Journal of Statistical and Econometric Methods, vol.3, no.1, 2014, 137-152
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2014

Comparing the Means of Two Log-Normal Distributions: A Likelihood Approach

L. Jiang[1], M. Rekkas[2] and A. Wong[3]

Abstract

The log-normal distribution is one of the most common distributions used for modeling skewed and positive data. In recent years, various methods for comparing the means of two independent log-normal distributions have been developed. In this paper a higher-order likelihood-based method is proposed. The method is applied to two real-life examples and simulation studies are used compare the accuracy of the proposed method to some existing methods.

Mathematics Subject Classification: 62F03

Keywords: Confidence interval; Coverage probability; Modified signed log-likelihood ratio statistic; r^*-formula

[1] Beijing Education Examination Authority, No. 9 Zhixin Road, Haidian District, Beijing, P.R. China, 100083, e-mail: jiangl@bjeea.cn

[2] Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6, e-mail: mrekkas@sfu.ca

[3] Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada, M3J 1P3, e-mail: august@yorku.ca

Article Info: *Received* : December 18, 2013. *Revised* : January 23, 2014.
Published online : February 7, 2014

1 Introduction

In real life applications, cases arise when random variables take on strictly positive values where the use of the normal distribution is not appropriate for statistical inference. The use of the log-normal distribution features prominently in these settings. For example, in medical research, the log-normal distribution is commonly used to model the incubation periods of certain diseases, and the survival times of cancer patients. In economics and business studies, the log-normal distribution is used to model income, firm size and stock prices. In this paper, we use higher-order likelihood-based analysis to derive a test statistic for comparing the means of two independent log-normal distributions. We compare the performance of this test statistic with three existing approaches in the literature. These approaches are due to Zhou et al. (1997), Abdollahnezhad et al. (2012) and the signed log-likelihood ratio statistic method. Simulation studies show that our proposed method outperforms the three existing methods.

Let X and Y be independently distributed with

$$\log X \sim N(\mu_x, \sigma_x^2) \quad \text{and} \quad \log Y \sim N(\mu_y, \sigma_y^2)\ .$$

The random variables X and Y are said to have a log-normal distribution with means $M_x = \exp\{\mu_x + \sigma_x^2/2\}$ and $M_y = \exp\{\mu_y + \sigma_y^2/2\}$, respectively. Then testing:

$$H_0 : M_x = M_y \qquad \text{vs} \qquad H_a : M_x < M_y$$

is equivalent to testing:

$$H_0 : \psi = 0 \qquad \text{vs} \qquad H_a : \psi < 0,$$

where $\psi = \log M_x - \log M_y = \log(M_x/M_y) = \mu_x - \mu_y + (\sigma_x^2 - \sigma_y^2)/2$. In this paper, we consider a slightly more general hypothesis:

$$H_0 : \psi = \psi_0 \qquad \text{vs} \qquad H_a : \psi < \psi_0\ . \tag{1}$$

Consider $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$ to be independent samples from $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$, respectively. Zhou et al. (1997) propose a likelihood-based test and show that the p-value for testing hypothesis (1) can be approximated by

$$p_Z(\psi_0) = \Phi(z), \tag{2}$$

where $\Phi()$ is the cumulative distribution function of $N(0,1)$ and

$$z = \frac{\hat{\mu}_x - \hat{\mu}_y + (s_x^2 - s_y^2)/2 - \psi_0}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m} + \frac{1}{2}\left(\frac{s_x^4}{n-1} + \frac{s_y^4}{m-1}\right)}}, \tag{3}$$

where $\hat{\mu}_x = \sum \log x_i/n$ and $\hat{\mu}_y = \sum \log y_i/m$ are the maximum likelihood estimators of μ_x and μ_y, respectively. The terms $s_x^2 = \sum(\log x_i - \hat{\mu}_x)^2/(n-1)$ and $s_y^2 = \sum(\log y_i - \hat{\mu}_y)^2/(m-1)$ are the unbiased estimators of σ_x^2 and σ_y^2, respectively. Moreover, a $(1-\alpha)100\%$ confidence interval for ψ is given by

$$\left(\hat{\psi} - z_{\alpha/2}\sqrt{\widehat{var}(\hat{\psi})},\quad \hat{\psi} + z_{\alpha/2}\sqrt{\widehat{var}(\hat{\psi})}\right), \tag{4}$$

where

$$\begin{aligned} \hat{\psi} &= \hat{\mu}_x - \hat{\mu}_y + (s_x^2 - s_y^2)/2 \\ \widehat{var}(\hat{\psi}) &= \frac{s_x^2}{n} + \frac{s_y^2}{m} + \frac{1}{2}\left(\frac{s_x^4}{n-1} + \frac{s_y^4}{m-1}\right) \end{aligned}$$

and $z_{\alpha/2}$ is the $(1-\alpha/2)100^{th}$ percentile of $N(0,1)$.

Abdollahnezhad et al. (2012) apply a generalized p-value approach to obtain the p-value for testing hypothesis (1). This generalized p-value is obtained using the following algorithm:

Step 1: For $k = 1$ to K, where K is reasonably large:

(a) generate u_x from χ^2_{n-1} and u_y from χ^2_{m-1}

(b) calculate

$$p_k = \Phi\left(\frac{\hat{\mu}_x - \hat{\mu}_y + \frac{n\hat{\sigma}_x^2}{2u_x} - \frac{m\hat{\sigma}_y^2}{2u_y} - \psi_0}{\sqrt{\frac{\hat{\sigma}_x^2}{u_x} + \frac{\hat{\sigma}_y^2}{u_y}}}\right), \tag{5}$$

where $\hat{\sigma}_x^2 = \sum(\log x_i - \hat{\mu}_x)^2/n$, and $\hat{\sigma}_y^2 = \sum(\log y_i - \hat{\mu}_y)^2/m$

Step 2: $p_A(\psi_0) = \sum_{k=1}^{K} p_k/K$.

The corresponding $(1-\alpha)100\%$ confidence interval for ψ using the method in Abdollahnezhad et al. (2012) is (ψ_L, ψ_U) such that $p(\psi_L) = 1 - \alpha/2$ and $p(\psi_U) = \alpha/2$. Note that this method requires extra computing time because for each calculation, we need to generate K values of u_x and K values of u_y respectively. Thus, obtaining confidence intervals will require substantial computing time, especially when K is large.

This paper proceeds as follows. In Section 2, we present the likelihood-based inference methods for a scalar parameter from an exponential family model. In Section 3 we apply these methods for inference on the difference of the means of two independent log-normal distributions. In Section 4 we present our numerical results and some concluding remarks are given in Section 5.

2 Likelihood-Based Inference for a Scalar Parameter from an Exponential Family Model

In this section we provide an overview of the third-order likelihood-based inference method we will use in this paper for testing the difference of means of two independent log-normal distributions. We begin by reviewing the standard first-order likelihood method. Consider a random sample $(y_1, \dots y_n)$ obtained from a distribution with density $f(y;\theta)$, where $\theta = (\psi, \lambda')'$ with ψ being the scalar interest parameter and λ the vector of nuisance parameters. The log-likelihood function based on the given sample is then

$$l(\theta) = l(\psi, \lambda) = \sum \log f(y_i; \theta).$$

From this function, the signed log-likelihood ratio statistic (r) can be obtained:

$$r \equiv r(\psi) = \text{sgn}(\hat{\psi} - \psi)\{2[l(\hat{\theta}) - l(\tilde{\theta})]\}^{1/2}, \tag{6}$$

where $\hat{\theta}$ represents the maximum likelihood estimator of θ satisfying the first-order conditions, $\partial l(\theta)/\partial\theta = 0$, and $\tilde{\theta}$ represents the constrained maximum likelihood estimator of θ for a given ψ. The constrained maximum likelihood estimator can be solved by maximizing the log-likelihood function subject to the constraint $\psi(\theta) = \psi$. The method of Lagrange multipliers can be used to solve this optimization problem. The function

$$H(\theta, \alpha) = l(\theta) + \alpha[\psi(\theta) - \psi] \tag{7}$$

is known as the Lagrangean, where α is termed the Lagrange multiplier. Then $(\tilde{\theta}, \hat{\alpha})$ is the solution from solving the equations

$$\frac{\partial H(\theta, \alpha)}{\partial \theta} = 0 \qquad \text{and} \qquad \frac{\partial H(\theta, \alpha)}{\partial \alpha} = 0$$

simultaneously. The tilted log-likelihood is defined as

$$\tilde{l}(\theta) = l(\theta) + \hat{\alpha}[\psi(\theta) - \psi]. \tag{8}$$

Note that the tilted log-likelihood function has the property that $\tilde{l}(\tilde{\theta}) = l(\tilde{\theta})$.

The signed log-likelihood ratio statistic in (6) is asymptotically distributed as $N(0,1)$ with rate of convergence $O(n^{-1/2})$ and is referred to as a first-order method. Hence, the p-value for testing

$$H_0 : \psi = \psi_0 \qquad \text{vs} \qquad H_a : \psi < \psi_0$$

using the signed log-likelihood ratio statistic method is

$$p_r(\psi_0) = \Phi(r(\psi_0)). \tag{9}$$

The $(1-\alpha)100\%$ confidence interval for ψ is

$$\{\psi : |r(\psi)| < z_{\alpha/2}\} . \tag{10}$$

It is well-known that the signed log-likelihood ratio statistic method does not perform satisfactorily well in small sample situations. Many improvements have been suggested in the literature. In particular, Barndorff-Nielsen (1986, 1991) proposes a modified signed log-likelihood ratio statistic, given by

$$r^* \equiv r^*(\psi) = r(\psi) + r(\psi)^{-1} \log \left\{ \frac{q(\psi)}{r(\psi)} \right\}, \tag{11}$$

where $r(\psi)$ is the signed log-likelihood ratio statistic given in (6) and $q(\psi)$ is a statistic that can be derived in various ways depending on the information at hand. It is shown in Barndorff-Nielsen (1986, 1991) that the modified signed log-likelihood ratio statistic is asymptotically distributed as $N(0,1)$ with rate of convergence $O(n^{-3/2})$ and is referred to as a third-order method. Hence the corresponding p-value is

$$p_{BN}(\psi_0) = \Phi(r^*(\psi_0)) \tag{12}$$

and the corresponding $(1-\alpha)100\%$ confidence interval for ψ is

$$\{\psi : |r^*(\psi)| < z_{\alpha/2}\} . \tag{13}$$

Note that the most difficult aspect of the methodology is obtaining $q(\psi)$.

Fraser and Reid (1995) provide a systematic approach to obtain $q(\psi)$ for a full rank exponential model with density

$$f(y;\theta) = \exp\{\varphi'(\theta)t(y) - c(\theta) + h(t(y))\},$$

where $\varphi(\theta)$ is the canonical parameter and $t(y)$ is the canonical variable. For this model, $q(\psi)$ takes the form

$$q \equiv q(\psi) = \text{sgn}(\psi(\hat{\theta}) - \psi)|\chi(\hat{\theta}) - \chi(\tilde{\theta})| \left[\widehat{var}\left(\chi(\hat{\theta}) - \chi(\tilde{\theta})\right)\right]^{-1/2}, \tag{14}$$

where $\chi(\theta)$ is the standardized maximum likelihood estimator recalibrated in the $\varphi(\theta)$ scale:

$$\chi(\theta) = \psi'_\theta(\tilde{\theta})\varphi_\theta^{-1}(\tilde{\theta})\varphi(\theta) \tag{15}$$

where

$$\varphi_\theta(\theta) = \frac{\partial\varphi(\theta)}{\partial\theta}\ .$$

The variance term in (14) is computed as

$$\widehat{var}\left(\chi(\hat{\theta}) - \chi(\tilde{\theta})\right) \approx \frac{\psi_\theta(\tilde{\theta})\tilde{\jmath}_{\theta\theta'}^{-1}(\tilde{\theta})\psi'_\theta(\tilde{\theta})|\tilde{\jmath}_{\theta\theta'}(\tilde{\theta})||\varphi_\theta(\tilde{\theta})|^{-2}}{|j_{\theta\theta'}(\hat{\theta})||\varphi_\theta(\hat{\theta})|^{-2}} \tag{16}$$

where $j_{\theta\theta'}(\hat{\theta})$ and $\tilde{\jmath}_{\theta\theta'}(\tilde{\theta})$ are defined as

$$j_{\theta\theta'}(\hat{\theta}) = -\left.\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\hat{\theta}}$$

and

$$\tilde{\jmath}_{\theta\theta'}(\tilde{\theta}) = -\left.\frac{\partial^2 \tilde{l}(\theta)}{\partial\theta\partial\theta'}\right|_{\theta=\tilde{\theta}}$$

and are referred to as the observed information matrix obtained from $l(\theta)$ and $\tilde{l}(\theta)$ respectively, and

$$\psi_\theta(\theta) = \frac{\partial\psi(\theta)}{\partial\theta}\ .$$

The idea of the method is to recalibrate (11) in the canonical parameter, $\varphi(\theta)$, scale. Since the signed log-likelihood ratio statistic is invariant to reparameterization, it remains the same as in (6). The only quantity that needs recalibration in the $\varphi(\theta)$ scale is therefore $q(\theta)$. For a detailed discussion of this derivation, see Fraser and Reid (1995).

3 Inference for the Difference of Means of Two Independent Log-Normal Distributions

We apply the modified signed log-likelihood ratio statistic method to our problem of interest. The log-likelihood function is:

$$\begin{aligned} l(\theta) &= -\frac{n}{2}\log\sigma_x^2 - \frac{n\mu_x^2}{2\sigma_x^2} + \frac{\mu_x}{\sigma_x^2}\sum_{i=1}^{n}\log x_i - \frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(\log x_i)^2 \\ &\quad -\frac{m}{2}\log\sigma_y^2 - \frac{m\mu_y^2}{2\sigma_y^2} + \frac{\mu_y}{\sigma_y^2}\sum_{j=1}^{m}\log y_j - \frac{1}{2\sigma_y^2}\sum_{j=1}^{m}(\log y_j)^2 \\ &= -\frac{n}{2}\log\sigma_x^2 - \frac{n\mu_x^2}{2\sigma_x^2} + \frac{\mu_x}{\sigma_x^2}t_1 - \frac{1}{2\sigma_x^2}t_3 \\ &\quad -\frac{m}{2}\log\sigma_y^2 - \frac{m\mu_y^2}{2\sigma_y^2} + \frac{\mu_Y}{\sigma_y^2}t_2 - \frac{1}{2\sigma_y^2}t_4, \end{aligned} \tag{17}$$

where

$$(t_1,t_2,t_3,t_4) = \left(\sum\log x_i, \sum\log y_i, \sum(\log x_i)^2, \sum(\log y_i)^2\right)$$

is a minimal sufficient statistic, $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)'$ and our parameter of interest is

$$\psi = \psi(\theta) = (\mu_x - \mu_y) + (\sigma_x^2 - \sigma_y^2)/2\ . \tag{18}$$

Moreover, the canonical parameter $\varphi(\theta)$ is

$$\varphi \equiv \varphi(\theta) = \left(\frac{\mu_x}{\sigma_x^2}, \frac{\mu_y}{\sigma_y^2}, \frac{1}{\sigma_x^2}, \frac{1}{\sigma_y^2}\right)'\ . \tag{19}$$

The first and second derivatives of $l(\theta)$ are given in the Appendix. By solving the first-order conditions, we have the maximum likelihood estimator of θ:

$$\hat\theta = (\hat\mu_x, \hat\mu_y, \hat\sigma_x^2, \hat\sigma_y^2)',$$

where

$$\begin{aligned} \hat\mu_x &= \frac{t_1}{n} = \bar t_1 \\ \hat\mu_y &= \frac{t_2}{m} = \bar t_2 \\ \hat\sigma_x^2 &= \frac{2}{n}\left[\frac{n}{2}(\bar t_1)^2 - n(\bar t_1)^2 + \frac{1}{2}t_3\right] = \frac{1}{n}\left[t_3 - n(\bar t_1)^2\right] \\ \hat\sigma_y^2 &= \frac{2}{m}\left[\frac{m}{2}(\bar t_2)^2 - m(\bar t_2)^2 + \frac{1}{2}t_4\right] = \frac{1}{m}\left[t_4 - m(\bar t_2)^2\right]. \end{aligned}$$

The corresponding observed information matrix is given by:

$$j_{\theta\theta}(\hat{\theta}) = \begin{pmatrix} \frac{n}{\hat{\sigma}_x^2} & 0 & 0 & 0 \\ 0 & \frac{m}{\hat{\sigma}_y} & 0 & 0 \\ 0 & 0 & \frac{n}{2\hat{\sigma}_x^4} & 0 \\ 0 & 0 & 0 & \frac{m}{2\hat{\sigma}_y^4} \end{pmatrix} . \quad (20)$$

The constrained maximum likelihood estimator is derived using the Lagrangean defined in (7):

$$H(\theta, \alpha) = l(\theta) + \alpha \left[(\mu_x - \mu_y) + \frac{1}{2}(\sigma_x^2 - \sigma_y^2) - \psi_0 \right], \quad (21)$$

with derivatives given by:

$$\begin{aligned} \frac{\partial H(\theta, \alpha)}{\partial \mu_x} &= \frac{\partial l(\theta)}{\partial \mu_x} + \alpha \\ \frac{\partial H(\theta, \alpha)}{\partial \mu_y} &= \frac{\partial l(\theta)}{\partial \mu_y} - \alpha \\ \frac{\partial H(\theta, \alpha)}{\partial \sigma_x^2} &= \frac{\partial l(\theta)}{\partial \sigma_x^2} + \frac{1}{2}\alpha \\ \frac{\partial H(\theta, \alpha)}{\partial \sigma_y^2} &= \frac{\partial l(\theta)}{\partial \sigma_y^2} - \frac{1}{2}\alpha \\ \frac{\partial H(\theta, \alpha)}{\partial \alpha} &= (\mu_X - \mu_Y) + \frac{1}{2}(\sigma_x^2 - \sigma_y^2) - \psi_0, \end{aligned}$$

from which $(\tilde{\theta}, \hat{\alpha})$ can be obtained. The tilted log-likelihood from (8) is given as:

$$\tilde{l}(\theta) = l(\theta) + \hat{\alpha} \left[(\mu_x - \mu_y) + \frac{1}{2}(\sigma_x^2 - \sigma_y^2) - \psi_0 \right] . \quad (22)$$

It is easy to show that $\tilde{l}(\tilde{\theta}) = l(\tilde{\theta})$ and $\tilde{j}_{\theta\theta'}(\tilde{\theta}) = j_{\theta\theta'}(\tilde{\theta})$.

Since the canonical parameter $\varphi(\theta)$ is given in (19), we have:

$$\varphi_\theta(\theta) = \begin{pmatrix} \frac{1}{\sigma_x^2} & 0 & -\frac{\mu_x}{\sigma_x^4} & 0 \\ 0 & \frac{1}{\sigma_y^2} & 0 & -\frac{\mu_y}{\sigma_y^4} \\ 0 & 0 & -\frac{1}{\sigma_x^4} & 0 \\ 0 & 0 & 0 & -\frac{1}{\sigma_y^4} \end{pmatrix},$$

and

$$\varphi_\theta^{-1}(\theta) = \begin{pmatrix} \sigma_x^2 & 0 & -\mu_x\sigma_x^2 & 0 \\ 0 & \sigma_y^2 & 0 & -\mu_y\sigma_y^2 \\ 0 & 0 & -\sigma_x^4 & 0 \\ 0 & 0 & 0 & -\sigma_y^4 \end{pmatrix}.$$

Moreover, the parameter of interest $\psi = \psi(\theta)$ is given in (18) and we have

$$\psi_\theta(\theta) = \begin{pmatrix} 1 & -1 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}'.$$

Hence, the recalibrated parameter $\chi(\theta)$ can be obtained by (15). Given the computed quantities above, we are able to construct the modified signed log-likelihood ratio statistic given in (11), with $q(\psi)$ defined as in (14).

It is important to note that theoretically, $(\tilde{\theta}, \hat{\alpha})$ is uniquely defined. However, special care must be taken when performing numerical calculations as standard optimization subroutines in standard statistical software may not converge to the true $\tilde{\theta}$, which results in a negative definite $\tilde{\jmath}_{\theta\theta'}(\tilde{\theta})$.

4 Numerical Results

We present two real-life examples followed by simulations to compare results obtained from the following methods: the Z-score method proposed by Zhou et al. (1997) (*Zhou*), the generalized test method proposed by Abdollahnezhad et al. (2012) (*Abdollahnezhad*), the signed log-likelihood ratio statistic method (r), and the proposed modified signed log-likelihood ratio statistic method (r^*).

The first example is discussed in Abdollahnezhad et al. (2012). They considered the amount of rainfall (in acre-feet) from 52 clouds, of which 26 were chosen at random and seeded with silver nitrate. The log-normal model is used and the summary statistics expressed in this paper's notation are:

$$n = 26,\ m = 26,\ t_1 = 133.484,\ t_2 = 103.74,\ t_3 = 749.2669,\ t_4 = 481.5226\ .$$

Let M_x and M_y be the mean rainfall for the seeded clouds and the mean rainfall for the unseeded clouds respectively. The p-values, obtained from the four methods discussed in this paper, for testing

$$H_0 : M_x = M_y \text{ vs } H_a : M_x > M_y \quad \Leftrightarrow \quad H_0 : \psi = 0 \text{ vs } H_a : \psi > 0\ ,$$

where ψ is the logarithm of the ratio of the log-normal means, are recorded in Table 1.

Table 1: p-values for testing $H_0 : \psi = 0$ vs $H_a : \psi > 0$ for the rainfall example

Method	p-value
Zhou	0.061
Abdollahnezhad	0.080
r	0.066
r^*	0.078

The 95% confidence interval for the ratio of the log-normal means are recorded in Table 2.

Table 2: 95% confidence interval for $\frac{M_x}{M_y}$ for the rainfall example

Method	95% Confidence interval
Zhou	(0.751, 11.342)
Abdollahnezhad	(0.600, 13.587)
r	(0.681, 12.150)
r^*	(0.606, 13.450)

From these tables it is clear that the four methods produce quite different inferential results.

The second example is a bioavailability study. A randomized, parallel-group experiment was conducted with 20 subjects to compare a new test formulation (x), with a reference formulation (y), of a drug product with a long half life. The data from this study is given in Table 3.

The Shapiro-Wilk tests for the normality on the log-transformed data give a p-value of 0.595 for the test formulation group and 0.983 for the reference formulation group. These tests suggest that the log-normal model is suitable for this data set. The p-values, obtained from the four methods discussed in

Table 3: Data for the bioavailability study

x	732.89	1371.97	614.62	557.24	821.39
	363.94	430.95	401.42	436.16	951.46
y	1053.63	1351.54	197.95	1204.72	447.20
	3357.66	567.36	668.48	842.19	284.86

this paper, for testing

$$H_0 : M_x = M_y \text{ vs } H_a : M_x \neq M_y \quad \Leftrightarrow \quad H_0 : \psi = 0 \text{ vs } H_a : \psi \neq 0 \ ,$$

where ψ is the logarithm of the ratio of the log-normal means, are recorded in Table 4.

Table 4: p-values for testing $H_0 : \psi = 0$ vs $H_a : \psi \neq 0$ for the bioavailability example

Method	p-value
Zhou	0.204
Abdollahnezhad	0.182
r	0.167
r^*	0.173

The 95% confidence interval for the ratio of the log-normal means are recorded in Table 5.

Table 5: 95% confidence interval for $\frac{M_x}{M_y}$ for the bioavailability example

Method	95% Confidence interval
Zhou	(0.339, 1.259)
Abdollahnezhad	(0.226, 1.236)
r	(0.451, 1.181)
r^*	(0.242, 1.120)

Again the results obtained from the four methods discussed in this paper are quite different.

To compare the accuracy of the four methods discussed in this paper, simulation studies are performed. 10,000 simulated samples from each combination of parameters are generated. For each generated sample, the 90% confidence interval for $\frac{M_x}{M_y}$ is calculated for each of the four methods discussed in this paper. The performance of a method is judged using the following criteria:

- the coverage probability (CP):
  Proportion of the true $\frac{M_x}{M_y}$ falling within the 90% confidence interval

- the lower tail error rate (LE):
  Proportion of the true $\frac{M_x}{M_y}$ falling below the lower limit of the 90% confidence interval

- the upper tail error rate (LE):
  Proportion of the true $\frac{M_x}{M_y}$ falling above the upper limit of the 905% confidence interval

- the average bias (AB):

$$AB = \frac{|LE - 0.025| + |UE - 0.025|}{2} .$$

The desired values are 0.9, 0.05, 0.05 and 0 respectively. These values reflect the desired properties of the accuracy and symmetry of the interval estimates of $\frac{M_x}{M_y}$. Results are recorded in Table 6. It is clear that the method by Zhou et al. (1997) and the signed log-likelihood ratio method do not give satisfactory results. The method by Abdollahnezhad et al. (2012) is an improvement of the other two methods but still it is not as good as the proposed modified signed log-likelihood ratio method.

Simulation studies for other combinations of parameters have also been performed and the same pattern is observed. These results are available from the authors upon request.

Table 6: Simulation results

n	μ_x	σ_x^2	m	μ_y	σ_y^2	Method	CP	LE	UE	AB
5	1.1	0.4	10	1.2	0.2	*Zhou*	0.859	0.045	0.096	0.0255
						Abdollahnezhad	0.916	0.049	0.036	0.0075
						r	0.851	0.063	0.087	0.0250
						r^*	0.895	0.053	0.052	0.0025
5	2.5	1.5	10	3.0	0.5	*Zhou*	0.855	0.016	0.129	0.0565
						Abdollahnezhad	0.909	0.051	0.040	0.0055
						r	0.847	0.048	0.105	0.0285
						r^*	0.898	0.049	0.053	0.0020
10	1.1	0.4	10	1.2	0.2	*Zhou*	0.886	0.047	0.067	0.0100
						Abdollahnezhad	0.919	0.044	0.037	0.0095
						r	0.878	0.058	0.064	0.0110
						r^*	0.900	0.051	0.049	0.0010
10	2.5	1.5	10	3.0	0.5	*Zhou*	0.889	0.023	0.088	0.0325
						Abdollahnezhad	0.912	0.048	0.040	0.0060
						r	0.876	0.050	0.074	0.0120
						r^*	0.901	0.051	0.049	0.0010

5 Discussion

In this paper, four methods are studied for comparing the means of two independent log-normal distributions. In terms of computation, the method by Zhou et al. (1997) is the simplest. The method by Abdollahnezhad et al. (2012) takes up the most computing time because for each calculation, we have to simulate K samples for u_x and also K samples for u_y. Both the signed log-likelihood ratio statistic method and the proposed modified signed log-likelihood statistic method require the constrained maximum likelihood estimator, $\tilde{\theta}$. Blindly applying standard optimization subroutines may result in $\tilde{\jmath}_{\theta\theta'}(\tilde{\theta})$ being a negative definite matrix. The two real-life examples illustrate that the four methods can give very different inferential results. Simulation studies show the proposed modified signed log-likelihood ratio statistic method to be superior to the other three methods. From a theoretical perspective, the proposed modified signed log-likelihood ratio statistic method has the advantage that it has a known rate of convergency, $O(n^{-3/2})$, whereas the signed log-likelihood ratio statistic method only has rate of convergency $O(n^{-1/2})$. The rate of convergency of the method by Zhou et al. (1997) and the method by Abdollahnezhad et al. (2012) are unknown. Thus the proposed modified signed log-likelihood ratio statistic method is recommended for comparing the means of two independent log-normal distributions.

Appendix

The first and second derivatives of $l(\theta)$ given in (17):

$$
\begin{aligned}
\frac{\partial l(\theta)}{\partial \mu_x} &= -\frac{n\mu_x}{\sigma_x^2} + \frac{1}{\sigma_x^2}t_1 \\
\frac{\partial l(\theta)}{\partial \mu_y} &= -\frac{m\mu_y}{\sigma_y^2} + \frac{1}{\sigma_y^2}t_2 \\
\frac{\partial l(\theta)}{\partial \sigma_x^2} &= -\frac{n}{2\sigma_x^2} + \frac{1}{\sigma_x^4}\left[\frac{n}{2}\mu_x^2 - t_1\mu_x + \frac{1}{2}t_3\right] \\
\frac{\partial l(\theta)}{\partial \sigma_y^2} &= -\frac{m}{2\sigma_y^2} + \frac{1}{\sigma_y^4}\left[\frac{m}{2}\mu_y^2 - t_2\mu_y + \frac{1}{2}t_4\right] \\
\frac{\partial^2 l(\theta)}{\partial \mu_x^2} &= -\frac{n}{\sigma_x^2} \\
\frac{\partial^2 l(\theta)}{\partial \mu_x \partial \mu_y} &= 0 \\
\frac{\partial^2 l(\theta)}{\partial \mu_x \partial \sigma_x^2} &= \frac{n\mu_s}{\sigma_s^4} - \frac{1}{\sigma_x^4}t_1 \\
\frac{\partial^2 l(\theta)}{\partial \mu_x \partial \sigma_y^2} &= 0 \\
\frac{\partial^2 l(\theta)}{\partial \mu_y^2} &= -\frac{m}{\sigma_y^2} \\
\frac{\partial^2 l(\theta)}{\partial \mu_y \partial \sigma_x^2} &= 0 \\
\frac{\partial^2 l(\theta)}{\partial \mu_y \partial \sigma_y^2} &= \frac{m\mu_y}{\sigma_y^4} - \frac{1}{\sigma_y^4}t_2 \\
\frac{\partial^2 l(\theta)}{\partial \sigma_x^2 \partial \sigma_x^2} &= \frac{n}{2\sigma_x^4} - \frac{2}{\sigma_x^6}\left[\frac{n}{2}\mu_x^2 - t_1\mu_x + \frac{1}{2}t_3\right] \\
\frac{\partial^2 l(\theta)}{\partial \sigma_x^2 \partial \sigma_y^2} &= 0 \\
\frac{\partial^2 l(\theta)}{\partial \sigma_y^2 \partial \sigma_y^2} &= \frac{m}{2\sigma_y^4} - \frac{2}{\sigma_y^6}\left[\frac{m}{2}\mu_y^2 - t_2\mu_y + \frac{1}{2}t_4\right].
\end{aligned}
$$

References

[1] K. Abdollahnezhad, M. Babanezhad and A. Jafari, Inference on Difference of Means of two Log-Normal Distributions: A Generalized Approach, *Journal of Statistical and Econometric Methods*, **1**, (2012), 125-131.

[2] O.E. Barndorff-Nielsen, Inference on full or partial parameters based on the standardized signed log likelihood ratio, *Biometrika*, **73**, (1986), 307-322.

[3] O.E. Barndorff-Nielsen, Modified signed log likelihood ratio, *Biometrika*, **78**, (1991), 557-563.

[4] D.A.S. Fraser and N. Reid, Ancillaries and third order significance, *Utilitas Mathemation*, **7**, (1995), 33-35.

[5] X. Zhou, S. Gao and S. Hui, Methods for Comparing the Means of Two Independent Log-Normal Samples, *Biometrics*, **53**, (1997), 1129-1135.