ST4234 Bayesian Statistics

Data Analysis Assignment

Tan Jun Yu (A0183140R)

Question 1 (a)

We have $y_i \sim N(\theta_j, \sigma_j^2)$ for known σ_j^2, $\theta_j \overset{\text{iid}}{\sim} N(\mu, \tau^2)$, and $p(\mu, \tau) \propto 1$. The unnormalized full posterior density is:

$$
\begin{aligned}
p(\theta, \mu, \tau | y, \sigma) &\propto p(\theta|\mu.\tau) \times p(\mu, \tau) \times p(y|\theta, \sigma) \\
&\propto \prod_{j=1}^{J} p(\theta_j | \mu, \tau) p(y_j | \theta_j, \sigma_j) \\
&\propto \prod_{j=1}^{J} \left(\frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{(\theta_j - \mu)^2}{2\tau^2} \right) \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{(y_j - \theta_j)^2}{2\sigma_j^2} \right) \right) \\
&\propto \prod_{j=1}^{J} \left(\frac{1}{\tau\sigma_j} \exp\left(-\frac{(\theta_j - \mu)^2}{2\tau^2} - \frac{(y_j - \theta_j)^2}{2\sigma_j^2} \right) \right)
\end{aligned}
$$

Question 1 (b)

From the full posterior, we can separate them into the posterior distributions of $p(\theta_j|\mu,\tau,y,\sigma)$ and $p(\mu,\tau|y,\sigma)$, by making a full square on θ_j similar to the process in Chapter 3. We also note that

$$\frac{1}{\sigma^2} + \frac{1}{\tau^2} = \frac{\tau^2 + \sigma_j^2}{\sigma_j^2\tau^2} \implies \sigma_j\tau = \sqrt{\frac{\tau^2 + \sigma_j^2}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}}$$

which will be of use later.

$$p(\theta,\mu,\tau|y,\sigma) \propto \prod_{j=1}^{J} \frac{1}{\tau\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{(\theta_j-\mu)^2}{\tau^2} + \frac{(y_j-\theta_j)^2}{\sigma_j^2}\right)\right\}$$

$$\propto \prod_{j=1}^{J} \frac{1}{\tau\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{\sigma_j^2(\theta_j-\mu)^2 + \tau^2(y_j-\theta_j)^2}{\tau^2\sigma_j^2}\right)\right\}$$

$$\propto \prod_{j=1}^{J} \frac{1}{\tau\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{\sigma_j^2(\theta_j^2 - 2\mu\theta_j - \mu^2) + \tau^2(y_j^2 - 2y_j\theta_j - \theta_j^2)}{\tau^2\sigma_j^2}\right)\right\}$$

$$\propto \prod_{j=1}^{J} \frac{1}{\tau\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{\sigma_j^2(\theta_j^2 + \tau^2) - 2\theta_j(\mu\sigma_j^2 + y_j^2) + \sigma_j^2\mu^2 + \tau^2 y_i^2}{\tau^2\sigma_j^2}\right)\right\}$$

$$\propto \prod_{j=1}^{J} \frac{1}{\tau\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{(\theta_j^2 + \tau^2)\left[\theta_j - \frac{\mu\sigma_j^2 + y_j\tau^2}{\sigma_j^2 + \tau^2}\right]^2 - \frac{(\mu\sigma_j^2 + y_j\tau^2)^2}{\sigma_j^2 + \tau^2} + \sigma_j^2\mu^2 + \tau^2 y_j^2}{\tau^2\sigma_j^2}\right)\right\}$$

$$\propto \prod_{j=1}^{J} \sqrt{\frac{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}{\tau^2 + \sigma_j^2}} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}\right)\left[\theta_j - \frac{\mu/\tau^2 + y_j/\sigma_j^2}{1/\tau^2 + 1/\sigma^2}\right]^2\right.$$

$$\left. -\frac{1}{2\tau^2\sigma_j^2} \times \frac{-\mu^2\sigma_j^4 - 2\mu\sigma_j^2 y_j\tau^2 - y_j^2\tau^4 + \sigma_j^4\mu^4 + \sigma_j^2\mu^2\tau^2 + \tau^2 y_i^2\sigma_j^2 + \tau^4 y_j^2}{\sigma_j^2 + \tau^2}\right\}$$

$$\propto \underbrace{\prod_{j=1}^{J} \sqrt{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}\right)\left[\theta_j - \frac{\mu/\tau^2 + y_j/\sigma_j^2}{1/\tau^2 + 1/\sigma^2}\right]^2\right\}}_{\theta_j|\mu,\tau,y,\sigma \sim N(\hat{\theta}_j, V_j)}$$

$$\times \underbrace{\prod_{j=1}^{J} \frac{1}{\sqrt{\tau^2 + \sigma_j^2}} \exp\left\{-\frac{1}{2}\frac{(\mu - y_j)^2}{\sigma_j^2 + \tau^2}\right\}}_{\phi\left(y_j|\mu, \sqrt{\sigma_j^2 + \tau^2}\right)}$$

where $\hat{\theta}_j = \frac{y_j/\sigma_j^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2}$, $V_j = \frac{1}{1/\sigma_j^2 + 1/\tau^2}$, and $\phi(y|\mu,\sigma)$ denotes the normal density with mean μ and standard deviation σ.

Question 1 (c)

```r
y <- c(28, 8, -3, 7, -1, 1, 18, 12)
sigma <- c(15, 10, 16, 11, 9, 11, 10, 18)

# defining the log posterior for lambda
logpost <- function(lambda, sigma, y){
  sum(-0.5*log(exp(2*lambda[2])+sigma^2) -
        ((lambda[1]-y)^2)/(2*(sigma^2+exp(2*lambda[2])))) +
        lambda[2]
}

# grids
lambda_1 <- seq(from = -18, to = 37, by = 0.1)
lambda_2 <- seq(from = -6, to = 4.1, by = 0.1)
z <- matrix(0, nrow = length(lambda_1), ncol = length(lambda_2))

for (i in 1:length(lambda_1)){
  for (j in 1:length(lambda_2)){
    lambda <- c(lambda_1[i], lambda_2[j])
    z[i,j] <- logpost(lambda, sigma, y)
  }
}

contour(x = lambda_1, y = lambda_2, z = z, col = "blue", nlevels = 40,
        xlab = expression(lambda[1]), ylab = expression(lambda[2]),
        cex.axis = 1.1, cex.lab = 1.3)
```
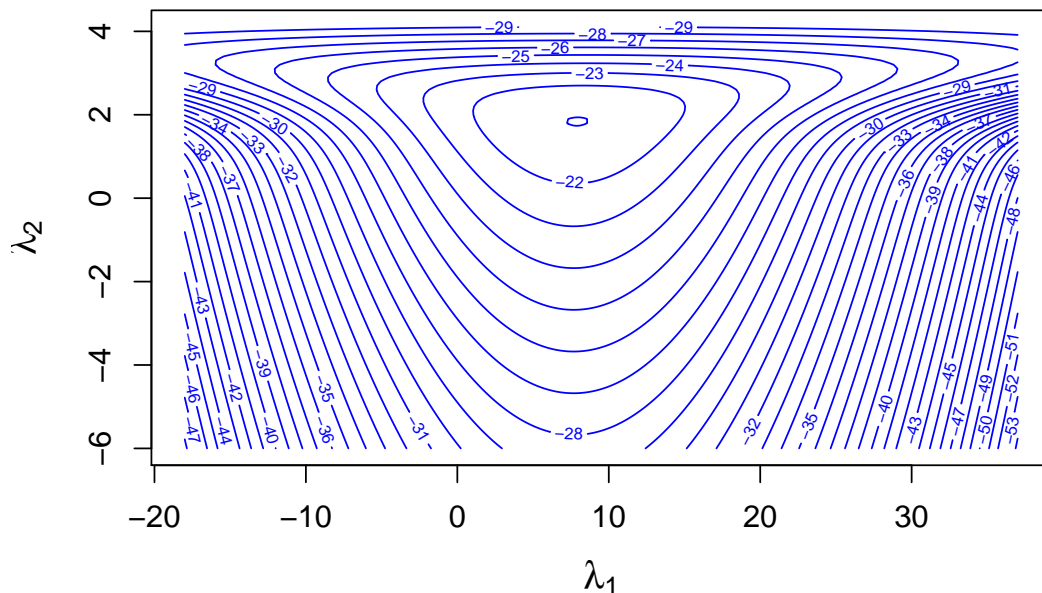

From the contour plot, the mode seems close to $(8, 2)$. We shall use this as a starting guess in `optim()` to find the posterior mode and covariance matrix.

```r
out <- optim(par = c(8, 2), fn = logpost, control = list(fnscale = -1),
             hessian = TRUE, sigma = sigma, y = y)
```

3

```
(post_mode <- out$par)
```

```
## [1] 7.926685 1.841525
```

```
(post_cov <- -solve(out$hessian))
```

```
##            [,1]      [,2]
## [1,] 22.3232882 0.1935228
## [2,]  0.1935228 0.5352576
```

The normal approximation to the posterior of (λ_1, λ_2) is

$$\lambda_1, \lambda_2 | \sigma, y \sim N\left(\begin{bmatrix} 7.926685 \\ 1.841525 \end{bmatrix}, \begin{bmatrix} 22.3232882 & 0.1935228 \\ 0.1935228 & 0.5352576 \end{bmatrix}\right)$$

Question 1 (d)

```r
library(LearnBayes)
library(coda)

set.seed(0183140)

iters <- 10^4
proposal <- list(var = post_cov, scale = 3)

# random walk metropolis
fit1 <- rwmetrop(logpost, proposal, start = post_mode, iters, sigma, y)

# overlaying last 5000 draws on contour plot of logpost
contour(x = lambda_1, y = lambda_2, z = z, col = "blue", nlevels = 40,
        xlab = expression(lambda[1]), ylab = expression(lambda[2]),
        cex.axis = 1.1, cex.lab = 1.3)
points(x = fit1$par[5001:iters,1], y = fit1$par[5001:iters,2], col = "red")
```
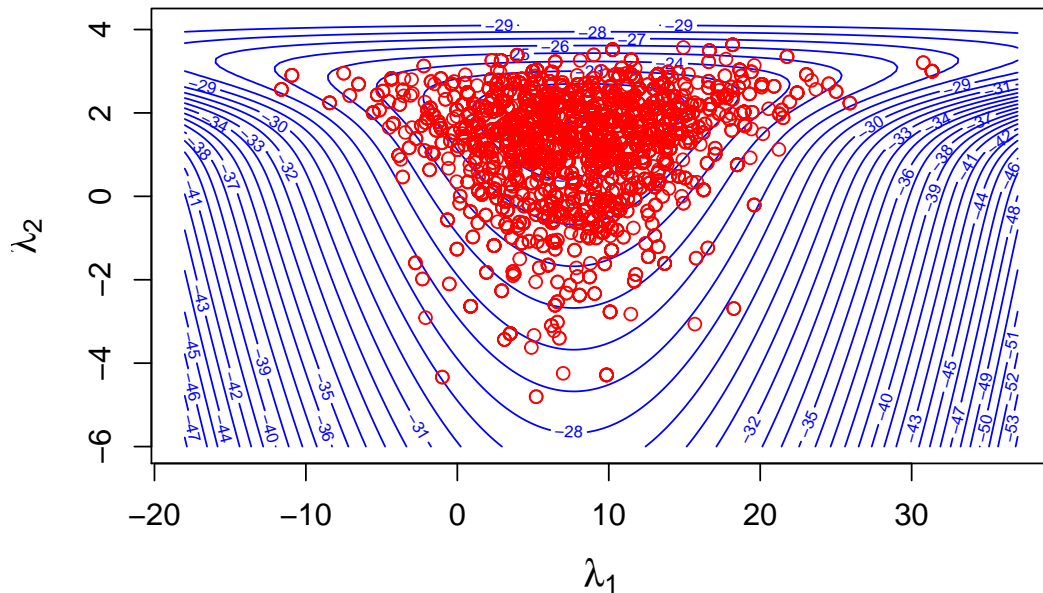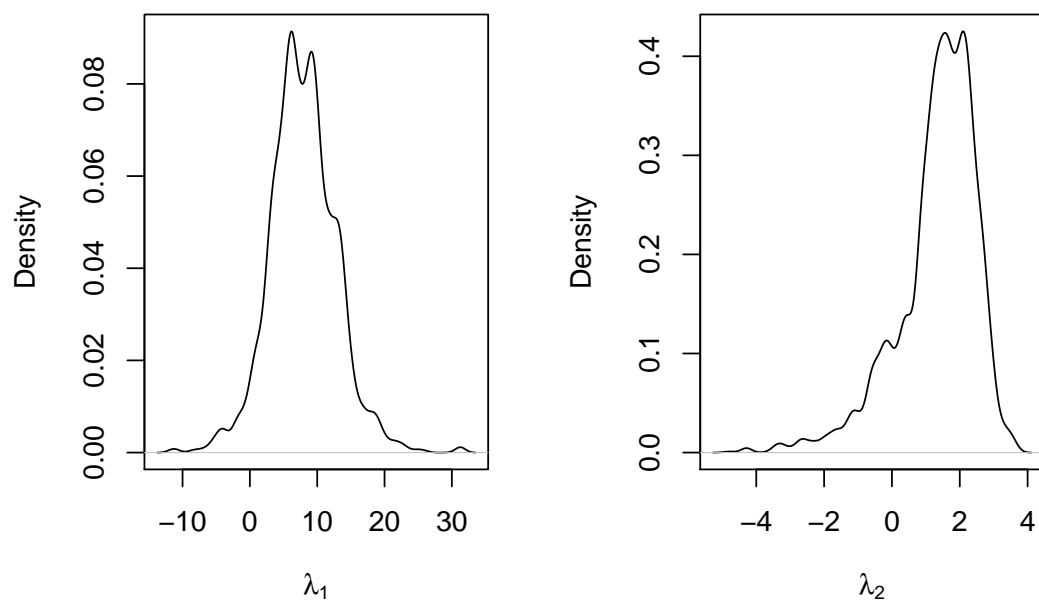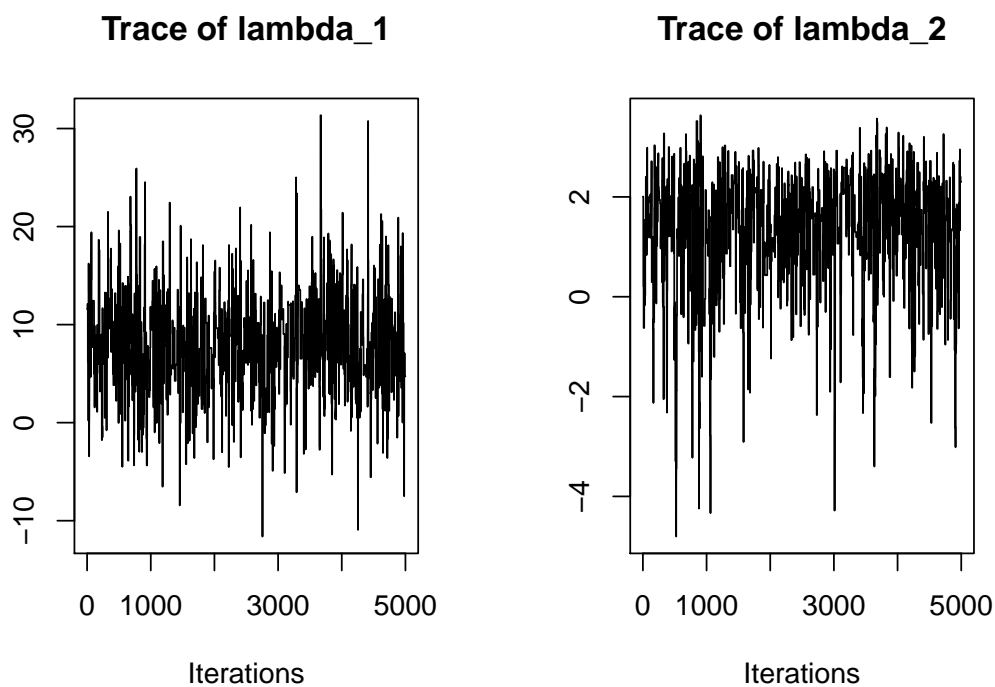


```r
# acceptance rate
fit1$accept
```

```
## [1] 0.1997
```

```r
par(mfrow=c(1,2))
plot(density(fit1$par[5001:iters,1]), main = "", xlab = expression(lambda[1]))
plot(density(fit1$par[5001:iters,2]), main = "", xlab = expression(lambda[2]))
```

The marginal posterior densities of λ_1 and λ_2 appear to be unimodal.

```r
mcmcobj1 <- mcmc(fit1$par[5001:iters,])
colnames(mcmcobj1) <- c("lambda_1", "lambda_2")
par(mfrow=c(1,2))
traceplot(mcmcobj1)
```


The traceplots of both λ_1 and λ_2 resembles random noise, generally showing great flunctuation.

This suggests that the samples of both λ_1 and λ_2 do not have high serial correlation/dependence and mixes well.

```
par(mfrow=c(1,2))
autocorr.plot(mcmcobj1, auto.layout = FALSE)
```
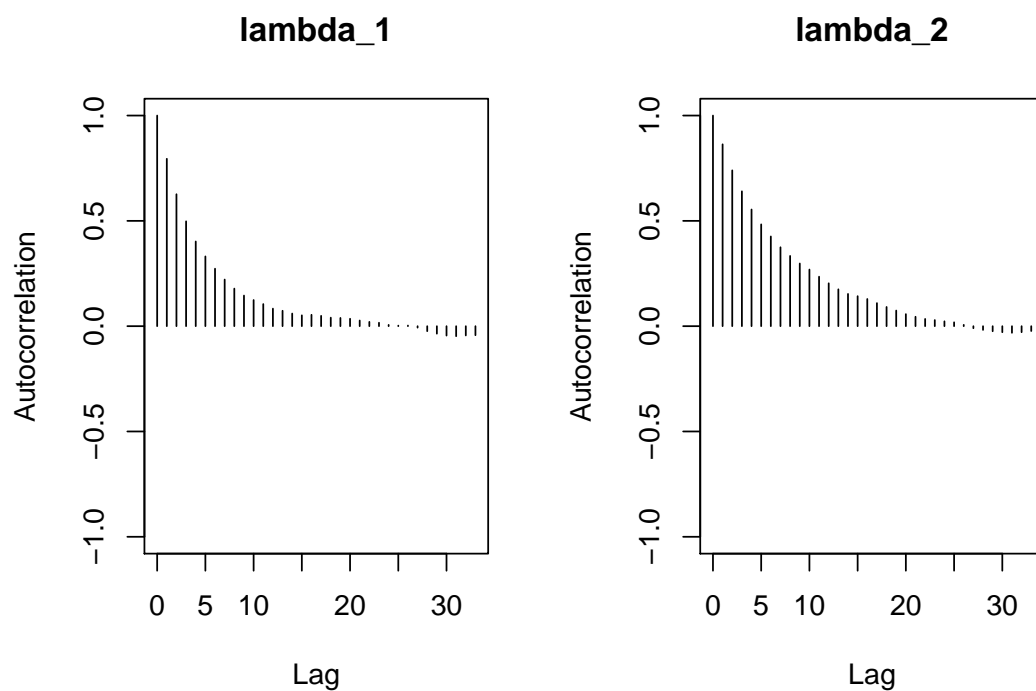
lambda_1

lambda_2

The autocorrelation plots show fast decay in both λ_1 and λ_2; autocorrelations are close to 1 for lag one but reduce quickly as a function of lag, indicating low serial dependence.

Question 1 (e)

```r
# the last 5000 MCMC samples
# contains (lambda_1, lambda_2)
lambda_samples <- fit1$par[5001:iters,]

# function to compute mean for each y_j
theta_hat <- function(lambda, y_j, sigma_j){
  ((y_j/sigma_j^2)+(lambda[,1]/exp(2*lambda[,2]))) /
  ((1/sigma_j^2)+(1/exp(2*lambda[,2])))
}

# function to compute variance for each y_j
V <- function(lambda, y_j, sigma_j){
  1 / (1/sigma_j^2 + 1/exp(2*lambda[,2]))
}

# drawing 5000 samples of theta_j
theta_samples <- function(lambda, y_j, sigma_j){
  rnorm(5000, mean = theta_hat(lambda, y_j, sigma_j),
        sd = sqrt(V(lambda, y_j, sigma_j)))
}

theta_mean <- rep(0, 8)
theta_sd <- rep(0,8)

# the joint posterior density of (theta_1,...,theta_j)
theta_all <- matrix(0, nrow = 5000, 8)

for (j in 1:8){
  thetas <- theta_samples(lambda_samples, y[j], sigma[j])
  theta_all[,j] <- thetas
  theta_mean[j] <- mean(thetas)
  theta_sd[j] <- sd(thetas)
}

(theta_dist <- cbind(theta_mean, theta_sd))
```

```
##       theta_mean theta_sd
## [1,]  10.984585 7.905697
## [2,]   7.759098 6.074982
## [3,]   6.270945 7.414442
## [4,]   7.668265 6.185918
## [5,]   5.228467 6.056489
## [6,]   6.310814 6.379558
## [7,]  10.395054 6.562673
## [8,]   8.380001 7.591385
```

Question 1 (f)

```r
# shrinkage function for each j
shrink_j <- function(lambda, sigma_j){
  (1/exp(lambda[,2]))^2 / ((1/exp(lambda[,2]))^2 + 1/sigma_j^2)
}

shrink <- rep(0, 8)

for (j in 1:8){
  shrink[j] <- mean(shrink_j(lambda_samples, sigma[j]))
}

data.frame(school = LETTERS[c(1:8)], shrink_size = shrink,
           rank_shrink = rank(shrink),
           rank_sigma = rank(sigma))
```

```
##   school shrink_size rank_shrink rank_sigma
## 1      A   0.8424852         6.0        6.0
## 2      B   0.7464852         2.5        2.5
## 3      C   0.8552654         7.0        7.0
## 4      D   0.7713269         4.5        4.5
## 5      E   0.7176384         1.0        1.0
## 6      F   0.7713269         4.5        4.5
## 7      G   0.7464852         2.5        2.5
## 8      H   0.8767254         8.0        8.0
```

In the above output, the rank is lowest for the lowest shrinkage/sigma values. From largest shrinkage to smallest shrinkage, the eight schools are: H, C, A, D, F, B, G, E.

We observe that shrinkage and sigma values for each school have the same rank. This is consistent with the shrinkage formula $B_j = \tau^{-2}/(t^{-2} + \sigma_j^{-2})$. Since the squared inverse of σ_j is in the denominator, B_j has a positive relationship with σ_j. This also means that the conditional posterior mean for schools with higher standard errors will be shrunk more towards the global mean.

Question 1 (g)

```r
prob <- c()

for (j in 2:8){
  prob[j] <- mean(sum(theta_all[,1] > theta_all[,j])) / nrow(theta_all)
}

data.frame(probability = prob)
```

```
##   probability
## 1          NA
## 2      0.6264
## 3      0.6734
## 4      0.6294
## 5      0.7264
## 6      0.6802
## 7      0.5212
## 8      0.6058
```

In the above output, the probability $P(\theta_1 > \theta_j)$ for $j = 2, \cdots, J$ is presented in the j-th row.

Question 2 (a)

We have:

$$y_{i,j}|\theta_j,\sigma^2 \overset{\text{indep}}{\sim} N(\theta_j,\sigma^2), \qquad \theta_j|\mu,\tau^2 \overset{\text{indep}}{\sim} N(\mu,\tau^2), \qquad \mu \sim N(\mu_0,\sigma_o^2)$$

$$\tau^2 \sim \text{Inv-Gamma}\left(\frac{a_1}{2},\frac{b_1}{2}\right), \quad \sigma^2 \sim \text{Inv-Gamma}\left(\frac{a_2}{2},\frac{b_2}{2}\right)$$

where

$$y_j = (y_{1,j},\cdots,y_{n_j,j}), \quad \bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{i,j}, \qquad \theta = (\theta_1,\cdots,\theta_j), \quad \bar{\theta} = \frac{1}{J}\sum_{j=1}^{J}\theta_j$$

The unnormalized full posterior density is:

$$p(\theta,\mu,\tau^2,\sigma^2|y) \propto p(\mu,\tau^2,\sigma^2)p(\theta|\mu,\tau^2,\sigma^2)p(y|\theta,\mu,\tau^2,\sigma^2)$$

$$\propto p(\mu)p(\tau^2)p(\sigma^2)\left(\prod_{j=1}^{J} p(\theta_j|\mu,\tau^2)\right)\left(\prod_{j=1}^{J}\prod_{i=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2)\right)$$

$$\propto \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}(\tau^2)^{-\frac{a_1}{2}-1}\exp\left\{-\frac{b_1}{2\tau^2}\right\}(\sigma^2)^{-\frac{a_2}{2}-1}\exp\left\{-\frac{b_2}{2\sigma^2}\right\}$$

$$\times \left(\prod_{j=1}^{J}\frac{1}{\tau}\exp\left\{-\frac{(\theta_j-\mu)^2}{2\tau^2}\right\}\right)\times\left(\prod_{j=1}^{J}\prod_{i=1}^{n_j}\frac{1}{\sigma}\exp\left\{-\frac{(y_{i,j}-\theta_j)^2}{2\sigma^2}\right\}\right)$$

$$\propto (\tau^2)^{-\frac{a_1+J}{2}-1}(\sigma^2)^{-\frac{a_2+\sum_{j=1}^{J}n_j}{2}-1}$$

$$\times \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}-\frac{b_1}{2\tau^2}-\frac{b_2}{2\sigma^2}-\frac{\sum_{j=1}^{J}(\theta_j-\mu)^2}{2\tau^2}-\frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2}{2\sigma^2}\right\}$$

$$\propto (\tau^2)^{-\frac{a_1+J}{2}-1}(\sigma^2)^{-\frac{a_2+\sum_{j=1}^{J}n_j}{2}-1}$$

$$\times \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}-\frac{b_1+\sum_{j=1}^{J}(\theta_j-\mu)^2}{2\tau^2}-\frac{b_2+\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2}{2\sigma^2}\right\}$$

Question 2 (b)

$$p(\theta_j|\mu,\tau^2,\sigma^2.y) \propto p(\theta_j|\mu,\tau^2) \prod_{j=1}^{n_j} p(y_{i,j}|\theta_j,\sigma^2)$$

$$\propto \exp\left\{-\frac{(\theta_j-\mu)^2}{2\tau^2}\right\} \prod_{j=1}^{n_j} \exp\left\{-\frac{(y_{i.j}-\theta_j)^2}{2\sigma^2}\right\}$$

$$\propto \exp\left\{-\frac{\theta_j^2-2\mu\theta_j^2}{2\tau^2}-\frac{n_j\theta_j^2-2\theta_j n_j\bar{y}_j}{2\sigma^2}\right\}$$

$$\propto \exp\left\{-\left(\frac{n_j}{2\sigma^2}+\frac{1}{2\tau^2}\right)\theta_j^2+\left(\frac{\mu}{\tau^2}+\frac{n_j\bar{y}_j}{\sigma^2}\right)\theta_j\right\}$$

$$\propto \exp\left\{-\frac{\tau^2 n_j+\sigma^2}{2\sigma^2\tau^2}\theta_j^2+\frac{2(\mu\sigma^2+n_j\bar{y}_j\tau^2)}{2\sigma^2\tau^2}\theta_j\right\}$$

$$\propto \exp\left\{-\frac{\tau^2 n_j+\sigma^2}{2\sigma^2\tau^2}\left(\theta^2-\frac{2(\mu\sigma^2+n_j\bar{y}_j\tau^2)}{\tau^2 n_j+\sigma^2}\right)\right\}$$

$$\propto \exp\left\{-\frac{\left(\theta_j-\frac{\mu\sigma^2+n_j\bar{y}_j\tau^2}{\tau^2 n_j+\sigma^2}\right)^2}{2\left(\frac{\sigma^2\tau^2}{\tau^2 n_j+\sigma^2}\right)}\right\}$$

$$\theta_j|\mu,\tau^2,\sigma^2,y \sim N\left(\frac{n_j\tau^2}{n_j\tau^2+\sigma^2}\bar{y}_j+\frac{\sigma^2}{n_j\tau^2+\sigma^2}\mu,\frac{\sigma^2\tau^2}{\tau^2 n_j+\sigma^2}\right)$$

$$p(\mu|\theta,\tau^2,\sigma^2,y) \propto p(\mu) \prod_{j=1}^{J} p(\theta_j|\mu,\tau^2)$$

$$\propto \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_o^2}\right\} \prod_{j=1}^{J} \exp\left\{-\frac{(\theta_j-\mu)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\frac{\mu^2-2\mu\mu_0}{2\sigma_o^2}\right\} \exp\left\{\frac{2J\mu\bar{\theta}-J\mu^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\left(\frac{J}{2\tau^2}+\frac{1}{2\sigma_0^2}\right)\mu^2+\left(\frac{\mu_0}{\sigma_0^2}+\frac{J\bar{\theta}}{\tau^2}\right)\mu\right\}$$

$$\propto \exp\left\{-\frac{J\sigma_0^2+\tau^2}{2\tau^2\sigma_0^2}\left(\mu^2-\frac{2(\mu_0\tau^2+J\bar{\theta}\sigma_0^2)}{J\sigma_0^2+\tau^2}\mu\right)\right\}$$

$$\propto \exp\left\{-\frac{J\sigma_0^2+\tau^2}{2\tau^2\sigma_0^2}\left(\mu-\frac{\mu_0\tau^2+J\bar{\theta}\sigma_0^2}{J\sigma_0^2+\tau^2}\right)^2\right\}$$

$$\propto \exp\left\{-\frac{\left(\mu-\frac{\mu_0\tau^2+J\bar{\theta}\sigma_0^2}{J\sigma_0^2+\tau^2}\right)^2}{2\left(\frac{\tau^2\sigma_0^2}{J\sigma_0^2+\tau^2}\right)}\right\}$$

$$\mu|\theta,\tau^2,\sigma^2,y \sim N\left(\frac{J\sigma_0^2}{J\sigma_0^2+\tau^2}\bar{\theta}+\frac{\tau^2}{J\sigma_0^2+\tau^2}\mu_0,\frac{\tau^2\sigma_0^2}{J\sigma_0^2+\tau^2}\right)$$

$$p(\tau^2|\theta, \mu, \sigma^2, y) \propto p(\tau^2) \prod_{j=1}^{J} p(\theta_j|\mu, \tau^2)$$

$$\propto (\tau^2)^{-\frac{a_1}{2}-1} \exp\left\{-\frac{b_1}{2\tau^2}\right\} (\tau^2)^{-\frac{J}{2}} \prod_{j=1}^{J} \exp\left\{-\frac{(\theta_j - \mu)^2}{2\tau^2}\right\}$$

$$\propto (\tau^2)^{-\frac{a_1+J}{2}-1} \exp\left\{-\frac{b_1 + \sum_{j=1}^{J}(\theta_j - \mu)^2}{2\tau^2}\right\}$$

$$\tau^2|\theta, \mu, \sigma^2, y \sim \text{Inv-Gamma}\left(\frac{a_1 + J}{2}, \frac{b_1 + \sum_{j=1}^{J}(\theta_j - \mu)^2}{2}\right)$$

$$p(\sigma^2|\theta, \mu, \tau^2, y) \propto p(\sigma^2|\theta, y)$$

$$\propto p(\sigma^2) \prod_{j=1}^{J} \prod_{i=1}^{n_j} p(y_{i,j}|\theta_j, \sigma^2)$$

$$\propto (\sigma^2)^{-\frac{a_2}{2}-1} \exp\left\{-\frac{b_2}{2\sigma^2}\right\} (\sigma^2)^{-\frac{1}{2}\sum_{j=1}^{J} n_j} \exp\left\{-\frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j} - \theta_j)^2}{2\sigma^2}\right\}$$

$$\propto (\sigma^2)^{-\frac{a_2}{2}-\frac{1}{2}\left(\sum_{j=1}^{J} n_j\right)-1} \exp\left\{-\frac{b_2 + \sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j} - \theta_j)^2}{2\sigma^2}\right\}$$

$$\sigma^2|\theta, \mu, \tau^2, y \sim \text{Inv-Gamma}\left(\frac{a_2 + \sum_{j=1}^{J} n_j}{2}, \frac{b_2 + \sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j} - \theta_j)^2}{2}\right)$$

Question 2 (c)

```r
# setwd("~/OneDrive/Y2S2/ST4234/Assignment")
y <- list()
for(i in 1:8) {
  y[[i]] <- as.vector(as.matrix(read.table(paste0("school",i,".txt"))))
}

set.seed(0183140)

mu_0 <- 7
sigma2_0 <- 5
a_1 <- 10
b_1 <- 20
a_2 <- 15
b_2 <- 30

iters <- 10^4

total_obs <- Reduce("+", lapply(y, length))

library(invgamma)
library(coda)

theta_draws <- matrix(0, nrow = iters, ncol = 8)
mu_draws <- rep(0, iters)
sigma2_draws <- rep(0, iters)
tau2_draws <- rep(0, iters)

# initialization
sigma2 <- 1
tau2 <- 1
mu <- 1

# Gibbs sampler
for (i in 1:iters){
  theta <- sapply(1:8, function(j){
    rnorm(1, mean = (mu*sigma2+length(y[[j]])*mean(y[[j]])*tau2)/
            (tau2*length(y[[j]])+sigma2),
          sd = sqrt((sigma2*tau2)/(tau2*length(y[[j]])+sigma2)))})
  theta_draws[i,] <- theta
  sigma2 <- rinvgamma(1, (a_2+total_obs)/2,
                      (b_2+sum(sapply(1:8,
                                      function(x){sum((y[[x]]-theta[x])^2)}))))/2)
  sigma2_draws[i] <- sigma2
  tau2 <- rinvgamma(1, (a_1+8)/2, (b_1+sum((theta-mu)^2))/2)
  tau2_draws[i] <- tau2
  mu <- rnorm(1, mean = (mu_0*tau2+8*mean(theta)*sigma2_0)/(8*sigma2_0+tau2),
              sd = sqrt((tau2*sigma2_0)/(8*sigma2_0+tau2)))
  mu_draws[i] <- mu
}

# discarding the first 5000 draws as burn-in
theta_draws <- theta_draws[5001:iters,]
mu_draws <- mu_draws[5001:iters]
sigma2_draws <- sigma2_draws[5001:iters]
tau2_draws <- tau2_draws[5001:iters]
```
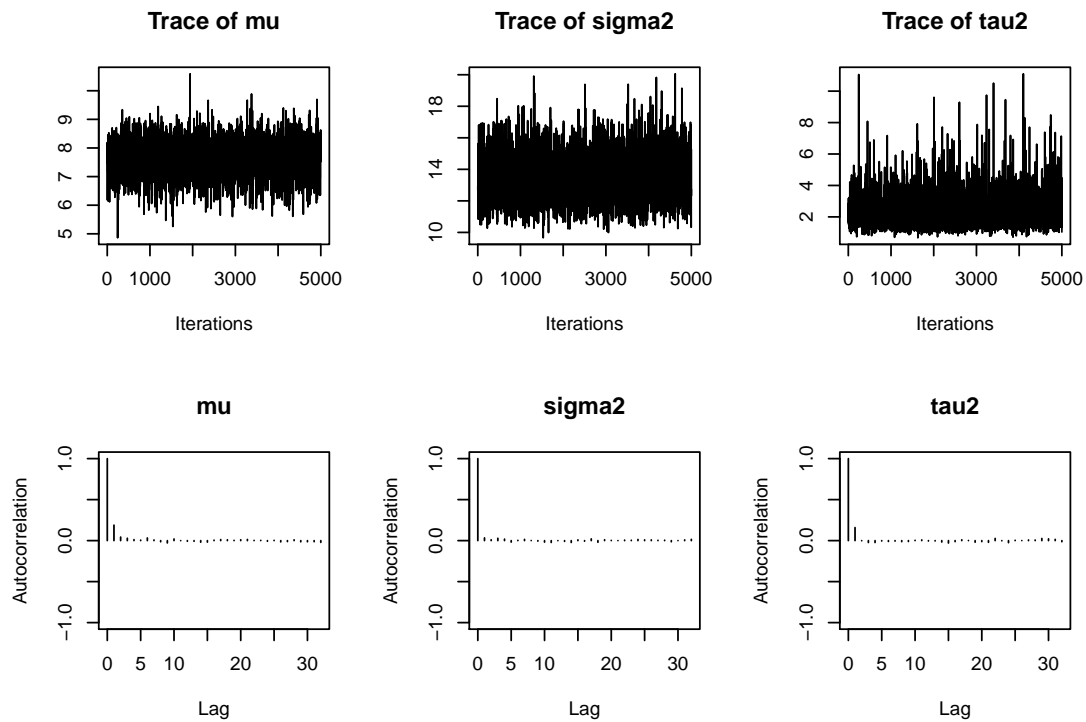
```
mcmcobj2 <- mcmc(cbind(mu_draws, sigma2_draws, tau2_draws))
colnames(mcmcobj2) <- c("mu", "sigma2", "tau2")

par(mfrow = c(2,3))
traceplot(mcmcobj2)
autocorr.plot(mcmcobj2, auto.layout = FALSE)
```


The traceplots of μ, σ^2 and τ^2 resemble random noise, showing rapid flunctuation. This suggests that each of μ, σ^2 and τ^2 do not have high serial correlation/dependence and mix well. Their autocorrelation plots show a very fast decay as a function of lag, indicating very low serial dependence.

Question 2 (d)

The posterior means for μ, σ^2 and τ^2 are:

```r
mean(mu_draws)
```

```
## [1] 7.607403
```

```r
mean(sigma2_draws)
```

```
## [1] 13.50969
```

```r
mean(tau2_draws)
```

```
## [1] 2.378122
```

The 95% HPD intervals are:

```r
HPDinterval(mcmcobj2)
```

```
##             lower     upper
## mu       6.4407684  8.756844
## sigma2  10.7481940 16.168939
## tau2     0.9580103  4.318181
## attr(,"Probability")
## [1] 0.95
```

```r
mu_grid <- seq(3, 11, by = 0.1)
plot(x = mu_grid, y = dnorm(mu_grid, mean = mu_0, sd = sqrt(sigma2_0)),
     type = "l", lwd = 2, xlab = expression(mu),
     ylab = "Density", col = "blue",
     ylim = c(0, 0.72))
points(density(mu_draws), type = "l", col = "red", lwd = 2)
legend("topleft", legend =
         c(expression(paste("p(",mu,")" %prop% N, "(",mu[0],",",sigma[0]^2,")")),
           expression(paste("marginal posterior of ", mu))),
       col=c("blue", "red"), lty = 1, lwd = 2, cex = 0.8)
```
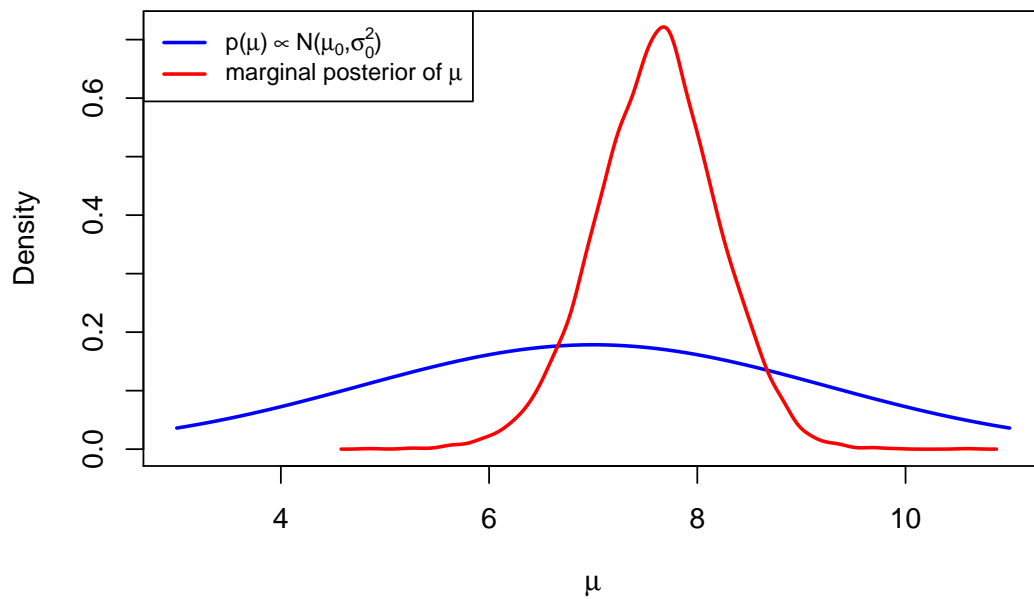


```r
data.frame(prior_mean = mu_0,
           posterior_mean = mean(mu_draws),
           sample_mean = sum(sapply(y, sum))/total_obs)
```

```
##   prior_mean posterior_mean sample_mean
## 1          7       7.607403    7.691278
```

After observing the data, the estimated global mean (7.6074033) from the marginal posterior of μ is slightly higher than what we expect from prior assumption of $\mu_0 = 7$. The marginal posterior is more peaked than the prior as it combines information from the data and the prior, hence containing more information from the data about the global mean that the prior. This posterior mean also lies between prior mean (7) and the sample mean (7.6458903), although it is closer to the latter, indicating that the prior is not very informative.

```
sigma2_grid <- seq(1, 20, by = 0.1)
plot(x = sigma2_grid, y = dinvgamma(sigma2_grid, a_2/2, b_2/2),
     type = "l", lwd = 2, xlab = expression(sigma^2),
     ylab = "Density", col = "blue", xlim = c(1, 20), ylim = c(0, 0.6))
points(density(sigma2_draws), type = "l", col = "red", lwd = 2)
legend("topright", legend =
        c(expression(paste("p(",sigma^2,")" %prop% "Inv-Gamma(",
                            a[2]/2,",",b[2]/2,")")),
            expression(paste("marginal posterior of ", sigma^2))),
        col=c("blue", "red"), lty = 1, lwd = 2, cex = 0.8)
```
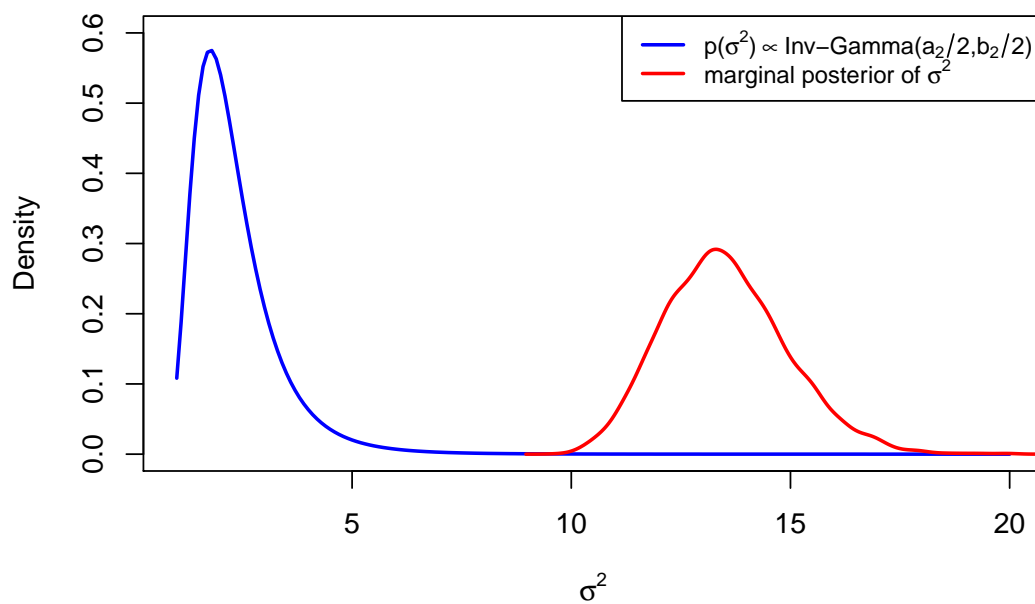


```
data.frame(prior_mean = b_2/(a_2-1),
           posterior_mean = mean(sigma2_draws),
           sample_mean = mean(sapply(y, var)))
```

```
##   prior_mean posterior_mean sample_mean
## 1   2.142857       13.50969    14.20669
```
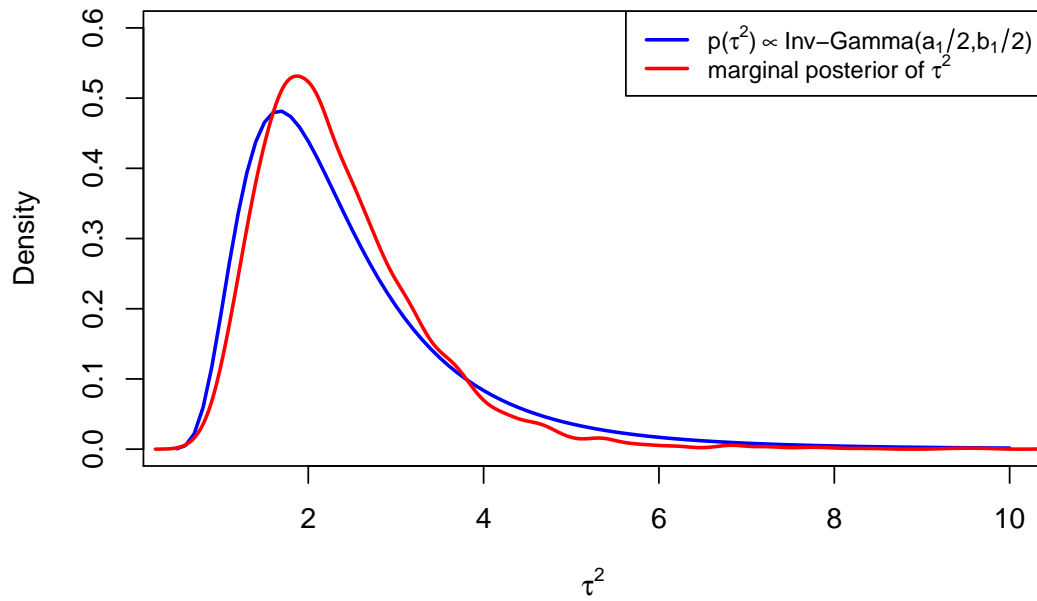
The data shows that the estimated within-group variance (13.5096919) is much larger than what we expect from our prior assumption (2.142857), suggesting that we have chosen a prior that is not informative / unsuitable for our model. The posterior mean of the within-group variance is between that of the prior and the sample, but very much closer to the latter. This suggests that we have dervied most of this information from the data, and should have chosen prior arguments that are more reflective of our samples.

```
tau2_grid <- seq(0.5, 10, by = 0.1)
plot(x = tau2_grid, y = dinvgamma(tau2_grid, a_1/2, b_1/2),
     type = "l", lwd = 2, xlab = expression(tau^2),
     ylab = "Density", col = "blue", ylim = c(0, 0.6))
points(density(tau2_draws), type = "l", col = "red", lwd = 2)
legend("topright", legend =
        c(expression(paste("p(",tau^2,")" %prop% "Inv-Gamma(",
                           a[1]/2,",",b[1]/2,")")),
          expression(paste("marginal posterior of ", tau^2))),
       col=c("blue", "red"), lty = 1, lwd = 2, cex = 0.8)
```


The data shows that the estimated between-group variance (2.3781218) is slightly larger than what we expect from our prior assumptions (mean = 2.222222). From the above plot, both densities are quite similar in shape, hence the prior is informative and is consistent with our data. Since both within-group and between-group variance are larger, we can say that total variation in the data is indeed larger than what we have expected it to be, based on the arguments in our priors.

Question 2 (e)

```r
tau2_prior <- rinvgamma(5000, a_1/2, b_1/2)
sigma2_prior <- rinvgamma(5000, a_2/2, b_2/2)
r_prior <- tau2_prior / (tau2_prior+sigma2_prior)

r_post <- tau2_draws /
  (tau2_draws + sigma2_draws)
plot(density(r_prior), type = "l", col = "blue", lwd = 2,
     main = "", ylim = c(0, 10))
points(density(r_post), type = "l", col = "red", lwd = 2)
legend("topright", legend = c("prior density of R", "posterior density of R"),
       col=c("blue", "red"), lty = 1, lwd = 2, cex = 0.8)
```


N = 5000   Bandwidth = 0.02288
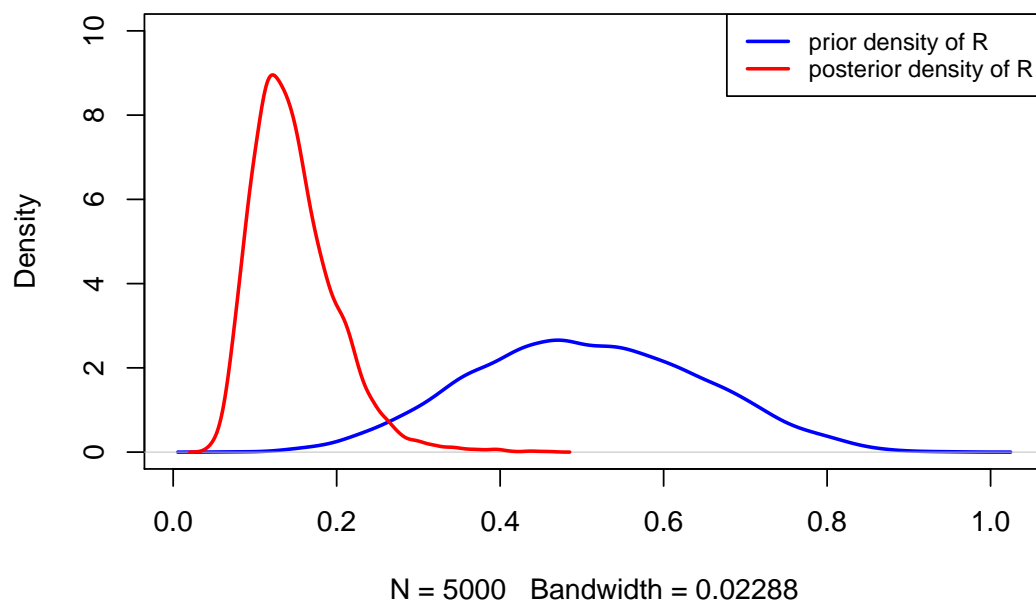
```r
data.frame(prior_mean = mean(r_prior),
           posterior_mean = mean(r_post))
```

```
##   prior_mean posterior_mean
## 1  0.5054923      0.1478072
```

$R = \frac{\tau^2}{\tau^2+\sigma^2}$ is the proportion of between-school variance to total variance; higher R values means the variation in the data can be explained to a greater extent by variation between schools. From Monte Carlo approximation based on assumed prior arguments, it is expected that between-school variation is attributed to about 50.4% of total variation (both within-school and between-school variation are approximately equal), but after having observed the data via Gibbs sampler, there is evidence to show that this value is only about 14.8%. As the peak of the posterior is much higher than the prior (which has a larger variance), this suggests that the prior is not very informative compared to the data.

Question 2 (f)

```r
sample_means <- sapply(y, mean)

post_means <- apply(theta_draws, 2, mean)

plot(x = post_means, y = sample_means, pch = 20,
     xlab = "Posterior Expectations", ylab = "Sample Averages")
mod <- lm(sample_means ~ post_means)
abline(mod, col = "blue")
```
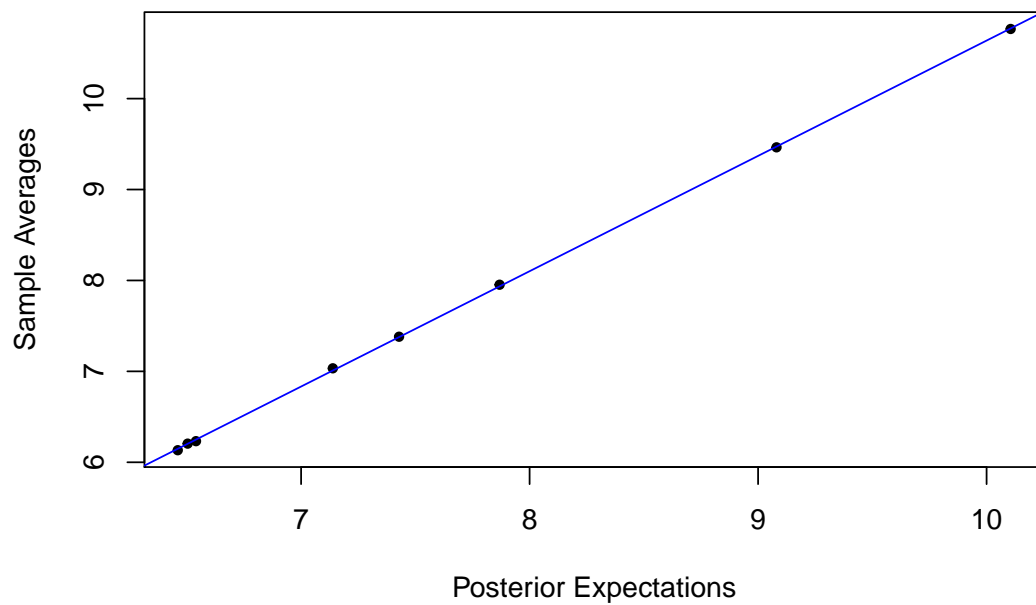

Posterior expectations of $\theta_1, \cdots, \theta_j$ have a positive linear relationship with sample averages $\bar{y}_1, \cdots, \bar{y}_j$.