Homework 2 Solution
of
STAT 632 Bayesian Statistics

Yanxun Xu

October 1, 2015

1 Problem 1

Solution:

(a) Let $X \sim exponential(\theta)$, then

$$P(x|\theta) = \theta e^{-\theta x}, x \geq 0.$$

The value of Fisher information can be obtained via

$$\begin{aligned}
I(\theta) &= -E[\frac{\partial^2}{\partial\theta^2}\log p(x|\theta)|\theta] \\
&= -E[\frac{\partial^2}{\partial\theta^2}\left[\log\theta + \log e^{-\theta x}\right]|\theta] \\
&= -E[\frac{\partial^2}{\partial\theta^2}\left[\log\theta - \theta x\right]|\theta] \\
&= -E[\frac{\partial}{\partial\theta}\left[\frac{1}{\theta} - x\right]|\theta] \\
&= -E[-\frac{1}{\theta^2}|\theta] \\
&= \frac{1}{\theta^2}.
\end{aligned}$$

Thus, $p_J(\theta) \propto \sqrt{I(\theta)} \propto \sqrt{\frac{1}{\theta^2}} \propto \frac{1}{\theta}$.

(b) The data likelihood of X is

$$p(X|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta n\bar{x}}.$$

Using the above Jeffrey's prior

$$p_J(\theta) \propto \frac{1}{\theta},$$

the posterior distribution of θ is

$$\begin{aligned}
p(\theta|X) &\propto p(X|\theta)p(\theta) \\
&\propto \theta^n e^{-\theta n\bar{x}}\frac{1}{\theta} \\
&\propto \theta^{n-1}e^{-\theta n\bar{x}}
\end{aligned}$$

$$= \quad gamma(n, n\bar{x}).$$

(c)

$$\begin{aligned} p(z|X) &= \int p(z, \theta|X) d\theta \\ &= \int p(z|\theta, X) p(\theta|X) d\theta \\ &= \int \theta e^{-\theta z} \theta^{n-1} e^{-\theta n \bar{x}} d\theta \\ &= \int \theta^n e^{\theta(-z - n\bar{x})} d\theta \\ &= \frac{\Gamma(n+1)}{(z + n\bar{x})^{n+1}}. \end{aligned}$$

(d) Write down the CDF of $p(z \mid x)$, then you can get any 95% of credible intervals.

2 Problem 2

Solution:

(a) Let $\theta \sim beta(a, b)$, then

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, 0 < \theta < 1.$$

Let $\psi = \log \frac{\theta}{1-\theta}$, then

$$\theta = h(\psi) = \frac{e^\psi}{1 + e^\psi}.$$

$$\begin{aligned} p_\psi(\psi) &= p_\theta(h(\psi)) \left| \frac{d}{d\psi} h(\psi) \right| \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{e^\psi}{1+e^\psi}\right)^{a-1} \left(1 - \frac{e^\psi}{1+e^\psi}\right)^{b-1} \frac{e^\psi(1+e^\psi) - e^\psi e^\psi}{(1+e^\psi)^2} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{e^{\psi(a-1)}}{(1+e^\psi)^{a+b-2}} \frac{e^\psi}{(1+e^\psi)^2} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{e^{\psi a}}{(1+e^\psi)^{a+b}}, -\infty < \psi < \infty. \end{aligned}$$

When $a = b = 1$, $p_\psi(\psi) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \frac{e^\psi}{(1+e^\psi)^{1+1}} = \frac{e^\psi}{(1+e^\psi)^2}$ and the corresponding plot is shown in Figure 1 (Left).

(b) Let $\theta \sim gamma(a, b)$, then

$$p(\theta) = b^a \frac{1}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, \theta \geq 0.$$

Let $\psi = \log \theta$, then

$$\theta = h(\psi) = e^\psi.$$

$$
\begin{aligned}
p_\psi(\psi) &= p_\theta(h(\psi))|\frac{d}{d\psi}h(\psi)| \\
&= b^a\frac{1}{\Gamma(a)}(e^\psi)^{a-1}e^{-be^\psi}e^\psi \\
&= b^a\frac{1}{\Gamma(a)}e^{\psi a-be^\psi}, -\infty < \psi < \infty.
\end{aligned}
$$

When $a = b = 1$, $p_\psi(\psi) = \frac{1}{\Gamma(1)}e^{\psi-e^\psi} = e^{\psi-e^\psi}$ and the corresponding plot is shown in Figure 1 (Right).

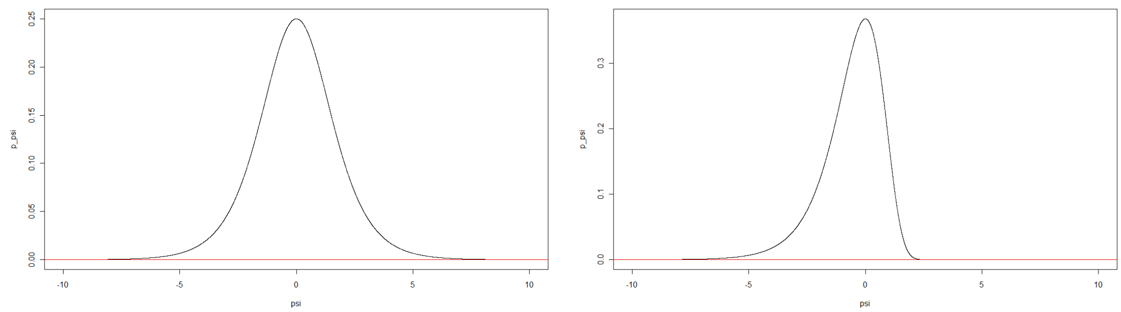

Figure 1: The plots of $p_\psi(\psi)$ against ψ.

3 Problem 3

Solution:

(a) Let $Y \sim binomial(n, \theta)$, then

$$
P(y|\theta) = \left(\begin{array}{c} n \\ y \end{array}\right)\theta^y(1-\theta)^{n-y}, 0 \le \theta \le 1.
$$

The value of Fisher information can be obtained via

$$
\begin{aligned}
I(\theta) &= -E[\frac{\partial^2}{\partial\theta^2}\log p(y|\theta)|\theta] \\
&= -E[\frac{\partial^2}{\partial\theta^2}\left[\log\left(\begin{array}{c} n \\ y \end{array}\right) + \log\theta^y + \log(1-\theta)^{n-y}\right]|\theta] \\
&= -E[\frac{\partial^2}{\partial\theta^2}\left[\log\left(\begin{array}{c} n \\ y \end{array}\right) + y\log\theta + (n-y)\log(1-\theta)\right]|\theta] \\
&= -E[\frac{\partial}{\partial\theta}\left[\frac{y}{\theta} - \frac{n-y}{(1-\theta)}\right]|\theta] \\
&= -E[-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}|\theta] \\
&= \frac{E[Y|\theta]}{\theta^2} + \frac{n-E[Y|\theta]}{(1-\theta)^2} \\
&= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2}
\end{aligned}
$$

$$\begin{aligned} &= \frac{n}{\theta} + \frac{n}{1-\theta} \\ &= \frac{n}{\theta(1-\theta)}. \end{aligned}$$

Thus, $p_J(\theta) \propto \sqrt{I(\theta)} \propto \sqrt{\frac{1}{\theta(1-\theta)}}$.

(b) In this situation,

$$P(y|\psi) = \left(\begin{array}{c} n \\ y \end{array} \right) e^{\psi y}(1+e^{\psi})^{-n}, -\infty < \psi < \infty.$$

The value of Fisher information can be obtained via

$$\begin{aligned} I(\psi) &= -E[\frac{\partial^2}{\partial \psi^2} \log p(y|\psi)|\psi] \\ &= -E[\frac{\partial^2}{\partial \psi^2} \left[\log \left(\begin{array}{c} n \\ y \end{array} \right) + \log e^{\psi y} + \log{(1+e^{\psi})^{-n}} \right] |\psi] \\ &= -E[\frac{\partial^2}{\partial \psi^2} \left[\log \left(\begin{array}{c} n \\ y \end{array} \right) + \psi y - n \log{(1+e^{\psi})} \right] |\psi] \\ &= -E[\frac{\partial}{\partial \psi} \left[y - n \frac{e^{\psi}}{1+e^{\psi}} \right] |\psi] \\ &= -E[-n \frac{e^{\psi}(1+e^{\psi}) - e^{\psi}e^{\psi}}{(1+e^{\psi})^2} |\psi] \\ &= -E[-n \frac{e^{\psi}}{(1+e^{\psi})^2} |\psi] \\ &= nE[\frac{e^{\psi}}{(1+e^{\psi})^2} |\psi] \\ &= n \frac{e^{\psi}}{(1+e^{\psi})^2}. \end{aligned}$$

Thus, $p_J(\psi) \propto \sqrt{I(\psi)} \propto \sqrt{n \frac{e^{\psi}}{(1+e^{\psi})^2}} \propto \frac{\sqrt{e^{\psi}}}{1+e^{\psi}}$.

(c) Let $\psi = \log \frac{\theta}{1-\theta}$, then

$$\theta = h(\psi) = \frac{e^{\psi}}{1+e^{\psi}}.$$

From the result of Problem 3.12(a), we have

$$p_J(\theta) \propto \sqrt{\frac{1}{\theta(1-\theta)}}.$$

$$\begin{aligned} p_J(\psi) &= p_J(\theta) |\frac{dh(\psi)}{d\psi}| \\ &= \sqrt{\frac{1}{\frac{e^{\psi}}{1+e^{\psi}}(1 - \frac{e^{\psi}}{1+e^{\psi}})}} \frac{e^{\psi}}{(1+e^{\psi})^2} \\ &= \frac{\sqrt{e^{\psi}}}{1+e^{\psi}}, \end{aligned}$$

which is as the same as the result of Problem 2.

4 Problem 4

Solution:

The data likelihood of Y is

$$\begin{aligned} p(Y|\mu,\sigma^2) &\propto \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i-\mu)^2\right) \\ &\propto (\frac{1}{\sigma^2})^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right). \end{aligned}$$

As $\mu|\sigma^2 \sim N(\mu_0, \frac{\sigma_0}{\kappa_0})$ and $\frac{1}{\sigma^2} \sim gamma(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2})$, the conjugate prior of μ and σ^2 can be written as,

$$\begin{aligned} p(\mu,\sigma^2) &= p(\mu|\sigma^2)p(\sigma^2) \\ &\propto (\frac{\sigma^2}{\tau_0})^{-\frac{1}{2}} \exp\left(-\frac{1}{2\frac{\sigma^2}{\tau_0}}(\mu-\mu_0)^2\right)(\sigma^2)^{-(\frac{\nu_0}{2}+1)} \exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right) \\ &\propto (\sigma^2)^{-(\frac{\nu_0+1}{2}+1)} \exp\left(-\frac{\tau_0}{2\sigma^2}\left(\frac{\nu_0\sigma_0^2}{\tau_0}+(\mu-\mu_0)^2\right)\right) \end{aligned}$$

The posterior distribution is

$$\begin{aligned} p(\mu,\sigma^2|Y) &\propto p(Y|\mu,\sigma^2)p(\mu,\sigma^2) \\ &\propto (\frac{1}{\sigma^2})^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)(\sigma^2)^{-(\frac{\nu_0+1}{2}+1)} \exp\left(-\frac{\tau_0}{2\sigma^2}\left(\frac{\nu_0\sigma_0^2}{\tau_0}+(\mu-\mu_0)^2\right)\right) \\ &\propto (\sigma^2)^{-(\frac{\nu_n+1}{2}+1)} \exp\left(-\frac{\tau_n}{2\sigma^2}\left(\frac{SS_n^2}{\tau_n}+(\mu-\mu_n)^2\right)\right), \end{aligned}$$

where $\tau_n = \tau_0 + n$, $\nu_n = \nu_0 + n$, $\mu_n = \frac{\tau_0\mu_0 + n\bar{y}}{\tau_n}$, and $SS_n = \nu_0\sigma_0^2 + \sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{\tau_0 n}{\tau_n}(\bar{y}-\mu_0)^2$.

The marginal posterior distribution $p(\mu|Y)$ is obtained integrating $p(\mu,\sigma^2|Y)$ over σ^2.

$$\begin{aligned} p(\mu|Y) &\propto \int_{\sigma^2} p(\mu,\sigma^2|Y)d\sigma^2 \\ &\propto \int_0^{\infty}(\sigma^2)^{-(\frac{\nu_n+1}{2}+1)} \exp\left(-\frac{\tau_n}{2\sigma^2}\left(\frac{SS_n^2}{\tau_n}+(\mu-\mu_n)^2\right)\right)d\sigma^2 \\ &\propto \left[1+\left(\frac{\mu-\mu_n}{\frac{SS_n}{\sqrt{\tau_n}}}\right)^2\right]^{-(\frac{\nu_n+1}{2})} \\ &\propto t_{\nu_n}(\mu_n, \frac{SS_n^2}{\nu_n\tau_n}). \end{aligned}$$

The marginal posterior distribution $p(\sigma^2|Y)$ is obtained integrating $p(\mu,\sigma^2|Y)$ over μ.

$$\begin{aligned} p(\sigma^2|Y) &\propto \int_{\mu} p(\mu,\sigma^2|Y)d\mu \\ &\propto \int_{-\infty}^{\infty}(\sigma^2)^{-(\frac{\nu_n+1}{2}+1)} \exp\left(-\frac{\tau_n}{2\sigma^2}\left(\frac{SS_n^2}{\tau_n}+(\mu-\mu_n)^2\right)\right)d\mu \end{aligned}$$

$$
\begin{aligned}
&\propto \quad (\sigma^2)^{-(\frac{\nu_n}{2}+1)} \exp\left(-\frac{SS_n^2}{2\sigma^2}\right) \\
&\propto \quad inverse - gamma(\frac{\nu_n}{2}, \frac{SS_n^2}{2}).
\end{aligned}
$$

Assume $\theta_0 = 2$, $\sigma_0^2 = 0.01$, $\nu_0 = 1$, and $\tau_0 = 2$, we simulated a Monte Carlo estimate of the marginal distribution. The comparison of the two results is shown in Figure 2, which suggests little difference between the two.

```
mu_0 <- 2
sigma_0 <- 0.01
nu_0 <- 1
tau_0 <- 2
y <- rnorm(10000, mean = 10, sd = 0.5)
n <- length(y)
bar_y <- mean(y)
s <- var(y)
tau_n <- tau_0 + n
nu_n <- nu_0 + n
mu_n<- (tau_0 * mu_0 + n * bar_y) / tau_n
sigma_n<-(nu_0 * sigma_0 + (n - 1) * s + tau_0 * n * (bar_y - mu_0)^2 / (tau_n)) / (nu_n)
sigma <- 1 / rgamma(10000, nu_n / 2, nu_n * sigma_n / 2)
mu <- rnorm(10000, mu_n, (sigma / tau_n)^0.5)
tt<-rt(10000, df = nu_n) * sqrt(sigma_n / tau_n) + mu_n
ggplot() + geom_freqpoly(aes(mu), color = "blue") + geom_freqpoly(aes(tt), color = "red")
```

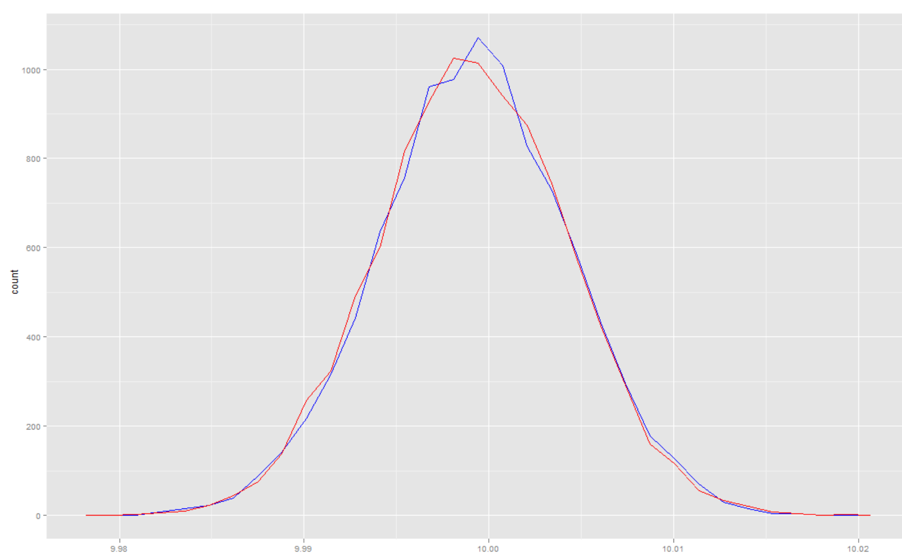

Figure 2: The marginal distributions of $p(\mu|Y)$ obtained by formula and Monte Carlo estimate.

5 Problem 5

Note: In (b) and (c), you need to carefully differentiate θ and $\tilde{Y}$. For whom having problems in this question, please talk to your classmates and ask for their codes.

Solution:

(a) Using the formula

$$\mu_n = \frac{\tau_0}{\tau_0 + n}\mu_0 + \frac{n}{\tau_0 + n}\bar{y},$$

we calculated the posterior means and using Monete Carlo sampling, we obtained the confidence intervals for the mean and the standard variance for the three data sets, as shown in Table 1.

Table 1: The data likelihood for each parameter

	School 1	School2	School 3
μ_n	9.292	6.949	7.812
$[\overline{\mu_n}, \underline{\mu_n}]$	$[7.777, 10.790]$	$[5.173, 8.691]$	$[6.195, 9.476]$
$[\overline{\sigma_n}, \underline{\sigma_n}]$	$[2.992, 5.156]$	$[3.345, 5.940]$	$[2.799, 5.136]$

(b) The posterior probability that $\theta_i < \theta_j < \theta_k$ for all six permutations are shown in Table 2.

Table 2: The data likelihood for each parameter

	$1<2<3$	$1<3<2$	$2<1<3$	$2<3<1$	$3<1<2$	$3<2<1$
Prob.	0.0047	0.0041	0.0425	0.7504	0.0162	0.1821

(c) The posterior probability that $\widetilde{Y}_i < \widetilde{Y}_j < \widetilde{Y}_k$ for all six permutations are shown in Table 3.

Table 3: The data likelihood for each parameter

	$\widetilde{Y}_1<\widetilde{Y}_2<\widetilde{Y}_3$	$\widetilde{Y}_1<\widetilde{Y}_3<\widetilde{Y}_2$	$\widetilde{Y}_2<\widetilde{Y}_1<\widetilde{Y}_3$	$\widetilde{Y}_2<\widetilde{Y}_3<\widetilde{Y}_1$	$\widetilde{Y}_3<\widetilde{Y}_1<\widetilde{Y}_2$	$\widetilde{Y}_3<\widetilde{Y}_2<\widetilde{Y}_1$
Prob.	0.1020	0.1021	0.1925	0.2602	0.1457	0.1970

(d) The posterior probability $P(\theta_1 > \theta_2, \theta_1 > \theta_3 | Y_1, Y_2, Y_3) = 0.8935$ and the posterior predictive probability $P(\widetilde{Y}_1 > \widetilde{Y}_2, \widetilde{Y}_1 > \widetilde{Y}_3 | Y_1, Y_2, Y_3) = 0.4577$

6 Problem 6

Solution:

(a) The conjugate prior distribution for the multinomial model is

$$p(\Theta) = Dirichlet(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

And the corresponding posterior distribution is

$$p(\Theta|Y) = Dirichlet(y_1 + \alpha_1, y_2 + \alpha_2, \cdots, y_n + \alpha_n).$$

According to one of the properties of the Dirichlet distribution, the marginal posterior distribution of θ_1 and θ_2 is also Dirichlet

$$p(\theta_1, \theta_2|Y) \propto \theta_1^{y_1+\alpha_1-1}\theta_2^{y_2+\alpha_2-1}(1-\theta_1-\theta_2)^{\sum_{j=3}^{J} y_j + \sum_{j=3}^{J} \alpha_j - 1}.$$

Let $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$ and $\beta = \theta_1 + \theta_2$. We have $\theta_1 = \alpha\beta$ and $\theta_2 = \beta - \alpha\beta$. Then, we can obtain that

$$\frac{\partial\theta_1}{\partial\alpha} = \beta,$$

$$\frac{\partial\theta_1}{\partial\beta} = \alpha,$$

$$\frac{\partial\theta_2}{\partial\alpha} = -\beta,$$

$$\frac{\partial\theta_2}{\partial\beta} = 1 - \alpha.$$

And, $J = \frac{\partial\theta_1}{\partial\alpha}\frac{\partial\theta_2}{\partial\beta} - \frac{\partial\theta_1}{\partial\beta}\frac{\partial\theta_2}{\partial\alpha} = \beta$. The joint pdf of α and β is given by

$$\begin{aligned}
p(\alpha,\beta|Y) &\propto p_{\theta_1,\theta_2}(\alpha\beta, \beta-\alpha\beta)|J| \\
&\propto (\alpha\beta)^{y_1+\alpha_1-1}(\beta-\alpha\beta)^{y_2+\alpha_2-1}(1-\beta)^{\sum_{j=3}^J y_j+\sum_{j=3}^J \alpha_j-1}\beta \\
&\propto \alpha^{y_1+\alpha_1-1}(1-\alpha)^{y_2+\alpha_2-1}\beta^{y_1+y_2+\alpha_1+\alpha_2-1}(1-\beta)^{\sum_{j=3}^J y_j+\sum_{j=3}^J \alpha_j-1} \\
&\propto beta(\alpha; y_1+\alpha_1, y_2+\alpha_2)beta(\beta; y_1+y_2+\alpha_1+\alpha_2, \sum_{j=3}^J y_j + \sum_{j=3}^J \alpha_j).
\end{aligned}$$

Thus, we have $p(\alpha|Y) \sim beta(y_1+\alpha_1, y_2+\alpha_2)$, which can be also derived from a $beta(\alpha;\alpha_1,\alpha_2)$ prior distribution and a binomial observation y_1 with sample size $y_1 + y + 2$.

(b) The conjugate prior distribution for the multinomial model is

$$p(\Theta) = Dirichlet(1, 1, 1).$$

And the corresponding posterior distributions before and after the debate are

$$p(\Theta|Y) = Dirichlet(294+1, 307+1, 38+1),$$

and

$$p(\Psi|Y) = Dirichlet(288+1, 332+1, 19+1).$$

Let $\alpha_1 = \frac{\theta_1}{\theta_1+\theta_2}$ and $\alpha_2 = \frac{\psi_1}{\psi_1+\psi_2}$, then according the the result of Problem 6(a), we have

$$p(\alpha_1|Y) = beta(295, 308),$$

and

$$p(\alpha_2|Y) = beta(289, 333).$$

Using the following code, we plotted the histogram of the posterior density for $\alpha_2 - \alpha_1$, as shown in Figure 3, and calculated the posterior probability that there was a shift toward Bush that was equal to 0.1922.

```
alpha_1 <- rbeta (10000, 295, 308)
alpha_2 <- rbeta (10000, 289, 333)
hist (alpha_2 - alpha_1, breaks = seq(-0.15, 0.15, by = 0.01))
mean(alpha_2 - alpha_1 > 0)
```

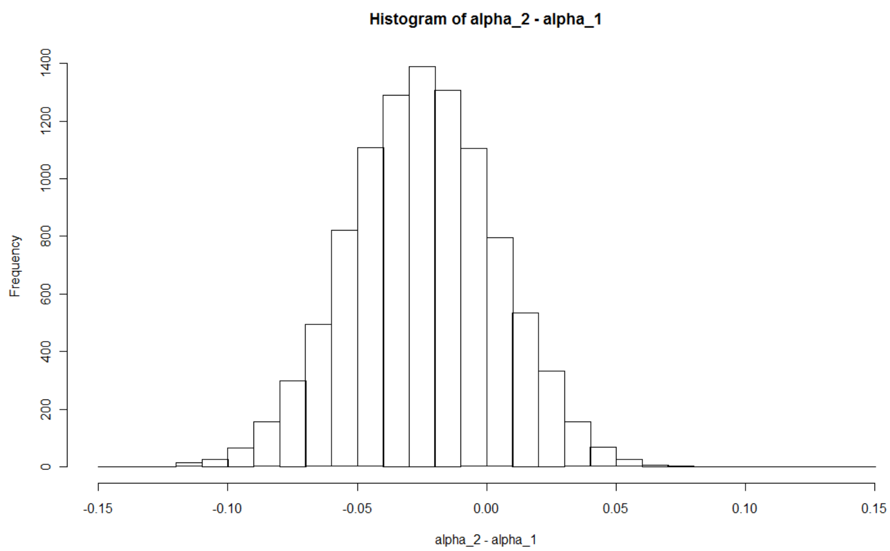

Figure 3: The histogram of the posterior density for $\alpha_2 - \alpha_1$.