วารสารวิชาการวิทยาศาสตร์
มหาวิทยาลัยราชภัฏจันทรเกษม

ช่วงความเชื่อมั่นบูตสแทร็ปสำหรับพารามิเตอร์ของการแจกแจงปัวซง-การิมา: กรณีศึกษาจำนวนพายุฝนฟ้าคะนองในประเทศสหรัฐอเมริกา

Bootstrap Confidence Intervals for the Parameter of the Poisson-Garima Distribution: A Case Study of the Number of Thunderstorms in the USA

วราฤทธิ์ พานิชกิจโกศลกุล[1]*

Wararit Panichkitkosolkul[1]*

[1] สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

[1] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University

บทคัดย่อ

มีการนำเสนอการแจกแจงปัวซง-การิมาสำหรับข้อมูลจำนวนนับ ซึ่งมีความสำคัญในหลายๆ สาขาวิชา เช่น วิทยาศาสตร์ชีวภาพ วิทยาศาสตร์การแพทย์ ประชากรศาสตร์ นิเวศวิทยา เทคโนโลยี เป็นต้น อย่างไรก็ตาม การประมาณค่าช่วงความเชื่อมั่นสำหรับพารามิเตอร์ของการแจกแจงดังกล่าวยังไม่มีการนำเสนอ งานวิจัยนี้ได้นำเสนอช่วงความเชื่อมั่นโดยใช้วิธีบูตสแทร็ปเปอร์เซ็นไทล์ วิธีบูตสแทร็ปพื้นฐาน และวิธีบูตสแทร็ปปรับค่าเอนเอียงและค่าเร่ง และพิจารณาค่าความน่าจะเป็นคุ้มรวมและความกว้างเฉลี่ยของช่วงความเชื่อมั่นด้วยวิธีการจำลองแบบมอนติคาร์โล ผลการวิจัยแสดงให้เห็นว่า ช่วงความเชื่อมั่นทุกวิธียังให้ค่าความน่าจะเป็นคุ้มรวมไม่เข้าใกล้ระดับนัยสำคัญที่กำหนด เมื่อตัวอย่างมีขนาดเล็กในทุก ๆ สถานการณ์ นอกจากนั้นเมื่อตัวอย่างมีขนาดใหญ่มากพอ ช่วงความเชื่อมั่นทุกวิธีจะผลการจำลองไม่แตกต่างกันมากนัก เมื่อพิจารณาในภาพรวมพบว่าช่วงความเชื่อมั่นบูตสแทร็ปปรับค่าเอนเอียงและค่าเร่งมีประสิทธิภาพมากกว่าช่วงความเชื่อมั่นวิธีอื่น ๆ ในทุกสถานการณ์ นอกจากนี้ ประสิทธิภาพของช่วงความเชื่อมั่นบูตสแทร็ปยังได้นำไปประยุกต์ใช้กับจำนวนพายุฝนฟ้าคะนองในประเทศสหรัฐอเมริกาซึ่งให้ผลลัพธ์สอดคล้องกับผลการจำลอง

คำสำคัญ: การประมาณค่าแบบช่วง, การแจกแจงปัวซง, พารามิเตอร์, วิธีบูตสแทร็ป, การจำลอง

Abstract

The Poisson-Garima distribution has been proposed for studying count data, which is of primary interest in several fields, such as biological science, medical science, demography, ecology and technology. However, estimating the confidence interval for its parameter has not yet been examined. In this study, confidence interval estimation based on the percentile, basic, and biased-corrected and accelerated bootstrap methods was examined in terms of their coverage probabilities and average lengths via Monte Carlo simulation. The results indicate that attaining the nominal confidence level using the bootstrap methods was not possible for small sample sizes regardless of the other settings. Moreover, when the sample size was large, the performances of the methods were not substantially different. Overall, the bias-corrected and accelerated bootstrap methods outperformed the others for all of the cases studied. Lastly, the efficacies of the bootstrap methods were illustrated by applying them to the number of thunderstorms in the USA, the results of which match those from the simulation study.

Keywords: interval estimation, Poisson distribution, parameter, bootstrap method, simulation

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

Received: *15 มีนาคม 2566* ***Revised:*** *19 มิถุนายน 2566* ***Accepted:*** *20 มิถุนายน 2566*

1. Introduction

A thunderstorm, also known as a lightning storm, is a storm characterized by the presence of lightning and its acoustic effect on the earth's atmosphere. Thunderstorms are usually accompanied by strong winds and can produce heavy rain, snow, sleet, or hail, and sometimes little or no precipitation (Wikipedia, 2023). It is estimated that there are as many as 40,000 thunderstorm occurrences each day world-wide, which translates into 14.6 million occurrences annually (National Weather Service, 2023). The USA certainly experiences its share of thunderstorms, and of the estimated 100,000 occurrences each year around 10% are classified as severe (Florida Division of Emergency Management, 2023). Figure 1 shows the average number of thunderstorm days each year throughout the USA from 1993-2018. It can be seen that the most frequent occurrence is in the southeastern states, with Florida having the highest number (80 to more than 105 days per year).

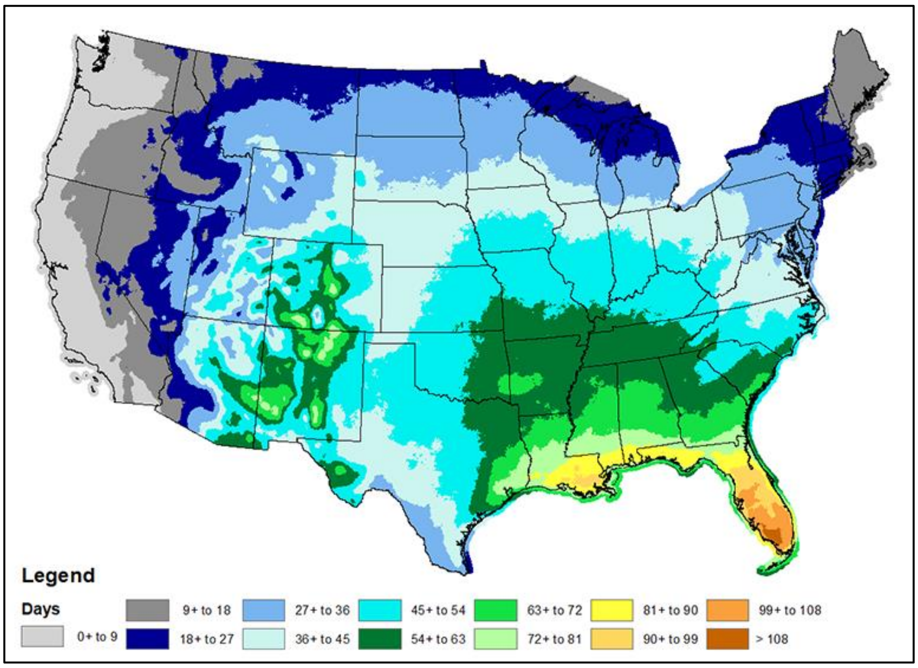

Figure 1. Average number of thunderstorm days each year throughout the USA in 1993-2018

Source: https://www.weather.gov/jetstream/tstorms_intro

It is well-known that that the number of events happening in a specific region of time and/or space has a Poisson distribution (Andrew and Michael, 2022). Data such as the number of thunderstorms or other meteorological events occurring in a particular locality over a given amount of time, the number of orders a firm will receive tomorrow, the number of calls attended per hour at a call center, the number of defects in a finished product etc., (Siegel, 2016) follow a Poisson distribution. Although the Poisson distribution is a basic model for the analysis of count data, its use is restricted due to the equality of its mean and variance (Equi-dispersion). A popular alternative when the count frequency data exhibit over-dispersion (the variance being greater than the mean) (Ong et al., 2021) is to use a mixed Poisson distribution in which it is assumed that the Poisson parameter is a random variable that has a single parameterized distribution (Tharshan and Wijekoon, 2022).

Recently, Shanker (2017) combined the Poisson and Garima distributions to produce the Poisson-Garima (PG) distribution and investigated its mathematical and statistical properties. The PG distribution arises from the Poisson distribution when Poisson parameter λ (the mean number of events) follows a Garima distribution. The method of moments and maximum likelihood estimation were both used to estimate the parameter of the PG distribution and when it was applied to two real data sets, it was more suitable than either the Poisson or Poisson-Lindley (Sankuran, 1970) distribution.

The Garima distribution is a lifetime continuous distribution introduced by Shanker (2016) with a probability density function (pdf) defined as

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

This distribution has been applied to model behavioral science data. Furthermore, Shanker (2016) also showed that the Garima distribution is a better model than either the exponential or Lindley (1958) distribution. The mathematical and statistical properties of the Garima distribution were established by Shanker (2016). Plots of the pdf of the Garima distribution with some specified values of parameter θ are shown in Figure 2.

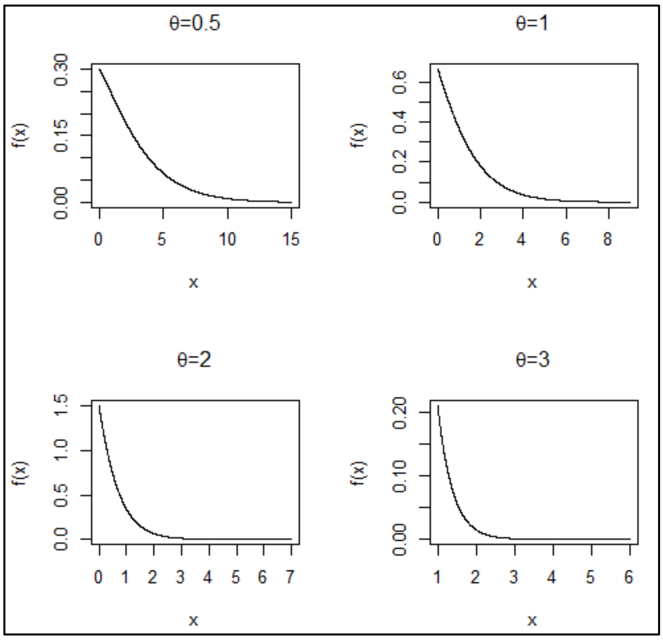

Figure 2. Plots of the pdf of the Garima distribution for $\theta=0.5$, 1, 2 or 3

In statistics, the confidence interval is a range of values that is likely to contain the true value of the population parameter of interest, which is a key output for many statistical analyses and has a critical role to play in the interpretation of parameter estimations (Tan and Tan, 2010). To the best of our knowledge, no research has been conducted on estimating the confidence interval for the parameter of the PG distribution. Bootstrap methods for estimating the confidence intervals of a parameter provide a way of quantifying the uncertainty in statistical inference based on a sample of data. The concept is to run a simulation study based on the actual data for estimating the likely extent of sampling error (Wood, 2004). The objective of the present study is to assess the efficiencies of three bootstrap methods, namely, the percentile bootstrap (PB), the basic bootstrap (BB), and the bias-corrected and accelerated bootstrap (BCa) to estimate the confidence interval for the parameter of the PG distribution. Because a theoretical comparison is not possible, we conducted a simulation study to compare their performances and used the result to determine the best-performing method based on their coverage probabilities and the average lengths.

2. Theoretical Background

The Poisson distribution is a discrete probability distribution that has the probability mass function (pmf) of a Poisson distribution as

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

where e is a constant approximately equal to 2.718282 and λ is a Poisson parameter. Let X be a random variable which follow the Poisson-Garima (PG) distribution with parameter θ, it is denoted as $X \sim \mathrm{PG}(\theta)$. Shanker (2017) defined the pmf of the PG distribution as

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

Plots of the pmf of the PG distribution with some specified parameter values θ are shown in Figure 3.

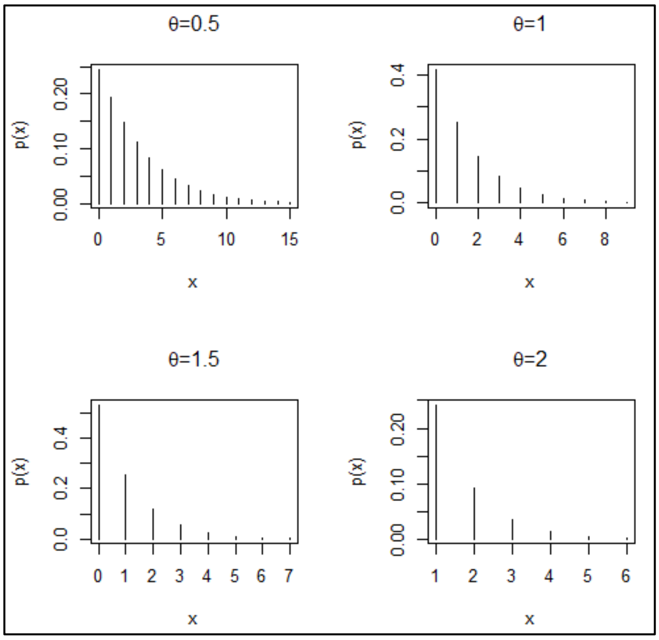

Figure 3. Plots of the pmf of the PG distribution for $\theta = 0.5$, 1, 1.5 or 2

The expected value and variance of X are as follows:

$$E(X) = \mu = \frac{\theta+3}{\theta(\theta+2)} \text{ and } \mathrm{var}(X) = \sigma^2 = \frac{\left(\theta^3+6\theta^2+12\theta+7\right)}{\theta^2(\theta+2)^2}.$$

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

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

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

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

where $\bar{x} = \sum_{i=1}^{n} x_i / n$ denotes the sample mean. Since the ML estimator for θ does not provide the closed-form solution, the non-linear equation can be solved by the numerical iteration methods such as Newton-Raphson, bisection, and Ragula-Falsi methods. In this paper, maxLik package (Henningsen and Toomet, 2011) with Newton-Raphson method was used for ML estimation in the statistical software R.

3. Bootstrap Confidence Interval Methods

In this paper, we focus on the three bootstrap confidence interval methods for the parameter of the PG distribution. The bootstrap methods, that are most popular in practice, are the percentile bootstrap, basic bootstrap, and bias-corrected and accelerated bootstrap methods (Chernick and LaBudde, 2011). In this study, boot package (Canty and Ripley, 2022) was used for estimating the bootstrap confidence intervals in the statistical software R.

3.1 Percentile bootstrap (PB) method

The percentile bootstrap confidence interval is the interval between the $(\alpha/2)\times 100$ and $(1-(\alpha/2))\times 100$ percentiles of the distribution of θ estimates obtained from resampling or the distribution of $\hat{\theta}^*,$ where θ represents a parameter of interest and α is the level of significance (e.g., α =0.05 for 95% confidence intervals) (Efron, 1982). A percentile bootstrap confidence interval for θ can be obtained as follows:

1) B random bootstrap samples are generated,

2) a parameter estimate $\hat{\theta}^*$ is calculated from each bootstrap sample,

3) all B bootstrap parameter estimates are ordered from the lowest to highest, and

4) the $(1-\alpha)100\%$ percentile bootstrap confidence interval is constructed as follows:

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

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

3.2 Basic bootstrap (BB) method

The basic bootstrap method is sometimes called the simple bootstrap method and is a method as easy to apply as the percentile bootstrap method. Suppose that the quantity of interest is θ and that the estimator of θ is $\hat{\theta}.$ The simple bootstrap method assumes that the distributions of $\hat{\theta}-\theta$ and $\hat{\theta}^* - \hat{\theta}$ are approximately the same (Meeker et al., 2017). The $(1-\alpha)100\%$ basic bootstrap confidence interval for θ is

$$CI_{BB} = \left[2\hat{\theta} - \hat{\theta}^*_{(s)}, 2\hat{\theta} - \hat{\theta}^*_{(r)}\right], \tag{5}$$

where the quantiles $\hat{\theta}^*_{(r)}$ and $\hat{\theta}^*_{(s)}$ are the same percentile of empirical distribution of bootstrap estimates $\hat{\theta}^*$ used in (4) for the percentile bootstrap method.

3.3 Bias-corrected and accelerated bootstrap (BCa) method

To overcome the over coverage issues in percentile bootstrap confidence intervals (Efron and Tibshirani, 1993), the BCa bootstrap method corrects for both bias and skewness of the bootstrap parameter estimates by incorporating a bias-correction factor and an acceleration factor (Efron and Tibshirani, 1993; Efron, 1987). The mathematical details of the BCa adjustment were provided in Chernick and LaBudde (2011) and Davison and Hinkley (1997). The bias-correction factor $\hat{z}_0$ is estimated as the proportion of the bootstrap estimates less than the original parameter estimate $\hat{\theta},$

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}^* \le \hat{\theta}\}}{B}\right), \tag{6}$$

where Φ^{-1} is the inverse function of a standard normal cumulative distribution function (e.g., $\Phi^{-1}(0.975) \approx 1.96).$ The acceleration factor $\hat{a}$ is estimated through jackknife resampling (i.e., "leave one out" resampling), which involves generating n replicates of the original sample, where n is the number of observations in the sample. The first jackknife replicate is obtained by leaving out the first case $(i=1)$ of the original sample, the second by leaving out the second case $(i=2),$ and so on, until n samples of size $n-1$ are obtained. For each of the jackknife resamples, $\hat{\theta}_{(-i)}$ is obtained. The average of these estimates is $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(-i)} / n.$ Then, the acceleration factor $\hat{a}$ is calculated as follow,

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

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

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

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

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

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

4. Simulation Study

The confidence intervals for the parameter of a PG distribution estimated via various bootstrap methods was considered in this study. Because a theoretical comparison is not possible, a Monte Carlo simulation study was designed using R version 4.2.2 (Ihaka and Gentleman, 1996) to cover cases with various sample sizes (n = 10, 30, 50, 100 and 500). To observe the effect of small and large variances, the true value of parameter (θ) was set as 0.1, 0.5, 1, 1.5, 2, or 2.5. The number of bootstrap replications (B) was set as 2000 because Ukoumunne et al. (2003) claimed that 2,000 bootstrap samples are sufficient to estimate the coverage probability for the 95% confidence intervals with a standard error of just under 0.5%. Bootstrap samples of size n were generated from the original sample and each simulation was repeated 1,000 times. Without loss of generality, the confidence level $(1-\alpha)$ was set at 0.95. The performances of the bootstrap methods were compared in terms of their coverage probabilities and average lengths; the one with a coverage probability greater than or close to the nominal confidence level (meaning that it contains the true value) and the shortest average length can be used to most accurately estimate the confidence interval for the parameter of interest for a particular scenario.

Table 1. Coverage probability and average length of the 95% confidence intervals for θ of the PG distribution.

n	θ	Coverage probability			Average length		
		PB	BB	BCa	PB	BB	BCa
10	0.1	0.857	0.846	0.865	0.1391	0.1395	0.1232
	0.5	0.868	0.862	0.882	0.9790	0.9827	0.8421
	1	0.896	0.839	0.889	2.7551	2.7413	2.5326
	1.5	0.938	0.822	0.916	4.4281	4.3673	3.9742
	2	0.974	0.752	0.969	5.5189	5.2922	5.4295
	2.5	0.964	0.668	0.972	5.6139	5.5344	5.4501
30	0.1	0.921	0.919	0.931	0.0684	0.0685	0.0650
	0.5	0.908	0.899	0.915	0.4083	0.4078	0.3822
	1	0.931	0.912	0.937	1.0184	1.0201	0.9358
	1.5	0.924	0.900	0.944	1.8834	1.8708	1.6632
	2	0.925	0.886	0.921	3.0302	3.0398	3.0255
	2.5	0.921	0.866	0.918	4.4960	4.5218	4.5366
50	0.1	0.930	0.916	0.928	0.0514	0.0515	0.0498
	0.5	0.932	0.920	0.933	0.3038	0.3032	0.2916
	1	0.942	0.920	0.948	0.6969	0.6970	0.6639
	1.5	0.933	0.917	0.933	1.2763	1.2764	1.1950
	2	0.928	0.922	0.925	1.9560	1.9593	1.9139
	2.5	0.909	0.899	0.908	2.8694	2.8717	2.8378
100	0.1	0.937	0.924	0.934	0.0359	0.0359	0.0353
	0.5	0.938	0.928	0.941	0.2061	0.2060	0.2021
	1	0.927	0.931	0.931	0.4905	0.4895	0.4776
	1.5	0.945	0.931	0.949	0.8232	0.8245	0.7980
	2	0.926	0.926	0.929	1.2497	1.2492	1.2339
	2.5	0.938	0.909	0.940	1.6926	1.6977	1.6718
500	0.1	0.961	0.959	0.958	0.0160	0.0160	0.0160
	0.5	0.947	0.937	0.944	0.0902	0.0904	0.0899
	1	0.948	0.947	0.945	0.2097	0.2094	0.2087
	1.5	0.949	0.945	0.949	0.3516	0.3516	0.3493
	2	0.950	0.924	0.949	0.5047	0.5044	0.5031
	2.5	0.938	0.932	0.940	0.6973	0.6966	0.6954

The results of the study are reported in Table 1. For n =10, the coverage probabilities of all three bootstrap methods tended to be less than 0.90 and so had not reach the nominal confidence level. Nevertheless, the BCa method outperformed the others in these scenarios. For n = 30 and 50, all of the methods once again provided coverage probabilities that were less than the nominal confidence level of 0.95. For $n \geq 100$, all of the bootstrap methods attained coverage probabilities close to the nominal confidence level and provided similarly average lengths. Therefore, as the sample size was increased, the coverage probabilities of the methods tended to increase and approach the nominal confidence level of 0.95. Moreover, the average lengths of the methods increased when the value of θ was increased because of the relationship between the variance and θ. Unsurprisingly, as the sample size was increased, the average lengths of all three bootstrap methods decreased, with the BCa method providing the shortest average length for all of the situations studied. Moreover, the average length of the BCa method was significantly different from the others when the sample size was small (n=10). In summary, the BCa method performs best in terms of coverage probability and average length for a large sample size ($n \geq 100$).

5. Empirical Application of the Confidence Interval Estimation Methods to Thunderstorm Data from USA

We used real-world count data to demonstrate the applicability of the bootstrap methods for estimating the confidence interval for the parameter of the PG distribution. Falls et al. (1971) reported the number of thunderstorm days per month at Cape Kennedy, Florida, USA for June, July, and August over 11 years from 1957 to 1967 (Table 2). The total number of thunderstorm days for June, July, and August, are 330, 341, 341, respectively. For the Chi-squared goodness-of-fit test (Turhan, 2020), the Chi-squared statistics and the p-values were calculated (Table 2). The results can be interpreted as PG distributions with $\hat{\theta}$=1.6906, 1.4726, and 1.5801 are suitable for the data-sets for June, July, and August, respectively. The 95% confidence interval estimation results for the parameter of the PG distribution are reported in Table 3. The results correspond with the simulation results in that the average lengths using the BCa methods were the shortest.

Table 2. The number of thunderstorms events per month for the month of June, July and August within 11-year period

Number of thunderstorms	June		July		August	
	Observed frequency	Expected frequency	Observed frequency	Expected frequency	Observed frequency	Expected frequency
0	187	183.011	177	179.438	185	186.226
1	77	82.91	80	86.656	89	86.024
2	40	36.597	47	40.743	30	38.708
3	17	15.846	26	18.782	24	17.082
≥ 4	9	11.636	11	15.381	13	12.96
Chi-squared statistic	1.5060		5.5272		4.8714	
p-value	0.5624		0.1370		0.1815	

6. Conclusions

Three bootstrap methods, namely PB, BB, and BCa, were proposed for estimating the confidence interval for the parameter of the PG distribution. When the sample sizes were small (n=10, 30 or 50), the coverage probabilities for all three were substantially lower than 0.95. When the sample size was sufficiently large ($n \geq 100$), the coverage probabilities and average lengths using all three bootstrap methods were not markedly different. According to our findings, the BCa method performed the best for almost all of the situations in both the simulation study and using real data-sets.

Table 3. The 95% confidence intervals and corresponding widths using all intervals for the parameter in the thunderstorms event examples for the month of June, July and August within 11-year period

Methods	June		July		August	
	Confidence intervals	Widths	Confidence intervals	Widths	Confidence intervals	Widths
PB	(1.4755, 1.9837)	0.5082	(1.3055, 1.6865)	0.3810	(1.3765, 1.8233)	0.4468
BB	(1.4135, 1.8975)	0.4840	(1.2592, 1.6334)	0.3742	(1.3362, 1.7678)	0.4316
BCa	(1.4639, 1.9415)	0.4776	(1.3051, 1.6744)	0.3693	(1.3800, 1.7947)	0.4147

7. References

Andrew, F.S., and Michael, R.W. (2022). *Practical business statistics*. 8th Ed. San Diego: Academic Press.

Canty, A., and Ripley, B. (2022). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-28.1.

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

David, F., and Johnson, N. (1952). The Truncated Poisson. *Biometrics, 8(4)*, 275–285. https://doi.org/10.2307/3001863

Davison, A.C., and Hinkley, D.V. (1997). *Bootstrap methods and their application*. 1st Ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511802843

Efron, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans, in *CBMS-NSF regional conference series in applied mathematics*, Philadelphia: SIAM. https://doi.org/10.1137/1.9781611970319

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(297), 171–185. https://doi.org/10.1080/01621459.1987.10478410

Efron, B., and Tibshirani, R.J. (1993). *An introduction to the bootstrap*. 1st Ed. New York: Chapman and Hall. https://doi.org/10.1007/978-1-4899-4541-9

Falls, L.W., Williford, W.O., and Carter, M.C. (1971). Probability distributions for thunderstorm activity at Cape Kennedy, Florida. *Journal of Applied Meteorology, 10*(1), 97–104. https://doi.org/10.1175/1520-0450(1971)010<0097:PDFTAA>2.0.CO;2

Florida Division of Emergency Management. (n.d.). *Thunderstorms*. https://www.floridadisaster.org/hazards/thunderstorms/.

Ghitany, M.E., Al-Mutairi, D.K., and Nadarajah, S. (2008). Zero-truncated Poisson-Lindley distribution and its application. *Mathematics and Computers in Simulation, 79*(3), 279–287. https://doi.org/10.1016/j.matcom.2007.11.021

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

Hussain, T. (2020). A zero truncated discrete distribution: Theory and applications to count data. *Pakistan Journal of Statistics and Operation Research, 16*(1), 167–190. https://doi.org/10.18187/pjsor.v16i1.2133

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/10.1080/10618600.1996.10474713

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

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. https://doi.org/10.1002/9781118594841

National Weather Service. (n.d.). *Introduction to Thunderstorms*. https://www.weather.gov/jetstream/tstorms_intro

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

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

Shanker, R. (2016). Garima distribution and its application to model behavioral science data. *Biometrics and Biostatistics International Journal, 4*(7), 275–281. https://doi.org/10.15406/bbij.2016.04.00116

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

Siegel, A.F. (2017). *Practical business statistics*. 7th Ed. London: Academic Press.

Tan, S.H., and 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., and Wijekoon, P. (2022). A new mixed Poisson distribution for over-dispersed count data: Theory and applications. *Reliability: Theory & Applications, 17*(1), 33–51.

Thunderstorm (n.d.). In *Wikipedia*. https://en.wikipedia.org/wiki/Thunderstorm

Turhan, N.S. (2020). Karl Pearson's chi-square tests. *Educational Research Review, 15*(9), 575–580. https://doi.org/10.5897/ERR2019.3817

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

Wood, M. (2004). Statistical inference using bootstrap confidence intervals. *Significance, 1*(4), 180–182. https://doi.org/10.1111/j.1740-9713.2004.00067.x