PROGRESS IN APPLIED SCIENCE AND TECHNOLOGY
Faculty of Science and Technology
Rajamangala University of Technology Thanyaburi
Prog Appl Sci Tech. 2023; 13(2):17-25
ISSN (Print): 2730-3012
ISSN (Online): 2730-3020
https://ph02.tci-thaijo.org/index.php/past

Research Article

Received: April 21, 2023
Revised: June 12, 2023
Accepted: July 04, 2023

DOI: 10.14456/past.2023.10

Bootstrap Methods for Estimating the Confidence Interval for the Index of Dispersion of the Zero-truncated Poisson-Shanker Distribution

Wararit Panichkitkosolkul
Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathumthani 12120, Thailand
E-mail: wararit@mathstat.sci.tu.ac.th

Abstract

The zero-truncated Poisson-Shanker distribution (ZTPS) has been introduced for count data, which is of primary interest in several fields. However, the construction of bootstrap confidence intervals for its index of dispersion (IOD) has not yet been studied. The bootstrap confidence intervals using the percentile, simple, biased-corrected, and accelerated bootstrap methods were proposed in this paper. A Monte Carlo simulation study was conducted to evaluate the performance of three bootstrap confidence intervals based on the coverage probability and average length of the bootstrap confidence intervals. The results indicate that attaining the nominal confidence level using the bootstrap methods was impossible for small sample sizes regardless of the other settings.

Moreover, when the sample size was large, the performances of all methods were not substantially different. The percentile bootstrap and the simple bootstrap methods perform well regarding coverage probability and average length for large sample sizes. However, calculating the percentile bootstrap method is easier than calculating the simple bootstrap method. In the end, real data sets from different fields were analyzed to verify the usefulness of the bootstrap confidence intervals. It is manifested that the results match those from the simulation study.

Keywords: Interval Estimation, Count Data, Shanker Distribution, Bootstrap Interval

1. Introduction

A discrete distribution is a probability distribution that depicts the occurrence of discrete (countable) outcomes (1). One of the discrete distributions is the Poisson distribution which measures the probability of an event happening a certain number of times within a given interval of time or space (2-3). Data such as the number of orders a firm will receive tomorrow, the number of defects in a finished product, the number of customers arriving at a checkout counter in a supermarket from 4 to 6 pm, etc. (4), follow a Poisson distribution.

The probability mass function (pmf) of a Poisson distribution is defined as

$$p(x;\lambda)=\frac{e^{-\lambda}\lambda^{x}}{x!},\ x=0,1,2,...,\ \lambda>0, \qquad (1.1)$$

where e is a constant approximately equal to 2.71828 and λ is the mean number of events within a given interval of time or space.

This probability model can analyze data containing zeros and positive values with low occurrence probabilities within a predefined time or area range (5). However, probability models can become truncated when a range of possible variable values is disregarded or impossible to observe. Indeed, zero truncation is often enforced when one wants to analyze the count data without zeros. Reference (6) developed the zero-truncated (ZT) Poisson (ZTP) distribution, which has been applied to datasets of the length of stay in the hospital, the number of fertile mothers who have experienced at least one child death, the number of children ever born to a sample of mothers over 40 years old, and the number of passengers in cars (7). Let be a ZT distribution, and then the pmf can be derived as

$$p(x;\theta)=\frac{p_0(x;\theta)}{1-p_0(0;\theta)},\ x=1,2,3,..., \tag{1.2}$$

where $p_0(x;\theta)$ and $p_0(0;\theta)$ are the pmf of the un-truncated distribution for any value of x and $x=0$, respectively. Reference (8) defined the pmf of the Poisson-Shanker (PS) distribution having as

$$p_0(x;\theta)=\frac{\theta^2}{\theta^2+1}\frac{x+(\theta^2+\theta+1)}{(\theta+1)^{x+2}},$$
$$x=0,1,2,...,\theta>0. \tag{1.3}$$

The mathematical and statistical properties of the PS distribution for modeling count data were established by (8). The PS distribution arises from the Poisson distribution when parameter λ follows the Shanker distribution proposed by (9) with probability density function (pdf)

$$f(\lambda;\theta)=\frac{\theta^2}{\theta^2+1}(\theta+\lambda)e^{-\theta\lambda},\lambda>0,\theta>0. \tag{1.4}$$

Reference (9) showed that the pdf in Eq. (1.4) is a better model than the exponential and Lindley (10) distributions for modeling lifetime data. Several distributions have been introduced as an alternative to the ZTP distribution for handling over-dispersion in data, such as ZT Poisson-Lindley (ZTPL) (11), ZT Poisson-Amarendra (ZTPA) (12), ZT Poisson-Akash (13) and ZT Poisson-Ishita (14) distributions.

Reference (8) proposed the ZTPS distribution and its properties, such as the moment, coefficient of variation, skewness, kurtosis, and the index of dispersion (IOD). The method of moments and the maximum likelihood have also been derived for estimating its parameter. Furthermore, when the ZTPS distribution was applied to real data set, it was more suitable than the ZTP and ZTPL distributions.

The IOD, like the coefficient of variation, is a normalized measure of the dispersion of a probability distribution. It is a measure used to quantify whether a set of observed occurrences are clustered or dispersed compared to a standard statistical model. It is defined as the ratio of the population variance σ^2 to the population mean μ; σ^2/μ. This index should typically only be used for data measured on a ratio scale. It is sometimes used for count data. Therefore, this measure can be used to assess whether observed count data can be modeled using a Poisson distribution. When the IOD is less than one, a dataset is said to be under-dispersed. On the other hand, if the IOD is larger than one, a dataset is said to be over-dispersed (15). Some distributions, most notably the Poisson distribution, have IOD = 1. The geometric and negative binomial distributions have IOD > 1, while the binomial distribution has IOD < 1, and the constant random variable has IOD = 0 (15).

To the best of our knowledge, no research has been conducted on estimating the confidence interval for the IOD of the ZTPS distribution. Bootstrap methods for estimating the confidence interval provide a way of quantifying the uncertainties in statistical inferences based on a sample of data. The concept is to run a simulation study based on the actual data to estimate the likely extent of sampling error (16). Therefore, the objective of the current study is to assess the efficiencies of three bootstrap methods, namely the percentile bootstrap (PB), the simple bootstrap (SB), and the bias-corrected and accelerated bootstrap (BCa), to estimate the confidence interval for the IOD of the ZTPS distribution. Because a theoretical comparison is impossible, we conducted a simulation study to compare their performance. We used the results to determine the best-performing method based on the coverage probability and the average length.

2. Theoretical Background

Compounding probability distributions is a sound and innovative technique to obtain new probability distributions to fit data sets not adequately fit by common parametric distributions. Reference (8) proposed a new compounding distribution by compounding the Poisson distribution with the Shanker distribution, as a more flexible model for analyzing statistical data is needed. The pmf of the PS distribution is given in Eq. (1.3).

Let X be a random variable which follow the ZTPS distribution with parameter θ, it is denoted as $X \sim \text{ZTPS}(\theta)$. Using Eqs. (1.2) and (1.3), the pmf of the ZTPS distribution can be obtained as

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

The plots of the pmf of the ZTPS distribution with some specified parameter values θ are shown in Figure 1.

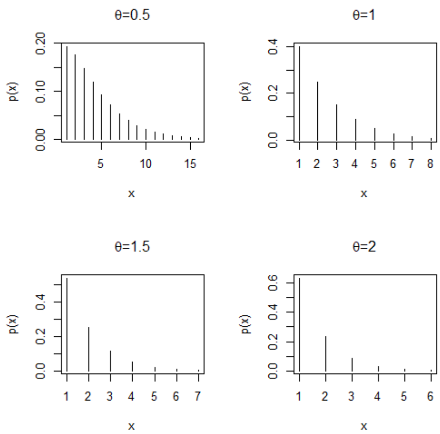

Figure 1 The plots of the pmf of the ZTPS distribution with θ =0.5, 1, 1.5 and 2.

The expected value, variance of X and the IOD are as follows:

$$E(X) = \mu = \frac{(\theta+1)\left(\theta^3+\theta^2+2\theta+2\right)}{\theta(\theta^3+\theta^2+2\theta+1)},$$

$$\operatorname{var}(X) = \sigma^2 = \frac{(\theta+1)\begin{pmatrix}\theta^6+2\theta^5+6\theta^4+9\theta^3\\ +10\theta^2+8\theta+2\end{pmatrix}}{\theta^2(\theta^3+\theta^2+2\theta+1)^2},$$

and

$$\kappa = \frac{\theta^6+2\theta^5+6\theta^4+9\theta^3+10\theta^2+8\theta+2}{\theta\left(\theta^6+2\theta^5+5\theta^4+7\theta^3+7\theta^2+6\theta+2\right)}. \quad (2.1)$$

The point estimator of θ is obtained by maximizing the log-likelihood function $\log L(x_i;\theta)$ or the logarithm of joint pmf of $X_1,\ldots,X_n$. Therefore, the maximum likelihood (ML) estimator for θ of the ZTPS distribution is derived by the following processes:

$$\frac{\partial}{\partial\theta}\log L(x_i;\theta) = \frac{\partial}{\partial\theta}\left[\begin{array}{l} n\log\left(\dfrac{\theta^2}{\theta^3+\theta^2+2\theta+1}\right) \\ -\sum_{i=1}^{n} x_i\log(\theta+1) \\ +\sum_{i=1}^{n}\log\left[x_i+(\theta^2+\theta+1)\right]\end{array}\right]$$
$$= \frac{2n}{\theta} - \frac{n\left(3\theta^2+2\theta+2\right)}{\left(\theta^3+\theta^2+2\theta+1\right)} - \frac{n\bar{x}}{\theta+1}$$
$$+\sum_{i=1}^{n}\frac{2\theta+1}{x_i+(\theta^2+\theta+1)}.$$

Solving the equation $\frac{\partial}{\partial\theta}\log L(x_i;\theta) = 0$ for θ, we have the non-linear equation

$$\frac{2n}{\theta} - \frac{n\left(3\theta^2+2\theta+2\right)}{\left(\theta^3+\theta^2+2\theta+1\right)} - \frac{n\bar{x}}{\theta+1}$$
$$+\sum_{i=1}^{n}\frac{2\theta+1}{x_i+(\theta^2+\theta+1)} = 0,$$

where $\bar{x} = \sum_{i=1}^{n} x_i / n$ denotes the sample mean. Since the ML estimator for θ does not provide the closed-form solution, the non-linear equation can be solved by the numerical iteration methods such as Newton-Raphson, bisection, and Ragula-Falsi methods. In this research, we use maxLik package (18) with Newton-Raphson method for ML estimation in the statistical software R.

The point estimator of the IOD can be estimated by replacing the parameter θ with the ML estimator for θ shown in Eq. (2.1). Therefore, the point estimator of the IOD ($\hat{\kappa}$) is given by

$$\hat{\kappa} = \frac{\hat{\theta}^6+2\hat{\theta}^5+6\hat{\theta}^4+9\hat{\theta}^3+10\hat{\theta}^2+8\hat{\theta}+2}{\hat{\theta}\left(\hat{\theta}^6+2\hat{\theta}^5+5\hat{\theta}^4+7\hat{\theta}^3+7\hat{\theta}^2+6\hat{\theta}+2\right)},$$

where $\hat{\theta}$ is the ML estimator for θ.

3. Bootstrap Methods

In this paper, we focus on the three bootstrap methods for estimating the confidence interval for the IOD of the ZTPS distribution. In practice, the popular bootstrap methods are the percentile, the simple and the bias-corrected and accelerated bootstrap methods. See the details of some bootstrap methods in (19-20).

3.1 Percentile Bootstrap (PB) Method

The percentile bootstrap confidence interval is the interval between the $(\alpha/2)\times 100$ and $(1-(\alpha/2))\times 100$ percentiles of the distribution of κ estimates obtained from resampling or the distribution of $\hat{\kappa}^*$, where κ represents a parameter of interest and α is the level of significance (e.g., α = 0.05 for 95% confidence intervals) (19, 21). A percentile bootstrap confidence interval for κ can be obtained as follows:

1) B random bootstrap samples are generated,

2) a parameter estimate $\hat{\kappa}^*$ is calculated from each bootstrap sample,

3) all B bootstrap parameter estimates are ordered from the lowest to highest, and

4) the $(1-\alpha)100\%$ percentile bootstrap confidence interval is given by

$$CI_{PB}=\left[\hat{\kappa}^*_{(r)},\hat{\kappa}^*_{(s)}\right], \tag{3.1}$$

where $\hat{\kappa}^*_{(r)}$ represents the r^{th} quantile of the set of ordered quantiles from lowest to highest, $\hat{\kappa}^*_{(s)}$ represents the s^{th} quantile of the same set, $r=\lceil(\alpha/2)B\rceil$, $s=\lceil(1-(\alpha/2))B\rceil$, where $\lceil x\rceil$ is the ceiling function of x, and α is the significance level. This study use B = 1,000 and α = 0.05; the quantile corresponding to the lower limit of confidence interval was $\hat{\kappa}^*_{(r)}=\hat{\kappa}^*_{(25)}$ (the 25th quantile) and that corresponding to the upper limit was $\hat{\kappa}^*_{(s)}=\hat{\kappa}^*_{(1975)}$ (the 1975th quantile).

3.2 Simple Bootstrap (SB) Method

The simple bootstrap method is sometimes called the basic bootstrap method and is a method as easy to apply as the percentile bootstrap method. Suppose that the quantity of interest is κ and that the estimator of κ is $\hat{\kappa}$. The simple bootstrap method assumes that the distributions of $\hat{\kappa}-\kappa$ and $\hat{\kappa}^*-\hat{\kappa}$ are approximately the same (22). The $(1-\alpha)100\%$ simple bootstrap confidence interval for κ is

$$CI_{SB}=\left[2\hat{\kappa}-\hat{\kappa}^*_{(s)},2\hat{\kappa}-\hat{\kappa}^*_{(r)}\right], \tag{3.2}$$

where the quantiles $\hat{\kappa}^*_{(r)}$ and $\hat{\kappa}^*_{(s)}$ are the same percentile of empirical distribution of bootstrap estimates $\hat{\kappa}^*$ used in Eq. (3.1) for the percentile bootstrap method.

3.3 Bias-Corrected and Accelerated Bootstrap (BCa) Method

To overcome the over coverage issues in percentile bootstrap confidence intervals, the BCa bootstrap method corrects for both bias and skewness of the bootstrap parameter estimates by incorporating a bias-correction factor and an acceleration factor (19, 21). The bias-correction factor $\hat{z}_0$ is estimated as the proportion of the bootstrap estimates less than the original estimate value $\hat{\kappa}$, $\hat{z}_0=\Phi^{-1}\left(\#\{\hat{\kappa}^*\le\hat{\kappa}\}/B\right)$, where Φ^{-1} is the inverse function of a standard normal cumulative distribution function (e.g., $\Phi^{-1}(0.975)\approx 1.96$) and $\#\{\hat{\kappa}^*\le\hat{\kappa}\}$ represent the number of times that $\hat{\kappa}^*$ is less than $\hat{\kappa}$ in each replication. The acceleration factor $\hat{a}$ is estimated through jackknife resampling (i.e., “leave one out” resampling), which involves generating n replicates of the original sample, where n is the number of observations in the sample. The first jackknife replicate is obtained by leaving out the first case $(i=1)$ of the original sample, the second by leaving out the second case $(i=2)$, and so on, until n samples of size $n-1$ are obtained. For each of the jackknife resamples, $\hat{\kappa}_{(-i)}$ is obtained. The average of these estimates is $\hat{\kappa}_{(\cdot)}=\sum_{i=1}^{n}\hat{\kappa}_{(-i)}/n$. Then, the acceleration factor $\hat{a}$ is calculated as follow,

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\kappa}_{(\cdot)} - \hat{\kappa}_{(-i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\kappa}_{(\cdot)} - \hat{\kappa}_{(-i)}\right)^2\right\}^{3/2}}.$$

The values of $\hat{z}_0$ and $\hat{a}$, the values α_1 and α_2 are calculated,

$$\alpha_1 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1-\hat{a}\left(\hat{z}_0 + z_{\alpha/2}\right)}\right\} \text{ and}$$

$$\alpha_2 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1-\hat{a}\left(\hat{z}_0 + z_{1-\alpha/2}\right)}\right\},$$

where $z_{\alpha/2}$ is the α quantile of the standard normal distribution (e.g. $z_{0.05/2} \approx -1.96$). Then, the $(1-\alpha)100\%$ BCa bootstrap confidence interval for κ is as follows

$$CI_{BCa} = \left[\hat{\kappa}^*_{(j)}, \hat{\kappa}^*_{(k)}\right], \quad (3.3)$$

where $j = \lceil \alpha_1 B \rceil$ and $k = \lceil \alpha_2 B \rceil$.

4. Simulation Study

The confidence interval for the IOD of a ZTPS distribution estimated via various bootstrap methods was considered in this study. Because a theoretical comparison is not possible, a Monte Carlo simulation study was designed using R version 4.2.2 (23) to cover cases with different sample sizes (n = 10, 30, 50, 100 and 500). To cover over-dispersion and under-dispersion cases, the true parameter (θ) was set as 0.25, 0.5, 1, 1.5 and 2, then the index of dispersion (κ) were 4.7105, 2.5146, 1.2667, 0.8072 and 0.5784, respectively. B =1000 bootstrap samples of size n were generated from the original sample and each simulation was repeated 1,000 times. This paper used the acceptance-rejection method (24) to generate random variate with PA distribution. Without loss of generality, the nominal confidence level $(1-\alpha)$ was set at 0.95. The performances of the bootstrap methods were compared in terms of their coverage probabilities and average lengths. The one with a coverage probability greater than or close to the nominal confidence level means that it contains the true value and can be used to precisely estimate the confidence interval for the index of dispersion.

Table 1 Coverage probability and average length of the 95% confidence intervals for the IOD of the ZTPS distribution.

n	θ	κ	Coverage probability			Average length		
			PB	SB	BCa	PB	SB	BCa
10	0.25	4.7105	0.868	0.860	0.880	3.6878	3.6845	3.8709
	0.5	2.5146	0.885	0.891	0.899	2.2686	2.2678	2.3697
	1	1.2667	0.866	0.839	0.893	1.4459	1.4462	1.5207
	1.5	0.8072	0.861	0.831	0.901	1.1167	1.1191	1.2086
	2	0.5784	0.881	0.861	0.928	0.9119	0.9127	1.0152
30	0.25	4.7105	0.927	0.926	0.925	2.2936	2.2884	2.3485
	0.5	2.5146	0.922	0.923	0.929	1.3644	1.3646	1.3943
	1	1.2667	0.935	0.926	0.940	0.9188	0.9200	0.9421
	1.5	0.8072	0.908	0.899	0.933	0.7185	0.7202	0.7462
	2	0.5784	0.924	0.909	0.939	0.5962	0.5963	0.6255
50	0.25	4.7105	0.939	0.932	0.934	1.7956	1.7948	1.8229
	0.5	2.5146	0.934	0.930	0.943	1.0749	1.0767	1.0920
	1	1.2667	0.938	0.934	0.936	0.7232	0.7238	0.7351
	1.5	0.8072	0.928	0.911	0.935	0.5733	0.5728	0.5857
	2	0.5784	0.920	0.903	0.940	0.4777	0.4770	0.4919
100	0.25	4.7105	0.945	0.941	0.941	1.2954	1.3005	1.3104
	0.5	2.5146	0.948	0.948	0.944	0.7731	0.7721	0.7785
	1	1.2667	0.935	0.937	0.940	0.5148	0.5153	0.5198
	1.5	0.8072	0.941	0.937	0.941	0.4143	0.4147	0.4198
	2	0.5784	0.944	0.923	0.947	0.3451	0.3443	0.3504

n	θ	κ	Coverage probability			Average length		
			PB	SB	BCa	PB	SB	BCa
500	0.25	4.7105	0.937	0.932	0.935	0.5795	0.5786	0.5790
	0.5	2.5146	0.948	0.951	0.951	0.3489	0.3489	0.3499
	1	1.2667	0.954	0.953	0.957	0.2325	0.2323	0.2329
	1.5	0.8072	0.939	0.940	0.942	0.1856	0.1853	0.1859
	2	0.5784	0.937	0.932	0.939	0.1555	0.1559	0.1562

The results of the study are reported in Table 1. For n = 10, 30 and 50, the coverage probabilities of the three methods tended to be less than 0.95 and so did not reach the nominal confidence level. All bootstrap methods had coverage probabilities close to the nominal confidence levels for large sample sizes $(n \geq 100)$. Additionally, the coverage probabilities of all bootstrap methods were not significantly different for these situations by testing with the Kruskal-Wallis test (25) (H = 1.559, p-value = 0.459). Thus, as the sample size was increased, the coverage probabilities of the methods tended to increase and approach 0.95.

Moreover, the average length of the methods decreased when the value of κ was decreased because of the relationship between the variance and κ (see Figure 2). As the sample size was increased, the average lengths decreased. For small sample sizes $(n \leq 50)$, the average lengths of the PB and SB methods were shorter than those of BCa method. For large sample sizes $(n \geq 100)$, the average lengths of all bootstrap methods were not significantly different by testing with the Kruskal-Wallis test (25) (H = 0.237, p-value = 0.888).

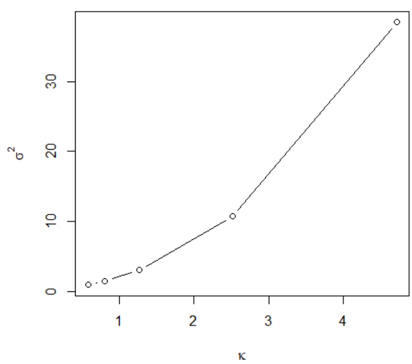

Figure 2 The relationship between the value of κ and variance.

5. Numerical Examples

We used three real-world examples to demonstrate the applicability of the bootstrap methods for estimating the confidence interval for the IOD of the ZTPS distribution.

5.1 The Unrest Events Example

The number of unrest events occurring in the southern border area of Thailand from July 2020 to October 2022 collected by the Southern Border Area News Summarises was used for this example (the total sample size is 28). The number of unrest events per month during this time period in the five southern provinces of Pattani, Yala, Narathiwat, Songkhla, and Satun is reported in Table 2. For the Chi-squared goodness-of-fit test (26), the Chi-squared statistic was 3.9112 and the p-value was 0.3113. Thus, a ZTPS distribution with $\hat{\theta} = 0.3058$ is suitable for this dataset. The estimator of the index of dispersion was 3.9311. Table 3 reported the 95% confidence intervals for the index of dispersion of the ZTPS distribution. The estimated parameter $\hat{\theta}$ is near to 0.5. The results correspond with the simulation results for $n = 30$ because the average lengths of the PB and SB methods were shorter than those of the BCa method.

Table 2 The number of unrest events in the southern border area of Thailand.

Number of Unrest Events	0	1	2	3	4	5	6	7	≥ 8
Observed Frequency	0	3	1	3	2	3	3	3	8
Expected Frequency	-	1.778	2.415	2.817	2.968	2.908	2.699	2.401	10.014

Table 3 The 95% confidence intervals and corresponding widths using all intervals for the index of dispersion in the unrest events example.

Methods	Confidence Intervals	Widths
PB	(3.0907, 4.7974)	1.7067
SB	(3.0430, 4.7706)	1.7276
BCa	(3.0870, 4.8198)	1.7328

5.2 Demographic Example

Table 4 shows the demographic data on the number of fertile mothers who have experienced at least one child death (27). The total sample size is 135. For Chi-squared goodness-of-fit test, the Chi-squared statistic was 3.1070 and the p-value was 0.2115. Thus, the ZTPS distribution with $\hat{\theta} =$ 1.9538 is suitable for this dataset. The point estimator of the index of dispersion is 0.5943. The 95% confidence intervals for the index of dispersion of the ZTPS distribution are reported in Table 5. The results correspond with the simulation results for $\kappa = 0.5784$ and $n = 100$ because the average lengths of the PB and SB methods were shorter than those of the BCa method.

Table 4 The number of fertile mothers who have experienced at least one child death.

Number of Child Deaths	0	1	2	3	≥ 4
Observed Frequency	0	89	25	11	10
Expected Frequency	-	83.7792	32.0125	12.0731	7.1352

Table 5 The 95% confidence intervals and corresponding widths using all intervals for the index of dispersion in the demographic example.

Methods	Confidence Intervals	Widths
PB	(0.4322, 0.7626)	0.3187
SB	(0.4272, 0.7551)	0.3279
BCa	(0.4414, 0.7723)	0.3309

5.3 Flower Heads Example

The third dataset, shown in Table 6, is the number of flower heads per the number of fly eggs reported by (28); the total sample size is 88. For the Chi-squared goodness-of-fit test, the Chi-squared statistic was 3.3591 and the p-value was 0.4996. Thus, a ZTPS distribution with $\hat{\theta} = 0.7335$ is suitable for this dataset. The point estimator of the index of dispersion is 1.7425. The 95% confidence intervals for the index of dispersion of the ZTPS distribution are reported in Table 7. The results correspond with the simulation results for $\kappa =$ 1.2667 and $n =$ 100 because the average lengths of the PB and SB methods were shorter than those of the BCa method.

Table 6 The number of flower heads as per the number of fly eggs.

Number of Fly Eggs	0	1	2	3	4	5	≥ 6
Observed Frequency	0	22	18	18	11	9	10
Expected Frequency	-	26.2820	19.7959	14.0933	9.6723	6.4694	11.6870

Table 7 The 95% confidence intervals and corresponding widths using all intervals for the index of dispersion in the flower heads example.

Methods	Confidence Intervals	Widths
PB	(1.4867, 2.0077)	0.5210
SB	(1.4832, 2.0072)	0.5240
BCa	(1.4903, 2.0176)	0.5273

6. Conclusions and Discussion

Three bootstrap confidence intervals, namely PB, SB, and BCa methods, of the IOD of the ZTPS distribution have been introduced in this study. Based on the simulation study, when the sample sizes were 10, 30, and 50, the coverage probabilities of all three were substantially lower than 0.95. When the sample size was large enough (i.e., $n \geq 100$), the coverage probabilities and average lengths using three bootstrap methods were not markedly different. According to our findings, the PB and SB methods provided the shortest average length for small sample sizes and parameter settings tested in the simulation study. The usefulness of the bootstrap confidence intervals was illustrated empirically using three applications to the number of unrest events in the southern border area of Thailand, the number of fertile mothers who have experienced at least one child death, and the number of flower heads as per the number of fly eggs. Therefore, the PB and SB methods are recommended to estimate the confidence interval for the IOD of the ZTPS distribution.

The limitation of the current study is that none of the bootstrap confidence intervals were exact. However, they would be consistent, meaning that the coverage probability approaches the nominal confidence level as the sample sizes get large. In addition, three bootstrap confidence intervals are computer intensive and not easy to compute. However, there are numerous available packages in R for computing the bootstrap confidence intervals, such as boot package (29), bootstrap package (30), semEff package (31), and BootES package (33). Since R is open-source, users are free to download these packages. Future research could focus on hypothesis testing for the IOD and the other confidence intervals to compare with the proposed bootstrap methods.

Acknowledgements

The author is thankful to the referees and the editor-in-chief for providing the useful comments which improved the earlier draft of the paper.

Declaration of conflicting interests

The authors declared that they have no conflicts of interest in the research, authorship, and this article's publication.

References

1. Zhang Z, Shafer D. Introductory statistics: Flatworld Knowledge; 2014.
2. Kissell R, Poserina J. Optimal sports math, statistics, and fantasy: Academic Press; 2017.
3. Andrew FS, Michael, RW. Practical business statistics: Academic Press; 2022.
4. Siegel AF. Practical business statistics: Academic Press; 2016.
5. Sangnawakij P. Confidence interval for the parameter of the zero-truncated Poisson distribution. J Appl Sci. 2021; 20(2):13-22.
6. David F, Johnson N. The truncated Poisson. Biometrics. 1952;8(4):275-85.
7. Hussain T. A zero truncated discrete distribution: Theory and applications to count data. Pak J Stat Oper Res. 2020;16(1):167-90.
8. Shanker R. The discrete Poisson-Shanker distribution. Jacobs J Biostat. 2016;1(1): 1-7.
9. Shanker R. Shanker distribution and its applications. Int J Stat Appl. 2015;5(6): 338-48;
10. Lindley DV. Fiducial distributions and Bayes' theorem. J R Stat Soc Ser B. 1958;20(1):102-7.
11. Ghitany ME, Al-Mutairi DK, Nadarajah S. Zero-truncated Poisson-Lindley distribution and its application. Math Comput Simul. 2008;79(3):279-87.

12. Shanker R. A zero-truncated Poisson-Amarendra distribution and its application. Int J Prob Stat. 2017a; 6(4):82-92.
13. Shanker R. Zero-truncated Poisson-Akash distribution and its applications. Am J Math Stat. 2017b;7(6):227-36.
14. Shukla KK, Shanker R, Tiwari MK. Zero-truncated Poisson-Ishita distribution and its application. J Sci Res. 2020;64(2):287-94.
15. Wikiwand, Index of dispersion. Available from: https://www.wikiwand.com/en/Index_of_dispersion (Accessed on 10 June 2023).
16. Wood M. Statistical inference using bootstrap confidence intervals. Significance. 2004;1(4):180-2.
17. Shanker R. A zero-truncated Poisson-Shanker distribution and its applications. Int J Stat Appl. 2017c;7(3):159-69.
18. Henningsen A, Toomet O. maxLik: A package for maximum likelihood estimation in R. Comput Stat. 2011; 26(3):443-58.
19. DiCiccio TJ, Efron B. Bootstrap confidence intervals. Stat Sci. 1996; 11(3):189-212.
20. Manoharan T, Arasan J, Midi H, Adam MB. Bootstrap intervals in the presence of left-truncation, censoring and covariates with a parametric distribution. Sains Malaysiana. 2017;46(12):2529-39.
21. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman and Hall; 1993.
22. Efron B. Better bootstrap confidence intervals. J Am Stat Assoc. 1987; 82(397):171-85.
23. Meeker WQ, Hahn GJ, Escobar LA. Statistical intervals: a guide for practitioners and researchers. New York: John Wiley and Sons; 2017.
24. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat. 1996;5(3):299-314.
25. Robert CP, Casella G. Monte Carlo statistical methods, New York: Springer; 2004.
26. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47(260):583-621.
27. Turhan NS. Karl Pearson's chi-square tests. Educ Res Rev. 2020;15(9):575-80.
28. Shanker R. Hagos F, Selvaraj, S. Yemane A. On zero-truncation of Poisson and Poisson-Lindley distributions and their application. Biom Biostat Int J. 2015; 2(6):168-81.
29. Finney DJ, Varley GC. An example of the truncated Poisson distribution, Biometrics. 1955;11(3):387-94.
30. Canty A, Ripley B. boot: bootstrap R (S-Plus) functions. R package version 1.3-28.1, 2022.
31. Kostyshak S. bootstrap: functions for the book "An introduction to the bootstrap". R package version, 2019.6, 2022.
32. Murphy MV. semEff: automatic calculation of effects for piecewise structural equation models. R package version 0.6.1, 2022.
33. Kirby KN, Gerlanc, D. BootES: an R package for bootstrap confidence intervals on effect sizes. Behav Res Methods. 2013;45(4):905-27.