UW Biostatistics Working Paper Series

12-7-2005

Interval Estimation for the Ratio and Difference of Two Lognormal Means

Yea-Hung Chen
University of Washington, yeahung@u.washington.edu

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

1 Introduction

Health research often gives rise to data that are positive and highly skewed. Moreover, in many naturally occurring situations, the data follow lognormal distributions. The assumption can and should be checked: we recommend the use of quantile plots and the Shapiro-Wilk test for normality. These should be applied to the log-transformed data, as the natural logarithm of lognormal data will follow a normal distribution.

Note that while the median is commonly regarded as a desirable summary for skewed data, it is not always a quantity of scientific interest. Hospital administrators, for instance, may be interested in characterizing mean health care costs in various patient subgroups. Note that the mean is intimately related to the total, which is a measure that can be used in this example to describe the cumulative "burden" patients place on the health care system. In general: if the scientific question involves inference on averages, totals, or rates, it would be sensible to conduct inference on population means rather than medians.

Numerous methods are available for estimating a single lognormal mean. These have been discussed and compared in some detail: see, for example, Reference [1], [2] and [3]. Unfortunately, methods for the two sample situation are not as well understood. While approaches for these settings are available, detailed comparisons are lacking. Information regarding methods for the difference of means is particularly difficult to find. Consequently, it is unclear which (if any) of the methods are most appropriate. Nor is it clear how the performance of these approaches might vary – potentially important considerations include the sample sizes, the population means, and the population variances.

In this paper we explore methods for estimating the ratio or difference of two lognormal means. Our focus is on confidence intervals, though the methods we discuss here could also be used to conduct hypothesis tests. The methods are: a traditional maximum likelihood approach, a bootstrap approach, two methods based on the log-likelihood ratio statistic, and a generalized pivotal approach. We have performed extensive simulation studies for these approaches, and provide the results here.

The approaches are summarized in Sections 2 and 3; simulation results are discussed in Sections 5 and 6. We close the paper with an illustrative example (Section 8).

2 Interval estimates for the ratio of means

Suppose we have two populations of interest. Let $W_{i1}, \ldots, W_{in_i}$ denote a random sample from population i (for $i = 1, 2$), and let $Y_{ij} = \log W_{ij}$. Assume that Y_{1j} and Y_{2j} are independently and normally distributed with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. Equivalently, assume W_{ij} has a lognormal distribution with mean $m_i = \exp(\mu_i + (1/2)\sigma_i^2)$. The ratio of the means, m_1/m_2, is:

$$m_1/m_2 = \exp(\mu_1 + \frac{1}{2}\sigma_1^2 - \mu_2 - \frac{1}{2}\sigma_2^2).$$

The natural logarithm of the ratio, which we will denote ψ, is:

$$\psi \equiv \log(m_1/m_2) = \mu_1 + \frac{1}{2} - \mu_2 - \frac{1}{2}\sigma_2^2. \tag{1}$$

Confidence intervals for m_1/m_2 may be obtained via various methods. We discuss five such approaches below. Note that for mathematical simplicity, we will focus on obtaining interval estimates for ψ. Any one of the confidence intervals for ψ can be exponentiated to obtain a confidence interval for m_1/m_2.

2.1 The maximum likelihood approach

Let $y_{i1}, \ldots, y_{in_i}$ denote observed values of the random variables $Y_{i1}, \ldots, Y_{in_i}$. The maximum likelihood (ML) estimates for μ_1, μ_2, σ_1^2 and σ_2^2 are:

$$\hat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \tag{2}$$

$$\hat{\sigma}_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i)^2, \qquad i = 1, 2, \tag{3}$$

and by the invariance property of ML estimation, the maximum likelihood estimate for ψ is:

$$\hat{\psi} = \hat{\mu}_1 - \hat{\mu}_2 + \frac{1}{2}(\hat{\sigma}_1^2 - \hat{\sigma}_2^2). \tag{4}$$

Let τ denote the standard error of $\hat{\psi}$, such that $\tau = \sqrt{\mathrm{Var}(\hat{\psi})}$. An estimate of τ is as follows:

$$\hat{\tau} = \left(\left(\frac{\partial\psi}{\partial\boldsymbol{\theta}}\right)' \hat{\boldsymbol{I}}^{-1}\frac{\partial\psi}{\partial\boldsymbol{\theta}}\right)^{1/2}, \tag{5}$$

where $\boldsymbol{I}$ denotes the information matrix and $\hat{\boldsymbol{I}}$ denotes its estimate:

$$\hat{\boldsymbol{I}} = \begin{pmatrix} n_1/\hat{\sigma}_1 & 0 & 0 & 0 \\ 0 & n_1/(2\hat{\sigma}_1^2) & 0 & 0 \\ 0 & 0 & n_2/\hat{\sigma}_2 & 0 \\ 0 & 0 & 0 & n_2/(2\hat{\sigma}_2^2) \end{pmatrix}. \tag{6}$$

The partial derivative of ψ with respect to $\boldsymbol{\theta}$ is:

$$\frac{\partial\psi}{\partial\boldsymbol{\theta}} = \begin{pmatrix} 1 & 1/2 & -1 & -1/2 \end{pmatrix}',$$

where $\boldsymbol{\theta}$ denotes the vector of parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$.

The distribution of $\hat{\psi}$ is asymptotically normal. Thus, a $100(1-\alpha)\%$ confidence interval for ψ can be given by:

$$[\hat{\psi} - z_{\alpha/2}\hat{\tau}, \hat{\psi} + z_{\alpha/2}\hat{\tau}] \tag{7}$$

where $z_{\alpha/2}$ denotes the $100(\alpha/2)$ percentile of the standard normal distribution.

Note that the maximum likelihood intervals assume that the parameter estimate has a distribution that is asymptotically normal; convergence to normality is likely to be poor in small sample settings.

2.2 A bootstrap approach

Zhou and Tu [4] have suggested a bootstrap approach for estimating ψ. It relies on the use of m bootstrap samples, where m is a "large" fixed number. The approach can be summarized by the following algorithm:

Compute, from the samples of interest, $\hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2, \hat{\psi}$ and $\hat{\tau}$ as defined in (2), (3), (4), and (5)

(For $j = 1$ to m)

Generate n_1 samples from $\mathrm{N}(\hat{\mu}_1, \hat{\sigma}_1^2)$ and n_2 samples from $\mathrm{N}(\hat{\mu}_2, \hat{\sigma}_2^2)$

Calculate, using the bootstrap sample, the estimates for ψ and τ, as defined in (4) and (5); denote these $\hat{\psi}_j$ and $\hat{\tau}_j$

Compute the test statistic $S_j = (\hat{\psi}_j - \hat{\psi})/\hat{\tau}_j$

(End loop)

Find the 100α and $100(1-\alpha)$ percentiles of $S_1, \ldots, S_m$; denote these $S_{(l)}$ and $S_{(u)}$, respectively.

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

$$[\hat{\psi} + S_{(l)}\hat{\tau}, \hat{\psi} + S_{(u)}\hat{\tau}]. \tag{8}$$

Note that unlike some bootstrap approaches, the above does in fact make parametric assumptions. Specifically, it assumes that the data are lognormally distributed. The method does not however explicitly state or assume the nature of the statistic's distribution; instead, the distribution is explored computationally via the use of the bootstrap samples.

2.3 The signed log-likelihood ratio approach

Methods based on the log-likelihood ratio statistic are also possible. Wu and colleagues [5] have proposed a signed log-likelihood method [6, 7] for ψ. The approach requires use of the log-likelihood function:

$$\ell(\boldsymbol{\theta}) = -n_1 \log \sqrt{2\pi} - n_2 \log \sqrt{2\pi} - n_1 \log \sigma_1 - n_2 \log \sigma_2 - \frac{1}{2\sigma_1^2}\sum_{j=1}^{n_1}(y_{1j} - \mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j} - \mu_2)^2. \quad (9)$$

It can be rewritten as a function of ψ:

$$\ell(\psi, \boldsymbol{\lambda}) = -n_1 \log \sqrt{2\pi} - n_2 \log \sqrt{2\pi} - n_1 \log \sigma_1 - n_2 \log \sigma_2 - \frac{1}{2\sigma_1^2}\sum_{j=1}^{n_1}(y_{1j} - (\psi - \frac{1}{2}\sigma_1^2 + \mu_2 + \frac{1}{2}\sigma_2^2))^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j} - \mu_2)^2, \quad (10)$$

where $\boldsymbol{\lambda}$ denotes the vector of nuisance parameters $(\mu_2, \sigma_1, \sigma_2)$.

The signed log-likelihood ratio statistic (SLLR), which we will denote r, is:

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi)(2\{\ell(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell(\psi, \hat{\boldsymbol{\lambda}}_\psi)\})^{1/2}, \quad (11)$$

where $\hat{\psi}$ and $\hat{\boldsymbol{\lambda}} = (\hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2)$ denote the maximum likelihood estimates. The expression $\hat{\boldsymbol{\lambda}}_\psi$ denotes "constrained" maximum likelihood estimates: they are the ML estimates of the nuisance parameters at a given value of ψ. These can be obtained computationally via, for instance, the `optim` function in the R programming environment [8].

The distribution of the r statistic approximates the standard normal to the first order (see for instance, Reference [7]). A $100(1-\alpha)\%$ confidence interval for ψ is thus given by the boundaries of the following region:

$$\{\psi; -z_{\alpha/2} \le r(\psi) \le z_{\alpha/2}\}. \quad (12)$$

2.4 A modified signed log-likelihood ratio approach

Barndorff-Nielssen [6, 7] has proposed a modified form of the r statistic which better approximates the standard normal. Wu and colleagues [5] have shown that the statistic can be used to create confidence intervals for ψ.

The modified r statistic depends on a quantity which we will denote u. Let $\boldsymbol{t}$ denote the vector of statistics $(\sum y_{1j}, \sum y_{2j}, \sum y_{1j}^2, \sum y_{2j}^2)$, and let $\boldsymbol{\omega}$ denote the reordered vector of parameters $(\psi, \boldsymbol{\lambda}) = (\psi, \mu_2, \sigma_1, \sigma_2)$. The statistic u is:

$$u(\psi) = \frac{|\, \ell_{;\boldsymbol{t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}}) - \ell_{;\boldsymbol{t}}(\psi, \hat{\boldsymbol{\lambda}}_\psi) \;\; \ell_{\boldsymbol{\lambda};\boldsymbol{t}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)\, |}{|\ell_{\boldsymbol{\omega};\boldsymbol{t}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})|} \times \left(\frac{|j_{\boldsymbol{\omega}\boldsymbol{\omega}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})|}{|j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)|}\right)^{1/2},$$

where $\ell_{;\boldsymbol{t}} = \partial\ell/\partial\boldsymbol{t}$, $\ell_{\boldsymbol{\lambda};\boldsymbol{t}} = \partial^2\ell/\partial\boldsymbol{\lambda}\partial\boldsymbol{t}$ and $\ell_{\boldsymbol{\omega};\boldsymbol{t}} = \partial^2\ell/\partial\boldsymbol{\omega}\partial\boldsymbol{t}$. The matrices $j_{\boldsymbol{\omega}\boldsymbol{\omega}}$ and $j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}$ can be found by:

$$\begin{aligned} j_{\boldsymbol{\omega}\boldsymbol{\omega}} &= -\frac{\partial^2 \ell}{\partial \boldsymbol{\omega}^2} \\ j_{\boldsymbol{\lambda}\boldsymbol{\lambda}} &= -\frac{\partial^2 \ell}{\partial \boldsymbol{\lambda}^2}. \end{aligned}$$

These quantities depend on ψ and $\boldsymbol{\lambda}$; $j_{\boldsymbol{\omega}\boldsymbol{\omega}}(\hat{\psi}, \hat{\boldsymbol{\lambda}})$ and $j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)$ denote the values of the matrices when evaluated at the appropriate estimates.

Note that $j_{\boldsymbol{\omega}\boldsymbol{\omega}}$ is conceptually analogous to $\boldsymbol{I}$ (defined in Section 2.1.1). The difference is that it uses the reordered vector of parameters $\boldsymbol{\omega}$. An estimate for the determinant of $j_{\boldsymbol{\omega}\boldsymbol{\omega}}$ is given by $4n_1^2 n_2^2/\hat{\sigma}_1^4\hat{\sigma}_2^4$. An expression for $j_{\boldsymbol{\lambda}\boldsymbol{\lambda}}$ (for solving the problem at hand) is provided by Wu and colleagues [5]; a theoretical discussion is provided by Barndorff-Nielssen [7].

The modified signed log-likelihood ratio statistic (MSLLR), which we will denote r^*, is:

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log\left(\frac{u(\psi)}{r(\psi)}\right). \tag{13}$$

The statistic has a distribution that approximates the standard normal to the third order [7]. A $100(1-\alpha)\%$ confidence interval for ψ is thus given by the boundaries of the following region:

$$\{\psi; -z_{\alpha/2} \leq r^*(\psi) \leq z_{\alpha/2}\}. \tag{14}$$

2.5 The generalized pivotal approach

Finally, we examine a fifth approach. This utilizes Weerahandi's notion of generalized confidence intervals and generalized pivotal quantities [9]. Weerahandi defines a generalized pivotal (GP) as a statistic that has a distribution free of unknown parameters and an observed value that does not depend on nuisance parameters.

Note that the generalized pivotal is allowed to be a function of nuisance parameters, whereas conventional pivotal quantities can only be a function of the sample and the parameter of interest. Weerahandi terms the confidence interval resulting from a GP a generalized confidence interval.

To find a $100\gamma\%$ generalized confidence interval, it is necessary to find a region C_γ of the pivotal space such that the probability that the pivotal quantity is in C_γ is equal to the confidence coefficient, γ. The generalized confidence interval is simply the region of the parameter space corresponding to C_γ.

Krishnamoorthy and Mathew [1] have proposed a generalized confidence interval approach for ψ. They use the following generalized pivotal quantity:

$$T_R = T_1 - T_2, \tag{15}$$

where

$$T_i = \hat{\mu}_i - \frac{Z_i}{U_i/\sqrt{n_i - 1}} \frac{\hat{\sigma}_i}{\sqrt{n_i}} + \frac{1}{2} \frac{\hat{\sigma}_i^2}{U_i^2/(n_i - 1)}, \qquad i = 1, 2, \tag{16}$$

and $Z_i \sim \mathrm{N}(0,1)$ and $U_i^2 \sim \chi^2_{n_i - 1}$.

Note that T_i can be rewritten as follows:

$$T_i = \hat{\mu}_i - \frac{\bar{Y}_i - \mu_i}{S_i/\sqrt{n_i}} \hat{\sigma}_i^2/\sqrt{n_i} + \frac{1}{2} \frac{\sigma_i^2}{S_i^2} \hat{\sigma}_i^2, \qquad i = 1, 2.$$

The above utilizes the substitutions $Z_i = \sqrt{n_i}(\bar{Y}_i - \mu_i)/\sigma_i$ and $U_i^2 = (n_i - 1)S_i^2/\sigma_i^2$, where $\bar{Y}_i$ and S_i^2 are defined as:

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

From the second expression of T_i it is apparent that the "observed value" of T_R, the value of T_R given the sample of interest, is ψ. Thus, a $100(1-\alpha)\%$ two-sided generalized confidence interval for ψ is simply the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of T_R.

In order to find the percentiles, it is necessary to understand the distribution of the statistic. Note that T_R depends only on the sample of interest and the normal and Chi-squared random variables. Consequently, it is possible to characterize the distribution of the pivotal computationally by randomly generating m other values of T_R, where m is some "large" number.

The approach can be summarized by the following algorithm:

(For $j = 1$ to m)
Generate values for Z_1, Z_2, U_1^2, and U_2^2
Calculate T_R
(End loop)
Order the m values of T_R; find the 100α and $100(1-\alpha)$ percentiles; denote these $T_{R(l)}$ and $T_{R(u)}$, respectively.

A $100(1-\alpha)\%$ confidence interval for ψ is simply:

$$[T_{R(l)}, T_{R(u)}]. \tag{17}$$

Note that while this approach assumes that the data are lognormally distributed, it does not explicitly state or assume the distribution of the pivotal quantity; rather, the distribution is approximated computationally.

3 Interval estimates for the difference of means

Ratios are relative comparisons. In some situations, the scientist may be more interested in absolute differences, such as the arithmetic difference of the means. For instance, given health care data for a cohort of senior citizens, we may be interested in determining whether the mean cost for elderly males exceeds the mean cost for females by more than \$5,000.

The difference of two lognormal means, which we will denote δ, is:

$$\delta \equiv m_1 - m_2 = \exp(\mu_1 + \frac{1}{2}\sigma_1^2) - \exp(\mu_2 + \frac{1}{2}\sigma_2^2). \tag{18}$$

We discuss below methods for constructing confidence intervals for δ.

3.1 The maximum likelihood approach

The maximum likelihood estimate for δ is:

$$\hat{\delta} = \exp(\hat{\mu}_1 + \frac{1}{2}\hat{\sigma}_1^2) - \exp(\hat{\mu}_2 + \frac{1}{2}\hat{\sigma}_2^2), \tag{19}$$

where $\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2$, and $\hat{\sigma}_2^2$ are as defined in (2) and (3).

Let υ denote the standard error of $\hat{\delta}$, such that $\upsilon = \sqrt{\mathrm{Var}(\hat{\delta})}$. An estimate for υ is:

$$\hat{\upsilon} = \left(h(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{I}}^{-1}h(\hat{\boldsymbol{\theta}})\right)^{1/2}, \tag{20}$$

where $\hat{\boldsymbol{\theta}}$ denotes the MLE of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{I}}$ is as defined in (6). The function h is defined as the partial derivative of δ with respect to $\boldsymbol{\theta}$:

$$h(\boldsymbol{\theta}) = \frac{\partial\delta}{\partial\boldsymbol{\theta}} = \begin{pmatrix} m_1 & \frac{1}{2}m_1 & -m_2 & -\frac{1}{2}m_2 \end{pmatrix}', \tag{21}$$

where $\boldsymbol{\theta}$ is the vector of parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$. Note that m_1 and m_2 depend on $\boldsymbol{\theta}$; $h(\hat{\boldsymbol{\theta}})$ denotes the value of h evaluated at $\hat{\boldsymbol{\theta}}$.

A $100(1-\alpha)\%$ confidence interval for δ can be given by:

$$[\hat{\delta} - z_{\alpha/2}\hat{\upsilon}, \hat{\delta} + z_{\alpha/2}\hat{\upsilon}]. \tag{22}$$

3.2 A bootstrap approach

A bootstrap approach for δ is also possible. It is entirely analagous to the approach described in Section 2.2, with $\hat{\delta}$ (19) used in place of $\hat{\psi}$, and $\hat{\upsilon}$ (20) used in place of $\hat{\tau}$.

3.3 The signed log-likelihood ratio approach

Wu and colleagues' SLLR approach [5] for estimating ψ can be modified to obtain confidence intervals for δ; the method is entirely analogous to that reviewed in Section 2.3. Note that the log-likelihood function can be rewritten as a function of δ:

$$\begin{aligned}\ell(\delta, \boldsymbol{\lambda}) \quad = \quad & -n_1 \log \sqrt{2\pi} - n_2 \log \sqrt{2\pi} - n_1 \log \sigma_1 - n_2 \log \sigma_2 \\ & - \frac{1}{2\sigma_1^2} \sum_{j=1}^{n_1} (y_{1j} - (\log\{\delta + \exp(\mu_2 + \frac{1}{2}\sigma_2^2)\} - \frac{1}{2}\sigma_1^2))^2 \\ & - \frac{1}{2\sigma_2^2} \sum_{j=1}^{n_2} (y_{2j} - \mu_2)^2, \end{aligned} \tag{23}$$

where $\boldsymbol{\lambda}$ is the vector of nuisance parameters $(\mu_2, \sigma_1, \sigma_2)$.

3.4 The modified signed log-likelihood approach

The MSLLR approach for ψ can be modified to obtain confidence intervals for δ. Unfortunately, the mathematical expressions for the quantities needed for this method are quite complex. We have evaluated the equations via the use of mathematical software, but found that the expressions involve a very large number of irreducible terms.

3.5 The generalized pivotal approach

Krishnamoorthy and Mathew [1] have proposed a generalized pivotal approach for δ. They use the following generalized pivotal:

$$T_D = \exp(T_1) - \exp(T_2), \tag{24}$$

where T_1 and T_2 are as defined in (16). The method is entirely identical to the approach described in Section 2.5, with T_D used in place of T_R.

4 Simulation methods

We conducted simulation studies for each of the methods described above (excluding the MSLLR approach for δ). For our simulations, we used 24 unique sets of parameter values. These include situations of varying n_i, μ_i, and σ_i. A complete list is provided in Table 1. The table also includes, for reference, the skewness coefficient (γ_i) of each of the lognormal distributions. Note that the skewness of a lognormal distribution depends only on σ_i^2 (the variance of the corresponding normal distribution).

For each simulation design, we randomly generated 10,000 sets of data. From each set of samples we then constructed a 95% confidence interval for the parameter of interest (the ratio or difference of means), using each of the methods above. For the bootstrap methods we used a bootstrap sample size of $m = 500$ and for the generalized pivotal computations we used $m = 10,000$ pivotal quantities.

All computer simulations were carried out in the R statistical programming environment [8]. Note that developing computational methods for the log-likelihood ratio methods is a non-trivial task. First (as discussed above), the methods require a maximization function in order to obtain the constrained ML estimates. We used the `optim` function in R. Secondly, some sort of algorithm is needed to find the appropriate bounds for ψ. For this task, we used the `uniroot` function in R. These approaches worked quite well.

Computation speed varied. The simulations for the likelihood ratio and bootstrap methods were particularly slow, requiring roughly two hours for each set. In comparison, our hardware required roughly ten minutes to produce 10,000 generalized confidence intervals and only one minute to produce 10,000 ML intervals.

5 Simulation results for the ratio of means

Results from the simulations for the ratio of means are presented in Table 2. Coverage, presented in the third column, is simply the percent of confidence intervals that included the true value. Coverage frequencies within half a percentage point of the nominal value (95%) are marked with parentheses.

The fifth column represents what one might call left error: the percent of time the confidence interval was to the left of m_1/m_2. The sixth column indicates right error: the percent of time the confidence interval was to the right of m_1/m_2. Note that left error plus right error is equal to the coverage error (100 minus the coverage).

Relative bias, presented in the last column, is simply a comparison of the error frequencies, and is defined as follows:

$$\text{relative bias} = \frac{(\text{right error}) - (\text{left error})}{(\text{right error}) + (\text{left error})}.$$

Note that this quantity is positive when the right error exceeds left error, and is negative when left error exceeds right error. The quantity takes on a value of 0 if left error equals right error; it is undefined when the coverage is equal to 100.

Our primary interest is the coverage frequencies. In this regard, the MSLLR and GP approaches are clear winners. Both resulted in excellent coverage frequencies, though the MSLLR approach appears to be more reliably accurate than the GP. In 12 simulations for the MSLLR method, there was only one setting in which the observed coverage was further than half a percentage point away from the nominal value (95%). In 12 simulations for the GP method, there were three settings in which the coverage error exceeded half a percentage point. These occured in the small sample settings, or settings in which there were fairly small samples from highly skewed distributions (Designs 1a, 2e, and 2g). Further simulations, not reported here, appear to confirm these observations.

Coverage frequencies for the maximum likelihood, bootstrap, and SLLR approaches were not satisfactory. The ML and bootstrap approaches resulted in fairly poor coverage, particularly in the small sample settings. Further simulations, not reported here, suggest that the coverage is yet worse when the sample originates from a distribution that is fairly highly skewed. The SLLR approach resulted in fairly good coverage, but the frequencies were noticeably worse in the small sample settings.

Neither the GP or MSLLR approach resulted in high bias. The GP approach appears to be slightly more biased than the MSLLR. The ML and bootstrap methods revealed strong left bias (bias toward the left) in Designs 1d and 1e, and strong right bias in the variations of Design 2. The SLLR approach resulted in similar patterns of bias, though the bias in each of those settings was never as large as that allowed by the ML or bootstrap methods. The implication is that one sided intervals obtained via the GP and MSLLR methods should result in accurate coverage, whereas one sided ML, bootstrap, and SLLR intervals may result in over coverage or under coverage.

6 Simulation results for the difference of means

Results from the simulations for δ are presented in Table 4. The results are fairly similar to those presented and discussed in Section 5; results for the SLLR and GP methods are nearly identical to those presented in the previous section.

Of the four methods examined in our simulation studies, the GP method clearly provided the most accurate coverage frequencies. In 9 of 12 simulations for the GP approach, the coverage error was less than half a percentage point. The ML and bootstrap approaches both revealed extremely poor coverage frequencies. The SLLR approach resulted in fairly accurate coverage, but the frequencies were noticeably worse in the small sample settings.

7 Discussion

For interval estimation of the ratio of means, we encourage use of the MSLLR approach. It results in highly accurate coverage frequencies in nearly all settings. For interval estimation for the difference of means, we recommend use of the generalized pivotal approach. This approach performs extremely well in most setting, though does appear to result in slightly worse coverage frequencies in small sample settings and/or settings when there are small samples from highly skewed distributions. If then, inference is only needed on relative differences, we recommend using the MSLLR approach to construct a confidence interval for the ratio of means (rather than the GP approach for the difference).

In this article, we have provided what we believe to be a fairly thorough and satisfactory comparison of the methods available for estimating two lognormal means. Nevertheless, we do not discourage further research. An important issue is that the methods as given above are inappropriate for data that include zero values. One possibility is to model the data as a mixture of a binomial and lognormal distribution (see, e.g., [4]). Using this approach, we have extended the methods discussed in this paper for data that also include zeros. The results of this exploration have been very encouraging.

Finally, we emphasize that statistical models should always be checked, whenever possible. Parametric approaches are sometimes criticized because they do not perform well when assumptions are violated. A recently published article [12] for instance, includes text that dissuades the use of lognormal approaches (it examines methods for estimating a single lognormal mean). The authors report the results of a series of simulation studies: these indicate that methods that assume lognormal data do not always perform well when the data in fact originate from a gamma distribution.

Unfortunately, the authors of that text fail to mention that the parametric methods they examined performed worst precisely in the settings in which the distribution was highly different from a lognormal curve. That is, they performed worst in situations in which model checks (such as the quantile plot and Shapiro-Wilk test) were likely to reveal violations of the distributional assumption.

It is of our opinion that robustness is only really an issue in situations when the scientist might incorrectly fail to reject the model assumption. Our own studies indicate that the MSLLR and GP approaches are both fairly robust in those sorts of settings. Details of this exploration can be obtained from the authors.

8 Illustrative example

We conclude our discussion by applying the techniques to a real example. We examine medical costs for patients with type I diabetes and patients being treated for diabetic ketoacidosis (DKA). The data are similar (but not identical) to those used in a study by Javor and colleagues [11]. For the present discussion, we are interested in determining whether the mean cost in the first group is equal to the mean cost in the second group.

First, we examined the data to check whether the data follow lognormal distributions. Quantile plots of the log-transformed data do not reveal any serious violations (Figure 1): the points fall fairly close to the quantile lines. We also performed a Shapiro-Wilk test for each of the two groups (for the log-transformed data). The tests do not provide evidence against lognormality: the p-value for the first group is 0.294; the p-value for the second group is 0.290. The lognormal model then, appears to be appropriate.

The mean cost in the first group is \$18,850.21; the mean cost in the second group is \$18,583.82. Interval estimates for the ratios and differences of the means are provided in Table 6. Note that these are presented only for purposes of comparison. We do not encourage using more than one approach in practice. Each of the intervals for the ratio of means include 1, and each of the intervals for the difference of means include 0. It appears then that the mean health costs are in fact equal.

9 Acknowledgments

This research was supported in part by AHRQ grant R01HS013105 and the Department of Veterans Affairs, Health Services Research and Development Service ECI 03-206-1. This report presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

References

[1] Krishnamoorthy K, Mathew T. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference* 2003; **115**:103-120.

[2] Taylor DJ, Kupper LL, Muller KE. Improved approximate confidence intervals for the mean of a lognormal random variable. *Statistics in Medicine* 2002; **21**:1443-1459.

[3] Wu J, Wong ACM, Jiang G. Likelihood-based confidence intervals for a log-normal mean. *Statistics in Medicine* 2003; **22**:1849-1860.

[4] Zhou X-H, Tu W. Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics and Data Analysis* 2000; **35**:201-210.

[5] Wu J, Jiang G, Wong ACM, Sun X. Likelihood analysis for the ratio of means of two independent log-normal distributions. *Biometrics* 2002; **58**:463-469.

[6] Barndorff-Nielsen OE. Inference on full and partial parameters, based on the standarized signed log-likelihood ratio. *Biometrika* 1986; **73**:307-322.

[7] Barndorff-Nielsen OE. Modified signed log likelihood ratio. *Biometrika* 1991; **78**:557-563.

[8] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2005.

[9] Weerahandi S. Generalized confidence intervals. *Journal of the American Statistical Association* 1993; **88**:899-905.

[10] Zhou X-H, Tu W. Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics* 2000; **56**:1118-1125.

[11] Javor KA, Kotsanos JG, McDonald RC, Baron AD, Kesterson JG, Tierney WM. Diabetic ketoacidosis charges relative to medical charges of adult patients with type I diabetes. *Diabetes Care* 1997; **20**:349-354.

[12] Briggs A, Nixon R, Dixon S, Thompson S. Parametric modelling of cost data: some simulation evidence. *Health Economics* 2005; **14**:421-428

Table 1: Sample sizes and parameter values used in the simulation studies.

design	n_1	n_2	μ_1	μ_2	σ_1^2	σ_2^2	γ_1	γ_2
1a	5	5	0	0	3	3	96.485	96.485
1b	25	25						
1c	50	50						
1d	5	25						
1e	25	50						
2a	5	5	0.75	0	0.5	2	2.939	23.732
2b	25	25						
2c	50	50						
2d	5	25						
2e	25	5						
2f	25	50						
2g	50	25						

Table 2: Results from the simulations for the ratio of means. Each simulation utilizes 10,000 95% confidence intervals.

design	method	coverage	left error	right error	relative bias
1a	ML	92.82	3.53	3.65	0.02
	bootstrap	97.79	1.05	1.16	0.05
	SLLR	91.32	4.23	4.45	0.03
	MSLLR	94.24	2.81	2.95	0.02
	GP	95.55	2.27	2.18	-0.02
1b	ML	95.54	2.12	2.34	0.05
	bootstrap	97.90	0.95	1.15	0.10
	SLLR	(94.73)	2.60	2.67	0.01
	MSLLR	(95.15)	2.34	2.51	0.04
	GP	(95.04)	2.36	2.60	0.05
1c	ML	(95.04)	2.35	2.61	0.05
	bootstrap	96.59	1.65	1.76	0.03
	SLLR	(94.68)	2.53	2.79	0.05
	MSLLR	(94.59)	2.64	2.77	0.02
	GP	(94.50)	2.67	2.83	0.03
1d	ML	87.18	12.39	0.43	-0.93
	bootstrap	88.06	11.78	0.16	-0.97
	SLLR	91.57	6.65	1.78	-0.58
	MSLLR	(94.66)	3.04	2.30	-0.14
	GP	(94.77)	3.14	2.09	-0.20
1e	ML	(94.96)	3.84	1.20	-0.52
	bootstrap	95.55	4.20	0.25	-0.89
	SLLR	(94.73)	3.01	2.26	-0.14
	MSLLR	(95.02)	2.49	2.49	0.00
	GP	(94.97)	2.66	2.37	-0.06

Table 3: *(continued)*

design	method	coverage	left error	right error	relative bias
2a	ML	85.72	0.96	13.32	0.87
	bootstrap	88.40	0.14	11.46	0.98
	SLLR	91.13	2.42	6.45	0.45
	MSLLR	(94.78)	2.31	2.91	0.11
	GP	(95.21)	1.74	3.05	0.27
2b	ML	93.57	0.68	5.75	0.79
	bootstrap	91.97	0.04	7.99	0.99
	SLLR	(94.62)	2.08	3.30	0.23
	MSLLR	(95.15)	2.39	2.46	0.01
	GP	(94.82)	2.31	2.87	0.11
2c	ML	93.69	1.05	5.26	0.67
	bootstrap	92.37	0.29	7.34	0.92
	SLLR	(94.66)	2.08	3.26	0.22
	MSLLR	(94.87)	2.29	2.84	0.11
	GP	(94.66)	2.28	3.06	0.15
2d	ML	93.05	3.09	3.86	0.11
	bootstrap	95.85	0.88	3.54	0.60
	SLLR	93.51	3.51	2.98	-0.08
	MSLLR	(94.85)	2.83	2.32	-0.10
	GP	(95.30)	2.02	2.68	0.14
2e	ML	82.35	0.54	17.11	0.94
	bootstrap	80.44	0.04	19.52	1.00
	SLLR	90.90	2.20	6.90	0.52
	MSLLR	(94.75)	2.55	2.70	0.03
	GP	94.09	2.88	3.03	0.03
2f	ML	(94.74)	1.11	4.15	0.58
	bootstrap	94.39	0.42	5.19	0.85
	SLLR	(94.96)	2.06	2.98	0.18
	MSLLR	(95.30)	2.16	2.54	0.08
	GP	(95.11)	2.07	2.82	0.15
2g	ML	91.95	0.66	7.39	0.84
	bootstrap	89.06	0.04	10.90	0.99
	SLLR	94.10	2.11	3.79	0.28
	MSLLR	(94.61)	2.47	2.92	0.08
	GP	94.29	2.40	3.31	0.16

Table 4: Results from the simulations for the difference of means. Each simulation utilizes 10,000 95% confidence intervals.

design	method	coverage	left error	right error	relative bias
1a	ML	99.01	0.43	0.56	0.13
	bootstrap	99.68	0.12	0.20	0.25
	SLLR	91.32	4.23	4.45	0.03
	GP	95.56	2.26	2.18	-0.02
1b	ML	99.93	0.03	0.04	0.14
	bootstrap	100.00	0.00	0.00	–
	SLLR	(94.73)	2.60	2.67	0.01
	GP	(95.05)	2.37	2.58	0.04
1c	ML	99.92	0.04	0.04	0.00
	bootstrap	100.00	0.00	0.00	–
	SLLR	(94.68)	2.53	2.79	0.05
	GP	(94.50)	2.67	2.83	0.03
1d	ML	98.66	1.26	0.08	-0.88
	bootstrap	99.93	0.06	0.01	-0.71
	SLLR	91.57	6.65	1.78	-0.58
	GP	(94.77)	3.14	2.09	-0.2
1e	ML	99.67	0.31	0.02	-0.88
	bootstrap	100.00	0.00	0.00	–
	SLLR	(94.73)	3.01	2.26	-0.14
	GP	(94.98)	2.65	2.37	-0.06

Table 5: *(continued)*

design	method	coverage	left error	right error	relative bias
2a	ML	87.15	0.15	12.70	0.98
	bootstrap	92.34	0.04	7.62	0.99
	SLLR	91.13	2.42	6.45	0.45
	GP	(95.19)	1.75	3.06	0.27
2b	ML	92.27	0.00	7.73	1.00
	bootstrap	85.82	0.00	14.18	1.00
	SLLR	(94.62)	2.08	3.30	0.23
	GP	(94.85)	2.31	2.84	0.10
2c	ML	92.70	0.03	7.27	0.99
	bootstrap	87.59	0.00	12.41	1.00
	SLLR	(94.66)	2.08	3.26	0.22
	GP	(94.66)	2.28	3.06	0.15
2d	ML	96.98	0.09	2.93	0.94
	bootstrap	97.05	0.00	2.95	1.00
	SLLR	93.51	3.51	2.98	-0.08
	GP	(95.30)	2.02	2.68	0.14
2e	ML	75.96	0.06	23.98	1.00
	bootstrap	68.59	0.01	31.40	1.00
	SLLR	90.89	2.21	6.90	0.51
	GP	94.10	2.88	3.02	0.02
2f	ML	95.13	0.07	4.80	0.97
	bootstrap	93.60	0.00	6.40	1.00
	SLLR	(94.96)	2.06	2.98	0.18
	GP	(95.09)	2.07	2.84	0.16
2g	ML	88.90	0.00	11.10	1.00
	bootstrap	80.00	0.00	20.00	1.00
	SLLR	94.10	2.11	3.79	0.28
	GP	94.28	2.41	3.31	0.16

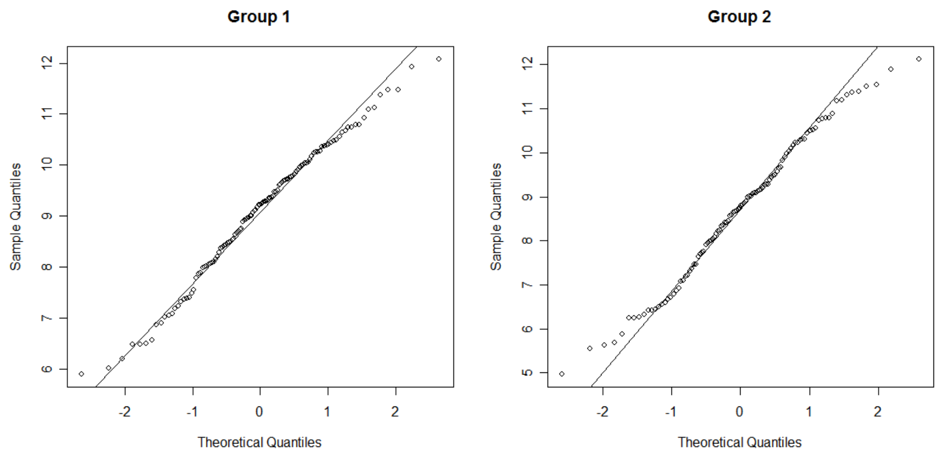

Figure 1: Normal quantile plots for the log-transformed medical data.

Table 6: 95% confidence intervals for the ratio and difference of the mean medical costs.

method	ratio	difference
ML	(0.53, 1.69)	(-14101.80, 11686.07)
bootstrap	(0.58, 1.70)	(-10692.69, 13527.26)
SLLR	(0.51, 1.67)	(-18429.75, 11360.78)
MSLLR	(0.50, 1.67)	−−
GP	(0.53, 1.58)	(-22637.45, 10937.92)