STATISTICA, anno LXXIII, n.4, 2013

ON ZERO-INFLATED LOGARITHMIC SERIES DISTRIBUTION AND ITS MODIFICATION

C. Satheesh Kumar

Department of Statistics, University of Kerala, Trivandrum-695 581, India.

E-mail: drcsatheeshkumar@gmail.com

A. Riyaz

Department of Statistics, University of Kerala, Trivandrum-695 581, India.

1. INTRODUCTION

The logarithmic series distribution (LSD), due to (Fisher *et.al.*, 1945), has been found applications in several areas of research such as biology, ecology, economics, inventory models, marine sciences etc. For details in this regard, refer (Johnson *et. al.*, 2005). Due to its practical suitability in case of data with long tails, several generalized forms of the LSD have been proposed in the literature. For example see Jain and Gupta (1973); Kempton (1975); Tripathi and Gupta (1985, 1988); Ong (2000) and Khang and Ong (2007). (Chatfield *et.al*, 1966) and Chatfield (1986) remarked that the LSD is likely to be a useful approximation to negative binomial distribution in certain cases. An important drawback of the LSD in certain practical situations is that it excludes the zero observation from its support. To mitigate this drawback, through this paper we consider a slightly modified form of LSD which has a non-negative support and call it as "the zero-inflated logarithmic series distribution (ZILSD)".

In section 2 we present the definition of the ZILSD and obtain some of its important properties. In section 3 we propose a modified version of the ZILSD which we named as "the modified zero-inflated logarithmic series distribution (MZILSD)" and derive expression for its probability generating function, probability mass function, recurrence relations for probabilities, raw moments and factorial moments. In section 4 we discuss the estimation of the parameters of the MZILSD and in section 5 we describe a test procedure for testing the additional parameter of the MZILSD by using generalized likelihood ratio test and Rao's score test. All these procedures are illustrated in section 6 with the help of certain real life data sets.

2. ZERO-INFLATED LOGARITHMIC SERIES DISTRIBUTION AND ITS PROPERTIES

In this section we present the definition and some properties of the zero-inflated logarithmic series distribution.

DEFINITION 2.1. *A non-negative integer valued random variable X is said to follow the zero inflated logarithmic series distribution (ZILSD) with parameter $\theta \in (0,1)$ if its probability mass function (pmf) $p_x = P(X = x)$ is the following, for $x = 0,1,2,...$*

$$p_x = \frac{\Lambda \theta^x}{x+1} \tag{2.1}$$

where $\Lambda = \theta\left[-\ln(1-\theta)\right]^{-1}$.

The probability generating function (pgf) of the ZILSD with pmf (2.1) is the following.

$$H(z) = -\Lambda z^{-1} \ln(1-\theta z), \tag{2.2}$$

which can also be written as

$$H(z) = \Lambda \, {}_2F_1(1, 1; 2; \theta z), \tag{2.3}$$

where

$${}_2F_1(a, b; c; z) = 1 + \sum_{r=1}^{\infty} \frac{(a)_r (b)_r}{(c)_r} \frac{z^r}{r!}$$

is the Gauss hypergeometric function, in which $(u)_r = \frac{\Gamma(u+r)}{\Gamma(r)} = u(u+1)\ldots(u+r-1)$ for $r \geq 1$. For details regarding Gauss hypergeometric function see Mathai and Haubold (2008) or Slater (1966). Note that the ZILSD with pmf (2.1) belongs to the generalized power series family of distributions. Since the pgf of ZILSD has representation given in (2.3), it is also a member of the Kemp family of distributions, studied by Kumar (2009). Hence from Kumar (2009) we have the following results.

RESULT 2.1. For $r \geq 1$, *an expression for rth factorial moment of ZILSD is*

$$\mu'_{[r]} = \frac{\Lambda r! \theta^r}{(r+1)} \, {}_2F_1(1+r, 1+r; 2+r; \theta). \tag{2.4}$$

Proof follows from the expression for rth factorial moment of Kemp family of distribution given in Kumar (2009).

RESULT 2.2. For $r \geq 1$, *an expression for r-th raw moment* μ'_r *of ZILSD is the following, in which* $S(r,m)$ *is the Stirling's number of the second kind.*

$$\mu'_r = \Lambda \sum_{m=0}^{r} \frac{m!}{(m+1)} S(r,m) \theta^m \, {}_2F_1(1+m, 1+m; 2+m; \theta) \tag{2.5}$$

Proof follows from equation (8) of Kumar (2009).

From the Result 3.2 and Result 3.3 of Kumar (2009) we obtain the following results. For $i = 0, 1, 2, \ldots$ we need the following notations for stating the results given below.

$$\underline{u} + i = (1+i, 1+i, 2+i), \tag{2.6}$$

$$D_i = \frac{(1+i)(1+i)}{(2+i)}, \tag{2.7}$$

and

$$\Lambda_i = {}_2F_1(1+i\,,1+i\,;2+i\,;\theta) \tag{2.8}$$

so that $\Lambda_0 = -\Lambda$.

RESULT 2.3. For $x \geq 0$, *an expression for raw moments of ZILSD is the following in which* $\mu'_x(\underline{u}) = 1$

$$\mu'_x(\underline{u}) = \theta D_0 \Lambda_0^{-1} \Lambda_1 \sum_{r=0}^{x} \binom{x}{r} \mu'_{x-j}(\underline{u}+1). \tag{2.9}$$

Proof follows from the Result 3.2 of Kemp family of distribution given in Kumar (2009).

RESULT 2.4. The following is a simple recurrence relation for factorial moments $\mu_{[x]}(\underline{u})$ *of the ZILSD, for* $x \geq 0$, *in which* $\mu_{[0]}(\underline{u}) = 1$.

$$\mu_{[x+1]}(\underline{u}) = \theta D_0 \Lambda_0^{-1} \Lambda_1 \mu_{[x]}(\underline{u}+1). \tag{2.10}$$

Proof follows from the Result 3.3 of Kemp family of distribution given in Kumar (2009).

RESULT 2.5. For $r \geq 0$, *the simple recurrence relation for probabilities of the ZILSD is the following.*

$$p_{x+1} = \theta \frac{x+1}{x+2} p_x \tag{2.11}$$

Proof is simple and hence omitted.

RESULT 2.6. The raw moments $\mu'_r = \mu'_r(\underline{u})$ *of the ZILSD satisfies the following relation:*

$$\mu'_{r+1} = \theta \frac{d\mu'_r}{d\theta} + \left[\frac{\Lambda}{1-\theta} - 1\right] \mu'_r \tag{2.12}$$

PROOF. By definition, we have

$$\mu'_r = \sum_{x=0}^{\infty} \frac{x^r}{-\ln(1-\theta)} \frac{\theta^{x+1}}{(x+1)}. \tag{2.13}$$

On differentiating both sides of (2.13) with respect to θ we get

$$\frac{d\mu_r'}{d\theta}=\sum_{x=0}^{\infty}\frac{x^r(x+1)}{-\ln(1-\theta)}\frac{\theta^x}{x+1}-\sum_{x=0}^{\infty}\frac{x^r}{[-\ln(1-\theta)]^2(1-\theta)}\frac{\theta^{x+1}}{x+1}$$
$$=\theta^{-1}\sum_{x=0}^{\infty}x^{r+1}p_x+\theta^{-1}\sum_{x=0}^{\infty}x^r p_x-\frac{1}{\left[-\ln(1-\theta)\right](1-\theta)}\sum_{x=0}^{\infty}x^r p_x$$
$$=\theta^{-1}\{\mu_{r+1}'+\left[1-\frac{\Lambda}{1-\theta}\right]\mu_r'\}, \tag{2.14}$$

which implies (2.12).

Now either by using (2.9) or by using (2.12) one can obtain the mean and variance of ZILSD, as given in the following result.

RESULT 2.7. *The mean and variance of ZILSD are the following:*

$$mean=\frac{\Lambda}{(1-\theta)}-1 \tag{2.15}$$

and

$$variance=\frac{\Lambda(\theta^2-\Lambda)}{(1-\theta)^2}+\frac{2\Lambda}{(1-\theta)}. \tag{2.16}$$

RESULT 2.8. *The central moments* μ_r *of the ZILSD satisfies the following relation.*

$$\mu_{r+1}=\theta\frac{d\mu_r}{d\theta}+\frac{r\Lambda(1-\Lambda)}{(1-\theta)^2}\mu_{r-1}. \tag{2.17}$$

PROOF. Let the random variable X follows the ZILSD with pmf (2.1). Now, by definition of central moments,

$$\mu_r=E\{[X-E(X)]^r\}$$
$$=\sum_{x=0}^{\infty}[x-E(X)]^r p_x$$
$$=\sum_{x=0}^{\infty}\frac{\left[x-\frac{\Lambda}{(1-\theta)}+1\right]^r}{-\ln(1-\theta)}\frac{\theta^{x+1}}{(x+1)}, \qquad 2.18$$

in the light of (2.15). Differentiating both sides of (2.18) with respect to θ to get the following, since $\Lambda=\theta\left[-\ln(1-\theta)\right]^{-1}$.

$$\frac{d\mu_r}{d\theta}=\sum_{x=0}^{\infty} r\frac{\left[x-\frac{\Lambda}{(1-\theta)}+1\right]^{r-1}}{-\ln(1-\theta)}\frac{\theta^{x+1}}{x+1}\left[\frac{-1}{(1-\theta)\left[-\ln(1-\theta)\right]}-\frac{\theta}{(1-\theta)^2\left[-\ln(1-\theta)\right]}+\right.$$

$$\left.\frac{\theta}{(1-\theta)^2\left[-\ln(1-\theta)\right]^2}\right]+\sum_{x=0}^{\infty}\left[x-\frac{\Lambda}{(1-\theta)}+1\right]^{r}\left[\frac{(x+1)\theta^x}{-\ln(1-\theta)}-\frac{\theta^{x+1}}{(1-\theta)[-\ln(1-\theta)]^2}\right]$$

$$=\frac{r(\Lambda-1)}{(1-\theta)\left[-\ln(1-\theta)\right]}\sum_{x=0}^{\infty}\left[x-\frac{\Lambda}{(1-\theta)}+1\right]^{r-1}p_x+\frac{1}{\theta}\sum_{x=0}^{\infty}\left[x-\frac{\Lambda}{(1-\theta)}+1\right]^{r+1}p_x$$

$$=\frac{r(\Lambda-1)}{(1-\theta)\left[-\ln(1-\theta)\right]}\mu_{r-1}+\frac{1}{\theta}\mu_{r+1}$$

which implies (2.17).

RESULT 2.9. *The maximum likelihood estimator* $\hat{\theta}$ *of* θ *of the ZILSD and the moment estimator* $\tilde{\theta}$ *of* θ *of the ZILSD coincides.*

PROOF. Let $X_1, X_2, ..., X_n$ be a random sample from a population following ZILSD with pmf (2.1). Then the log-likelihood function of the sample is

$$\log L = n\log\left[\frac{1}{-\ln(1-\theta)}\right]+\sum_{j=1}^{n}(x_j+1)\log\theta-\log\left[(x_1+1)(x_2+1)...(x_n+1)\right]. \quad (2.19)$$

On differentiating (2.19) with respect to θ and equating to zero, we get the following

$$\frac{\partial\log L}{\partial\theta}=\frac{-n}{\left[-\ln(1-\theta)\right](1-\theta)}+\frac{\sum_{j=1}^{n}(x_j+1)}{\theta}=0,$$

which on simplification gives

$$\frac{\hat{\theta}}{\left[-\ln(1-\hat{\theta})\right](1-\hat{\theta})}-1=\bar{x}, \quad (2.20)$$

where

$$\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_j .$$

The moment estimator $\tilde{\theta}$ of θ of the ZILSD is obtained by equating the mean of ZILSD as given in (2.15) with the sample mean $\bar{X}$. That is, $\tilde{\theta}$ is the solution of the following equation.

$$\frac{\tilde{\theta}}{\left[-\ln(1-\tilde{\theta})\right](1-\tilde{\theta})}-1=\overline{x} \tag{2.21}$$

Equation (2.20) and (2.21) shows that maximum likelihood estimator $\hat{\theta}$ and moment estimator $\tilde{\theta}$ of the ZILSD coincide each other, since both estimators are solution of the same equation.

3. MODIFIED ZERO-INFLATED LOGARITHMIC SERIES DISTRIBUTION

In this section, first we obtain a modified version of ZILSD as in the following.

Let X be a random variable following ZILSD with pgf (2.3). Consider reals $\theta_1 > 0,\ \theta_2 > 0$ such that $\theta = \theta_1 + \theta_2 < 1$. Define $\eta = \frac{\theta_1}{\theta}$ and $\{Y_n, n \geq 1\}$ be a sequence of independent and identically distributed random variables each with pgf

$$A(z) = \eta\, z + (1-\eta)z^2 . \tag{3.1}$$

Assume that $\{Y_n, n \geq 1\}$ and X are independent. Define $S_0 = 0$ and $S_X = \sum_{n=1}^{X} Y_n$, then the pgf of Y_n is

$$\begin{aligned} H(z) &= G\{A(z)\} \\ &= -\Lambda(\theta_1 z - \theta_2 z^2)^{-1} \ln(1-\theta_1 z - \theta_2 z^2) . \end{aligned} \tag{3.2}$$

A distribution with pgf (3.2) we call "the modified zero-inflated logarithmic series distribution (MZILSD)" with two parameters θ_1 and θ_2 $(\theta_1, \theta_2 > 0$ and $\theta_1 + \theta_2 < 1)$. Clearly when $\theta_2 = 0$, (3.2) reduces to the pgf of ZILSD as given in (2.2).

Let Y be a random variable following MZILSD with pgf (3.2). Now, we obtain the following results in the light of the series representations:

$$\sum_{x=0}^{\infty}\sum_{r=0}^{\infty} A(r,x) = \sum_{x=0}^{\infty}\sum_{r=0}^{x} A(r,x-r) \tag{3.3}$$

and

$$\sum_{x=0}^{\infty}\sum_{r=0}^{\infty} B(r,x) = \sum_{x=0}^{\infty}\sum_{r=0}^{\left[\frac{x}{m}\right]} B(r,x-mr) , \tag{3.4}$$

in which $[a]$ is the whole part of a.

RESULT 3.1. For $x = 0,1,2,...,$ *the probability mass function* $q_x = P(Y = x)$ *of the MZILSD with pgf (3.2) is the following, in which* $\Lambda = \left[-\ln(1-\theta_1-\theta_2)\right]^{-1}(\theta_1+\theta_2)$.

$$q_x = \Lambda \sum_{r=0}^{\left[\frac{x}{2}\right]} \frac{(x-r)!}{(x-r+1)} \frac{\theta_1^{x-2r}}{(x-2r)!} \frac{\theta_2^r}{r!} \tag{3.5}$$

PROOF. From (3.2) we have the following.

$$H(z) = \sum_{x=0}^{\infty} q_x z^x \tag{3.6}$$

$$= \Lambda \frac{-\ln(1-\theta_1 z-\theta_2 z^2)}{(\theta_1 z+\theta_2 z^2)} \tag{3.7}$$

On expanding the logarithmic function in (3.7), to obtain

$$H(z) = \Lambda \sum_{x=1}^{\infty} \frac{(\theta_1 z+\theta_2 z^2)^{x-1}}{x}$$

$$= \Lambda \sum_{x=0}^{\infty} \frac{(\theta_1 z+\theta_2 z^2)^{x}}{(x+1)}$$

$$= \Lambda \sum_{x=0}^{\infty} \sum_{r=0}^{x} \frac{x!}{(x+1)} \frac{\theta_1^{x-r}}{(x-r)!} \frac{\theta_2^{r}}{r!} z^{x+r}, \tag{3.8}$$

by binomial theorem. Now by applying (3.3) in (3.8), we get

$$H(z) = \Lambda \sum_{x=0}^{\infty} \sum_{r=0}^{\infty} \frac{(x+r)!}{(x+r+1)x!r!} \theta_1^{x} \theta_2^{r} z^{x+2r}, \tag{3.9}$$

which implies the following, in the light of (3.4).

$$H(z) = \Lambda \sum_{x=0}^{\infty} \sum_{r=0}^{\left[\frac{x}{2}\right]} \frac{(x-r)!}{(x-r+1)} \frac{\theta_1^{x-2r}}{(x-2r)!} \frac{\theta_2^{r}}{r!} z^{x} \tag{3.10}$$

On equating the coefficient of z^x on the right hand side expressions of (3.6) and (3.10) we get (3.5)

RESULT 3.2. For $r \geq 0$ *the following is a simple recurrence relation for probabilities* $q_r = q_r(\underline{u})$ *of the MZILSD, in which for any positive integer r,* $q_{-r}(\underline{u}) = 0$.

$$(r+1)q_{r+1}(\underline{u}) = D_0 \Lambda_0^{-1} \Lambda_1 \left[\theta_1 q_r(\underline{u}+1) + 2\theta_2 q_{r-1}(\underline{u}+1)\right] \tag{3.11}$$

in which $\Lambda_i = {}_2F_1(1+i\,, 1+i\,; 2+i\,; \theta_1+\theta_2)$, *for* $i = 0, 1$.

PROOF. The pgf of the MZILSD given in (3.2) can be written as

$$H(z) = \Lambda \, {}_2F_1(1\,,1\,;2\,;\theta_1 z + \theta_2 z^2)\,. \tag{3.12}$$

On differentiating (3.12) with respect to z, we get

$$\sum_{r=0}^{\infty} r\, q_r(\underline{u}) z^{r-1} = D_0 \Lambda_0^{-1} \left(\theta_1 + 2\theta_2 z\right) {}_2F_1(2\,,2\,;3\,;\theta_1 z + \theta_2 z^2)\,. \tag{3.13}$$

From (3.12) and (3.13) we obtain the following.

$$\Lambda_1 \sum_{r=0}^{\infty} q_r(\underline{u})\, z^r = {}_2F_1(2\,,2\,;3\,;\theta_1 z + \theta_2 z^2) \tag{3.14}$$

Substituting (3.14) in (3.13) we have

$$\begin{aligned}\sum_{r=0}^{\infty} r\, q_r(\underline{u}) z^{r-1} &= D_0 \Lambda_0^{-1} \Lambda_1 (\theta_1 + 2\theta_2 z) \sum_{r=0}^{\infty} q_r(\underline{u}+1) z^r \\ &= D_0 \Lambda_0^{-1} \Lambda_1 \left[\theta_1 \sum_{r=0}^{\infty} q_r(\underline{u}+1) z^r + 2\theta_2 \sum_{r=0}^{\infty} q_r(\underline{u}+1) z^{r+1} \right].\end{aligned} \tag{3.15}$$

On equating the coefficient of z^r on both sides of (3.15) we obtain (3.11).

RESULT 3.3. The following is a simple recurrence relation for raw moments $\mu_r(\underline{u})$ of MZILSD, for $r \geq 0$.

$$\mu_{r+1}(\underline{u}) = D_0 \Lambda_0^{-1} \Lambda_1 \sum_{j=0}^{r} \binom{r}{j} (\theta_1 + 2^{j+1}) \theta_2 \mu_{r-j}(\underline{u}+1) \tag{3.16}$$

PROOF. The characteristic function of MZILSD with pgf given in (3.12) has the following representation. For z in R and $i = \sqrt{-1}$,

$$\varphi_Y(z) = \sum_{r=0}^{\infty} \mu_r(\underline{u}) \frac{(iz)^r}{r!} \tag{3.17}$$

$$= \Lambda_0^{-1}\, {}_2F_1(1\,,1\,;2;\,\theta_1 e^{iz} + \theta_2 e^{2iz})\,. \tag{3.18}$$

On differentiating the right side expressions of (3.17) and (3.18) with respect to z, we get

$$\sum_{r=0}^{\infty} \mu_r(\underline{u}) \frac{(iz)^{r-1}}{(r-1)!} = D_0 \Lambda_0^{-1} (\theta_1 e^{iz} + 2\theta_2 e^{2iz})\, {}_2F_1(2\,,2\,;3\,;\theta_1 e^{iz} + \theta_2 e^{2iz})\,. \tag{3.19}$$

From (3.17) and (3.18) we obtain the following.

$$\Lambda_1 \sum_{r=0}^{\infty} \mu_r(\underline{u}) \frac{(iz)^r}{r!} = {}_2F_1(2\,,2\,;3\,;\theta_1 e^{iz} + \theta_2 e^{2iz}) \tag{3.20}$$

Equations (3.19) and (3.20) together leads to

$$\sum_{r=0}^{\infty}\mu_r(\underline{u})\frac{(iz)^{r-1}}{(r-1)!} = D_0\Lambda_0^{-1}\Lambda_1(\theta_1 e^{iz} + 2\theta_2 e^{2iz})\sum_{r=0}^{\infty}\mu_r(\underline{u}+1)\frac{(iz)^r}{r!}$$

$$= D_0\Lambda_0^{-1}\Lambda_1\left[\theta_1\sum_{j=0}^{\infty}\frac{(iz)^j}{j!} + 2\theta_2\sum_{j=0}^{\infty}\frac{(2iz)^j}{j!}\right]\sum_{r=0}^{\infty}\mu_r(\underline{u}+1)\frac{(iz)^r}{r!}. \tag{3.21}$$

Equating the coefficient of $(r!)^{-1}(iz)^r$ on both sides of (3.21) we get (3.16).

RESULT 3.4. The following is a simple recurrence relation for factorial moments $\mu_{[r]}(\underline{u})$ *of MZILSD for* $r \geq 1$, *in which* $\mu_{[0]}(\underline{u}) = 1$.

$$\mu_{[r+1]}(\underline{u}) = D_0\Lambda_0^{-1}\Lambda_1\left(\theta_1 + 2\theta_2\right)\mu_{[r]}(\underline{u}) + D_0\Lambda_0^{-1}\Lambda_1 2\theta_2\mu_{[r-1]}(\underline{u}) \tag{3.22}$$

PROOF. From (3.12) we have the following factorial moment generating function $\psi(z)$ of the MZILSD

$$\psi(z) = \sum_{r=0}^{\infty}\mu_{[r]}(\underline{u})\frac{z^r}{r!} \tag{3.23}$$

$$= \Lambda_0^{-1}{}_2F_1\left[1\,,1\,;2\,;\theta_1(1+z)+\theta_2(1+z)^2\right]. \tag{3.24}$$

On differentiating the right hand side expression of (3.23) and (3.24) with respect to z, we obtain

$$\sum_{r=0}^{\infty}\mu_{[r]}(\underline{u})\frac{z^{r-1}}{(r-1)!} = D_0\Lambda_0^{-1}\left[\theta_1 + 2\theta_2(1+z)\right]{}_2F_1\left[2\,,2\,;3\,;\theta_1(1+z)+\theta_2(1+z)^2\right] \tag{3.25}$$

From (3.23) and (3.24) we obtain the following

$$\Lambda_1\sum_{r=0}^{\infty}\mu_r(\underline{u}+1)\frac{z^r}{r!} = \Lambda_0^{-1}{}_2F_1\left[2\,,2\,;3\,;\theta_1(1+z)+\theta_2(1+z)^2\right] \tag{3.26}$$

Equations (3.24) and (3.25) together lead to

$$\sum_{r=0}^{\infty}\mu_{[r]}(\underline{u})\frac{z^{r-1}}{(r-1)!} = D_0\Lambda_0^{-1}\Lambda_1\left[\theta_1 + 2\theta_2(1+z)\right]\sum_{r=0}^{\infty}\mu_{[r]}(\underline{u}+1)\frac{z^r}{r!}$$

$$= D_0\Lambda_0^{-1}\Lambda_1(\theta_1 + 2\theta_2)\sum_{r=0}^{\infty}\mu_{[r]}(\underline{u}+1)\frac{z^r}{x!} + 2D_0\Lambda_0^{-1}\Lambda_1\,\theta_2\sum_{r=0}^{\infty}\mu_{[r]}(\underline{u}+1)\frac{z^{r+1}}{r!} \tag{3.27}$$

On equating the coefficient of $(x!)^{-1}z^x$ on both sides of (3.27) we get (3.22).

RESULT 3.5. The mean and variance of MZILSD are given below.

$$E(Y) = D_0\Lambda_0^{-1}\Lambda_1(\theta_1 + 2\theta_2)$$

$$Var(Y) = D_0\Lambda_0^{-2}(D_0\Lambda_0\Lambda_1 - D_0\Lambda_1^2)(\theta_1 + 2\theta_2)^2 + D_0\Lambda_0^{-1}\Lambda_1(\theta_1 + 4\theta_2)$$

Proof is simple and hence omitted.

4. ESTIMATION

Here we estimate the parameters θ_1 and θ_2 of the MZILSD by the method of maximum likelihood and illustrated by using certain real life data sets.

Let $a(x)$ be the observed frequency of x events and let m be the highest value of x. Then the likelihood function of the sample is

$$L = \prod_{x=0}^{m} \left[q_x\right]^{a(x)}, \tag{4.1}$$

where q_x is the pmf of the MZILSD as given in (3.5). Now taking logarithm on both sides of (4.1), we have

$$\begin{aligned}\log L &= \sum_{x=0}^{m} a(x)\log(q_x)\\ &= \sum_{x=0}^{m} a(x)\left[\log\Lambda + \log\Psi(x\,;\theta_1;\theta_2\,)\right],\end{aligned} \tag{4.2}$$

where Λ is as given in (3.5) and

$$\Psi(x;\theta_1;\theta_2\,) = \sum_{r=0}^{\left[\frac{x}{2}\right]} \frac{(x-r)!}{(x-r+1)}\frac{\theta_1^{x-2r}}{(x-2r)!}\frac{\theta_2^r}{r!}.$$

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ denote the maximum likelihood estimators of the parameter θ_1 and θ_2 respectively of the MZILSD. On differentiating (4.2) partially with respect to the parameters θ_1 and θ_2 and equating to zero, we get the following likelihood equations.

$$\sum_{x=0}^{m} a(x)\left[\frac{1}{(\theta_1+\theta_2)} - \frac{1}{\left[-\ln(1-\theta_1-\theta_2)\right](1-\theta_1-\theta_2)} + \frac{\sum_{r=0}^{\left[\frac{x}{2}\right]} \frac{(x-r)!}{(x-r+1)}\frac{\theta_1^{x-2r-1}}{(x-2r-1)!}\frac{\theta_2^r}{r!}}{\Psi(x;\theta_1;\theta_2\,)}\right] = 0 \tag{4.3}$$

and

$$\sum_{x=0}^{m} a(x)\left[\frac{1}{(\theta_1+\theta_2)}-\frac{1}{\left[-\ln(1-\theta_1-\theta_2)\right](1-\theta_1-\theta_2)}+\frac{\sum_{r=1}^{\left[\frac{x}{2}\right]}\frac{(x-r)!}{(x-r+1)}\frac{\theta_1^{x-2r}}{(x-2r)!}\frac{\theta_2^{r-1}}{(r-1)!}}{\Psi(x;\theta_1;\theta_2)}\right]=0 \quad (4.4)$$

The likelihood equations do not always have a solution because the MZILSD is not a regular model. Therefore, when likelihood equations do not always have a solution the maximum of the likelihood function attained at the border of the domain of parameters. We obtained the second order partial derivatives of $\log q_x$ with respect to parameters θ_1 and θ_2 and by using MATHCAD we observed that these equations give negative values for all $\theta_1 > 0$ and $\theta_2 \geq 0$ such that $\theta_1 + \theta_2 < 1$.Thus the density of the MZILSD is log-concave and have maximum likelihood estimates of the parameters θ_1 and θ_2 are unique (cf. Puig, 2003). Now on solving these two likelihood equations (4.3) and (4.4) by using some mathematical software such as MATHLAB, MATHCAD, MATHEMATICA etc. one can obtain the maximum likelihood estimates of the parameters θ_1 and θ_2 of the MLSD.

5. TESTING OF THE HYPOTHESIS

In this section we discuss certain test procedures for testing the significance of the additional parameter θ_2 of the MZILSD by using generalized likelihood ratio test and Rao's efficient score test. Here the null hypothesis is $H_0 : \theta_2 = 0$ against the alternative hypothesis $H_1 : \theta_2 \neq 0$.

In case of generalized likelihood ratio test, the test statistic is

$$-2\log\lambda = 2[\log L(\hat{\underline{\theta}};x) - \log L(\hat{\underline{\theta}}^{*};x)], \quad (5.1)$$

where $\hat{\underline{\theta}}$ is the maximum likelihood estimator of $\underline{\theta} = (\theta_1, \theta_2)$ with no restrictions, and $\hat{\underline{\theta}}^{*}$ is the maximum likelihood estimator of $\underline{\theta}$ when $\theta_2 = 0$. The test statistic $-2\log\lambda$ given in (5.1) is asymptotically distributed as χ^2 with one degree of freedom (for details see Rao, 1973).

In case of Rao's efficient score test, the test statistic is

$$T = V'\phi^{-1}V , \quad (5.2)$$

where

$$V' = \left(\frac{1}{\sqrt{n}}\frac{\partial \log L}{\partial \theta_1}, \frac{1}{\sqrt{n}}\frac{\partial \log L}{\partial \theta_2}\right)$$

and ϕ is the Fisher information matrix. The test statistic given in (5.2) follows chi-square distribution with one degree of freedom (for details see Rao, 1973).

TABLE 1

Observed frequencies and computed values of expected frequencies of the PD, the NBD, the ZLSD and the MZILSD by method of maximum likelihood for the first data set.

x	Observed frequency	PD	NBD	ZILSD	MIZLSD
0	83	40.97	90.288	64.512	82.656
1	18	51.50	18.00	27.36	18.144
2	13	32.37	9.648	15.552	16.56
3	9	13.56	6.192	9.936	8.208
4	7	4.26	4.32	6.768	5.76
5	7	1.07	3.168	4.752	3.744
6	2	0.23	2.448	3.456	2.592
≥ 7	5	0.04	9.936	11.664	6.336
Total	144	144	144	144	144
Estimation of parameters		$m = 1.257$	$\hat{p} = 0.181$ $\hat{r} = 0.32$	$\hat{\theta}_1 = 0.85$	$\hat{\theta}_1 = 0.44$ $\hat{\theta}_2 = 0.27$
Chi-square values		120.383	9.179	12.899	2.513
Degrees of freedom		3	3	5	4
P-values		< 0.00001	0.027	0.024	0.642

6. DATA ILLUSTRATIONS AND CONCLUDING REMARKS

For numerical illustrations, we have considered the following two data sets.

(i) Data on the number of food stores in the Ljubljuna taken from Douglas (1980)

(ii) Data on accidents experienced by 414 machinists in three months taken from Greenwood and Yule, (1920)

We have fitted the models - the Poisson distribution (PD), the negative binomial distribution (NBD), the ZILSD and the MZILSD. Based on the computed chi-square values and P - values given in the Table 1 and Table 2, it can be observed that the MZILSD gives a better fit to both data sets compared to the existing models - the PD, the NBD and the ZILSD.

TABLE 2

Observed frequencies and computed values of expected frequencies of the PD, the NBD, the ZLSD and the MZILSD by method of maximum likelihood for the second data set.

Accidents per machinist	No. of machinists	PD	NBD	ZILSD	MZILSD
0	296	261.6	314.226	268.272	296.01
1	74	120.1	68.724	81.972	74.106
2	26	27.6	20.7	33.534	26.082
3	8	4.2	6.624	15.318	9.936
≥ 4	10	0.5	3.726	14.904	7.886
Total	414	414	414	414	414
Estimation of parameters		$m = 0.457$	$\hat{p} = 0.622$ $\hat{r} = 0.58$	$\hat{\theta}_1 = 0.61$	$\hat{\theta}_1 = 0.50$ $\hat{\theta}_2 = 0.0099$
Chi-square values		26.457	8.474	10.44	0.957
Degrees of freedom		1	1	3	2
P-values		< 0.0001	0.004	0.015	0.62

In order to test the significance of the additional parameter θ_2 of the MZILSD, we have computed the values of $\log L(\underline{\hat{\theta}}^*;x)$, $\log L(\underline{\hat{\theta}}^*;x)$ and the test statistic given in (5.1) for the MZILSD for both data sets and presented in Table 3.

TABLE 3
Computed values of the test static for generalized likelihood ratio test.

	$\log L(\hat{\underline{\theta}}^*; x)$	$\log L(\hat{\underline{\theta}}; x)$	Test statistic
Data set 1	-96.59	-93.557	6.066
Data set 2	-165.45	-162.706	5.488

Further, we have computed the values of T for first data set as T_1 and the second data set as T_2, as given below.

$$T_1 = \begin{pmatrix} 0.703 & 12.481 \end{pmatrix} \begin{bmatrix} 0.068 & -0.064 \\ -0.064 & 0.065 \end{bmatrix} \begin{pmatrix} 0.703 \\ 12.481 \end{pmatrix} = 8.979 \qquad (6.1)$$

$$T_2 = \begin{pmatrix} -0.192 & 28.74 \end{pmatrix} \begin{bmatrix} 0.016 & -0.015 \\ -0.015 & 0.015 \end{bmatrix} \begin{pmatrix} -0.192 \\ 28.74 \end{pmatrix} = 12.662 \qquad (6.2)$$

The critical values for both the tests – the generalized likelihood ratio test and Rao's efficient score test – with significance level $\alpha = 0.05$ and degrees of freedom one is 3.84. Thus, based on the values of test statistic computed in Table 3 and equations (6.1) and (6.2), we reject the null hypothesis $H_0 : \theta_2 = 0$ in all the cases. Hence we conclude that the additional parameter θ_2 of the MZILSD is significant in case of both the data sets considered in the paper.

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to the Editor and the anonymous Referee for carefully reading the paper and for valuable comments.

REFERENCES

C. CHATFIELD, A. S. C. EHRENBERG, G. J. GOODHARDT (1966). *Progress on a simplified model of Stationary purchasing behavior (with discussion).* Journal of the Royal Statistical Society, Series A, 129, pp. 317-367.

C. CHATFIELD (1986). *Discrete distributions and purchasing models.* Communication in Statistics-Theory and Methods, 15, pp. 697-708.

J. B. DOUGLAS (1980). *Analysis with standard contagious distribution*. Statistical distributions in scientific work (ed. G. P. Patil), International cooperative publishing house, Fairland, pp. 241-262.

R. A. FISHER, A. S. CORBERT, C.B. WILLIAMS (1943).*The relation between the number of species and the number of individuals in a random sample of an animal population*. Journal of Animal Ecology, 12, pp. 42-58.

M. GREENWOOD, G. U. YULE (1920). *An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents.* Journal of Royal Statistical Society, 83, pp. 255-279.

G. C. JAIN, R. P. GUPTA (1973). *A logarithmic series type distribution*. Trabaios de Estadistica, 24, pp.99- 105.

N. L. JOHNSON, A. W. KEMP, A. W. KOTZ (2005). *Univariate Discrete Distributions*. Wiley, New York.

T. F. KANG, S. H. ONG (2007). *A new generalization of the logarithmic distribution arising from the inverse trinomial distribution*. Communication in Statistics–Theory and Methods, 36, pp. 3 – 21.

R. A. KEMPTON (1975). *A generalized form Fisher's logarithmic series*. Biometrika, 62, pp. 29-38.

C. S. KUMAR (2009). *Some properties of Kemp family of distributions.* Statistica, 69, pp. 311-316.

A. M. MATHAI, H. J. HAUBOLD (2008). *Special Functions for Applied Scientists*. Springer, New York.

S. H. ONG (2000). *On a generalization of the log-series distribution.* Journal of Applied Statistical Science, 10(1), PP. 77- 88.

P. PUIG (2003). *Characterizing additively closed discrete models by a property of their MLEs, with an application to generalized Hermite distribution*. Journal of American Statistical Association, 98, pp. 687-692.

C. R. RAO (1973). *Linear Statistical Inference and its Applications.* John Wiley, New York.

L. J. SLATER (1966). *Generalized Hypergeometric Functions*. Cambridge university press, Cambridge.

R. C. TRIPATHI, R. C. GUPTA (1985). *A generalization of the log-series distribution*. Communication in Theory and Methods, 14, pp. 1779-1799.

R. C. TRIPATHI, R. C. GUPTA (1988). *Another generalization of the logarithmic series and the geometric distribution.* Communication in Statistics-Theory and Methods, 17, pp. 1541-1547.

SUMMARY

On zero - inflated logarithmic series distribution and its modification

Here we consider a zero-inflated logarithmic series distribution (ZILSD) and study some of its properties. A modified form of ZILSD is also developed and derived several important aspects of it. The parameters of modified ZILSD are estimated by method of maximum likelihood and illustrated the procedure using certain real life data sets. Further, certain test procedures are suggested.