Vol.10.Issue.4.2022 (Oct-Dec)
©KY PUBLICATIONS

http://www.bomsr.com

Email:editorbomsr@gmail.com

RESEARCH ARTICLE

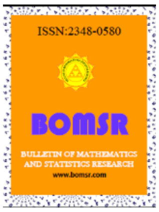
BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

SOME PROPERTIES OF ZERO-INFLATED POISSON MODEL

Dr. Chakrala Sreelatha
Assistant Professor, Department of Statistics, Rajendra University, Bolangir-767002
Email: srilathastats@gmail.com
DOI:10.33329/bomsr.10.4.61



ABSTRACT

In many applications, count outcomes are fairly common and often these count data have a large number of zeros. Zero-inflated models are useful for analyzing such data. The number of zeros is observed to be significantly higher than the predicted number by the standard assumed model. Such models are called zero-inflated models. This model shows the two parameter exponential family. One parameter is zero-inflation and the other parameter is Poisson model. In this paper we discuss the structural properties namely, mean, variance, P.G.F, M.G.F, C.F, Moments, F.M.G.F, moments recurrence relations of zero-inflated Poisson model were obtained. And also discuss the maximum likelihood estimation and the method of moment estimation of parameters in a zero-inflated Poisson series model. The maximum likelihood and moment estimators are asymptotically compared. Here we use the application of these distributions for data modeling of road traffic accidents and that these distributions have more appropriate approximation then corresponding distributions by using chi-square goodness of fit.

Keywords: Zero-inflated model, maximum likelihood estimation, method of moments, Simulation study.

1. Introduction

Zero-inflated models are the models obtained from existing models by appropriately mixing existing model with a singular distribution at zero. Zero-inflation indicates that a data set contains an excessive number of zeros. The word inflation is used to emphasize that the probability mass at the point zero exceeds than the one allowed under a standard parametric family of discrete distribution.

Inflation may occur at any of the support point. Neyman (1939) and Feller (1945) first introduced the concept of zero-inflation to address the problem of excess zeros. Since then, there have been extensive studies related to the development of Zero-Inflated Poisson (ZIP) modeling. Gupta et al. (1995) have considered inflated distributions at the point zero and studied the structural properties of the inflated distribution. Murat and Szynal (1998) have extended the results of Gupta et al. (1995) to the discrete distributions inflated at any point in the support 's' . Farewell and Sprott (1988) analyzed a data set on premature ventricular contractions where the distribution turns out to be inflated binomial. Famoye et al. (2005) have studied a rich family of generalized Poisson regression models inflated at ‘k’ and they have further provided a score test for inflation at ‘k’. Deng and Paul (2000) have obtained score test for zero inflation in generalized discrete linear models. The test is illustrated using the binomial and Poisson model. Ridout et al. (2001) have provided a score test for testing a zero-inflated Poisson regression model against zeroinflated negative binomial alternatives. Poisson regression model provides a standard framework for the analysis of the count data. In the regression context, either or both of these parameters may depend on covariates. Jansakul et al. (2002) have modified the score test to the regression case where the zero mixing probability is dependent on covariates. They extended Van Den Broek’s (1995) test to the more general situation, where the zero probability is allowed to depend on covariates. Deng and Paul (2005) deal with the class of zero-inflated over dispersed generalized linear models and propose procedures based on score tests for selecting a model that fits such data.

The present paper, deals with zero-inflated Poisson model, which is define through a zero-inflated power series distribution. The zero-inflated Poisson series distribution contains two parameters. The first parameter indicates inflation of zero and the other parameter is that of Poisson series distribution. In the present study, we focus on the structural properties of zero-inflated Poisson series model.

2. Zero-inflated power series distributions

If the data set contains excess number of zero counts, a mixture assigning a mass of (1-π) to the extra zeros and a mass of π to the power series distribution, it leads to the zero-inflated power series distribution. The probability mass function of ZIPSD is given by

$$P(Y=y)=\begin{cases}\pi+(1-\pi)\dfrac{a_0}{g(\lambda)}, & y=0\\[2ex](1-\pi)\dfrac{a_y\lambda^y}{g(\lambda)}, & y=1,2,3,.....\end{cases}$$

The mean and variance of Y is given by

$$E(Y)=(1-\pi)\lambda\frac{g'(\lambda)}{g(\lambda)}$$

$$Var(Y)=(1-\pi)\lambda\left\{\lambda\frac{g''(\lambda)}{g(\lambda)}+\frac{g'(\lambda)}{g(\lambda)}-(1-\pi)\lambda\left[\frac{g'(\lambda)}{g(\lambda)}\right]^2\right\}$$

3. Zero-inflated Poisson distribution model

By using power series model from the Poisson distribution

$$f(\lambda) = e^{\lambda} = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$$

The probability mass function of zero-inflated Poisson series is given by

$$P(Y = y) = \begin{cases} \pi + (1-\pi)e^{-\lambda} & for\ y = 0 \\ (1-\pi)\dfrac{e^{-\lambda}\lambda^y}{y!} & for\ y = 1,2,3,.... \end{cases}$$

The distribution of Y could be represented as Y~ZIP(π,λ). We would consider ZIP distribution as a mixture of a standard Poisson(λ) distribution and a degenerate distribution with all mass at zero. In fact, when there is no zero-inflation π=0, the zero-inflation simple becomes the ordinary Poisson distribution.

So $f'(\lambda) = \dfrac{\partial f}{\partial \lambda} = e^{\lambda}$ & $f''(\lambda) = \dfrac{\partial^2 f}{\partial \lambda^2} = e^{\lambda}$

The mean is given by

$$\begin{aligned} E(Y) &= (1-\pi)\lambda \frac{f'(\lambda)}{f(\lambda)} \\ &= (1-\pi)\lambda \frac{e^{\lambda}}{e^{\lambda}} \\ &= (1-\pi)\lambda \end{aligned}$$

The variance is given by

$$\begin{aligned} Var(Y) &= (1-\pi)\lambda^2 \frac{f''(\lambda)}{f'(\lambda)} + (1-\pi)\lambda \frac{f'(\lambda)}{f(\lambda)} - (1-\pi)^2 \lambda^2 \left[\frac{f'(\lambda)}{f(\lambda)}\right]^2 \\ &= (1-\pi)\lambda^2 \frac{e^{\lambda}}{e^{\lambda}} + (1-\pi)\lambda \frac{e^{\lambda}}{e^{\lambda}} - (1-\pi)^2 \lambda^2 \left[\frac{e^{\lambda}}{e^{\lambda}}\right]^2 \\ &= (1-\pi)\lambda^2 + (1-\pi)\lambda - (1-\pi)^2\lambda^2 \\ &= (1-\pi)\lambda[\lambda + 1 - (1-\pi)\lambda] \\ &= (1-\pi)\lambda[\lambda + 1 - \lambda + \pi\lambda] \\ &= (1-\pi)\lambda[1+\pi\lambda] \end{aligned}$$

4. The recurrence relation for the central moments of ZIP is given by

$$\mu_{r+1} = \lambda\left[\frac{\partial}{\partial\lambda}\mu_r + r\mu_{r-1}\frac{\partial}{\partial\lambda}\mu_1^{'}\right] + \pi(-\mu_1^{'})^{r+1} - \left[\mu_r - \pi(-\mu_1^{'})^r\right]\left[\mu_1^{'} - \frac{\mu_1^{'}}{(1-\pi)}\right]$$

But $\mu_1^{'} = E(Y) = (1-\pi)\lambda$ so we can write the ZIP as fallows

$$= \lambda \left[\frac{\partial}{\partial \lambda} \mu_r + (1-\pi) r \mu_{r-1} \right] + \pi(-(1-\pi)\lambda)^{r+1} - \pi\lambda[\mu_r - \pi(-(1-\pi)\lambda)^r]$$

If r=1

$$\mu_2 = (1-\pi)\lambda + \pi((1-\pi)\lambda)^2 - \pi^2(1-\pi)\lambda^2$$
$$= (1-\pi)\lambda + \pi(1-\pi)\lambda^2((1-\pi)-\pi)$$
$$= (1-\pi)\lambda[1+\pi\lambda]$$

The moment generating function of ZIP is given by

$$M_Y^{(t)} = \pi + (1-\pi)\frac{f(\lambda e^t)}{f(\lambda)}$$
$$= \pi + (1-\pi)e^{(\lambda e^t - \lambda)}$$
$$= \pi + (1-\pi)e^{\lambda(e^t - 1)}$$

The r^{th} moment is obtained from the r^{th} derivative of $M_Y^{(t)}$ w. r. t.'t' then t=0 i.e.,

$$\mu_r^{'} = \frac{\partial^r M_Y^{(t)}}{\partial t^r}|_{t=0}$$

If r=1 we have

$$\mu_1^{'} = \frac{\partial}{\partial t}\left[\pi + (1-\pi)e^{\lambda(e^t-1)} \right]_{t=0}$$
$$= \left[(1-\pi)\lambda e^t e^{\lambda(e^t-1)} \right]_{t=0}$$
$$= \left[(1-\pi)\lambda e^0 e^{\lambda(e^0-1)} \right]$$
$$= (1-\pi)\lambda$$

If r=2 we have

$$\mu_2^{'} = (1-\pi)\lambda \frac{\partial}{\partial t} e^{\lambda(e^t-1)+t} |_{t=0}$$
$$= \left[(1-\pi)\lambda(\lambda e^t + 1)e^{\lambda(e^t-1)+t} \right]_{t=0}$$
$$= (1-\pi)\lambda(\lambda+1)$$
$$= (1-\pi)(\lambda^2+\lambda)$$
$$= (1-\pi)\lambda^2 + (1-\pi)\lambda$$

The variance is given by

$$\mu_2 = \mu_2^{'} - (\mu_1^{'})^2$$
$$= (1-\pi)\lambda^2 + (1-\pi)\lambda - (1-\pi)^2\lambda^2$$
$$= (1-\pi)\lambda(1+\pi\lambda)$$

5. Factorial moment generating function of ZIP is given by

$$M_Y^{(t)} = \pi + (1-\pi)\frac{f(\lambda+\lambda t)}{f(\lambda)}$$
$$= \pi + (1-\pi)e^{\lambda t}$$

The r^th factorial moment is obtained from the r^th derivative of $M_Y^{(t)}$ w. r. t.'t' then t=0 i.e.,

$$\mu_r = \frac{\partial^r M_Y^{(t)}}{\partial t^r}|_{t=0}$$

If r=1

$$\mu_1 = \frac{\partial}{\partial t}\left[\pi + (1-\pi)e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda e^{0}\right]$$
$$= (1-\pi)\lambda$$

If r=2

$$\mu_2 = \frac{\partial}{\partial t}\left[(1-\pi)\lambda e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda^2 e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda^2 e^{0}\right]$$
$$= (1-\pi)\lambda^2$$

If r=3

$$\mu_3 = \frac{\partial}{\partial t}\left[(1-\pi)\lambda^2 e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda^3 e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda^3 e^{0}\right]$$
$$= (1-\pi)\lambda^3$$

If r=4

$$\mu_4 = \frac{\partial}{\partial t}\left[(1-\pi)\lambda^3 e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda^4 e^{\lambda t}\right]_{t=0}$$
$$= \left[(1-\pi)\lambda^4 e^{0}\right]$$
$$= (1-\pi)\lambda^4$$

The recurrence relationship between factorial moments of ZIP is given by $\mu_r = \lambda\mu_{r-1} \quad if\ r \geq 1$

6. The cumulant generating function of ZIP is given by

$$K_y^{(t)} = \log M_y^{(t)}$$
$$= \log[\pi + (1-\pi)e^{\lambda(e^t-1)}]$$

The r[th] cumulant of the distribution is obtained from the r[th] derivative of $K_Y^{(t)}$ w. r. t.'t' then t=0 i.e.,

$$K_r = \frac{\partial^r K_y^{(t)}}{\partial t^r}\Big|_{t=0}$$

If r=1, we have

$$K_1 = \frac{\partial K_y^{(t)}}{\partial t}\Big|_{t=0}$$
$$= \frac{\partial}{\partial t}\log[\pi + (1-\pi)e^{\lambda(e^t-1)}]_{t=0}$$
$$= \left[\frac{(1-\pi)\lambda e^{\lambda(e^t-1)+t}}{\pi + (1-\pi)e^{\lambda(e^t-1)}}\right]_{t=0}$$
$$= \frac{(1-\pi)\lambda}{\pi + (1-\pi)} = (1-\pi)\lambda$$

If r=2, we have

$$K_2 = \frac{\partial}{\partial t}\left[\frac{(1-\pi)\lambda e^{\lambda(e^t-1)+t}}{\pi + (1-\pi)e^{\lambda(e^t-1)}}\right]_{t=0}$$
$$= \left[\frac{(1-\pi)\lambda(\lambda e^t+1)e^{\lambda(e^t-1)+t}[\pi+(1-\pi)e^{\lambda(e^t-1)}] - [(1-\pi)\lambda e^{\lambda(e^t-1)+t}]^2}{[\pi+(1-\pi)e^{\lambda(e^t-1)}]^2}\right]$$
$$= \frac{(1-\pi)\lambda(\lambda+1)(\pi+(1-\pi)) - (1-\pi)^2\lambda^2}{(\pi+(1-\pi))^2}$$
$$= (1-\pi)\lambda(\lambda+1) - (1-\pi)^2\lambda^2$$
$$= (1-\pi)\lambda(\lambda+1-(1-\pi)\lambda)$$
$$= (1-\pi)\lambda(1+\lambda\pi)$$

7. The probability generating function for ZIP is given by

$$G_y(s) = \pi + (1-\pi)\frac{f(\lambda s)}{f(\lambda)}$$
$$= \pi + (1-\pi)e^{\lambda(s-1)}$$
$$G_y^{'}(s) = (1-\pi)\lambda\frac{e^{\lambda s}}{e^{\lambda}}$$
$$G_y^{''}(s) = (1-\pi)\lambda^2\frac{e^{\lambda s}}{e^{\lambda}}$$

If s=1, we have

$$G_y^{'}(1) = (1-\pi)\lambda \frac{e^{\lambda}}{e^{\lambda}}$$
$$= (1-\pi)\lambda$$
$$G_y^{''}(1) = (1-\pi)\lambda^2 \frac{e^{\lambda}}{e^{\lambda}}$$
$$= (1-\pi)\lambda^2$$

The mean and variance are

$$E(Y) = G_y^{'}(1) = (1-\pi)\lambda$$
$$Var(Y) = G_y^{''}(1) + G_y^{'}(1) - [G_y^{'}(1)]^2$$
$$= (1-\pi)\lambda^2 + (1-\pi)\lambda - [(1-\pi)\lambda]^2$$
$$= (1-\pi)\lambda(1+\pi\lambda)$$

8. Moment estimation of zero-inflated Poisson model

The first and second rth moment for ZIP are given by

$$\mu_1^{'} = E(X) = (1-\pi)\lambda$$
$$\mu_2^{'} = E(X^2) = (1-\pi)\lambda^2 + (1-\pi)\lambda$$

The first and second rth sample moment for ZIP are given by

$$s_1^{'} = \frac{\Sigma x_i}{n} = \bar{x} = (1-\pi)\lambda \qquad (1)$$
$$s_2^{'} = \frac{\Sigma x_i^2}{n} = (1-\pi)\lambda^2 + (1-\pi)\lambda \qquad (2)$$

From (1) and (2) we have

$$\frac{\Sigma x_i^2}{n} = \bar{x}\lambda + \bar{x}$$
$$\Sigma x_i^2 = \bar{x}\lambda n + \bar{x}n$$
$$\Sigma x_i^2 - \bar{x}n = \bar{x}\lambda n$$
$$\lambda = \frac{\Sigma x_i^2 - \bar{x}n}{\bar{x}n} = \frac{\Sigma x_i^2}{\bar{x}n} - 1$$

Substituting the value of $\hat{\lambda}$ in equation (1) we get

$$\bar{x} = (1-\pi)\lambda$$
$$\bar{x} = \lambda - \pi\lambda$$
$$\pi = \frac{\lambda - \bar{x}}{\lambda} = 1 - \frac{\bar{x}}{\lambda}$$
$$\hat{\pi} = 1 - \frac{\bar{x}^2 n}{\Sigma x_i^2 - \bar{x}n}$$

9. Maximum likelihood estimation for ZIP model is given by

By using power series model we have a random sample of size 'n', taking the values $x_1,x_2,\ldots..,x_n$ obsereved from ZIP distribution. The likelihood function is given by

$$L(\lambda,\pi;x)=\prod_{i=1}^{n}\left\{\pi+(1-\pi)e^{-\lambda}\right\}^{1-a_i}\left\{(1-\pi_i)\frac{e^{-\lambda}\lambda^{x_i}}{x_i\,!}\right\}^{a_i}\qquad \lambda,\pi>0$$

Where $a_i=0$ if $x_i=0$ & $a_i=1$ if $x_i=1,2,3,\ldots..$

The corresponding log-likelihood function is given by

$$\log L(\lambda,\pi;x)=\eta_0\log(\pi+(1-\pi)e^{-\lambda})+\sum_{i=1}^{n}a_i\log(1-\pi_i)$$

$$-\lambda\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_ix_i\log(\lambda)-\sum_{i=1}^{n}a_i\log x_i\,!$$

Here η_0 = number of x_i' s are equal to zero in the sample

Therefore MLE's of λ and π is obtained by maximizing Log L (λ, π ; x) w.r.to λ & π respectively.

$$\frac{\partial\log L}{\partial\pi}=\frac{\eta_0(1-e^{-\lambda})}{\pi+(1-\pi)e^{-\lambda}}-\frac{\sum_{i=1}^{n}a_i}{(1-\pi)}\qquad (3)$$

$$\frac{\partial\log L}{\partial\lambda}=\frac{\eta_0(-(1-\pi)e^{-\lambda})}{\pi+(1-\pi)e^{-\lambda}}+\frac{\sum_{i=1}^{n}a_ix_i}{\lambda}-\sum_{i=1}^{n}a_i\qquad (4)$$

From equation (3) & (4) equating to zero then we get π and λ

$$\frac{\sum_{i=1}^{n}a_i}{(1-\pi)}=\frac{\eta_0(1-e^{-\lambda})}{\pi+(1-\pi)e^{-\lambda}}\qquad (5)$$

$$\frac{\eta_0(1-\pi)e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}}+\sum_{i=1}^{n}a_i=\frac{\sum_{i=1}^{n}a_ix_i}{\lambda}\qquad (6)$$

From (5) we have

$$\sum_{i=1}^{n}a_i\pi+\sum_{i=1}^{n}a_i(1-\pi)e^{-\lambda}=\eta_0(1-e^{-\lambda})-\pi\eta_0(1-e^{-\lambda})\qquad (7)$$

Here we take $\eta-\eta_0=\sum_{i=1}^{n}b_i$

Substituting this in equation (7)

$$(\eta-\eta_0)\pi+(\eta-\eta_0)(1-\pi)e^{-\lambda}=\eta_0(1-e^{-\lambda})-\pi\eta_0(1-e^{-\lambda})$$
$$\pi[(\eta-\eta_0)-(\eta-\eta_0)e^{-\lambda}+\eta_0(1-e^{-\lambda})]=\eta_0(1-e^{-\lambda})-(\eta-\eta_0)e^{-\lambda}$$
$$\pi(\eta-\eta e^{-\lambda})=\eta_0-\eta e^{-\lambda}$$
$$\hat{\pi}=\frac{\eta_0-\eta e^{-\hat{\lambda}}}{\eta(1-e^{-\hat{\lambda}})}$$
$$\hat{\pi}=\frac{\eta_0 e^{\hat{\lambda}}-\eta}{\eta(e^{\hat{\lambda}}-1)}$$

&

$$\frac{\sum_{i=1}^{n}a_i x_i}{\lambda}=\frac{\eta_0(1-\pi)e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}}+\sum_{i=1}^{n}a_i$$

$\hat{\pi}$ in above equation, then we get

$$\frac{\sum_{i=1}^{n}a_i x_i}{\lambda}=\frac{\eta_0\left[1-\left(\frac{\eta_0 e^{\lambda}-\eta}{\eta(e^{\lambda}-1)}\right)\right]e^{-\lambda}}{\left(\frac{\eta_0 e^{\lambda}-\eta}{\eta(e^{\lambda}-1)}\right)+\left(1-\left(\frac{\eta_0 e^{\lambda}-\eta}{\eta(e^{\lambda}-1)}\right)\right)e^{-\lambda}}+(\eta-\eta_0)$$

$$\frac{\sum_{i=1}^{n}a_i x_i}{\lambda}=\frac{\eta_0\left(\frac{\eta e^{\lambda}-\eta-\eta_0 e^{\lambda}+\eta}{\eta(e^{\lambda}-1)}\right)e^{-\lambda}}{\left(\frac{\eta_0 e^{\lambda}-\eta}{\eta(e^{\lambda}-1)}\right)+\left(\frac{\eta e^{\lambda}-\eta-\eta_0 e^{\lambda}+\eta}{\eta(e^{\lambda}-1)}\right)e^{-\lambda}}+(\eta-\eta_0)$$

$$\frac{\sum_{i=1}^{n}a_i x_i}{\lambda}=\frac{(\eta-\eta_0)}{(e^{\lambda}-1)}+(\eta-\eta_0)$$

$$\frac{\sum_{i=1}^{n}a_i x_i}{\lambda}=(\eta-\eta_0)\left(\frac{1}{(e^{\lambda}-1)}+1\right)$$

$$\frac{\sum_{i=1}^{n}a_i x_i}{(\eta-\eta_0)}=\frac{\lambda e^{\lambda}}{(e^{\lambda}-1)}$$

$$\frac{(\eta-\eta_0)\Sigma x_i}{(\eta-\eta_0)}=\frac{\lambda e^{\lambda}}{(e^{\lambda}-1)}$$

$$\overline{x}=\frac{\hat{\lambda}e^{\hat{\lambda}}}{(e^{\hat{\lambda}}-1)}$$

Where $\bar{x}$ is the sample mean. To obtain $\hat{\lambda}$ using the Newton-Raphson iteration.

$$\bar{x}\left(\frac{e^{\hat{\lambda}}-1}{e^{\hat{\lambda}}}\right)=\hat{\lambda}$$
$$\bar{x}\left(1-\frac{1}{e^{\hat{\lambda}}}\right)=\hat{\lambda}$$
$$\bar{x}\left(1-e^{-\hat{\lambda}}\right)=\hat{\lambda}$$

10. Compound zero-inflated Poisson model:

Let $S_N = X_1+X_2+.....+X_N$ where X_i's are random variables, with N being a zero-inflated Poisson random variable. Then S_N is said to have a compound zero-inflated Poisson model. Suppose N is zero-inflated Poisson with parameter(λ,π). Then the PGF of N is given by

$$F(s)=\pi+(1-\pi)e^{\lambda(s-1)}$$

And the PGF of S_N is given by

$$F(G_x(s))=\pi+(1-\pi)e^{\lambda(G_x(s)-1)}$$

The mean and variance is given by

$$\begin{aligned}E(S_N)&=E(N)E(X_i)\\&=(1-\pi)\lambda E(X_i)\\Var(S_N)&=E(N)Var(X_i)+\left[E(X)\right]^2 VarN\\&=(1-\pi)\lambda Var(X_i)+(1-\pi)\lambda(1+\pi\lambda)\left[E(X)\right]^2\\&=(1-\pi)\lambda\left[Var(X_i)+(1+\pi\lambda)\left[E(X)\right]^2\right]\end{aligned}$$

Here we can find the more cases

Case(i) If X is zero truncated geometric geometric distribution with parameter p, then

$$G_X(s)=\frac{ps}{1-qs}$$

$$\begin{aligned}F[G_X(s)]&=\pi+(1-\pi)e^{\lambda\left(\frac{ps}{1-qs}-1\right)}\\&=\pi+(1-\pi)e^{\lambda\left(\frac{s-1}{1-qs}\right)}\end{aligned}$$

This is the PGF of zero-inflated Poisson zero-inflated Poisson zero-truncated geometric distribution. The mean and variance is given by

$$\begin{aligned}E(S_N)&=E(N)E(X_i)\\&=\frac{(1-\pi)\lambda}{p}\\Var(S_N)&=E(N)VarX_i+[E(X)]^2VarN\\&=(1-\pi)\lambda VarX_i+(1-\pi)\lambda(1+\pi\lambda)[E(X)]^2\\&=(1-\pi)\lambda\left\{\frac{1-p}{p^2}+\frac{1-\pi\lambda}{p^2}\right\}\\&=(1-\pi)\lambda\left(\frac{1-p+1-\pi\lambda}{p^2}\right)\end{aligned}$$

Case(ii) If X_i is binomial with parameter (n,p) then

$$G_X(s) = (q + ps)^n$$
$$= (1 - p + ps)^n$$
$$F[G_X(s)] = \pi + (1-\pi)e^{\lambda\left((1-p+ps)^n - 1\right)}$$

This is the PGF of zero-inflated Poisson-Binomial model. The mean and variance is given by

$$E(S_N) = E(N)E(X_i)$$
$$= (1-\pi)\lambda np$$
$$Var(S_N) = E(N)VarX_i + [E(X)]^2 VarN$$
$$= (1-\pi)\lambda VarX_i + (1-\pi)\lambda(1+\pi\lambda)[E(X)]^2$$
$$= (1-\pi)\lambda np\left\{1 - p + (1+\pi\lambda)np\right\}$$

Case(iii) If X_i is Negative binomial with parameter (α,p) then

$$G_X(s) = \left(\frac{p}{1-(1-p)s}\right)^{\alpha}$$
$$F[G_X(s)] = \pi + (1-\pi)e^{\lambda\left(\left(\frac{p}{1-(1-p)s}\right)^{\alpha} - 1\right)}$$

This is the PGF of zero-inflated Poisson-negative binomial model. The mean and variance is given by

$$E(S_N) = E(N)E(X_i)$$
$$= (1-\pi)\lambda\frac{\alpha q}{p}$$
$$Var(S_N) = E(N)VarX_i + [E(X)]^2 VarN$$
$$= (1-\pi)\lambda\frac{\alpha q}{p^2} + (1-\pi)\lambda(1+\pi\lambda)\frac{\alpha^2 q^2}{p^2}$$
$$= (1-\pi)\lambda\left\{\frac{\alpha q}{p^2} + (1+\pi\lambda)\frac{\alpha^2 q^2}{p^2}\right\}$$

Case(iv) If X_i is the logarithmic series model with parameter (p), then

$$G_X(s) = \left(\frac{\ln(1-ps)}{\ln(1-p)}\right)$$
$$F[G_X(s)] = \pi + (1-\pi)e^{\lambda\left(\frac{\ln(1-ps)}{\ln(1-p)} - 1\right)}$$

This is the PGF of zero-inflated Poisson logarithemic series model. The mean and variance is given by

$$E(S_N) = E(N)E(X_i)$$
$$= \frac{(1-\pi)\lambda p}{-(1-p)\log(1-p)}$$
$$Var(S_N) = E(N)VarX_i + [E(X)]^2 VarN$$
$$= (1-\pi)\lambda\left\{\frac{-p^2 + p\log(1-p)}{\left[(1-p)\log(1-p)\right]^2} + (1+\pi\lambda)\left[\frac{p}{-(1-p)\log(1-p)}\right]^2\right\}$$
$$= (1-\pi)\lambda\left\{\frac{-p^2 + p\log(1-p) + (1+\pi\lambda)p^2}{\left[(1-p)\log(1-p)\right]^2}\right\}$$

Case(iv) If X_i is the Poisson model with parameter (λ), then

$$F(s) = e^{\lambda(s-1)}$$
$$F(G_x(s)) = \pi + (1-\pi)e^{\lambda(e^{\lambda(s-1)}-1)}$$

This is the PGF of zero-inflated Poisson-Poisson model. The mean and variance is given by

$$E(S_N) = E(N)E(X_i)$$
$$= (1-\pi)\lambda^2$$
$$Var(S_N) = E(N)Var(X_i) + \left[E(X)\right]^2 VarN$$
$$= (1-\pi)\lambda^2 + (1-\pi)\lambda(1+\pi\lambda)\lambda^2$$
$$= (1-\pi)\lambda^2\left[1 + (1+\pi\lambda)\lambda\right]$$

11. Asymptotic Relative Efficiency

The method of moment estimators are discussed above based on that a ZIP model the first and second theoretical moments are arrived as fallows

$$E(X) = (1-\pi)\lambda = s_1^{'} \; and \; E(X^2) = (1-\pi)\lambda^2(1+\pi\lambda) = s_2^{'}$$

The method of moment estimators of π and λ are respectively

$$\hat{\lambda} = \frac{s_2^{'}}{s_1^{'}} - 1 = \frac{\Sigma x_i^2}{\bar{x}n} - 1$$

$$\hat{\pi} = 1 - \frac{(s_1^{'})^2}{(s_2^{'} - s_1^{'})} = 1 - \frac{\bar{x}^2 n}{\Sigma x_i^2 - \bar{x}n}$$

It is easy to see that $P(s_1^{'} = 0) = \left\{\pi + (1-\pi)e^{-\lambda}\right\}^n \to 0, \, as \; n \to \infty$ .similarly, $P(s_1^{'} = s_2^{'}) \to 0, \, as \; n \to \infty$. Hence the problem of division by zero in these MMEs doesn't arise when n is sufficiently large. The fisher information matrix are

$$\Sigma = \begin{bmatrix} \dfrac{1-e^{-\lambda}}{(1-\pi)(\pi+(1-\pi)e^{-\lambda})} & \dfrac{e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}} \\ \dfrac{e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}} & \dfrac{(1-\pi)(\pi+(1-\pi)e^{-\lambda}+\pi\lambda)}{\lambda(\pi+(1-\pi)e^{-\lambda})} \end{bmatrix}$$

The asymptotic relative efficiency of MMEs with respect to MLEs of the parameters compared analytically in ZIP model. The MLEs of the parameters π and λ based on the ZIPS model. According to the above estimators of the parameters are asymptotically normally distributions. Hence the asymptotic relative efficiencies of the estimators are compared analytically. Since the zero-inflated Poisson series model belongs to two parameter exponential family, the MLEs of π and λ are asymptotically normal and

$$\sqrt{n}\begin{pmatrix} \hat{\lambda}_{mle} \\ \hat{\pi}_{mle} \end{pmatrix} \rightarrow Z^{'} \sim N\left(\begin{pmatrix} \lambda \\ \pi \end{pmatrix}, \Sigma^{-1}\right), as\ n \rightarrow \infty.$$

Hence the asymptotic relative efficiency of $\hat{\lambda}_m$ with respect to $\hat{\lambda}_{mle}$ is

$$ARE = (\hat{\lambda}_{mle}, \hat{\lambda}_m) = \frac{AV(\lambda_m)}{AV(\lambda_{mle})} = 1$$

Therefore, the MMEs and the MLEs of λ are asymptotically equally efficient. Similarly results in the case of π.

12. Heavy vehicle traffic accident data:

This data set consists of the accident frequencies for the heavy vehicle traffic accident data collected for the year 2010 from the National Highway No.6 commonly refer to as NH-6 or G.E. Road (Great Eastern Road) in India. We used a ZIP regression model to estimate the accident frequencies for the heavy vehicle traffic accident data, as shown in below table. Using our ZIP model gives accurate estimates for the given accident data frequencies using the ME and MLE methods for estimating its parameters, as shown in below Table from the goodness of fit p-values.

Table 1: Estimates of the heavy vehicle traffic accident data frequencies.

No.of accidents	Observed freq.,	Expected frequencies	
		ME ZIP	MLE ZIP
0	55	65	55
1	26	7	16
2	4	8	13
3	3	7	8
4	3	5	3
≥5	1	4	1
Total	96	96	96
Model parameters	Chi-square	58.4456	31.60577
	d.f	3	3
	p-value	0.0001	0.0001

Conclusion

We introduce the zero-inflated generalized Poisson distributions. Its structural properties were studied namely its mean, variance , probability generating function, moment generating function, and the factorial moment generating functions, and also moment's recurrence relations were obtained. Here we can also deals the compound zero-inflated Poisson cases. And also discuss the maximum likelihood estimation and the method of moment estimation of parameters in a zero-inflated Poisson series model. The maximum likelihood and moment estimators are asymptotically equally efficient. Here we use the application of these distributions for data modeling of road traffic accidents and that these distributions have more appropriate approximation then corresponding distributions by using chi-square goodness of fit.

References

[1]. Agresti A.(1996), An Introduction to Categorical Data Analysis , wiley.

[2]. Cameron A.C. and Trivedi P.K. (1998). Regression Analysis of Count Data. Cambridge University Press.

[3]. Cheung, Y. B. (2002). Zero-inflated models for regression analysis of count data: A study of growth and development. *Statistics in Medicine*, *21*, 1462–1469.

[4]. Dagne, G. A. (2010). Bayesian semiparametric zero-inflated Poisson model for longitudinal count data. *Mathematical Biosciences*, *224*, 126–130.

[5]. Gupta P. L., Gupta R. L. and Tripathi R. C. (2004). Score test for Zero Inflated Generalized Poisson Regression Model. Comm. Statist - Theory Meth. 33 (No.1), pp 47-64.

[6]. Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). Univariate Discrete Distributions, Third Edition, John Wiley and Sons; New Jersey.

[7]. Jonson, N. L., Kotz, S. and Kemp, A. W. "Univariate Discrete Distributions". Wiley Series in Probability and Mathematical Statistics.2 nd edition, (1992).

[8]. Jansakul, N. and Hinde, J. P. 2002. Score tests for zero-inflated Poisson models. Computational Statistics and Data Analysis. 40: 75-96.

[9]. Kolev, N., Minkova, L.D. and Neytchev, P. (2000). Inflated-Parameter Family of Generalized Power Series Distributions and Their Application in Analysis of Overdispersed Insurance Data. ARCH Research Clearing House, 2, 295-320.

[10]. Mccullagh P. and Nelder J.A. (1989). Generalized Linear Models. Chapman and Hall, London.

[11]. Nanjundan, G. and Naika, T. R. (2013). Estimation of Parameters in a Zero-inflated Power Series Model. International Journal of Statistika and Mathematika, 6(3), 109-114.

[12]. Patil, M. K. and Shirke, D. T. (2007). Testing parameter of the power series distribution of a zero inflated power series model. Statistical Methodology, 4(4), 393-406.

[13]. Yip, P., "Inference about the mean of Poisson distribution in the presence of a nuisance parameter", Austral. J.Statist., 30 (3), 299-306, 1988.