International Journal of Mathematical Archive-8(3), 2017, 92-98
Available online through www.ijma.info ISSN 2229 – 5046

A COMPARATIVE STUDY ON ZERO-TRUNCATED GENERALIZED POISSON-LINDLEY AND ZERO-TRUNCATED POISSON-LINDLEY DISTRIBUTIONS

SAMUEL ADEWALE ADEROJU*[1]
[1]Department of Statistics and Mathematical Sciences,
Kwara State University, Malete, P. M. B. 1530, Ilorin, Kwara State, Nigeria.

EMMANUEL TEJU JOLAYEMI[2]
[2]Department of Statistics, University of Ilorin, P. M. B. 1515, Ilorin, Nigeria.

ABDULWASIU OPEYEMI IBRAHIM[3]
[3]Department of Statistics and Mathematical Sciences,
Kwara State University, Malete, P. M. B. 1530, Ilorin, Kwara State, Nigeria.

(Received On: 13-01-17; Revised & Accepted On: 23-03-17)

ABSTRACT

In this paper, the zero-truncated generalized Poisson-Lindley (ZTGPL) distribution has been proposed and its properties studied. The maximum likelihood method is used to obtain the estimators of its parameters through R-software. The comparative study with zero-truncated Poisson (ZTP) and zero-truncated Poisson-Lindley (ZTPL) distributions is done with two datasets. The proposed distribution is characterized by two parameters and flexible to account for both over- and under-dispersion in structurally non-zero count data. The statistic (chi square) is used to check its goodness-of-fit.

Key words: *Generalized Poisson-Lindley distribution; goodness of fit; Poisson distribution; Poisson-Lindley distribution; Zero-Truncated distribution.*

1. INTRODUCTION

In probability theory, zero-truncated discrete distributions are distribution that can handle the set of positive integers that are structurally non-zero counts.

The Lindley distribution and the Poisson-Lindley (PL) distributions have been generalized by many researchers. Baghestani *et al.* (2014) applied a Generalized Poisson-Lindley distribution initially introduced by Mahmudi and Zakerzadeh (2010). They derived the Generalized Poisson-Lindley distribution by compounding Poisson distribution with the generalized Lindley distribution introduced by Zakerzadeh and Dolati (2010). Shanker and Mishra. (2013a) introduced two parameter Lindley distribution. Shanker and Mishra (2014) obtained a two parameter Poisson-Lindley distribution by compounding Poisson distribution with the two parameter Lindley distribution they introduced (Shanker and Mishra, 2013a). Bhati *et al.* (2015) introduced a new generalized Poisson-Lindley distribution by compounding Poisson distribution with a two parameter generalized Lindley distribution. A quasi Poisson-Lindley distribution has been introduced by Shanker and Mishra (2015) by compounding Poisson distribution with a quasi Lindley distribution introduced by Shanker and Mishra (2013b). Ghitany *et al.* introduced a one-parameter zero-truncated Poisson-Lindley (ZTPL) distribution.

Zero-truncated Poisson (ZTP) and zero-truncated Poisson-Lindley (ZTPL) distributions have been compared and studied using graphs for different values of their parameter by Shanker *et al.* (2015). A general expression for the *r*th factorial moment of ZTPL distribution has been obtained and the first four moments about origin has been given. A very simple and easy method for finding moments of ZTPL distribution has been suggested. Both ZTP and ZTPL distributions have been fitted to a number of data sets from different fields to study their goodness of fits and superiority of one over the other.

Corresponding Author: Emmanuel Teju Jolayemi[2]
[2]Department of Statistics, University of Ilorin, P. M. B. 1515, Ilorin, Nigeria.

The rest of this paper is organized as follows: In section two, we review on other distributions under -consideration. In sections three and four we present the ZTGPL distribution with its properties and the maximum likelihood method for estimating the parameters respectively. In section five, we apply the ZTGPL with other distributions to two real datasets and show that this distribution provides an excellent fit to these datasets. Finally, some concluding remarks were made in section six.

2. REVIEW ON OTHER DISTRIBUTIONS UNDER CONSIDERATION

2.1Zero-truncated Poisson (ZTP) distribution

The probability mass function (pmf) of a ZTP variable as given independently by Plackett (1953) and Johnson *et al.* (1969) as:

$$P\left(Y=y\right)=\begin{cases}\dfrac{\lambda^{y}}{y!(e^{\lambda}-1)}, & if\ y=1,2,\ldots\\ 0, & e.w\end{cases} \tag{1}$$

The difference with the standard Poisson distribution lies in the correction factor $\left(1-e^{-\lambda}\right)^{-1}$, which reflects the fact that a value of zero (0) cannot occur.

The moment generating function is obtained to be:

$$M_{y}(t)=\frac{e^{\lambda t}-1}{e^{\lambda}-1} \tag{2}$$

Moreover, the basic parameters such as the mean and variance respectively are:

$$\mu=\frac{\lambda e^{\lambda}}{e^{\lambda}-1} \quad and \quad \sigma^{2}=\frac{\lambda e^{\lambda}}{e^{\lambda}-1}\left[1-\frac{\lambda}{e^{\lambda}-1}\right]$$

2.2 Zero-truncated Poisson-Lindley (ZTPL) distribution

The ZTPL was studied by Shanker *et al.* (2015) and the pmf defined as:

$$P(Y=y)=\begin{cases}\dfrac{\theta^{2}}{\theta^{2}+3\theta+1}\dfrac{y+\theta+2}{\left(\theta+1\right)^{y}}, & if\ y=1,2,\ldots;\theta>0\\ 0\ \ , & e.w\end{cases} \tag{3}$$

Moreover, the basic parameters such as the mean and variance respectively are as given by Ghitany *et al.* (2008):

$$\mu=\frac{\left(\theta+1\right)^{2}\left(\theta+2\right)}{\theta\left(\theta+3\theta+1\right)} \quad and \quad \sigma^{2}=\frac{\left(\theta+1\right)^{2}\left(\theta^{2}+6\theta^{2}+10\theta+2\right)}{\theta^{2}\left(\theta^{2}+3\theta+1\right)^{2}}$$

Ghitany *et al.* (2008) showed that the MLE $\hat{\theta}$ of θ is consistent and asymptotically normal.

3. ZERO-TRUNCATED GENERALIZED POISSON-LINDLEY DISTRIBUTION

Baghestani *et al.* (2014) derived a generalized Poisson-Lindley (GPL) distribution by compounding Poisson distribution with the generalized Lindley distribution introduced by Zakerzadeh and Dolati (2010). For details on its various properties, see Baghestani *et al.* (2014). However, the probability mass function of the GPL distribution is

$$P(y_i;\alpha,\beta)=\begin{cases}\left\{\dfrac{\Gamma(y_i+\alpha)\beta^{\alpha+1}}{y_i!\Gamma(\alpha+1)(\beta+1)^{y_i+\alpha+1}}\left[\alpha+\dfrac{\alpha+y_i}{(\beta+1)}\right],\right. & for\ y_i=1,\ 2,\ \ldots;\alpha,\beta>0\\ 0, & e.w\end{cases} \tag{4}$$

Hence, the Zero-truncated version of the distribution, which we refer to as Zero-truncated Generalized Poisson-Lindley (ZTGPL) distribution can be derived as

$$f\left(y_i;\alpha,\beta\right)=\frac{P\left(y_i;\alpha,\beta\right)}{1-P\left(0;\alpha,\beta\right)}, \qquad y_i=1,2,\ldots,n \tag{5}$$

Where, $P(0;\alpha,\beta)=\frac{(\alpha-1)!}{\alpha!}\frac{\beta^{\alpha+1}}{(1+\beta)^{\alpha+1}}\left(\alpha+\frac{\alpha}{1+\beta}\right)\equiv\frac{(2+\beta)\beta^{\alpha+1}}{(1+\beta)^{\alpha+2}}$ and $f(y_i;\alpha,\beta)$ is the ZTGPL.

Therefore, the pmf of the ZTGPL distribution is derived as

$$f(y_i;\alpha,\beta)=\frac{\Gamma(y_i+\alpha)\beta^{\alpha+1}}{y_i!\Gamma(\alpha+1)(\beta+1)^{y_i+\alpha+1}}\left[\alpha+\frac{\alpha+y_i}{(\beta+1)}\right]\div\left\{1-\frac{(2+\beta)\beta^{\alpha+1}}{(1+\beta)^{\alpha+2}}\right\}$$

$$=\frac{\beta^{1+\alpha}(1+\beta)^{-1-y_i-\alpha}\left(\alpha+\frac{\alpha+y_i}{(\beta+1)}\right)(-1+y_i+\alpha)!}{y_i!\left(1-\frac{\beta^{1+\alpha}(1+\beta)^{-1-\alpha}\left(\alpha+\frac{\alpha}{1+\beta}\right)(\alpha-1)!}{\alpha!}\right)\alpha!}$$

$$=\frac{\beta^{1+\alpha}(1+\beta)^{-y_i}\left(y_i+\alpha(2+\beta)\right)(y_i+\alpha-1)!}{y_i!\left((1+\beta)^{2+\alpha}\alpha!-\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!\right)},\quad y_i=1,2,\ldots,n \tag{6}$$

Therefore, the p.m.f of ZTGPL as obtained in equation 6, can be written as

$$f(y_i;\alpha,\beta)=\begin{cases}\frac{\beta^{1+\alpha}(1+\beta)^{-y_i}\left(y_i+\alpha(2+\beta)\right)(y_i+\alpha-1)!}{y_i!\left((1+\beta)^{2+\alpha}\alpha!-\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!\right)}, & y_i=1,2,\ldots,n\,;\quad \alpha,\beta>0\\ 0\qquad, & elsewhere\end{cases}$$

Hence, this is the p.m.f of ZTGPL with parameters α and β

3.1 Generating Functions of ZTGPL distribution

The Probability Generating function is

$$P_y(t)=E(t^y)=\sum_{y=1}^{\infty}t^y f(y;\alpha,\beta)$$

$$=\sum_{y=1}^{\infty}t^y\frac{\beta^{1+\alpha}(1+\beta)^{-y_i}\left(y_i+\alpha(2+\beta)\right)(y_i+\alpha-1)!}{y_i!\left((1+\beta)^{2+\alpha}\alpha!-\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!\right)}$$

$$=\frac{\beta^{1+\alpha}\left(\frac{1-t+\beta}{1+\beta}\right)^{-\alpha}\left(-\left(2+3\beta+\beta^2\right)\left(-1+\left(\frac{1-t+\beta}{1+\beta}\right)^{\alpha}\right)+t\left(-1+2\left(\frac{1-t+\beta}{1+\beta}\right)^{\alpha}+\beta\left(-1+\left(\frac{1-t+\beta}{1+\beta}\right)^{\alpha}\right)\right)\right)\alpha!}{(-1+t-\beta)\left(\alpha\beta^{1+\alpha}(2+\beta)(-1+\alpha)!-(1+\beta)^{2+\alpha}\alpha!\right)} \tag{7}$$

3.2 Moments of ZTGPL distribution

The first four moments are derived as

$$\mu_1'=\frac{\beta^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{-(\alpha+1)}(1+\alpha+\alpha\beta)\alpha!}{\alpha!(1+\beta)^{2+\alpha}-\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!} \tag{8}$$

$$\mu_2'=\frac{(1+\alpha)\beta^{-2+\alpha}\left(\frac{\beta}{1+\beta}\right)^{-\alpha}(1+\beta)(2+\alpha+\alpha\beta)\alpha!}{\alpha!(1+\beta)^{2+\alpha}-\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!} \tag{9}$$

$$\mu_3^{'} = \frac{\left(2+3\alpha+\alpha^2\right)\beta^{-3+\alpha}\left(\frac{\beta}{1+\beta}\right)^{-\alpha}(1+\beta)(3+\alpha+\alpha\beta)\alpha!}{\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!-\alpha!(1+\beta)^{2+\alpha}} \tag{10}$$

$$\mu_4^{'} = \frac{\left(6+11\alpha+6\alpha^2+\alpha^3\right)\beta^{-4+\alpha}\left(\frac{\beta}{1+\beta}\right)^{-\alpha}(1+\beta)(4+\alpha+\alpha\beta)\alpha!}{\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!-\alpha!(1+\beta)^{2+\alpha}} \tag{11}$$

The mean and variance of the distribution can therefore be derived as;

$$\mu = \frac{\beta^{\alpha}\left(\frac{\beta}{1+\beta}\right)^{-(\alpha+1)}(1+\alpha+\alpha\beta)}{(1+\beta)^{2+\alpha}-\alpha\beta^{1+\alpha}(2+\beta)} \tag{12}$$

$$\sigma^2 = \frac{\left(\beta^{\alpha-2}\left(\frac{\beta}{1+\beta}\right)^{-2\alpha}(1+\beta)\alpha!\left(-\beta^{\alpha}(1+\beta)(1+\alpha+\alpha\beta)^2\alpha!-(1+\alpha)\left(\frac{\beta}{1+\beta}\right)^{\alpha}(2+\alpha+\alpha\beta)\left(\alpha\beta^{2+\alpha}(2+\beta)(\alpha-1)!-(1+\beta)^{2+\alpha}\alpha!\right)\right)\right)}{\left(\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)-\alpha!(1+\beta)^{2+\alpha}\right)^2} \tag{13}$$

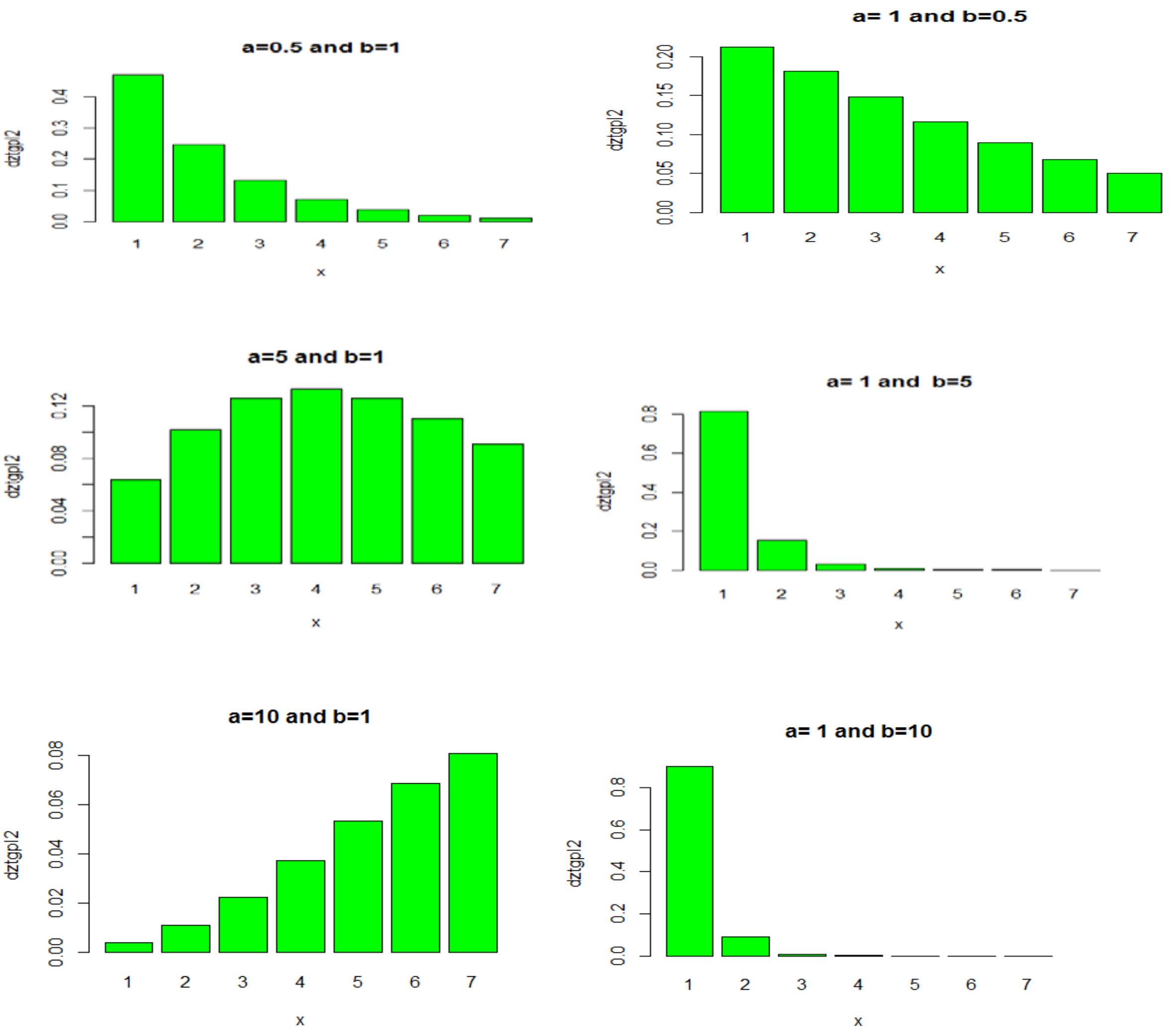

Figure-1: *Graph of probability function of ZTGPL distribution for different values of the parameters (a represents α and b represents β).*

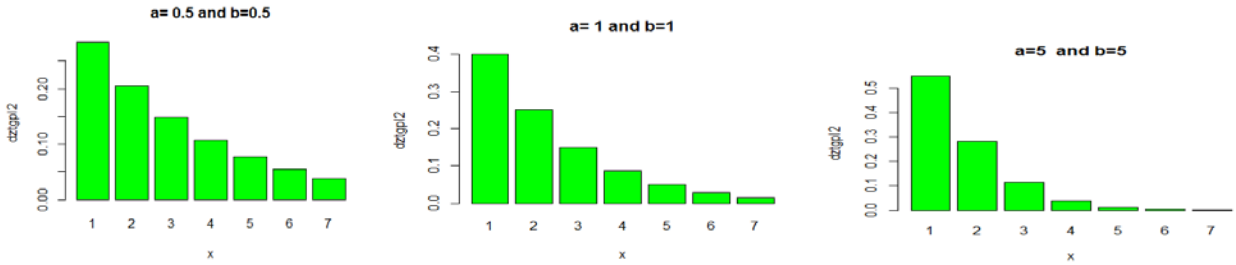

Figure-2: *Graph of probability function of ZTGPL distribution for equal values of the parameters (a represents α and b represents β).*

In figure 1, as α→0 while β remain around 1 the distribution skewed to the right but as α→∞ while β remain around 1 the distribution skewed to the left. However, when α remains around 1 and β → ∞, the distribution right tail disappears gradually. In figure 2, as α and β is simultaneously increase, the distribution right tail disappears gradually.

4. ESTIMATION OF PARAMETERS OF ZTGPL DISTRIBUTION

The likelihood function of the ZTGPL distribution is given as;

$$L(\alpha,\beta;y_i)=\prod_{i=1}^{n}\frac{\beta^{\alpha+1}(1+\beta)^{-y_i}\left(y_i+\alpha(2+\beta)\right)(y_i+\alpha-1)!}{y_i!\left[(1+\beta)^{2+\alpha}\alpha!-\alpha\beta^{\alpha+1}(2+\beta)(\alpha-1)!\right]} \quad (14)$$

The log-likelihood function therefore is;

$$\ell=n(\alpha+1)\log\beta-\sum_{i=1}^{n}y_i\log(\beta+1)+\sum_{i=1}^{n}\log(y_i+2\alpha+2\beta)+\sum_{i=1}^{n}\log(y_i+\alpha-1)!$$
$$-\sum_{i=1}^{n}\log(y_i)!-n\log\left[(1+\beta)^{2+\alpha}\alpha!-\alpha\beta^{1+\alpha}(2+\beta)(\alpha-1)!\right] \quad (15)$$

We are to find the first and second partial derivative of equation (15) with respect to each parameter and equate them to zero as:

$$\frac{\partial\ell}{\partial\alpha}=0,\quad \frac{\partial^2\ell}{\partial\alpha^2}=0,\quad \frac{\partial\ell}{\partial\beta}=0,\quad \text{and}\quad \frac{\partial^2\ell}{\partial\beta^2}=0$$

However, the equations do not have closed form. Therefore, the maximum likelihood estimates (MLEs) of ZTGPL distribution cannot be solved analytically, an iterative methods such as Fisher Score Algorithm, Bisection method Regula-Falsi method or *Newton-Raphson* (NR) iterative method, as implemented by Jolayemi (1990).

We obtained the MLEs of the parameters by direct maximization of the log-likelihood function using ***"optim"*** routine of R software (R Development Core Team, 2016) with ***"L-BFGS-B"*** method. This can as well be done by using **PROC NLMIXED** in SAS.

5. APPLICATION

We would want to compare the distributions with two real datasets so as to establish their goodness-of-fit.

Example 1: Immunogold assay data

The data is taking from Mathews et al (1993), who gave counts of sites with 1, 2, 3, 4 and 5 particles from immunogold assay data. The sample mean and variance are 1.576 and 0.7897, respectively.

X	Obs.Freq	ZTP	ZTPL	ZTGPL
1	122	115.86	124.77	121.89
2	50	57.39	46.76	50.07
3	18	18.95	17.07	17.74
4	4	4.69	6.11	5.78
5	4	0.93	2.15	1.78
Total	**198**	**198**	**198**	**198**

MLE		$\lambda = 0.9906$	$\theta = 2.18307$	$\alpha = 2.3326$ $\beta = 3.3571$
Loglik		205.9477	204.6221	204.3334
X^2 ***(P-value)***		2.329 (0.51)	0.511 (0.77)	0.0296 (0.98)
df		2	2	1

Table-1: *The results of the ZTGPL distribution and other distributions for immunogold assay data*

From the results in table 1 above, the three distributions fit the data well. However, ZTGPL distribution outperforms the ZTPL and ZTP distributions.

Example 2: Mortality data

The data is taking from Shanker *et al.* (2015), who gave counts of mothers of the rural area having at least one live birth and one neonatal death.

X (# of neonatal deaths)	Obs. Freq	ZTP	ZTPL	ZTGPL
1	409	399.7	408.1	409.6
2	88	102.3	89.4	87.3
3	19	17.5	19.3	19.6
4	5	2.2	4.1	4.5
5	1	0.3	1.1	1.0
Total	**522**	**198**	**198**	**198**
MLE		$\lambda = 0.5121$	$\theta = 4.1996$	$\alpha = 0.7857$ $\beta = 3.8696$
Loglik		352.1068	349.2384	349.2242
X^2 ***(P-value)***		3.464 (0.06)	0.145 (0.93)	0.055 (0.82)
df		1	2	1

Table-2: *The number of mothers of the rural area having at least one live birth and one neonatal death.*

Considering the mortality data in table 2, the value of the Chi-square test of ZTGPL distribution is smaller than that of the ZTPL distribution but the p-value is otherwise. That is due to the difference in the number of parameters; the ZTPL distribution is one parameter distribution while the ZTGPL distribution is a two parameter distribution. However, the two distributions perform better than the ZTP distribution. Several other examples, which are not reported in this article show that the ZTGPL distribution is a good alternative to the ZTPL and zero-truncated Poisson-Gamma distributions.

6. CONCLUSION

In this paper the ZTGPL distribution has been proposed and its properties studied. The maximum likelihood method is used to obtain the estimators of its parameters through R-software. The two datasets that have been studied previously with ZTP and ZTPL distributions were used to study its goodness of fit to count data. The proposed distribution is characterized by two parameters and flexible to account for both over- and under-dispersion in structurally non-zero count data. The statistic (chi square) is used to check its goodness-of-fit.

REFERENCE

1. Baghestani A.R, Rahimzadeh M and Pourhoseingholi M.A. (2014). Generalized Poisson-Lindley Distribution in Promotion Time Cure Distribution. Pak.j.stat.oper.res. Vol.X No.4, pp361-368
2. Cullen, M. J., Walsh, J., Nicholson, L. V., and Harris, J. B. (1990). Ultrastructural localization of dystrophin in human muscle by using gold immune labelling, Proceedings of the Royal Society of London, 20, 197-210.
3. Dutta, P. and Borah, M. (2015). Zero-modified Poisson-Lindley distribution. (accepted for publication in the Journal " *Assam Statistical Review*").
4. Ghitany M.E, Al-Mutairi D.K and Nadarajah S. (2008) Zero-truncated Poisson-Lindley distribution and its Applications. Mathematics and Computers in Simulation 79(3): 279-287.
5. Jolayemi, E.T.(1990). Distribution selection for One-Dimensional Multinomials. Biom.J.32: 827-834.
6. Mahmudi .E and Zakerzadeh .H, 2010. Generalized Poisson–Lindley Distribution. Commun. Statist-Theor. Method. 39, 1785–1798.
7. Matthews, J.N.S. and Appleton, D.R. (1993). An application of the truncated Poisson distribution to immunogold assay. *Biometrics*, 49, 617–621.

8. Plackett R.L (1953) The truncated Poisson- distribution. Biometrics 9: 185-188
9. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
10. Shanker R, Fesshaye H, Selvaraj S, Yemane A (2015) On Zero-Truncation of Poisson and Poisson-Lindley Distributions and Their Applications. Biom Biostat Int J 2(6): 00045. DOI: 10.15406/bbij.2015.02.00045.
11. Shanker R and Mishra A (2015) A quasi Poisson-Lindley distribution (submitted).
12. Shanker R and Mishra A (2014) A two-parameter Poisson-Lindley distribution. International Journal of Statistics and Systems 9(1): 79- 85.
13. Shanker R and Mishra A (2013a) A two-parameter Lindley distribution. Statistics in Transition new Series 14(1): 45-56.
14. Shanker R and Mishra A (2013b) A quasi Lindley distribution. African journal of Mathematics and Computer Science Research 6(4): 64-71.
15. Zakerzadeh.H and Dolati .A .(2010). Generalized Lindley distribution, J. Math. Exten, 3, 13-25.

APPENDIX:

R function for the parameters estimation:

```
mlefn=function(par,x,n)
{
α=par[1]
β=par[2]

ll=(α+1)*log(β)-x*log(β+1)+log(x+2*α+α*β)+lfactorial(x+α-1)-lgamma(x+1)-log(factorial(α)*(β+1)^(2+α)-
α*(2+β)*factorial(α-1)*(β^(α+1)))
-sum(ll)
}
optim(c(0.05,0.05),mlefn,method="L-BFGS-B",lower=c(0.05,0.05),upper=c(Inf,Inf),x=data,n)
```

Source of support: Nil, Conflict of interest: None Declared.