In order to prove that the function $g(w)$ given in (5) is a proper p.m.f, consider

$$\begin{aligned} \sum_{w=0}^{\infty} g(w) &= \pi + (1-\pi)\phi(1;\rho;-\nu) + (1-\pi)\sum_{w=1}^{\infty}\frac{\nu^w}{(\rho)_w}\phi(1+w;\rho+w;-\nu) \\ &= \pi + (1-\pi)\sum_{w=0}^{\infty}\frac{\nu^w}{(\rho)_w}\phi(1+w;\rho+w;-\nu) \\ &= \pi + (1-\pi)\sum_{w=0}^{\infty} g_1(w), \end{aligned}$$

where $g_1(w) = \frac{\nu^w}{(\rho)_w}\phi(1+w;\rho+w;-\nu)$ is the p.m.f of the alternative hyper-Poisson distribution given in (2). Thus, $\sum_{w=0}^{\infty} g(w) = 1$.

Clearly, when $\rho = 1$, the ZIAHPD reduces to the ZIPD with p.m.f given in (1) and when $\pi = 0$, the ZIAHPD reduces to the AHPD with p.m.f given in (2).

Illustrations of p.m.f of ZIAHPD for different values of π, ρ and ν are given in Figures 1-4. Now we obtain the following results.

Result 2.1. *The probability generating function (p.g.f) H(t) of the ZIAHPD with p.m.f (5) is the following.*

$$H(t) = \pi + (1-\pi)\xi_1(1,-;\rho;t\nu,-\nu) \tag{6}$$

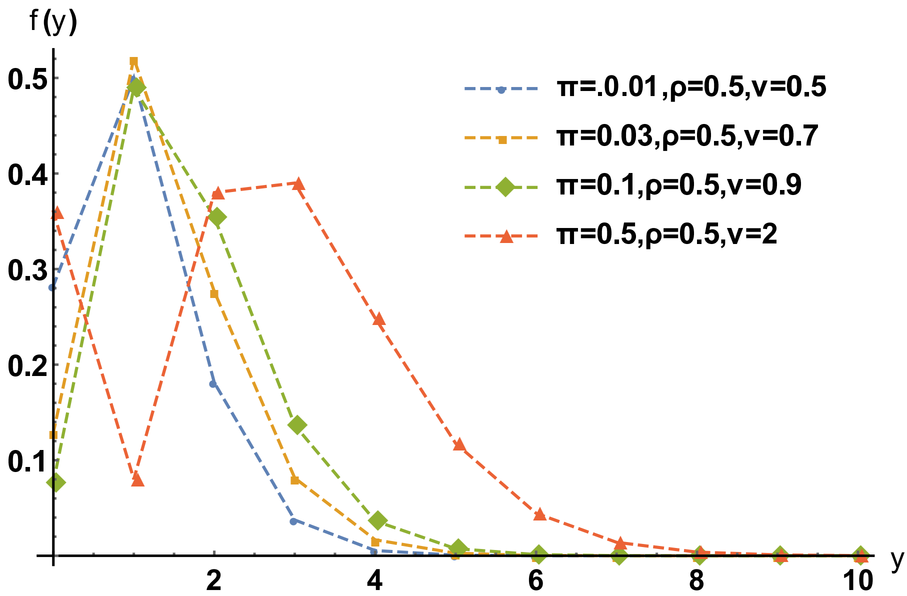

Figure 1 – Plots of probability mass functions of ZIAHPD for different values of π *and* ν.

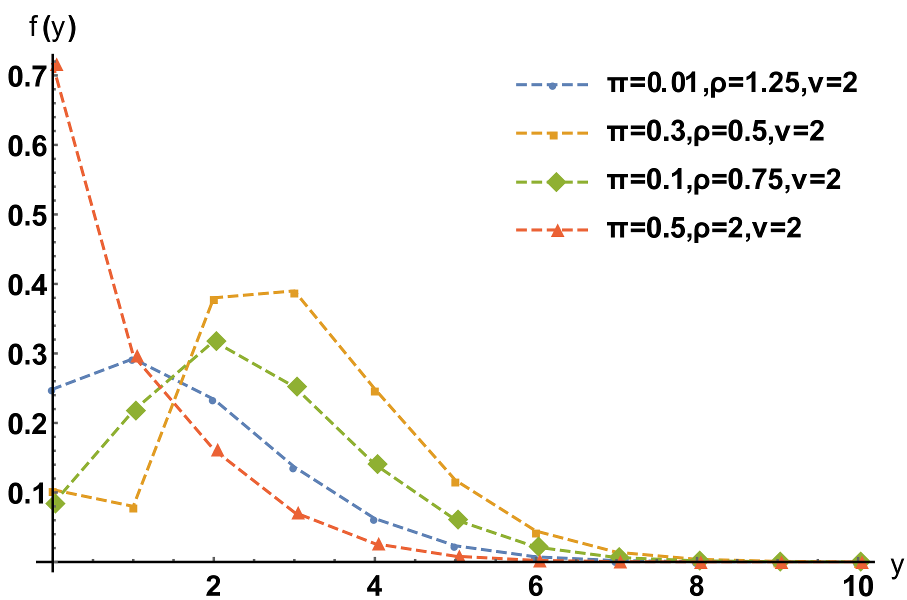

Figure 2 – Plots of probability mass functions of ZIAHPD for different values of π *and* ρ.

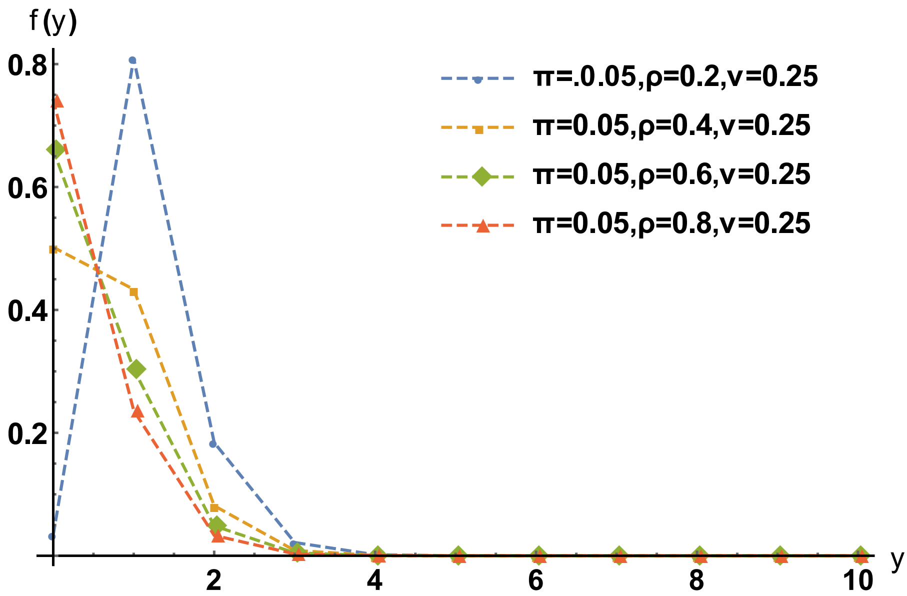

Figure 3 – Plots of probability mass functions of ZIAHPD for different values of ρ.

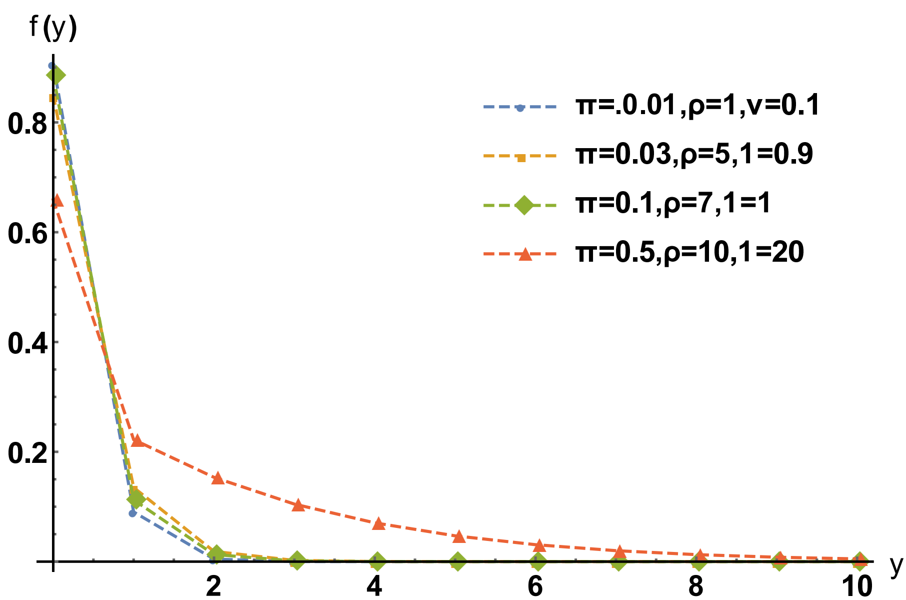

Figure 4 – Plots of probability mass functions of ZIAHPD for different values of π, ρ *and* ν.

where $\xi_1(a,b;c;x,y)$ is the Horn-Appel function which is a generalized hypergeometric function in two variables.

PROOF. By definition, the p.g.f of the ZIAHPD having p.m.f (5) is given by

$$\begin{aligned}
H(t) &= \sum_{w=0}^{\infty} g(w)t^w \qquad (7)\\
&= [\pi+(1-\pi)\phi(1;\rho;-\nu)]+\sum_{w=1}^{\infty}(1-\pi)t^w\frac{\nu^w}{(\rho)_w}\phi(1+w;\rho+w;-\nu)\\
&= \pi+(1-\pi)\phi(1;\rho;-\nu)+(1-\pi)\sum_{w=1}^{\infty}\sum_{r=0}^{\infty}\frac{(t\nu)^w}{(\rho)_w}\frac{(1+w)_r\,(-\nu)^r}{(\rho+w)_r\;r!}\\
&= \pi+(1-\pi)\phi(1;\rho;-\nu)+(1-\pi)\left(\sum_{w=0}^{\infty}\sum_{r=0}^{\infty}\frac{(t\nu)^w}{(\rho)_w}\frac{(1+w)_r\,(-\nu)^r}{(\rho+w)_r\;r!}\right) \qquad (8)\\
&- (1-\pi)\sum_{r=0}^{\infty}\frac{(1)_r\,(-\nu)^r}{(\rho)_r\;r!},
\end{aligned}$$

which implies the following in the light of (4).

$$H(t)=\pi+(1-\pi)\,\phi(1;\rho;-\nu)+(1-\pi)\,\xi_1(1,-;\rho;t\nu,-\nu)-(1-\pi)\,\phi(1;\rho;-\nu). \qquad (9)$$

On simplifying (9), gives (6). □

RESULT 2.2. *For $r \geq 1$, an expression for the factorial moments $\mu_{[r]}$ of the ZIAHPD with p.m.f (5) is the following.*

$$\mu_{[r]}=(1-\pi)\frac{\nu^r\,r!}{(\rho)_r} \qquad (10)$$

PROOF. The factorial moment generating function $F(t)$ of the ZIAHPD with p.g.f (6) is

$$\begin{aligned}
F(t) &= H(1+t)\\
&= \pi+(1-\pi)\xi_1(1+r,-;\rho+r;(t+1)\nu,-\nu). \qquad (11)
\end{aligned}$$

On differentiating (11) r times with respect to t and putting $t=1$, we get

$$F^{(r)}(1)=(1-\pi)\frac{\nu^r\;r!}{(\rho)_r}\xi_1(1+r,-;\rho+r;\nu,-\nu), \qquad (12)$$

which on simplication gives (10), since $\xi_1(1+r,-;\rho+r;\nu,-\nu)=1.$. □

By using Result 2.2, we obtain the following corollaries.

COROLLARY 2. *Mean and Variance of the ZIAHPD are*

$$Mean = (1-\pi)\frac{\nu}{\rho} = \vartheta, say$$

$$and$$

$$Variance = \vartheta\left[1+\nu\frac{(\pi(1+\rho)-(1-\rho))}{\rho(1+\rho)}\right].$$

COROLLARY 3. *The third and fourth moments of the ZIAHPD are, respectively*

$$\mu_3' = E(W^3) = \frac{(1-\pi)\nu}{\rho}\left(\frac{6\nu^2}{(\rho+1)(\rho+2)} + \frac{6\nu}{\rho+1} + 1\right) \tag{13}$$

and

$$\mu_4' = E(W^4) = \frac{(1-\pi)\nu}{\rho}\left(\frac{24\nu^3}{(\rho+1)(\rho+2)(\rho+3)} + \frac{36\nu^2}{(\rho+1)(\rho+2)} + \frac{14\nu}{\rho+1} + 1\right). \tag{14}$$

By using Corollary 2 and Corollary 3, we have computed measures of skewness and kurtosis with the help of Mathematica software and plotted the values in Figure 5, Figure 6, Figure 7 and Figure 8. From the Figures, it can be seen that the distribution enjoys positively and negatively skewed nature and both platykurtic and leptokurtic behaviour.

In the light of Corollary 2, we have the following important result, which depicts the nature of dispersion of the distribution.

RESULT 2.3. *The ZIAHPD becomes under-dispersed when* $\rho < \frac{1-\pi}{1+\pi}$, *over-dispersed when* $\rho > \frac{1-\pi}{1+\pi}$ *and equi-dispersed when* $\rho = \frac{1-\pi}{1+\pi}$.

Next we deal certain recurrence formulae for probabilities, raw moments and factorial moments of the ZIAHPD.

RESULT 2.4. *The probabilities* $g_w(\pi,\rho^*,\nu) = g(w)$ *of the ZIAHPD satisfies the following recursion formula, in which* $\rho^* = (1,\rho)$

$$g_1(\pi,\rho^*,\nu) = \nu\,\rho^{-1}\{g_0(\pi,\rho^*+1,\nu)-\pi\},\ for\ w=0 \tag{15}$$

and

$$(w+1)g_{w+1}(\pi,\rho^*,\nu) = \nu\,\rho^{-1}\{g_w(\pi,\rho^*+1,\nu)\},\ for\ w>0. \tag{16}$$

PROOF. The p.g.f of the ZIAHPD can be written as

$$H(t) = \pi + (1-\pi)\xi_1(1,-;\rho;t\nu,-\nu)$$

$$= \sum_{w=0}^{\infty} t^w g_w(\pi,\rho^*,\nu). \tag{17}$$

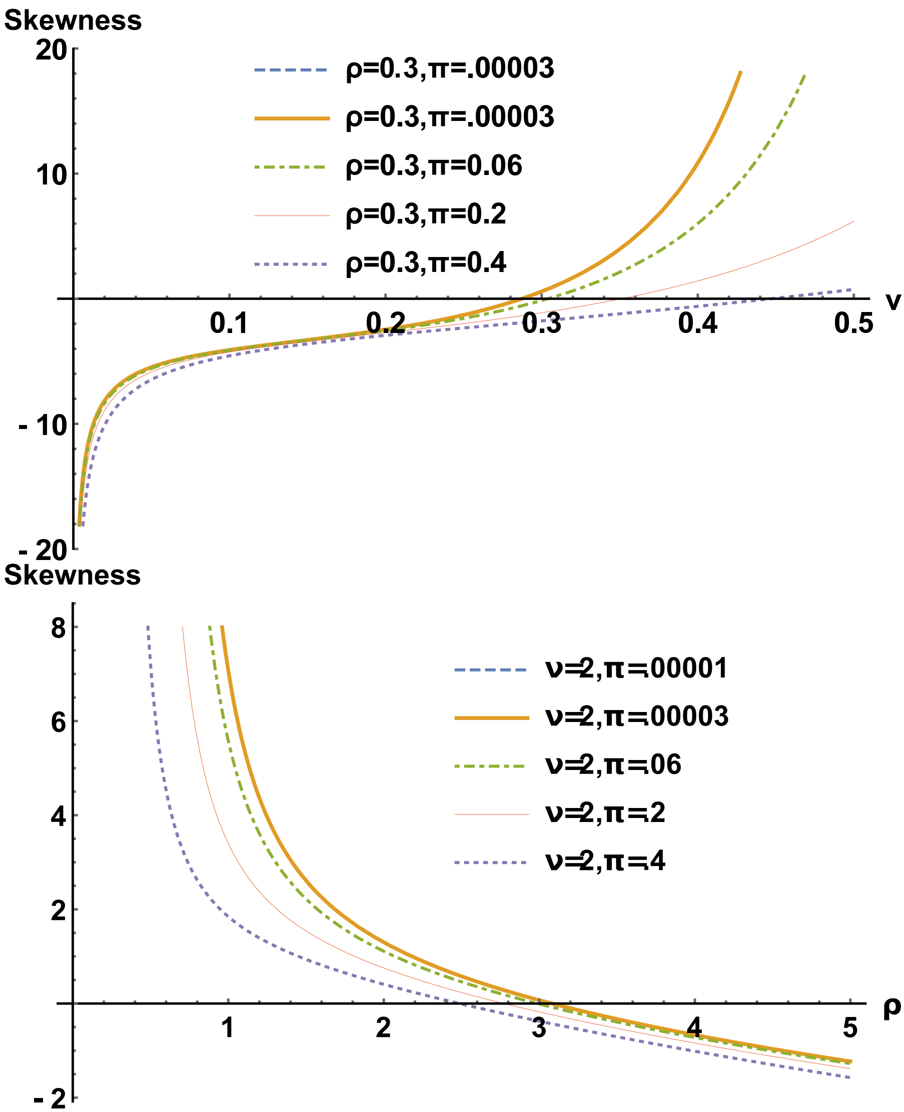

Figure 5 – Plots of skewness of ZIAHPD for particular values of its parameters.

On differentiating (17) with respect to t, we have the following.

$$\sum_{w=0}^{\infty}(w+1)g_{w+1}(\pi,\rho^*,\nu)t^w=(1-\pi)\frac{\nu}{\rho}\xi_1(2,-;\rho+1;t\nu,-\nu). \tag{18}$$

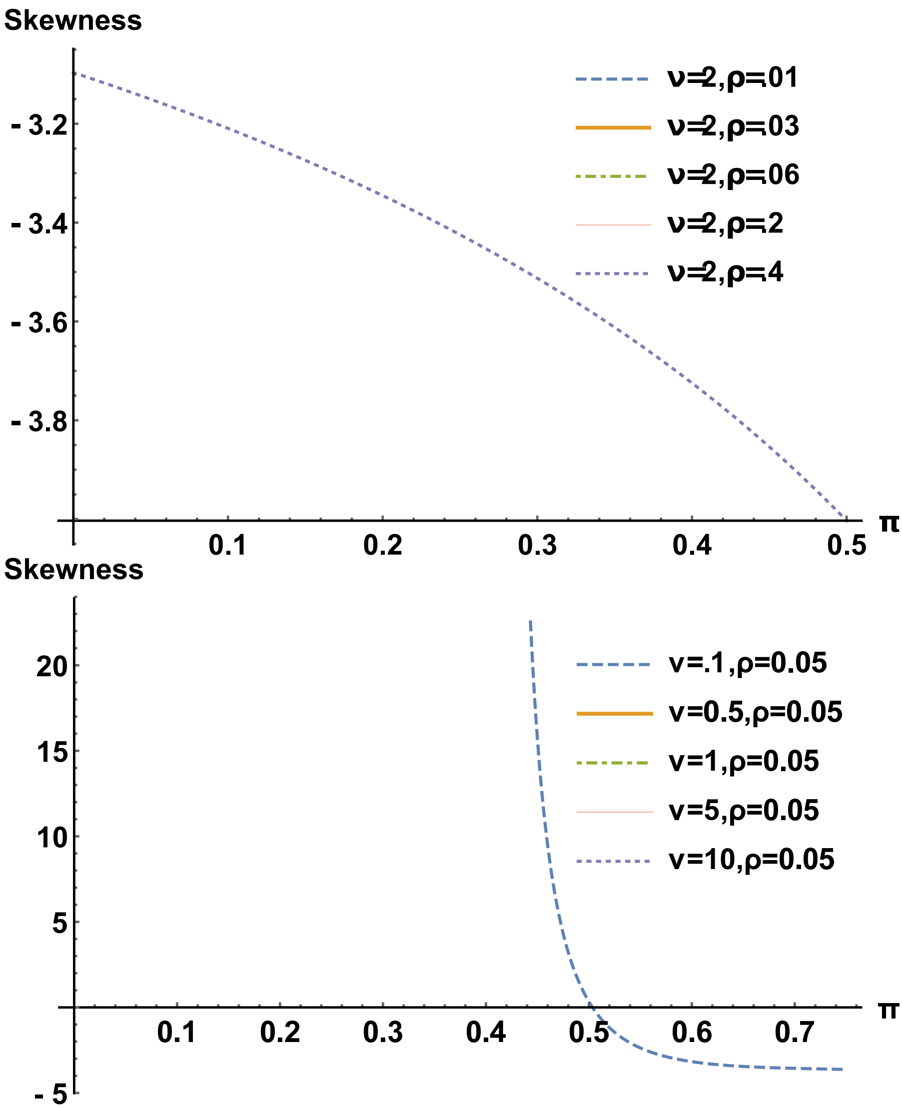

Figure 6 – Plots of skewness of ZIAHPD for particular values of its parameters.

Also, from (17) we have

$$(1-\pi)\xi_1(2,-;\rho+1;t\nu,-\nu)=\sum_{w=0}^{\infty} t^w g_w(\pi,\rho^*+1,\nu)-\pi. \tag{19}$$

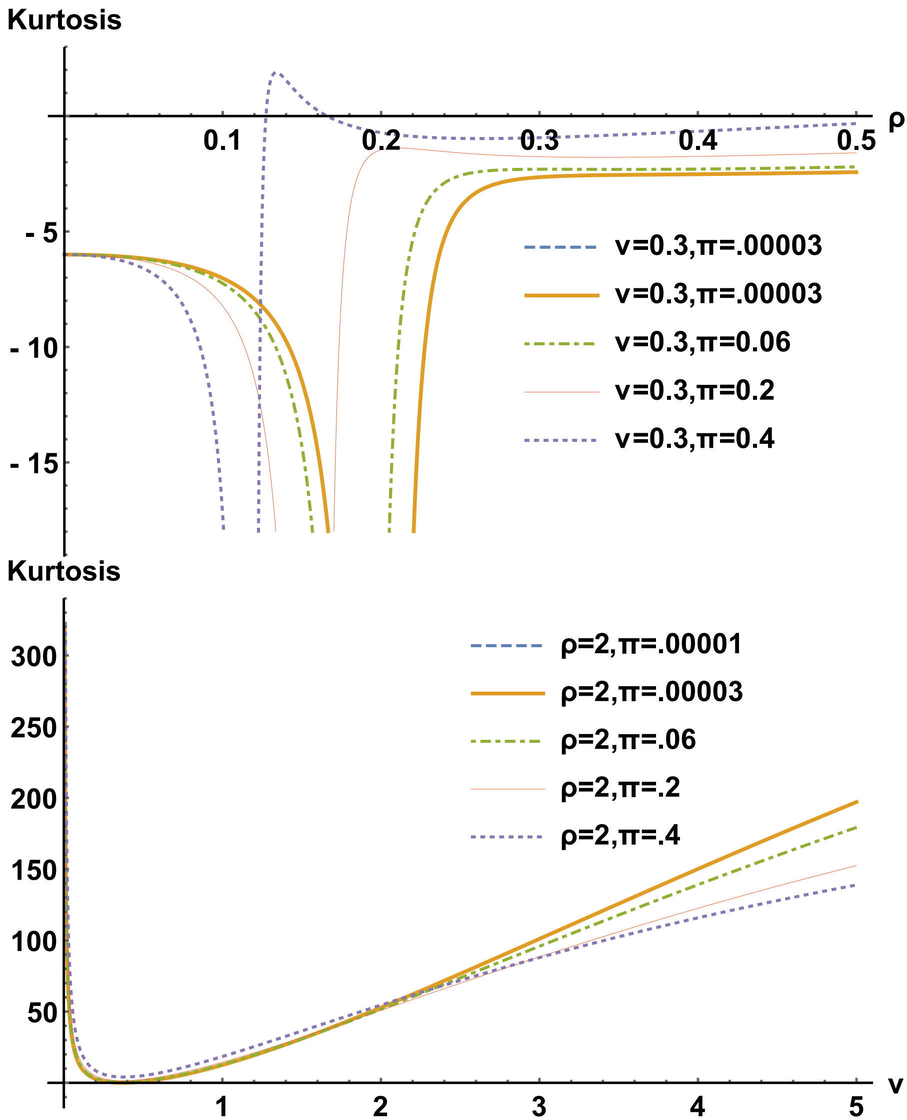

Figure 7 – Plots of kurtosis of ZIAHPD for particular values of its parameters.

Combining relation (18) and (19) together lead the following.

$$\sum_{w=0}^{\infty}(w+1)g_{w+1}(\pi,\rho^*,\nu)t^w = \frac{\nu}{\rho}\left[\sum_{w=0}^{\infty}t^w g_w(\pi,\rho^*+1,\nu)-\pi\right]. \tag{20}$$

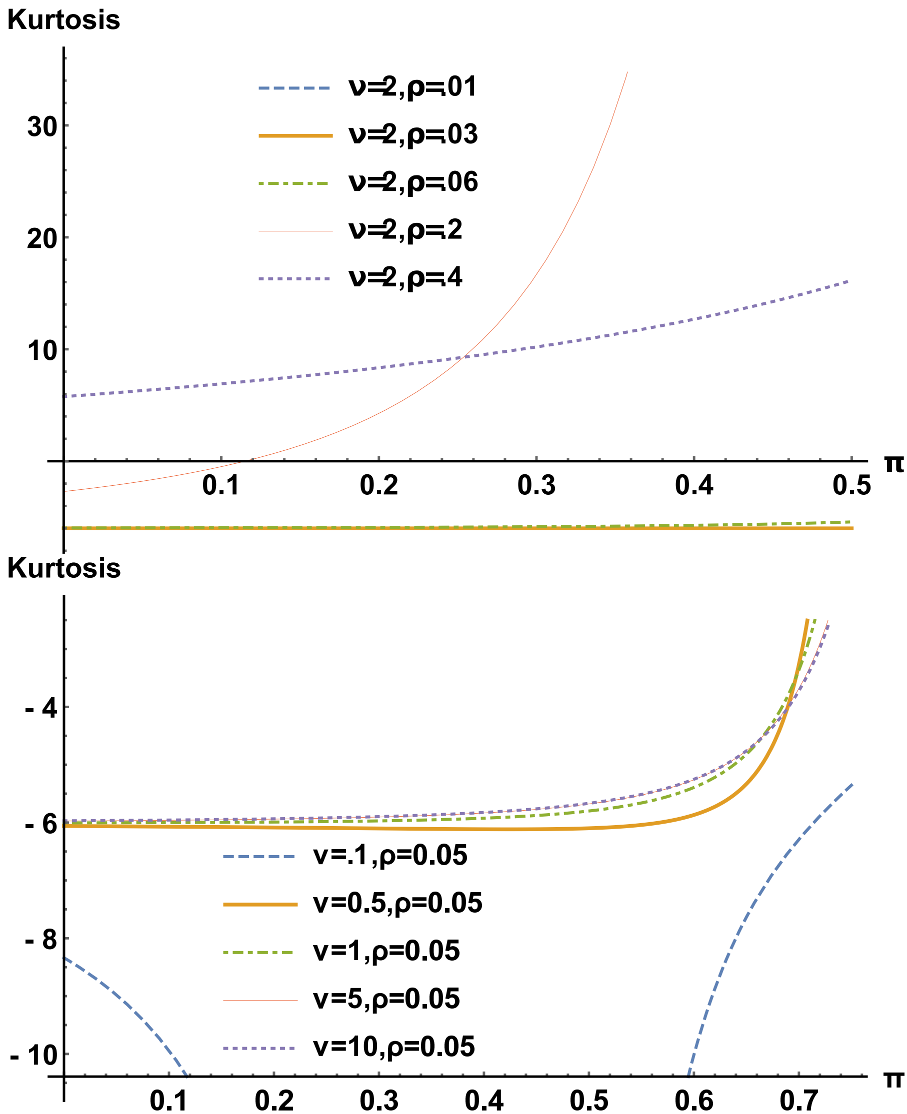

Figure 8 – Plots of kurtosis of ZIAHPD for particular values of its parameters.

Now, equating the coefficients of t^0 on both sides of (20) we get (15), and on equating the coefficients of t^w on both sides of (20), we get (16). □

RESULT 2.5. *For $r \geq 0$, the raw moments $\mu_r(\rho^*) = \mu_r$ of the ZIAHPD satisfies the* *following recursion formula*

$$\mu_{r+1}(\rho^*) = \nu\,\rho^{-1}\left\{\sum_{k=0}^{r}\binom{r}{k}\mu_{r-k}(\rho^*+1) - \pi\right\} \tag{21}$$

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

$$\begin{aligned}\psi(t) &= \sum_{r=0}^{\infty}\mu_r(\rho^*)\frac{(it)^r}{r!}\\ &= H(e^{it})\\ &= \pi + (1-\pi)\xi_1(1,-;\rho;\nu e^{it},-\nu).\end{aligned} \tag{22}$$

Differentiating (22) with respect to t, we get

$$(1-\pi)\frac{\nu e^{it}}{\rho}\xi_1(2,-;\rho+1;\nu e^{it},-\nu) = \sum_{r=1}^{\infty}\mu_r(\rho^*)\frac{(it)^{r-1}}{(r-1)!}. \tag{23}$$

Using (23), we obtain

$$(1-\pi)\xi_1(2,-;\rho+1;\nu e^{it},-\nu) = \sum_{r=0}^{\infty}\mu_r(\rho^*+1)\frac{(it)^r}{r!} - \pi. \tag{24}$$

Combining (23) and (24), we obtain the following.

$$\begin{aligned}\sum_{r=0}^{\infty}\mu_{r+1}(\rho^*)\frac{(it)^r}{r!} &= \nu\rho^{-1}e^{it}\left[\sum_{r=0}^{\infty}\mu_r(\rho^*+1)\frac{(it)^r}{r!} - \pi\right]\\ &= \nu\rho^{-1}\left[\sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\frac{(it)^k}{k!}\mu_r(\rho^*+1)\frac{(it)^r}{r!} - \pi\sum_{r=0}^{\infty}\frac{(it)^r}{r!}\right]\\ &= \nu\rho^{-1}\left[\sum_{r=0}^{\infty}\sum_{k=0}^{r}\frac{(it)^k}{k!}\mu_{r-k}(\rho^*+1)\frac{(it)^{r-k}}{(r-k)!} - \pi\sum_{r=0}^{\infty}\frac{(it)^r}{r!}\right]\\ &= \nu\rho^{-1}\left[\sum_{r=0}^{\infty}\sum_{k=0}^{r}\frac{(it)^r}{k!(r-k)!}\mu_{r-k}(\rho^*+1) - \pi\sum_{r=0}^{\infty}\frac{(it)^r}{r!}\right].\end{aligned} \tag{25}$$

Now, equating the coefficients of $(r!)^{-1}(it)^r$ on both sides of (25), we get (21). □

RESULT 2.6. *For $r \geq 1$, the factorial moments $\mu_{[r]}(\rho^*) = \mu_{[r]}$ of the ZIAHPD satisfies the following recursion formula, in which $\mu_{[1]}(\rho^*) = \vartheta$*

$$\mu_{[r+1]}(\rho^*) = \nu\rho^{-1}\left\{\mu_{[r]}(\rho^*+1) - \pi\right\} \tag{26}$$

PROOF. By using (6) the factorial moment generating function $F(t)$ of the ZIAHPD can be written as

$$\begin{aligned} F(t) &= G(1+t) \\ &= \pi + (1-\pi)\xi_1(1,-;\rho;\nu(1+t),-\nu) \\ &= \sum_{r=0}^{\infty} \mu_{[r]}(\rho^*)\frac{t^r}{r!}. \end{aligned} \tag{27}$$

On differentiating (27) with respect to t, we have

$$(1-\pi)\nu\rho^{-1}\,\xi_1(2,-;\rho+1;(1+t)\nu,-\nu) = \sum_{r=1}^{\infty} \mu_{[r]}(\rho^*)\frac{t^{r-1}}{(r-1)!}. \tag{28}$$

By using (27), we obtain the following from (28).

$$\sum_{r=0}^{\infty} \mu_{[r+1]}(\rho^*)\frac{t^r}{r!} = \nu\rho^{-1}\left[\sum_{r=0}^{\infty} \mu_r(\rho^*+1)\frac{t^r}{r!} - \pi\right]. \tag{29}$$

Finally, on equating the coefficients of $(r!)^{-1}t^r$ on both sides of (29), we get (26). □

Model Identifiability: Let W be a discrete random variable having p.m.f $p(w) = P(W = w)$ of the form $p(w) = \pi_1 p_1(w) + \pi_2 p_2(w) + \ldots + \pi_g p_g(w)$, where for each $j = 1,2,\ldots,g$; $\pi_j > 0$ such that $\sum_{j=1}^{g}\pi_j = 1$, $p_j(w) \geq 0$ and $\sum_{j=1}^{g} p_j(w) = 1$. Then, we say that W has a mixture distribution and $p(w)$ is a finite mixture of distributions. The parameters $\pi_1, \pi_2, \ldots, \pi_g$ are known as mixing weights which, $p_1, p_2, \ldots, p_g$ are the components of the mixture. We denote Ψ as the collection of all distinct parameters occuring in the components and Θ as the complete collection of all distinct parameters occuring in the mixture model.

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

$$f(w_j;\psi) = f(w_j;\psi^*) \tag{30}$$

if and only if

$$\psi = \psi^* \tag{31}$$

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

parametric family.
Let

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

and

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

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 $f(w_j;\psi)\equiv f(w_j;\psi^*)$ if and only if $g=g*$ and we can permute the component labels so that $\pi_i=\pi_i^*$ and $f_i(w_j;\beta_i)=f_i(w_j;\beta_i^*)$ 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 condition that is identifiable is that the distribution function of convex combination of mixture densities is linearly independent over the field of real numbers.*

DEFINITION 4. *A random variable W is said to have g component mixture model of zero-inflated alternative hyper-Poisson distribution if it has the following p.m.f $p(w)$, in which $0\leq\pi_j\leq 1$, for $j=1,2,...,k$, $\sum_{j=1}^{k}\pi_j=1$ and $w=0,1,2,...$*

$$f(w)=\sum_{j=1}^{k}\pi_j q_j(w) \tag{32}$$

where

$$q_j(w)=\frac{\nu_j^w}{\left(\rho_j\right)_w}\ \phi(1+w;\rho_j+w;-\nu_j) \tag{33}$$

with $\nu_j>0$ and $\rho_j>0$ for each $j=1,2,\ldots,k$.

RESULT 2.7. *The identifiability condition for ZIAHPD with p.m.f $g(w)$ given in (5) is $\nu_r\neq\nu_s$ and $\rho_r\neq\rho_s$ for positive integers r and s assuming values from $1,2,\ldots,k$ with $r\neq s$ and $w=0,1,2,...$*

PROOF. Consider the equation

$$a_1G_1(w)+a_2G_2(w)=0 \tag{34}$$

where a_1 and a_2 be any two arbitrary real numbers, $G_1(w) = \sum_{j=1}^{w} g(j)$ and $G_2(w) = \sum_{j=1}^{w} h(j)$, for $w = 0, 1, \ldots$ in which h_j obtained from g_j by replacing ν_j by σ_j and ρ_j by τ_j. Assume that for each $j = 1, 2$, $\nu_j \neq \sigma_j$ and $\rho_j \neq \tau_j$. Thus,

$$G_1(w) = \pi + (1-\pi)\phi(1+j;\rho+j;-\nu) + \sum_{j=1}^{w}(1-\pi)\frac{\nu_j^j}{\left(\rho_j\right)_j}\phi(1+j;\rho+j;-\nu) \quad (35)$$

$$G_2(w) = \pi + (1-\pi)\phi(1+j;\tau+j;-\sigma) + \sum_{j=1}^{w}(1-\pi)\frac{\sigma_j^j}{\left(\tau_j\right)_j}\phi(1+j;\tau+j;-\sigma) \quad (36)$$

Combining (32), (33) and (34), we have

$$a_1\sum_{j=0}^{w}\frac{\nu_1^j}{(\rho_1)_j}\phi(1+j;\rho_1+j;-\nu_1) = a_2\sum_{j=0}^{w}\frac{\sigma_1^j}{(\tau_1)_j}\phi(1+j;\tau_1+j;-\sigma_1) \quad (37)$$

$$a_1\sum_{j=0}^{w}\frac{\nu_2^j}{(\rho_2)_j}\phi(1+j;\rho_2+j;-\nu_2) = a_2\sum_{j=0}^{w}\frac{\sigma_2^j}{(\tau_2)_j}\phi(1+j;\tau_2+j;-\sigma_2) \quad (38)$$

Eliminate a_1 using (32) and (33), we have $a_2 = 0$. From (32), we obtain $a_1 = 0$ shows that G_1 and G_2 are linearly independent. □

3. Maximum Likelihood Estimation

Here we consider the estimation by the method of maximum likelihood for estimating the parameters π, ρ and ν of the ZIAHPD. For any $w = 0, 1, 2, \ldots$, let $A(w)$ be the observed frequency of w events and let p be the highest value of w observed. Then the likelihood function of the sample is given by

$$L(\theta; w) = \prod_{w=0}^{p}[g(w)]^{A(w)},$$

where $g(w)$ is the p.m.f of the ZIAHPD given in (5).
Now $L(\theta; w)$ can be written as

$$L(\theta; w) = (g_0)^s \prod_{w=1}^{p}(g_1(w))^{A(w)},$$

where $s = A(0)$, g_0 is the p.m.f of the ZIAHPD when $w = 0$ and $g_1(.)$ is the p.m.f of the distribution when $w = 1, 2, \ldots$.

Then the log-likelihood function can be written as

$$\begin{aligned}\ln L(\theta;w) &= s\ln[\pi+(1-\pi)\phi(1;\rho;-\nu)] \\ &\quad +\sum_{w=1}^{p}A(w)\ln\left\{(1-\pi)\frac{\nu^{w}}{(\rho)_{w}}\phi(1+w;\rho+w;-\nu)\right\}.\end{aligned} \tag{39}$$

Assume that $\hat{\pi}$, $\hat{\rho}$ and $\hat{\nu}$ be the maximum likelihood estimators of the parameters π, ρ and ν of the ZIAHPD. Now, on differentiating the log-likelihood function (39) with respect to π, ρ and ν and equating to zero, we obtain the following likelihood equations:

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

implies

$$s\frac{1-\phi(1;\rho;-\nu)}{\pi+(1-\pi)\phi(1;\rho;-\nu)}-(1-\pi)\sum_{w=1}^{p}A(w)=0, \tag{40}$$

$$\frac{\partial \log L}{\partial \rho}=0$$

$$\frac{s(1-\pi)}{\pi+(1-\pi)\phi(1;\rho;-\nu)}\sum_{k=0}^{\infty}\frac{(-\nu)^{k}\Gamma(\rho)}{\Gamma(\rho+k)}[\psi(\rho)-\psi(\rho+k)]+\sum_{w=1}^{p}A(w) \tag{41}$$

$$\left\{\sum_{k=1}^{\infty}\frac{(1+w)_{k}}{k!}(-\nu)^{k}\frac{\Gamma(\rho+w)}{\Gamma(\rho+w+k)}[\psi(\rho+w)-\psi(\rho+w+k)]-\psi(\rho)\right\}=0$$

and

$$\frac{\partial \log L}{\partial \nu}=0$$

implies

$$\frac{(-s)(1-\pi)\phi(2;\rho+1;-\nu)}{\rho(\pi+(1-\pi)\phi(1;\rho;-\nu))}+\sum_{w=0}^{p}A(w)\left[\frac{w}{\nu}-\frac{\phi(2+w;\rho+w+1;-\nu)}{(\rho+w)\phi(1+w;\rho+w;-\nu)}\right]=0 \tag{42}$$

in which $\psi(\rho)=\frac{\partial}{\partial\rho}log\Gamma(\rho)$.

On solving the likelihood equations (40), (41) and (42) with the help of Mathematical software, say Mathematica, one can obtain the maximum likelihood estimators of the parameters of the proposed distribution.

4. Testing

In order to test the significance of the inflation parameter π of the ZIAHPD, we adopt the following generalized likelihood ratio test (GLRT) procedure. Here the null hypothesis is

$$H_0 : \pi = 0 \; Vs \; H_1 : \pi \neq 0.$$

The test statistic suggested in the case of GLRT is given by

$$-2 \ln \psi = 2(\iota_1 - \iota_2), \tag{43}$$

where, $\iota_1 = \ln L(\hat{\theta}; w)$, where $\hat{\theta}$ is the maximum likelihood estimator for $\theta = (\pi, \rho, \nu)$ with no restrictions, and $l n L(\hat{\theta}^*; w)$, in which $\hat{\theta}^*$ is the maximum likelihood estimator for θ under the null hypothesis H_0. The test statistic defined in (43) is asymptotically distributed as χ^2 with one degree of freedom.

5. Data Illustration

In this Section, we consider certain real life data sets for illustrating the methods discussed in Sections 3 and 4. The first data set is on the distribution of data of Corn borers per hill (Rodriguez-Avi *et al.*, 2007) and the second data is on the distribution of currency and banking crisis (Giles *et al.*, 2010). We have fitted the ZIAHPD to all these data sets and considered the fitting of the models - HPD, AHPD, ZIPD, ZINBD, ZIGPD, ZIHPD for comparison. For comparing the models we computed the values of χ^2, AIC, BIC and AICc. The numerical results obtained are presented in Tables 1 and 2. Based on the computed values of χ^2, AIC, BIC and AICc as presented in Tables 1 and 2, one can observe that the ZIAHPD gives a better fit to all these data sets considered.

We have also calulated the values of the test statistic given in (43) and are included in Table 3. The critical value of the test having 5% level of significance and degree of freedom one is 3.84, so that the null hypothesis is rejected in all the cases. Thus, we conclude that the additional parameter π in the model is significant.

We have also plotted the observed frequency curves of the data sets along with the fitted densities corresponding to the HPD, AHPD, ZIPD, ZINBD, ZIGPD, ZIHPD and the ZIAHPD. From Tables: 1, 2 and Figures: 9, 10, it can be seen that all these models are not given best fit to the data sets while the ZIAHPD only gives best fit based on the P-value and Chi-square value. And also the values of information measures like AIC, BIC and AICc support the factor that the ZIAHPD can be considered as a suitable model compared to the other models discussed in the paper.

6. Simulation

Since it is difficult to compare the theoretical performances of estimators of the parameters of the ZIAHPD obtained by the method of maximum likelihood, we have

TABLE 1

Distribution of data of Corn borers per hill (Rodriguez-Avi et al., 2007) and the expected frequencies computed using hyper-Poisson, alternative hyper-Poisson, zero-inflated Poisson, zero-inflated negative binomial, zero-inflated generalized Poisson, zero-inflated hyper-Poisson and ZIAHPD.

Count	Observed frequency	HPD	AHPD	ZIPD	ZINBD	ZIGPD	ZIHPD	ZIAHPD
0	47	49.778	29.4	21.3	71.80	70.6	41.59	45.0
1	23	48.848	57.71	14.4	28.8	15.09	48.15	23.5
2	27	16.59	24.722	20.5	12.40	26.44	21.47	29.5
3	9	3.31	5.99	24.8	5.24	7.847	6.29	12.9
4	7	0.474	1.042	22.3	1.26	0.0232	1.25	5.2
5	7	1.0	1.1357	16.7	0.5	4.64×10^{-8}	0.025	3.9
Total	120	120	120	120	120	120	120	120
df		1	2	3	1	1	1	2
Estimates		ρ=0.5 ν=0.5	ρ=0.55 ν=0.59	ρ=3.5 π=0.1	ρ=4.22 ν=0.83 π=0.26	ρ=3.15 $\nu=3.7\times10^{-7}$ π=0.37	ρ=0.5 ν=0.7 π=0.0.06	ρ=0.22 ν=0.92 π=0.69
χ^2-value		155.99	97.48	64.40	63.49	41.13	44.45	4.57
p-value		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.102
AIC		721.12	612.15	499.84	603.08	1195.08	688.23	489.58
BIC		720.7	612.05	506.46	602.45	1194.92	687.61	488.96
AICc		725.12	616.46	503.84	612.08	1207.08	700.23	501.58

TABLE 2

Distributions of currency and banking crises (Giles et al., 2010) and the expected frequencies computed using hyper-Poisson, alternative hyper-Poisson, zero-inflated Poisson, zero-inflated negative binomial, zero-inflated generalized Poisson, zero-inflated hyper-Poisson and ZIAHPD.

Count	Observed frequency	HPD	AHPD	ZIPD	ZINBD	ZIGPD	ZIHPD	ZIAHPD
0	45	94.11	50.7	68.88	98.22	80.94	54.57	49.33
1	44	53.614	77.02	60.49	31.8	39.36	58.932	39.9
2	19	14.10	30.66	27.28	13.66	30.88	34.38	23.2
3	17	3.43	7.184	8.17	9.19	15.778	11.62	20.66
4	19	1.711	1.19	1.834	7.86	0.0418	6.85	15.2
5	13	0.0318	0.155	0.33	5.34	0.0000102	0.54	10.43
6	6	0.0026	0.0585	0.014	0.63	4.02×10^{-11}	0.0866	4.2
7	4	0.000195	0.03248	0.0022	0.36	1.7×10^{-18}	0.0216	4.08
Total	167	167	167	167	167	167	167	167
df		3	1	1	1	1	1	1
Estimates		ρ=0.92 ν=0.49	ρ=0.59 ν=0.59	ρ=0.9 π=0.01	ρ=0.7 ν=0.7 π=0.1	ρ=3.67 ν=0.0000002 π=0.25	ρ=0.63 ν=0.89 π=0.06	ρ=0.20 ν=0.89 π=0.67
χ^2-value		588.81	313.68	243.92	109.22	37.17	32.47	3.05
p-value		0.00010	0.00010	0.00010	0.00010	0.00010	0.00010	0.25
AIC		1605.6	1261.3	1179.94	1083.9	3205.6	1106.96	936.8
BIC		1605.8	1261.4	1180.1	1084.14	3211.6	1107.22	937.1
AICc		1608.06	1263.7	1182.3	1091.4	3205.8	1112.98	942.8

TABLE 3
Calculated value of the test statistic in case of the generalised likelihood ratio test.

	$\ln L(\hat{\theta}^*; w)$	$\ln L(\hat{\theta}; w)$	Test statistic
Data set 1	-304.23	-241.79	124.88
Data set 2	-628.67	-465.43	326.46

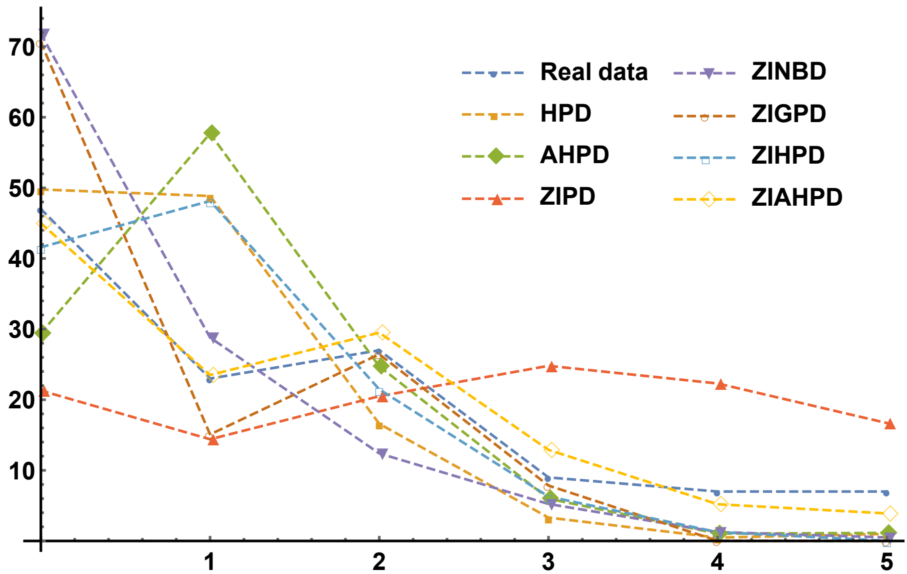

Figure 9 – Frequency curves corresponding to data set 1

attempted a brief simulation study. We have computed the absolute bias and standard errors in case of each simulated samples. The simulation results are summarised in Table 4 corresponding to the sample of sizes 100, 200 and 500 for the following two parameter sets.

1. $\pi = 0.80$, $\rho = 1.2$, $\nu = 0.90$ (over-dispersion)

2. $\pi = 0.60$, $\rho = 0.20$, $\nu = 0.90$ (under-dispersion)

Probability plots corresponding to the simulated data sets in case of parameter set 1 and parameter set 2 are as given in Figures: 11, 12.

From Table 4 and Figure 11 and 12, it can be observed that as the sample size increases, both absolute bias and standard errors of the parameter sets are in decreasing order among all the cases.

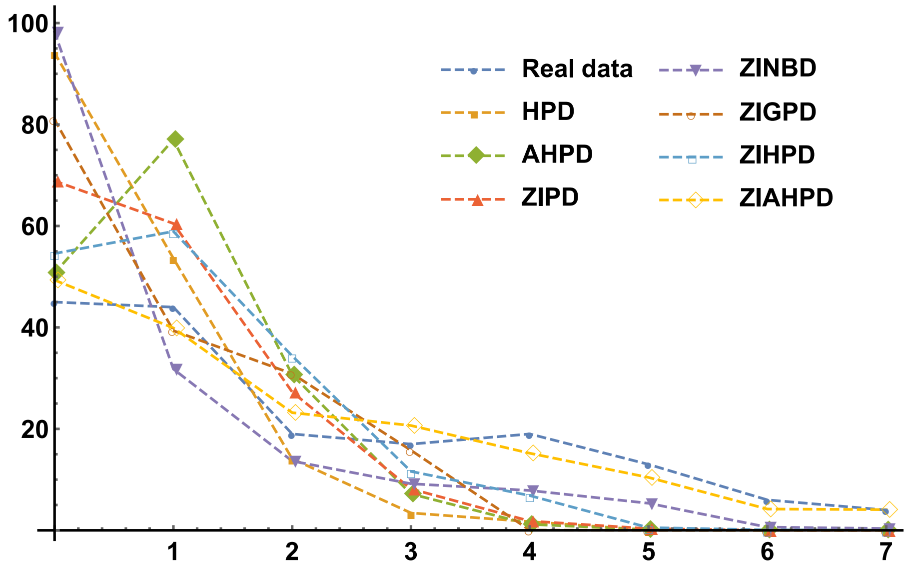

Figure 10 – Frequency curves corresponding to data set 2

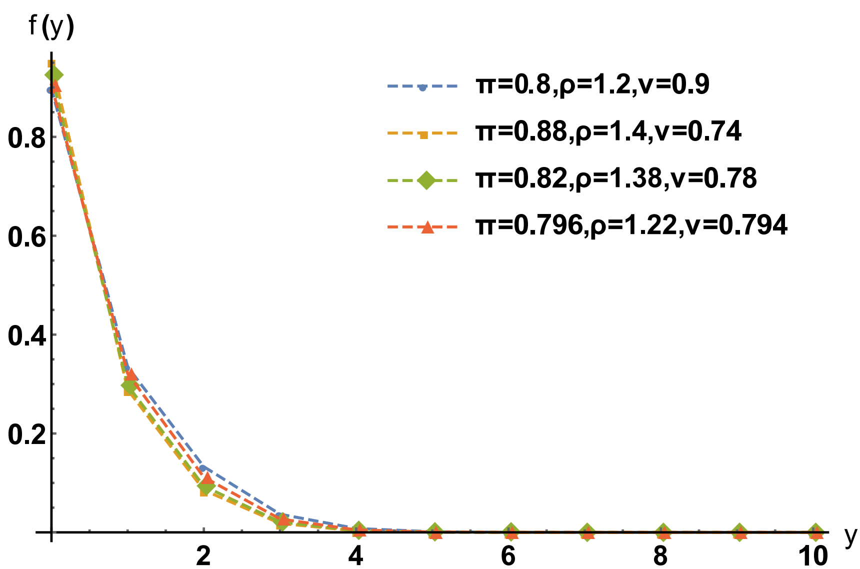

Figure 11 – Probability plots corresponding to parameter set 1.

TABLE 4
Absolute bias and standard errors (given in paranthesis) of each estimators of the ZIAHPD obtained by the method of maximum likelihood in case of parameter sets 1 and 2.

Parameter set	Sample size	MLE $\hat{\pi}$	MLE $\hat{\rho}$	MLE $\hat{\nu}$
1	$n = 100$	0.08 (0.024)	0.22 (0.08)	0.16 (0.05)
	$n = 200$	0.02 (0.017)	0.18 (0.07)	0.12 (0.04)
	$n = 500$	0.004 (0.009)	0.02 (0.033)	0.11 (0.019)
2	$n = 100$	0.12 (0.14)	0.22 (0.07)	0.18 (0.17)
	$n = 200$	0.044 (0.037)	0.12 (0.032)	0.096 (0.075)
	$n = 500$	0.030 (0.025)	0.04 (0.0046)	0.056 (0.036)

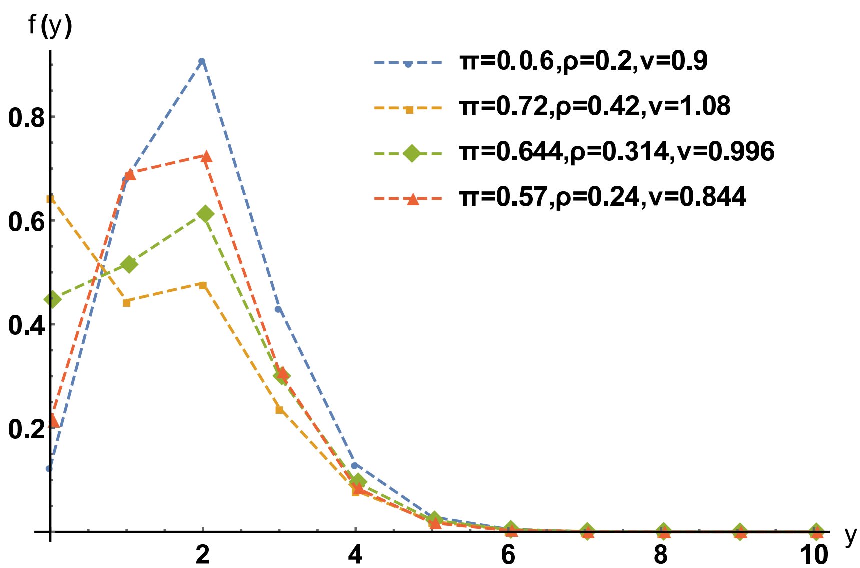

Figure 12 – Probability plots corresponding to parameter set 2.

7. Concluding Remarks

This paper introduces a zero-inflated version of the alternative hyper-Poisson distribution namely 'the zero-inflated alternative hyper-Poisson distribution (ZIAHPD)'. Important statistical properties of the distribution such as moments, generating functions, recursion formulae etc have been studied. The estimation of the parameters of the distribution have been attempted using maximum likelihood estimation. In addition, the generalized likelihood ratio test procedure is constructed for testing the significance of the inflation parameter. Two real life data applications are considered for illustrating the usefulness of the proposed model compared to the existing models such as HPD, AHPD, ZIPD, ZINBD, ZIGPD and ZIHPD. It can be seen that all these models are not given best fit to the data sets while the ZIAHPD only gives best fit based on the Chi-square value and P-value. The values of information measures like AIC, BIC and AICc also reveals the factor that the ZIAHPD can be considered as a suitable model compared to the other models discussed in the paper. A brief simulation study has been carried out for assessing the efficiency of the estimation procedure discussed in the paper. Several characteristic properties as well as inferential aspects of the new model and related regression models are yet to study, those results we wish to include in another publication.

Acknowledgements

The authors are highly thankful to the Editor in Chief 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

G. D. Barriga, F. Louzada (2014). *The zero-inflated Conway–Maxwell–Poisson distribution: Bayesian inference, regression modeling and influence diagnostic*. Statistical Methodology, , no. 21, pp. 23–34.

K. L. Bassil, D. C. Cole, R. Moineddin, W. Lou, A. M. Craig, B. Schwartz, E. Rea (2010). *The relationship between temperature and ambulance response calls for heat-related illness in Toronto, Ontaria*. Journal of Epidemiology and Community Health, , no. 65, pp. 829–831.

D. Bohning (1998). *Zero-inflated Poisson models and C.A.MAN: a tutorial collection*. Biometrical Journal, , no. 40, pp. 833–843.

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

D. E. Giles, *et al.* (2010). *Hermite regression analysis of multi-modal count data*. Economics Bulletin, 30, no. 4, pp. 2936–2945.

A. Goraski (1977). *Distribution z-Poisson*. Institut de statistique de l'Universite de Paris, 12, pp. 45–53.

V. Hasselblad (1969). *Estimation of finite mixtures of distributions from the exponential family*. Journal of the American Statistical Association, 64, pp. 1495–1471.

C. KEMP (2002). *q-analogues of the hyper-Poisson distribution*. Journal of Statistical Planning and Inference, 101, no. 1-2, pp. 179–183.

C. S. KUMAR, B. U. NAIR (2012). *An alternative hyper-Poisson distribution*. Statistica, 72, no. 3, pp. 357–369.

C. S. KUMAR, R. RAMACHANDRAN (2019). *On zero-inflated hyper-Poisson distribution and its applications*. Communications in Statistics-Simulation and Computation, (Ahead of Print), pp. 1–14.

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

B. G. LEROUX, M. L. PUTERMAN (1992). *Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models*. Biometrics, 48, pp. 545–558.

D. C. MARTIN, S. KATTI (1965). *Fitting of certain contagious distributions to some available data by the maximum likelihood method*. Biometrics, 21, pp. 34–48.

A. M'KENDRICK (1925). *Applications of mathematics to medical problems*. Proceedings of the Edinburgh Mathematical Society, 44, pp. 98–130.

J. RODRIGUEZ-AVI, A. CONDE-SANCHEZ, A. SAEZ-CASTILLO, M. OLMO-JIMENEZ (2007). *Gaussian hypergeometric probability distributions for fitting discrete data*. Communications in Statistics—Theory and Methods, 36, no. 3, pp. 453–463.

A. SAEZ-CASTILLO, A. CONDE-SANCHEZ (2013). *A hyper-Poisson regression model for overdispersed and underdispersed count data*. Computational Statistics & Data Analysis, 61, pp. 148–157.

S. Z. SIM, R. C. GUPTA, S. H. ONG (2018). *Zero-inflated Conway-Maxwell Poisson distribution to analyze discrete data*. The international journal of biostatistics, 14, no. 1.

S. SINGH (1962). *Note on inflated Poisson-distribution*. In *Journal of the Indian Statistical Association*. vol. 33, pp. 140–144.

D. M. TITTERINGTON, A. SMITH, U. MAKOV (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

J. VAN DEN BROEK (1995). *A score test for zero inflation in a Poisson distribution*. Biometrics, 51, pp. 738–743.

SUMMARY

Here we develop a zero-inflated version of the alternative hyper-Poisson distribution and discuss its important statistical properties such as probability generating function, expressions for mean, variance, factorial moments, skewness, kurtosis, recursion formula for probabilities, raw moments and factorial moments. Then the maximum likelihood estimation of the parameters of the zero-inflated alternative hyper-Poisson distribution is discussed and certain test procedures are constructed for testing the significance of the inflation parameter. All the procedures are illustrated with the help of certain real life data sets. Moreover, a brief simulation study is carried out for assessing the performances of the maximum likelihood estimators of the parameters of the proposed distribution.

Keywords: Confluent hypergeometric function; Count data models; Generalized likelihood ratio test; Horn-Appel function; Maximum likelihood estimation; Model selection; Simulation.