Sains Malaysiana 52(6)(2023): 1879-1888
http://doi.org/10.17576/jsm-2023-5206-22

Bootstrap Methods for Estimating the Confidence Interval for the Parameter of the Zero-Truncated Poisson-Sujatha Distribution and Their Applications

(Kaedah *Bootstrap* untuk Menganggar Selang Keyakinan untuk Parameter Taburan Poisson-Sujatha Terpangkas Sifar dan Aplikasinya)

WARARIT PANICHKITKOSOLKUL*

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, 12121 Pathumthani, Thailand

Received: 29 December 2022/Accepted: 12 June 2023

ABSTRACT

Numerous phenomena involve count data containing non-zero values and the zero-truncated Poisson-Sujatha distribution can be used to model such data. However, the confidence interval estimation of its parameter has not yet been examined. In this study, confidence interval estimation based on percentile, simple, biased-corrected and accelerated bootstrap methods, as well as the bootstrap-t interval, was examined 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. Overall, the bias-corrected and accelerated bootstrap approach outperformed the others, even for small sample sizes. Last, the bootstrap methods were used to calculate the confidence interval for the zero-truncated Poisson-Sujatha parameter via three numerical examples, the results of which match those from the simulation study.

Keywords: Bootstrap interval; count data; interval estimation; Poisson-Sujatha distribution; simulation

ABSTRAK

Banyak fenomena melibatkan data bilangan yang mengandungi nilai bukan sifar dan taburan Poisson-Sujatha terpangkas sifar boleh digunakan untuk memodelkan data tersebut. Walau bagaimanapun, anggaran selang keyakinan parameternya masih belum diperiksa. Dalam kajian ini, anggaran selang keyakinan berdasarkan kaedah persentil, mudah, pembetulan berat sebelah dan dipercepatkan, serta selang *bootstrap*-t, telah diperiksa dari segi kebarangkalian liputan dan panjang selang purata melalui simulasi Monte Carlo. Keputusan menunjukkan bahawa mencapai tahap keyakinan nominal menggunakan kaedah bootstrap tidak mungkin untuk saiz sampel yang kecil tanpa mengira tetapan lain. Selain itu, apabila saiz sampel adalah besar, prestasi kaedah tidak jauh berbeza. Secara keseluruhannya, pendekatan bootstrap yang diperbetulkan berat sebelah dan dipercepatkan mengatasi prestasi yang lain, walaupun untuk saiz sampel yang kecil. Terakhir, kaedah *bootstrap* digunakan untuk mengira selang keyakinan bagi parameter Poisson-Sujatha terpangkas sifar melalui tiga contoh berangka, yang hasilnya sepadan dengan kajian simulasi.

Kata kunci: Anggaran selang; data bilangan; selang *Bootstrap*; simulasi; taburan Poisson-Sujatha

INTRODUCTION

The Poisson distribution is a discrete distribution that measures the probability of a given number of events happening in specific regions of time or space (Andrew & Michael 2022; Kissell & Poserina 2017). Data such as the number of orders a firm will receive tomorrow, the number of people who will apply for a job tomorrow, the number of defects in a finished product, the number of confirmed COVID-19 cases per day, the number of bacteria in a higher organism, follow a Poisson distribution (Siegel 2016).

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 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 (Sangnawakij 2021). 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 (1952) developed the zero-truncated (ZT) Poisson (ZTP) distribution, which has been applied to datasets of the length of stay in hospitals, the number of published journal articles in various disciplines, the number of children ever born to a sample of mothers over 40 years old, and the number of passengers in cars (Hussain 2020). A ZT distribution's pmf can be derived as

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

where $p_0(x;\theta)$ is the pmf of the un-truncated distribution. Shanker (2016a) defined the pmf of the Poisson-Sujatha (PS) distribution having as

$$p_0(x;\theta)=\frac{\theta^3}{\theta^2+\theta+2}\frac{x^2+(\theta+4)x+(\theta^2+3\theta+4)}{(\theta+1)^{x+3}}, \\ x=0,1,2,...,\theta>0. \tag{3}$$

The mathematical and statistical properties of the PS distribution for modeling biological science data were established by Shanker (2016a). The PS distribution arises from the Poisson distribution when parameter λ follows the Sujatha distribution proposed by Shanker (2016b) with porbability density function (pdf)

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

Shanker (2016b) showed that the pdf in (4) is a better model than the exponential and Lindley distribution (Lindley 1958) 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) (Ghitany, Al-Mutairi & Nadarajah 2018), ZT Poisson-Amarendra (ZTPA) (Shanker 2017a), ZT Poisson-Akash (Shanker 2017b) and ZT Poisson-Ishita (Shukla, Shanker & Tiwari 2020) distributions.

Shanker et al. (2015) proposed the ZTPS 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 ZTPS distribution was applied to real data set, it was more suitable than the ZTP and ZTPL distributions.

To the best of our knowledge, no research has been conducted on estimating the confidence interval for the parameter of the ZTPS distribution. It is essential to note that the score function of ZTPS 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 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 for estimating the likely extent of sampling error (Wood 2004). Therefore, the objective of the current study was to assess the efficiencies of four bootstrap methods, namely the percentile bootstrap (PB), the simple bootstrap (SB), the bias-corrected and accelerated bootstrap (BCa), and the bootstrap-t (B-t), to estimate the confidence interval for the parameter of the ZTPS distribution. Additionally, 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 (Chernick & LaBudde 2011). In light of the impossibility of a theoretical comparison of these bootstrap confidence intervals, we conduct a simulated study to evaluate their relative merits. Moreover, the bootstrap confidence intervals had been compared via a simulation study in several studies (Flowers-Cano et al. 2018; Jung et al. 2019; Reiser et al. 2017). In this study, a Monte Carlo simulation study is conducted to compare their performance and used the results to determine the best performing method based on the coverage probability and the average length.

THEORETICAL BACKGROUND

Compounding of probability distributions is a sound and innovative technique to obtain new probability distributions to fit data sets not adequately fit by common parametric distributions. Shanker et al. (2016) proposed a new compounding distribution by compounding Poisson distribution with Sujatha distribution, as there is a need to find more flexible model for analyzing statistical data. The pmf of the PS distribution is given in Equation (3).

Let X be a random variable which follow the ZTPS distribution with parameter θ, it is denoted as XZTPS(θ). Using Equations (2) and (3), the pmf of the ZTPS distribution can be obtained as

$$p(x;\theta) =$$

$$\frac{\theta^3}{\theta^4+4\theta^3+10\theta^2+7\theta+2}\,\frac{x^2+(\theta+4)x+\left(\theta^2+3\theta+4\right)}{(\theta+1)^x},\quad x=1,2,3,...,$$

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

The expected value and variance of X are as follows:

$$E(X)=\mu=\frac{\theta^5+5\theta^4+15\theta^3+25\theta^2+20\theta+6}{\theta\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)}$$

and

$$\mathrm{var}(X)=\sigma^2=$$

$$\frac{\theta^9+10\theta^8+58\theta^7+210\theta^6+503\theta^5+760\theta^4+686\theta^3+352\theta^2+96\theta+12}{\theta^2\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)^2}.$$

By equating the population mean to the corresponding sample mean, the method of moment (MOM) estimator $\tilde{\theta}$ of θ is the solution of the following non-linear equation;

$$(1-\bar{x})\theta^5+(5-4\bar{x})\theta^4+(15-10\bar{x})\theta^3+(25-7\bar{x})\theta^2+$$

$$(20-2\bar{x})\theta+6=0,$$

where $\bar{x}=\sum_{i=1}^{n}x_i/n$ denotes the sample mean. Since the MOM 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, bisection method and Ragula-Falsi method. In this research, we use rootSolve package (Soetaert 2021) with Newton-Raphson method for MOM estimation in the statistical software R (Ihaka & Gentleman 1996).

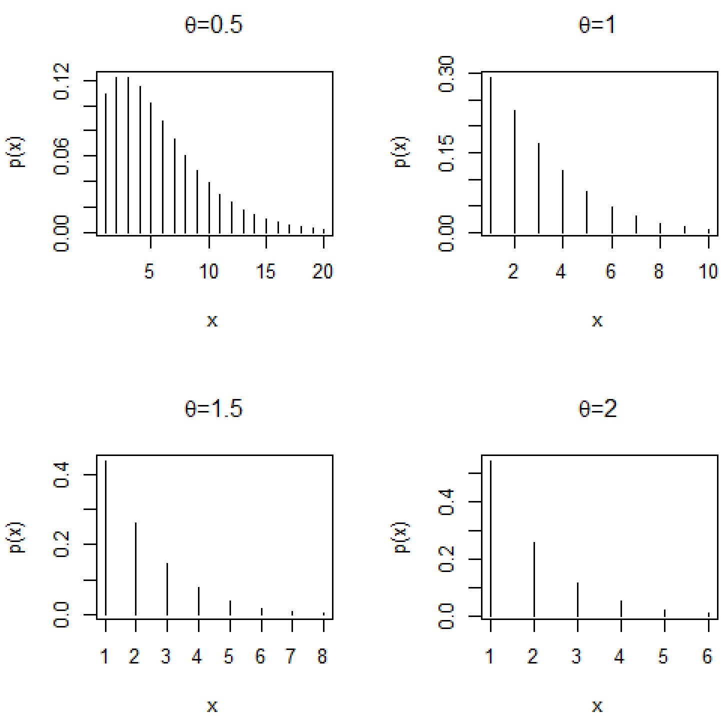

FIGURE 1. The plots of the mass function of the ZTPS distribution with θ=0.5, 1, 1.5 and 2

The maximum likelihood (ML) estimator of θ is obtained by maximizing the log-likelihood function $\log L(x_i;\theta)$ or the logarithm of joint pmf of $X_1, X_2, ..., X_n$. Therefore, the ML estimator for θ of the ZTPS distribution is derived by the following processes:

$$\frac{\partial}{\partial\theta}\log L(x_i;\theta) = \frac{\partial}{\partial\theta}\left[\begin{array}{l} n\log\left(\dfrac{\theta^3}{\theta^4+4\theta^3+10\theta^2+7\theta+2}\right) - \sum_{i=1}^{n} x_i \log(\theta+1) \\ +\sum_{i=1}^{n}\log\left[x_i^2+(\theta+4)x_i+\left(\theta^2+3\theta+4\right)\right] \end{array}\right]$$

$$= \frac{3n}{\theta} - \frac{n\left(4\theta^3+12\theta^2+20\theta+7\right)}{\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)} - \frac{n\bar{x}}{\theta+1} +$$

$$\sum_{i=1}^{n}\frac{x_i+(2\theta+3)}{x_i^2+(\theta+4)x_i+\left(\theta^2+3\theta+4\right)}.$$

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

$$\frac{3n}{\theta} - \frac{n\left(4\theta^3+12\theta^2+20\theta+7\right)}{\left(\theta^4+4\theta^3+10\theta^2+7\theta+2\right)} - \frac{n\bar{x}}{\theta+1} +$$

$$\sum_{i=1}^{n}\frac{x_i+(2\theta+3)}{x_i^2+(\theta+4)x_i+\left(\theta^2+3\theta+4\right)} = 0.$$

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. In this research, we use maxLik package (Henningsen & Toomet 2011) with Newton-Raphson method for ML estimation in the statistical software R.

BOOTSTRAP METHODS

In this study, we focus on four bootstrap methods for estimating confidence interval for the parameter of the ZTPS distribution. In practice, the popular bootstrap methods are the percentile bootstrap, the simple bootstrap, the bias-corrected and accelerated bootstrap, and the bootstrap-t methods. The computer-intensive bootstrap methods described in this study provide alternative for constructing approximate confidence intervals without having to make an assumption about the underlying distribution (Meeker et al. 2017). See the details of some bootstrap methods in DiCiccio and Efron (1996) and Manoharan et al. (2017).

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 θ estimates obtained from resampling or the distribution of $\hat{\theta}^*$, where θ represents a parameter of interest and α is the level of significance (e.g., α = 0.05 for 95% confidence intervals) (Efron 1982). A percentile bootstrap confidence interval for θ can be obtained as follows:

1) B random bootstrap samples are generated,
2) a parameter estimate $\hat{\theta}^*$ is computed 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:

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

where $\hat{\theta}^*_{(\alpha)}$ denotes the α^{th} percentile of the distribution of $\hat{\theta}^*$ and $0 \le r < s \le 100$. For example, a 95% percentile bootstrap confidence interval with 1000 bootstrap samples is the interval between the 2.5 percentile value and the 97.5 percentile value of the 1000 bootstrap parameter estimates.

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 θ and that the estimator of θ is $\hat{\theta}$. The simple bootstrap method assumes that the distributions of $\hat{\theta}-\theta$ and $\hat{\theta}^*-\hat{\theta}$ are approximately the same (Meeker et al. 2017). The $(1-\alpha)100\%$ simple bootstrap confidence interval for θ is

$$CI_{SB} = \left[2\hat{\theta}-\hat{\theta}^*_{(s)}, 2\hat{\theta}-\hat{\theta}^*_{(r)}\right], \tag{6}$$

where the quantiles $\hat{\theta}^*_{(r)}$ and $\hat{\theta}^*_{(s)}$ are the same percentile of empirical distribution of bootstrap estimates $\hat{\theta}^*$ used in (5) for the percentile bootstrap method.

BIAS-CORRECTED AND ACCELERATED BOOTSTRAP (BCa) 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 (Efron 1987; Efron & Tibshirani

1993) to overpower the over coverage issues in percentile bootstrap confidence intervals (Efron & Tibshirani 1993). The bias-correction factor $\hat{z}_0$ is estimated as the proportion of the bootstrap estimates less than the original parameter estimate $\hat{\theta}$,

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

where Φ^{-1} is the inverse function of a standard normal cumulative distribution function (e.g., $\Phi^{-1}(0.975) \approx 1.96$). 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 $(i = 2)$, until n samples of size $n-1$ are obtained. For each of the jackknife resamples, $\hat{\theta}_{(-i)}$ is obtained. The average of these estimates is

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

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

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

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

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

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

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 θ is as follows

$$CI_{BCa} = \left[\hat{\theta}^*_{(\alpha_1)}, \hat{\theta}^*_{(\alpha_2)}\right], \tag{7}$$

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

BOOTSTRAP-t (B-t) METHOD

Suppose that the quantity of interest is θ and that from the given data one can compute the estimate $\hat{\theta}$ and $s.e.(\hat{\theta})$, a corresponding estimate of the standard error of $\hat{\theta}$. Then, the bootstrap estimates $\hat{\theta}_j$ and their corresponding estimated standard errors $s.e.(\hat{\theta}^*_j)$ are computed from each bootstrap sample $j = 1, 2, ..., B$. From these, the bootstrap-t (studentized) statistics

$$R^*_j = \frac{\hat{\theta}^*_j - \hat{\theta}}{s.e.(\hat{\theta}^*_j)}, \quad j = 1, 2, ..., B,$$

are computed. The $(1-\alpha)100\%$ B-t confidence interval for θ is

$$CI_{B-t} = \left[\hat{\theta} - t^*_r \times s.e.(\hat{\theta}), \hat{\theta} - t^*_s \times s.e.(\hat{\theta})\right], \tag{8}$$

where $r = 1-(\alpha/2)$ and $s = \alpha/2$, and t^*_q denotes the q quantile of the distribution of R^*_j.

SIMULATION STUDY

The confidence interval for the parameter of a ZTPS distribution estimated via various bootstrap methods was considered in this study. Due to the unavailability of a direct theoretical comparison, a Monte Carlo simulation study was designed using R version 4.2.2 to cover cases with different sample sizes (= 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, and 2 (the variance of the random variables decreases as the value of θ increases). $B = 1{,}000$ bootstrap samples of size n were generated from the original sample and each simulation was repeated 1,000 times. Without loss of generality, the confidence level $(1-\alpha)$ was set at 0.95. In this study, the confidence intervals were computed based on both of the MOM and ML estimators. 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 parameter of interest. The bootstrap confidence interval that satisfies the criterion is the best in comparison.

The coverage probabilities and average lengths of four bootstrap confidence intervals based on the MOM and ML estimators are reported in Tables 1 and 2, respectively. In the simulation results, the coverage probabilities and average lengths of them based on the MOM and ML estimators are similar for all situations.

Therefore, the results of both cases are described at once. For $n = 10$ and 30, the coverage probabilities of the all methods tended to be less than 0.95 and so did not reach the nominal confidence level. Nevertheless, the BCa method outperformed the other in these scenarios. For $n = 50$, the PB, BCa, and B-t methods attained coverage probabilities close to the nominal confidence level. For $n \geq 100$, all four methods performed similarly well in terms of coverage probability and average length. Thus, as the sample size was increased, the coverage probabilities of the methods tended to increase and approach 0.95. Moreover, the average lengths of the methods increased when the value of θ was increased because of the relationship between the variance and θ.

Unsurprisingly, as the sample size was increased, the average lengths of the four methods decreased. Although the average length of the bootstrap-t method was the shortest when the sample size was small, it provided a poor coverage probability value significantly below the nominal confidence level.

The performances of the four methods differed when the variance of the distribution was small (i.e., $\mathrm{var}(X) = 2.26, 1.37$ for $\theta = 1.5, 2$, respectively) and n was small (i.e., $n \leq 30$); the PB and BCa approaches outperformed the others in terms of coverage probability. For a small sample size, a larger variance (i.e., $\mathrm{var}(X) = 58.59, 16.41$ for $\theta = 0.25, 0.50$, respectively) provided similar performances from all four methods.

TABLE 1. Coverage probability and average length of the 95% confidence intervals for θ of the ZTPS distribution (based on MOM estimator)

n	θ	Coverage probability				Average length			
		PB	SB	BCa	B-t	PB	SB	BCa	B-t
10	0.25	0.882	0.885	0.885	0.891	0.215	0.215	0.202	0.187
	0.5	0.875	0.878	0.887	0.881	0.514	0.513	0.461	0.419
	1	0.862	0.868	0.894	0.844	1.521	1.534	1.181	1.005
	1.5	0.891	0.857	0.918	0.817	3.295	3.286	2.190	1.956
	2	0.879	0.840	0.913	0.819	5.026	5.008	3.264	3.083
30	0.25	0.926	0.921	0.933	0.925	0.117	0.117	0.114	0.113
	0.5	0.925	0.927	0.926	0.926	0.266	0.266	0.255	0.251
	1	0.913	0.914	0.931	0.916	0.670	0.668	0.622	0.605
	1.5	0.922	0.911	0.921	0.909	1.225	1.223	1.090	1.052
	2	0.921	0.897	0.930	0.901	2.003	2.002	1.687	1.596
50	0.25	0.923	0.933	0.934	0.929	0.091	0.091	0.090	0.089
	0.5	0.940	0.923	0.936	0.938	0.203	0.202	0.197	0.195
	1	0.943	0.927	0.946	0.940	0.494	0.494	0.475	0.466
	1.5	0.930	0.932	0.939	0.924	0.897	0.899	0.843	0.826
	2	0.936	0.923	0.945	0.927	1.386	1.386	1.266	1.234
100	0.25	0.952	0.949	0.951	0.952	0.064	0.064	0.063	0.063
	0.5	0.949	0.917	0.939	0.938	0.148	0.149	0.145	0.147
	1	0.955	0.961	0.964	0.962	0.344	0.344	0.337	0.335
	1.5	0.931	0.930	0.928	0.932	0.600	0.600	0.582	0.578
	2	0.939	0.926	0.937	0.938	0.907	0.907	0.867	0.858
500	0.25	0.958	0.954	0.957	0.956	0.028	0.028	0.028	0.028
	0.5	0.954	0.956	0.951	0.954	0.064	0.064	0.064	0.065
	1	0.948	0.955	0.953	0.954	0.153	0.153	0.153	0.152
	1.5	0.948	0.974	0.948	0.982	0.267	0.267	0.270	0.265
	2	0.951	0.940	0.950	0.933	0.404	0.404	0.397	0.398

TABLE 2. Coverage probability and average length of the 95% confidence intervals for θ of the ZTPS distribution (based on ML estimator)

n	θ	Coverage probability				Average length			
		PB	SB	BCa	B-t	PB	SB	BCa	B-t
10	0.25	0.891	0.876	0.909	0.899	0.211	0.212	0.200	0.187
	0.5	0.883	0.884	0.904	0.884	0.511	0.514	0.471	0.421
	1	0.871	0.872	0.880	0.833	1.612	1.647	1.469	1.084
	1.5	0.897	0.851	0.884	0.815	3.007	3.040	2.774	1.952
	2	0.938	0.814	0.942	0.797	4.375	4.363	4.265	3.171
30	0.25	0.928	0.942	0.938	0.930	0.117	0.117	0.115	0.113
	0.5	0.924	0.922	0.930	0.929	0.265	0.264	0.257	0.251
	1	0.927	0.908	0.931	0.929	0.654	0.653	0.624	0.599
	1.5	0.906	0.913	0.921	0.911	1.239	1.235	1.155	1.071
	2	0.922	0.895	0.935	0.907	2.025	2.031	1.840	1.626
50	0.25	0.923	0.921	0.920	0.918	0.089	0.089	0.088	0.088
	0.5	0.934	0.929	0.941	0.939	0.202	0.203	0.199	0.197
	1	0.929	0.916	0.935	0.931	0.484	0.485	0.473	0.462
	1.5	0.934	0.928	0.937	0.931	0.897	0.894	0.859	0.829
	2	0.938	0.912	0.942	0.932	1.385	1.385	1.312	1.239
100	0.25	0.941	0.934	0.948	0.946	0.064	0.064	0.063	0.063
	0.5	0.930	0.936	0.933	0.925	0.141	0.141	0.140	0.139
	1	0.942	0.946	0.935	0.940	0.343	0.343	0.337	0.336
	1.5	0.954	0.931	0.950	0.947	0.595	0.594	0.584	0.574
	2	0.951	0.928	0.950	0.944	0.900	0.902	0.879	0.856
500	0.25	0.949	0.940	0.944	0.950	0.028	0.028	0.028	0.028
	0.5	0.944	0.942	0.944	0.946	0.062	0.063	0.062	0.062
	1	0.945	0.939	0.947	0.942	0.149	0.148	0.148	0.148
	1.5	0.939	0.935	0.937	0.937	0.255	0.256	0.255	0.255
	2	0.947	0.951	0.946	0.946	0.390	0.391	0.387	0.386

NUMERICAL EXAMPLES

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

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 (http://summarise.wbns.oas.psu.ac.th) 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 3. For the Chi-squared goodness-of-fit test (Turhan 2020), the Chi-squared statistic was 4.2384 and the p-value was 0.7522. Thus, a ZTPS distribution with $\hat{\theta} = 0.4252$ is suitable for this dataset. Table 4 reported the 95% confidence intervals for the parameter of the ZTPS 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 BCa and B-t methods were shorter than those of the PB and SB methods. According to the simulation results, the coverage probability should be 0.92-0.93.

DEMOGRAPHIC EXAMPLE

Table 5 shows the demographic data on the number of fertile mothers who have experienced at least one child death (Shanker et al. 2015). The total sample size is 135. For the Chi-squared goodness-of-fit test (Turhan 2020), the Chi-squared statistic was 3.5906 and the p-value was 0.3092. Thus, a ZTPS distribution with $\hat{\theta} = 2.5254$ is suitable for this dataset. The 95% confidence intervals for the parameter of the ZTPS distribution are reported in Table 6. The results correspond with the simulation results for $n = 100$ and $\theta = 2$ because the average lengths of the BCa and B-t methods were shorter than those of the PB and SB methods. According to the simulation results, the coverage probability should be 0.94.

TABLE 3. 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	2.3069	2.7231	2.8944	2.8675	2.7018	2.4518	2.1611	9.8934

TABLE 4. The 95% confidence intervals and corresponding widths using all intervals for the parameter in the unrest events example

Methods	Confidence intervals	Widths
PB	(0.4625, 0.7381)	0.2756
SB	(0.4072, 0.6785)	0.2713
BCa	(0.4590, 0.7180)	0.2590
B-t	(0.4505, 0.7039)	0.2534

TABLE 5. 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

MIGRATION EXAMPLE

Table 7 reports the number of hourseholds having at least one migrant (Singh & Yadava 1971). This dataset contains 590 observations. For the Chi-squared goodness-of-fit test (Turhan, 2020), the Chi-squared statistic was 0.9053 and the p-value was 0.9232. Thus, a ZTPS distribution with $\hat{\theta} = 2.7223$ is suitable for this dataset. Table 8 reports the 95% confidence intervals for the parameter of the ZTPS distribution. The results correspond with the simulation results for $n = 500$ and θ 2 because the average lengths of the BCa and B-t methods are shorter than those of the PB and SB methods. According to the simulation results, the coverage probability should be 0.95.

TABLE 6. The 95% confidence intervals and corresponding widths using all intervals for the parameter in the demographic example

Methods	Confidence intervals	Widths
PB	(2.0716, 3.2597)	1.1881
SB	(1.8031, 2.9829)	1.1798
BCa	(2.0229, 3.1924)	1.1695
B-t	(2.0390, 3.1623)	1.1233

TABLE 7. The number of hourseholds having at least one migrant

Number of migrants	1	2	3	4	≥ 5
Observed frequency	375	143	49	17	6
Expected frequency	378.3141	137.8003	48.7321	16.7743	8.3791

TABLE 8. The 95% bootstrap confidence intervals and corresponding widths using all intervals for the parameter in the migrantion example

Methods	Confidence intervals	Widths
PB	(2.4684, 3.0343)	0.5659
SB	(2.4135, 2.9737)	0.5602
BCa	(2.4567, 3.0098)	0.5531
B-t	(2.4697, 3.0193)	0.5496

Conclusion and Discussion

Herein, we propose four bootstrap methods, namely PB, SB, BCa, and B-t, to estimate the confidence interval for the parameter of the ZTPS distribution. When the sample size was 10 or 30, the coverage probabilities of all four were substantially lower than 0.95. When the sample size was large enough $(n \geq 100)$, the coverage probabilities and average lengths using four bootstrap methods were not markedly different. According to our findings, the BCa method performed the best even for small sample sizes and parameter settings tested in both the simulation study and using real data sets.

Acknowledgements

The author would like to thank editor and the reviewers for the valuable comments and suggestions to improve this paper. The author is grateful for the financial support provided by the Thammasat University under Grant no. TUFT 15/2566.

References

Andrew, F.S. & Michael, R.W. 2022. *Practical Business Statistics*. 8th ed. San Diego: Academic Press.

Chernick, M.R. & Labudde, R.A. 2011. *An Introduction to Bootstrap Methods with Applications to R*. 1st ed. New Jersey: John Wiley and Sons.

David, F. & Johnson, N. 1952. The truncated Poisson. *Biometrics* 8(4): 275-285.

DiCiccio, T.J. & Efron, B. 1996. Bootstrap confidence intervals. *Statistical Science* 11(3): 189-212.

Efron, B. 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82(297): 171-185.

Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: SIAM.

Efron, B. & Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. 1st ed. New York: Chapman and Hall.

Flowers-Cano, R.S., Ortiz-Gómez, R., León-Jiménez, J.E., Rivera, R.L. & Cruz, L.A.P. 2018. Comparison of bootstrap confidence intervals using Monte Carlo simulations. *Water* 10(2): 166. https://doi.org/10.3390/w10020166

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

Henningsen, A. & Toomet, O. 2011. maxLik: A package for maximum likelihood estimation in R. *Computational Statistics* 26(3): 443-458.

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

Ihaka, R. & Gentleman, R. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3): 299-314.

Jung, K., Lee, J., Gupta, V. & Cho, G. 2019. Comparison of bootstrap confidence interval methods for GSCA using a Monte Carlo simulation. *Frontiers in Psychology* 10: 2215. https://doi.org/10.3389/fpsyg.2019.02215

Kissell, R. & Poserina, J. 2017. *Optimal Sports Math, Statistics, and Fantasy*. 1st ed. London: Academic Press.

Lindley, D.V. 1958. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society: Series B* 20(1): 102-107.

Manoharan, T., Arasan, J., Midi, H. & Adam, M.B. 2017. Bootstrap intervals in the presence of left-truncation, censoring and covariates with a parametric distribution. *Sains Malaysiana* 46(12): 2529-2539.

Meeker, W.Q., Hahn, G.J. & Escobar, L.A. 2017. *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. New Jersey: John Wiley and Sons.

Reiser, M., Yao, L., Wang, X., Wilcox, J. & Gray, S. 2017. A comparison of bootstrap confidence intervals for multi-level longitudinal data using Monte-Carlo simulation. In *Monte-Carlo Simulation-Based Statistical Modeling*. ICSA Book Series in Statistics, edited by Chen, D.G. & Chen, J. Singapore: Springer.

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

Shanker, R. 2017a. A zero-truncated Poisson-Amarendra distribution and its application. *International Journal of Probability and Statistics* 6(4): 82-92.

Shanker, R. 2017b. Zero-truncated Poisson-Akash distribution and its applications. *American Journal of Mathematics and Statistics* 7(6): 227-236.

Shanker, R. 2016a. The discrete Poisson-Sujatha distribution. *International Journal of Probability and Statistics* 5(1): 1-9.

Shanker, R. 2016b. Sujatha distribution and its applications. *Statistics in Transition-new Series* 17(3): 391-410.

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

Shukla, K.K., Shanker, R. & Tiwari, M.K. 2020. Zero-truncated Poisson-Ishita distribution and its application. *Journal of Scientific Research* 64(2): 287-294.

Siegel, A.F. 2016. *Practical Business Statistics*. 7th ed. London: Academic Press.

Singh, S.N. & Yadava, K.N. 1971. Trends in rural out-migration at household level. *Rural Demography* 8(1): 53-61.

Soetaert, K. 2021. Rootsolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. R Package 1.8.2.3.

Turhan, N.S. 2020. Karl Pearson's chi-square tests. *Educational Research Review* 15(9): 575-580.

Wood, M. 2004. Statistical inference using bootstrap confidence intervals. *Significance* 1(4): 180-182.

*Corresponding author; email: wararit@mathstat.sci.tu.ac.th