Current Applied Science and Technology Vol. 24 No. 5

Research article

Bootstrap Confidence Intervals for the Poisson-Pranav Distribution Parameter with an Application to COVID-19 Data

Wararit Panichkitkosolkul*

Department of Mathematics and Statistics, Thammasat University,Pathum Thani, Thailand

Curr. Appl. Sci. Technol. 2024, Vol. 24 (No. 5), e0260107; https://doi.org/10.55003/cast.2024.260107

Received: 15 September 2023, Revised: 9 February 2024, Accepted: 25 March 2024, Published: 2 May 2024

Abstract

The Poisson distribution is commonly applied in statistical modeling to represent the count of events, assuming that the events are independent and occur at a constant rate. This assumption, however, may not always hold true in real-life situations. The Poisson distribution may be inadequate if the underlying rate of occurrence is not constant. The mixed Poisson distribution is proposed to solve this limitation because it permits the rate parameter of the Poisson distribution to be random rather than fixed. The Poisson-Pranav distribution, which is classified as a type of mixed Poisson distribution, has been commonly utilized to analyze count data that exhibits over-dispersion over many domains. However, no research has been done into building confidence intervals for the parameter of the Poisson-Pranav distribution using the bootstrap method. Using Monte Carlo simulation, the coverage probabilities and average lengths of the percentile bootstrap (PB), basic bootstrap (BB), and biased-corrected and accelerated (BCa) bootstrap's interval-estimation performances were compared. The bootstrap method was not appropriate for achieving the desired nominal confidence level with small sample sizes. In addition, the performance of the bootstrap confidence intervals did not differ significantly when the sample size was increased considerably. In each case studied, the BCa bootstrap confidence intervals performed better than the others. The effectiveness of bootstrap confidence intervals was demonstrated by applying them to the number of COVID-19 deaths in Belgium. The computations substantially supported the proposed bootstrap confidence intervals.

Keywords

estimation;

discrete distribution;

count data;

mixture distribution;

resampling method

---

*Corresponding author: Tel.: (+66) 5644440-59 ext. 2101 Fax: (+66) 5644489
E-mail: wararit@mathstat.sci.tu.ac.th

1. Introduction

The Poisson distribution is a discrete distribution that measures the probability of a given number of events occurring in specific regions of time or space [1, 2]. This distribution is used to describe data such as the number of confirmed COVID-19 cases per day, the number of COVID-19 deaths per week, etc. The Poisson distribution is frequently employed to model the number of occurrences under the assumption that events are independent and occur at a constant rate. In actual situations, however, this assumption may not always hold true. The Poisson distribution may be inadequate if the underlying rate of occurrence is not constant. This limitation is overcome by the use of mixed Poisson distribution, which permits the rate parameter of the Poisson distribution to be random rather than fixed. In a mixed Poisson distribution, the parameter controlling the mean rate of events can itself be a random variable, following a certain distribution. This allows for flexibility in modeling scenarios where the rate of events varies among different subpopulations or over time.

Recent work by Shukla and Shanker [3] investigated the mathematical and statistical properties of the Poisson-Pranav (PP) distribution, which is a type of mixed Poisson distribution. Combining the Poisson distribution and the Pranav distribution [4] yields this distribution. When the Poisson parameter λ follows a Pranav distribution, the Poisson distribution gives rise to the PP distribution [3]. When applied to two real-world data sets, the PP distribution was the most appropriate choice when compared to the Poisson, Poisson-Lindley [5], Poisson-Akash [6], and Poisson-Ishita [7] distributions. The Pranav distribution was first proposed by Shukla [4] as a continuous distribution for modeling lifetimes. It is characterized by a probability density function (pdf) denoted as

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

It is a composite of an exponential distribution with a scale parameter and a gamma distribution with a shape parameter 4 and a scale parameter θ, with respective proportions $\theta^4/(\theta^4+6)$ and $6/(\theta^4+6)$. Data on lifetimes in engineering and medicine have been modeled using this distribution. Shukla [4] demonstrated that the Pranav distribution was a more accurate representation of reality when compared to other distributions such as the exponential, Lindley [8], Akash [9], Ishita [10], Sujatha [11], and Shanker [12] distributions. The key statistical properties of the Pranav distribution were also discussed by Shukla [4]. In this paper, Figure 1 depicts plots of the pdf of the Pranav distribution for various values of parameter θ.

A confidence interval (CI) is a range of values that contain the true parameter of population. It is the primary output for statistical inferences and plays a crucial role in interpreting parameter estimates. In the literature reviews, several studies applied CIs to real-world applications. In 2022, Chankham *et al*. [13] suggested CIs for the coefficient of variation of an inverse Gaussian distribution. These CIs were the generalized CI, the adjusted generalized CI, the bootstrap percentile CI, the fiducial CI, and the fiducial highest posterior density CI. The efficacy of these CIs was demonstrated using PM 2.5 data from Bangkok, Thailand. Ye *et al*. [14] introduced CIs for variance component functions in the skew-normal unbalanced heteroscedastic one-way classification random effects model. These CIs were obtained using both the bootstrap technique and the generalized approach. The approaches mentioned were illustrated with two real examples of the annual average concentrations of fine particulate matter and nitrogen dioxide.

According to the literature review, no research has been conducted to estimate the CI for the PP distribution parameter. Bootstrap methods provide a robust approach to quantifying

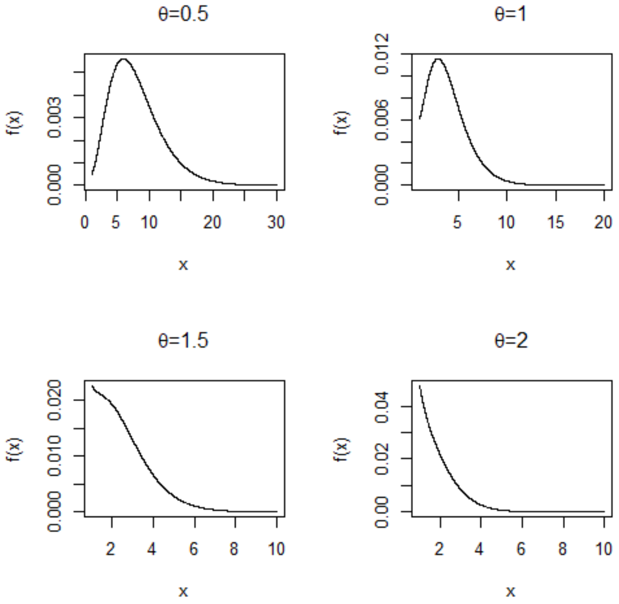

Figure 1. Plots of the pdf of the Pranav distribution for θ = 0.5, 1, 1.5 and 2

uncertainty and constructing CIs without relying on stringent distributional assumptions. These approaches allow for the quantification of uncertainty in statistical inference by utilizing resampling techniques on a given sample of data. Several bootstrap CIs are a powerful and widely used method for estimating CIs for the parameter, and it often performs well in practice. The use of the bootstrap method, in particular, may not be the best approach for all situations. It depends on the specific characteristics of the data and the parameter of interest in each case [15]. Therefore, it is essential to propose bootstrap CIs and evaluate their performances when the data follow the PP distribution. This study examines three different approaches for constructing bootstrap CIs: the percentile bootstrap (PB), the basic bootstrap (BB), and the bias-corrected and accelerated (BCa) bootstrap.

2. Materials and Methods

2.1 Point parameter estimation

Let X be a random variable from the population that has a PP distribution with parameter θ. The probability mass function (pmf) of the PP distribution was identified by Shukla and Shanker [3] as

$$p(x;\theta)=\frac{\theta^4}{\theta^4+6}\frac{\left[x^3+6x^2+11x+\theta(\theta+1)^3+6\right]}{(\theta+1)^{x+4}},\quad x=0,1,2,\ldots,\quad \theta>0. \tag{2}$$

Figure 2 displays pmf plots of the PP distribution for a range of θ values. The expected values such as mean, variance, and dispersion index of the random variable X are as follows:

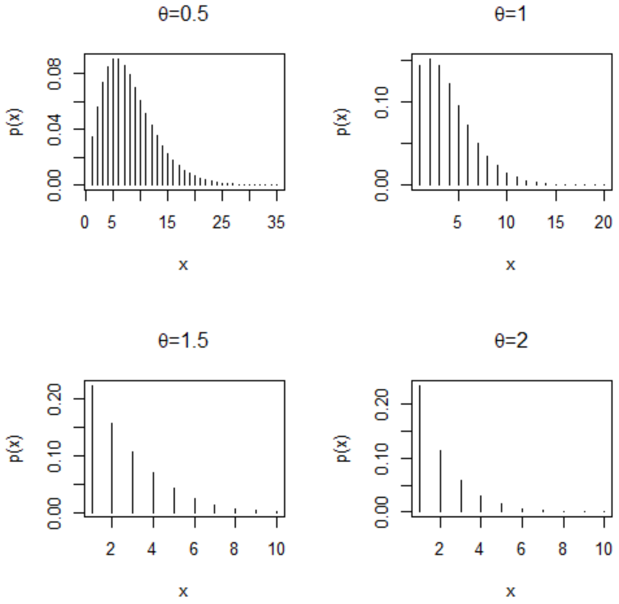

Figure 2. Plots of the pmf of the PP distribution for θ = 0.5, 1, 1.5 and 2

$$E(X)=\mu=\frac{\theta^4+24}{\theta\left(\theta^4+6\right)},\quad \operatorname{var}(X)=\sigma^2=\frac{\left(\theta^9+\theta^8+30\theta^5+840\theta^4+144\theta+144\right)}{\theta^2\left(\theta^4+6\right)^2},\ \text{and}$$

$$DI(X)=\frac{\operatorname{var}(X)}{E(X)}=\tau=\frac{\left(\theta^9+\theta^8+30\theta^5+840\theta^4+144\theta+144\right)}{\theta\left(\theta^4+6\right)\left(\theta^4+24\right)}.$$

Maximizing the log-likelihood function $\log L(x_i;\theta)$ yields the point estimator of θ. As a result, the following procedures are used to arrive at the ML estimator for θ:

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L(x_i;\theta)&=\frac{\partial}{\partial\theta}\left[n\log\left(\frac{\theta^4}{\theta^4+6}\right)-\sum_{i=1}^{n}(x_i+4)\log(\theta+1)+\sum_{i=1}^{n}\log\left(x_i^3+6x_i^2+11x_i+\theta(\theta+1)^3+6\right)\right]\\&=\frac{12n}{\theta(\theta^4+6)}-\frac{n(\overline{x}+4)}{\theta+1}+\sum_{i=1}^{n}\frac{4\theta^3+9\theta^2+6\theta+1}{\left[x_i^3+6x_i^2+11x_i+\theta(\theta+1)^3+6\right]}.\end{aligned}$$

We obtain the non-linear equation by solving for θ in the equation $\frac{\partial}{\partial\theta}\log L(x_i;\theta)=0$:

$$\frac{12n}{\theta(\theta^4+6)}-\frac{n(\overline{x}+4)}{\theta+1}+\sum_{i=1}^{n}\frac{4\theta^3+9\theta^2+6\theta+1}{\left[x_i^3+6x_i^2+11x_i+\theta(\theta+1)^3+6\right]}=0,$$

where $\overline{x}=n^{-1}\sum_{i=1}^{n}x_i$. Due to the lack of a closed-form solution provided by the ML estimator for θ, numerical iteration techniques can be used to solve the resulting non-linear problem. In order to estimate ML parameters, this study utilized the statistical program R [16] and its maxLik package [17] along with the Newton-Raphson technique.

2.2 Bootstrap confidence intervals

Confidence intervals (CIs) are derived using a parametric estimator that estimates the standard errors of a quantity of interest, denoted as ϕ. The two-sided CI for ϕ is derived by subtracting or adding the standard error multiplied by a critical value. The assumption made in this calculation is that the estimator of ϕ follows an approximately normal distribution [18]. Nonetheless, there are a number of instances in which the assumption of normality is violated. The consequence is that CIs may be wider or narrower than they should be, leading to incorrect levels of confidence in the estimates. When faced with these scenarios, or when estimating the standard error presents significant difficulties, an alternative approach involves utilizing techniques founded on the bootstrap method [19]. The bootstrap methods discussed in this study offer a different method to make approximate CIs without making assumptions about the underlying distribution [20]. Furthermore, the score function of PP distribution is complicated, and the maximum likelihood has no closed form. Therefore, likelihood-based, and score CIs have no closed forms. In such cases, finding the CIs can be challenging; alternative methods, such as numerical techniques or resampling methods like the bootstrap, can be utilized.

This paper focuses on the three bootstrap CIs for the PP distribution parameter. In practice, the most prevalent bootstrap CIs are the percentile bootstrap (PB), the basic bootstrap (BB), and the bias-corrected and accelerated (BCa) bootstrap CIs [21]. In this investigation, the statistical software R's boot package [22] was used to estimate the bootstrap CIs.

2.2.1 Percentile bootstrap (PB) confidence interval

The PB CI is a non-parametric method that estimates the uncertainty surrounding a population parameter by resampling the original sample. It is particularly useful when the underlying distribution of the data is unknown or complicated [23]. The stages for obtaining a PB CI for θ are as follows:

1) Collect the sample data: Begin with the initial sample data, which represents a subset of the population. Consider that the sample contains n observations.

2) Resampling with replacement: The bootstrap method involves resampling from the original sample with replacement. The n observations are selected from the original sample, allowing for the possibility that the same observation can be selected multiple times.

3) Calculate the statistics: For each bootstrap sample, the statistics of interest (e.g., parameter, mean, median, etc.) is calculated. A distribution of the statistic under repeated resampling is obtained.

4) Arrange bootstrap statistics: In order to construct a CI, it is necessary to arrange the bootstrap statistics in ascending order and afterwards select the relevant percentiles. For example, if we wanted a 95% confidence interval, we would select the 2.5th percentile as the lower bound and the 97.5th percentile as the upper bound. The $(1-\alpha)100\%$ two-sided PB CI for θ is created as follows:

$$CI_{PB} = \left[\hat{\theta}^*_{(r)}, \hat{\theta}^*_{(s)}\right], \tag{3}$$

where the notation $\hat{\theta}^*_{(r)}$ is the r^{th} quantile of a collection of the parameter estimate $\hat{\theta}^*$ arranged in ascending order, while $\hat{\theta}^*_{(s)}$ is the s^{th} quantile of the aforementioned collection, $r = \lceil (\alpha/2)B \rceil$, $s = \lceil (1-(\alpha/2))B \rceil$, where $\lceil x \rceil$ stands for the ceiling function of x, and $1-\alpha$ is the confidence level. This study utilized $\alpha = 0.05$ and $B = 2{,}000$; the two quantiles related to the lower and upper bounds of the PB two-sided CI were $\hat{\theta}^*_{(r)} = \hat{\theta}^*_{(50)}$ (the 50th quantile) and $\hat{\theta}^*_{(s)} = \hat{\theta}^*_{(1950)}$ (the 1950th quantile).

2.2.2 Basic bootstrap (BB) confidence interval

The BB CI is a straightforward approach that does not involve more complex modifications or adjustments to the bootstrap process. It is sometimes referred to as the simple bootstrap CI. The BB CI focuses on the variability of the statistic itself rather than explicitly considering the distribution's tails. Assume that we consider the parameter θ and the estimator of θ is $\hat{\theta}$. When $\hat{\theta}^*$ is the bootstrap estimate of θ based on the bootstrap sample, the BB CI implies that the distributions of $\hat{\theta}-\theta$ and $\hat{\theta}^*-\hat{\theta}$ are roughly equivalent [20]. The $(1-\alpha)100\%$ two-sided BB CI for θ is

$$CI_{BB} = \left[2\hat{\theta}-\hat{\theta}^*_{(s)},\, 2\hat{\theta}-\hat{\theta}^*_{(r)}\right], \tag{4}$$

where the quantiles $\hat{\theta}^*_{(r)}$ and $\hat{\theta}^*_{(s)}$ represent the same percentile of the empirical distribution of bootstrap estimates $\hat{\theta}^*$ that are utilized in Equation (3) to calculate the PB CI.

2.2.3 Bias-corrected and accelerated (BCa) bootstrap confidence ınterval

The BCa bootstrap CI is an advanced technique used to improve the accuracy of PB and BB CI estimations when dealing with small sample sizes or when the distribution of the data is skewed. With a correction for bias and a correction for skewness, the BCa bootstrap CI corrects for both bias and skewness in the distribution of the bootstrap statistics [23]. This reduces several problems that come up with the PB and BB CIs, such as bias in the estimate of the population parameter, inaccurate coverage, and narrow or unreliable confidence intervals. The calculation of BCa bootstrap CI commonly involves the utilization of influence statistics derived from jackknife simulations.

Davison and Hinkley [24] and Chernick and LaBudde [21] described the mathematical particulars of the BCa adjustment. The bias correction factor, denoted $\hat{z}_0$, is as follows: $\hat{z}_0 = \Phi^{-1}\left(\sum_{i=1}^{B} I\left(\hat{\theta}^*_i < \hat{\theta}\right)/B\right)$, where $\Phi^{-1}(\cdot)$ is the inverse function of the cumulative standard normal distribution and $I(\cdot)$ is the indicator function. The skewness or acceleration adjustment is calculated via jackknife resampling, which entails generating n replicates of the initial set of data, where n is the sample size. The initial jackknife replication is obtained by omitting the first case $(i=1)$ from the initial sample, the second by omitting the second case $(i=2)$, etc., until a total of n samples, each with a size of $n-1$, are generated. From jackknife replicates, we obtain the value of $\hat{\theta}_{(-i)}, i = 1, 2, \ldots, n.$ The acceleration factor $\hat{a}$ is given by $\hat{a} = \dfrac{\sum_{i=1}^{n} \left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(-i)} \right)^3}{6 \left\{ \sum_{i=1}^{n} \left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(-i)} \right)^2 \right\}^{3/2}}$, where $\hat{\theta}_{(\cdot)} = n^{-1} \sum_{i=1}^{n} \hat{\theta}_{(-i)}$. The values of α_1 and α_2 are calculated with the values of $\hat{z}_0$ and $\hat{a}$, $\alpha_1 = \Phi \left\{ \hat{z}_0 + \dfrac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}\left(\hat{z}_0 + z_{\alpha/2} \right)} \right\}$ and $\alpha_2 = \Phi \left\{ \hat{z}_0 + \dfrac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}\left(\hat{z}_0 + z_{1-\alpha/2} \right)} \right\}$, where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution. Then, the $(1-\alpha)100\%$ two-sided BCa bootstrap CI for θ is given in equation (5):

$$CI_{BCa} = \left[\hat{\theta}^*_{(j)}, \hat{\theta}^*_{(k)} \right], \tag{5}$$

where $j = \lceil \alpha_1 B \rceil$ and $k = \lceil \alpha_2 B \rceil$. The stages of construction for the bootstrap CIs are shown in Figure 3.

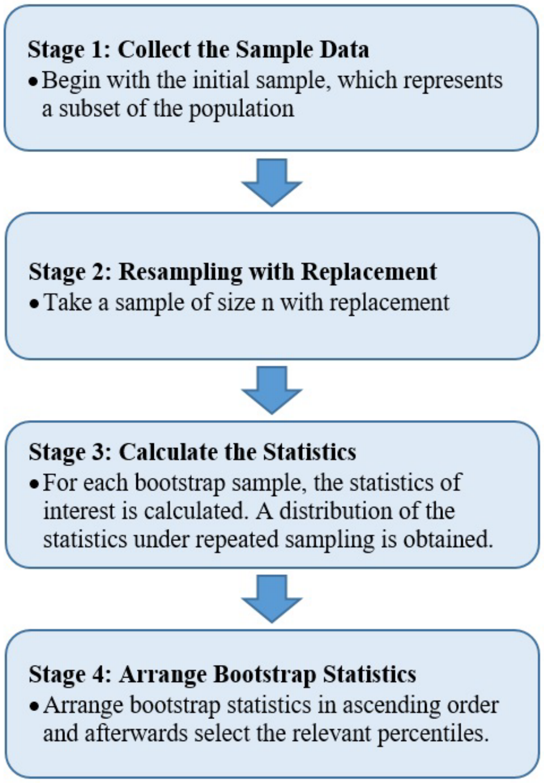

Figure 3. Stages of construction for the bootstrap CI

3. Results and Discussion

3.1 Simulation study and results

In this study, a two-sided CI estimation for the PP distribution parameter was performed using three bootstrapping methods, which were illustrated in equations (3)-(5). By using statistical software R, a Monte Carlo simulation was designed to encompass cases for n = 10, 30, 50, 100, and 500. The true parameter (θ) was set to 0.1, 0.3, 0.5, 0.8, 1.0, 1.5 and 2.0. The number of bootstrap samples (B) was fixed at 2,000, as per the study of Ukoumunne *et al*. [25]. The nominal confidence level $(1-\alpha)$ was 0.95 and the number of replications was set to 1,000. The performances of the bootstrap CIs were compared with the estimated coverage probabilities and average lengths. In this study, we concluded that a coverage probability was greater than or equal to the nominal confidence level when the estimated coverage probability was greater than or equal to 0.939 by using the one-proportion z-test with a significance level of 0.05. Moreover, the bootstrap CI with the minimum average length could be used to estimate the parameter more precisely. The R source code for the simulation study is available at https://codeocean.com/capsule/4346018/tree.

The simulation study's findings are presented in Table 1. For $n=10$, the estimated coverage probabilities of all three bootstrap CIs were found to be below 0.90, suggesting that they did not achieve the nominal confidence level of 0.95. In these circumstances, however, the BCa bootstrap CI demonstrated superior performance compared to the other methods. For $n=30$, all bootstrap CIs again yielded estimated coverage probabilities lower than 0.95. For $n\geq 50$, each bootstrap CI obtained coverage probabilities near the nominal confidence level of 0.95 and had average lengths that were comparable. As a simulation result, the estimated coverage probabilities of the bootstrap CIs tended to rise and get closer to 0.95 as the sample size rose. Additionally, as θ increased, the average lengths of the bootstrap CIs tended to increase. This can be attributed to the relationship between the variance and θ. As expected, the average lengths of all the bootstrap CIs decreased as the sampler size increased, with the BCa bootstrap CI offering the shortest average length across all cases. In addition, when the sample size was small (n = 10), there was a significant difference in the average length of the BCa bootstrap CI compared to the others. In brief, the BCa bootstrap CI demonstrated superior performance when applied to moderate and large sample sizes $(n\geq 50)$.

3.2 Application to real data

Coronaviruses encompass a vast array of viruses capable of inducing a diverse range of illnesses, spanning from mild ailments like the common cold to more severe afflictions. The emergence of a novel coronavirus (COVID-19) was seen in the city of Wuhan, China, during the year 2019. This was an extremely novel coronavirus that had never been detected in humans. The World Health Organization (WHO) declared that the coronavirus disease 2019 (COVID-19) was a pandemic [26]. To prevent the virus from proliferating further, a global effort was necessary. The pandemic affected a large geographical area and an exceptionally high proportion of the population. The H1N1 flu pandemic, which was documented in 2009, was an earlier global pandemic.

Many scientists have conducted research on the COVID-19 pandemic, and they have developed models to analyze data and provide predictions regarding the expected number of cases. The aim of using these models is to assist nations in making informed decisions regarding prevention strategies. For example, Maleki *et al*. [27] employed an autoregressive time series model utilizing a

Table 1. Estimated coverage probability and average length of the 95% two-sided bootstrap CIs for the PP distribution parameter

n	θ	Coverage Probability			Average Length		
		PB	**BB**	**BCa**	**PB**	**BB**	**BCa**
10	0.1	0.913	0.903	0.910	0.0656	0.0666	0.0800
	0.3	0.881	0.882	0.886	0.2170	0.2167	0.2122
	0.5	0.892	0.856	0.881	0.3666	0.3667	0.3619
	0.8	0.903	0.866	0.891	0.5488	0.5492	0.5509
	1.0	0.899	0.902	0.895	0.6779	0.6775	0.6765
	1.5	0.938	0.923	0.947*	1.4531	1.4244	1.3298
	2.0	0.931	0.878	0.920	3.2792	3.3833	3.1137
30	0.1	0.925	0.941*	0.936	0.0378	0.0378	0.0373
	0.3	0.925	0.923	0.925	0.1229	0.1228	0.1212
	0.5	0.927	0.907	0.927	0.2162	0.2162	0.2133
	0.8	0.921	0.922	0.920	0.3308	0.3309	0.3291
	1.0	0.930	0.925	0.932	0.3902	0.3902	0.3893
	1.5	0.950*	0.954*	0.958*	0.6034	0.6036	0.5908
	2.0	0.925	0.914	0.940	1.2785	1.2860	1.1632
50	0.1	0.945*	0.938	0.946*	0.0291	0.0290	0.0288
	0.3	0.948*	0.939*	0.949*	0.0947	0.0948	0.0938
	0.5	0.938	0.914	0.939*	0.1641	0.1642	0.1630
	0.8	0.951*	0.941*	0.947*	0.2595	0.2587	0.2580
	1.0	0.941*	0.945*	0.940*	0.3064	0.3065	0.3061
	1.5	0.950*	0.965*	0.953*	0.4514	0.4518	0.4465
	2.0	0.939*	0.944*	0.952*	0.8293	0.8276	0.7815
100	0.1	0.950*	0.953*	0.951*	0.0205	0.0204	0.0204
	0.3	0.955*	0.953*	0.954*	0.0675	0.0674	0.0671
	0.5	0.968*	0.951*	0.963*	0.1171	0.1172	0.1166
	0.8	0.937	0.934	0.937	0.1855	0.1854	0.1854
	1.0	0.953*	0.946*	0.957*	0.2186	0.2188	0.2186
	1.5	0.945*	0.940*	0.946*	0.3109	0.3110	0.3093
	2.0	0.928	0.951*	0.945*	0.5430	0.5432	0.5300
500	0.1	0.939*	0.946*	0.943*	0.0092	0.0092	0.0092
	0.3	0.953*	0.961*	0.957*	0.0300	0.0300	0.0300
	0.5	0.949*	0.952*	0.943*	0.0527	0.0527	0.0527
	0.8	0.959*	0.956*	0.955*	0.0832	0.0829	0.0831
	1.0	0.953*	0.953*	0.951*	0.0985	0.0987	0.0985
	1.5	0.944*	0.945*	0.940*	0.1376	0.1378	0.1374
	2.0	0.956*	0.953*	0.954*	0.2265	0.2266	0.2259

* indicates that the empirical coverage probability is greater than or equal to the nominal confidence level.

two-piece scale mixture normal distribution to forecast and validate COVID-19 cases. Nesteruk [28] examined the daily incidence of COVID-19 cases in China through the application of the mathematical model (Susceptible, Infected, and Recovered model). Almongy *et al*. [29] proposed a novel model for analyzing the mortality rates due to COVID-19 in Italy, Mexico, and the Netherlands. Iranzo and Pérez-González [30] proposed two kinds of epidemiological models invoked during the COVID-19 crisis. Ega and Ngeleja [31] developed a deterministic mathematical model that incorporated compartments for both human-to-human transmission and transmission via contaminated surfaces.

Using an actual data set, the application of bootstrap CIs for the parameter of the PP distribution was demonstrated in this study. The data represents the number of COVID-19 daily deaths in Belgium from December 1, 2022, to March 31, 2023 (https://lookerstudio.google.com/embedreporting/c14a5cfc-cab7-4812-848c-0369173148ab/page/ZwmOB). The total number of observations is 90 and they are listed below:

1 3 3 10 9 6 6 5 6 9 10 5 18 10 9 9 9 13 11 15 11 13 13 13 13 11 18 8
16 17 10 6 12 13 12 10 9 12 5 13 10 8 9 5 9 4 3 5 6 4 5 1 5 3 4 8 4 6
0 3 2 6 4 2 4 9 9 7 3 6 6 8 3 7 5 10 7 7 9 5 11 12 6 6 12 7 7 10 7 10

Table 2 presents the descriptive statistics for the given data set.

Table 2. Descriptive statistics of the number of COVID-19 daily deaths in Belgium

Min	Mean	Median	Var	Skewness	Kurtosis	Q1	Q3	Max
0.0	7.9	7.5	15.37	0.4037	2.8026	5.0	10.0	18.0

The distribution comparison is made between the fitting performance of the PP distribution and the following mixed Poisson distributions:

- For the Poisson-Adya distribution [32], its pmf is

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

- For the Poisson-Akash distribution [6], Its pmf is

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

- For the Poisson-Aradhana distribution [33], its pmf is

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

- For the Poisson-Garima [34] distribution, its pmf is

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

- For the Poisson-Lindley distribution [5], its pmf is

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

- For the Poisson-Shanker distribution [35], its pmf is

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

- For the Poisson-Sujatha distribution [36], its pmf is

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

The estimation of the parameters of the mixed Poisson distributions was performed using the ML approach. We considered the log-likelihood (log L), Akaike's information criterion (AIC) [37], and Bayesian information criterion (BIC) [38] for model comparison. In Table 3, estimates of the parameters, their standard errors (SE), and measures of goodness of fit are provided for this data set.

The AIC and BIC values in Table 3 illustrate that the Poisson-Pranav (PP) distribution provided an adequate fit to the number of COVID-19 daily deaths as compared with other distributions. The dispersion index (DI) for this data was 3.6942. The dataset demonstrated the phenomenon of over-dispersion because the DI was greater than 1 [39]. Table 4 presents the 95% two-sided bootstrap CIs for the PP distribution parameter. We concluded that the PP distribution parameter lay between 0.4547 and 0.5522, based on the BCa bootstrap CI. Therefore, it was inferred that the interval (7.1608, 8.75033) contained the average or population mean (μ) of the number of daily deaths due to COVID-19. Additionally, the correspondence between the application to COVID-19 data and the simulation results was observed, as the BCa bootstrap CI exhibited the shortest length in comparison to other intervals.

Table 3. The ML estimates, SE, and goodness of fit measures for the number of COVID-19 daily deaths in Belgium

Distributions	Estimates (SE)	Log L	AIC	BIC
Poisson-Pranav	0.5013 (0.0317)	-251.6858	505.3716	507.8714
Poisson-Adya	0.3602 (0.0245)	-256.1517	514.3034	516.8032
Poisson-Akash	0.3622 (0.0250)	-256.7396	515.4792	517.9790
Poisson-Aradhana	0.3370 (0.0235)	-258.6242	519.2484	521.7482
Poisson-Garima	0.1906 (0.0186)	-275.5591	553.1182	555.6180
Poisson-Lindley	0.2285 (0.0189)	-266.6528	535.3056	537.8054
Poisson-Shanker	0.2435 (0.0196)	-263.3568	528.7136	531.2134
Poisson-Sujatha	0.3476 (0.0241)	-257.8473	517.6946	520.1944

Table 4. The 95% two-sided bootstrap CIs and lengths for the parameter in the number of COVID-19 daily deaths in Belgium

Methods	Confidence Intervals	Lengths
PB	(0.4548, 0.5540)	0.0992
BB	(0.4484, 0.5461)	0.0977
BCa	(0.4547, 0.5522)	0.0975

3.3 Discussion

Based on the simulation results, all three bootstrap CIs functioned well in all scenarios with large sample sizes ($n \geq 50$). Average lengths did not significantly change, and coverage probabilities were close to or near the nominal confidence level. However, when dealing with small sample sizes ($n = 10$ and 30), all three bootstrap CIs displayed coverage probabilities that were below the nominal confidence level. For the average length comparison, the BCa bootstrap CI provided the shortest average length in almost all cases. Furthermore, when the sample size increased and θ became smaller, the average lengths of all the bootstrap CIs decreased. The results of this study are similar to those of previous studies. This is because there are theoretical reasons to preference of BCa bootstrap CIs. Moreover, the BCa bootstrap CIs provided the shortest average lengths, so they can be used to estimate CIs for the parameter of the PP distribution in multiple domains with precision and effectiveness. Our methodology has the potential to assist physicians and government health agencies in tracking the daily mortality rate of COVID-19 in some countries and issuing alerts when the daily death toll of COVID-19 reaches defined thresholds. Furthermore, it is important to note that while knowing the number of COVID-19 deaths is crucial, it should be considered in conjunction with other data, such as the number of cases, hospitalizations, and vaccination rates in order to get a comprehensive understanding of the pandemic's status. Our findings provide valuable insights into estimating the parameters, including the population mean of the number of COVID-19 daily deaths, which can facilitate more effective pandemic response strategies such as better COVID-19 vaccine management and improved ways of dealing with the future outbreaks of COVID-19.

This study was limited by the fact that none of the bootstrap CIs were exact. However, it is worth noting that these intervals consistently converged towards a coverage probability of $1-\alpha$ as the sample sizes increased. With a small sample size, we have limited information about the underlying population. This means that resampling from a small dataset may not adequately capture the variability present in the population, leading to biased or unreliable estimates. Bootstrap CIs may end up too narrow or too wide, leading to incorrect inferences about the parameter. In addition, three bootstrap CIs are computationally intensive and difficult to calculate. Nevertheless, there exists a variety of R packages that can be utilized for the computation of bootstrap CIs. These packages include the boot package [22], the bootstrap package [40], the semEff package [41], and the BootES package [42]. Since statistical software R is an open-source programming language, users are allowed to download the above-mentioned packages.

4. Conclusions

In the previous studies, there was no research that studied the bootstrap CIs for the parameter of the Poisson-Pranav (PP) distribution. In this study, we aimed to evaluate the performance of percentile bootstrap (PB), basic bootstrap (BB), and bias-corrected and accelerated (BCa) bootstrap approaches for the PP distribution parameter. The advantages of bootstrap CIs are their robustness and flexibility, and their ability to make inferences without assuming a specific data distribution. They work well with non-Gaussian data, and in situations where traditional parametric methods are inappropriate. Additionally, they provide a model-free approach to estimate the sampling distribution of statistics and are relatively simple to implement. The performance of these approaches was evaluated by comparing their estimated coverage probability and average length using simulated data. The results show that the sample size (n) appeared to have a substantial impact on the performances of the bootstrap CIs. The estimated coverage probabilities for three bootstrap CIs were significantly less than 0.95 in case of the small sample sizes $(n=10$ and 30). When the sample size reached a sufficient magnitude $(n \geq 50),$ there were no significant disparities observed in the estimated coverage probabilities and average lengths of the three bootstrap CIs. Based on our research outcomes, it was observed that the BCa bootstrap CI exhibited superior performance across a wide range of scenarios, as evidenced by both the simulation study and the application to COVID-19 data. Thus, BCa bootstrap CI can be recommended for constructing CIs for the PP distribution parameter in cases of large sample sizes.

In further studies, it would be beneficial to focus on the comparative analysis of alternative CI estimations when compared to the bootstrap CIs presented in this research. The construction of CIs for functions of other parameters, such as the population mean and dispersion index, is of interest. Additionally, there is a lack of statistical theoretical research regarding hypothesis testing for the PP distribution parameter. The bootstrap CIs studied in this paper can be applied to other distributions. These topics may be subject to further investigation in subsequent studies. Moreover, the number of deaths for other epidemics should be studied by using the bootstrap CIs for the parameter of the PP distribution.

References

[1] Kissell, R. and Poserina, J., 2017. *Optimal Sports Math, Statistics, and Fantasy.* London: Academic Press, pp. 71-89.

[2] Siegel, A.F. and Wagner M.R., 2022. *Practical Business Statistics*. San Diego: Academic Press, pp. 190-192.

[3] Shukla, K.K. and Shanker, R., 2022. On statistical properties and applications of Poisson-Pranav distribution. *Biometrics and Biostatistics International Journal*, 11(3), 93-98.

[4] Shukla, K.K., 2018. Pranav distribution with properties and its applications. *Biometrics and Biostatistics International Journal*, 7(3), 244-254.

[5] Sankaran, M., 1970. The discrete Poisson-Lindley distribution. *Biometrics,* 26(1), 145-149, https://doi.org/10.2307/2529053.

[6] Shanker, R., 2017. The discrete Poisson Akash distribution. *International Journal of Probability and Statistics*, 6(1), 1-10, https://doi.org/10.5923/j.ijps.20170601.01.

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

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

[9] Shanker, R., 2015. Akash distribution and its applications. *International Journal of Probability and Statistics*, 4(3), 65-75, https://doi.org/10.5923/j.ijps.20150403.01.

[10] Shanker, R. and Shukla, K.K., 2017. Ishita distribution and its applications. *Biometrics and Biostatistics International Journal*, 5(2), 39-46.

[11] Shanker, R., 2016. Sujatha distribution and its applications. *Statistics in Transition-New Series*, 17(3), 391-410.

[12] Shanker, R., 2015. Shanker distribution and its applications. *International Journal of Statistics and Applications*, 5(6), 338-348.

[13] Chankham, W., Niwitpong, S.-A., and Niwitpong, S., 2022. Measurement of dispersion of PM 2.5 in Thailand using confidence intervals for the coefficient of variation of an inverse Gaussian distribution. *PeerJ*, 10, http://doi.org/10.7717/peerj.12988.

[14] Ye, R., Du, W. and Lu, Y., 2023. Bootstrap inference for skew-normal unbalanced heteroscedastic one-way classification random effects model. *Journal of Statistical Computation and Simulation*, 93(15), 2672-2702.

[15] InfluentialPoints.com, 2023. *Bootstrap Confidence Intervals*. [online] Available at: https://influentialpoints.com/Training/bootstrap_confidence_intervals-principles-properties-assumptions.htm.

[16] Ihaka, R. and Gentleman, R., 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314, https://doi.org/10.2307/1390807.

[17] Henningsen, A. and Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3), 443-458, https://doi.org/10.1007/s00180-010-0217-1.

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

[19] van den Boogaard, H.F.P. and Hall, M.J., 2004. The construction of confidence intervals for frequency analysis using resampling techniques: a supplementary note. *Hydrology and Earth System Sciences*, 8(6), 1174-1178, https://doi.org/10.5194/hess-8-1174-2004.

[20] Meeker, W.Q., Hahn, G.J. and Escobar, L.A., 2017. *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. New Jersey: John Wiley and Sons, pp. 245-266.

[21] Chernick, M.R. and LaBudde, R.A., 2011. *An Introduction to Bootstrap Methods with R.* Singapore: John Wiley and Sons, pp. 76-98.

[22] Canty, A. and Ripley, B., 2022. *Boot: Bootstrap R (S-Plus) Functions (Version 1.3-28.1)* [online] Available at: https://cran.r-project.org/web/packages/boot/boot.pdf.

[23] Efron, B., 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

[24] Davison, A.C. and Hinkley, D.V., 1997. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press, pp. 203-208.

[25] Ukoumunne, O.C., Davison, A.C., Gulliford, M.C. and Chinn, S., 2003. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. *Statistics in Medicine,* 22(24), 3805-3821, https://doi.org/10.1002/sim.1643.

[26] Cucinotta, D., and Vanelli, M., 2020. WHO declares COVID-19 a pandemic. *Acta Biomedica*, 91(1), 157-160, https://doi.org/10.23750/abm.v91i1.9397.

[27] Maleki, M., Mahmoudi, M.R., Wraith D. and Pho, K.-H., 2020. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease*, 37, https://doi.org/10.1016/j.tmaid.2020.101742.

[28] Nesteruk, I., 2020. Statistics-based predictions of coronavirus epidemic spreading in mainland China. *Innovative Biosystems and Bioengineering*, 4(1),13-18, https://doi.org/10.20535/ibb. 2020.4.1.195074.

[29] Almongy, H.M., Alshenawy, F.Y., Almetwally, E.M. and Abdo, D.A., 2021. Applying transformer insulation using Weibull extended distribution based on progressive censoring scheme. *Axioms*, 10(2), https://doi.org/10.3390/axioms10020100.

[30] Iranzo, V. and Pérez-González, S., 2021. Epidemiological models and COVID-19: a comparative view. *History and Philosophy of the Life Sciences*, 43(3), https://doi.org/10.1007/s40656-021-00457-9.

[31] Ega, T.T. and Ngeleja, R.C., 2022. Mathematical model formulation and analysis for COVID-19 transmission with virus transfer media and quarantine on arrival. *Computational and Mathematical Methods*, 2022, https://doi.org/10.1155/2022/2955885.

[32] Shanker, R. and Shukla, K.K., 2022. The Poisson-Adya distribution. *Biometrics and Biostatistics International Journal*, 11(3), 100-103, https://doi.org/10.15406/bbij.2022.11.00361.

[33] Shanker, R., 2017. The discrete Poisson-Aradhana distribution. *Turkiye Klinikleri Journal of Biostatistics*, 9(1), 12-22, https://doi.org/10.5336/biostatic.2017-54834.

[34] Shanker, R., 2017. The discrete Poisson-Garima distribution. *Biometrics and Biostatistics International Journal*, 5(2), 48-53, https://doi.org/10.15406/bbij.2017.05.00127.

[35] Shanker, R., Fesshaya, H., Shanker, R., Leonida, T.A. and Sium, S., 2017. On discrete Poisson-Shanker distribution and its applications. *Biometrics & Biostatistics International Journal*, 5(1), https://doi.org/10.15406/bbij.2017.05.00121.

[36] Shanker, R. and Fesshaya, H. 2016. On Poisson-Sujatha distribution and its applications to model count data from biological sciences. *Biomarkers and Biostatistics International Journal,* 3(4), 100-106, https://doi.org/10.15406/bbij.2016.03.00069.

[37] Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov and B.F. Csaki, eds. *Second International Symposium on Information Theory*. Budapest: Academiai Kiado, pp. 199-213.

[38] Schwarz, G.E., 1978. Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.

[39] Cox, D.R. and Lewis, P.A.W., 1966. *The Statistical Analysis of Series of Events*. London: Methuen, pp.71-72.

[40] Kostyshak, S., 2022. *Bootstrap: Functions for the Book "An Introduction to the Bootstrap"*. R Package Version 2019.6.

[41] Murphy, M.V., 2022. *semEff: Automatic Calculation of Effects for Piecewise Structural Equation Models.* R Package Version 0.6.1.

[42] Kirby, K.N. and Gerlanc, D., 2013. BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, 45(3), 905-927, https://doi.org/10.3758/s13428-013-0330-5.