

Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 5, No. 2, April, 2021

EMERGING SCIENCE JOURNAL



Bayesian Confidence Intervals for Coefficients of Variation of PM10 Dispersion

Warisa Thangjai¹, Sa-Aat Niwitpong^{2*}, Suparat Niwitpong²

¹ Department of Statistics, Faculty of Science, Ramkhamhaeng University, Bangkok, 10240 Thailand

² Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800 Thailand

Abstract

Herein, we propose the Bayesian approach for constructing the confidence intervals for both the coefficient of variation of a log-normal distribution and the difference between the coefficients of variation of two log-normal distributions. For the first case, the Bayesian approach was compared with large-sample, Chi-squared, and approximate fiducial approaches via Monte Carlo simulation. For the second case, the Bayesian approach was compared with the method of variance estimates recovery (MOVER), modified MOVER, and approximate fiducial approaches using Monte Carlo simulation. The results show that the Bayesian approach provided the best approach for constructing the confidence intervals for both the coefficient of variation of a log-normal distribution and the difference between the coefficients of variation of two log-normal distributions. To illustrate the performances of the confidence limit construction approaches with real data, they were applied to analyze real PM10 datasets from the Nan and Chiang Mai provinces in Thailand, the results of which are in agreement with the simulation results.

Keywords:

Bayesian Approach;
Coefficient of Variation;
Difference; Log-normal Distribution;
Monte Carlo Simulation.

Article History:

Received:	20	December	2020
Revised:	22	February	2021
Accepted:	02	March	2021
Published:	01	April	2021

1- Introduction

The log-normal distribution is important in statistics because log-transformation of skewed data can bring symmetry to the data distribution. Therefore, this distribution is used to model data in real-life applications and has been used to measure the size of living tissue, certain physiological characteristics, and rainfall. Furthermore, it has been used to analyze data in medical, pharmaceutical, economics, and hydrological studies. Specifically, the log-normal distribution has been used to analyze the survival times of breast and ovarian cancer patients [1], pharmacokinetic parameters [2], extreme values of daily rainfall [3], and air pollution.

In 2019, the air quality in Thailand was worse than in previous years because of smog exacerbated by human activities. One of its components, Particulate Matter (PM), which is a complex mixture of particles and liquid droplets, has steadily increased. PM is categorized by particle size: PM2.5 and PM10 with diameters of 2.5 μm and 10 μm or less, respectively. PM is produced from various sources and activities, including construction work, landfill, agriculture, wildfires, waste burning, industrial sources, and motor vehicles. PM10 is of particular interest because high levels can lead to poor health and respiratory disease. The deleterious effects of PM10 exposure are chest pain, general respiratory discomfort due to lung tissue damage, cancer, and premature death. Especially, its effects on children, the elderly, and people with chronic lung disease are of primary concern.

* **CONTACT:** Sa-aat.n@sci.kmutnb.ac.th

DOI: <http://dx.doi.org/10.28991/esj-2021-01264>

© 2021 by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The problem of interval estimation for the parameters of a log-normal distribution has received considerable attention in the literature. For instance, Krishnamoorthy and Mathew [4], Hannig et al. [5], Tian and Wu [6], Harvey and van der Merwe [7], and Lin and Wang [8] presented confidence intervals for functions of the means of log-normal distributions. Rao and D'Cunha [9] proposed a novel approach to obtain inference using the median of a log-normal distribution. Thangjai et al. [10], Nam and Kwon [11], Hasan and Krishnamoorthy [12], Niwitpong et al. [13], Thangjai and Niwitpong [14], and Thangjai et al. [15] suggested confidence intervals for functions of the coefficient of variation of a log-normal distribution. Furthermore, Thangjai and Niwitpong [16, 17] introduced approaches to estimate the confidence intervals for functions of the signal-to-noise ratio of log-normal distributions.

The coefficient of variation is defined as the standard deviation divided by the mean. It has been used as a measure of precision within and between data series. For example, Tsim et al. [18] used the coefficient of variation to analyze blood samples taken from different laboratories. Faupel-Badger et al. [19] used the coefficient of variation to compare the concentrations of an estrogen metabolite measured using two methods. In environmental studies, the coefficient of variation has often been used for measuring the daily air quality [20]. Pollution levels in different areas can be compared using the coefficient of variation: larger coefficient of variation values indicate greater dispersion whereas smaller ones indicate lower risk. The difference between the air pollution levels in two areas can be analyzed by comparing their coefficients of variation.

Niwitpong [21] proposed confidence intervals for the coefficient of variation of log-normal distribution with restricted parameter space. Nam and Kwon [11] and Hasan and Krishnamoorthy [12] constructed the confidence intervals for the ratio of coefficients of variation of two log-normal distributions. Thangjai et al. [22] proposed Bayesian confidence intervals for the coefficient of variation and the difference between the coefficients of variation of two normal distributions. For k coefficients of variation, Ng [23] estimated confidence intervals for the common coefficient of variation of log-normal populations. Thangjai et al. [10] presented simultaneous confidence intervals for the differences between the coefficients of variation of log-normal distributions. Nam and Kwon [11] proposed the method of variance estimates recovery (MOVER) approach for constructing a confidence interval for the ratio of coefficients of variation of two log-normal distributions, which Hasan and Krishnamoorthy [12] later modified. Although the MOVER approach is easy to compute using an exact formula, it is based on the initial confidence interval for a single parameter of interest. In addition, Hasan and Krishnamoorthy [12] presented the approximate fiducial approach for constructing the confidence interval for the ratio of coefficients of variation. Although this approach was very simple, it was based on simulated data.

In statistics, classical and Bayesian inference are fundamentally different. In classical inference, the parameter is unknown but fixed and its value is based on the observed values in a sample. The Bayesian approach uses a prior distribution based on the experimenter's belief that is updated with the sample information. Subsequently, the posterior distribution is used to update the prior by using Bayes' rule. In this study, the Bayesian approach was applied for constructing confidence intervals for the coefficient of variation and the difference between the coefficients of variation of log-normal distributions. The Bayesian approach is based on combining the likelihood function and the prior distribution. Depending on the choice of prior distribution, we show that the Bayesian approach has equal or better coverage accuracy and shorter average lengths than the classical approaches.

The rest of this article is organized as follows. In Section 2, confidence intervals for the coefficient of variation of a log-normal distribution are presented, while those for the difference between the coefficients of variation of log-normal distributions are given in Section 3. In Section 4, the results of simulation studies are presented. PM10 datasets from Nan and Chiang Mai provinces are used to illustrate the performances of the confidence intervals in Section 5, and concluding remarks are presented in Section 6.

2- Confidence Intervals for the Coefficient of Variation

If $X = \ln(Y)$ follows a normal distribution with mean μ and variance σ^2 , then Y follows a log-normal distribution with parameters μ and σ^2 . The mean and variance of Y are $\mu_Y = \exp(\mu + (\sigma^2/2))$ and $\sigma_Y^2 = (\exp(\sigma^2) - 1) \cdot (\exp(2\mu + \sigma^2))$, respectively. The coefficient of variation of Y , is defined as a ratio of standard deviation and mean of Y , is expressed as $\tau = \sqrt{\exp(\sigma^2) - 1}$. Since the coefficient of variation of the log-normal distribution depends on parameter σ^2 only. While the coefficient of variation of the normal distribution, is defined as σ/μ , depends on mean μ and variance σ^2 .

Let \bar{X} and S^2 be unbiased estimators of μ and σ^2 , respectively. Also, let \bar{x} and s^2 be observed values of \bar{X} and S^2 , respectively.

2-1- Classical Confidence Intervals for the Coefficient of Variation

In this section, there are three approaches for interval estimation of the coefficient of variation. The three approaches are large-sample, chi-squared, and approximate fiducial approaches.

2-1-1- Large-sample Confidence Intervals for the Coefficient of Variation

Large-sample confidence intervals are constructed based on two variances. First, the variance estimate of $\ln(\hat{\tau})$ is given by Nam and Kwon [11]. Let $\hat{\tau} = \sqrt{\exp(\hat{\sigma}^2) - 1}$ be the estimator of τ , where $\hat{\sigma}^2 = (n - 1)S^2/n$. The large-sample confidence interval for $\ln(\tau)$ based on the variance estimate of Nam and Kwon [11] is obtained by

$$[L_{NK}, U_{NK}] = [\ln(\hat{\tau}) - z_{1-\alpha/2} \sqrt{\hat{V}\hat{a}r(\ln(\hat{\tau}))}, \ln(\hat{\tau}) + z_{1-\alpha/2} \sqrt{\hat{V}\hat{a}r(\ln(\hat{\tau}))}], \quad (1)$$

Where $Z_{1-\alpha/2}$ denotes the $100(1 - \alpha/2)$ -th percentile of the standard normal distribution and $\hat{V}\hat{a}r(\ln(\hat{\tau})) = \hat{\sigma}^2(1 + \hat{\tau}^2)^2/(2n\hat{\tau}^4)$.

Therefore, the $100(1 - \alpha)\%$ large-sample confidence interval for the coefficient of variation τ based on the variance estimate of Nam and Kwon [11] is defined as:

$$CI_{\tau,LS1} = [L_{\tau,LS1}, U_{\tau,LS1}] = [\exp(L_{NK}), \exp(U_{NK})] \quad (2)$$

Second, the variance estimate of $\hat{\tau}$ is proposed by Thangjai et al. [10]. Let $\hat{\tau} = \sqrt{\exp(\hat{\sigma}^2) - 1}$ be the estimator of τ , where $\hat{\sigma}^2 = S^2$. Therefore, the $100(1 - \alpha)\%$ large-sample confidence interval for the coefficient of variation τ based on the variance estimate of Thangjai et al. [10] is defined as:

$$CI_{\tau,LS2} = [L_{\tau,LS2}, U_{\tau,LS2}] = [\hat{\tau} - z_{1-\alpha/2} \sqrt{\hat{V}\hat{a}r(\hat{\tau})}, \hat{\tau} + z_{1-\alpha/2} \sqrt{\hat{V}\hat{a}r(\hat{\tau})}], \quad (3)$$

Where $Z_{1-\alpha/2}$ denotes the $100(1 - \alpha/2)$ -th percentile of the standard normal distribution and $\hat{V}\hat{a}r(\hat{\tau}) = \hat{\sigma}^4(\exp(2\hat{\sigma}^2))/(2(n - 1)(\exp(\hat{\sigma}^2) - 1))$.

2-1-2- Chi-squared Confidence Interval for the Coefficient of Variation

Following Niwitpong [21], the $100(1 - \alpha)\%$ Chi-squared confidence interval for the coefficient of variation τ is defined as:

$$CI_{\tau,CHI} = [L_{\tau,CHI}, U_{\tau,CHI}] = \left[\sqrt{\exp\left(\frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2}\right) - 1}, \sqrt{\exp\left(\frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2}\right) - 1} \right], \quad (4)$$

Where $\chi_{n-1,1-\alpha/2}^2$ and $\chi_{n-1,\alpha/2}^2$ denote the $100(1 - \alpha/2)$ -th and $100(\alpha/2)$ -th percentiles of the Chi-squared distribution with $n - 1$ degrees of freedom.

2-1-3- Approximate Fiducial Confidence Interval for the Coefficient of Variation

According to Hasan and Krishnamoorthy [12], let:

$$T_{0.5} = \exp\left(\frac{(n-1)S^2}{\chi_{n-1,0.5}^2}\right) - 1, \quad (5)$$

$$T_{\alpha/2} = \exp\left(\frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2}\right) - 1, \quad (6)$$

and

$$T_{1-\alpha/2} = \exp\left(\frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2}\right) - 1, \quad (7)$$

Where $\chi_{n-1,0.5}^2$, $\chi_{n-1,\alpha/2}^2$ and $\chi_{n-1,1-\alpha/2}^2$ denote the $100(0.5)$ -th, $100(\alpha/2)$ -th and $100(1 - \alpha/2)$ -th percentiles of the Chi-squared distribution with $n - 1$ degrees of freedom, respectively.

Therefore, the $100(1 - \alpha)\%$ approximate fiducial confidence interval for the coefficient of variation τ is defined as:

$$CI_{\tau,AF} = [L_{\tau,AF}, U_{\tau,AF}] = \left[\sqrt{T_{0.5} - \sqrt{(T_{0.5} - T_{1-\alpha/2})^2}}, \sqrt{T_{0.5} + \sqrt{(T_{0.5} - T_{\alpha/2})^2}} \right], \quad (8)$$

2-2- Bayesian Confidence Interval for the Coefficient of Variation

In classical approach, the parameter θ is unknown, but it is fixed. Let X_1, X_2, \dots, X_n be a random sample from population indexed by parameter θ . Let x_1, x_2, \dots, x_n be observed values of X_1, X_2, \dots, X_n . Then the value of θ is known. In Bayesian approach, the parameter θ is considered to be a quantity. The prior distribution, based on the experimenter's belief, is updated with the sample information. The posterior distribution is updated prior with the use of Bayes' Rule: see Casella and Berger [24].

Let \bar{X} be the sample mean and let S^2 be the sample variance. Also, \bar{x} and s^2 are the observed values of \bar{X} and S^2 , respectively.

In this paper, a random sample X_1, X_2, \dots, X_n is drawn from a normal population with parameter $\theta = (\mu, \sigma^2)$. The likelihood function is:

$$L(\mu, \sigma^2 | data) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\mu - \bar{x})^2]\right\}, \quad (9)$$

The logarithm of the likelihood can be written as

$$\ln(L) = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2}[(n-1)s^2 + n(\mu - \bar{x})^2], \quad (10)$$

Therefore, the Fischer information matrix is;

$$F(\mu, \sigma^2) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2} \left(\frac{1}{\sigma^2}\right)^2 \end{bmatrix}, \quad (11)$$

Following the Fischer information matrix, the Jeffreys Independence prior is;

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2), \quad (12)$$

In this paper is interested in the coefficient of variation. Then a flat prior for coefficient of variation is defined by;

$$p(\theta) = \frac{p(\sigma)}{p(\mu)}, \quad (13)$$

Hence, the Jeffreys Independence prior is defined by;

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}, \quad (14)$$

Let $\mu | \sigma^2, x$ be the conditional posterior distribution for μ given σ^2 and x . The $\mu | \sigma^2, x$ is normal distribution with mean $\hat{\mu}$ and variance σ^2/n . It can written as;

$$\mu | \sigma^2, x \sim N\left(\hat{\mu}, \frac{\sigma^2}{n}\right), \quad (15)$$

Furthermore, let $\sigma^2 | x$ be the posterior distribution for σ^2 given x . It is inverse gamma distribution with shape parameter $(n-1)/2$ and scale parameter $(n-1)s^2/2$. That is;

$$\sigma^2 | x \sim IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right), \quad (16)$$

The posterior distribution is used to make statements for parameter which is considered a random quantity. Then the posterior distribution of the coefficient of variation can be used as a point estimate of coefficient of variation. It is denoted by:

$$\tau_{BS} = \sqrt{\exp(\sigma^2) - 1}, \quad (17)$$

Where σ^2 is simulated through Monte Carlo simulation from the posterior distribution defined as in Equation 16.

A smallest confidence interval with a specified coverage probability can be obtained using the Bayesian criteria. A highest posterior density region is used to obtain the shortest confidence interval. The posterior density region consists of the values of the parameter which is highest.

Therefore, the $100(1 - \alpha)\%$ Bayesian confidence interval for the coefficient of variation τ is defined as;

$$CI_{\tau, BS} = [L_{\tau, BS}, U_{\tau, BS}], \quad (18)$$

Where $L_{\tau, BS}$ and $U_{\tau, BS}$ are the lower and upper limits of the highest posterior density region of τ_{BS} , respectively.

Since the Bayesian confidence interval in Equation 18 can be estimated using a computational procedure given in the following algorithm.

Algorithm 1.

Step 1: Generate $\sigma^2 | x \sim IG((n-1)/2, (n-1)s^2/2)$;

Step 2: Calculate the value of τ_{BS} as given in Equation 17;

Step 3: Repeat the step 1 - step 2 for q times and obtain $\tau_{BS}^1, \tau_{BS}^2, \dots, \tau_{BS}^q$;

Step 4: Calculate $L_{\tau,BS}$ and $U_{\tau,BS}$ from the $100(\alpha/2)$ -th and $100(1 - \alpha/2)$ -th percentiles of $\tau_{BS}^1, \tau_{BS}^2, \dots, \tau_{BS}^q$.

The coverage probabilities and average lengths of the confidence intervals for the coefficient of variation can be approximated via Monte Carlo simulations using the following algorithm.

Algorithm 2. For a given n, μ, σ , and τ :

Step 1: Generate x from $N(\mu, \sigma^2)$;

Step 2: Calculate \bar{x} and s^2 ;

Step 3: Construct $CI_{\tau,LS1(h)} = [L_{\tau,LS1(h)}, U_{\tau,LS1(h)}]$;

Step 4: Construct $CI_{\tau,LS2(h)} = [L_{\tau,LS2(h)}, U_{\tau,LS2(h)}]$;

Step 5: Construct $CI_{\tau,CHI(h)} = [L_{\tau,CHI(h)}, U_{\tau,CHI(h)}]$;

Step 6: Construct $CI_{\tau,AF(h)} = [L_{\tau,AF(h)}, U_{\tau,AF(h)}]$;

Step 7: Construct $CI_{\tau,BS(h)} = [L_{\tau,BS(h)}, U_{\tau,BS(h)}]$;

Step 8: If $L_{(h)} \leq \tau \leq U_{(h)}$ set $p_{(h)} = 1$, else $p_{(h)} = 0$;

Step 9: Calculate $U_{(h)} - L_{(h)}$.

Step 10: Repeat the step 1 to step 9 for a large number of times (say, M times) and calculate coverage probability

$$\sum_{h=1}^M p_{(h)} / M \text{ and average length } \sum_{h=1}^M (U_{(h)} - L_{(h)}) / M.$$

3- Confidence Intervals for Difference between the Coefficients of Variation

Assume that $X_1 = \ln(Y_1)$ is a random sample of size n_1 from normal distribution with mean μ_1 and variance σ_1^2 . Since \bar{X}_1 is the sample mean and S_1^2 is the sample variance. Also, $X_2 = \ln(Y_2)$ is a random sample of size n_2 from normal distribution with mean μ_2 and variance σ_2^2 . And \bar{X}_2 and S_2^2 are the sample mean and sample variance, respectively. The coefficients of variation of Y_1 and Y_2 are defined as $\tau_1 = \sqrt{\exp(\sigma_1^2) - 1}$ and $\tau_2 = \sqrt{\exp(\sigma_2^2) - 1}$. Therefore, the difference between the coefficients of variation is $\delta = \tau_1 - \tau_2 = \sqrt{\exp(\sigma_1^2) - 1} - \sqrt{\exp(\sigma_2^2) - 1}$.

3-1- Classical Confidence Intervals for Difference between the Coefficients of Variation

Three approaches are used to construct the confidence intervals for the difference between the coefficients of variation of log-normal distributions.

3-1-1- MOVER Confidence Interval for Difference between the Coefficients of Variation

The variance estimate of $\ln(\hat{\tau})$ given in Nam and Kwon [11] is used. The estimator of difference between the coefficients of variation is obtained by;

$$\hat{\delta} = \hat{\tau}_1 - \hat{\tau}_2 = \sqrt{\exp(\hat{\sigma}_1^2) - 1} - \sqrt{\exp(\hat{\sigma}_2^2) - 1}, \quad (19)$$

Where $\hat{\sigma}_1^2 = (n_1 - 1)S_1^2 / n_1$ and $\hat{\sigma}_2^2 = (n_2 - 1)S_2^2 / n_2$.

Donner and Zou [25] proposed the idea of the MOVER approach for the confidence interval for difference between the parameters using the confidence interval $[l_i, u_i]$, where $i = 1, 2$. From Equation 3, the confidence intervals for the coefficient of variation of Y_1 and Y_2 are defined by;

$$[l_1, u_1] = [\hat{\tau}_1 - z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\tau}_1)}, \hat{\tau}_1 + z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\tau}_1)}]; \tag{20}$$

and

$$[l_2, u_2] = [\hat{\tau}_2 - z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\tau}_2)}, \hat{\tau}_2 + z_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\tau}_2)}], \tag{21}$$

Where $Z_{1-\alpha/2}$ denotes the $100(1 - \alpha/2)$ -th percentile of the standard normal distribution, $\widehat{Var}(\hat{\tau}_1) = \exp(\hat{\sigma}_1^2(1 + \hat{\tau}_1^2)^2 / (2n_1\hat{\tau}_1^4))$, and $\widehat{Var}(\hat{\tau}_2) = \exp(\hat{\sigma}_2^2(1 + \hat{\tau}_2^2)^2 / (2n_2\hat{\tau}_2^4))$.

The lower and upper limits of the confidence interval for the difference between the coefficients of variation based on MOVER approach are;

$$L_{\delta,MOVER} = \hat{\tau}_1 - \hat{\tau}_2 - \sqrt{(\hat{\tau}_1 - l_1)^2 + (u_2 - \hat{\tau}_2)^2}, \tag{22}$$

and

$$U_{\delta,MOVER} = \hat{\tau}_1 - \hat{\tau}_2 + \sqrt{(u_1 - \hat{\tau}_1)^2 + (\hat{\tau}_2 - l_2)^2}, \tag{23}$$

Therefore, the $100(1 - \alpha)\%$ MOVER confidence interval for difference between the coefficients of variation δ is defined as;

$$CI_{\delta,MOVER} = [L_{\delta,MOVER}, U_{\delta,MOVER}], \tag{24}$$

Where $L_{\delta,MOVER}$ and $U_{\delta,MOVER}$ are defined in Equations 22 and 23, respectively.

3-1-2- Modified MOVER Confidence Interval for Difference between the Coefficients of Variation

Suppose that the coefficient of variation estimators of Y_1 and Y_2 are defined as $\hat{\tau}_1 = \sqrt{\exp(\hat{\sigma}_1^2) - 1}$ and $\hat{\tau}_2 = \sqrt{\exp(\hat{\sigma}_2^2) - 1}$, where $\hat{\sigma}_1^2 = S_1^2$ and $\hat{\sigma}_2^2 = S_2^2$. This paper is interested in the difference between the coefficients of variation. The estimator of difference between the coefficients of variation of Y_1 and Y_2 is $\hat{\delta} = \hat{\tau}_1 - \hat{\tau}_2$.

The concept of the MOVER approach is modified to construct the confidence interval for difference between the coefficients of variation. It is called the modified MOVER approach. Using the Equation 4, the confidence intervals for the coefficient of variation of Y_1 and Y_2 are defined by $[l_1^*, u_1^*]$ and $[l_2^*, u_2^*]$, respectively, are defined by;

$$[l_1^*, u_1^*] = \left[\sqrt{\exp\left(\frac{(n_1 - 1)S_1^2}{\chi_{n_1-1,1-\alpha/2}^2}\right) - 1}, \sqrt{\exp\left(\frac{(n_1 - 1)S_1^2}{\chi_{n_1-1,\alpha/2}^2}\right) - 1} \right], \tag{25}$$

and

$$[l_2^*, u_2^*] = \left[\sqrt{\exp\left(\frac{(n_2 - 1)S_2^2}{\chi_{n_2-1,1-\alpha/2}^2}\right) - 1}, \sqrt{\exp\left(\frac{(n_2 - 1)S_2^2}{\chi_{n_2-1,\alpha/2}^2}\right) - 1} \right], \tag{26}$$

Where $\chi_{n_i-1,1-\alpha/2}^2$ and $\chi_{n_i-1,\alpha/2}^2$ denote the $100(1 - \alpha/2)$ -th and $100(\alpha/2)$ -th percentiles of the Chi-squared distribution with $n_i - 1$ degrees of freedom for $i = 1, 2$.

Applying the MOVER approach, the lower and upper limits of the confidence interval for the difference between the coefficients of variation based on modified MOVER approach are

$$L_{\delta,MMOVER} = \hat{\tau}_1 - \hat{\tau}_2 - \sqrt{(\hat{\tau}_1 - l_1^*)^2 + (u_2^* - \hat{\tau}_2)^2}, \tag{27}$$

and

$$U_{\delta,MMOVER} = \hat{\tau}_1 - \hat{\tau}_2 + \sqrt{(u_1^* - \hat{\tau}_1)^2 + (\hat{\tau}_2 - l_2^*)^2}, \tag{28}$$

Therefore, the $100(1 - \alpha)\%$ modified MOVER confidence interval for difference between the coefficients of variation δ is defined as;

$$CI_{\delta,MMOVER} = [L_{\delta,MMOVER}, U_{\delta,MMOVER}], \tag{29}$$

Where $L_{\delta,MMOVER}$ and $U_{\delta,MMOVER}$ are defined in Equations 27 and 28, respectively.

3-1-3- Approximate Fiducial Confidence Interval for Difference between the Coefficients of Variation

Suppose that $T_{0.5}$, T_α , and $T_{1-\alpha}$ are the modified normal based approximations given in Krishnamoorthy [26]. For $i = 1, 2$, the approximations for the coefficient of variation of Y_i are defined by;

$$T_{i;0.5} = \exp\left(\frac{(n_i - 1)S_i^2}{\chi_{n_i-1,0.5}^2}\right) - 1, \tag{30}$$

$$T_{i;\alpha/2} = \exp\left(\frac{(n_i - 1)S_i^2}{\chi_{n_i-1,\alpha/2}^2}\right) - 1, \tag{31}$$

and

$$T_{i;1-\alpha/2} = \exp\left(\frac{(n_i - 1)S_i^2}{\chi_{n_i-1,1-\alpha/2}^2}\right) - 1, \tag{32}$$

Where $\chi_{n_i-1,0.5}^2$, $\chi_{n_i-1,\alpha/2}^2$ and $\chi_{n_i-1,1-\alpha/2}^2$ denote the 100(0.5)-th, 100($\alpha/2$)-th and 100($1 - \alpha/2$)-th percentiles of the Chi-squared distribution with $n_i - 1$ degrees of freedom, respectively.

Applying Hasan and Krishnamoorthy [12] and the MOVER approach, the lower and upper limits for difference between the coefficients of variation are obtained by;

$$L_{\delta,AF} = \sqrt{T_{1;0.5} - T_{2;0.5} - \sqrt{(T_{1;0.5} - T_{1;1-\alpha/2})^2 + (T_{2;0.5} - T_{2;\alpha/2})^2}}, \tag{33}$$

and

$$U_{\delta,AF} = \sqrt{T_{1;0.5} - T_{2;0.5} + \sqrt{(T_{1;0.5} - T_{1;\alpha/2})^2 + (T_{2;0.5} - T_{2;1-\alpha/2})^2}}, \tag{34}$$

Therefore, the 100($1 - \alpha$)% approximate fiducial confidence interval for difference between the coefficients of variation δ is defined as;

$$CI_{\delta,AF} = [L_{\delta,AF}, U_{\delta,AF}], \tag{35}$$

Where $L_{\delta,AF}$ and $U_{\delta,AF}$ are defined in Equations 33 and 34, respectively.

3-2- Bayesian Confidence Interval for Difference between the Coefficients of Variation

For $i = 1, 2$, let $X_i = (X_{i1}, X_{i2}, \dots, X_{in_i})$ be the random samples from normal populations with parameter $\theta_i = (\mu_i, \sigma_i^2)$. Also, let $x_i = (x_{i1}, x_{i2}, \dots, x_{in_i})$ be the observed values of X_i , where $i = 1, 2$. The \bar{x}_i and s_i^2 are observed values of \bar{X}_i and S_i^2 , respectively, where \bar{X}_i is sample mean and S_i^2 is sample variance. The likelihood function of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ | data is;

$$L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | data) \propto \left(\frac{1}{\sigma_1^2}\right)^{n_1/2} \exp\left\{-\frac{1}{2\sigma_1^2}[(n_1 - 1)s_1^2 + n_1(\mu_1 - \bar{x}_1)^2]\right\} \times \left(\frac{1}{\sigma_2^2}\right)^{n_2/2} \exp\left\{-\frac{1}{2\sigma_2^2}[(n_2 - 1)s_2^2 + n_2(\mu_2 - \bar{x}_2)^2]\right\}, \tag{36}$$

The logarithm of the likelihood is;

$$\ln(L) = -\frac{n_1}{2} \ln(\sigma_1^2) - \frac{1}{2\sigma_1^2}[(n_1 - 1)s_1^2 + n_1(\mu_1 - \bar{x}_1)^2] - \frac{n_2}{2} \ln(\sigma_2^2) - \frac{1}{2\sigma_2^2}[(n_2 - 1)s_2^2 + n_2(\mu_2 - \bar{x}_2)^2], \tag{37}$$

From the logarithm of the likelihood, the Fischer information matrix is;

$$F(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \begin{bmatrix} \frac{n_1}{\sigma_1^2} & 0 & 0 & 0 \\ 0 & \frac{n_2}{\sigma_2^2} & 0 & 0 \\ 0 & 0 & \frac{n_1}{2} \left(\frac{1}{\sigma_1^2}\right)^2 & 0 \\ 0 & 0 & 0 & \frac{n_2}{2} \left(\frac{1}{\sigma_2^2}\right)^2 \end{bmatrix}, \tag{38}$$

Therefore, the Jeffreys Independence prior is;

$$p(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = p(\mu_1, \mu_2) p(\sigma_1^2, \sigma_2^2), \tag{39}$$

For $i = 1, 2$, the flat prior for coefficient of variation is defined by;

$$p(\theta_i) = \frac{p(\sigma_i)}{p(\mu_i)}, \quad (40)$$

The Jeffreys Independence prior is;

$$p(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \frac{1}{\sigma_1^2 \left(\frac{1}{\sigma_2^2} \right)}, \quad (41)$$

The joint posterior distribution for μ_1 , μ_2 , σ_1^2 , and σ_2^2 is given by

$$p(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | data) \propto \prod_{i=1}^2 \left(\frac{1}{\sigma_i^2} \right)^{(n_i+2)/2} \exp \left\{ -\frac{1}{2\sigma_i^2} [(n_i - 1)s_i^2 + n_i(\mu_i - \bar{x}_i)^2] \right\}, \quad (42)$$

Therefore, the conditional posterior distribution for μ_i given σ_i^2 and x_i is given by;

$$\mu_i | \sigma_i^2, x_i \sim N(\hat{\mu}_i, \frac{\sigma_i^2}{n_i}), \quad (43)$$

Where $i = 1, 2$.

The posterior distribution for σ_i^2 given x_i is obtained by;

$$\sigma_i^2 | x_i \sim IG\left(\frac{n_i - 1}{2}, \frac{(n_i - 1)s_i^2}{2}\right), \quad (44)$$

Where $i = 1, 2$.

The posterior distribution for the difference between the coefficients of variation is denoted by;

$$\delta_{BS} = \sqrt{\exp(\sigma_1^2) - 1} - \sqrt{\exp(\sigma_2^2) - 1}, \quad (45)$$

Where σ_1^2 and σ_2^2 are simulated through Monte Carlo simulation from the posterior distribution defined as in Equation 44.

Therefore, the $100(1 - \alpha)\%$ Bayesian confidence interval for difference between the coefficients of variation δ is defined as;

$$CI_{\delta,BS} = [L_{\delta,BS}, U_{\delta,BS}], \quad (46)$$

Where $L_{\delta,BS}$ and $U_{\delta,BS}$ are the lower and upper limits of the highest posterior density region of δ_{BS} , respectively.

The computational procedure for constructing the Bayesian confidence interval in Equation 46 is presented in Algorithm 3.

Algorithm 3.

Step 1: Generate $\sigma_i^2 | x_i \sim IG((n_i - 1) / 2, (n_i - 1)s_i^2 / 2)$, where $i = 1, 2$;

Step 2: Calculate the value of δ_{BS} as given in Equation 45;

Step 3: Repeat the step 1 - step 2 for q times and obtain $\delta_{BS}^1, \delta_{BS}^2, \dots, \delta_{BS}^q$;

Step 4: Calculate $L_{\delta,BS}$ and $U_{\delta,BS}$ from the $100(\alpha/2)$ -th and $100(1 - \alpha/2)$ -th percentiles of $\delta_{BS}^1, \delta_{BS}^2, \dots, \delta_{BS}^q$.

The following algorithm is given to calculate the coverage probabilities and average lengths of the confidence intervals for difference between the coefficients of variation.

Algorithm 4. For a given $n_1, n_2, \mu_1, \mu_2, \sigma_1, \sigma_2$, and δ :

Step 1: Generate x_1 from $N(\mu_1, \sigma_1^2)$ and generate x_2 from $N(\mu_2, \sigma_2^2)$;

Step 2: Calculate $\bar{x}_1, \bar{x}_2, s_1^2$ and s_2^2 ;

Step 3: Construct $CI_{\delta,MOVER(h)} = [L_{\delta,MOVER(h)}, U_{\delta,MOVER(h)}]$;

Step 4: Construct $CI_{\delta,MMOVER(h)} = [L_{\delta,MMOVER(h)}, U_{\delta,MMOVER(h)}]$;

Step 5: Construct $CI_{\delta,AF(h)} = [L_{\delta,AF(h)}, U_{\delta,AF(h)}]$;

Step 6: Construct $CI_{\delta,BS(h)} = [L_{\delta,BS(h)}, U_{\delta,BS(h)}]$;

Step 7: If $L_{(h)} \leq \delta \leq U_{(h)}$ set $p_{(h)} = 1$, else $p_{(h)} = 0$;

Step 8: Calculate $U_{(h)} - L_{(h)}$;

Step 9: Repeat the step 1 to step 8 for a large number of times (say, M times) and calculate coverage probability and average length.

4- Simulation Studies

A simulation study with 10,000 replications and 2,500 repetitions of the Bayesian calculation was conducted to estimate the performance of the confidence intervals based on the Bayesian method and existing confidence intervals. A comparison of their performances in terms of coverage probability and average length is presented in Figure 1. A coverage probability of greater than or equal to the nominal confidence level of 0.95 and the shortest average length are the criteria for the best-performing confidence interval.

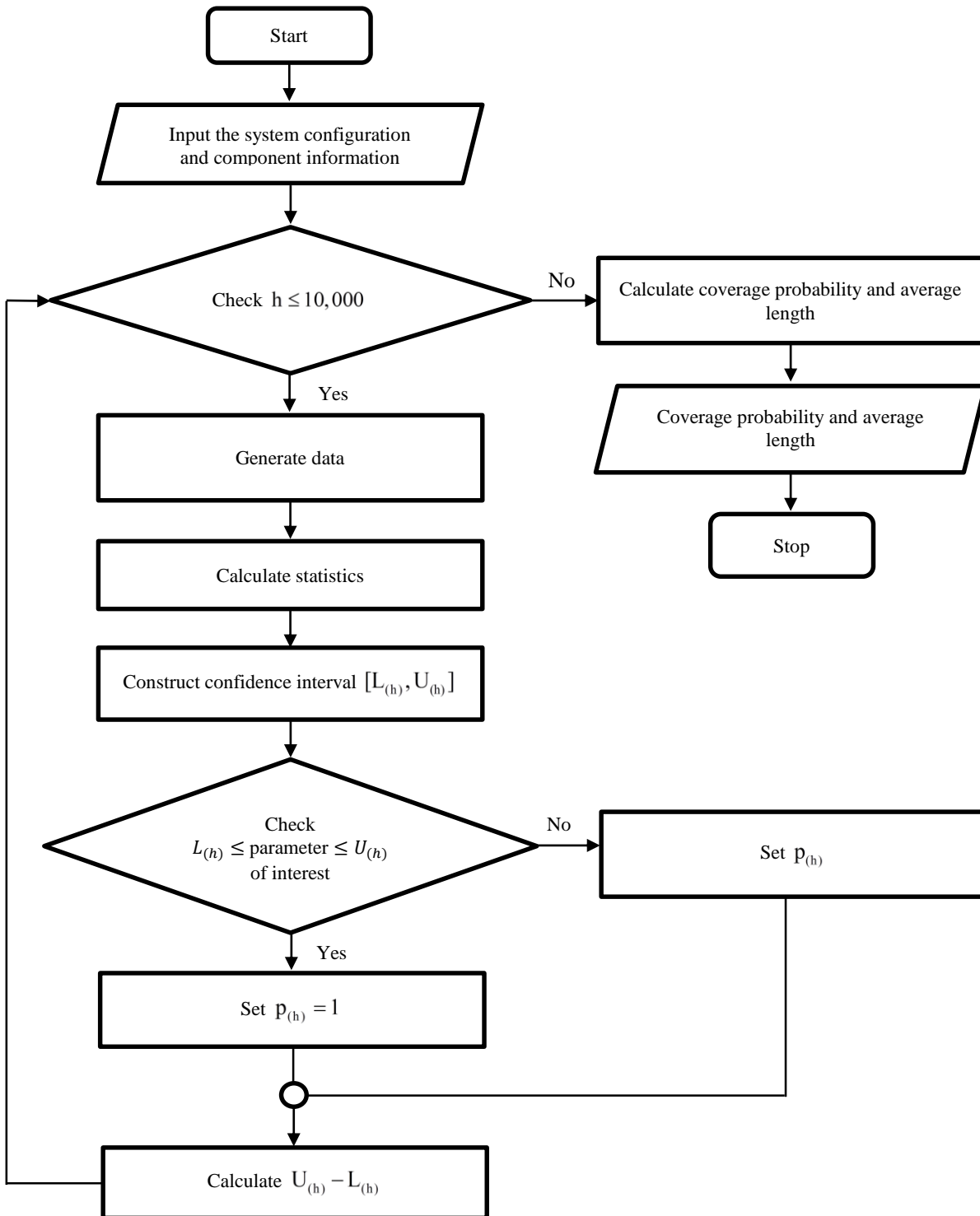


Figure 1. Monte Carlo simulation flowchart.

For the coefficient of variation of a single log-normal distribution, a random sample with sample size n was generated from a normal distribution with mean $\mu = 1$ and standard deviation $\sigma = 0.1, 0.7, 1.2,$ and 1.6 ; the results are given in Table 1. For a large sample size, the coverage probabilities of the confidence intervals of Nam and Kwon [11] and Thangjai et al. [10] were close to 1.00 and 0.95, respectively, for $\sigma \leq 1.2$ and both were less than the nominal confidence level of 0.95 for $\sigma > 1.2$. Although the Bayesian, Chi-squared, and approximate fiducial confidence intervals provided coverage probabilities close to the nominal confidence level of 0.95 for all values of σ , the Bayesian confidence interval obtained the shortest average length in all cases. Therefore, the Bayesian confidence interval was better than the other methods for constructing the confidence intervals for the coefficient of variation of a log-normal distribution.

Table 1. The coverage probabilities (CP) and average lengths (AL) of 95% two-sided confidence intervals for the coefficient of variation of log-normal distribution.

n	Approach	σ							
		0.1		0.7		1.2		1.6	
		CP	AL	CP	AL	CP	AL	CP	AL
4	LS1	1.0000	4.1747	1.0000	4.3016	0.9896	5.0059	0.7080	105.2728
	LS2	0.9818	0.1612	0.9835	1.6097	0.9612	5.4197	0.9388	15.3395
	CHI	0.9519	0.3095	0.9456	30002	0.9480	1.2146e+29	0.9470	6.0419e+50
	AF	0.9519	0.3095	0.9456	30002	0.9480	1.2146e+29	0.9470	6.0419e+50
	BS	0.9482	0.2392	0.9467	602.4821	0.9495	9.6344e+16	0.9476	2.1661e+32
8	LS1	1.0000	4.0455	1.0000	4.1128	0.9832	4.3784	0.7608	5.1539
	LS2	0.9557	0.1056	0.9725	1.0538	0.9582	3.5480	0.9339	10.0420
	CHI	0.9444	0.1346	0.9520	2.5814	0.9560	232.5419	0.9479	1014553
	AF	0.9444	0.1346	0.9520	2.5814	0.9560	232.5419	0.9479	1014553
	BS	0.9433	0.1208	0.9516	1.7921	0.9554	38.6583	0.9511	16648.0000
10	LS1	1.0000	4.0201	1.0000	4.0750	0.9883	4.2778	0.7982	4.7939
	LS2	0.9591	0.0931	0.9680	0.9293	0.9583	3.1291	0.9401	8.8562
	CHI	0.9504	0.1122	0.9478	1.6902	0.9506	26.5996	0.9504	36232.9300
	AF	0.9504	0.1122	0.9478	1.6902	0.9506	26.5996	0.9504	36232.9300
	BS	0.9506	0.1029	0.9504	1.3329	0.9515	11.5776	0.9490	1353.3980
12	LS1	1.0000	4.0032	1.0000	4.0499	0.9894	4.2206	0.8232	4.6353
	LS2	0.9538	0.0842	0.9652	0.8406	0.9549	2.8303	0.9374	8.0108
	CHI	0.9470	0.0979	0.9517	1.3316	0.9523	12.4236	0.9476	424.2455
	AF	0.9470	0.0979	0.9517	1.3316	0.9523	12.4236	0.9476	424.2455
	BS	0.9458	0.0912	0.9539	1.1136	0.9556	7.1811	0.9511	101.4081
15	LS1	1.0000	3.9864	1.0000	4.0238	0.9933	4.1583	0.8506	4.4747
	LS2	0.9544	0.0746	0.9614	0.7451	0.9577	2.5088	0.9373	7.1008
	CHI	0.9505	0.0840	0.9524	1.0386	0.9501	6.7495	0.9462	68.9924
	AF	0.9505	0.0840	0.9524	1.0386	0.9501	6.7495	0.9462	68.9924
	BS	0.9470	0.0793	0.9564	0.9105	0.9507	4.7040	0.9491	31.4261
20	LS1	1.0000	3.9697	1.0000	3.9980	0.9957	4.0972	0.8849	4.3223
	LS2	0.9532	0.0641	0.9571	0.6396	0.9564	2.1536	0.9403	6.0953
	CHI	0.9503	0.0697	0.9488	0.8066	0.9468	4.0851	0.9453	23.5648
	AF	0.9503	0.0697	0.9488	0.8066	0.9468	4.0851	0.9453	23.5648
	BS	0.9475	0.0666	0.9509	0.7343	0.9514	3.2299	0.9469	14.7982
30	LS1	1.0000	3.9531	1.0000	3.9722	0.9996	4.0375	0.9244	4.1806
	LS2	0.9521	0.0519	0.9542	0.5177	0.9529	1.7432	0.9445	4.9337
	CHI	0.9503	0.0548	0.9484	0.6008	0.9496	2.5382	0.9509	10.2423
	AF	0.9503	0.0548	0.9484	0.6008	0.9496	2.5382	0.9509	10.2423
	BS	0.9477	0.0531	0.9515	0.5650	0.9562	2.2044	0.9504	7.8984

The coefficient of variation of log-normal distribution depends on the value of σ^2 only and is independent of μ . To simplify matters, the population means (μ_1, μ_2) were given the same value ($\mu_1 = \mu_2 = 1$) when testing the confidence

intervals for the difference between the coefficients of variation of log-normal distributions. Four confidence intervals constructed using the MOVER, modified MOVER, approximate fiducial, and Bayesian approaches were compared (Tables 2 and 3). The results indicate that the MOVER confidence interval was conservative since the coverage probabilities were in the range from 0.98 to 1.00 for $(\sigma_1, \sigma_2) = (0.1, 0.3), (0.1, 0.7), (0.3, 0.7), (0.3, 0.9),$ and $(0.4, 1.2)$. However, the MOVER confidence interval performed better than the others when the sample sizes were small and the values of (σ_1, σ_2) were large. Meanwhile, the Bayesian confidence interval performed better than the others in terms of the coverage probability and average length.

Table 2. The CP and AL of 95% two-sided confidence intervals for difference between the coefficients of variation of log-normal distributions as function of parameters.

(n_1, n_2)	Approach	(σ_1, σ_2)							
		(0.1, 0.3)		(0.3, 0.7)		(0.4, 1.2)		(0.5, 1.6)	
		CP	AL	CP	AL	CP	AL	CP	AL
(4,8)	MOVER	1.0000	5.8198	1.0000	5.8726	0.9905	6.0767	0.9092	6.7016
	MMOVER	0.9611	0.6596	0.9628	3.9692	0.9597	217.7059	0.9572	12342172
	AF	0.9560	0.6426	0.9581	3.9056	0.9571	217.5643	0.9561	12342171
	BS	0.9780	0.6038	0.9894	3.2132	0.9825	59.5095	0.9807	33278.1700
(5,10)	MOVER	1.0000	5.7634	1.0000	5.8067	0.9931	5.9633	0.9135	6.3768
	MMOVER	0.9604	0.4948	0.9582	2.2795	0.9559	27.9188	0.9540	2448.4490
	AF	0.9565	0.4834	0.9553	2.2374	0.9548	27.8399	0.9529	2448.3010
	BS	0.9727	0.4599	0.9804	2.0027	0.9756	14.3499	0.9688	337.3556
(10,5)	MOVER	1.0000	5.7681	1.0000	5.8386	0.9875	6.2159	0.9071	10.1079
	MMOVER	0.9549	0.9219	0.9531	82.6247	0.9502	1.0649e+13	0.9502	7.1655e+24
	AF	0.9543	0.9160	0.9522	82.6017	0.9501	1.0649e+13	0.9505	7.1655e+24
	BS	0.9563	0.7139	0.9613	9.2918	0.9548	54697391	0.9552	2.3896e+14
(7,12)	MOVER	1.0000	5.7088	1.0000	5.7443	0.9960	5.8731	0.9145	6.2009
	MMOVER	0.9575	0.3888	0.9553	1.5815	0.9541	12.7382	0.9518	332.6372
	AF	0.9556	0.3821	0.9527	1.5568	0.9538	12.7001	0.9513	332.5826
	BS	0.9660	0.3648	0.9711	1.4170	0.9691	7.8683	0.9653	88.4367
(10,10)	MOVER	1.0000	5.6906	1.0000	5.7297	0.9936	5.8837	0.9129	6.3096
	MMOVER	0.9547	0.4124	0.9544	1.8389	0.9510	46.5689	0.9487	9118.4900
	AF	0.9527	0.4081	0.9521	1.8225	0.9517	46.5477	0.9493	9118.4640
	BS	0.9593	0.3793	0.9664	1.5237	0.9588	13.7952	0.9540	613.5661
(10,15)	MOVER	1.0000	5.6652	1.0000	5.6932	0.9975	5.7934	0.9323	6.0299
	MMOVER	0.9547	0.3134	0.9511	1.1802	0.9497	6.8152	0.9547	63.5273
	AF	0.9532	0.3096	0.9498	1.1662	0.9496	6.7967	0.9550	63.5051
	BS	0.9603	0.2966	0.9682	1.0749	0.9621	4.9215	0.9589	29.8344
(20,20)	MOVER	1.0000	5.6168	1.0000	5.6369	0.9987	5.7100	0.9403	5.8787
	MMOVER	0.9519	0.2424	0.9561	0.8641	0.9488	4.1163	0.9499	23.0779
	AF	0.9517	0.2411	0.9568	0.8592	0.9483	4.1110	0.9506	23.0726
	BS	0.9546	0.2316	0.9645	0.7997	0.9533	3.2893	0.9536	14.7232
(20,30)	MOVER	1.0000	5.6041	1.0000	5.6184	1.0000	5.6674	0.9607	5.7748
	MMOVER	0.9527	0.1948	0.9494	0.6594	0.9475	2.5841	0.9521	10.3191
	AF	0.9520	0.1936	0.9492	0.6550	0.9476	2.5793	0.9520	10.3145
	BS	0.9535	0.1884	0.9583	0.6281	0.9570	2.2667	0.9556	7.9961
(30,30)	MOVER	1.0000	5.5923	1.0000	5.6058	0.9999	5.6545	0.9640	5.7600
	MMOVER	0.9502	0.1876	0.9508	0.6339	0.9494	2.5885	0.9496	10.2430
	AF	0.9494	0.1870	0.9502	0.6314	0.9497	2.5860	0.9495	10.2406
	BS	0.9495	0.1816	0.9566	0.6019	0.9555	2.2590	0.9491	7.9242

Table 3. The CP and AL of 95% two-sided confidence intervals for difference between the coefficients of variation of log-normal distributions as function of sample sizes.

(σ_1, σ_2)	Approach	(n_1, n_2)							
		(5,5)		(10,10)		(10,20)		(30,40)	
		CP	AL	CP	AL	CP	AL	CP	AL
(0.1,0.7)	MOVER	1.0000	5.9054	1.0000	5.7244	1.0000	5.6696	1.0000	5.5947
	MMOVER	0.9487	88.3996	0.9489	1.7185	0.9485	0.8220	0.9486	0.5000
	AF	0.9487	88.3867	0.9489	1.7161	0.9484	0.8199	0.9481	0.4997
	BS	0.9590	9.2059	0.9543	1.3623	0.9536	0.7520	0.9488	0.4775
(0.3,0.9)	MOVER	0.9994	5.9936	0.9999	5.7672	1.0000	5.6931	1.0000	5.6058
	MMOVER	0.9563	81962.3600	0.9476	4.2557	0.9524	1.5728	0.9461	0.8719
	AF	0.9548	81962.3000	0.9470	4.2408	0.9525	1.5608	0.9455	0.8700
	BS	0.9716	339.8361	0.9597	2.9684	0.9634	1.4138	0.9516	0.8230
(0.5,1.4)	MOVER	0.9635	6.9065	0.9618	6.0404	0.9865	5.8194	0.9978	5.6657
	MMOVER	0.9525	8.0748e+16	0.9462	226.2264	0.9497	9.2485	0.9478	3.7011
	AF	0.9513	8.0748e+16	0.9457	226.1946	0.9483	9.2237	0.9478	3.6979
	BS	0.9703	200005367179	0.9565	52.8499	0.9616	6.8973	0.9568	3.2364
(0.3,1.5)	MOVER	0.9517	7.6320	0.9418	6.1432	0.9717	5.8507	0.9924	5.6817
	MMOVER	0.9510	9.6776e+19	0.9497	725.8446	0.9527	13.6975	0.9523	5.0227
	AF	0.9514	9.6776e+19	0.9498	725.8370	0.9527	13.6916	0.9520	5.0220
	BS	0.9603	1.9951e+12	0.9534	111.5663	0.9539	9.4326	0.9555	4.2822
(0.9,1.8)	MOVER	0.6441	9241.6420	0.7121	6.8877	0.8209	6.1364	0.8988	5.8026
	MMOVER	0.9566	1.9579e+37	0.9496	149904	0.9537	67.4141	0.9495	14.9772
	AF	0.9541	1.9579e+37	0.9491	149903.8000	0.9529	67.2834	0.9497	14.9633
	BS	0.9820	4.6013e+24	0.9652	6082.4660	0.9718	41.0790	0.9565	11.9328
(0.4,0.5)	MOVER	1.0000	5.8818	1.0000	5.7130	1.0000	5.6690	1.0000	5.5928
	MMOVER	0.9570	5.0863	0.9554	1.1138	0.9549	0.7881	0.9503	0.3973
	AF	0.9495	5.0159	0.9515	1.0913	0.9527	0.7741	0.9493	0.3938
	BS	0.9888	3.5582	0.9752	1.0292	0.9713	0.7424	0.9589	0.3873

5- Empirical Application

In this section, the performances of the existing and Bayesian confidence intervals were compared using real datasets. The Bayesian confidence interval was first computed using 2,500 repetitions via Monte Carlo simulation. The datasets comprising PM10 data from 24 March 2019 to 17 April 2019 reported by the Pollution Control Department for the Nan and Chiang Mai provinces in Thailand are given in Table 4 and statistics based on them are summarized in Table 5. Figures 2 and 3 show histograms and normal QQ-plots of the data, respectively. The Akaike Information Criterion (AIC) results in Table 6 indicate that the datasets can be fitted to log-normal distributions.

For the coefficient of variation of a log-normal distribution, the PM10 data for the Nan province had a coefficient of variation for the log-normal distribution of $\hat{\tau} = 0.2930$. The large-sample confidence intervals were $CI_{\tau,LS1} = [-1.6893, 2.2752]$ and $CI_{\tau,LS2} = [0.2048, 0.3812]$ with interval lengths of 3.9645 and 0.1764, respectively, while those of the Chi-squared, approximate fiducial, and Bayesian confidence intervals were $CI_{\tau,CHI} = [0.2308, 0.4292]$, $CI_{\tau,AF} = [0.2308, 0.4292]$, and $CI_{\tau,BS} = [0.2292, 0.4088]$ with interval lengths of 0.1984, 0.1984, and 0.1796, respectively. For the PM10 data for Chiang Mai province, the coefficient of variation of the log-normal distribution was $\hat{\tau} = 0.2656$. The 95% confidence intervals for the large-sample approaches were $CI_{\tau,LS1} = [-1.7154, 2.2467]$ and $CI_{\tau,LS2} = [0.1879, 0.3433]$ with interval lengths of 3.9621 and 0.1554, respectively, while those using the Chi-squared, approximate fiducial, and Bayesian approaches were $[0.2103, 0.3838]$, $[0.2103, 0.3838]$, and $[0.2034, 0.3641]$ with interval lengths of 0.1735, 0.1735, and 0.1607, respectively.

The difference between the coefficients of variation of log-normal distributions of the PM10 data for the Nan and Chiang Mai provinces was $\hat{\delta} = 0.0274$. The MOVER, modified MOVER, and approximate fiducial, and Bayesian confidence intervals were $CI_{\delta,MOVER} = [-2.7751, 2.8299]$, $CI_{\delta,MMOVER} = [-0.1036, 0.1716]$, $CI_{\delta,AF} = [-0.1021, 0.1698]$, and $CI_{\delta,BS} = [-0.0937, 0.1693]$ with interval lengths of 5.6050, 0.2752, 0.2719, and 0.2630, respectively.

The results show that the confidence intervals in both scenarios covered the population coefficient of variation and the difference between the population coefficients of variation, respectively, with the Bayesian confidence interval having the shortest length in both cases. In concert with the simulation results, the Bayesian confidence interval can be

suggested for constructing the confidence intervals for the coefficient of variation and the difference between the coefficients of variation of log-normal distributions.

Table 4. PM10 data of the Nan and Chiang Mai provinces ($\mu\text{g}/\text{m}^3$).

Nan province					Chiang Mai province				
224	134	138	148	190	227	170	164	105	128
145	232	136	144	127	156	262	167	112	103
114	199	100	155	116	138	146	166	123	94
107	176	90	178	126	125	191	142	139	96
80	130	126	254		113	184	117	138	98

Source: Pollution Control Department (<http://air4thai.pcd.go.th/webV2/download.php>).

Table 5. Sample statistics of the Nan and Chiang Mai provinces.

Statistics	Nan province	Chiang Mai province
n	24	25
\bar{y}	148.7083	144.1600
s_y	44.9662	41.2580
\bar{x}	4.9603	4.9355
s_x	0.2931	0.2665
\hat{t}	0.2930	0.2656

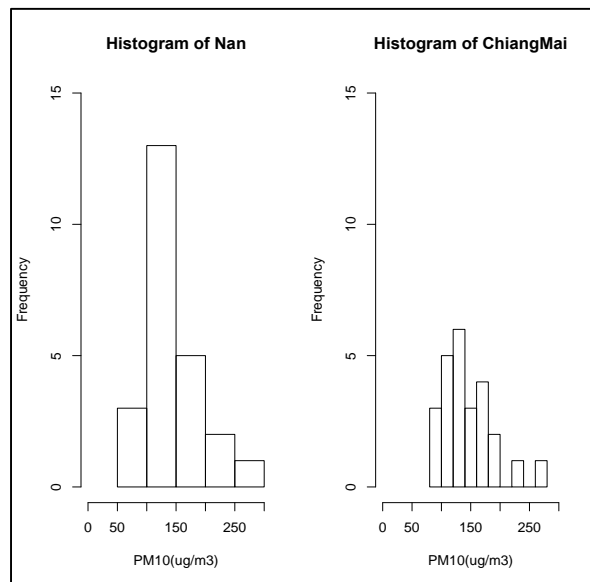


Figure 2. Histogram plots of PM10 data of the Nan and Chiang Mai provinces.

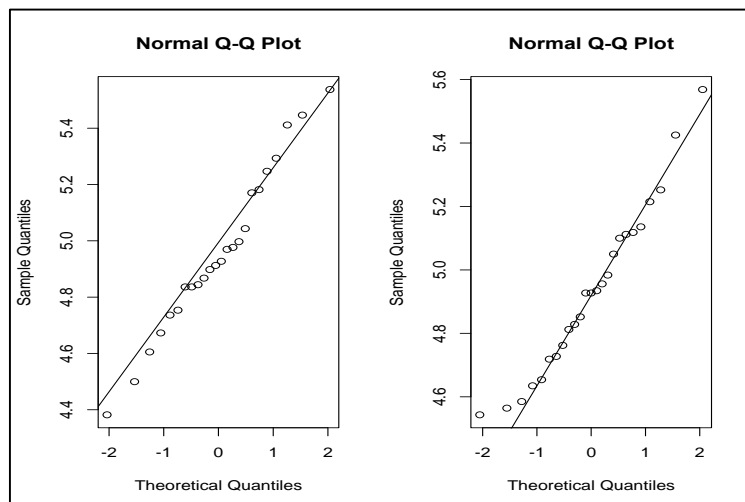


Figure 3. The normal QQ-plots of log-PM10 data of the Nan and Chiang Mai provinces.

Table 6. The minimum Akaike Information Criterion (AIC) values of the Nan and Chiang Mai provinces.

Distribution	Nan province	Chiang Mai province
Normal	253.7713	259.9186
Log-Normal	250.2824	254.5765
Gamma	250.9095	255.8663
Exponential	289.0954	299.5462

6- Discussion

Classical and Bayesian inference are fundamentally different in statistics, and we evaluated the performances of the confidence intervals for the coefficient of variation and the difference of the coefficients of variation of log-normal distributions using both approaches. For the coefficient of variation of a log-normal distribution, four classical confidence intervals constructed via two large-sample approaches based on the two variances definitions of Thangjai et al. [10] and Nam and Kwon [11], the Chi-squared approach of Niwitpong [21], and the approximate fiducial approach of Hasan and Krishnamoorthy [12]. The classical confidence intervals were derived use formulas whereas the Bayesian approach was based on a simulation technique.

For the difference between the coefficients of variation of log-normal distributions, the three classical confidence intervals: MOVER by using the variance of Nam and Kwon [11], modified MOVER by using the Chi-squared approach of Niwitpong [21], and the approximate fiducial approach with modified normal-based approximations were compared with the Bayesian confidence interval. The results in this investigation were similar to those of Harvey and van der Merwe [7], Rao and D'Cunha [9], Thangjai and Niwitpong [14], and Thangjai et al. [22].

7- Conclusion

In this study, the Bayesian approach was used to construct confidence intervals for the coefficient of variation of a log-normal distribution and the difference between the coefficients of variation of two log-normal distributions, both of which were then compared with several classical approaches. For the first scenario, although the coverage probabilities of all of the confidence intervals were close to the nominal confidence level, the Bayesian confidence interval provided the shortest average length in all cases. For the difference between the coefficients of variation of two log-normal distributions, the Bayesian confidence interval was once again the best in terms of the coverage probability and average length.

The performances of all of the approaches were appraised by application to real PM10 data from the Nan and Chiang Mai provinces in Thailand. As with the results of the simulation study, the Bayesian approach was better than the others in terms of average length. Therefore, the Bayesian approach is recommended for constructing the confidence intervals for the coefficient of variation and the difference between the coefficients of variation of log-normal distributions.

8- Declarations

8-1- Author Contributions

Conceptualization, W.T., S.N. and S.N.; methodology, W.T., S.N. and S.N.; formal analysis, W.T., S.N. and S.N.; writing—original draft preparation, W.T., S.N. and S.N.; writing—review and editing, W.T., S.N. and S.N. All authors have read and agreed to the published version of the manuscript.

8-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

8-3- Funding

This research was funded by Faculty of Applied Science, King Mongkut's University of Technology North Bangkok. Grant No. 641079.

8-4- Conflicts of Interest

The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

9- References

- [1] Royston, P. "The Lognormal Distribution as a Model for Survival Time in Cancer, With an Emphasis on Prognostic Factors." *Statistica Neerlandica* 55, no. 1 (March 2001): 89–104. doi:10.1111/1467-9574.00158.
- [2] Lacey, L. F., O. N. Keene, J. F. Pritchard, and A. Bye. "Common Noncompartmental Pharmacokinetic Variables: Are They Normally or Log-Normally Distributed?" *Journal of Biopharmaceutical Statistics* 7, no. 1 (January 1, 1997): 171–178. doi:10.1080/10543409708835177.
- [3] Sharma, Mohita Anand, and Jai Bhagwan Singh. "Use of probability distribution in rainfall analysis." *New York Science Journal* 3, no. 9 (2010): 40-49.
- [4] Krishnamoorthy, K, and Thomas Mathew. "Inferences on the Means of Lognormal Distributions Using Generalized p-Values and Generalized Confidence Intervals." *Journal of Statistical Planning and Inference* 115, no. 1 (July 2003): 103–121. doi:10.1016/s0378-3758(02)00153-2.
- [5] Hannig, Jan, E. Lidong, Amany Abdel-Karim, and Hari Iyer. "Simultaneous fiducial generalized confidence intervals for ratios of means of lognormal distributions." *Austrian Journal of Statistics* 35, no. 2&3 (2006): 261-269.
- [6] Tian, Lili, and Jianrong Wu. "Inferences on the Common Mean of Several Log-Normal Populations: The Generalized Variable Approach." *Biometrical Journal* 49, no. 6 (June 2007): 944–951. doi:10.1002/bimj.200710391.
- [7] Harvey, J., and A.J. van der Merwe. "Bayesian Confidence Intervals for Means and Variances of Lognormal and Bivariate Lognormal Distributions." *Journal of Statistical Planning and Inference* 142, no. 6 (June 2012): 1294–1309. doi:10.1016/j.jspi.2011.12.006.
- [8] Lin, S. H., and R. S. Wang. "Modified Method on the Means for Several Log-Normal Distributions." *Journal of Applied Statistics* 40, no. 1 (November 5, 2012): 194–208. doi:10.1080/02664763.2012.740622.
- [9] Rao, K. Aruna, and Juliet Gratia D’Cunha. "Bayesian Inference for Median of the Lognormal Distribution." *Journal of Modern Applied Statistical Methods* 15, no. 2 (November 1, 2016): 526–535. doi:10.22237/jmasm/1478003400.
- [10] Thangjai, Warisa, Sa-Aat Niwitpong, and Suparat Niwitpong. "Simultaneous Fiducial Generalized Confidence Intervals for All Differences of Coefficients of Variation of Log-Normal Distributions." *Integrated Uncertainty in Knowledge Modelling and Decision Making* (2016): 552–561. doi:10.1007/978-3-319-49046-5_47.
- [11] Nam, Jun-mo, and Deukwoo Kwon. "Inference on the Ratio of Two Coefficients of Variation of Two Lognormal Distributions." *Communications in Statistics - Theory and Methods* 46, no. 17 (May 17, 2017): 8575–8587. doi:10.1080/03610926.2016.1185118.
- [12] Hasan, Md Sazib, and K. Krishnamoorthy. "Improved Confidence Intervals for the Ratio of Coefficients of Variation of Two Lognormal Distributions." *Journal of Statistical Theory and Applications* 16, no. 3 (2017): 345. doi:10.2991/jsta.2017.16.3.6.
- [13] Niwitpong, Sa-aat, Suparat Niwitpong, and Warisa Thangjai. "Simultaneous Confidence Intervals for All Differences of Coefficients of Variation of Log-Normal Distributions." *Hacettepe Journal of Mathematics and Statistics* (December 31, 2019): 1–17. doi:10.15672/hujms.454804.
- [14] Thangjai, Warisa, and Suparat Niwitpong. "Comparing Particulate Matter Dispersion in Thailand Using the Bayesian Confidence Intervals for Ratio of Coefficients of Variation." *Statistics in Transition New Series* 21, no. 5 (2020): 41–60. doi:10.21307/stattrans-2020-054.
- [15] Thangjai, Warisa, Sa-Aat Niwitpong, and Suparat Niwitpong. "Confidence Intervals for the Common Coefficient of Variation of Rainfall in Thailand." *PeerJ* 8 (September 21, 2020): e10004. doi:10.7717/peerj.10004.
- [16] Thangjai, Warisa, and Sa-Aat Niwitpong. "Confidence Intervals for the Signal-to-Noise Ratio and Difference of Signal-to-Noise Ratios of Log-Normal Distributions." *Stats* 2, no. 1 (February 27, 2019): 164–173. doi:10.3390/stats2010012.
- [17] Thangjai, Warisa, and Sa-Aat Niwitpong. "Confidence Intervals for Common Signal-to-Noise Ratio of Several Log-Normal Distributions." *Iranian Journal of Science and Technology, Transactions A: Science* 44, no. 1 (November 19, 2019): 99–107. doi:10.1007/s40995-019-00793-3.
- [18] Tsim, Y.L., Yip, S.P., Tsang, K.S., Li, K.F. and Wong, H.F. "Haematogoy and Serology." *Annual Report, Honk Kong Medical Technology Association Quality Assurance Programme*, (1991): 25-40.
- [19] Faupel-Badger, Jessica M., Barbara J. Fuhrman, Xia Xu, Roni T. Falk, Larry K. Keefer, Timothy D. Veenstra, Robert N. Hoover, and Regina G. Ziegler. "Comparison of Liquid Chromatography-Tandem Mass Spectrometry, RIA, and ELISA Methods for Measurement of Urinary Estrogens." *Cancer Epidemiology Biomarkers & Prevention* 19, no. 1 (January 2010): 292–300. doi:10.1158/1055-9965.epi-09-0643.
- [20] Lipfert, F.W. "Air Pollution and Community Health: A Critical Review and Data Sourcebook." *Van Nostrand Reinhold*, New York, USA (1994).

- [21] Niwitpong, Sa-aat. "Confidence Intervals for Coefficient of Variation of Lognormal Distribution with Restricted Parameter Space." *Applied Mathematical Sciences* 7 (2013): 3805–3810. doi:10.12988/ams.2013.35251.
- [22] Thangjai, Warisa, Sa-Aat Niwitpong, and Suparat Niwitpong. "A Bayesian Approach for Estimation of Coefficients of Variation of Normal Distributions." *Sains Malaysiana* 50, no. 1 (January 31, 2021): 261–278. doi:10.17576/jsm-2021-5001-25.
- [23] Ng, C. K. "Inference on the common coefficient of variation when populations are lognormal: a simulation-based approach." *Journal of Statistics: Advances in Theory and Applications* 11, no. 2 (2014): 117-134.
- [24] Casella, G. and Berger, R.L. "Statistical Inference." Duxbury, California, USA (2002).
- [25] Donner, Allan, and GY Zou. "Closed-Form Confidence Intervals for Functions of the Normal Mean and Standard Deviation." *Statistical Methods in Medical Research* 21, no. 4 (September 8, 2010): 347–359. doi:10.1177/0962280210383082.
- [26] Krishnamoorthy, K. "Modified Normal-Based Approximation to the Percentiles of Linear Combination of Independent Random Variables with Applications." *Communications in Statistics - Simulation and Computation* 45, no. 7 (June 23, 2014): 2428–2444. doi:10.1080/03610918.2014.904342.