Thai Journal of Mathematics : (2020) 193–202
Special Issue: The 14th IMT-GT ICMSA 2018
http://thaijmath.in.cmu.ac.th
Online ISSN 1686-0209

Zero-Truncated Bell Distribution

Watchareewan Chuncharoenkit † and Winai Bodhisuwan †,1

†Department of Statistics, Faculty of Science,
Kasetsart University, Bangkok 10900, Thailand
e-mail : watchareewan.ch@gmail.com and fsciwnb@ku.ac.th

Abstract : In this paper, the zero-truncated Bell distribution is introduced. It is modified from Bell distribution. Its probability mass function is derived. Some properties of the distribution are presented, e.g., mean and variance. Maximum likelihood estimation is employed for model parameter estimation. Finally, the proposed distribution is applied to fit with a real data set. The finding result of fitting the proposed distribution to real data set is included. It shows that the proposed distribution has the smallest minus log-likelihood where fitting to the real data set among zero-truncated Poisson and zero-truncated Poisson-Shanker distributions.

Keywords : Zero-truncated, Count data, Bell distribution, Maximum likelihood estimation
2000 Mathematics Subject Classification : 62E10; 62F10

1 Introduction

Count data is referred to the number of observations, events and items that are enumerated per period or area, so it is nonnegative integer values [1]. Moreover, count data has a role in many fields, including public health, medicine, epidemiology, biology, ecology, agriculture, social science, econometric, and engineering [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12].

In count data analysis, Poisson distribution is played central role. Poisson distribution is fitted with the data which its mean equals variance, it is called equidispersion [1, 7]. In practice, the mean is not equal the variance, if the mean is less than the variance called overdispersion whereas if the mean is greater than the variance called underdispersion [1].

Bell distribution is a special case of multiple Poisson process that has one parameter. The importance of Bell distribution's properties is the small value of the parameter in Bell distribution caused approaching Poisson distribution [13].

In count data analysis, sometime zero counts are excluded in the analysis, it is so called zero-truncated distribution [9]. Zero-truncated distribution is the distribution that separating zero counts from the data and analyzing the other accordingly [1,6]. Some examples of zero-truncated distributions are zero-truncated Poisson (ZTP) [2], zero-truncated negative binomial (ZTNB) [14], zero-truncated Poisson-Shanker (ZTPS) [11] and zero-truncated negative binomial-generalized exponential (ZTNB-GE) distributions [10].

Sometime, zero counts are not interested, e.g. the length of stay of a patient in the hospital [1,6], the time that a whale stays at the surface before submerging per hour [15], the number of injured from the

[1] Corresponding author.

accident on the road per month [15], etc., so the zero-truncated distribution is proposed.

In this research, we proposed zero-truncated Bell (ZTB) distribution. The characteristics (mean and variance) and the plots of its probability mass function (pmf) are provided. The maximum likelihood estimation (MLE) is used to estimate the parameter. Finally, we will show the efficiency of ZTB distribution comparing with ZTP and ZTPS distributions with the real data set.

2 Preliminaries

2.1 Zero-Truncated Distribution

In count data analysis, we can consider into 2 categoricals, truncated part and un-truncated part. In this research, we truncate the data that no observation (zero-truncated). The probability mass function (pmf) of the zero-truncated distribution can be written as

$$f(x) = \frac{g(x)}{1-g(0)}; \quad x = 1, 2, 3, ... \tag{2.1}$$

where $g(x)$ is the pmf of the un-truncated distribution [10, 16, 17, ,18, 19].

In the literature, there are many zero-truncated distribution e.g. zero-truncated Poisson (ZTP) distribution [2], Zero-truncated negative binomial (ZTNB) distribution [14].

The ZTP distribution was introduced in 1952 by David and Johnson [2]. If X is distributed as ZTP with parameter λ, then its pmf is

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{(1-e^{-\lambda})x!}; \quad x = 1, 2, 3, ... \tag{2.2}$$

In the last century, there were many authors investigated and extended the ZTP distribution, e.g., several goodness of fit tests for the ZTP distribution were presented [20] although, the ZTP distribution is modified and applied, e.g., zero-truncated bivariate Poisson distribution [21], truncated generalized Poisson distribution [22], truncated generalized Poisson distribution [5], intervened truncated Poisson distribution [18], extension of truncated Poison distribution [23], zero-truncated Poisson-Lindley distribution [24], zero-truncated Poisson-Sujatha distribution [9], zero-truncated Poisson-Shanker (ZTPS) distribution [11].

In addition, ZTP distribution is appropriate to fit the real data, e.g., the number of defective tablets in phamaceutical products [4], the number of gall-cells in knapweed gall-fly in flower-heads of black knapweed [3], the number of emissions of particles [25], the number of drug users [8]. ZTP distribution also has a role in regression model e.g. fitting with the adenomatous polyps [26], fitting with the illegal immigrants in four large cities in the Netherlands [27].

Zero-truncated Poisson-Shanker (ZTPS) distribution was introduced in 2017 by Shanker et al. [11]. If X is distributed as ZTPS with parameter λ, then its pmf is

$$f(x;\lambda) = \frac{\lambda^2}{\lambda^3+\lambda^2+2\lambda+1}\cdot\frac{x+(\lambda^2+\lambda+1)}{(\lambda+1)^{x}}; \quad x = 1, 2, 3, ... \tag{2.3}$$

Zero-truncated negative binomial (ZTNB) distribution was introduced in 1955 by Sampfort [14]. If X is distributed as ZTNB with parameter n and p, then its pmf is

$$f(x;n,p) = \binom{x+n-1}{n}\frac{p^{n}(1-p)^{x}}{1-p^{n}}; \quad x = 1, 2, 3, ... \tag{2.4}$$

In the last century, there were many authors investigated and extended the ZTNB distribution, e.g., the zero-truncated symmetrical bivariate negative binomial distribution (SBNBD) [28]. In addition, there were many authors applied the ZTNB distribution, e.g., fitting the ZTNB distribution with the real data to propose appropriate regression [29].

Zero-truncated negative binomial-generalized exponential (ZTNB-GE) distribution was introduced in 2016 by Bodhisuwan [10]. If X is distributed as ZTNB-GE with parameter r, α and β then its pmf is

$$f(x;r,\alpha,\beta)=\frac{1}{1-g(0)}\binom{r+x-1}{x}\sum_{j=0}^{\infty}\binom{x}{j}\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)};\quad x=1,2,3,... \tag{2.5}$$

where $r>0$, $\alpha>0$ and $\beta>0$, $g(0)=\frac{\Gamma(\alpha+1)\Gamma(1+\frac{1+r}{\beta})}{\Gamma(\alpha+\frac{r}{\beta}+1)}$, and $\Gamma(\cdot)$ is the gamma function.

Zero-truncated negative binomial-Erlang (ZTNB-EL) distribution was introduced in 2017 by Bodhisuwan et al [12]. If X is distributed as ZTNB-EL with parameter r, c and k, then its pmf is

$$f(x;r,c,k)=\frac{\binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j(\frac{c}{c+r+j})^k}{1-(\frac{c}{c+r})^k};\quad x=1,2,3,... \tag{2.6}$$

where $r>0$, $c>0$ and $k>0$.

2.2 Bell distribution

Bell distribution was introduced in 2018 by Castellares et al. [13] which is improved from Bell numbers [30, 31] and multiple Poisson process [32]. If X is distributed as Bell with parameter λ, then its pmf is

$$f(x;\lambda)=\frac{\lambda^x e^{(-e^{\lambda})+1}B_x}{x!} \tag{2.7}$$

where $\lambda>0$ and $B(x)$ is the Bell numbers [13, 30, 31].

2.3 The Maximum Likelihood Estimation

In the early twentieth century, Bu Sir Ronald Fisher developed method of parameter estimation, so he proposed his method named maximum likelihood estimation (MLE). It is a one of method for estimating parameters in statistical models. It gave the absolute criterion named likelihood function for the parameter [33, 34, 35].

Let $X=(x_1,\ x_2,...,\ x_n)$ be an independent and identically distributed (iid) random sample from a population with probability mass function $f(x_i|\lambda_1,\lambda_2,...,\lambda_k)$, $i=1,2,...,n$. The likelihood function of the sample is

$$\mathcal{L}(\lambda_1,\lambda_2,...,\lambda_k;x)=\prod_{i=1}^{n}f(x_i|\lambda_1,\lambda_2,...,\lambda_k). \tag{2.8}$$

The maximum likelihood estimators of λ_j, denoted by $\hat{\lambda}_j$, $j=1,2,...,k$ are those values of the parameters that maximize the likelihood function with respect to the parameter λ_j. That is,

$$\mathcal{L}(\hat{\lambda}_1,\hat{\lambda}_2,...,\hat{\lambda}_k;x)=\max_{(\lambda_1,\lambda_2,...,\lambda_k)\in\mathbb{R}^k}\mathcal{L}(\lambda_1,\lambda_2,...,\lambda_k;x). \tag{2.9}$$

In general, the MLE gave maximum likelihood function of one or several variables, so the methods of calculus is used in most situations. If the likelihood function is differentiable, possible candidates for the maximum likelihood estimators are the values $\hat{\lambda}_1,\hat{\lambda}_2,...,\hat{\lambda}_k$ that solves

$$\frac{\partial}{\partial \lambda_j} \mathcal{L}(\lambda_1, \lambda_2, ..., \lambda_k; x) = 0, \quad j = 1, 2, ..., k. \tag{2.10}$$

Taking the natural logarithm (log) to the likelihood function, Eq. (2.10) is easier to work, we will get the log-likelihood function, denoted by

$$\ell(\boldsymbol{\lambda}; x) = \log \mathcal{L}(\lambda_1, \lambda_2, ..., \lambda_k; x) = \sum_{i=1}^{n} \log f(x_i|\lambda_1, \lambda_2, ..., \lambda_k), \tag{2.11}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_k)^T$.

Because the natural logarithm function is increasing, the maximum value of the likelihood function will occur at the same point as that of the log-likelihood function. The calculus-based procedure to find the maximum likelihood estimators based log-likelihood function can be summarized as

$$\frac{\partial}{\partial \lambda_j} \ell(\boldsymbol{\lambda}; x) = \frac{\partial}{\partial \lambda_j} \log \mathcal{L}(\lambda_1, \lambda_2, ..., \lambda_k; x) = 0, \quad j = 1, 2, ..., k. \tag{2.12}$$

3 Main Results

3.1 The pmf of zero-truncated Bell distribution

Theorem 3.1. *Let X be the ZTB random variable with parameter λ. It is denoted $X \sim ZTB(\lambda)$. Then, the pmf of X can be obtained by*

$$f(x; \lambda) = \frac{\lambda^x e^{(-e^\lambda)+1} B_x}{x!} \cdot \frac{1}{1 - e^{(-e^\lambda)+1}}; \quad x = 1, 2, 3, ... \tag{3.1}$$

where $\lambda > 0$.

Proof. From, equation (2.1), substituting the pmf of the Bell distribution in (2.7), the pmf of ZTB distribution can be obtained by

$$f(x) = \frac{\lambda^x e^{(-e^\lambda)+1} B_x}{x!} \quad , \tag{3.2}$$

and

$$f(0) = 1 - e^{(-e^\lambda)+1} \quad , \tag{3.3}$$

then the pmf for the ZTB distribution can be expressed as equation (3.1). □

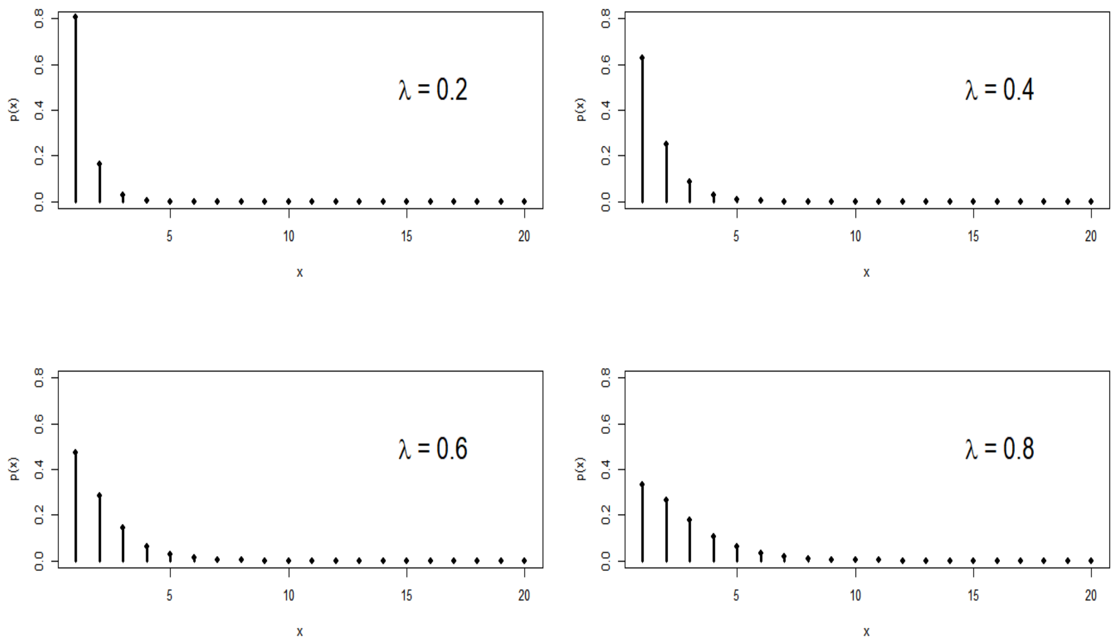

Figure 1: Some pmf plots of ZTB distribution with specified parameter λ.

Theorem 3.2. *If $X \sim ZTB(\lambda)$, then the mean and variance of X can be obtained respectively by*

$$E(X) = \frac{\lambda e^{\lambda}}{1 - e^{(-e^{\lambda})+1}} \tag{3.4}$$

and

$$Var(X) = \frac{\lambda e^{\lambda}(1+\lambda)}{1 - e^{(-e^{\lambda})+1}} - \frac{\lambda^2 e^{2\lambda}}{(1 - e^{(-e^{\lambda})+1})^2}. \tag{3.5}$$

Proof. Let $X \sim ZTB(\lambda)$, then the mean of X is

$$E(X) = \sum_{x=0}^{\infty} x f(x) = \frac{1}{1 - e^{(-e^{\lambda})+1}} \sum_{x=0}^{\infty} x\left(\frac{\lambda^x e^{(-e^{\lambda})+1} B_x}{x!}\right), \tag{3.6}$$

where B_x is Bell numbers, x=1,2,3,... Since $X \sim Bell(\lambda)$, the mean of Y is

$$E(Y) = \sum_{y=0}^{\infty} y f(y) = \sum_{y=0}^{\infty} y\left(\frac{\lambda^y e^{(-e^{\lambda})+1} B_y}{y!}\right) = \lambda e^{\lambda} \tag{3.7}$$

then

$$E(X) = \frac{E(Y)}{1 - e^{(-e^{\lambda})+1}} = \frac{\lambda e^{\lambda}}{1 - e^{(-e^{\lambda})+1}}. \tag{3.8}$$

Since

$$E(X^2) = \frac{1}{1 - e^{(-e^{\lambda})+1}} \sum_{x=0}^{\infty} x^2\left(\frac{\lambda^x e^{(-e^{\lambda})+1} B_x}{x!}\right) = \frac{E(Y^2)}{1 - e^{(-e^{\lambda})+1}}. \tag{3.9}$$

From $Var(X) = E(X^2) - (E(X))^2$, then

$$Var(X) = \frac{\lambda e^{\lambda}(1+\lambda)}{1-e^{(-e^{\lambda})+1}} - \frac{\lambda^2 e^{2\lambda}}{(1-e^{(-e^{\lambda})+1})^2}. \tag{3.10}$$

□

3.2 Parameter Estimation of the ZTB distribution

MLE method is used to obtain the estimated parameter λ of ZTB distribution. If $X \sim ZTB(\lambda)$, then the likelihood function of parameter λ is

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{(-e^{\lambda})+1} B_{x_i}}{x_i!} \cdot \frac{1}{1-e^{(-e^{\lambda})+1}} \tag{3.11}$$

and, the associated log-likelihood function is

$$\ell = (\sum_{i=1}^{n} \log(\frac{\lambda^{x_i} e^{(-e^{\lambda})+1} B_{x_i}}{x_i!})) - n\log(1-e^{(-e^{\lambda})+1}). \tag{3.12}$$

To find the optimal values of λ, we will differentiate Eq. (3.12) with respect to λ, that is

$$\frac{d}{d\lambda}\ell = \frac{d}{d\lambda}[(\sum_{i=1}^{n} \log(\frac{\lambda^{x_i} e^{(-e^{\lambda})+1} B_{x_i}}{x_i!})) - n\log(1-e^{(-e^{\lambda})+1})]. \tag{3.13}$$

We can find the MLE solution by using numerical optimization with the optim function in R language [36].

3.3 Application

The ZTB distribution has been fitted to the real data set, that is the number of sites with particles from Immunogold data reported by Mathews and Appleton (1993) [9, 37]. Performances of ZTP, ZTPS and ZTB distributions are compared by using minus log-likelihood (-LL). Anderson-Darling (AD) test statistic is used to evaluate goodness of fit test of the ZTP, ZTPS and ZTB distributions [20]. We found that the ZTB distribution has a smallest minus log-likelihood.

Table 1: Observed and expected frequencies for the number of sites with particles from Immunogold data reported by Mathews and Appleton (1993).

Number of Sites with Particles	Observed Frequency	Expected Freqency		
		ZTP	ZTPS	ZTB
1	122	115.8614	125.1286	121.0010
2	50	57.3876	46.3400	51.0709
3	18	18.9499	16.9569	17.9629
4	4	4.6931	6.1448	5.6862
5	4	0.9298	2.2088	1.6640
Estimated parameter		$\lambda = 0.9906$	$\lambda = 2.0313$	$\lambda = 0.4221$
-Log-likelihood		205.9477	204.7125	204.3499
AD-statistic		0.5159	0.1119	0.0337
p-value		0.4603	0.8747	0.9818

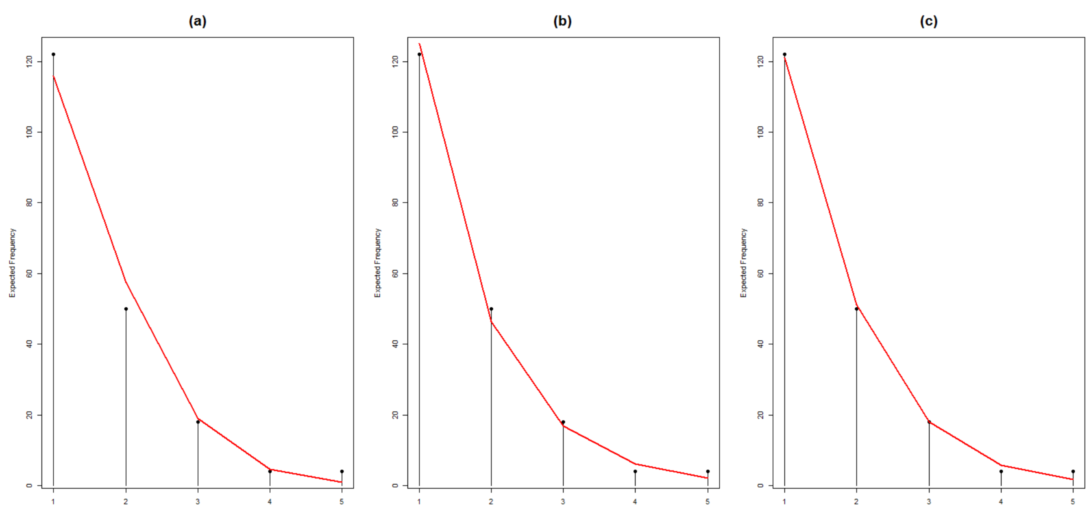

Figure 2: Results of fitting distribution (a) ZTP, (b) ZTPS and (c) ZTB distributions to real data set.

4 Conclusion

In this paper, zero-truncated Bell (ZTB) distribution is proposed to be an alternative distribution for count data analysis which the response variable does not contain zero values. Some properties of the distribution are presented. The parameter estimation method using the maximum likelihood estimation (MLE) is discussed. The ZTB distribution performed the best in applying with the real data set that is the number of sites with particles from Immunogold data. A goodness of fit test based on the Anderson-Darling (AD) test is employed. The result shows that ZTB distribution is the smallest minus loglikelihood and the smallest Anderson-Darling test when it is comparing with ZTP and ZTPS distributions, it means the ZTB distribution is the best fit for this real data set.

Acknowledgements : We would like to thank the Department of Statistics, Faculty of Science, Kasetsart University. The first author would like to thank the Science Achievement Scholarship of Thailand (SAST) for financial support.

References

[1] J. M. Hilbe, Modeling Count Data, Cambridge University Press, New York, 2014.

[2] F. N. David and N. L. Johnson, The truncated Poisson, Biometrics, 8 (1952) 275-285.

[3] D.J. Finney and G. C. Varley, An example of the truncated Poisson distribution, Biometrics, 11 (1955) 387-394.

[4] N. R. Bohidar, On truncated Poisson distribution for determining mean number of defectives in pharmaceutical products, Drug Development and Industrial Pharmacy, 18 (1992) 1373-1393.

[5] D. P. M. Scollnik, On the analysis of the truncated generalized Poisson distribution using a Baysian method, Astin Bulletin, 28 (1998) 135-152.

[6] A. Zuur, E. Ieno, N. Walker, A. Saveliev, and G.Smith, Mixed effects models and extensions in ecology with r, Springer, New York, 2009.

[7] C. Jornsatian, Statistical analysis of zero inflated negative binomial-beta exponential distribution, In proceedings of the 54 th Kasetsart University Annual Conference, Bangkok, Thailand, 2-5 February (2016) 1-8.

[8] C.Viwatwongkasem, EM algorithm for truncated and censored Poisson likelihoods, In proceedings of International Electrical Engineering Congress (iEECON), Chiangmai, Thailand, 2-4 March (2016) 240-243.

[9] R. Shanker and H. Fesshaye, On zero-truncation of Poisson, Poisson-Likndley and Poisson-Sujatha distributions and their applications, Biometric and Biostatistics International Journal, 3 (2016) 1-13.

[10] R. Bodhisuwan, A mathematical model of count data analysis based on ZTNB-GE distribution, In proceedings of the 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA), Banda Aceh, Indonesia, 4-9 October (2016) 113-116.

[11] R. Shanker, A zero-truncated Poisson-Shanker distribution and its applications, International Journal of Statistics and Applications, 7 (2017) 159-169.

[12] W. Bodhisuwan, C. Pudprommarat, R. Bodhisuwan, and L. Saothayanun, Zero-truncated negative binomial-Erlang distribution, In proceedings of the 13 th International Conference on Mathematics, Statistics, and Their Applications (ICMSA), Kedah, Malaysia, 4-7 December (2017) 1-6.

[13] F. Castellares, S.L.P. Ferrari and A. J. Lemonte, On the Bell distribution and its associated regression model for count data, Applied Mathematical Modeling, 56 (2018) 172-185.

[14] M. R. Sampford, The truncated negative binomial, Biometrics, 42 (1955) 58-69.

[15] S. Sookkua, C. Wongoutong and W. Bodhisuwan, Zero truncated negative binomial-beta exponential distribution, In proceedings of the 3 rd Ratchathani University National Conference (RTUNC), Ubonratchani, Thailand, (2018) 229-237.

[16] P. R. Rider, Truncated Poisson distributions, Journal of the American Statistical Association, 48 (1953) 826-830.

[17] K. Subrahmaniam, A note on estimation in the truncated Poisson, Biometrika, 52 (1965) 279-281.

[18] M. L. Huang and K. Y. Fung, Intervened truncated Poisson distribution, The Indian Journal of Statistics, 51 (1989) 302-310.

[19] J. T. Grogger and R. T. Carson, Models for truncated counts, Journal of Applied Ecconometrics, 6(1991) 225-238.

[20] D. J. Best, J. C. W. Rayner and O. Thas, Goodness of fit for the zero-truncated Poisson distribution, Journal of Statistical Computation and Simulation, 77 (2007) 585-591.

[21] M. A. Hamdan, Estimation in the truncated bivariate Poisson distribution, Technometrics, 14 (1972) 37-45.

[22] P. C. Consul and F. Famoye, The truncated generalized Poisson distribution and its estimation, Communications in Statistics-Theory and Methods, 18 (1989) 3635-3648.

[23] A. C. Cohen, An extension of a truncated Poisson distribution, Biometrics, 16 (1960) 446-450.

[24] M. E. Ghitany, D. K. Al-Mutairi and S. Nadarajah, Zero-truncated Poisson-Lindley distribution and its applications, Mathematics and Computers in Simulation, 79 (2008) 279-287.

[25] P. G. Moore, The estimation of the Poisson parameter from a truncated distribution, Biometrika, 39 (1952) 247-251.

[26] T. Xie and M. Aickin, A truncated Poisson regression model with applications to occurance of adenomatous polyps, Statistics in Medicine, 16 (1997) 1845-1857.

[27] M. J. L. F. Cruyff and P. G. M. van der Heijan, Point and interval estimation of the population size using the truncated Poisson regression model, Statistical Modelling, 3 (2003) 305-322.

[28] A. K. Sinha and R. Kumar, The zero-truncated symmetrical bivariate negative binomial distribution, American Journal of Mathematical and Management Sciences, 21 (2001) 57-68.

[29] C. T. Arrabel, K. P. dos S. Silva and L. N. Bandeira, Zero-truncated negative binomial applied to nonlinear data, Journal of Biostatistics, 11 (2014) 55-67.

[30] E. T. Bell, Exponential polynomials, Annals of Mathematics, 35 (1934) 258-277.

[31] E. T. Bell, Exponential numbers, The American Mathematical Monthly, 41 (1934) 411-419.

[32] L. J*á*nossy, A. R*é*nyi and J. Acz*é*l, On composed Poisson distributions: 1, Acta Mathematica Hungarica, 1 (1950) 209-224.

[33] G. C. White and L. E. Eberhardt, Statistical analysis of deer and elk Pellet-group data, The Journal of Wildlife Management, 44 (1980) 121-131.

[34] J. X. Pan and K. T. Fang, Growth curve models and statistical diagnostics, Springer, New York, 2002.

[35] S. Sangpoom and W. Bodhisuwan, The discrete asymmetric Laplace distribution, Journal of Statistical Theory and Practice, 10 (2016) 73-86.

[36] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (2018).

[37] J. N. S. Mathews and D. R. Appleton, An application of the truncated Poisson distribution to Immunogold assay, Biometrics, 49 (1993) 617-621.