RESEARCH ARTICLE

Simultaneous Confidence Intervals for Ratios of the Percentiles of the Delta-Lognormal Distribution

Usanee Janthasuwan, Suparat Niwitpong, Sa-Aat Niwitpong*

Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Abstract Statistical percentiles are extremely important and valuable measurement tools in various contexts. They offer significant advantages when it comes to estimating a wide range of values across different fields, making them essential in statistical analysis and research. This importance motivates researchers to study the construction of simultaneous confidence intervals for ratios of percentiles of the delta-lognormal distribution. This research introduces four distinct methods for constructing such confidence intervals: the fiducial generalized confidence interval, the Bayesian-based highest posterior density credible interval using Jeffreys prior, Jeffreys rule prior, and Uniform prior. All methods are subjected to evaluation and comparison through coverage probabilities and expected lengths in Monte Carlo simulations. The simulation study shows that, overall, the Bayesian-based highest posterior density credible interval works better and more accurately than the fiducial generalized confidence interval. To further validate the findings of the simulation study, these methods are applied to real-world daily rainfall data collected from a river basin in Thailand.

Keywords: Delta-lognormal distribution, fiducial generalized confidence interval, highest posterior density credible interval, Percentiles.

Introduction

Percentiles are a statistical measure that indicates the location, or position, of a value relative to the entire set of data. When dealing with multiple datasets, comparing percentiles alone may not provide a comprehensive understanding of the relationship between the groups. However, when we calculate the ratio of percentiles between these datasets, a clearer and more informative picture emerges regarding how these groups compare in terms of their distribution or specific values. This ratio represents the relative magnitude or proportion between the two percentiles and provides a standardized measure of their relationship. It enables a more nuanced comparison and deeper understanding of the interplay between these values. Furthermore, percentiles are advantageous because they offer robustness against outliers. Outliers, or extreme values that deviate substantially from the overall trend of the data, can have a significant impact on other statistical measurements. However, percentiles are less affected by outliers due to their reliance on the ranking of values. This statistical measure is particularly useful in fields where proportional comparisons are important, such as market research. It helps identify relative differences in performance, market shares, or growth rates. In healthcare, the ratio of percentiles can be used to compare health indicators or medical measurements between different populations. In finance, it can be used to assess risk and return. In economics, it is often used to analyse income or wealth distribution. Additionally, the ratio of percentiles also plays a significant role in academic assessments and educational research; see, e.g., [1]–[2].

Data encompassing both zero values and positive values can be effectively modelled using the delta-lognormal distribution. In this distribution, the number of zero observations with zero values follows a binomial distribution with a probability represented by ζ_i^*, where $0 < \zeta_i^* < 1$. Meanwhile, the positive

values follow a lognormal distribution with a probability represented by ζ_i. This statistical technique was initially introduced by Aitchison [3], and subsequently, many researchers have applied this delta-lognormal distribution to diverse datasets, demonstrating its versatility in various contexts, particularly in the environmental and medical aspects; see, e.g., [4]–[9].

Numerous researchers have delved into the realm of simultaneous confidence intervals (SCIs) for ratios of parameters across a diverse array of distributions. For instance, Malley [10] presented a method for forming SCIs of all ratios of linear forms of the mean vectors given more groups of multivariate normal samples. Kumar *et al.* [11] investigated k independent exponential populations with varying scale and location parameters, where these parameters may not be known, and proposed a method to derive a set of SCIs for all ratios in relation to the largest scale parameter. Then, Rublik [12] offered an explicit formula for constructing SCIs for ratios of variances across multiple populations. Next, Sadooghi-Alvandi and Malekzadeh [13] introduced a novel parametric bootstrap method for creating SCIs for the ratios of means in various lognormal distributions. Schaarsch-midt and Djira [14] used the multivariate t quantiles of SCIs to come up with Fieller-type intervals for ratios of fixed effect parameters in mixed models. After that, Maneerat and Niwitpong [15] constructed SCIs for all possible pairwise ratios of variances of several zero-inflated lognormal models. Finally, Zhang *et al.* [16] investigated SCIs for the ratios of means of zero-heavy log-normal populations and proposed SCIs based on the Bonferroni adjustment principle. In the same year, Kaewprasert *et al.* [17] developed SCIs for the ratio of means in multiple delta-gamma populations. The study applied these intervals to estimate the ratio means for natural rainfall datasets from six regions in Thailand during September, which represents the peak of the rainy season.

The importance of percentile ratios serves as a powerful tool, enriching our comprehension of data and enabling meaningful comparisons. When comparing the two populations using the difference between the percentiles, there might not be much of a difference. This results in problems with making clear inferences, and it is difficult to conclude. Therefore, simultaneously comparing the ratios of percentiles is a more accurate option than simultaneously comparing the differences between the percentiles when examining multiple populations. Importantly, no researchers studied have used simultaneous confidence intervals (SCIs) for ratios of percentiles within the context of the delta-lognormal distribution. Consequently, the main objective of this study is the reconstruction of SCIs for ratios of percentiles in delta-lognormal populations. We have presented four methods for constructing SCIs: Fiducial generalized confidence interval (FGCI), Bayesian-based highest posterior density credible interval based on Jeffreys prior (HPD.J), Jeffreys Rule prior (HPD.JR), and Uniform prior (HPD.U). The efficiency of these methods is compared using coverage probabilities along with expected lengths. Crucially, we will apply all five methods to the daily rainfall data collected along the river basin during the period from June 20th to July 9th, 2023 [18].

Materials and Methods

The delta-lognormal (DLN) distribution has three parameters: mean (μ_i), variance (σ_i^2), and probability of zero observations (ζ_i^*), which are represented as $\Delta\left(\mu_i, \sigma_i^2, \zeta_i^*\right)$, where $\zeta_i^* = 1 - \zeta_i$ and ζ_i is the probability of non-zero observations. Suppose that $A_{ij} = \left(A_{i1}, A_{i2}, \ldots, A_{im_i}\right)$; $i = 1, 2, \ldots k$ and $j = 1, 2, \ldots, m_i$ be a non-negative random sample for *k* independent of the DLN distribution. The distribution function of the DLN distribution is

$$G\left(a_{ij}; \mu_i, \sigma_i^2, \zeta_i^*\right) = \begin{cases} \zeta_i^* & ; a_{ij} = 0 \\ \zeta_i^* + \zeta_i F\left(a_{ij}; \mu_i, \sigma_i^2\right) & ; a_{ij} > 0 \end{cases}$$

according to Tian and Wu [19], where $F\left(a_{ij}; \mu_i, \sigma_i^2\right)$ is the distribution function of the lognormal distribution. The number of zero observed values, denoted by $m_{i(0)} \sim Binomial(m_i, \zeta_i^*)$, has a binomial distribution, and the number of non-zero observed values, $m_{i(1)}$, has a log-normal distribution, where $m_i = m_{i(0)} + m_{i(1)}$. Let $B_{ij} = \ln\left(A_{ij}\right)$ be the normal distribution, denoted by $B_{ij} \sim N(\mu_i, \sigma_i^2)$. Assume that the mean and variance are $\overline{B}_{i(1)}$ and $C_{i(1)}^2$, respectively, obtained from the log-transformed

positive observations. The observed values of $\overline{B}_{i(1)}$ and $C^2_{i(1)}$ can also be given as $\overline{b}_{i(1)}$ and $c^2_{i(1)}$,

respectively. The maximum likelihood estimator of parameter μ_i, σ^2_i, and ζ_i are $\overline{b}_{i(1)} = \dfrac{1}{m_{i(1)}} \displaystyle\sum_{j=1}^{m_{i(1)}} b_{ij}$,

$c^2_{i(1)} = \dfrac{1}{m_{i(1)}-1} \displaystyle\sum_{j=1}^{m_{i(1)}} \left(b_{ij} - \overline{b}_{i(1)} \right)^2$, and $\hat{\zeta}_i = \dfrac{m_{i(1)}}{m_i}$, respectively. The estimator of the *p*th quantile of

the DLN distribution is $\hat{q}_{pi} = \exp\left(\hat{\eta}_{pi} \right)$, where $\hat{\eta}_{pi} = \overline{B}_{i(1)} + \Phi^{-1}\left(\dfrac{p_i - \hat{\zeta}^*_i}{1 - \hat{\zeta}^*_i} \right) C_{i(1)}$. Since we are

interested in constructing the SCIs for all pairwise ratios between the percentiles (ω_{il}), then

$$\hat{\omega}_{il} = \frac{\hat{q}_{pi}}{\hat{q}_{pl}} = \exp\left(\hat{\eta}_i - \hat{\eta}_l \right)$$

where Φ is the standard normal distribution function, $i, l = 1, 2, \ldots, k$, and $i \neq l$.

Fiducial Generalized Confidence Interval

Hannig *et al.* [20] presented the concept of the FGCI. They highlighted that this interval relies on the fiducial generalized pivotal quantities (FGPQ), which are a subset of GPQ. Thus, the FGCI method builds upon the foundation of FGPQs. In accordance with Thangjai *et al.* [8], let p_i be the percentiles and

$$\Phi^{-1}\left(Q_i \right) = \Phi^{-1}\left(\frac{p_i - Q_{\zeta^*_i}}{1 - Q_{\zeta^*_i}} \right), \tag{1}$$

where $Q_{\zeta^*_i} = M^{-1}\left(X_i O_i;\, m_{i(0)} + \dfrac{1}{2},\, m_{i(1)} + \dfrac{1}{2} \right)$ is the quartile function of the beta distributions, while

$O_i = M\left(p_i;\, m_{i(0)} + \dfrac{1}{2},\, m_{i(1)} + \dfrac{1}{2} \right)$ is the function of beta distribution and X_i is the standard uniform

distribution. We can find the FGPQ of μ_i as

$$Q_{\mu i} = \overline{b}_{i(1)} - \frac{Z_i}{\sqrt{V_{i(1)}}} \sqrt{\frac{\left(m_{i(1)} - 1 \right) c^2_{i(1)}}{m_{i(1)}}} \tag{2}$$

and we can obtain the FGPQ of σ^2_i as

$$Q_{\sigma^2_i} = \frac{\left(m_{i(1)} - 1 \right) c^2_{i(1)}}{\chi^2_{m_{i(1)}-1}}. \tag{3}$$

Afterwards, from equation (1) to equation (3), we can calculate the FGPQ of η_{pi} as follows:

$$Q_{\eta_{pi}} = Q_{\mu i} + \sqrt{\frac{Q_{\sigma^2_i}}{m_{i(1)}}} \left(\frac{Z_i + \Phi^{-1}\left(Q_i \right) \sqrt{m_{i(1)}}}{\sqrt{V_{i(1)}}} \right),$$

where $Z_i \sim N(0,1)$ and $V_i \sim \sqrt{\dfrac{\chi^2_{m_{i(1)}}}{(m_{i(1)}-1)}}$. Therefore, the FGCI for the ratio of two independent

percentiles can be expressed as $Q_{\omega_{il}} = \exp\left(Q_{\eta_{pi}} - Q_{\eta_{pl}} \right)$. Consequently, $(1 - \varepsilon)100\%$ the two-sided

SCI for ω_{il} are

$$CI^{FGCI}_{\omega_{il}} = \left[Q_{\omega_{il}}\left(\varepsilon/2\right), Q_{\omega_{il}}\left(1-\varepsilon/2\right) \right], \tag{4}$$

where $Q_{\omega_{il}}\left(\varepsilon\right)$ as the $100\,\varepsilon$ th percentile of $Q_{\omega_{il}}$.

Bayesian-based HPD

In Bayesian statistics, HPD stands for highest posterior density, and it is a technique used to construct parameter estimates. It is a method of estimating the parameter values in Bayesian analysis in the form of the interval with the highest probability density. In this research, we will use the following prior distribution:

Jeffreys Prior

The prior density based on the square root of the Fisher information matrix was proposed by Jeffreys

[21]. Consequently, Jeffreys prior is defined as $P_J\left(\sigma_i^2\right) \propto \dfrac{1}{\sigma_i^2}$. The posterior distribution merges the

likelihood function and the prior distribution. Bayesian confidence intervals rely on the posterior distribution. It's a conditional distribution tied to observed sample values, used for statements about

parameters treated as random. As a result, the posterior densities of σ_i^2 and μ_i are

$$\sigma_{J.i}^2 \big| a_{ij} \sim Inv-Gamma\left(\frac{m_{i(1)}-1}{2}, \frac{(m_{i(1)}-1)c_{i(1)}^2}{2}\right)$$

and $\mu_{J.i}\big|\sigma_i^2, a_{ij} \sim N\left(\bar{b}_{i(1)}, \dfrac{\sigma_{J.i}^2}{m_{i(1)}}\right)$, respectively. According to Thangjai *et al.* [9], the probability

distributions for ζ_i^* is

$$H_{\zeta_i^*}^J = M^{-1}\left(X_i O_i;\ m_{i(0)}+\frac{1}{2},\ m_{i(1)}+\frac{1}{2}\right).$$

We can find the posterior distributions of η_{pi} as

$$\eta_{pi}^J = \mu_{J.i} + \sqrt{\frac{\sigma_{J.i}^2}{m_{i(1)}}}\left(\frac{Z_i + \Phi^{-1}\left(\lambda_i\right)\sqrt{m_{i(1)}}}{\sqrt{V_{i(1)}}}\right), \tag{5}$$

where $\lambda_i = \dfrac{p_i - H_{\zeta_i^*}^J}{1 - H_{\zeta_i^*}^J}$. By using equation (5), we can obtain the posterior distribution of ω_{il} as

$\omega_{il}^J = \exp\left(\eta_{pi}^J - \eta_{pl}^J\right)$. As a result, the $\left(1-\varepsilon\right)100\%$ two-sided SCI for ω_{il} is based on the HPD-

Jeffreys prior, which is provided by

$$CI^{HPD.J}_{\omega_{il}^J} = \left[L^{HPD.J}_{\omega_{il}^J}, U^{HPD.J}_{\omega_{il}^J} \right], \tag{6}$$

where $L^{HPD.J}_{\omega_{il}^J}$ and $U^{HPD.J}_{\omega_{il}^J}$ are the lower and upper bounds of the HPD interval of ω_{il}^J, respectively.

Jeffreys Rule Prior

Lee [22] introduced the Jeffreys rule prior, which is imposed as $P_{JR}\left(\sigma_i^2\right) \propto \dfrac{1}{\sigma_i^3}$. According to Maneerat

et al. [6], the posterior densities of σ_i^2 and μ_i are

$$\sigma_{JR.i}^2\big|a_{ij} \sim Inv-Gamma\left(\frac{m_{i(1)}+1}{2},\frac{(m_{i(1)}+1)c_{i(1)}^2}{2}\right), \quad \text{and} \quad \mu_{JR.i}\big|\sigma_i^2,a_{ij} \sim N\left(\bar{b}_{i(1)},\frac{\sigma_{JR.i}^2}{m_{i(1)}}\right),$$

respectively. We can calculate the posterior distributions of ζ_i^* and η_{pi} as

$$H_{\zeta_i^*}^{JR} = M^{-1}\left(X_i O_i;\, m_{i(0)}+\frac{1}{2},\, m_{i(1)}+\frac{3}{2}\right) \text{and}\, \eta_{pi}^{JR} = \mu_{JR.i} + \sqrt{\frac{\sigma_{JR.i}^2}{m_{i(1)}}}\left(\frac{Z_i + \Phi^{-1}(\xi_i)\sqrt{m_{i(1)}}}{\sqrt{V_{i(1)}}}\right),$$

where $\xi_i = \dfrac{p_i - H_{\zeta_i^*}^{JR}}{1 - H_{\zeta_i^*}^{JR}}$. It is possible to get the posterior distribution for ω_{il} as

$$\omega_{il}^{JR} = \exp\left(\eta_{pi}^{JR} - \eta_{pl}^{JR}\right).$$

Accordingly, the $(1-\varepsilon)100\%$ two-sided SCI for ω_{il} is based on the HPD-Jeffreys rule prior, which is made obtainable by

$$CI_{\omega_{il}^{JR}}^{HPD.JR} = \left[L_{\omega_{il}^{JR}}^{HPD.JR}, U_{\omega_{il}^{JR}}^{HPD.JR}\right], \tag{7}$$

where $L_{\omega_{il}^{JR}}^{HPD.JR}$ and $U_{\omega_{il}^{JR}}^{HPD.JR}$ are the lower and upper bounds of the HPD interval of ω_{il}^{JR}, respectively.

Uniform Prior

According to Kalkur and Rao [23], the uniform priors of ζ_i^* and σ_i^2 have a relationship that is proportional to 1. Therefore, the $P_U\left(\zeta_i^*,\sigma_i^2\right) \propto 1$ value serves as the uniform prior for a DLN distribution. According to Yosboonruang, *et al.* [24], the posterior distributions of σ_i^2 and μ_i are

$$\sigma_{U.i}^2\big|a_{ij} \sim Inv-Gamma\left(\frac{m_{i(1)}-2}{2},\frac{(m_{i(1)}-2)c_{i(1)}^2}{2}\right)$$

and $\mu_{U.i}\big|\sigma_i^2,a_{ij} \sim N\left(\bar{b}_{i(1)},\frac{\sigma_{U.i}^2}{m_{i(1)}}\right)$, respectively. We can determine the posterior distributions of ζ_i^* , η_{pi} , and ω_{il} as follows:

$$H_{\zeta_i^*}^{U} = M^{-1}\left(X_i O_i;\, m_{i(0)}+1,\, m_{i(1)}+1\right),$$

$$\eta_{pi}^{U} = \mu_{U.i} + \sqrt{\frac{\sigma_{U.i}^2}{m_{i(1)}}}\left(\frac{Z_i + \Phi^{-1}(v_i)\sqrt{m_{i(1)}}}{\sqrt{V_{i(1)}}}\right),$$

and $\omega_{il}^{U} = \exp\left(\eta_{pi}^{U} - \eta_{pl}^{U}\right)$, where $v_i = \dfrac{p_i - H_{\zeta_i^*}^{U}}{1 - H_{\zeta_i^*}^{U}}$.

Subsequently, the $(1-\varepsilon)100\%$ two-sided SCI for ω_{il} is based on the HPD-Uniform prior, that is supplied by

$$CI^{HPD.U}_{\omega^{U}_{il}} = \left[L^{HPD.U}_{\omega^{U}_{il}}, U^{HPD.U}_{\omega^{U}_{il}} \right],$$

(8)

where $L^{HPD.U}_{\omega^{U}_{il}}$ and $U^{HPD.U}_{\omega^{U}_{il}}$ are the lower and upper bounds of the HPD interval of ω^{U}_{il}, respectively.

Results and Discussion

In this simulation study, we conducted a performance comparison of four methods, which are denoted as FGCI, HPD.J, HPD.JR, and HPD.U, as described in equations (4), (6), (7), and (8), respectively. This comparison was carried out using the statistical software R through Monte Carlo simulations. We evaluated and compared the coverage probabilities (CPs) that were greater than or equal to the nominal confidence level of 0.95, along with the shortest expected lengths (ELs). To conduct the simulations, we generated 10,000 sets of random samples from the DLN distribution and 5,000 pivotal quantities for both FGCI and HPD. This research provides a flowchart that explains the simulation study process. It is shown in Figure 1. The variables in the study were set as follows: The number of samples k was either 3 or 5.

Additionally, we configured the sample sizes and parameters as follows: 30, 50, 100, and 200; $\mu_i = 0$;

$\sigma_i^2 = 0.5$ and 1.0; $\zeta_i = 0.1$, 0.3, and 0.5.

Based on the simulation results in Table 1 and Figures 2 and 3, when $k = 3$, the FGCI method provides CPs that exceed the specified values and exhibits superior performance compared to other methods when the sample size is the same and sample sizes are equal to 30 and 50. As for the HPD.U method, it yields CPs that are close to the desired values and are stable in all cases studied, while also outperforming other methods when the sample sizes are equal to 100 and 200. Meanwhile, the HPD.J method produces CPs that are nearly in line with the specified values in almost all cases studied and offers better Els than both the FGCI and HPD.U methods when the sample sizes are not equal. Additionally, the HPD.JR method, despite yielding shorter Els than the HPD.J and HPD.U methods in all cases, provides CPs that are lower than the nominal confidence level of 0.95 in almost all cases as well.

Based on the simulation results in Table 2 and Figures 4 and 5, when $k = 5$ and the sample sizes are equal, it was observed that for the CPs with fixed and specified criteria in the HPD.U and HPD.J methods. Then, the FGCI method yields CPs that closely approximate the nominal confidence level of 0.95, especially when the sample sizes are 30 and 50. On the other hand, the HPD.JR method consistently produces CPs that are lower than the specified criteria in almost all cases studied. When assessing the Els values, it became apparent that the FGCI method yielded the shortest intervals, followed by HPD.JR, HPD.J, and, finally, HPD.U. However, when the sample sizes are not equal but $k = 5$, it was found that, in all cases studied, the FGCI method still results in CPs that are lower than the nominal confidence level of 0.95. On the other hand, in the HPD.J, HPD.JR, and HPD.U methods, the majority of the CPs are close to and aligned with the specified criteria. Comparing the Els values, it is evident that HPD.JR consistently yields the shortest intervals among all methods, followed by HPD.J.

Importantly, when considering the term of σ_i^2, it was observed that $k = 0.5$ outperforms $\sigma_i^2 = 1.0$. This implies that a smaller value of sigma leads to more accurate and reliable results for the statistical analyses under consideration. Regarding the term ζ_i, an increase in the value of ζ_i results in higher Els values. This indicates that a larger ζ_i tends to lead to wider confidence intervals, which may result in less precise estimations. Conversely, as the sample size increases, all methods consistently exhibit a decreasing trend in Els values. This suggests that larger sample sizes generally lead to more precise estimations and narrower confidence intervals, as indicated in Tables 1–2 and Figures 2–5.
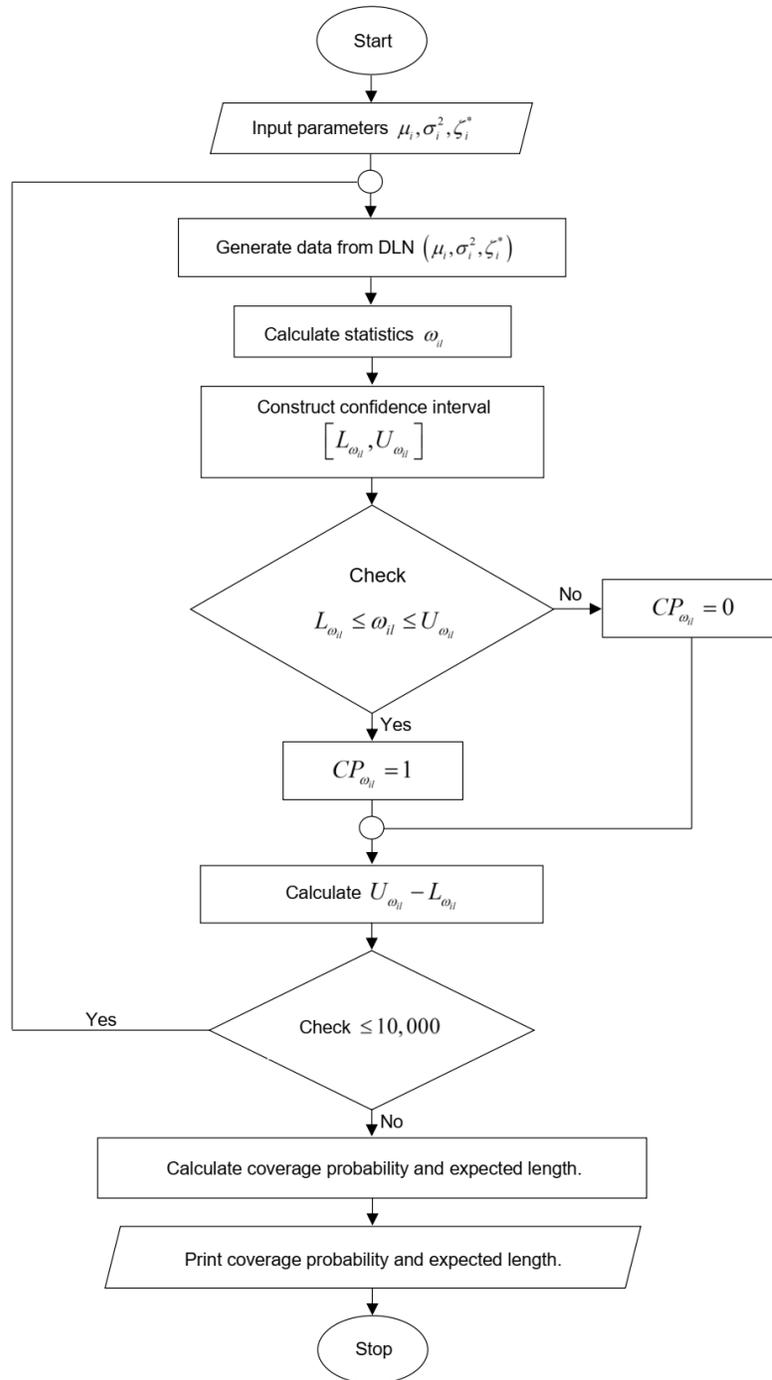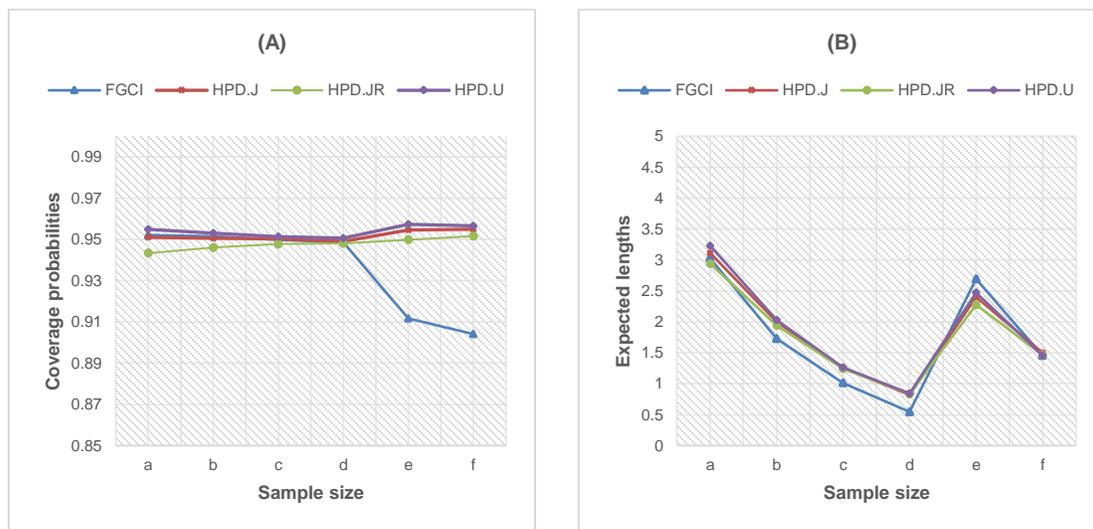
Figure 1. A flowchart of the simulation study

Table 1. The coverage probabilities and expected lengths for the 95% SCIs for $k = 3$

$m_1 : ... : m_3$	$\sigma_1^2 : ... : \sigma_3^2$	$\zeta_1 : ... : \zeta_3$	Coverage probabilities				Expected lengths			
			FGCI	HPD.J	HPD.JR	HPD.U	FGCI	HPD.J	HPD.JR	HPD.U
30^3	0.5^3	0.1^3	**0.9515**	**0.9536**	0.9463	**0.9564**	*1.7893*	1.9430	1.8713	1.9809
		0.3^3	**0.9537**	**0.9522**	0.9444	**0.9554**	*2.1625*	2.2947	2.1792	2.3617
		0.5^3	**0.9547**	**0.9521**	0.9439	**0.9571**	*2.5610*	2.6419	2.4713	2.7497
	1.0^3	0.1^3	**0.9539**	**0.9518**	0.9464	**0.9548**	*2.9736*	3.3134	3.0063	3.2009
		0.3^3	**0.9527**	**0.9512**	0.9432	**0.9552**	*3.7971*	3.8805	3.6463	4.0138
		0.5^3	**0.9509**	0.9472	0.9374	**0.9520**	*4.8131*	4.7413	4.3733	4.9822
50^3	0.5^3	0.1^3	**0.9507**	**0.9504**	0.9468	**0.9519**	*1.1620*	1.3618	1.3350	1.3747
		0.3^3	**0.9506**	**0.9507**	0.9460	**0.9536**	*1.3331*	1.5595	1.5190	1.5800
		0.5^3	**0.9508**	**0.9512**	0.9456	**0.9549**	*1.5243*	1.7711	1.7129	1.8049
	1.0^3	0.1^3	0.9490	0.9487	0.9440	**0.9503**	1.7835	2.0674	2.0237	*2.0880*
		0.3^3	**0.9504**	**0.9509**	0.9457	**0.9529**	*2.1042*	2.4337	2.3664	2.4684
		0.5^3	**0.9505**	0.9474	0.9416	**0.9507**	*2.4361*	2.7263	2.6876	2.8442
100^3	0.5^3	0.1^3	**0.9515**	**0.9514**	0.9487	**0.9518**	*0.7287*	0.9021	0.8940	0.9060
		0.3^3	0.9477	0.9496	0.9463	**0.9503**	0.8130	1.0107	0.9993	*1.0166*
		0.5^3	0.9496	0.9492	0.9462	**0.9502**	0.9082	1.1315	1.1150	*1.1404*
	1.0^3	0.1^3	0.9488	0.9495	0.9475	**0.9509**	1.0687	1.3192	1.3069	*1.3250*
		0.3^3	0.9475	0.9452	0.9424	**0.9500**	1.2006	1.4874	1.6494	*1.4960*
		0.5^3	0.9489	**0.9500**	0.9480	**0.9517**	1.3548	1.6830	1.6582	*1.6367*
200^3	0.5^3	0.1^3	**0.9520**	**0.9508**	**0.9508**	**0.9519**	*0.4845*	0.6186	0.6160	0.6199
		0.3^3	0.9485	**0.9489**	0.9448	**0.9505**	0.5326	0.6839	0.6802	*0.6857*
		0.5^3	0.9489	0.9495	0.9474	**0.9502**	0.5923	0.7641	0.7589	*0.7669*
	1.0^3	0.1^3	0.9492	0.9476	0.9466	**0.9500**	0.6963	0.8877	0.8838	*0.8896*
		0.3^3	0.9481	0.9483	0.9471	**0.9500**	0.7715	0.9901	0.9847	*0.9926*
		0.5^3	0.9490	**0.9501**	0.9490	**0.9517**	0.8588	*1.1069*	1.0994	1.1111
30:50:100	0.5^3	0.1^3	0.9147	**0.9537**	0.9481	**0.9557**	1.5332	*1.4934*	1.4480	1.5155
		0.3^3	0.9121	**0.9529**	0.9477	**0.9563**	1.8924	*1.7831*	1.7092	1.8252
		0.5^3	0.9059	**0.9590**	**0.9514**	**0.9627**	2.3099	2.1032	*1.9888*	2.1757
	1.0^3	0.1^3	0.9134	**0.9528**	0.9485	**0.9544**	2.6495	*2.3838*	2.2994	2.4233
		0.3^3	0.9135	**0.9556**	**0.9506**	**0.9585**	3.4281	2.9502	*2.8033*	3.0339
		0.5^3	0.9043	**0.9571**	**0.9506**	**0.9603**	4.5129	3.7060	*3.4564*	3.7105
50:100:200	0.5^3	0.1^3	0.9065	**0.9537**	**0.9514**	**0.9549**	0.9477	1.0191	*1.0034*	1.0256
		0.3^3	0.9055	**0.9547**	**0.9517**	**0.9560**	1.0948	1.1679	*1.1454*	1.1792
		0.5^3	0.9009	**0.9572**	**0.9524**	**0.9585**	1.2721	1.3406	*1.3076*	1.3601
	1.0^3	0.1^3	0.9104	**0.9534**	**0.9508**	**0.9540**	1.4987	1.5293	*1.5041*	1.5400
		0.3^3	0.9067	**0.9547**	**0.9514**	**0.9559**	1.7737	1.7892	*1.7512*	1.8083
		0.5^3	0.9010	**0.9595**	**0.9551**	**0.9607**	2.1584	2.1430	*2.0836*	2.1789

Notes: Bold indicates CP values greater than or equal to 0.95, and slant indicates the shortest EL values. $m^k = m_1 : m_2 : ... : m_k$

Figure 2. Graphs comparing the performance of the proposed methods for $k = 3$ with respect to the (A) CPs and (B) ELs for various sample sizes (a=30^3, b=50^3, c=100^3, d=200^3, e=30:50:100, and f=50:100:200)
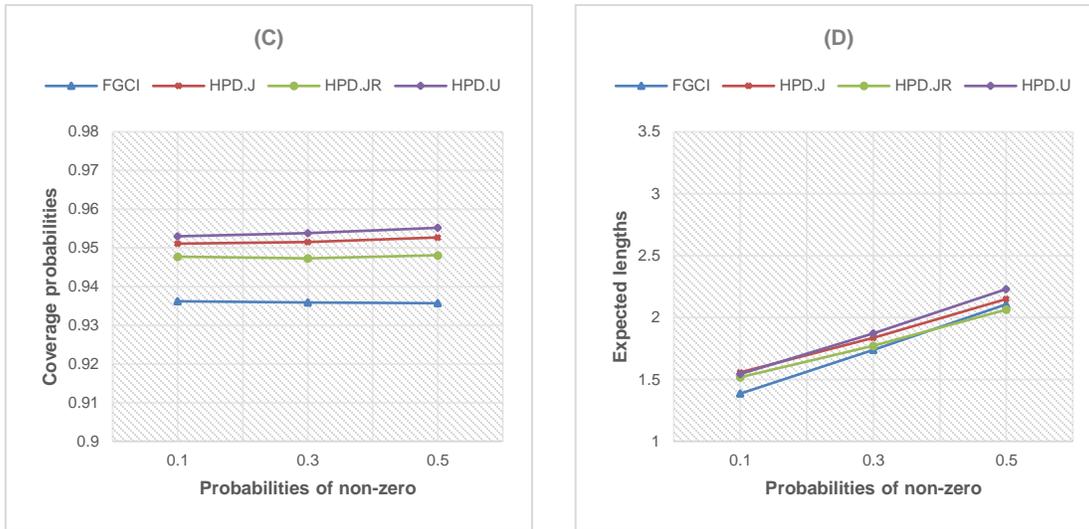
Figure 3. Graphs comparing the performance of the proposed methods for $k = 3$ with respect to the (C) CPs and (D) ELs for probabilities of non-zero

Table 2. The coverage probabilities and expected lengths for the 95% SCIs for $k = 5$

$m_1 : ... : m_3$	$\sigma_1^2 : ... : \sigma_3^2$	$\zeta_1 : ... : \zeta_3$	Coverage probabilities				Expected lengths			
			FGCI	HPD.J	HPD.JR	HPD.U	FGCI	HPD.J	HPD.JR	HPD.U
30^5	0.5^3	0.1^5	**0.9519**	**0.9516**	0.9458	**0.9548**	*1.7748*	1.9261	1.8551	1.9633
		0.3^5	**0.9526**	**0.9521**	0.9440	**0.9555**	*2.1647*	2.2964	2.1800	2.3630
		0.5^5	**0.9519**	**0.9509**	0.9416	**0.9558**	*2.5838*	2.6663	2.4927	2.7752
	1.0^3	0.1^5	**0.9521**	**0.9501**	0.9442	**0.9530**	*2.9687*	3.1310	3.0023	3.1986
		0.3^5	**0.9508**	**0.9501**	0.9426	**0.9540**	*3.8255*	3.9165	3.6877	4.0483
		0.5^5	**0.9533**	**0.9513**	0.9418	**0.9560**	*4.8470*	4.7316	4.4188	5.0326
50^5	0.5^3	0.1^5	**0.9509**	**0.9511**	0.9474	**0.9535**	*1.1607*	1.3603	1.3334	1.3734
		0.3^5	**0.9518**	**0.9512**	0.9465	**0.9534**	*1.3315*	1.5558	1.5160	1.5771
		0.5^5	**0.9521**	**0.9517**	0.9459	**0.9550**	*1.5264*	1.7737	1.7159	1.8070
	1.0^3	0.1^5	0.9498	0.9499	0.9466	**0.9517**	1.7921	2.0796	2.0362	*2.1005*
		0.3^5	**0.9507**	0.9489	0.9448	**0.9514**	*2.0951*	2.4198	2.3536	2.4542
		0.5^5	**0.9531**	**0.9504**	0.9447	**0.9530**	*2.4638*	2.8102	2.7090	2.8676
100^5	0.5^3	0.1^5	**0.9512**	**0.9506**	0.9487	**0.9514**	*0.7290*	0.9024	0.8944	0.9064
		0.3^5	0.9485	0.9481	0.9455	**0.9500**	0.8112	1.0077	0.9962	*1.0135*
		0.5^5	0.9497	**0.9513**	0.9482	**0.9523**	0.9093	1.1335	1.1170	*1.1426*
	1.0^3	0.1^5	0.9498	0.9497	0.9478	**0.9505**	1.0696	1.3203	1.3080	*1.3260*
		0.3^5	0.9495	**0.9509**	0.9489	**0.9521**	1.2030	*1.4906*	1.4732	1.4992
		0.5^5	0.9498	**0.9509**	0.9470	**0.9513**	1.3580	*1.6870*	1.6619	1.7007
200^5	0.5^3	0.1^5	0.9486	0.9489	0.9484	**0.9509**	0.4849	0.6191	0.6164	*0.6204*
		0.3^5	0.9475	0.9480	0.9472	**0.9500**	0.5324	0.6832	0.6796	*0.6850*
		0.5^5	0.9492	0.9493	0.9483	**0.9508**	0.5916	0.6303	0.7579	*0.7658*
	1.0^3	0.1^5	0.9476	0.9486	0.9478	**0.9502**	0.0447	0.8913	0.8873	*0.8932*
		0.3^5	0.9486	0.9486	0.9471	**0.9508**	0.7700	0.9880	0.9827	*0.9907*
		0.5^5	0.9498	**0.9504**	0.9491	**0.9510**	0.8585	*1.1071*	1.0996	1.1112

Table 2. (*continue*)

$m_1 : \ldots : m_3$	$\sigma_1^2 : \ldots : \sigma_3^2$	$\zeta_1 : \ldots : \zeta_3$	Coverage probabilities				Expected lengths			
			FGCI	HPD.J	HPD.JR	HPD.U	FGCI	HPD.J	HPD.JR	HPD.U
30^2:50:100^2	0.5^3	0.1^5	0.9146	**0.9536**	0.9491	**0.9560**	1.5226	*1.4943*	1.4482	1.5175
		0.3^5	0.9129	**0.9552**	0.9492	**0.9578**	1.8730	*1.7796*	1.7035	1.8220
		0.5^5	0.9094	**0.9569**	**0.9523**	**0.9601**	2.2844	2.0906	*1.9745*	2.1643
	1.0^3	0.1^5	0.9131	**0.9526**	0.9489	**0.9550**	2.6306	*2.3918*	2.3065	2.4337
		0.3^5	0.9120	**0.9532**	0.9473	**0.9559**	3.4056	*2.9627*	2.8124	3.0492
		0.5^5	0.9080	**0.9560**	0.9517	**0.9592**	4.4443	3.6706	*3.4184*	3.8355
50^2:100:200^2	0.5^3	0.1^5	0.9082	**0.9512**	0.9486	**0.9523**	0.9493	*1.0273*	1.0112	1.0343
		0.3^5	0.9001	**0.9558**	**0.9520**	**0.9572**	1.0873	1.1241	*1.1005*	1.1357
		0.5^5	0.9010	**0.9566**	**0.9530**	**0.9588**	1.2792	1.3564	*1.3218*	1.3770
	1.0^3	0.1^5	0.9086	**0.9529**	**0.9502**	**0.9543**	1.4961	1.5405	*1.5146*	1.1551
		0.3^5	0.9061	**0.9556**	**0.9524**	**0.9576**	1.7767	1.8102	*1.7703*	1.8310
		0.5^5	0.9008	**0.9571**	**0.9536**	**0.9592**	2.1374	2.1401	*2.0785*	2.1778

Notes: Bold indicates CP values greater than or equal to 0.95, and slant indicates the shortest EL values. $m^k = m_1 : m_2 : \ldots : m_k$
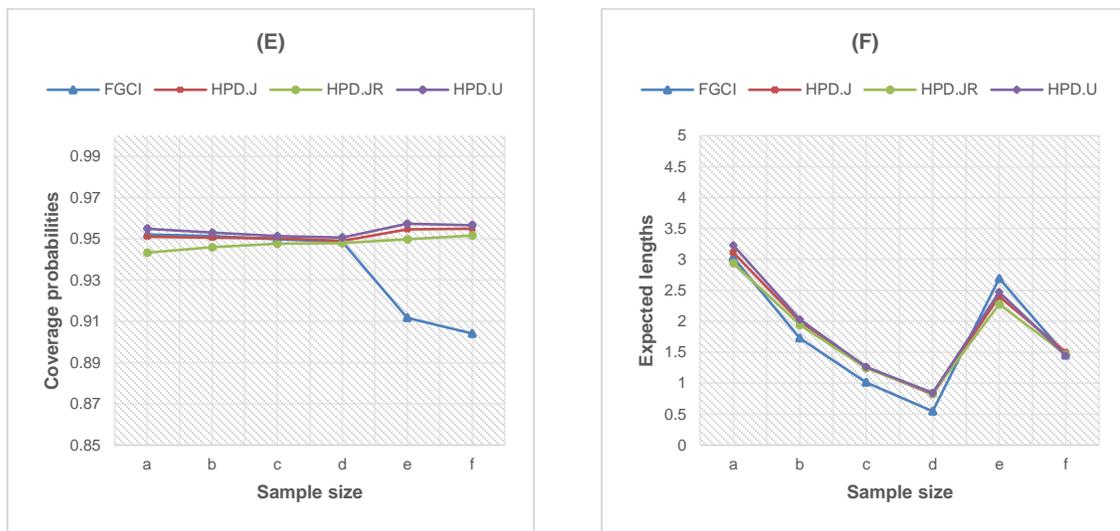

Figure 4. Graphs comparing the performance of the proposed methods for $k = 5$ with respect to the (E) CPs and (F) ELs for various sample sizes (a=30^5, b=50^5, c=100^5, d=200^5, e=30^2:50:100^2, and f=50^2:100:200^2)
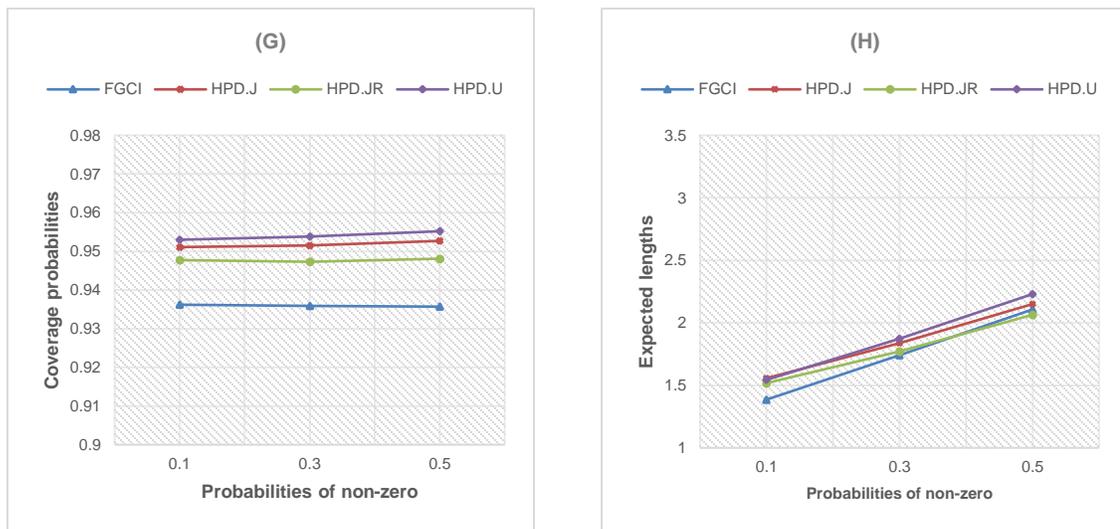

Figure 5. Graphs comparing the performance of the proposed methods for $k = 5$ with respect to the (G) CPs and (H) ELs for probabilities of non-zero

Application

Rainfall data holds immense significance when it comes to environmental management, particularly in the context of river basin management. It is a pivotal component of the larger water cycle and contributes significantly to maintaining the delicate balance of our atmosphere. Therefore, data on weather conditions and water resources becomes increasingly essential to support economic and social activities, including agriculture, industry, transportation, energy production, and water resource management. For Thailand, rainfall measurements are collected as point data at various monitoring stations for both rainfall and weather conditions within different agencies. The collected rainfall data represents only a snapshot of the weather conditions at those specific measurement locations. In this research, we utilize daily rainfall data from June 20th to July 9th, 2023, within five river basins: The Mae Klong Basin has four monitoring stations: Kanchanaburi, Thong Pha Phum, Ratchaburi, and Umphang. The Sakae Krang Basin has three monitoring stations: Kamphaeng Phet, Nakhon Sawan, and Uthai Thani. The Tha Chin Basin has three monitoring stations: Nakhon Pathom, Suphan Buri, and U Thong (agricultural). The North Mekong Basin has three monitoring stations: Phayao, Chiang Rai, and Chiang Rai (agricultural). The Ping Basin has five monitoring stations: Kamphaeng Phet, Lamphun, Tak, Bhumibol Dam, and Chiang Mai. Table 3 displays the recorded rainfall data, while Figures 6 and 7 illustrate the histogram and Normal quantile-quantile (Q-Q) plot. In addition, the goodness of fit for the distribution is assessed using the minimum Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). The results are presented in Table 4. The analysis indicates that the lognormal model achieved the lowest AIC and BIC values, suggesting that the lognormal distribution is the most appropriate fit for this dataset. Table 5 presents descriptive statistics for the nonzero rainfall dataset in five river basins. Based on the results in Table 6, which shows the SCIs for ratios of the percentiles for the rainfall dataset in five river basins, it was found that the HPD.JR method is the most suitable approach. Following that, the second-best method is the HPD.J method. Moreover, these findings are consistent with the simulation results.

Table 3. Data on rainfall by the river basin, June 20–July 9, 2023

Mae Klong				Sakae Krang			Tha Chin			North Mekong			Ping				
0.0	0.5	8.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.1	0.0	0.0	0.0	0.0
0.0	11.1	1.1	0.0	0.0	8.7	0.0	0.0	0.0	0.0	0.6	0.0	1.8	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	4.2	0.2	5.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	4.2	0.0
0.0	6.3	18.0	0.0	0.0	0.3	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.3
0.0	6.3	0.0	0.0	0.0	1.0	0.1	0.0	0.0	0.0	33.4	35.8	26.0	0.0	0.0	0.7	0.0	14.3
0.0	1.6	0.0	0.0	0.0	0.0	0.2	0.0	2.7	0.0	2.4	6.1	6.1	0.0	0.4	0.0	0.0	6.9
0.0	1.8	0.0	11.8	0.0	2.2	0.7	0.0	0.4	0.0	0.7	0.0	2.2	0.0	0.1	0.0	11.8	12.4
0.0	2.9	0.0	1.0	0.0	0.0	0.6	0.8	0.0	0.0	2.1	0.0	0.0	0.0	0.0	5.0	1.0	0.0
0.0	0.2	34.3	0.6	1.5	0.0	0.1	0.6	0.0	1.7	0.0	47.6	2.3	1.5	0.7	1.0	0.6	34.4
73.5	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	4.8	0.0	2.8	2.0	0.0	0.0	0.0	0.0	0.4
0.7	0.0	9.7	13.3	0.8	0.4	0.0	0.4	3.2	2.6	54.0	11.2	1.2	0.8	0.7	0.0	0.0	0.0
0.0	1.8	0.0	0.0	5.4	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	5.4	33.8	5.9	0.0	5.9
4.9	22.1	19.8	3.2	8.6	0.4	5.9	2.1	0.3	2.6	0.5	0.0	0.0	8.6	0.1	7.6	0.2	7.6
0.0	13.4	0.0	5.0	26.5	0.1	0.0	0.0	0.0	0.0	19.3	0.0	0.0	26.5	5.2	29.9	1.8	29.9
0.0	30.5	0.0	3.0	5.3	0.5	5.0	0.3	4.2	1.8	0.0	0.8	0.0	5.3	8.8	0.5	12.4	0.5
15.5	0.9	0.3	2.1	0.0	0.0	0.0	3.3	0.0	9.8	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0
0.0	0.2	0.0	2.5	0.0	0.0	0.0	0.0	1.7	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.8	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	3.5	0.0	3.5
112.7	5.6	0.0	2.3	0.0	2.8	0.3	0.0	2.0	39.4	0.0	1.6	0.6	0.0	36.7	0.0	0.0	0.0
2.0	0.0	0.0	2.2	0.0	0.0	0.0	0.6	0.0	0.3	0.0	0.6	0.2	0.0	0.0	0.0	0.0	0.0

Table 4. Data on rainfall by the river basin, June 20–July 9, 2023

| Model	Mae Klong		Sakae Krang		Tha Chin		North Mekong		Ping	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Normal	376.98	380.46	175.99	178.65	176.24	178.68	227.80	230.40	323.99	327.51
Lognormal	**266.85**	**270.32**	**105.20**	**107.86**	**105.40**	**107.84**	**162.45**	**165.04**	**244.61**	**248.13**
Cauchy	307.39	310.87	135.40	138.06	123.24	125.68	186.22	188.81	302.12	305.65
Logistic	350.32	353.79	161.95	164.61	152.77	155.20	223.85	226.44	316.05	319.57
Exponential	286.73	288.47	119.32	120.65	116.49	117.71	179.11	180.40	258.27	260.03
Gamma	276.50	279.97	112.31	114.97	116.19	118.62	171.07	173.67	246.66	250.18
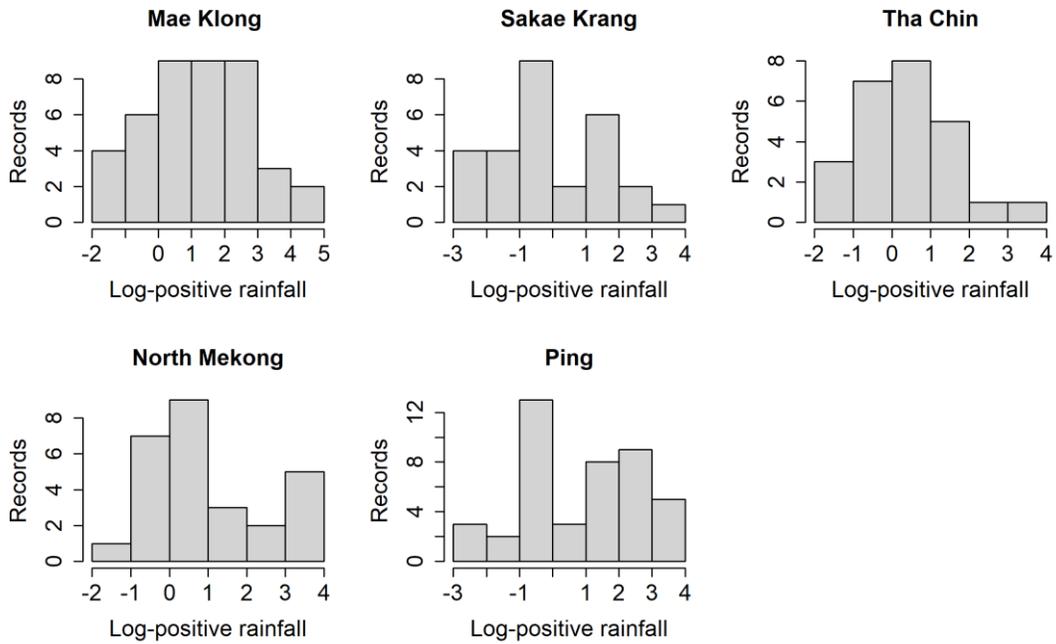
Figure 6. Histogram graph of the log-transformed positive rainfall datasets
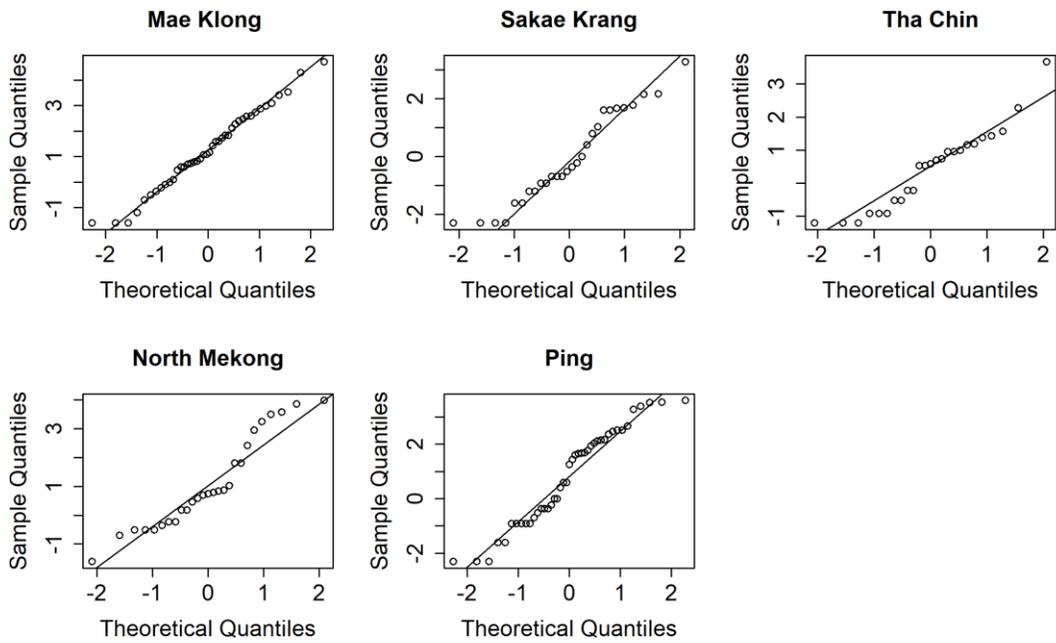

Figure 7. Normal Q-Q of the log-transformed positive rainfall datasets

Table 5. The summary of statistics for the non-zero rainfall datasets in five river basins

River basin	m_i	$m_{i(1)}$	$m_{i(0)}$	$\hat{\zeta}_i^*$	$\hat{\eta}_i$
Mae Klong	80	42	38	0.4750	3.3291
Sakae Krang	60	28	32	0.5333	1.9196
Tha Chin	60	25	35	0.5833	1.8625
North Mekong	60	21	33	0.5500	3.0102
Ping	100	43	57	0.5700	2.9203

Table 6. The SCIs for ratios of the percentiles for the rainfall datasets in five river basins

Methods	Confidence interval for ω_{il}		Length of intervals
	Lower	Upper	
FGCI	0.4251	6.0960	5.6708
HPD.J	0.2894	5.5554	5.2659
HPD.JR	**0.2702**	**5.1721**	**4.9019**
HPD.U	0.2778	5.5783	5.3005

Conclusions

In conclusion, this study contributes to the understanding of confidence interval construction methods for SCIs for ratios of the percentiles of the delta-lognormal distribution, which were proposed using four methods: FGCI, HPD.J, HPD.JR, and HPD.U. The research defines diverse sample sizes, parameter values, and Monte Carlo simulations to ensure comprehensive analysis. Additionally, the efficiency of these methods has been compared using cumulative coverage probability and the expected length of the confidence intervals. The results of the simulation study revealed that, for equal sample sizes, the FGCI method demonstrated better performance when dealing with small sample sizes. However, as the sample size increases, the HPD.U method shows superior efficiency compared to the other methods. Furthermore, for unequal sample sizes, all three HPD methods exhibit stability in terms of coverage probability; notably, HPD.JR consistently yields the shortest expected length across almost all cases studied. Importantly, it becomes evident that with increasing sample sizes, the confidence intervals tend to become more efficient. Overall, based on the findings, the HPD method outperforms the FGCI method. Consequently, the HPD method is recommended for constructing simultaneous confidence intervals for ratios of the percentiles of the delta-lognormal distribution.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

References

[1] Huang, L. F. (2017). Approximated non parametric confidence regions for the ratio of two percentiles. *Communications in Statistics-Theory and Methods*, *46*(8), 4004-4015.

[2] Arachchige, C. N., Cairns, M., & Prendergast, L. A. (2021). Interval estimators for ratios of independent quantiles and interquantile ranges. *Communications in Statistics-Simulation and Computation*, *50*(*12*), 3914-3930.

[3] Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, *50*(*271*), 901-908.

[4] Zhou, X. H., & Tu, W. (2000). Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics*, *56*(*4*), 1118-1125.

[5] Hasan, M. S., & Krishnamoorthy, K. (2018). Confidence intervals for the mean and a percentile based on zero-inflated lognormal data. *Journal of Statistical Computation and Simulation*, *88*(*8*), 1499-1514.

[6] Maneerat, P., Niwitpong, S. A., & Niwitpong, S. (2020). A Bayesian approach to construct confidence intervals for comparing the rainfall dispersion in Thailand. *PeerJ*, *8*, e8502.

[7] Yosboonruang, N., Niwitpong, S. A., & Niwitpong, S. (2021). Simultaneous confidence intervals for all pairwise differences between the coefficients of variation of rainfall series in Thailand. *PeerJ*, *9*, e11651.

[8] Thangjai, W., Niwitpong, S. A., & Niwitpong, S. (2022). Estimation of common percentile of rainfall datasets in Thailand using delta-lognormal distributions. *PeerJ*, *10*, e14498.

[9] Thangjai, W., Niwitpong, S. A., Niwitpong, S., & Smithpreecha, N. (2023). Confidence interval estimation for the ratio of the percentiles of two delta-lognormal distributions with application to rainfall data. *Symmetry*, *15*(4), 794.

[10] Malley, J. D. (1982). Simultaneous confidence intervals for ratios of normal means. *Journal of the American Statistical Association*, *77*(*377*), 170-176.

[11] Kumar, N., Misra, N., & Gill, A. N. (1994). Simultaneous confidence intervals for all ratios to the best: The

exponential distribution. *IEEE Transactions on Reliability*, *43*(1), 61-64.

[12] Rublík, F. (2010). A note on simultaneous confidence intervals for ratio of variances. *Communications in Statistics—Theory and Methods*, *39*(6), 1038-1045.

[13] Sadooghi-Alvandi, S. M., & Malekzadeh, A. (2014). Simultaneous confidence intervals for ratios of means of several lognormal distributions: a parametric bootstrap approach. *Computational Statistics & Data Analysis*, *69*, 133-140.

[14] Schaarschmidt, F., & Djira, G. D. (2016). Simultaneous confidence intervals for ratios of fixed effect parameters in linear mixed models. *Communications in Statistics-Simulation and Computation*, *45*(*5*), 1704-1717.

[15] Maneerat, P., & Niwitpong, S. A. (2021). Multiple comparisons of precipitation variations in different areas using simultaneous confidence intervals for all possible ratios of variances of several zero-inflated lognormal models. *PeerJ*, *9*, e12659.

[16] Zhang, Q., Xu, J., Zhao, J., Liang, H., & Li, X. (2022). Simultaneous confidence intervals for ratios of means of zero-inflated log-normal populations. *Journal of Statistical Computation and Simulation*, *92*(*6*), 1113-1132.

[17] Kaewprasert, T., Niwitpong, S. A., & Niwitpong, S. (2022). Simultaneous Confidence Intervals for the Ratios of the Means of Zero-Inflated Gamma Distributions and Its Application. *Mathematics*, *10*(24), 4724.

[18] Meteorological. (2023). The climate of Thailand, *Thai Meteorological Development Bureau*. Retrieved from https://swocrf.rid.go.th/.

[19] Tian, L., & Wu, J. (2006). Confidence intervals for the mean of lognormal data with excess zeros. *Biometrical Journal*, *48*(1), 149-156.

[20] Hannig, J., Iyer, H., & Patterson, P. (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, *101*(473), 254-269.

[21] Jeffreys, H. (1961). Small corrections in the theory of surface waves. *Geophysical Journal International*, *6*(1), 115-117.

[22] Lee, P.M. (2012). *Bayesian statistics: An introduction*, John Wiley & Sons Ltd, UK.

[23] Kalkur, T. A., & Rao, A. (2017). Bayes estimator for coefficient of variation and inverse coefficient of variation for the normal distribution. *International Journal of Statistics and Systems*, *12*(4), 721-732.

[24] Yosboonruang, N., Niwitpong, S. A., & Niwitpong, S. (2019). Measuring the dispersion of rainfall using Bayesian confidence intervals for coefficient of variation of delta-lognormal distribution: a study from Thailand. *PeerJ*, *7*, e7344.