*Vietnam Journal of Science and Technology* ***62*** *(6) (2024) 1173-1184*
doi:10.15625/2525-2518/18056

# Bootstrap methods for estimating the confidence interval for the parameter of zero-truncated Poisson-Garima distribution and their application

**Wararit Panichkitkosolkul[1, 2, *], Chainarong Kesamoon[1], Sirichan Vesarachasart[1]**

[1]*Department of Mathematics and Statistics, Thammasat University, Klong Luang, Pathumthani 12121 Thailand*

[2]*Thammasat University Research Unit in Mathematical Sciences and Applications, Pathum Thani, 12120, Thailand*

[*]Emails*: wararit@mathstat.sci.tu.ac.th*

Received: 1 February 2023; Accepted for publication: 25 August 2023

**Abstract.** Zero-truncated count data are of significant importance across various fields, including biological sciences, medical sciences, demography, and ecology. Recently, a new distribution has been introduced for this type of data, namely the zero-truncated Poisson-Garima (ZTPG) distribution. However, the confidence interval (CI) for its parameter has yet to be thoroughly analysed. This paper investigates CI estimation using the percentile bootstrap (PB), simple bootstrap (SB), bias-corrected, and accelerated (BCa) bootstrap methods, as well as the bootstrap-t CI, with a focus on coverage rate and average width assessed through Monte Carlo simulations. The simulation results show that the bootstrap methods could not achieve the nominal confidence level for very small sample sizes, irrespective of other settings. Additionally, the bootstrap methods did not exhibit a notable difference in performance when the sample size was large. It is clear that the BCa bootstrap approaches outperformed the alternatives, even with small sample sizes. Finally, bootstrap methods were applied to compute the CIs for the ZTPG parameter in two real-world applications, with the CI widths closely matching the average widths from the simulation study.

*Keywords:* bootstrap interval, discrete distribution, interval estimation, parameter, Poisson-Garima distribution.

*Classification numbers*: 4.6.2., 4.6.4.

## 1. INTRODUCTION

One of the discrete probabilities is the Poisson distribution, which characterizes the likelihood of a given number of events occurring within specified intervals of space or time [1 - 2]. Examples of data fitting this distribution include the hourly count of patients visiting an emergency department, the number of goals scored by a soccer team during a match, the monthly frequency of minor earthquakes in a region, and the daily volume of spam emails received by a particular email account [3]. Based on the Poisson distribution, the probability mass function (pmf) is defined as

$$p(y;\eta)=\frac{\exp(-\eta)\eta^{y}}{y!},\quad y=0,1,2,...,\ \eta>0, \tag{1}$$

where $\exp(\cdot)$ is the exponential function and $\eta$ is both the mean and variance of the distribution. The parameter $\eta$ represents the expected number of occurrences within a specified time or space interval. The Poisson distribution is particularly suitable for analyzing datasets that contain both zero and positive values, especially when these events occur infrequently over a fixed interval [4]. The equi-dispersion assumption, requiring the mean and variance of the random variable to be equivalent, limits the use of the Poisson distribution as a fundamental model for count data analysis. Quite the reverse, count data frequently convey over-dispersion, with a larger variance compared to the mean [5]. Flawed analyses and faulty conclusions can result from the use of the Poisson distribution to over-dispersion data [6]. Using a mixed Poisson distribution [7] with a Poisson parameter presumed to be a random variable with a single parameterized distribution [8] is a common substitute when the count data exhibit over-dispersion.

The Poisson and Garima distributions were recently joined by Shanker [9] to create the Poisson-Garima (PG) distribution. Focusing on its mathematical and statistical attributes was assessed. The PG distribution was developed by assuming that the Poisson parameter $\eta$ follows a Garima distribution [10]. The PG distribution demonstrates greater suitability compared to the Poisson and Poisson-Lindley [11] distributions when applied to two real datasets. The pmf of the PG distribution is expressed in Eq. (2),

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

The Garima distribution, a continuous lifetime distribution, was first introduced by Shanker [10] and has a probability density function (pdf) as defined in Eq. (3),

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

Comprising a scale parameter $\theta$ and gamma distribution with a shape parameter of 2 as well as scale parameter $\theta$ with proportions of $(\theta+1)/(\theta+2)$ and $1/(\theta+2),$ respectively, it is a combination of exponential distribution. Moreover, the Garima distribution is a more suitable model than the exponential, Lindley [12], Akash [13], Aradhana [14] and Sujatha [15] distributions for modeling behavioral science data, as shown by Shanker [10]. Comprising several statistical properties of the Garima distribution have been considered [10]. Figure 1 illustrates plots of the Garima distribution with various stipulated parameter values.

Probability distributions can be truncated when certain ranges of potential values are either ignored or unobservable. Zero-truncation (ZT) is often used in the analysis of count data that excludes zeros. The ZT Poisson (ZTP) distribution had been developed by David and Johnson [16] in 1952. This distribution has found practical applications in various datasets, such as the number of hospital admissions per patient, considering only patients admitted at least once; the number of purchases made by customers who have made at least one purchase; and the number of goals scored by players who scored at least once during a season [17]. The pmf of the ZT distribution can be derived as shown in Eq. (4),

$$p(x;\theta) = \frac{p_0(x;\theta)}{1-p_0(0;\theta)}, \quad x = 1,2,3,..., \tag{4}$$

where $p_0(x;\theta)$ is the pmf of the untruncated distribution. Numerous ZT distributions have been proposed as alternatives to the ZTP distribution for focusing on over-dispersion in count data, including the ZT Poisson-Lindley (ZTPL) [18], ZT Poisson-Sujatha [19], and ZT Poisson-Akash [20] distributions. Shanker and Shukla [21] established the ZT Poisson-Garima (ZTPG) distribution, providing essential insights into its statistical properties. The method of moments and maximum likelihood estimation method are statistical techniques developed for point-estimating its parameter. The application of the ZTPG distribution to real-world data demonstrated its superior suitability compared to the ZTP and ZTPL distributions.

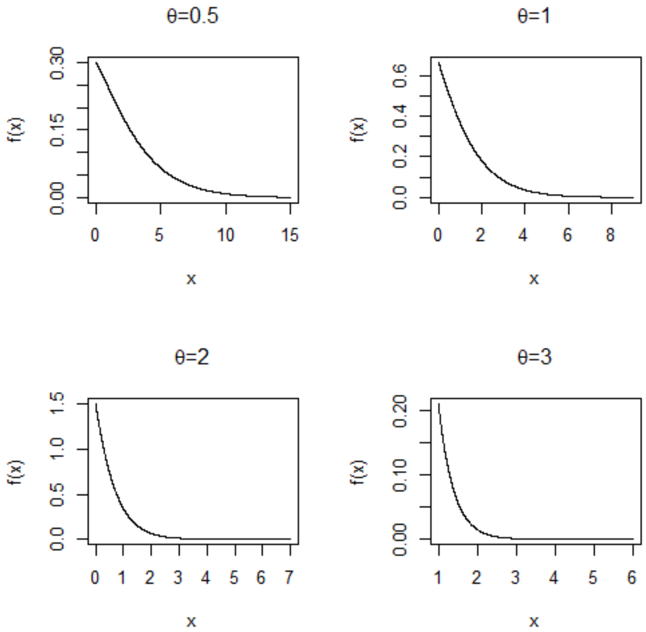


*Figure 1*. Plots of the Garima distribution's pdf for $\theta = 0.5$, 1, 2, and 3.

The confidence interval (CI), a vital result for many statistical investigations and an essential component in the interpretation of parameter estimations, is an array of values in statistics and probability that probably include the true value of the population parameter of interest [22]. Attempts expected to calculate bootstrap CIs for the ZTPG distribution parameter are not covered in the reviewed literature. A way of quantifying the vagueness in statistical inference based on a sample of data is offered by the bootstrap CIs for estimating the parameter. The idea is to estimate the prospective weight of sampling error by conducting a simulation study using the real data [23]. Comprising the percentile bootstrap (PB), the simple bootstrap (SB), the bias-corrected and accelerated (BCa) bootstrap and bootstrap-t (B-t) methods to assess the parameter of the ZTPG distribution, this study aims to examine the efficacies of four bootstrap CIs. Furthermore, bootstrap CIs will not be precise though they will be reliable. Thus, as the sample size increases, the confidence level approaches $1-\alpha$ [24]. This research undertook a Monte Carlo simulation study to contrast the performance levels. The simulation results and real-world applications were used to establish the best-performing approach according to the coverage rate and average width.

## 2. ZERO-TRUNCATED POISSON-GARIMA DISTRIBUTION AND THE PARAMETER ESTIMATION

The technique of combining probability distributions represents a creative and efficient method for developing new distributions that can more accurately reflect the characteristics of datasets which are not well described by classical statistical distributions. By combining different distributions, this method enhances the ability of the resulting distribution, making it more capable of capturing complex data behaviors, such as heavy tails, skewness, or multimodal patterns. This flexibility is particularly valuable in fields where data exhibit unique or irregular features that standard distributions fail to address, thereby improving the overall fit and predictive performance of the distribution. Shanker and Shukla [21] proposed a novel compound distribution by integrating the Poisson and Garima distributions, aiming to fulfill the demand for a more adaptable distribution in statistical data analysis. The pmf of the PG distribution is provided in Eq. (2).

Let $X$ be a random variable that follows the ZTPG distribution with parameter $\theta,$ denoted as $X \sim$ ZTPG ($\theta$). Using Eqs. (2) and (4), the pmf of ZTPG distribution is given by

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

Figure 2 presents plots of the pmf of the ZTPG distribution for some values of the parameter $\theta$. The expected value and variance of ZTPG random variable are given, respectively, as follows:

$$E(X)=\mu=\frac{(\theta+1)(\theta^2+4\theta+3)}{\theta\left(\theta^2+4\theta+2\right)}\quad \text{and}\quad \operatorname{var}(X)=\sigma^2=\frac{(\theta+1)\left(\theta^4+9\theta^3+26\theta^2+25\theta+7\right)}{\theta^2\left(\theta^2+4\theta+2\right)^2}.$$

The estimation of parameter is achieved by finding the value that maximizes the log-likelihood function, which is equivalent to the natural logarithm of the joint pmf function of $X_1,...,X_n$. Therefore, the maximum likelihood (ML) estimator for $\theta$ in the ZTPG distribution is derived using the derivative:

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L(x_i;\theta)&=\frac{\partial}{\partial\theta}\left[n\log\left(\frac{\theta}{\theta^2+4\theta+2}\right)-\sum_{i=1}^{n}x_i\log(\theta+1)+\sum_{i=1}^{n}\log\left[\theta x_i+\left(\theta^2+3\theta+1\right)\right]\right]\\&=\frac{n}{\theta}-\frac{2n\left(\theta+2\right)}{\left(\theta^2+4\theta+2\right)}-\frac{n\bar{x}}{\theta+1}+\sum_{i=1}^{n}\frac{x_i+2\theta+3}{\theta x_i+\left(\theta^2+3\theta+1\right)}.\end{aligned}$$

To solve the equation $\frac{\partial}{\partial\theta}\log L(x_i;\theta)=0$ for $\theta,$ we obtain the following non-linear equation

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

where $\bar{x}$ denotes the sample mean. As the ML estimator for parameter $\theta$ cannot be expressed in a closed-form solution, numerical iterative methods are employed to solve the non-linear equation. In this study, the maxLik package [25] in R [26] was used, specifically applying the Newton-Raphson method for ML estimation.

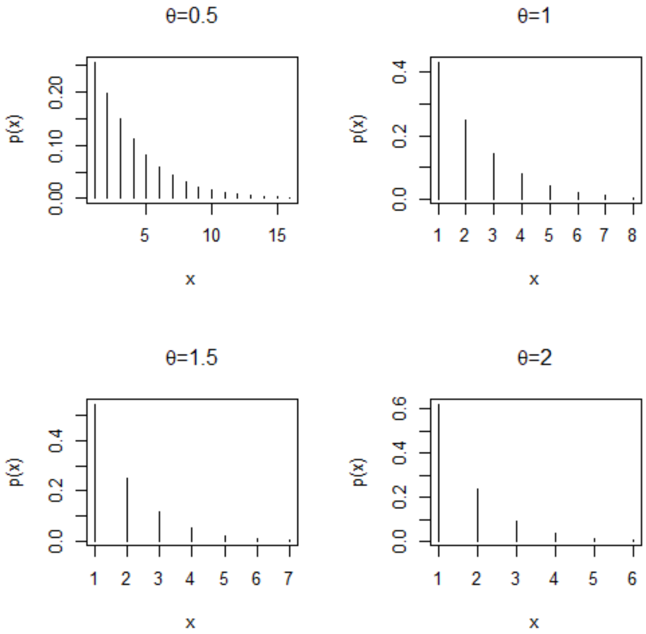


*Figure 2.* Plots of the ZTPG distribution's pmf for $\theta = 0.5, 1, 1.5,$ and 2.

## 3. BOOTSTRAP CONFIDENCE INTERVEL METHODS

A parametric estimator of standard errors is used to calculate CIs by adding or subtracting the standard error multiplied by a critical value, assuming normality of the estimator [27]. However, when the normality assumption violates or standard error estimation is problematic, bootstrap methods can be employed [28]. This paper focuses on bootstrap techniques for estimating the interval of the ZTPG distribution parameter [29].

### 3.1. Percentile bootstrap (PB) method

The PB CI is derived from the $1-\alpha$ percentiles of the bootstrap distribution of the parameter estimates, denoted by $\hat{\theta},$ where $\theta$ denotes the parameter of the ZTPG distribution and $1-\alpha$ is the nominal confidence level [30]. The PB CI for $\theta$ is obtained through the following steps:

1) Generate Bootstrap Samples: Draw $B$ random bootstrap samples from the fundamental distribution with substitution, where $B$ represents the amount of bootstrap replications.

2) Estimate the Parameter: Calculate the parameter estimate $\hat{\theta}^*$ for each of the $B$ bootstrap samples.

3) Order the Estimates: Arrange the $B$ bootstrap parameter estimates in ascending order.

4) Construct the PB CI: Identify the $\alpha/2$ and $1-(\alpha/2)$ quantiles of the ordered bootstrap estimates to form the PB CI. The $(1-\alpha)100\%$ PB CI is given by

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

where $\hat{\theta}^*_{(r)}$ represents the $r^{\text{th}}$ quantile of the set of ordered quantiles from the lowest to highest, $\hat{\theta}^*_{(s)}$ is the $s^{\text{th}}$ quantile of the same set, $r = \lceil (\alpha/2)B \rceil$, $s = \lceil (1-(\alpha/2))B \rceil$, where $\lceil \cdot \rceil$ is the ceiling function, and $\alpha$ is the significance level. This paper identified $B$ = 1,000 and $\alpha$ = 0.05; the quantile consistent with the lower threshold of the CI was $\hat{\theta}^*_{(r)} = \hat{\theta}^*_{(25)}$ and that conforming to the upper threshold was $\hat{\theta}^*_{(s)} = \hat{\theta}^*_{(975)}$.

### 3.2. Simple bootstrap (SB) method

The SB method, also known as the basic bootstrap method, is straightforward to apply, similar to the PB method. Let $\theta$ be the quantity of interest, and let $\hat{\theta}$ represent its estimator. The SB method operates under the assumption that the distributions of $\hat{\theta} - \theta$ and $\hat{\theta}^* - \hat{\theta}$ are approximately equivalent [29]. The $(1-\alpha)100\%$ SB CI for $\theta$ is expressed in Eq. (6),

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

where $\hat{\theta}^*_{(r)}$ and $\hat{\theta}^*_{(s)}$ represent the same percentiles of the sample-based distribution of bootstrap estimates $\hat{\theta}^*$ as those used in Eq. (5) for the PB method. The SB method adjusts the interval by centering it around the observed estimate, thus aiming to improve the accuracy of the CI.

### 3.3. Bias-corrected and accelerated (BCa) bootstrap method

The BCa bootstrap CI incorporates two key adjustments: a bias-correction factor and an acceleration factor. These elements are designed to correct for bias and adjust for skewness in the distribution of bootstrap estimates. This method addresses the tendency of the PB CI to overestimate the true coverage rate, as noted in previous studies [31 - 32]. The estimator of the bias-correction factor, denoted as $\hat{z}_0$, is calculated by determining the proportion of bootstrap samples where the estimated parameter is less than the original observed estimate:

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

where $\Phi^{-1}$ is the inverse cumulative distribution function of the standard normal distribution. The estimator of the acceleration factor, denoted as $\hat{a}$, is computed through jackknife resampling, or "leave-one-out" resampling, which generates $n$ replicates of the initial sample, where $n$ is the sample size. The first jackknife replicate omits the first observation $X_1$, the second omits the second observation $X_2$, and so forth, resulting in $n$ samples, each of size $n-1$. An estimate $\hat{\theta}_{(-i)}$ is obtained for each jackknife resample, with their average denoted by $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(-i)} / n$. The estimator of the acceleration factor, $\hat{a}$, is then calculated as follow:

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

Using $\hat{z}_0$ and $\hat{a}$, the values $\alpha_1$ and $\alpha_2$ 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. The $(1-\alpha)100\%$ BCa bootstrap CI for $\theta$ is then constructed as follows:

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

where $\hat{\theta}^*_{\lceil \alpha_1 B\rceil}$ and $\hat{\theta}^*_{\lceil \alpha_2 B\rceil}$ are the corresponding quantiles from the sorted bootstrap estimates, and $\lceil\cdot\rceil$ denotes the ceiling function. When $\hat{z}_0 = 0$ and $\hat{a} = 0$, the BCa bootstrap CI is equal to the PB CI.

### 3.4. Bootstrap-t (B-t) method

Let $\theta$ be the parameter of interest. We compute an estimate $\hat{\theta}$ and an estimate of its standard error, $\hat{\sigma}_{\hat{\theta}}$ from the given data. Then, for each bootstrap sample, we calculate the bootstrap estimates $\hat{\theta}^*$ and estimated standard errors of the bootstrap estimator, $\hat{\sigma}_{\hat{\theta}^*}$. The bootstrap-t statistics are computed as:

$$t^* = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}_{\hat{\theta}^*}}.$$

The $(1-\alpha)100\%$ B-t CI for $\theta$ is then given by

$$CI_{B-t} = \left(\hat{\theta} - t^*_{1-\alpha/2}\hat{\sigma}_{\hat{\theta}}, \hat{\theta} - t^*_{\alpha/2}\hat{\sigma}_{\hat{\theta}}\right), \tag{8}$$

where $t^*_{\alpha/2}$ and $t^*_{1-(\alpha/2)}$ are the $\alpha/2$ and $1-(\alpha/2)$ quantiles of the distribution of $t^*$.

## 4. SIMULATION STUDY

This study focused on evaluating CIs for the ZTPG distribution parameter, using four distinct bootstrap methods. A Monte Carlo simulation was conducted using R version 4.2.2 [26] to cover cases with different sample sizes ($n$ = 10, 30, 50, 100, and 500) owing to the inaccessibility of a direct theoretical comparison. To examine the performance of the bootstrap methods under various conditions of data, the ZTPG parameter $\theta$ was set at different levels—0.1, 0.2, 0.5, 0.7, and 1.0 assessing a comprehensive assessment of the bootstrap methods across situations involving both small and large variances. For each situation, 1,000 bootstrap samples of size $n$ were generated from an initial pseudo-random sample, with the entire simulation process repeated 1,000 times to ensure robust results. The nominal confidence level $1-\alpha$ was fixed at 0.95 for evaluating the accuracy and reliability of the bootstrap CIs. The performance of the bootstrap CIs was evaluated based on two key criteria: coverage rate and average width. The bootstrap CI with a coverage rate near or exceeding the nominal confidence level reliably includes the true parameter value and is suitable for precise CI estimation of the parameter. Thus, a shorter average width apparently suggests that the bootstrap CI is suitable for the particular situation where the coverage rate is identical.

*Table 1.* Coverage rate and average width of the 95 % CIs for $\theta$ of the ZTPG distribution.

| $n$ | $\theta$ | Coverage rate | | | | Average width | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | PB | SB | BCa | B-t | PB | SB | BCa | B-t |
| 10 | 0.1 | 0.876 | 0.886 | 0.896 | 0.902 | 0.1399 | 0.1402 | 0.1208 | 0.1080 |
| | 0.2 | 0.866 | 0.873 | 0.881 | 0.880 | 0.3078 | 0.3083 | 0.2607 | 0.2247 |
| | 0.5 | 0.866 | 0.853 | 0.890 | 0.865 | 1.0258 | 1.0311 | 0.8239 | 0.8014 |
| | 0.7 | 0.900 | 0.830 | 0.899 | 0.858 | 1.5715 | 1.5515 | 1.3544 | 1.2417 |
| | 1.0 | 0.899 | 0.822 | 0.882 | 0.866 | 2.5515 | 2.6133 | 2.3927 | 2.2552 |
| 30 | 0.1 | 0.891 | 0.927 | 0.908 | 0.911 | 0.0698 | 0.0698 | 0.0661 | 0.0649 |
| | 0.2 | 0.914 | 0.905 | 0.922 | 0.922 | 0.1422 | 0.1422 | 0.1345 | 0.1313 |
| | 0.5 | 0.926 | 0.917 | 0.933 | 0.929 | 0.4125 | 0.4136 | 0.3835 | 0.3699 |
| | 0.7 | 0.904 | 0.902 | 0.913 | 0.914 | 0.6369 | 0.6382 | 0.5881 | 0.5586 |
| | 1.0 | 0.918 | 0.894 | 0.926 | 0.920 | 1.0256 | 1.0219 | 0.9331 | 0.8623 |
| 50 | 0.1 | 0.919 | 0.919 | 0.921 | 0.925 | 0.0519 | 0.0518 | 0.0501 | 0.0499 |
| | 0.2 | 0.942 | 0.908 | 0.935 | 0.932 | 0.1068 | 0.1067 | 0.1031 | 0.1021 |
| | 0.5 | 0.926 | 0.922 | 0.932 | 0.930 | 0.3061 | 0.3063 | 0.2937 | 0.2878 |
| | 0.7 | 0.943 | 0.917 | 0.944 | 0.938 | 0.4614 | 0.4630 | 0.4399 | 0.4287 |
| | 1.0 | 0.926 | 0.915 | 0.928 | 0.918 | 0.7369 | 0.7338 | 0.6945 | 0.6714 |
| 100 | 0.1 | 0.932 | 0.941 | 0.939 | 0.939 | 0.0358 | 0.0359 | 0.0354 | 0.0352 |
| | 0.2 | 0.944 | 0.930 | 0.944 | 0.939 | 0.0756 | 0.0757 | 0.0743 | 0.0741 |
| | 0.5 | 0.949 | 0.935 | 0.944 | 0.946 | 0.2094 | 0.2093 | 0.2049 | 0.2033 |
| | 0.7 | 0.949 | 0.943 | 0.948 | 0.947 | 0.3128 | 0.3126 | 0.3051 | 0.3023 |
| | 1.0 | 0.951 | 0.932 | 0.947 | 0.946 | 0.4911 | 0.4907 | 0.4775 | 0.4707 |
| 500 | 0.1 | 0.954 | 0.953 | 0.951 | 0.955 | 0.0159 | 0.0159 | 0.0159 | 0.0159 |
| | 0.2 | 0.937 | 0.943 | 0.940 | 0.945 | 0.0329 | 0.0329 | 0.0329 | 0.0328 |
| | 0.5 | 0.943 | 0.952 | 0.940 | 0.949 | 0.0918 | 0.0918 | 0.0915 | 0.0914 |
| | 0.7 | 0.943 | 0.952 | 0.940 | 0.949 | 0.0918 | 0.0918 | 0.0915 | 0.0914 |
| | 1.0 | 0.951 | 0.935 | 0.942 | 0.943 | 0.2116 | 0.2110 | 0.2101 | 0.2095 |

The simulation results, presented in Table 1, indicate that for $n=10,$ the coverage rates of all four bootstrap methods were generally below 0.90, failing to meet the nominal confidence level of 0.95. Despite this, the BCa bootstrap method exhibited superior performance compared to the others in these scenarios. For sample sizes of $n=30$ and 50, while all methods still fell short of the nominal confidence level, the coverage rates of the BCa bootstrap and B-t methods showed no significant differences. When the sample size increased to $n\geq 100,$ all methods achieved coverage rates near the nominal level, performing comparably well in terms of both coverage rate and average width. Conversely, a coverage rate that was similar to the nominal confidence level of 0.95 was found using the BCa bootstrap approach. As the sample size increased, the coverage rates for all bootstrap methods showed an upward trend, gradually converging towards the nominal confidence level of 0.95. Additionally, the average widths of the CIs expanded as the parameter value increased. This pattern is attributed to the direct relationship between the parameter value and the variance, where larger parameter values typically lead to greater variability, thereby resulting in wider CIs.

As expected, increasing the sample size led to a decrease in the average widths of the CIs for all four bootstrap methods. The B-t method produced the shortest average width when the sample size was small. However, this method provided a poor coverage rate markedly lower than the nominal confidence level of 0.95, indicating its inadequacy for reliable interval estimation in smaller samples. For all sample sizes, it was shown that the average widths gained using the PB and SB approaches did not differ significantly. In summary, the BCa bootstrap method performed best in terms of coverage rate even with small sample sizes.

## 5. NUMERICAL EXAMPLES

For appraising the CI for the parameter of the ZTPG distribution with two real-life samples, this study established the suitability of bootstrap methods. In addition to these two data sets, the projected bootstrap methods can be used on other count data fitted to the ZTPG distribution.

### 5.1. Demographic application

Table 2 presents data on the count of mothers with children who have experienced at least one child loss [33], based on a total sample of 135 individuals. The Chi-squared goodness-of-fit test [34] yielded a Chi-squared statistic of 3.1924 and a p-value of 0.2027, suggesting that the ZTPG distribution with parameter estimator $\hat{\theta} = 2.0011$ is appropriate for modeling this dataset. Table 3 reports the 95 % CIs for the parameter of the ZTPG distribution. These results are consistent with the Monte Carlo simulation results where the average widths of the BCa bootstrap and B-t CIs were shorter than those obtained using the PB and SB methods.

*Table 2.* The count of mothers with children who have experienced at least one child loss.

| Number of child deaths | 1 | 2 | 3 | ≥ 4 |
|---|---|---|---|---|
| Observed frequency | 89 | 25 | 11 | 10 |
| Expected frequency | 83.5907 | 32.1380 | 12.1364 | 7.1349 |

*Table 3.* The 95 % CIs for the parameter and their corresponding widths in the demographic application.

| Methods | Confidence intervals | Widths |
|---|---|---|
| PB | (1.5778, 2.7108) | 1.1330 |
| SB | (1.2878, 2.4245) | 1.1367 |
| BCa | (1.5384, 2.6671) | 1.1287 |
| B-t | (1.5431, 2.6090) | 1.0659 |

### 5.2. European red mites application

Garman [35] provided data on the European red mite number on apple leaves, presented in the second dataset of Table 4, with a total sample size of 80. The Chi-squared goodness-of-fit test [34] yielded a Chi-squared statistic of 2.3791 and a p-value of 0.4975, indicating that the ZTPG distribution, with an estimated parameter value of $\hat{\theta} = 1.1087$, is a suitable model for this dataset. The 95 % CIs for the ZTPG distribution parameter are presented in Table 5. The results correspond with the Monte Carlo simulation findings since the widths of the BCa bootstrap and B-t methods were shorter than those of the PB and SB methods.

*Table 4.* The European red mite number on apple leaves.

| European red mite number | 1 | 2 | 3 | 4 | ≥ 5 |
|---|---|---|---|---|---|
| Observed frequency | 38 | 17 | 10 | 9 | 6 |
| Expected frequency | 36.5747 | 20.2298 | 10.9617 | 5.8471 | 6.3867 |

*Table 5.* The 95 % CIs for the parameter and their corresponding widths in the European red mites application.

| Methods | Confidence intervals | Widths |
|---|---|---|
| PB | (0.8806, 1.4666) | 0.5859 |
| SB | (0.7491, 1.3368) | 0.5877 |
| BCa | (0.8669, 1.4454) | 0.5785 |
| B-t | (0.8595, 1.4129) | 0.5534 |

## 6. CONCLUSIONS AND DISCUSSION

In this study, we introduce the bootstrap methods-namely, the PB, SB, BCa bootstrap, and B-t methods-for constructing CIs for the parameter of the ZTPG distribution. The Monte Carlo simulation results indicate that sample size plays a critical role in the performance of these bootstrap CIs. Specifically, when the sample sizes were 10, 30, and 50, the coverage rates for all four methods were noticeably lower than the confidence level of 0.95, highlighting their limitations in smaller sample situations. When the sample size was sufficiently large $(n \geq 100)$, the coverage rates and average widths showed minimal differences. Based on our findings, the BCa bootstrap method demonstrated superior performance, maintaining reliable coverage rates and reasonable interval lengths even with smaller sample sizes and across various parameter settings. This robustness was evident both in the simulation study and when applied to two real-world datasets. The findings in this study offer simulation results that agreed with the study of Flowers-Cano *et al*. [27], which used a Monte Carlo Simulation to contrast the handling of four bootstrap CIs. The coverage rates of the BCa bootstrap CI were nearly constantly higher than those gained with the other CIs, as suggested by their results.

The other CIs for comparison with the anticipated bootstrap CIs should be the emphasis of future research. Of particular significance are the methods used to build the CIs for the population mean, the coefficient of variation and others. Further, no research exists concerning hypothesis testing for the ZTPG distribution parameter. For other distributions, the bootstrap CIs reviewed in this work could be used. Future studies could address these concerns.

***Acknowledgements.*** The authors would like to express their gratitude to the editor and reviewers for their valuable comments and suggestions, which have significantly contributed to improving this paper. This study was supported by the Thammasat University Research Unit in Mathematical Sciences and Applications.

***Authorship contribution statement.*** Wararit Panichkitkosolkul: Conceptualization, Methodology, Investigation, Writing - original draft preparation, Project administration, Funding acquisition. Chainarong Kesamoon: Software, Data curation. Sirichan Vesarachasart: Validation, Writing - review & editing

***Declaration of competing interest.*** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper**.**

## REFERENCES


1. Kissell R., Poserina J. - Optimal sports math, statistics, and fantasy, Academic Press, London, 2017.
2. Andrew F. S., Michael R. W. - Practical business statistics, Academic Press, London, 2022.
3. Montgomery D. C., Runger G. C. - Applied statistics and probability for engineers, John Wiley & Sons, Noida, 2003.
4. Sangnawakij P. - Confidence interval for the parameter of the zero-truncated Poisson distribution, J. Appl. Sci. **20** (2021) 13-22.
5. Hougaard P., Lee M. L. T., Whitmore G. A. - Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, Biometrics **53** (1997) 1225-1238.
6. McElduff F. C. - Models for discrete epidemiological and clinical data, Ph.D. dissertation, University College London, 2012.
7. Ong S. H., Low Y. C., Toh K. K. - Recent developments in mixed Poisson distributions, ASM Sci. J. (2021) https://doi.org/10.32802/asmscj.2020.464.
8. Tharshan R., Wijekoon P. - A new mixed Poisson distribution for over-dispersed count data: Theory and applications, Reliab.: Theory Appl. **17** (2022) 33-51.
9. Shanker R. - The discrete Poisson-Garima distribution, Biom. Biostat. Int. J. **5** (2017) 1-7.
10. Shanker R. - Garima distribution and its application to model behavioral science data, Biom. Biostat. Int. **4** (2016) 1-9.
11. Sankaran M. - The discrete Poisson-Lindley distribution, Biometrics **26** (1970) 145-149.
12. Lindley D. V. - Fiducial distributions and Bayes' theorem, J. R. Stat. Soc. B. **20** (1958) 102-107.
13. Shanker R. - Akash distribution and its applications, Int. J. Stat. Appl. **4** (2015b) 65-75.
14. Shanker R. - Aradhana distribution and its applications, Int. J. Stat. Appl. **6** (2016a) 23-34.
15. Shanker R. - Sujatha distribution and its applications, Stat. Transit. **17** (2016c) 1-20.
16. David F., Johnson N. - The truncated Poisson, Biometrics **8** (1952) 275-285.
17. Hussain T. - A zero truncated discrete distribution: Theory and applications to count data, Pak. J. Stat. Oper. Res. **16** (2020) 167-190.
18. Ghitany M. E., Al-Mutairi D. K., Nadarajah S. - Zero-truncated Poisson-Lindley distribution and its application, Math. Comput. Simul. **79** (2008) 279-287.
19. Shanker R., Hagos F. - Zero-truncated Poisson-Sujatha distribution with applications, J. Ethiopian Stat. Assoc. **24** (2015) 55-63.
20. Shanker R. - Zero-truncated Poisson-Akash distribution and its applications, Am. J. Math. Stat. **7** (2017) 227-236.

21. Shanker R., Shukla K. K. - Zero-truncated Poisson-Garima distribution and its applications, Biostat. Biom. **3** (2017) 1-6.
22. Tan S. H., Tan S. B. - The correct interpretation of confidence intervals, Proceedings of Singapore Healthcare **19** (2010) 276-278.
23. Wood M. - Statistical inference using bootstrap confidence intervals, Signif. **1** (2004) 180-182.
24. Chernick M. R., LaBudde R. A. - An introduction to bootstrap methods to R, John Wiley & Sons, New York, 2014.
25. Henningsen A., Toomet O. - maxLik: A package for maximum likelihood estimation in R, Comput. Stat. **26** (2011) 443-458.
26. Ihaka R., Gentleman R. - R: A language for data analysis and graphics, J. Comput. Graph. Stat. **5** (1996) 299-314.
27. Flowers-Cano R. S., Ortiz-Gómez R., León-Jiménez J. E., López Rivera R., Perera Cruz L. A. - Comparison of bootstrap confidence intervals using Monte Carlo simulations, Water **10** (2018) 166, https://doi.org/10.3390/w.10020166.
28. Van den Boogaard H. F. P., Hall M. J. - The construction of confidence intervals for frequency analysis using resampling techniques: A supplementary note, Hydrol. Earth Syst. Sci. **8** (2004) 1174-1178.
29. Meeker W. Q., Hahn G. J., Escobar L. A. - Statistical intervals: A guide for practitioners and researchers, John Wiley and Sons, New York, 2017.
30. Efron B. - The Jackknife, the bootstrap, and other resampling plans, in CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, SIAM, Philadelphia, 1982.
31. Efron B., Tibshirani R. J. - An introduction to the bootstrap, Chapman and Hall, New York, 1993.
32. Efron B. - Better bootstrap confidence intervals, J. Am. Stat. Assoc. **82** (1987) 171-185.
33. Shanker R. Fesshaye H., Selvaraj S., Yemane A. - On zero-truncation of Poisson and Poisson-Lindley distributions and their applications, Biom. Biostat. Int. J. **2** (2015) 168-181.
34. Turhan N. S. - Karl Pearson’s chi-square tests, Educ. Res. Rev. **15** (2020) 575-580.
35. Garman P. - The European red mites in Connecticut apple orchards, Connecticut. Agri. Exper. Station. Bull. **252** (1923) 103-125.