ช่วงความเชื่อมั่นบูตสแทร็ปสำหรับค่าเฉลี่ยของการแจกแจงปัวซง-อิชิตาตัดค่าศูนย์: กรณีศึกษาจำนวนเหตุการณ์ความไม่สงบในจังหวัดชายแดนใต้ของไทย

Bootstrap Confidence Intervals for the Mean of Zero-truncated Poisson-Ishita Distribution: A Case Study of the Number of Unrest Events in the Southern Border Area of Thailand

วราฤทธิ์ พานิชกิจโกศลกุล[1]
E-mail: wararit@mathstat.sci.tu.ac.th
โทรศัพท์: 086-892-9579

บทคัดย่อ

ในหลายสถานการณ์สนใจข้อมูลจำนวนนับที่ไม่มีค่าศูนย์ เช่น จำนวนสินค้าที่อยู่ในตะกร้าสินค้าของลูกค้าในแถวคอยของแคชเชียร์ที่ซูเปอร์มาร์เกต จำนวนวันที่ผู้ป่วยในพักรักษาในโรงพยาบาล เมื่อไม่นานมานี้มีการนำเสนอการแจกแจงปัวซง-อิชิตาที่มีการตัดปลายที่ค่าศูนย์ (ZTPID) สำหรับข้อมูลดังกล่าว แต่การอนุมานเชิงสถิติโดยเฉพาะการประมาณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยยังไม่ได้นำเสนอ ดังนั้น บทความนี้จึงได้นำเสนอช่วงความเชื่อมั่นแบบเปอร์เซ็นต์ไทล์บูตสแทร็ป ช่วงความเชื่อมั่นแบบบูตสแทร็ปอย่างง่าย และช่วงความเชื่อมั่นแบบบูตสแทร็ปที่ปรับค่าเอนเอียง และเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่น โดยประมาณค่าความน่าจะเป็นคุ้มรวมและความกว้างเฉลี่ยของช่วงความเชื่อมั่นด้วยวิธีการจำลองแบบมอนติคาร์โล กำหนดค่าพารามิเตอร์และค่าเฉลี่ยให้มีค่าหลากหลาย ซึ่งแสดงถึงความแปรปรวนของข้อมูลว่ามีค่าน้อยจนถึงมีค่ามาก ผลการวิจัยแสดงให้เห็นว่า ช่วงความเชื่อมั่นทุกวิธียังให้ค่าความน่าจะเป็นคุ้มรวมไม่เข้าใกล้ระดับนัยสำคัญที่กำหนดในกรณีที่ตัวอย่างมีขนาดเล็กสำหรับทุก ๆ สถานการณ์ เมื่อขนาดตัวอย่างใหญ่มากพอ ช่วงความเชื่อมั่นทุกวิธีจะผลการจำลองไม่แตกต่างกันมากนัก ในภาพรวมพบว่าช่วงความเชื่อมั่นแบบบูตสแทร็ปที่ปรับค่าเอนเอียงมีประสิทธิภาพมากกว่าช่วงความเชื่อมั่นวิธีอื่น ๆ ถึงแม้ว่าตัวอย่างจะมีขนาดเล็กก็ตาม นอกจากนี้ ช่วงความเชื่อมั่นแต่ละวิธีได้นำมาประยุกต์ใช้กับจำนวนเหตุการณ์ความไม่สงบในจังหวัดชายแดนใต้ของไทย

คำสำคัญ: การประมาณค่าแบบช่วง, การแจกแจงปัวซง-อิชิตาที่มีการตัดปลายที่ค่าศูนย์, ค่าเฉลี่ย, วิธีบูตสแทร็ป

Abstract

Many situations interact with count data without zero values, such as the number of items in a shopper's basket at a supermarket checkout line and the length of hospital stay, in days. Recently, the zero-truncated Poisson-Ishita distribution (ZTPID) has been proposed for such data, but its statistical inference, especially confidence interval estimation for the mean, has not been proposed. In this paper, the percentile, simple, and biased-corrected and accelerated bootstrap confidence intervals are proposed and compared the performance in terms of coverage probability and average length, which are estimated from the Monte Carlo simulation method. The parameter values and the means of ZTPID are varied, resulting in populations with variances ranging from small to large values. The results indicate that small sample sizes are inadequate to attain the nominal level of confidence for all settings and bootstrap methods. When a sample size is large enough, all confidence intervals do not substantially differ. Overall, it is observed that the biased-corrected and accelerated bootstrap confidence interval outperforms the other confidence intervals, even with small sample sizes. Additionally, each of the bootstrap confidence intervals is estimated for the number of unrest events in the southern border area of Thailand.

Keywords: Interval estimation, zero-truncated Poisson-Ishita distribution, mean, bootstrap method

[1] อาจารย์ประจำสาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ อำเภอคลองหลวง จังหวัดปทุมธานี 12120

Introduction

The Poisson distribution is a discrete probability distribution that measures the probability of an event happening a certain number of times within a given interval of time or space. (Kissell and Poserina, 2017; Andrew and Michael, 2022). Some random variables might follow a Poisson distribution: the number of orders your firm receives tomorrow, the number of calls the firm receives next week for help concerning an "easy-to-assemble" toy, the number of defects in a finished product, the number of customers arriving at a checkout counter in a supermarket from 3 to 6 p.m., etc. (Siegel, 2016).

The probability mass function (p.m.f.) of a Poisson distribution is defined as

$$p(x;\theta)=\frac{e^{-\theta}\theta^{x}}{x!},\quad x=0,1,2,\dots,\ \theta>0, \tag{1}$$

where e is a constant approximately equal to 2.71828 and θ is the parameter of the Poisson distribution. This probability model is usually used in analysis of data containing zero and positive events that have low probabilities of occurrence within some definite time or area range (Sangnawakij, 2021). However, probability models are truncated when a range of possible values for the variables is either disregarded or impossible to observe. The models' zero truncation is another truncation event in which one tries to simulate count data without zero. David and Johnson (1952) developed the zero-truncated Poisson (ZTP) distribution, which can be found in several datasets, including the length of hospital stay, which is recorded as a minimum of one day, the number of journal articles published in various disciplines, the number of children ever born to a sample of mothers over 40 years old, and the number of occupants in passenger cars (Hussain, 2020). The zero-truncated distribution's p.m.f. can be represented as

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

where $p_0(x;\theta)$ is the p.m.f. of the un-truncated distribution. Many distributions have been introduced as an alternative to zero-truncated Poisson distribution on handling the over-dispersion on data, such as zero-truncated Poisson-Lindley (ZTPL) distribution (Ghitany et al., 2008), zero-truncated Poisson-Sujatha (ZTPS) distribution (Shanker and Hagos, 2015) and zero-truncated Poisson-Akash (ZTPA) distribution (Shanker, 2017b).

Recently, Shukla et al. (2020) proposed the zero-truncated Poisson-Ishita (ZTPI) distribution and its applications. The moment, coefficient of variation, skewness, kurtosis and the index of dispersion of ZTPID had been proposed. The method of moments and the method of maximum likelihood had also derived for estimating its parameter. Furthermore, ZTPI distribution was applied on two real data sets to test its goodness of fit. It was more suitable than ZTP, ZTPL, ZTPS and ZTPA distributions.

In the review literature, there is no research study for estimating the bootstrap confidence intervals for the mean of ZTPI distribution. Bootstrap confidence intervals provide a way of quantifying the uncertainties in the inferences that can be drawn from a sample of data. The concept is to use a simulation, based on the actual data, to estimate the likely extent of sampling error (Wood, 2004). Therefore, the objective of the paper is to study the efficiency of bootstrap confidence intervals for the population mean of ZTPI distribution in three methods, namely, percentile bootstrap (PB), simple bootstrap (SB), and bias-corrected and accelerated (BCa) bootstrap methods. Because a theoretical comparison is not possible, we conduct a simulation study to compare the performance of these bootstrap confidence intervals, and use these results to suggest a bootstrap confidence interval with coverage probability that attained a nominal confidence level and short average length for practitioners.

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. Shukla and Shanker (2019) proposed a new compounding distribution by compounding Poisson distribution with Ishita distribution, as there is a need to find more flexible model for analyzing statistical data. The p.m.f. of the Poisson-Ishita distribution is given by

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

Let X be a random variable which follow ZTPI distribution with parameter θ, it is denoted as $X \sim \mathrm{ZTPI}(\theta)$. Using Equations (2) and (3), the p.m.f. of ZTPI distribution can be obtained as

$$p(x;\theta) = \frac{\theta^3}{\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2} \frac{x^2+3x+(\theta^3+2\theta^2+\theta+2)}{(\theta+1)^x}, \quad x=1,2,3,\ldots,\ \theta>0. \tag{4}$$

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

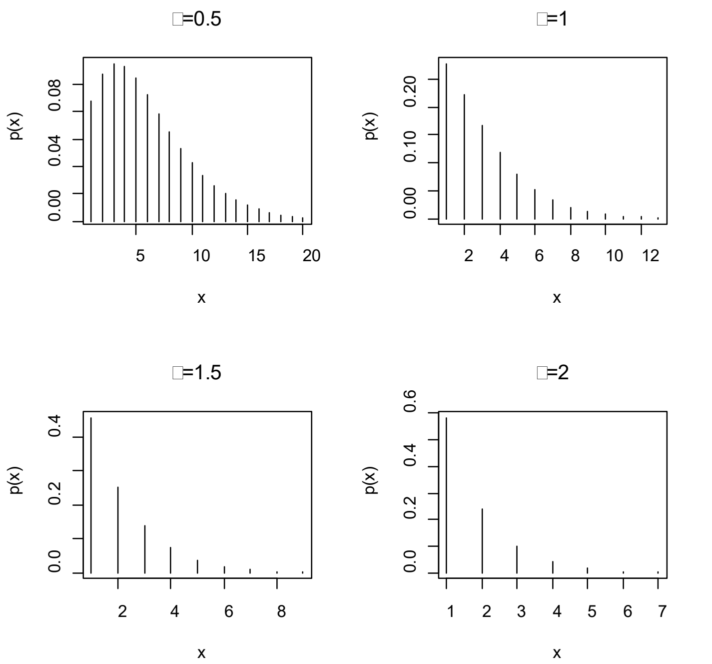

Figure 1 The plots of the mass function of the ZTPI distribution with θ =0.5, 1, 1.5 and 2

The expected value and variance of X are as follows:

$$E(X) = \mu = \frac{\theta^6+3\theta^5+3\theta^4+7\theta^3+18\theta^2+18\theta+6}{\theta(\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2)} \tag{5}$$

and

$$\mathrm{var}(X) = \sigma^2 = \frac{(\theta+1)(\theta^{10}+4\theta^9+6\theta^8+27\theta^7+69\theta^6+98\theta^5+136\theta^4+208\theta^3+180\theta^2+72\theta+12)}{\theta^2(\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2)^2}.$$

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

$$\frac{\partial}{\partial\theta}\log L(x_i;\theta)=\frac{\partial}{\partial\theta}\left[\begin{array}{l} n\log\left(\dfrac{\theta^3}{\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2}\right)-\sum_{i=1}^{n}x_i\log(\theta+1) \\ +\sum_{i=1}^{n}\log\left[x_i^2+3x_i+\left(\theta^3+2\theta^2+\theta+2\right)\right] \end{array}\right]$$

$$=\frac{3n}{\theta}-\frac{n\left(5\theta^4+8\theta^3+3\theta^2+12\theta+6\right)}{\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2}-\frac{n\overline{x}}{\theta+1}+\sum_{i=1}^{n}\frac{\left(3\theta^2+4\theta+1\right)}{x_i^2+3x_i+\left(\theta^3+2\theta^2+\theta+2\right)}.$$

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

$$\frac{3n}{\theta}-\frac{n\left(5\theta^4+8\theta^3+3\theta^2+12\theta+6\right)}{\theta^5+2\theta^4+\theta^3+6\theta^2+6\theta+2}-\frac{n\overline{x}}{\theta+1}+\sum_{i=1}^{n}\frac{\left(3\theta^2+4\theta+1\right)}{x_i^2+3x_i+\left(\theta^3+2\theta^2+\theta+2\right)}=0,$$

where $\overline{x}=\sum_{i=1}^{n}x_i/n$ 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, bisection method and Ragula-Falsi method. In this paper, we use maxLik package (Henningsen and Toomet, 2011) for ML estimation in the statistical software R.

The point estimator of the population mean μ can be estimated by replacing the parameter θ with the ML estimator for θ shown in Equation (5). Therefore, the point estimator of the population mean μ is given by

$$\hat{\mu}=\frac{\hat{\theta}^6+3\hat{\theta}^5+3\hat{\theta}^4+7\hat{\theta}^3+18\hat{\theta}^2+18\hat{\theta}+6}{\hat{\theta}\left(\hat{\theta}^5+2\hat{\theta}^4+\hat{\theta}^3+6\hat{\theta}^2+6\hat{\theta}+2\right)},$$

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

Bootstrap Confidence Interval Methods

In this paper, we focus on the three bootstrap confidence interval methods that are most popular in practice: percentile bootstrap, simple bootstrap, and bias-corrected and accelerated bootstrap confidence intervals.

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 μ estimates obtained from resampling or the distribution of $\hat{\mu}^*$, where μ 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 μ can be obtained as follows:

1) B random bootstrap samples are generated,

2) a parameter estimate $\hat{\mu}^*$ 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:

$$CI_{PB}=\left[\hat{\mu}^*_{(r)},\hat{\mu}^*_{(s)}\right], \tag{6}$$

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

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

$$CI_{SB} = \left[2\hat{\mu}-\hat{\mu}^{*}_{(s)}, 2\hat{\mu}-\hat{\mu}^{*}_{(r)}\right], \tag{7}$$

where the quantiles $\hat{\mu}^{*}_{(r)}$ and $\hat{\mu}^{*}_{(s)}$ are the same percentile of empirical distribution of bootstrap estimates $\hat{\theta}^{*}$ used in (6) for the percentile bootstrap method.

3. Bias-corrected and accelerated (BCa) bootstrap method

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

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

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 $(i=2)$, and so on, until n samples of size $n-1$ are obtained. For each of the jackknife resamples, $\hat{\mu}_{(-i)}$ is obtained. The average of these estimates is

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

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

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

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

$$\alpha_1 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1-\hat{a}\left(\hat{z}_0 + z_{\alpha/2}\right)}\right\} \text{ and } \alpha_2 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1-\hat{a}\left(\hat{z}_0 + z_{1-\alpha/2}\right)}\right\},$$

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

$$CI_{BCa} = \left[\hat{\mu}^{*}_{(\alpha_1)}, \hat{\mu}^{*}_{(\alpha_2)}\right], \tag{8}$$

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

Simulation Study

In this study, the bootstrap confidence intervals for the mean of ZTPI distribution are determined. Because a theoretical comparison is not possible, a Monte Carlo simulation study was designed using R version 4.2.1 statistical software (Ihaka and Gentleman, 1996) and conducted to compare the performances of three bootstrap confidence intervals for the mean in a ZTPI distribution. The study was designed to cover cases with different sample sizes, as n = 10, 25, 50, 75 and 100, reflecting small to large samples. To observe the effect of small and large variances, the true parameter (θ) was given by 0.25, 0.5, 0.75, 1 and 2, and the population means μ are 12.0523, 6.0968, 4.1094, 3.1111 and 1.7182, respectively. It shows that the mean and variance of random variables will decrease as the value of θ increases. The $B=1000$ bootstrap samples of size n are generated from the original sample and repeated the simulation 5000 times. Without loss of generality, the confidence level $(1-\alpha)$ was set at 0.95. The bootstrap confidence intervals were compared in terms of their coverage probabilities and the average lengths of their performances. A bootstrap confidence interval which has a coverage probability greater than or close to the nominal confidence level means that it contains the true value with a given probability. In other words, it can precisely estimate the parameter of interest. The bootstrap confidence interval that satisfies the criterion is the best in comparison.

The results of the study are reported in Table 1. When the sample size is only 10, the coverage probabilities tend to be less than 0.90, except in a few cases where the values of μ are less or equal to 3.1111 for BCa bootstrap method. The nominal confidence level of SB method is difficult to reach in circumstances where $\mu=1.7182$ and $n=10.$ Generally, as sample size increases, the coverage probability tends to increase and approach 0.95. The average length also obviously increases when the value of μ increases; this is because of the relationship between the variance and μ value. Unsurprisingly, as sample size increases, the average length falls. It can be as small as approximately 0.8239 when μ is at 1.7182 and the sample size is 25; the largest average length, 5.9211, occurs when $\mu=12.0523$ and $n=25$ in the case of BCa method. Furthermore, the average lengths of PB method are similar to those of SB method in all situations.

When three types of confidence intervals are compared, they can differ when the variance of the distribution is small, i.e., $\mathrm{var}(X)=5.0185, 1.2602$ for $\mu=3.1111,\ 1.7182,$ respectively, and n is small, i.e., $n=25$; the BCa bootstrap method outperform the PB and SB methods in terms of coverage probability. Given the same sample sizes, scenarios with larger variances, i.e., $\mathrm{var}(X)=59.3808,\ 17.2077,\ 8.4358$ for $\mu=12.0523,\ 6.0968,\ 4.1094$ respectively, lead to the conclusion that all confidence intervals perform approximately the same.

Numerical Example

We used a real-world example to demonstrate the application of the bootstrap confidence intervals for the mean of the ZTPI distribution established in the preceding section.

The number of unrest events occurred in the southern border area of Thailand was collected by the Southern Border Area News Summarizes (SBAN Summarizes) (http://summarise.wbns.oas.psu.ac.th) in July 2020 to August 2022. The number of unrest events per month in five southern provinces of Pattani, Yala, Narathiwat, Songkhla, and Satun provinces was reported in Table 2; the total sample size is 26. For the chi-square goodness-of-fit test (Turhan, 2020), the chi-square statistic was 2.5298 and the p-value was 0.9248. It was found that these data fitted well to the ZTPI distribution with parameter $\hat{\theta}$ of 0.4500. The point estimator of the population mean is 6.7575. Table 3 reported the 95% bootstrap confidence intervals for the mean of the ZTPI distribution. The estimated parameter $\hat{\theta}$ is between 0.25 and 0.5. The results correspond with the simulation results with $n=25$ because the average lengths in PB and SB methods are shorter than in BCa bootstrap method. According to the simulation results, the coverage probability is expected to be 0.92.

Table 1. Coverage probability and average length of the 95% bootstrap confidence intervals for μ in the zero-truncated Poisson-Ishita distribution

n	θ	μ	Coverage probability			Average length		
			PB	SB	BCa	PB	SB	BCa
10	2	1.7182	0.8596	0.8124	0.9354	1.2000	1.2003	1.3755
	1	3.1111	0.8946	0.8666	0.9096	2.5049	2.5059	2.6504
	0.75	4.1094	0.8860	0.8706	0.8910	3.2019	3.1999	3.3505
	0.5	6.0968	0.8814	0.8652	0.8878	4.5973	4.5999	4.8053
	0.25	12.0523	0.8838	0.8740	0.8872	8.5869	8.5910	8.9717
25	2	1.7182	0.9134	0.8948	0.9378	0.8239	0.8239	0.8773
	1	3.1111	0.9220	0.9056	0.9306	1.6627	1.6635	1.7086
	0.75	4.1094	0.9276	0.9174	0.9294	2.1769	2.1772	2.2318
	0.5	6.0968	0.9244	0.9170	0.9272	3.1100	3.1124	3.1889
	0.25	12.0523	0.9212	0.9152	0.9268	5.7835	5.7859	5.9211
50	2	1.7182	0.9310	0.9204	0.9400	0.5950	0.5946	0.6146
	1	3.1111	0.9432	0.9352	0.9420	1.2053	1.2057	1.2248
	0.75	4.1094	0.9402	0.9362	0.9400	1.5667	1.5678	1.5877
	0.5	6.0968	0.9430	0.9370	0.9424	2.2455	2.2473	2.2779
	0.25	12.0523	0.9356	0.9334	0.9358	4.1784	4.1760	4.2344
75	2	1.7182	0.9334	0.9222	0.9426	0.4883	0.4887	0.5002
	1	3.1111	0.9416	0.9364	0.9408	0.9908	0.9903	1.0022
	0.75	4.1094	0.9436	0.9390	0.9434	1.2886	1.2887	1.3004
	0.5	6.0968	0.9404	0.9380	0.9418	1.8493	1.8475	1.8651
	0.25	12.0523	0.9466	0.9422	0.9474	3.4443	3.4450	3.4761
100	2	1.7182	0.9396	0.9286	0.9454	0.4237	0.4237	0.4311
	1	3.1111	0.9436	0.9398	0.9444	0.8605	0.8602	0.8672
	0.75	4.1094	0.9416	0.9358	0.9444	1.1208	1.1205	1.1286
	0.5	6.0968	0.9466	0.9456	0.9462	1.6084	1.6074	1.6187
	0.25	12.0523	0.9416	0.9420	0.9382	2.9885	2.9857	3.0071

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	3	3	3	8
Expected frequency	1.8586	2.3890	2.6657	2.7161	2.5995	2.3772	2.1001	9.2937

Table 3. The 95% bootstrap confidence intervals and corresponding widths using all intervals for the population mean in the unrest events example

Methods	Confidence intervals	Widths
PB	(5.1071, 8.4197)	3.3126
SB	(5.0725, 8.3636)	3.2911
BCa	(5.0794, 8.6045)	3.5251

Conclusions and Discussion

The bootstrap confidence intervals of the mean of the zero-truncated Poisson-Ishita distribution are investigated in this study. At $n=10,$ all coverage probabilities are substantially lower than 0.90. A sample size of 25 is still insufficient to achieve the nominal confidence level for all $\theta's$ and bootstrap confidence intervals. When the sample size is large enough, i.e., greater than or equal to 50, the coverage probabilities from three bootstrap methods, as well as the average length, are not markedly different. According to our findings, the BCa bootstrap method performs best even with small sample sizes as long as the variance of ZTPID is not too large.

Acknowledgements

The author would like to thank the reviewers for the valuable comments and suggestions to improve this paper.

References

Andrew, F. S., & Michael, R. W. (2022). **Practical business statistics**. Academic Press.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. **Biometrika**. 74(3), 457–468.

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

Donna, L. M., Wilson, W. J., & Freund, R. J. (2022). **Statistical methods**. Academic Press.

Efron, B. (1982). **The Jackknife, the bootstrap, and other resampling plans**, in CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38 (Philadelphia, PA: SIAM).

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

Efron, B., & Tibshirani, R. J. (1993). **An Introduction to the bootstrap**. Chapman and Hall.

Finney, D. J., & Varley, G. C. (1955). An example of the truncated Poisson distribution. **Biometrics**. 11(3), 387–394.

Garman, P. (1923). The European red mites in Connecticut apple orchards. **Connecticut Agricultural Experiment Station Bulletin**. 252, 103–125.

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.

Kissell, R., & Poserina, J. **Optimal sports math, statistics, and fantasy**. Academic Press.

Letson, D., & McCullough, B. D. (1998). Better confidence intervals: The double bootstrap with no pivot. American **Journal of Agricultural Economics**. 80(3), 552–559.

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

McCullough, B. D., & Vinod, H. D. (1998). Implementing the double bootstrap. **Computational Economics**. 12(1), 79–95.

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). **Statistical intervals: A guide for practitioners and researchers**. John Wiley and Sons.

Nankervis, J. C. (2005). Computational algorithms for double bootstrap confidence intervals. **Computational Statistics and Data Analysis**. 49(2), 461–475.

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). The discrete Poisson-Akash distribution. **International Journal of Statistics and Probability**. 6(1), 1–10.

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

Shanker, R. (2015). Akash distribution and its applications. **International Journal of Statistics and Probability**. 4(3), 65–75.

Shanker, R., & Hagos, F. (2015). Zero-truncated Poisson-Sujatha distribution with applications. **Journal of Ethiopian Statistical Association**. 24, 55–63.

Shanker, R., & Shukla, K. K. (2017). Ishita distribution and its application to model lifetime data. **Biometrics and Biostatistics International Journal**. 5(2), 1–9.

Shukla, K. K., & Shanker, R. (2019). A discrete Poisson-Ishita distribution and its applications. **International Journal of Economics and Statistics**. 20(2), 109–122.

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

Siegel, A. F. (2016). **Practical business statistics**. Academic Press.

Simonoff, J. S. (2003). **Analyzing categorical data**. Springer.

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.