A Zero-Truncated Discrete Lindley Distribution with Applications

Simon Sium and Rama Shanker*

Department of Statistics, College of Science, Eritrea Institute of Technology, Asmara, Eritrea

Abstract: A zero-truncated discrete Lindley distribution has been proposed and studied. The generating functions, moments about origin and moments about the mean have been obtained. Natures of coefficient of variation, skewness, kurtosis, and the index of dispersion of the distribution have been explained. It has been observed that both the method of moment and the method of maximum likelihood estimation results into the same equation and hence give the same estimator of the parameter. Applications of the distribution have been explained through three examples of observed real datasets from biological sciences and demography and its goodness of fit has been found quite satisfactory over zero-truncated Poisson distribution and zero-truncated Poisson-Lindley distribution.

Keywords: Zero-truncated distribution, Discrete lindley distribution, Moments based measures, Estimation, Applications.

1. INTRODUCTION

Let $P_0(x;\theta)$ is the original discrete distribution. Then the zero-truncated version of $P_0(x;\theta)$ can be defined as

$$P(x;\theta)=\frac{P_0(x;\theta)}{1-P_0(0;\theta)}\quad ;x=1,2,3,... \tag{1.1}$$

Generally, zero-truncated distributions are used to model count data when the data originate from a mechanism generating excluding zero counts. Zero excluding data can be generated if the recording mechanism is not activated unless at least one event occurs. There are many real life situations where the frequency of zero counts cannot be observed in the random experiment. For example, if a survey is related to the number of household having at least one migrant, then the frequency of zero migrant is not possible. Similarly if the data is regarding number of counts of flower heads having at least one fly egg, then the frequency of zero fly egg is not possible. Best *et al.* [1] have detailed discussion on applications and goodness of fit for the zero-truncated Poisson distribution. Coleman and James [2], Finney and Varley [3], Mathews and Appleton [4], are some among others who have discussed various applications of zero-truncated distributions particularly zero-truncated Poisson distribution.

Recently Berhane and Shanker [5] introduced a discrete Lindley distribution (DLD), a discrete analogue of continuous Lindley distribution, using infinite series approach of discretization and defined by its probability mass function (pmf)

$$P_0(x;\theta)=\frac{\left(e^{\theta}-1\right)^2}{e^{2\theta}}(1+x)e^{-\theta x};\ x=0,1,2,3,....,\theta>0 \tag{1.2}$$

Berhane and Shanker [5] have obtained moments and moments based measures of DLD and discussed the behaviors of coefficient of variation, skewness, kurtosis and index of dispersion. They have shown that both the method of moment and the method of maximum likelihood give the same estimator of the parameter θ. Further, Berhane and Shanker [5] have discussed its applications to model count data from biological sciences and showed that it gives better fit than Poisson and Poisson-Lindley distributions. The Lindley distribution, introduced by Lindley [6] is defined by its probability density function (pdf)

$$f(x;\theta)=\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}\quad ;x>0,\ \theta>0 \tag{1.3}$$

Statistical properties, estimation of parameter using maximum likelihood and application of Lindley distribution for waiting time data in a bank have been discussed by Ghitany *et al.* [7]. Shanker *et al.* [8] have comparative study on modeling of several lifetime datasets from engineering and biomedical sciences using one parameter Lindley and exponential distributions and observed that in many datasets exponential distribution gives better fit than Lindley distribution and in some datasets Lindley distribution gives better fit than exponential distribution. Sankaran [9] proposed Poisson-Lindley distribution (PLD), a Poisson mixture of Lindley distribution, defined by its pmf

*Address correspondence to this author at the Department of Statistics, College of Science, Eritrea Institute of Technology, Asmara, Eritrea;
E-mail: cimonesium@gmail.com ; shankerrama2009@gmail.com

$$P_1(x;\theta)=\frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}}\quad ;x=0,1,2,...,\theta>0 \qquad (1.4)$$

Sankaran [9] has obtained its moments and discussed estimation of parameter using both the method of moment and the maximum likelihood and applications to model count data. Shanker and Hagos [10] have detailed study on applications of PLD in biological sciences and concluded that PLD is a suitable choice over Poisson distribution to model count data in biological sciences because in general biological sciences data are over-dispersed and PLD is also an over-dispersed distribution.

The pmf of zero-truncated Poisson- Lindley distribution (ZTPLD) given by

$$P_2(x;\theta)=\frac{\theta^2}{\theta^2+3\theta+1}\frac{x+\theta+2}{(\theta+1)^x}\quad ;x=1,2,3,....,\theta>0 \qquad (1.5)$$

has been introduced by Ghitany *et al.* [11]. Statistical properties, estimation of parameter using maximum likelihood and method of moment along with applications for count data excluding zero counts have been discussed by Ghitany *et al.* [11]. Shanker *et al.* [12] have comparative study on applications of zero-truncated Poisson distribution (ZTPD) and ZTPLD in demography, biological sciences and social sciences and showed that ZTPLD gives better fit than ZTPD in demography and biological sciences whereas ZTPD gives better fit than ZTPLD in social sciences.

The pmf of zero-truncated Poisson-distribution (ZTPD) is given by

$$P_3(x;\theta)=\frac{e^{-\theta}\,\theta^x}{(1-e^{-\theta})\,x!}\quad ;x=1,2,3,....,\theta>0 \qquad (1.6)$$

In this paper, a zero-truncated discrete Lindley distribution (ZTDLD) has been suggested. Its generating functions, moments about origin and moments about mean have been obtained. Expressions for coefficient of variation, skewness, kurtosis, and index of dispersion have been given and their behaviors have been studied numerically and graphically for varying values of parameter. Estimation of its parameter has been discussed using the method of moment and the maximum likelihood. Finally, applications of ZTDLD to three observed real datasets have been given to test its goodness of fit over ZTPD and ZTPLD.

2. A ZERO-TRUNCATED DISCRETE LINDLEY DISTRIBUTION

The pmf of zero-truncated discrete Lindley distribution (ZTDLD), using (1.1) and (1.2), can be expressed as

$$P_4(x;\theta)=\frac{(e^\theta-1)^2}{2e^\theta-1}(1+x)\,e^{-\theta x}\quad ;x=1,2,3,...,\ \theta>0 \qquad (2.1)$$

The behavior of the ZTDLD for varying values of parameter θ has been shown in Figure **1**.

It is obvious that $\frac{P_4(x+1;\theta)}{P_4(x;\theta)}=\frac{1}{e^\theta}\left(1+\frac{1}{x+1}\right)$ is a decreasing function of $x>0$ and hence $P_4(x;\theta)$ is log-concave. This shows that ZTDLD is unimodal, has increasing failure rate (IFR), and hence increasing failure rate average (IFRA). It can be also shown that it is new better than used (NBU), new better than used in expectation (NBUE), and has decreasing mean residual life (DMRL). The relationships of these aging concepts are available in Barlow and Proschan [13].

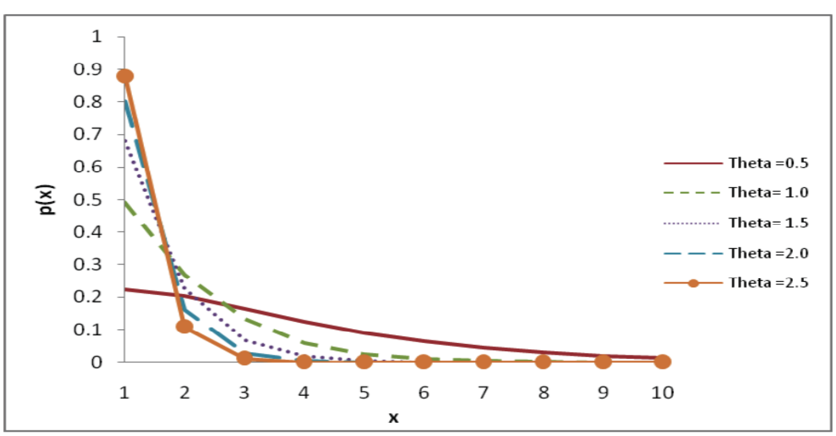


Figure 1: Behavior of ZTDLD for varying values of parameter θ .

3. MOMENTS AND MOMENTS BASED MEASURES

The generating functions namely probability generating function (G (t)) and the moment generating function (M (t)) of ZTDLD can be obtained as

$$G(t)=\frac{(e^{\theta}-1)^2}{2e^{\theta}-1}\frac{(2te^{\theta}-t^2)}{\left(e^{\theta}-t\right)^2} \quad \text{for } t\neq e^{\theta}, \tag{3.1}$$

and

$$M(t)=\frac{(e^{\theta}-1)^2}{2e^{\theta}-1}\frac{(2e^{t}e^{\theta}-e^{2t})}{\left(e^{\theta}-e^{t}\right)^2} \quad \text{for } t\neq \theta. \tag{3.2}$$

It can be easily verified that the mgf in (3.2) is infinitely differentiable with respect to t, since it involves exponential terms of its argument. This means that all moments about origin μ_r' $,r\geq 1$ of ZTDLD can be obtained from its mgf. The first four moments about origin of ZTDLD can thus be obtained as

$$\mu_1'=\frac{2e^{2\theta}}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)}$$

$$\mu_2'=\frac{2e^{2\theta}\left(e^{\theta}+2\right)}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)^2}$$

$$\mu_3'=\frac{2e^{2\theta}\left(e^{2\theta}+7e^{\theta}+4\right)}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)^3}$$

$$\mu_4'=\frac{2e^{2\theta}\left(e^{3\theta}+18e^{2\theta}+33e^{\theta}+8\right)}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)^4}$$

Now using the relationship between moments about the mean and moments about the origin, the moments about the mean of ZTDLD are thus given by

$$\mu_2=\sigma^2=\frac{2e^{2\theta}\left(3e^{\theta}-2\right)}{\left(2e^{\theta}-1\right)^2\left(e^{\theta}-1\right)^2}$$

$$\mu_3=\frac{2e^{2\theta}\left(6e^{3\theta}+e^{2\theta}-9e^{\theta}+4\right)}{\left(2e^{\theta}-1\right)^3\left(e^{\theta}-1\right)^3}$$

$$\mu_4=\frac{2e^{2\theta}\left(12e^{5\theta}+94e^{4\theta}-153e^{3\theta}+52e^{2\theta}+15e^{\theta}-8\right)}{\left(2e^{\theta}-1\right)^4\left(e^{\theta}-1\right)^4}$$

Finally, the coefficient of variation (C.V), coefficient of Skewness $\left(\sqrt{\beta_1}\right)$, and coefficient of Kurtosis $\left(\beta_2\right)$ and index of dispersion $\left(\gamma\right)$ of ZTDLD are obtained as

$$C.V.=\frac{\sigma}{\mu_1'}=\frac{\sqrt{2\left(3e^{3\theta}-2e^{2\theta}\right)}}{2e^{2\theta}}=\frac{e^{-\theta}\sqrt{\left(6e^{\theta}-4\right)}}{2}$$

$$\sqrt{\beta_1}=\frac{\mu_3}{\left(\mu_2\right)^{3/2}}=\frac{2e^{2\theta}\left(6e^{3\theta}+e^{2\theta}-9e^{\theta}+4\right)}{\left\{2\left(3e^{3\theta}-2e^{2\theta}\right)\right\}^{3/2}}$$

$$\beta_2=\frac{\mu_4}{\mu_2^{\,2}}=\frac{\left(12e^{5\theta}+94e^{4\theta}-153e^{3\theta}+52e^{2\theta}+15e^{\theta}-8\right)}{e^{2\theta}\left(3e^{\theta}-2\right)^2}$$

$$\gamma=\frac{\sigma^2}{\mu_1'}=\frac{3e^{\theta}-2}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)}.$$

Table 1: Behavior of Descriptive Statistics of ZTDLD for Values of the Parameter θ

θ	Mean	Variance	C.V	Skewness	Kurtosis	ID
0.5	3.6477	7.2107	0.7361	1.5919	6.6360	1.9768
1.0	1.9386	1.5651	0.6454	1.8303	7.6043	0.8074
1.5	1.4489	0.5981	0.5338	2.1346	8.9225	0.4128
2.0	1.2405	0.2842	0.4298	2.5493	10.9373	0.2291
2.5	1.1361	0.1502	0.3412	3.1131	14.1750	0.1322
3.0	1.0793	0.0841	0.2687	3.8673	19.4675	0.0779
3.5	1.0469	0.0486	0.2107	4.8620	28.1679	0.0465
4.0	1.0281	0.0287	0.1647	6.1610	42.4978	0.0279
4.5	1.0169	0.0171	0.1286	7.8466	66.1152	0.0168
5.0	1.0102	0.0103	0.1003	10.0249	105.0488	0.0102

The behaviors of mean, variance, C.V, Skewness, Kurtosis and index of dispersion of ZTDLD for some selected values of parameter θ are shown numerically in Table **1**.

It is clear from Table **1** that mean, variance, coefficient of variation and index of dispersion of ZTDLD are decreasing while coefficients of skewness and kurtosis of ZTDLD are increasing for increasing values of parameter θ. The behavior of coefficient of variation, skewness, kurtosis, index of dispersion and relationship between mean and variance of ZTDLD for varying values of parameter θ are presented in Figure **2**.

The following Table **2** presents the dispersion (over-dispersed, equi-dispersed, and under-dispersed) of ZTDLD and ZTPLD for values of parameter θ.

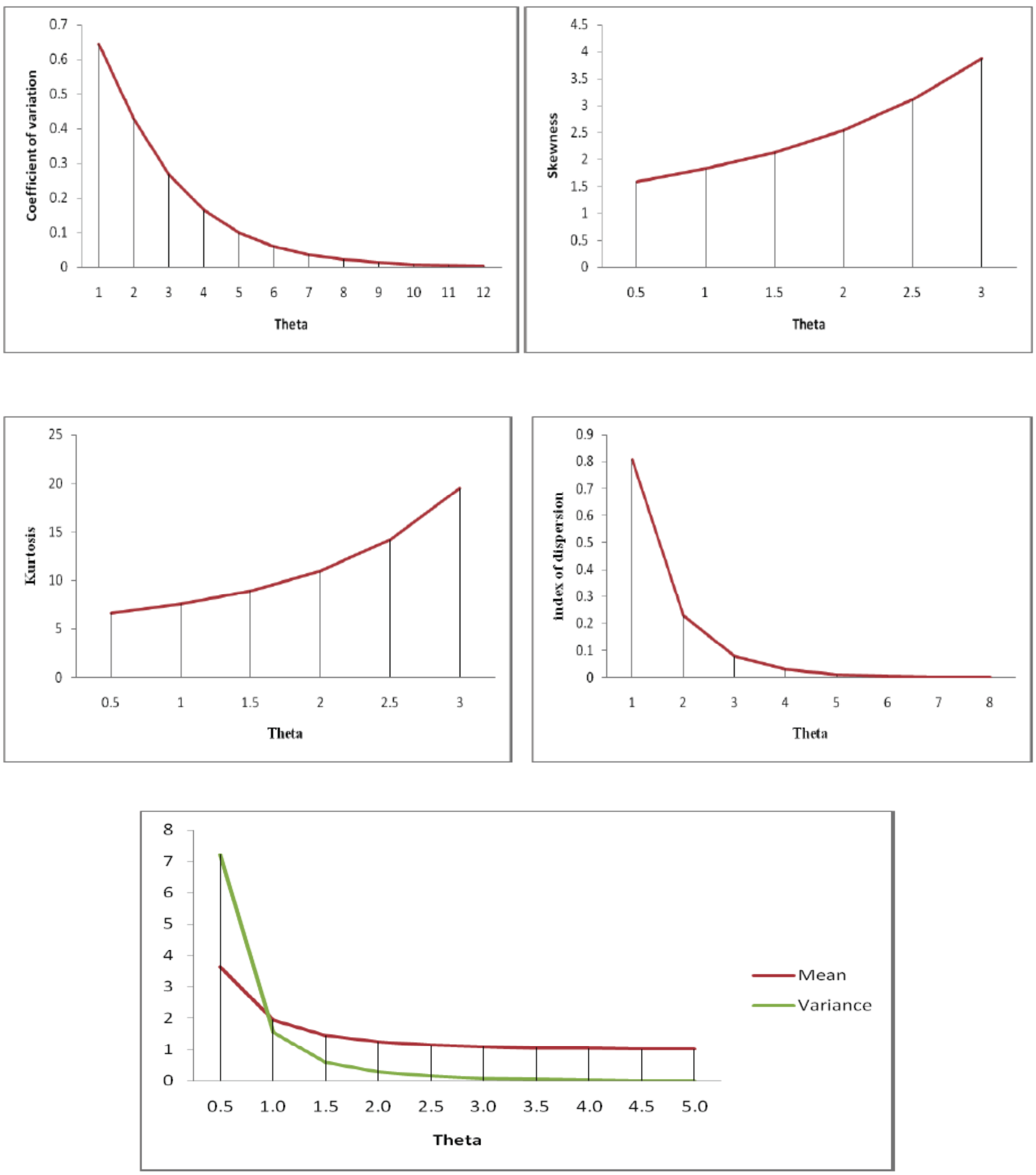


Figure 2: Behavior of coefficient of variation, skewness, kurtosis, index of dispersion and relationship between mean and variance of ZTDLD for varying values of parameter θ.

Table 2: Dispersion of ZTDLD and ZTPLD

Distributions	Over-dispersion $(\mu < \sigma^2)$	Equi-dispersion $(\mu = \sigma^2)$	Under-dispersion $(\mu > \sigma^2)$
ZTDLD	$\theta < 0.8612$	$\theta = 0.8612$	$\theta > 0.8612$
ZTPLD	$\theta < 1.2586$	$\theta = 1.2586$	$\theta > 1.2586$

4. ESTIMATION

4.1. Method of Moment Estimate (MOME)

Since ZTDLD has one parameter, equating the population mean to the corresponding sample mean, the MOME $\tilde{\theta}$ of parameter θ of ZTDLD results in the solution of the following non linear equation

$$2\left(\bar{x}-1\right)e^{2\theta}-3\bar{x}\,e^{\theta}+\bar{x}=0\,,$$

where $\bar{x}$ being the sample mean.

4.2. Maximum Likelihood Estimate (MLE)

Assuming $\left(x_1, x_2, ..., x_n\right)$ be a random sample of size n from the ZTDLD, the likelihood function L of the ZTDLD can be expressed as

$$L=\left(\frac{(e^{\theta}-1)^2}{2e^{\theta}-1}\right)^n \prod_{i=1}^{n}\left(1+x_i\right)e^{-n\theta\bar{x}}\,.$$

The log likelihood function is thus given by

$$\ln L = n\left\{2\ln\left(e^{\theta}-1\right)-\ln\left(2e^{\theta}-1\right)\right\}-\sum_{i=1}^{n}\ln\left(1+x_i\right)-n\theta\bar{x}$$

The maximum likelihood estimate (MLE) $\hat{\theta}$ of the parameter θ of ZTDLD is the solution of the following log likelihood equation

$$\frac{d\ln L}{d\theta}=\frac{2ne^{\theta}}{e^{\theta}-1}-\frac{2ne^{\theta}}{2e^{\theta}-1}-n\bar{x}=0$$

This gives $2\left(\bar{x}-1\right)e^{2\theta}-3\bar{x}\,e^{\theta}+\bar{x}=0$, where $\bar{x}$ being the sample mean.

It is obvious that both the method of moment and the method of maximum likelihood give the same estimating equation and hence the same estimator of the parameter θ. This non-linear equation can be solved by any numerical iteration methods such as Newton- Raphson method, Bisection method, Regula – Falsi method etc. In this paper, Newton- Raphson method has been used to solve above equation.

5. GOODNESS OF FIT TO REAL DATASETS

In this section applications of ZTDLD has been discussed with three examples of observed real datasets and its goodness of fit has been compared with that of ZTPD and ZTPLD. Note that maximum likelihood estimate of the parameter has been used to fit ZTPD, ZTPLD and ZTDLD. The three examples of real datasets are consisting of two from biological sciences and one from demography. The dataset in Table **3** is regarding the number of counts of flower heads as per the number of fly eggs and are available in Finney and Varley [3]. The dataset in Table **4** is regarding the number of snowhares counts captured over 7 days and are available in Keith and Meslow [14]. The dataset in Table **5** is due to Singh and Yadav [15] regarding the number of households having at least one migrant from households according to the number of migrants. The mean and variance of the datasets in Tables **3**, **4**, and **5** are $\left(\bar{X}=3.03409, \sigma^2=1.08946\right)$, $\left(\bar{X}=1.57576, \sigma^2=0.48113\right)$, $\left(\bar{X}=1.54746, \sigma^2=0.16092\right)$ respectively. Since the mean is greater than the variance, the datasets in Tables **3**, **4** and **5** are under-dispersed. It is evident from the values of Chi-square $\left(\chi^2\right)$ that ZTDLD gives better fit than ZTPD and ZTPLD and thus ZTDLD can be well thought out zero-truncated discrete distribution for modeling count data excluding zero count over ZTPD and ZTPLD. Further, the fitted plots of ZTPD, ZTPLD and ZTDLD reveal that ZTDLD gives much closer fit for the considered datasets.

The fitted plot of ZTPD, ZTPLD, and ZTDLD along with original observed frequencies for datasets in Tables **3**, **4**, and **5** are given in Figure **3**.

Table 3: The Numbers of Counts of Flower Heads as Per the Number of Fly Eggs Available in Finney and Varley [3]

Number of Fly Eggs	Observed Frequency	Expected Value		
		ZTPD	ZTPLD	ZTDLD
1	22	15.3	26.8	24.9
2	18	21.9	19.8	20.4
3	18	20.8	13.9	14.9
4	11	14.9	9.5	10.2
5	9	8.5	6.4	6.7
6	6	4.1	4.2	4.3
7	3	1.7	2.7	2.7
8	0	0.6	1.7	1.6
9	1	0.3	3.0	2.3
Total	88	88.0	88.0	88.0
ML Estimate		$\hat{\theta} = 2.8604$	$\hat{\theta} = 0.7186$	$\hat{\theta} = 0.6042$
χ^2		6.677	3.743	2.257
d.f.		4	4	4
p-value		0.1540	0.4419	0.8125

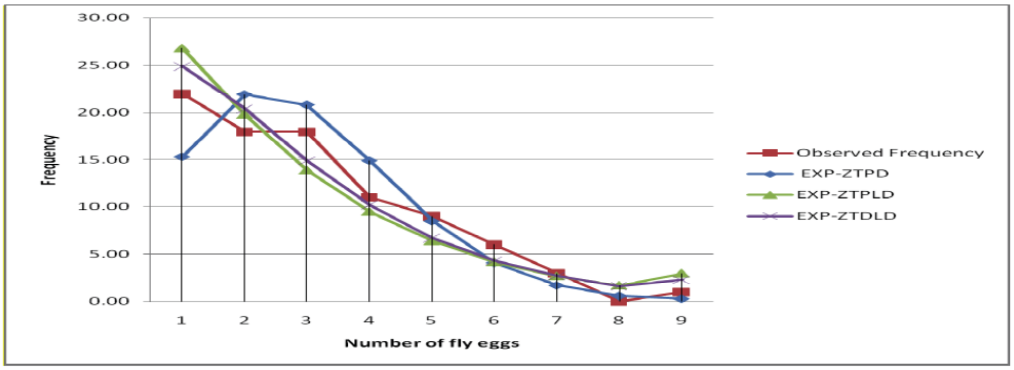
Table 4: The Number of Snowshoe Hares Counts Captured over 7 Days Available in Keith and Meslow [14]

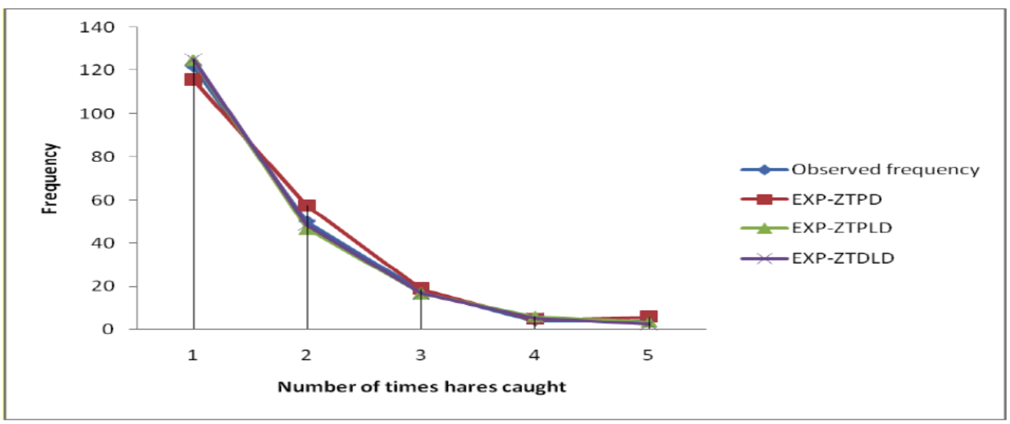
Number of Times Hares Caught	Observed Frequency	Expected Value		
		ZTPD	ZTPLD	ZTDLD
1	122	115.8	124.7	125.1
2	50	57.4	46.7	48.4
3	18	18.9	17.0	16.7
4	4	4.7	6.1	5.4
5	4	5.9	3.5	2.4
Total	198	198.0	198.0	198.0
ML Estimate		$\hat{\theta} = 0.9906$	$\hat{\theta} = 2.1831$	$\hat{\theta} = 1.3189$
χ^2		2.140	0.617	0.243
d.f.		2	2	2
P-value		0.3430	0.7345	0.9703

Table 5: Number of Households having at Least one Migrant According to the Number of Migrants, Available in Singh and Yadav [15]

Number of Migrants	Observed Frequency	Expected Value		
		ZTPD	ZTPLD	ZTDLD
1	375	354.0	379.0	372.9
2	143	167.7	137.2	144.3
3	49	52.9	48.4	49.7
4	17	12.5	16.7	16.0
5	2	2.4	5.7	5.0
6	2	0.4	1.9	1.5
7	1	0.1	0.6	0.4
8	1	0.0	0.5	0.2

Total	590	590.0	590.0	590.0
ML Estimate		$\hat{\theta} = 0.9475$	$\hat{\theta} = 2.2848$	$\hat{\theta} = 1.3546$
χ^2		8.922	1.138	0.258
d.f.		2	3	3
P-value		0.0115	0.7679	0.9924





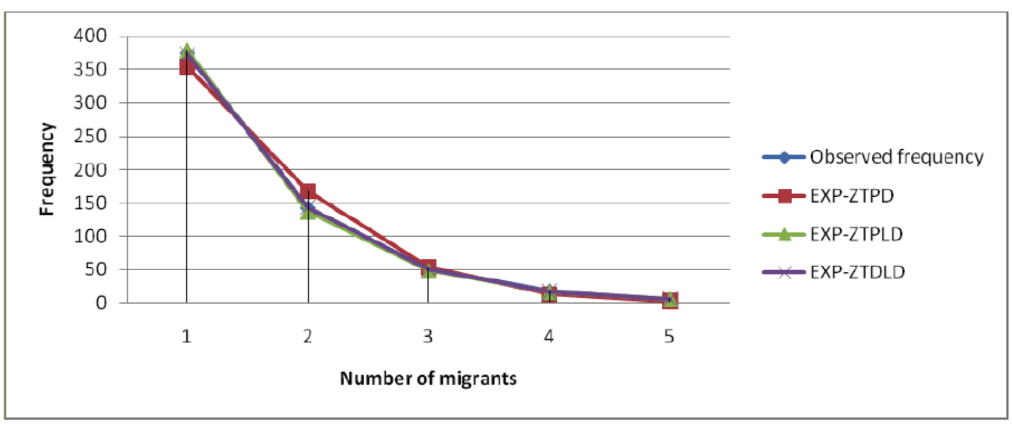


Figure 3: The fitted plot of ZTPD, ZTPLD, and ZTDLD along with original observed frequencies for datasets in Tables **3**, **4**, and **5**.

CONCLUSIONS

This paper proposes a zero-truncated discrete Lindley distribution (ZTDLD) which is a zero-truncation of discrete Lindley distribution. A discrete Lindley distribution is the discrete analogue of continuous Lindley distribution using infinite series approach of discretization. Its moments and moments based measures have been obtained. The behaviors of mean, variance, coefficient of variation, skewness, kurtosis and index of dispersions for varying values of parameter have been discussed both numerically and graphically. Both the method of moment and the method of maximum likelihood give the same estimate of the parameter of ZTDLD. Three examples of observed real datasets have been given to test the goodness of fit of ZTDLD over ZTPD and ZTPLD and found that ZTDLD gives much closer fit over these distributions. Since ZTDLD is competing well with ZTPD and ZTPLD, it can be one of the important zero-truncated distributions in statistics literature.

ACKNOWLEDGEMENTS

Authors are grateful to the Editor-In-Chief of the Journal and an anonymous reviewer for constructive comments for the improvement in the quality and presentation of the paper.

REFERENCES

[1] Best DJ, Rayner JCW and Thas O. Goodness of fit for the zero-truncated Poisson distribution, Journal of Statistical Computation and Simulation 2007; 77(7): 585-591. https://doi.org/10.1080/10629360600569329

[2] Coleman JS and James J. The equilibrium size distribution of freely forming groups, Sociometry 1961; 24: 36-45. https://doi.org/10.2307/2785927

[3] Finney DJ and Varley GC. An example of the truncated Poisson distribution, Biometrics 1955; 11: 387-94. https://doi.org/10.2307/3001776

[4] Mathews JNS and Appleton DR. An application of the truncated Poisson distribution to immunogold assay, Biometrics 1993; 49: 617-621. https://doi.org/10.2307/2532574

[5] Berhane A and Shanker R. A discrete Lindley distribution with Applications in Biological Sciences, Biometrics and Biostatistics International Journal 2018; 7(2): 1-5.

[6] Lindley DV. Fiducial distributions and Bayes' theorem, Journal of the Royal Statistical Society, Series B 1958; 20: 102-107

[7] Ghitany ME, Atieh B and Nadarajah S. Lindley distribution and Its Application, Mathematics Computing and Simulation 2008 a; 78: 493-506. https://doi.org/10.1016/j.matcom.2007.06.007

[8] Shanker R, Hagos F and Sujatha S. On modeling of lifetimes data using exponential and Lindley distributions, Biometrics and Biostatistics International journal 2015; 2(5): 1-9.

[9] Sankaran M. The discrete Poisson-Lindley distribution, Biometrics 1970; 26: 145-149. https://doi.org/10.2307/2529053

[10] Shanker R and Hagos F. On Poisson-Lindley distribution and Its Applications to biological sciences, Biometrics & Biostatistics International Journal 2015; 2(4): 1-5. https://doi.org/10.15406/bbij.2015.02.00036

[11] Ghitany ME, Al-Mutairi DK and Nadarajah S. Zero-truncated Poisson- Lindley distribution and its Applications, Mathematics and Computers in Simulation 2008 b; 79(3): 279-287. https://doi.org/10.1016/j.matcom.2007.11.021

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

[13] Barlow RE and Proschan F. Statistical Theory of Reliability and Life Testing, Silver Spring, MD (1981)

[14] Keith LB and Meslow EC. Trap response by snowshoe hares, Journal of Wildlife Management 1968; 32: 795-801. https://doi.org/10.2307/3799555

[15] Singh SN and Yadav RC. Trends in rural out-migration at ousehold level, Rural Demography 1971; 8: 53-61.

Received on 12-03-2018 Accepted on 03-04-2018 Published on 21-05-2018