Interdisciplinary Research Review
Volume 18, Number 4, Pages 13 – 22

Bootstrap Methods for Estimating the Confidence Interval for the Index of Dispersion of the Zero-Truncated Poisson-Amarendra Distribution

Wararit Panichkitkosolkul[1]*
[1]Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathumthani 12120, Thailand

Abstract

The zero-truncated count data is of primary interest in several areas such as biological science, medical science, demography, ecology, etc. Recently, the zero-truncated Poisson-Amarendra distribution has been proposed for such data. However, the confidence interval estimation of the index of dispersion has not yet been examined. This paper examined confidence interval estimation based on percentile, simple, biased-corrected, and accelerated bootstrap methods in terms of coverage probability and average interval length via Monte Carlo simulation. The results indicate that attaining the nominal confidence level using the bootstrap methods was not possible for small sample sizes regardless of the other settings. Moreover, when the sample size was large, the performances of the methods were not substantially different. However, the percentile bootstrap and the simple bootstrap methods provided the shortest average lengths for small sample sizes. Last, the bootstrap methods were used to calculate the confidence intervals for the index of dispersion of the zero-truncated Poisson-Amarendra distribution via two numerical examples, the results of which match those from the simulation study.

Keywords: interval estimation, count data, Amarendra distribution, Bootstrap interval, simulation

Article history: Received 29 January 2023, Revised 01 April 2023, Accepted 27 April 2023

1. Introduction

The Poisson distribution is a discrete probability distribution that measures the probability of an event occurring a certain number of times within a given interval of time or space [1-2]. 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 9 to 11 AM., the number of births per day, etc. [3], follow a Poisson distribution.

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

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

where e is a constant approximately equal to 2.71828 and λ is the shape parameter which indicates the mean number of events within a given interval of time or space. This probability model can be used to analyze data containing zeros and positive values that have low occurrence probabilities within a predefined time or area range [4]. However, probability models can become truncated when a range of possible values for the variables is either disregarded or impossible to observe. Indeed, zero truncation is often enforced when one wants to analyze count data without zeros. David and Johnson [5] developed the zero-truncated (ZT) Poisson (ZTP) distribution, which has been applied to datasets of the length of stay in hospitals, 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 [6]. A ZT distribution's pmf can be derived as

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

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. Shanker [7] defined the pmf of the Poisson-Amarendra (PA) distribution as

*Corresponding author; e-mail: wararit@mathstat.sci.tu.ac.th

$$p_0(x;\theta)=\frac{\theta^4}{\left(\theta^3+\theta^2+2\theta+6\right)}\frac{\left(x^3+(\theta+7)x^2+\left(\theta^2+5\theta+15\right)x+(\theta^3+4\theta^2+7\theta+10)\right)}{(\theta+1)^{x+4}},\ x=0,1,2,...,\theta>0. \quad (2)$$

The mathematical and statistical properties of the PA distribution for modeling biological science data were established by Shanker [7]. The PA distribution arises from the Poisson distribution when parameter λ follows the Amarendra distribution proposed by Shanker [8] with probability density function (pdf)

$$f(\lambda;\theta)=\frac{\theta^4}{\theta^3+\theta^2+2\theta+6}\left(1+\lambda+\lambda^2+\lambda^3\right)e^{-\theta\lambda},\ \lambda>0,\theta>0. \quad (3)$$

Shanker [8] showed that the pdf in (3) is a more suitable model than the exponential, Lindley [9] and Sujatha distributions [10] for modeling lifetime data from biomedical sciences and engineering. 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-Sujatha (ZTPS) [12] and ZT Poisson-Akash (ZTPAk) distributions [13].

Shanker [14] proposed the ZT Poisson-Amarendra (ZTPA) distribution and its properties, such as the moment, coefficient of variation, skewness, kurtosis, and the index of dispersion. The method of moments and the maximum likelihood have also been derived for estimating its parameter. Furthermore, when the ZTPA distribution was applied to real data, it was more suitable than the ZTP, ZTPL, and ZTPS distributions.

The index of dispersion [15], 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. If the count data follows a Poisson distribution, then the mean and variance should be equal and the index of dispersion is 1. If the counts follow a geometric or negative binomial, then the index of dispersion should be greater than 1. If the counts follow a binomial distribution, the index of dispersion should be less than 1 [16].

The relevance of the index of dispersion is that it has a value of 1 when the probability distribution of the number of occurrences in an interval is a Poisson distribution. Thus, the measure can be used to assess whether observed data can be modeled using a Poisson process. When the index of dispersion is less than 1, a dataset is said to be under-dispersed, this condition can relate to patterns of occurrence that are more regular than the randomness associated with a Poisson process. For example, regular, periodic events will be under-dispersed. If the index of dispersion is larger than 1, a dataset is said to be over-dispersed [16].

To the best of our knowledge, no research has been conducted on estimating the confidence interval for the index of dispersion of the ZTPA distribution. It is essential to note that the score function of ZTPA distribution is complicated, and the maximum likelihood estimator has no closed form. Therefore, likelihood-based, score, and Wald-type confidence intervals have no closed forms. In such cases, finding these confidence intervals can be challenging; alternative methods, such as numerical techniques or resampling methods like the bootstrap method, can be utilized. Bootstrap methods for estimating confidence intervals provide a way of quantifying the uncertainties in statistical inference based on a sample of data. The concept is to run a simulation study based on the actual data for estimating the likely extent of sampling error [17]. 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) methods, to estimate the confidence interval for the index of dispersion of the ZTPA distribution. In addition, none of the bootstrap confidence intervals will be exact (i.e., the actual confidence level is exactly equal to the nominal confidence level $1-\alpha$) but they will all be consistent, meaning that the confidence level approaches $1-\alpha$ as the sample size gets large [18]. In light of the impossibility of a theoretical comparison of these bootstrap confidence intervals, we conduct a simulated study to evaluate their relative merits.

2. Theoretical Background

To obtain novel probability distributions, compounding of probability distributions is an innovative approach to fit data sets inadequately fit by common distributions. As there is a need to find more suitable model for analyzing statistical data, Shanker [7] proposed a new compounding distribution by compounding Poisson distribution with Amarendra distribution [8]. The pmf of the PA distribution is given in Eq. (3).

Let X be a random variable which follows the ZTPA distribution with parameter $\theta,$ it is denoted as $X \sim \text{ZTPA}(\theta).$ Using Eqs. (2) and (3), the pmf of ZTPA distribution can be obtained as

$$p(x;\theta)=\frac{\theta^4}{\begin{pmatrix}\theta^6+5\theta^5+14\theta^4+41\theta^3+\\ 45\theta^2+26\theta+6\end{pmatrix}}\frac{\begin{pmatrix}x^3+(\theta+7)x^2+\left(\theta^2+5\theta+15\right)x+\\ (\theta^3+4\theta^2+7\theta+10)\end{pmatrix}}{(\theta+1)^x},\ x=1,2,3\ldots,\ \theta>0.$$

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

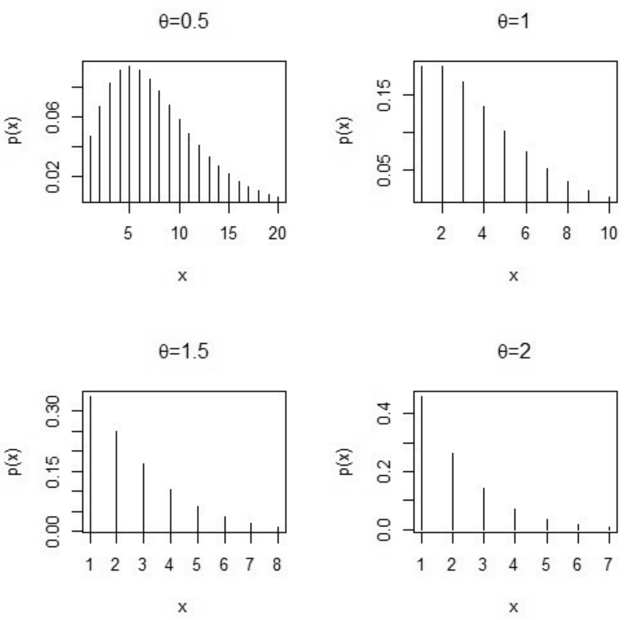

Figure 1. The plots of the pmf of the ZTPA distribution with $\theta = 0.5,\ 1,\ 1.5$ and 2.

The expected value, variance and index of dispersion of X are as follows:

$$E(X)=\frac{\theta^7+6\theta^6+20\theta^5+64\theta^4+141\theta^3+170\theta^2+102\theta+24}{\theta\left(\theta^6+5\theta^5+14\theta^4+41\theta^3+45\theta^2+26\theta+6\right)},$$

$$\operatorname{var}(X)=\frac{\begin{pmatrix}\theta^{13}+12\theta^{12}+79\theta^{11}+420\theta^{10}+1749\theta^9+5486\theta^8+13461\theta^7+24780\theta^6+\\ 31990\theta^5+27898\theta^4+16176\theta^3+6108\theta^2+1392\theta+144\end{pmatrix}}{\theta^2\left(\theta^6+5\theta^5+14\theta^4+41\theta^3+45\theta^2+26\theta+6\right)^2},$$

and

$$ID(X)=\kappa=\frac{\begin{pmatrix}\theta^{13}+12\theta^{12}+79\theta^{11}+420\theta^{10}+1749\theta^{9}+5486\theta^{8}+13461\theta^{7}+24780\theta^{6}+\\ 31990\theta^{5}+27898\theta^{4}+16176\theta^{3}+6108\theta^{2}+1392\theta+144\end{pmatrix}}{\theta\begin{pmatrix}\theta^{6}+5\theta^{5}+14\theta^{4}+41\theta^{3}+\\ 45\theta^{2}+26\theta+6\end{pmatrix}\begin{pmatrix}\theta^{7}+6\theta^{6}+20\theta^{5}+64\theta^{4}+141\theta^{3}+\\ 170\theta^{2}+102\theta+24\end{pmatrix}}. \quad (4)$$

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 $X_1,...,X_n$. Thus, the maximum likelihood (ML) estimator for θ of the ZTPA distribution is derived by the following processes:

$$\frac{\partial}{\partial\theta}\log L(x_i;\theta)=\frac{\partial}{\partial\theta}\begin{bmatrix}n\log\left(\dfrac{\theta^4}{\theta^6+5\theta^5+14\theta^4+41\theta^3+45\theta^2+26\theta+6}\right)-\sum_{i=1}^{n}x_i\log(\theta+1)\\ +\sum_{i=1}^{n}\log\left[x_i^3+(\theta+7)x_i^2+(\theta^2+5\theta+15)x_i+\left(\theta^3+4\theta^2+7\theta+10\right)\right]\end{bmatrix}$$

$$=\frac{4n}{\theta}-\frac{n\begin{pmatrix}6\theta^5+25\theta^4+56\theta^3+\\ 123\theta^2+90\theta+26\end{pmatrix}}{\begin{pmatrix}\theta^6+5\theta^5+14\theta^4+41\theta^3\\ +45\theta^2+26\theta+6\end{pmatrix}}-\frac{n\bar{x}}{\theta+1}+\sum_{i=1}^{n}\frac{x_i^2+(2\theta+5)x_i+\left(3\theta^2+8\theta+7\right)}{\begin{pmatrix}x_i^3+(\theta+7)x_i^2+\left(\theta^2+5\theta+15\right)x_i\\ +(\theta^3+4\theta^2+7\theta+10)\end{pmatrix}}.$$

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

$$\frac{4n}{\theta}-\frac{n\begin{pmatrix}6\theta^5+25\theta^4+56\theta^3+\\ 123\theta^2+90\theta+26\end{pmatrix}}{\begin{pmatrix}\theta^6+5\theta^5+14\theta^4+41\theta^3\\ +45\theta^2+26\theta+6\end{pmatrix}}-\frac{n\bar{x}}{\theta+1}+\sum_{i=1}^{n}\frac{x_i^2+(2\theta+5)x_i+\left(3\theta^2+8\theta+7\right)}{\begin{pmatrix}x_i^3+(\theta+7)x_i^2+\left(\theta^2+5\theta+15\right)x_i\\ +(\theta^3+4\theta^2+7\theta+10)\end{pmatrix}}=0,$$

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 method, Ragula-Falsi method, and bisection method. In this paper, we use maxLik package [19] with Newton-Raphson method for ML estimation in the statistical software R.

The point estimator of the index of dispersion κ can be estimated by replacing the parameter θ with the ML estimator for θ shown in Eq. (4). Therefore, the point estimator of the index of dispersion κ is given by

$$\hat{\kappa}=\frac{\begin{pmatrix}\hat{\theta}^{13}+12\hat{\theta}^{12}+79\hat{\theta}^{11}+420\hat{\theta}^{10}+1749\hat{\theta}^{9}+5486\hat{\theta}^{8}+13461\hat{\theta}^{7}+24780\hat{\theta}^{6}+\\ 31990\hat{\theta}^{5}+27898\hat{\theta}^{4}+16176\hat{\theta}^{3}+6108\hat{\theta}^{2}+1392\hat{\theta}+144\end{pmatrix}}{\hat{\theta}\begin{pmatrix}\hat{\theta}^{6}+5\hat{\theta}^{5}+14\hat{\theta}^{4}+41\hat{\theta}^{3}+\\ 45\hat{\theta}^{2}+26\hat{\theta}+6\end{pmatrix}\begin{pmatrix}\hat{\theta}^{7}+6\hat{\theta}^{6}+20\hat{\theta}^{5}+64\hat{\theta}^{4}+141\hat{\theta}^{3}+\\ 170\hat{\theta}^{2}+102\hat{\theta}+24\end{pmatrix}},$$

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

3. Bootstrap Methods

In this paper, we focus on three most common bootstrap methods for estimating the confidence interval for the index of dispersion that are most popular in practice: percentile bootstrap, simple bootstrap, and bias-corrected and accelerated bootstrap methods. The computer-intensive bootstrap methods described in this study provide alternative for constructing approximate confidence intervals for the index of dispersion without having to make an assumption about the underlying distribution [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) [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 constructed as follows:

$$CI_{PB} = \left[\hat{\kappa}^*_{(r)}, \hat{\kappa}^*_{(s)}\right], \qquad (5)$$

where $\hat{\kappa}^*_{(\alpha)}$ denotes the α^{th} percentile of the distribution of $\hat{\kappa}^*$ and $0 \le r < s \le 100$. For example, a 95% percentile bootstrap confidence interval with 2,000 bootstrap samples is the interval between the 2.5 percentile value and the 97.5 percentile value of the 2,000 bootstrap parameter estimates.

3.2 Simple Bootstrap (SB) Method

The simple bootstrap method is a method as easy to apply as the percentile bootstrap method. It is sometimes called the basic 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 [20]. The $(1-\alpha)100\%$ simple bootstrap confidence interval for κ is

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

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. (5) for the percentile bootstrap method.

3.3 Bias-Corrected and Accelerated (BCa) Bootstrap Method

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 [22-23] to overpower the over coverage cases in percentile bootstrap confidence intervals [22]. The bias-correction factor $\hat{z}_0$ is estimated as the proportion of the bootstrap estimates less than the original parameter estimate $\hat{\kappa}$,

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\kappa}^* \le \hat{\kappa}\}}{B}\right),$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function (e.g., $\Phi^{-1}(0.975)\approx 1.96$). The acceleration factor $\hat{a}$ is computed through jackknife resampling (i.e., “leave one out” resampling), which associates generating n replicates of the original sample, where n is the number of observations in the sample. Firstly, we obtain the first jackknife replicate by leaving out the first case $(i=1)$ of the original sample. Secondly, the second jackknife replicate is obtained by leaving out the second case $(i=2)$, and so on, until n samples of size $n-1$ are obtained. $\hat{\kappa}_{(-i)}$ is obtained for each of the jackknife resamples. The average of these estimates is

$$\hat{\kappa}_{(\cdot)} = \frac{\sum_{i=1}^{n} \hat{\kappa}_{(-i)}}{n}.$$

Then, the acceleration factor $\hat{a}$ is estimated as follow,

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

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

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

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

where $z_{\alpha/2}$ is the $\alpha/2$ 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

$$CI_{BCa} = \left[\hat{\kappa}^*_{(\alpha_1)}, \hat{\kappa}^*_{(\alpha_2)}\right],$$

where $\hat{\kappa}^*_{(\alpha)}$ denotes the α^{th} percentile of the distribution of $\hat{\kappa}^*$.

4. Simulation Study

The confidence interval for the index of dispersion of a ZTPA 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 [24] to cover cases with different sample sizes (n = 10, 30, 50, 100 and 500). To observe the effect of small and large variances, the true parameter (θ) was set as 0.25, 0.5, 1, 1.5, 2, and 2.5, then the index of dispersion (κ) were 5.083, 3.004, 1.780, 1.245, 0.923, and 0.709, respectively. The number of bootstrap replications (B) was set as 2,000 because Ukoumunne and Davison [25] claimed that 2,000 bootstrap samples are sufficient to estimate the coverage probability for the 95% confidence intervals with a standard error of just under 0.5%. Bootstrap samples of size n were generated from the original sample and each simulation was repeated 1,000 times. 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.

The results of the study are reported in Table 1. For $n = 10$, the coverage probabilities of the three methods tended to be less than 0.95 and so did not reach the nominal confidence level. For $n = 30$, all of the methods once again provided coverage probabilities that were less than the nominal confidence level of 0.95. All bootstrap methods had coverage probabilities close to the nominal confidence levels for large sample sizes $(n \geq 50)$. Additionally, the coverage probabilities of all bootstrap methods were not significantly different for these situations. 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. Unsurprisingly, as the sample size was increased, the average lengths decreased. For small sample sizes $(n \leq 30)$, the average lengths of the PB and SB methods were shorter than those of BCa method. For large sample sizes $(n \geq 50)$, the average lengths of all bootstrap methods were not significantly different.

Table 1. Coverage probability and average length of the 95% bootstrap confidence intervals for the index of dispersion of the ZTPA distribution

n	θ	κ	Coverage probability			Average length		
			PB	SB	BCa	PB	SB	BCa
10	0.25	5.083	0.916	0.912	0.908	2.5321	2.5340	2.6197
	0.5	3.004	0.910	0.918	0.909	1.5674	1.5654	1.6075
	1	1.780	0.891	0.926	0.900	1.1912	1.1934	1.2079

n	θ	κ	Coverage probability			Average length		
			PB	**SB**	**BCa**	**PB**	**SB**	**BCa**
	1.5	1.245	0.886	0.882	0.913	1.0738	1.0727	1.0853
	2	0.923	0.870	0.856	0.910	0.9787	0.9763	0.9980
	2.5	0.709	0.874	0.830	0.923	0.8916	0.8946	0.9338
30	0.25	5.083	0.937	0.930	0.936	1.5667	1.5663	1.5893
	0.5	3.004	0.950	0.953	0.941	0.9352	0.9360	0.9450
	1	1.780	0.941	0.946	0.952	0.7144	0.7146	0.7190
	1.5	1.245	0.937	0.942	0.942	0.6582	0.6585	0.6617
	2	0.923	0.916	0.924	0.938	0.6208	0.6206	0.6260
	2.5	0.709	0.924	0.911	0.943	0.5792	0.5798	0.5890
50	0.25	5.083	0.942	0.938	0.947	1.2280	1.2297	1.2385
	0.5	3.004	0.931	0.933	0.934	0.7336	0.7328	0.7375
	1	1.780	0.946	0.955	0.950	0.5567	0.5564	0.5586
	1.5	1.245	0.944	0.949	0.948	0.5166	0.5172	0.5183
	2	0.923	0.940	0.940	0.947	0.4893	0.4892	0.4913
	2.5	0.709	0.943	0.936	0.946	0.4568	0.4569	0.4619
100	0.25	5.083	0.943	0.945	0.942	0.8728	0.8728	0.8767
	0.5	3.004	0.947	0.947	0.940	0.5220	0.5223	0.5234
	1	1.780	0.948	0.952	0.953	0.3950	0.3952	0.3963
	1.5	1.245	0.942	0.942	0.944	0.3697	0.3695	0.3701
	2	0.923	0.949	0.952	0.952	0.3523	0.3517	0.3530
	2.5	0.709	0.942	0.937	0.947	0.3283	0.3287	0.3306
500	0.25	5.083	0.938	0.940	0.939	0.3945	0.3949	0.3948
	0.5	3.004	0.967	0.966	0.965	0.2352	0.2352	0.2352
	1	1.780	0.944	0.948	0.944	0.1779	0.1777	0.1780
	1.5	1.245	0.947	0.946	0.949	0.1658	0.1658	0.1657
	2	0.923	0.937	0.930	0.938	0.1584	0.1587	0.1586

n	θ	κ	Coverage probability			Average length		
			PB	SB	BCa	PB	SB	BCa
	2.5	0.709	0.943	0.942	0.943	0.1485	0.1487	0.1488

5. Numerical Example

We used two real-world examples to demonstrate the applicability of the bootstrap methods for estimating confidence interval for the index of dispersion of the ZTPA distribution.

5.1 The Number of Unrest Events

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 Summary was used for this example (the sample size was 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. This study used the chi-square goodness-of-fit test for checking whether the sample data is likely to be from a specific theoretical distribution [26]. The chi-square statistic was 3.9112 and the p-value was 0.6887. Thus, a ZTPA distribution with $\hat{\theta}$=0.5707 is suitable for this dataset. The point estimator of the index of dispersion is 2.7261. Table 3 reported the 95% confidence intervals for the index of dispersion of the ZTPA distribution. The estimated parameter $\hat{\theta}$ is approximately 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	1	2	3	4	5	6	7	≥ 8
Observed frequency	3	1	3	2	4	3	4	8
Expected frequency	1.7775	2.4149	2.8174	2.9684	2.9084	2.6987	2.4008	10.0140

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	(2.2753, 3.1634)	0.8881
SB	(2.2754, 3.1770)	0.9016
BCa	(2.2862, 3.1999)	0.9137

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-square goodness-of-fit test, the chi-square statistic was 3.5737 and the p-value was 0.1675. Thus, a ZTPA distribution with $\hat{\theta}$= 2.9563 is suitable for this dataset. The point estimator of the index of dispersion is 0.5720. The 95% confidence intervals for the index of dispersion of the ZTPA distribution are reported in Table 5. The results correspond with the simulation results for κ = 1.245 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	1	2	3	≥ 4
Observed frequency	89	25	11	10
Expected frequency	83.4756	32.3839	12.2451	6.8953

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	(0.4279, 0.7125)	0.2846
SB	(0.4323, 0.7121)	0.2798
BCa	(0.4377, 0.7300)	0.2923

6. Conclusions and Discussion

Herein, we propose three bootstrap methods, namely PB, SB, and BCa, to estimate the confidence interval of the index of dispersion of the ZTPA distribution. When the sample sizes were 10 and 30, the coverage probabilities of all three were substantially lower than 0.95. When the sample size was large enough (i.e., $n \geq 50$), 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 both the simulation study and using real data sets. Our findings provided the simulation results which are correspondent with the study of Jung et al. [28]. They compared three bootstrap confidence intervals for generalized structured component analysis (GSCA) using a Monte Carlo Simulation. They found that the PB method produced confidence intervals closer to the desired level of coverage than the other methods. Future research could focus on the other approaches to compare with the bootstrap methods.

References

[1] R. Kissell, J. Poserina, Optimal sports math, statistics, and fantasy, London, UK: Academic Press, 2017.

[2] F. S. Andrew, R. W. Michael, Practical business statistics, London, UK: Academic Press, 2022.

[3] A. F. Siegel, Practical business statistics, London, UK: Academic Press, 2016.

[4] P. Sangnawakij, Confidence interval for the parameter of the zero-truncated Poisson distribution, The Journal of Applied Science 20(2) (2021) 13 – 22.

[5] F. David, N. Johnson, The truncated Poisson, *Biometrics 8*(4) (1952) 275 – 285.

[6] T. Hussain, A zero truncated discrete distribution: Theory and applications to count data, Pakistan Journal of Statistics and Operation Research 16(1) (2020) 167 – 190.

[7] R. Shanker, The discrete Poisson-Amarendra distribution, International Journal of Statistical Distributions and Applications 2(2) (2016a) 14 – 21.

[8] R. Shanker, Amarendra distribution and its applications, American Journal of Mathematics and Statistics 6(1) (2016b) 44 – 56.

[9] D. V. Lindley, Fiducial distributions and Bayes' theorem, Journal of the Royal Statistical Society, Series B 20(1) (1958) 102 – 107.

[10] R. Shanker, Sujatha distribution and its applications, Statistics in Transition-New Series 17(3) (2016c) 1 – 20.

[11] M. E. Ghitany, D. K. Al-Mutairi, S. Nadarajah, Zero-truncated Poisson-Lindley distribution and its application, Mathematics and Computers in Simulation 79(3) (2008) 279 – 287.

[12] R. Shanker, F. Hagos, (2016). On zero-truncation of Poisson, Poisson-Lindley and Poisson-Sujatha distributions and their applications, Biometrics and Biostatistics International Journal 6(4) (2016) 125 – 135.

[13] R. Shanker, Zero-truncated Poisson-Akash distribution and its applications, American Journal of Mathematics and Statistics 7(6) (2017a) 227 – 236.

[14] R. Shanker, A zero-truncated Poisson-Amarendra distribution and its application, International Journal of Probability and Statistics 6(4) (2017b) 82 – 92.

[15] D.R. Cox, P.A.W. Lewis, The statistical analysis of series of events. London: Methuen.

[16] Wikiwand, Index of dispersion. Available from: https://www.wikiwand.com/en/Index_of_dispersion (Accessed on 24 March 2023).

[17] M. Wood, Statistical inference using bootstrap confidence intervals, Significance 1(4) (2004) 180 – 182.

[18] M. R. Chernick, R. A. Labudde, An introduction to bootstrap methods with applications to R. New Jersey: John Wiley and Sons, 2011.

[19] A. Henningsen, O. Toomet, maxLik: A package for maximum likelihood estimation in R, Computational Statistics 26(3) (2011) 443 – 458.

[20] W. Q. Meeker, G. J. Hahn, L. A. Escobar, Statistical intervals: A guide for practitioners and researchers. New York, USA: John Wiley and Sons, 2017.

[21] B. Efron, The Jackknife, the bootstrap, and other resampling plans, in CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, Philadelphia, USA: SIAM, 1982.

[22] B. Efron, R. J. Tibshirani, An introduction to the bootstrap, New York, USA: Chapman and Hall, 1993.

[23] B. Efron, Better bootstrap confidence intervals, Journal of the American Statistical Association 82(397) (1987) 171 – 185.

[24] R. Ihaka, R. Gentleman, R: A language for data analysis and graphics, Journal of Computational and Graphical Statistics 5(3) (1996) 299 – 314.

[25] O. C. Ukoumunne, A. C. Davison, M. C. Gulliford, S. Chinn, Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient, Statistics in Medicine 22(24) (2003) 3805 – 3821.

[26] N. S. Turhan, Karl Pearson's chi-square tests, Educational Research Review 15(9) (2020) 575 – 580.

[27] R. Shanker, F. Hagos, S. Selvaraj, A. Yemane, On zero-truncation of Poisson and Poisson-Lindley distributions and their applications, Biometrics and Biostatistics International Journal *2(6)* (2015) 168 – 181.

[28] K. Jung, J. Lee, V. Gupta, G. Cho, Comparison of bootstrap confidence interval methods for GSCA using a Monte Carlo simulation. Frontiers in Psychology 10 (2019), https://doi.org/10.3389/fpsyg.2019.02215.