South African Statistical Journal
2018, Vol. 52, No. 1, 29–55

A BAYESIAN APPROACH TO INFERENCE ON THE VARIANCE OF LOGNORMAL DATA

Justin Harvey [1]
Department of Statistics and Actuarial Science, University of Stellenbosch
e-mail: *jharvey@sun.ac.za*

Abrie J. van der Merwe
Department of Mathematical Statistics and Actuarial Science, University of the Free State

Krishnamoorthy, Mathew and Ramachandran (2006) developed a method to draw inference on the mean and variance of one or more lognormal distributions. Their method was based on generalised confidence intervals (frequentist methods). In this article we focus on the variance of the lognormal distribution and implement a Bayesian approach and obtain credibility intervals to compare the performance of four different non-informative prior distributions. This is done by means of various Monte Carlo simulation studies as well as practical examples. The accuracy (coverage) and efficiency (interval length) of some of the Bayesian priors, particularly for the highest posterior density (HPD) credibility intervals will be illustrated in these simulation studies and examples. It can be observed that the frequentist approach is equivalent to the Bayesian approach, when using the Independence Jeffreys prior. Even so, the Bayesian approach offers some additional benefits, namely, through the calculation of the HPD intervals. Hypothesis testing and practical applications are also presented. Further results comparing various estimators of the lognormal variance are derived and evaluated. The usefulness of the Bayesian approach is also illustrated in its ability to easily modify the method to account for the possibility of zero-valued observations. This is something for which there is (to our knowledge) currently no frequentist method available and serves to highlight the usefulness of the Bayesian approach.

Key words: Bayes, Lognormal variance, Monte Carlo simulation, Non-informative prior.

1. Introduction

Lognormally distributed data occur frequently in practice. For example, in studies of the treatment of the human immunodeficiency virus (HIV) important variables, such as the viral load, are often found to be normally distributed after taking an appropriate log-transformation (Chu, Gange, Li, Hoover, Liu, Chmiel and Jacobson, 2010). In addition, it has been established that occupational exposure data generally follow this distribution (Krishnamoorthy et al., 2006). The dust level in coal mines was found to be approximately lognormally distributed by Oldham (1953) and since then this distribution has been used to describe various other types of workplace exposure data to harmful pollutants (refer to Krishnamoorthy et al., 2006, for additional references).

However, the application is not limited to only HIV and occupational exposure setting. Variables such as medical costs arising from patient care have been found to be lognormally distributed. Both

[1] Corresponding author.
MSC2010 subject classifications. 62C10, 62F15.

McDonald, Blemis, Tierny and Martin (1988) and Zhou, Gao and Hui (1997) analysed data on the effects of race on medical costs for patients with Type I diabetes.

In many of these cases, the objective is to compare the mean level of the data (e.g. the mean level of the pollutant or medical costs). However, in some situations it may be of interest to draw inference on the variance of the distribution (Krishnamoorthy et al., 2006). These are situations where we would like to compare the distribution of the data rather than simply the mean values from two independent data sets (the case of dependent lognormal distributions was discussed in Harvey and van der Merwe, 2012). According to Krishnamoorthy et al. (2006) there are no readily available procedures available for computing confidence intervals for the variance. They proposed methods based on generalised p-values and generalised confidence intervals for addressing this problem. The authors in particular applied the procedures they developed to address situations involving patient and worker exposure data.

In this paper we look at a similar situation in terms of the distribution of the data. We are primarily concerned with Bayesian inference for the variance of the distribution as opposed to the mean of the distribution. In Bayesian analysis the choice of the prior distribution used is important in analyzing the data. In this paper we will propose the use of a number of non-informative prior distributions, such as the Independence Jeffreys prior, the Jeffreys Rule prior, the probability-matching prior as well as the reference prior. A simulation study is undertaken to examine the performance of these various prior distributions. Credibility intervals (Bayesian confidence intervals) will be developed based on these different choices of prior distributions. In this way we present an alternative to the frequentist method (generalised confidence interval). A few basic examples will be given to illustrate the use of the Bayesian methods.

Another problem that is often encountered in real-life data sets is the possibility of zero-valued data. It is not possible for the lognormal distribution to have zero-valued observations, but in practical settings the situation may arise where some zero-valued observations are present and the non-zero-valued observations follow a lognormal distribution. For the assessment of the extent of variability among health care costs or among exposure measurements, confidence intervals or tests concerning the variance of lognormally distributed data with zero-valued observations ($\widetilde{\sigma}^2$) becomes necessary. Krishnamoorthy et al. (2006) made inference about the lognormal variance, while Bebu and Mathew (2008) obtained confidence intervals for the ratio of variances in the case of the bivariate lognormal distribution. However, as far as we know no procedures are known for computing confidence intervals for $\widetilde{\sigma}^2$.

This article begins with a description of the setting and the generalised confidence interval method proposed by Krishnamoorthy et al. (2006). This is followed by an introduction to the Bayesian analysis of the problem and a discussion of the non-informative prior distributions. A few simulation studies are then performed to compare the performance of the various prior distributions to the frequentist approach. Some small examples are then presented to illustrate the application of the method. The case of the zero-valued observations will then be described and implemented in order to show the flexibility of the Bayesian approach.

Table 1. Definitions.

$x_1 \ldots x_n$	Sample from a lognormal distribution
$y_1 \ldots y_n$	Logged data; $y_i = \ln(x_i)$, $i = 1, \ldots, n$
μ_l	Population mean of the logged data
σ_l	Population standard deviation of the logged data
μ	Mean of the lognormal distribution $\mu = \exp(\mu_l + \frac{1}{2}\sigma_l^2)$
σ^2	Variance of the lognormal distribution $\sigma^2 = \exp(2\mu_l + \sigma_l^2)[\exp(\sigma_l^2) - 1]$
$\overline{y}$	Sample mean of the logged data, the observed value of $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$
s^2	Sample standard deviation of the logged data, the observed value of $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

2. Description of the setting and the generalised p-value and confidence interval approach

We will use the same setting, as described in Krishnamoorthy et al. (2006). Let X denote the lognormally distributed measurement. Then

$$Y = \ln(X) \sim N\left(\mu_l,\ \sigma_l^2\right).$$

The mean of the lognormal distribution (say μ) is then given by

$$\mu = \exp(\eta) = \exp\left(\mu_l + \frac{1}{2}\sigma_l^2\right)$$

and the variance by

$$\sigma^2 = \exp(2\mu_l + \sigma_l^2)\left[\exp(\sigma_l^2) - 1\right].$$

Table 1 supplies the relevant definitions for the setting (as used in Krishnamoorthy et al., 2006). The aim is to determine the amount of variation in the measurements. To do this will require the calculation of confidence (or credibility) intervals. For the generalised confidence interval approach this is determined by calculating a generalised pivot statistic. According to Krishnamoorthy et al. (2006) this can be defined as:

$$T_{\sigma^2} = \exp\left(2\left(\overline{y} - \frac{Z}{V\sqrt{n-1}}\frac{s}{\sqrt{n}}\right) + \frac{s^2}{V^2/(n-1)}\right)\left[\exp\left(\frac{s^2}{V^2/(n-1)}\right) - 1\right],$$

where the independent random variables Z and V are defined as

$$Z = \frac{\sqrt{n}\left(\overline{Y} - \mu_l\right)}{\sigma_l} \sim N(0,1), \qquad V^2 = \frac{(n-1)S^2}{\sigma_l^2} \sim \chi_{n-1}^2.$$

It can be noted (when comparing to later relevant sections) that the above method is equivalent to the Bayesian approach, specifically when using the Independence Jeffreys prior distribution. The generalised confidence interval approach seems to offer a simpler approach to implement than the Monte Carlo methods that are required for the Bayesian approach. Nevertheless, the Bayesian approach is useful in that it allows the calculation of HPD intervals. The HPD intervals offer substantial improvements in coverage and interval length, as will be illustrated in later sections.

Using these definitions, Krishnamoorthy and Mathew (2003) gave the following algorithm for computing the generalised confidence interval:

1. For a given data set of logged observations, determine $\overline{y}$ and s^2.
2. For $i = 1$ to 10 000:
   (a) Generate $Z \sim N(0,1)$.
   (b) Generate $V^2 \sim \chi^2_{n-1}$.
   (c) Set $T_{\sigma^2 i} = \exp\left(2(\overline{y} - \frac{Z}{V\sqrt{n-1}}\frac{s}{\sqrt{n}}) + \frac{s^2}{V^2/(n-1)}\right)\left[\exp\left(\frac{s^2}{V^2/(n-1)}\right) - 1\right]$.
3. The $100(1-\alpha)$th percentile of $T_{\sigma^2 1}, \ldots, T_{\sigma^2 10\,000}$, denoted by $T_{\sigma^2, 1-\alpha}$, is the generalised upper confidence limit for σ^2.

3. Description of the Bayesian approach and derivation of prior distributions

Given the previous description of the setting, the likelihood function can be written as

$$L\left(\mu_l, \sigma_l^2\right) \propto \left(\sigma_l^2\right)^{-\frac{1}{2}\upsilon} \exp\left\{-\frac{\upsilon s^2}{2\sigma_l^2}\right\}\left(\frac{n}{2\sigma_l^2 \pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{n(\mu_l - \overline{y})^2}{2\sigma_l^2}\right\},$$

that is

$$L\left(\mu_l, \sigma_l^2\right) \propto L\left(\sigma_l^2\right) L\left(\mu_l | \sigma_l^2\right), \tag{1}$$

where

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad \upsilon s^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2$$

and $\upsilon = n - 1$.

We are interested in Bayesian confidence intervals (i.e. credibility intervals) for the variance of the distribution, σ^2.

3.1 The Independence Jeffreys prior

Consider the first prior distribution: $p_{IJ}(\mu_l, \sigma_l^2) \propto \sigma_l^{-2}$. According to Box and Tiao (1973) in most cases it is appropriate to take location parameters to be distributed independently of scale parameters. Using the argument in Section 1.3.2 of Box and Tiao (1973) the above prior distribution follows.

Combining this prior distribution with the likelihood given by (1) results in the following posterior distribution:

$$p\left(\mu_l, \sigma_l^2 \,|\, \boldsymbol{y}\right) = p_I\left(\sigma_l^2 \,|\, \boldsymbol{y}\right) p\left(\mu_l \,|\, \sigma_l^2, \boldsymbol{y}\right),$$

where

$$p_I\left(\sigma_l^2 | \boldsymbol{y}\right) = \left(\frac{\upsilon s^2}{2}\right)^{\frac{1}{2}\upsilon} \frac{1}{\Gamma\left(\frac{\upsilon}{2}\right)}\left(\sigma_l^2\right)^{-\frac{1}{2}(\upsilon+2)} \exp\left\{-\frac{\upsilon s^2}{2\sigma_l^2}\right\} \tag{2}$$

for $\sigma_l^2 > 0$, which is an inverse gamma distribution and

$$\mu_l \,|\, \sigma_l^2, \boldsymbol{y} \sim N\left(\overline{y}, \frac{1}{n}\sigma_l^2\right).$$

From (2) it follows that

$$\frac{\upsilon s^2}{\sigma_l^2} \sim \chi^2_\upsilon.$$

To obtain credibility intervals for this Bayesian procedure Monte Carlo simulation is applied and is described next.

3.2 Simulation procedure

Based on the previous derivations, the following algorithm was implemented in R to simulate from the posterior distribution. For given μ_l, σ_l^2 and n the procedure is as follows:

1. Since σ_l^2 is known, in accordance with Krishnamoorthy and Mathew (2003), set $\mu_l = -\frac{1}{2}\sigma_l^2$, or as specified by the simulation study or data set.

2. Calculate the following: $\widehat{\mu}_l = \overline{y}$ and $m = \nu\widehat{\sigma}_l^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2$, where $\nu = n - 1$. This can either be calculated from given data or can be simulated (in the case of a simulation study). In the latter case, since we are only interested in the sufficient statistics these can be simulated directly, namely: $\overline{Y} \sim N(\mu_l, \sigma_l^2/n)$ and $m/\sigma_l^2 \sim \chi^2_\nu$ and therefore, $m = \sigma_l^2(\chi^2_\nu)$.

3. Given these calculated values we can simulate μ_l and σ_l^2 from their their posterior distributions as follows:

   (a) $\sigma_l^2 = m/\chi^2_\nu$.

   (b) $\mu_l|\sigma_l^2 \sim N(\widehat{\mu}_l, \sigma_l^2/n)$.

   (c) $\eta = \mu_l + \frac{1}{2}\sigma_l^2$.

   (d) For this experiment/sample, simulate 10 000 values of η.

4. Sort them in ascending order such that $\eta^*_{(1)} \leq \eta^*_{(2)} \leq \cdots \leq \eta^*_{(10\,000)}$.

5. Let $K_1 = [\frac{\alpha}{2} \times 10\,000]$ and $K_2 = [(1 - \frac{\alpha}{2}) \times 10\,000]$, where $[a]$ denotes the largest integer not greater than a.

6. $\{\eta^*_{(K_1)}, \eta^*_{(K_2)}\}$ is then a $100\,(1-\alpha)\,\%$ Bayesian confidence interval for η.

7. Repeat the procedure for 10 000 experiments (in the case of a simulation study).

Using these methods we can simulate μ_l and σ_l^2 from their respective posterior distributions and obtain a 10 000 values of $\sigma^2 = \exp(2\mu_l + \sigma_l^2)\,[\exp(\sigma_l^2) - 1]$. These can then be used to obtain credibility intervals for σ^2.

An additional advantage of the Bayesian approach is the calculation of highest posterior density (HPD) credibility intervals. It is possible to form many credibility intervals from observations simulated from the posterior distribution. The HPD interval is then the credibility interval among all possibilities that results in the shortest length, while still retaining the required credibility level. HPD intervals are therefore, not possible with the generalised confidence interval approach.

3.3 Other prior distributions

Other non-informative prior distributions were also derived and the procedure was repeated. The simulation procedure is similar to that for posterior distribution (2) except that (given the choice of prior distributions) σ_l^2 will have a different posterior distribution.

3.3.1 Jeffreys Rule prior

The Jeffreys Rule prior is the square root of the determinant of the Fisher Information matrix and is given as $p_{JR}(\sigma_l^2) \propto \sigma_l^{-3}$. A complete derivation can be found in Harvey (2012). In this case the we could simulate from the posterior distribution of σ_l^2 by noting that $\sigma_l^2 = m/\chi^2_{\nu+1}$.

3.3.2 Probability-matching prior distribution

In addition to the two previously mentioned priors by Jeffreys, both the reference and probability-matching prior distributions were derived.

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 $o(n^{-1})$, 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 following 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,

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

and

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

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 probability-matching prior for σ^2 is given by (refer to Harvey, 2012)

$$p_M\left(\mu_l, \sigma_l^2\right) \propto \sigma_l^{-3} \sqrt{\frac{2\left\{\exp(\sigma_l^2) - 1\right\}^2}{\left\{2\exp(\sigma_l^2) - 1\right\}^2} + \sigma_l^2}. \tag{3}$$

The derivation is given in the appendix to this article. From (1) it follows that if we multiply (1) by (3) then we have the following posterior distribution:

$$p_M\left(\mu_l, \sigma_l^2 \,|\, \boldsymbol{y}\right) \propto p_M\left(\sigma_l^2 \,|\, \boldsymbol{y}\right) \times p_M\left(\mu_l \,|\, \sigma_l^2, \boldsymbol{y}\right),$$

where $\mu_l|\sigma_l^2, \boldsymbol{y} \sim N(\overline{y}, \sigma_l^2/n)$ and

$$p_M\left(\sigma_l^2 \,|\, \boldsymbol{y}\right) \propto L\left(\sigma_l^2\right) \times p_M\left(\mu_l, \sigma_l^2\right). \tag{4}$$

Using these results one is able to simulate from the posterior distribution by first simulating σ_l^2 from (4) using the Acceptance-Rejection method as described in Rizzo (2007).

3.3.3 Reference prior distribution

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 reference prior for σ^2 is given by (refer to Harvey, 2012):

$$p_R\left(\mu_l, \sigma_l^2\right) \propto \sigma_l^{-1}\sqrt{\left\{\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right\}^2 + \frac{2}{\sigma_l^2}}.$$

The derivation is available in the appendix to this article. The simulation procedure will be similar to that of the probability-matching prior.

In summary, the use of these four prior distributions results in the following posterior distributions of σ_l^2 as previously described for the Independence Jeffreys prior distribution:

$$p_{IJ}\left(\sigma_l^2 \,\middle|\, \boldsymbol{y}\right) \propto \frac{\left(\frac{\upsilon s}{2}\right)^{\frac{\upsilon}{2}} \sigma_l^{-\frac{1}{2}(\upsilon+2)} \exp\left(-\frac{\upsilon s}{2\sigma_l}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)},$$

$$p_{JR}\left(\sigma_l^2 \,\middle|\, \boldsymbol{y}\right) \propto \frac{\left(\frac{\upsilon s}{2}\right)^{\frac{\upsilon}{2}} \sigma_l^{-\frac{1}{2}(\upsilon+3)} \exp\left(-\frac{\upsilon s}{2\sigma_l}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)},$$

$$p_M\left(\sigma_l^2 \,\middle|\, \boldsymbol{y}\right) \propto \sigma_l^{-\frac{1}{2}\upsilon} \exp\left(-\frac{\upsilon s}{2\sigma_l}\right) \sigma^{-3}\sqrt{\frac{2\left\{\exp(\sigma_l^2)-1\right\}^2}{\left\{2\exp(\sigma_l^2)-1\right\}^2} + \sigma_l^2}$$

and

$$p_R\left(\sigma_l^2 \,\middle|\, \boldsymbol{y}\right) \propto \sigma^{-\frac{1}{2}\upsilon} \exp\left(-\frac{\upsilon s}{2\sigma_l}\right) \sigma_l^{-1}\sqrt{\left\{\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right\}^2 + \frac{2}{\sigma_l^2}}.$$

Table 2. Results for various prior distributions.

	Equal-tail intervals			HPD intervals		
Prior	**Lower**	**Upper**	**Length**	**Lower**	**Upper**	**Length**
Independence Jeffreys	3.7246	554.497	550.772	1.5494	229.647	228.097
Jeffreys Rule	3.3337	253.727	250.394	1.5496	120.675	119.125
PMP	2.8121	85.472	82.660	1.4988	50.461	48.962
REF	3.2913	322.017	318.725	1.5394	146.464	144.925

4. Example 1 – Single sample

In this example the following was done:

1. Take the following initial values: $n = 10$, $\nu s^2 = 6$, $\overline{y} = 1$.
2. Simulate from the four posterior distributions (as previously described).
3. Simulate $t(\mu_l, \sigma_l^2) = Var(X)$ for all four prior distributions above.
4. Calculate 95% credibility intervals (equal-tailed and HPD).
5. Calculate the generalised confidence interval, according to the previously-described method.

For the generalised confidence intervals the methods (as previously described) are only comparable to the equal-tailed confidence intervals in the Bayesian context.

The results are given in Table 2. In terms of interval length, we can see from the above results that the Bayesian prior distribution that performs the best is the Jeffrey's Rule prior. The probability-matching prior appears to have the shortest interval length, however, as will be shown in later results this prior distribution suffers from insufficient coverage. A more complete simulation study (including interval coverage) will be presented later in this article, but the above examples serves to illustrate the usefulness and accuracy of the Bayesian methodology.

The difference between the various prior distributions (and thus the reason for the difference in performance) is based on the form of the posterior distributions, more specifically, the $p(\sigma_l^2|\boldsymbol{y})$ portion of the posterior. Figure 1 illustrates the posterior distributions for the different prior distributions:

From Figure 1 we can see that the two Jeffreys priors are similar. The probability-matching prior results in a distribution with less weight in the tail region. This explains the credibility intervals with a shorter length but less than adequate coverage, as will be shown later in the simulation study. With particular reference to the two Jeffreys priors it is evident that the Independence Jeffreys prior has a heavier distribution tail. This would explain the better coverage (as seen in the previous example and the simulation study) and the wider interval lengths of the Independence Jeffreys prior.

5. Simulation study examining coverage

In addition to the above example a simulation study was performed for a wider range of parameter settings. These parameter values were chosen arbitrarily and are outlined in Table 3. In each case 10 000 samples (supplying the sufficient statistics $\overline{y}$ and νs^2) and for each sample 10 000 simulated observations from the posterior distribution (in the Bayesian case) were used. The results are

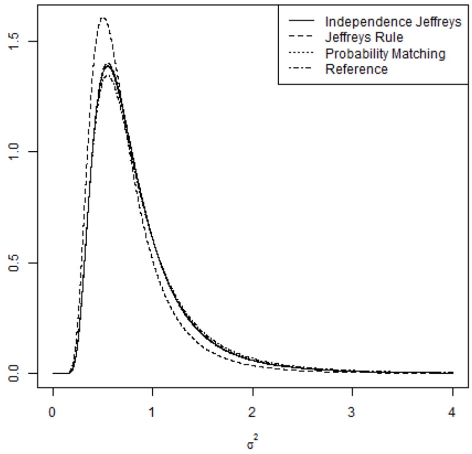

Figure 1. Posterior distributions of σ^2 under different choices of priors.

Table 3. Parameter values.

Parameters	**Parameter values chosen**
μ_l	0; 1; 2
σ_l^2	0.25; 2.25
n	10; 20

presented in Table 4. From Table 4 the coverage of the Bayesian methodology is adequate (at least for the Independence Jeffreys and Jeffreys Rule prior distributions). The probability-matching prior in particular, however, suffers from substantial undercoverage, even though it results in the shortest interval lengths. This may be due to the fact that the probability-matching prior was developed for use in one-sided credibility intervals. An advantage of the Bayesian approach is the determination of HPD credibility intervals. It can be seen that the coverage is better for these HPD intervals (as opposed to the equal-tail intervals) and particularly so for the larger sample size.

In terms of interval length, the shortest lengths are observed for the probability-matching prior and, to a lesser extent, the reference prior. However, as previously observed, the coverage of these intervals is not sufficient. The HPD intervals calculated for the Bayesian approach reduce the interval length, while improving coverage for the Independence Jeffreys prior distribution.

In summary, it appears as though the simplest prior distribution, the Independence Jeffreys approach, through the calculation of HPD intervals results in the most efficient intervals with the requisite coverage. Not all of these non-informative prior distributions are similar in performance though.

6. Example 2 – An application to hypothesis testing

As an illustration of the application of the above, the following simple example is given, which is an extension of Example 1. Suppose we wish to compare the variances from two lognormal distributions

Table 4. Simulation study results.

Method	σ_l^2	μ_l	Coverage / length	EQ		HPD	
				$n = 10$	$n = 20$	$n = 10$	$n = 20$
Independence Jeffreys	0.25	0	Coverage	94.7000	94.7600	94.8700	94.9400
Jeffreys Rule				93.7800	94.4900	91.2700	93.5000
PMP				87.7600	91.5800	81.5600	87.9800
REF				92.4800	93.8400	89.2000	92.1000
Independence Jeffreys	0.25	0	Length	5.4988	1.1972	2.8917	0.9517
Jeffreys Rule				3.0271	1.0168	1.8300	0.8239
PMP				1.3341	0.7757	0.9462	0.6447
REF				3.0992	1.0063	1.7988	0.8107
Independence Jeffreys	0.25	2	Coverage	95.3200	94.5500	95.0300	95.0600
Jeffreys Rule				94.1300	94.7800	91.7700	93.3100
PMP				87.8700	91.3200	81.7200	87.5200
REF				92.6000	93.5500	89.5000	91.6700
Independence Jeffreys	0.25	2	Length	296.0200	64.8010	156.5900	51.5260
Jeffreys Rule				162.9400	55.5430	99.0710	45.0290
PMP				75.2320	41.7820	52.9360	34.7870
REF				179.3900	54.2280	102.9100	43.7200
Independence Jeffreys	1	0	Coverage	94.8600	95.0100	95.0900	95.1300
Jeffreys Rule				94.0400	94.3200	91.7700	92.6100
PMP				86.7200	91.5500	80.7200	87.1200
REF				93.1700	94.5500	91.4500	93.3700
Independence Jeffreys	1	0	Length	2.5100×10^7	277.4500	2.5600×10^5	112.9800
Jeffreys Rule				4.2900×10^6	181.9100	5.5600×10^4	78.5340
PMP				840.1100	66.0740	209.7300	35.6650
REF				1.9200×10^4	221.9200	4.0335×10^4	95.2960
Independence Jeffreys	1	2	Coverage	94.9100	95.0800	94.8700	95.3000
Jeffreys Rule				94.1800	94.3100	92.1600	92.8400
PMP				87.2000	92.0700	80.9500	87.2000
REF				93.0700	94.6700	91.3000	93.6500
Independence Jeffreys	1	2	Length	3.9400×10^8	1.4400×10^4	6.9100×10^6	6.0368×10^3
Jeffreys Rule				1.7400×10^8	8.7958×10^3	2.1600×10^6	4.0086×10^3
PMP				7.8900×10^4	3.5034×10^3	2.0700×10^4	1.8866×10^3
REF				1.1800×10^6	1.2500×10^4	2.6400×10^5	5.0902×10^3
Independence Jeffreys	2.25	0	Coverage	95.0700	95.2600	95.1700	95.3300
Jeffreys Rule				93.4700	94.1900	90.9300	92.8600
PMP				87.6200	91.0000	81.5500	86.7700
REF				94.6500	94.1200	92.9500	93.7200
Independence Jeffreys	2.25	0	Length	5.5000×10^{22}	3.5000×10^8	1.0000×10^{18}	8.9100×10^6
Jeffreys Rule				1.3000×10^{16}	1.2400×10^7	3.1000×10^{12}	9.6600×10^5
PMP				2.9400×10^5	5.1200×10^4	8.9800×10^4	1.8700×10^4
REF				1.4200×10^6	1.8800×10^5	4.9000×10^5	7.0400×10^4
Independence Jeffreys	2.25	2	Coverage	94.8700	95.1500	95.1000	94.9000
Jeffreys Rule				94.0800	94.3400	91.7400	92.5700
PMP				86.9700	91.8700	80.5700	87.0200
REF				94.2200	94.8000	92.6700	94.3000
Independence Jeffreys	2.25	2	Length	5.3000×10^{24}	2.8900×10^9	3.0000×10^{19}	1.2600×10^8
Jeffreys Rule				2.6000×10^{21}	5.4800×10^9	9.6000×10^{16}	2.9400×10^8
PMP				1.6700×10^7	2.4200×10^6	5.2400×10^6	8.3100×10^5
REF				7.7500×10^7	9.5200×10^6	2.6900×10^7	3.4400×10^6

Table 5. 95% credibility intervals for the ratio $\psi_1 = Var(X_1)/Var(X_2)$.

	Equal-tail intervals			HPD intervals		
Prior	**Lower**	**Upper**	**Length**	**Lower**	**Upper**	**Length**
Independence Jeffreys	0.8834	181.918	181.035	0.1828	74.491	74.308
Jeffreys Rule	0.8313	87.565	86.733	0.1694	42.670	42.500
PMP	0.7599	32.773	32.013	0.2277	19.478	19.250
REF	0.8412	113.374	112.553	0.1748	49.976	49.801

Table 6. 95% credibility interval for the difference $\psi_2 = Var\,(X_1) - Var\,(X_2)$.

	Equal-tail intervals			HPD intervals		
Prior	**Lower**	**Upper**	**Length**	**Lower**	**Upper**	**Length**
Independence Jeffreys	−0.5419	568.080	568.622	−9.1488	225.357	234.506
Jeffreys Rule	−0.8138	256.634	257.448	−6.2640	123.744	130.008
PMP	−1.0809	82.061	83.142	−4.2851	49.650	53.936
REF	−0.7407	322.310	323.051	−7.8606	141.330	149.191

in some way. In other words, we would like to test the following hypothesis in some way:

$$H_0 : \sigma^2_{(1)} = \sigma^2_{(2)}$$
$$H_1 : \sigma^2_{(1)} \neq \sigma^2_{(2)}.$$

This could be performed by simulating observations from the two distributions and determine either the ratio or difference between the variances. Intervals (confidence or credibility) could then be calculated and interpreted accordingly. For example, if the interval for the ratio includes 1 then the variances do not differ significantly from each other and similarly if the interval for the difference includes the point 0.

To illustrate this, a similar methodology was repeated to obtain samples from each distribution (refer to Example 1 for a detailed explanation). The objective is to calculate credibility intervals for the difference between the sample variances or the ratio between the sample variances. In so doing one is actually testing the hypothesis that the variances are equal. In this example the following was done:

1. Take the following initial values: $n_1 = 10$, $\nu_1 s_1^2 = 6$, $\overline{y}_1 = 1$, and for the second sample $n_2 = 30$, $\nu_2 s_2^2 = 8$, $\overline{y}_2 = 1$.

2. Using the prior distributions described previously simulate the posterior distribution of:

   (a) $\psi_1 = Var(X_1)/Var(X_2)$,

   (b) $\psi_2 = Var\,(X_1) - Var(X_2)$.

The results are given in Tables 5 and 6. Note that in this example, all methods agree that there is no difference between the two populations. We can see a similar trend as was identified in the simulation study previously. The Independence Jeffreys prior results in the largest interval length, but the use of HPD intervals minimises this as a disadvantage. In particular, all HPD intervals

Table 7. Summary statistics of medical charges data.

Data	Group	Sample mean	Sample variance
Original	African American	\$18 850	26897^2
	White	\$18 584	30694^2
Logged	African American	9.06694	1.824
	White	8.69306	2.692

Table 8. 95% credibility intervals for the ratio $\psi_1 = Var(X_1)/Var(X_2)$.

	Equal-tail intervals			HPD intervals		
Prior	Lower	Upper	Length	Lower	Upper	Length
Independence Jeffreys	0.043 386 50	1.934 705	1.891 318	0.005 331 481	1.468 957	1.463 625
Jeffreys Rule	0.045 771 65	1.949 739	1.903 967	0.010 692 140	1.476 621	1.465 928
PMP	0.042 889 35	1.923 892	1.881 003	0.003 103 380	1.472 247	1.469 143
REF	0.042 056 40	1.987 440	1.945 384	0.004 123 816	1.503 702	1.499 578

decrease the interval length considerably for both the ratio and difference between the variances. It seems that the probability-matching and reference priors give the best results, but again, it is worth remembering (from the simulation study) that this may come at a risk of insufficient coverage. For the probability-matching prior, this may be due to the fact that the prior was developed for use in one-sided credibility intervals.

7. Example 3 – Medical charges

The following example comes from Gupta and Li (2006) and concerns the effect of race on medical charges or expenses. The population consisted of patients with Type I diabetes who had received treatment on two or more occasions between 1 January 1993 and 30 June 1994. One hundred and nineteen (119) African American and 106 white patients were sampled and their medical charges were obtained. In both groups, it was found (using the Shapiro-Wilks test) that the log-transformed data were normally distributed at a 5% significance level ($p = 0.15$ and $p = 0.12$ for African American and white patients respectively). The summary statistics from the sample data are given in Table 7.

Gupta and Li (2006) found that the means did not differ significantly from each other. However, upon further analysis it was found that the variances between the groups were not equal. Here we apply the Bayesian methods to this problem of comparisons of the variances of the medical charges.

The results are given in Tables 8 and 9, for both the difference and ratio of the variances.

All Bayesian priors in Tables 8 and 9 agree that there is no difference between the variance in medical charges for the two patient groups, which is in contrast to what was observed by Gupta and Li (2006).

8. Zero values

For the assessment of the extent of variability among health care costs or among exposure measurements, confidence intervals or tests concerning the variance $\widetilde{\sigma}^2$ of lognormally distributed data

Table 9. 95% credibility interval for the difference $\psi_2 = Var\,(X_1) - Var\,(X_2)$.

	Equal-tail intervals			HPD intervals		
Prior	**Lower**	**Upper**	**Length**	**Lower**	**Upper**	**Length**
Independence Jeffreys	-4.47×10^{10}	2.91×10^{9}	4.76×10^{10}	-3.30×10^{10}	6.16×10^{9}	3.92×10^{10}
Jeffreys Rule	-4.07×10^{10}	2.73×10^{9}	4.35×10^{10}	-3.04×10^{10}	5.42×10^{9}	3.58×10^{10}
PMP	-4.40×10^{10}	2.93×10^{9}	4.70×10^{10}	-3.24×10^{10}	6.15×10^{9}	3.86×10^{10}
REF	-4.73×10^{10}	3.06×10^{9}	5.04×10^{10}	-3.52×10^{10}	6.45×10^{9}	4.17×10^{10}

with zero-valued observations becomes necessary. Krishnamoorthy et al. (2006) made inference about the lognormal variance, while Bebu and Mathew (2008) obtained confidence intervals for the ratio of variances in the case of the bivariate lognormal distribution. However, as far as we know no procedures are known for computing confidence intervals for $\widetilde{\sigma}^2$. This further illustrates the usefulness of the Bayesian methodology.

Non-zero observations are assumed to follow a lognormal distribution, but the inclusion of the possibility of zero values necessitates the addition of a binomial parameter to model this possibility. The Independence Jeffreys and Jeffreys Rule prior distributions were once again analysed in this situation. What follows is a brief description of the setting and then the corresponding results.

Let

$$f(x) = \begin{cases} \delta & \text{for } x = 0, \\ (1-\delta)\,\frac{1}{\sigma_l\sqrt{2\pi}x}\exp\left(-\frac{(\ln x - \mu_l)^2}{2\sigma_l^2}\right) & \text{for } x > 0. \end{cases}$$

Then,

$$E\,(X) = (1-\delta)\exp\left(\mu_l + \frac{1}{2}\sigma_l^2\right),$$
$$E\left(X^2\right) = (1-\delta)\exp\left(2(\mu_l + \sigma_l^2)\right)$$

and

$$Var\,(X) = (1-\delta)\exp\left(2(\mu_l + \sigma_l^2)\right)\left\{\exp(\sigma_l^2) - (1-\delta)\right\}.$$

Note that δ is the probability of observing a zero-valued observation and $0 \leq \delta \leq 1$. The non-zero-valued observations are then assumed to be lognormally distributed, where the logged observations are normally distributed with mean μ_l and variance σ_l^2. A more complete specification of the resulting likelihood can be found in Harvey (2012). The following prior distributions were used:

- Independence Jeffreys prior:

$$p_I(\delta)p_I(\mu_l\,,\,\sigma_l^2) \propto \delta^{-\frac{1}{2}}(1-\delta)^{-\frac{1}{2}}\sigma_l^{-2}.$$

- Jeffreys Rule prior:

$$p_D(\delta)p_D(\mu_l\,,\,\sigma_l^2) \propto \delta^{-\frac{1}{2}}(1-\delta)^{\frac{1}{2}}\sigma_l^{-3}.$$

In the above prior distributions, $p_I(\delta) \propto \delta^{-1/2}(1-\delta)^{-1/2}$ is the prior proposed by Jeffreys (1961) for the binomial parameter. It can be observed (refer to Harvey (2012) for more detail) that the posterior distribution of δ is a beta distribution and is independently distributed of μ_l and σ_l^2.

Table 10. Parameter settings for the simulation study.

Parameters	Parameter values chosen
δ	0.1; 0.2; 0.3
μ_l	0; 2
σ_l^2	0.25; 1; 2.25
n	10 (except where δ = 0.3); 20; 50

Specifically, this beta distribution is given as $B(n_0 + \frac{1}{2},\ n_1 + \frac{1}{2})$ for the Independence Jeffreys prior and $B(n_0 + \frac{1}{2},\ n_1 + \frac{3}{2})$ for the Jeffreys Rule prior, where n_0 is the number of zero-valued observations and n_1 is the number of non-zero-valued observations in the sample. Note that $n = n_0 + n_1$. Further results from Harvey (2012) indicate that the posterior distribution $\mu_l|\sigma_l^2, data \sim N(\overline{y}, \sigma_l^2/n_1)$ and the posterior distribution $\sigma_l^2|data$ is an inverse gamma distribution. Using the simulation methodology described earlier in this article as well as methods described in Harvey (2012) for simulating from a beta distribution for the zero-valued observations, credibility intervals were calculated for a single sample. Both equal-tailed and HPD intervals were calculated. These were calculated for various combinations of the four parameters, as in Table 10.

The results are presented in Table 11.

From the results in Table 11 it is clear that with both priors the HPD intervals are a considerable improvement, particularly in terms of interval length, on the standard equal-tailed intervals. Thus, the flexibility of a Bayesian approach to handling these situations is evident.

The results do not clearly define which prior distribution is better suited. Small sample size situations with small σ_l^2 values favour the use of the Independence Jeffreys prior. While offering better coverage it does so at the expense of wider interval widths. As the value of σ_l^2 increases this tendency reverses with regards to the interval width, with the Independence Jeffreys prior still offering the better coverage. As the proportion of zero-valued observations increases the width of the credibility intervals also increase. Thus, with respect to coverage, the Independence Jeffreys prior seems better suited and with regards to interval width the prominence is evident, except in small sample size settings.

9. Example 4 – An application to hypothesis testing including zero values

The following example is similar to that of Example 2 presented earlier, except that the possibility of zero values is also included. Suppose we wish to test the following hypothesis:

$$H_0 : \widetilde{\sigma}^2_{(1)} = \widetilde{\sigma}^2_{(2)}$$
$$H_1 : \widetilde{\sigma}^2_{(1)} \neq \widetilde{\sigma}^2_{(2)}.$$

Again, this could be performed by simulating observations from the two distributions and determine either the ratio or difference between the variances and determining whether the intervals contains 1 or 0, respectively.

In this example the following was done:

1. Take the following initial values: $n_{10} = 5,\ \ n_{11} = 25,\ \nu_1 s_1^2 = 6,\ \overline{y}_1 = 1$ and for the second sample $n_{20} = 7,\ n_{21} = 30,\ \nu_2 s_2^2 = 8,\ \overline{y}_2 = 1$.

Table 11. 95% credibility intervals for $Var(X)$ with zero values included.

					ET			HPD		
δ	n	μ	**Result**	**Prior**	σ_l^2=0.25	σ_l^2=1	σ_l^2=2.25	σ_l^2=0.25	σ_l^2=1	σ_l^2=2.25
0.1	10	0	Coverage	IJ	95.07	95.05	95.34	95.43	95.05	95.45
0.1	10	0	Coverage	JR	94.35	93.53	93.50	92.20	91.60	91.20
0.1	10	0	Length	IJ	82.87	1.40×10^{16}	3.30×10^{35}	7.09	3.30×10^{10}	2.40×10^{24}
0.1	10	0	Length	JR	10.16	3.60×10^{10}	5.10×10^{39}	3.19	7.95×10^{7}	7.50×10^{27}
0.1	10	2	Coverage	IJ	95.06	94.86	94.93	95.47	94.65	95.29
0.1	10	2	Coverage	JR	93.97	93.98	93.47	92.43	91.97	90.84
0.1	10	2	Length	IJ	1371.90	2.70×10^{16}	2.00×10^{180}	303.44	1.90×10^{11}	1.60×10^{88}
0.1	10	2	Length	JR	296.05	4.12×10^{10}	5.10×10^{29}	152.70	6.78×10^{7}	3.80×10^{22}
0.1	20	0	Coverage	IJ	95.12	94.53	95.05	95.59	95.12	95.50
0.1	20	0	Coverage	JR	94.72	94.52	94.32	93.82	93.00	92.92
0.1	20	0	Length	IJ	1.65	1465.40	1.64×10^{9}	1.27	330.75	4.26×10^{7}
0.1	20	0	Length	JR	1.37	358.68	1.57×10^{8}	1.08	128.53	6.89×10^{6}
0.1	20	2	Coverage	IJ	95.30	95.23	95.52	95.80	94.95	95.63
0.1	20	2	Coverage	JR	94.50	94.38	94.70	93.45	93.13	93.10
0.1	20	2	Length	IJ	90.33	2.02×10^{5}	2.50×10^{11}	69.40	2.70×10^{4}	2.79×10^{9}
0.1	20	2	Length	JR	75.41	1.95×10^{4}	9.28×10^{9}	59.40	7074.4	3.74×10^{8}
0.1	50	0	Coverage	IJ	94.99	95.00	95.33	95.37	94.92	95.45
0.1	50	0	Coverage	JR	95.28	94.93	94.74	94.92	94.20	93.77
0.1	50	0	Length	IJ	0.62	27.362	6647.40	0.57	20.02	2964.50
0.1	50	0	Length	JR	0.60	24.152	4975.80	0.55	17.91	2335.90
0.1	50	2	Coverage	IJ	95.03	95.05	94.51	95.49	94.85	94.71
0.1	50	2	Coverage	JR	95.10	94.83	94.70	94.69	93.99	93.75
0.1	50	2	Length	IJ	34.28	1460.60	4.16×10^{5}	31.27	1070.30	1.82×10^{5}
0.1	50	2	Length	JR	32.46	1337.30	2.67×10^{5}	29.72	990.61	1.25×10^{5}
0.2	10	0	Coverage	IJ	95.12	95.07	95.06	95.04	94.75	95.32
0.2	10	0	Coverage	JR	93.48	93.11	93.40	91.16	90.76	91.28
0.2	10	0	Length	IJ	9.90×10^{15}	6.40×10^{45}	1.00×10^{105}	6.83×10^{5}	6.90×10^{26}	6.80×10^{64}
0.2	10	0	Length	JR	50.23	1.50×10^{18}	1.10×10^{43}	4.62	1.50×10^{11}	4.40×10^{28}
0.2	10	2	Coverage	IJ	95.60	95.29	94.77	94.81	95.06	95.36
0.2	10	2	Coverage	JR	94.02	93.47	93.50	91.85	91.67	91.52
0.2	10	2	Length	IJ	1.00×10^{182}	9.90×10^{52}	1.00×10^{291}	1.30×10^{49}	6.10×10^{30}	6.00×10^{129}
0.2	10	2	Length	JR	7.40×10^{4}	1.40×10^{25}	4.10×10^{41}	367.61	5.20×10^{15}	3.40×10^{28}
0.2	20	0	Coverage	IJ	95.57	94.91	95.10	95.67	94.64	95.35
0.2	20	0	Coverage	JR	94.72	94.53	94.54	93.00	92.88	92.98
0.2	20	0	Length	IJ	1.99	2729.80	1.50×10^{14}	1.44	469.51	1.60×10^{11}
0.2	20	0	Length	JR	1.56	1328.80	1.00×10^{11}	1.20	277.71	8.13×10^{8}
0.2	20	2	Coverage	IJ	94.80	94.67	95.05	95.00	94.79	95.10
0.2	20	2	Coverage	JR	94.15	94.24	94.09	92.76	92.97	92.47
0.2	20	2	Length	IJ	104.79	3.03×10^{5}	1.30×10^{18}	77.28	3.61×10^{4}	1.70×10^{14}
0.2	20	2	Length	JR	83.83	8.69×10^{4}	7.40×10^{11}	64.48	1.61×10^{4}	6.24×10^{9}
0.2	50	0	Coverage	IJ	95.22	94.74	95.12	95.50	95.09	95.08
0.2	50	0	Coverage	JR	95.05	94.88	95.12	94.63	93.95	94.06
0.2	50	0	Length	IJ	0.67	30.77	1.15×10^{4}	0.61	21.60	4463.10
0.2	50	0	Length	JR	0.63	26.43	8758.90	0.58	18.89	3418.20
0.2	50	2	Coverage	IJ	95.15	95.39	95.32	95.25	95.25	95.45
0.2	50	2	Coverage	JR	94.65	94.90	94.61	94.07	93.93	93.68
0.2	50	2	Length	IJ	36.60	1640.70	6.32×10^{5}	33.20	1152.30	2.46×10^{5}
0.2	50	2	Length	JR	34.67	1428.00	4.46×10^{5}	31.58	1021.50	1.84×10^{5}
0.3	20	0	Coverage	IJ	95.22	95.04	94.87	94.56	95.08	95.19
0.3	20	0	Coverage	JR	94.15	94.29	94.12	92.94	92.78	92.69
0.3	20	0	Length	IJ	2.41	6.41×10^{4}	2.70×10^{21}	1.64	3010.50	5.40×10^{14}
0.3	20	0	Length	JR	1.77	5833.80	2.70×10^{13}	1.30	688.14	4.10×10^{10}
0.3	20	2	Coverage	IJ	95.39	95.18	94.79	95.14	95.13	95.11
0.3	20	2	Coverage	JR	94.52	93.86	94.34	92.83	92.83	92.74
0.3	20	2	Length	IJ	133.20	2.21×10^{7}	1.80×10^{27}	89.54	3.27×10^{5}	1.20×10^{19}
0.3	20	2	Length	JR	100.49	5.00×10^{6}	3.60×10^{17}	72.48	1.80×10^{5}	1.20×10^{13}
0.3	50	0	Coverage	IJ	95.36	95.24	95.09	95.50	95.19	95.04
0.3	50	0	Coverage	JR	94.53	94.67	94.76	94.07	93.88	94.03
0.3	50	0	Length	IJ	0.72	36.28	3.76×10^{4}	0.64	24.00	1.03×10^{4}
0.3	50	0	Length	JR	0.67	30.93	1.97×10^{4}	0.60	20.86	6224.70
0.3	50	2	Coverage	IJ	95.22	94.69	95.03	95.31	94.88	95.13
0.3	50	2	Coverage	JR	94.75	94.68	94.84	94.49	94.35	94.01
0.3	50	2	Length	IJ	39.30	2029.40	2.39×10^{6}	35.20	1337.20	6.36×10^{5}
0.3	50	2	Length	JR	36.78	1738.80	9.99×10^{5}	33.13	1171.60	3.28×10^{5}

Table 12. 95% credibility intervals for the ratio $\psi_1 = Var(X_1)/Var(X_2)$.

Prior	Equal-tail intervals			HPD intervals		
	Lower	**Upper**	**Length**	**Lower**	**Upper**	**Length**
Independence Jeffreys	0.347 279 2	2.582 510	2.235 231	0.238 225 1	2.159 510	1.921 285
Jeffreys Rule	0.344 921 9	2.528 715	2.183 793	0.233 901 7	2.099 736	1.865 834

Table 13. 95% credibility interval for the difference $\psi_2 = Var\,(X_1) - Var\,(X_2)$.

Prior	Equal-tail intervals			HPD intervals		
	Lower	**Upper**	**Length**	**Lower**	**Upper**	**Length**
Independence Jeffreys	−5.166 438	4.931 006	10.097 440	−5.269 847	4.750 744	10.020 590
Jeffreys Rule	−5.124 059	4.422 148	9.546 207	−5.258 599	4.165 873	9.424 472

2. Using the prior distributions described previously simulate the posterior distribution of:
   (a) $\psi_1 = Var(X_1)/Var(X_2)$.
   (b) $\psi_2 = Var\,(X_1) - \,Var(X_2)$.

The results are given in Tables 12 and 13.

In this example, both methods agree that there is no difference between the two populations. We can see a similar trend as was identified in the simulation study previously. The Independence Jeffreys prior results in the largest interval length, but the use of HPD intervals minimises this as a disadvantage. In particular, all HPD intervals decrease the interval length considerably for both the ratio and difference between the variances. The Independence Jeffreys prior, due to its superior coverage (as indicated in the simulation study) is the preferred choice here.

10. Other estimators for the variance $\sigma^2 = \exp(2\mu_l + \sigma_l^2)[\exp(\sigma_l^2) - 1]$

Zellner (1971) derived two estimators, conditional on the parameter σ_l^2, for the mean $\mu = \exp(\mu_l + \frac{1}{2}\sigma_l^2)$. First, he considered the class of estimators $(k)\exp(\overline{Y})$ where $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and k a constant. He showed that the estimator $\mu^* = \exp(\overline{Y} + \frac{1}{2}\sigma_l^2 - \frac{3}{2}\sigma_l^2/n)$ is a minimum means square error estimator of μ. By using a relative quadratic error loss function it is also noted that μ^* is a minimum posterior expected loss estimator. He further indicated that $\widetilde{\mu} = \exp(\overline{Y} + \frac{1}{2}\sigma_l^2 - \frac{1}{2}\sigma_l^2/n)$ is a minimum variance unbiased estimator of μ.

We considered the class of estimators, $(k)\exp(2\overline{Y})$, for the variance σ^2. In the section that follows as well as the derivations in the appendix, $\sigma^2 = \delta$. This is in no way related to the value of δ used in previous sections and is used for notational simplicity. In the appendix it is shown that

$$\delta^* = \exp\left\{2\overline{Y} + \sigma_l^2\left(1 - \frac{6}{n}\right)\right\}\left[\exp(\sigma_l^2) - 1\right]$$

is a minimum mean square error estimator as well as a minimum posterior expected loss estimator of $\delta = \sigma^2$. It is also shown that

$$\widetilde{\delta} = \exp\left\{2\overline{Y} + \sigma_l^2\left(1 - \frac{2}{n}\right)\right\}\left[\exp(\sigma_l^2) - 1\right]$$

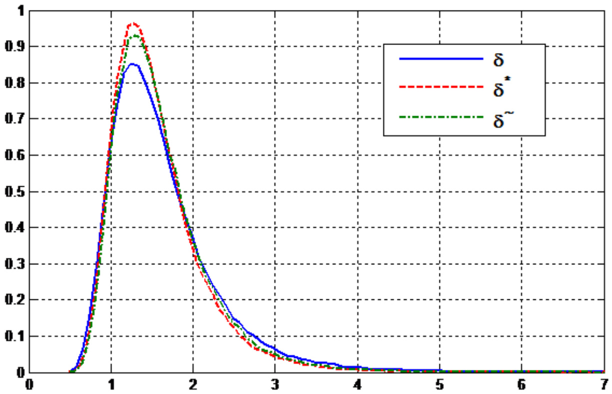

Figure 2. Distributions of δ, δ^* and $\widetilde{\delta}$ – Example 1.

Table 14. Distribution summary of other estimators – Example 1.

	Mean	Median	Mode	Variance	95% HPD interval
δ	1.6360	1.4862	1.2820	0.4678	(0.670 – 2.935)
δ^*	1.5593	1.4441	1.2800	0.3173	(0.718 – 2.645)
$\widetilde{\delta}$	1.5967	1.4710	1.2840	0.3493	(0.719 – 2.727)

is a minimum variance unbiased estimator of σ^2 for a given σ_l^2.

The following three examples illustrate the application of these results. In all three examples $Y \sim N(\mu_l, \sigma_l^2)$ and $X = \exp(Y)$. In the first instance parameter values are assumed as follows: $\mu_l = 1$, $\sigma_l^2 = 0.2$ and $n = 31$. The chosen values result in a "true" value for $\delta = \sigma^2 = \exp(2\mu_l + \sigma_l^2)[\exp(\sigma_l^2) - 1] = 1.9982$, against which the simulated results can be compared. Furthermore, if we observe $\overline{y} = 0.9972$ and $s^2 = 0.1560$ then we have "true" values for $\delta^* = 1.9116$ and $\widetilde{\delta} = 1.9615$ against which to compare the simulated results. Figure 2 presents the simulated distributions of these estimators (as obtained using simulations A or B).

Table 14 presents summary statistics and HPD intervals based on the above simulated distributions.

In addition to this, Table 15 indicates the coverage and length of equal-tailed (EQ) and highest posterior density (HPD) credibility intervals for the different estimators, using 50 000 simulated observations.

Based on the results in Tables 14 and 15 it is evident that δ^* results in the shortest interval length. For the equal-tailed intervals however, this comes at the cost of less than adequate coverage. However, the advantage of the Bayesian approach is the possibility of obtaining of HPD intervals. For the HPD

Table 15. Coverage and interval length of other estimators – Example 1.

Method	EQ coverage	EQ mean length	HPD coverage	HPD mean length
δ	0.9506	4.1683	0.9518	3.6251
δ^*	0.9241	3.3264	0.9525	2.9672
$\widetilde{\delta}$	0.9243	3.5396	0.9525	3.1403

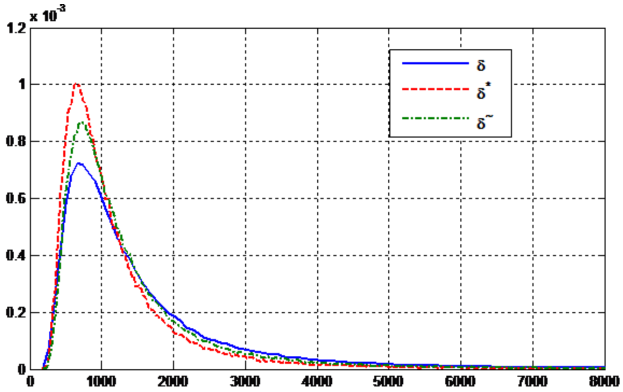

Figure 3. Distributions of δ, δ^* and $\widetilde{\delta}$ – Example 2.

Table 16. Distribution summary of other estimators – Example 2.

	Mean	Median	Mode	Variance	95% HPD interval
$\delta(=\sigma^2)$	1739.5	1107.1	689.1	1.3005×10^7	(195.6 – 4673.7)
δ^*	1274.2	933.5	648.7	2.3884×10^6	(252.3 – 3085.8)
$\widetilde{\delta}$	1485.6	1047.1	712.3	4.4859×10^6	(260.1 – 3702.3)

intervals the coverage is similar (and acceptable) for all estimators and δ^* still results in the most efficient intervals.

In the next example, the assumed parameter values are changed to $\mu_l = 3$, $\sigma_l^2 = 1$ and $n = 30$. This results in a "true" value of $\delta = 1\,884.3$ and (if $\overline{y} = 2.9182$ and $s^2 = 0.8422$) $\delta^* = 1.9116$ and $\widetilde{\delta} = 1.9615$. For the sake of comparison, the distributions are once again simulated together with the summary statistics obtained previously and the results are shown in Figure 3 and Table 16.

From the results in Figure 3 and Table 16 it can be noticed in both cases that the estimator with the narrowest HPD intervals is δ^*. However, this comes at the cost of coverage (for at least the equal-tailed intervals), as shown in the previous simulation example.

11. Predictive Distributions of Future Responses

Sinha (1989) obtained the predictive density of a single future lognormal variable in his article on pg 75, equation 2.4. In this article we extend on this by obtaining the predictive density of a future sample variance of a lognormal distribution. This is the situation described below and a complete derivation can be found in the appendix.

Consider a future sample of m observations from a $N(\mu_l, \sigma_l^2)$ population, denoted as $Y_{1f}, Y_{2f}, \ldots, Y_{mf}$. The future sample mean and variance are respectively defined as:

$$\overline{Y}_f = \frac{1}{m}\sum_{j=1}^{m} Y_{jf}, \qquad s_f^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(Y_{jf} - \overline{Y}_f\right)^2.$$

Furthermore, we know that

$$\frac{(m-1)\,s_f^2}{\sigma_l^2} \sim \chi_{m-1}^2$$

independently of

$$\overline{Y}_f|\mu_l, \sigma_l^2 \sim N\left(\mu_l, \frac{\sigma_l^2}{m}\right).$$

Thus, the maximum likelihood estimator of a future sample lognormal variance is

$$W = \exp(2\overline{y}_f + s_f^2)\left[\exp(s_f^2) - 1\right] = \exp(2\overline{y}_f)\left\{\exp(s_f^2)\left[\exp(s_f^2) - 1\right]\right\} = \exp(2\overline{y}_f)\{a\},$$

where $a = \exp(s_f^2)\,[\exp(s_f^2)\ - 1]$. Using this description, it has been shown in the appendix that the distribution of W is given by the following density function:

$$f\left(w \mid \widetilde{\mu}, \widetilde{\sigma}^2, a\right) = \frac{1}{w\sqrt{2\pi\widetilde{\sigma}^2}}\exp\left\{-\frac{1}{2\widetilde{\sigma}^2}\left(\ln\left(\frac{w}{a}\right)\ - \widetilde{\mu}\right)^2\right\}\ , \quad 0 \leq w \leq \infty.$$

Using this result, the unconditional density function, $f(w \mid data)$ can be simulated as follows:

Simulation A

1. Simulate $\sigma_l^2 = (n-1)\,s^2/\chi_{n-1}^2$.
2. Given σ_l^2, simulate $\mu_l \sim N(\overline{y}, \sigma_l^2/n)$.
3. Calculate $\widetilde{\sigma}^2 = 4\sigma_l^2/m$ and $\widetilde{\mu} = 2\mu_l$.
4. Simulate $s_f^2 = \sigma_l^2\chi_f^2/f$, where $f = m-1$.
5. Calculate $a = \exp(s_f^2)[\exp(s_f^2) - 1]$.
6. Substitute $\widetilde{\mu}$, $\widetilde{\sigma}^2$ and a into the expression for $f(w \mid \widetilde{\mu}, \widetilde{\sigma}^2, a)$ and draw/determine the density function.
7. Repeat steps 1-6 a large number of, say l (= 100 000 or 1 000 000), times and calculate the average of the l densities (Rao-Blackwell method) to obtain $f(w \mid data)$.
8. Obtain the mean, median, mode, variance and 95% confidence intervals.

Similarly, $W = \exp(2\overline{y}_f + s_f^2)\ [\exp(s_f^2)\ - 1]$ can be directly simulated as follows:

Simulation B

1. Simulate $\sigma_l^2 = (n-1)s^2/\chi_{n-1}^2$. Given σ_l^2, simulate $\mu_l \sim N(\overline{y}, \sigma_l^2/n)$.
2. Using these values simulate $s_f^2 = \sigma_l^2\chi_f^2/f$ and $\overline{y}_f|\mu_l, \sigma_l^2 \sim N(\mu_l, \sigma_l^2/m)$, where $f = m-1$ and m is the size of the future sample.
3. Substitute $\overline{y}_f$ and s_f^2 in W.
4. Repeat steps 1-3 l (= 100 000) times and obtain l observations of W.

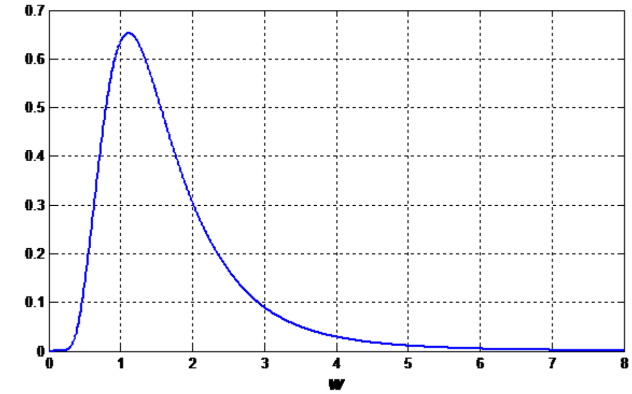

Figure 4. Density of W.

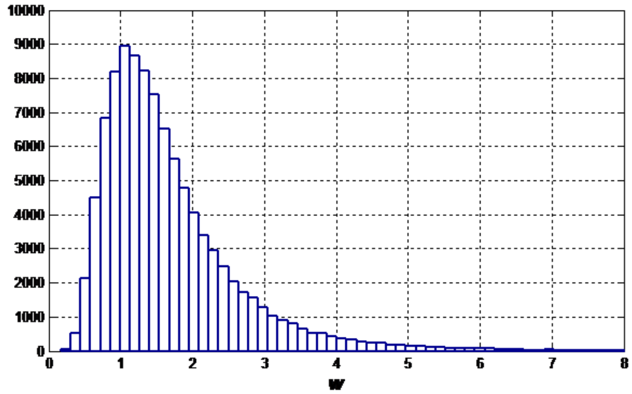

Figure 5. Histogram of values simulate from W.

5. Plot a histogram to illustrate the unconditional distribution of W.

6. Determine the mean, median, mode, variance and 95% confidence interval. This will coincide with the values obtained in step 8 of simulation A.

7. If the bars of the histogram are smoothed then it will reflect the density function obtained in simulation A.

The following example serves to illustrate the above results.

Consider the following data: $n = 31$, $\overline{y} = 0.9972$, $s^2 = 0.1560$ and $m = 31$. Simulations A and B produced the results in Figures 4 and 5 and Table 17 (based on $l = 100\,000$ simulations).

It can be noted that both the simulated distribution and the histogram-based method result in similar summary measures.

Table 17. Distribution Summary (based on 100 000 simulations).

Method	Mean	Median	Mode	Variance	95% interval
Distribution	1.6726	1.4333	1.110	0.9206	(0.567 – 4.285)
Histogram	1.6914	1.4410	1.075	1.0733	(0.575 – 4.297)

12. Conclusion

In this article we have implemented a Bayesian approach to answer various questions regarding the variance of a lognormal distribution. For one of the prior distributions, namely the Independence Jeffreys prior, the method is identical to the frequentist approach of generalised confidence intervals, as proposed by Krishnamoorthy et al. (2006). Various non-informative prior distributions were compared in the Bayesian approach. Simulation studies were performed to determine the performance and examples were presented. In addition, other estimators of the lognormal variance were derived. These are similar to a class of estimators that Zellner (1971) evaluated for the mean of the lognormal distribution, except that here the same class of estimators is applied to the variance.

The Bayesian methods perform well in this setting. The Independence Jeffreys and the Jeffreys Rule prior in particular seem to offer similar results. Interval length for the former prior is higher, but coverage is better in general. Although the equal-tail credibility intervals resulted in wider intervals, an advantage of the Bayesian paradigm is in the calculation of HPD intervals and thereby increase the efficiency and in some cases the accuracy and coverage of the intervals.

The probability-matching and reference prior distributions were not suited to this setting in comparison to the other prior distributions. Specifically for the probability-matching prior the interval length is less than the other methods and priors, but this comes at the expense of poor coverage.

One of the biggest advantages to using the Bayesian paradigm to answer questions such as these is the flexibility. For example, when considering the possibility of the presence of zero-valued observations, there is no known frequentist method currently available. In addition, the flexibility of the Bayesian approach was also illustrated by deriving and simulating the predictive density of the variance of a future sample of observations. The Bayesian methodology is able to handle the increased complexity quite easily.

Appendix

A.1 Derivation of the probability-matching prior distribution for the variance of a lognormal distribution

We know that the Fisher information matrix for $\boldsymbol{\theta} = (\mu_l, \sigma_l^2)$ per unit observation is given by

$$F(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\sigma_l^2} & 0 \\ 0 & \frac{1}{2\sigma_l^4} \end{bmatrix}.$$

Define

$$t(\boldsymbol{\theta}) = \exp(2\mu_l + \sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\}.$$

Now,

$$\frac{\partial t(\boldsymbol{\theta})}{\partial \mu_l} = 2\exp(2\mu_l + \sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\},$$

$$\frac{\partial t(\boldsymbol{\theta})}{\partial \sigma_l^2} = \exp(2\mu_l + \sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\} + \exp(2\mu_l + \sigma_l^2)\exp(\sigma_l^2),$$

$$\nabla_t'(\boldsymbol{\theta}) = \left[\begin{array}{cc} \frac{\partial t(\boldsymbol{\theta})}{\partial \mu_l} & \frac{\partial t(\boldsymbol{\theta})}{\partial \sigma_l^2} \end{array}\right] = \exp(2\mu_l + \sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\}\left[\begin{array}{cc} 2 & \frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1} \end{array}\right].$$

Using the Fisher Information matrix (refer to Harvey, 2012),

$$\begin{aligned}
\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta}) &= \nabla_t'(\boldsymbol{\theta})\sigma_l^2\left[\begin{array}{cc} 1 & 0 \\ 0 & 2\sigma_l^2 \end{array}\right] \\
&= \sigma_l^2\exp(2\mu_l + \sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\}\left[\begin{array}{cc} 2 & \frac{2\sigma_l^2\left\{2\exp(\sigma_l^2)-1\right\}}{\exp(\sigma_l^2)-1} \end{array}\right], \\
\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta}) &= \sigma_l^2\exp(4\mu_l + 2\sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\}^2\left[4 + \frac{2\sigma_l^2\left\{2\exp(\sigma_l^2) - 1\right\}^2}{\left\{\exp(\sigma_l^2) - 1\right\}^2}\right], \\
\left\{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})\right\}^{1/2} &= \sigma_l\exp(2\mu_l + \sigma_l^2)\left\{\exp(\sigma_l^2) - 1\right\}2\left[1 + \frac{\sigma_l^2\left\{2\exp(\sigma_l^2) - 1\right\}^2}{2\left\{\exp(\sigma_l^2) - 1\right\}^2}\right]^{1/2}, \\
\gamma(\boldsymbol{\theta}) = \left[\begin{array}{cc} \gamma_1(\boldsymbol{\theta}) & \gamma_2(\boldsymbol{\theta}) \end{array}\right] &= \frac{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})}{\left\{\nabla_t'(\boldsymbol{\theta})F^{-1}(\boldsymbol{\theta})\nabla_t(\boldsymbol{\theta})\right\}^{1/2}} \\
&= \frac{2\sigma_l}{2\left[1 + \frac{\sigma_l^2\left\{2\exp(\sigma_l^2)-1\right\}^2}{2\left\{\exp(\sigma_l^2)-1\right\}^2}\right]^{1/2}}\left[\begin{array}{cc} 1 & \frac{\sigma_l^2\left\{2\exp(\sigma_l^2)-1\right\}}{\exp(\sigma_l^2)-1} \end{array}\right].
\end{aligned}$$

Therefore, for the prior $p_M(\mu, \sigma_l^2)$ to be a probability-matching prior the following differential equation must be satisfied:

$$\frac{\partial}{\partial \mu}\left[\gamma_1(\boldsymbol{\theta})p_m(\mu, \sigma_l^2)\right] + \frac{\partial}{\partial \sigma_l^2}\left[\gamma_2(\boldsymbol{\theta})p_m(\mu, \sigma_l^2)\right] = 0.$$

That is,

$$\begin{aligned}
\varphi(\boldsymbol{\theta}) &= \frac{2\sigma_l}{2\left[1 + \frac{\sigma_l^2\left\{2\exp(\sigma_l^2)-1\right\}^2}{2\left\{\exp(\sigma_l^2)-1\right\}^2}\right]^{1/2}} \times \frac{\sigma_l^2\left\{2\exp(\sigma_l^2) - 1\right\}}{\exp(\sigma_l^2) - 1} \\
&= \frac{\sigma_l^3\left\{2\exp(\sigma_l^2) - 1\right\}}{\left[\frac{2\left\{\exp(\sigma_l^2)-1\right\}^2 + \sigma_l^2\left\{2\exp(\sigma_l^2)-1\right\}^2}{2\left\{\exp(\sigma_l^2)-1\right\}^2}\right]^{1/2}}\frac{1}{\exp(\sigma_l^2) - 1} \\
&= \frac{\sigma_l^3\left\{2\exp(\sigma_l^2) - 1\right\}\sqrt{2}}{\left\{2\left\{\exp(\sigma_l^2) - 1\right\}^2 + \sigma_l^2\left\{2\exp(\sigma_l^2) - 1\right\}^2\right\}^{1/2}}.
\end{aligned}$$

Therefore, the following prior distribution will satisfy the differential equation:

$$p_M(\boldsymbol{\theta}) = p_M(\mu_l, \sigma_l^2) \propto \sigma_l^{-3}\left\{\frac{2\left(\exp(\sigma_l^2) - 1\right)^2}{\left\{2\exp(\sigma_l^2) - 1\right\}^2} + \sigma_l^2\right\}^{1/2}.$$

A.2 Derivation of the reference prior distribution for the variance of a lognormal distribution

Using the same definition for $F(\boldsymbol{\theta})$ as in the previous derivation, additionally define

$$t(\boldsymbol{\theta}) = \exp\left(2\mu_l + \sigma_l^2\right)\left\{\exp\left(\sigma_l^2\right) - 1\right\}.$$

Then

$$\ln(t(\boldsymbol{\theta})) = \left(2\mu_l + \sigma_l^2\right) + \ln\left\{\exp\left(\sigma_l^2\right) - 1\right\}$$

$$\therefore\ \mu_l = \frac{1}{2}\ln(t(\boldsymbol{\theta})) - \frac{1}{2}\sigma_l^2 - \frac{1}{2}\ln\left\{\exp\left(\sigma_l^2\right) - 1\right\}$$

$$\therefore\ \frac{\partial\mu_l}{\partial t(\boldsymbol{\theta})} = \frac{1}{2t(\boldsymbol{\theta})}$$

$$\therefore\ \frac{\partial\mu_l}{\partial\sigma_l^2} = -\frac{1}{2} - \frac{1}{2}\left(\frac{\exp\left(\sigma_l^2\right)}{\exp\left(\sigma_l^2\right) - 1}\right) = -\frac{1}{2}\left(1 + \frac{\exp\left(\sigma_l^2\right)}{\exp\left(\sigma_l^2\right) - 1}\right) = -\frac{1}{2}\left(\frac{2\exp\left(\sigma_l^2\right) - 1}{\exp\left(\sigma_l^2\right) - 1}\right).$$

Let

$$A = \frac{\partial\left(\mu_l, \sigma_l^2\right)}{\partial\left\{t(\boldsymbol{\theta}), \sigma_l^2\right\}} = \begin{bmatrix} \frac{\partial\mu_l}{\partial t(\boldsymbol{\theta})} & \frac{\partial\mu_l}{\partial\sigma_l^2} \\ \frac{\partial\sigma_l^2}{\partial t(\boldsymbol{\theta})} & \frac{\partial\sigma_l^2}{\partial\sigma_l^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2t(\boldsymbol{\theta})} & -\frac{1}{2}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right) \\ 0 & 1 \end{bmatrix}.$$

Hence, the Fisher information matrix under the reparametrization $\left(t(\boldsymbol{\theta}),\ \sigma_l^2\right)$ is given by:

$$F\left(t(\boldsymbol{\theta}), \sigma_l^2\right) = A' F\left(\mu_l, \sigma_l^2\right) A = \begin{bmatrix} \frac{1}{2t(\boldsymbol{\theta})} & 0 \\ -\frac{1}{2}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right) & 1 \end{bmatrix} \frac{1}{\sigma_l^2}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2\sigma_l^2} \end{bmatrix} A$$

$$= \frac{1}{\sigma_l^2}\begin{bmatrix} \frac{1}{2t(\boldsymbol{\theta})} & 0 \\ -\frac{1}{2}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right) & \frac{1}{2\sigma_l^2} \end{bmatrix}\begin{bmatrix} \frac{1}{2t(\boldsymbol{\theta})} & -\frac{1}{2}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right) \\ 0 & 1 \end{bmatrix}$$

$$= \frac{1}{\sigma_l^2}\begin{bmatrix} \frac{1}{4t^2(\boldsymbol{\theta})} & -\frac{1}{4}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right)\frac{1}{t(\boldsymbol{\theta})} \\ -\frac{1}{4}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right)\frac{1}{t(\boldsymbol{\theta})} & \frac{1}{4}\left(\frac{2\exp(\sigma_l^2)-1}{\exp(\sigma_l^2)-1}\right)^2 + \frac{1}{2\sigma_l^2} \end{bmatrix}.$$

We require the prior distribution that maximised the entropy between the prior and posterior distributions. This prior is derived as

$$p_R(t(\boldsymbol{\theta}), \sigma_l^2) \propto \frac{1}{t(\boldsymbol{\theta})}\frac{1}{\sigma_l}\sqrt{\frac{1}{4}\left(\frac{2\exp(\sigma_l^2) - 1}{\exp(\sigma_l^2) - 1}\right)^2 + \frac{1}{2\sigma_l^2}},$$

$$\therefore\ p_R(\mu_l, \sigma_l^2) = p_R(t\left(\boldsymbol{\theta}), \sigma_l^2\right)\left|\frac{\partial t(\boldsymbol{\theta})}{\partial\mu_l}\right|.$$

Since $\partial t(\boldsymbol{\theta})/\partial\mu_l = 2t(\boldsymbol{\theta})$, it follows that

$$p_R(\mu_l, \sigma_l^2) \propto \frac{1}{\sigma_l}\sqrt{\frac{1}{4}\left(\frac{2\exp(\sigma_l^2) - 1}{\exp(\sigma_l^2) - 1}\right)^2 + \frac{1}{2\sigma_l^2}}.$$

A.3 Derivation of the Estimators δ^* and $\widetilde{\delta}$ for σ^2

$$Var\,(X) = \sigma^2 = \delta = \exp(2\mu_l + \sigma_l^2)\ \left[\exp(\sigma_l^2)\ - 1\right] = \exp(2\mu_l + \sigma_l^2)\, g(\sigma_l^2).$$

Consider the class of estimators $\delta^* = (k)\exp(2\overline{Y})$, where, for a given σ_l^2, $\overline{Y} \sim N(\mu_l, \frac{\sigma_l^2}{n})$. The mean squared error associated with δ^* is

$$E(\delta^* - \sigma^2)^2 = E(ke^{2\overline{Y}} - \sigma^2)^2 = E\left(ke^{2\overline{Y}} - \exp(2\mu_l + \sigma_l^2)\, g(\sigma_l^2)\right)^2$$
$$= E\left(k^2e^{4\overline{Y}} - 2ke^{2\overline{Y}}\exp(2\mu_l + \sigma_l^2)\, g(\sigma_l^2) + \exp(4\mu_l + 2\sigma_l^2)\, g^2(\sigma_l^2)\right).$$

Since

$$E(e^{2\overline{Y}}) = \exp\left(2\mu_l + \frac{2\sigma_l^2}{n}\right) \quad \text{and} \quad E(e^{4\overline{Y}}) = \exp\left(4\mu_l + \frac{8\sigma_l^2}{n}\right),$$

it follows that

$$E(\delta^* - \sigma^2)^2 = k^2\exp\left(4\mu_l + \frac{8\sigma_l^2}{n}\right)$$
$$- 2k\exp\left(2\mu_l + \frac{2\sigma_l^2}{n}\right)\exp(2\mu_l + \sigma_l^2)\, g(\sigma_l^2) + \exp(4\mu_l + 2\sigma_l^2)g^2(\sigma_l^2).$$

Differentiating with respect to k and setting equal to zero,

$$\frac{dE(\delta^* - \sigma^2)^2}{dk} = 2k\exp\left(4\mu_l + \frac{8\sigma_l^2}{n}\right) - 2\exp\left(2\mu_l + \frac{2\sigma_l^2}{n}\right)\exp(2\mu_l + \sigma_l^2)\, g(\sigma_l^2) = 0.$$

Therefore,

$$k = \exp\left\{\sigma_l^2\left(1 - \frac{6}{n}\right)\right\} g(\sigma_l^2),$$

and

$$\delta^* = \exp\left\{2\overline{Y} + \sigma_l^2\left(1 - \frac{6}{n}\right)\right\}\ \left[\exp(\sigma_l^2)\ - 1\right]$$

is a minimum mean square error estimator of σ^2.

To prove that δ^* is also a minimum posterior expected loss estimator of σ^2, consider the relative quadratic loss function

$$L = \left(\frac{\sigma^2 - \widetilde{\widetilde{\delta}}}{\sigma^2}\right)^2 = \left(1 - \frac{\widetilde{\widetilde{\delta}}}{\sigma^2}\right)^2 = \left(1 - \widetilde{\widetilde{\delta}}\exp(-2\mu_l - \sigma_l^2)\, g^{-1}(\sigma_l^2)\right)^2$$
$$= 1 - 2\widetilde{\widetilde{\delta}}\exp(-2\mu_l - \sigma_l^2)\, g^{-1}(\sigma_l^2) + \widetilde{\widetilde{\delta}}^2\exp(-4\mu_l - 2\sigma_l^2)\, g^{-2}(\sigma_l^2).$$

Since

$$\mu_l|\sigma_l^2, data\ \sim N\left(\overline{y}, \frac{\sigma_l^2}{n}\right),$$

it follows that

$$E\left(\exp\left(-2\mu_l\right)\right) = \exp\left(-2\overline{y} + \frac{2\sigma_l^2}{n}\right) \quad \text{and} \quad E\left(\exp\left(-4\mu_l\right)\right) = \exp\left(-4\overline{y} + \frac{8\sigma_l^2}{n}\right).$$

Therefore,

$$E\left(L\right) = 1 - 2\tilde{\tilde{\delta}}\exp\left(-2\overline{y} + \frac{2\sigma_l^2}{n} - \sigma_l^2\right) g^{-1}(\sigma_l^2) + \tilde{\tilde{\delta}}^2 \exp\left(-4\overline{y} + \frac{8\sigma_l^2}{n} - 2\sigma_l^2\right) g^{-2}(\sigma_l^2).$$

Differentiating with respect to $\tilde{\tilde{\delta}}$ and setting equal to zero we obtain

$$\frac{dE\left(L\right)}{d\tilde{\tilde{\delta}}} = -2\exp\left(-2\overline{y} + \frac{2\sigma_l^2}{n} - \sigma_l^2\right) g^{-1}(\sigma_l^2) + 2\tilde{\tilde{\delta}}\exp\left(-4\overline{y} + \frac{8\sigma_l^2}{n} - 2\sigma_l^2\right) g^{-2}(\sigma_l^2) = 0,$$

so that

$$\tilde{\tilde{\delta}} = \exp\left(2\overline{y} + \sigma_l^2\left(1 - \frac{6}{n}\right)\right) \left[\exp(\sigma_l^2) - 1\right] = \delta^*.$$

Therefore δ^* is also a minimum posterior expected loss estimator.

Also, since $E(\exp(2\overline{Y})) = \exp(2\mu_l + \sigma_l^2/n)$, for a given σ^2 it follows that

$$\tilde{\delta} = \exp\left(2\overline{Y} + \sigma_l^2\left(1 - \frac{2}{n}\right)\right) \left[\exp(\sigma_l^2) - 1\right]$$

is an unbiased estimator of

$$\sigma^2 = \exp\left(2\mu_l + \sigma_l^2\right) \left\{\exp(\sigma_l^2) - 1\right\}.$$

A.4 Derivation of the Predictive Density

Consider the setting described in the article regarding the variance of a future sample. It was indicated that the maximum likelihood estimator of a lognormal variance of a future sample is

$$W = \exp(2\overline{y}_f + s_f^2)\left[\exp(s_f^2) - 1\right] = \exp(2\overline{y}_f)\left\{\exp(s_f^2)\left[\exp(s_f^2) - 1\right]\right\} = \exp(2\overline{y}_f)\left\{a\right\}.$$

It is apparent that $\overline{y}_f$ will be distributed independently of $\exp(s_f^2)\,[\exp(s_f^2) - 1]$ since $\overline{y}_f$ is distributed independently of s_f^2.

Let

$$2\overline{y}_f = z.$$

Since

$$2\overline{y}_f \sim N\left(2\mu_l, \frac{4\sigma_l^2}{m}\right) = N(\tilde{\mu}, \tilde{\sigma}^2),$$

it follows that

$$f\left(z|\tilde{\mu}, \tilde{\sigma}^2\right) = \frac{\exp\left\{-\frac{(z-\tilde{\mu})^2}{2\tilde{\sigma}^2}\right\}}{\sqrt{2\pi\tilde{\sigma}^2}}.$$

Let $\tilde{Y} = e^Z$. Then, since $\ln \tilde{y} = z$ and $dz/d\tilde{y} = 1/\tilde{y}$, it follows that

$$f\left(\tilde{y} \mid \widetilde{\mu},\ \widetilde{\sigma}^2\right) = \frac{1}{\tilde{y}\sqrt{2\pi\widetilde{\sigma}^2}}\exp\left\{-\frac{1}{2\widetilde{\sigma}^2}(\ln \tilde{y}\ - \widetilde{\mu})^2\right\}.$$

To find the distribution of W, the conditional distribution first needs to be derived. Now, since $W = a\tilde{Y}$ one has that $\tilde{y} = w/a$ and $d\tilde{y}/dw = 1/a$. Therefore,

$$f\left(w \mid \widetilde{\mu}, \widetilde{\sigma}^2, a\right) = \frac{1}{w\sqrt{2\pi\widetilde{\sigma}^2}}\exp\left\{-\frac{1}{2\widetilde{\sigma}^2}\left(\ln\left(\frac{w}{a}\right) - \widetilde{\mu}\right)^2\right\}, \quad 0 \leq w < \infty.$$

References

Bebu, I. and Mathew, T. (2008). Comparing the means and variances of a bivariate lognormal distribution. *Stat. Med.*, **27**, 2684–2696.

Berger, J. O. and Bernardo, J. M. (1992). On the development of the reference prior method. *In* Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (Editors) *Bayesian Statistics*, volume 4. Oxford University Press, Oxford, 35–60.

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

Chu, H., Gange, S. J., Li, X., Hoover, D. R., Liu, C., Chmiel, J. S., and Jacobson, L. P. (2010). The effect of HAART on HIV RNA trajectory among treatment-naïve men and women. *Epidemiology*, **21** (Supplement), S25–S34.

Datta, G. S. and Ghosh, J. K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, **82**, 37–45.

Gupta, R. C. and Li, X. (2006). Statistical inference for the common mean of two log-normal distributions and some applications in reliability. *Comput. Stat. Data Anal.*, **50**, 3141–3164.

Harvey, J. (2012). *Bayesian inference for the lognormal distribution*. Unpublished Ph.D. dissertation, Bloemfontein, South Africa.

Harvey, J. and van der Merwe, A. J. (2012). Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions. *J. Stat. Plan. Infer.*, **142**, 1294–1309.

Jeffreys, H. (1961). *Theory of Probability*. Third edition. Oxford University Press, Oxford.

Krishnamoorthy, K. and Mathew, T. (2003). Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *J. Stat. Plan. Infer.*, **115**, 103–121.

Krishnamoorthy, K., Mathew, T., and Ramachandran, G. (2006). Generalized p-values and confidence intervals: A novel approach for analyzing lognormally distributed exposure data. *J. Occup. Environ. Hyg.*, **3**, 642–650.

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

Oldham, P. (1953). The nature of the variability of dust concentrations at the coal face. *Br. J. Ind. Med.*, **10**, 227–234.

Pearn, W. L. and Wu, C. W. (2005). Process capability assessment for index Cpk based on the Bayesian approach. *Metrika*, **61**, 211–234.

RIZZO, M. (2007). *Statistical Computing with R*. CRC Press, Boca Raton.

SINHA, S. K. (1989). Bayesian inference about the prediction/credible intervals and reliability function for log-normal distribution. *Indian Statist. Assoc.*, **27**, 73–78.

ZELLNER, A. (1971). Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression. *J. Am. Stat. Assoc.*, **66**, 327–330.

ZHOU, X.-H., GAO, S., AND HUI, S. L. (1997). Methods for comparing the means of two independent log-normal samples. *Biometrics*, **53**, 1129.

Manuscript received 2016-12-01, revised 2017-10-20, accepted 2017-11-09.