Introduction to Bayesian Data Analysis
Tutorial 7

(1) Problem 8.1 (Hoff) Components of variance: Consider the hierarchical model where

$$\theta_1, ..., \theta_m | \mu, \tau^2 \stackrel{\text{iid}}{\sim} \text{normal}(\mu, \tau^2)$$

$$y_{1,j}, ..., y_{n_j,j} | \theta_j, \sigma^2 \stackrel{\text{iid}}{\sim} \text{normal}(\theta_j, \sigma^2)$$

For this problem, we will eventually compute the following:

$$Var[y_{i,j}|\theta_j, \sigma^2], Var[\bar{y}_{\cdot,j}|\theta_j, \sigma^2], Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$$

$$Var[y_{i,j}|\mu, \tau^2], Var[\bar{y}_{\cdot,j}|\mu, \tau^2], Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$$

First let's use our intuition to guess at the answers:

(a) Which do you think is bigger, $Var[y_{i,j}|\theta_j, \sigma^2]$ or $Var[y_{i,j}|\mu, \tau^2]$? To guide your intuition, you can interpret the first as the variability of the Y's when sampling from a fixed group, and the second as the variability in first sampling a group, then sampling a unit from within the group.

(b) Do you think $Cov[y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2]$ is negative, positive, or zero? Answer the same for $Cov[y_{i_1,j}, y_{i_2,j}|\mu, \tau^2]$. You may want to think about what $y_{i_2,j}$ tells you about $y_{i_1,j}$ if θ_j is known, and what it tells you when θ_j is unknown.

(c) Now compute each of the six quantities above and compare to your answers in a) and b).

(d) Now assume we have a prior $p(\mu)$ for μ. Using Bayes' rule show that

$$p(\mu|\theta_1, ..., \theta_m, \sigma^2, \tau^2, \mathbf{y}_1, ..., \mathbf{y}_m) = p(\mu|\theta_1, ..., \theta_m, \tau^2)$$

Interpret in words what this means.

(2) Problem 8.2 (Hoff) Sensitivity analysis: In this exercise we will revisit the study from Exercise 5.2, in which 32 students in a science classroom were randomly assigned to one of two study methods, A and B, with $n_A = n_B = 16$. After several weeks of study, students were examined on the course material, and the scores summarized by $\{\bar{y}_A = 75.2, s_A = 7.3\}$ and $\{\bar{y}_B = 77.5, s_B = 8.1\}$. We will estimate $\theta = \mu + \delta$ and $\theta_B = \mu - \delta$ using the two-sample model and the prior distributions of Section 8.1.

(a) Let $\mu \sim N(75, 100)$, $1/\sigma^2 \sim \text{Gamma}(1, 100)$ and $\delta \sim N(\delta_0, \tau_0^2)$. For each combination of $\delta_0 \in \{-4, -2, 0, 2, 4\}$ and $\tau_0^2 \in \{10, 50, 100, 500\}$ obtain the posterior distribution of μ, δ and σ^2 and compute

(i) $Pr(\delta < 0|\mathbf{Y})$

(ii) a 95% posterior confidence interval for δ

(iii) the prior and posterior correlation of θ_A and θ_B

(b) Describe how you might use these results to convey evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

(3) Problem 8.3 (Hoff) The files `school1.dat` through `school8.dat` give weekly hours spent on homework for students sampled from eight diffferent schools. We want to obtain posterior distributions for the true means for the eight different schools using a hierarchical normal model with the following prior parameters: $\mu_0 = 7, \gamma_0^2 = 5, \tau_0^2 = 10, \eta_0 = 2, \sigma_0^2 = 15, \nu_0 = 2$

(a) Run a Gibbs sampling algorithm to approximate the posterior distribution of $\{\boldsymbol{\theta}, \sigma^2, \mu, \tau^2\}$. Assess the convergence of the Markov chain, and find the effective sample size for $\{\sigma^2, \mu, \tau^2\}$. Run the chain long enough so that the effective sample sizes are all above 1,000.

(b) Compute posterior means and 95% confidence regions for $\{\sigma^2, \mu, \tau^2\}$. Also, compare the posterior densities to the prior densities, and discuss what was learned from the data.

(c) Plot the posterior density for $R = \frac{\tau^2}{\sigma^2+\tau^2}$ and compare it to a plot of the prior density on R. Describe the evidence for between-school variation.

(d) Obtain the posterior probability that θ_7 is smaller than θ_6, as well as the posterior probability that θ_7 is the smallest of all the θ's.

(e) Plot the sample averages $\bar{y}_1, ..., \bar{y}_8$ against the posterior expectations of $\theta_1, ..., \theta_8$, and describe the relationship. Also compute the sample mean of all observations and compare it to the posterior mean of μ. Estimate the shrinkage effect for each school.