Bayesian Confidence Intervals for the Ratio of Means of Lognormal Data with Zeros

J. Harvey[a,b]
&
A.J. van der Merwe[b]

[a]Centre for Statistical Consultation
Department of Statistics and Actuarial Science
University of Stellenbosch
Private Bag X1
Stellenbosch
South Africa

[b]Department of Mathematical Statistics and Actuarial Science
University of the Free State
P.O. Box 339
Bloemfontein
9300
South Africa

Abstract

The lognormal distribution is currently used extensively to describe the distribution of positive random variables. This is especially the case with data pertaining to occupational health and other biological data. One particular application of the data is statistical inference with regards to the mean of the data and even more specifically the ratio between two means from two different lognormal distributions with different parameters. An added specification to the problem is the analysis of data with zero observations, where the non-zero data has a lognormal distribution. In this paper we consider a variety of issues from a Bayesian perspective, namely, the problem of constructing Bayesian confidence intervals for the ratio of the means of two independent populations that contain both lognormally distributed data and zero observations. An extensive simulation study is conducted to evaluate the coverage accuracy, interval width and relative bias of the proposed method. In addition, since the Bayesian procedure is evaluated the choice of which prior distribution to use becomes an important consideration. Three different prior distributions (independence Jeffreys' prior, the Jeffreys-rule prior, namely, the square root of the determinant of the Fisher Information matrix and uniform prior) are evaluated and compared to determine which give the best coverage with the most efficient interval width. Finally, this analysis is a Bayesian adaptation of the maximum likelihood and bootstrap methods originally proposed by Zhou and Tu (2000) and also the generalized confidence intervals used by Krishnamoorthy and Mathew (2003). The simulation results indicate that for the analysed priors the Bayesian procedure gives the correct coverage probability and is in general better than the maximum likelihood and bootstrap procedures. In addition to this, reference and probability-matching priors were derived and applied to rainfall data.

Keywords: Bayesian procedure; Lognormal; Zero observations; Monte Carlo simulation; Credibility intervals; Coverage probabilities.

Introduction

Lognormally distributed data presents itself in a number of scientific fields. However, the previously mentioned problem, as proposed by Zhou and Tu (2000) was in response to a demand in the field of diagnostic medical testing. The Ambulatory Care Group (ACG) was used to divide patients into distinct populations based on the burden of their medical illness. The ACG system provides a method of measuring the health status of the patient as well as the health resources they are likely to consume (Starfield, Weiner, Mumford, Steinwachs, 1991; Weiner, Starfield, Mumford., 1991). The diagnostic testing charges were then obtained for each patient. Since some of the patients had no diagnostic testing charges this resulted in zero observations. However, the testing charges for patients with data can be modeled with a lognormal distribution.

For testing the equality of means from two skewed populations Zhou, Gao and Hui (1997) and Zhou, Melfi and Hui (1997) first proposed a Z-score method for populations that have lognormal distributions. Zhou and Tu (1999) then extended the scenario by proposing a likelihood ratio test for instance when the populations contain both zero and non-zero observations. Furthermore, Zhou and Gao (1997) did propose confidence intervals for the one-sample lognormal mean, but until Zhou and Tu (2000) no confidence intervals had been proposed for methods that compare the means of two populations. Even though Zhou & Tu (1999) provided a test of whether the means were the same, if they were indeed found to be different the method they proposed did not add any additional information on the relative differences and magnitudes of the two population means, as a confidence interval would indeed.

Tian and Wu (2007), Tian (2005) and Tian and Wu (2006) variously proposed a frequentist method that is in effect rather similar to the method proposed here. However, this method was not suggested for the case of the ratio of two means from a lognormal population containing zero values, but rather only from a single mean.

As mentioned, Zhou and Tu (2000) considered the problem of constructing confidence intervals for the ratio of two means of independent populations that contain both lognormal and zero observations. The context of the proposed techniques is as described earlier concerning the excess charges of diagnostic testing. For the purposes of the analysis a maximum likelihood method and a two-stage bootstrap method was used. An extensive simulation study was conducted to ascertain the coverage accuracy, interval width and relative bias of the proposed methods. The focus was also on inferences about the overall population means, including zero costs. The results for different methods indicated that for confidence interval estimation of the ratio of the two different population means when the two skewness coefficients are the same the maximum likelihood based method had a better coverage accuracy than the bootstrap method. However, when the skewness coefficients are not the same the bootstrap method provides better results in terms of coverage accuracy than what the maximum likelihood based methods did.

In this paper we take a Bayesian approach to the problem, that is, constructing credibility intervals (Bayesian confidence intervals) for the ratio of the means from two independent lognormal populations that contain zero and non-zero observations. Depending on the choice of prior distribution it will be shown that the Bayesian procedure has better coverage accuracy than both the maximum likelihood and bootstrap methods. In the next section we begin with a formulation of the model and a specification of all parameters and distributions of interest. In further sections we compare the performance of the method for different prior distributions by conducting a simulation study to assess the following quantities of the proposed credibility intervals in pre-defined finite sample sizes (the same as those used by Zhou and Tu [2000]):

- Coverage accuracy
- Interval width
- Bias

Description of the Setting

The lognormal distribution in itself does not allow for zero values to be included in the data. This suggests an interesting setting, namely the analysis of data that contains both zero and non-zero values, with the non-zero values being lognormally distributed.

The Case of Zero-Valued Observations

Model Formulation

From the specification of the problem in the Introduction we can assume that the populations of interest contain both zero and non-zero (positive observations) and we furthermore assume that the probability of obtaining a zero observation from the *j*-th population (*j* = 1,2) is δ_j where $0 \le \delta_j \le 1$. Furthermore, we assume that the non-zero observations are distributed lognormally with mean μ_j and variance σ_j^2. Now, let $X_{1j}, X_{2j}, ..., X_{nj}$ be a random sample from the j^{th} population and let $M_j = E(X_{ij})$. From this preliminary setting specification we wish to construct credibility intervals for the ratio of the means, M_1 and M_2, of the two populations. As in Zhou and Tu (2000) we assume that in the j^{th} sample the non-zero observations come first: $X_{ij} > 0$, and $\ln(X_{ij}) \mid n_{j1} \sim N(\mu_j, \sigma_j)$, for $i = 1, ..., n_{j1}$. In addition, $X_{ij} = 0$, for $i = n_{j1} + 1, ..., n_j$ and $n_{j0} = n_j - n_{j1} \sim Bin(n_j, \delta_j)$. From this it follows that the mean of the j^{th} population, which is a function of μ_j, σ_j^2 and δ_j, is given by:

$$M_j = (1 - \delta_j)\exp(\mu_j + \frac{1}{2}\sigma_j^2).$$

To compare the two population means we will construct credibility intervals (Bayesian confidence intervals) for the ratio of the means:

$$\frac{M_1}{M_2} = \frac{(1-\delta_1)\exp(\mu_1 + \frac{1}{2}\sigma_1^2)}{(1-\delta_2)\exp(\mu_2 + \frac{1}{2}\sigma_2^2)}.$$

Intervals Based on a Bayesian Procedure

Denote $y_{ij} = \ln X_{ij}$ and $\boldsymbol{\theta} = [\delta_1 \quad \mu_1 \quad \sigma_1^2 \, \delta_2 \quad \mu_2 \quad \sigma_2^2]'$ then the likelihood function is given by:

$$L(\boldsymbol{\theta} \mid data) \propto \prod_{j=1}^{2} \{\delta_j^{n_{j0}} (1-\delta_j)^{n_{j1}} \prod_{i=1}^{n_{j1}} (\frac{1}{\sigma_j^2})^{\frac{1}{2}} \exp[-\frac{(y_{ij}-\mu_j)^2}{2\sigma_j^2}]\} \tag{1}$$

The choice of prior to be used in this setting will be discussed in further sections. Given the previous specification of the likelihood, the Fisher Information Matrix in our case can be written as:

$$I(\boldsymbol{\theta}) = -E\left\{\frac{\partial^2}{\partial^2\boldsymbol{\theta}} lnL(\boldsymbol{\theta}|data)\right\}$$

Therefore,

$$I(\boldsymbol{\theta}) = diag\left[\frac{n_1}{\delta_1(1-\delta_1)} \quad \frac{n_1(1-\delta_1)}{\sigma_1^2} \quad \frac{n_1(1-\delta_1)}{2\sigma_1^4} \quad \frac{n_2}{\delta_2(1-\delta_2)} \quad \frac{n_2(1-\delta_2)}{\sigma_2^2} \quad \frac{n_2(1-\delta_2)}{2\sigma_2^4}\right] \tag{2}$$

1 Independence Jeffreys Prior:

Since $\boldsymbol{\theta}$ is unknown the prior

$$p(\boldsymbol{\theta}) \propto \prod_{j=1}^{2} \sigma_j^{-2} \delta_j^{-1/2} (1-\delta_j)^{-1/2} \tag{3}$$

will be specified for the unknown parameters. This is known as the independence Jeffreys prior. In (3) we have assumed μ_j and σ_j^2, for $j = 1,2$ to be independently distributed, *a priori*, with μ_j and $\log\sigma_j^2$ each uniformly distributed. See Zellner (1971) and Box and Tiao (1973) for further discussion. The prior $p(\delta_j) \propto \delta_j^{-\frac{1}{2}}(1-\delta_j)^{-\frac{1}{2}}$ is the one proposed by Jeffreys (1967) for the binomial parameter. Combining the likelihood function (1) and the prior density function (3) the joint posterior density function can be written as:

$$P(\boldsymbol{\theta} \mid data) = \prod_{j=1}^{2} \left\{ \begin{array}{l} \dfrac{1}{B(n_{j0}+0.5; n_{j1}+0.5)} \delta_j^{n_{j0}-\frac{1}{2}} (1-\delta_j)^{n_{j1}-\frac{1}{2}} \\ \times \left(\dfrac{2\pi\sigma_j^2}{n_{j1}} \right)^{-\frac{1}{2}} \exp\left[-\dfrac{n_{j1}}{2\sigma_j^2} (\mu_j - \hat{\mu}_j)^2 \right] \left(\dfrac{\nu_{j1}\hat{\sigma}_j^2}{2} \right)^{\frac{1}{2}\nu_{j1}} \left(\dfrac{(\sigma_j^2)^{-\frac{1}{2}(\nu_{j1}+2)} \exp\left[-\dfrac{\nu_{j1}\hat{\sigma}_j^2}{2\sigma_j^2} \right]}{\Gamma\left(\dfrac{\nu_{j1}}{2} \right)} \right) \end{array} \right\} \quad (4)$$

where $\hat{\mu}_j = \frac{1}{n_{j1}} \sum_{i=1}^{n_{j1}} y_{ij}$, $\nu_{j1} = n_{j1} - 1$,

$\hat{\sigma}_j^2 = \frac{1}{\nu_{j1}} \sum_{i=1}^{n_{j1}} (y_{ij} - \hat{\mu}_j)^2$ and $B(m,n) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}$.

From (4) it follows that the posterior distribution of δ_j is a Beta distribution (specifically $B\left(n_{j0} + \frac{1}{2}; n_{j1} + \frac{1}{2}\right)$) and δ_j is independently distributed of μ_j and σ_j^2, where the conditional posterior distribution of μ_j is normal:

$$\mu_j \mid \sigma^2, data \sim N\left(\hat{\mu}_j, \frac{\sigma_j^2}{n_{j1}} \right) \quad (5)$$

and for σ_j^2, the posterior density function is an Inverted Gamma density, specifically:

$$P(\sigma_j^2 \mid data) = \left(\frac{\nu_{j1}\hat{\sigma}_j^2}{2} \right)^{\frac{1}{2}\nu_{j1}} \left(\frac{(\sigma_j^2)^{-\frac{1}{2}(\nu_{j1}+2)} \exp\left[-\frac{\nu_{j1}\hat{\sigma}_j^2}{2\sigma_j^2} \right]}{\Gamma\left(\frac{\nu_{j1}}{2} \right)} \right). \quad (6)$$

From equation (6) it follows that $\tau_j^* = \frac{\nu_{j1}\hat{\sigma}_j^2}{\sigma_j^2}$ has a chi-square distribution with ν_{j1} degrees of freedom. From classical statistics (if $\hat{\sigma}_j^2$is considered to be random) it is well known that τ_j^* is also distributed chi-square with ν_{j1} degrees of freedom. This agreement between classical and Bayesian statistics is only true if the prior $p(\sigma_j^2) \propto \sigma_j^{-2}$ is used. If some other prior distributions are used, for example $p(\sigma_j^2) \propto \sigma_j^{-3}$ or $p(\sigma_j^2) \propto constant$, then the posterior of τ_j^* will still be a chi-square distribution but the degrees of freedom will be different.

The method proposed here to find the Bayesian credibility intervals for $D = \ln M_1 - \ln M_2$, the log of the ratio of the population means, is through Monte Carlo simulation. Since $\ln M_j = \ln(1-\delta_j) + \mu_j + \frac{\sigma_j^2}{2}$ (j = 1,2), standard routines can be used in the simulation procedure.

Simulation Procedure:
The following simulation was obtained from the preceding theory using the MATLAB® package:

1. Simulation of σ_j^2 can be obtained from (6) in the following way:
   a. Simulate τ_j^* from a $\chi^2_{v_{j1}}$ distribution, as the sum of v_{j1} squared independent normal random variables.
   b. Calculate $\sigma_j^{2^*} = \frac{v_{j1}\hat{\sigma}_j^2}{\tau_j^{*2}}$
2. Given $\sigma_j^{2^*}$, simulate μ_j^* from (5).
3. A simulated value of δ_j (a Beta random variable), namely δ_j^*, can easily be obtained by using the rejection method (Rice (1995) p. 91). The rejection method is commonly used to generate random variables from a density function, especially when the inverse of the cumulative distribution function (CDF) cannot be found in closed form. However, for the purposes of this analysis Beta-distributed random variables were simulated using built-in MATLAB® functions.
4. Substitute the simulated values $\sigma_j^{2^*}$, μ_j^* and δ_j^* into the expression for D to obtain D^*, a simulated value for the log-ratio of the population means.
5. Repeat steps 1. to 4. l times to obtain l simulated values, $D_1^*, D_2^*, ..., D_l^*$. Sort then in ascending order such that $D_{(1)}^* \le D_{(2)}^* \le ... \le D_{(l)}^*$.
6. Let $K_1 = \left[\frac{\alpha}{2}l\right]$ and $K_2 = \left[\left(1-\frac{\alpha}{2}\right)l\right]$ where $[a]$ denotes the largest integer not greater than a.
7. $\{D_{(k_1)}^*, D_{(k_2)}^*\}$ is then a $100(1-\alpha)\%$ Bayesian confidence interval for $\ln\left(\frac{M_1}{M_2}\right)$.
8. The resulting Bayesian confidence interval for the ratio $\frac{M_1}{M_2}$ is $\left(\exp(D_{(k_1)}^*), \exp(D_{(k_2)}^*)\right)$

Alternate Prior Distributions – Jeffreys-Rule Prior:
As mentioned in the Abstract and Introduction to this document, one of the objectives was to compare the Bayesian procedure for different choices of prior distributions for $\boldsymbol{\theta}$,

the unknown parameters. In the previous two sections we discussed the analysis methods using Jeffreys' non-informative prior and the resulting simulation technique respectively.

In this and subsequent sections different choices of prior distributions will be discussed in an effort to eventually compare the results. The choice of density applied in this section is the square root of the determinant of the Fisher Information Matrix, which is an adaptation of the Jeffreys' rule used in the previous section.

Since $\boldsymbol{\theta}$ is unknown this prior becomes

$$p(\boldsymbol{\theta}) \propto \prod_{j=1}^{2} \sigma_j^{-3} \delta_j^{-1/2} (1-\delta_j)^{1/2} \qquad (7)$$

This was derived from $|I(\boldsymbol{\theta})|^{1/2}$, which was defined in (2). In (6) we have assumed μ_j and σ_j^2, for $j=1,2$ to be independently distributed, *a priori*, with μ_j and $\log\sigma_j^2$ each uniformly distributed. Combing the likelihood function (1) and the prior density function (7) it follows that τ_j^* has a chi-square distribution with ν_{j1+1} degrees of freedom and the posterior distribution of μ_i given σ_j^2 is as defined in equation (5). From (7) it is also clear that the posterior distribution of δ_j is a Beta distribution (specifically $B\left(n_{j0}+\frac{1}{2}, n_{j1}+\frac{3}{2}\right)$) and δ_j is distributed independently of μ_i and σ_j^2.

A similar simulation procedure to the one previously described can be used with the following differences:

1. Simulation of σ_j^2 can be obtained in the following way:
   a. Simulate τ_j^* from a $\chi^2_{\nu_{j1+1}}$ distribution, as the sum of $\nu_{j1}+1$ squared independent normal random variables.
   b. Calculate $\sigma_j^{2^*} = \dfrac{(\nu_{j1})\hat{\sigma}_j^2}{\tau_j^{*2}}$
   c. Simulate δ_j from a $B\left(n_{j0}+\frac{1}{2}, n_{j1}+\frac{3}{2}\right)$ distribution.

Furthermore, the simulation procedure is similar to the procedure described previously.

Alternate Prior Distributions – Constant (Uniform) Prior:

Since $\boldsymbol{\theta}$ is unknown this prior becomes

$$p(\boldsymbol{\theta}) \propto const \qquad (8)$$

In (8) we have assumed μ_j and σ_j^2, for $j=1,2$ to be independently distributed, *a priori*, with μ_j and σ_j^2 each uniformly distributed. Combing the likelihood function (1) and the prior density function (8) it follows that τ_j^* has a chi-square distribution with ν_{j1-1} degrees of freedom, the posterior distribution of δ_j is a Beta distribution (specifically $B\left(n_{j0}+1; n_{j1}+1\right)$) and δ_j is independently distributed of μ_j and σ_j^2. Furthermore, the simulation procedure is similar to the procedure described previously.

Simulation Study

As was done in Zhou and Tu (2000) we will use computer simulations to study the operating characteristics of the proposed Bayesian confidence interval procedure in finite sample sizes. Random sample sizes containing both zero and lognormal observations are generated using the following different sample sizes:

Table 1
Sample Sizes Analysed by Monte Carlo Simulation Techniques

n_1	n_2
10	10
25	25
50	50
100	100
10	25
25	10
25	50

Zero proportions with different skewness coefficients are also considered. Based on these generated samples the credibility intervals (or Bayesian confidence intervals [BCI's]) are constructed. The following additional characteristics are reported:

- coverage probabilities
- average interval lengths
- coverage error (target coverage – actual coverage),
- percentages of under-coverage on both sides ($\%BCI < \theta$ and $\%BCI > \theta$)
- relative bias $\frac{|\%BCI < \theta - \%BCI > \theta|}{(\%BCI < \theta + \%BCI > \theta)}$.

As was in Zhou and Tu (2000) the nominal significance level of $\alpha = 0.05$ will be used and for each parameter setting, $C = 10000$ random samples are simulated to ensure that the margin of error is less than 0.005 with 95% confidence. l is taken to be 1000.

In the following table the parameter settings used in the simulation study are presented (the skewness coefficients for samples 1 and 2 are reported under headings γ_1 and γ_2):

Table 2
Parameter Settings used in the Simulation Study

Design	σ_1^2	σ_2^2	δ_1	δ_2	γ_1	γ_2
1	3.0	1.0	0.0	0.0	96.4851	6.1849
2	4.0	4.0	0.0	0.0	414.3593	414.3593
3	3.0	1.0	0.1	0.1	100.9809	6.1763
4	2.0	0.5	0.0	0.1	23.7323	2.6848
5	2.0	0.5	0.1	0.2	24.5572	2.5806

The results from the simulation study performed by Zhou and Tu (2000) for the Maximum Likelihood and Bootstrap methods have been supplied as well for the purposes of comparison. For the purposes of brevity only a summary of the average results of the five designs are presented here.

Discussion of Results for the Simulation Studies

As mentioned, Zhou and Tu (2000) presented the results for the ML and Bootstrap methods and compared only these. They ascertained that when the two population skewness coefficients are the same the ML-based method results in better coverage probabilities in comparison with the stated nominal level. However, it is found that the ML-based method is more biased than the bootstrap method, as evidenced by a larger relative bias. This was particularly evident when the sample sizes were not the same. The ML method tends to cover too many observations on the left and too few on the right.

When the two skewness coefficients are not the same the results indicate better coverage accuracy for the bootstrap method. This method is also less biased than the ML-based method.

However, the objective of this report was to compare these results against results obtained from a Bayesian-based simulation study using a specifically chosen set of prior distributions and to evaluate the performance of each prior distribution against both the other distributions and the results obtained by Zhou and Tu (2000), overall.

The following table presents the summary statistics of the results in both the Zhou and Tu (2000) simulation study and the Bayesian simulation study.

Table 3
Summary Results for Simulation Studies

Design	Method	Coverage Probability	Coverage Error	Average Length	$\%CI < \theta$	$\%CI > \theta$	Relative Bias
1	ML	0.9285	0.0215	2.3625	0.0636	0.0080	0.6678
	Bootstrap	0.9266	0.0234	2.6252	0.0393	0.0341	0.0857
	Prior 1	0.9550	-0.0050	3.2020	0.0234	0.0216	0.1126
	Prior 2	0.9473	0.0027	2.8889	0.0226	0.0301	0.2028
	Prior 3	0.9603	-0.0103	4.3057	0.0280	0.0117	0.4019
2	ML	0.9476	0.0024	3.8491	0.0278	0.0246	0.3417
	Bootstrap	0.9369	0.0131	3.9506	0.0324	0.0307	0.0287
	Prior 1	0.9494	0.0006	5.4961	0.0249	0.0257	0.1409
	Prior 2	0.9429	0.0071	4.8849	0.0271	0.0300	0.1506
	Prior 3	0.9590	-0.0090	7.6650	0.0207	0.0460	0.2535
3	ML	0.9237	0.0263	2.5127	0.0675	0.0088	0.7266
	Bootstrap	0.9298	0.0202	2.7003	0.0393	0.0308	0.1269
	Prior 1	0.9486	0.0014	3.5759	0.0257	0.0514	0.1422
	Prior 2	0.9446	0.0093	3.1424	0.0227	0.0327	0.2508
	Prior 3	0.9569	-0.0069	7.3283	0.0289	0.0143	0.3953
4	ML	0.9274	0.0226	1.7269	0.0624	0.0102	0.6828
	Bootstrap	0.9294	0.0206	1.8569	0.0380	0.0326	0.0716
	Prior 1	0.9564	-0.0064	2.2691	0.0211	0.0224	0.0608
	Prior 2	0.9513	-0.0013	2.0566	0.0206	0.0230	0.1699
	Prior 3	0.9619	-0.0157	3.0409	0.0224	0.0119	0.3122
5	ML	0.9274	0.0226	1.8666	0.0619	0.0108	0.6650
	Bootstrap	0.9346	0.0154	1.9795	0.0349	0.0305	0.0596
	Prior 1	0.9524	-0.0024	2.5810	0.0241	0.0234	0.1119
	Prior 2	0.9450	0.0050	2.2786	0.0243	0.0307	0.1356
	Prior 3	0.9627	-0.0127	4.6573	0.0250	0.0123	0.4981
Overall	ML	0.9309	0.0191	2.4636	0.0566	0.0125	0.6168
	Bootstrap	0.9315	0.0185	2.6225	0.0368	0.0317	0.0745
	Prior 1	0.9524	-0.0024	3.4248	0.0239	0.0289	0.1137
	Prior 2	0.9462	0.0046	3.0503	0.0235	0.0293	0.1819
	Prior 3	0.9601	-0.0109	5.3994	0.0250	0.0192	0.3722

From the overall summary statistics we see that the choices of prior distributions have better coverage than both the ML-based and bootstrap methods. However, this does not provide the full picture.

Coverage Probabilities

As evident from the above summary table, it is apparent that the ML and Bootstrap methods are comparable in terms of the coverage probability. The ML method, as noted by Zhou and Tu (2000), gives better coverage for designs 1, 2 and 3, *i.e.*, when the skewness coefficients of the two populations are the same. Otherwise the Bootstrap method offers superior coverage. However, the coverage probabilities overall for the ML and Bootstrap methods were 0.9309 and 0.9315 respectively.

The three Bayesian methods considered here all provide better coverage than the ML and Bootstrap methods proposed. However, at least one of the Bayesian methods, the method of the constant or uniform prior distribution, results in over-coverage, with an overall coverage ratio of 0.9601. Naturally, this will imply a larger coverage error when compared with the other prior distributions used, which is ultimately due to a larger average interval length. But this will be discussed further in later sections.

Overall, the best prior distribution to be used in terms of coverage probability was the independence Jeffreys prior. In terms of the literature, Box and Tiao (1973), this would be the natural choice of prior distribution in this setting and thus, its accuracy compared to the other prior distributions should be expected. The overall coverage probability was 0.9524 compared to 0.9462 for the Jeffreys-rule prior described previously.

However, the Jeffreys Rule appears to be nearly as good as the Independence Jeffreys prior. The prior tends to under cover, but not by much at all. What is particularly positive is that even though there is slight under coverage, the average interval length is shorter for the Jeffreys Rule prior.

A point of interest is that the coverage of the Bayesian methods does not appear to be affected by the skewness coefficients of the different designs.

The better coverage probabilities are as a direct result of the increased average interval lengths for the Bayesian methods. However, this is discussed in more comprehensively in subsequent sections.

Coverage Error

Overall, the coverage error for the Bootstrap method is better than if compared to the ML method. The only possible exception to this overall figure is perhaps the case when the population skewness coefficients are similar. However, this is by no means concrete.

For the Bayesian methods the overall coverage error was better for all choices of prior distributions, as opposed to the ML and Bootstrap methods. For the independence Jeffreys prior the coverage error appears smallest, thereby reinforcing the observation of the better coverage probability. This error for the uniform prior appears to increase when the population skewness coefficients are different.

Average Length

Firstly, results by Zhou and Tu (2000) indicate that the Bootstrap method results in intervals with longer interval lengths. Overall, the interval length for the Bootstrap method was 2.6225 compared to the 2.4636 of the ML method. As previously mentioned, coverage probability and average interval length are related. Thus, we would expect the average interval length for the Bootstrap method to be greater since it provides better probability of coverage. However, the average interval length for both these methods appears to be related to the population skewness coefficients in the following way: when the coefficients are the same (designs 1, 2 and 3) the average interval lengths are distinctly larger for the Bootstrap method, particularly when sample sizes are small.

Overall, when analyzing the results from the Bayesian methods it is apparent that the interval lengths are larger. Once again this would be expected due to the previously mentioned relationship between the coverage probabilities and the interval length. As with the methods proposed by Zhou and Tu (2000), the average interval length decreases when the population skewness coefficients are different. The independence Jeffreys prior and the Jeffreys-rule prior produced average interval lengths of 3.4248 and 3.0503 respectively. One way to obtain smaller Bayesian intervals with the correct coverage probability is to assign proper priors to the unknown parameters. The assignment of proper priors to specific parameters must however, be justifiable from a practical point of view. We also tested other types of non-informative prior (reference and probability-matching priors), but did not achieve any improvement on the independence Jeffreys and Jeffreys rule priors. In the next section we will, however, apply these two priors (reference and probability-matching priors) to rainfall data.

Lastly, as was mentioned previously, the constant or uniform prior tends to over-cover. This inefficiency is accurately portrayed by the average interval length, namely: 5.3994.

Coverage on Left and Right and Bias

As mentioned in previous sections, the results obtained by Zhou and Tu (2000) indicate that the ML method covers too many observation on the left and too many on the right. The only exception to this is Design 2. Overall, the Bootstrap method had a better spread.

The Bayesian methods employed indicate a much more equal spread of observations above and below. The uniform prior is the only possible exception. Thus, although the average interval lengths are greater for the Bayesian case the spread of the interval appears better, *i.e.*, the Bayesian methods overall tend to cover as many observations on the right as on the left.

In addition, even though the relative bias in the last column seems to be better for the Bootstrap method than the Bayesian methods employed, the exact bias ($|\%BCI > \theta - \%BCI < \theta|$) for the four cases are as follows:

Bootstrap:	0.0051
Independence Jeffreys prior:	0.0050
Jeffreys-rule prior:	0.0058
Constant Prior:	0.0058

Thus, the exact bias is comparable for all models.

Example – Rainfall Data

For the purposes of comparison of the different methods, an example was chosen using raw data obtained from the South African Weather Service. The data consisted of the monthly rainfall totals for the cities of Bloemfontein and Kimberley, two South African cities, over a period of 69 to 70 years of measurement. However, these two cities are both located in relatively arid regions and are characterised by mainly summer rainfall. For that reason, the winter months of June do contain some rainfall data, but also contain many years where the total monthly rainfall data was zero. Probability plots as well as the Shapiro-Wilk (1965) test indicate that the lognormal distribution is a better fit than the normal distribution. The data can be summarised as follows:

Table 4
Summary of the Rainfall Data

City	Parameter	Value
Bloemfontein	Number of Years of Available Data	70
	Number of Zero Valued Observations	18
	Mean of Log-Transformed Data	1.9578
	Variance of Log-Transformed Data	2.1265
Kimberley	Number of Years of Available Data	69
	Number of Zero Valued Observations	10
	Mean of Log-Transformed Data	1.0526
	Variance of Log-Transformed Data	3.1589

In order to compare the results, both the Maximum Likehood and Bootstrap methods of Zhou and Tu (2000) were applied to the data.

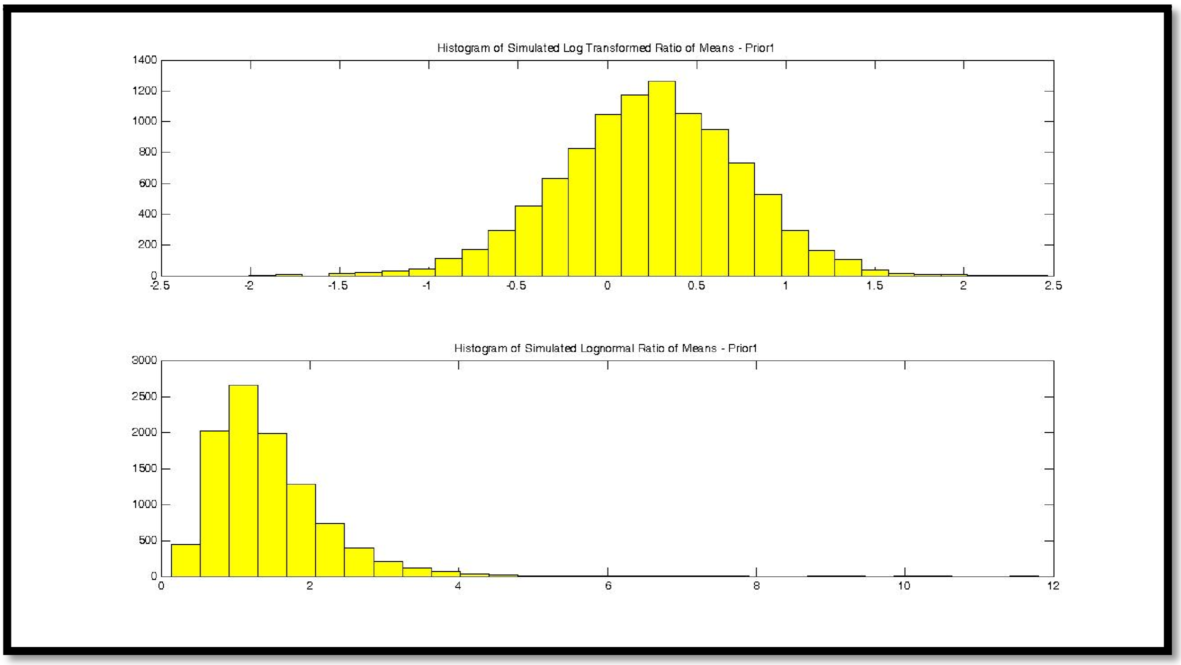
In addition to the maximum likelihood and boostrap confidence intervals the confidence intervals were obtained using the Bayesian methods described in the preceding text and for the following priors: Independence Jeffery's Prior (Prior 1 in the table), the Jeffery's Rule Prior (Prior 2), the constant prior (Prior 3) , the Reference Prior (Prior 4) and the Probability Matching Prior (Prior 5). Priors 4 and 5 will be derived in the next section. The results are presented in the following table:

Table 5
Summary of Results for the Rainfall Data

	Maximum Likelihood	Bootstrap	Prior 1	Prior 2	Prior 3	Reference Prior	Probability Matching Prior
Lower Limit of Logged Data	-0.6914	-0.6404	-0.8019	-0.7942	-0.8624	-0.8188	-0.7794
Upper Limit of Logged Data	1.1880	1.2444	1.2013	1.1954	1.2248	1.2104	1.1982
Lower Limit	0.5009	0.5271	0.4485	0.4520	0.4221	0.4410	0.4587
Upper Limit	3.2805	3.4710	3.3245	3.3048	3.4035	3.3548	3.3141

The following graphs also illustrate these results:

Prior 1



From Table 5 it is clear that the intervals for the seven methods are for practical purposes the same.

Probability-matching and Reference Priors for $M_e = e^{\mu+\frac{1}{2}\sigma^2}$, the Mean of the Lognormal Distribution

Probability-matching and reference priors often lead to procedures with good frequency properties while retaining the Bayesian flavor. The fact that the resulting posterior intervals of level $1-\alpha$ are also good frequentist intervals at the same level is a very desirable situation.

The Probability-matching prior for $M_e = e^{\mu+\frac{1}{2}\sigma^2}$

Datta and Ghosh (1995) derived the differential equation that a prior must satisfy if the posterior probability of a one-sided credibility interval (Bayesian confidence interval) for a parametric function and its frequentist probability agree up to $0(n^{-1})$, where n is the sample size. They proved that the agreement between the posterior probability and the frequentist probability holds if and only if the differential equation

$$\sum_{\alpha=1}^{m} \frac{\partial}{\partial \boldsymbol{\theta}_\alpha} \left\{ \eta_\alpha(\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \right\} = 0$$

is satisfied, where $p(\boldsymbol{\theta})$ is the probability-matching prior distribution for $\boldsymbol{\theta}$, the vector of unknown parameters.

Also,

$$\nabla_t = \left[\frac{\partial}{\partial \theta_1} t(\boldsymbol{\theta}), \ldots, \frac{\partial}{\partial \theta_m} t(\boldsymbol{\theta}) \right]'$$

and

$$\eta(\boldsymbol{\theta}) = \frac{F^{-1}(\boldsymbol{\theta}) \nabla_t(\boldsymbol{\theta})}{\sqrt{\nabla_t'(\boldsymbol{\theta}) F^{-1}(\boldsymbol{\theta}) \nabla_t(\boldsymbol{\theta})}} = \left[\eta_1(\boldsymbol{\theta}), \ldots, \eta_m(\boldsymbol{\theta}) \right]'.$$

It is clear that $\eta'(\boldsymbol{\theta}) F(\boldsymbol{\theta}) \eta(\boldsymbol{\theta}) = 1$ for all $\boldsymbol{\theta}$ where $F^{-1}(\boldsymbol{\theta})$ is the inverse of $F(\boldsymbol{\theta})$. $F(\boldsymbol{\theta})$ is the Fisher information matrix of $\boldsymbol{\theta}$ and $t(\boldsymbol{\theta})$ is the parameter of interest.

The following theorem can now be stated:

Theorem 1

For the mean, $M_e' = e^{\mu + \frac{1}{2}\sigma^2}$, of the lognormal distribution, the probability matching prior is given by:

$$p_P(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \sqrt{1 + \frac{2}{\sigma^2}}. \qquad (9)$$

Proof: The proof is given in the appendix.
Multiplying (9) by the likelihood function in equation (1) it follows that:

$$p_P(\sigma_j^2|data) \propto (\sigma_j^2)^{-\frac{1}{2}(\nu_{j1}+2)} \left(1 + \frac{2}{\sigma_j^2}\right)^{\frac{1}{2}} exp\left[-\frac{\nu_{j1}\hat{\sigma}_j^2}{2\sigma_j^2}\right]$$

for $j = 1,2$

(10)

Simulation from (10) can be obtained using the rejection method. Simulation of μ_j and δ_j are as before.

The Reference prior for $M_e = e^{\mu+\frac{1}{2}\sigma^2}$

The determination of reasonable, non-informative priors in multiparameter problems is not easy; common non-informative priors, such as Jeffreys' prior, can have features that have an unexpectedly dramatic effect on the posterior distribution. In recognition of this problem Berger and Bernardo (1992) proposed the *reference prior* approach to the development of non-informative priors. As in the case of the Jeffreys and probability-matching priors, the reference prior method is derived from the Fisher information matrix. Reference priors depend on the group ordering of the parameters. Berger and Bernardo (1992) suggested that multiple groups, ordered in terms of inferential importance, are allowed, with the reference prior being determined through a succession of analyses for the implied conditional problems. They particularly recommend the reference prior based on having each parameter in its own group, i.e. having each conditional reference prior be only one dimensional.

As mentioned by Pearn and Wu (2005) the reference prior maximises the difference in information (entropy) about the parameter provided by the prior and posterior distributions. In other words, the reference prior is derived in such a way that it provides as little as possible information about the parameter.

The following theorem can now be stated.

Theorem 2

For the mean, $M_e = e^{\mu+\frac{1}{2}\sigma^2}$, of the lognormal distribution, the reference prior relative to the ordered parameterisation (μ, σ^2) is given by:

$$p_R(\mu, \sigma^2) \propto \frac{1}{\sigma}\sqrt{1+\frac{2}{\sigma^2}}. \quad (11)$$

<u>Proof:</u> The proof is given in the appendix.
The posterior distribution of σ_j^2 is now

$$p_P(\sigma_j^2|data) \propto (\sigma_j^2)^{-\frac{1}{2}(\nu_{j1}+1)} \left(1 + \frac{2}{\sigma_j^2}\right)^{\frac{1}{2}} exp\left[-\frac{\nu_{j1}\hat{\sigma}_j^2}{2\sigma_j^2}\right]$$

for $j = 1,2$

(12)

Appendix

Proof of Theorem 1

The probability matching prior is derived from the inverse of the Fisher information matrix. Now

$$F^{-1}(\boldsymbol{\theta}) = F^{-1}(\mu,\sigma^2) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}.$$

and

$$t(\boldsymbol{\theta}) = e^{\mu+\frac{1}{2}\sigma^2}$$

from which it follows that

$$\frac{\partial t(\boldsymbol{\theta})}{\partial\mu} = e^{\mu+\frac{1}{2}\sigma^2} \text{ and } \frac{\partial t(\boldsymbol{\theta})}{\partial\sigma^2} = \frac{1}{2}e^{\mu+\frac{1}{2}\sigma^2}.$$

Also

$$\nabla_t'(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial t(\boldsymbol{\theta})}{\partial\mu} & \frac{\partial t(\boldsymbol{\theta})}{\partial\sigma^2} \end{bmatrix} = e^{\mu+\frac{1}{2}\sigma^2}\begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}$$

and

$$\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta}) = e^{\mu+\frac{1}{2}\sigma^2}\begin{bmatrix} \sigma^2 & \sigma^4 \end{bmatrix}$$

$$\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta}) = e^{2\mu+\sigma^2}\left(\sigma^2 + \frac{1}{2}\sigma^4\right)$$

and

$$\sqrt{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})} = e^{\mu+\frac{1}{2}\sigma^2}\sqrt{\sigma^2 + \frac{1}{2}\sigma^4}.$$

Therefore

$$\eta'(\boldsymbol{\theta}) = \frac{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})}{\sqrt{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})}} = [\eta_1(\boldsymbol{\theta}) \;\; \eta_2(\boldsymbol{\theta})] = \frac{1}{\sqrt{\sigma^2+\frac{1}{2}\sigma^4}}[\sigma^2 \quad \sigma^4]$$

For a prior $p_P(\boldsymbol{\theta}) = p_P(\mu,\sigma^2)$ to be a probability matching prior, the differential equation

$$\frac{\partial}{\partial\mu}[\eta_1(\boldsymbol{\theta}) \;\; p_P(\boldsymbol{\theta})] + \frac{\partial}{\partial\sigma^2}[\eta_2(\boldsymbol{\theta}) \;\; p_P(\boldsymbol{\theta})] = 0$$

must be satisfied. If we take

$$p_P(\mu,\sigma^2) \propto \frac{1}{\sigma^2}\sqrt{1+\frac{2}{\sigma^2}}$$

then the differential equation will be satisfied.

Proof of Theorem 2

The Fisher information matrix of $\boldsymbol{\theta} = \left[\mu, \sigma^2\right]$ per unit observation is given by:

$$F(\boldsymbol{\theta}) = F(\mu,\sigma^2) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{bmatrix}.$$

The parameter of interest is the mean of the lognormal distribution

$$t(\boldsymbol{\theta}) = e^{\mu + \frac{1}{2}\sigma^2}$$

Define $A = \dfrac{\partial(\mu,\sigma^2)}{\partial(t(\boldsymbol{\theta}),\sigma^2)} = \begin{bmatrix} 1/t(\boldsymbol{\theta}) & -1/2 \\ 0 & 1 \end{bmatrix}.$

Hence, the Fisher information matrix under the reparameterisation $(t(\boldsymbol{\theta}),\sigma^2)$ is given by

$$F(t(\boldsymbol{\theta}),\sigma^2) = A'F(\mu,\sigma^2)A = \begin{bmatrix} \dfrac{1}{t^2(\boldsymbol{\theta})\sigma^2} & \dfrac{-1}{2t(\boldsymbol{\theta})\sigma^2} \\ \dfrac{-1}{2t(\boldsymbol{\theta})\sigma^2} & \dfrac{1}{4\sigma^2} + \dfrac{1}{2\sigma^4} \end{bmatrix}.$$

Following the notation of Berger and Bernardo (1992), the functions $h_j, (j = 1,2)$, which are needed to calculate the reference prior for the group ordering $(t(\boldsymbol{\theta}),\sigma^2)$, can be obtained from $F(t(\boldsymbol{\theta}),\sigma^2)$ as follows:

$$h_1^{\frac{1}{2}} = \left| \frac{1}{t^2(\boldsymbol{\theta})\sigma^2} - \left(\frac{-1}{2t(\boldsymbol{\theta})\sigma^2}\right)^2 \left(\frac{1}{4\sigma^2} + \frac{1}{2\sigma^4}\right)^{-1} \right|^{\frac{1}{2}} = \frac{1}{t(\boldsymbol{\theta})}\left(\frac{1}{\sigma^2} - \frac{1}{2+\sigma^2}\right)^{\frac{1}{2}}$$

and

$$h_2^{\frac{1}{2}} = \left[\frac{1}{2\sigma^2}\left(\frac{1}{2} + \frac{1}{\sigma^2}\right)\right]^{\frac{1}{2}}.$$

Therefore, the reference prior relative to the ordered parameterisation $(t(\boldsymbol{\theta}),\sigma^2)$ is given by

$$p_R(t(\boldsymbol{\theta}),\sigma^2) \propto \frac{1}{t(\boldsymbol{\theta})}\frac{1}{\sigma}\sqrt{1 + \frac{2}{\sigma^2}}.$$

In the (μ,σ^2) parameterisation this corresponds to

$$p_R(\mu,\sigma^2) \propto \frac{1}{t(\boldsymbol{\theta})}\frac{1}{\sigma}\sqrt{1 + \frac{2}{\sigma^2}}\,(t(\boldsymbol{\theta})) \propto \frac{1}{\sigma}\sqrt{1 + \frac{2}{\sigma^2}}.$$

This is the same result derived by Roman (2008).

BIBLIOGRAPHY

Berger, J.O., Bernardo, J.M. (1992). "On the development of the Reference Prior Method," *Bayesian Statistics 4: Proceedings of the Fourth Valencia Meeting.* Eds. Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. Oxford, U.K., Clarendon Press, pp. 35 – 60.

Box, G.E.P., Tiao, G.C. (1973). "Bayesian Inference in Statistical Analysis," *Reading, MA*, Addison-Wesley.

Datta, G.S., Ghosh, J.K. (1995). "On priors providing Frequentist Validity of Bayesian Inference," *Biometrika*, 82, pp. 37 – 45.

Krishnamoorthy, K., Mathew, T. (2003). "Inferences on the Means of Lognormal Distributions using Generalized P-values and Generalized Confidence Intervals." *Journal of Statistical Planning and Inference*, 115, pp. 103 - 121.

Jeffreys, H. (1967). "Theory of Probability," *Oxford University Press*, London.

Pearn, W.L., Wu, C.W. (2005). "Process Capability Assessment for Index C_{pk} based on the Bayesian Approach," *Metrika*, 61, p. 211

Rice, A.J. (1995). "Mathematical Statistics and Data Analysis – 2nd ed.," *Duxbury Press*, Belmont, California.

Roman, L.P. (2008). "Reference prior distributions for prediction problems," *Current technical report*.

Shapiro, S.S., Wilk, M.B. (1965). "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika*, 52: pp. 591 - 611.

Starfield, B., Weiner, J., Mumford, L., Steinwachs, D. (1991). "Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management." *Health Serv. Res*. 26: pp. 53 - 74.

Tian, L., (2005). "Inferences on the Mean of Zero-inflated Lognormal Data: the Generalized Variable Approach." *Statistics in Medicine*, pp. 3223 - 3232.

Tian, L., Wu, J., (2006). "Confidence Intervals for the Mean of Lognormal Data with Excess Zeros." *Biometrical Journal*, 48 no. 1: pp. 149 - 156.

Weiner, J., Starfield, B., Mumford, L. (1991) "Development and Application of a Population-Oriented Measure of Ambulatory Care Casemix." *Medical Care*. 29: pp. 452 - 472

Zellner, A. (1971). "An Introduction to Bayesian Inference in Econometrics," *Wiley*, New York, NY.

Zhou, X.H., Gao, S. (1997). "Confidence intervals for the lognormal mean," *Statistical Medicine*, 16, 783 – 790.

Zhou, X.H., Gao, S., Hui, S.L. (1997). "Methods for comparing the means of two independent lognormal samples," *Biometrics*, 53, pp. 1129 – 1135.

Zhou, X.H., Melfi, C.A., Hui, S.L. (1997b). "Methods for Comparison of Cost Data." *Annal of Internal Medicine*, 127: pp. 752 - 756.

Zhou, X.H., Tu, W. (1999). "Comparison of Several Independent Population Means when their Samples Contain Lognormal and Possibly Zero Observations" *Biometrics*, 55, pp. 645 - 651.

Zhou, X.H., Tu, W. (2000). "Interval estimation for the ratio in means of log-normally distributed medical costs with zero values," *Computational Statistics and Data Analysis*, 35, pp. 201 – 210.