Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:87
https://doi.org/10.28919/cmbn/8725
ISSN: 2052-2541

ZERO-TRUNCATED POISSON-BILAL DISTRIBUTION AND ITS APPLICATIONS TO BIOLOGICAL SCIENCES

THANASATE AKKANPHUDIT*

Research Administration Center Siam Technology College, Bangkok 10600, Thailand

Abstract: This paper proposes a zero-truncated version of the Poisson-Bilal distribution called the zero-truncated Poisson-Bilal (ZTPB) distribution. The ZTPB distribution with one parameter can be used to model extremely right-skewed and one-inflated count data sets. The method of maximum likelihood is used to estimate the parameter in the model. Its application to some real data sets from biological science has been given. Based on the goodness of fit, the ZTPB distribution gives a better fit than some zero-truncated distributions with one parameter. Based on the simulation study, the maximum likelihood estimator gives the estimate value is close to the true value.

Keywords: zero-truncated data; Poisson-Bilal distribution; MLE; biological sciences; count data.

2020 AMS Subject Classification: 97K80, 97K50, 91G70.

1. INTRODUCTION

In probability theory, zero-truncated distributions are certain discrete distributions that support the set of positive integers but exclude the value zero from their support. Zero-truncated distributions are suitable models for modeling data when the data originates from a mechanism that generates data excluding the zero value. In other words, these distributions only assign probabilities to

---

*Corresponding author
E-mail address: thanasatea@siamtechno.ac.th
Received June 25, 2024

positive integer values. The adjustment of the probability mass function ensures that the sum or integral over the positive values equals one, causing the truncation at zero [1, 2].

In the biological sciences, a zero-truncated distribution can be useful when analyzing count data. A zero-truncated distribution is a probability distribution that models the count of data excluding the value zero. In other words, it assumes that the minimum possible value for the data is one or more than one. Count data are often used in the biological sciences to study the number of events or occurrences of a particular phenomenon, such as the number of cells, organisms, mutations, or gene expressions. However, zero values are sometimes impossible due to a variety of factors, such as experimental design or biological constraints. For instance, studying the number of offspring produced by organisms may require zero-truncation, as the inclusion of at least one offspring in the data is a requirement. Studies involving the number of bacterial colonies on a petri dish also apply zero-truncation, as unfavorable growth conditions prevent the observation of colonies. The Poisson distribution is the most commonly encountered zero-truncated distribution. The Poisson distribution models the number of events occurring in a fixed time or space interval. However, in some situations, a zero value is impossible or not observed. For instance, if we count the number of phone calls received in a day, we are unlikely to observe zero calls, as we expect at least some calls. In such cases, the zero-truncated Poisson (ZTP) distribution is more appropriate than the Poisson distribution [2, 3, 4, 5].

It is worth noting that zero-truncated distributions can also be defined for continuous random variables. In such cases, we adjust the probability density function to exclude the zero value from the support. Zero-truncated distributions with one parameter were proposed, such as the zero-truncated Poisson-Lindley (ZIPL) distribution [1, 6], the zero-truncated Poisson-Akash (ZTPA) distribution [2], the zero-truncated Poisson-Sujatha (ZTPS) distribution [7, 8], the zero-truncated Poisson-Garima (ZTPG) distribution [9], the zero-truncated Poisson-Ishita (ZTPI) distribution [10], etc. Modelers use these distributions to model count data where zero values are either impossible or unobserved.

In 2020, a new one-parameter discrete distribution called the Poisson-Bilal distribution was proposed for modeling the over-dispersed count data sets, which is proposed by [11]. The Poisson-Bilal distribution has one parameter and simple forms for its probability density, cumulative distribution functions, moments, and probability-generating functions. Moreover, the Poisson-Bilal distribution is a more flexible alternative to analyzing count data with overdispersion.

This paper proposes a zero truncation of the Poisson-Bilal distribution. Some statistical properties have been given. We have discussed the method of maximum likelihood for parameter estimation. The proposed distribution has been applied to a real dataset to test its goodness of fit over some zero-truncated distributions. In addition, simulation study for parameter estimation are provided.

2. PRELIMINARIES

In this section, the Poisson-Bilal and zero-truncated distributions are described. The Poisson-Bilal (PB) distribution has one parameter, which is obtained by compounding the Poisson distribution and the Bilal distribution [11]. Its probability mass function (pmf) is

$$g_0(x;\theta) = 6\theta\left[\frac{1}{(2\theta+1)^{x+1}} - \frac{1}{(3\theta+1)^{x+1}}\right] \text{ for } x = 0,1,2,\ldots \text{ and } \theta > 0. \tag{1}$$

Its corresponding cumulative density function (cdf) is

$$G_0(x) = 1 + \frac{2}{(3\theta+1)^{x+1}} - \frac{3}{(2\theta+1)^{x+1}}. \tag{2}$$

The factorial moment of order r about the origin of the PB distribution is

$$\mu_{\mathrm{PB}[r]} = \frac{(3^{r+1} - 2^{r+1})\Gamma(r+1)}{(6\theta)^r} \text{ for } r = 1,2,3,\ldots. \tag{3}$$

From (3), we have the mean and variance of the PB distribution as follows:

$$E_{\mathrm{PB}}(X) = \frac{5}{6\theta} \text{ and } V_{\mathrm{PB}}(X) = \frac{30\theta+13}{36\theta^2}. \tag{4}$$

Suppose $g_0(x;\mathbf{\Theta})$ is the original distribution with a parameter vector of $\mathbf{\Theta}$. Then the pmf of the zero-truncated version of $g_0(x;\mathbf{\Theta})$ can be defined as

$$f(x;\mathbf{\Theta}) = \frac{g_0(x;\mathbf{\Theta})}{1 - g_0(0;\mathbf{\Theta})} \text{ for } x = 1,2,3,\ldots \tag{5}$$

where $g_0(0;\mathbf{\Theta})$ is the original distribution for $x = 0$ [1, 2].

Some zero-truncated distribution that has one parameter are shown in Table 1.

Table 1. Some zero-truncated distributions.

Distribution	pmf	Mean	Authors
ZTP	$\frac{\theta^x}{x!(e^\theta-1)}$	$\frac{\theta e^\theta}{e^\theta-1}$	[2, 4, 5]
ZTPL	$\frac{\theta^2(x+\theta+2)}{(\theta^2+3\theta+1)(\theta+1)^x}$	$\frac{(\theta+1)^2(\theta+2)}{\theta(\theta^2+3\theta+1)}$	[1, 6]
ZTPS	$\frac{\theta^3\left[x^2+(\theta+4)x+(\theta^2+3\theta+4)\right]}{(\theta^4+4\theta^3+10\theta^2+7\theta+2)(\theta+1)^x}$	$\frac{\theta^5+5\theta^4+15\theta^3+25\theta^2+20\theta+6}{\theta(\theta^4+4\theta^3+10\theta^2+7\theta+2)}$	[7, 8]
ZTPA	$\frac{\theta^3\left[x^2+3x+(\theta^2+2\theta+3)\right]}{(\theta^4+2\theta^3+7\theta^2+6\theta+2)(\theta+1)^x}$	$\frac{\theta^5+3\theta^4+9\theta^3+19\theta^2+18\theta+6}{\theta(\theta^4+2\theta^3+7\theta^2+6\theta+2)}$	[2]
ZTPG	$\frac{\theta\left[\theta x+(\theta^2+3\theta+1)\right]}{(\theta^2+4\theta+2)(\theta+1)^x}$	$\frac{(\theta+1)(\theta^2+4\theta+3)}{\theta(\theta^2+4\theta+2)}$	[9]
ZTPI	$\frac{\theta^3\left[x^2+3x+(\theta^3+2\theta^2+\theta+2)\right]}{(\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2)(\theta+1)^x}$	$\frac{\theta^6+3\theta^5+3\theta^4+7\theta^3+18\theta^2+18\theta+6}{\theta(\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2)}$	[10]

3. MAIN RESULTS

This section proposes the zero-truncated version of the PB distribution and its properties. Next, the parameter estimation of the parameter of the proposed distribution is introduced. Finally, applications to some biological data are illustrated for the proposed distribution and other distributions.

3.1 A zero-truncated Poisson-Bilal distribution

A new zero-truncated distribution is obtained by taking the zero-truncated version of the BP distribution by proposed [11]. Using equations (1) and (5), the pmf of zero-truncated Poisson-Bilal (ZTPB) distribution can be obtained as

$$f(x;\theta)=\frac{6\theta}{5\theta+1}\left[\frac{3\theta+1}{(2\theta+1)^x}-\frac{2\theta+1}{(3\theta+1)^x}\right] \text{ for } x=1,2,3,\ldots \text{ and } \theta>0. \tag{6}$$

Its corresponding cdf is

$$F(x;\theta)=1+\frac{1}{5\theta+1}\left[\frac{2(2\theta+1)}{(3\theta+1)^{x}}-\frac{3(3\theta+1)}{(2\theta+1)^{x}}\right]. \tag{7}$$

The probability generating function of the ZTPB distribution can be obtained as follows

$$E(s^{x})=\frac{6\theta}{5\theta+1}\left[\frac{(3\theta+1)s}{2\theta+1-s}-\frac{(2\theta+1)s}{3\theta+1-s}\right].$$

The moment generating function the ZTPB distribution are respectively

$$M_{X}(t)=\frac{6\theta}{5\theta+1}\left[\frac{(3\theta+1)e^{t}}{2\theta+1-e^{t}}-\frac{(2\theta+1)e^{t}}{3\theta+1-e^{t}}\right].$$

The mean and variance of the ZTPB distribution are respectively

$$E_{\mathrm{ZTPB}}(X)=\frac{5(2\theta+1)(3\theta+1)}{6\theta(5\theta+1)} \text{ and } V_{\mathrm{ZTPB}}(X)=\frac{(2\theta+1)(3\theta+1)(95\theta+13)}{36\theta^{2}(5\theta+1)^{2}}.$$

Now, Figure 1 shows the pmf shapes of the ZTPB for different parameter values of $\theta>0.$ The ZTPB distribution can be used to model extremely right-skewed and one-inflated count data sets.

3.2 Parameter estimation of the ZTPB distribution

The maximum likelihood (ML) method is considered to estimate the parameter θ of the ZTPB distribution. Let $\mathbf{x}=(x_1,\ldots,x_n)^T$ be the vector of the observed value. The likelihood function of the ZTPB distribution is given by

$$L(\mathbf{x};\theta)=\frac{6^{n}\theta^{n}}{(5\theta+1)^{n}}\prod_{i=1}^{n}\left[\frac{3\theta+1}{(2\theta+1)^{x_i}}-\frac{2\theta+1}{(3\theta+1)^{x_i}}\right] \text{ for } x_i=1,2,3,\ldots. \tag{8}$$

Its log-likelihood function is

$$\log L(\mathbf{x};\theta)=n\log 6+n\log\theta-n\log(5\theta+1)+\sum_{i=1}^{n}\log\left[\frac{3\theta+1}{(2\theta+1)^{x_i}}-\frac{2\theta+1}{(3\theta+1)^{x_i}}\right]. \tag{9}$$

Taking the first order derivative of equation (9) with respect to $\theta,$ we have

$$\frac{d}{d\theta}\log L(\mathbf{x};\theta)=\frac{n}{\theta}-\frac{5n}{5\theta+1}+\sum_{i=1}^{n}\left\{\frac{1}{(3\theta+1)^{x_i+1}-(2\theta+1)^{x_i+1}}\left[3(3\theta+1)^{x_i}-2(2\theta+1)^{x_i}\right.\right.\;.$$

$$\left.\left.+\frac{3x_i(2\theta+1)^{x_i+1}}{(3\theta+1)}-\frac{2x_i(3\theta+1)^{x_i+1}}{(2\theta+1)}\right]\right\}\;. \tag{10}$$

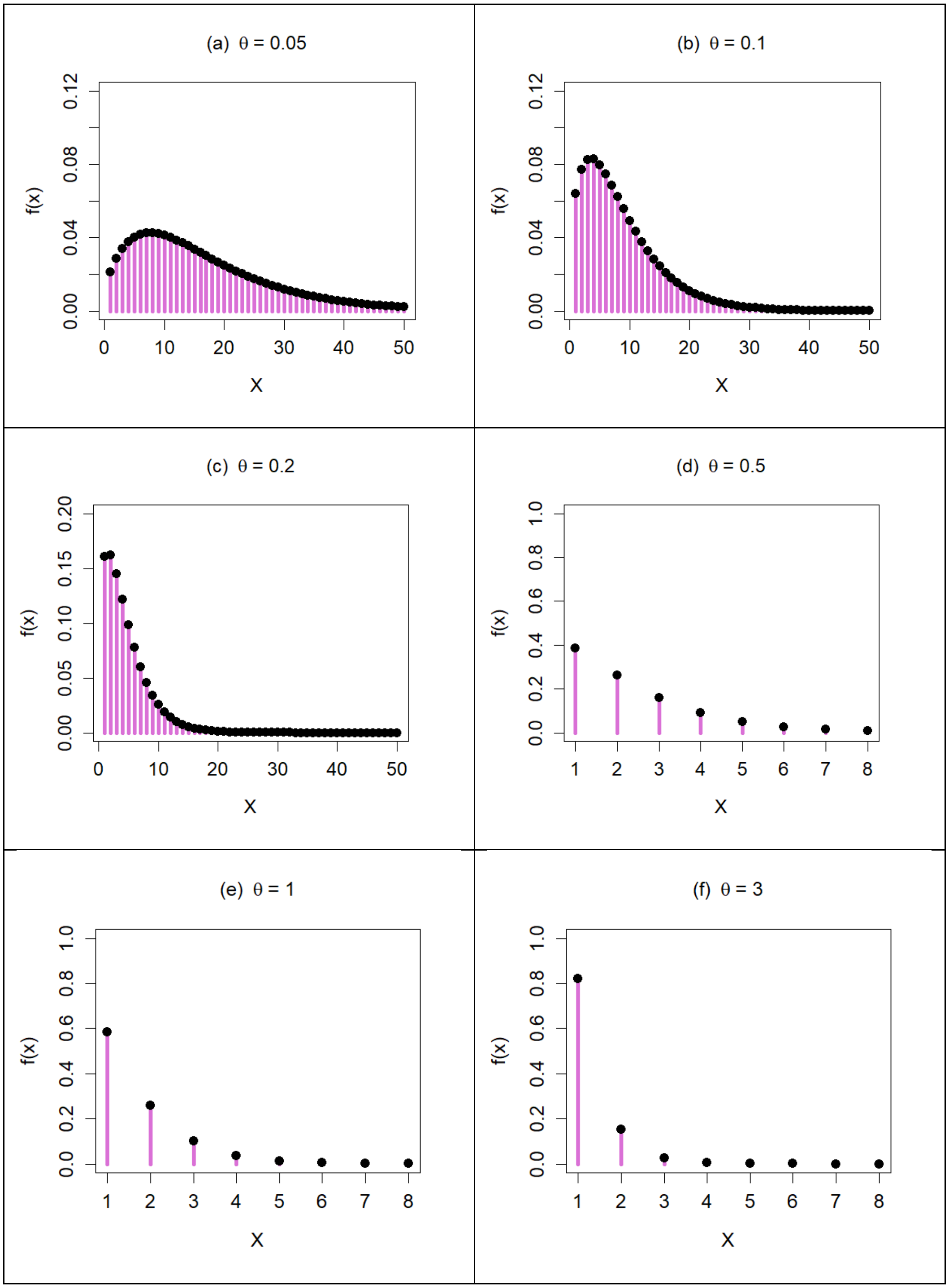


Figure 1. The probability mass function shapes of the ZTPB distribution for selected parameter values.

The ML estimator can be obtained by equating equation (10) to zero and solving for the parameter. However, as seen from equation (10), obtaining the explicit form of the ML method is impossible. Therefore, equation (10) has to be solved using numerical methods. In this paper, the direct maximization of the log-likelihood function, given in equation (9), using the statistical software R [12] with the *nlm* function in the **stats** package.

3.3 Applications of the ZTPB distribution to biological sciences

This section considers the real data set to fit with the ZTPB distribution. The proposed ZTPB distribution has been fitted to several data sets using the ML estimate. It is the number of egg cells on a flower head. The eggs are easily seen and counted when the flower head is split open [2, 9, 13, 14]. The number of empty flower heads is omitted because many cause irrelevant to the investigation may secure that no eggs are laid, which the data shows in Table 2.

The ZTPB distribution has been compared with other distributions, including the ZTP, ZTPL, ZTPS, ZTPA, ZTPG, and ZTPI distributions. The criteria for model selection are the lowest of the chi-square value and the highest of p-value based on the chi-square test for goodness of fit test. The results are shown in Table 2. The ML estimate of the ZTP, ZTPL, ZTPS, and ZTPA distributions have the value corresponding to the ML estimate of [2] for decimal four digits. The ZTPI distribution's ML estimate corresponds to the ML estimate of [9] for decimal four digits. In addition, the ZTPG and ZTPB are 0.6292 and 0.3330, respectively. These results illustrate that the ZTPB distribution has the minimum chi-square value and the maximum p-value based on the chi-square test. Figure 2 displays the observed frequencies and the expected frequencies of the distributions. The expected frequencies of the ZTPB distribution are close to the observed frequencies. Therefore, the proposed distribution can be applied to data that exclude zero.

Table 2. Observed and expected frequencies for the number of counts of flower heads as per the number of fly eggs (X) reported by [13].

X	Observed frequency	Expected frequency						
		ZTP	ZTPL	ZTPS	ZTPA	ZTPG	ZTPI	ZTPB
1	22	15.6	27.4	25.7	25.7	27.7	25.5	24.7
2	18	22.4	20.2	20.4	20.2	19.7	20.2	20.5
3	18	21.3	14.3	15.0	15.0	13.8	15.0	15.1
4	11	15.2	9.7	10.4	10.5	9.5	10.5	10.5
5	9	8.7	6.5	6.9	7.0	6.5	7.1	7.0
6	6	4.2	4.3	4.5	4.5	4.4	4.6	4.5
7	3	1.7	2.8	2.8	2.8	2.9	2.9	2.9
8	0	0.6 }6.8	1.8 }11.9	1.7 }11.6	1.7 }11.6	1.9 }12.8	1.8 }11.7	1.8 }12.2
9	1	0.2	1.1	1.0	1.0	1.3	1.1	1.1
> 9	0	0.1	1.9	1.6	1.6	2.3	1.3	1.9
$\hat{\theta}$		2.8604	0.7186	0.9814	1.0215	0.6292	1.0141	0.3330
χ^2		9.15	3.40	2.10	1.98	3.85	1.86	1.76
df		4	4	4	4	4	4	4
p-value		0.0575	0.4932	0.7174	0.7394	0.4292	0.7615	0.7798

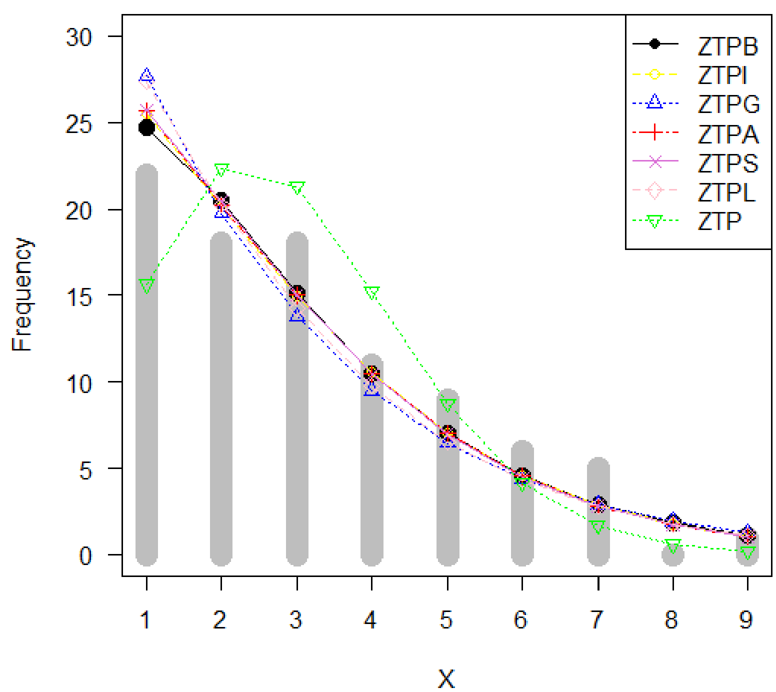
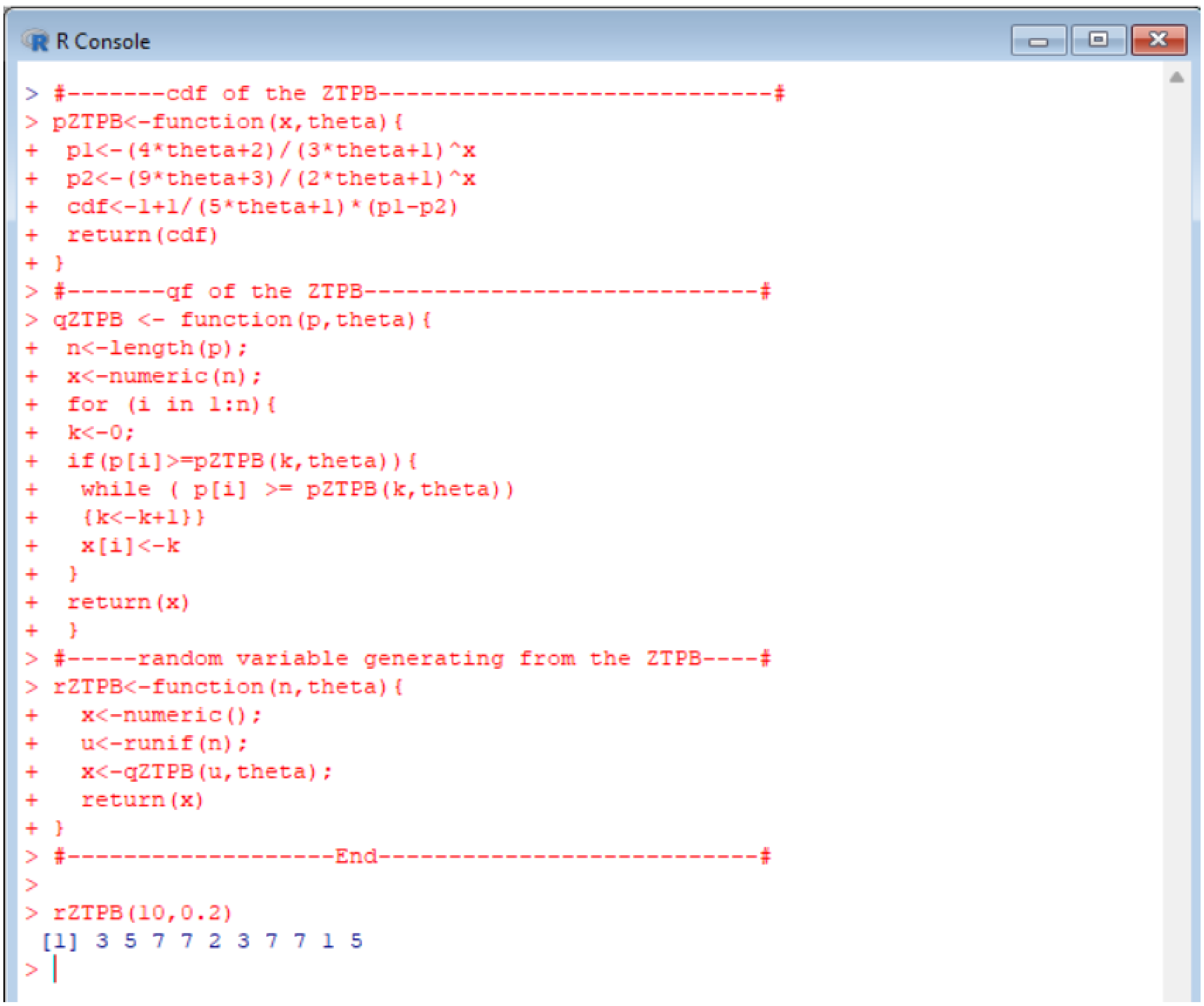


Figure 2. Observed and expected frequencies of the number of fly eggs.

3.4 Simulation study

Based on the cdf in (7), let $F(x;\theta)=U$ where U is a uniform random variable on [0, 1], then the quantile function (qf) of the ZTPB distribution is $Q(u;\theta)=F^{-1}(u;\theta)$. The qf has no closed form solution, so we have to use a numerical technique to get the quantile. Under the study of the R code in [15, 16], we applied it for generating a ZTPB random variables according to the following [12]:



```
R Console
> #-------cdf of the ZTPB---------------------------#
> pZTPB<-function(x,theta){
+  p1<-(4*theta+2)/(3*theta+1)^x
+  p2<-(9*theta+3)/(2*theta+1)^x
+  cdf<-1+1/(5*theta+1)*(p1-p2)
+  return(cdf)
+ }
> #-------qf of the ZTPB---------------------------#
> qZTPB <- function(p,theta){
+  n<-length(p);
+  x<-numeric(n);
+  for (i in 1:n){
+  k<-0;
+  if(p[i]>=pZTPB(k,theta)){
+   while ( p[i] >= pZTPB(k,theta))
+   {k<-k+1}}
+   x[i]<-k
+  }
+  return(x)
+  }
> #-----random variable generating from the ZTPB----#
> rZTPB<-function(n,theta){
+   x<-numeric();
+   u<-runif(n);
+   x<-qZTPB(u,theta);
+   return(x)
+ }
> #------------------End--------------------------#
>
> rZTPB(10,0.2)
 [1] 3 5 7 7 2 3 7 7 1 5
>
```

Figure 3. R code for generating a ZTPB random variables.

This section demonstrates how well the ML estimator θ of the ZTPB distribution works. We achieve this by selecting random samples of sizes ($n=25$, 50, 100, 200, 300, 400, and 500) and values θ of 0.1, 0.333, 1, and 3. With 1,000 replications, consider each case and sample size. The estimators are compared in We express the estimates in terms of the mean square errors (MSE) of the estimator. Table 3 summarizes the values of the ML estimator's average estimates (AEs), absolute biases (ABs), standard deviation (SD), and MLE. According to Table 3, the results are as follows: (i) As sample sizes increase for large n, the estimated values of the suggested estimators become closer to the true parameter. (ii) For a larger sample size, the ABs value approaches zero. (iii) The SD and MSE values are decreasing as the sample size values are increasing for all cases considered. Thus, the ML estimator give the estimate value θ is close to the true value.

Table 3. The values of AEs, Abs, SD, and MSE of ML esitmator θ of the ZTPB distribuiton.

θ	0.333	$n=25$	$n=50$	$n=100$	$n=200$	$n=300$	$n=400$	$n=500$
0.1	AEs	0.104069	0.103439	0.100316	0.100290	0.100259	0.100129	0.100039
	Abs	0.004069	0.003439	0.000316	0.000290	0.000259	0.000129	0.000039
	SD	0.020226	0.012849	0.011806	0.005536	0.004576	0.004046	0.003561
	MSE	0.000426	0.000177	0.000139	0.000031	0.000021	0.000016	0.000013
0.333	AEs	0.345841	0.345505	0.346757	0.334639	0.332256	0.337094	0.333359
	Abs	0.012841	0.012505	0.013757	0.001639	0.000744	0.004094	0.000359
	SD	0.073517	0.054101	0.035345	0.023343	0.019938	0.015930	0.015046
	MSE	0.005570	0.003083	0.001439	0.000548	0.000398	0.000271	0.000227
1	AEs	1.090683	1.076939	1.022459	1.016086	1.012447	1.006160	1.002920
	Abs	0.090683	0.076939	0.022459	0.016086	0.012447	0.006160	0.002920
	SD	0.378996	0.248425	0.154480	0.108973	0.086592	0.072789	0.063426
	MSE	0.151861	0.067635	0.024369	0.012134	0.007653	0.005336	0.004031
3	AEs	3.483113	3.383034	3.063518	3.050376	3.044667	3.026258	3.020740
	Abs	0.483113	0.383034	0.063518	0.050376	0.044667	0.026258	0.020740
	SD	1.659486	1.242107	0.729983	0.478139	0.398387	0.354286	0.310847
	MSE	2.987291	1.689546	0.536910	0.231155	0.160707	0.126208	0.097056

4. CONCLUSIONS

The Poisson-Bilal distribution arises first by mixing the Poisson and Bilal distributions proposed by [11]. Next, we modify the Poisson-Bilal distribution by applying the zero-probability Poisson-Bilal distribution. That modification is called the zero-truncated method. Therefore, the proposed zero-truncation of the Poisson-Bilal distribution is called the zero-truncated Poisson-Bilal (ZTPB) distribution. Parameter estimation has been discussed using the method of maximum likelihood estimation. We apply the ZTPB distribution to a real data set from various biological science examples, demonstrating its superior goodness of fit compared to zero-truncated Poisson, zero-truncated Poisson-Lindley, zero-truncated Poisson-Sujatha, zero-truncated Poisson-Akash, zero-truncated Poisson-Garima, and zero-truncated Poisson-Ishita distributions. Based on the simulation study, the ML estimator give the estimate value θ is close to the true value.

ACKNOWLEDGMENTS

Author is grateful to the Editor-In-Chief and the anonymous reviewer for fruitful comments on the paper which improved the presentation of the paper.

CONFLICT OF INTERESTS

The author declares that there is no conflict of interests.

REFERENCES

[1] F.N. David, N.L. Johnson, The truncated Poisson, Biometrics. 8 (1952), 275-285. https://www.jstor.org/stable/3001863.

[2] R. Shanker, Zero-truncated Poisson-Akash distribution and its applications, Amer. J. Math. Stat. 7 (2017), 227-236.

[3] R. Shanker, H. Fesshaye, S. Selvaraj, et al. On zero-truncation of Poisson and Poisson-Lindley distributions and their applications, Biom. Biostat. Int. J. 2 (2015), 00045.

[4] A.C. Cohen, Estimating the parameter in a conditional Poisson distribution, Biometrics. 16 (1960), 203-211. https://www.jstor.org/stable/2527552.

[5] J. Singh, A characterization of positive Poisson distribution and its statistical application, SIAM J. Appl. Math. 34 (1978), 545-548. https://doi.org/10.1137/0134043.

[6] M.E. Ghitany, D.K. Al-Mutairi, S. Nadarajah, Zero-truncated Poisson-Lindley distribution and its application,

Math. Computers Simul. 79 (2008), 279-287. https://doi.org/10.1016/j.matcom.2007.11.021.

[7] R. Shanker, F. Hagos, Zero-truncated Poisson-Sujatha distribution with applications, J. Ethiopian Stat. Assoc. 24 (2015), 55-63.

[8] R. Shanker, H. Fesshaye, On zero-truncation of Poisson, Poisson-Lindley and Poisson-Sujatha distributions and their applications, Biom. Biostat. Int. J. 3 (2016), 00072.

[9] R. Shanker, K.K. Shukla, Zero-truncated Poisson-Garima distribution and its applications, Biostat. Biom. Open Access. J. 3 (2017), 555605. https://doi.org/10.19080/bboaj.2017.03.555605.

[10] K.K. Shukla, R. Shanker, M.K. Tiwari, Zero-truncated Poisson-Ishita distribution and its applications, J. Sci. Res. 64 (2020), 287-294. https://doi.org/10.37398/jsr.2020.640240.

[11] E. Altun, A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models, Math. Slovaca. 70 (2020), 979-994. https://doi.org/10.1515/ms-2017-0407.

[12] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. (2023), https://www.R-project.org.

[13] D.J. Finney, G.C. Varley, An example of the truncated Poisson distribution, Biometrics, 11 (1955), 387-394. https://www.jstor.org/stable/3001776.

[14] A.N. Hilma, D. Lestari, S. Devila, Zero-truncated Poisson-Lindley distribution, in: Proceeding of ICSA, pp. 40-50, (2019).

[15] A. Rattanalertnusorn, S. Aryuyuen, The zero-truncated discrete transmuted generalized inverse Weibull distribution and its applications, Songklanakarin J. Sci. Technol. 43 (2021), 1140-1151. https://doi.org/10.14456/SJST-PSU.2021.149.

[16] A. Rattanalertnusorn, S. Aryuyuen, An R implementation for estimating the parameters of the discrete transmuted generalized inverse Weibull distribution, Int. J. Math. Comput. Sci. 17 (2022), 1427-1438.