Introduction to Bayesian Data Analysis

Tutorial 5 - Solutions

(1) We will use posterior predictive checking to evaluate the appropriateness of the normal sampling model.

Recall $\theta|y_1,...,y_n,\sigma^2 \sim \text{Norm}(\mu_n, \sigma^2/\kappa_n)$ and $\sigma^2|y_1,...,y_n \sim \text{Inv}-\text{Gamma}(\nu_n/2, \nu_n\sigma_n^2/2)$ where $\kappa_n = \kappa_0 + n = 1 + 45 = 46$; $\mu_n = \frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_n} = \frac{2+45\times\bar{y}}{46}$; $\nu_0 = \nu_0 + n = 46$; $\sigma_n^2 = \frac{1}{\nu_n}[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n}(\bar{y}-\mu_0)^2] = \frac{1}{46}[1 + 44s^2 + \frac{45}{46}(\bar{y}-2)^2]$.

For s=1,...,10000

sample $\theta^{(s)}, (\sigma^2)^{(s)} \sim p(\theta, \sigma^2|y_1,...,y_n)$, sample $\tilde{y}^{(s)} \sim p(\tilde{y}|\theta,\sigma^2)$. That is, we create 10000 posterior predictive samples.

We want to check different aspects of the normal sampling model, and so will use the following test quantities: sample mean; sample variance; sample skewness; 90% and 97.5% quantiles , the probability that a women has more than 3 children (note: this list is not exhaustive, you can consider other test quantities, but bear in mind to test tail area probabilities and asymmetry of the predictive data distribution, not just location and spread)

```
k0<-1; mu0<-2 ; nu0<-1 ; s20<-1

n<-45
set.seed(1)
y<-sample(CHILDS,n)
ybar<-mean(y) ; s2<-var(y)

kn<-k0+n ; nun<-nu0+n
mun<- (k0*mu0 + n*ybar)/kn
s2n<- (nu0*s20 +(n-1)*s2 +k0*n*(ybar-mu0)^2/(kn))/(nun)
```

```
###

S<-10000
s2.postsample <- 1/rgamma(S, (nu0+n)/2, s2n*(nu0+n)/2 )
theta.postsample <- rnorm(S, mun, sqrt(s2.postsample/(k0+n)))


#posterior predictive draws
y.pred.stat<-NULL
for (s in 1:10000){
y.pred<-rnorm(n,theta.postsample[s], sqrt(s2.postsample[s]))
y.pred.stat<-rbind(y.pred.stat,c(mean(y.pred),var(y.pred),skewness(y.pred),
           quantile(y.pred,c(0.9,0.975)),mean(y.pred>3)))
}

#posterior predictive probabilities

> mean(y.pred.stat[,1]>ybar)
[1] 0.4857
> mean(y.pred.stat[,2]>s2)
[1] 0.4439
> mean(y.pred.stat[,3]>skewness(y))
[1] 0.0044
> mean(y.pred.stat[,4]>quantile(y,0.90))
[1] 0.4608
> mean(y.pred.stat[,5]>quantile(y,0.975))
[1] 0.0217
> mean(y.pred.stat[,6]>mean(y>3))
[1] 0.962
> 1-mean(y.pred.stat[,6]>mean(y>3))
[1] 0.038
```

Based on the posterior predictive checks, the normal sampling model is reasonable for inference on the population mean, and population variance. But the normal sampling model is not appropriate to capture the skewness or upper tail area probabilities of the data. The normal-model inference will under estimate the number of people in the right tail of the distribution, and so will provide poor estimates of the percentage of people with large numbers of children.

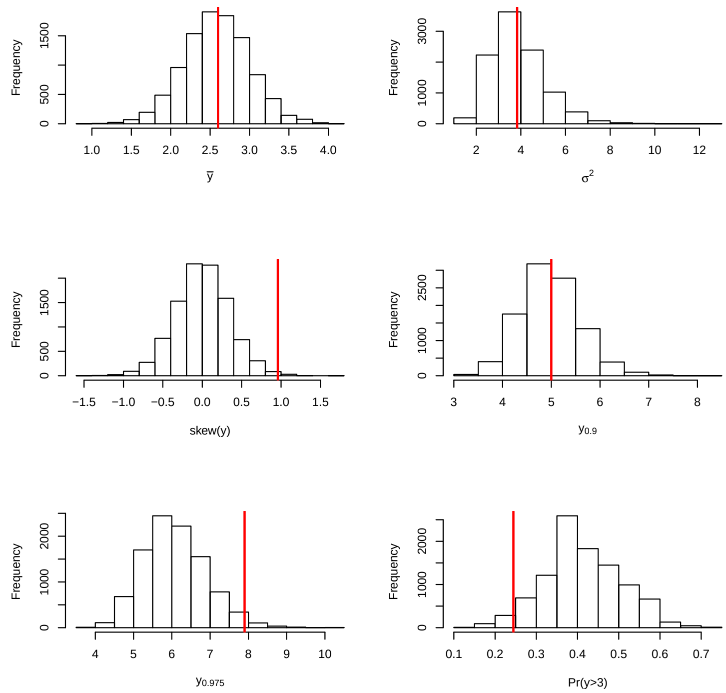


Figure 1: Realized vs posterior predictive distributions for various test quantities

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

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

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 :

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

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 $s = 1, \ldots, S$, 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] 7.256887
> quantile(theta.postsample1, c(0.025, 0.975))
     2.5%     97.5%
 1.830615 12.583310
> mean(sqrt(s2.postsample1))
[1] 3.379885
```

```
> quantile(sqrt(s2.postsample1), c(0.025, 0.975))
    2.5%    97.5%
1.389650 9.191462
>
> #School 2 posterior summaries
> mean(theta.postsample2)
[1] 6.048284
> quantile(theta.postsample2, c(0.025, 0.975))
     2.5%     97.5%
 2.034965 10.224398
> mean(sqrt(s2.postsample2))
[1] 2.538676
> quantile(sqrt(s2.postsample2), c(0.025, 0.975))
    2.5%    97.5%
1.025610 6.764966
>
> #School 3 posterior summaries
> mean(theta.postsample3)
[1] 6.50059
> quantile(theta.postsample3, c(0.025, 0.975))
     2.5%     97.5%
 1.072242 12.214839
> mean(sqrt(s2.postsample3))
[1] 3.378639
> quantile(sqrt(s2.postsample3), c(0.025, 0.975))
    2.5%    97.5%
1.390883 9.177575
```

Based on the posterior summaries, the mean and standard deviation estimates are lowest for School 2. The estimates for Schools 1 and 3 are approximately the same.

(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] 0.1325
```

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

```
> mean(theta.postsample1 < theta.postsample3 &
               theta.postsample3 < theta.postsample2)
[1] 0.0852
```

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

```
> mean(theta.postsample2 < theta.postsample1 &
                  theta.postsample1< theta.postsample3)
[1] 0.1855
```

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

```
> mean(theta.postsample2 < theta.postsample3 &
               theta.postsample3< theta.postsample1)
[1] 0.2498
```

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

```
> mean(theta.postsample3 < theta.postsample1 &
                 theta.postsample1< theta.postsample2)
[1] 0.11
```

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

```
> mean(theta.postsample3 < theta.postsample2 &
                  theta.postsample2< theta.postsample1)
[1] 0.237
```

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] 0.1585
   ```

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

   ```
   > mean(y1.pred < y3.pred & y3.pred < y2.pred)
   [1] 0.1114
   ```

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

   ```
   > mean(y2.pred < y1.pred & y1.pred < y3.pred)
   [1] 0.1716
   ```

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

   ```
   > mean(y2.pred < y3.pred & y3.pred < y1.pred)
   [1] 0.2042
   ```

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

   ```
   > mean(y3.pred < y1.pred & y1.pred < y2.pred)
   [1] 0.1336
   ```

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

   ```
   > mean(y3.pred < y2.pred & y2.pred < y1.pred)
   [1] 0.2207
   ```

   The highest posterior predictive probability is for the ordering $\tilde{Y}_3 < \tilde{Y}_2 < \tilde{Y}_1$. This is different from the ordering of the population parameters with the highest posterior probability in part (b). 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.

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

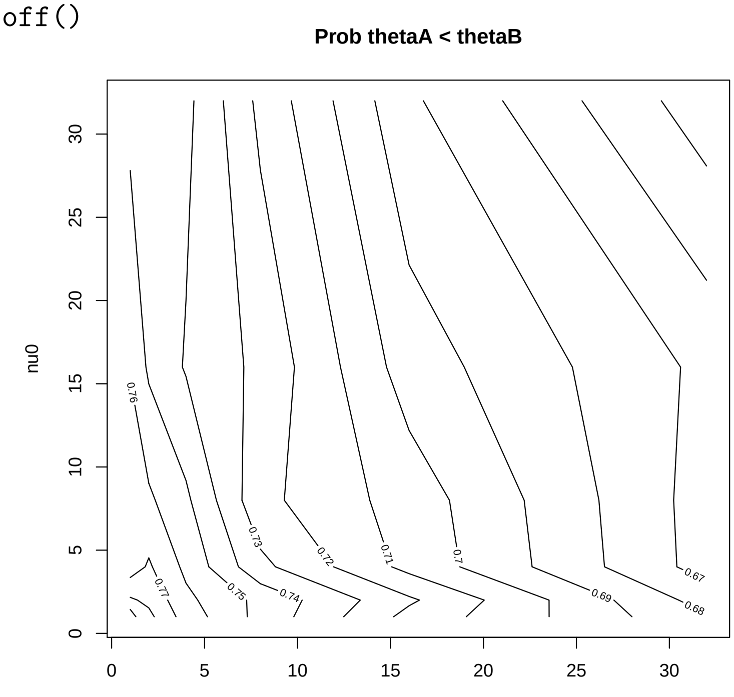
```
> mean(theta.postsample1 > theta.postsample2 &
             theta.postsample1 > theta.postsample3)
[1] 0.4868
> mean(y1.pred > y2.pred & y1.pred > y3.pred)
[1] 0.4249
> #or
> theta.pr <- rep(0, S)
>  Y.pr <- rep(0, S)
>  for (i in 1:S) {
+      theta.max.i <- max(theta.postsample2[i], theta.postsample3[i])
+      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] 0.4868
> mean(Y.pr)
[1] 0.4249
```

There is close to 50% 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 just over 40% probability, we predict given the data that the student from School 1 will spend the most amount of time.

(3) 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)
```

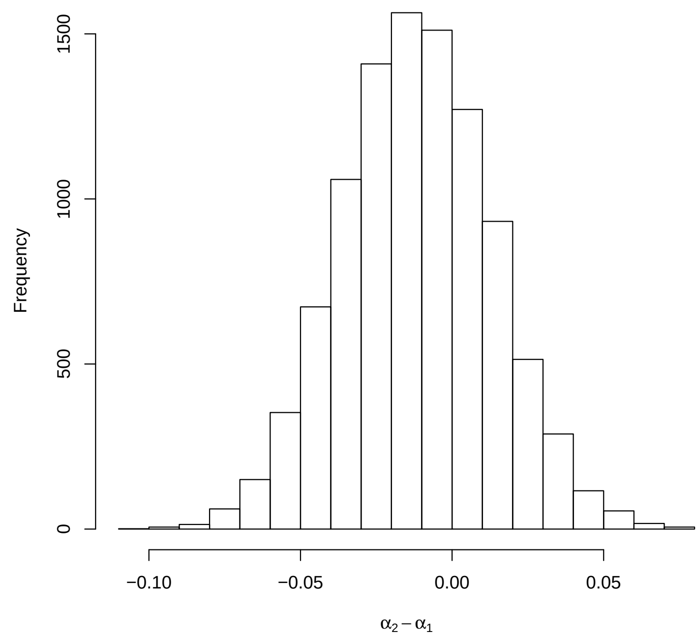
```
+  }
+ }
> result
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
[1,] 0.7789 0.7814 0.7646 0.7669 0.7593 0.7467
[2,] 0.7825 0.7778 0.7713 0.7617 0.7483 0.7464
[3,] 0.7654 0.7624 0.7577 0.7519 0.7391 0.7427
[4,] 0.7467 0.7472 0.7326 0.7228 0.7274 0.7174
[5,] 0.7169 0.7216 0.7070 0.7054 0.6951 0.6818
[6,] 0.6810 0.6757 0.6659 0.6656 0.6676 0.6443
> 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.

(4) 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 '$\text{data}_1$' 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] 0.3199
```