Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics
Turkiye Klinikleri J Biostat. 2021;13(1):1-12

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2020-76457

A Zero-truncated Discrete Quasi Lindley Distribution with Applications

Uygulamalarla Sıfır Kesikli Yarı Lindley Dağıtımı

Rama SHANKER[a], Kamlesh Kumar SHUKLA[b]

[a]Department of Statistics, Assam University, Silchar, Assam, INDIA
[b]Department of Statistics, College of Science, EIT, Asmara, ERITREA

ABSTRACT Objective: Zero-truncated distributions are used to model count data when the data originate from a mechanism 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 boy child, then the frequency of zero boy children is not possible. Objective of this research is to propose a zero-truncated discrete quasi Lindley distribution and study its behavior and properties both theoretically and graphically. **Material and Methods:** The definition of zero-truncation has been used to derive the zero-truncated discrete quasi Lindley distribution. All the properties of the proposed discrete distribution have been derived. **Results:** The generating functions, moments about origin and moments about the mean of zero-truncated discrete quasi Lindley distribution have been obtained. Natures of coefficient of variation, skewness, kurtosis, and the index of dispersion of the distribution have been explained. Maximum likelihood estimation has been discussed for estimating the parameters of the distribution. **Conclusion:** Applications of the distribution have been explained through four examples of observed real datasets from biological sciences and demography and its goodness of fit has been found quite satisfactory over zero-truncated Poisson-Lindley distribution, zero-truncated discrete Lindley distribution and zero-truncated quasi Poisson-Lindley distribution.

Keywords: Zero-truncated distribution; discrete Lindley distribution; moments based measures; estimation; applications

ÖZET Amaç: Veriler sıfır sayımlarını dışlayan bir mekanizmadan kaynaklandığında sıfır kesikli dağılımlar kullanılır. En az bir olay meydana gelmedikçe kayıt mekanizması etkinleştirilmezse sıfır dışlama verileri üretilebilir. Rastgele deneyde sıfır sayım sıklığının gözlemlenemediği birçok gerçek yaşam durumu vardır. Örneğin, bir anket en az bir erkek çocuğu olan hane halkı sayısıyla ilgiliyse, sıfır erkek çocuk sıklığı mümkün değildir. Bu araştırmanın amacı, sıfır kesikli bir Lindley dağılımını önermek ve davranışını ve özelliklerini hem teorik hem de grafiksel olarak incelemektir. **Gereç ve Yöntemler:** Sıfır kesikli yarı Lindley dağılımı elde etmek için sıfır kesikli tanımı kullanılmıştır. Önerile kesikli dağılımın tüm özellikleri elde edilmiştir. **Bulgular ve Sonuç:** Üreten fonksiyonlar, başlangıç ile ilgili momentler ve sıfır kesikli yarı Lindley dağılımının ortalaması ile ilgili momentler elde edilmiştir. Dağılımın değişkenlik katsayısı, çarpıklığı, basıklığı ve dağılım indeksi açıklanmıştır. Dağılım parametrelerini tahmin etmek için maksimum olasılık tahmini tartışılmıştır. **Sonuç:** Dağılım uygulamaları dört biyolojik bilimlerden gözlemlenen gerçek veri kümelerinin örnekleri ve demografi ve uyum iyiliği, sıfır kesikli Poisson-Lindley dağılımı, sıfır kesikli ayrık Lindley dağılımı ve sıfır kesilmiş yarı Poisson Lindley dağılımına göre oldukça tatmin edici bulunmuştur.

Anahtar kelimeler: Sıfır kesikli dağılım; ayrık Lindley dağılımı; momente dayalı ölçümler; tahmin; uygulamalar

Zero-truncated distributions are used to model count data when the data originate from a mechanism excluding zero counts. Data excluding zero counts 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

Correspondence: Kamlesh Kumar SHUKLA
Department of Statistics, College of Science, EIT, Asmara, ERITREA/ERİTRE
E-mail: kkshukla22@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 23 May 2020 ***Received in revised form:*** 18 Dec 2020 ***Accepted:*** 26 Jan 2021 ***Available online:*** 29 Apr 2021

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. have detailed discussion on applications and goodness of fit of zero-truncated Poisson distribution in biological sciences.[1] Coleman and James, Finney and Varley, Matthews and Appleton, are some others who have discussed various applications of zero-truncated distributions particularly zero-truncated Poisson distribution.[2-4]

Assuming $P_0(x;\theta)$ as the probability mass function (pmf) of a discrete distribution, the zero-truncated version of $P_0(x;\theta)$ can be defined as

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

Recently, Sium and Shanker obtained zero-truncated discrete Lindley distribution (ZTDLD) defined by its pmf[5]

$$P_1(x;\theta)=\frac{\left(e^{\theta}-1\right)^2}{2e^{\theta}-1}(1+x)e^{-\theta x} \quad ; x=1,2,3,...,\ \theta>0 \tag{1.2}$$

Note that ZTDLD is the zero-truncated version of a discrete Lindley distribution (DLD) proposed by Berhane and Shanker and defined by its pmf[6]

$$P_2(x;\theta)=\frac{\left(e^{\theta}-1\right)^2}{e^{2\theta}}(1+x)e^{-\theta x};\ x=0,1,2,3,....,\theta>0 \tag{1.3}$$

The first four moments about origin of ZTDLD obtained by Sium and Shanker are[5]

$$\mu_1'=\frac{2e^{2\theta}}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)}, \qquad \mu_2'=\frac{2e^{2\theta}\left(e^{\theta}+2\right)}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)^2}$$

$$\mu_3'=\frac{2e^{2\theta}\left(e^{2\theta}+7e^{\theta}+4\right)}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)^3}, \quad \mu_4'=\frac{2e^{2\theta}\left(e^{3\theta}+18e^{2\theta}+33e^{\theta}+8\right)}{\left(2e^{\theta}-1\right)\left(e^{\theta}-1\right)^4}$$

The DLD is a discrete analogue of continuous Lindley distribution introduced by Lindley.[7] Berhane and Shanker have obtained moments and moments based measures of DLD and discussed the behaviors of coefficient of variation, skewness, kurtosis and index of dispersion. Further, Berhane and Shanker have discussed its applications to model count data from biological sciences and showed that it gives better fit than both Poisson distribution and Poisson-Lindley distribution (PLD).[6] The PLD, a Poisson mixture of Lindley distribution introduced by Sankaran, is defined by its pmf [7,8]

$$P_3(x;\theta)=\frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}} \ ; x=0,1,2,...,\theta>0 \tag{1.4}$$

Sankaran 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.[8] Shanker and Fesshaye showed that PLD is an over-dispersed distribution and hence several applications of PLD have been discussed in biological sciences because in general biological sciences data are over-dispersed.[9]

The Lindley distribution, proposed by Lindley, is defined by its probability density function (pdf)[7]

$$f_1(x;\theta)=\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x} \quad ;x>0,\ \theta>0 \tag{1.5}$$

Ghitany et al. have proposed zero-truncated PLD (ZTPLD) defined by its pmf

$$P_4(x;\theta)=\frac{\theta^2}{\theta^2+3\theta+1}\frac{x+\theta+2}{(\theta+1)^x} \quad ;x=1,2,3,....,\theta>0 \tag{1.6}$$

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 them.[10] Shanker et al. 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.[11]

Shanker and Mishra proposed a quasi PLD (QPLD) having parameters θ and α and defined by its pmf[12]

$$P_5(x;\theta)=\frac{\theta}{\alpha+1}\frac{\theta x+(\alpha\theta+\theta+\alpha)}{(\theta+1)^{x+2}};\ x=0,1,2,...,\ \theta>0,\alpha>-1 \tag{1.7}$$

Note that PLD is a particular case of QPLD for $\alpha=\theta$. Various statistical properties, estimation of parameters using method of moments and method of maximum likelihood and applications of QPLD have been discussed by Shanker and Mishra.[12] It has been shown by Shanker and Mishra that QPLD gives much closer fit than the PLD.[12] The zero-truncated quasi PLD (ZTQPLD) has been studied by Shanker and Shukla.[13] QPLD is a Poisson mixture of quasi Lindley distribution (QLD) introduced by Shanker and Mishra and defined by its pdf [14]

$$f_2(x;\theta,\alpha)=\frac{\theta}{\alpha+1}(\alpha+\theta x)e^{-\theta x}; \quad x>0,\ \theta>0,\ \alpha>-1 \tag{1.7}$$

Lindley distribution is a particular case of QLD at $\alpha=\theta$. Shanker and Mishra has studied various statistical properties, estimation of parameters and applications of QLD.[14]

Recently, Shanker et al. obtained a discrete QLD (DQLD) defined by its pmf [15]

$$P_6(x;\theta,\alpha)=\frac{(e^\theta-1)^2}{e^\theta(\alpha e^\theta+\theta-\alpha)}(\alpha+\theta x)e^{-\theta x};\ x=0,1,2,...,\ \theta>0,\alpha>0 \tag{1.8}$$

Note that DQLD is based on the infinite series approach of discretization of QLD given in (1.7).[15]

In this paper, a zero-truncated discrete quasi Lindley distribution (ZTDQLD) has been proposed. 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. Maximum likelihood estimation has been discussed for estimating the parameters of the distribution. Applications of the distribution have been explained through four examples of observed real datasets from biological sciences and demography and its goodness of fit has been found quite satisfactory and competing well with ZTPLD, ZTDLD and ZTQPLD.

A ZERO-TRUNCATED DISCRETE QUASI LINDLEY DISTRIBUTION

The pmf of ZTDQLD, using (1.1) and (1.8), can be expressed as

$$P_7(x;\theta,\alpha)=\frac{\left(e^{\theta}-1\right)^2}{(\theta+\alpha)e^{\theta}-\alpha}(\alpha+\theta x)e^{-\theta x}\quad ;x=1,2,3,...,\ \theta>0 \tag{2.1}$$

At $\alpha=\theta$, it reduces to ZTDLD. The behavior of the ZTDQLD for varying values of parameter θ has been shown in Figure 1.

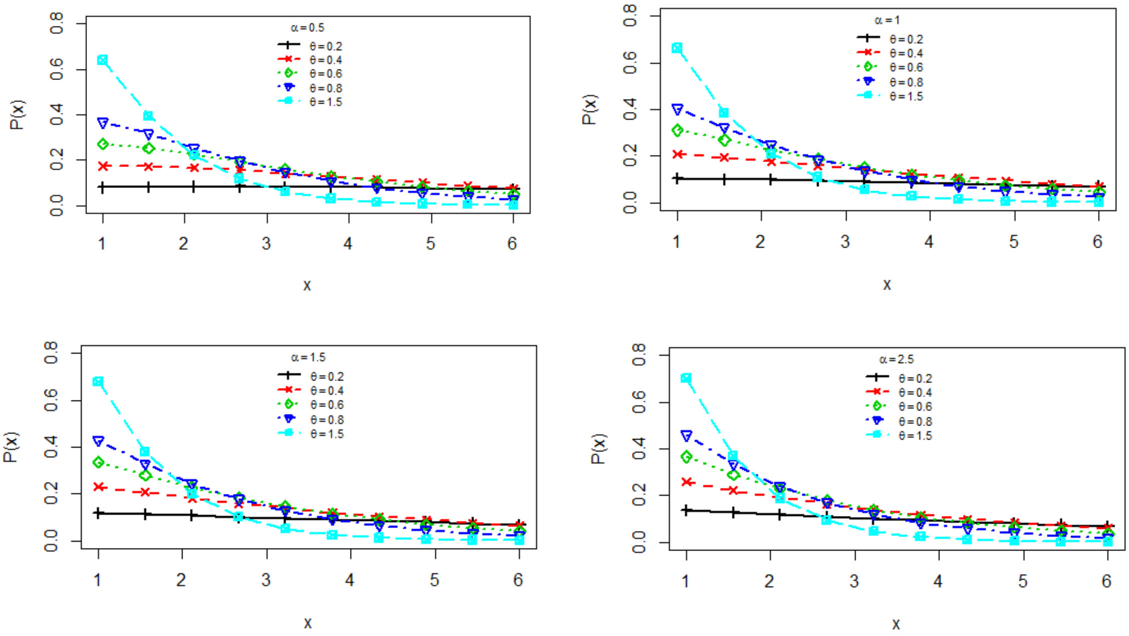

FIGURE 1: Behavior of ZTDQLD for varying values of parameters θ and α.

STATISTICAL PROPERTIES

UNIMODALITY

It is obvious that $\frac{P_7(x+1;\theta,\alpha)}{P_7(x;\theta,\alpha)}=\left(1+\frac{\theta}{\alpha+\theta x}\right)e^{-\theta}$ is a decreasing function of $x\geq 0$ and hence $P_7(x;\theta,\alpha)$ is log-concave. This shows that ZTDQLD is unimodal, has increasing failure rate (IFR), and hence IFR average (IFRA). It can be also shown that it is new better than used (NBU), NBU in expectation (NBUE), and has decreasing mean residual life (DMRL). The relationships of these aging concepts can be seen in Barlow and Proschan.[16]

GENERATING FUNCTIONS

The probability generating function $G(t)$ of ZTDQLD can be obtained as

$$G(t)=E\left(t^X\right)=\frac{\left(e^{\theta}-1\right)^2}{(\theta+\alpha)e^{\theta}-\alpha}\sum_{x=1}^{\infty}t^x(\alpha+\theta x)e^{-\theta x}$$

$$= \frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\sum_{x=1}^{\infty}(\alpha+\theta x)\left(\frac{t}{e^{\theta}}\right)^{x}$$

$$= \frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\left[\alpha\sum_{x=1}^{\infty}\left(\frac{t}{e^{\theta}}\right)^{x}+\theta\sum_{x=1}^{\infty}x\left(\frac{t}{e^{\theta}}\right)^{x}\right]$$

$$= \frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\frac{t(\theta+\alpha)e^{\theta}-\alpha t^{2}}{\left(e^{\theta}-t\right)^{2}};t\neq e^{\theta}$$

The moment generating function $M(t)$ of ZTDQLD can thus be expressed as

$$M(t)=\frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\frac{e^{t}\left\{(\theta+\alpha)e^{\theta}-\alpha e^{t}\right\}}{\left(e^{\theta}-e^{t}\right)^{2}}\ ;t\neq\theta .$$

It can be easily verified that the moment generating function is infinitely differentiable with respect to t, and hence all moments about origin μ_r' , $r\geq 1$ of ZTDQLD can be obtained.

MOMENTS AND MOMENTS BASED MEASURES

The r th moment about origin μ_r' of ZTDQLD can be obtained as

$$\mu_r' = E\left(X^{r}\right)=\frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\sum_{x=1}^{\infty}x^{r}(\alpha+\theta x)e^{-\theta x}$$

$$= \frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\left[\alpha\sum_{x=1}^{\infty}x^{r}\left(e^{-\theta}\right)^{x}+\theta\sum_{x=1}^{\infty}x^{r+1}\left(e^{-\theta}\right)^{x}\right] \tag{4.1}$$

Taking $r=1,2,3$ and 4 in (4.1) and simplifying the tedious algebraic expression, the first four moments about origin of ZTDQLD can be expressed as

$$\mu_1' = \frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\left[\alpha\sum_{x=1}^{\infty}x\left(e^{-\theta}\right)^{x}+\theta\sum_{x=1}^{\infty}x^{2}\left(e^{-\theta}\right)^{x}\right]$$

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

$$\mu_2' = \frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\left[\alpha\sum_{x=1}^{\infty}x^{2}\left(e^{-\theta}\right)^{x}+\theta\sum_{x=1}^{\infty}x^{3}\left(e^{-\theta}\right)^{x}\right]$$

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

$$\mu_3'=\frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\left[\alpha\sum_{x=1}^{\infty}x^{3}\left(e^{-\theta}\right)^{x}+\theta\sum_{x=1}^{\infty}x^{4}\left(e^{-\theta}\right)^{x}\right]$$

$$=\frac{(\theta+\alpha)e^{4\theta}+(11\theta+3\alpha)e^{3\theta}+(11\theta-3\alpha)e^{2\theta}+(\theta-\alpha)e^{\theta}}{\left(e^{\theta}-1\right)^{3}\left\{(\theta+\alpha)e^{\theta}-\alpha\right\}}$$

$$\mu_4'=\frac{\left(e^{\theta}-1\right)^{2}}{(\theta+\alpha)e^{\theta}-\alpha}\left[\alpha\sum_{x=1}^{\infty}x^{4}\left(e^{-\theta}\right)^{x}+\theta\sum_{x=1}^{\infty}x^{5}\left(e^{-\theta}\right)^{x}\right]$$

$$=\frac{(\theta+\alpha)e^{5\theta}+(26\theta+10\alpha)e^{4\theta}+66\theta e^{3\theta}+(26\theta-10\alpha)e^{2\theta}+(\theta-\alpha)e^{\theta}}{\left(e^{\theta}-1\right)^{4}\left\{(\theta+\alpha)e^{\theta}-\alpha\right\}}$$

Using the relationship between moments about the mean and moments about the origin, the moments about the mean of ZTDQLD are obtained as

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

$$\mu_3=\frac{\left\{\begin{array}{l}\left(2\theta^3+5\theta^2\alpha+4\theta\alpha^2+\alpha^3\right)e^{5\theta}+\left(2\theta^3+3\theta^2\alpha-\theta\alpha^2-2\alpha^3\right)e^{4\theta}-9\left(\theta\alpha^2+\theta^2\alpha\right)e^{3\theta}\\+\left(2\alpha^3+\theta^2\alpha+5\theta\alpha^2\right)e^{2\theta}+\left(\theta\alpha^2-\alpha^3\right)e^{\theta}\end{array}\right\}}{\left(e^{\theta}-1\right)^{3}\left\{(\theta+\alpha)e^{\theta}-\alpha\right\}^{3}}$$

$$\mu_4=\frac{\left\{\begin{array}{l}\left(2\theta^4+7\theta^3\alpha+9\theta^2\alpha^2+5\theta\alpha^3+\alpha^4\right)e^{7\theta}+\left(20\theta^4+64\theta^3\alpha+71\theta^2\alpha^2+30\theta\alpha^3+3\alpha^4\right)e^{6\theta}\\+\left(2\theta^4-54\theta^3\alpha-134\theta^2\alpha^2-99\theta\alpha^3-21\alpha^4\right)e^{5\theta}+9\left(18\theta^2\alpha^2-16\theta^3\alpha+68\theta\alpha^3+34\alpha^4\right)e^{4\theta}\\+\left(37\theta^2\alpha^2+15\theta\alpha^3-\theta^3\alpha-21\alpha^4\right)e^{3\theta}+\left(3\alpha^4-\theta^2\alpha^2-18\theta\alpha^3\right)e^{2\theta}+\left(\alpha^4-\theta\alpha^3\right)e^{\theta}\end{array}\right\}}{\left(e^{\theta}-1\right)^{4}\left\{(\theta+\alpha)e^{\theta}-\alpha\right\}^{4}}$$

At $\alpha=\theta$, these moments reduce to the corresponding moments of ZTDLD, obtained by Sium and Shanker.[5]

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

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

$$\sqrt{\beta_1} = \frac{\mu_3}{\left(\mu_2\right)^{3/2}} = \frac{\left\{\begin{array}{l}\left(2\theta^3+5\theta^2\alpha+4\theta\alpha^2+\alpha^3\right)e^{5\theta}+\left(2\theta^3+3\theta^2\alpha-\theta\alpha^2-2\alpha^3\right)e^{4\theta}-9\left(\theta\alpha^2+\theta^2\alpha\right)e^{3\theta}\\+\left(2\alpha^3+\theta^2\alpha+5\theta\alpha^2\right)e^{2\theta}+\left(\theta\alpha^2-\alpha^3\right)e^{\theta}\end{array}\right\}}{\left\{\left(2\theta^2+3\theta\alpha+\alpha^2\right)e^{3\theta}-2\left(\alpha^2+\theta\alpha\right)e^{2\theta}+\left(\alpha^2-\theta\alpha\right)e^{\theta}\right\}^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}} = \frac{\left\{\begin{array}{l}\left(2\theta^4+7\theta^3\alpha+9\theta^2\alpha^2+5\theta\alpha^3+\alpha^4\right)e^{7\theta}+\left(20\theta^4+64\theta^3\alpha+71\theta^2\alpha^2+30\theta\alpha^3+3\alpha^4\right)e^{6\theta}\\+\left(2\theta^4-54\theta^3\alpha-134\theta^2\alpha^2-99\theta\alpha^3-21\alpha^4\right)e^{5\theta}+9\left(18\theta^2\alpha^2-16\theta^3\alpha+68\theta\alpha^3+34\alpha^4\right)e^{4\theta}\\+\left(37\theta^2\alpha^2+15\theta\alpha^3-\theta^3\alpha-21\alpha^4\right)e^{3\theta}+\left(3\alpha^4-\theta^2\alpha^2-18\theta\alpha^3\right)e^{2\theta}+\left(\alpha^4-\theta\alpha^3\right)e^{\theta}\end{array}\right\}}{\left\{\left(2\theta^2+3\theta\alpha+\alpha^2\right)e^{3\theta}-2\left(\alpha^2+\theta\alpha\right)e^{2\theta}+\left(\alpha^2-\theta\alpha\right)e^{\theta}\right\}^{2}}$$

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

The behaviors of mean, variance, C.V, skewness, kurtosis and index of dispersion of ZTDQLD for some selected values of parameters θ and α are shown graphically in Figure 2. As the values of parameter α increases and for $\theta<1$, the coefficient of variation increases slowly and stabilizes at 1. Similarly, for increasing values of parameter α and for $\theta\geq 1$, the coefficient of variation decreases and stabilizes at 1. As the values of parameters α and θ increases, the coefficient of skewness increases but for increasing values of θ, the graph of coefficient of skewness shifts upward. This means that higher the values of θ, coefficient of skewness shifts upward. As the values of parameters α and θ increases, the coefficient of kurtosis increases slowly but for fixed values of α and increasing values of θ it shifts downward. In case of index of dispersion, as the values of parameters α and θ increases, the index of dispersion shifts downward for increasing values of θ.

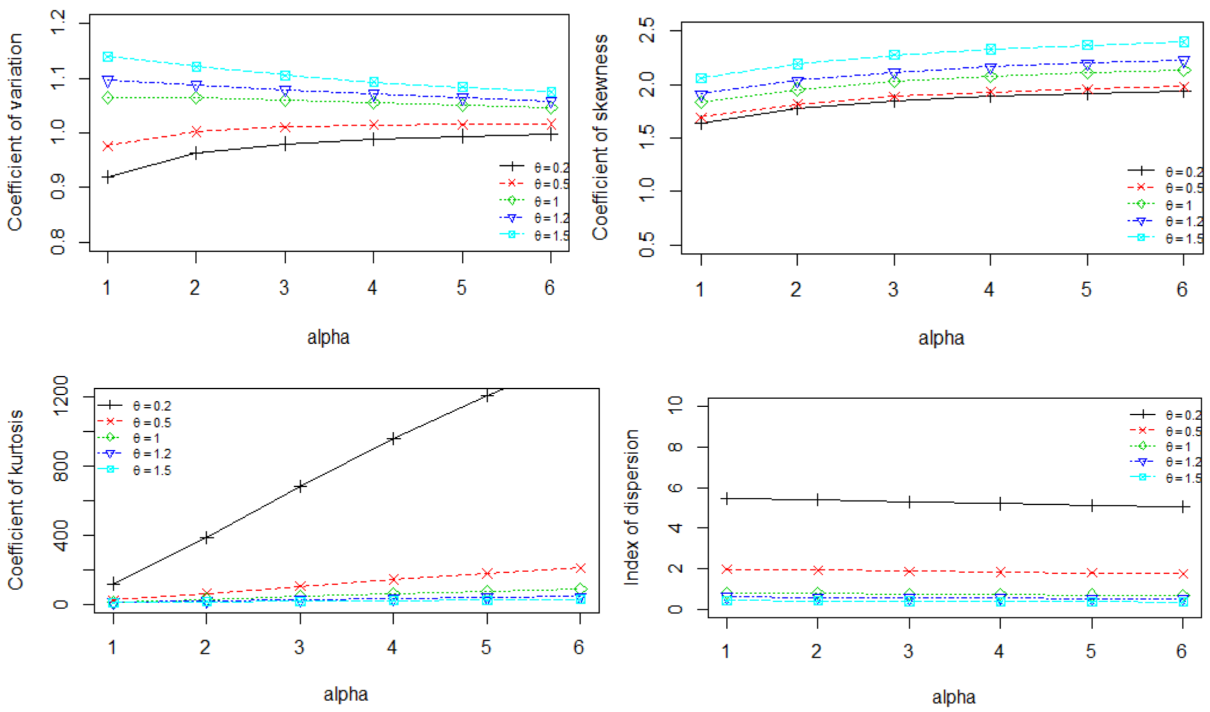

FIGURE 2: Behaviors of mean, variance, C.V, skewness, kurtosis and index of dispersion of ZTDQLD for varying values of parameters θ and α .

ESTIMATION OF PARAMETERS USING MAXIMUM LIKELIHOOD

Assuming $(x_1, x_2, ..., x_n)$ be a random sample from the ZTDQLD, the natural log-likelihood function L of the ZTDQLD can be expressed as

$$\ln L = n\left[2\ln\left(e^{\theta}-1\right)-\ln\left\{\left(\theta+\alpha\right)e^{\theta}-\alpha\right\}\right]+\sum_{x=1}^{k} f_x \ln\left(\alpha+\theta x\right)-n\theta\bar{x} ,$$

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

The maximum likelihood estimate (MLE) $\left(\hat{\theta},\hat{\alpha}\right)$ of the parameter $\left(\theta,\alpha\right)$ of ZTDQLD is the solution of the following log likelihood equations

$$\frac{\partial \ln L}{\partial \theta}=\frac{2ne^{\theta}}{e^{\theta}-1}-\frac{n\left(\theta+\alpha+1\right)e^{\theta}}{\left(\theta+\alpha\right)e^{\theta}-\alpha}+\sum_{x=1}^{k}\frac{x f_x}{\alpha+\theta x}-n\bar{x}=0$$

$$\frac{\partial \ln L}{\partial \alpha}=-\frac{n\left(e^{\theta}-1\right)}{\left(\theta+\alpha\right)e^{\theta}-\alpha}+\sum_{x=1}^{k}\frac{f_x}{\alpha+\theta x}=0,$$

Since these two log likelihood equations are non-linear in nature and cannot be expressed in closed form, it is hard to solve these equations analytically. However, these two equations can be solved iteratively using R-software till sufficiently close estimates of $\hat{\theta}$ and $\hat{\alpha}$ are obtained.

Fisher scoring method can also be used to solve the above non-linear equations.

The 2×2 observed information matrix for Fisher scoring method of ZTDQLD can be presented as,

$$\begin{pmatrix}\hat{\theta}\\ \hat{\alpha}\end{pmatrix} \sim \left[\begin{pmatrix}\theta\\ \alpha\end{pmatrix}, \begin{pmatrix}\frac{\partial^2 \ln L}{\partial\theta^2} & \frac{\partial^2 \ln L}{\partial\theta\,\partial\alpha}\\ \frac{\partial^2 \ln L}{\partial\alpha\,\partial\theta} & \frac{\partial^2 \ln L}{\partial\alpha^2}\end{pmatrix}\right]$$

The inverse of the information matrix results in the well-known variance-covariance matrix. The 2×2 approximate Fisher information matrix corresponding to the above observed information matrix is given by

$$I^{-1} = -E\begin{bmatrix}\frac{\partial^2 \ln L}{\partial\theta^2} & \frac{\partial^2 \ln L}{\partial\theta\,\partial\alpha}\\ \frac{\partial^2 \ln L}{\partial\alpha\,\partial\theta} & \frac{\partial^2 \ln L}{\partial\alpha^2}\end{bmatrix}$$

The solution of the Fisher information matrix will yield asymptotic variance and covariance of the MLEs for $\left(\hat{\theta},\hat{\alpha}\right)$. The approximate $100\left(1-\alpha\right)\%$ confidence intervals for $\left(\theta,\alpha\right)$ respectively are $\hat{\theta} \pm Z_{\frac{\alpha}{2}} \frac{\sigma_{\theta\theta}}{n}$, and $\hat{\alpha} \pm Z_{\frac{\alpha}{2}} \frac{\sigma_{\alpha\alpha}}{n}$ where Z_{α} is the upper 100αth percentile of the standard normal distribution.

GOODNESS OF FIT TO REAL DATASETS

The applications of ZTDQLD have been discussed with four examples of observed real datasets and its goodness of fit has been compared with that of ZTPLD, ZTDLD, and ZTQPLD. The goodness of fit of ZTPLD, ZTDLD, and ZTQPLD are based on MLEs of the parameters. The dataset in Table 1 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 2 is regarding the number of counts of sites with particles from immunogold data reported by Matthews and Appleton.[4] The dataset in Table 3 is regarding the number of snow hares counts captured over 7 days and are available in Keith and Meslow.[17] The dataset in Table 4 is due to Singh and Yadava regarding the number of households having at least one migrant from households according to the number of migrants.[18] It is clear from the goodness of fit of ZTDQLD that it competes well with ZTQPLD. Further, one important conclusion from the goodness of fit of the ZTDQLD is that in majority of datasets it gives the same fit as the fit given by ZTQPLD but in few datasets the fit given by ZTDQLD is better than the fit given by ZTQPLD and this is due to the nature of datasets. Since both ZTDQLD and ZTQPLD are two-parameter discrete distributions and are useful for over-dispersed datasets, the advantage of ZTDQLD over ZTQPLD is that ZTDQLD is suitable for data having higher over-dispersion.

TABLE 1: The numbers of counts of flower heads as per the number of fly eggs, Finney and Varley.[3]

Number of fly eggs	Observed frequency	Expected value			
		ZTPLD	ZTDLD	ZTQPLD	ZTDQLD
1	22	26.8	24.9	20.8	20.7
2	18	19.8	20.4	22.2	22.2
3	18	13.9	14.9	16.9	16.9
4	11	9.5	10.2	11.3	11.3
5	9	6.4	6.7	7.0	7.0
6	6	4.2	4.3	4.4	4.2
7	3	2.7	2.7	2.4	2.4
8	0	1.7	1.6	1.4	1.4
9	1	3.0	2.3	1.6	1.9
Total	88	88.0	88.0	88.0	88.0
Maximum likelihood estimate		$\hat{\theta}=0.7186$	$\hat{\theta}=0.6042$	$\hat{\theta}=1.02087$ $\hat{\alpha}=-0.57630$	$\hat{\theta}=0.70353$ $\hat{\alpha}=-0.09904$
χ^2		3.743	2.257	1.52	1.52
df		4	4	3	3
p value		0.4419	0.8125	0.6822	0.6822

ZTPLD: Zero-truncated Poisson-Lindley distribution; ZTDLD: Zero-truncated discrete Lindley distribution; ZTQPLD: Zero-truncated quasi Poisson-Lindley distribution; ZTDQLD: Zero-truncated discrete quasi Lindley distribution.

TABLE 2: The number of counts of sites with particles from immunogold data, Matthews and Appleton.[4]

Number of sites with particles	Observed frequency	Expected value			
		ZTPLD	ZTDLD	ZTQPLD	ZTDQLD
1	122	124.7	125.1	121.9	121.9
2	50	46.7	48.4	50.1	50.1
3	18	17.0	16.7	17.7	17.7
4	4	6.1	5.4	5.8	5.8
5	4	3.5	2.4	2.5	2.5
Total	198	198.0	198.0	198.0	198.0
Maximum likelihood estimate		$\hat{\theta}=2.1831$	$\hat{\theta}=1.3189$	$\hat{\theta}=2.89975$ $\hat{\alpha}=-0.24999$	$\hat{\theta}=1.36088$ $\hat{\alpha}=0.90350$
χ^2		0.617	0.243	0.038	0.038
df		2	2	1	1
p value		0.7345	0.9703	0.8454	0.8454

ZTPLD: Zero-truncated Poisson-Lindley distribution; ZTDLD: Zero-truncated discrete Lindley distribution; ZTQPLD: Zero-truncated quasi Poisson-Lindley distribution; ZTDQLD: Zero-truncated discrete quasi Lindley distribution.

TABLE 3: The number of snowshoe hares counts captured over 7 days.[17]

Number of times hares caught	Observed frequency	Expected frequency			
		ZTPD	ZTPLD	ZTQPLD	ZTDQLD
1	184	174.6	182.6	183.0	182.8
2	55	66.0	55.3	54.8	54.9
3	14	16.6	16.4	16.3	16.3
4	4	3.2	4.8	4.8	4.8
5	4	0.6	1.9	2.1	2.5
Total	261	261.0	261.0	261.0	261.0
Maximum likelihood estimate		$\hat{\theta}=0.7563$	$\hat{\theta}=2.8639$	$\hat{\theta}=2.5525$ $\hat{\alpha}=9.9802$	$\hat{\theta}=1.29032$ $\hat{\alpha}=12.85054$
χ^2		2.464	0.615	0.506	0.40
df		1	2	1	1
p value		0.1165	0.7353	0.4768	0.5271

ZTPD: Zero-truncated Poisson distribution; ZTPLD: Zero-truncated Poisson-Lindley distribution; ZTQPLD: Zero-truncated quasi Poisson-Lindley distribution; ZTDQLD: Zero-truncated discrete quasi Lindley distribution.

TABLE 4: Number of households having at least one migrant according to the number of migrants.[18]

Number of migrants	Observed frequency	Expected value			
		ZTPLD	ZTDLD	ZTQPLD	ZTDQLD
1	375	379.0	372.9	376.3	376.3
2	143	137.2	144.3	140.2	140.2
3	49	48.4	49.7	49.0	49.0
4	17	16.7	16.0	16.5	16.5
5	2	5.7	5.0	5.4	5.3
6	2	1.9	1.5	1.7	1.7
7	1	0.6	0.4	0.5	0.5
8	1	0.5	0.2	0.4	0.5
Total	590	590.0	590.0	590.0	590.0
Maximum likelihood estimate		$\hat{\theta}=2.2848$	$\hat{\theta}=1.3546$	$\hat{\theta}=2.56866$ $\hat{\alpha}=0.74218$	$\hat{\theta}=1.27192$ $\hat{\alpha}=2.58896$
χ^2		1.138	0.258	0.575	0.575
df		3	3	2	2
p value		0.7679	0.9924	0.7501	0.7501

ZTPLD: Zero-truncated Poisson-Lindley distribution; ZTDLD: Zero-truncated discrete Lindley distribution; ZTQPLD: Zero-truncated quasi Poisson-Lindley distribution; ZTDQLD: Zero-truncated discrete quasi Lindley distribution.

CONCLUSION

This paper proposes a ZTDQLD. A discrete quasi Lindley distribution (DQLD) is a discrete analogue based on infinite series approach of discretization of quasi Lindley distribution (QLD). Its moments and moments based measures have been obtained. The behaviors of mean, variance, and coefficient of variation, skewness, kurtosis and index of dispersions for varying values of parameters have been discussed graphically. Maximum likelihood estimation has been discussed. Four examples of observed real datasets have been given to test the goodness of fit of ZTDQLD over ZTPLD, ZTDLD and ZTQPLD. Since ZTDQLD is competing well with ZTPLD, ZTDLD and ZTQPLD, it can be one of the important zero-truncated distributions in statistics literature.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Rama Shanker;* ***Design:*** *Rama Shanker;* ***Control/Supervision:*** *Rama Shanker;* ***Data Collection and/or Processing:*** *Kamlesh Kumar Shukla;* ***Analysis and/or Interpretation:*** *Kamlesh Kumar Shukla;* ***Literature Review:*** *Kamlesh Kumar Shukla;* ***Writing the Article:*** *Rama Shanker;* ***Critical Review:*** *Kamlesh Kumar Shukla;* ***References and Fundings:*** *Rama Shanker;* ***Materials:*** *Kamlesh Kumar Shukla.*

REFERENCES

1. Best DJ, Rayner JCW, Thas O. Goodness of fit for the zero-truncated Poisson distribution. Journal of Statistical Computation and Simulation. 2007;77(7):585-91. [Crossref]
2. Coleman JS, James J. The equilibrium size distribution of freely-forming groups. Sociometry. 1961;24(1):36-45. [Crossref]
3. Finney DJ, Varley GC. An example of the truncated Poisson distribution. Biometrics. 1955;11(3):387-94. [Crossref]
4. Matthews JN, Appleton DR. An application of the truncated Poisson distribution to immunogold assay. Biometrics. 1993;49(2):617-21. [Crossref] [PubMed]
5. Sium S, Shanker R. A zero truncated discrete Lindley distribution with Applications. International Journal of Statistics in Medical and Biological Research. 2018;2:8-15. [Link]
6. Berhane A, Shanker R. A discrete Lindley distribution with applications in biological sciences. Biom Biostat Int J. 2018;7(1):48-52. [Crossref]
7. Lindley DV. Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society, Series B. 1958;20(1):102-7. [Crossref]
8. Sankaran M. 275. Note: The discrete Poisson-Lindley distribution. Biometrics. 1970;26(1):145-9. [Crossref]
9. Shanker R, Fesshaye H. On poisson-Lindley distribution and its applications to biological sciences. Biom Biostat Int J. 2015;2(4):103-7. [Crossref]
10. Ghitany ME, Al-Mutairi DK, Nadarajah S. Zero-truncated Poisson-Lindley distribution and its applications. Mathematics and Computers in Simulation. 2008;79(3):279-87. [Crossref]
11. Shanker R, Fesshaye H, Selvaraj S, Yemane A. On zero-truncation of Poisson and Poisson-Lindley distributions and their applications. Biom Biostat Int J. 2015;2(6):168-81. [Crossref]
12. Shanker R, Mishra A. A quasi Poisson-Lindley distribution. Journal of the Indian Statistical Association. 2016;54(1&2):113-25. [Link]
13. Shanker R, Shukla KK. On zero-truncated quasi Poisson-Lindley distribution and its applications. International Journal of Mathematics and Statistics. 2018;19(2):14-25. [Link]
14. Shanker R, Mishra A. A quasi Lindley distribution. AJMCSR. 2013;6(4):64-71. [Link]
15. Shanker R, Berhane B, Shukla KK. A discrete quasi Lindley distribution. Journal of Probability and Statistical Sciences. 2019;17(1):27-41.
16. Barlow RE, Proschan F. Statistical Theory of Reliability and Life Testing: Probability Models. 2nd ed. Silver Spring, Md.: To Begin With; 1981.
17. Keith LB, Meslow EC. Trap response by snowshoe hares. The Journal of Wildlife Management. 1968;32(4):795-801. [Crossref]
18. Singh SN, Yadava KN. Trends in rural out-migration at household level. Rural Demogr. 1981;8(1):53-61. [PubMed]