Likelihood Analysis for the Ratio of Means of Two Independent Log-Normal Distributions

Jianrong Wu,[1] Guoyong Jiang,[2,*] A. C. M. Wong,[3] and Xiang Sun[4]

[1]Division of Public Health Science, Fred Hutchinson Cancer Research Center,
Seattle, Washington 98109, U.S.A.
[2]Cephalon, Inc., West Chester, Pennsylvania 19380, U.S.A.
[3]SASIT, Atkinson Faculty of Professional and Liberal Studies, North York, Ontario M3J 1P3, Canada
[4]Department of Biostatistics, School of Public Health, University of Michigan,
Ann Arbor, Michigan 48109, U.S.A.
*email: jjiang@cephalon.com

SUMMARY. Existing methods for comparing the means of two independent skewed log-normal distributions do not perform well in a range of small-sample settings such as a small-sample bioavailability study. In this article, we propose two likelihood-based approaches—the signed log-likelihood ratio statistic and modified signed log-likelihood ratio statistic—for inference about the ratio of means of two independent log-normal distributions. More specifically, we focus on obtaining p-values for testing the equality of means and also constructing confidence intervals for the ratio of means. The performance of the proposed methods is assessed through simulation studies that show that the modified signed log-likelihood ratio statistic is nearly an exact approach even for very small samples. The methods are also applied to two real-life examples.

KEY WORDS: Log-normal; Ratio of means; r^*-Formula; Signed log-likelihood ratio statistic; Z-score test.

1. Introduction

The statistical analysis regarding the ratio of means of two independent log-normal distributions is often of interest in biomedical research. For example, in a bioavailability study, the relative potency of a new drug to that of a standard one is expressed in terms of the ratio of means, and analysts often need to construct a confidence interval for this ratio or to test the null hypothesis that the ratio is one, i.e., the mean outcomes of the two products are the same (Berger and Hsu, 1996; Chow and Liu, 2000). In such a study, the primary endpoints of interest are estimates of the pharmacokinetic parameters specifying the bioavailability of a drug, such as the area under the plasma concentration-time curve (AUC) and the maximum plasma concentration ($C_{\max}$). One of the difficulties commonly encountered is that these data are positively skewed. Zhou, Gao, and Hui (1997) also presented a case where they compared the hospital charges of two groups of patients, and medical charge data are frequently skewed significantly toward higher cost patients. In these situations, a log transformation is often considered in order to normalize the distribution of the original data. A common approach is to perform testing procedures and to construct confidence intervals for the difference of means of the log-transformed outcome variables and to report the resulting p-values for the null hypothesis based on the original outcomes and the back-transformed confidence intervals for the ratio of means. As pointed out by Zhou et al. (1997), however, the null hypoth-

esis based on the log-transformed outcomes is not equivalent to the one based on the original outcomes when the variances of the log-transformed outcome variables are unequal. In view of this, Zhou et al. (1997) proposed two methods for correctly testing the null hypothesis. One is a Z-score test and the other is a nonparametric bootstrap approach. Their simulation results show that the Z-score test is the best among all five tests considered in their paper. However, as our simulation results show in Section 3, the Z-score test does not perform well in a range of small-sample settings. In this article, we propose two likelihood-based methods—the signed log-likelihood ratio test and the modified signed log-likelihood ratio test (Barndorff-Nielsen and Cox, 1994). We demonstrate the high performance of the proposed methods in small-sample settings.

This article is organized as follows. In Section 2, we introduce the likelihood-based inference methods, i.e., the signed log-likelihood ratio statistic and the modified signed log-likelihood ratio statistic, which we will employ to test the null hypothesis of equal means and to construct confidence intervals for the ratio of means of two log-normal distributions. In Section 3, we report some simulation results to demonstrate the accuracy of the proposed methods. Two examples are examined in Section 4, one of which is a small-sample bioavailability study and the other is a medical charge study. Some final remarks are recorded in Section 5, and the derivation

of the modified signed log-likelihood ratio statistic is given in Appendix.

2. Likelihood-Based Methods

Let x_i be the outcome variable of the ith subject in the first group $(i = 1, \ldots, n)$ and y_j be the outcome variable of the jth subject in the second group $(j = 1, \ldots, m)$. Their corresponding means are α_1 and α_2, respectively. Assume that the logarithms of x_i and y_j are independently and normally distributed with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, i.e.,

$$\log x_i \sim \mathrm{N}\left(\mu_1, \sigma_1^2\right), \qquad \log y_j \sim \mathrm{N}\left(\mu_2, \sigma_2^2\right).$$

The null hypothesis of interest is

$$\mathrm{H}_0 \colon \alpha_1 = \alpha_2,$$

where $\log \alpha_k = \mu_k + \sigma_k^2/2, k = 1, 2$. Note that testing the null hypothesis H_0 is not equivalent to testing the null hypothesis $\mathrm{H}_0^* \colon \mu_1 = \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$. Similarly, a back-transformed confidence interval as discussed in Section 1 is actually for $e^{\mu_1 - \mu_2}$, not for the ratio of means α_1/α_2 when $\sigma_1^2 \neq \sigma_2^2$.

A Z-score test proposed by Zhou et al. (1997) for testing H_0 is

$$Z = \frac{\hat{\mu}_1 - \hat{\mu}_2 + \frac{1}{2}\left(S_1^2 - S_2^2\right)}{\sqrt{\dfrac{S_1^2}{n} + \dfrac{S_2^2}{m} + \dfrac{1}{2}\left(\dfrac{S_1^4}{n-1} + \dfrac{S_2^4}{m-1}\right)}},$$

where $\hat{\mu}_1 = \Sigma \log x_i/n$ and $\hat{\mu}_2 = \Sigma \log y_j/m$ are the maximum likelihood estimators of μ_1 and μ_2, respectively, and $S_1^2 = \Sigma \left(\log x_i - \hat{\mu}_1\right)^2/(n-1)$ and $S_2^2 = \Sigma \left(\log y_j - \hat{\mu}_2\right)^2/(m-1)$ are the unbiased estimators of σ_1^2 and σ_2^2, respectively. According to the results in Zhou et al. (1997), when n and m are both large and when H_0 is true, the statistic Z is approximately distributed as the standard normal distribution. Our simulation results in Section 3, however, show that the distribution of the Z-statistic is skewed when the sample size is small. Hence, the Z-score test is not suitable for small-sample data.

We propose two likelihood-based methods for small-sample inference purposes. Suppose the joint log-likelihood function based on sample data is $\ell(\theta) = \ell(\psi, \lambda)$, where $\theta = (\psi, \lambda)$, ψ is a parameter of interest and λ is a vector nuisance parameter. One can make inferences about ψ based on the signed log-likelihood ratio statistic

$$r \equiv r(\psi) = \mathrm{sgn}(\hat{\psi} - \psi)\left\{2[\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_\psi)]\right\}^{1/2},$$

where $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$ denotes the maximum likelihood estimator of $\theta = (\psi, \lambda)$ and $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$ denotes the constrained maximum likelihood estimator of θ for a fixed ψ. It is well known that r is approximately distributed as a standard normal to the first order (Cox and Hinkley, 1974). For testing the null hypothesis $\mathrm{H}_0 \colon \psi = \psi_0$, a two-sided p-value can be obtained from $r = r(\psi)$ by

$$p\text{-value} = 2P(r > |r_0|) \approx 2\{1 - \Phi(|r_0|)\},$$

where $\Phi(\cdot)$ is the standard normal distribution function and $r_0 = r(\psi_0)$ is the observed value of r under H_0. Furthermore,

the approximate $100(1 - \alpha)\%$ confidence interval for ψ can be obtained from

$$\{\psi; \ |r(\psi)| \leq z_{\alpha/2}\},$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of the standard normal distribution.

Note that the first-order approximation is not very accurate, especially when the sample size is small (see Pierce and Peters, 1992). There exist various ways to improve the accuracy of this approximation by adjusting the signed log-likelihood ratio statistic. In this article, we consider the modified signed log-likelihood ratio statistic, known as the r^*-formula, introduced by Barndorff-Nielsen (1986, 1991), which has the form

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

where $u \equiv u(\psi)$ is a statistic given in the Appendix. Barndorff-Nielsen (1986, 1991) showed that r^* is approximately distributed as a standard normal to the third order. Therefore, the p-value and confidence limits based on r^* are highly accurate. The two-sided p-value for testing the null hypothesis H_0 is

$$p\text{-value} = 2P(r^* > |r_0^*|) \approx 2\{1 - \Phi(|r_0^*|)\},$$

where $r_0^* = r^*(\psi_0)$ is the observed value of r^* under H_0. Moreover, the $100(1 - \alpha)\%$ confidence interval for ψ can be obtained from the r^*-formula as

$$\{\psi; \ |r^*(\psi)| \leq z_{\alpha/2}\}.$$

Consider two independent log-normal distributions and let the difference of two log means be the parameter of interest. In other words, $\psi = \log \alpha_1 - \log \alpha_2 = \mu_1 - \mu_2 + (1/2)(\sigma_1^2 - \sigma_2^2)$ is the parameter of interest and $\lambda = (\mu_2, \sigma_1, \sigma_2)$ is a vector nuisance parameter. Let (ψ_L, ψ_U) be a $100(1 - \alpha)\%$ confidence interval for ψ. Then (e^{ψ_L}, e^{ψ_U}) is the corresponding $100(1 - \alpha)\%$ confidence interval for the ratio of means of the two independent log-normal distributions, α_1/α_2.

For this problem, the log-likelihood function $\ell(\theta) = \ell(\psi, \lambda)$ is

$$\begin{aligned}
\ell(\psi, \lambda) = {}&-n\log\sigma_1 - m\log\sigma_2 \\
&+ \left\{\psi + \mu - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)\right\}\frac{1}{\sigma_1^2}t_1 \\
&+ \frac{\mu}{\sigma_2^2}t_2 - \frac{1}{2\sigma_1^2}t_3 - \frac{1}{2\sigma_2^2}t_4 \\
&- \left\{\psi + \mu - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)\right\}^2\frac{n}{2\sigma_1^2} - \frac{m\mu^2}{2\sigma_2^2}, \quad (2)
\end{aligned}$$

where $\mu = \mu_2$ and $t = (t_1, t_2, t_3, t_4) = (\Sigma \log x_i, \Sigma \log y_j, \Sigma (\log x_i)^2, \Sigma (\log y_j)^2)$ is a minimum sufficient statistic. It can be shown that the maximum likelihood estimator $\hat{\theta} = (\hat{\psi}, \hat{\lambda}) = (\hat{\psi}, \hat{\mu}, \hat{\sigma}_1, \hat{\sigma}_2)$ is

$$\hat{\mu} = \bar{t}_2,$$

$$\hat{\sigma}_1^2 = \frac{1}{n}\left(t_3 - n\bar{t}_1^2\right),$$

$$\hat{\sigma}_2^2 = \frac{1}{m}\left(t_4 - m\bar{t}_2^2\right),$$

$$\hat{\psi} = \bar{t}_1 - \hat{\mu} + \frac{1}{2}\left(\hat{\sigma}_1^2 - \hat{\sigma}_2^2\right),$$

Table 1

Parameter configurations for the simulation study

Design	μ_1	σ_1^2	μ_2	σ_2^2	ψ
1	1.1	0.4	1.2	0.2	0
2	2.5	1.5	3.0	0.5	0
3	1.1	2.4	2.2	0.2	0
4	2.5	3.5	4.0	0.5	0
5	$1.0 + 1.0/n^{1/2}$	0.6	1.0	0.4	$0.1 + 1.0/n^{1/2}$
6	$1.0 + 1.5/n^{1/2}$	0.6	1.0	0.4	$0.1 + 1.5/n^{1/2}$
7	$1.0 + 2.5/n^{1/2}$	0.6	1.0	0.4	$0.1 + 2.5/n^{1/2}$

where $\bar{t}_1 = t_1/n$ and $\bar{t}_2 = t_2/m$. Furthermore, for a fixed value of ψ, the constrained maximum likelihood estimator $\hat{\lambda}_\psi = (\hat{\mu}_\psi, \hat{\sigma}_{1\psi}, \hat{\sigma}_{2\psi})$ is defined by the following recursive equations:

$$\hat{\mu}_\psi \left(\frac{n}{\hat{\sigma}_{1\psi}^2} + \frac{m}{\hat{\sigma}_{2\psi}^2} \right) = \frac{t_1}{\hat{\sigma}_{1\psi}^2} + \frac{t_2}{\hat{\sigma}_{2\psi}^2}$$
$$- \frac{n}{\hat{\sigma}_{1\psi}^2} \left\{ \psi - \frac{1}{2} \left(\hat{\sigma}_{1\psi}^2 - \hat{\sigma}_{2\psi}^2 \right) \right\},$$

$$\frac{1}{\hat{\sigma}_{1\psi}^2} \left(t_3 - 2t_1\hat{\xi}_\psi + n\hat{\xi}_\psi^2 \right) = n + t_1 - n\hat{\xi}_\psi,$$

$$\frac{1}{\hat{\sigma}_{2\psi}^2} \left(t_4 - 2t_2\hat{\mu}_\psi + m\hat{\mu}_\psi^2 \right) = m - t_1 \frac{\hat{\sigma}_{2\psi}^2}{\hat{\sigma}_{1\psi}^2} + n\hat{\xi}_\psi \frac{\hat{\sigma}_{2\psi}^2}{\hat{\sigma}_{1\psi}^2},$$

where $\hat{\xi}_\psi = \psi + \hat{\mu}_\psi - \frac{1}{2}(\hat{\sigma}_{1\psi}^2 - \hat{\sigma}_{2\psi}^2)$. Hence, the Gauss–Seidel iteration can be used to find the constrained maximum likelihood estimator $\hat{\lambda}_\psi$. It can also be calculated by using the S-plus function *nlmin* (MathSoft, 1998). The signed log-likelihood ratio statistic r can be simplified to be

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi) \left\{ n \log \frac{\hat{\sigma}_{1\psi}^2}{\hat{\sigma}_1^2} + m \log \frac{\hat{\sigma}_{2\psi}^2}{\hat{\sigma}_2^2} \right.$$

$$+ (t_1 - n\hat{\xi}_\psi) \left(1 - \frac{\hat{\sigma}_{2\psi}^2}{\hat{\sigma}_{1\psi}^2} \right) \right\}^{1/2}. \quad (3)$$

The details for the derivation of r^* are given in the Appendix.

3. Simulation Studies

Table 1 gives the parameter configurations of seven designs that we will consider in this section. Our aim is to assess the coverage probabilities of the Z-, r-, and r^*-intervals. Moreover, the type I error rates and powers of these three tests are also examined.

Our first simulation study is to compare the coverage probabilities of the three methods using designs 1 and 2 of Table 1. Note that these two designs have the same parameter configurations as those considered in Zhou et al. (1997). They represent increasing differences in skewness between two independent log-normal distributions. We also use two different settings in small sample sizes: (i) $(n, m) = (5, 10)$ and (ii) $(n, m) = (10, 10)$. For each of the four possible combinations of sample size (n, m) and design parameters, we have generated 20,000 samples from two independent log-normal distributions and have calculated the two-sided 90% confidence intervals for ψ with equal tail probabilities. The simulated empirical coverage probabilities, upper and lower error probabilities, and average lengths for each method are given in Table 2. The criteria for comparison are (a) coverage probability or coverage error of a confidence interval, which is defined as the absolute difference between the coverage probability and nominal value, and (b) upper and lower error probabilities and their symmetry, where the upper error probability is the percentage of the intervals falling below the true parameter and the lower error probability is the percentage of the intervals falling above the true parameter.

From Table 2, we observe that the coverage probabilities are lower than the nominal value, or equivalently, the coverage errors are relatively large for both the Z- and r-intervals. Furthermore, for these two intervals, the biases of the upper

Table 2

Coverage probabilities, coverage errors, error probabilities and average lengths of two-sided 90% confidence intervals for the three methods

Sample size	Design	Method	Coverage probability	Coverage error	Upper error probability	Lower error probability	Average length
(5, 10)	1	Z	0.859	0.041	0.096	0.045	1.129
		r	0.851	0.049	0.087	0.063	1.154
		r^*	0.895	0.005	0.052	0.053	1.578
	2	Z	0.855	0.045	0.129	0.016	2.643
		r	0.847	0.053	0.105	0.048	2.788
		r^*	0.898	0.002	0.053	0.049	4.505
(10, 10)	1	Z	0.886	0.014	0.067	0.047	0.873
		r	0.878	0.022	0.064	0.058	0.878
		r^*	0.900	0.000	0.049	0.051	0.979
	2	Z	0.889	0.011	0.088	0.023	1.924
		r	0.876	0.024	0.074	0.050	1.984
		r^*	0.901	0.001	0.049	0.051	2.365

Table 3
Type I error rates of the three tests

Sample size	Design	Test statistic	Nominal level			
			0.01	0.025	0.05	0.1
(5, 10)	1	Z	0.043	0.068	0.098	0.149
		r	0.025	0.053	0.089	0.153
		r^*	0.011	0.027	0.055	0.105
	2	Z	0.069	0.098	0.131	0.187
		r	0.034	0.065	0.107	0.180
		r^*	0.012	0.029	0.055	0.107
(10, 10)	1	Z	0.023	0.041	0.069	0.120
		r	0.017	0.037	0.066	0.123
		r^*	0.010	0.026	0.051	0.101
	2	Z	0.036	0.060	0.088	0.143
		r	0.020	0.044	0.077	0.139
		r^*	0.011	0.027	0.053	0.103

error probabilities are large and they are extremely asymmetric in small sample-size cases ($n = 5$, $m = 10$). Also, as the difference in skewness between the two distributions increases, the asymmetry becomes serious.

In contrast, the r^*-interval gives nearly the exact coverage probabilities, or equivalently, nearly zero coverage errors. Moreover, its upper and lower error probabilities are close to the nominal ones and they are much more symmetric than those of the Z- and r-intervals.

Overall, based on the comparison criteria examined in this article, the r^*-interval performs much better than the Z- and r-intervals. Note that the r^*-interval has the longest average length whereas the Z-interval has the shortest average length. However, both the Z- and r-intervals are too short to ensure their coverage probabilities close to the nominal value.

Some further simulations have been conducted to explore how the coverage probabilities change as a function of ψ and how the nuisance parameters affect the coverage probabilities as were recorded in Wu et al. (2001). The simulation results showed that, if we fix the nuisance parameters and only change the parameter of interest, ψ, the coverage probabilities remain relatively stable, i.e., the simulation results were consistent with those given in Table 2. To explore how the nuisance parameters affect the coverage probabilities, we performed simulations for designs 3 and 4 from Table 1. Designs 3 and 4 are obtained by increasing the variance σ_1^2 of designs 1 and 2 by two and keeping the parameter of interest, ψ, unchanged. Furthermore, we observed from the simulation results that the coverage probabilities of Z have decreased. However, the coverage probabilities of r and r^* remain relatively constant and r^* gives almost the exact coverage probabilities even if we increase the difference of skewness between the two distributions by a large amount.

We have also conducted some simulations for the comparison of the type I error rates and powers of the three tests. In estimating the type I error rates, we have used the same four possible combinations of sample size and design parameters (designs 1 and 2) as in Table 2. For each of the four possible

combinations, we have simulated 20,000 data sets. For each data set, we have calculated the values of the Z-, r-, and r^*-statistics under the null hypothesis H_0: $\psi = 0$. The proportion of those values that falls below the value -1.96 represents the empirical type I error rate corresponding to the nominal level 0.025 of the standard normal distribution. Table 3 records the empirical type I error rates of Z, r, and r^* corresponding to the nominal levels 0.01, 0.025, 0.05, and 0.1, respectively. It can be seen that, among all the tests compared, r^* has the empirical type I error rates that are the closest to the preset nominal levels and the Z-statistic has the worst performance among the three test statistics. One can also observe that the worst performance of Z and r occurs in the setting of a small sample size and large skewness. When the sample size increases and/or skewness decreases, the performance of both Z and r improves.

In simulating powers of the tests, we have considered designs 5–7 from Table 1, where H_0: $\psi = 0$ is not true. The proportion of rejecting H_0 based on a test represents the empirical power of this test. For a prespecified significance level $\alpha = 0.05$ and sample sizes $n = m$ from 10 to 25, Table 4 records simulation results based on 20,000 replications for the empirical powers and empirical type I error rates of the three tests for designs 5–7 in Table 1. The parameter values used to obtain these empirical type I error rates are also from designs 5–7 but by setting ψ to be the true value, i.e., $\mu_1 - \mu_2 + \frac{1}{2}(\sigma_1^2 - \sigma_2^2)$. The results in Table 4 show that r has the largest power among the three tests and the power of r^* is the smallest among the three tests. However, from the empirical type I error rates, we can see that both r and Z are too liberal and, in contrast, that the r^*-test nearly achieves the desired nominal level $\alpha = 0.05$. Thus, to be safe for maintaining the type I error rate, we suggest using r^*.

4. Two Real-Life Examples

In this section, we will illustrate our method using two real-life examples. The first example is a bioavailability study in which a randomized, parallel-group experiment was conducted with

Table 4
Powers of the three tests with $\alpha = 0.05$

Sample size	Design	Z	r	r^*
10	5	0.217 (0.066)[a]	0.237 (0.065)	0.200 (0.051)
	6	0.367 (0.066)	0.385 (0.065)	0.329 (0.051)
	7	0.705 (0.066)	0.697 (0.065)	0.623 (0.051)
15	5	0.237 (0.063)	0.258 (0.062)	0.235 (0.051)
	6	0.390 (0.063)	0.407 (0.062)	0.376 (0.051)
	7	0.728 (0.063)	0.728 (0.062)	0.688 (0.051)
25	5	0.266 (0.059)	0.281 (0.058)	0.269 (0.051)
	6	0.431 (0.059)	0.444 (0.058)	0.426 (0.051)
	7	0.766 (0.059)	0.769 (0.058)	0.751 (0.051)

[a] Empirical power, with empirical type I error rate in parentheses.

20 subjects to compare a new test formulation (x) with a reference formulation (y) of a drug product with a long half-life. Among other statistical analyses, testing the equality of the means of $C_{\max}$ and constructing a confidence interval for the ratio of means of $C_{\max}$ of the two formulations are of great importance in determining if the two formulations have different bioavailability. The $C_{\max}$ data from this study are presented in Table 5 ($n = m = 10$). The sample means are $\bar{x} = 668.20$ and $\bar{y} = 997.56$, and the sample standard deviations are $S_x = 314.86$ and $S_y = 913.43$. After the log transformation, the sample means are $\hat{\mu}_1 = 6.417$ and $\hat{\mu}_2 = 6.601$ and the sample standard deviations are $S_1 = 0.429$ and $S_2 = 0.817$. The QQ-plots for the original data and log-transformed data are given in Figure 1. These plots suggest that the distributions of $C_{\max}$ data are highly 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 0.595 for the test formulation group and a p-value of 0.983 for the reference formulation group, while the same tests on the original data give a p-value of 0.099 for the test formulation group and a p-value of 0.005 for the reference formulation group. Therefore, the log transformation normalizes the data. The F-test for equal variances of the log-transformed data between the two groups gives a p-value of 0.034, and therefore the log transformation does not stabilize the variances. In testing equal means of $C_{\max}$ between the two formulations, we compute the two-sided p-value of r^* to be 0.173 and the 95% confidence interval for the ratio of means of $C_{\max}$ to be (0.242, 1.200). For comparison purposes, we also computed the two-sided p-values of r and the Z-score test to be 0.167 and 0.203, respectively, and the corresponding 95% confidence intervals to be (0.295, 1.181) and (0.339, 1.259), respectively.

The second example is a medical charge study of patient data from the Regenstrief Medical System (McDonald et al., 1988; Zhou et al., 1997) on effects of race on medical charges of patients with type I diabetes who had received inpatient

or outpatient care on at least two occasions during the period from January 1, 1993, through June 30, 1994. One of the questions of interest is whether the average medical charge for African American patients is the same as that for white patients. The data set consists of 119 African American patients and 106 white patients. The distributions of medical charges are skewed significantly toward higher cost patients for both groups. The QQ plots for the original data and log-transformed data are given in Zhou et al. (1997). As they reported, the Shapiro–Wilk tests for normality on the log-transformed data give a p-value of 0.15 for the African American patient group and a p-value of 0.15 for the white patient group. Thus, the log transformation normalizes the data. Since the F-test for equal variances of log-transformed data gives a p-value of 0.04, the log transformation does not stabilize the variances. After the log transformation, the sample means and sample standard deviations are 9.067 and 1.351, respectively, for the African American group and 8.693 and 1.641, respectively, for the white group. For testing equal means of medical charges between the two groups, our r^*-test gives the two-sided p-value of 0.83, while the r- and Z-score tests give 0.85 and 0.84, respectively. Due to large sample sizes, all three tests give almost the same p-values.

5. Conclusions

In this article, we have proposed two small-sample likelihood-based methods for testing the hypothesis of equal means and for calculating confidence intervals for the ratio of means of two log-normal distributions. Simulation studies show that the proposed method based on the r^*-formula gives essentially exact coverage probabilities and is almost an exact test even for small samples. The calculations involved are simple because the maximum likelihood estimator has an explicit form and the constrained maximum likelihood estimator can easily be obtained by the S-plus function $nlmin$. The method can be applied to randomized, parallel-group bioavailability studies and other settings of skewed log-normal observations.

Table 5
$C_{\max}$ data of a parallel-group experiment

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
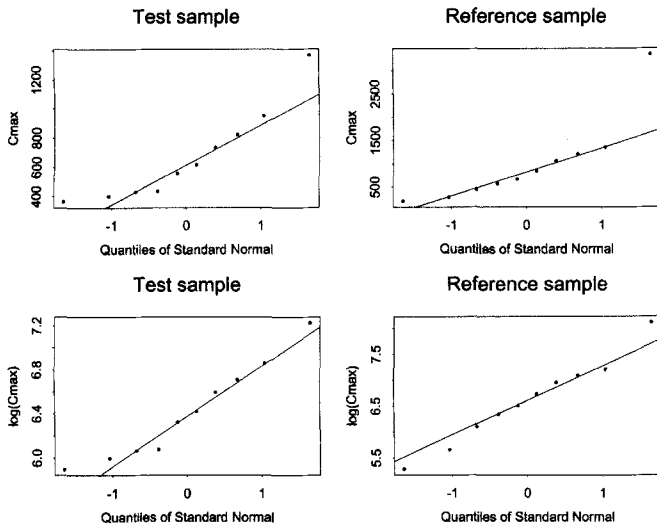
Figure 1. Quantile plots of C_{max} and $\log(C_{max})$ data.

RÉSUMÉ

Les méthodes existantes de comparaison des moyennes de deux distributions log-normales asymétriques indépendantes n'ont pas de bonnes performances dans certains contextes d'échantillons de petites tailles, comme par exemple dans des études de bio-disponibilité sur petits échantillons. Dans cet article nous proposons deux approches, basées sur la vraisemblance—le rapport des log-vraisemblances signé et le rapport modifié des log-vraisemblances signé—pour effectuer une inférence sur le rapport des moyennes de deux distributions log-normales indépendantes. Nous nous intéressons particulièrement à l'obtention des p-valeurs pour le test d'égalité des moyennes, et également à la construction d'intervalles de confiance pour le rapport des moyennes.

REFERENCES

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

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

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics.* London: Chapman and Hall.

Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests, and equivalence confidence sets. *Statistical Science* **11,** 283–315.

Chow, S. C. and Liu, J. P. (2000). *Design and Analysis of Bioavailability and Bioequivalence Studies,* 2nd edition. New York: Marcel Dekker.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

MathSoft. (1998). *S-plus 5 for Unix Guide to Statistics.* Seattle, Washington: MathSoft.

McDonald, C. J., Blevins, L., Tierney, W. M., and Martin, D. K. (1988). The Regenstrief medical records. *MD Computing* **5,** 34–47.

Pierce, D. A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *Journal of the Royal Statistical Society, Series B* **54,** 701–737.

Wu, J., Jiang, G., Wong, A., and Sun, X. (2001). *Likelihood-based inference for the ratio of means of two independent log-normal distributions.* Technical Report, York University, North York, Ontario, Canada.

Zhou, X. H., Gao, S. J., and Hui, S. L. (1997). Methods for comparing the means of two independent log-normal samples. *Biometrics* **53,** 1129–1135.

APPENDIX

Derivation of the r^*-Formula for Two Independent Log-Normal Distributions

Suppose that the log likelihood $\ell(\psi, \lambda; x, y)$ based on the sample (x, y) can be rewritten as $\ell(\psi, \lambda; \hat{\theta})$ with a fixed ancillary statistic; then the statistic u required for computing r^* is given by

$$u(\psi) = \frac{\left| \ell_{;\hat{\theta}}\left(\hat{\psi}, \hat{\lambda}\right) - \ell_{;\hat{\theta}}\left(\psi, \hat{\lambda}_\psi\right) \; \ell_{\lambda;\hat{\theta}}\left(\psi, \hat{\lambda}_\psi\right) \right|}{\left\{ \left| j_{\theta\theta}\left(\hat{\psi}, \hat{\lambda}\right) \right| \left| j_{\lambda\lambda}\left(\psi, \hat{\lambda}_\psi\right) \right| \right\}^{1/2}}, \qquad (4)$$

where the sample-space derivatives are defined as

$$\ell_{;\hat{\theta}}(\psi, \lambda) = \frac{\partial}{\partial \hat{\theta}} \ell\left(\psi, \lambda; \hat{\theta}\right),$$

the mixed derivatives as

$$\ell_{\lambda;\hat{\theta}}(\psi, \lambda) = \frac{\partial}{\partial \lambda} \ell_{;\hat{\theta}}(\psi, \lambda),$$

and $j_{\theta\theta}(\hat{\psi}, \hat{\lambda})$ is the observed information matrix and $j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)$ is the observed nuisance information matrix (Barndorff-Nielsen, 1991).

Since the two-sample log-normal model is a full-rank exponential model, the log-likelihood function based on the sample data (x, y) is only related to a minimum sufficient statistic $t = (t_1, t_2, t_3, t_4)$ and is given in (2). There is a one-to-one transformation between the maximum likelihood estimator $\hat{\theta} = (\hat{\psi}, \hat{\mu}, \hat{\sigma}_1, \hat{\sigma}_2)$ and the minimum sufficient statistic t, and the transformation Jacobian matrix is $\partial\hat{\theta}/\partial t$. Hence, the sample-space derivatives with respect to $\hat{\theta}$ in the formula for u can be derived based on the sample-space derivatives with respect to t. By using the identity $j_{\theta\theta}(\hat{\theta}) = \ell_{\theta;\hat{\theta}}(\hat{\theta})$ (Barndorff-Nielsen and Cox, 1994) and by canceling the determinant of the transformation Jacobian matrix, one can show that u reduces to the following form:

$$u(\psi) = \frac{\left| \ell_{;t}\left(\hat{\psi}, \hat{\lambda}\right) - \ell_{;t}\left(\psi, \hat{\lambda}_\psi\right) \; \ell_{\lambda;t}\left(\psi, \hat{\lambda}_\psi\right) \right|}{\left| \ell_{\theta;t}\left(\hat{\psi}, \hat{\lambda}\right) \right|}$$

$$\times \left\{ \frac{\left| j_{\theta\theta}\left(\hat{\psi}, \hat{\lambda}\right) \right|}{\left| j_{\lambda\lambda}\left(\psi, \hat{\lambda}_\psi\right) \right|} \right\}^{1/2},$$

where the sample-space derivatives $\ell_{;t}(\theta) = \partial\ell(\theta;t)/\partial t$ and mixed derivatives $\ell_{\lambda;t}(\theta) = \partial^2\ell(\theta;t)/\partial\lambda\partial t$ are given by

$$\ell_{;t}(\theta) = \left(\frac{1}{\sigma_1^2}\left\{\psi + \mu - \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right)\right\}, \frac{\mu}{\sigma_2^2}, -\frac{1}{2\sigma_1^2}, -\frac{1}{2\sigma_2^2}\right)'$$

and

$$\ell_{\lambda;t}(\theta) = \begin{pmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{2}{\sigma_1^3}(\psi + \mu + \frac{1}{2}\sigma_2^2) & \dfrac{\sigma_2}{\sigma_1^2} \\ \dfrac{1}{\sigma_2^2} & 0 & -\dfrac{2\mu}{\sigma_2^3} \\ 0 & \dfrac{1}{\sigma_1^3} & 0 \\ 0 & 0 & \dfrac{1}{\sigma_2^3} \end{pmatrix},$$

respectively. The determinants of the observed information matrix and mixed derivative matrix are given by $|j_{\theta\theta}(\hat\psi, \hat\lambda)| = 4n_1^2n_2^2/\hat\sigma_1^4\hat\sigma_2^4$ and $|\ell_{\theta;t}(\hat\theta)| = 1/\hat\sigma_1^5\hat\sigma_2^5$, respectively. The observed nuisance information matrix is

$$j_{\lambda\lambda}\left(\psi, \hat\lambda_\psi\right) = \begin{pmatrix} \dfrac{n}{\hat\sigma_{1\psi}^2} + \dfrac{m}{\hat\sigma_{2\psi}^2} & -\dfrac{n}{\hat\sigma_{1\psi}} - \dfrac{2n}{\hat\sigma_{1\psi}^3}(\hat\xi_\psi - \bar t_1) & \dfrac{n\hat\sigma_{2\psi}}{\hat\sigma_{1\psi}^2} - \dfrac{2m}{\hat\sigma_{2\psi}^3}(\hat\mu_\psi - \bar t_2) \\ -\dfrac{n}{\hat\sigma_{1\psi}} - \dfrac{2n}{\hat\sigma_{1\psi}^3}(\hat\xi_\psi - \bar t_1) & n + \dfrac{2n}{\hat\sigma_{1\psi}^2} & -\dfrac{n\hat\sigma_{2\psi}}{\hat\sigma_{1\psi}} - \dfrac{2n\hat\sigma_{2\psi}}{\hat\sigma_{1\psi}^3}(\hat\xi_\psi - \bar t_1) \\ \dfrac{n\hat\sigma_{2\psi}}{\hat\sigma_{1\psi}^2} - \dfrac{2m}{\hat\sigma_{2\psi}^3}(\hat\mu_\psi - \bar t_2) & -\dfrac{n\hat\sigma_{2\psi}}{\hat\sigma_{1\psi}} - \dfrac{2n\hat\sigma_{2\psi}}{\hat\sigma_{1\psi}^3}(\hat\xi_\psi - \bar t_1) & \dfrac{2m}{\hat\sigma_{2\psi}^2} + \dfrac{n\hat\sigma_{2\psi}^2}{\hat\sigma_{1\psi}^2} + \dfrac{4n}{\hat\sigma_{1\psi}^2}(\hat\xi_\psi - \bar t_1) \end{pmatrix}.$$

Therefore, the statistic $u(\psi)$ can be calculated from (4). Thus, with the log-likelihood function given in (2) and $r(\psi)$ and $u(\psi)$ defined in (3) and (4), respectively, $r^*(\psi)$ can be obtained from (1).