Confidence Intervals for the Ratio of Means of Two Independent Log-Normal Distributions

Lapasrada Singhasomboon, Wararit Panichkitkosolkul, Andrei Volodin

Department of Mathematics and Statistics, Thammasat University, Pathum Thani, Thailand
Department of Mathematics and Statistics, University of Regina, Saskatchewan, Canada

Abstract— In this paper, we investigate confidence intervals for the ratio of means of two independent log-normal distributions. The normal approximation (NA) approach was proposed. We compared the proposed with another approaches, the ML, GCI, and MOVER. The performance of these approaches were evaluated in terms of coverage probabilities and interval widths. The Simulation studies and results showed that the GCI and MOVER approaches performed similar in terms of the coverage probability and interval width for all sample sizes. The ML and NA approaches provided the coverage probability close to nominal level for large sample sizes. However, our proposed method provided the interval width shorter than other methods. Overall, our proposed is conceptually simple method. We recommend that our proposed approach is appropriate for large sample sizes because it is consistently performs well in terms of the coverage probability and the interval width is typically shorter than the other approaches. Finally, the proposed approaches are illustrated using a real-life example.

KeyWords: Confidence intervals, Log-normal, Normal approximation, Simulation

Received: February 8, 2021. Revised: March 5, 2021. Accepted: March 8, 2021. Published: March 17, 2021.

1. Introduction

The log-normal distribution is important in describing positively skewed data. Therefore, the log-normal distribution is used as a model in various real life applications, for example in medicine where latency periods (the time between the infection and the first symptoms) are log-normal as in [1], in environmental engineering where the probability distribution of contaminant concentrations are often modeled by the log-normal as in [2], in Atmospheric Science, many atmospheric physical and chemical properties are modeled by the log-normal as in [3], and in economics where it can be used to model markets, for example, incomes asin [4] and closing prices on stocks as in [5].

Let the random variable $X$ follow a log-normal distribution with parameter $\mu$ and $\sigma^2$. Then the random variable $Y = \ln(X)$ follows the normal distribution $N(\mu, \sigma^2)$. The probability density function (pdf) of $X$ is

$$f_X(x) = \frac{1}{x\sqrt{2\pi}\sigma^2} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}; x > 0, \mu \in \mathbb{R}, \sigma > 0.$$

The mean and variance of the log-normal distribution are

$$\tau = \exp\left(\mu + \frac{\sigma^2}{2}\right) \text{ and } \tau_2 = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2),$$

respectively.

The inference of mean of the log-normal distribution are frequently interested. Many studies have examined to estimate the confidence interval of the mean as in [6]-[8].

When the interested data follow a skewed distribution, a log-transformation can be used to normalize the distribution of the original. The two well-known methods, the t-test and the Wilcoxon test have been used to study for comparing the means of two independent log-normal distributions in [9]-[12]. However, the Wilcoxon test and the t-test had type I error rates that were very different from the nominal levels when the two population variances were not equal. Then, Reference [13] proposed two new methods: one is a traditional maximum likelihood test which is based on the parametric procedure and the other is the bootstrap test which is based on the nonparametric procedure to overcome this problem. Their purpose is deriving the new methods to testing the difference of the means of two independent log-normal distributions and compare with the t-test and the Wilcoxon test in terms of type I error and power of the test. Their simulation study showed that the traditional maximum likelihood test was the best in terms of the type I error rate and power of the test when the
This paper is organized as follows. Section II, we first review the existing approaches for the problem of constructing confidence intervals for ratio of two log-normal means, and then we proposed a Normal approximation (NA) approach for this problem. In Section III, simulation studies and results are conducted to study the performance of the NA, and the results are compared with those of other approaches in terms of coverage probabilities and interval widths, a real-life example is analyzed for illustration purposes is presented in section IV. In section V, the discussion is presented. Finally, section VI is provided conclusion.

2. Preliminaries

Suppose that \( X_{i1}, X_{i2}, \ldots, X_{in_i} \) be two independent random samples from log-normal population with parameters \( \mu_i \) and \( \sigma_i^2, i = 1,2 \)

The mean of the i-th log-normal distribution is

\[
M_i = \exp \left( \mu_i + \frac{\sigma_i^2}{2} \right)
\]

The problem of interest is to constructing confidence intervals for ratio of means of the two log-normal populations; that is

\[
\theta = \frac{\mu_1}{\mu_2} = \exp(\eta_1 - \eta_2),
\]

where \( \eta_i = \mu_i + \frac{\sigma_i^2}{2}, i = 1,2 \)

The random variables \( Y_{ij} = \ln(X_{ij}), i = 1,2, j = 1, ..., n_i \) follow the normal distribution \( N(\mu_i, \sigma_i^2) \). Based on this fact, the following unbiased maximum likelihood estimators (MLEs) for \( \mu_i \) and \( \sigma_i^2 \) respectively, are well known

\[
\hat{\mu} = \bar{Y} = \frac{1}{n_t} \sum_{j=1}^{n_t} Y_{ij} \quad \text{and} \quad \hat{\sigma}_i^2 = S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2
\]

In this Section, we first review the three of existing approaches for the problem confidence interval estimation of the ratio of means of two log-normal populations. Then, we present the new approach to this problem using the Normal approximation (NA).

2.1. Traditional maximum likelihood (ML)

Zhou, Gao, and Hui [13] proposed a traditional maximum likelihood (ML) approach for testing the difference log-normal means. Thus, this method can be extended to construct the confidence interval for the ratio of two log-normal means for this study.

Firstly, they derived the point estimator for testing the difference log-normal means. The maximum likelihood method was used. Then, the maximum likelihood estimators for \( \eta_1 - \eta_2 \) is given by

\[
\hat{\eta}_1 - \hat{\eta}_2 = \left( \bar{Y}_1 + \frac{S_1^2}{2} \right) - \left( \bar{Y}_2 + \frac{S_2^2}{2} \right).
\]
Note that $\overline{Y}_i$ and $S_i^2$ are independent, $\overline{Y}_i$ is distributed as $N\left(\mu_i, \frac{S_i^2}{n_i}\right)$ and $S_i^2$ is distributed as $\chi^2$ with $n_i - 1$ degrees of freedom.

Secondly, they obtained the variance of the difference log-normal means

$$\text{Var}(\eta_1 - \eta_2) = \text{Var}\left[(\overline{Y}_1 + \frac{S_1^2}{2}) - (\overline{Y}_2 + \frac{S_2^2}{2})\right] = \frac{S_1^4}{n_1} + \frac{S_2^4}{n_2} + \frac{S_1^2 S_2^2}{2(n_1 - 1)} + \frac{S_1^2 S_2^2}{2(n_2 - 1)}$$

After estimating the variance for $\text{Var}(\eta_1 - \eta_2)$ by substituting estimates $S_1^2$ and $S_2^2$ for $\sigma_1^2$ and $\sigma_2^2$, respectively. They applied the central limit theorem and asymptotic property of the maximum likelihood estimator to build the confidence interval for the difference means, $\eta_1 - \eta_2$.

Then, the confidence interval (CI) for the difference means, $\eta_1 - \eta_2$ takes the form

$$CI = \left(\overline{Y}_1 + \frac{S_1^2}{2} - \overline{Y}_2 - \frac{S_2^2}{2}\right) \pm z_{\alpha/2} \sqrt{\frac{S_1^4}{n_1} + \frac{S_2^4}{n_2} + \frac{S_1^2 S_2^2}{2(n_1 - 1)} + \frac{S_1^2 S_2^2}{2(n_2 - 1)}}$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile value of the standard normal distribution.

Finally, taking the exponentiation into the lower and upper limits of the CI for $\eta_1 - \eta_2$.

Therefore, the $100(1 - \alpha)%$ two-sided CI for ratio of two log-normal means $\overline{Y}_1/\overline{Y}_2$ based on the ML method is given by

$$c_{\alpha/2} = \exp\left(\overline{Y}_1 + \frac{S_1^2}{2} - \overline{Y}_2 - \frac{S_2^2}{2}\right) \pm z_{\alpha/2} \sqrt{\frac{S_1^4}{n_1} + \frac{S_2^4}{n_2} + \frac{S_1^2 S_2^2}{2(n_1 - 1)} + \frac{S_1^2 S_2^2}{2(n_2 - 1)}}$$

### 2.2. Generalized confidence interval (GCI)

The basic idea of the Generalized confidence interval (GCI) was originally introduced by Weerahandi [15]. Let $X$ be a random sample whose distribution involve the $\theta$, the parameter of interest, and $\lambda$, a nuisance parameter, and let $x$ denote the observed value of $X$.

To find confidence interval for $\theta$, it is first need to find 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 $\theta, \lambda$ and should be satisfy the following conditions:

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

Krishnamoorthy and Mathew [16] defined a GCI for $\eta_1 - \eta_2$ as

$$R_{\eta_1 - \eta_2} = R_{\eta_1} - R_{\eta_2}$$

This result is based on the observation that

$$R_{\eta_i} = \overline{Y}_i - \mu_i = \frac{S_i}{\sqrt{n_i / n_i}} R_i + \frac{1}{2} \frac{S_i^2}{\sqrt{n_i}}$$

$$= \overline{Y}_i - \frac{z_{\alpha/2}}{\sqrt{n_i}} \frac{S_i}{\sqrt{n_i}} + \frac{1}{2} \frac{S_i^2}{\sqrt{n_i}}$$

for $i = 1, 2$.

where $Z_i$ and $U_i$ are independent and $Z_i \sim N(0,1), U_i = \frac{(n_i - 1)S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}$.

We note first of all that the second expression suggests that the distribution of $R_{\eta_i}$ is free of unknown parameters.

Secondly, the first expression is equal to $\eta_i$ if we substitute $\overline{Y}_i$ and $S_i^2$ for $\eta_1$ and $\eta_2$.

Therefore, the generalized confidence interval for $\theta$ may be obtained using the following algorithm:

**The algorithm:**

1. For a given data set, compute the sample means and sample variances $\overline{Y}_1, \overline{Y}_2, S_1^2$ and $S_2^2$ using the log-transform data.
2. For $i = 1$ to $m$
   3. Generate value for $Z_i, Z_{-i} U_i, U_{-i}$ from the standard normal distribution and the chi-squared distribution with $n - 1$ degree of freedom, respectively.
   4. compute $R_{\theta_1 - \theta_2} = R_{\theta_1} - R_{\theta_2}$ as in (1).
   5. End loop for $i$
6. Order the $m$ values for $R_{\theta_1 - \theta_2}$ and compute the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of $R_{\theta_1 - \theta_2}$ denoted by $R_{\theta_1 - \theta_2}(\alpha/2)$ and $R_{\theta_1 - \theta_2}(1 - \alpha/2)$, respectively.
7. The $(1 - \alpha)100$% two-sided generalized confidence interval for $\theta$ is obtained by take the exponentiated for $R_{\theta_1 - \theta_2}(\alpha/2)$ and $R_{\theta_1 - \theta_2}(1 - \alpha/2)$.

### 2.3 Method of Variance Estimates Recovery (MOVER)

Zou and Donner [17] offered a Method of Variance Estimates Recovery (MOVER) for a confidence interval construction. This method provides a closed-form CI and easy to compute. The idea of deriving the closed-form interval is based on only estimates confidence limits for a linear combination of parameters from confidence limits for the individual or single parameters based on the recovery of variance estimates.

To describe the MOVER concept, let $\hat{\theta}_i, i = 1, 2$ be point estimates and assume that $\theta_1$ and $\theta_2$ are independently distributed. In general, the Wald’s confidence interval for $\theta_1 + \theta_2$ is given by

$$(L, U) = \left(\hat{\theta}_1 + \hat{\theta}_2 - z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2)}, \hat{\theta}_1 + \hat{\theta}_2 + z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2)}\right)$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile value from the standard normal distribution.

The Wald’s CI does not perform well in small sample sizes. Its performance can be improved by obtaining $\text{Var}(\hat{\theta}_i)$ for individual parameter $\hat{\theta}_i, i = 1, 2$ at the neighborhood of the confidence limits $L$ and $U$ respectively.

Let $(l_1, u_1)$ and $(l_2, u_2)$ be confidence intervals for $\theta_1$ and $\theta_2$ respectively, where

$$l_i = \hat{\theta}_i - z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta}_i)}, u_i = \hat{\theta}_i + z_{\alpha/2} \sqrt{\text{Var}(\hat{\theta}_i)}.$$
which yield the estimated variances is \[\text{var}(\hat{\theta}_i) = \frac{(\hat{\theta}_1 - l_1)^2}{z_{a/2}^2}, \text{var}(\hat{\theta}_j) = \frac{(\hat{\theta}_2 - l_2)^2}{z_{a/2}^2}.\]

The confidence limits \(L\) have plausible values \(l_1 + l_2\) as the minimum value and \(U\) or \(u_1 + u_2\) as the maximum value for \(\hat{\theta}_1 + \hat{\theta}_2\), respectively. This implies that constructing of confidence interval for \(\theta_1 + \theta_2\) they substitute corresponding variance estimators, \(\text{var}(\hat{\theta}_i)\) in the confidence limits \(L\) and \(U\), respectively.

Therefore, the two-sided confidence interval \((L, U)\) for \(\theta_1 + \theta_2\) given by

\[L = \hat{\theta}_1 + \hat{\theta}_2 - z_{a/2} \sqrt{\text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2)} = \hat{\theta}_1 + \hat{\theta}_2 - \left((\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2\right)^{1/2},\]

and

\[U = \hat{\theta}_1 + \hat{\theta}_2 + z_{a/2} \sqrt{\text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2)} = \hat{\theta}_1 + \hat{\theta}_2 + \left((\hat{\theta}_1 - l_1)^2 + (\hat{\theta}_2 - l_2)^2\right)^{1/2},\]

Considering the construction of confidence interval for \(\theta_1 - \theta_2\), we can apply this concept by writing \(\theta_1 - \theta_2\) as \(\theta_1 + (-\theta_2)\) and recognizing that the CI for \(-\theta_2\) is \((-u_2, -l_2)\). Then, the two-sided confidence interval \((L, U)\) for \(\theta_1 - \theta_2\) is given by

\[(L, U) = (\hat{\theta}_1 + \hat{\theta}_2 - z_{a/2} \sqrt{\text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2)}) = \hat{\theta}_1 - \hat{\theta}_2 \pm (u_1 - l_1)^2 + (u_2 - l_2)^2,\]

where \(\hat{\theta}_1, \hat{\theta}_2\) are the observed values of \(\theta_1, \theta_2\), respectively.

Similarly, for the difference two log-normal means \(\eta_1 - \eta_2\), the estimator is

\[\hat{\eta}_1 - \hat{\eta}_2 = (\hat{Y}_1 + \hat{Y}_2) - (\hat{Y}_1 + \hat{Y}_2).\]

They let \(\eta_1 = \mu_1 + \frac{\sigma_1^2}{2}\) and \(\eta_2 = \mu_2 + \frac{\sigma_2^2}{2}\) and consider in construction of confidence interval for \(\theta_1 - \theta_2\).

Then, the \((1 - \alpha)100\%\) two-sided confidence interval for \(\eta_1 - \eta_2\) based on the MOVER is as follows

\[\text{CI} = [\mu_1 - \eta_2 - \sqrt{(\hat{\eta}_1 - \hat{\eta}_2)^2 + (\hat{\eta}_1 - \hat{\eta}_2)^2}, \mu_1 - \eta_2 - \sqrt{(\hat{\eta}_1 - \hat{\eta}_2)^2 + (\hat{\eta}_1 - \hat{\eta}_2)^2}],\]

where

\[L = \bar{Y} + \frac{s^2}{2} - \sqrt{\frac{2\sigma_1^2 s^2}{n_1} + \frac{(n_1 - 1)s^2}{2n_1}},\]

and

\[U = \bar{Y} + \frac{s^2}{2} + \sqrt{\frac{2\sigma_1^2 s^2}{n_1} + \frac{(n_1 - 1)s^2}{2n_1}}\]

for \(i = 1, 2\).

Finally, taking exponential function into the lower and upper limits of the CI for the difference two log-normal means, \(\eta_1 - \eta_2\).

Therefore, the confidence interval for ratio of two log-normal means, \(\theta\) based on the MOVER is given by

\[\text{CI}_{\text{heaven}} = \exp[\mu_1 - \mu_2 - \sqrt{(\bar{Y}_1 - \bar{Y}_2)^2 + (\bar{Y}_1 - \bar{Y}_2)^2}],\]

where the lower limit is \(\exp[\mu_1 - \mu_2 - \sqrt{(\bar{Y}_1 - \bar{Y}_2)^2 + (\bar{Y}_1 - \bar{Y}_2)^2}]\) and the upper limit is \(\exp[\mu_1 - \mu_2 + \sqrt{(\bar{Y}_1 - \bar{Y}_2)^2 + (\bar{Y}_1 - \bar{Y}_2)^2}]\).

### 2.4 Normal approximation method (NA)

The parameter of interest is \(\theta = \exp\left(\mu_1 + \frac{\sigma_1^2}{2} - \mu_2 + \frac{\sigma_2^2}{2}\right)\)

and the suggested plug-in estimator \(\hat{\theta} = \exp\left(\bar{Y}_1 + \frac{s_1^2}{2} - \bar{Y}_2 - \frac{s_2^2}{2}\right)\).

Since the ratio of means, \(\theta\), is a complex function. The exact mean and exact variance may be difficult to obtain.

Then in this study, we would like to propose the CI for \(\theta\) based on the normal approximation method with the estimated mean and the variance by using the Delta method and also apply the Delta method for proving the asymptotic normality of this estimator.

In the Delta method, function \(f(V_1, V_2, ..., V_k)\) of \(k\) variables is expanded into the Taylor series at the point...
It is possible to prove that $\sqrt{n} \text{Remainder} \to 0$ in probability as the sample sizes $n_1, n_2 \to \infty$. For details we refer to the famous monograph in [19].

In our case we have two samples $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ from independent log-normal populations with parameters $\mu_1, \sigma_1^2$ and $\mu_2, \sigma_2^2$, respectively. The random variables $Y_{ij} = \ln(X_{ij})$, $j = 1, \ldots, n_i$ follow the normal distribution $N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Based on this fact, the following estimators of $\mu_1$ and $\sigma_1^2$, respectively, are considered:

\[
\hat{\mu}_1 = \bar{Y}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} Y_{ij} \quad \text{and} \quad \hat{\sigma}_1^2 = S_1^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (Y_{ij} - \bar{Y}_1)^2, \quad i = 1, 2.
\]

Note that all four estimators under consideration $\bar{Y}_1, \bar{Y}_2, S_1^2$ and $S_2^2$ are independent because sample mean and sample variance for a normal population are independent and, moreover, the populations are independent.

We also need the following facts:

1. $E(\bar{Y}_1) = \mu_1$ and $E(S_1^2) = \sigma_1^2$.
2. $\text{Var}(\bar{Y}_1) = \frac{\sigma_1^2}{n_1}$ and $\text{Var}(S_1^2) = \frac{2 \sigma_1^4}{n_1 - 1}, \quad i = 1, 2$.

If we denote the basic statistics

\[
V_1 = \bar{Y}_1 - \bar{V}_2 = S_1^2 - S_2^2, \quad V_3 = \bar{V}_2, \quad \text{and} \quad V_4 = S_2^2,
\]

then $\hat{\theta} = g(V_1, V_2, V_3, V_4)$, where the function

\[
g(n_1, n_2, v_2, v_4) = \exp \left(\frac{n_1}{2} v_2 - v_4 \right)
\]

Partial derivatives are as follows:

\[
\frac{\partial g}{\partial v_1} = \exp \left(\frac{n_1}{2} v_2 - v_4 \right) = g(v_1, v_2, v_3, v_4),
\]

\[
\frac{\partial g}{\partial v_2} = \frac{1}{2} g(v_1, v_2, v_3, v_4),
\]

\[
\frac{\partial g}{\partial v_3} = -g(v_1, v_2, v_3, v_4),
\]

\[
\frac{\partial g}{\partial v_4} = -\frac{1}{2} g(v_1, v_2, v_3, v_4).
\]

Remind that

\[
\theta = g(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = \exp \left(\mu_1 + \frac{\sigma_1^2}{2} - \mu_2 - \frac{\sigma_2^2}{2} \right),
\]

then the values of the partial derivatives at the point of means are:

\[
\frac{\partial g}{\partial \mu_1} = \theta,
\]

\[
\frac{\partial g}{\partial \mu_2} = -\theta,
\]

\[
\frac{\partial g}{\partial \sigma_1^2} = \theta,
\]

\[
\frac{\partial g}{\partial \sigma_2^2} = -\theta.
\]

The linear terms of the Taylor expansion of the statistics $\hat{\theta}$ take form

\[
\hat{\theta} = g(\bar{Y}_1, \bar{Y}_2, S_1^2, S_2^2) = \theta + \frac{\partial g}{\partial \mu_1} \frac{\mu_1 - \mu_2}{2} + \frac{\partial g}{\partial \mu_2} \frac{(\bar{Y}_1 - \bar{Y}_2)}{2} - \frac{\partial g}{\partial \sigma_1^2} \frac{(S_1^2 - S_2^2)}{2} - \frac{\partial g}{\partial \sigma_2^2} \frac{(S_1^2 - S_2^2)}{2}.
\]

Therefore, the statistic $\hat{\theta}$ is asymptotically normal with the mean

\[
\text{Asymptotic Mean} = \theta
\]

Variance (remind that all four statistics $\bar{Y}_i$ and $S_i^2, i = 1, 2$ are independent):

\[
\text{Asymptotic Variance} = \tau^2
\]

\[
= \text{Var} \left( \theta \left( 1 + \frac{(\bar{Y}_1 - \bar{Y}_2)}{2} \right) \right) \left( 1 - \frac{(S_1^2 - S_2^2)}{2} \right).
\]

\[
= \theta^2 \left( \frac{\sigma_1^4}{n_1} + \frac{\sigma_2^4}{n_2} + \frac{1}{4} \frac{\sigma_1^4}{n_1 - 1} + \frac{1}{4} \frac{\sigma_2^4}{n_2 - 1} \right)
\]

\[
= \theta^2 \left( \frac{\sigma_1^4}{n_1} + \frac{\sigma_2^4}{n_2} + \frac{1}{2(n_1 - 1)} + \frac{1}{2(n_2 - 1)} \right).
\]

To obtain the plug-in estimator of the variance, we substitute estimations for $\mu_1$ and $\sigma_1^2$, respectively, then

\[
\hat{\tau}^2 = \hat{\theta}^2 \left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2(n_1 - 1)} + \frac{1}{2(n_2 - 1)} \right)
\]

In the following we will deal with the second component of this formula, which we can call the variance component. Hence we use the formula

\[
\tau^2 = \hat{\theta}^2 \left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2(n_1 - 1)} + \frac{1}{2(n_2 - 1)} \right).
\]

If the sample sizes for both sampling schemes tend to infinity, then

\[
P(|\theta - \hat{\theta}| \leq z_{\alpha/2} \tau) \sim 1 - \alpha,
\]

where $z_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. Replacing $\tau^2$ by its plug-in estimators $\hat{\tau}^2$ presented in above, we obtain the same asymptotic equality.

Simple algebra shows that if the sample sizes in both sample schemes tend to infinity, then the intervals with the following end-points,

\[
\hat{\theta} \left( 1 \pm \frac{z_{\alpha/2} \hat{\tau}}{\sqrt{2}} \right)
\]

are the asymptotic $(1 - \alpha)$-confidence sets for the ratio of log-normal means $\theta$.

3. Simulation Studies and Results

A simulation study is performed to evaluate the coverage probability and interval width of the NA in comparison to three existing approaches: ML, MOV, and GCI for constructing 95% confidence intervals for ratio of means of two independent log-normal distributions. We use the nominal level $\alpha = 0.05$ and $N = 10,000$ simulated samples for each parameter setting. All simulations were carried out in R statistical software.
For our simulations, we used the combination of parameter values as follows. Without loss of generality, we set $\mu_1 = \mu_2 = 0$ and three values of $\sigma^2$ were 0.3, 0.5 and 1.0. Three values for $n_i$ were 10, 50 and 100, $i = 1, 2$.

Table I. Coverage probability and interval width of the approaches for constructing 95% confidence intervals for ratio of two log-normal means for homogeneity variances and balance designs

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$\sigma^2$</th>
<th>Methods</th>
<th>(n1,n2)=(10,10)</th>
<th>(n1,n2)=(50,50)</th>
<th>(n1,n2)=(100,100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3 0.3</td>
<td>ML</td>
<td>0.9780</td>
<td>0.9090</td>
<td>0.9742</td>
<td>0.9900</td>
</tr>
<tr>
<td></td>
<td>MOVER</td>
<td>0.9900</td>
<td>0.9212</td>
<td>0.9900</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>0.9910</td>
<td>0.9930</td>
<td>0.9960</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.9040</td>
<td>0.9323</td>
<td>0.9700</td>
<td>0.9999</td>
</tr>
<tr>
<td>0.5 0.5</td>
<td>ML</td>
<td>0.9510</td>
<td>0.8790</td>
<td>0.9580</td>
<td>0.9820</td>
</tr>
<tr>
<td></td>
<td>MOVER</td>
<td>0.9770</td>
<td>0.9660</td>
<td>0.9910</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>0.9510</td>
<td>0.9660</td>
<td>0.9990</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.9040</td>
<td>0.9323</td>
<td>0.9700</td>
<td>0.9999</td>
</tr>
<tr>
<td>1.0 1.0</td>
<td>ML</td>
<td>0.9500</td>
<td>0.8790</td>
<td>0.9580</td>
<td>0.9820</td>
</tr>
<tr>
<td></td>
<td>MOVER</td>
<td>0.9770</td>
<td>0.9660</td>
<td>0.9910</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>0.9510</td>
<td>0.9660</td>
<td>0.9990</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.9040</td>
<td>0.9323</td>
<td>0.9700</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

We report the results of the coverage probability and interval width of all approaches for CIs for ratio of means of two log-normal distributions. Table I for homogeneity variances and balance designs, and Table II for both homogeneous and heterogeneous variances under unbalance designs.

Table I for the homogeneity variances and balance designs, when $\left(\sigma^2_1, \sigma^2_2\right) = (0.3,0.3), (0.5,0.5)$, the coverage probability of the MOVER and GCI were greater than the nominal level for small to moderate sample sizes. For sample sizes were large, the coverage probability of all approaches performed well in terms of the coverage probability. However, the interval width of the NA was always shorter than other approaches. When $\left(\sigma^2_1, \sigma^2_2\right) = (1.0,1.0)$, the coverage probability of all approaches were performed well for small to moderate sample sizes. The NA approach performed well in terms of the interval width.

Table II. Coverage probability and Interval width of four approaches for constructing 95% confidence intervals for ratio of two log-normal means for both homogeneous and heterogeneous variances under unbalance designs

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$\sigma^2$</th>
<th>Methods</th>
<th>(n1,n2)=(50,10)</th>
<th>(n1,n2)=(100,10)</th>
<th>(n1,n2)=(100,50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3 0.3</td>
<td>ML</td>
<td>0.9340</td>
<td>0.8350</td>
<td>0.9190</td>
<td>0.7811</td>
</tr>
<tr>
<td></td>
<td>MOVER</td>
<td>0.9930</td>
<td>1.1195</td>
<td>0.9930</td>
<td>1.1466</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>0.9560</td>
<td>0.9043</td>
<td>0.9510</td>
<td>0.8485</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.9380</td>
<td>0.8100</td>
<td>0.9220</td>
<td>0.7601</td>
</tr>
<tr>
<td>0.5 0.5</td>
<td>ML</td>
<td>0.9340</td>
<td>1.1661</td>
<td>0.9190</td>
<td>1.0662</td>
</tr>
<tr>
<td></td>
<td>MOVER</td>
<td>0.9930</td>
<td>1.6502</td>
<td>0.9910</td>
<td>1.5878</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>0.9560</td>
<td>1.2201</td>
<td>0.9470</td>
<td>1.1336</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.9380</td>
<td>1.1031</td>
<td>0.9220</td>
<td>1.0336</td>
</tr>
<tr>
<td>1.0 1.0</td>
<td>ML</td>
<td>0.9340</td>
<td>1.6607</td>
<td>0.9180</td>
<td>1.5580</td>
</tr>
<tr>
<td></td>
<td>MOVER</td>
<td>0.9930</td>
<td>2.3399</td>
<td>0.9840</td>
<td>2.2456</td>
</tr>
<tr>
<td></td>
<td>GCI</td>
<td>0.9560</td>
<td>1.6481</td>
<td>0.9460</td>
<td>1.5060</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>0.9380</td>
<td>1.5034</td>
<td>0.9140</td>
<td>1.4054</td>
</tr>
</tbody>
</table>

Table II for both homogeneous and heterogeneous variances under unbalance designs. When the sample sizes were small to moderate $(n_1,n_2) = (50,10),(100,10)$, all approaches performed well in terms of the coverage probability, except when $(\sigma^2_1, \sigma^2_2) = (0.5,1.0)$ the coverage probability of the ML was less than the nominal level Furthermore, the NA approach always performed well in terms of the interval width for all values of the variances.

4. A Real Life Example

According to Tai et al. [20], the illustrative example deals with survival times in months for patients who died from a particular cancer. Data of the first group were constructed for 184 patients who had limited stage small-cell lung cancer (LC) and the survival time for the second group were constructed from 38 patients who died of cervical cancer (CC). After taking the natural logarithm, the calculated statistics for these two group of log-transformed measurements are as follows Table III.

Table III. Sample size, sample means and sample variances for the two data sets

<table>
<thead>
<tr>
<th>Groups</th>
<th>$n_1$</th>
<th>$\mu_1$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small-Cell Lung cancer (LC)</td>
<td>184</td>
<td>2.8591</td>
<td>0.2461</td>
</tr>
<tr>
<td>Cervical cancer (CC)</td>
<td>38</td>
<td>3.3900</td>
<td>0.6454</td>
</tr>
</tbody>
</table>

The histogram and corresponding density plots, Q-Q plot and fitted density for log-transformed data for both groups were shown in Fig. 1(a)-(b).
Fig. 1 (a) The histogram and corresponding density plots for both groups, (b) and the Q-Q plot and fitted density for log-transformed data for both groups.

Moreover, To confirm that the log-normal distributions appropriate for two datasets. The Shapiro-Wilk test was used and p-values for the first group was 0.5306 and for the second group was 0.1186. These results indicate that the log-normal models is adequate for two datasets.

To compare the ratio of two log-normal means, the 95% confidence interval (lower, upper) and the average width by using the ML, MOVER, GCI, and NA approaches are demonstrated in Table IV.

<table>
<thead>
<tr>
<th>Approach</th>
<th>ML</th>
<th>MOVER</th>
<th>GCI</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% CI</td>
<td>(0.3552, 0.5529)</td>
<td>(0.2791, 0.9124)</td>
<td>(0.3366, 0.6348)</td>
<td>(0.3356, 0.6223)</td>
</tr>
<tr>
<td>Average width</td>
<td>0.2977</td>
<td>0.6333</td>
<td>0.2982</td>
<td>0.2932</td>
</tr>
</tbody>
</table>

This is in agreement with our simulation study which indicated that the NA approach is shorter interval width than other approaches.

5. Discussion

In this paper, we would like to identify potential methods that can be recommended to practitioners for constructing confidence interval for the ratio of means of two independent log-normal distributions. From the simulation results presented in Table I and II.

In Table I for homogeneity variances and balance designs, when small to moderate sample sizes the GCI and MOVER approaches were better than other approaches in terms of the coverage probability and interval width. The NA approach always shortest interval width especially when sample sizes were larger and variance were decrease. For both homogeneous and heterogeneous variances and unbalance designs were shown in Table II. When small to moderate sample sizes, MOVER leads to coverage probabilities closer to the nominal level while GCI approach are better than the MOVER in terms of interval width. For large sample sizes, NA has better coverage probability and interval width than other approaches.

6. Conclusion

The purpose of the paper is to propose a simple approach for the problem of the CI for ratio of means of two log-normal distributions and compare performance with the existing approaches. A simulation studies and results were conduct to compare the performance of these methods in terms of the coverage probability and interval width to indicate which method is optimal under difference condition of the population of two independent log-normal distributions. The results indicate that the GCI and MOVER approaches perform well for all situations. However, when he sample size were large, the NA approach performed better than other approaches in terms of interval width.

Acknowledgment

The authors thank the reviewers for suggestions which lead to considerable improvements in the manuscript.

References


