A Note on the Applications of some Zero Truncated Distributions

M. A. Umar*; K. Jimoh†; W. B. Yahya

Department of Statistics,
University of Ilorin, Ilorin, Nigeria.

†Department of Physical Sciences,
Al-Hikmah University, Adeta, Ilorin, Nigeria.
E-mail: mumarad90@gmail.com*

Abstract **— Poisson and its related distributions are suitable statistical models for the situations where events occur independently at random. But sufficient conditions are given for the models to be applicable. These conditions are rarely satisfied completely in some instances, especially in Biological, Medical and other applied sciences where zero doesn't count. These may be modeled using a Zero-truncated distribution. Zero-truncated distributions are certain discrete distributions having support the set of positive integers. They are applicable for the situations when the data to be modeled originate from a mechanism that generates data excluding zero counts. Among others the Zero-truncated Poisson (ZTPD), Zero-truncated Poisson – Lindley (ZTPLD), Zero-truncated Negative Binomial (ZTNBD), Zero-truncated Negative Binomial-Erlang (ZTNB-ELD), Zero-truncated Negative Binomial – Beta-Exponential (ZTNB-BED), Zero-truncated Poisson – Exponential-Gamma (ZTPEGD) and Zero-truncated Poisson – Exponential (ZTPED) distributions. In this paper, these distributions were applied to real-life datasets from different fields of study. Their Goodness-of-fit are discussed based on their values of logliklihood ($-2\log Lik$), Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). These were achieved by estimating the parameters of the distributions using the real-life datasets considered. ZTPEGD distribution gave a closer fit to the Stillbirths dataset, ZTNB-ELD distribution had the minimum values of AIC and BIC for the Methamphetamine treatments dataset and ZTNB-BED distribution gave a closer fit to the Accident dataset with minimum values of $-2loglik$ as compared to the competing distributions. On the basis of this, it is concluded that these distributions can serve as important alternatives to other models that are often considered for modelling real life data without zero counts.**

.

Keywords - *Zero-truncated, Negative Binomial, Erlang, Beta-Exponential, Poisson-Exponential, Poisson – Exponential-Gamma, Poisson-Lindley, Goodness-of-fit.*

I. INTRODUCTION

The Poisson and its mixture distributions are suitable statistical models for the situations where events occur independently at random. According to Best et al., (2007), sufficient conditions are given for the models to be applicable.

For fitting Poisson distribution to count data, equality of means and variance should be satisfied. But this condition is rarely satisfied completely in some instances, especially in biological and medical sciences (Shanker et al., 2015) where zero doesn't count. These may be modeled using a Zero-truncated Poisson distribution. For instance, among women of reproductive ages that experience stillbirths, the number of prenatal deaths is not Poisson distributed because it cannot be zero (i.e. no birth was still). The distribution may be modeled using Zero-truncated distributions.

Zero-truncated distributions are certain discrete distributions having support the set of positive integers (Shanker & Shukla, 2018). These distributions are applicable for the situations when the data to be modeled originate from a mechanism that generates data excluding zero counts.

Zero-truncated Poisson distribution (ZTPD) also known as the Conditional Poisson distribution (Cohen, 1960) or the Positive Poisson distribution (Singh, 1978) is one of these distributions. It is the conditional probability distribution of a Poisson-distributed random variable given that the value of the random variable is not zero. Thus, it is impossible for a Zero-truncated Poisson random variable to be zero. As a result of this, a number of works focused on proposing methods that modeled and give quite satisfactory fit in the datasets. Among these are: the Zero-truncated

Poisson Distribution (ZTPD, David & Johnson, 1952; Placket, 1953), Zero-truncated Poisson – Lindley Distribution (ZTPLD, David & Johnson, 1952), Zero-truncated Negative Binomial Distribution (ZTNBD, Claude et al, 2014; Cruyff & Van Der Heijden, 2008), Zero-truncated Negative Binomial Erlang Distribution (ZTNB-EL, Bodhisuwan et al, 2017), Zero-truncated Negative Binomial – Beta-Exponential Distribution (ZTNB-BED, Sookua et al, 2018), Zero-truncated Poisson – Exponential-Gamma (ZTPEGD, Umar, 2019) and Zero-truncated Poisson – Exponential Distribution (ZTPED, Umar, 2019) among others.

The probability mass function of Zero-truncated Poisson-Lindley distribution (ZTPLD) was derived by David & Johnson (1952), and Ghitaney et al (2008) showed that the distribution can also arise from the size biased Poisson distribution (SBPD). Shanker et al., (2015) observed that ZTPLD gives much closer fit than ZTPD in almost all cases of Biological Sciences, Mortality and Migration datasets while ZTPD has an advantage over ZTPLD for modeling Zero-truncated count data in Social Sciences (Shanker & Hagos, 2016a, 2016b).

Bodhisuwan et al., (2017) introduced the Zero-truncated Negative Binomial-Erlang distribution. The probability mass function and some important properties were derived and the parameters estimated using the Maximum Likelihood Estimation. The distribution was applied to a real-life dataset; the number of methamphetamine in Bangkok, Thailand. It was shown that the distribution provided a better fit than the Zero-truncated Poisson, Zero-truncated Negative Binomial (Sampford, 1955), Zero-truncated Generalized Negative Binomial (Ding & Noop, 2015) and Zero-truncated Poisson-Lindley distributions for this data. Sookkua et al., (2018) derived a Zero-truncated Negative Binomial – Beta-Exponential distribution. Some mathematical properties of the distribution were derived accordingly. Its parameters were estimated using the Maximum Likelihood Estimate. The distribution was applied to a real life dataset from the Department of highways, Ministry of Transport, Bangkok, Thailand. The distribution was found to be better than the ZTPD and ZTNBD.

Analysis and modeling of real life count data is crucial in applied sciences, community health and other fields of knowledge. This has attracted the attention and interest of researchers all over the world. Thus, number methods were introduced to facilitate better modeling and significant progress (Asad et al, 2018). These methods have been shown to perform better than one another in the various fields tested. However, this work is carried out to apply some of these methods to real life count datasets especially from community health and other fields of knowledge.

II. Methodology

The Zero-truncated Poisson distribution (ZTPD) is defined (David & Johnson, 1952; Plackett, 1953) as:

$$P(x;\theta) = \frac{\theta^x}{(e^{\theta}-1)x!};\ 1,2,3,\dots,\theta > 0 \qquad (1)$$

An extension of this distribution and estimation when zeros are excluded were discussed by Cohen (1960a, 1960b).
The Zero-truncated Negative Binomial - Erlang distribution (ZTNB-ELD) is defined (Bodhisuwan et al, 2017) as:

$$f(x;r,c,k) = \frac{\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\left(\frac{c}{c+r+j}\right)^k}{1-\left(\frac{c}{c+r}\right)^k},\ x = 1,2,3,\dots,r,c,k > 0 \qquad (2)$$

The Zero-truncated Negative Binomial – Beta-Exponential distribution (ZTNB-BED) is defined (Sookua et al, 2018) as:

$$f(x;r,a,b,c) = \frac{\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j B\left(b+\frac{r+j}{c},a\right)}{1-\frac{B\left(b+\frac{r}{c},a\right)}{B(a,b)}},\ x = 1,2,3,\dots,r,a,b,c > 0 \qquad (3)$$

The Zero-truncated Poisson – Exponential-Gamma distribution is defined (Umar, 2019) as;

$$P(x;\alpha,\theta) = \frac{\theta^2(\theta+1)^{\alpha}x! + \theta^{\alpha}(\theta+1)\Gamma(\alpha+x)}{\{\Gamma(\alpha)(\theta+1)[(\theta+1)^{\alpha}-\theta^{\alpha}]+\theta(\theta+1)^{\alpha}\}(\theta+1)^x x!};\ x = 1,2,3,\dots,\alpha > 0,\theta > 0 \qquad (4)$$

The expression in (5) reduces to the Zero-truncated Poisson-Lindley distribution, (Ghitaney et al., 2008; Shanker & Hagos, 2016) when $\alpha = 2$ and a Zero-truncated Poisson-Exponential (Umar, 2019) when $\alpha = 1$.
That is;

$$P(x;2,\theta) = \frac{\theta^2(x+\theta+2)}{(\theta^2+3\theta+1)(\theta+1)^x};\ x = 1,2,3,\dots,\theta > 0 \qquad (5)$$

and

$$P(x;1,\theta) = \frac{\theta}{(\theta+1)^x};\ x = 1,2,3,\dots,\theta > 0 \qquad (6)$$

The graphs of these distributions for different values of their parameter(s), important properties, parameter(s) estimation, goodness-of-fit and applications were discussed accordingly.

The Zero-truncated Poisson-Lindley distribution (ZTPLD) has been studied (Shanker et al., 2015) as a special case of the Zero-truncated Two-Parameter Poisson-Lindley Distribution, ZTTPPLD, (Shanker & Shukla, 2017a), Zero-truncated New Quasi Poisson-Lindley Distribution, ZTNQPL, (Shanker & Shukla, 2017b) and Zero-truncated Discrete Two-Parameters Poisson-Lindley Distribution, ZTDTPPLD, (Shanker & Shukla, 2018) and these distributions were found to perform better than Zero-truncated Poisson and Zero-truncated Poisson – Lindley distributions in different fields of knowledge.

III. Applications

In this section, the goodness-of-fit of the models above is discussed with an application to real-life datasets. The parameters of the models were estimated using the MLE method while the goodness-of-fit was evaluated using the Akaike Information Criterion (AIC, Akaike, 1974), Bayesian Information Criterion (BIC, Schwarz, 1978) and –2logLik with their respective statistics given below.

$$AIC = -2\ln L + 2k \quad (7)$$

$$BIC = -2\ln L + k\ln n \quad (8)$$

where k is the number of parameters and n is the sample size. The distribution that has a lower value of these criteria is judged to be the best among others.

3.1 Data Description

Dataset 1: These data contain the number of mothers with at least one Stillbirth in the Urban areas of Nigeria (NDHS, 2013).

Dataset 2: These data present the number of Methamphetamine from the Office of the Narcotics Control Board (ONCB), Thailand in a Bangkok metropolitan region (Bodhisuwan et al, 2017).

Dataset 3: The data consist of the number of injured from an accident on major road in Bangkok, Thailand, collected from the Department of Highways, Ministry of Transport, Thailand (Sookua et al, 2018).

Table 1: Observed and Expected number of Mothers with at least one stillbirth in the Urban areas of Nigeria

Number of Stillbirths	Observed number of Mothers	Expected Frequency			
		ZTPD	ZTPED	ZTPLD	ZTPEGD
1	453	456.9	451.1	451.5	453.3
2	139	157.6	141.8	143.5	132.7
3	49	36.3	44.6	44.3	45.6
4	12	6.3	14.0	13.5	16.5
5	5	0.9	4.4	4.1	6.1
TOTAL	658	658	655.9	656.9	654.2
ML Estimates		$\hat{\theta} = 0.69$	$\hat{\theta} = 2.18$	$\hat{\theta} = 3.21$	$\hat{\alpha} = 0.03$ $\hat{\theta} = 1.69$
-2logLik		223.74	161.98	183.26	149.68
AIC		225.74	163.98	185.26	153.68
BIC		230.23	168.47	189.75	162.66

The observed and expected Numbers of mothers with at least one Stillbirth in the Urban areas of Nigeria are presented in table 1. The expected frequencies are given by the ZTPD, ZTPED, ZTPLD and ZTPEGD distributions. It can be seen that the ZTPEGD gave much closer fit than than the ZTPD, ZTPLD and ZTPED. This has also been confirmed by the values of AIC, BIC and the following graph.

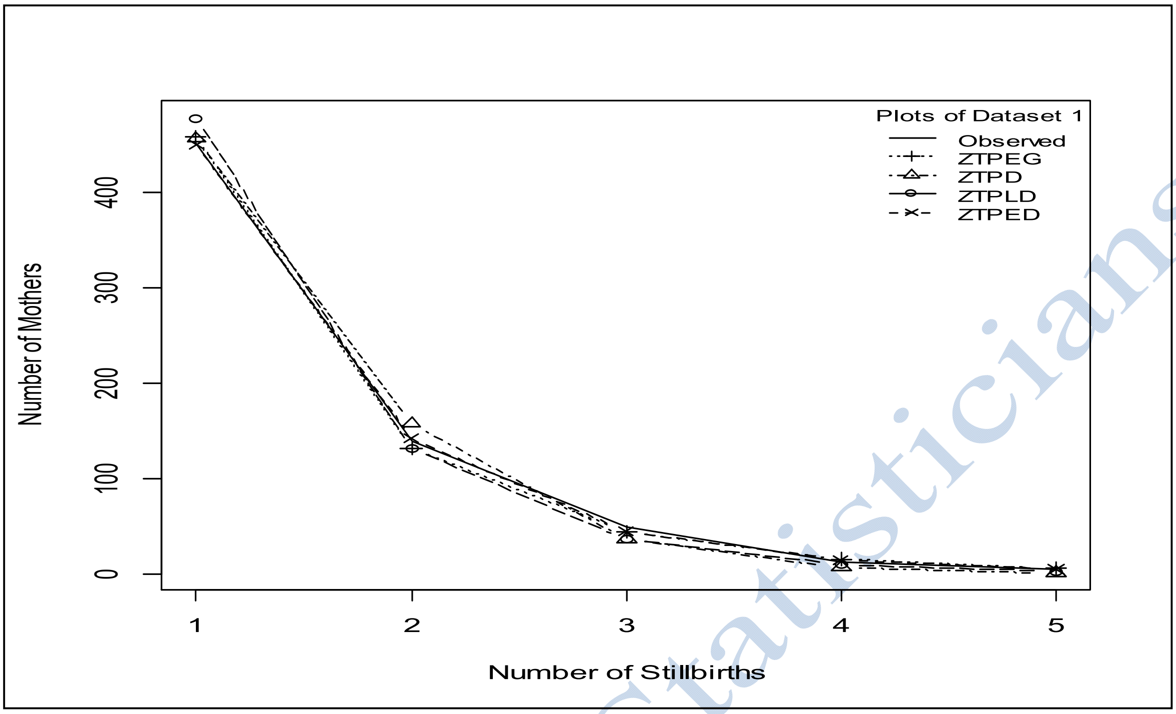

Figure 1: Graph of the observed and expected number of Mothers with at least one stillbirth

Table 2: Observed and Expected Frequencies of the Methamphetamine dataset

The number Treatments	Observed Frequency	Expected Frequency				
		ZTPD	ZTPED	ZTPLD	ZTNBD	ZTNB-ELD
1	3114	2027.3	3105.2	3114.4	2982.5	3071.4
2	163	942.7	222.6	214.7	311.7	229.7
3	23	292.2	16.0	14.8	43.0	34.0
4	20	67.9	1.1	1.0	6.6	7.2
5	9	12.6	0.8	0.7	1.2	1.9
6	3	2.0	0.0	0.0	0.3	0.6
7	3	0.3	0.0	0.0	0.0	0.2
8	3	0.0	0.0	0.0	0.0	0.1
9	4	0.0	0.0	0.0	0.0	0.0
10	3	0.0	0.0	0.0	0.0	0.0
TOTAL	3345	3345	3345.7	3345.6	3345.3	3345.1
ML Estimates		$\hat{\theta} = 0.93$	$\hat{\theta} = 12.95$	$\hat{\theta} = 14.34$	$\hat{r} = 0.03$ $\hat{p} = 0.79$	$\hat{r} = 0.51$ $\hat{c} = 0.79$ $\hat{k} = 0.05$
-2logLik		2880.90	2700.89	2689.32	2509.99	2219.24
AIC		2882.90	2702.89	2691.32	2513.99	2225.24
BIC		2889.01	2709.00	2697.44	2517.04	2229.82

In table 2, the estimates of the ZTPD, ZTPED, ZTPLD, ZTNBD and ZTNB-ELD distributions are obtained using the observed number of Methamphetamine treatments dataset. The goodness-of-fit of these distributions has been compared and presented in the table and garaphically as follows.

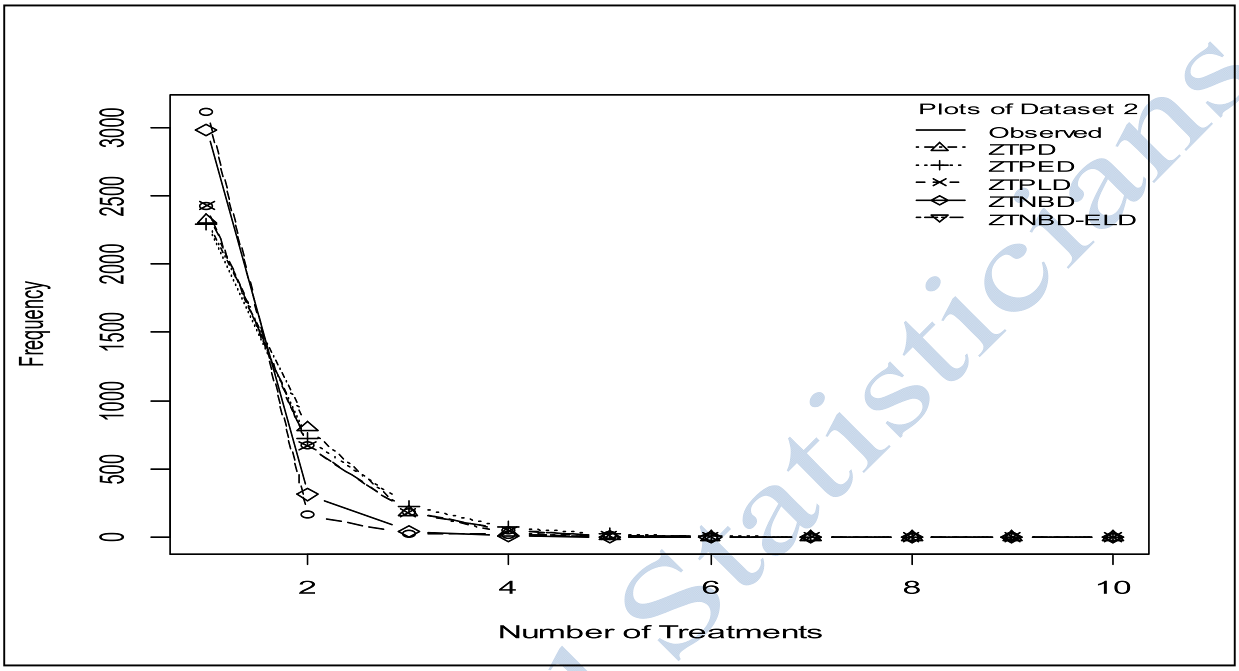

Figure 2: Graph of the observed and expected number of Methamphetamine Treatments

Table 3 contains the estimates of the ZTPD, ZTPED, ZTPLD, ZTNBD and ZTNB-ELD distributions obtained from the number of injured in an accident. The goodness-of-fit for these distributions has been compared and presented in the table and garaphically in figure 3.

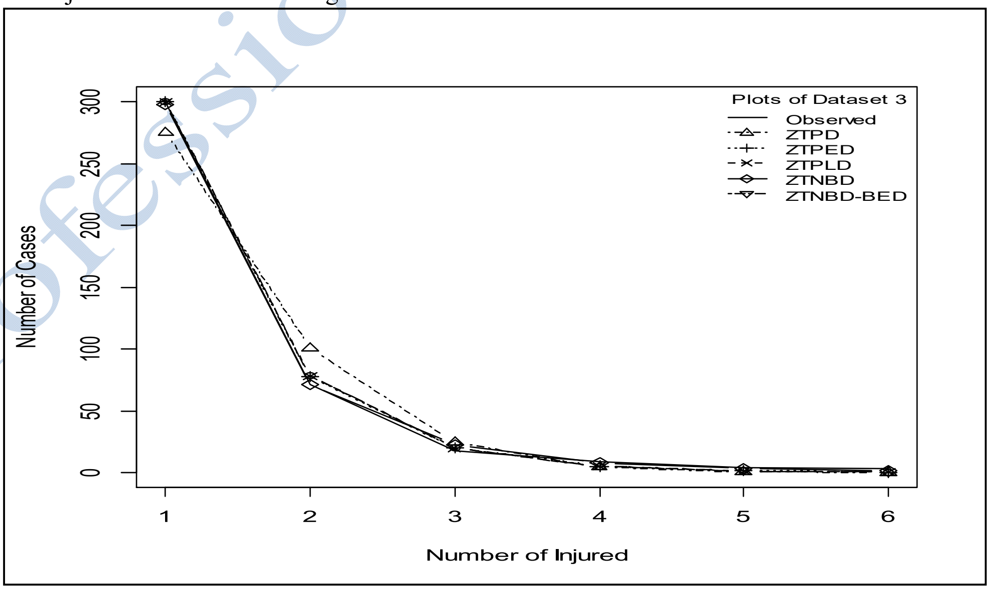

Figure 3: Graph of the observed and expected number of Injured in the Accident cases

Table 3: Observed and Expected Frequencies of the Accident dataset

The number Injured	Observed Frequency	Expected Frequency				
		ZTPD	ZTPED	ZTPLD	ZTNBD	ZTNB-BED
1	300	275.0	300.3	300.0	298.0	299.8
2	71	100.4	77.6	78.2	71.1	69.9
3	18	24.4	20.1	20.0	22.6	21.7
4	9	4.5	5.2	5.1	8.1	7.8
5	4	0.7	1.3	1.3	3.1	3.1
6	3	0.1	0.3	0.3	1.2	1.4
TOTAL	405	405.1	404.8	404.9	404.1	403.7
ML Estimates		$\hat{\theta} = 0.73$	$\hat{\theta} = 2.87$	$\hat{\theta} = 3.43$	$\hat{r} = 0.01$ $\hat{p} = 0.52$	$\hat{r} = 0.15$ $\hat{a} = 8.40$ $\hat{b} = 3.68$ $\hat{c} = 2.61$
-2logLik		715.32	688.72	687.36	679.23	679.37
AIC		717.32	690.72	689.36	683.23	687.37
BIC		721.32	694.72	693.85	691.24	703.39

IV. Conclusion

The Zero-truncated distributions considered in this study were estimated and compared using real life datasets. It is obvious that the expected frequencies given by the ZTPEG distribution are closer to the observed frequencies of the Stillbirths dataset, ZTNB-EL distribution has the minimum values of AIC and BIC for the Methamphetamine treatments dataset and ZTNB-BE distribution gave more closer fit to the Accident dataset with minimum values of $-2loglik$ as compared to the competing distributions. These distributions can therefore be considered important alternatives to modeling real life datasets.

References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, AC- **19***, 716-723.

Asad, A., Qaisar, R., Muhammad, Z. and Muhammad, T. J. (2018). A Quasi Lindley Pareto distribution. *Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Science.* ***55***(2): 32 – 40.

Best, D. J., Rayner, J. C. W. and Thas, O. (2007). Goodness of fit for the Zero-truncated Poisson distribution. *Journal of Statistical Computation and Simulation. 77*(7), 585-591. http://www.tandf.co.uk/journals//doi:10.1080/10629360600569329

Bodhisuwan, W., Pudprommarat, C., Bodhisuwan, R. and Saothayanun, L. (2017). Zero-Truncated Negative Binomial - Erlang Distribution. *AIP Conference Proceedings **1905**, 050011 (2017)*; American Institute of Physics. https://doi.org/10.1063/1.5012230

Claude, T. A., Karina, P. S. S. & Lucas, N. B. (2014). Zero-Truncated Negative Binomial applied to Nonlinear Data. *JP Journal of Biostatistics, **11***(1); 55-67.

Cohen, A. C. (1960). Estimating parameters in a conditional Poisson distribution. *Biometrics. **16***: 203–211. http://doi:10.2307/2527552

Cohen, A. C. (1960a). An extension of a truncated Poisson distribution. *Biometrics **16***(3): 446-450.

Cohen, A. C. (1960b). Estimation in a truncated Poisson distribution when zeros and some ones are missing. *Journal of American Statistical Association **55***: 342-348.

Cruyff, M. J. L. F. & Van Der Heijden, P. G. M. (2008). Point and interval estimation of the population size using a Zero-truncated Negative Binomial Regression Model, *Biometrics, **50***, 1035-1050.

David, F., N. & Johnson, N. L. (1952). The truncated Poisson. *Biometrics **8***: 275-285.

Ghitany, M. E., Mutairi, D. K. & Nadarajah, S. (2008). Zero-truncated Poisson-Lindley distribution and its Applications. *Mathematics and Computers in Simulation*, ***79***(3): 279–287.

Plackett, R. L. (1953). The truncated Poisson distribution. *Biometrics **9***(4): 485-488.

Schwarz, G. (1978). Noop. *The Annals of Statistics **6***, 461–464

Shanker, R. and Hagos, F. (2016a): On Zero-truncation of Poisson, Poisson-Lindley and Poisson-Sujatha Distributions and their Applications, *Biometrics & Biostatistics International Journal*, ***3***(5), 1-13.

Shanker, R. and Shukla, K. K. (2017a). A Zero-truncated Two-Parameter Poisson-Lindley distribution with an Application to Biological Science. *Turkey Klinikleri Journal of Biostatistics*. ***9***(2): 85-95.

Shanker, R. and Shukla, K. K. (2017b). A Zero-truncated New Quasi Poisson-Lindley distribution and its Applications. *The Statistician*. ***66***(2): 33-46.

Shanker, R. and Shukla, K. K. (2018). A Zero-truncated Discrete Two-Parameter Poisson-Lindley distribution with Applications. *Journal of Institute of Science and Technology*. ***22***(2): 76-85.

Shanker, R., Hagos, F., Sujatha, S. and Abrehe, Y. (2015). On Zero-truncation of Poisson and Poisson-Lindley distributions and Their Applications. *Biometrics & Biostatistics International Journal, **2***(6): 1-14.

Singh, J. (1978). A characterization of positive Poisson distribution and its application. *SIAM Journal on Applied Mathematics*. ***34***: 545–548. http://doi:10.1137/0134043

Sookkua, S., Wongoutong, C. and Bodhisuwan, W. (2018). Zero-Truncated Negative Binomial-Beta Exponential Distribution. *RTUNC 3rd National Conference May 25, 2018.* Ubonratchathani, Thailand. Pp 229 – 237.

Umar, M. A. (2019). A Zero-truncated Poisson – Exponential-Gamma Distribution and its Applications. *An M.Sc. dissertation Submitted to the Department of Statistics, University of Ilorin, Ilorin, Nigeria*. Unpublished.