ProbStat Forum, Volume 11, April 2018, Pages 77–80
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

Characterization of a Family of Zero-inflated Discrete Distributions

G.Nanjundan, Sadiq Pasha

Department of Statistics, Bangalore University, Bangalore 560 056, India.

Abstract. New distributions arise in literature when the existing ones become inadequate to model emerging situations. Zero-inflated discrete distributions are such examples. These have found fruitful applications in many real life situations in the recent past. Especially there is a sudden surge of applications of these distributions in count regression models.

1. Introduction

Let $p_1(x,\theta)$ be a probability mass function (pmf) with the support on the set of nonnegative integers and $p_0(x)$ be the pmf of a distribution degenerate at 0. Here θ may be real or a vector. If a random variable X has the pmf given by

$$p(x,\theta,\varphi) = \begin{cases} \varphi + (1-\varphi)p_1(x,\theta), & x = 0 \\ (1-\varphi)p_1(x,\theta), & x = 1,\ 2,\ 3, \ldots \end{cases} \tag{1}$$

where $0 < \varphi < 1$, then X has a zero-inflated distribution. Note that $p(x,\theta,\varphi) = \varphi p_0(x) + (1-\varphi)p_1(x,\theta)$ is a mixture of $p_0(x)$ and $p_1(x,\theta)$.

If $p_1(x,\theta)$ is the pmf of a binomial distribution then X is said to have a zero-inflated binomial distribution. Similarly, we have zero-inflated Poisson and zero-inflated negative binomial distributions. Also, let us denote that $X \sim ZIB(n,p,\varphi)$ or $X \sim ZIP(\theta,\varphi)$ or $X \sim ZINB(r,p,\varphi)$ according as $p_1(x,\theta)$ is the pmf of binomial (n,p) or Poisson (θ) or negative binomial (r,p) distribution.

If a random variable X has the pmf specified in (1) with $p_1(x,\theta) = \frac{a(x)\theta^x}{g(\theta)}$, $x \in N$, $\theta > 0$ where $N = \{0,1,2,...\}$, then X is said to have a zero-inflated power series distribution. It is well known that binomial, Poisson, and negative binomial are particular cases of power series distribution.

Zero-inflated models are appropriate for data sets with excess zeros. Lambert (1992) has discussed the relevance of zero-inflated Poisson model in the context of a manufacturing process. If the manufacturing equipment is properly calibrated, the defects may be nearly impossible. On the other hand, when it

2010 *Mathematics Subject Classification.* 60F05, 60G50.
Keywords. Zero-inflated binomial, zero-inflated Poisson, zero-inflated negative binomial distributions, probability generating function, linear differential equation.
Received: 08 October 2016; Revised: 31 May 2018, Accepted: 10 June 2018
Financial support may be acknowledged
Email addresses: nanzundan@gmail.com (G.Nanjundan), sadiqbujc@gmail.com (Sadiq Pasha)

is miscalibrated, the defects may occur according to a Poisson law. Hence the number of defects of the manufacturing process follows a zero-inflated Poisson model. Also, Xie and Goh (1993) have shown that a zero-inflated Poisson model is more appropriate than the regular Poisson model for a near zero-defect manufacturing environment.

Zero-inflated models have been found appropriate for insurance claim count data too. If a general insurance policy is under the usual detectable agreement, a claim will not be recorded and honored unless the loss exceeds a prescribed detectable limit. Also, under the no claim discount, which is widely adopted by the automoblie insurance, the policy holders may seldom claim if their losses are small. These lead to excess zero claims. Yip and Yau (2005) have used ZIP and ZINB distributions to model insurance claim count data having excess zeros. Further, they have adopted Pearson's $\chi 2$ statistic, Akaike information and Bayesian information criteria for testing goodness of fit and model selection.

In insurance, the hunger for bonus is a well known phenomenon. Under this phenomenon, the insureds do not report all of their losses in order to gain their bonus for next year's premiums. Boucher et al. (2009) have employed zero-inflated distribuitions to model insurance panel count data. They have also illustrated zero-inflated models based on the data on the claims reported to a Spanish insurance company.

Nanjundan (2011) has characterized a subfamily of power series distributions through a differential equation satisfied by the probability generating functions (pgfs) of the distributions. Nanjundan and Sadiq Pasha (2015a, 2015b) have characterized zero-inflated Poisson and zero-inflated binomial distributions through a linear differential equation. Along the same lines, Suresh et al (2015) have identified a linear differential equation that characterizes the zero-inflated negative binomial distribution.

A subfamily of zero-inflated discrete distributions is characterized in this paper by a linear differential equation. Let $f(s)$, $0 \leq s \leq 1$, denote the pgf of X. It can be readily seen that when
(i) $X \sim ZIB(n, p, \varphi)$, $f(s) = \varphi + (1 - \varphi)(q + ps)^n$,
(ii) $X \sim ZIP(\theta, \varphi)$, $f(s) = \varphi + (1 - \varphi)e^{\theta(s-1)}$, and
(iii)$X \sim ZINB(r, p, \varphi)$, $f(s) = \varphi + (1 - \varphi)\frac{p^r}{(1-qs)^r}$.

2. Characterization

The following theorem characterizes a subfamily of zero-inflated power series distributions.

Theorem 2.1. *Let X be a nonnegative integer valued random variable with $P(X = k) = p_k, k = 0, 1, 2, ...$ and $p_k > 0$ at least for $k = 0, 1$. The pgf $f(s)$ of X satisfies*

$$f(s) = a + b(c + ds)f^{'}(s) \tag{2}$$

where $0 < a < 1, b, c, d$ are constants and $f^{'}(s)$ is the first derivative of $f(s)$, if and only if the distribution of X is zero-inflated-Poisson, binomial, or negative binomial.

Proof. It is straight forward to verify that the pgfs of these zero-inflated distributions satisfy (2).

i) If $X \sim ZIP(\theta, \varphi)$, then its pgf $f(s)$ satisfies (2) with $a = \varphi$, $b = 1/\theta$, $c = 1$, and $d = 0$.
ii) When $X \sim ZIB(n, p, \varphi)$, then its pgf $f(s)$ satisfies (2) with $a = \varphi$, $b = 1/np$, $c = q$, and $d = p$.
iii) If $X \sim ZINB(r, p, \varphi)$, then its pgf $f(s)$ satisfies (2) with $a = \varphi$, $b = 1/rq$, $c = 1$, $d = -q$.

Suppose that the pgf $f(s)$ of X satisfies the linear differential equation specified in (2). Now let us have a close look at the possible values of b, c, d and their consequences.

1) If $b = 0$, then we get $f(s) = a$, $\forall s \in [0, 1]$. Hence $f(1) = 1 = a$, which is not possible because

$0 < a < 1$. Therefore $b \neq 0$.

2) Let $b \neq 0$ and $c = 0$, then (2) reduces to $f(s) = a + bdsf'(s)$, $\forall s \in [0, 1]$. We get $f'(s) = \frac{f(s)-a}{bds}$ and $f'(0) = p_1 = \infty$, which is not admitted. Hence $c \neq 0$.

3) Let $b, c \neq 0$. If $d = 0$, then (2) becomes $f(s) = a + bcsf'(s)$, $0 \leq s \leq 1$.

The solution of this linear differential equation is given by

$$f(s) = a + ke^{\frac{s}{bc}}, o \leq s \leq 1,$$

where k is an arbitrary constant. Since $f(1) = 1$, $a + ke^{\frac{1}{bc}} = 1$. This implies that $k = (1-a)e^{\frac{-1}{bc}}$. Therefore, $f(s) = a + (1-a)e^{\frac{1}{bc}(s-1)}$, $0 \leq s \leq 1$. If $bc < 0$, then $f(0) = p_0 > 1$ which is not possible and hence $bc > 0$. Hence X has the pgf of $ZIP \sim (\frac{1}{bc}, a)$.

4) Let $b \neq 0$, $c \neq 0$, $d \neq 0$, then by solving the differential equation (2), we get $f(s) = a + k(c + ds)^{\frac{1}{bd}}, 0 \leq s \leq 1$, where k is an arbitrary constant. Since $f(1) = a + k(c + d)^{\frac{1}{bd}} = 1$, we get $k = (1-a)\frac{1}{(c+d)^{\frac{1}{bd}}}$. Therefore $f(s) = a + (1-a)(\frac{c+ds}{c+d})^{\frac{1}{bd}}$. If $c + d = 0$, then $f(s)$ does not define a pgf. Hence $c + d \neq 0$.

Further, (2) can be expressed as

$$f(s) = a + (1-a)(c^* + d^*s)^{\frac{1}{bd}}, \tag{3}$$

where $c^* = \frac{c}{c+d}$, $d^* = \frac{d}{c+d}$, and $c^* + d^* = 1$. Since $0 < a < 1$ and $0 < f(0) = a + (1-a)(c^*)^{\frac{1}{bd}} = po < 1$, $0 < c^* < 1$ and hence $0 < d^* < 1$. In turn c, $d > 0$. Thus case (4) reduces to $b \neq 0$ and c, $d > 0$.

4a) Let c, $d > 0$ and $b > 0$. Take $\frac{1}{bd} = N > 0$ to be an integer. Then,

$$f(s) = a + (1-a)(c^* + d^*s)^N,$$

which is the pgf of $ZIB(N, d^*, a)$.

4b) Let c, $d > 0$ and $b < 0$. Take $\frac{1}{bd} = N < 0$ to be an integer. Then, $f(s)$ turns out to be the pgf of $ZINB(N, c^*, a)$.

Now it remains to verify whether $\frac{1}{bd}$ can be a fraction with $b \neq 0$. Note that $f(s)$ in (3) can be expressed as

$$f(s) = a + (1-a)(c^*)^{\frac{1}{bd}}(1 + \frac{d^*}{c^*}s)^{\frac{1}{bd}}. \tag{4}$$

The expansion of the factor $(1 + \frac{d^*}{c^*}s)^{\frac{1}{bd}}$ in the RHS of (4) is a power series in s with some terms being negative when $\frac{1}{bd}$ is a fraction. Since $(1 + \frac{d^*}{c^*}s)^{\frac{1}{bd}}$ is a part of a pgf, $\frac{1}{bd}$ cannot be a fraction. This completes the proof of the theorem. $\square$

Acknowledgments

The authors are very grateful to Dr. R. Vasudeva and anonymous referee for their fruitful suggestions.

References

Boucher, J. P., Denuit, M., and Guillen, M. (2009). Number of accidents or number of claims? An approach with zero-inflated Poisson models for panel data, *The Journal of Risk and Insurance*, Vol. 76, No. 4, 821 - 846.

Lambert, D. (1992). Zero-inflated Poisson regression with an application to defects in manufacturing, *Technometrics*, Vol. 34, No. 1, 1 - 14.

Nagesh, S., Nanjundan, G., Suresh, R., and Sadiq Pasha (2015). A characterization of zero-inflated geometric model, *Int. Journal of Mathematics Trends and Technology*, Vol. 23, No. 1, 71 - 73.

Nanjundan, G. (2011). A characterization of the members of a subfamily of power series distributions, *Applied Mathematics*, Vol. 2, 750 - 51.

Nanjundan, G. and Sadiq Pasha (2015a). A note on the characterization of zero-inflated Poisson distribution, *Open Journal of Statistics*, Vol. 5, 140 - 142.

Nanjundan, G. and Sadiq Pasha (2015b). A characterization of zero-inflated binomial model, *Int. Journal of Mathematics and Computer Research*, Vol. 3, No. 11, 1187 - 1190.

Suresh, R., Nagesh, S., Nanjundan G., and Sadiq Pasha (2015).On a characterization of a zero-inflated negative binomial distribution, *Open Journal of Statistics*, Vol. 5, 511 - 513.

Xie, M. and Goh, T. N. (1993). SPC of a near zero-defect process subject to random shocks, *Quality and Reliability Engineering International*, Vol. 9, 89 - 93.

Yip, K. C. H. and Yau, K. K. W. (2005). On modeling claim frequency data in general insurance with extra zeros, *Insurance: Mathematics and Economics*, Elsevier, Vol. 36, 153 - 163.