Non-Inferiority Hypothesis Testing in Two-Arm Trials with Log-normal Data

by

LAHIRU WICKRAMASINGHE

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
In partial fulfilment of the requirements of the Degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg

Abstract

In health related studies, non-inferiority tests are used to demonstrate that a new treatment is not worse than a currently existing treatment by more than a pre-specified margin. In this thesis, we discuss three approaches; a Z-score approach, a generalized p-value approach and a Bayesian approach, to test the non-inferiority hypotheses in two-arm trials for ratio of log-normal means. The log-normal distribution is widely used to describe the positive random variables with positive skewness which is appealing for data arising from studies with small sample sizes. We demonstrate the approaches using data arising from an experimental aging study on cognitive penetrability of posture control. We also examine the suitability of three methods under various sample sizes via simulations. The results from the simulation studies indicate that the generalized p-value and the Bayesian approaches reach an agreement approximately and the degree of the agreement increases when the sample sizes increase. However, the Z-score approach can produce unsatisfactory results even under large sample sizes.

Keywords: Generalized p-value, Log-normal, Monte Carlo, Non-inferiority, Simulation.

Acknowledgment Page

Foremost I would like to thank my advisor, Dr. Saman Muthukumarana, whose invaluable support, patience, guidance and monitorship has helped me a lot over the last few years.

I would also like to thank my committee, Dr. Saumen Mandal and Dr. Lisa Lix for their time and dedication in helping me to complete this thesis. The suggestions and the support are appreciated to make my thesis a success.

Many thanks go to the staff, the faculty and the colleagues in the Department of Statistics, University of Manitoba. Also, I would like to thank the Department of Statistics, University of Manitoba for the financial support given to me throughout my M.Sc. study.

Finally and most especially, I would like to thank my mother and father who supported me and believed in me. You both taught me many valuable lessons on my life and I know that I have got the best mother and father in the world. I love you both a lot.

Dedication Page

This work is dedicated to my mother, father and sister who have supported me with heart and soul.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Hypothesis Testing

Hypothesis testing is a method for testing a claim about a parameter in a population using the data in a sample. Suppose that we have a random sample $X_1, X_2, \ldots, X_n$ from a distribution $f(x|\theta)$, θ is a parameter with parameter space Θ and we want to determine whether $\theta \in \Theta_1$ or $\theta \in \Theta_2$ where

$$\Theta_1 \cup \Theta_2 = \Theta \qquad \text{and} \qquad \Theta_1 \cap \Theta_2 = \emptyset.$$

The null hypothesis (H_0), is a statement about the population parameter which the researcher tries to disprove and the alternative hypothesis (H_1), is a statement that directly contradicts the null hypothesis. More specifically, we write this test as

$$H_0 : \theta \in \Theta_1 \qquad \text{vs.} \qquad H_1 : \theta \in \Theta_2. \tag{1.1}$$

There are two main approaches, the frequentist approach and the Bayesian approach, for testing the hypotheses in (1.1). We describe the basics of these two methods below. More details and the methods for testing the hypotheses can be found in Lehmann et al. (2005).

1.1.1 Frequentist Approach

In this approach, four main steps are followed to test the hypotheses.

1. State the hypotheses (H_0 and H_1).
2. Set the decision criteria.
3. Compute the test statistic, $t(\boldsymbol{x})$.
4. Make the decision.

In frequentist approach, we assume that H_0 is true and the observed values in the sample, serve as the evidence against H_0. The sample values are used either to reject or not to reject H_0 and two types of errors (type I and type II) may occur based on the decision. If we decide to reject H_0 when it is true, then we made a type I error whereas if we decide not to reject H_0 when it is false, then we made a type II error. Ideally, we need to carry out the test, by minimizing both types of errors. Unfortunately, we can not control both errors simultaneously. The rejection region of a test, R is defined as

$$R = \{t(\boldsymbol{x});\ \text{observed values } \boldsymbol{x} \text{ would lead to the rejection of } H_0\}.$$

Then the probability of type I error is,

$$P(\text{type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true}) = P(R | H_0 \text{ is true }).$$

Since we assume that H_0 is true, we fixed the probability of type I error in advance and we try to minimize the probability of type II error. This is called level of significance (significance level) and denoted by α. This is the largest probability of

type I error. Before we start testing the hypotheses, we set the level of significance (α) for the test.

The test statistic ($t(\boldsymbol{x})$) is to determine the likelihood of obtaining sample outcomes if the null hypothesis was true. The test statistic tells us how far the statistic from the parameter which is stated in the null hypothesis. The strength of evidence in support of a null hypothesis is measured by the p-value. The p-value is the probability of observing a test statistic as extreme as $t(\boldsymbol{x})$, assuming the null hypothesis is true. To make the decision, we compare the p-value with the level of significance (α), and decide to reject the null hypothesis if the p-value $\leq \alpha$. Otherwise we fail to reject H_0.

Another approach is to use the region of acceptance. The region of acceptance (A) is a set of observed values, where the null hypothesis is not rejected. The set of values outside the region of acceptance is called the region of rejection. If the test statistic ($t(\boldsymbol{x})$) falls within the A, the null hypothesis is not rejected, whereas if $t(\boldsymbol{x})$ falls outside the A, the null hypothesis is rejected.

1.1.2 Bayesian Approach

The main idea behind Bayesian hypothesis testing is to use the Bayesian approache for determining the strength of evidence. In the frequentist approach for hypothesis testing, inferences are based on the probability of data conditioning on the null hypothesis, whereas the Bayesian approach is based on the probability of hypotheses conditioning on the data. Also, the frequentist approach tests one hypothesis, the null hypothesis (H_0) against its alternative hypothesis (H_1). But, the Bayesian

approach can test multiple hypotheses and takes advantage of prior information. The Bayesian hypothesis testing is fairly straightforward and gives a Bayesian version of the standard p-value once the posterior distribution is obtained.

In testing the hypothesis in (1.1), first we need to find the posterior distribution of the unknown parameter θ either analytically or by simulation. Given the posterior distribution and the data ($\boldsymbol{x}$), we simply calculate $P(\theta \in \Theta_0|\boldsymbol{x})$ and decide in favor of the null hypothesis (H_0) if $P(\theta \in \Theta_0|\boldsymbol{x})$ is sufficiently large. In order to make a better comparison of the hypotheses, the prior and posterior information can also be combined in a ratio (called Bayes factor) unless the posterior probability, $(P(\theta \in \Theta_0|\boldsymbol{x}))$ is close to 0 or 1. We discuss the Bayes factor below.

Bayes Factor

The Bayes factor (B) for comparing two hypotheses is the ratio of the marginal likelihoods of the data under H_0 and H_1. We define the prior probabilities of hypotheses as follows

$$\pi_0 = P(\theta \in \Theta_0) \qquad \text{and} \qquad \pi_1 = P(\theta \in \Theta_1).$$

Then the prior odds on H_0 against H_1 is $\dfrac{\pi_0}{\pi_1}$. We also define the posterior probabilities of hypotheses

$$p_0 = P(\theta \in \Theta_0|\boldsymbol{x}) \qquad \text{and} \qquad p_1 = P(\theta \in \Theta_1|\boldsymbol{x}).$$

Then the posterior odds on H_0 against H_1 is $\dfrac{p_0}{p_1}$.

If the hypotheses are simple, that is for each hypothesis, θ can take only one value,

$$\Theta_0 = \{\theta_0\} \qquad \text{and} \qquad \Theta_1 = \{\theta_1\}.$$

Then we know that

$$p_0 \propto \pi_0 p(\boldsymbol{x}|\theta_0) \qquad \text{and} \qquad p_1 \propto \pi_1 p(\boldsymbol{x}|\theta_1).$$

So the posterior odds on H_0 against H_1 is $\dfrac{p_0}{p_1} = \dfrac{\pi_0}{\pi_1}\dfrac{p(\boldsymbol{x}|\theta_0)}{p(\boldsymbol{x}|\theta_1)}$.

Then the Bayes factor for simple hypotheses is

$$B = \frac{p(\boldsymbol{x}|\theta_0)}{p(\boldsymbol{x}|\theta_1)} = \frac{(p_0/p_1)}{(\pi_0/\pi_1)}.$$

If the hypotheses are composite, that is for each hypothesis, θ can take more than one value, the prior probabilities of $\theta \in \Theta_0$ and $\theta \in \Theta_1$ are $p_0(\theta)$ and $p_1(\theta)$ respectively. Then, we define

$$p_0(\theta) = \frac{p(\theta)}{\pi_0} \quad \text{for} \quad \theta \in \Theta_0 \qquad \text{and} \qquad p_1(\theta) = \frac{p(\theta)}{\pi_1} \quad \text{for} \quad \theta \in \Theta_1.$$

The marginal likelihoods of data under H_0 and H_1 are $\int_{\theta\in\Theta_0} p(\boldsymbol{x}|\theta)p_0(\theta)\,d\theta$ and $\int_{\theta\in\Theta_1} p(\boldsymbol{x}|\theta)p_1(\theta)\,d\theta$ respectively.

Then the Bayes factor is

$$B = \frac{\int_{\theta\in\Theta_0} p(\boldsymbol{x}|\theta)p_0(\theta)\,d\theta}{\int_{\theta\in\Theta_1} p(\boldsymbol{x}|\theta)p_1(\theta)\,d\theta}.$$

We know that

$$\begin{aligned} p_0 &= P(\theta \in \Theta_0 | \boldsymbol{x}) \\ &= \int_{\theta \in \Theta_0} p(\theta|\boldsymbol{x})\, d\theta \\ &\propto \int_{\theta \in \Theta_0} p(\boldsymbol{x}|\theta) p(\theta)\, d\theta \end{aligned}$$

$$= \pi_0 \int_{\theta \in \Theta_0} p(\boldsymbol{x}|\theta) p_0(\theta)\, d\theta.$$

Similarly

$$p_1 \propto \int_{\theta \in \Theta_1} p(\boldsymbol{x}|\theta) p_1(\theta)\, d\theta.$$

So the posterior odds on H_0 against H_1

$$\frac{p_0}{p_1} = \frac{\pi_0}{\pi_1} \frac{\int_{\theta \in \Theta_0} p(\boldsymbol{x}|\theta) p_0(\theta)\, d\theta}{\int_{\theta \in \Theta_1} p(\boldsymbol{x}|\theta) p_1(\theta)\, d\theta}.$$

Then the Bayes factor for composite hypotheses is

$$B = \frac{\int_{\theta \in \Theta_0} p(\boldsymbol{x}|\theta) p_0(\theta)\, d\theta}{\int_{\theta \in \Theta_1} p(\boldsymbol{x}|\theta) p_1(\theta)\, d\theta} = \frac{(p_0/p_1)}{(\pi_0/\pi_1)}.$$

So the Bayes factor is

$$B = \frac{\text{Posterior odds}}{\text{Prior odds}}. \tag{1.2}$$

The Bayes factor provides a scale of evidence in favour of one hypothesis versus another. Table 1.1 provides the possible interpretations for the Bayes factor. More details about the Bayes factor can be found in Stauffer (2008).

Table 1.1: The scale of evidence for the Bayes factor

B	Interpretation
$B < 0.1$	Strong evidence for H_1
$0.1 \leq B < 0.33$	Moderate evidence for H_1
$0.33 \leq B < 1$	Weak evidence for H_1
$1 \leq B < 3$	Weak evidence for H_0
$3 \leq B < 10$	Moderate evidence for H_0
$10 \leq B$	Strong evidence for H_0

1.2 Non-Inferiority Hypothesis

In health related clinical trials, the "arm" is a set or a group of subjects who receives a specific treatment, or no treatment. The treatment could be a drug or a certain physical activity. The group sizes are pre-defined before the trial begins. Here our interest is on two-arm trails (2 groups). The examples of two-arm study are a comparison of a treatment versus a placebo or a new drug versus the existing drug.

Nowadays, in many clinical trials, the notion of equivalence and non-inferiority tests have become effective standard procedure. In classical or traditional hypothesis tests, well established theory and methods are used to show whether two products or treatments are significantly differ from each other. The null hypothesis will be rejected if the test statistic is sufficiently large compared to the critical value or if the p-value of the test statistic is sufficiently small compared to the level of significance α. However, in pharmaceutical industry, when a new product or a treatment is introduced, the main goal is often to demonstrate that the new product or treatment is either equivalent or superior to the current or the old product or the treatment. In equivalence studies, the objective is to demonstrate that the effect of the new

product or treatment is equivalent to the effect of the current or the old product or the treatment. Practically it is impossible to show the exact equivalence since an infinite sample size is often required. Then the researchers have introduced the equivalence margin, denoted by δ, which defines the range of values for which the effect of the new product or treatment is to be considered equivalent to the old product or the treatment. If the effects of the two treatments or the products differ by more than the equivalence margin in either direction, then equivalence does not hold. For example, we write the hypotheses for testing that the new drug Y is equivalent to the existing drug X as

$$H_0 : |\mu_X - \mu_Y| \geq \delta \qquad \text{vs.} \qquad H_1 : |\mu_X - \mu_Y| < \delta \qquad \delta \to 0,$$

where μ_X and μ_Y are the population means of the existing drug and the new drug respectively.

In non-inferiority studies, the objective is to demonstrate the new treatment or the product is not worse than the existing treatment or the product by more than a pre-specified, small amount (δ). If you consider the above example, the null hypothesis is that the new drug (Y) is inferior to the existing drug (X) by at least a certain pre-specified amount (δ). The alternative is that the new drug (Y) is not inferior to the existing drug (X) by less than that pre-specified amount (δ). More specifically,

$$H_0 : \mu_X - \mu_Y \geq \delta \qquad \text{vs.} \qquad H_1 : \mu_X - \mu_Y < \delta \quad \text{where } \delta > 0.$$

Here δ is called the non-inferiority margin and it is defined to be strictly positive. A detailed discussion on equivalence and non-inferiority tests are given in Snapinn (2000). We summarize the hypotheses associated with different studies in Table 1.2.

Table 1.2: Hypotheses associated with different types of studies

Type of study	**Null hypothesis**	**Alternative hypothesis**
Classical	There is no difference between new and old	There is a difference between new and old
Equivalence	The new and old are not equivalent	The new and old are equivalent
Non-inferiority	The new is inferior to the old	The new is not inferior to the old

There is a well established literature for testing non-inferiority in two-arm trials with normal data. An overview of recent methods can be found in Snapinn (2000), Walker at el. (2011), Munk at el. (2005) and Wellek (2003). Koti (2013) also discuss the non-inferiority hypothesis testing in comparing the ratio of two normal means. In this thesis, we consider the non-inferiority hypothesis testing with log-normal data. We review the basic properties of log-normal distribution in Section 1.3.

1.2.1 Choice of Non-Inferiority Margin

The determination of the non-inferiority margin (δ), is the most critical step in non-inferiority hypothesis testing. The non-inferiority margin is the degree of inferiority of the new treatments to the existing treatment and is selected prior to the trial. A small value of δ determines a narrower non-inferiority region, and makes it more difficult to establish non-inferiority.

It is important to know that the trials have assay sensitivity, that is to distinguish an effective treatment from a less effective or ineffective treatment. The non-inferiority trials are dependent on the expected effect of the existing treatment. If the existing

treatment has no effect at all in the non-inferiority trial, then finding even a very small difference between the new treatment and the existing treatment provides no evidence that the new treatment is effective. Thus, assay sensitivity is an essential property of a non-inferiority clinical trial. It can be quite difficult to specify an appropriate non-inferiority margin. There are two basic approaches, which were stated in the International Conference on Harmonization guidance E10 on Choice of Control Group and Related Issues in Clinical Trials (ICH E10). The first approach states that non-inferiority margin cannot be greater than the smallest expected effect size of the existing treatment. The second approach states that the determination of the non-inferiority margin should be based on both statistical reasoning and historical evidence.

1.3 Log-normal Distribution

The log-normal distribution is appropriate for the random variables that are positive, and also positively skewed (right skewed). The log-normal distributions are commonly used to model the data from many scientific disciplines: geology and mining, human medicine, environment, microbiology, plant physiology, ecology, food science and economics. Eckhard et al. (2001) have mentioned several examples for each field such as the concentration of elements in the earth's crust and their radioactivity, the latent period of infectious diseases, the distribution of chemicals in the environment, the abundance of species in plant and animal communities and the distribution of income in a city.

1.3.1 Properties of Log-normal Distribution

Let $X(0 < x < \infty)$ be a positive random variable that follows a log-normal distribution. Then

$$Y = \ln(X) \sim N(\mu, \sigma^2).$$

Since X and Y are connected by the relation $Y = \ln(X)$, the cumulative distribution function (cdf) and the probability density function (pdf) are given by

$$\begin{aligned}
F_X(x) &= \begin{cases} P(X \leq x) & \text{for} \quad x > 0 \\ 0 & \text{Otherwise} \end{cases} \\
&= \begin{cases} P(\ln\,(X) \leq \ln\,(x)) & \text{for} \quad x > 0 \\ 0 & \text{Otherwise} \end{cases} \\
&= \begin{cases} P(Y \leq \ln\,(x)) & \text{for} \quad x > 0 \\ 0 & \text{Otherwise} \end{cases} \\
&= \begin{cases} F_Y(\ln(x)) & \text{for} \quad x > 0 \\ 0 & \text{Otherwise,} \end{cases}
\end{aligned}$$

$$\begin{aligned}
f_X(x) &= \frac{d}{dx} F_Y(\ln\,(x)) \\
&= \begin{cases} \frac{1}{x} f_Y(\ln\,(x)) & \text{for} \quad x > 0 \\ 0 & \text{Otherwise} \end{cases} \\
&= \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right] & ; \quad x > 0, -\infty < \mu < \infty, \sigma^2 \geq 0 \\ 0 & \text{Otherwise.} \end{cases}
\end{aligned}$$

Note that μ is the location parameter or the log mean and σ is the scale parameter or the log standard deviation on the log-transformation. It is important that X cannot get zero values since the transformation $Y = \ln(X)$ is not defined at $X = 0$.

Figure 1.1: Shape of log-normal distribution for different σ values

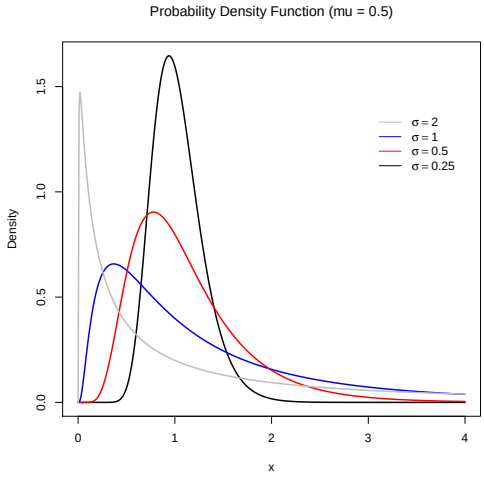


Figure 1.1 shows the density curves for different values of the shape parameter (σ). The log-normal distribution is right skewed distribution and the skewness increases as the value of σ increases for a given μ. Also, if σ is greater than 1, the density curve rises very sharply in the beginning, peaks out early and decreases sharply. Figure 1.2 shows the density curves for different values of the location parameter (μ). The skewness increases as the value of μ decreases for a given σ. Figure 1.3 shows the density curves for different values of the location parameter (μ) and the shape parameter (σ). The skewness decreases as the value of μ increases and the value of σ decreases.

Figure 1.2: Shape of log-normal distribution for different μ values

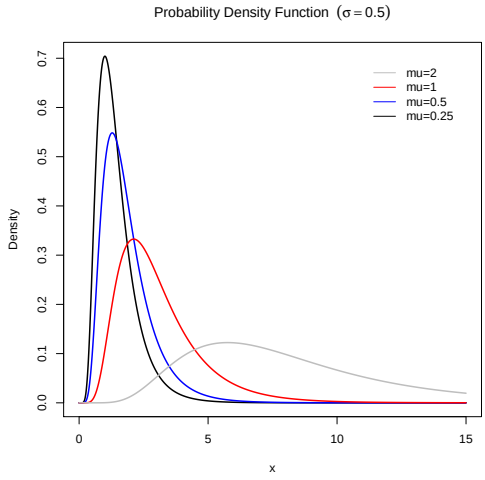
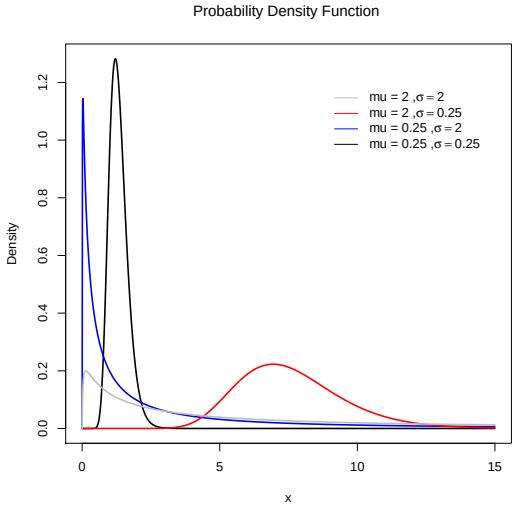


Figure 1.3: Shape of log-normal distribution for different μ and σ values



The mean of the log-normal distribution is given by

$$E(X) = E(e^Y)$$

$$= \int_{-\infty}^{\infty} e^y \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy$$

$$= \int_{-\infty}^{\infty} e^{\mu+t} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}} dt \qquad \text{[Apply transformation t = y -}\mu\text{]}$$

(1.3)

$$= e^{\mu} \int_{-\infty}^{\infty} e^{t} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}} dt$$

$$= e^{\mu} \int_{-\infty}^{\infty} e^{\frac{\sigma^2}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(t-\sigma^2)^2}{2\sigma^2}} dt$$

$$= e^{\mu} e^{\frac{\sigma^2}{2}}$$

$$= e^{\mu + \frac{\sigma^2}{2}}. \tag{1.4}$$

Note that the mean of the log-normal distribution is a function of μ and σ^2.

Also,

$$E(X^2) = E(e^{2Y})$$

$$= \int_{-\infty}^{\infty} e^{2y} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy$$

$$= \int_{-\infty}^{\infty} e^{2(\mu+t)} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}} dt \qquad \text{[Apply transformation t = y -}\mu\text{]}$$

$$= e^{2\mu} \int_{-\infty}^{\infty} e^{2t} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{2\sigma^2}} dt$$

$$= e^{2\mu} \int_{-\infty}^{\infty} e^{2\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(t-2\sigma^2)^2}{2\sigma^2}} dt$$

$$= e^{2\mu} e^{2\sigma^2}$$

$$= e^{2(\mu+\sigma^2)}. \tag{1.5}$$

The variance of the log-normal distribution is given by

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= e^{2(\mu+\sigma^2)} - [e^{\mu+\frac{1}{2\sigma^2}}]^2$$

$$= e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$$

$$= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \tag{1.6}$$

The variance of a log-normal distribution is also a function of μ and σ^2. The median and the mode of the log-normal distribution are e^{μ} and $e^{\mu-\sigma^2}$ respectively. This implies that the log-normal distribution is positively skewed (Mean>Median).

1.3.2 Estimators

In statistics, there are two types of point estimators, method of moments estimators and maximum likelihood estimators. In this section, we discuss these estimators for log-normal parameters.

The Method of Moments

We know that,

$$E(X) = e^{\mu+\frac{\sigma^2}{2}} \qquad \text{and} \qquad E(X^2) = e^{2(\mu+\sigma^2)}.$$

The first sample moment $= \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$.

The second sample moment $= \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$.

Then the first and second sample moments are obtained by solving

$$e^{\hat{\mu}+\frac{\hat{\sigma}^2}{2}} = \bar{X} \qquad \text{and}$$

$$e^{2(\hat{\mu}+\hat{\sigma}^2)} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

The estimators of μ and σ^2 are given by

$$\hat{\mu} = 2\ln\bar{X} - \frac{1}{2}\ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) \quad \text{and} \quad \hat{\sigma}^2 = \ln\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - 2\ln\bar{X}.$$

Maximum Likelihood

Let $x_1, x_2, \ldots, x_n$ be a observed sample from the log-normal distribution. Then the likelihood function of the data is given by

$$\begin{aligned} L(\boldsymbol{x}|\mu,\sigma^2) &= \prod_{i=1}^{n} f(x_i|\mu,\sigma^2) \\ &= \prod_{i=1}^{n}\left(\frac{1}{x_i\sigma\sqrt{2\pi}}\exp\left[-\frac{(\ln(x_i)-\mu)^2}{2\sigma^2}\right]\right) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\left(\frac{1}{\prod_{i=1}^{n} x_i}\right)\exp\left[-\frac{\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{2\sigma^2}\right]. \end{aligned}$$

It is more convenient to work with the logarithm of the likelihood function. So

$$l = \ln L(\boldsymbol{x}|\mu,\sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \sum_{i=1}^{n}\ln(x_i) - \frac{\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{2\sigma^2}.$$

Then the the logarithm of likelihood function is differentiated with respect to the parameters μ and σ^2.

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} \ln(x_i) - n\mu\right) \qquad \text{and}$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{-n\sigma^2 + \sum_{i=1}^{n}(\ln(x_i) - \mu)^2}{2(\sigma^2)^2}.$$

Then the estimators of μ and σ^2 are obtained by equating the above two equations to zero. This gives

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln(x_i)}{n} \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(\ln(x_i) - \hat{\mu}\right)^2}{n}. \tag{1.7}$$

Note that $\hat{\mu}$ and $\hat{\sigma}^2$ are the sample mean and the variance of the transformed data in the log scale.

1.4 Non-inferiority Hypothesis Testing with Log-normal Data

The mean of the log-normal distribution is a function of both μ and σ^2, and it is difficult to obtain the exact or optimum tests for testing hypotheses in 1.1. Zhou et al. (1997) introduced a Z-score approach for the problem of testing the equality of two log-normal means. We can use this approach for testing non-inferiority hypotheses.

Let X_1 and X_2 are two independent log-normal random variables distributed as

$$X_1 \sim \text{log-N}(\mu_1, \sigma_1^2) \qquad \text{and} \qquad X_2 \sim \text{log-N}(\mu_2, \sigma_2^2).$$

Suppose that $X_{1i}, i = 1, 2, \ldots, n_1$ and $X_{2j}, j = 1, 2, \ldots, n_2$ denote random samples from X_1 and X_2, respectively. Also, let $Y_{1i} = \ln(X_{1i}), i = 1, 2, \ldots, n_1$ and $Y_{2j} = \ln(X_{2j}), j = 1, 2, \ldots, n_2$.

The maximum likelihood estimators for μ_1 and μ_2 are

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n_1} \ln(X_{1i})}{n_1} = \bar{Y_1} \qquad \text{and} \qquad \hat{\mu}_2 = \frac{\sum_{j=1}^{n_2} \ln(X_{2j})}{n_2} = \bar{Y_2}.$$

The unbiased estimators of σ_1^2 and σ_2^2 are

$$\hat{\sigma}_1^2 = S_1^2 = \frac{\sum_{i=1}^{n_1} \left(\ln(X_{1i}) - \hat{\mu}_1\right)^2}{n_1 - 1} \qquad \text{and} \qquad \hat{\sigma}_2^2 = S_2^2 = \frac{\sum_{j=1}^{n_2} \left(\ln(X_{2j}) - \hat{\mu}_2\right)^2}{n_2 - 1}.$$

Let $E(X_1) = \exp(\eta_1) = \exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right)$ and $E(X_2) = \exp(\eta_2) = \exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right)$. Then the problem of our interest is to compare the ratio of two log-normal means for non-inferiority. More specifically we test

$$H_0 : \frac{E(X_1)}{E(X_2)} \geq \delta \qquad \text{vs.} \qquad H_1 : \frac{E(X_1)}{E(X_2)} < \delta \qquad \text{where } \delta > 1$$

$$H_0 : \frac{\exp(\eta_1)}{\exp(\eta_2)} \geq \exp(\delta^{'}) \qquad \text{vs.} \qquad H_1 : \frac{\exp(\eta_1)}{\exp(\eta_2)} < \exp(\delta^{'}) \qquad \text{where} \quad \delta = \exp(\delta^{'})$$

$$H_0 : \exp(\eta_1 - \eta_2) \geq \exp(\delta^{'}) \quad \text{vs.} \quad H_1 : \exp(\eta_1 - \eta_2) < \exp(\delta^{'}) \quad \text{where} \quad \delta = \exp(\delta^{'})$$

which is equivalent to

$$H_0 : \ln(\exp(\eta_1 - \eta_2)) \geq \ln(\exp(\delta^{'})) \qquad \text{vs.} \qquad H_1 : \ln(\exp(\eta_1 - \eta_2)) < \ln(\exp(\delta^{'}))$$

$$H_0 : \eta_1 - \eta_2 \geq \delta' \qquad \text{vs.} \qquad H_1 : \eta_1 - \eta_2 < \delta' \qquad \text{where } \delta' > 0. \tag{1.8}$$

So a test statistic can be constructed from $\hat{\eta}_1 - \hat{\eta}_2$, where $\hat{\eta}_i = \hat{\mu}_i + S_i^2/2$, for $i = 1, 2$.

Since $\hat{\mu}_i$ and S_i^2 are independent, $\hat{\mu}_i$ is distributed as $N(\mu_i, \sigma_i^2/n_i)$ and $(n_i - 1)S_i^2/\sigma_i^2$ is distributed as $\chi^2_{n_i-1}$, for $i = 1, 2$. Then we have

$$\text{var}(\hat{\eta}_1 - \hat{\eta}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^4}{2(n_1 - 1)} + \frac{\sigma_2^2}{n_2} + \frac{\sigma_2^4}{2(n_2 - 1)}.$$

Replacing $\sigma_i^2, i = 1, 2$ by the unbiased estimator $S_i^2, i = 1, 2$, then a test statistic for testing H_0 is given by

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2 + \frac{1}{2}(S_1^2 - S_2^2) - \delta'}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2}\left(\frac{S_1^4}{n_1 - 1} + \frac{S_2^4}{n_2 - 1}\right)}}. \tag{1.9}$$

When n_1 and n_2 are both large, the distribution of Z is approximately standard normal under H_0. The following algorithm can be used to calculate the type I error and power of Z-score approach

Algorithm

1. Specify $n_1, n_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \delta'$ and $0 < \alpha < 1$.

2. For $l = 1$ to m, generate $x_{11}, x_{12}, \ldots, x_{1n_1}$ from log-N(μ_1,σ_1^2) and $x_{21}, x_{22}, \ldots, x_{2n_2}$ from log-N(μ_2,σ_2^2). Set $y_{1i} = \ln(x_{1i}), i = 1, 2, \ldots, n_1$ and $y_{2i} = \ln(x_{2i}), i = 1, 2, \ldots, n_2$.

3. Compute $\bar{Y}_1, \bar{Y}_2, S_1^2$ and S_2^2, where $\bar{Y}_i = \dfrac{1}{n_i}\displaystyle\sum_{j=1}^{n_i} y_{ij}$ and $S_i^2 = \dfrac{1}{n_i - 1}\displaystyle\sum_{j=1}^{n_i}(y_{ij} - \bar{Y}_i)^2$, $i = 1, 2$.

4. Compute

$$Z_l = \frac{\bar{Y}_1 - \bar{Y}_2 + \frac{1}{2}(S_1^2 - S_2^2) - \delta'}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{1}{2}\left(\frac{S_1^4}{n_1 - 1} + \frac{S_2^4}{n_2 - 1}\right)}}.$$

5. Let $K_l = 1$ if $Z_l < Z_\alpha$, else $K_l = 0$.

   End l loop.

6. Then compute $\bar{K} = \frac{1}{m}\sum_{l=1}^{m} K_l$.

$\bar{K}$ provides the simulated type I error when Z_l is computed under H_0. On the other hand, when Z_l is computed under H_1., $\bar{K}$ provides the simulated power of the test.

The Z-score approach is recommended for large samples ($n > 25$) but for small samples the power and the type I error are too conservative or too liberal. On the other hand, a generalized p-value approach was introduced by Krishnamoorthy et al. (2003) for log-normal data. In this thesis, we develop a Bayesian approach for comparing two log-normal means and compare the performance of the approach with the Z-score approach and the generalized p-value approach.

1.5 Organization of the Thesis

In this thesis, we discuss non-inferiority hypothesis testing in two-arm trials with log-normal data. In Chapter 1, we give an introduction to hypothesis testing, non-inferiority hypothesis testing, the properties of log-normal distribution and the

Z-score test approach for comparing two log-normal means. In Chapter 2, we discuss the development of a generalized p-value approach for non-inferiority hypothesis testing. We then propose a Bayesian approach for non-inferiority hypothesis testing in Chapter 3. In Chapter 4, we apply the methods proposed in this thesis on a data set arising from an experimental aging study on cognitive penetrability of posture control. We also assess the suitability of the methods for various sample sizes via simulation studies. In Chapter 5, we provide the conclusions based on the methods proposed in the thesis and discuss some future research directions.

Chapter 2

Inference Using a Generalized P-value Approach

2.1 Generalized P-value

The generalized p-value was first introduced by Weerahandi and Tsui (1989) as an extended version of the classical p-value. In classical p-value method, there are few challenges: difficult to find the suitable test statistic, difficult to find the sampling distribution of the test statistic and involves many nuisance parameters. The nuisance parameters are unknown parameters which are required to construct a realistic model but there is no interest in making inferences about them. As a result, the exact solution may not exist, hence the approximate solution with restrictions may suggest. The generalized p-value method is used in order to overcome these challenges in the classical p-value method and to relax some of the limitations.

A comprehensive discussion of the generalized p-value is given in the book by Weerahandi (1995). We provide a brief overview of the approach focusing on non-

inferiority hypotheses below.

Let $\mathbf{X}$ be a random variable with parameter of interest ω and nuisance parameter β. Suppose, we are interested in testing $H_0 : \omega \leq \omega_0$ vs. $H_1 : \omega > \omega_0$, where ω_0 is a specified value. Let $\mathbf{x}$ be the observed value of $\mathbf{X}$, and β be a nuisance parameter. Tsui and Weerahandi (1989) defined a generalized pivotal quantity (GPQ), G($\mathbf{X}$, $\mathbf{x}$, ω, β), as a function of ($\mathbf{X}$, $\mathbf{x}$, ω, β), satisfying the following conditions:

i For a fixed $\mathbf{x}$ and $\omega = \omega_0$, the distribution of G($\mathbf{X}$, $\mathbf{x}$, ω_0, β) is free of nuisance parameter β.

ii The observed value of G($\mathbf{X}$, $\mathbf{x}$, ω, β), g_{obs} = G($\mathbf{x}$, $\mathbf{x}$, ω, β) is free of any unknown parameters ω and β.

If in addition to condition (i) and (ii), P(G($\mathbf{X}$, $\mathbf{x}$, ω_0, β) $\geq g_{obs}|\omega$) is stochastically non-decreasing in ω for fixed $\mathbf{x}$ and β, then GPQ is the generalized test function (GTF) for testing H_0 against H_1. If G is stochastically non-decreasing in ω, then the generalized p-value (GPV) for the test of H_0 against H_1 is defined as

$$p = \sup_{\omega \leq \omega_0} P(G \geq g_{obs}|\omega) = P(G \geq g_{obs}|\omega_0).$$

Note that the computation of GPV is possible since the distribution of G($\mathbf{X}$, $\mathbf{x}$, ω, β) is free of the nuisance parameter β and g_{obs} = G($\mathbf{x}$, $\mathbf{x}$, ω_0, β) , is computed from the data alone. For a nominal level α, one typically rejects the null hypothesis if $p < \alpha$. In many situations, the GTF is the GPQ minus the parameter of interest; i.e., G($\mathbf{X}$, $\mathbf{x}$, ω, β) - ω or a standardized version of that. In the following section, we adapt the above definition of GPV to develop the required GTF for the non-inferiority hypotheses in (1.7).

2.2 Inference on Two Log-normal Distributions

Krishnamoorthy and Mathews (2003) proposed an algorithm for testing hypotheses on log-normal means. We extend the approach and implement the algorithm in R for testing the non-inferiority hypotheses in (1.7).

Let $X_{1i}, i = 1, 2, \ldots, n_1$, and $X_{2i}, i = 1, 2, \ldots, n_2$, denote the random samples from the log-normal distributions of X_1 and X_2 respectively. Also, let $Y_{1i} = \ln(X_{1i}), i = 1, 2, \ldots, n_1$ and $Y_{2i} = \ln(X_{2i}), i = 1, 2, \ldots, n_2$. Then define

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

Let $\bar{y}_1, \bar{y}_2, s_1^2$ and s_2^2 denote the observed values of $\bar{Y}_1, \bar{Y}_2, S_1^2$ and S_2^2 respectively.

Let

$$\begin{aligned} T_{3i} &= \bar{y}_i - \frac{\bar{Y}_i - \mu_i}{S_i/\sqrt{n_i}} \times \frac{s_i}{\sqrt{n_i}} + \frac{\sigma_i^2}{2S_i^2}s_i^2 \\ &= \bar{y}_i - \frac{Z_i}{U_i/\sqrt{n_i - 1}}\frac{s_i}{\sqrt{n_i}} + \frac{s_i^2}{2U_i^2/(n_i - 1)} \qquad i = 1, 2, \end{aligned} \tag{2.1}$$

where $Z_i = \dfrac{\bar{Y}_i - \mu_i}{\sigma_i/\sqrt{n_i}} \sim N(0, 1)$ and $U_i^2 = (n_i - 1)\dfrac{S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}$ for $i = 1, 2$ and these random variables are independent. Define the generalized test function

$$T_3 = T_{31} - T_{32} - (\eta_1 - \eta_2)$$

and let

$$T_4 = T_{31} - T_{32}$$

so that $T_3 = T_4 - (\eta_1 - \eta_2)$. Thus the generalized p-value for testing the hypotheses in (1.7) is given by

$$P(T_3 \geq 0 | \eta_1 - \eta_2 = \delta^{'}) = P(T_4 \geq \delta^{'}). \tag{2.2}$$

We compute the generalized p-value using the following algorithm.

Algorithm 1

1. For the given data sets $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$, set $y_{1i} = \ln(x_{1i}), i = 1, 2, \ldots, n_1$ and $y_{2i} = \ln(x_{2i}), i = 1, 2, \ldots, n_2$.

2. Compute $\bar{y}_1, \bar{y}_2, s_1^2$ and s_2^2, where $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ and $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$, $i = 1, 2$.

3. For $i = 1$ to 2,

   For $j = 1$ to m,

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

   Set

$$T_{ij} = \bar{y}_i - \frac{\bar{Y}_i - \mu_i}{S_i / \sqrt{n_i}} \times \frac{s_i}{\sqrt{n_i}} + \frac{\sigma_i^2}{2S_i^2} s_i^2$$

$$= \bar{y}_i - \frac{Z_i}{U_i / \sqrt{n_i - 1}} \frac{s_i}{\sqrt{n_i}} + \frac{s_i^2}{2U_i^2 / (n_i - 1)}$$

   where $Z_i = \dfrac{\bar{Y}_i - \mu_i}{\sigma_i / \sqrt{n_i}} \sim N(0, 1)$ and $U_i^2 = (n_i - 1)\dfrac{S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}$ for $i = 1, 2$ and these random variables are independent.

End j loop.

End i loop.

4. Define the generalized test function $T_{3j} = T_{1j} - T_{2j}$.

5. Let $K_j = 1$ if $T_{3j} \geq \delta^{'}$ else $K_j = 0$.

6. Then compute $\bar{K} = \dfrac{1}{m}\sum_{j=1}^{m} K_j$.

$\bar{K}$ is the Monte Carlo simulated generalized p-value. Note that the generalized test function satisfies the three requirements.

1. At the observed values, $\bar{Y}_1 = \bar{y}_1, \bar{Y}_2 = \bar{y}_2, S_1^2 = s_1^2$ and $S_2^2 = s_2^2$, T_3 is free on any unknown parameters.

2. For fixed η_1 and η_2, the distribution of T_3 does not depend on any unknown parameters.

3. The distribution of T_3 is stochastically decreasing in $\eta_1 - \eta_2$, i.e $P(T_3 \geq c)$ is a decreasing function of $\eta_1 - \eta_2$.

We compute the type I error and the power of the test using the following algorithm.

Algorithm 2

1. Specify $n_1, n_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \delta^{'}$ and $0 < \alpha < 1$.

2. For $i = 1$ to m_1,

   Generate $\bar{y}_{1i}$ form $N(\mu_1, \sigma_1^2/n_1)$ and $\bar{y}_{2i}$ form $N(\mu_2, \sigma_2^2/n_2)$.

   Generate Q_{1i} from $\chi^2_{n_1-1}$, Q_{2i} from $\chi^2_{n_2-1}$ and set $s^2_{1i} = \dfrac{\sigma_1^2 Q_{1i}}{n_1 - 1}, s^2_{2i} = \dfrac{\sigma_2^2 Q_{2i}}{n_2 - 1}$.

3. For $j = 1$ to m_2,

   Generate $Z_{1j} \sim N(0,1), Z_{2j} \sim N(0,1), U^2_{1j} \sim \chi^2_{n_1-1}$ and $U^2_{2j} \sim \chi^2_{n_2-1}$.

   Set

$$T_{1j} = \bar{y}_{1i} - \frac{Z_{1j}}{U_{1j}/\sqrt{n_1 - 1}} \frac{s_{1i}}{\sqrt{n_1}} + \frac{s^2_{1i}}{2U^2_{1j}/(n_1 - 1)} \qquad \text{and}$$

$$T_{2j} = \bar{y}_{2i} - \frac{Z_{2j}}{U_{2j}/\sqrt{n_2 - 1}} \frac{s_{2i}}{\sqrt{n_2}} + \frac{s^2_{2i}}{2U^2_{2j}/(n_2 - 1)}.$$

   End j loop.

4. Set $T_{3j} = T_{1j} - T_{2j}$.

5. Get the $100(\alpha)^{th}$ percentile from T_{3j} , and denoted it as T_{4i}.

   End i loop.

6. Let $K_i = 1$ if $T_{4i} < \delta^{'}$ else $K_i = 0$.

7. Then compute $\bar{K} = \dfrac{1}{m_1} \sum_{i=1}^{m_1} K_i$.

$\bar{K}$ provides the simulated type I error when T_{1j} and T_{2j} are computed under null hypothesis. On the other hand, when T_{1j} and T_{2j} are computed under alternative hypothesis, $\bar{K}$ provides the simulated power of the test.

Chapter 3

Inference Using a Bayesian Approach

3.1 Ingredients of Bayesian Analysis

Bayes' theorem for conditional probability provides the foundation for Bayesian statistical inference. The Bayes' theorem was originally published by Thomas Bayes in 1763. The Bayesian statistical inference uses your beliefs about the unknown parameters before the data are available (prior), along with the data (likelihood function) to calculate the posterior distribution of the unknown parameters.

Let θ is a parameter of interest and $\mathbf{X}$ represents the data arising from a random sample. Then the posterior distribution of θ,

1. if θ is discrete,

$$P(\theta|\mathbf{X}) = \frac{P(\mathbf{X}|\theta) \times P(\theta)}{\sum P(\mathbf{X}|\theta) \times P(\theta)}$$

$$Posterior(\theta|\mathbf{X}) = \frac{Likelihood(\mathbf{X}|\theta) \times Prior(\theta)}{\sum Likelihood(\mathbf{X}|\theta) \times Prior(\theta)}$$

2. if θ is continuous,

$$P(\theta|\mathbf{X}) = \frac{P(\mathbf{X}|\theta) \times P(\theta)}{\int P(\mathbf{X}|\theta) \times P(\theta)\, d\theta}$$

$$Posterior(\theta|\mathbf{X}) = \frac{Likelihood(\mathbf{X}|\theta) \times Prior(\theta)}{\int Likelihood(\mathbf{X}|\theta) \times Prior(\theta)\, d\theta}.$$

The denominator is the normalizing constant that ensures the probability distributions sum (discrete case) or integrate (continuous case) to one as required by the definition of a probability function. Since the denominator does not depend on θ, it does not provide inferential information about θ. Therefore, the un-normalized posterior distribution of θ is

$$P(\theta|\mathbf{X}) \propto P(\mathbf{X}|\theta) \times P(\theta)$$

$$Posterior(\theta|\mathbf{X}) \propto Likelihood(\mathbf{X}|\theta) \times Prior(\theta).$$

The prior distribution of θ represents the initial understanding, or belief about the unknown parameters, before the data analysis. If the sample size is small or available data provide little information about the parameters of interest, the prior distribution becomes more important. Mainly there are two types of priors. If little information is known a priori about the unknown parameters, these type of priors are called non-informative priors. The most commonly used non-informative prior is the uniform distribution, assuming equal probability for the unknown parameters within a realistic range. If enough information is available, the choice of non-informative priors are irrelevant, because these priors have a minimal impact on the posterior distributions of unknown parameters. On the other hand, if we have definite information about the unknown parameters, then we take those information into account as priors. These are called informative priors and have a strong impact

on the posterior.

A prior distribution that sums or integrates to a finite number is called a proper prior. On the other hand, if the area under the prior density is infinity, it is called an improper prior as summarized below.

	P(.) is discrete	P(.) is continuous
Proper Prior	$\sum_{\theta=0}^{\infty} P(\theta) < \infty$	$\int_{-\infty}^{\infty} P(\theta) < \infty$
Improper Prior	$\sum_{\theta=0}^{\infty} P(\theta) = \infty$	$\int_{-\infty}^{\infty} P(\theta) = \infty$

The improper priors can lead to the proper posterior distribution with sum or integral equal to 1, if the likelihood of the dataset has a bounded range for the support. Sometimes, improper priors lead to improper posterior distribution, where we can not make any inferences. More details on Bayesian analysis can be found in Gelman at el. (1996). We now describe the Bayesian concepts relevant to log-normal distribution.

Suppose the random variable X follows the log-normal distribution so that $Y = \ln(X) \sim N(\mu, \sigma^2)$. The probability density function (pdf) is given by

$$f(x|\mu,\sigma^2) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(\ln(x)-\mu)^2}{2\sigma^2}\right] & ; \quad x > 0, -\infty < \mu < \infty, \sigma^2 > 0 \\ 0 & \text{Otherwise.} \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from the log-normal distribution and the likelihood function is given by

$$L(\boldsymbol{x}|\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i|\mu,\sigma^2)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{x_i \sigma \sqrt{2\pi}} \exp \left[-\frac{(\ln(x_i) - \mu)^2}{2\sigma^2} \right] \right)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \left(\frac{1}{\prod_{i=1}^{n} x_i} \right) \exp \left[-\frac{\sum_{i=1}^{n} (\ln(x_i) - \mu)^2}{2\sigma^2} \right].$$

The choice for a prior distribution on the location parameter (μ) is the normal distribution. The obvious choice for a prior on the scale parameter (σ^2) is the inverse gamma distribution, since not only it is conjugate, but also the support of inverse gamma distribution is restricted to positive real numbers. In notation, we write the prior distributions of unknown parameters μ and σ^2 as follows

$$\mu \sim \mathrm{N}(\mu_0, \sigma_0^2) \quad \text{and} \quad \sigma^2 \sim \text{Inv-Gamma}(\alpha, \beta).$$

Then,

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[-\frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right]$$

$$\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-2(\alpha+1)} \exp \left(-\frac{\beta}{\sigma^2} \right).$$

Here, the parameters $\mu_0, \sigma_0^2, \alpha$ and β are assigned fixed values, so they are called hyper parameters. One can also add another level to the hierarchy as hyper priors. The hyper priors provide more information but also they add complexity to the model.

Assuming prior independence, the joint prior distribution $\boldsymbol{\theta} = (\mu, \sigma^2)$ is

$$\begin{aligned}
\pi(\boldsymbol{\theta}) &= \pi(\mu) \times \pi(\sigma^2) \\
&= \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right] \times \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-2(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right) \\
&= \frac{\beta^\alpha \, \sigma^{-2(\alpha+1)}}{\sqrt{2\pi}\sigma_0\Gamma(\alpha)} \exp\left[-\left(\frac{(\mu-\mu_0)^2}{2\sigma_0^2} + \frac{\beta}{\sigma^2}\right)\right]. \qquad (3.1)
\end{aligned}$$

Then the posterior distribution is as follows:

$$\begin{aligned}
&\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto L(\boldsymbol{x}|\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}) \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \left(\frac{1}{\prod_{i=1}^n x_i}\right) \exp\left[-\frac{\sum_{i=1}^n (\ln(x_i)-\mu)^2}{2\sigma^2}\right] \\
&\qquad\qquad \times \frac{\beta^\alpha \, \sigma^{-2(\alpha+1)}}{\sqrt{2\pi}\sigma_0\Gamma(\alpha)} \exp\left[-\left(\frac{(\mu-\mu_0)^2}{2\sigma_0^2} + \frac{\beta}{\sigma^2}\right)\right] \\
&= \frac{\beta^\alpha \, \sigma^{-(2\alpha+n+2)}}{2\pi^{(n+1)/2}\sigma_0\Gamma(\alpha)\prod_{i=1}^n x_i} \exp\left[-\frac{1}{2}\left(\frac{(\mu-\mu_0)^2}{\sigma_0^2} + \frac{2\beta + \sum_{i=1}^n (\ln(x_i)-\mu)^2}{\sigma^2}\right)\right]. \qquad (3.2)
\end{aligned}$$

It is difficult to identify the distribution in (3.2) using distribution theory since it is not in the closed form. This leads to obtain the distribution in (3.2) using Markov Chain Monte Carlo (MCMC) simulations. We briefly describe the MCMC in Section 3.2.

3.2 Markov Chain Monte Carlo (MCMC)

Metropolis and Ulam (1949) and Metropolis et al. (1953) were the first to introduce Markov Chain Monte Carlo (MCMC) methods to evaluate integral quantities. These methods were generalized by Hastings (1970). Geman and Geman (1984) introduced the most widely used MCMC technique, Gibbs Sampling.

In a simple non-hierarchical posterior distribution, it is often easy to sample from the distribution directly, but when the complexity increases it is difficult or impossible to sample from that distribution directly. For more complicated posterior distributions, first obtain samples from the marginal posterior distributions of the hyper-parameters, then simulate the other parameters conditioned on the data and the other simulated parameters. This process was called Monte Carlo simulation.

According to Karlin and Taylor (1975), a stochastic process is a collection of random variables X_t, defined on a known state space indexed by ordered set $T\{X_t; t \in T\}$. The state space is the range of possible values that X_t can take. A Markov Chain is a stochastic process with the property that the probability of making a transition (transition probability) to a new state is dependent only on the current state of the process. More specifically, we write it as

$$p(X_t = j | X_0 = i_0, X_1 = i_1, \ldots, X_{t-1} = i) = p(X_t = j | X_{t-1} = i).$$

In Markov chain simulation, we simulate a Markov process on the state space of θ, which converges to a stationary distribution, is the desired posterior distribution $p(\theta|\boldsymbol{X})$. In MCMC, we use Markov chain simulation to obtain the desired posterior distribution $p(\theta|\boldsymbol{X})$. The primary distinction between Monte Carlo simulation and

MCMC is that in Monte Carlo simulation produces a set of independent simulated values whereas in MCMC the simulated values are dependent on the preceding values.

3.2.1 Gibbs Sampling

The basic idea behind Gibbs sampling is to draw set of full conditional distributions of unknown parameters $\boldsymbol{\theta}$ and produce a Markov chain that cycle through these conditional distributions to produce the joint posterior distribution of interest.

Suppose we know the joint posterior distribution of $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_k\}$, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$. To find the full conditional posterior distributions θ_i, follow those steps.

1. Ignore all the constants in the joint posterior distribution.
2. Drop everything else that does not depend on θ_i.
3. If other parameters depend on θ_i, assume those parameters as constants and figure out the distribution of θ_i using distribution theory.

Then the steps of Gibbs sampling algorithm is given below.

1. Choose the initial values : $\boldsymbol{\theta}^{[0]} = \{\theta_1^{[0]}, \theta_2^{[0]}, \ldots, \theta_k^{[0]}\}$.
2. At each iteration j $(j = 1, 2, \ldots)$ draw values from the full conditional distributions.
   $\theta_1^{[j]} \sim \pi(\theta_1|\theta_2^{[j-1]}, \theta_3^{[j-1]}, \ldots, \theta_{k-1}^{[j-1]}, \theta_k^{[j-1]})$
   $\theta_2^{[j]} \sim \pi(\theta_2|\theta_1^{[j]}, \theta_3^{[j-1]}, \ldots, \theta_{k-1}^{[j-1]}, \theta_k^{[j-1]})$

$\vdots$

$\theta_{k-1}^{[j]} \sim \pi(\theta_{k-1}|\theta_1^{[j]}, \theta_2^{[j]}, \ldots, \theta_{k-2}^{[j]}, \theta_k^{[j-1]})$

$\theta_k^{[j]} \sim \pi(\theta_k|\theta_1^{[j]}, \theta_2^{[j]}, \ldots, \theta_{k-2}^{[j]}, \theta_{k-1}^{[j]})$.

3. Increment j and repeat the step 2 until the convergence.

After the process reaches convergence, all the simulated values can be treated as arising from the desired posterior distribution.

To simulate the values from the posterior distribution in (3.2), first we need to find the full conditional distributions of μ and σ^2.

The full conditional distribution of μ is

$$\pi(\mu|\boldsymbol{x}, \sigma^2) \propto \exp\left[-\frac{1}{2}\left(\frac{(\mu-\mu_0)^2}{\sigma_0^2} + \frac{2\beta + \sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{\sigma^2}\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\mu^2\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right) - 2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n}\ln(x_i)}{\sigma^2}\right)\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)\left(\mu^2 - 2\mu\frac{\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n}\ln(x_i)}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)\left(\mu - \frac{\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n}\ln(x_i)}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}\right)^2\right].$$

Then,

$$\pi(\mu|\boldsymbol{x},\sigma^2) \sim \mathrm{N}\left(\frac{\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^{n}\ln(x_i)}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}, \frac{1}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}\right). \tag{3.3}$$

The full conditional distribution of σ^2 is

$$\pi(\sigma^2|\boldsymbol{x},\mu) \propto \sigma^{-2(\alpha+n/2+1)} \exp\left[-\frac{1}{2}\left(\frac{2\beta+\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{\sigma^2}\right)\right].$$

Then,

$$\pi(\sigma^2|\boldsymbol{x},\mu) \sim \text{Inv-Gamma}\left(\alpha+\frac{n}{2}, \frac{2\beta+\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{2}\right). \tag{3.4}$$

We proposed a Gibbs sampling algorithm using these full conditionals as follows.

1. Give μ, σ^2 starting values, say $\mu^{[0]}$, $\sigma^{2[0]}$.

2. Generate $\mu^{[1]}$, $\sigma^{2[1]}$ from the full conditionals as describe earlier

$$\mu^{[1]} \sim \pi(\mu|\boldsymbol{x},\sigma^{2[0]}) \qquad \text{and} \qquad \sigma^{2[1]} \sim \pi(\sigma^2|\boldsymbol{x},\mu^{[1]}).$$

3. Then repeat the previous step recursively using a loop replacing the starting values.

4. The chain of values produced by this procedure is a Markov chain, and it converges to its equilibrium distribution which is $\pi(\mu,\sigma^2|\boldsymbol{x})$.

We implement Gibbs sampling using R in Chapter 4. If the full set of conditional posterior distributions are not in closed form or it is difficult or impossible to obtain

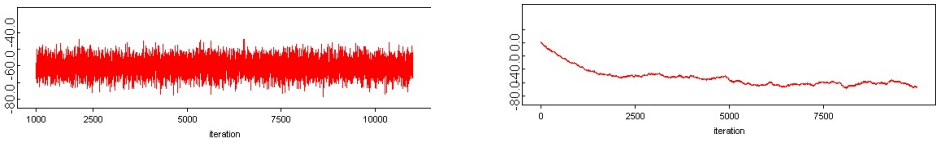
the closed forms then we can use the Metropolis-Hasting algorithm to sample from the desired posterior distribution.

The above Gibbs sampling algorithm generates chains of simulated values from the desired posterior distribution. The simulated values are independent and identically distributed from the posterior distribution and the chains should converge to the desired posterior distribution of interest. Therefore convergence assessment is an important step in MCMC. We describe the convergence assessment in the next section.

3.3 Convergence

One way to check that the chain has converged is to see how well the chain is mixing around the parameter space. The trace plots and history plots can be used to visually inspect the mixing of the chain.

Figure 3.1: Trace plot of one chain



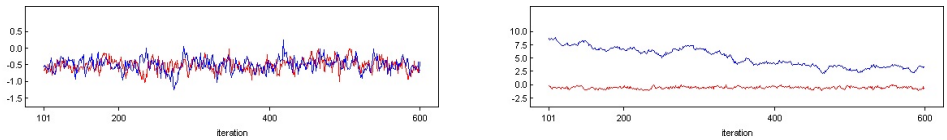
(a) Good Mixing (b) Bad Mixing

Figure 3.1 shows the trace plots of one chain. Figure 3.1 (a) shows good mixing around the parameter space but the mixing is not good in Figure 3.1 (b) . For Figure 3.1 (a), the convergence looks reasonable, but for Figure 3.1 (b) the chain has clearly not reached convergence.

Figure 3.2 shows the trace plots of two chains and each chain is in different colour.

Figure 3.2 (a) shows good mixing around the parameter space and the chains are overlapping one another but in Figure 3.2 (b), there is no overlapping of the chains. For Figure 3.2 (a), the convergence looks reasonable, but for Figure 3.1 (b) the chain has clearly not reached convergence. So the convergence assesment should be done using multiple chains for a proper assessment.

Figure 3.2: Trace plot of two chains



(a) Good Mixing (b) Bad Mixing

3.3.1 Initial Values

The choice of the initial values for the prior distributions of the chains is an essential part of the simulation specification. It is best to try with several initial values for the prior parameters of the chains, look at the mixing of the chains and observe whether those initial values lead to the desired posterior distribution of interest.

Usually, if we can determine the mode, the method of moment or the maximum likelihood estimates we pick the initial values that are close to these estimates. If this is not possible, particularly in complex distributions with high dimensions, we randomly pick the initial values from the state space. Sometimes, we can find the previous works with the same data, so we can use the values associated with previous studies as the initial values.

3.3.2 Burn-In Period

The observations obtained after the chains have converged to the posterior will be very useful on making inferences on unknown parameters. The chains will take some time to converge to the posterior distribution. Therefore, the early observations are not a good representation of the posterior distribution. So we discard the early observations and we assume the remainder of the observations are sampled from the desired posterior distribution.

To calculate the length of the burn-in period, we use the Brooks-Gelman diagnostics (1998). The Brooks-Gelman diagnostics calculates the corrected potential scale reduction factor (R_c). The following steps are used to calculate the corrected potential scale reduction factor (R_c) for each unknown parameter.

- Run $m \geq 2$ chains of length $2n$ with different initial values.
- Discard the first n observations in each chain.
- Calculate the within-chain and between-chain variance.

$$\text{Within Chain Variance } W = \frac{1}{m}\sum_{j=1}^{m} s_j^2 \quad \text{where} \quad s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_j)^2,$$

$\bar{\theta}_j = \frac{1}{n}\sum_{i=1}^{n}\theta_{ij}$, s_j^2 is the variance of the j^{th} chain, θ_{ij} is the i^{th} iteration of θ in chain j and W is the mean of the variances of each chain.

$$\text{Between Chain Variance } B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j - \bar{\bar{\theta}})^2 \quad \text{where} \quad \bar{\bar{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_j.$$

- Calculate the estimated variance of the parameter as a weighted sum of the within-chain and between-chain variance.

$$\hat{Var}(\theta) = \left(1 - \frac{1}{n}\right) W + \left(\frac{1}{n}\right) B.$$

If the starting points are over-dispersed then $\hat{Var}(\theta)$ will over-estimate. Then the pooled posterior variance estimate ($\hat{V}$) is defined as

$$\hat{V} = \hat{Var}(\theta) + \frac{B}{mn}.$$

- The potential scale reduction factor ($\hat{R}$) is defined as

$$\hat{R} = \frac{\hat{V}}{W}.$$

- In order to correctly account for the sampling variability, the correction factor $\frac{d+3}{d+1}$ is used, where d is the degree of freedom and is estimated by the method of moments.

$$d = \frac{2 * \hat{V}^2}{\text{Var}(\hat{V})}.$$

- Then the corrected potential scale reduction factor ($\hat{R}_c$) is defined as

$$\hat{R}_c = \frac{d+3}{d+1}\hat{R} = \frac{(d+3)\hat{V}}{(d+1)W}.$$

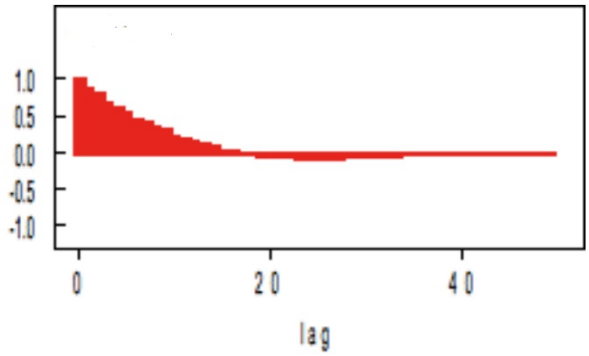
If $\hat{R}_c$ is close to 1, we can conclude that each of the m set of n simulated observations is close to the target distribution.

3.3.3 Thinning

After the chain has settled down to the posterior distribution, we obtain a random sample from the posterior distribution. Even after the burn-in phase, the adjacent observations may be correlated. Thinning is a process to make the observations more nearly independent, hence a random sample from the posterior distribution can be obtained. To check the level of dependence, we look at the autocorrelation plot which plots the autocorrelations with respect to the time differences (lag). The autocorrelation measures the correlation over time differences in a sequence of observations.

The correlation at lag 0 is always 1. If rest of the correlations are very close to zero, the sequence of observations is a random sample. For example, Figure 3.3 shows the first 18 autocorrelations are substantial (autocorrelation > 0.1). Then we obtain a random sample by recording every 20^{th} observation in the sequence. It is important to note that thinning does not improve the quality of the estimate, but provide a good sample representing the posterior distribution.

Figure 3.3: Autocorrelation plot



3.4 Prediction

In Bayesian approach, the predictions of future observations are based on a predictive distribution, which refer to the distribution of data, averaged over all possible parameter values. There are two types of predictive distributions known as

- Prior predictive distribution
- Posterior predictive distribution.

For a new data value x_{new} the prior predictive distribution is defined as

1. If θ is discrete

$$p(x_{new}) = \sum_{\theta} p(x_{new}, \theta) d\theta = \sum_{\theta} p(\theta) p(x_{new}|\theta) d\theta.$$

2. If θ is continuous

$$p(x_{new}) = \int_{\theta} p(x_{new}, \theta) d\theta = \int_{\theta} p(\theta) p(x_{new}|\theta) d\theta.$$

It is called prior because it is not conditional on data and, predictive because it is a prediction for a new data value x_{new}.

After collecting the data $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$, the posterior predictive distribution of a new data value x_{new} is

1. If θ is discrete

$$p(x_{new}|\boldsymbol{X}) = \sum_{\theta} p(x_{new}, \theta|\boldsymbol{X}) d\theta$$

$$= \sum_{\theta} \frac{p(x_{new}, \theta|\boldsymbol{X})}{p(\theta|\boldsymbol{X})} p(\theta|\boldsymbol{X}) d\theta$$

$$= \sum_{\theta} p(x_{new}|\theta, \boldsymbol{X}) p(\theta|\boldsymbol{X}) d\theta$$

$$= \sum_{\theta} p(x_{new}|\theta) p(\theta|\boldsymbol{X}) d\theta \quad [x_{new} \text{ and } \boldsymbol{X} \text{ are independent}].$$

2. If θ is continuous

$$p(x_{new}|\boldsymbol{X}) = \int_{\theta} p(x_{new}|\theta) p(\theta|\boldsymbol{X}) d\theta.$$

It is called posterior because it is conditional on data and, predictive because it is a prediction for a new data value x_{new}.

We now derive the prior and posterior predictive distribution of a new value y using the joint prior distribution in (3.1) and the posterior distribution in (3.2).

The prior predictive probability of a new value y is given by

$$f(y) = \int_{-\infty}^{\infty} \int_{0}^{\infty} f(y|\theta) \pi(\theta) \, d\sigma^2 d\mu$$

$$= \int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2} \right]$$

$$\times \frac{\beta^{\alpha} \, \sigma^{-2(\alpha+1)}}{\sqrt{2\pi}\sigma_0 \Gamma(\alpha)} \exp\left[-\left(\frac{(\mu - \mu_0)^2}{2\sigma_0^2} + \frac{\beta}{\sigma^2} \right) \right] d\sigma^2 d\mu$$

$$\propto \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{\sigma^{-2(\alpha+\frac{1}{2}+1)}}{y} \exp\left[-\left(\frac{(\ln y-\mu)^2}{2\sigma^2}+\frac{(\mu-\mu_0)^2}{2\sigma_0^2}+\frac{\beta}{\sigma^2}\right)\right] d\sigma^2 d\mu$$

$$= \frac{1}{y}\int_{-\infty}^{\infty} \exp\left[-\left(\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\right]$$

$$\times \int_{0}^{\infty} \sigma^{-2(\alpha+\frac{1}{2}+1)} \exp\left[-\left(\frac{(\ln y-\mu)^2+2\beta}{2\sigma^2}\right)\right] d\sigma^2 d\mu$$

$$\propto \frac{1}{y}\int_{-\infty}^{\infty} \exp\left[-\left(\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\right] \frac{1}{[(\ln y-\mu)^2+2\beta]^{(\alpha+\frac{1}{2})}}\, d\mu. \tag{3.5}$$

It is difficult to calculate the integral in (3.5). So $f(y)$ can not be derived directly and we use a Monte Carlo method to sample from $f(y)$. To sample from $f(y)$, follow the following steps.

For $j = 1$ to m,

1. Generate μ_j from $p(\mu)$.

2. Generate σ_j^2 from $p(\sigma^2)$.

3. Generate y_j^* from $p(x|\mu_j, \sigma_j^2)$.

End of j loop.

$y_1^*, y_2^*, \ldots, y_m^*$ are an independent and identically(iid) sample from $f(y)$.

The posterior predictive probability of a future observation y, given the observed

data $\boldsymbol{x}$ is,

$$f(y|\boldsymbol{x}) = \int_{-\infty}^{\infty}\int_{0}^{\infty} f(y|\theta)\pi(\theta|\boldsymbol{x})\, d\sigma^2 d\mu$$

$$= \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right]$$

$$\times \frac{\beta^{\alpha}\ \sigma^{-(2\alpha+n+2)}}{2\pi^{(n+1)/2}\sigma_0\Gamma(\alpha)\prod_{i=1}^{n} x_i} \exp\left[-\frac{1}{2}\left(\frac{(\mu-\mu_0)^2}{\sigma_0^2} + \frac{2\beta+\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{\sigma^2}\right)\right] d\sigma^2 d\mu$$

$$\propto \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{\sigma^{-2(\alpha+\frac{n}{2}+\frac{1}{2}+1)}}{y}$$

$$\times \exp\left[-\frac{(\ln y-\mu)^2 + 2\beta + \sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right] d\sigma^2 d\mu.$$

$$= \frac{1}{y}\int_{-\infty}^{\infty} \exp\left[-\left(\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\right] \int_{0}^{\infty} \sigma^{-2(\alpha+\frac{n}{2}+\frac{1}{2}+1)}$$

$$\times \exp\left[-\left(\frac{(\ln y-\mu)^2 + 2\beta + \sum_{i=1}^{n}(\ln(x_i)-\mu)^2}{2\sigma^2}\right)\right] d\sigma^2 d\mu$$

$$\propto \frac{1}{y}\int_{-\infty}^{\infty} \exp\left[-\left(\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\right] \frac{1}{\left[(\ln y-\mu)^2 + 2\beta + \sum_{i=1}^{n}(\ln(x_i)-\mu)^2\right]^{(\alpha+\frac{n}{2}+\frac{1}{2})}}\, d\mu. \tag{3.6}$$

Again, the distribution in (3.6) is not in the closed form. So $f(y|\boldsymbol{x})$ can not be derived directly and we use a Monte Carlo method to sample from $f(y|\boldsymbol{x})$. To sample from $f(y|\boldsymbol{x})$, follow the following steps.

For $j = 1$ to m,

1. Generate μ_j from $p(\mu, \sigma^2|\boldsymbol{x})$.

2. Generate σ_j^2 from $p(\mu, \sigma^2|\boldsymbol{x})$.

3. Generate y_j^* from $p(x|\mu_j, \sigma_j^2)$.

End of j loop.

$y_1^*, y_2^*, \ldots, y_m^*$ are an independent and identically(iid) sample from $f(y|\boldsymbol{x})$.

The sample from the posterior predictive distribution can be compared to the observed data to assess model fit. If the model fits well, the predictive distribution should be relatively likely to the original data. On the other hand, if there is a large variation between observed data and the data from the predictive distribution, it indicates that the model performs poorly. We perform an analysis using predictive simulations in Chapter 4.

Chapter 4

Data Analysis

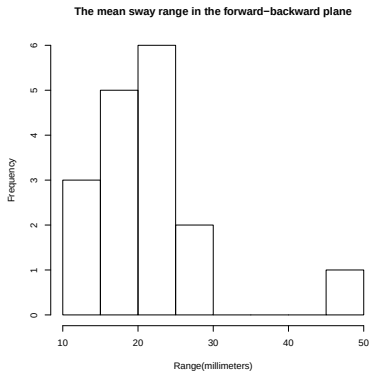
We illustrate the methods discussed in Chapter 2 and 3 using a published data arising from an experimental aging study from Teasdale et al. (1993). The data set consists of the mean sway range in the forward-backward plane and the mean sway range in the side-to-side plane. Note that forward-backward plane and side-to-side plane are two methods of physical treatments on penetrability of posture control. Nine elderly and eight young adults participated in the experiment. All the participants stood barefoot on the force platform with feet together and arms along the body. They were asked to maintain an upright stable posture on the force platform and the sway range in the forward-backward plane and the sway range in the side-to-side plane were calculated. For each participant, 24 trials were given and for each trial, the participants need to maintain an upright stable posture on the force platform for 20 seconds. The mean sway ranges in the forward-backward plane and in the side-to-side plane were calculated and the data are given below with $n_1 = 17$ and $n_2 = 17$.

The mean sway range (in millimeters) in the forward-backward plane (X_1): 19, 30,

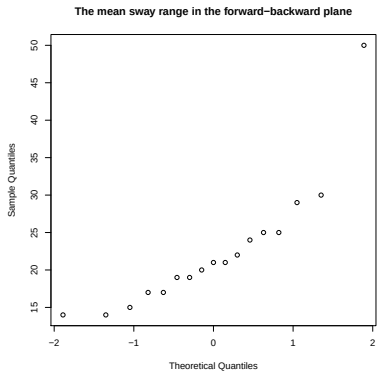
20, 19, 29, 25, 21, 24, 50, 25, 21, 17, 15, 14, 14, 22, 17.

The mean sway range (in millimeters) in the side-to-side plane (X_2): 14, 41, 18, 11, 16, 24, 18, 21, 37, 17, 10, 16, 22, 12, 14, 12, 18.

Figure 4.1: Graphical summaries of mean sway range in forward-backward plane

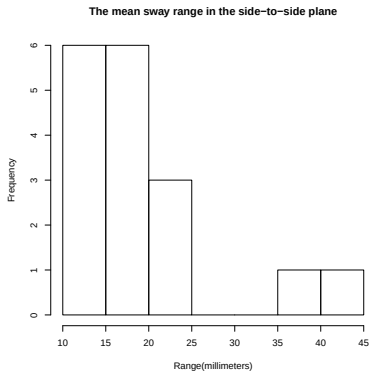


(a) Histogram

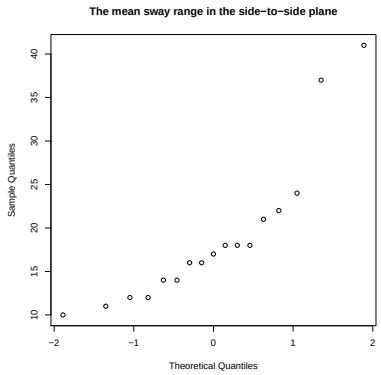


(b) Q-Q plot

Figure 4.2: Graphical summaries of mean sway range in side-to-side plane



(a) Histogram



(b) Q-Q plot

Figure 4.1 and Figure 4.2 show that the histogram and Q-Q plot of the mean sway range in the forward-backward plane and the mean sway range in the side-to-side plane respectively. Both histogram and Q-Q plot show that the distributions of the mean sway range in the forward-backward plane and the side-to-side plane are

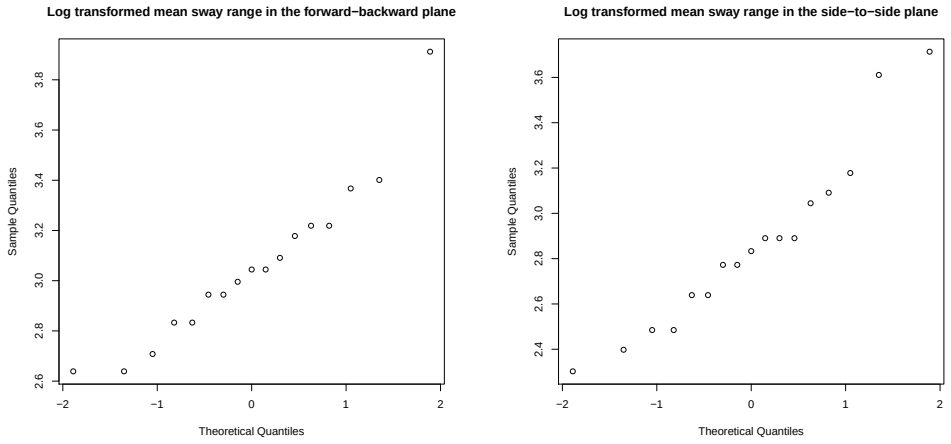
positively skewed (right skewed).

Table 4.1: Descriptive statistics

	Forward-backward	Side-to-side
Mean	22.47	18.88
Median	21	17
Standard Deviation	8.54	8.53

Table 4.1 also indicates that the distributions of the mean sway range in the forward-backward plane and the side-to-side plane are positively skewed (Mean > Median).

Figure 4.3: Q-Q plots for log-transformed mean sway range



(a) Forward-backward (b) Side-to-side

Figure 4.3 shows the Q-Q plots for log transformed mean sway range in the forward-backward plane and the side-to-side plane and the distributions of the transformed data are more nearly normal. This suggests that the distribution of the mean sway range in the forward-backward plane and the side-to-side plane might follow log-normal distribution, but without more information, we cannot predict accurately. We use the Kolmogorov-Smirnov (KS) test to assess the appropriateness of the

log-normal distribution for the data as follows,

H_0 : The mean sway range follows a log-normal distribution

H_a : The mean sway range does not follow a log-normal distribution.

Table 4.2: KS test statistic and p-value

	Forward-backward	Side-to-side
KS test statistic	0.1262	0.1747
p-value	0.9494	0.6773

Table 4.2 shows that for both planes, we don't have sufficient evidence to reject the log-normal fit. So we can assume that the log-normal model adequately describes the mean sway ranges in the forward-backward plane and the side-to-side plane.

Table 4.3 gives the method of moment estimates (MME) and the maximum likelihood estimates (MLE) of log-normal parameters for the mean sway ranges in the forward-backward plane and the side-to-side plane. The MME's and the MLE's of μ and σ^2 are very close for both planes.

Table 4.3: The MME's and the MLE's of parameters

	Forward-backward	Side-to-side
MME of μ	3.05	2.85
MME of σ^2	0.13	0.18
MLE of μ	3.04	2.86
MLE of σ^2	0.11	0.15

For illustration purpose, we assume that a practically meaningful test is provided by

$\delta' = 0.01$ millimeters. Accordingly the hypotheses to be tested are given below

$$H_0 : \frac{E(X_1)}{E(X_2)} \geq 1.01 \qquad \text{vs.} \qquad H_a : \frac{E(X_1)}{E(X_2)} < 1.01$$

$$H_0 : \eta_1 - \eta_2 \geq 0.01 \qquad \text{vs.} \qquad H_a : \eta_1 - \eta_2 < 0.01. \tag{4.1}$$

We obtain the generalized p-value of 0.86 using the algorithm describe in Chapter 3. On the other hand, the Z-score approach gives a p-value of 0.92. All these results lead to the conclusion that the data do not provide sufficient evidence to indicate that the ratio of the means of the mean sway ranges in the forward-backward plane and the side-to-side plane is less than 1.01.

We now illustrate the Bayesian approach discussed in Chapter 3. Note that the prior distribution of location parameter (μ) is normal distribution. When no information is available on μ, a usual choice for the prior mean of location parameter (μ_0) is zero. Note that this prior choice centers the prior belief of μ around zero. To make this prior diffuse, we can set the prior variance of location parameter (σ_0^2) to a large value, (for example 10000). The prior distribution of scale parameter (σ^2) is inverse-gamma distribution. We set α and β to a small value, $\alpha = \beta = 0.01$.

One can also take an empirical Bayes approach in selecting prior parameters. For example one can set the prior variance to a very small value (0.01). The mean (μ) and the variance (σ^2) of the mean sway range in the forward-backward plane are 3.06 and 0.10 respectively. The prior distribution of the location parameter (μ) is the normal distribution. We have set prior mean of location parameter (μ_0) to 3.06 and the prior variance of location parameter (σ_0^2) to 0.01. The prior distribution of

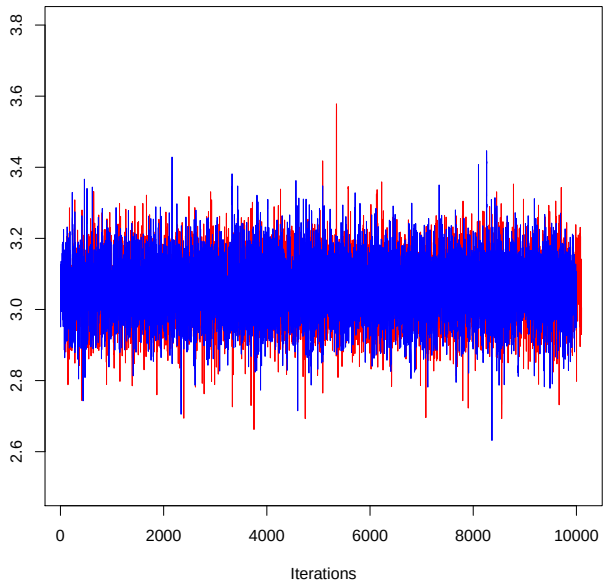
scale parameter (σ^2) is inverse-gamma distribution. We have set prior mean of scale parameter to 0.10 and the prior variance of scale parameter to 0.01.

$$E(\sigma^2) = \frac{\beta}{\alpha - 1} = 0.10 \qquad \text{and} \qquad Var(\sigma^2) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = 0.01.$$

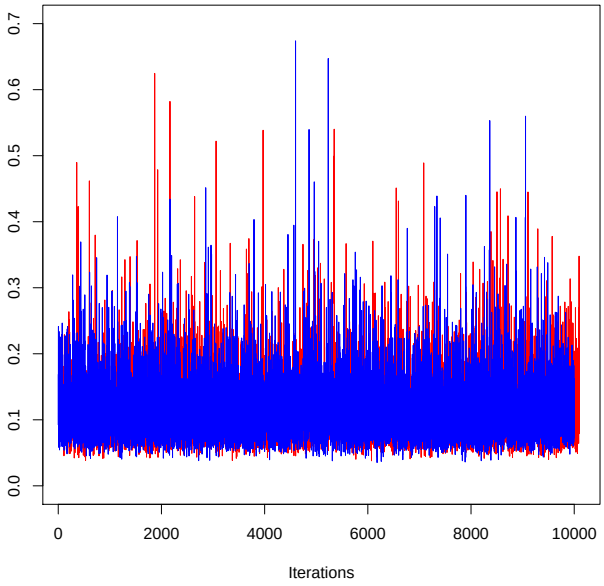
This will give $\alpha = 3$ and $\beta = 0.2$. Similarly, for the mean sway range in the side-to-side plane, the values of hyper parameters are $\mu_0 = 3.04, \sigma_0^2 = 0.01, \alpha = 4.3$ and $\beta = 0.5$. We perform the Bayesian analysis under these prior settings. We obtain two random samples (two chains) with 10,000 observations from the posterior distributions for μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane. So we have run two chains for 14,000 iterations with a burn-in of 4,000.

The first and easiest way to check the convergence of the MCMC algorithm is to visually inspect the trace plots or history plots of μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane. Figure 4.4 and Figure 4.5 show the history plots of μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane. Here we have run two chains simultaneously and each chain shows in different colour (red or blue). The plots do not show any particular patterns and all the chains appear to be overlapping one another and mixing well.

Figure 4.4: History plots of MCMC sample for forward-backward plane

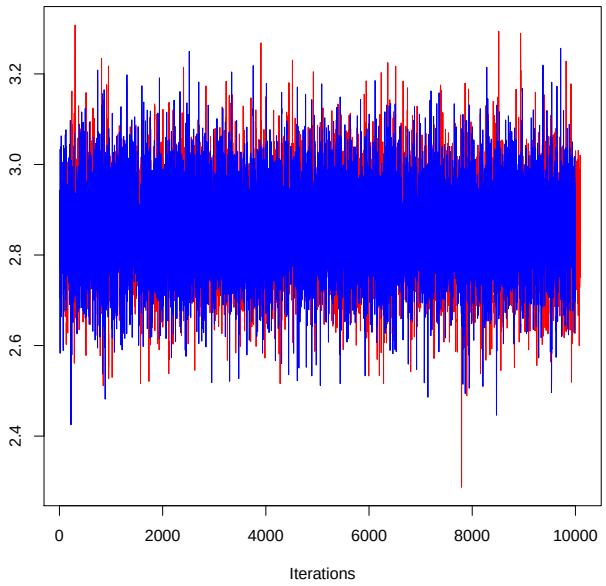


(a) History plot of μ

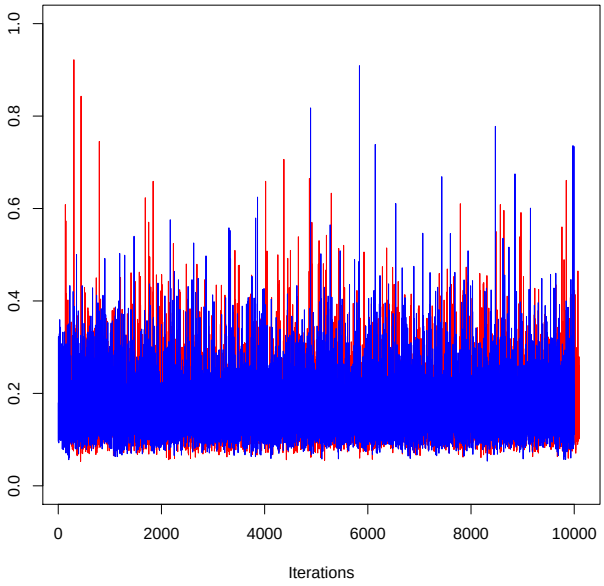


(b) History plot of σ^2

Figure 4.5: History plots of MCMC sample for side-to-side plane



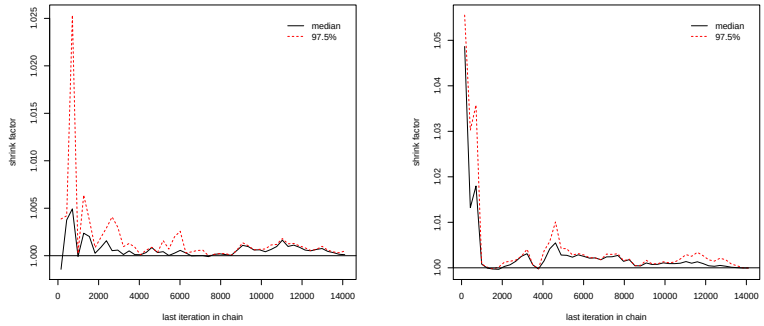
(a) History plot of μ



(b) History plot of σ^2

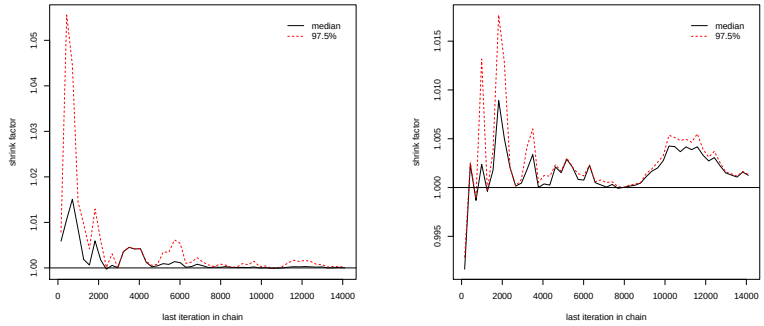
To have a more precise view on convergence, we use the Brooks-Gelman diagnostics which calculates the Brooks-Gelman convergence statistics (R_c) as describe in section 3.3.2. Figure 4.6 and Figure 4.7 plot the Brooks-Gelman diagnostic for μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane for two chains of 14000 iterations. For the convergence, R_c should be close to 1 and R_c is indeed close to 1, roughly after 3000 iterations. So we have considered a burn-in period of 4000 iterations in obtaining posterior summaries.

Figure 4.6: Gelman-Rubin diagnostic for forward-backward plane



(a) Gelman-Rubin diagnostic of μ (b) Gelman-Rubin diagnostic of σ^2

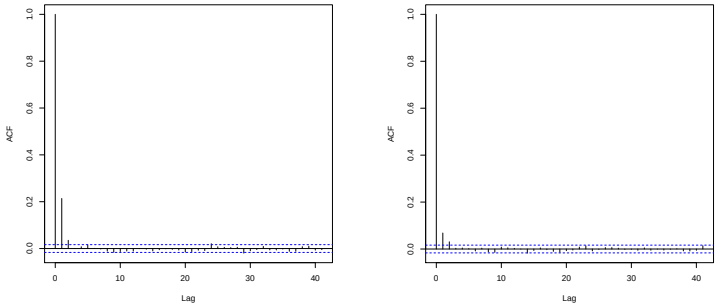
Figure 4.7: Gelman-Rubin diagnostic for side-to-side plane



(a) Gelman-Rubin diagnostic of μ (b) Gelman-Rubin diagnostic of σ^2

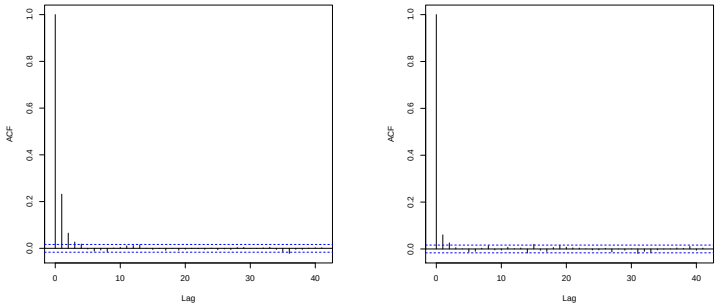
The next step is to obtain random samples after the chains are settled down to the posterior distribution. Figure 4.8 and Figure 4.9 show the autocorrelations by lag for μ and σ^2 of the mean sway ranges in the forward-backward plane and the side-to-side plane for two chains of 6000 iterations. The correlation at lag 0 is 1 and the rest of the correlations are very close to zero. It indicates that we have a random sample from the posterior distribution and thinning is not needed in this case.

Figure 4.8: Autocorrelation plots for forward-backward plane



(a) Autocorrelation plot of μ (b) Autocorrelation of σ^2

Figure 4.9: Autocorrelation plots for side-to-side plane



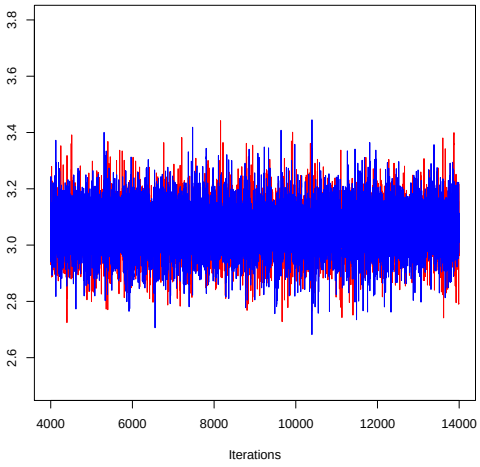
(a) Autocorrelation of μ (b) Autocorrelation of σ^2

Table 4.4: Posterior estimates of the parameters

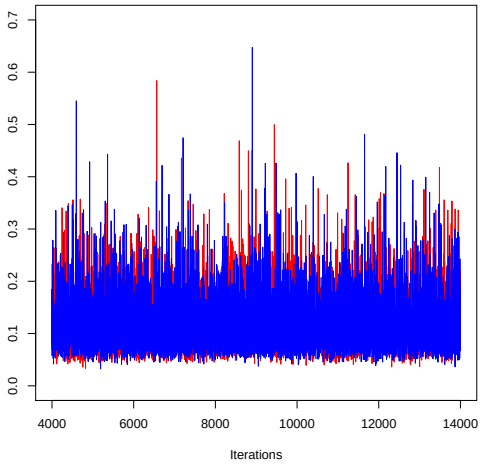
	Parameter	Estimate	MC Error	SD	5%*SD
Forward-Backward	μ	3.05	0.00062	0.082	0.0041
	σ^2	0.12	0.00037	0.046	0.0023
Side-to-Side	μ	2.84	0.00067	0.101	0.0051
	σ^2	0.16	0.00052	0.070	0.0035

Table 4.4 reports the posterior summaries, Monte Carlo error (MC error) and sample standard deviation for μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane. The MC errors of each parameters of both planes less than 5% of the sample standard deviations. We have obtained the random samples with 10,000 independent and identically distributed observations from the posterior distributions.

Figure 4.10: History plots of MCMC sample after burn-in and thinning for forward-backward plane

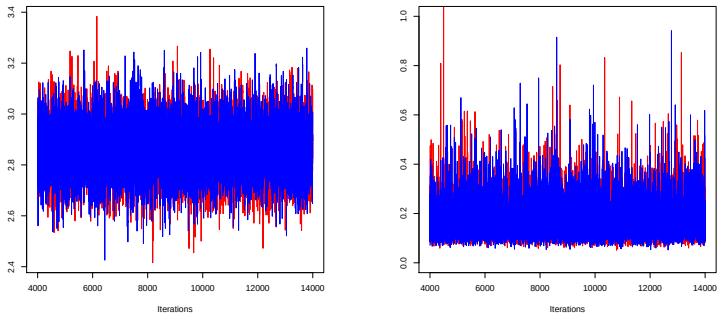


(a) History plot of μ



(b) History plot of σ^2

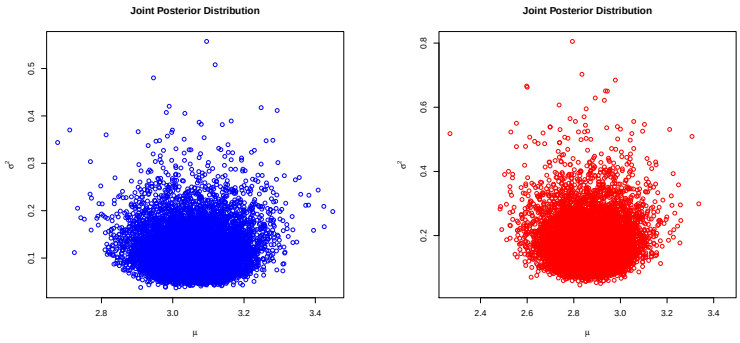
Figure 4.11: History plots of MCMC sample after burn-in and thinning for side-to-side plane



(a) History plot of μ (b) History plot of σ^2

Figure 4.10 and Figure 4.11 show the history plots of μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane after burn-in. Figure 4.12 shows the joint posterior distribution of μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane for the 10,000 simulations. Figure 4.13 combines both joint posterior distribution of μ and σ^2 of the mean sway range in the forward-backward plane and the side-to-side plane into one graph for comparison purpose. The plot indicates very little difference in two distributions.

Figure 4.12: Joint posterior distribution of each plane



(a) Forward-backward plane (b) Side-to-side plane

Figure 4.13: Joint posterior distribution of both planes

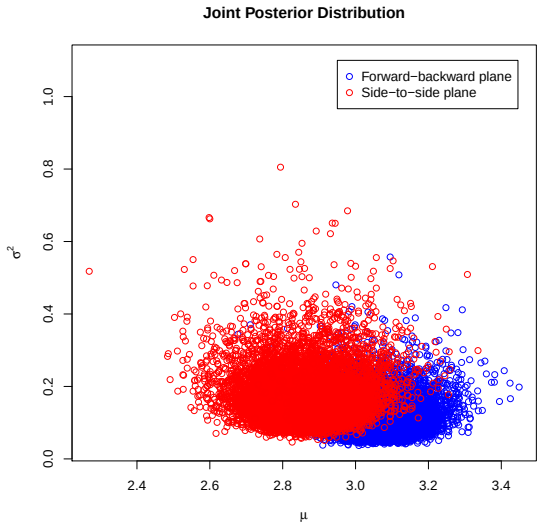
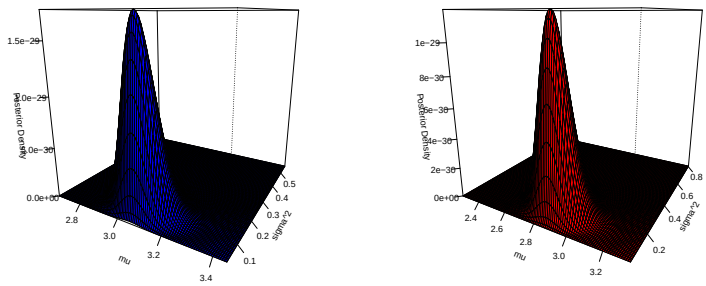


Figure 4.14 shows the perspective plots for the joint posterior of the mean sway range in the forward-backward plane and the side-to-side plane. These perspective plots make 3D plots of Posterior distributions of a surface over μ-σ^2 plane. The plots clearly show that the joint posterior is unimodal indicating that MCMC analysis is more meaningful. Therefore we can now make inferences on η_1 and η_2.

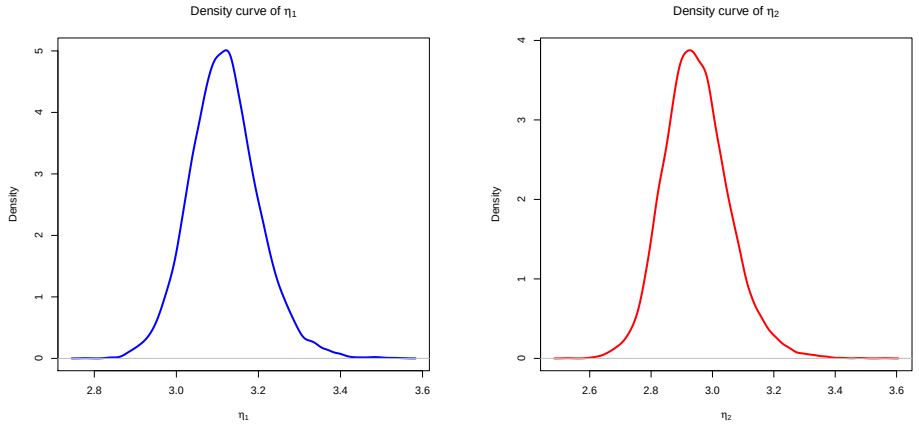
Figure 4.14: Perspective plots of the posterior distribution



(a) Forward-backward plane (b) Side-to-side plane

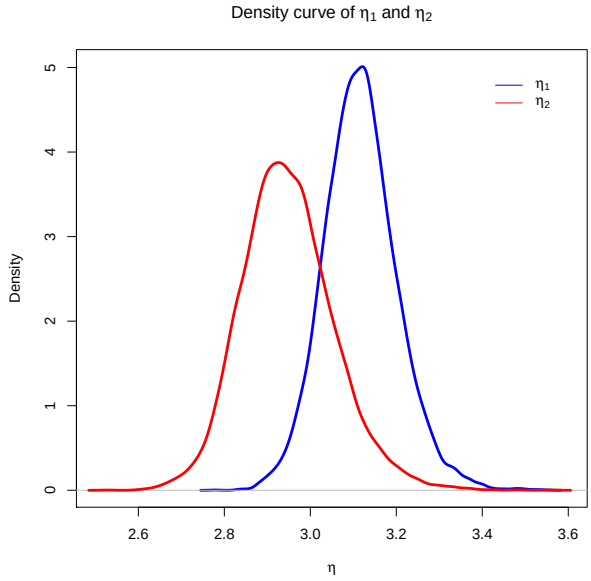
Note that our interest is on η_1 and η_2. Figure 4.15 shows the density curves of η_1 and η_2. Figure 4.16 combines both density curves of η_1 and η_2 into one graph.

Figure 4.15: Density curves of η_1 and η_2



(a) Forward-backward plane (b) Side-to-side plane

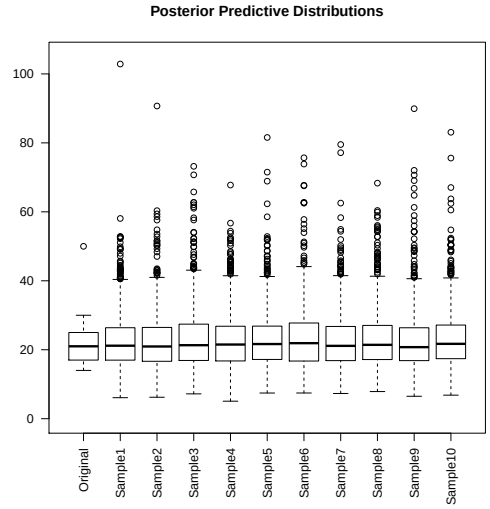
Figure 4.16: Density curve of η_1 and η_2 in one graph



We computed Bayesian probability under H_0 ($P(H_0|\boldsymbol{x})$) for testing hypotheses in (4.1) and the value is 0.88. Since $P(H_0|\boldsymbol{x})$ is sufficiently large, we decide in favor of Null Hypothesis (H_0). This leads to the conclusion that the data do not provide the sufficient evidence to indicated that the ratio of the means of the mean sway ranges in the forward-backward plane and the side-to-side plane is less than 1.01. We also examine the prediction ability of the model using posterior predictive distribution described in Section 3.4.

Figure 4.17 shows the posterior predictions of 10 samples with 10,000 observations for the mean sway ranges in the forward-backward plane. The first boxplot represents the original sample and the next 10 boxplots represent the predicted samples. We compare the last 10 boxplots (predicted samples) with the first boxplot (original sample) and the observed dataset appears to be consistent with the generated datasets.

Figure 4.17: Posterior prediction for forward-backward plane



Furthermore, Table 4.5 compares the descriptive statistics of the predicted samples with the original sample of the mean sway range in the forward-backward plane. The mean and median of each predicted samples are approximately equal to the mean and median of the original sample. But the inter-quartile range of predicted samples are slightly higher than the inter-quartile range of the original sample. Figure 4.17 and Table 4.5 lead to the conclusion that there is no major difference between predicted samples and the original sample of the mean sway range in the forward-backward plane.

Table 4.5: Descriptive statistics of posterior prediction for forward-backward plane

	Mean	Median	Inter-quartile range (IQR)
Original	22.47	21	8.00
Sample1	22.52	21.17	9.38
Sample2	22.36	20.95	9.81
Sample3	23.02	21.32	10.48
Sample4	22.84	21.51	10.03
Sample5	22.87	21.65	9.64
Sample6	23.10	21.91	11.03
Sample7	22.60	21.13	9.90
Sample8	23.05	21.42	9.86
Sample9	22.60	20.75	9.51
Sample10	23.07	21.71	9.70

Similarly, Figure 4.18 shows the posterior predictions of 10 samples with 10,000 observations for the mean sway range in the side-to-side plane. The first boxplot represents the original sample and the next 10 boxplots represent the predicted samples. We have compared the last 10 boxplots (predicted samples) with the first

boxplot (original sample) and it seems that no difference can be claimed.

Table 4.6 compares the descriptive statistics of the predicted samples with the original sample of the mean sway range in the side-to-side plane. The mean and median of each predicted samples are approximately equal to the mean and median of the original sample. But the inter-quartile range of predicted samples are higher than the inter-quartile range of the original sample. Figure 4.17 and Table 4.5 lead to the conclusion that there is no major difference between predicted samples and the original sample of the mean sway range in the side-to-side plane.

Figure 4.18: Posterior prediction for side-by-side plane

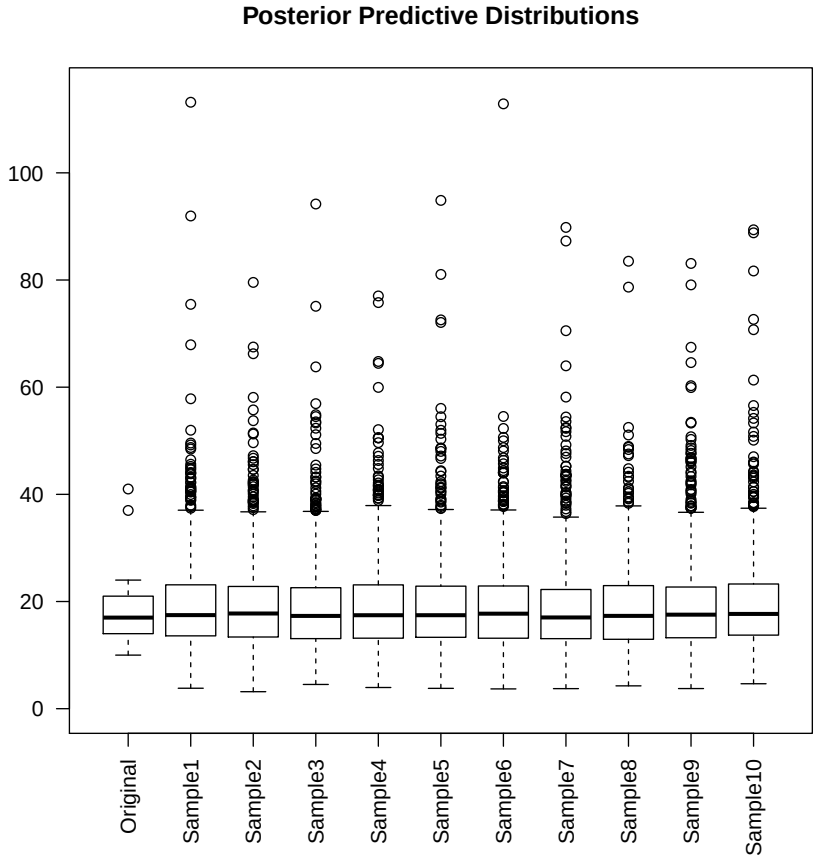


Table 4.6: Descriptive statistics of posterior prediction for side-by-side plane

	Mean	Median	Inter-quartile range (IQR)
Original	18.88	17	8.53
Sample1	19.28	17.46	9.51
Sample2	19.26	17.77	9.43
Sample3	18.88	17.31	9.49
Sample4	18.99	17.44	9.94
Sample5	19.04	17.45	9.52
Sample6	19.06	17.73	9.73
Sample7	18.88	17.03	9.17
Sample8	18.79	17.32	10.00
Sample9	19.17	17.55	9.45
Sample10	19.61	17.68	9.53

We also conducted various simulation studies to compare the performance of three methods; Z-score, generalized p-value and Bayesian approaches. For Z-score and generalized p-value methods, the type I error probabilities for various combinations of $n_1, n_2, \mu_1, \sigma_1^2$ and σ_2^2 are reported in Table 4.7. For the simulations, we have taken $\mu_2 = 0$, without loss of generality. The type I error probability corresponds to the parameter combinations satisfying $\mu_1 + \frac{1}{2}\sigma_1^2 - \mu_2 + \frac{1}{2}\sigma_2^2 = 0.01$ with the significance level 5%.

For smaller values of n_1 and n_2 ($n_1 < 25$ and $n_2 < 25$), Z-score test has type I error probabilities which are too small, whereas the test based on generalized p-value is extremely satisfactory for controlling type I error probability. For $H_a : \eta_1 - \eta_2 < 0.01$, the Z-score test type I error probabilities which are too small when $n_1 > n_2$, and they are too large when $n_1 < n_2$. The histogram in Figure 4.19 (d), is skewed to the right and the corresponding type I error probability to the Z-score test is too small. On the other hand, the histogram of Z-score statistics in Figure 4.19 (f), is

skewed to the left and the corresponding type I error probability to the Z-score test is too large. But, the test based on the generalized p-value controls type I error quite satisfactorily, regardless of n_1 and n_2.

The power for various combinations of $n_1, n_2, \mu_1, \sigma_1^2$ and σ_2^2 are reported in Table 4.8. For $H_a : \eta_1 - \eta_2 < 0.01$, when $n_1 > n_2$, the Z-score test has a smaller power compared to the generalized p-value test. On the other hand when $n_1 < n_2$, the Z-score test has a larger power compared to the generalized p-value test. Also, there are situations where the power of the Z-score test is smaller than the significance level of 0.05 (last row in Table 4.8), indicating that Z-score test is biased.

Figure 4.19: Histogram of the Z-score statistic, the numbers in parenthesis represent $(n_1, \mu_1, \sigma_1^2, n_2, \mu_2, \sigma_2^2)$

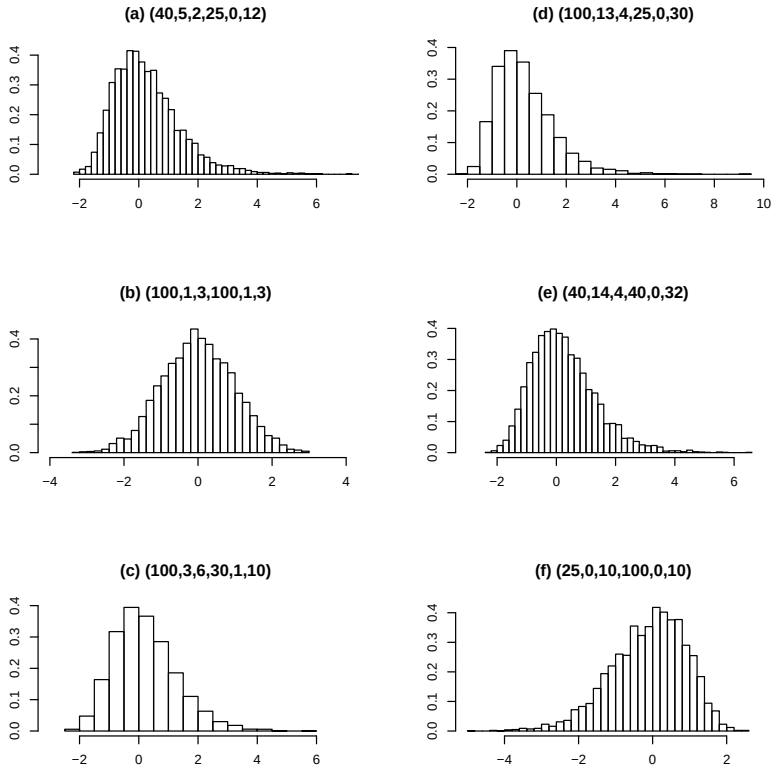


Table 4.7: Size of the generalized p-value test and the Z-score test at 5% significance level when $\mu_2 = 0$ and $H_0 : \eta_1 - \eta_2 \geq 0.01$ vs. $H_a : \eta_1 - \eta_2 < 0.01$

n_1	n_2	μ_1	σ_1^2	σ_2^2	Size	
					Generalized p-value	Z-Score
4	4	1.01	2	4	0.0338	0.0098
		0.01	3	3	0.0374	0.0375
		5.01	2	12	0.0344	0.0001
		0.01	12	12	0.0398	0.0166
10	10	1.01	2	4	0.0427	0.0133
		0.01	3	3	0.0464	0.0388
		5.01	2	12	0.0424	0.0004
		0.01	12	12	0.0489	0.0260
25	25	0.01	1	1	0.0454	0.0436
		0.01	5	5	0.0510	0.0447
		0.01	10	10	0.0488	0.0382
		0.01	100	100	0.0505	0.0342
		2.01	4	8	0.0463	0.0199
		4.01	8	16	0.0473	0.0137
40	25	0.01	1	1	0.0516	0.0357
		0.01	5	5	0.0530	0.0340
		0.01	10	10	0.0533	0.0257
25	40	0.01	1	1	0.0490	0.0542
		0.01	5	5	0.0471	0.0558
		0.01	10	10	0.0502	0.0583
40	25	5.01	2	12	0.0483	0.0079
25	40	5.01	2	12	0.0489	0.0154
40	40	8.01	4	20	0.0469	0.0148
		14.01	4	32	0.0457	0.0121
100	25	0.01	1	1	0.0530	0.0293
		0.01	5	5	0.0514	0.0160
		0.01	10	10	0.0482	0.0118

n_1	n_2	μ_1	σ_1^2	σ_2^2	Size	
					Generalized p-value	Z-Score
25	100	0.01	1	1	0.0425	0.0666
		0.01	5	5	0.0478	0.0819
		0.01	10	10	0.0486	0.0847
100	25	13.01	4	30	0.0494	0.0058
25	100	13.01	4	30	0.0445	0.0262
100	100	13.01	4	30	0.0514	0.0225
200	50	0.01	1	1	0.0534	0.0356
		0.01	5	5	0.0502	0.0259
		0.01	10	10	0.0498	0.0228
50	200	0.01	1	1	0.0439	0.0535
		0.01	5	5	0.0532	0.0711
		0.01	10	10	0.0477	0.0739
200	200	24.01	12	60	0.0500	0.0319
		40.01	2	82	0.0491	0.0272

Zhou et al. claimed that, when both n_1 and n_2 are large, the distribution of the Z-score statistic is approximately standard normal under H_0. The histogram of the Z-score statistic in Figure 4.19 indicate that the distribution of Z-score statistic is skewed when both sample sizes are large as 100 but (n_1, μ_1, σ_1^2) is not approximately equal to (n_2, μ_2, σ_2^2). So the normal approximation is appropriate only when both samples are very large and (n_1, μ_1, σ_1^2) is approximately equal to (n_2, μ_2, σ_2^2).

Table 4.9 shows the generalized p-value and the Bayesian probabilities for testing hypothesis (4.1). The generalized p-value has performed well for both large and small samples. The Bayesian probabilities also perform very well and indicate a very good agreement in conclusion based on the Bayes factor.

Table 4.8: Power of the generalized p-value test and the Z-score test at 5% significance level when $\mu_2 = 0$ and $H_0 : \eta_1 - \eta_2 \geq 0.01$ vs. $H_a : \eta_1 - \eta_2 < 0.01$

n_1	n_2	μ_1	σ_1^2	σ_2^2	Power	
					Generalized p-value	Z-Score
4	4	0	4	12	0.1592	0.0347
		0	4	20	0.2650	0.0358
		3	4	12	0.0525	0.0018
		4	1	11	0.0693	0.0002
10	10	0	4	12	0.3961	0.229
		0	4	20	0.6753	0.4149
		3	4	12	0.0849	0.0056
		4	1	11	0.1113	0.0041
25	25	1	1	5	0.3760	0.2302
		1	5	9	0.1536	0.0925
		1	10	14	0.1045	0.0658
		0	2	4	0.3582	0.3136
		0	7	9	0.1369	0.1066
		0	1	4	0.7568	0.6807
40	25	1	1	5	0.4051	0.2365
		1	5	9	0.1766	0.0821
		1	10	14	0.1111	0.0539
25	40	1	1	5	0.4656	0.3780
		1	5	9	0.1600	0.1566
		1	10	14	0.1003	0.1059
40	25	1	4	9	0.3050	0.1716
		1	9	14	0.1598	0.0831
25	40	1	4	9	0.3005	0.2746
		1	9	14	0.1527	0.1489
100	25	1	1	5	0.4187	0.2344
		1	5	9	0.2003	0.0777
		1	10	14	0.1250	0.0400
25	100	1	1	5	0.6665	0.6602
		1	5	9	0.1895	0.2485
		1	10	14	0.1097	0.1724
100	25	0	1	2	0.4192	0.2740
		0	1	3	0.7050	0.5452
25	100	0	1	2	0.4006	0.4751
		0	1	3	0.8155	0.8511

n_1	n_2	μ_1	σ_1^2	σ_2^2	Power	
					Generalized p-value	Z-Score
100	25	1	4	7	0.1294	0.0434
		1	9	12	0.0871	0.0262
25	100	1	4	7	0.1235	0.1608
		1	9	12	0.0783	0.1246
50	200	1	1	4	0.5116	0.5043
		1	5	8	0.1495	0.1860
		1	10	14	0.1645	0.2168
200	50	1	1	4	0.3175	0.2130
		1	5	8	0.1469	0.0765
		1	10	14	0.1646	0.0954
		0	1	2	0.6126	0.5150
		0	1	3	0.9030	0.8490
50	200	0	1	2	0.6831	0.7067
		0	1	3	0.9801	0.9847
300	25	0.8	12	18	0.0589	0.0228

Table 4.9: The generalized p-value and the Bayesian probabilities under H_0 and H_a when $\mu_2 = 0$ and $H_0 : \eta_1 - \eta_2 \geq 0.01$ vs. $H_a : \eta_1 - \eta_2 < 0.01$

n_1	n_2	μ_1	σ_1^2	σ_2^2	Generalized p-value	$P(H_0\|\boldsymbol{x})$	$P(H_1\|\boldsymbol{x})$	Bayes Factor
4	4	0	14	3	0.7334	0.9561	0.0439	23.6883
4	4	0	3	14	0.0121	0.3709	0.6291	0.6413
4	4	0.01	3	3	0.6160	0.7212	0.2788	2.8136
25	25	0	3	0	0.9994	1.0000	0.0000	∞
25	25	0	3	14	0.0001	0.0000	1.0000	0
25	25	0.01	3	3	0.3425	0.9154	0.0846	11.7689
25	4	0	3	0	0.9998	0.9977	0.0023	471.8105
25	4	0	3	14	0.0001	0.2085	0.7915	0.2865
25	4	0.01	3	3	0.3305	0.8804	0.1196	8.0065
4	25	0	3	0	0.9916	0.7483	0.2517	3.2336
4	25	0	3	14	0.0249	0.0324	0.9676	0.0364
4	25	0.01	3	3	0.8359	0.9298	0.0702	14.4061
100	100	0	3	0	1.0000	1.0000	0.0000	∞
100	100	0	3	14	0.0000	0.0000	1.0000	0
100	100	0.01	3	3	0.1671	0.8388	0.1612	5.6596

Chapter 5

Conclusion

In this thesis, we derived a Bayesian inference procedure for testing non-inferiority hypotheses for the ratio of two independent log-normal means. In non-inferiority studies, the objective is to demonstrate the new treatment or the product is not worse than the existing treatment or the product by more than a pre-specified, small amount (δ). The determination of the non-inferiority margin (δ), is a critical step in non-inferiority hypothesis testing and we have stated two basic approaches in section 1.2.1.

The mean of the log-normal distribution is a function of both μ and σ^2 and it is difficult to obtain exact or optimum tests for testing non-inferiority hypotheses. Zhou et al. (1997) have introduced the Z-score approach for the problem of testing the equality of two log-normal means and we have extended this idea for testing non-inferiority hypotheses. This approach is recommended for large samples but for small samples the power and the type I error rates are too conservative or too liberal. Although, the samples are large, still there are situations where the Z-score test is unsatisfactory. But, the Z-score test performs well in terms of power and type

I error rates when the μ's and σ^2's close to each other under large sample sizes.

There are also few challenges with the classical p-value method in hypothesis testing: such as difficulty finding the suitable test statistics, difficulty finding the sampling distribution of the test statistic and many nuisance parameters. To overcome these challenges, Weerahandi and Tsui (1989) introduced the concept of the generalized p-value approach. We extended the generalized p-value approach for testing non-inferiority hypotheses. The generalized p-value approach controls type I error rates quite satisfactorily, regardless of n_1 and n_2. The generalized p-value approach performed well for both large and small samples but the Bayesian approach also showed a good agreement with conclusion in non-inferiority hypothesis testing.

We remark that further works need to be done in order to quantify the evidence of the strength of superiority of each method discussed in the thesis. This requires developing new summary measures for comparing performance of the methods.

Future research can be done to extend this non-inferiority hypothesis testing concept to multi-arm trials. For instance, two-arm non-inferiority trials exhibit some major challenges in the design and analysis and sometimes many of the trials fail to show a statistically significant difference between the new treatment and an existing treatment. It may be difficult or impossible to evaluate whether the trials have adequate assay sensitivity. The inclusion of placebo arm attempts to answer whether the new treatment is as effective as active control given that the new treatment is effective compared to placebo. More specifically, we would like to extend the approach described by Gamalo et al. (2013) for log-normal data. As another direction of future research, we would like to extend the Bayesian approach to accommodate

potential covariates into the non-inferiority hypotheses testing mechanism. This will require utilizing generalized linear models within a Bayesian framework.

Bibliography

[1] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

[2] S. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.

[3] L. Eckhard, A. Werner, and A. Markus. Log-normal distributions across the sciences: Keys and clues. *American Institute of Biological Sciences*, 51:341–352, 2001.

[4] M. Gamalo, S. Muthukumarana, P. Ghosh, and R. Tiwari. A generalized p-value approach for assesing non-inferiority in a three-arm trial. *Statistical Methods in Medical Research*, 22:261–277, 2013.

[5] A. Gelman, J.B. Carlin, H. S. Stern, and D. B. Rublin. *Bayesian Data Analysis*. Chapman and Hall, 1996.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[7] W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[8] S. Karlin and H. Taylor. *A First Course in Stochastic Processes*. San Diego: Academic Press, 1st edition, 1975.

[9] K. Koti. New tests for assessing non-inferiority and equivalence from survival data. *Open Journal of Statistics*, 3(2):55–64, 2013.

[10] K. Krishnamoorthy and T. Mathews. Inferences on the means of lognormal distributions using generalized p-value and generalized confidence interval. *Journal of Statistical Planning and Inference*, 115:103–121, 2003.

[11] E. Lehmann and J. Romano. *Testing Statistical Hypotheses*. Springer, 3rd edition, 2005.

[12] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21:1087–1091, 1953.

[13] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949.

[14] A. Munk and H. Trampisch. Therapeutic equivalence - clincal issues and statistical methodology in non-inferiority trials. *Biometrical Journal*, 47:7–9, 2005.

[15] International Conference on Harmonization. Guidance on choice of control group and related design and conduct issues in clinical trials. Food and Drug Administration, July 2000.

[16] S. Snapinn. Noninferiority trials. *Current Allergy Reports Control Trials Cardiovascular Medicine*, 1:19–21, 2000.

[17] H. B. Stauffer. *Contemporary Bayesian and Frequentist Statistical Research Methods for Natural Resource Scientists*. A John Wiley & Sons, Inc., 2008.

[18] N. Teasdale, J. LaRue, C. Bard, and M. Fleury. On the cognitive penetrability of posture control. *Experimental Aging Research*, 19:1–13, 1993.

[19] E. Walker and A. Nowacki. Understanding equivalence and non-inferiority test. *Journal of General Internal Medicine*, 26:1992–1996, 2011.

[20] S. Weerahandi. *Exact Statistical Methods for Data Analysis*. New York:Springer-Verlag, 1995.

[21] S. Weerahandi and K. Tsui. Generalized p-value in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84:602–607, 1989.

[22] S. Wellek. *Testing Statistical Hypotheses of Equivalence*, volume 22. Chapman and Hall/CRC, 2003.

[23] X. Zhou, S. Gao, and S. Hui. Methods for comparing the means of two independent log-normal samples. *Biometrics*, 53:1129–1135, 1997.

Appendix A

R codes

```
#Checking Log-normality

install.packages("nortest")

library("nortest")

install.packages("MASS")

library("MASS")

w<-matrix(1,nrow=4,ncol=2)

#Reading data

Data=read.table("Data3.txt")

Data1=Data[,2]

Data2=Data[,3]

#Histograms

pdf("hist1.pdf")

hist(Data1,xlab="Range(millimeters)", main=

"The mean sway range in the forward-backward plane")
```

```
dev.off()
pdf("hist2.pdf")
hist(Data2,xlab="Range(millimeters)", main=
"The mean sway range in the side-to-side plane")
dev.off()

#Anderson-Darling test
w[1,1]=round(ad.test(Data1)$p.value,3)
w[2,1]=round(ad.test(Data2)$p.value,3)

#Kolmogorov-Smirnov test
w[1,2]=round(lillie.test(Data1)$p.value,3)
w[2,2]=round(lillie.test(Data2)$p.value,3)

#Log transformation
logtrans1=log(Data1)
logtrans2=log(Data2)


w[3,1]=round(ad.test(logtrans1)$p.value,3)
w[4,1]=round(ad.test(logtrans2)$p.value,3)
w[3,2]=round(lillie.test(logtrans1)$p.value,3)
w[4,2]=round(lillie.test(logtrans2)$p.value,3)

write.table(w,sep="\t",file="Output.txt",row.names=c("Original F-B",
 "Original S-S","log transformed F-B","log transformed S-S"),
 col.names=c("AD Test","KS Test"))
```

```
pdf("qqplot1.pdf")
qqnorm(Data1, main="The mean sway range in the forward-backward plane")
dev.off()
pdf("qqplot2.pdf")
qqnorm(Data2,main="The mean sway range in the side-to-side plane ")
dev.off()
pdf("qqplot3.pdf")
qqnorm(logtrans1,main="Log transformed mean sway range in the
forward-backward plane ")
dev.off()
pdf("qqplot4.pdf")
qqnorm(logtrans2, main="Log transformed mean sway range in the
side-to-side plane ")
dev.off()

fitdistr(Data1,"lognormal")
ks.test(Data1,"plnorm", 3.05962417,0.30818322)

fitdistr(Data2,"lognormal")
ks.test(Data2,"plnorm", 2.86094253,0.37597521)

##############################################################################

#Calculating Size using GP-Test and Z-score

m1=m2=10000
```

```
delta=0.01
alpha=0.05
z=vector("numeric",m1)
k=vector("numeric",m1)
d=vector("numeric",m2)

Zscorefun<-function(n1,n2,mean1,var1,mean2,var2){
  for(i in 1:m1){
    X1=rlnorm(n1,mean1,sqrt(var1))
    Y1=log(X1)

    X2=rlnorm(n2,mean2,sqrt(var2))
    Y2=log(X2)

    Ybar1=mean(Y1)
    S21=var(Y1)
    S41=(var(Y1))^2
    Ybar2=mean(Y2)
    S22=var(Y2)
    S42=(var(Y2))^2

    z[i]= ((Ybar1-Ybar2)+((S21 /2 )-(S22 /2)))/sqrt((S21/n1)+
    (S22/n2)+(S41/(2*(n1-1))) +(S42/(2*(n2-1))))+delta

    if(z[i] < qnorm(alpha)){
      k[i]=1
    }
    else{
```

```
      k[i]=0
    }
  }
    return(k)
}

Tablefun<-function(n1,n2,mean1, var1, mean2, var2){

  W<-matrix(1,nrow=1,ncol=7)

  B=Zscorefun(n1,n2,mean1, var1, mean2, var2)
  z=sum(B)/m1
  #U.CI1=quantile(A[,1],probs=0.95)
  #U.CI2=quantile(A[,1],probs=0.99)

  W[1,1]=n1
  W[1,2]=n2
  W[1,3]=mean1
  W[1,4]=var1
  W[1,5]=mean2
  W[1,6]=var2
  W[1,7]=z

  return(W)
}

na=c(4,4,4,4,10,10,10,10,25,25,25,25,25,25,40,40,40,25,25,25,40,25,
40,40,100,100,100,25,25,25,100,25,100,200,200,200,50,50,50,200,200)
```

```
nb=c(4,4,4,4,10,10,10,10,25,25,25,25,25,25,25,25,25,40,40,40,25,40,
40,40,25,25,25,100,100,100,25,100,100,50,50,50,200,200,200,200,200)
mua=c(1,0,5,0,1,0,5,0,0,0,0,0,2,4,0,0,0,0,0,0,0,5,5,8,14,0,0,0,0,
0,0,13,13,13,0,0,0,0,0,0,0,24,40)
mua=mua+0.01
sigmaa=c(2,3,2,12,2,3,2,12,1,5,10,100,4,8,1,5,10,1,5,10,2,2,4,
4,1,5,10,1,5,10,4,4,4,1,5,10,1,5,10,12,2)
mub=rep(0,41)
sigmab=c(4,3,12,12,4,3,12,12,1,5,10,100,8,16,1,5,10,1,5,10,12,
12,20,32,1,5,10,1,5,10,30,30,30,1,5,10,1,5,10,60,82)

D<-matrix(1,nrow=41,ncol=7)

for(i in 1:41){
  D[i,]=Tablefun(na[i],nb[i],mua[i],sigmaa[i],mub[i],sigmab[i])
}
write.table(D,sep="\t",file="ZSizedelta.txt",row.names=FALSE,
col.names=c("n1","n2","mu_1","sigma2_1","mu_2","sigma2_2","Z-score"))


#####################################################################

#Calculating power using GP-Test and Z-score

m1=m2=10000
delta=0.01
alpha=0.05
z=vector("numeric",m1)
```

```
k=vector("numeric",m1)
d=vector("numeric",m2)

Zscorefun<-function(n1,n2,mean1,var1,mean2,var2){
  for(i in 1:m1){
    X1=rlnorm(n1,mean1,sqrt(var1))
    Y1=log(X1)

    X2=rlnorm(n2,mean2,sqrt(var2))
    Y2=log(X2)

    Ybar1=mean(Y1)
    S21=var(Y1)
    S41=(var(Y1))^2
    Ybar2=mean(Y2)
    S22=var(Y2)
    S42=(var(Y2))^2
    #for(j in 1:m2){
    z[i]= ((Ybar1-Ybar2)+((S21 /2 )-(S22 /2)))/
    sqrt((S21/n1)+(S22/n2)+(S41/(2*(n1-1)))+
    (S42/(2*(n2-1))))+delta
    #}
    #q1[i]=quantile(d,prob=alpha)
    if(z[i] < qnorm(alpha)){
      k[i]=1
    }
    else{
```

```
      k[i]=0
    }
  }

  return(k)
}

Tablefun<-function(n1,n2,mean1, var1, mean2, var2){

  W<-matrix(1,nrow=1,ncol=7)

  B=Zscorefun(n1,n2,mean1, var1, mean2, var2)
  z=sum(B)/m1
  #U.CI1=quantile(A[,1],probs=0.95)
  #U.CI2=quantile(A[,1],probs=0.99)

  W[1,1]=n1
  W[1,2]=n2
  W[1,3]=mean1
  W[1,4]=var1
  W[1,5]=mean2
  W[1,6]=var2
  W[1,7]=z

  return(W)
}

na=c(4,4,4,4,10,10,10,10,25,25,25,25,25,25,40,40,40,25
,25,25,40,40,25,25,100,100,100,25,25,25,100,100,25,
```

```
25,100,100,25,25,50,50,50,200,200,200,200,200,50,50,25)
nb=c(4,4,4,4,10,10,10,10,25,25,25,25,25,25,25,25,25,40,
40,40,25,25,40,40,25,25,25,100,100,100,25,25,100,100,
25,25,100,100,200,200,200,50,50,50,50,50,200,200,300)
mua=c(0,0,3,4,0,0,3,4,1,1,1,1,0,0,0,1,1,1,1,1,1,1,1,1,1,
1,1,1,1,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0.8)
sigmab=c(12,20,12,11,12,20,12,11,5,9,14,4,9,4,5,9,14,5
,9,14,9,14,9,14,5,9,14,5,9,14,2,3,2,3,7,12,7,12,4,8,14,
4,8,14,2,3,2,3,14)
mub=rep(0,49)
sigmaa=c(4,4,4,1,4,4,4,1,1,5,10,2,7,1,1,5,10,1,5,10,4,
9,4,9,1,5,10,1,5,10,1,1,1,1,4,9,4,9,1,5,10,1,5,10,1,1,
1,1,12)

D<-matrix(1,nrow=49,ncol=7)
for(i in 1:49){
  D[i,]=Tablefun(na[i],nb[i],mua[i],sigmaa[i],mub[i],
  sigmab[i])
}
write.table(D,sep="\t",file="ZPower.txt",row.names=FALSE,
col.names=c("n1","n2","mu_1","sigma2_1","mu_2","sigma2_2",
"Z-score"))

#####################################################################

#Calculating Generalized p-value
```

```
local({r <- getOption("repos")
       r["CRAN"] <- "http://probability.ca/cran/"
       options(repos=r)})

# Set a vector of strings: package names to use (and install,
if necessary)
pkg_list = c('MCMCpack')

for (pkg in pkg_list)
{
  # Try loading the library.
  if ( ! library(pkg, logical.return=TRUE, character.only=TRUE) )
  {
    # If the library cannot be loaded, install it; then load.
    install.packages(pkg)
    library(pkg, character.only=TRUE)
  }
}

mu01=mu02=100
sigma01=sigma02=50

alpha1=alpha2=2
beta1=beta2=200

m2=m1=10000
```

```
delta=-0.01
alphapow=0.05

k=vector("numeric",m1)
T1=vector("numeric",m1)
T2=vector("numeric",m1)
T3=vector("numeric",m1)

UCIfun<-function(n1,n2,mean1, var1, mean2, var2){

  X1=rlnorm(n1,mean1,sqrt(var1))
  Y1=log(X1)

  X2=rlnorm(n2,mean2,sqrt(var2))
  Y2=log(X2)

  Ybar1=mean(Y1)
  s21=var(Y1)
  s1=sqrt(s21)
  eta1=mean1+(var1/2)

  Ybar2=mean(Y2)
  s22=var(Y2)
  s2=sqrt(s22)
  eta2=mean2+(var2/2)

  A=matrix(0,nrow=m1,ncol=4)
  C=matrix(0,nrow=m2,ncol=1)
```

```
for(i in 1:m1){
  Z1=rnorm(1,0,1)
  U21=rchisq(1,n1-1,0)
  U1=sqrt(U21)

  Z2=rnorm(1,0,1)
  U22=rchisq(1,n2-1,0)
  U2=sqrt(U22)

  T1[i]= Ybar1 - (Z1/(U1/sqrt(n1-1)))*(s1/sqrt(n1)) +
   s21/(2*U21/(n1-1))
  T2[i]= Ybar2 - (Z2/(U2/sqrt(n2-1)))*(s2/sqrt(n2)) +
   s22/(2*U22/(n2-1))

  T3[i]= T2[i] - T1[i]

  if(T3[i] <= delta){
    k[i]=1
  }
  else{
    k[i]=0
  }
}

for(j in 1:m1){
  A[j,1]=T1[j]
  A[j,2]=T2[j]
```

```
  A[j,3]=T3[j]
  A[j,4]=k[j]
}

mux1=numeric(m2)
sigma2x1=numeric(m2)
mux2=numeric(m2)
sigma2x2=numeric(m2)

sigma2x1[1]=sigma2x2[1]=4

for(i in 1:m2){

  b1=(1/sigma01)+(n1/sigma2x1[i])
  a1= ((mu01/sigma01)+(sum(Y1)/sigma2x1[i]))/b1
  b2=(1/sigma02)+(n2/sigma2x2[i])
  a2= ((mu02/sigma02)+(sum(Y2)/sigma2x2[i]))/b2

  mux1[i]=rnorm(1,a1,b1)
  mux2[i]=rnorm(1,a2,b2)

  c1=alpha1+(n1/2)
  d1=((sum((log(X1)-mux1[i])^2))+2*beta1)/2
  c2=alpha2+(n2/2)
  d2=((sum((log(X2)-mux2[i])^2))+2*beta2)/2

  sigma2x1[i]= rinvgamma(1,c1,d1)
  sigma2x1[i+1]=sigma2x1[i]
```

```
    sigma2x2[i]= rinvgamma(1,c2,d2)
    sigma2x2[i+1]=sigma2x2[i]
  }

  sigma2x1=sigma2x1[1:m2]
  sigma2x2=sigma2x2[1:m2]

  eta1=mux1+(sigma2x1 / 2)
  eta2=mux2+(sigma2x2 / 2)
  etadiff=eta1-eta2

  for(i in 1:m2){
    C[i]=etadiff[i]
  }


  return(cbind(A,C))
}

Tablefun<-function(n1,n2,mean1, var1, mean2, var2){

  W<-matrix(1,nrow=1,ncol=9)

  A=UCIfun(n1,n2,mean1, var1, mean2, var2)
  p=sum(A[,4])/m1
  kbay = subset(A[,5],subset=A[,5]>=delta)
  bayHo = length(kbay)/length(A[,5])
  bayH1=1-bayHo
```

```
  W[1,1]=n1
  W[1,2]=n2
  W[1,3]=mean1
  W[1,4]=var1
  W[1,5]=mean2
  W[1,6]=var2
  W[1,7]=p
  W[1,8]=bayHo
  W[1,9]=bayH1
  return(W)
}

W<-matrix(1,nrow=18,ncol=9)

W[1,]=Tablefun(4,4,0,14,0,3)
W[2,]=Tablefun(4,4,0,3,0,14)
W[3,]=Tablefun(4,4,0.01,3,0,3)
W[4,]=Tablefun(25,25,0,3,0,0)
W[5,]=Tablefun(25,25,0,3,0,14)
W[6,]=Tablefun(25,25,0.01,3,0,3)
W[7,]=Tablefun(25,4,0,3,0,0)
W[8,]=Tablefun(24,4,0,3,0,14)
W[9,]=Tablefun(25,4,0.01,3,0,3)
W[10,]=Tablefun(4,25,0,3,0,0)
W[11,]=Tablefun(4,25,0,3,0,14)
W[12,]=Tablefun(4,25,0.01,3,0,3)
```

```
W[13,]=Tablefun(100,100,0,3,0,0)
W[14,]=Tablefun(100,100,0,3,0,14)
W[15,]=Tablefun(100,100,0.01,3,0,3)
W[16,]=Tablefun(4,4,1,2,0,4)
W[17,]=Tablefun(25,25,1,2,0,4)
W[18,]=Tablefun(100,100,1,2,0,4)

write.table(W,sep="\t",file="Pvalue.txt",row.names=FALSE,
col.names=c("n1","n2","mu_1","sigma2_1","mu_2","sigma2_2",
"Pvalue","Bayesian_Ho","Bayesian_H1"))

############################################################

#Z-score histograms

m1=m2=5000
delta=0.01
alpha=0.05
z=vector("numeric",m1)
k=vector("numeric",m1)
d=vector("numeric",m2)

Zscorefun<-function(n1,n2,mean1,var1,mean2,var2){
  for(i in 1:m1){
    X1=rlnorm(n1,mean1,sqrt(var1))
    Y1=log(X1)

    X2=rlnorm(n2,mean2,sqrt(var2))
```

```
    Y2=log(X2)

    Ybar1=mean(Y1)
    S21=var(Y1)
    S41=(var(Y1))^2
    Ybar2=mean(Y2)
    S22=var(Y2)
    S42=(var(Y2))^2

    z[i]= ((Ybar1-Ybar2)+((S21 /2 )-(S22 /2)))/
    sqrt((S21/n1)+(S22/n2)+(S41/(2*(n1-1)))
    +(S42/(2*(n2-1))))+delta

  }
      return(z)
}

pdf("Zhist.pdf")
par(mfcol=c(3,2))
hist(Zscorefun(40,25,5,2,0,12),breaks=40,freq=FALSE,xlim=
c(-3,7),xlab=" ", ylab=" ", main="(a) (40,5,2,25,0,12) ")
hist(Zscorefun(100,100,1,3,1,3),breaks=40,freq=FALSE,xlim=
c(-4,4),xlab=" ", ylab=" ", main="(b) (100,1,3,100,1,3) ")
hist(Zscorefun(100,30,3,6,1,10),breaks=30,freq=FALSE,xlim=
c(-3,7),xlab=" ", ylab=" ", main="(c) (100,3,6,30,1,10) ")
hist(Zscorefun(100,25,13,4,0,30),breaks=40,freq=FALSE,xlim=
c(-2,10),xlab=" ", ylab=" ", main="(d) (100,13,4,25,0,30) ")
```

```
hist(Zscorefun(40,40,14,4,0,32),breaks=40,freq=FALSE,xlim=
c(-3,7),xlab=" ", ylab=" ", main="(e) (40,14,4,40,0,32) ")
hist(Zscorefun(25,100,0,10,0,10),breaks=30,freq=FALSE,xlim=
c(-5,3),xlab=" ", ylab=" ", main="(f) (25,0,10,100,0,10) ")
dev.off()

#####################################################################

#Gibbs sampling

install.packages("MCMCpack")
library(MCMCpack)

#Initial values
mu0=10
sigma02=50

alpha=2
beta=200

n=100
m=10000


calP<-function(inmu,insigma2){
  mu=numeric(m)
  sigma2=numeric(m)
  k=numeric(m)
```

```
x=rlnorm(n,meanlog=inmu,sdlog=sqrt(insigma2))

sigma2[1]=4

for(i in 1:m){

  b=(1/sigma02)+(n/sigma2[i])
  a= ((mu0/sigma02)+(sum(log(x))/sigma2[i]))/b

  mu[i]=rnorm(1,a,b)

  c=alpha+(n/2)
  d=((sum((log(x)-mu[i])^2))+2*beta)/2
  sigma2[i]= rinvgamma(1,c,d)
  sigma2[i+1]=sigma2[i]
}

sigma2=sigma2[1:m]
eta0=mu0+(sigma02 / 2)

eta=mu+(sigma2 / 2)
k = subset(eta,subset=eta<=eta0)
pValue = length(k)/length(eta)
pValue

crediblesetdiff1 = quantile(eta,c(0.025, 0.975),names=FALSE)
crediblesetdiff1

crediblesetdiff2 = quantile(eta,c(0.05, 0.95),names=FALSE)
```

```
  crediblesetdiff2


  return(c(pValue,crediblesetdiff1,crediblesetdiff2))
}

a<-seq(0,20,10)
b<-seq(10,100,10)

p<-matrix(1,nrow=length(a)*length(b),ncol=7)

for(i in 1:length(a)){
  for(j in 1:length(b)){
    n=(i-1)*length(b)+j
    p[n,3]=round(calP(a[i],b[j])[1],3)
    p[n,4]=round(calP(a[i],b[j])[2],3)
    p[n,5]=round(calP(a[i],b[j])[3],3)
    p[n,6]=round(calP(a[i],b[j])[4],3)
    p[n,7]=round(calP(a[i],b[j])[5],3)
    p[n,1]=a[i]
    p[n,2]=b[j]
  }
}

write.table(p,sep="\t",file="Output.txt",row.names=FALSE,
col.names=c("mu","sigma2","pvalue","Lower CI","Upper CI",
"Lower CI","Upper CI"))
```