science, engineering and health studies

Research Article

Science, Engineering and Health Studies
https://li01.tci-thaijo.org/index.php/sehs
ISSN (Online): 2630-0087

Analyzing meteorological data with bootstrap-based confidence intervals for Poisson-Rani distribution parameter estimation

Wararit Panichkitkosolkul

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand

***Corresponding author**:*
Wararit Panichkitkosolkul
wararit@mathstat.sci.tu.ac.th

Received: *14 June 2024*
Revised: *24 October 2024*
Accepted: *9 November 2024*
Published: *11 November 2025*

Citation:
Panichkitkosolkul, W. (2025). Analyzing meteorological data with bootstrap-based confidence intervals for Poisson-Rani distribution parameter estimation. Science, Engineering and Health Studies, 19, 25020005.

ABSTRACT

The Poisson distribution is commonly used when events are assumed to be independent and occur at a consistent rate. This may not be generally applicable, and the Poisson distribution is not appropriate in situations where the underlying rate of occurrence displays variability. A mixed Poisson distribution such as the Poisson-Rani distribution permits the rate parameter to be random instead of constant. Bootstrap-based confidence intervals (CIs) were developed for the Poisson-Rani distribution parameter in this study. The percentile bootstrap (PB), basic bootstrap (BB), and bias-corrected and accelerated (BCa) bootstrap methods were compared for empirical coverage probabilities and expected lengths by the Monte Carlo simulation using the RStudio program with sample sizes of 10, 30, 50, 100, 500, and 1,000. The parameter values (θ) were set at 0.1, 0.3, 0.5, 0.8, 1, 1.5, and 2 with 1,000 replications. The simulation results suggested that the bootstrap-based CIs required improvement to attain the nominal confidence level for small sample sizes. No significant differences were detected in the performances of bootstrap-based CIs when evaluating large sample sizes, with the BCa bootstrap CI exhibiting superior performance compared to the others. The application of bootstrap-based CIs to meteorological data yielded comparable results to the simulation study.

Keywords: count data; mixed Poisson distribution; rainy days; resampling method; statistical inference

1. INTRODUCTION

Analyzing meteorological data is essential to understanding and managing weather and climate-related phenomena, which have far-reaching impacts on various aspects of human life, the environment, and the economy (Clarke et al., 2022), and enable informed decision-making to adapt to the challenges posed by a changing climate. Numerous studies have analyzed meteorological data. Brammer (2020) assessed climate change in Bangladesh, while Zhang et al. (2023) analyzed the influencing factors of forest fires using different combustibles, meteorological and climatic factors, and spatial and temporal distribution characteristics. This study determined the number of rainy days in a week, which followed a Poisson distribution as a mathematical model that calculates the probability of a specific number of events happening during a given time or space interval. The input describes a distribution that assumes events occur at a constant average rate and are independent of time (Kissell & Poserina, 2017). However, in actual situations, this assumption may not always hold true, and if the rate parameter is not constant a Poisson distribution may be inappropriate. When the Poisson parameter is assumed to be a random variable, a mixed

Poisson distribution can be used (Tharshan & Wijekoon, 2022). Several mixed Poisson distributions were reviewed in this paper including the Poisson-Lindley (Sankaran, 1970), Poisson-Akash (Shanker, 2017a), Poisson-Ishita (Shukla & Shanker, 2019), and Poisson-Prakaamy (Shukla & Shanker, 2020) distributions. Ahmad et al. (2021) proposed the Poisson-Rani (PR) distribution and studied its statistical properties. The PR distribution is obtained using the Rani distribution as the underlying distribution for the Poisson parameter, representing the average number of events. When applied to two real data sets, the PR distribution worked better than either the Poisson or Poisson-Lindley (Sankaran, 1970) distributions.

The Rani distribution is a continuous lifetime distribution with a probability density function (pdf) defined in Equation 1:

$$f(x;\theta)=\frac{\theta^5}{\theta^5+24}(\theta+x^4)e^{-\theta x}, x>0,\ \theta>0. \qquad (1)$$

This distribution is a combination of an exponential distribution with a scale parameter θ and a gamma $(5,\theta)$ distribution with proportions $\theta^5/(\theta^5+24)$, and $24/(\theta^5+24)$, respectively. Shanker (2017b) showed that the Rani distribution outperformed the Akash (Shanker, 2015a), Rama (Shanker, 2017c), Akshaya (Shanker, 2017d), Shanker (Shanker, 2015b), Amarendra (Shanker, 2016a), Aradhana (Shanker, 2016b), Sujatha (Shanker, 2016c), Devya (Shanker, 2016d), Lindley (Lindley, 1958), and exponential distributions in terms of model fit. Figure 1 depicts plots of the Rani distribution pdf with specified parameter values.

A confidence interval (CI) is a statistical tool that establishes a range of values in which the true population parameter is expected to lie, based on the available sample data and a confidence level (Tan & Tan, 2010). A CI value can communicate the uncertainty of statistical estimates and assist researchers and decision-makers in making informed determinations regarding the underlying population characteristics. Traditional interval estimation methods for the PR distribution such as Wald-type, likelihood-based, and Bayesian CIs each have distinct advantages and limitations. The Wald-type CI is straightforward to compute but often unreliable, especially with small sample sizes. Likelihood-based CI offers greater accuracy and better coverage but can be computationally demanding, while Bayesian CI allows the incorporation of prior information but is sensitive to the choice of prior and may also require substantial computational resources. The bootstrap method presents several advantages over these traditional approaches due to its high flexibility because it does not rely on large sample approximations and can be effectively applied in situations where traditional methods may fail or prove less reliable. Despite these benefits, the bootstrap method also has drawbacks including significant computational demands and the need for a large number of resamples to achieve stable results.

No previous studies have used bootstrap methods to calculate CIs for the PR distribution parameter. The fundamental idea behind bootstrap CIs is to estimate the uncertainty or variability in a statistic, such as the mean, median, variance, or any other parameter of interest, by repeatedly resampling the observed data with replacement and then calculating the statistic of interest for each resampled dataset (Chernick & LaBudde, 2011). The purpose of this resampling technique is to estimate the sampling distribution of the statistic without relying on specific assumptions about the distribution of the population.

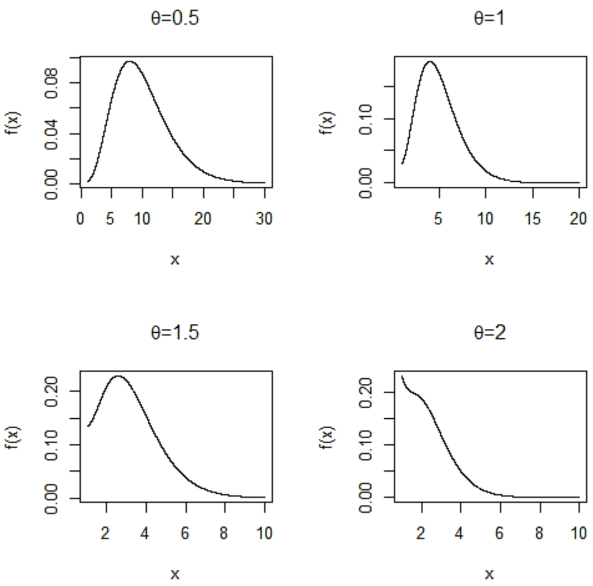

Figure 1. Plots of the Rani distribution pdf for θ = 0.5, 1, 1.5, and 2

This study evaluated the accuracy of three bootstrap-based CIs as percentile bootstrap (PB), basic bootstrap (BB), and bias-corrected and accelerated (BCa) bootstrap to estimate the PR distribution parameter.

2. MATERIALS AND METHODS

2.1 Point estimation for the Poisson-Rani distribution parameter

The Poisson distribution probability mass function (pmf) for a random variable Y can be represented in Equation 2:

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

The expected value and variance of Y are both equal to the parameter λ. Ong et al. (2021) and Tharshan and Wijekoon (2022) provided in-depth descriptions of the formation of mixed Poisson distributions. The Poisson-Rani (PR) distribution is based on the assumption that the Poisson parameter, denoted by λ, follows the Rani distribution.

If X has a PR distribution (Ahmad et al., 2021), its pmf is given by

$$p(x;\theta)=\frac{\theta^5[x^4+10x^3+35x^2+50x+24+\theta(\theta+1)^4]}{(\theta^5+24)(\theta+1)^{x+5}},\ x=0,1,2,\ldots,\ \theta>0.$$

Figure 2 shows the plots of the PR distribution pmf with several parameter values θ. The mean (or the first central moment) and variance (or the second central moment) of X are given by:

$$E(X)=\mu=\frac{\theta^5+120}{\theta(\theta^5+24)} \text{ and } var(X)=\sigma^2=\frac{(\theta^{11}+\theta^{10}+144\theta^6+528\theta^5+2880\theta+2880)}{\theta^2(\theta^5+24)^2}.$$

The log-likelihood function $log\,L\,(x_i\,;\theta)$ is maximized to obtain the point estimator of θ. Therefore, the maximum likelihood (ML) estimator for θ of the PR distribution can be derived by the following processes:

$$\frac{\partial}{\partial\theta}log\,L\,(x_i\,;\theta) = \frac{\partial}{\partial\theta}\left[\begin{array}{r}5n\,log(\theta) - n\,log(\theta^5+24) + \sum_{i=1}^{n} log[x_i^4+10x_i^3+35x_i^2+50x_i+24+\theta(\theta+1)^4] \\ -\sum_{i=1}^{n}(x_i+5n)\,log(\theta+1)\end{array}\right]$$
$$= \frac{5n}{\theta} - \frac{5n\theta^4}{\theta^5+24} + \sum_{i=1}^{n}\frac{(5\theta+1)(\theta+1)^3}{[x_i^4+10x_i^3+35x_i^2+50x_i+24+\theta(\theta+1)^4]} - \sum_{i=1}^{n}\frac{(x_i+5n)}{\theta+1}.$$

The subsequent nonlinear equation can be obtained by solving the equation $\frac{\partial}{\partial\theta}log\,L\,(x_i\,;\theta) \overset{\text{set}}{=} 0$ for θ,

$$\frac{5n}{\theta} - \frac{5n\theta^4}{\theta^5+24} + \sum_{i=1}^{n}\frac{(5\theta+1)(\theta+1)^3}{[x_i^4+10x_i^3+35x_i^2+50x_i+24+\theta(\theta+1)^4]} - \sum_{i=1}^{n}\frac{(x_i+5n)}{\theta+1} \overset{\text{set}}{=} 0.$$

There is no exact mathematical solution for the ML estimator of the parameter θ, and numerical iteration methods are used to solve the corresponding non-linear equation (Nwry et al., 2021). This study used the maxLik package (Henningsen & Toomet, 2011) to perform ML estimation using the Newton-Raphson technique in the RStudio program (RStudio Team, 2021).

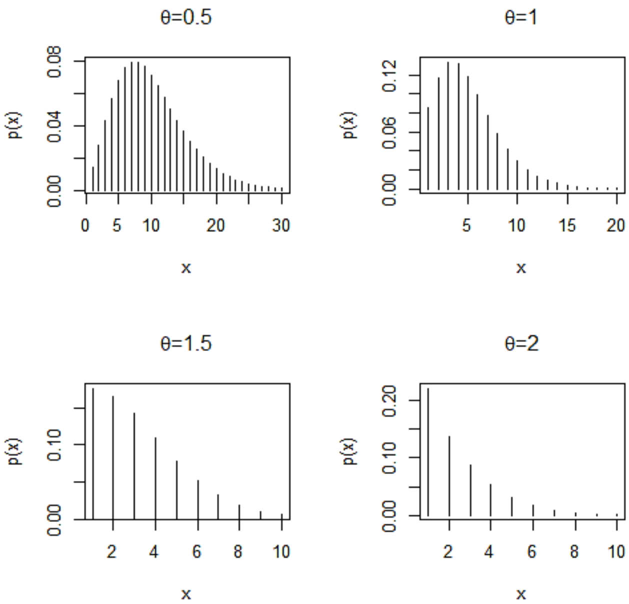

Figure 2. Plots of the PR distribution pmf for θ = 0.5, 1, 1.5, and 2

2.2 Bootstrap-based confidence intervals

CIs are commonly utilized in statistics and research to make inferences about populations when data collection from the entire population is not feasible. The CI calculation implies that the parameter distribution of the estimator is approximately normal (Ukoumunne et al., 2003). However, in some instances, the assumption of normality is violated, and estimating the standard error is difficult. One potential approach is to employ techniques rooted in the bootstrap method (van den Boogaard & Hall, 2004). This paper used bootstrap methods to construct approximate CIs that did not depend on assumptions regarding the underlying distribution (Meeker et al., 2017). The boot package (Canty & Ripley, 2022) was used to estimate the bootstrap CIs within the RStudio program.

2.2.1 PB confidence interval

The PB CI is a non-parametric method that estimates the uncertainty surrounding a population parameter by resampling the original sample. This is particularly useful when the underlying distribution of the data is unknown or complicated (Efron, 1982). The procedure for obtaining a PB CI for θ is 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 observations are selected from the original sample, allowing for the possibility that the same observation can be selected multiple times.

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

4) Generate the confidence interval. To construct a CI, the bootstrap statistics must be arranged in ascending order and the relevant percentiles selected. To obtain a

95% CI, the 2.5th percentile is selected 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 in Equation 3,

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

where $\hat{\theta}^*_{(\alpha)}$ denotes the α^{th} percentile of the distribution of the parameter estimate $\hat{\theta}^*$ and $0 \le r < s \le 100$. Hence, $r = [(\alpha/2)B]$, $s = [(1-(\alpha/2))B]$, where $[x]$ stands for the ceiling function of x, and $1-\alpha$ is the confidence level. A 95% PB two-sided CI is the interval between the 2.5 percentile value and the 97.5 percentile value of the 2,000 bootstrap parameter estimates. The two quantiles related to the lower and upper limits of the PB two-sided CI are $\hat{\theta}^*_{(r)} = \hat{\theta}^*_{(50)}$ (the 50th quantile) and $\hat{\theta}^*_{(s)} = \hat{\theta}^*_{(1950)}$ (the 1950th quantile).

2.2.2 BB confidence interval

The BB CI is a straightforward method that does not require complex adjustments or modifications to the bootstrap procedure. The BB CI focuses on the variability of the statistic itself rather than explicitly considering the tails of the distribution. Assume that 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 (Meeker et al., 2017). The $(1-\alpha)100\%$ two-sided BB CI for θ is

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

where $\hat{\theta}^*_{(r)}$ and $\hat{\theta}^*_{(s)}$ are, in ascending order, the r^{th} and s^{th} quantiles of a collection of parameter estimates $\hat{\theta}^*$ that are utilized in Equation 3 to calculate the PB CI.

2.2.3 BCa bootstrap confidence interval

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 data distribution is skewed. The BCa bootstrap CI corrects for both bias and skewness in the distribution of the bootstrap statistics (Efron, 1982). This reduces several problems 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 the BCa bootstrap CI commonly involves using statistics derived from jackknife simulations.

Davison and Hinkley (1997) and Chernick and LaBudde (2011) provided mathematical details for the BCa adjustment. They denoted the bias correction factor $\hat{z}_0$ as

$$\hat{z}_0 = \Phi^{-1}\left(\frac{1}{B}\sum_{i=1}^{B} I\left(\hat{\theta}_i^{\,*} < \hat{\theta}\right)\right)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative standard normal probability function and $I(\cdot)$ is the indicator function defined as $I\left(\hat{\theta}_i^{\,*} < \hat{\theta}\right) = 1$, if $\hat{\theta}_i^{\,*} < \hat{\theta}$ and $I\left(\hat{\theta}_i^{\,*} < \hat{\theta}\right) = 0$, if $\hat{\theta}_i^{\,*} \ge \hat{\theta}$. 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. From jackknife replicates, we obtain the value of $\hat{\theta}_{(-i)}, i = 1,2,\ldots,n$. The acceleration factor $\hat{a}$ is given by:

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

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 $\hat{a}$,

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

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 can be computed as

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

where $j = [\alpha_1 B]$ and $k = [\alpha_2 B]$.

3. RESULTS

Using the RStudio program (RStudio Team, 2021), three bootstrap-based CIs for the PR distribution parameter were considered in this study based on a Monte Carlo simulation. Small, medium, and large sample sizes were determined as n = 10, 30, 50, 100, 500, and 1,000 to reflect different data scenarios encountered in the simulation study. The parameter (θ) was set at 0.1, 0.3, 0.5, 0.8, 1, 1.5, and 2 to ensure a comprehensive simulation study across different potential values. The value for B, representing the bootstrap number of replications, was set at 2,000. A series of bootstrap samples, each consisting of n observations, was constructed from the initial sample following a PR distribution using 1,000 replications. The nominal confidence level, denoted as $1-\alpha$, was set to 0.95. The empirical coverage probability is the proportion of replicates for which the bootstrap-based CI contained the known parameter. The performances of the bootstrap-based CIs were evaluated by investigating their empirical coverage probabilities and expected lengths. Using the one-proportion z-test to test $H_0: CP \ge 0.95$ versus $H_a: CP < 0.95$ at a significance level of 0.05, where CP denotes the coverage probability of the CI, we concluded that the empirical coverage probability was greater than or equal to 0.95 if it was greater than or equal to 0.939. The bootstrap-based CI with the shortest expected length was used to obtain a more precise parameter estimation.

The simulation results are reported in Table 1 and Figures 3 and 4. For n = 10, 30, 50, and 100 all bootstrap-based CIs provided empirical coverage probabilities of less than 0.95. For these cases, the empirical coverage probabilities of all bootstrap-based CIs were significantly different from 0.95 using a one-sample t-test to test $H_0: \mu_{CP} = 0.95$ versus $H_a: \mu_{CP} \ne 0.95$, where μ_{CP} denotes the population mean of the coverage probability. This analysis was conducted using IBM SPSS Statistics, yielding a sample mean of the empirical coverage probability of 0.9304, a standard deviation of 0.0219, a t statistic of 8.196, and a p-value approximately equal to 0.000. Under these circumstances, the BCa bootstrap CI demonstrated superior performance compared to the other methods in terms of expected length.

The bootstrap-based CIs for n = 500 and n = 1,000 achieved empirical coverage probabilities close to the nominal confidence level and were not significantly different from

0.95, as determined by a one-sample t-test using IBM SPSS Statistics (sample mean of empirical coverage probability = 0.9482, SD = 0.0076, t statistic = 0.9310, p-value = 0.3650). The BCa bootstrap CI had an empirical coverage probability closer to 0.95, and as the sample size increased, coverage probabilities further approached 0.95.

The expected length of bootstrap-based CIs increased with an increase in the parameter value. As the sample size increased, the expected lengths of all three CIs decreased, with the BCa bootstrap CI providing the shortest expected length for all investigated situations. However, the expected lengths of the three bootstrap-based CIs were not greatly different using one-way ANOVA to test $H_0: \mu_{CP(PB)} = \mu_{CP(BB)} = \mu_{CP(BCa)}$, where $\mu_{CP(PB)}$, $\mu_{CP(BB)}$, and $\mu_{CP(BCa)}$ denote the population mean of the coverage probability of CI_{PB}, CI_{BB}, and CI_{BCa}, respectively via IBM SPSS Statistics (F statistic = 0.0049, p-value = 0.9951). The BCa bootstrap CI provided good empirical coverage probability and the shortest expected length for almost all situations.

Table 1. Empirical coverage probability and expected length of the 95% two-sided bootstrap-based CIs for the PR distribution parameter

n	θ	Empirical coverage probability			Expected length		
		PB	BB	BCa	PB	BB	BCa
10	0.1	0.904	0.891	0.909	0.0563	0.0564	0.0557
	0.3	0.892	0.891	0.896	0.1911	0.1910	0.1885
	0.5	0.886	0.897	0.896	0.3363	0.3352	0.3319
	0.8	0.913	0.863	0.904	0.5247	0.5242	0.5268
	1.0	0.887	0.859	0.883	0.6168	0.6161	0.6211
	1.5	0.911	0.921	0.912	0.8821	0.8866	0.8764
	2.0	0.934	0.942*	0.944*	2.1249	2.1296	1.9161
30	0.1	0.945*	0.935	0.945*	0.0333	0.0332	0.0329
	0.3	0.922	0.914	0.925	0.1098	0.1099	0.1085
	0.5	0.936	0.925	0.942*	0.1967	0.1966	0.1945
	0.8	0.932	0.916	0.928	0.3262	0.3263	0.3238
	1.0	0.944*	0.922	0.942*	0.3864	0.3862	0.3859
	1.5	0.943*	0.955*	0.948	0.4905	0.4903	0.4898
	2.0	0.937	0.956*	0.946*	0.7363	0.7374	0.7058
50	0.1	0.931	0.934	0.936	0.0258	0.0259	0.0257
	0.3	0.939*	0.936	0.939*	0.0848	0.0850	0.0842
	0.5	0.947*	0.953*	0.945*	0.1526	0.1525	0.1510
	0.8	0.941*	0.937*	0.944*	0.2531	0.2529	0.2517
	1.0	0.934	0.913	0.932	0.3025	0.3024	0.3015
	1.5	0.938	0.949*	0.940*	0.3783	0.3780	0.3780
	2.0	0.939*	0.968*	0.953*	0.5362	0.5352	0.5254
100	0.1	0.938	0.935	0.932	0.0183	0.0183	0.0183
	0.3	0.951*	0.945*	0.951*	0.0596	0.0597	0.0595
	0.5	0.945*	0.949*	0.953*	0.1078	0.1078	0.1072
	0.8	0.948*	0.940*	0.943*	0.1794	0.1794	0.1788
	1.0	0.941*	0.934	0.938	0.2198	0.2196	0.2196
	1.5	0.943*	0.952*	0.943*	0.2678	0.2681	0.2682
	2.0	0.934	0.958*	0.942*	0.3622	0.3628	0.3598
500	0.1	0.940*	0.938	0.936	0.0082	0.0082	0.0082
	0.3	0.942*	0.946*	0.939*	0.0268	0.0268	0.0267
	0.5	0.957*	0.954*	0.955*	0.0480	0.0480	0.0479
	0.8	0.938	0.939*	0.930	0.0811	0.0812	0.0810
	1.0	0.949*	0.948*	0.951*	0.0996	0.0997	0.0997
	1.5	0.944*	0.941*	0.941*	0.1201	0.1201	0.1200
	2.0	0.958*	0.958*	0.958*	0.1580	0.1582	0.1580
1,000	0.1	0.941*	0.949*	0.944*	0.0058	0.0058	0.0058
	0.3	0.948*	0.952*	0.951*	0.0190	0.0190	0.0190
	0.5	0.948*	0.952*	0.952*	0.0338	0.0339	0.0339
	0.8	0.950*	0.956*	0.953*	0.0577	0.0576	0.0576
	1.0	0.942*	0.941*	0.944*	0.0705	0.0704	0.0704
	1.5	0.955*	0.959*	0.953*	0.0846	0.0846	0.0846
	2.0	0.956*	0.959*	0.959*	0.1115	0.1115	0.1113

Note: The asterisk (*) indicates that the empirical coverage probability was greater than or equal to 0.939

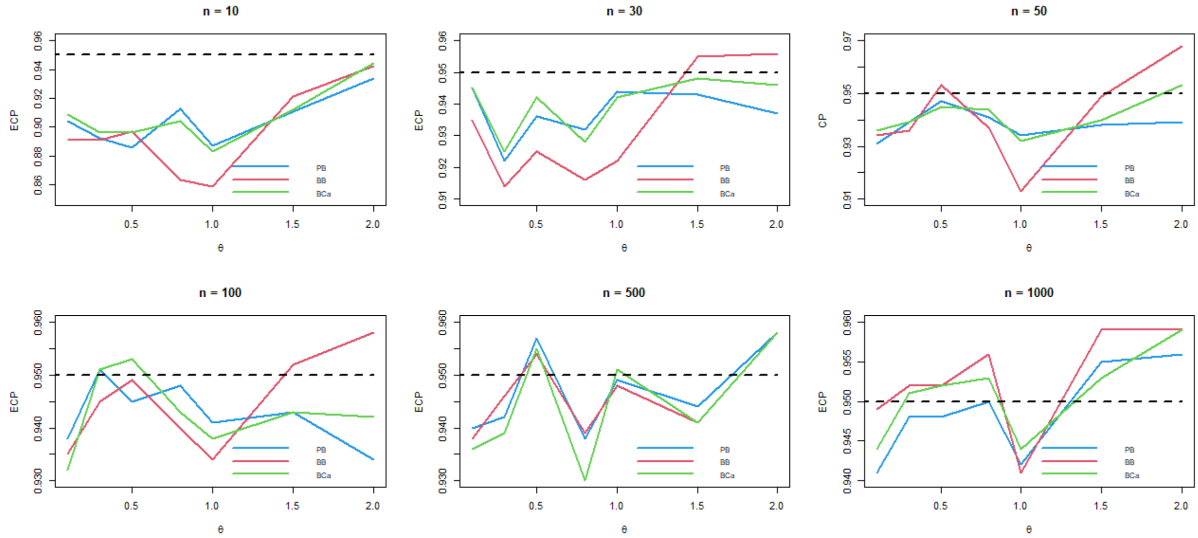

Figure 3. Plots of the empirical coverage probabilities of the CIs for θ of the PR Distribution

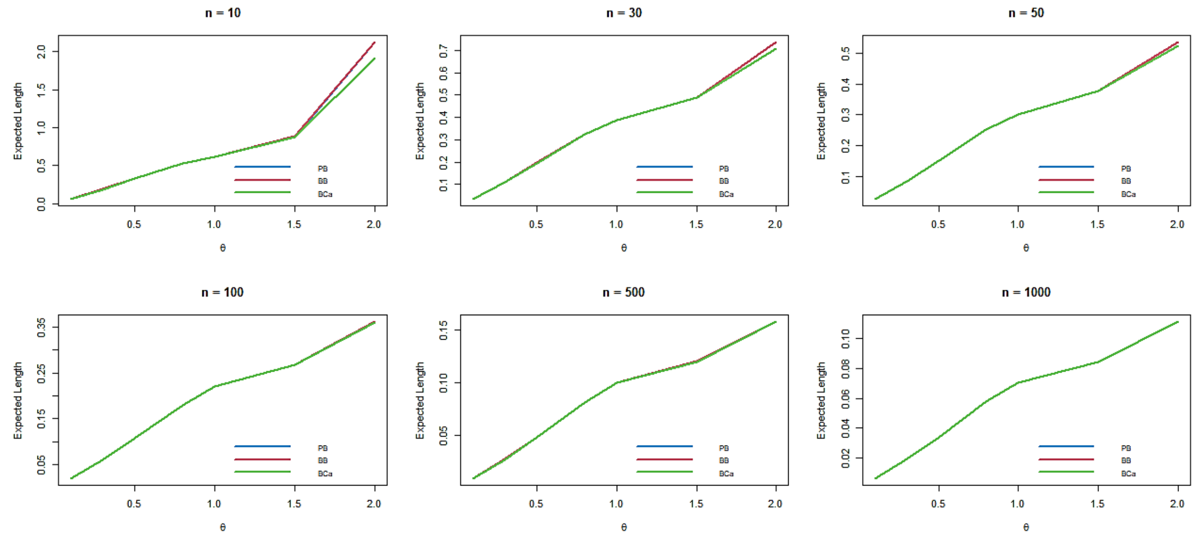

Figure 4. Plots of the expected lengths of the CIs for θ of the PR distribution

4. DISCUSSION

The Thai Meteorological Department collects data on the number of rainy days in a week at meteorological stations located in the central region of Thailand. The data, comprising 49 observations, recorded from July 1–7, 2019 are listed as follows in ascending order: 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6. Some descriptive statistics of the data set are reported in Table 2.

Table 2. Descriptive statistics of the data sets

Min	Mean	Median	SD	Q1	Q3	Max
0	2.653	3	2.037	0	4	6

The performance adequacy of the PR distribution was compared to the following alternative distributions:

- The Poisson-Shanker (PS) distribution (Shanker et al., 2017). Its pmf was

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

- The Poisson-Lindley (PL) distribution (Sankaran, 1970). Its pmf was

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

- The Poisson-Sujatha (PSj) distribution (Shanker, 2016e). Its pmf was

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

- The Poisson-Akash (PA) distribution (Shanker, 2017a). Its pmf was

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

All the distribution parameters were estimated using the ML technique. We considered the log-likelihood (log L), Akaike's information criterion (AIC) (Wasinrat & Choopradit, 2023; Akaike, 1974), and the Bayesian information criterion (BIC) or Schwarz information criterion (Wasinrat & Choopradit, 2023; Schwarz, 1978) for model comparison. The AIC and BIC statistics were defined as AIC = $2k-2\log\hat{L}$ and BIC = $2k\,log(n)-2log\,\hat{L}$, where k is the number of estimated parameters in the model and $\hat{L}$ is the maximized value of the likelihood function for the model. Estimates of the parameter, their standard errors (SE), and measures of goodness of fit for this data set are shown in Table 3.

Table 3. ML estimates, SE, AIC, and BIC for the number of rainy days in a week for the central region, Thailand

Distribution	Estimate (SE)	Log L	AIC	BIC
PR	1.5069 (0.0956)	-99.7069	<u>201.4138</u>	<u>203.3056</u>
PS	0.6376 (0.0740)	-102.7967	207.5934	209.4852
PL	0.6081 (0.0788)	-103.0545	208.1090	210.0008
PSj	0.8764 (0.0939)	-101.8970	205.7940	207.6858
PA	0.9041 (0.0922)	-101.5895	205.1790	207.0708

Note: The underline represents the minimum AIC and BIC

By employing the Kolmogorov-Smirnov (K-S) test (Wilcox, 2021; Sukkasem, 2010) for fitting the PR distribution, we obtained the expected frequencies shown in Table 4, a K-S statistic of 0.7937, and a corresponding p-value of 0.7980. Consequently, a PR distribution with $\hat{\theta}$ = 1.5069 was selected as appropriate for this data set. Table 5 presents the 95% two-sided bootstrap-based CIs for the PR distribution parameter and their lengths. This application was consistent with the simulated results, as the expected lengths of all bootstrap-based CIs were similar.

Table 4. The number of rainy days in a week for the central region of Thailand

Number of rainy days	Observed frequencies	Expected frequencies
0–1	14	18.7758
2–3	18	14.9901
4–5	12	9.1002
6–7	5	6.1338

Table 5. The 95% two-sided bootstrap-based CIs and lengths of the number of rainy days in a week

Method	Confidence interval	Length
PB	(1.3371, 1.6898)	0.3527
BB	(1.3292, 1.6732)	0.3440
BCa	(1.3373, 1.6893)	0.3520

5. CONCLUSION

No previous research has studied bootstrap-based CIs for the PR distribution parameter. This study evaluated the performance of PB, BB, and BCa bootstrap approaches for the parameter of the PR distribution. The advantages of bootstrap-based CIs are their robustness, flexibility, and ability to make inferences without assuming a specific data distribution. They work well with non-Gaussian data, and situations where traditional parametric methods are inappropriate, providing a model-free approach to estimate the sampling distribution of statistics and they are also relatively simple to implement. The performances of these three approaches were evaluated by comparing their empirical coverage probability and expected length using simulated data. Results showed that the bootstrap-based CIs were significantly influenced by the sample size (n). When the sample sizes were 10, 30, 50, and 100 the empirical coverage probabilities for all three bootstrap-based CIs differed from 0.95. For large sample sizes (n = 500 and 1,000), the empirical coverage probabilities of all bootstrap-based CIs showed no substantial deviations from the 95% confidence level, and the expected lengths of all bootstrap-based CIs remained relatively consistent. The simulation results showed that the BCa bootstrap CI outperformed the others in almost all cases, both in the simulated research and when using a real data set. Furthermore, when we applied these proposed methods to the meteorological data, specifically to the number of rainy days in a week in the central region of Thailand, the results were consistent with our simulation findings. The BCa bootstrap CI provided a narrower interval, particularly evident in the small sample size of 49 observations. The expected lengths of the CIs were slightly longer in the real data application compared to the simulation, likely due to the real-world variability present in the data. However, the overall consistency between the simulation results and the real data application underscored the reliability of the BCa bootstrap CI for practical use, even when dealing with real-world data that may exhibit more complexity and variability than simulated data. Table 6 displays the estimated probability and 95% CIs for the number of rainy days in a week in July for the central region of Thailand.

Table 6. Estimated probability and 95% CIs of the number of rainy days in a week in July for the central region of Thailand

Number of rainy days	Estimated probability (95% CI)
0	0.2063 (0.1386, 0.2910)
1	0.1769 (0.1484, 0.2007)
2	0.1649 (0.1587, 0.1606)
3	0.1410 (0.1236, 0.1494)
4	0.1088 (0.0876, 0.1250)
5	0.0769 (0.0573, 0.0952)
6	0.0507 (0.0351, 0.0676)
7	0.0317 (0.0204, 0.0453)

One drawback of this study was that none of the CIs based on bootstrapping yielded exact results; however, they demonstrated consistency because the empirical coverage probability approached the nominal confidence level as sample sizes increased. Our methodology showed potential to assist environmental scientists and government agencies in managing agriculture and water resources. Monitoring weekly rainfall patterns helps to identify areas at risk and take measures to prevent disasters, while monitoring is essential for assessing and managing water quality. Our findings provide valuable insights into estimating parameters including the population mean of the number of rainy days, which can inform decision-making, support ecosystem health, and contribute to the safety and well-being of communities.

This study had limitations because the three bootstrap-based CIs were computationally difficult and time-consuming. RStudio program provides numerous utilities for computing bootstrap-based CIs such as the boot package (Canty & Ripley, 2022), the bootstrap package (Kostyshak, 2022), the semEff package (Murphy, 2022), and the BootES package (Kirby & Gerlanc, 2013).

Future studies should focus on how alternative CI estimations compare to the bootstrap-based CIs presented in this research. The construction of CIs for functions of parameters such as the population mean and dispersion index is of interest. There is also a lack of statistical theoretical research regarding hypothesis testing for the PR distribution parameter. The bootstrap-based CIs studied in this paper can be applied to other distributions. These topics may be the subjects for further investigation in subsequent studies. Moreover, other meteorological data should study using bootstrap-based CIs for PR distribution parameter.

ACKNOWLEDGMENTS

The author would like to thank the editor and the reviewers for their valuable comments and suggestions on how to improve this paper. This study was supported by the Thammasat University Research Fund, Contract No. TUFT 0004/2567.

REFERENCES

Ahmad, A., Jallal, M., Mudasir, S., Ahmad, A., & Tripathi, R. (2021). The discrete Poisson-Rani distribution with properties and applications. *International Journal of Open Problems in Computer Science and Mathematics*, *14*(3), 63–78.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Brammer, H. (2020). Problems with analysing meteorological data for assessing climate change: Examples from Bangladesh. *International Journal of Environmental Studies*, *77*(6), 905–915. https://doi.org/10.1080/00207233.2020.1848763

Canty, A., & Ripley, B. (2022). *Boot: Bootstrap R (S-Plus) functions* (R Package Version 1.3-28.1) [Computer software]. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=boot

Chernick, M. R., & LaBudde, R. A. (2011). *An introduction to bootstrap methods with applications to R*. John Wiley & Sons.

Clarke, B., Otto, F., Stuart-Smith, R., & Harrington, L. (2022). Extreme weather impacts of climate change: An attribution perspective. *Environmental Research: Climate*, *1*(1), Article 012001. https://doi.org/10.1088/2752-5295/ac6e7d

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. SIAM.

Henningsen, A., & 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

Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, *45*(4), 905–927. https://doi.org/10.3758/s13428-013-0330-5

Kissell, R., & Poserina, J. (2017). *Optimal sports math, statistics, and fantasy*. Academic Press.

Kostyshak, S. (2022). *Bootstrap: Functions for the book an introduction to the bootstrap* (R package version 2019.6) [Computer software] Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=bootstrap

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *20*(1), 102–107. https://doi.org/10.1111/j.2517-6161.1958.tb00278.x

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

Murphy, M. V. (2022). *semEff: Automatic calculation of effects for piecewise structural equation models* (R Package Version 0.6.1) [Computer software] Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=semEff

Nwry, A. W., Kareem, H. M., Ibrahim, R. B., & Mohammed, S. M. (2021). Comparison between bisection, Newton and secant methods for determining the root of the non-linear equation using MATLAB. *Turkish Journal of Computer and Mathematics Education*, *12*(14), 1115–1122.

Ong, S.-H., Low, Y.-C., & Toh, K.-K. (2021). Recent developments in mixed poisson distributions. *ASM Science Journal*, *14*. https://doi.org/10.32802/asmscj.2020.464

RStudio Team. (2021). *RStudio: Integrated development for R*. [Computer software]. RStudio, PBC.

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

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. http://www.jstor.org/stable/2958889

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

Shanker, R. (2015b). Shanker distribution and its applications. *International Journal of Statistics and Applications*, *5*(6), 338–348. https://doi.org/10.5923/j.statistics.20150506.08

Shanker, R. (2016a). Amarendra distribution and its applications. *American Journal of Mathematics and Statistics*, *6*(1), 44–56. https://doi.org/10.5923/j.ajms.20160601.05

Shanker, R. (2016b). Aradhana distribution and its applications. *International Journal of Statistics and Applications*, *6*(1), 23–34. https://doi.org/10.5923/j.statistics.20160601.04

Shanker, R. (2016c). Sujatha distribution and its applications. *Statistics in Transition. New Series*, *17*(3), 391–410. https://doi.org/10.21307/stattrans-2016-029

Shanker, R. (2016d). Devya distribution and its applications. *International Journal of Statistics and Applications*, *6*(4), 189–202. https://doi.org/10.5923/j.statistics.20160604.01

Shanker, R. (2016e). The discrete Poisson-Sujatha distribution. *International Journal of Probability and Statistics*, *5*(1), 1–9. https://doi.org/10.5923/j.ijps.20160501.01

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

Shanker, R. (2017b). Rani distribution and its application. *Biometrics & Biostatistics International Journal*, *6*(1), 256–265. https://doi.org/10.15406/bbij.2017.06.00155

Shanker, R. (2017c). Rama distribution and its application. *International Journal of Statistics and Applications*, *7*(1), 26–35. https://doi.org/10.5923/j.statistics.20170701.04

Shanker, R. (2017d). Akshaya distribution and its application. *American Journal of Mathematics and Statistics*, *7*(2), 51–59. https://doi.org/10.5923/j.ajms.20170702.01

Shanker, R., Fesshaye, H., Shanker, R., Leonida, T. A., & Sium, S. (2017). On discrete Poisson-Shanker distribution and its applications. *Biometrics & Biostatistics International Journal*, *5*(1), 6–14. https://doi.org/10.15406/bbij.2017.05.00121

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

Shukla, K. K., & Shanker, R. (2020). The Poisson-Prakaamy distribution and its applications. *Aligarh Journal of Statistics*, *40*, 137–150.

Sukkasem, J. (2010). The modified Kolmogorov-Smirnov one-sample test statistic. *Thailand Statistician*, *8*(2), 143–155. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34284

Tan, S. H., & Tan, S. B. (2010). The correct interpretation of confidence intervals. *Proceedings of Singapore Healthcare*, *19*(3), 276–278. https://doi.org/10.1177/201010581001900316

Tharshan, R., & Wijekoon, P. (2022). A new mixed Poisson distribution for over-dispersed count data: Theory and applications. *Reliability: Theory and Applications*, *17*(1), 33–51. https://doi.org/10.24412/1932-2321-2022-167-33-51

Ukoumunne, O. C., Davison, A. C., Gulliford, M. C., & 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

van den Boogaard, H. F. P., & 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

Wasinrat, S., & Choopradit, B. (2023). The poisson inverse Pareto distribution and its application. *Thailand Statistician*, *21*(1), 110–124. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/248027

Wilcox, R. R. (2021). *Introduction to robust estimation and hypothesis testing* (5th ed.). Academic Press.

Zhang, B., Cai, D., Ai, S., Wang, H., & Zuo, X. (2023). Research on the influencing factors and prevention measures of long-term forest fire risk in Northeast China. *Ecological Indicators*, *155*, Article 110965. https://doi.org/10.1016/j.ecolind.2023.110965