Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal. Vol. 3, Issue 1 (2010), 65 – 74
ISSN 2070-5948, DOI 10.1285/i20705948v3n1p65

SOME ASPECTS OF ZERO-MODIFIED DISTRIBUTIONS OF ORDER *k*

Rupak Gupta[1*], Kishore K. Das[2]

[1]Faculty of Science & Technology, The Icfai university, Tripura, India
[2]Department of Statistics, Gauhati University, Guwahati, Assam, India

Received 14 September 2009; Accepted 16 January 2010
Available online 28 February 2010

Abstract: *Under certain simplifying assumptions, zero-modified distribution of order k has been introduced and defined. In this paper, an attempt is made to study different zero-modified distributions of order k such as Binomial, Poisson, Geometric, Negative Binomial and logarithmic series. Also an inflated generalized Poisson distribution of order k has been defined and studied. Some of their properties also have been discussed.*

Keywords: *Zero-modified distribution, Zero-modified distributions of order k, Zero-modified Binomial, Poisson, Negative binomial, Geometric, Logarithmic series distributions of order k, inflated distribution of order k.*

Mathematics subject classification: *62E15, 62E17.*

1. Introduction

Empirical distributions obtained in the course of experimental investigations often have an excess of zeroes compared with a Poisson distribution with the same mean. This has been a major motivating force behind the development of many distributions that have been used as models in applied statistics. The phenomenon can arise as the result of clustering ; distributions with clustering interpretations often do indeed exhibit the feature that the proportion of observations in the zero class in greater than $e^{\bar{x}}$, where $\bar{x}$ is the observed mean.
A very simple alternative to use of a cluster model is just to add an arbitrary proportion of zeroes, decreasing the remaining frequencies in the appropriate manner. Thus a combination of the original distribution with probability mass function (pmf) P_x, $x = 0,1,2,.....$ together with the degenerate distribution with all probability concentrated at the origin, and gives a finite mixture distribution with:

* Corresponding Author. Email: rupak_icfaitech@yahoo.co.uk

$$P_r[X = 0] = w + (1-w)P_0$$

$$P_r[X = x] = (1-w)P_x \; ; \quad x \geq 1$$

A mixture of this kind is referred to as a zero-modified distribution or as a distribution with added zeroes. Another epithet is inflated distribution [14 & 17].
The probability function (pf) of the generalized Poisson distribution is defined for $a > 0$, $|\lambda| < 1$ [Consul and Jain, 1973] as:

$$P(x) = \begin{cases} \frac{1}{x!} a(a + x\lambda)^{x-1}; (a + x\lambda) > 0, x = 0,1,2,... \\ 0; Otherwise \end{cases}$$

The probability function (pf) of a class of generalized Poisson distribution for $a > 0, |\lambda| < 1$ [13] is :

$$P(x) = \frac{a + x\lambda \;\; ^{x+s} e^{-x\lambda}}{x! K \;\; a; s; \lambda}; \quad a + x\lambda \;\; > 0, x = 0,1,2,...$$

where $K \;\; a; s; \lambda \;\; = \sum_{x \geq 0} \frac{1}{x!} \;\; a + x\lambda \;\; ^{x+s} \; e^{-x\lambda}$

The probability generating function (pgf) of a class of generalized Poisson distribution [8] is:

$$G(u) = \frac{K \;\; at; s; \lambda t}{t^s K \;\; a; s; \lambda}$$, where $t = ue^{\lambda \;\; t-1}$

This distribution attempt to take into account errors in recording a variable which in reality does have a generalized Poisson distribution. Suppose that the zero-class alone is misrecorded. Then the probability mass function is:

$$P_r[X = 0] = w + (1-w)e^{-a}$$

$$P_r[X = x] = (1-w)\frac{a(a + x\lambda)^{x-1} e^{-(a+x\lambda)}}{x!}; x \geq 1$$

where $0 < w < 1$, i.e., there is over reporting. The distribution is known as the generalized Poisson with zeroes or the zero inflated generalized Poisson [12]. The Probability generating function is therefore:

$$H(u) = w + (1-w)e^{a(t-1)} \quad ; \quad \text{where } t = ue^{\lambda \;\; t-1}$$

Feller (1968) introduced the discrete distribution of order *k*, when he extended the notion of success to a success run of length *k*. A *Run* is usually defined as an uninterrupted sequence of like symbols ('S' or 'F'). Thus the distributions associated with runs of *k* like outcomes are distribution of order *k*.

The only explicit relationship between distributions of order *k* and that of order 1 is that "*Every distribution of order 1 is the usual corresponding discrete distribution"*.

2. Zero-modified distributions of order k

2.1 Zero-modified Binomial distribution of order k

Hirano (1986) and Philippou and Makri (1986) gave exactly the probability function of Binomial distribution of order *k*, denoted by $B_k(n,p)$ as:

$$B_k\ n,p;x\ = P(X=x) = \sum_{m=0}^{k-1}\ \sum_{x_1+2x_2+...+kx_k=n-m-kx} \binom{x_1+x_2+...+x_k+x}{x_1,x_2,...,x_k,x} p^n \left(\frac{q}{p}\right)^{x_1+x_2+...+x_k}; \quad \text{for } x=0,1,2,...,\left[\frac{n}{k}\right] \qquad (1)$$

Some properties of the binomial distribution of order *k* was studied by Feller (1968), Hirano (1986), Philippou and Makri (1986) and Aki and Hirano (1988). Aki and Hirano (1989) gave the calculation of the probability function, the first and the second derivatives of the probability function with respect to the parameter which is necessary for getting the MLE based on independent observations. Putting $x=0$ in (1), we get:

$$B_k\ n,p;0\ = \sum_{m=0}^{k-1}\ \sum_{x_1+2x_2+...+kx_k=n-m} \binom{x_1+x_2+...+x_k}{x_1,x_2,...,x_k} p^n \left(\frac{q}{p}\right)^{x_1+x_2+...+x_k}$$
$$= I_0 \text{, say}$$

and

$$B_k\ n,p;x\ = \sum_{m=0}^{k-1}\ \sum_{x_1+2x_2+...+kx_k=n-m-kx} \binom{x_1+x_2+...+x_k+x}{x_1,x_2,...,x_k,x} p^n \left(\frac{q}{p}\right)^{x_1+x_2+...+x_k} \quad ; \text{ for } x=0,1,2,...,\left[\frac{n}{k}\right]$$
$$= I \text{, say}$$

So, a finite mixture of the distribution having the probability mass functions as:

$$P_r[X=0]=B_k(n,p;0)=w+(1-w)I_0$$

$$P_r[X=x]=B_k(n,p;x)=(1-w)I;\ x=1,2,3,...,\left[\frac{n}{k}\right]$$

is known as zero-modified binomial distribution of order *k*.

2.2 Zero-modified Poisson distribution of order k

Philippou (1983) introduced Poisson distribution of order *k* denoted by $P_k(\lambda;x)$ with the probability function as:

$$P_k(\lambda,x)=P\left[X=x\right]=\sum_{x_1,x_2,..,x_k} e^{-k\lambda}\frac{\lambda^{x_1+x_2+..+x_k}}{x_1!x_2!..x_k!};x=0,1,2,... \tag{2}$$

where the summation is over all non-negative integers $x_1,x_2,...,x_k$ such that $x_1+2x_2+...+kx_k=x$. He also showed that $\sum_{x=0}^{\infty}P(X=x)=1$ and studied some of its characteristics. Putting $x=0$ in (2), we have:

$$P_k(\lambda,0)=e^{-k\lambda}=I_0\text{, say}$$

$$P_k(\lambda,x)=\sum_{x_1+2x_2+..+kx_k=x} e^{-k\lambda}\frac{\lambda^{x_1+x_2+..+x_k}}{x_1!x_2!..x_k!};\ x=0,1,2,...$$
$$=I\text{, say}$$

Now a finite mixture of the distribution having the probability mass functions as:

$$P_r[X=0]=P_k(\lambda;0)=w+(1-w)I_0$$

$$P_r[X=x]=P_k(\lambda;x)=(1-w)I;\ x=1,2,3,...$$

is known as zero-modified Poisson distribution of order *k*.

2.3 Zero-modified Geometric distribution of order k

Let *k* be a positive integer. Suppose we are given independent trials with success probability *p*. The distribution of the number of trials until the first occurrence of the *k*-th consecutive success is called the geometric distribution of order *k* and is denoted by $G_k(p)$. Philippou et al. (1983) called it the geometric distribution of order *k* and derived its exact probability function. Since then exact distribution theory for so called discrete distributions of order *k* has been extensively developed. They also showed that the mean of $G_k(p)$ is monotonously decreasing and hence obtained the moment estimate which was determined uniquely. The probability function of $G_k(p)$ is given by:

$$G_k(p;x) = P \quad X = x \quad = \sum_{x_1+2x_2+...+kx_k=x-k} \binom{x_1+x_2+...+x_k}{x_1, x_2, ..., x_k} p^x \left(\frac{q}{p}\right)^{x_1+x_2+...+x_k}$$

$$; x = k, k+1, k+2, ...$$

$$= I \text{, say}$$

Putting $x = k$ we get:

$$G_k(p;k) = p^k = I_k \text{, say}$$

So, a finite mixture of the distribution having the probability mass functions as:

$$G_k(p;k) = w + \quad 1-w \quad I_k$$

$$G_k(p;x) = \quad 1-w \quad I \qquad ; \qquad x = k+1, k+2, ...$$

is known as zero-modified Geometric distribution of order *k.*

2.4 Zero-modified Negative binomial distribution of order k

The *type I negative binomial distribution of order k* is also called the *type I waiting time distribution of order k* . It is the waiting-time distribution for *b* runs of successes of length *k*, given non-overlapping counting, the pattern of the successes and failures that have occurred becomes irrelevant and counting begins all over again. This distribution is the *b*-fold convolution of geometric distribution of order *k* (Philippou, 1984) and therefore has the probability mass function:

$$NB_k(r, p; x) = P \quad X = x \quad = \sum_{x_1+2x_2+...+kx_k=x-kr} \binom{x_1+x_2+...+x_k+r-1}{x_1, x_2, ..., x_k, r-1} p^x \left(\frac{q}{p}\right)^{x_1+x_2+...+x_k}$$

$$; x = kr, kr+1, ...$$

$$= I \text{, say}$$

Substituting $x = kr$ in the above expression, we have:

$$NB_k(r, p; kr) = p^{kr} = I_k \text{, say}$$

Therefore, a finite mixture of the distribution having the probability mass functions as:

$$NB_k(r, p; kr) = w + \quad 1-w \quad I_{kr}$$

$$NB_k(r, p; kr) = \quad 1-w \quad I \qquad ; \quad x = kr+1, kr+2, ...$$

2.5 Zero-modified logarithmic series distribution of order k

A *logarithmic series distribution of order k* is obtained as a limiting form of a left-truncated type I negative binomial distribution of order *k*; the probability mass function can be expressed as:

$$LS_k(p;x) = P\left(X = x\right) = \sum_{x_1+2x_2+\ldots+kx_k=x} \frac{\left(x_1+x_2+\ldots+x_k-1\right)!}{-k\log p\left(x_1!x_2!\ldots.x_k!\right)} p^x \left(\frac{q}{p}\right)^{x_1+x_2+\ldots+x_k} \quad ;\; x=1,2,\ldots.$$

$$= I_k\text{, say}$$

Substituting $x = 1$ in the above expression, we have:

$$LS_k(p;1) = \frac{1-p}{-k\log p} = I_0\text{, say}$$

Therefore, a finite mixture of the distribution having the probability mass function as:

$$LS_k(p;1) = w + \left(1-w\right) I_k$$

$$LS_k(p;x) = \left(1-w\right) I_0 \;;\quad x = 2,3,4,\ldots..$$

2.6 A class of Inflated Generalized Poisson distribution of order k

Gupta et al. (2008) defined Generalized Poisson distribution of order *k* having the pf as:

$$P(X=x) = \sum_{\substack{x_1,x_2,\ldots,x_k \ni \\ x_1+2x_2+\ldots+kx_k=x}} \frac{e^{-k[a+\lambda\sum_{i=1}^{k}x_i]} a[a+\lambda\sum_{i=1}^{k}x_i]^{\sum_{i=1}^{k}x_i-1}}{\prod_{i=1}^{k}x_i} \;;\; [a+\lambda\sum_{i=1}^{k}x_i] > 0, a>0, |\lambda|<1 \quad ;\quad x=0,1,\ldots \tag{3}$$

and the pf of a class of generalized Poisson distribution of order *k* has the pmf:

$$P(X=x) = \sum_{\substack{x_1,x_2,\ldots,x_k \ni \\ x_1+2x_2+\ldots+kx_k=x}} \frac{e^{-\lambda\sum_{i=1}^{k}x_i} [a+\lambda\sum_{i=1}^{k}x_i]^{\sum_{i=1}^{k}x_i+s}}{\prod_{i=1}^{k}x_i K(a;s;\lambda)}$$

where, $$K(a;s;\lambda) = \sum_{x_1,x_2,\ldots,x_k\ge 0} \frac{e^{-\lambda\sum_{i=1}^{k}x_i} [a+\lambda\sum_{i=1}^{k}x_i]^{\sum_{i=1}^{k}x_i+s}}{\prod_{i=1}^{k}x_i} \tag{4}$$

Definition 1: The probability function of a class of Inflated Generalized Poisson distribution of order *k* has the following form:

$$P_r[X=0]=w+(1-w)\frac{a^s}{K(a;s;\lambda)}$$
$$P_r[X=x]=(1-w)\frac{e^{-\lambda\sum_{i=1}^{k}x_i}[a+\lambda\sum_{i=1}^{k}x_i]^{\sum_{i=1}^{k}x_i+s}}{\prod_{i=1}^{k}x_iK(a;s;\lambda)}\quad;\ \sum_{i=1}^{k}x_i\geq 1 \tag{5}$$

where $K(a;s;\lambda)$ is defined in (4) is a general form of a class of exponential sums of order *k.* If we put *k* =1 in this expression, we get a class of exponential sums [13].
The probability generating function of a class of Inflated Generalized Poisson distribution of order *k* is:

$$H(u)=w+(1-w)\frac{K(at;s;\lambda t)}{t^sK(a;s;\lambda)} \tag{6}$$

where $K(a;s;\lambda)$ is defined in (4) and

$$K(at;s;\lambda t)=\sum_{x_1,x_2,\dots,x_k\geq 0}\frac{e^{-\lambda t\sum_{i=1}^{k}x_i}[a+\lambda t\sum_{i=1}^{k}x_i]^{\sum_{i=1}^{k}x_i+s}}{\prod_{i=1}^{k}x_i} \tag{7}$$

Putting different values of '*s'* in (5) and (6) we may obtain different forms of Inflated GPDs of order *k* and their probability generating functions. A few forms of Inflated GPDs of order *k* and their pgfs are given below:

(*i*) Substituting $s=-1$, the pf (5) reduces to the pf of Inflated or Zero-modified GPD I of order *k*:

$$P\ X=0\ =w+\ 1-w\ \frac{a^{-1}}{k\ \ a;-1;\lambda}$$

$$P\ X=x\ =\ 1-w\ \frac{\left(a+\lambda\sum_{i=1}^{k}x_i\right)^{\sum_{i=1}^{k}x_i-1}e^{-\lambda\sum_{i=1}^{k}x_i}}{\prod_{i=1}^{k}x_i!k\ \ a;-1;\lambda};\qquad \sum_{i=1}^{k}x_i\ \geq\ 1$$

where $k(a;-1;\lambda)=\sum_{x_1,x_2,\ldots,x_k\geq 0}\frac{1}{\prod_{i=1}^{k}x_i!}\left[a+\lambda\sum_{i=1}^{k}x_i\right]^{\sum_{i=1}^{k}x_i-1}e^{-\lambda\sum_{i=1}^{k}x_i}$

and the probability generating function of Inflated GPD I of order k is:

$$H(u)=w+(1-w)e^{a(t-1)}$$

And accordingly the mean and variance are:

Mean = $(1-w)\dfrac{a}{1-\lambda}$

Variance = $(1-w)\dfrac{a+(1-\lambda)a^2w}{(1-\lambda)^3}$

(*ii*) Assuming $s=-2$ in (5), we can obtain the probability generating function of Inflated or Zero-modified GPD II of order *k*:

$$P(X=0)=w+(1-w)\frac{a^{-2}}{k(a;-2;\lambda)}$$

$$P(X=x)=(1-w)\frac{\left(a+\lambda\sum_{i=1}^{k}x_i\right)^{\sum_{i=1}^{k}x_i-2}e^{-\lambda\sum_{i=1}^{k}x_i}}{\prod_{i=1}^{k}x_i!\,k(a;-2;\lambda)}; \qquad \sum_{i=1}^{k}x_i\geq 1$$

where $k(a;-2;\lambda)=\sum_{x_1,x_2,\ldots,x_k\geq 0}\frac{1}{\prod_{i=1}^{k}x_i!}\left[a+\lambda\sum_{i=1}^{k}x_i\right]^{\sum_{i=1}^{k}x_i-2}e^{-\lambda\sum_{i=1}^{k}x_i}$

Therefore the probability generating function of Inflated GPD II of order *k* is:

$$H(u)=w+(1-w)e^{a(t-1)}\frac{a+\lambda-a\lambda t}{a+\lambda-a\lambda}$$

Hence, $E(X)=(1-w)\dfrac{a^2}{a+\lambda-a\lambda}$

and $Var(X) = \dfrac{1-w}{(1-\lambda)(a+\lambda-a\lambda)^2}\left[a^3+a^2\lambda+wa^4(1-\lambda)\right]$

(iii) Considering $s=0$ in (5), we may obtain the probability function of Inflated GPD II of order k:

$$P(X=0) = w + (1-w)\frac{1}{k(a;0;\lambda)}$$

$$P(X=x) = (1-w)\frac{\left(a+\lambda\sum_{i=1}^{k}x_i\right)^{\sum_{i=1}^{k}x_i} e^{-\lambda\sum_{i=1}^{k}x_i}}{\prod_{i=1}^{k}x_i!\,k(a;0;\lambda)}$$

where $k(a;0;\lambda) = \sum_{x_1,x_2,\ldots,x_k\ge 0}\dfrac{1}{\prod_{i=1}^{k}x_i!}\left[a+\lambda\sum_{i=1}^{k}x_i\right]^{\sum_{i=1}^{k}x_i} e^{-\lambda\sum_{i=1}^{k}x_i}$

and the probability generating function of GPD II of order k is given by:

$$H(u) = w + (1-w)\frac{(1-\lambda)e^{a(t-1)}}{1-\lambda t}$$

And hence, $E(X) = (1-w)\dfrac{a+\lambda-a\lambda}{(1-\lambda)^2}$

$$Var(X) = \frac{1-w}{(1-\lambda)^4}\left[(a+\lambda-a\lambda)+2\lambda^3+\lambda^2+w(a+\lambda-a\lambda)^2\right]$$

3. Discussion and Conclusions

In this paper discussions have been made about some zero modified distributions of order k and also an Inflated Poisson distribution of order k has been studied along with some properties.

Acknowledgement

The authors are highly thankful to the editor and two anonymous referees for their valuable suggestions and critical comments

References

[1]. Aki, S. Hirano, K. (1988). Some characteristics of the binomial distribution of order k and related distributions. *Statistical Theory and Data Analysis II*, (ed. K. Matusita), 11-222, North Holand.

[2]. Aki, S., Kuboki, H., Hirano, K. (1984). On discrete distributions of order k. *Ann. Inst. Statist. Math.*, 36, 431-440.

[3]. Aki, S., Hirano, K. (1988). Some characteristics of the binomial distribution of order k and related distributions. Statistical Theory and Data Analysis II, in *2nd Pacific Area Statistical Conference*, eds. K. Matusita, Amsterdam: North-Holland, 211-222.

[4]. Aki, S., Hirano, K. (1989). Estimation of parameters in the discrete distributions of order k. *Ann. Inst. Statist. Math.*, 41, 47-61.

[5]. Aki, S., Hirano, K. (1994). Distributions of numbers of failures and successes until the consecutive k successes. *Ann. Inst. Statist. Math.*, 46, 193-202.

[6]. Consul, P.C., Jain, G.C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15 (4), 791-799.

[7]. Consul, P.C. (1989). *Generalized Poisson distributions: Properties and applications*. New York: Maecel Dekker.

[8]. Chakraborty, S. (2001). *Some aspects of discrete probability distribution*. Ph.D thesis, Tezpur University.

[9]. Feller, W. (1968). *An Introduction to probability theory and its applications*, Vol. I, 3rd edition. New York: Wiley.

[10]. Gupta, R., Das, K.K. (2008). Some Aspects of a class of Quasi Binomial distributions of order k. *The ICFAI University Journal of Computational Mathematics*, 1 (1), 53-59.

[11]. Hirano, K., (1986). Some properties of the distributions of order k, in *Fibonacci Numbers and Their Applications*, eds. A. N. Philippou, G. E. Bergum and A. F., Horadam, 43-53, Reidel, Dordrecht.

[12]. Johnson, N.L. Kemp, A.W. Kotz, S. (2005). *Univariate discrete distributions, third edition*. New York: John Wiley & Sons INC.

[13]. Nandi, S.B. Nath, D.C., Das, K.K. (1999). A class of generalized Poisson distributions. *Statistica*, anno LIX, n.4, 1999.

[14]. Pandey, K.N. (1965). On generalized inflated Poisson distribution. *Banaras Hindu University Journal of Research*, 15, 157 – 162.

[15]. Philippou, A.N., Georghiou, C., Philippou, G.N. (1983). A generalized Geometric distribution and some of its properties. *Statist. Probab. Lett.*, 1,171-175.

[16]. Philippou, A.N., Makri, F.S. (1986). Successes, runs, and longest runs. *Statistics & Probability Letters*, 4, 102 – 105.

[17]. Singh, S.N. (1963). A note on Inflated Poisson distribution. *Journal of the Indian Statistical Association*, 1, 140 – 144.