Introduction to Bayesian Data Analysis
Tutorial 4 Solutions

(1) For each of the 3 school data sets, we will consider the following model:

$$\begin{aligned} y_i, \dots, y_n | \theta, \sigma^2 &\sim N(\theta, \sigma^2) \\ \theta | \sigma^2 &\sim N(\mu_0 = 5, \sigma^2/\kappa_0 = \sigma^2) \\ 1/\sigma^2 &\sim \text{gamma}(\nu_0/2 = 2/2 = 1, \sigma_0 \nu_0/2 = (4 \times 2)/2 = 4) \end{aligned}$$

Thus the joint posterior distribution can be written as:

$$p(\theta, \sigma^2 | \mathbf{y}) = p(\theta | \mathbf{y}, \sigma^2) p(\sigma^2 | \mathbf{y}).$$

Where :

$$\begin{aligned} p(\sigma^2 | \boldsymbol{y}) &= \text{inv} - \text{gamma}(\nu_n/2, \nu_n \sigma_n^2/2) \\ p(\theta | \boldsymbol{y}, \sigma^2) &= N(\mu_n, \sigma^2/\kappa_n) \\ \nu_n &= \nu_0 + n \\ \kappa_n &= \kappa_0 + n \\ \sigma_n^2 &= \frac{1}{\nu_n}[\nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n}(\bar{y} - \mu_0)^2] \\ \mu_n &= (\kappa_0 \mu_0 + n\bar{y})/\kappa_n \end{aligned}$$

The joint posterior can be explored through Monte-Carlo sampling. First create S independent samples of σ^2 by drawing from the conditional posterior distribution $p(\sigma^{2(s)}|\boldsymbol{y})$ for $s = 1, \dots, S$, and then create S independent samples of θ by drawing from the conditional posterior distribution $p(\theta|\boldsymbol{y}, \sigma^{2(s)})$.

(a)
```
> y1 <- read.table("school1.dat", header = F)
> y2 <- read.table("school2.dat", header = F)
> y3 <- read.table("school3.dat", header = F)
>  mu0 <- 5
>  k0 <- 1
>  s20 <- 4
>  nu0 <- 2
>  n1 <- length(y1)
>  ybar1 <- mean(y1)
>  s21 <- var(y1)
>  n2 <- length(y2)
>  ybar2 <- mean(y2)
>  s22 <- var(y2)
```

```
>  n3 <- length(y3)
>  ybar3 <- mean(y3)
>  s23 <- var(y3)
>
>  #School 1 Quantities
>   kn1 <- k0 + n1
>  nun1 <- nu0 + n1
>  mun1 <- (k0 * mu0 + n1 * ybar1)/kn1
>  s2n1 <- (nu0 * s20 + (n1 - 1) * s21 + k0 * n1 *
            (ybar1 -  mu0)^2/(kn1))/(nun1)

>  #School 2 Quantities
>   kn2 <- k0 + n2
>  nun2 <- nu0 + n2
>  mun2 <- (k0 * mu0 + n2 * ybar2)/kn2
> s2n2 <- (nu0 * s20 + (n2 - 1) * s22 + k0 * n2 *
           (ybar2 - mu0)^2/(kn2))/(nun2)
>
> #School 3 Quantities
>  kn3 <- k0 + n3
>  nun3 <- nu0 + n3
>  mun3 <- (k0 * mu0 + n3 * ybar3)/kn3
>  s2n3 <- (nu0 * s20 + (n3 - 1) * s23 + k0 * n3 *
        (ybar3 -  mu0)^2/(kn3))/(nun3)
>
> #a)
>  S <- 10000
>  #School 1 Monte Carlo Sampling
>
> s2.postsample1 <- 1/rgamma(S, (nu0 + n1)/2, s2n1 * (nu0 + n1)/2)
> theta.postsample1 <- rnorm(S, mun1, sqrt(s2.postsample1/(k0 + n1)))
>
>  #School 2 Monte Carlo Sampling
>  s2.postsample2 <- 1/rgamma(S, (nu0 + n2)/2, s2n2 * (nu0 + n2)/2)
>  theta.postsample2 <- rnorm(S, mun2, sqrt(s2.postsample2/(k0 + n2)))
>
>  #School 3 Monte Carlo Sampling
>   s2.postsample3 <- 1/rgamma(S, (nu0 + n3)/2, s2n1 * (nu0 + n1)/2)
> theta.postsample3 <- rnorm(S, mun3, sqrt(s2.postsample3/(k0 + n3)))
```

```
> #School 1 posterior summaries
> mean(theta.postsample1)
[1] 9.281013
> quantile(theta.postsample1, c(0.025, 0.975))
     2.5%     97.5%
 7.743142 10.795338
> mean(sqrt(s2.postsample1))
[1] 3.908459
> quantile(sqrt(s2.postsample1), c(0.025, 0.975))
    2.5%    97.5%
3.007581 5.212050
> #School 2 posterior summaries
> mean(theta.postsample2)
[1] 6.948754
> quantile(theta.postsample2, c(0.025, 0.975))
    2.5%    97.5%
5.147193 8.779580
> mean(sqrt(s2.postsample2))
[1] 4.407719
> quantile(sqrt(s2.postsample2), c(0.025, 0.975))
    2.5%    97.5%
3.360833 5.915965
> #School 3 posterior summaries
> mean(theta.postsample3)
[1] 7.812453
> quantile(theta.postsample3, c(0.025, 0.975))
    2.5%    97.5%
5.884712 9.725384
> mean(sqrt(s2.postsample3))
[1] 4.351671
> quantile(sqrt(s2.postsample3), c(0.025, 0.975))
    2.5%    97.5%
3.234402 5.944782
```

Based on the posterior summaries, the mean estimate is lowest for School 2, and highest for School 1. School 1 has the lowest posterior mean estimate for σ^2. Interval widths are largest for School 3 which makes sense given the smaller sample size of School 3.

(b) We want $P(\theta_i < \theta_j < \theta_k|\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$

1. $i = 1, j = 2, k = 3$

```
> mean(theta.postsample1 < theta.postsample2 &
               theta.postsample2 < theta.postsample3)
[1] 0.0071
```

2. $i = 1, j = 3, k = 2$

```
> mean(theta.postsample1 < theta.postsample3 &
               theta.postsample3 < theta.postsample2)
[1] 0.0029
```

3. $i = 2, j = 1, k = 3$

```
> mean(theta.postsample2 < theta.postsample1 &
                  theta.postsample1< theta.postsample3)
[1] 0.1075
```

4. $i = 2, j = 3, k = 1$

```
> mean(theta.postsample2 < theta.postsample3 &
               theta.postsample3< theta.postsample1)
[1] 0.6335
```

5. $i = 3, j = 1, k = 2$

```
> mean(theta.postsample3 < theta.postsample1 &
                 theta.postsample1< theta.postsample2)
[1] 0.0157
```

6. $i = 3, j = 2, k = 1$

```
> mean(theta.postsample3 < theta.postsample2 &
                  theta.postsample2< theta.postsample1)
[1] 0.2333
```

The highest posterior probability is for the ordering $\theta_2 < \theta_3 < \theta_1$ which is consistent with the results in part (a).

(c) We want $P(\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k | \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$

```
y1.pred<-rnorm(S,theta.postsample1,sqrt(s2.postsample1))
y2.pred<-rnorm(S,theta.postsample2,sqrt(s2.postsample2))
y3.pred<-rnorm(S,theta.postsample3,sqrt(s2.postsample3))
```

1. $i = 1, j = 2, k = 3$
   ```
   > mean(y1.pred < y2.pred & y2.pred < y3.pred)
   [1] 0.1079
   ```
2. $i = 1, j = 3, k = 2$
   ```
   > mean(y1.pred < y3.pred & y3.pred < y2.pred)
   [1] 0.091
   ```
3. $i = 2, j = 1, k = 3$
   ```
   > mean(y2.pred < y1.pred & y1.pred < y3.pred)
   [1] 0.1973
   ```
4. $i = 2, j = 3, k = 1$
   ```
   > mean(y2.pred < y3.pred & y3.pred < y1.pred)
   [1] 0.247
   ```
5. $i = 3, j = 1, k = 2$
   ```
   > mean(y3.pred < y1.pred & y1.pred < y2.pred)
   [1] 0.1449
   ```
6. $i = 3, j = 2, k = 1$
   ```
   > mean(y3.pred < y2.pred & y2.pred < y1.pred)
   [1] 0.2119
   ```

   The highest posterior predictive probability is for the ordering $\tilde{Y}_2 < \tilde{Y}_3 < \tilde{Y}_1$, but the posterior predictive probability is not that much greater than for the ordering $\tilde{Y}_3 < \tilde{Y}_2 < \tilde{Y}_1$. In posterior predictions, we are also allowing for predictive variance σ^2 from the sampling model in addition to the uncertainty in the population mean θ (given the observed data set), so we do not necessarily expect the ordering of the schools to remain the same or posterior predictive probabilities to be as large as their posterior probability counterparts.

(d) We want $P(\theta_1 > \max(\theta_2, \theta_3)|\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$, and $P(\tilde{Y}_1 > \max(\tilde{Y}_2, \tilde{Y}_3)|\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$

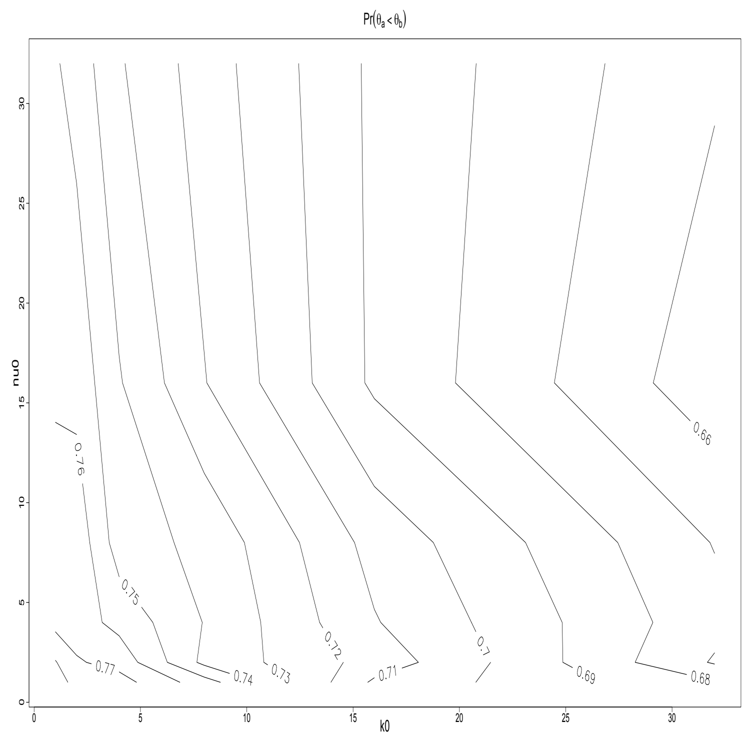
```
> mean(theta.postsample1 > theta.postsample2 & theta.postsample1 > theta.postsampl
[1] 0.8747
> mean(y1.pred > y2.pred & y1.pred > y3.pred)
[1] 0.4576
> #or
> theta.pr <- rep(0, S)
> Y.pr <- rep(0, S)
> for (i in 1:S) {
+      theta.max.i <- max(theta.postsample2[i], theta.postsample3[i])
+      theta.pr[i] <- 1 * (theta.postsample1[i] > theta.max.i)
+      Y.max.i <- max(y2.pred[i], y3.pred[i])
+      Y.pr[i] <- 1 * (y1.pred[i] > Y.max.i)
+  }
> mean(theta.pr)
[1] 0.8668
> mean(Y.pr)
[1] 0.4589
```

There is close to 90% probability that the average amount of time spent studying homework is highest for students from School 1. If we were to sample a new student from each school and record the amount of time each student spends studying homework during exam period, with approximately 46% probability, we predict given the data that the student from School 1 will spend the most amount of time.

(2) The idea is that we want to check the sensitivity in relationship between the two population means θ_A and θ_B to the prior "sample sizes" k_0 and ν_0.

```
>  mu0 <- 75
>  k0 <- c(1,2,4,8,16,32)
>  nu0<-c(1,2,4,8,16,32)
>  s20 <- 100
>
>  n1 <- 16
> ybar1<-75.2
> s21<-7.3^2
>  n2 <- 16
>  ybar2 <-77.5
>  s22 <- 8.1^2
>
>  S<-10000
>  result<-matrix(0,6,6)
>  for (i in 1:6){
+     for ( j in 1:6){
+  #Group A
+  kn1 <- k0[i] + n1
+  nun1 <- nu0[j] + n1
+  mun1 <- (k0[i] * mu0 + n1 * ybar1)/kn1
+  s2n1 <- (nu0[j] * s20 + (n1 - 1) * s21 + k0[i] * n1 *
       (ybar1 -  mu0)^2/(kn1))/(nun1)
+  s2.postsample1 <- 1/rgamma(S, (nu0[j] + n1)/2, s2n1 * (nu0[j] + n1)/2)
+ theta.postsample1 <- rnorm(S, mun1, sqrt(s2.postsample1/(k0[i] + n1)))
+
+  #Group B
+    kn2 <- k0[i] + n2
+  nun2 <- nu0[j] + n2
+  mun2 <- (k0[i] * mu0 + n2 * ybar2)/kn2
+ s2n2 <- (nu0[j] * s20 + (n2 - 1) * s22 + k0[i] * n2 *
       (ybar2 - mu0)^2/(kn2))/(nun2)
+  s2.postsample2 <- 1/rgamma(S, (nu0[j] + n2)/2, s2n2 * (nu0[j] + n2)/2)
+  theta.postsample2 <- rnorm(S, mun2, sqrt(s2.postsample2/(k0[i] + n2)))
+
+  result[i,j]<-mean(theta.postsample1<theta.postsample2)
+  }
+ }
```

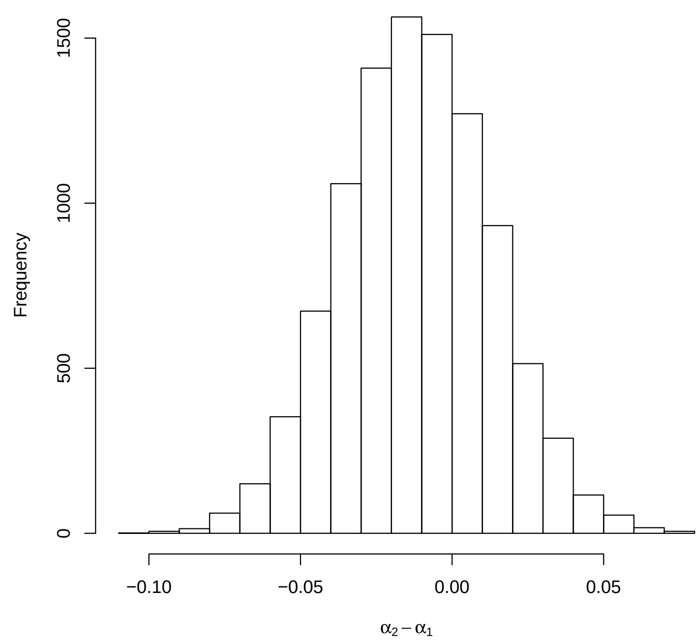
```
> result
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
[1,] 0.7851 0.7807 0.7667 0.7658 0.7581 0.7511
[2,] 0.7764 0.7711 0.7648 0.7667 0.7568 0.7459
[3,] 0.7740 0.7663 0.7568 0.7449 0.7408 0.7311
[4,] 0.7544 0.7375 0.7396 0.7373 0.7205 0.7151
[5,] 0.7082 0.7161 0.7107 0.7064 0.6882 0.6879
[6,] 0.6807 0.6690 0.6732 0.6695 0.6538 0.6615
> pdf("HW_Fig2.pdf")
> contour(k0, nu0, result, main = "Prob thetaA < thetaB",
         xlab = "k0", ylab = "nu0")
> dev.off()
```



The posterior probability $Pr(\theta_A < \theta_B|\mathbf{y}_A, \mathbf{y}_B)$ is more sensitive to κ_0 (the prior sample size for the prior mean), than ν_0 (the prior sample size for the prior variance). We see the posterior evidence for $\theta_A < \theta_B$ decreases as the amount of prior evidence for the sample mean increases, but still remains above 0.5 (which corresponds to the belief that A and B are equally effective methods). That is, the data provide evidence that Study A is more effective to people of a variety of prior opinions.

(3) Let $\boldsymbol{\theta}_j$ be the vector of parameters $(\theta_{j1}, \theta_{j2}, \theta_{j3})$ representing the proportion of Candidate X supporters, Candidate Y supporters, and those with no opinion, at the time of survey j. As a relatively non-informative prior distribution, we use $\boldsymbol{\theta}_j \sim \text{Dirichlet}(1, 1, 1)$ as our prior to obtain the posterior distribution after the pre-debate survey data are collected, that is, $\boldsymbol{\theta}_1|\text{data}_1 \sim \text{Dirichlet}(1 + 294, 1 + 307, 1 + 38)$. The distribution $\boldsymbol{\theta}_1|\text{data}_1$ can form our prior to derive the posterior distribution given the post-survey debate data, that is $\boldsymbol{\theta}_2|\text{data}_1, \text{data}_2 \sim \text{Dirichlet}(1 + 294 + 288, 1 + 307 + 332, 1 + 38 + 19)$, where 'data$_1$' refers to the survey counts pre-debate and 'data$_2$' refers to the survey counts post-debate.

If α_j is the proportion of voters who preferred Candidate X, out of those who had a preference for Candidate X or Y at the time of survey j, then $\alpha_j = \frac{\theta_{j1}}{\theta_{j1}+\theta_{j2}}$. We can simulate draws from the posterior distribution of $\alpha_2 - \alpha_1$ by drawing S values (where S is large, say S=10000) from the posterior distributions of θ_1 and θ_2 and directly computing $\alpha_2 - \alpha_1$ from each draw.



We can approximate the posterior probability that there was a shift towards Candidate X by counting the proportion of the simulated values for which $\alpha_2 - \alpha_1 > 0$. This is 0.32.

```
> mean(alpha2>alpha1)
[1] 0.3083
```