University of the Free State

Department of Mathematical Statistics and Actuarial Science

Estimation of a Common Mean in Biological Research - Bayesian Approach

By

E. du Plessis

A.J. van der Merwe

Abstract

In this paper three methods of estimating a common mean, μ, for normal populations are discussed and compared. The first two methods are similar in concept and are based on pivotal quantinties. Krishnamoorthy and Lu (2003) based their procedure on weighted linear combinations of pivotal quantities, while Lin and Lee (2005) proposed a new generalized pivotal based on the best linear unbiased estimator of μ. The third method is a Bayesian technique which was proposed by Gregurich and Broemeling in 1997. With the Bayesian technique various prior distributions were considered.

A simulation study was done and the expected interval lengths and coverage probabilities of the three methods are compared. The three methods are illustrated by using two examples. The results show that the Bayesian technique is just as good for estimating the common mean as the method based on the best unbiased estimator of μ. The fact that the Bayesian procedure has good frequentist properties is a very desirable situation. Some statisticians argue that frequency calculations are an important part of applied Bayesian statistics (Rubin, 1984).

Keys Words: Common mean, Generalized Variables, Metadata

1. Introduction

In statistics, meta-analysis allows for the results of different studies, addressing related research hypotheses, to be combined (The ASAP, 2007). Identification of a common measure is the simplest form of meta-analysis. In these simple cases the results from the meta-analysis might simply be a weighted average of the common measure in the different studies. Meta-analysis is commonly used in social, behavioural and biological sciences as well as in clinical trials (Krishnamoorthy and Lu, 2003).

There is however a problem when using meta-analysis to combine the results of independent studies. This problem is due to the fact that different methods, instruments and laboratories are used in each study. The different populations may have a common measure, but other parameters are highly likely to differ from population to population. Due to the heterogeneity of parameters in the different studies, it may be difficult to combine the results.

2. Literature Study

The problem which will be addressed in this paper can be described in the following way: We are considering k different normal populations. These populations have a common mean μ and possibly different variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$. Each of the k samples consists of n_i observations, $i = 1,2,\dots,k$. Let $\bar{X}_i$ denote the mean of the sample drawn from the i^{th} population, defined as

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\ . \tag{2.1}$$

Also let S_i^2 denote the sample variance, as calculated from the i^{th} sample

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2 . \tag{2.2}$$

The information contained in these independent samples must now be combined in some way, with the objective to estimate the value of μ. There are a large number of papers already written on the point estimation of μ. Graybill and Deal (1959) were the first to show that for the case where $k = 2$ the combined estimator

$$\hat{\mu}_{GD} = \frac{\frac{n_1}{s_1^2}\bar{X}_1 + \frac{n_2}{s_2^2}\bar{X}_2}{\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}}, \tag{2.3}$$

has a smaller variance than either of the sample means, given that both n_1 and n_2 are greater or equal to 11. Since then, there were many improvements and extensions to $\hat{\mu}_{GD}$ given in (2.3). A lot of the work done since 1959 propose confidence intervals centred around $\hat{\mu}_{GD}$.

Krishnamoorthy and Lu (2003) based their procedure on weighted linear combinations of pivotal quantities. The procedure, as proposed by Krishnamoorthy and Lu (2003) can be described as the following:

If σ_1^2 and σ_2^2 were known, then $\hat{\mu}_{GD}$, given in (2.3), can be written as:

$$\hat{\mu}_{GD} = \frac{\frac{n_1}{\sigma_1^2}\bar{X}_1 + \frac{n_2}{\sigma_2^2}\bar{X}_2}{\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}}. \tag{2.4}$$

If μ was also known, then (2.4) will exactly be equal to μ,

$$\hat{\mu}_{GD} = \frac{\frac{n_1}{\sigma_1^2}\mu + \frac{n_2}{\sigma_2^2}\mu}{\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}} = \mu. \tag{2.5}$$

However, μ, σ_1^2 and σ_2^2 are unknown and therefore these values must be estimated. Instead of estimating, Krishnamoorthy and Lu (2003) simulated these values.

From classical statistics and by making use of the fact that we are working with a common mean, that is $\mu = \mu_1 = \mu_2$, it follows that $\bar{X}_i \sim N\left(\mu, \frac{\sigma_i^2}{n_i}\right), i = 1,2$. In other words, the sample mean can be written as:

$$\bar{X}_i = \mu + \frac{\sigma_i}{\sqrt{n_i}} Z_i, \text{ where } Z_i \sim N(0,1) \text{ for given } \sigma_i \ \ (i = 1, 2). \tag{2.6}$$

By changing the subject of (2.6) to μ, the population mean can be written as

$$\mu = \bar{X}_i - \frac{\sigma_i}{\sqrt{n_i}} Z_i, \text{ where } Z_i \sim N(0,1). \tag{2.7}$$

By substituting μ, as in (2.7), into (2.5), we obtain

$$\hat{\mu}_{GD} = \frac{\frac{n_1}{\sigma_1^2}(\bar{X}_1 - \frac{\sigma_1}{\sqrt{n_1}} Z_1) + \frac{n_2}{\sigma_2^2}(\bar{X}_2 - \frac{\sigma_2}{\sqrt{n_2}} Z_2)}{\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}} \text{ for given } \sigma_i \ \ (i = 1, 2). \tag{2.8}$$

In practice the variance components are not known, but it is well known from classical statistics that

$$\frac{(n_i-1)s_i^2}{\sigma_i^2} = \frac{V_i}{\sigma_i^2} \sim \chi^2_{n_i-1}. \tag{2.9}$$

By letting Q_i be equal to (2.9), and changing the subject of the formula to σ_i^2, it is found that the population variance can be written in the following way:

$$\sigma_i^2 = \frac{V_i}{Q_i} \Rightarrow \sigma_i = \sqrt{\frac{V_i}{Q_i}}. \tag{2.10}$$

Once the population variance, as given in (2.10), is substituted into (2.8), we have:

$$\hat{\mu}_{GD} = \frac{\frac{n_1}{\sigma_1^2}(\bar{X}_1 - \sqrt{\frac{V_1}{Q_1 n_1}}\, Z_1) + \frac{n_2}{\sigma_2^2}(\bar{X}_2 - \sqrt{\frac{V_2}{Q_2 n_2}}\, Z_2)}{\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}}. \tag{2.11}$$

Again by using (2.9) and letting it be equal to U_i, a generalized variable for estimating the unknown variances is obtained

$$\frac{(n_i-1)s_i^2}{\sigma_i^2} = \frac{V_i}{\sigma_i^2} = U_i \sim \chi^2_{n_i-1},$$

$$\frac{1}{\sigma_i^2} = \frac{U_i}{V_i}. \tag{2.12}$$

Here we note that two different χ^2-variables, Q_i and U_i were constructed for σ_i^2. Krishnamoorthy and Lu (2003) used two different variables for the variance component, although these variables are both related to the sample sums of squares $\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2$. Numerical studies showed that the confidence intervals which were formed by using the same chi-squared variable were too narrow. By using two chi-squared variables, the resulting confidence intervals are somewhat conservative or almost exact.

Finally after substituting the precision, as written in (2.12), in (2.11) it follows that:

$$\hat{\mu}_{GD} = \frac{\frac{n_1 U_1}{V_1}(\bar{X}_1 - \sqrt{\frac{V_1}{Q_1 n_1}}\, Z_1) + \frac{n_2 U_2}{V_2}(\bar{X}_2 - \sqrt{\frac{V_2}{Q_2 n_2}} Z_2)}{\frac{n_1 U_1}{V_1} + \frac{n_2 U_2}{V_2}}. \tag{2.13}$$

The estimate of the common mean given (2.13) is used when working with only two populations. If we have k populations, the expression becomes:

$$\hat{\mu}_{GD} = \frac{\sum_{i=1}^{k} \frac{n_i U_i}{V_i}\left(\bar{X}_i - Z_i\sqrt{\frac{V_i}{Q_i n_i}}\right)}{\sum_{j=1}^{k} \frac{n_j U_j}{V_j}},$$

$$= \frac{\sum_{i=1}^{k} \frac{n_i U_i \bar{X}_i}{V_i} - \sum_{i=1}^{k} U_i Z_i \sqrt{\frac{n_i}{Q_i V_i}}}{\sum_{j=1}^{k} \frac{n_j U_j}{V_j}} \tag{2.14}$$

By using equation (2.14) it is possible to simulate values for the common mean, μ. This is done in order to obtain point estimates and confidence intervals of μ. We can simplify equation (2.14), by noting that

$$\frac{Z_i}{\sqrt{\frac{\chi^2_{n_i-1}}{n_i - 1}}} \sim t_{n_i-1} = t_i. \tag{2.15}$$

By using (2.15), we can write equation (2.14) as:

$$\hat{\mu}_{GD} = \frac{\sum_{i=1}^{k} \frac{n_i U_i}{V_i}\left(\bar{X}_i - \frac{t_i s_i}{\sqrt{n_i}}\right)}{\sum_{j=1}^{k} \frac{n_j U_j}{V_j}}. \tag{2.16}$$

By defining W_i as

$$W_i = \frac{\frac{n_i U_{1i}}{V_i}}{\sum_{j=1}^{k} \frac{n_j U_{2j}}{V_j}} \tag{2.17}$$

and defining T_i as

$$T_i = \bar{X}_i - \frac{t_i s_i}{\sqrt{n_i}}, \tag{2.18}$$

we can simply write equation (2.16) as

$$\hat{\mu}_{GD} = \sum_{i=1}^{k} W_i T_i \,. \tag{2.19}$$

Krishnamoorthy and Lu (2003) computed the coverage probabilities and expected lengths of the 95% confidence intervals for the common mean μ, as well as the power for hypothesis testing numerically by using a Monte Carlo method. The expected length of the confidence interval was calculated as the average length of 1000 confidence intervals.

Krishnamoorthy and Lu (2003) compared their method, of estimating a confidence interval for a common mean of two normal populations, with two other methods. The first of these methods was done by Fairweather in 1972. The other method to which Krishnamoorthy and Lu compared their method, of estimating a common mean, is the method due to Jordan and Krishnamoorthy in 1996.

Both these methods, Fairweather's, as well as Jordan and Krishnamoorthy's, are exact methods. This means that a 95% confidence interval will give an exact coverage of 95%. Krishnamoorthy and Lu (2003) found that their method of estimating the common mean of two normal populations give shorter confidence intervals as well as coverage around 95% each time.

Furthermore, Krishnamoorthy and Lu compared the power of their method to that of Zhou and Mathew (1993), Fisher's test and Mathew et. al (1993). According to Krishnamoorthy and Lu, the hypothesis tests are scale invariant. Therefore the power of the tests only depend on the variances via the ratio of the two population variances, $\frac{\sigma_2^2}{\sigma_1^2}$ and $\delta = (\mu - \mu_0)$ where μ_0 is the value of the mean under the null hypothesis. When working with only two populations it is sufficient to only study the power properties for different values of the square root of this ratio, namely $\frac{\sigma_2}{\sigma_1}$.

Krishnamoorthy and Lu compared their method to four others which can be used for hypothesis testing. From the results they obtained, they could not show that any one of the five methods was uniformly most powerful. Although their method is not uniformly most powerful, their method's powers are close to the highest power obtain by the comparison tests. There are also cases for which this method had much higher power than the other four tests.

Due to these results mentioned above, we can use the generalized variable method as a benchmark for the next methods. This method constructed shorter confidence intervals, while still showing good coverage. When using this method for hypothesis testing we are sure to have power close to the highest, if not the highest.

The next method for estimating the common mean of normal populations which will be discussed also refers to the work done by Fairweather (1972), Jordan and Krishnamoorthy

(1996) and Krishnamoorthy and Lu (2003). This method, which is similar in principal to that of Krishnamoorthy and Lu (2003), was derived by Lin and Lee in 2005.

This procedure is also based on pivotal quantities. These quantities are based on the best unbiased estimator of μ. Krishnamoorthy and Lu (2003) as well as Lin and Lee (2005) based their work on concepts of generalized p-values and generalized confidence intervals. The major difference between these two works is the difference in pivotal quantities they used.

In the simulation of $\sigma_i^2 = \frac{V_i}{Q_i}$ in (2.10) and $\sigma_i^2 = \frac{V_i}{U_i}$ in (2.12) Krishnamoorthy and Lu (2003) used two different chi-squared random variables Q_i and U_i, even though both are related to the same $V_i = (n_i - 1)S_i^2$. Lin and Lee (2005) on the other hand used only $U_i \sim \chi^2_{n_i-1}$ in the simulation procedure.

Again referring to the problem as explained at the start of this section, Lin and Lee (2005) used the following:

To start, again we know $\bar{X}_i \sim N\left(\mu, \frac{\sigma_i^2}{n_i}\right)$ as well as S_i^2 is defined as in (2.9). These two sufficient statistics, $\bar{X}_i$ and S_i^2, are mutually independent. When the variances, σ_i^2, are known, the best linear unbiased estimator for μ is

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \bar{X}_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}. \tag{2.20}$$

As in (2.5) if μ was known, this estimate would exactly be equal to the common mean,

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \mu}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}} = \mu\,. \tag{2.21}$$

As before, we substitute μ in (2.21) with (2.7),

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}\left(\bar{X}_i - \frac{\sigma_i}{\sqrt{n_i}} Z_i\right)}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}} \text{ for given } \sigma_i. \tag{2.22}$$

If the variance components are known, we can write the estimate of the common mean as

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \overline{X}_i}{\sigma_i^2} - \sum_{i=1}^{k} \sqrt{\frac{n_i}{\sigma_i^2}} Z_i}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}, \tag{2.23}$$

where Z_i is normally distributed with a zero mean and a variance of one, for $i = 1,2, \ldots, k$.

Again if the variance components are known, it follows that

$$\sum_{i=1}^{k} \left(\sqrt{\frac{n_i}{\sigma_i^2}} Z_i \right) \sim N\left(0, \sum_{i=1}^{k} \frac{n_i}{\sigma_i^2} \right), \tag{2.24}$$

from which we have

$$\sum_{i=1}^{k} \left(\sqrt{\frac{n_i}{\sigma_i^2}} Z_i \right) = \sqrt{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}\ Z, \tag{2.25}$$

with $Z \sim N(0,1)$.

By substituting (2.25) in (2.23), we find

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \overline{X}_i}{\sigma_i^2} - \sqrt{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}\ Z}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}. \tag{2.26}$$

Finally we need to account for the fact that variance components are unknown. By letting U_i be equal to (2.9), and changing the subject of the formula to σ_i^2, it is found that

$$\sigma_i^2 = \frac{V_i^2}{U_i}. \tag{2.27}$$

After substituting (2.27) in (2.28) we obtain the generalized pivotal quantity as used Lin and Lee (2005),

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \overline{X}_i U_i}{V_i} - \sqrt{\sum_{i=1}^{k} \frac{n_i U_i}{V_i}}\ Z}{\sum_{j=1}^{k} \frac{n_j U_j}{V_j}}. \tag{2.28}$$

Lin and Lee (2005) compared five different methods of estimating the common mean of two normal populations. The comparison is done on the basis of the expected length of a 95% confidence interval for the common mean, μ, as well as the power performances. The methods that are compared are the following:

- Classical method.
- Fairweather (1972).

- Jordan and Krishnamoorthy (1996).
- Krishnamoorthy and Lu Young (2003).
- Lin and Lee (2005).

When looking at expected lengths and the coverage of the intervals, Lin and Lee found the following. The classical method does not allow for heteroscedasticity, therefore it delivered the worst results among all the methods. All four of the remaining methods does allow for heteroscedasticity. As mentioned earlier, it is already known that Krishnamoorthy and Lu's method gives shorter intervals than Fairweather as well as Jordan and Krishnamoorthy.

The method proposed by Lin and Lee however outperformed the other methods in the sense that this method delivered the shortest confidence intervals. These intervals are even shorter than the confidence intervals obtained by Krishnamoorthy and Lu (2003). Lin and Lee (2005) also showed that for a 95% confidence interval, this method based on the best linear estimator of μ, gives coverage around 95%.

When comparing the power performances of the various methods, Lin and Lee showed that their procedure produced the highest power uniformly. None of the other four tests used for the comparison of the powers performed better than the generalized variable method as proposed by Lin and Lee. From these results it is clear that generalized variable method that is based on the best unbiased estimator of the common mean performs better than the generalized variable which is based on weighted linear combinations of pivotal quantities.

Chang and Pal (2008) concentrated on testing the common mean of several normal distributions. They proposed three new methods for testing the common mean of normal populations, two which are based on the Graybill-Deal estimator and the third method uses maximum likelihood estimates. Chang and Pal specifically compared their methods to that of Krishnamoorty and Lu (2003) as well as Lin and Lee (2005). These three methods delivered results which are about the same as the results obtained by Lin and Lee's procedure. Lin and Lee's method however is more powerful than any of these methods.

A Bayesian technique might also be adopted to estimate the common mean of several normal populations (Gregurich and Broemeling, 1997). A meta-analysis study's results rely heavily on the prior information available. Bayesian methods allow for the inclusion of prior information and therefore it was suggested that Bayesian methods should be used in meta-analysis (Gregurich and Broemeling, 1997).

A complete Bayesian analysis is done to determine the posterior distribution of the common mean. This can be done by making use of the Gibbs sampler. The Gibbs sampler provides a way for sampling from the joint distributions with more than one variable. According to StatSoft (2011), the Gibbs sampler is based on the notion: "[T]o sample from a joint distribution just sample repeatedly from its one-dimensional conditionals given whatever you've seen at the time."

Gregurich and Broemeling (1997) begin with the same setting as Krishnamoorthy and Lu. They start with two normal populations with a common mean, $\mu = \mu_1 = \mu_2$, and variances which are not necessarily equal, σ_1^2 and σ_2^2. A random sample, of size n_i, from each of these populations is obtained. Let X_{ij} denote the sample, where $i = 1, 2$ and $j = 1,2, \dots, n_i$, therefore; $X_{ij} \sim N(\mu, \sigma_i^2)$.

The likelihood function can therefore be written as:

$$\begin{aligned} L(\mu, \sigma_1^2, \sigma_2^2 | data) \propto & \left(\frac{1}{\sigma_1^2}\right)^{\frac{1}{2}n_1} \exp\left\{-\frac{1}{2\sigma_1^2}[(n_1-1)s_1^2 + n_1(\mu - \overline{x}_1)^2]\right\} \\ & \times \left(\frac{1}{\sigma_2^2}\right)^{\frac{1}{2}n_2} \exp\left\{-\frac{1}{2\sigma_2^2}[(n_2-1)s_2^2 + n_2(\mu - \overline{x}_2)^2]\right\} \end{aligned} \tag{2.29}$$

where $\overline{x}_i$ is the sample mean and s_i^2 is the sample variance, defined as

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(x_{ij} - \overline{x}_i\right)^2 \text{ where } i = 1, 2. \tag{2.30}$$

Gregurich and Broemeling (1997) used the Jeffreys Independence Prior, which can be derived in the following way:

We start by taking the natural logarithm of the likelihood, given in (2.29),

$$\begin{aligned} \ell = \ln L = & -\frac{n_1}{2}\ln\sigma_1^2 - \frac{1}{2\sigma_1^2}[(n_1-1)s_1^2 + n_1(\mu - \overline{x}_1)^2] \\ & -\frac{n_2}{2}\ln\sigma_2^2 - \frac{1}{2\sigma_2^2}[(n_2-1)s_2^2 + n_2(\mu - \overline{x}_2)^2] \end{aligned} \tag{2.31}$$

From this we need to derive the Fischer Information matrix. Therefore we need to find the second derivative of ℓ with respect to each possible pair of parameters.

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n_1}{\sigma_1^2} - \frac{n_2}{\sigma_2^2} \tag{2.32}$$

$$\frac{\partial^2 \ell}{\partial \mu\, \partial \sigma_1^2} = \frac{n_1(\mu - \overline{x}_1)}{(\sigma_1^2)^2} \qquad \frac{\partial^2 \ell}{\partial \mu\, \partial \sigma_2^2} = \frac{n_2(\mu - \overline{x}_2)}{(\sigma_2^2)^2} \tag{2.33}$$

$$\frac{\partial^2 \ell}{\left(\partial \sigma_1^2\right)^2} = \frac{n_1}{2}\left(\frac{1}{\sigma_1^2}\right)^2 - \left(\frac{1}{\sigma_1^2}\right)^3 \left[(n_1 - 1)s_1^2 + n_1(\mu - \overline{x}_1)^2\right] \tag{2.34}$$

$$\frac{\partial^2 \ell}{\left(\partial \sigma_2^2\right)^2} = \frac{n_2}{2}\left(\frac{1}{\sigma_2^2}\right)^2 - \left(\frac{1}{\sigma_2^2}\right)^3 \left[(n_2 - 1)s_2^2 + n_2(\mu - \overline{x}_2)^2\right] \tag{2.35}$$

$$\frac{\partial^2 \ell}{\partial \sigma_1^2\, \partial \sigma_2^2} = 0 \qquad \frac{\partial^2 \ell}{\partial \sigma_2^2\, \partial \sigma_1^2} = 0 \tag{2.36}$$

Now to find the Fischer Information matrix, we need to find the expected values or these second derivatives given above. The negative of the expected values are the entries of the Fischer information matrix.

$$F(\mu, \sigma_1^2, \sigma_2^2) = \begin{bmatrix} \frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2} & 0 & 0 \\ 0 & \frac{n_1}{2}\left(\frac{1}{\sigma_1^2}\right)^2 & 0 \\ 0 & 0 & \frac{n_2}{2}\left(\frac{1}{\sigma_2^2}\right)^2 \end{bmatrix} \tag{2.37}$$

From the Fischer Information matrix we can calculate the Jeffreys Independence Prior as:

$$p(\mu, \sigma_1^2, \sigma_2^2) = p(\mu)p(\sigma_1^2, \sigma_2^2) \tag{2.38}$$

We choose a flat prior for the common mean that is

$$p(\mu) \propto constant. \tag{2.39}$$

Now we can calculate the joint prior for the variances as the following by using the Fischer information matrix,

$$p(\,\sigma_1^2,\sigma_2^2) \propto \begin{vmatrix} \frac{n_1}{2}\left(\frac{1}{\sigma_1^2}\right)^2 & 0 \\ 0 & \frac{n_2}{2}\left(\frac{1}{\sigma_2^2}\right)^2 \end{vmatrix}^{\frac{1}{2}} \tag{2.40}$$

Finally we find the Jeffreys Independence prior to be equal to

$$p(\mu,\sigma_1^2,\sigma_2^2) \propto \sigma_1^{-2}\sigma_2^{-2}. \tag{2.41}$$

This prior is usually used when there is a limited amount of prior information available (Gregurich and Broemeling, 1997).

It can also be shown that (2.41) is a reference prior for any ordering of the parameters μ, σ_1^2 and σ_2^2. According to Berger and Bernardo (1992) the reference prior theory has been the most successful technique for deriving objective priors. The derivation of reference priors depend on the parameters of interest, the grouping of the parameters and the ordering of the groupings.

The joint posterior distribution for μ, σ_1^2 and σ_2^2 is obtained by multiplying the prior, given in (2.41), with the likelihood, given in (2.29). The joint posterior distribution for μ, σ_1^2 and σ_2^2 is then given by

$$p(\mu,\sigma_1^2,\sigma_2^2|data) \propto \prod_{i=1}^{2}\left(\frac{1}{\sigma_i^2}\right)^{\frac{1}{2}(n_i+2)} \exp\left\{-\frac{1}{2\sigma_i^2}[(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2]\right\}. \tag{2.42}$$

From the joint posterior distribution it is possible to derive the conditional posterior distributions, of σ_1^2 and σ_2^2 given μ, as:

$$p(\sigma_i^2|\mu, data) = C_i\left(\frac{1}{\sigma_i^2}\right)^{\frac{1}{2}(n_i+2)} \exp\left\{-\frac{1}{2\sigma_i^2}[(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2]\right\}, \tag{2.43}$$

where

$$C_i = \left(\frac{(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2}{2}\right)^{\frac{1}{2}n_i}\frac{1}{\Gamma\left(\frac{n_i}{2}\right)}. \tag{2.44}$$

From equation (2.43) the distribution of $(\sigma_i^2|\mu, data)$ can be recognized as an Inverse Gamma distribution. From this it follows that

$$\frac{(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2}{\sigma_i^2} \sim \chi_{n_i}^2\,, \tag{2.45}$$

which can be rewritten with σ_i^2 the subject of the formula as

$$\sigma_i^2 = \frac{(n_i - 1)s_i^2 + n_i(\mu - \overline{x}_i)^2}{\chi^2_{n_i}}. \tag{2.46}$$

Also the conditional posterior densities of σ_1^2 and σ_2^2 are independent.

When working with only two populations, the conditional posterior distribution of μ given the variance components σ_1^2 and σ_2^2 is a Normal distribution,

$$\mu|\sigma_1^2, \sigma_2^2, data \sim N\left\{\frac{\frac{n_1}{\sigma_1^2}\bar{x}_1 + \frac{n_2}{\sigma_2^2}\bar{x}_2}{\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}}, \left(\frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}\right)^{-1}\right\}. \tag{2.47}$$

If we have k populations, the conditional posterior distribution of μ given the variance components σ_i^2 is still a Normal distribution:

$$\mu|\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2, data \sim N\left\{\frac{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}\bar{x}_i}{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}}, \left(\sum_{i=1}^k \frac{n_i}{\sigma_i^2}\right)^{-1}\right\}. \tag{2.48}$$

Now the Gibbs sampler can be used to simulate from the joint posterior distribution given in (2.42), since it is easy to simulate from the conditional distributions defined in (2.45) and (2.47) (Gregurich and Broemeling, 1997). The Gibbs sampler basically comes down to simulating a large number of values for each of the parameters, μ, σ_1^2 and σ_2^2, given the values of the other parameters as simulated at each step. These values can be easily simulated from (2.46) and (2.47).

From these simulated values of μ, we can estimate confidence intervals for the common mean. Again the average length of the confidence intervals can be compared to that observed with the other methods.

Equation (2.48) can also be written as

$$\begin{aligned}\mu &= \frac{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}\bar{x}_i}{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}} - Z\left(\sum_{i=1}^k \frac{n_i}{\sigma_i^2}\right)^{-\frac{1}{2}} \\ &= \frac{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}\bar{x}_i - Z\left(\sum_{i=1}^k \frac{n_i}{\sigma_i^2}\right)^{\frac{1}{2}}}{\sum_{i=1}^k \frac{n_i}{\sigma_i^2}}\end{aligned} \tag{2.49}$$

where $Z \sim N(0,1)$.

If the substitution $\sigma_i^2 = \frac{V_i}{U_i}$ is made where $U_i \sim \chi^2_{n_i-1}$, then equation (2.49) becomes

$$\mu = \frac{\sum_{i=1}^{k} \frac{n_i U_i}{V_i} \bar{x}_i - Z \left(\sum_{i=1}^{k} \frac{n_i U_i}{V_i}\right)^{\frac{1}{2}}}{\sum_{i=1}^{k} \frac{n_i U_i}{V_i}} \tag{2.50}$$

which is exactly the same as equation (3.4) in Lin and Lee (2005).

The two procedures (Bayesian method and Lin and Lee's procedure) are therefore very much the same. The differecnce is that in the case of Lin and Lee the variance components are simulated as

$$\sigma_i^2 = \frac{V_i}{\chi^2_{n_i-1}} \quad (i = 1,2, \dots, k) \tag{2.51}$$

while for the Bayesian approach

$$\sigma_i^2 = \frac{V_i + n(\mu - \overline{x}_i)^2}{\chi^2_{n_i}} \quad (i = 1,2, \dots, k) \tag{2.52}$$

is used in the Gibbs sampler.

3. Methodology

The first two methods which were discussed are very similar in concept, while the last method is totally a different approach to this problem. In order to compare the three methods we will consider two samples of sizes n_1 and n_2. These samples are drawn from normal populations with the common mean, $\mu = 0$ and each sample has a known variance, σ_1^2 and σ_2^2, respectively.

The first method, done by Krishnamoorthy and Lu (2003), can easily be used to calculate confidence intervals. Once the sample is known, we calculate the sample mean and sample variance. After this we simulate from a χ^2-distribution and a t-distribution as specified in (2.17) and (2.18). Finally, to calculate an estimate for the common mean we multiply these two simulated values as in (2.19).

To generalize these estimates to confidence intervals we repeat this process 5000 times in order to obtain 5000 estimates of the common mean. These estimates are then sorted and a 95% confidence interval is obtained from these estimates. The lower limit of the interval is the 125th value (5000×0.025), while the upper limit is the 4875th value (5000×0.975), of these sorted estimates. To estimate the interval length we calculate the difference between the upper and the lower limits. We repeated this process 1000 times and calculated the expected interval length as the mean of the 1000 intervals lengths.

It is not only important that we compare the various interval lengths, but also the accuracy of the intervals. This can be measured by the coverage of the calculated intervals. Once we calculated the 1000 confidence intervals, we count how many of these intervals actually contain the common mean, μ. Since we calculate 95% confidence intervals, we would expect that about 950 of these intervals actually do contain the value of μ.

Another way to compare coverage is to compare the type I errors of the different methods. The type I error shows the probability that the confidence interval does not contain the true value of the common mean. As we calculated 95% confidence intervals, we expect that only 50 of the 1000 confidence intervals would not contain the value of μ. The observed type I errors of the different methods, as well as the expected interval lengths, are compared in Tables 4.1 to 4.4, which are given in the next section.

When using this method for hypothesis testing, we can compare the power of this method to that of other known methods. The power of a hypothesis test is defined as the probability of correctly rejecting the null hypothesis. We can estimate the power for a specific method through simulation. The hypothesis that we are interested in testing is the following:

$$H_0: \mu = 0 \quad vs. H_1: \mu \neq 0 \tag{3.1}$$

To test this hypothesis we consider two samples of size n_1 and n_2 respectively. We use the method as explained in equations (2.17) to (2.19) to estimate the common mean. This process is repeated 5000 times. The next step is to calculate the generalized p-value for testing (3.1), as

$$\mathrm{p-value} = 2 \times \min\left\{P\left(T < \mu_0 \middle| \bar{\mathrm{x}}, \mathrm{v}\right), P\left(T > \mu_0 \middle| \bar{\mathrm{x}}, \mathrm{v}\right)\right\} \tag{3.2}$$

where $P(T < \mu_0|\overline{\mathrm{x}}, \mathrm{v})$ can be calculated as the proportion of estimates that are smaller than μ_0, which in this case is zero. In the same way, $P(T > \mu_0|\overline{\mathrm{x}}, \mathrm{v})$, can be estimated as the proportion of estimates that are larger than μ_0.

Once we have the p-value, we can compare that to the value of α. If the p-value is smaller than α, we would reject the null hypothesis. In order to calculate the power, the procedure as just explained is repeated 2500 times. This is effectively the same as doing a hypothesis test 2500 times. From this we count the number of times that the null hypothesis would be rejected and this is divided by 2500 to convert it to a probability of rejecting the null hypothesis. This probability is what we call the power of a hypothesis test. The estimated powers for testing (3.1) are compared in Tables 4.6 and 4.7 in the next section.

The second method, done by Lin and Lee (2005), is similar in concept to that of Krishnamoorthy and Lu. Again the sample is known and we calculate the sample mean and sample variance. After this we simulate from a χ^2-distribution as specified in (2.9) and (2.27). Then we simulate from a standard normal distribution in (2.25). Finally, to calculate an estimate for the common mean we use (2.28) and substitute these two values which were just calculated.

This process is repeated 5000 times to obtain estimates of the common mean. The same steps, as described for the first method, are used to calculate a 95% confidence interval from these estimates. Again we calculate the average interval length of 1000 intervals as well as the type I error, as explained previously. These values are also given in Tables 4.1 to 4.4 in the next section. The power for hypothesis testing can be calculated in exactly the same way was done for the first method. The power performance of this method is compared to that of the two other methods in Tables 4.5 and 4.6 in the next section. As mentioned before the major difference between these two simulation procedures is the difference in the pivotal quantities they used.

The last method, done by Gregurich and Broemeling (1997), is very different in concept from the other two methods. The Bayesian technique is used to find the joint posterior distribution of the three parameters, μ, σ_1^2 and σ_2^2, given in (2.42). Then we use the Gibbs sampler to simulate from this distribution. That comes down to simulating from a Normal distribution and two independent χ^2-distributions, as specified in (2.47) and (2.46), respectively.

To start the simulation we need an estimate of the common mean. The estimate we use to start the simulation is the average of the two sample means. This value is necessary to simulate the first estimates of the variances of the two populations. These variances are simulated by using (2.46), the conditional posterior distribution of the variances. After the first variances are simulated we can use that to simulate a value of μ from the conditional posterior distribution of the mean.

In the next simulation of the variances we use the value of the mean that was simulated in the previous step. This process of simulating the mean and the variances is repeated 5100 times. The first 100 simulated values are discarded and the remaining 5000 is used to determine the confidence interval. The confidence interval is determined in the same way as done for the previous two methods. The estimates of the common mean are sorted and a 95% credibility interval (Bayesian confidence interval) is determined. The lower limit is again the 125th value, while the upper limit is also the 4875th value of the sorted estimates.

This method of finding a confidence interval is used to estimate an equal tail confidence interval. Another Bayesian interval that can be used is the highest posterior density confidence interval. As the name suggests, the interval will contain the part of the distribution with the highest density, therefore this method of interval construction would lead to the shortest confidence interval which can be constructed.

To construct these intervals we start with the ordered estimates of the common mean. Next we compute 95% credible intervals with a different lower limit each time. In this case we chose lower limits from the 75th to the 175th values. The upper limit changes according to the lower limit. The upper limit is always 4750 (5000×0.95) values further than the lower limit. This ensures that we calculate a 95% confidence interval. This means that we calculate 100 different confidence intervals. To determine the highest posterior density interval, we find the shortest interval from these 100 intervals. The type I error is determined in the same manner as for the first two methods.

Gregurich and Broemeling (1997) used the Jeffreys Independence Prior, given in (2.41). We will also consider two other priors, namely

$$p(\mu,\sigma_1^2,\sigma_2^2) \propto \sigma_1^{-1}\sigma_2^{-1} \tag{3.3}$$

and

$$p(\mu,\sigma_1^2,\sigma_2^2) \propto \sigma_1^{-3}\sigma_2^{-3}. \tag{3.4}$$

The joint posterior distribution, as given in (2.42), will change according to the specific prior distribution used. When we use

$$p(\mu,\sigma_1^2,\sigma_2^2) \propto \sigma_1^{-i}\sigma_2^{-i} \tag{3.5}$$

as a prior distribution the joint posterior distribution becomes

$$p(\mu,\sigma_1^2,\sigma_2^2|data) \propto \prod_{i=1}^{2}\left(\frac{1}{\sigma_i^2}\right)^{\frac{1}{2}(n_i+i)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2\right]\right\}. \tag{3.6}$$

The conditional posterior distribution of $(\mu|\sigma_1^2,\sigma_2^2)$ will be unaffected by the specific choice of the prior distribution and this will remain as given in (2.47). The choice of prior distribution will however change the conditional posterior distribution of σ_1^2 and σ_2^2 given μ. When we use the prior distribution as specified in (3.7), the conditional posterior distribution of σ_i^2 would still be an Inverse Gamma distribution:

$$p(\sigma_i^2|\mu,data) = C_i\left(\frac{1}{\sigma_i^2}\right)^{\frac{1}{2}(n_i+i)} \exp\left\{-\frac{1}{2\sigma_i^2}\left[(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2\right]\right\}, \tag{3.7}$$

where

$$C_i = \left(\frac{(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2}{2}\right)^{\frac{1}{2}(n_i+i-2)} \frac{1}{\Gamma\left(\frac{n_i+i-2}{2}\right)}. \tag{3.8}$$

From this it again follows that

$$\frac{(n_i-1)s_i^2+n_i(\mu-\overline{x}_i)^2}{\sigma_i^2} \sim \chi^2_{n_i+i-2}, \tag{3.9}$$

This means that instead of simulating two independent $\chi^2_{n_i}$ values, as in (2.46), we simulated two independent $\chi^2_{n_i+i-2}$ values and proceed with the other calculations as usually.

Once again we need to calculate the power of the hypothesis test given in (3.1). The power is again calculated in the exact same way as was done for the first method. For the Bayesian method we only calculated the power when using the Jeffreys Independence Prior and not the other two priors specified in (3.5) and (3.6). The power performance is also compared to that of the other two methods in Tables 4.5 and 4.6.

The results given in the tables in the next section are only based on two samples, but as explained previously, it is easy to extend all three of these methods to incorporate more than two samples. This is illustrated in the two examples which follow the next sections.

4. Results

The tables given below compare the three methods of estimating a common mean for two normal populations. Each of the tables has different sample sizes, indicated above each table. In Table 4.1 to 4.4 each column gives the average length of the 95% confidence intervals as constructed by the different methods. The type I error is also indicated in brackets each time. Each confidence interval was constructed by using 5000 estimates for the common mean. These averages are based on 1000 confidence intervals.

In Tables 4.1 to 4.4 the different pairs of variances which were used to draw the samples are indicated in the first column of each table. The column labelled K and Lu (2003) gives the results which were obtained by using the method as proposed by Krishnamoorthy and Lu. The next column labelled Lin and Lee (2005) was completed by the use of the second method we described, due to Lin and Lee. The last three columns were obtained by the Bayesian technique. We used three different priors in each case and each column represents a prior. The three priors are defined below:

Prior I: $p_1(\mu, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-2}\sigma_2^{-2}$
Prior II: $p_2(\mu, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-1}\sigma_2^{-1}$
Prior III: $p_3(\mu, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-3}\sigma_2^{-3}$

For each of the three priors we calculated two types of confidence intervals. The equal tailed and the highest posterior density (HPD) confidence intervals.

The shortest average interval length is indicated in each row of the first four tables. This value is shaded on the condition that the coverage is near 95%. This condition is also considered since the confidence intervals obtained by using Prior III are the shortest confidence intervals overall, but the coverage is much worse than that of the other methods. The coverage is about 93% if Prior III is used.

Tables 4.5 and 4.6 compare the powers of the three different methods when testing the hypothesis given in (3.1). The rows labelled K and L corresponds to the results obtained by using the method as proposed by Krishnamoorthy and Lu in 2003. The first column indicated the various ratios of variances that were used to calculate the power of the hypothesis tests. The other columns indicate the different values used for the populations’ common mean.

Table 4.2: Average length of 95% confidence intervals with $n_1 = 10$ and $n_2 = 10$.

$n_1 = 10$, $n_2 = 10$	K and Lu (2003)	Lin and Lee (2005)	Prior I		Prior II		Prior III	
(σ_1^2, σ_2^2)			Equal tail	HPD	Equal tail	HPD	Equal tail	HPD
(5,5)	2.266 (4.1)	2.038 (5.4)	2.059 (4.9)	2.047 (4.9)	2.201 (4.7)	2.174 (5.2)	1.935 (6.7)	1.909 (6.8)
(5,10)	2.626 (4.1)	2.385 (5.1)	2.372 (5.0)	2.358 (5.1)	2.529 (5.2)	2.494 (5.2)	2.270 (6.9)	2.210 (7.9)
(5,15)	2.782 (4.9)	2.539 (5.4)	2.565 (5.3)	2.549 (5.6)	2.748 (4.7)	2.709 (5.3)	2.425 (7.2)	2.391 (7.2)
(5,20)	2.880 (4.9)	2.657 (5.3)	2.654 (5.3)	2.636 (5.4)	2.843 (4.7)	2.802 (5.3)	2.513 (8.3)	2.476 (9.2)
(5,30)	2.988 (4.5)	2.790 (5.1)	2.799 (5.1)	2.781 (5.4)	2.987 (4.5)	2.941 (4.8)	2.614 (6.7)	2.576 (7.1)
(5,40)	3.024 (4.3)	2.828 (5.2)	2.872 (5.4)	2.852 (5.3)	3.032 (5.4)	2.984 (5.5)	2.720 (6.4)	2.681 (7.5)
(5,50)	3.046 (4.8)	2.888 (5.3)	2.915 (4.7)	2.893 (4.9)	3.147 (4.5)	3.099 (4.6)	2.759 (6.3)	2.717 (6.8)
(5,100)	3.100 (5.3)	3.033 (5.6)	2.995 (4.6)	2.973 (4.7)	3.300 (4.7)	3.249 (5.2)	2.821 (6.1)	2.778 (6.9)
(5,250)	3.116 (4.5)	3.110 (4.9)	3.071 (4.9)	3.051 (5.4)	3.298 (4.5)	3.276 (4.0)	2.876 (6.5)	2.857 (7.3)
(5,500)	3.119 (5.2)	3.153 (5.4)	3.104 (5.3)	3.085 (5.4)	3.315 (4.5)	3.304 (4.8)	2.881 (6.9)	2.863 (7.1)
(5,1000)	3.140 (4.5)	3.164 (4.9)	3.119 (5.4)	3.101 (5.4)	3.342 (5.2)	3.322 (5.2)	2.914 (6.0)	2.897 (6.1)

Table 4.3: Average length of 95% confidence intervals with $n_1 = 15$ and $n_2 = 15$.

$n_1 = 15$, $n_2 = 15$	K and Lu (2003)	Lin and Lee (2005)	Prior I		Prior II		Prior III	
(σ_1^2, σ_2^2)			Equal tail	HPD	Equal tail	HPD	Equal tail	HPD
(5,5)	1.748 (4.4)	1.641 (5.1)	1.665 (5.4)	1.656 (6.2)	1.725 (5.3)	1.703 (5.4)	1.598 (7.7)	1.579 (8.2)
(5,10)	2.027 (5.1)	1.911 (5.4)	1.912 (5.0)	1.902 (5.4)	1.999 (4.8)	1.973 (5.0)	1.845 (6.9)	1.822 (7.4)
(5,15)	2.161 (4.5)	2.031 (5.1)	2.046 (5.3)	2.034 (5.5)	2.130 (4.5)	2.102 (4.7)	1.961 (7.1)	1.936 (7.0)
(5,20)	2.214 (4.5)	2.106 (5.1)	2.126 (5.4)	2.114 (4.8)	2.205 (4.4)	2.174 (4.6)	2.035 (7.1)	2.008 (7.2)
(5,30)	2.293 (4.3)	2.199 (4.8)	2.192 (5.4)	2.179 (5.3)	2.267 (4.5)	2.237 (4.8)	2.136 (5.8)	2.107 (5.9)
(5,40)	2.316 (5.3)	2.244 (5.5)	2.274 (4.8)	2.260 (5.1)	2.356 (4.3)	2.323 (4.9)	2.190 (7.5)	2.162 (7.2)
(5,50)	2.361 (4.3)	2.285 (4.7)	2.292 (5.1)	2.279 (5.2)	2.422 (4.5)	2.388 (4.6)	2.202 (7.6)	2.170 (8.1)
(5,100)	2.419 (4.5)	2.373 (5.1)	2.363 (5.0)	2.349 (5.0)	2.430 (4.6)	2.396 (4.7)	2.279 (6.2)	2.245 (7.0)
(5,250)	2.424 (5.4)	2.412 (5.2)	2.399 (4.9)	2.385 (5.1)	2.513 (4.5)	2.499 (4.6)	2.304 (6.0)	2.292 (6.0)
(5,500)	2.432 (4.6)	2.428 (4.5)	2.421 (5.3)	2.408 (5.4)	2.533 (4.4)	2.519 (4.5)	2.330 (5.9)	2.317 (6.1)
(5,1000)	2.434 (4.5)	2.431 (4.6)	2.435 (5.4)	2.422 (5.4)	2.545 (4.9)	2.531 (5.1)	2.332 (5.7)	2.320 (6.2)

Table 4.4: Average length of 95% confidence intervals with $n_1 = 10$ *and* $n_2 = 30$ *.*

$n_1 = 10$, $n_2 = 30$	K and Lu (2003)	Lin and Lee (2005)	Prior I		Prior II		Prior III	
(σ_1^2, σ_2^2)			Equal tail	HPD	Equal tail	HPD	Equal tail	HPD
(5,5)	1.498 (5.4)	1.418 (5.5)	1.420 (5.0)	1.413 (5.0)	1.460 (5.4)	1.442 (5.5)	1.382 (6.4)	1.366 (7.2)
(5,10)	1.916 (4.7)	1.788 (4.8)	1.806 (5.3)	1.796 (5.5)	1.856 (5.5)	1.832 (5.6)	1.740 (7.1)	1.717 (7.7)
(5,15)	2.159 (4.6)	2.011 (5.1)	2.028 (5.3)	2.017 (5.6)	2.095 (5.4)	2.068 (5.5)	1.946 (6.9)	1.921 (7.4)
(5,20)	2.317 (4.9)	2.165 (5.6)	2.185 (5.4)	2.173 (5.7)	2.287 (5.3)	2.257 (5.7)	2.074 (6.2)	2.047 (6.7)
(5,30)	2.532 (4.3)	2.357 (5.0)	2.366 (5.4)	2.352 (5.5)	2.491 (5.0)	2.455 (5.0)	2.264 (7.2)	2.233 (8.0)
(5,40)	2.639 (4.7)	2.468 (5.5)	2.523 (5.2)	2.506 (5.3)	2.623 (4.6)	2.586 (5.2)	2.390 (5.9)	2.357 (6.7)
(5,50)	2.732 (4.7)	2.582 (5.4)	2.617 (5.0)	2.598 (5.1)	2.736 (5.3)	2.697 (5.3)	2.447 (8.6)	2.412 (9.2)
(5,100)	2.954 (4.5)	2.808 (5.6)	2.802 (5.4)	2.708 (5.5)	2.986 (5.0)	2.939 (5.4)	2.693 (6.5)	2.653 (7.2)
(5,250)	3.050 (4.6)	3.017 (4.9)	2.976 (4.9)	2.954 (5.2)	3.247 (4.3)	3.221 (4.4)	2.795 (7.4)	2.777 (7.4)
(5,500)	3.130 (4.7)	3.106 (5.2)	3.051 (5.2)	3.030 (5.5)	3.319 (4.3)	3.295 (4.8)	2.863 (6.8)	2.845 (6.8)
(5,1000)	3.155 (4.8)	3.114 (4.5)	3.069 (4.6)	3.049 (4.9)	3.321 (4.6)	3.298 (4.7)	2.905 (5.6)	2.887 (6.3)

Table 4.5: Average length of 95% confidence intervals with $n_1 = 30$ *and* $n_2 = 10$ *.*

$n_1 = 30$, $n_2 = 10$	K and Lu (2003)	Lin and Lee (2005)	Prior I		Prior II		Prior III	
(σ_1^2, σ_2^2)			Equal tail	HPD	Equal tail	HPD	Equal tail	HPD
(5,5)	1.499 (5.6)	1.417 (5.6)	1.425 (4.5)	1.418 (4.9)	1.467 (4.3)	1.448 (4.5)	1.383 (6.0)	1.365 (6.1)
(5,10)	1.581 (4.9)	1.515 (5.5)	1.523 (4.6)	1.515 (4.8)	1.543 (4.9)	1.524 (5.1)	1.493 (6.2)	1.475 (6.1)
(5,15)	1.620 (5.3)	1.562 (5.4)	1.568 (5.3)	1.560 (5.2)	1.602 (4.5)	1.581 (4.8)	1.527 (6.3)	1.507 (6.6)
(5,20)	1.629 (4.8)	1.586 (5.0)	1.574 (5.4)	1.566 (5.5)	1.617 (4.8)	1.597 (5.0)	1.549 (5.7)	1.530 (6.1)
(5,30)	1.644 (4.9)	1.608 (5.0)	1.597 (5.1)	1.589 (5.1)	1.650 (5.3)	1.629 (5.7)	1.578 (6.0)	1.559 (6.6)
(5,40)	1.643 (4.7)	1.621 (5.3)	1.616 (4.9)	1.608 (5.0)	1.650 (4.5)	1.629 (5.0)	1.575 (5.7)	1.554 (6.6)
(5,50)	1.660 (4.5)	1.624 (5.0)	1.624 (5.3)	1.616 (5.5)	1.661 (5.4)	1.639 (5.5)	1.578 (5.6)	1.559 (5.9)
(5,100)	1.669 (4.7)	1.641 (5.2)	1.629 (5.0)	1.621 (5.5)	1.672 (5.1)	1.650 (5.4)	1.606 (5.9)	1.586 (6.2)
(5,250)	1.665 (4.8)	1.676 (5.3)	1.651 (4.5)	1.642 (4.8)	1.698 (4.2)	1.689 (4.6)	1.619 (6.7)	1.611 (6.8)
(5,500)	1.659 (5.7)	1.679 (5.1)	1.652 (5.2)	1.644 (5.2)	1.702 (4.5)	1.698 (4.7)	1.627 (5.8)	1.618 (6.0)
(5,1000)	1.661 (5.5)	1.665 (5.2)	1.660 (4.4)	1.652 (4.7)	1.809 (4.7)	1.801 (5.2)	1.635 (6.1)	1.627 (6.5)

Table 4.6: Powers of the tests for testing $H_0: \mu = 0$ versus $H_1: \mu \neq 0$ with $\alpha = 0.05$ with $n_1 = 9$ and $n_2 = 9$

$n_1 = 9$ $n_2 = 9$		μ					
$\frac{\sigma_2^2}{\sigma_1^2}$	**Tests**	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**
1	**K and L**	0.04	0.11	0.29	0.59	0.83	0.96
	Lin and Lee	0.07	0.14	0.37	0.67	0.89	0.98
	Bayesian	0.07	0.15	0.39	0.69	0.88	0.98
4	**K and L**	0.04	0.09	0.20	0.41	0.62	0.83
	Lin and Lee	0.06	0.11	0.25	0.49	0.69	0.88
	Bayesian	0.06	0.11	0.26	0.48	0.70	0.88
9	**K and L**	0.05	0.08	0.18	0.36	0.60	0.75
	Lin and Lee	0.05	0.09	0.22	0.41	0.65	0.80
	Bayesian	0.05	0.09	0.21	0.41	0.64	0.80
16	**K and L**	0.05	0.08	0.19	0.36	0.57	0.75
	Lin and Lee	0.05	0.10	0.22	0.39	0.60	0.78
	Bayesian	0.05	0.09	0.21	0.40	0.61	0.78

Table 4.7: Powers of the tests for testing $H_0: \mu = 0$ versus $H_1: \mu \neq 0$ with $\alpha = 0.05$ with $n_1 = 15$ and $n_2 = 10$

$n_1 = 15$ $n_2 = 10$		μ					
$\frac{\sigma_2^2}{\sigma_1^2}$	**Tests**	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1**
1	**K and L**	0.04	0.13	0.42	0.77	0.95	0.99
	Lin and Lee	0.06	0.17	0.48	0.82	0.97	1.00
	Bayesian	0.06	0.18	0.49	0.82	0.97	1.00
4	**K and L**	0.05	0.12	0.31	0.63	0.86	0.96
	Lin and Lee	0.05	0.14	0.34	0.68	0.88	0.97
	Bayesian	0.05	0.13	0.35	0.67	0.88	0.98
9	**K and L**	0.04	0.12	0.32	0.59	0.83	0.95
	Lin and Lee	0.05	0.13	0.34	0.62	0.85	0.96
	Bayesian	0.05	0.11	0.34	0.61	0.85	0.96
16	**K and L**	0.05	0.12	0.29	0.57	0.82	0.95
	Lin and Lee	0.05	0.13	0.34	0.62	0.85	0.96
	Bayesian	0.05	0.13	0.32	0.59	0.84	0.96

5. Illustrative Examples

We will use two examples to illustrate how the Bayesian method compares to the generalized variable methods. The first example is given in Meier (1953) and the second example is given in Eberhardt et al. (1989). Both these examples can also be found in Krishnamoorthy and Lu (2003) as well as Lin and Lee (2005).

5.1 . Example 1

The mean percentage of albumin, μ, in the plasma protein of normal human subjects were estimated by four different experiments. The results obtained from these four experiments together with the sample size for each of the experiments are given in Table 5.1 below. The reported variances in this example differ only slightly.

Table 5.1: Percentage of albumin in plasma protein estimated by four experiments

Experiment	n_i	$\bar{x}$	s^2
A	12	62.3	12.986
B	15	60.3	7.840
C	7	59.5	33.433
D	16	61.5	18.513

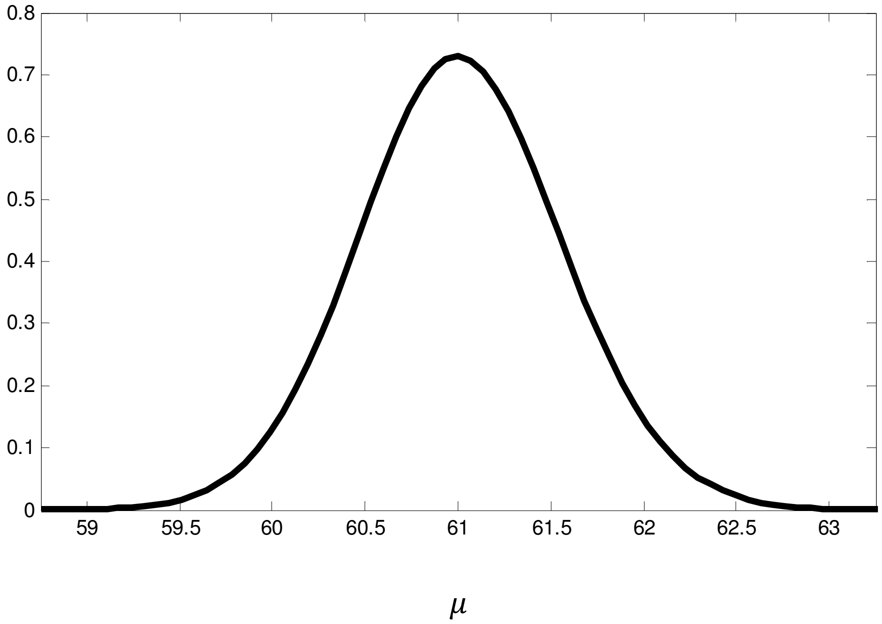
The next table compares 95% confidence intervals that were estimated by the three methods that were discussed in this paper.

Table 5.2: Estimated intervals for μ

Method		**Interval**	**Width**
Krishnamoorthy and Lu (2003)		(59.76, 62.23)	2.44
Lin and Lee (2005)		(59.92, 62.10)	2.18
Bayesian Method	**Equal Tail**	(59.91, 62.09)	2.18
	HPD	(59.91, 62.08)	2.17

From the results given in Table 5.2 we see that the Bayesian method is just as good as and even better than the Generalized Variable method by Lin and Lee. The equal tailed confidence interval is basically the same as the interval obtained by using the method proposed by Lin and Lee. The highest posterior density confidence interval is just a little bit shorter than the equal tailed interval. The shortest interval is therefore obtained by using the Bayesian method and using the highest posterior density confidence interval.

Figure 5.1: Posterior distribution of the average percentage of albumin in plasma protein in normal human subjects



5.2 . Example 2

The mean selenium content in non-fat milk powder was estimated by four different analytical methods. The results obtained from the four analytical methods together with the sample size for each of the methods are given in Table 5.3 below. The reported variances in this example differ significantly.

Table 5.3: Selenium in non-fat milk powder

Method	n_i	$\bar{x}$	s^2
Atomic absorption spectrometry	8	105.0	85.711
Neutron activation Instrumental	12	109.75	20.748
Neutron activation Radiochemical	14	109.5	2.729
Isotope dilution mass spectrometry	8	113.25	33.640

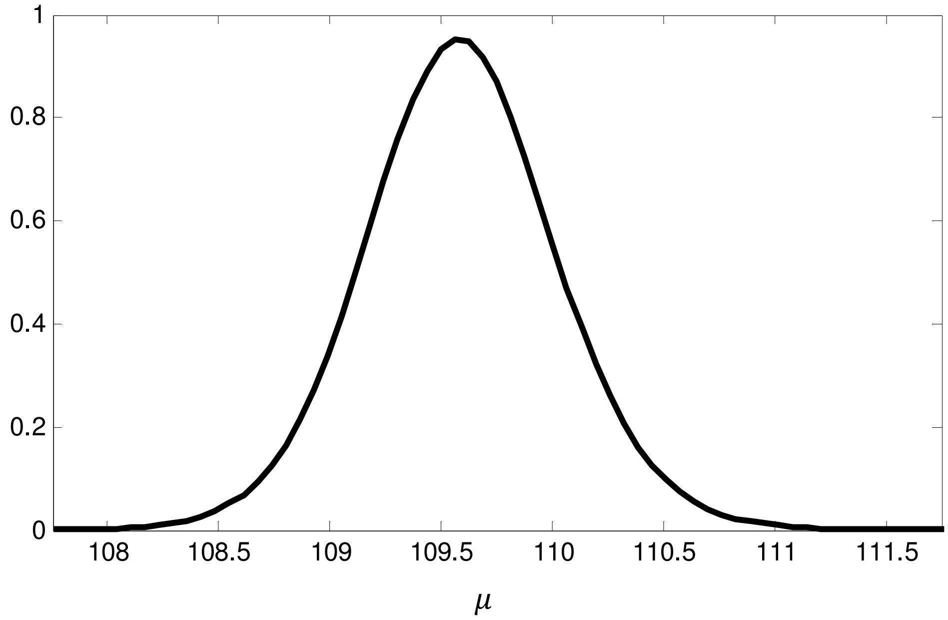
The next table compares 95% confidence intervals that were estimated by the three methods that were discussed in this paper.

Table 5.4: Estimated intervals for μ

Method		Interval	Width
Krishnamoorthy and Lu (2003)		(108.67, 110.53)	1.86
Lin and Lee (2005)		(108.75, 110.51)	1.76
Bayesian Method	**Equal Tailed**	(108.73, 110.47)	1.74
	HPD	(108.73, 110.46)	1.73

In Table 5.4 we see that again from the three methods, the one proposed by Krishnamoorthy and Lu gives the widest interval. The Bayesian technique delivers a somewhat shorter interval than the interval obtained by using Lin and Lee's method. Once more the highest posterior density confidence interval is slightly shorter than the equal tailed interval and again the shortest overall.

Figure 5.2: Posterior distribution of the average percentage of albumin in plasma protein in normal human subjects



6. Conclusion

A lot of work has been done in the area of estimating the common mean of normal populations. A large number of the methods proposed to estimate the common mean is based on the work done by Graybill and Deal in 1959. Many confidence intervals are centred about the estimate given in (2.3).

Krishnamoorthy and Lu (2003) developed a method based on weighted linear combinations of pivotal quantities. The procedure which Lin and Lee used to estimate a confidence interval, also based on pivotal quantities of the best unbiased estimator of μ. Both these methods followed from the estimator which Graybill and Deal developed. The third approach is a Bayesian method. Gregurich and Broemeling (1997) used the Independence Jeffreys prior to form confidence intervals for the common mean of several normal populations.

All of these are fairly easy to use and can be extended to incorporate the information contained in three or more samples to estimate a common mean. Therefore the method which one would prefer to use would only depend on the results obtained by the method.

From Tables 4.1 to 4.4 given above we can see that the first method, the one by Krishnamoorthy and Lu, does not have a single shaded value. Although this method shows coverage near 95%, the intervals produced by this method are wider than the intervals produced by the other two methods. From this result we can conclude that the other two methods are better for estimating the common mean of normal populations.

In Tables 4.1 to 4.4 we see that the columns for Prior II and Prior III also do not have a single shaded value. In the case of Prior II we see that the coverage is near 95%, but once more these intervals are wider than some of the other methods. When looking at the column for Prior III we see that all of these intervals are narrower that any of the other in each specific row. However, none of these intervals has coverage near 95%. While the intervals are narrower, they are in fact too narrow and show a much higher type I error. Therefore none of these values are shaded as these intervals would not be the best to use in analysis.

As seen from the first four tables above, all of the shaded values were calculated by only two methods. The simulation procedure by Lin and Lee as well as the Bayesian technique with the Jeffreys' independence prior delivered the best results. Both these methods have coverage around 95%. About half of the shortest intervals are from the Bayesian method, using the highest posterior density intervals rather than the equal tail intervals.

From these results we can conclude that the Bayesian method is just as good as the method from Lin and Lee, which they based on pivotal quantities. Both these methods deliver confidence intervals which are significantly narrower than the confidence intervals obtained by using the method which Krishnamoorthy and Lu developed.

In all four of the first tables we see that the Bayesian method delivers the shortest confidence intervals when the differences in the two population's variances get very large. In each table the last row's shaded value is the one from the Bayesian method when using Prior I and working with the highest posterior density confidence intervals. This shows that we would prefer to use the Bayesian method when there are significant differences in the population variances.

From the results given in Tables 4.5 and 4.6 we see that both the Bayesian method and the generalized variables method which is based on the best unbiased estimator of μ are uniformly more powerful than the method proposed by Krishnamoorthy and Lu. However we cannot say either of these is uniformly more powerful than the other.

The Bayesian method and the method proposed by Lin and Lee show powers which are very similar. In the cases where there is a difference in power between these two methods it is at most only a 0.02 difference. In most of the cases given in Tables 4.5 and 4.6 we see that these methods give the same power for testing the hypothesis as given in (3.1). We can also see that Both Lin and Lee's method and the Bayesian method show a slight tendency to be liberal tests when the ratio of the variances is smaller than or equal to four. When the true mean is equal to zero, we tend to reject more than 5% of the times when working with a 95% hypothesis test.

From the examples given in the previous section we can conclude that the Bayesian method performs just as well as the method by Lin and Lee. The equal tail intervals are basically the same as the Lin and Lee's intervals, while the highest posterior density interval are slightly shorter.

7. References

Berger, J. and Bernardo, J.O. (1992) 'On the Developement of Reference Priors', in *Bayesian Statistics*, 4th edition, Oxford University Press.

Chang, C.-H. and Pal, N. (2008) 'Testing on the common mean of several normal distributions', *Computational Statistics and Data Analysis*, no. 53, July, pp. 321-333.

Eberhardt, K.R., Reeve, C.P. and Spiegelman, C.H. (1989) 'A minimax approach to combining mean with practical examples', *Chemometrics ans Intelligent Laboratory Systems*, vol. 5, no. 2, January, pp. 129-148.

Fairweather, W.R. (1972) 'A Method of Obtaining an Exact Confidence Interval for the Common mean of Several Normal Population', *Applied Statistics*, vol. 21, no. 3, pp. 229-233.

Graybill, F.A. and Deal, R.B. (1959) 'Combining Unbiased Estimators', *Biometrics*, vol. 15, no. 4, pp. 543-550.

Gregurich, M.A. and Broemeling, L.D. (1997) 'A Bayesian analysis for estimating the common mean of independent normal populations using the Gibbs sampler', *Communications in Statistics - Simulation and Computation*, vol. 1, no. 26, pp. 35-51.

Krishnamoorthy, K. and Lu, Y. (2003) 'Inferences on the Common Mean of Several Normal Populations Based on the Generalized Variable Method', *Biometrics*, vol. 59, no. 2, pp. 237-247.

Lin, S.-H. and Lee, J.C. (2005) 'Generalized inferences on the common mean of several normal populations', *Journal of Statistical Planning and Inference*, vol. 134, no. 2, October, pp. 568-582.

Meier, P. (1953) 'Variance of a weighted mean', *Biometrics*, vol. 9, pp. 59-73.

Rubin, D.B. (1984) 'Bayesian justifiable and relevant frequency calculation fir the applied statistician', *The Annals of Statistics*, no. 12, pp. 1151-1172.

StatSoft (2011) *Statistics Glossary: G*, 2 May, [Online], Available: http://www.statsoft.com [12 May 2011].

The ASAP (2007) *Dictionary M: The ASAP*, 2 October, [Online], Available: http://anxiety-panic.com/dictionary/en-main.htm [13 April 2011].