# Stat 289 Homework 3 - Due Monday, April 12 (6:10 pm)

**You are required to submit a hard copy document in class with answers to the following questions. This document must be generated using the LaTeX typesetting language. It should also include any requested graphics for each question, integrated with the text for that question. Do not just put all graphics at the end of the document!**

**In addition to the hardcopy document, please also submit the LaTeX source code (.tex file) to me by email (stroud@gwu.edu). Please do all computational work using the software package R, and submit your R code to me by email as a separate plain text file. Please do not consult with other students on this assignment.**

For Questions 1-7, you will need the SAT coaching dataset given in Table 1 below. This dataset comes from a randomized experiment to test the effects of 8 different SAT coaching programs. For each program $i$, an observed effect $y_i$ was measured as well as a standard error $\sigma_i$ of the treatment effect. We want to analyze these treatments under a hierarchical model where

$$\begin{aligned} y_i &\sim \text{Normal}(\theta_i, \sigma_i^2) \\ \theta_i &\sim \text{Normal}(\mu, \tau^2) \end{aligned}$$

We will assume that $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_n^2)$ are all known without error. We are interested in posterior inference for each $\theta_i$, which is the underlying true treatment effect for program $i$, as well as the common treatment mean $\mu$ and variance $\tau^2$.

1. Verify that using a flat prior of $p(\mu, \tau) \propto 1$ corresponds to a prior for $(\mu, \tau^2)$ of $p(\mu, \tau^2) \propto \tau^{-1}$. With this prior, write out the full posterior distribution of the unknown parameters $\mu$, $\tau^2$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$.

2. The first approach is based on the realization that we can actually integrate out each $\theta_i$ from the above distribution, giving us the marginal distribution of our treatment effects $y_i$:

   $$y_i \sim \text{Normal}(\mu, \sigma_i^2 + \tau^2)$$

   Write out the marginal posterior distribution of $p(\mu, \tau^2 | \mathbf{y}, \boldsymbol{\sigma}^2)$, and evaluate this posterior distribution over a grid of values of $\mu$ and and $\tau^2$. Use a grid of (0,10) for $\tau^2$.

3. Use the grid sampling method to get 1000 samples from $p(\mu, \tau^2 | \mathbf{y}, \boldsymbol{\sigma}^2)$. Calculate the mean, median and 95% posterior interval for $\mu$ and $\tau^2$.

4. Write out the distribution $p(\theta_i | \mu, \tau^2, \mathbf{y}, \boldsymbol{\sigma}^2)$. Use the 1000 samples of $\mu$ and $\tau^2$ to draw 1000 samples of each $\theta_i$. Calculate the mean of each $\theta_i$ and compare to the observed treatment effects $y_i$.

5. For Questions 5-7, consider $\tau^2$ to be fixed and equal to $\tau^2 = \text{median}(\tau^2)$ that you calculated in Question 3. We now use a Gibbs sampler to draw samples from the posterior distribution $p(\mu, \boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\sigma}^2)$. Remember that $\tau^2$ is now a known quantity in this question, so we do not need to calculate the posterior distribution for it. Give the conditional distributions of the remaining unknown parameters given the other parameters.

6. Use a Gibbs sampler based on the conditional distributions from the previous question to obtain 1000 samples from $p(\mu, \boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\sigma}^2)$. Make sure that you only use samples after convergence and that your 1000 samples have low autocorrelation. Calculate the mean and posterior interval for each $\theta_i$.

7. Use your samples of $\boldsymbol{\theta}$ to calculate an estimate of the posterior probability that school A offers the best program.

   Questions 8-11 are based on the bicycle data in Table 2 below. We will focus only on the first two rows of the table (the residential streets with bike routes). We want to model the total amount of traffic $y_i$ on each street (eg. $y_1 = 74$) as follows:

$$\begin{aligned} y_i &\sim \text{Poisson}(\theta_i) \\ \theta_i &\sim \text{Gamma}(\alpha, \beta) \end{aligned}$$

8. Write out the posterior distribution of our unknown variables $p(\alpha, \beta, \boldsymbol{\theta}|\mathbf{y})$ using a flat prior distribution for $\alpha$ and $\beta$, $p(\alpha, \beta) \propto 1$.

9. We want to use a Gibbs sampler to get samples from $p(\alpha, \beta, \boldsymbol{\theta}|\mathbf{y})$. What is the conditional distribution of $\theta_i$ given all the other parameters?

10. The conditional distributions $p(\alpha, \beta|\mathbf{y})$ and $p(\beta|\alpha, \mathbf{y})$ are not easy to sample from, so you instead need to use a Metropolis (or Metropolis-Hastings) step to obtain samples from $p(\alpha, \beta|\boldsymbol{\theta}, \mathbf{y})$. Choose a proposal distribution for $\alpha$ and a proposal distribution for $\beta$.

11. Combine the algorithm from Question 10 with the conditional distribution from Question 9 to form a Gibbs sampler, which iterates between sampling:

    1. sampling $\theta_i|\alpha, \beta, \mathbf{y}$ for each $i$.
    2. sampling $\alpha, \beta|\boldsymbol{\theta}, \mathbf{y}$.

    Use this algorithm to obtain 1000 samples from $p(\alpha, \beta, \boldsymbol{\theta}|\mathbf{y})$. Make sure that you only use samples after convergence and that your 1000 samples have low autocorrelation. Calculate means and 95% posterior intervals for $\alpha$ and $\beta$.

12. Give a 95% predictive interval for the total traffic on a new residential street with a bike lane.

| School | Estimated treatment effect, $y_j$ | Standard error of effect estimate, $\sigma_j$ |
|---|---|---|
| A | 28 | 15 |
| B | 8 | 10 |
| C | -3 | 16 |
| D | 7 | 11 |
| E | -1 | 9 |
| F | 1 | 11 |
| G | 18 | 10 |
| H | 12 | 18 |

Table 1: *Observed effects of special preparation on SAT-V scores in eight randomized experiments. Estimates are based on separate analyses for the eight experiments. From Rubin (1981).*

| Type of street | Bike route? | Counts of bicycles/other vehicles |
|---|---|---|
| Residential | yes | 16/58, 9/90, 10/48, 13/57, 19/103, 20/57, 18/86, 17/112, 35/273, 55/64 |
| Residential | no | 12/113, 1/18, 2/14, 4/44, 9/208, 7/67, 9/29, 8/154 |
| Fairly busy | yes | 8/29, 35/415, 31/425, 19/42, 38/180, 47/675, 44/620, 44/437, 29/47, 18/462 |
| Fairly busy | no | 10/557, 43/1258, 5/499, 14/601, 58/1163, 15/700, 0/90, 47/1093, 51/1459, 32/1086 |
| Busy | yes | 60/1545, 51/1499, 58/1598, 59/503, 53/407, 68/1494, 68/1558, 60/1706, 71/476, 63/752 |
| Busy | no | 8/1248, 9/1246, 6/1596, 9/1765, 19/1290, 61/2498, 31/2346, 75/3101, 14/1918, 25/2318 |

Table 2: *Counts of bicycles and other vehicles in one hour in each of 10 city blocks in each of six categories. (The data for two of the residential blocks were lost.) For example, the first block had 16 bicycles and 58 other vehicles, the second had 9 bicycles and 90 other vehicles, and so on. Streets were classified as 'residential,' 'fairly busy,' or 'busy' before the data were gathered.*