Pak. J. Statist.
2018 Vol. 34(8), 33-46

MODELING CESAREAN VERSUS NATURAL BIRTH DATA WITH SUSPICIOUS ZERO COUNTS SHOULD BE A PRELUDE TO HEALTHCARE COST ANALYSIS

Anwar Hassan[1§], **Peer Bilal Ahmad**[2] and **Ramalingam Shanmugam**[3]

[1] Department of Statistics & Operations Research, King Saud University Riyadh, Saudi Arabia. Email: anwar.hassan2007@gmail.com

[2] Department of Mathematical Sciences, Islamic University of Science & Technology, Awantipora (J&K), India. Email: peerbilal@yahoo.co.in

[3] School of Health Administration, Texas State University-San Marcos San Marcos, TX 78666, USA. Email: rs25@txstate.edu

[§] Corresponding author

ABSTRACT

Skyrocketing healthcare cost is an alarming concern not only in the United States of America but also in many developing and developed nations. A prelude to any discussions of financial matters of healthcare is to secure the best fitting probability pattern of the data. For an example, why not consider and compare the probability patterns in the random number of cesarean (X) versus the random number of natural (Y) childbirths in a day of the hospital data. A public opinion is that all cesarean births are really warranted and the healthcare cost could be considerably reduced if it can be eliminated. Their aggregate $N = X + Y$ denotes the total number of births in a day of the chosen hospital. The general expression for the correlation coefficient, ρ, between X and Y reveals when N, follows a Zero Inflated Modified Power Series Distribution (ZIMPSD) introduced by Gupta, *et al.* (1995). Special expressions for Zero Inflated Generalized Negative Binomial Distribution (ZIGNBD) and Zero Inflated Generalized Poisson Distribution (ZIGPD) are obtained as particular cases.

KEYWORDS

Zero Incidence, Inflated Modified Power Series Distribution (ZIMPSD), Zero Inflated Generalized Negative Binomial Distribution (ZIGNBD), Zero Inflated Generalized Poisson Distribution (ZIGPD), correlation coefficient (ρ), Genetics, Count models.

1. MOTIVATION AND LITERATURE SURVEY

In this beginning stage of the 21st century, the much-needed reform in healthcare operations is felt in all nations of the world. Suppose of interest is the relationship between the random number of cesarean births (X) and the random number of natural births (Y) in a maternity wing of a hospital. Notice that X is a binomial random variable with parameters N (which is the total number of births in a day in maternity wing of the chosen hospital) and p. The correlation coefficient 'ρ' between X and $Y = N - X$ provides a lot more insight in these two phenomena (see Rao, *et al.* (1973) for details). If N is constant (due to the capacity of the maternity wing in the hospital), it is evident that $\rho = -1$. However, the problem under consideration in this paper is to investigate the value of ρ when N itself is a random variable.

This type of discussion is not new in statistical research and falls under the framework of damaged models in the literature. Kojima and Kelleher (1962) showed that the aggregate of the number of boys and girls in a family follows a negative binomial distribution. Gupta (1976) studied the relationship between the number of boys and girls in a family, between the number of animals trapped and entrapped in an animal trapping experiment, between the surviving and non-surviving eggs laid by an insect. He obtained a general expression for their correlation coefficient, ρ, when N, the family size (see Gupta's (1974) for details) follows a modified power series distribution (MPSD). Gupta (1976) tabulated the value of ρ for Jain and Consul's (1971) Generalized Negative Binomial Distribution (GNBD), Consul and Jain's (1973) Generalized Poisson Distribution (GPD) and Jain and Gupta's (1973) Generalized Logarithmic Series Distribution (GLSD). Earlier, Rao (1981) investigated the correlation between the number of boys and girls for a given family size N. Interestingly, Rao (1981) proved that the correlation is positive (or negative) depending only on whether the function log f (θ) of MPSD is convex (or concave). Hasnain and Ahmad (2011) derived a distribution of the mean of k-independent sample correlation coefficient each of which is based on '*n*' pairs of observations. Hasnain and Ahmad (2013) also derived the distribution of weighted mean of two correlation coefficients and obtained its moments using Bassel function and confluent hyper geometric series function. Shanmugam, *et al.* (2006) obtained an expression for the correlation between the random number of epileptic and healthy children in family whose size follows a size-biased modified power series distribution.

In certain discrete data, the observed count for $X = 0$ is significantly higher than what the fitted model predicts; see Neyman (1939) and Feller (1943). To realize this, why not consider two machines. One of which (say machine I) is perfect and does not produce any defective item. The other (say machine II) produces defective items according to a Poisson distributional pattern. The aggregate number of items produced by both machines is only recorded without knowing whether a specific item is produced by machine I or by machine II. In this application, the observed number of Zeros is seen to be inflated. Farewell and Sprott (1988) analyze a data set on premature ventricular contractions where the distribution turns out to be inflated binomial. Yip (1988) described a similar situation with an inflated Poisson distribution dealing with the number of insects per leaf. Martine, *et al.* (2005) and Kuhnert, *et al.* (2005) make the distinction between different types of zeros in the ecological setting. Rodrigues (2003) and Gosh, *et al.* (2006) developed Bayesian Zero Inflated Poisson models for the cross-sectional data, using Markov chain Monte Carlo (MCMC) with data augmentation to obtain posterior samples. Kolev, *et al.* (2000) studied the inflated-parameter family of generalized power series distributions and their application in analysis of overdispersed insurance data. Patil and Shirke (2007) studied testing parameter of the power series distribution of a zero-inflated power series model. Patil and Shirke (2011) also studied equality of inflation parameters of two zero-inflated power series distributions. It appears that majority of the study in the literature is restricted to Poisson distribution and its extension to multivariate set up. Relatively less has been reported for the family of distributions containing other distributions.

In this paper, we consider several cases of Zero Inflated Modified Power Series Distributions (ZIMPSD). The Zero Inflated Modified Power Distributions were introduced as a versatile class (see Gupta, Gupta and Tripathi (1995) for additional details). The general expression for the correlation coefficient, ρ, is obtained when N, the aggregate has a Zero Inflated Modified Power Series Distribution (ZIMPSD).Then, specialized results for Zero Inflated Generalized Negative Binomial and Zero Inflated Generalized Poisson Distributions are discussed as particular cases. In Section 4, the results are illustrated by using the analysis of number of cesarean versus natural births in a day of a hospital. Finally, in Section 5, some conclusions and recommendations are provided to analyze such data sets more accurately.

2. DERIVATION OF A GENERAL EXPRESSION FOR CORRELATION COEFFICIENT 'ρ'

Let N be a discrete random variable having Zero Inflated Modified Power Series Distributions (ZIMPSD) given by

$$P[N=x]=\begin{cases}\varphi+\dfrac{(1-\varphi)a(0)}{f(\theta)}; x=0\\ \dfrac{(1-\varphi)a(x)(g(\theta))^{x}}{f(\theta)}; x=1,2,3,........\end{cases} \tag{2.1}$$

where $0<\varphi\leq 1$, the normalizing constant, $f(\theta)=\sum_{x} a(x)(g(\theta))^{x}$ and $g(\theta)$ are positive, finite and differentiable and coefficients $a\left(x\right)$ are non-negative and free of θ. With $\varphi=0$ in (2.1), Gupta's (1974) Modified Power Series Distributions (MPSD) is obtained and its probability function is

$$P[Z=z]=\frac{\mathrm{a(z)}\left(\mathrm{g}(\theta)\right)^{\mathrm{z}}}{\mathrm{f}(\theta)}, z\in T \tag{2.2}$$

where $a\left(z\right)>0$, $g(\theta)$ and $f(\theta)$ are positive, finite and differentiable and T is a subset of set of non-negative integers. Gupta *et al.* (1995) obtained the mean and variance of (2.1) and they are respectively

$$E(N)=(1-\varphi)E(Z) \tag{2.3}$$

and $$V(N)=(1-\varphi)\left[V(Z)+\varphi\left(E(Z)\right)^{2}\right] \tag{2.4}$$

where $E\left(Z\right)$ is the mean of (2.2), that is,

$$E(Z)=\frac{f'(\theta)g(\theta)}{f(\theta)g'(\theta)} \tag{2.5}$$

Similarly, the $V(Z)$ is the variance of (2.2) and it is given by

$$V(Z)=\frac{g(\theta)}{g'(\theta)}\frac{\partial}{\partial\theta}E(Z)=\frac{g(\theta)}{g'(\theta)}E'(Z) \tag{2.6}$$

where $E'(Z)=\frac{\partial}{\partial\theta}E(Z)$.

As discussed in the introduction, let X be a binomial random variable with parameters N and p, then the correlation coefficient, ρ, between X and $Y=N-X$ is given by

$$\rho=\frac{\left[p(1-p)\right]^{½}\left[V(N)-E(N)\right]}{\left[pV(N)+\left(1-p\right)E(N)\right]^{1/2}\left[\left(1-p\right)V(N)+p\,E(N)\right]^{½}} \tag{2.7}$$

(See Rao, *et al.* (1973) for the derivation). When $p=1/2$, the expression (2.7) reduces to

$$\rho=\frac{V(N)-E(N)}{V(N)+E(N)} \tag{2.8}$$

Substituting (2.3) and (2.4) in (2.7), we get

$$\rho=\frac{\left[p\,(1-p)\right]^{½}\left\{(1-\varphi)\left[V(Z)+\varphi(E(Z))^2\right]-(1-\varphi)E(Z)\right\}}{\left\{p(1-\varphi)\left[V(Z)+\varphi\,(E(Z))^2\right]+(1-p)(1-\varphi)E(Z)\right\}^{½}\left\{(1-p)(1-\varphi)\left[V(Z)+\varphi(E(Z))^2\right]+p\,(1-\varphi)E(Z)\right\}^{½}}$$

$$=\frac{\left[p\,(1-p)\right]^{½}\left\{\left[V(Z)+\varphi\,(E(Z))^2\right]-E(Z)\right\}}{\left\{p\left[V(Z)+\varphi\,(E(Z))^2\right]+(1-p)\,E(Z)\right\}^{½}\left\{(1-p)\left[V(Z)+\varphi\,(E(Z))^2\right]+p\,E(Z)\right\}^{½}}$$

That is,

$$\rho\ =\frac{\left[p\,(1-p)\right]^{½}\left\{V(Z)-\left[1-\varphi\,E(Z)\right]E(Z)\right\}}{\left\{pV(Z)+\left[p\,\varphi\,E(Z)+(1-p)\right]E(Z)\right\}^{½}\left\{(1-p)V(Z)+\left[(1-p)\,\varphi\,E(Z)+p\right]E(Z)\right\}^{½}} \tag{2.9}$$

Substituting (2.5) and (2.6) in (2.9), it simplifies further to.

$$\rho = \frac{\left[p(1-p)\right]^{1/2}\left\{\frac{g(\theta)}{g'(\theta)}E'(Z)-\left(1-\varphi E(Z)\right)\frac{g(\theta)f'(\theta)}{g'(\theta)f(\theta)}\right\}}{\left\{p\frac{g(\theta)}{g'(\theta)}E'(Z)+\left[p\varphi E(Z)+(1-p)\right]\frac{g(\theta)f'(\theta)}{g'(\theta)f(\theta)}\right\}^{1/2}\left\{(1-p)\frac{g(\theta)}{g'(\theta)}E'(Z)+\left[(1-p)\varphi E(Z)+p\right]\frac{g(\theta)f'(\theta)}{g'(\theta)f(\theta)}\right\}^{1/2}}$$

$$= \frac{\left[p(1-p)\right]^{1/2}\left\{f(\theta)E'(Z)-\left(1-\varphi E(Z)\right)f'(\theta)\right\}}{\left\{pf(\theta)E'(Z)+\left[p\varphi E(Z)+(1-p)\right]f'(\theta)\right\}^{1/2}\left\{(1-p)f(\theta)E'(Z)+\left[(1-p)\varphi E(Z)+p\right]f'(\theta)\right\}^{1/2}} \quad (2.10)$$

With $\varphi = 0$ in (2.10), one immediately gets the results obtained by Gupta (1976). When $p = 1/2$ in (2.10) indicating equally likely occurrence, it gives

$$\rho = \frac{\left\{f(\theta)E'(Z)-\left[1-\varphi E(Z)\right]f'(\theta)\right\}}{\left\{f(\theta)E'(Z)+\left[\varphi E(Z)+1\right]f'(\theta)\right\}} \quad (2.11)$$

Furthermore, with $\varphi = 0$, in (2.11), we get

$$\rho = \frac{\left[f(\theta)E'(Z)-f'(\theta)\right]}{\left[f(\theta)E'(Z)+f'(\theta)\right]} \quad (2.12)$$

With a reparametrization: $\lambda = \left\{(1-\varphi)\left(1-a(0)/f(\theta)\right)\right\}$, the model (2.1) transforms to

$$P\left[N=x\right] = \begin{cases} 1-\lambda; & x=0 \\ \dfrac{\lambda}{\left(1-a(0)/f(\theta)\right)}\dfrac{a(x)\left(g(\theta)\right)^x}{f(\theta)}; & x=1,2,3,..... \end{cases} \quad (2.13)$$

Suppose that $x_1, x_2, ..., x_n$ be a random sample from (2.13) and let n_i be the number of observations which are equal to i so that $n = \sum_{i=0}^{\infty} n_i$. Gupta, *et al.* (1995) showed that MLE of φ is

$$\hat{\varphi} = 1 - \frac{\hat{\lambda}\, f(\hat{\theta})}{f(\hat{\theta})-a(0)} \quad (2.14)$$

3. CORRELATION COEFFICIENT ‘ ρ ’ FOR A CLASS OF ZERO INFLATED MPSD’S.

In this section, we investigate Zero Inflated cases of several Modified Power Series Distributions.

3.1 Zero Inflated Generalized Poisson Distribution (ZIGPD):

For this case, the probability function is

$$P\left[N=x\right]=\begin{cases}\varphi+(1-\varphi)e^{-\theta} & ,\ x=0\\ \dfrac{(1-\varphi)\left(1+x\beta\right)^{x-1}\left(\theta e^{-\beta\theta}\right)^{x}}{x!\ \ e^{\theta}}, & x=1,2,......\end{cases}$$

Where $\beta\geq 0$ and $\theta>0$, $\left|\beta\theta\right|<1$. In our framework, $g(\theta)=\theta e^{-\beta\theta}$, $f(\theta)=e^{\theta}$, $a(x)=\dfrac{(1+x\beta)^{x-1}}{x!}$, $g'(\theta)=e^{-\beta\theta}(1-\beta\theta)$, $\mathrm{f}'(\theta)=\mathrm{e}^{\theta}$, $E(Z)=\dfrac{g(\theta)f'(\theta)}{g'(\theta)f(\theta)}=\dfrac{\theta}{1-\beta\theta}$ and $E'(Z)=\dfrac{1}{(1-\beta\theta)^{2}}$.

Substituting these in (2.10), we get

$$\rho=\frac{\left[p(1-p)\right]^{½}\left\{\beta\theta(2-\theta\beta)+\varphi\theta(1-\theta\beta)\right\}}{\left\{1-2(1-p)\theta\beta+(1-p)\theta^{2}\beta^{2}+p\varphi\theta(1-\theta\beta)\right\}^{½}\left\{1-2p\theta\beta+p\theta^{2}\beta^{2}+(1-p)\varphi\theta(1-\theta\beta)\right\}^{½}} \tag{3.1}$$

With $\varphi=0$ in (3.1), the results of Gupta (1976) can be seen. When $p=½$, we get

$$\rho=\frac{\left\{\beta\theta(2-\theta\beta)+\varphi\theta(1-\theta\beta)\right\}}{\left\{2-2\theta\beta+\theta^{2}\beta^{2}+\varphi\theta(1-\theta\beta)\right\}} \tag{3.2}$$

Special Cases:

i) For the Zero Inflated Poisson Distribution (that is, when $\beta=0$), the general expression for ρ reduces to

$$\rho=\frac{\left[p(1-p)\right]^{½}\left\{\varphi\theta\right\}}{\left\{1+p\varphi\theta\right\}^{\frac{1}{2}}\left\{1+(1-p)\varphi\theta\right\}^{½}}. \tag{3.3}$$

When $p=½$, it simplifies to

$$\rho=\frac{\{\varphi\theta\}}{\{2+\varphi\theta\}}. \tag{3.4}$$

ii) For the Zero Inflated Borel Tanner distribution (see Haight and Breuer (1960) for its descriptions) $(\beta = 1)$, the general expression for ρ is given by

$$\rho = \frac{\left[p(1-p)\right]^{½}\left\{\theta(2-\theta)+\varphi\theta(1-\theta)\right\}}{\left\{1-2(1-p)\theta+(1-p)\theta^2+p\varphi\theta(1-\theta)\right\}^{½}\left\{1-2p\theta+p\theta^2+(1-p)\varphi\theta(1-\theta)\right\}^{½}} \tag{3.5}$$

When $p = ½$, the correlation coefficient is

$$\rho = \frac{\left\{\theta(2-\theta)+\varphi\theta(1-\theta)\right\}}{\left\{2-2\theta+\theta^2+\varphi\theta(1-\theta)\right\}}\,. \tag{3.6}$$

3.2 Zero Inflated Generalized Negative Binomial Distribution (ZIGNBD):

For this case, the probability function is

$$P[N=x]=\begin{cases}\varphi+(1-\varphi)(1-\theta)^m & ;x=0\\[2mm] \dfrac{(1-\varphi)m\Gamma(m+\beta x)}{x!\Gamma(m+\beta x-x+1)}\dfrac{\left[\theta(1-\theta)^{\beta-1}\right]^x}{(1-\theta)^{-m}}, & x=1,2,3,....\end{cases}$$

where $0<\theta<1$, $|\theta\beta|<1$, $\beta=0\, or\, \beta\geq 1$. In our framework, $g(\theta)=\theta(1-\theta)^{\beta-1}$,

$f(\theta)=(1-\theta)^{-m}$, $a(x)=\dfrac{m\Gamma(m+\beta x)}{x!\Gamma(m+\beta x-x+1)}$ $g'(\theta)=(1-\theta)^{\beta-2}(1-\theta\beta)$,

$f'(\theta)=m(1-\theta)^{-m-1}$, $E(Z)=\dfrac{f'(\theta)g(\theta)}{g'(\theta)f(\theta)}=\dfrac{m\theta}{(1-\theta\beta)}$, $E'(Z)=\dfrac{m}{(1-\theta\beta)^2}$.

Using these in (2.10), we get

$$\rho = \frac{\left[p(1-p)\right]^{½}\left\{\theta(2\beta-1)-\theta^2\beta^2+\varphi m\theta(1-\theta\beta)\right\}}{\left\{1-\theta\left(p+2\beta(1-p)\right)+(1-p)\theta^2\beta^2+p\varphi m\theta(1-\theta\beta)\right\}^{½}\left\{1-\theta\left[(1-p)+2\beta p\right]+p\theta^2\beta^2+(1-p)\varphi m\theta(1-\theta\beta)\right\}^{½}} \tag{3.7}$$

If we put $\varphi = 0$ in (3.7), the result of Gupta (1976) is seen. Now when $p = ½$, we have

$$\rho = \frac{\left\{\theta(2\beta-1)-\theta^2\beta^2+\varphi m\theta(1-\theta\beta)\right\}}{\left\{2-\theta(1+2\beta)+\theta^2\beta^2+\varphi m\theta(1-\theta\beta)\right\}}\,. \tag{3.8}$$

Special cases:

i) For the Zero Inflated Generalized Geometric Series Distribution (m=1) (see Mishra 1982), the general expression for ‘ ρ ’ is given by

$$\rho = \frac{\left[p(1-p)\right]^{1/2}\left\{\theta(2\beta-1)-\theta^2\beta^2+\varphi\theta(1-\theta\beta)\right\}}{\left\{1-\theta\left(p+2\beta(1-p)\right)+(1-p)\theta^2\beta^2+p\varphi\theta(1-\theta\beta)\right\}^{1/2}\left\{1-\theta\left[(1-p)+2\beta p\right]+p\theta^2\beta^2+(1-p)\varphi\theta(1-\theta\beta)\right\}^{1/2}} \tag{3.9}$$

With $p = 1/2$, it reduces to

$$\rho = \frac{\left\{\theta(2\beta-1)-\theta^2\beta^2+\varphi\theta(1-\theta\beta)\right\}}{\left\{2-\theta(1+2\beta)+\theta^2\beta^2+\varphi\theta(1-\theta\beta)\right\}} \tag{3.10}$$

ii) For the Zero Inflated Geometric Series Distribution ($m = 1,\ \beta = 1$), note that the expression for ‘ ρ ’ is

$$\rho = \frac{\left[p(1-p)\right]^{1/2}\left\{\theta-\theta^2+\varphi\theta(1-\theta)\right\}}{\left\{1-\theta\left(p+2(1-p)\right)+(1-p)\theta^2+p\varphi\theta(1-\theta)\right\}^{1/2}\left\{1-\theta\left[(1-p)+2p\right]+p\theta^2+(1-p)\varphi\theta(1-\theta)\right\}^{1/2}} \tag{3.11}$$

When $p = 1/2$, it reduces to

$$\rho = \frac{\left\{\theta-\theta^2+\varphi\theta(1-\theta)\right\}}{\left\{2-3\theta+\theta^2+\varphi\theta(1-\theta)\right\}} \tag{3.12}$$

iii) For the Zero Inflated Binomial Distribution (that is, when $\beta = 0$), the general expression for ‘ ρ ’ is

$$\rho = \frac{\left[p(1-p)\right]^{1/2}\left\{-\theta+\varphi m\theta\right\}}{\left\{1-\theta p+p\varphi m\theta\right\}^{1/2}\left\{1-\theta(1-p)+(1-p)\varphi m\theta\right\}^{1/2}} \tag{3.13}$$

Now for $p = 1/2$ in (3.13), it is then

$$\rho = \frac{\left\{-\theta+\varphi m\theta\right\}}{\left\{2-\theta+\varphi m\theta\right\}} \tag{3.14}$$

iv) For the Zero Inflated Negative Binomial Distribution (that is, when $\beta = 1$), the expression for ‘ ρ ’ is

$$\rho = \frac{\left[p(1-p)\right]^{½}\left\{\theta(1-\theta)+\varphi m\theta(1-\theta)\right\}}{\left\{1-\theta\left(p+2(1-p)\right)+(1-p)\theta^2+p\varphi m\theta(1-\theta)\right\}^{½}\left\{1-\theta\left[(1-p)+2p\right]+p\theta^2+(1-p)\varphi m\theta(1-\theta)\right\}^{½}} \tag{3.15}$$

When $p = ½$ in (3.15), it reduces to

$$\rho = \frac{\left\{\theta(1-\theta)+\varphi m\theta(1-\theta)\right\}}{\left\{2-3\theta+\theta^2+\varphi m\theta(1-\theta)\right\}}. \tag{3.16}$$

4. ANALYSIS OF NUMBER OF CESAREAN VERSUS NATURAL BIRTHS IN A DAY OF A HOSPITAL

In this section, we illustrate the results of the previous section. A concern among the hospital administrators and insurance policy makers is that the number of cesarean births is increasing compared to natural births. The underlying model of the cesarean births data is key to address this issue. The real data in Table1 are from a hospital in Austin, Texas and are selected to check whether the models of this article are fitting well the data. For this set of data *(that is, the random variable N denoting births per day at the hospital), we applied Poisson Distribution (PD), Negative Binomial Distribution (NBD), Generalized Poisson Distribution (GPD) and Zero-Inflated Generalized Poisson Distribution (ZIGPD). For estimating parameters of Poisson, we used the maximum likelihood estimate (MLE) method but in NBD, there are two parameters (r and θ), the parameter ‘ r ’ is obtained by using moment method of estimation and after obtaining the parameter ‘ r ’ the second parameter θ is obtained by MLE method. Similarly there are two parameters of GPD, θ and β, the parameter β is obtained by moment method of estimation and the parameter θ is obtained by MLE method. The parameters of Zero-Inflated GPD are obtained by MLE method. Finally a correlation coefficient between the random number of cesarean births and the random number of natural births in a maternity wing of a hospital whose data is given in Table1 is obtained.

Let Y denotes the natural births; X the cesarean births and N the total (natural + cesarean) births in the hospital per day.

Table 1

Total Births (N)	**Frequency** $f(N)$	**Natural Births** (Y)	**Frequency** $f(Y)$	**Cesarean Births** (X)	**Frequency** $f(X)$
0	16	0	3	0	13
1	20	1	6	1	14
2	17	2	8	2	9
3	10	3	5	3	5
4	10	4	8	4	2
5	8	5	5	5	3
6 and more	13	6 and more	12	6 and more	1

To see whether the ordinary or the Zero-Inflated Generalized Poisson Distribution is suitable for this data set, we compare the observed proportion of Zero's with the expected proportion of Zero's as predicted by the ordinary Generalized Poisson Distribution. The observed proportion of Zeros is 0.17. For the expected proportion, we use the moment estimator of β given by $\hat{\beta} = \sqrt{\frac{S^2}{\bar{x}^3}} - \frac{1}{\bar{x}}$. After obtaining, the value of β by moment method of estimation, the parameter θ is obtained by the method of maximum likelihood estimation. The MLE of θ for GPD is given by $\hat{\theta} = \frac{\bar{x}}{(1+\beta\bar{x})}$.

For the data given in Table 1, $\bar{x} = 2.574$ and $S^2 = 4.076$. The estimated values are $\hat{\beta} = 0.100$, $\hat{\theta} = 2.047$. Therefore, $\hat{P}_0 = P(X=0) = e^{-\hat{\theta}} = 0.13$. Thus, the predicted proportion of Zeros is low as compared to the observed proportion of Zero's, thereby suggesting the Zero-Inflated Generalized Poisson Distribution as an appropriate model**.** We fitted the Zero-Inflated Generalized Poisson Distribution (ZIGPD) with β known. The other two parameters of Zero-Inflated GPD i.e. θ and φ are obtained by maximum likelihood estimation. A value of β was chosen first as proposed by the method of moments for GPD, then θ and φ are estimated by the method of maximum likelihood. The maximum likelihood estimates obtained by Gupta *et al*. (1995) for Zero-Inflated GPD are given as

$$\frac{\theta e^{\theta}}{(1-\theta\beta)\left(e^{\theta}-1\right)} = \frac{\sum_{i-1}^{\infty} i n_i}{n-n_0} \tag{4.1}$$

and

$$\hat{\varphi} = \left(\frac{n_0}{n} - e^{-\hat{\theta}}\right) \Big/ \left(1-e^{-\hat{\theta}}\right). \tag{4.2}$$

Now from the above data, $n = 94$, $n_0 = 16$ and $\sum_{i=1}^{\infty} in_i = 242$. Assuming $\beta = 0.100$ (Known) as proposed by moment method for GPD. Using these values, we get the following equation

$$f(\theta) = \theta e^{\theta} - 3.103\left[e^{\theta} - 0.103\theta e^{\theta} + 0.103\theta - 1\right].$$

After using the Newton Raphson method for solving the above equation, we get the estimated value of the parameter θ as $\hat{\theta} = 2.135$. Using the value of $\hat{\theta}$ in (4.2), we obtain the estimated value of φ as $\hat{\varphi} = 0.0589$. Therefore we have $\hat{\theta} = 2.135$, $\hat{\varphi} = 0.0589$ and $\beta = 0.100$ (known).

Table 2
Observed and Expected Number of Births (Cesarean + Natural) in the Hospital (Goodness of Fit)

Total Births (N)	**Observed Frequency**	**Expected Frequency**			
		NBD	**PD**	**GPD**	**ZIGPD**
0	16	12.4	7.2	12.3	16.0
1	20	20.1	18.4	20.3	17.9
2	17	20.1	23.7	20.2	18.5
3	10	15.8	20.4	15.8	15.0
4	10	10.8	13.1	10.7	10.6
5	8	8.0	6.9	7.8	8.5
≥6	13	6.8	4.3	6.9	7.5
Total	94	94	94	94	94
Mean	2.574				
S.D	2.019				
χ^2_{cal}		9.365	27.400	9.197	6.131
d.f.		4	4	4	3
P-value		0.0526	0.00001	0.0564	0.1054
Estimates					
$\hat{\theta}$		0.6315	2.574	2.047	2.135
$\hat{r}$		4.41			
$\hat{\beta}$				0.100	0.100
$\hat{\varphi}$					0.0589

It is evident from the Table2 that after fitting the Poisson, Negative Binomial, Generalized Poisson and Zero-Inflated Generalized Poisson Distributions to the observed distribution, the Zero-Inflated GPD provides better fit to the observed distribution than PD, NBD and GPD.

Now, suppose we are interested in finding the relationship between the random number of cesarean births $\left(X\right)$ and the random number of natural births $\left(Y\right)$ in a maternity wing of a hospital whose data is given in Table1. Notice that X is a binomial random variable with parameters N (which is the total number of births in a day in maternity wing of the chosen hospital) and p, note that $\hat{p}=\frac{76}{242}=0.314$. Assume that the parameter N is itself a random variable following Zero-Inflated Generalized Poisson Distribution as discussed above. After using the expression for ‘ ρ ’ given in (3.1) and the estimates obtained above for different parameters θ, β and φ, the value of correlation coefficient ρ is obtained as $\hat{\rho}=0.308$.

5. CONCLUSIONS AND REMARKS

The Zero-Inflated Generalized Distributions studied in this paper are generalization of the classical distributions. It takes into account the extra zeros present in the data than those predicted by the model. The analysis presented shows that the Zero-Inflated Generalized Poisson Distribution provides better fit to the observed distribution than Poisson distribution, Negative Binomial Distribution and Generalized Poisson Distribution. It is, therefore, recommended that in order to obtain more accurate results, the model should be adjusted for the number of zeros. The analysis also shows that the inflated distributions has great applications in health care data which is almost inflated in one way or the other and therefore the inflated distributions should be tried for such data in order to get the better results. The paper also gives the insight about how to find the relationship between the two random variables, where one of the variables follows a binomial distribution with parameters N and p, and the parameter N is itself a random variable following zero-inflated distribution.

ACKNOWLEDGEMENT

The authors appreciate and thank the referee and the editor for many helpful comments and suggestions, which substantially simplified the paper.

REFERENCES

1. Consul, P.C. and Jain, G.C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4), 791-799.
2. Farewell, V.T. and Sprott, D.H. (1988). The use of a mixture model in the analysis of count data. *Biometrics*, 44, 1191-1194.
3. Feller, W. (1943). On a general class of “Contagious” distributions. *Annals of Mathematical Statistics*, 14, 389-400.

4. Ghosh, S.K., Mukhopadhyay, P. and Lu, J.C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, 136, 1360-1375.
5. Gupta, P.L., Gupta, R.C. and Tripathi, R.C. (1995). Inflated modified power series distribution with applications. *Commun. Statist. Theory Meth*., 24(9), 2355-2374.
6. Gupta, R.C. (1974). Modified power series distribution and some of its applications. *Sankhya*, 36, Series B, 288-298.
7. Gupta, R.C. (1976). Some applications of modified power series distribution in genetics. *Sankhya*, 38(2), Series B, 187-191.
8. Haight, F.A. and Breuer, M.A. (1960). The Borel-Tanner distribution. *Biometrika*, 47, 143-146.
9. Hasnain, S.A. and Ahmad, M. (2011). Distribution of mean of correlation coefficients. *Pakistan Journal of Statistics*, 27(2), 187-196.
10. Hasnain, S.A. and Ahmad, M. (2013). On the moments of weighted mean of two sample correlation coefficients. *Pakistan Journal of Statistics*, 29(1), 79-82.
11. Jain, G.C. and Consul, P.C. (1971). A generalized negative binomial distribution. *SIAM Jour. Appl. Math.*, 21, 501-513.
12. Jain, G.C. and Gupta, R.P. (1973). A logarithmic type distribution. *Trabjos de Estadistica*, 24, 99-105.
13. Kojima, Kem-Ichi and Kelleher, T.M. (1962). Survival of mutant genes. *The American Naturalist*, XCVI, 329-346.
14. Kolev, N. Minkova, L. and Neytchev, P. (2000). Inflated-Parameter family of generalized power series distributions and their applications in analysis of overdispersed insurance data. *ARCH Research Clearing House*, 2, 295-320.
15. Kuhnert, P.M., Martin, T.G., Mengersen, K. and Possingham, H.P. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: A Bayesian approach using expert opinion. *Environmetrics*, 16, 717-747.
16. Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J. and Possingham, H.P. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8, 1235-1246.
17. Mishra, A. (1982). A generalization of geometric series distribution. *J. Bihar Math. Soc.*, 6, 18-22.
18. Neyman, J. (1939). On a class of contagious distributions applicable in entomology and bacteriology. *Annals of Math. Stat.*, 10, 35-57.
19. Patil, M.K. and Shirke, D.T. (2007). Testing parameter of the power series distributions of a zero inflated power series model. *Statistical Methodology*, 4, 393-406.
20. Patil, M.K. and Shirke, D.T. (2011). Tests for equality of inflation parameters of two zero-inflated power series distributions. *Communications in Statistics: Theory and Methods*, 40(14), 2539-2553.
21. Rao, B.R., Mazumdar, S., Waller, J.H. and Li, C.C. (1973). Correlation between the numbers of two types of children in a family. *Biometrics*, 29, 271-279.
22. Rao, B.R. (1981). Correlation between the numbers of two types of children in a family with the modified power series distribution for the family size. *Communications in Statistics: Theory and Methods*, 10(3), 249-254.

23. Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions. *Communications in Statistics: Theory and Methods*, 32, 281-289.
24. Shanmugam, R., Hassan, A. and Ahmad, P.B. (2006). Correlation between the number of epileptic and healthy children in family size that follows a size-biased modified power series distribution. *Journal of Modern Applied Statistical Methods*, 5(2), 458-463.
25. Yip, P. (1988). Inference about the mean of a Poisson distribution in presence of a nuisance parameter. *Austral. J. Statist.*, 30(3), 299-306.