International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 4 Issue 9 || November. 2016 || PP-45-48

An Extension to the Zero-Inflated Generalized Power Series Distributions

Rafid Saeed Abdulrazak Alshkaki
Ahmed Bin Mohammed Military College, Doha, Qatar

ABSTRACT: *In many sampling involving non negative integer data models, the number of zeros is observed to be significantly higher than the expected number in the assumed model. Such models are called zero-inflated models. These models are recently cited in literature in various fields of science including; engineering, natural, social and political sciences. The class of zero-inflated power series distributions was recently considered and studied. Members of the class of generalized power series distributions are some of the well-known discrete distributions such as; the Poisson, binomial, negative binomials, as well as most of their modified forms. In this paper an extension to class of zero-inflated power series distributions was introduced, namely, the zero-one inflated case, and its structural properties were studied. In particular, its moments and some of its moment's recurrence relations were obtained along with some of its generating functions. Some of these results were shown in term of the zero-inflated cases as well.*

Keywords: *Zero and Zero-One Inflated Models, Power Series Distributions, non-negative integer sampling, Moments.*

I. INTRODUCTION

In many sampling involving non-negative integer data models, the number of zeros is observed to be significantly higher than the predicated number by the standard assumed distribution. Hence, as a result of mis-specifying the proper statistical distribution, leads to the so called the zero-inflated distributions. These models are recently cited in literature in various fields of science including; engineering, natural, social and political sciences.

The class of zero-inflated power series distributions (ZIPSDs) was recently considered and studied. Well-known discrete distributions such as, the Poisson, binomial, negative binomial and the logarithmic, are members of this class.

Gupta et al [1] studied some structural properties and the distribution of the sum of independent the ZIPSD, as well as estimation of its parameters. Ridout et al [2] considered the problem of modelling count data with excess zeros, called them zero-modified distributions. Kolev et al [3] proposed the ZIPSDs for modeling of correlated count data which exhibit overdispersion. Patil and Shirke [4] provided three asymptotic test for testing the parameters of the ZIPSDs. Murat and Szynal [5] studied some structural properties and estimation of the parameters of the ZIPSDs. Bhattacharya et al [6] considered a Bayesian test for testing the ZIPSD as a mixture of a power series distribution (PSD) and a degenerate distribution at zero. Nanjundan and Naika [7] considered estimation the parameters of ZIPSDs. Edwin [8] studied the construction and estimation of the ZIPSDs, and considered some its applications.

In this paper, we introduce in Section 2, the definition of the class of discrete PSDs with some of its members, properties; its mean, variance, probability generating function (PGF). Then, in Section 3, we introduce the class of ZIPSD, its definition with some of its properties including the mean, variance, and the PGF. In Section 4 we introduce the definition of the zero-one inflated power series distribution (ZOIPSD), and it's mean and variance, PGF, moment generating function(MGF) and the factorial moment generating(FMGF), and moment recurrences relations, were given in Section 5. Some of these results are shown in term of the standard and the zero-inflated cases as well.

II. POWERSERIES DISTRIBUTIONS

Let $g(\theta) = \sum_{x \in T} a_x \theta^x$ be a power series for $\theta \epsilon \Omega = \{\theta; 0 < \theta < \omega\}$, where ω is the radius of convergence of $g(\theta)$, and $a_x \geq 0$ for all $x \in T \subseteq I = \{0,1,2,\ldots\}$. Then the discrete random variable (rv) X having probability mass function (pmf);

$$P(X = x) = \begin{cases} \frac{a_x \theta^x}{g(\theta)}, & x \in T \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

is said to have a PSD with range T and parameter space Ω. The function and $g(\theta)$ is called the series defining function (sdf). We will denote that by writing $X \sim PSD(\theta, g(\theta))$.

The discrete distributions; Poisson, binomial, negative binomial, and the logarithmic series distributions, are well-known member of the PSD's with sdf's; e^{θ}, $(1+\theta)^n$, $(1-\theta)^{-k}$, and $-\ln(1-\theta)$, respectively. See Abdulrazak and Patil [9], [10] and Johnson et al [11] for further details.
The mean of the rv X having $\mathrm{PSD}(\theta, g(\theta))$ can be easily seen to be:

$$E(X) = \theta \frac{g'(\theta)}{g(\theta)},$$

and its variance is given by;

$$\mathrm{Var}(X) = \theta^2 \frac{g''(\theta)}{g(\theta)} + \theta \frac{g'(\theta)}{g(\theta)} - \left[\theta \frac{g'(\theta)}{g(\theta)}\right]^2$$

Similarly, the PGF of the rv X, $G_X(t)$, is given by;

$$G_X(t) = E(t^X) = \frac{g(\theta t)}{g(\theta)}$$

III. ZERO-INFLATED POWER SERIES DISTRIBUTIONS

Let the rv $X \sim \mathrm{PSD}(\theta, g(\theta))$.as given in (1) with $T = \{0, 1, 2, \dots\}$, and let $\alpha \in (0,1)$ be an extra proportion added to the proportion of zero of the rv X, then the rv Y defined by;

$$P(Y = y) = \begin{cases} \alpha + (1-\alpha)\frac{a_0}{g(\theta)}, & y = 0 \\ (1-\alpha)\frac{a_y\,\theta^y}{g(\theta)} & y = 1, 2, 3, \dots \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

is said to have a ZIPSD, and we will denote that by writing $Y \sim \mathrm{ZIPSD}(\theta, g(\theta); \alpha)$. The mean of Y is given by;

$$E(Y) = (1-\alpha)\,\theta\, \frac{g'(\theta)}{g(\theta)}$$
$$= (1-\alpha)\, E(X)$$

and its variance of is given by;

$$\mathrm{Var}(Y) = (1-\alpha)\,\theta \left\{\theta \frac{g''(\theta)}{g(\theta)} + \frac{g'(\theta)}{g(\theta)} - (1-\alpha)\,\theta \left[\frac{g'(\theta)}{g(\theta)}\right]^2\right\}$$
$$= (1-\alpha)\mathrm{Var}(X) + \alpha(1-\alpha)\,\theta^2 \left[\frac{g'(\theta)}{g(\theta)}\right]^2$$
$$= (1-\alpha)[\mathrm{Var}(X) + \alpha\,(E(X))^2]$$

The PGF of Y is given by;

$$G_Y(t) = \alpha + (1-\alpha)\frac{g(\theta t)}{g(\theta)}$$

IV. ZERO-ONE INFLATED POWER SERIES DISTRIBUTIONS

The following table represents the fetal movement data set that was collected, Leroux and Puterman [12], in a study of breathing and body movements in fetal lambs designed to examine the possible change in the amount of pattern of fetal activity during the last two thirds of the gestation period.

Number of movements	0	1	2	3	4	5	6	7
Number of intervals	182	41	12	2	2	0	0	1

Edwin [8] considered thatthe number of fetal movement with zero was inflated and studied this case as a ZIPSD model.
If we look at the above frequencies, we notice that the number of fetal movement with ones is also inflated. This suggests an extension to the zero-one inflated models.
Let the rv $X \sim \mathrm{PSD}(\theta, g(\theta))$ as given in (1) with $T = \{0, 1, 2, \dots\}$, and let $\alpha \in (0,1)$ be an extra proportion added to the proportion of zero of the rv X, $\beta \in (0,1)$ be an extra proportion added to the proportion of ones of the rv X, such that $0 < \alpha + \beta < 1$, then the rv Z defined by;

$$P(Z=z)=\begin{cases}\alpha+(1-\alpha-\beta)\frac{a_0}{g(\theta)}, & z=0\\ \beta+(1-\alpha-\beta)\frac{a_1\theta}{g(\theta)}, & z=1\\ (1-\alpha-\beta)\frac{a_z\theta^z}{g(\theta)}, & z=2,3,4,\ldots\\ 0 & \text{otherwise,}\end{cases} \qquad (3)$$

and is zero otherwise, is said to have a ZOIPSD with sdf $g(\theta)$, and inflated parameters α and β. We will denoted that by writing$Z\sim \mathrm{ZOIPSD}(\theta, g(\theta);\alpha,\beta)$.
Note that, if $\beta\rightarrow 0$, then (3) reduces to the form of the ZIPSD. Similarly, the case with $\alpha\rightarrow 0$ and $\beta\rightarrow 0$, reduces to the standard case of PSD.
Although, it does not fit the nature of the supposed model, the inflation parameters α and β may also take negative values providing that $\alpha\in\left(\max\left\{-1,-(1-\beta)\frac{p_0}{1-p_0}\right\},0\right)$ and $\beta\in\left(\max\left\{-1,-(1-\alpha)\frac{p_1}{1-p_1}\right\},0\right)$ without violating that (3) is a pmf. This situation represents the excluding proportion of zero's and one's, respectively, from the standard model given by (1).

V. MOMENTS AND SOME STRUCTURAL PROPERTIES

Let the rvs $X\sim \mathrm{PSD}(\theta, g(\theta))$, $Y\sim \mathrm{ZIPSD}(\theta, g(\theta);\alpha)$ and $Z\sim \mathrm{ZOIPSD}(\theta, g(\theta);\alpha,\beta)$, then the mean of the rv Z is given by;

$$E(Z)=\beta+(1-\alpha-\beta)\,\theta\frac{g'(\theta)}{g(\theta)} \qquad (4)$$

Note that;

$$E(Z)=E(Y)+\beta\left[1-\theta\frac{g'(\theta)}{g(\theta)}\right]$$
$$E(Z)=E(Y)+\beta[1-E(X)]$$

Similarly,

$$E[Z(Z-1)]=(1-\alpha-\beta)\theta^2\frac{g''(\theta)}{g(\theta)}$$

Hence, the variance Z is given by;

$$\mathrm{Var}(Z)=\beta(1-\beta)+(1-\alpha-\beta)\theta\left\{\theta\frac{g''(\theta)}{g(\theta)}+(1-2\beta)\frac{g'(\theta)}{g(\theta)}-(1-\alpha-\beta)\theta\left[\frac{g'(\theta)}{g(\theta)}\right]^2\right\}$$

The PGF of Z is given by;

$$G_Z(s)=E[s^Z]=\alpha+\beta s+(1-\alpha-\beta)\frac{g(\theta s)}{g(\theta)}$$
$$=G_Y(s)+\beta(s-1)\frac{g(\theta s)}{g(\theta)}$$

The MGF of the rv Z is given by;

$$M_Z(t)=E[e^{tZ}]=\alpha+\beta e^t+(1-\alpha-\beta)\frac{g(\theta e^t)}{g(\theta)}$$
$$=M_Y(t)+\beta\left[e^t-\frac{g(\theta e^t)}{g(\theta)}\right]$$

Similarly, the FMGF of Z can be shown to be;

$$M_{[Z]}(t)=E([1+t]^Z)$$
$$=\alpha+\beta(1+t)+(1-\alpha-\beta)\frac{g(\theta+\theta t)}{g(\theta)}$$
$$=M_Y(t)+\beta\left[(1+t)-\frac{g(\theta+\theta t)}{g(\theta)}\right]$$

Let μ_r' be the rth moment of the rv Z then for r=1, 2, 3, …;

$$\mu_r'=E(Z^r)=\beta+(1-\alpha-\beta)\sum_{z=1}^{\infty}z^r\frac{a_z\theta^z}{g(\theta)}$$

Therefore,

$$\mu_r'-\beta=\sum_{z=1}^{\infty}(1-\alpha-\beta)z^r\frac{a_z\theta^z}{g(\theta)} \qquad (5)$$

It follows that;

$$\frac{d}{d\theta}\mu_r^{'} = (1-\alpha-\beta)\sum_{z=1}^{\infty} z^r a_z \frac{d}{d\theta}\{\theta^z[g(\theta)]^{-1}\}$$

$$= \frac{1}{\theta}\sum_{z=1}^{\infty} z^{r+1}\,(1-\alpha-\beta)\frac{a_z\theta^z}{g(\theta)} - \frac{g'(\theta)}{g(\theta)}\sum_{z=1}^{\infty} z^r\,(1-\alpha-\beta)\frac{a_z\theta^z}{g(\theta)}$$

And using (5), we have that;

$$\frac{d}{d\theta}\mu_r^{'} = \frac{1}{\theta}\left(\mu_{r+1}^{'} - \beta\right) - \frac{g^{'}(\theta)}{g(\theta)}\left(\mu_r^{'} - \beta\right) \qquad (6)$$

It follows, with the use of (4), that (6) can be written as;

$$\mu_{r+1}^{'} = \beta + \theta\frac{d}{d\theta}\mu_r^{'} + \left(\frac{\mu_1^{'} - \beta}{1-\alpha-\beta}\right)\left(\mu_r^{'} - \beta\right)$$

VI. CONCLUSIONS

We introduced the zero-one inflated generalized power series distributions as an extension to class of zero-inflated generalized power series distributions. Its structural properties were studied, namely, its mean, variance, probability generating function, moment generating function, and the factorial moment generating functions. These results were shown in term of the zero-inflated cases as well. Its moment's recurrence relations were obtained also. Further studies for this class of distributions need to done, namely, estimation and testing of its parameters.

REFERENCES

[1]. Gupta, P. L., Gupta, R. C. and Tripathi, R. C. (1995). Inflated modified power series distributions with applications. Communications in Statistics - Theory and Methods, 24(9), 2355-2374.

[2]. Ridout, M., Dem´etrio, C. G. B. and Hinde, J. (1998). Models for count data with many zeros. International Biometric Conference, Cape Town, December 1998.

[3]. 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.

[4]. 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.

[5]. Murat, M. and Szynal, D. (2007). Non–zero inflated modified power series distributions. Communications in Statistics - Theory and Methods, 27(12), 3047-3064.

[6]. Bhattacharya, A., Clarke, B. S., and Datta, G. S. (2008). A Bayesian Test For Excess Zeros in a Zero-Inflated Power Series Distribution. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen,, Institute of Mathematical Statistics, Beachwood, Ohio, USA, 89-104

[7]. 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.

[8]. Edwin, T. K. (2014) Power series distribution sand zero-inflated models. Ph.D. Thesis, University of Nairobi.

[9]. Abdul-Razak, R. S. and Patil, G. P. (1986). Power series distributions and their conjugates in stochastic modeling and Bayesian inference. Communication in Statistics, Part l - Theory and Methods, 15(3), 623-641.

[10]. Abdul-Razak, R. S. and Patil, G. P. (1994). Some stochastic characteristics of the power series distributions. Pakistan Journal of Statistics, 10(1), 189-203.

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

[12]. Leroux, B. G., and Puterman,M. L. (1992). Maximum-Penalized-Likelihood Estimation For Independent And Markov-Dependent Mixture Models. Biometrics, 545-558.