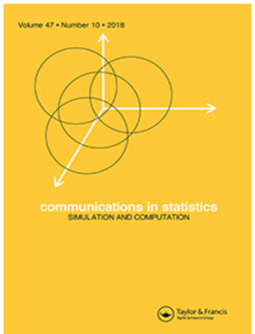

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Confidence intervals for the ratio of medians of two independent log-normal distributions

Lapasrada Singhasomboon , Wararit Panichkitkosolkul & Andrei Volodin

To cite this article: Lapasrada Singhasomboon , Wararit Panichkitkosolkul & Andrei Volodin (2020): Confidence intervals for the ratio of medians of two independent log-normal distributions, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2020.1812649

To link to this article: https://doi.org/10.1080/03610918.2020.1812649

Published online: 28 Aug 2020.

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2020.1812649

Confidence intervals for the ratio of medians of two independent log-normal distributions

Lapasrada Singhasomboon[a], Wararit Panichkitkosolkul[a], and Andrei Volodin[b]

[a]Department of Mathematics and Statistics, Thammasat University, Phathum Thani, Thailand; [b]Department of Mathematics and Statistics, University of Regina, Saskatchewan, Canada

ABSTRACT
We focus on the construction of confidence intervals for the ratios of medians of two independent, log-normal distributions based on the normal approximation (NA) approach, the method of variance estimate recovery (MOVER), and the generalized confidence interval (GCI) approach. We also compare the performance of the three confidence intervals in terms of the coverage probabilities, and average lengths, using Monte Carlo simulations. The results show that the GCI confidence interval is generally preferred in terms of coverage probabilities; however, the average length for the GCI is always wider than for other approaches. The NA and MOVER approaches could be recommended on the basis of the specific values of μ_i, σ_i^2 and/or sample sizes. The confidence intervals are illustrated using real data examples.

ARTICLE HISTORY
Received 9 April 2020
Accepted 16 August 2020

KEYWORDS
Central tendency; Generalized confidence interval; Interval estimation; Normal approximation; simulation; Skew distribution; Variance estimates recovery

1. Introduction

In many applications, such as medicine, biology, exposure, pollution, economics, finance, reliability, survival and meteorology data analysis, measurements are often right-skewed. In these data, analyses are usually assumed by log-normal models. The log-normal distribution is common in many application areas. Estimating the parameters of the log-normal distribution is an interesting problem. There are many studies about approaches for constructing confidence intervals for log-normal distributions; for example, the interval for a single log-normal mean has been addressed multiple times in the literature (e.g., see Land 1972; Angus 1988, 1994; Zhou and Gao 1997, etc.) The interval for the ratio, or difference of two log-normal means is addressed in Zhou and Tu (2000); Wu et al. (2002); Krishnamoorthy and Mathew (2003), etc. The interval estimation for the mean of several log-normal distributions is discussed in Baklizi and Ebrahem (2005); Behboodian and Jafari (2006); Tian and Wu (2007); Lin and Wang (2013); Malekzadeh and Kharrati-Kopaei (2018). However, log-normal distributions that follow right-skewed data typically have extremely low measurements, which affect the median less than the mean. Thus, in this situation, the median is a more meaningful central tendency measure than the mean.

Some authors have considered the median of the log-normal distribution. Zellner (1971) proposed a Bayesian and non-Bayesian estimator for the parameters of the mean and median of the log-normal distribution. Rao and D'Cunha (2016) proposed the Bayes credible interval for the median of the log-normal distribution, and compared the interval based on the MLE. The conclusion was that the Bayes credible interval has a shorter average length compared to the one interval. To our knowledge, there is no research paper on the confidence interval for medians of two

CONTACT Lapasrada Singhasomboon lapasrada.shsb@hotmail.com Department of Mathematics and Statistics, Thammasat University, Phathum Thani, Thailand.

log-normal distributions. In this article, we take this new challenge by investigating the new aspect of confidence intervals construction. We propose the Normal Approximation (NA), the method of variance estimates recovery (MOVER) and the generalized confidence interval (GCI) approaches to construct confidence intervals for the ratio of medians for two independent, log-normal distributions. We assess these three confidence intervals using the coverage probabilities and their expected lengths. Typically, we prefer a confidence interval with a coverage probability of at least the nominal level $(1-\alpha)$; its expected length is short.

The description of notation and the log-normal model, followed by three confidence interval approaches for constructing confidence intervals for the ratio medians of two independent, log-normal distributions are discussed in Sec. 2. A simulation study comparing the proposed interval is presented in Sec. 3, with Sec. 3.1 containing the discussion and the results. In Sec. 4, the proposed CI construction approaches are illustrated with PM2.5 datasets. Finally, the concluding remarks are given in Sec. 5.

2. Methods

Let$X_{ij}; i=1,2, j=1,\ldots,n_i$. be random variables from two independent, log-normal distributions with the following parameters: the meansμ_i and variancesσ_i^2, respectively, are denoted as $X_{ij} \sim LN(\mu_i, \sigma_i^2)$.

The probability density function is

$$f(X) = \frac{1}{x_{ij}\sqrt{2\pi\sigma_i{}^2}} e^{-\frac{1}{2}\left(\frac{\ln x_{ij}-\mu_i}{\sigma_i}\right)^2}$$

where $x_{ij} > 0, -\infty < \mu_i < \infty, \sigma_i{}^2 > 0$ and $i=1,2, j=1,\ldots,n_i$.

We know that the logarithm of X_{ij}. i.e., $Y_{ij} = \ln(X_{ij}) \sim N(\mu_i, \sigma_i^2)$ is normally distributed. Therefore, the unbiased estimators (and MLE) for $\mu_i, \sigma_i^2, i=1,2$ are

$$\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

$$S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_i\right)^2, i=1,2$$

It is well-known that the medians of two independent, log-normal distribution can be calculated as follows:

$$m_1 = e^{\mu_1}$$

and

$$m_2 = e^{\mu_2}$$

which may be applied to obtain the ratio medians of two independent log-normal(ψ) Inferences are made on

$$\psi = \frac{m_1}{m_2} = e^{(\mu_1-\mu_2)} \tag{1}$$

By using the plug-in estimator, the unbiased (and MLE) point estimator of ψ is

$$\hat{\psi} = e^{(\bar{y}_1-\bar{y}_2)} \tag{2}$$

where $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, i=1,2$.

In the rest of this section, we address constructing the confidence interval for ψ by the Normal approximation, Generalized Confidence Interval and MOVER approaches.

2.1. The normal approximation confidence interval approach

Consider the following point estimator of ψ as in Eq. (2):

$$\hat{\psi} = e^{(\bar{y}_1 - \bar{y}_2)}.$$

In the Normal Approximation approach, the main statistical tool we use to obtain an asymptotically normal and limiting distribution of an estimator $\hat{\psi}$ is the famous Delta method. It can be explained briefly in the following way.

Let $g(v_1, v_2)$ be a differentiable scalar function of two variables. Consider an estimator $T = g(V_1, V_2)$, which is a function of two other basic statistics V_1 and V_2. Usually, statistics V_1 and V_2 have a simple form, and it is known that they are jointly asymptotically normal. The asymptotic distribution of an estimator, T, can be found by the Delta method, which is a procedure of stochastic representation of T.

Now, we apply the Delta method to prove its asymptotically normality as $n_i \to \infty$ and to find the asymptotic mean and variance for the estimator ψ.

In the Delta method, function g is used to expand into the Taylor series at the point $\mu_1 = E(V_1)$ and $\mu_2 = E(V_2)$:

$$g(V_1, V_2) = g(\mu_1, \mu_2) + \frac{\partial g(\mu_1, \mu_2)}{\partial v_1}(V_1 - \mu_1) + \frac{\partial g(\mu_1, \mu_2)}{\partial v_2}(V_2 - \mu_2) + \text{Remainder}.$$

Note that it is possible to prove that $\sqrt{n}\text{R}emainder \to 0$ in probability as the sample size $n_1, n_2 \to \infty$.

For our case, $g(v_1, v_2) = e^{v_1 - v_2}$ and $V_1 = \bar{Y}_1, V_2 = \bar{Y}_2$. The statistic $V_1 = \bar{Y}_1$ is normally distributed as $N\left(\bar{Y}_1, {S_1^2}/{n_1}\right)$ and the statistic $V_2 = \bar{Y}_2$ is normally distributed as $N\left(\bar{Y}_2, {S_2^2}/{n_2}\right)$.

To calculate partical derivatives,

$$\frac{\partial g(v_1, v_2)}{\partial v_1} = e^{v_1 - v_2} \text{ and } \frac{\partial g(v_1, v_2)}{\partial v_2} = -e^{v_1 - v_2}.$$

Hence,

$$\frac{\partial g(\mu_1, \mu_2)}{\partial v_1} = e^{\mu_1 - \mu_2} \text{ and } \frac{\partial g(\mu_1, \mu_2)}{\partial v_2} = -e^{\mu_1 - \mu_2}.$$

By the Taylor series,

$$\begin{aligned} g(V_1, V_2) &\approx e^{\mu_1 - \mu_2} + e^{\mu_1 - \mu_2}(\hat{\mu}_1 - \mu_1) - e^{\mu_1 - \mu_2}(\hat{\mu}_2 - \mu_2) \\ &\approx e^{\mu_1 - \mu_2}(\hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2 + 1) \end{aligned}.$$

From this, we take the expectation and variance on the both sides, we obtained the asymptotic mean and variance for the estimator ψ_{NA} are given by

$$\mu_{\hat{\psi}_{NA}} = e^{\mu_1 - \mu_2} \text{ and } \sigma^2_{\hat{\psi}_{NA}} = e^{2(\mu_1 - \mu_2)}\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right),$$

respectively.

To sum up, as sample sizes $n_1, n_2 \to \infty$, we have the $\hat{\psi}_{NA}$ is approximately normal with a mean $\mu_{\hat{\psi}_{NA}}$ and variance of the form $\sigma^2_{\hat{\psi}_{NA}}$. Obviously, values μ_1, μ_2, σ_1^2 and σ_2^2 are unknown when we estimate the parameter function ψ_{NA} having only samples in our hands. In this case, we use the plug-in estimators of $\hat{\mu}_{\hat{\psi}_{NA}}$ and $\hat{\sigma}^2_{\hat{\psi}_{NA}}$ as follows:

$$\hat{\mu}_{\hat{\psi}_{NA}} = e^{\bar{y}_1 - \bar{y}_2} \text{ and } \hat{\sigma}^2_{\hat{\psi}_{NA}} = e^{2(\bar{y}_1 - \bar{y}_2)}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$$

where $\bar{y}_i$ and s_i^2 are the observed values of $\bar{Y}_i$ and S_i^2, respectively.

Therefore, the $(1-\alpha)100\%$ two-sided approximate confidence interval for the ψ based on NA appraoch is given by

$$CI_{NA} = \left(\hat{\psi}_{NA} - z_{\alpha/2}\sqrt{e^{2(\bar{y}_1-\bar{y}_2)}\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)}, \hat{\psi}_{NA} + z_{\alpha/2}\sqrt{e^{2(\bar{y}_1-\bar{y}_2)}\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)} \right) \quad (3)$$

where $z_{\alpha/2}$ is the $(\alpha/2)$-th quantile value from the standard normal distribution.

2.2. The confidence interval based on the MOVER approach

Zou (2008) and Zou and Donner (2008) introduced the method of variance estimates recovery (MOVER) for constructing a confidence interval for a linear combination of the parameters estimated by confidence limits for each component of the parameters.

Recall in Eq. (1) that the ratio median of two independent log-normal is

$$\psi = \frac{m_1}{m_2} = e^{(\mu_1-\mu_2)}.$$

The logarithm of the ratio, which we will denote as θ, is

$$\theta = \ln(\psi) = \mu_1 - \mu_2. \quad (4)$$

We start by constructing the confidence interval for $\theta = \ln(\psi) = \mu_1 - \mu_2$. Then, we take the exponent to obtain a confidence interval for the ratio of medians: $\psi = e^{(\mu_1-\mu_2)}$.

Let $\mu_i, i = 1, 2$ as each component parameters of the θ, and let L_i and U_i be the lower and upper limits of the interval for $\mu_i, i = 1, 2$, respectively.

Then, we have

$$L_i = \bar{y}_i - z_{\alpha/2}\sqrt{\mathrm{var}(\bar{y}_i)}, U_i = \bar{y}_i + z_{\alpha/2}\sqrt{\mathrm{var}(\bar{y}_i)} \quad (5)$$

where $\bar{Y}_i \sim N\left(\bar{Y}_i, {S_i^2}/{n_i}\right)$ and $z_{\alpha/2}$ is the $(\alpha/2)$-th quantile value from the standard normal distribution.

Following Zou (2008) and Zou and Donner (2008), the $(1-\alpha)100\%$ two-sided confidence interval for the parameters θ, based on the MOVER approach, is given by

$$CI = \left[\bar{y}_1 - \bar{y}_2 - \sqrt{(\bar{y}_1 - L_1)^2 + (U_2 - \bar{y}_2)^2}, \bar{y}_1 - \bar{y}_2 + \sqrt{(U_1 - \bar{y}_1)^2 + (\bar{y}_2 - L_2)^2}\right] \quad (6)$$

Therefore, we will take the exponent in Eq. (6) to obtain a confidence interval for the ψ as follows:

$$CI_{MOVER} = \exp\left[\bar{y}_1 - \bar{y}_2 - \sqrt{(\bar{y}_1 - L_1)^2 + (U_2 - \bar{y}_2)^2}, \bar{y}_1 - \bar{y}_2 + \sqrt{(U_1 - \bar{y}_1)^2 + (\bar{y}_2 - L_2)^2}\right] \quad (7)$$

where L_i, U_i are defined by Eq. (5).

2.3. The generalized confidence interval approach

Weerahandi (1993) introduced the generalized inference approach for testing a confidence interval in the situation where the parameter of interest consists of several component parameters, such as nuisance parameter(s). The basic concept of the GCI is as follows. Let $X = (X_1, X_2, \ldots, X_n)$ be a random sample from the probability density function $f(x; \theta, \lambda)$, where θ is the parameter of interest and λ is a nuisance parameter(s). Let $x = (x_1, x_2, \ldots, x_n)$ denote the observed value of $X = (X_1, X_2, \ldots, X_n)$. To obtain the confidence limits for the parameter of interest, θ, we first need to construct a generalized pivotal quantity $R(X; x, \theta, \lambda)$, which is a function of the random sample X. The observed data, x, and the unknown parameters, θ, λ, must satisfy the following conditions:

(i) The distribution of $R(X; x, \theta, \lambda)$ is free of unknown parameters;
(ii) The observed value of $R(X; x, \theta, \lambda)$ is equal to the parameter of interest(θ)

Similar to the MOVER approach, we start by constructing the confidence interval for$\theta = \ln(\psi) = \mu_1 - \mu_2$. Finally, we take the exponent to obtain a confidence interval for the ratio of medians(ψ).

The generalized pivotal quantity of $\theta = \mu_1 - \mu_2$ is defined as follows:

$$R_\theta = R_{\mu_1} - R_{\mu_2}, \tag{8}$$

where R_{μ_i} according to Krishnamoorthy and Mathew (2003), is given by:

$$\begin{aligned} R_{\mu_i} &= \bar{y}_i - \frac{\bar{Y}_i - \mu_i}{S_i/\sqrt{n_i}} \frac{s_i}{\sqrt{n_i}} \\ &= \bar{y}_i - \frac{Z_i}{\sqrt{n_i}} \frac{s_i\sqrt{n_i - 1}}{U_i} i = 1, 2 \end{aligned}$$

where $\bar{y}_i$ and s_i^2 are the observed values of $\bar{Y}_i$ and S_i^2, respectively. Z_i and U_i are independent, where $Z_i = \frac{(\bar{Y}_i - \mu_i)}{\sigma_i/\sqrt{n_i}} \sim N(0, 1)$ and $U_i^2 = \frac{(n_i - 1)S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}, i = 1, 2.$

It is easy to verify that R_θ satisfies the above two conditions. The last expression suggests that the distribution of R_θ is free of unknown parameters. In the first expression, substituting $\bar{y}_i$ and s_i^2 for $\bar{Y}_i$ and S_i^2 is equal to θ. So, the $(1 - \alpha)100\%$ two-sided generalized confidence interval for θ is simply the $R_\theta(\alpha/2), R_\theta(1 - \alpha/2)$ of percentile of R_θ. Next, we will take the exponent to obtain a generalized confidence interval for the ψ.

Therefore, the $(1 - \alpha)100\%$ two-sided generalized confidence interval for ψ, based on the GCI approach, is given by

$$CI_{GCI} = \exp[R_\theta(\alpha/2), R_\theta(1 - \alpha/2)]. \tag{9}$$

Constructing the GCI for the ψ can be summarized by the following algorithm:

Algorithm:
for a given $\bar{y}_1, \bar{y}_2, s_1^2, s_2^2$
for i = 1 to m:
Generate the value for Z_1, Z_2, U_1^2, U_2^2 from the standard normal distribution and the chi-squared distribution with n-1 degree of freedom, respectively.
Calculate R_θ as $R_{\mu_1} - R_{\mu_2}$
End loop for i
The $(1 - \alpha)100\%$ two-sided generalized confidence interval for ψ is then obtained by taking the exponent of the $100(\alpha/2)$ percentile of R_θ defined by $R_\theta(\alpha/2)$ and $100(1 - \alpha/2)$ percentile of R_θ defined by $R_\theta(1 - \alpha/2)$.

3. Simulation study

In the simulation studies, we evaluate the performance of the proposed CI construction approaches. We estimated the coverage probabilities and average length through Monte Carlo simulation with the R statistical software. For the parameter configurations, we have generated 10,000 random samples from two independent, log-normal populations, with the parameters μ_i and $\sigma_i^2, i = 1, 2$. Numerical results on the coverage probabilities and average length of the 95%

Table 1. Coverage probabilities and Average length of 95% CIs for ψ when $(\sigma_1^2, \sigma_2^2) = (0.20, 0.20)$.

		Coverage probabilities			Average length		
Parameters	($n1$,$n2$)	NA	MOVER	GCI	NA	MOVER	GCI
$(\mu_1 = 1, \mu_2 = 1)$	(25)	0.9439	0.9433	0.9552	0.4308	0.4342	0.4551
	(25,40)	0.9435	0.9447	0.9532	0.3866	0.3890	0.4048
	(25,50)	0.9457	0.9477	0.956	0.3729	0.3751	0.3899
	(25,100)	0.9411	0.9449	0.9538	0.3373	0.3389	0.3530
	(25,40)	0.9442	0.9472	0.9552	0.3861	0.3885	0.4043
	(40)	0.9468	0.9487	0.9539	0.3391	0.3408	0.3504
	(40, 50)	0.9483	0.944	0.9495	0.3231	0.3246	0.3326
	(40,100)	0.9468	0.9479	0.9538	0.2836	0.2846	0.2911
	(25, 50)	0.9429	0.9442	0.9516	0.3727	0.3749	0.3897
	(40, 50)	0.9465	0.9441	0.9505	0.3229	0.3243	0.3326
	(50)	0.9452	0.9475	0.9524	0.3038	0.3050	0.3116
	(50,100)	0.9437	0.9453	0.9484	0.2627	0.2634	0.2681
	(100,25)	0.9465	0.9467	0.9544	0.3382	0.3399	0.3539
	(100,40)	0.9454	0.947	0.953	0.2830	0.2840	0.2905
	(100,50)	0.9475	0.9466	0.9507	0.2628	0.2636	0.2683
	(100,100)	0.948	0.9498	0.9509	0.2148	0.2152	0.2173
$(\mu_1 = 3, \mu_2 = 3)$	(25)	0.9414	0.9433	0.9532	0.4294	0.4328	0.4537
	(25,40)	0.9429	0.9433	0.9506	0.3879	0.3904	0.4062
	(25,50)	0.9433	0.9446	0.9519	0.3717	0.3739	0.3888
	(25,100)	0.9428	0.9414	0.9511	0.3396	0.3412	0.3554
	(25,40)	0.9452	0.9451	0.9549	0.3869	0.3893	0.4050
	(40)	0.944	0.9456	0.9513	0.3405	0.3422	0.3517
	(40,50)	0.9453	0.9461	0.9525	0.3223	0.3237	0.3318
	(40,100)	0.9484	0.95	0.9546	0.2838	0.2847	0.2914
	(25,50)	0.9419	0.9436	0.9529	0.3721	0.3742	0.3891
	(40,50)	0.9455	0.947	0.9528	0.3215	0.3229	0.3310
	(50)	0.9479	0.9488	0.9526	0.3038	0.3050	0.3115
	(50,100)	0.9456	0.9468	0.9499	0.2632	0.2640	0.2687
	(100,25)	0.9391	0.9415	0.9501	0.3379	0.3395	0.3535
	(100,40)	0.9474	0.9479	0.9513	0.2835	0.2845	0.2910
	(100,50)	0.9487	0.9494	0.9517	0.2632	0.2639	0.2686
	(100,100)	0.9473	0.9473	0.9488	0.2149	0.2154	0.2174

two-sided confidence interval for ψ of two independent log-normal distributions when equal $(\sigma_1^2, \sigma_2^2) = (0.20, 0.20), (\sigma_1^2, \sigma_2^2) = (0.30, 0.30)$ and unequal $(\sigma_1^2, \sigma_2^2) = (0.20, 0.30)$ are reported in Tables 1, 2, and 3, respectively, for various values $\mu_i = 1, 3; i = 1, 2,$ and sample sizes (n_1, n_2) varying from small to large under equal and unequal sample sizes. In practical applications, the values of (n_1, n_2) are usually unequal, so we consider unequal sample sizes in our simulation studies and also include the PM2.5 datasets, which can be seen in Sec. 4.

3.1. Discussion

Based on Table 1, we can conclude that:

(i) The coverage probability of the GCI approach is always greater than the nominal level, regardless of n_i, μ_i, σ_i^2 values. However, the average length of this approach is wider than other approaches.

(ii) For $\sigma_1^2 = \sigma_2^2$, and for small sample sizes, the GCI approach is recommended. For moderate sample sizes, the average length of all proposed NA, MOVER, GCI approaches perform well, except $(\mu_1 = 1, \mu_2 = 1, \sigma_1^2 = 0.20, \sigma_2^2 = 0.20)$; the NA performs better than the GCI and MOVER approaches. For large sample sizes, the NA approach is recommended.

(iii) For $\sigma_1^2 \neq \sigma_2^2$, for moderate to large sample sizes, the NA approach is quite satisfactory. However, the GCI approach is recommended for all sample sizes.

Table 2. Coverage probabilities and Average length of 95% CIs for ψ when $(\sigma_1^2, \sigma_2^2) = (0.30, 0.30)$.

		Coverage probabilities			Average length		
Parameters	(n1,n2)	NA	MOVER	GCI	NA	MOVER	GCI
$(\mu_1 = 1, \mu_2 = 1)$	(25)	0.9373	0.9433	0.9532	0.6104	0.6200	0.6505
	(25,40)	0.9415	0.9433	0.9506	0.5514	0.5585	0.5814
	(25,50)	0.9415	0.9446	0.9519	0.5280	0.5342	0.5558
	(25,100)	0.942	0.9415	0.9511	0.4825	0.4872	0.5077
	(25,40)	0.945	0.9451	0.9549	0.5497	0.5567	0.5795
	(40)	0.9435	0.9457	0.9513	0.4838	0.4885	0.5022
	(40,50)	0.9444	0.9461	0.9525	0.4572	0.4613	0.4729
	(40,100)	0.9478	0.95	0.9546	0.4024	0.4051	0.4146
	(25,50)	0.941	0.9436	0.9529	0.5285	0.5348	0.5562
	(40,50)	0.9452	0.947	0.9528	0.4559	0.4599	0.4716
	(50)	0.9466	0.9489	0.9526	0.4309	0.4343	0.4436
	(50,100)	0.944	0.9469	0.9499	0.3731	0.3753	0.3820
	(100,25)	0.938	0.9416	0.9501	0.4793	0.4841	0.5042
	(100,40)	0.948	0.9479	0.9513	0.4019	0.4047	0.4139
	(100,50)	0.9469	0.9494	0.9517	0.3731	0.3752	0.3819
	(100,100)	0.9465	0.9473	0.9488	0.3045	0.3057	0.3086
$(\mu_1 = 3, \mu_2 = 3)$	(25)	0.9381	0.9422	0.9512	0.6111	0.6208	0.6510
	(25,40)	0.9386	0.9422	0.9505	0.5493	0.5563	0.5790
	(25,50)	0.9445	0.9462	0.954	0.5287	0.5350	0.5566
	(25,100)	0.9424	0.9433	0.9504	0.4813	0.4861	0.5064
	(25,40)	0.945	0.9488	0.9575	0.5491	0.5561	0.5788
	(40)	0.9393	0.9435	0.9516	0.4821	0.4868	0.5006
	(40,50)	0.9438	0.9441	0.9499	0.4577	0.4617	0.4734
	(40,100)	0.949	0.9494	0.9541	0.4028	0.4056	0.4153
	(25,50)	0.9448	0.946	0.9539	0.5277	0.5339	0.5555
	(40,50)	0.9449	0.9477	0.9543	0.4571	0.4611	0.4725
	(50)	0.9457	0.9485	0.9534	0.4312	0.4345	0.4440
	(50,100)	0.9465	0.9507	0.9538	0.3725	0.3746	0.3813
	(100,25)	0.9413	0.9436	0.952	0.4812	0.4860	0.5064
	(100,40)	0.9471	0.948	0.9526	0.4027	0.4055	0.4149
	(100,50)	0.9464	0.9472	0.9523	0.3733	0.3755	0.3822
	(100,100)	0.9516	0.9515	0.9514	0.3042	0.3054	0.3082

4. An applications to real data

Air pollution, especially particulate matter (PM) from vehicles, has a major impact on the health of people who live in urban centers and near traffic. The World Bank studies the health effects of particulate matter air pollution in Bangkok. In one of their studies, they found that there were 4,000–5,500 premature deaths each year in urban centers, with hospital admissions for respiratory diseases related to the levels of particulate matter (Pollution Control Department, Ministry of Science Technology and Environment, Bangkok, Thailand 1999). The top causes of air pollution in Thailand are (1) vehicular emissions in cities, (2) biomass burning and transboundary haze in rural and border areas, and (3) industrial discharges in concentrated industrialized zones. Therefore, in this example, we study the PM2.5 mass concentrations in Bangkapi and Dindaeng areas, which are located in urban centers with busy roads representing a high traffic site. We compare the ratio of medians of the measurements in both areas with our proposed CI. The PM2.5 mass concentration measurements (ug/m^3) were recorded simultaneously in the areas by the Pollution Control Department (PCD) every fourth day at 9:00 AM local time from March 2019 to February 2020. The data can be found in http://aqmthai.com/aqi.php.

The data sets from this study are as follows

The PM2.5 mass concentration (ug/m^3) in Bangkapi: 25,22,16,32,31,20,28,17,18,14,13,19,20,16, 10,17,34,23,24,21,18,7,11,19,9,6,5,14,7,18,14,12,9,8,16,14,11,7,9,7,11,12,11,8,8,15,15,7,18,15,36,48,47, 27,14,23,18,24,24,21,13,36,39,24,30,29,17,21,32,38,32,28,25,31,18,34,49,24,32,56,16,16,41,21,39,12,20, 50,56, 31

Table 3. Coverage probabilities and Average length of 95% CIs for ψ when $(\sigma_1^2, \sigma_2^2) = (0.30, 0.20)$.

		Coverage probabilities			Average length		
Parameters	($n1$,$n2$)	NA	MOVER	GCI	NA	MOVER	GCI
$(\mu_1 = 1, \mu_2 = 1)$	(25)	0.9408	0.9451	0.9544	0.5278	0.5340	0.5602
	(25,40)	0.9418	0.9437	0.9537	0.4928	0.4979	0.5201
	(25,50)	0.9379	0.9386	0.9486	0.4798	0.4845	0.5060
	(25,100)	0.9357	0.9384	0.9478	0.4553	0.4594	0.4805
	(25,40)	0.9433	0.9443	0.9529	0.4572	0.4612	0.4784
	(40)	0.9472	0.9473	0.9527	0.4165	0.4195	0.4314
	(40,50)	0.9487	0.9485	0.9546	0.4026	0.4054	0.4162
	(40,100)	0.9386	0.943	0.9479	0.3717	0.3739	0.3837
	(25,50)	0.9422	0.9445	0.9515	0.4300	0.4333	0.4485
	(40,50)	0.9479	0.9475	0.9534	0.3889	0.3913	0.4009
	(50)	0.9446	0.9462	0.9495	0.3721	0.3743	0.3825
	(50,100)	0.9421	0.9443	0.9482	0.3389	0.3406	0.3474
	(100,25)	0.9452	0.9462	0.9547	0.3716	0.3738	0.3870
	(100,40)	0.9517	0.9511	0.9549	0.3221	0.3235	0.3300
	(100,50)	0.944	0.9438	0.9477	0.3048	0.3060	0.3108
	(100,100)	0.9475	0.9461	0.9492	0.2633	0.2640	0.2666
$(\mu_1 = 3, \mu_2 = 3)$	(25)	0.9418	0.9454	0.9548	0.5279	0.5341	0.5602
	(25,40)	0.9417	0.9411	0.9495	0.4932	0.4983	0.5205
	(25,50)	0.9419	0.9433	0.9517	0.4809	0.4857	0.5073
	(25,100)	0.9373	0.938	0.9492	0.4545	0.4586	0.4798
	(25,40)	0.9476	0.9474	0.9546	0.4571	0.4611	0.4781
	(40)	0.9458	0.9457	0.9515	0.4171	0.4201	0.4321
	(40,50)	0.9455	0.9456	0.951	0.4020	0.4047	0.4156
	(40,100)	0.9443	0.9453	0.9507	0.3721	0.3743	0.3841
	(25,50)	0.9392	0.942	0.95	0.4303	0.4337	0.4487
	(40,50)	0.9496	0.9505	0.9548	0.3887	0.3912	0.4008
	(50)	0.9471	0.9463	0.9506	0.3726	0.3748	0.3829
	(50,100)	0.9456	0.9455	0.9505	0.3406	0.3423	0.3491
	(100,25)	0.9419	0.9413	0.9495	0.3715	0.3736	0.3870
	(100,40)	0.9456	0.9453	0.9483	0.3225	0.3239	0.3302
	(100,50)	0.9463	0.9462	0.9503	0.3039	0.3051	0.3099
	(100,100)	0.9495	0.952	0.9534	0.2633	0.2641	0.2666

The PM2.5 mass concentration (ug/m^3) in Dindaeng: 33,24,18,38,32,24,28,20,21,21,14,16,20, 20,21,32,23,23,19,19,9,10,20,11,10,9,15,10,17,13,15,14,12,17,15,14,11,11,10,15,14,11,10,14,15,14,11, 18,19,35,48,54,29,17,25,19,23,24,20,14,37,38,24,30,27,22,23,27,39,54,44,35,26,41,27,39,51,31,40,60,24, 23,53,26,44,15,20,50,53,29

Figure 1 shows the QQ-plots for the original data and log-transformed data. These plots show that the distribution of PM2.5 concentrations are positively skewed, and the logarithmically transformed data are approximately symmetric. The Shapiro-Wilk tests for the normality on the log-transformed data give a p-value of 0.3718 for the Bangkapi area and 0.0610 for the Dindaeng area, while the same tests on the original data give a p-value of 4.55e-05 for the Bangkapi area and 4.118e-06 for the Dindaeng area. So, the log- transformation normalizes the data, and the summary statistics of the log-transformed of the PM2.5 concentration data are given in Table 4.

Table 5 gives the 95% two-sided confidence intervals for the ψ based on the proposed CI construction approaches. This result shows that the NA confidence interval has a shorter length size than the MOVER and GCI, respectively. The results are consistent with the results of our simulation study (see Table 3). As previously mentioned, the NA approach is consistent for this example; we obtain a 95% CI for the ratio of medians that is equal to (0.7267, 0.9915). The results can be interpreted as the median (average) of the PM2.5 mass concentration in the Bangkapi area is less than the Dindaeng area for the period from March 2019 to February 2020. It also means that the chances of death of people living in the Bangkapi area affected by PM2.5 might be less than in the Dindaeng area.

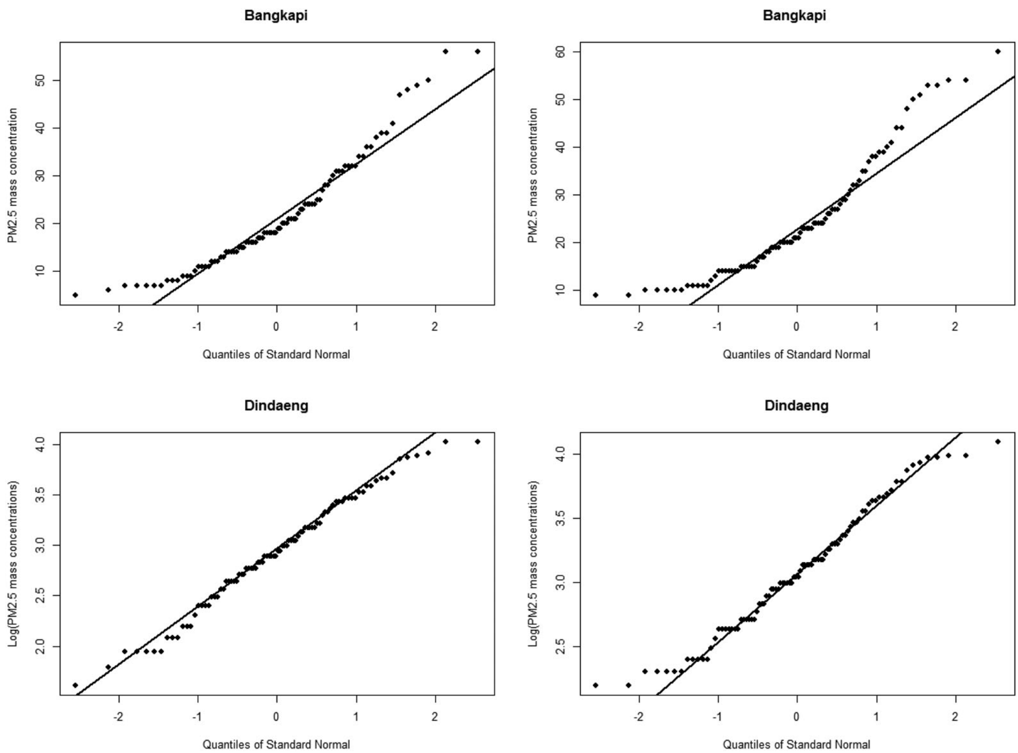

Figure 1. Quantile plots of PM2.5 mass concentration data and log of PM2.5 mass concentration data of both areas.

Table 4. The summary statistics of the log-transformed of the PM2.5 mass concentration data of both areas.

Areas	n_i	μ_i	σ_i^2
Bangkapi	90	2.9286	0.3145
Dindaeng	90	3.0805	0.2417

Table 5. The 95% confidence intervals for the ψ based on the proposed CI construction approaches.

Methods	Confidence interval	Length
NA	(0.7267, 0.9915)	0.2647
MOVER	(0.7364, 1.0022)	0.2658
GCI	(0.7355, 1.0034)	0.2679

5. Concluding remarks

In this paper, we focus on the construction of confidence intervals for the ratio of medians of two independent, log-normal distributions based on the NA, MOVER and GCI approaches. The performances of the proposed CIs are compared in terms of the coverage probabilities (CP) and average length (AL) in our simulation study sections. The results show that the GCI confidence interval can be preferred generally in terms of CP; however, the AL is widest for all situations. The NA approach is recommended for moderate to large sample sizes when μ_i and σ_i^2 are small values. In addition, the MOVER approach may not be reliable when σ_i^2 has unequal values for all sample sizes.

Acknowledgments

We would also like to thank the editor and referees for their valuable suggestions on the revision of this paper.

Funding

The research of the last author is partially supported by Center of the Volga Region Federal District (project number 075-02-2020-1478). The first author gratefully acknowledges the financial support provided by Science Achievement Scholarship of Thailand.

References

Angus, J. E. 1988. Inferences on the lognormal mean for complete samples. Communications *in Statistics - Simulation and Computation* 17 (4):1307–31. doi:10.1080/03610918808812726.

Angus, J. E. 1994. Bootstrap one-sided confidence intervals for the lognormal mean. The *Statistician* 43 (3): 395–401. doi:10.2307/2348575.

Baklizi, A., and M. Ebrahem. 2005. Interval estimation of common lognormal mean of several populations. *Journal of Probability and Statistical Science* 3 (1):1–16.

Behboodian, J., and A. Jafari. 2006. Generalized inference for the common mean of several lognormal populations. *Journal of Statistical Theory and Applications* 5 (3):240–59.

Krishnamoorthy, K., and T. Mathew. 2003. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference* 115 (1):103–20. doi:10.1016/S0378-3758(02)00153-2.

Land, C. E. 1972. An evaluation of approximate confidence interval estimation methods for lognormal means. *Technometrics* 14 (1):145–58. doi:10.1080/00401706.1972.10488891.

Lin, S. H., and R. S. Wang. 2013. Modified method on the means for several log-normal distributions. *Journal of Applied Statistics* 40 (1):194–208. doi:10.1080/02664763.2012.740622.

Malekzadeh, A., and M. Kharrati-Kopaei. 2018. Inferences on the common mean of several heterogeneous log-normal distributions. *Journal of Applied Statistics* 33:1–18.

Pollution Control Department, Ministry of Science Technology and Environment, Bangkok, Thailand. 1999. Report on situation and management of air pollution and Noise in Thailand 1998. Bangkok, Thailand: Ministry of Science Technology and Environment.

Rao, K. A., and J. G. D'Cunha. 2016. Bayesian inference for median of the lognormal distribution. *Journal of Modern Applied Statistical Methods* 15 (2):526–35. doi:10.22237/jmasm/1478003400.

Tian, L., and J. Wu. 2007. Inferences on the common mean of several log-normal populations: The generalized variable approach. *Biometrical Journal. Biometrische Zeitschrift* 49 (6):944–51. doi:10.1002/bimj.200710391.

Weerahandi, S. 1993. Generalized confidence intervals. Journal of the American Statistical Association 88 (423): 899–905. doi:10.1080/01621459.1993.10476355.

Wu, J., G. Jiang, A. C. M. Wong, and X. Sun. 2002. Likelihood analysis for the ratio of means of two independent log-normal distributions. *Biometrics* 58 (2):463–69. doi:10.1111/j.0006-341x.2002.00463.x.

Zellner, A. 1971. Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression. *Journal of the American Statistical Association* 66 (334):327–30. doi:10.1080/01621459.1971.10482263.

Zhou, X. H., and S. Gao. 1997. Confidence intervals for the lognormal mean. *Statistics in Medicine* 16 (7):783–90. doi:10.1002/(SICI)1097-0258(19970415)16:7<783::AID-SIM488>3.0.CO;2-2.

Zhou, X.-H., and W. Tu. 2000. Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics & Data Analysis* 35 (2):201–10. doi:10.1016/S0167-9473(00)00009-8.

Zou, G. Y. 2008. On the estimation of additive interaction by use of the four-by-two table and beyond. *American Journal of Epidemiology* 168 (2):212–24. doi:10.1093/aje/kwn104.

Zou, G. Y., and A. Donner. 2008. Construction of confidence limits about effect measures: A general approach. *Statistics in Medicine* 27 (10):1693–702. doi:10.1002/sim.3095.