ST4234: Bayesian Statistics

Data Analysis Assignment, AY 19/20

Instructions:

1. This data analysis assignment consists of **TWO (2)** problems. The total marks of this assignment is **70**=35+35.

2. Copying of solutions is **strictly prohibited**.

3. Please upload your work in **a single pdf file** to the LumiNUS folder **"Project submission"**. The deadline for submission is

   Singapore Time (GMT+8) 12pm noon, Wednesday, 29 April, 2020

   For those students who are in different time zones, please convert this deadline to your current time zone and strictly follow this deadline.

4. Please write clearly **your student number and your name** in the your submission.

5. If your student number is XXX, please name your file

   XXX.pdf

   For example, if your student number is A0012345R, your file should be named **A0012345R.pdf**.

6. You must include your R codes, outputs, and figures in your submitted file. **Your codes must be executable in R or RStudio and produce the outputs in your file. Otherwise, your codes will be subject to mark deduction.**

7. **All random seeds in your R codes must be set as your student number without the letters.** For example, if your student number is A0012345R, your R codes to generate random numbers should be preceded by `set.seed(0012345)`. **Otherwise, your codes will be subject to mark deduction.**

8. You are allowed to scan your handwritings and combine files. Your file should be well organized. **Unrecognizable handwriting will be subject to mark deduction.** Ideally, you should compress the final pdf file to no more than **5Mb**, without severely compromising the quality of images in your file.

9. **No hard copy** will be accepted. **No late submission** will be accepted (i.e. marks for your project = zero).

Question 1. (35 marks) Gelman et al. (2003) describe the results of independent experiments to determine the effects of special coaching programs on SAT scores. This dataset can be found in Section 5.5, Table 5.2 in our reference book *Bayesian Data Analysis third edition (BDA3)*. There are $J = 8$ schools in this experiment. For the jth experiment ($j = 1, \ldots, J$), one observes an estimated coaching effect y_j with associated standard error σ_j; the values of the effects and standard errors are displayed in the table below. We only observe $\boldsymbol{y} = \{y_1, \ldots, y_J\}$ and $\boldsymbol{\sigma} = \{\sigma_1, \ldots, \sigma_J\}$, instead of the original full dataset.

School	Treatment effect y_j	Standard error σ_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

We model this dataset as follows. The coaching effect y_j is normally distributed with mean θ_j and known variance σ_j^2, independently across $j = 1, \ldots, J$. $\theta_1, \ldots, \theta_J$ are drawn independently from a normal population with mean μ and variance τ^2. The vector of parameters (μ, τ) is assigned a uniform prior $p(\mu, \tau) \propto 1$. The objective of modeling in this way is to combine the coaching estimates in some way to obtain improved estimates of the true effects θ_j.

Analyze this dataset by answering the following questions:

(a) (3 marks) Based on the description above, write down an expression for the unnormalized full posterior density $p(\boldsymbol{\theta}, \mu, \tau | \boldsymbol{y}, \boldsymbol{\sigma})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$.

(b) (6 marks) Write down detailed derivations for the following posterior distributions:

- Conditional on (μ, τ), θ_j's have independent posterior distributions:

$$\theta_j | \mu, \tau, \boldsymbol{y}, \boldsymbol{\sigma} \sim \mathrm{N}(\hat{\theta}_j, V_j), \quad \text{where } \hat{\theta}_j = \frac{y_j/\sigma_j^2 + \mu/\tau^2}{1/\sigma_j^2 + 1/\tau^2}, \;\; V_j = \frac{1}{1/\sigma_j^2 + 1/\tau^2}.$$

- The marginal posterior density of (μ, τ) is given by

$$p(\mu, \tau|\boldsymbol{y}, \boldsymbol{\sigma}) \propto \prod_{j=1}^{J} \phi\left(y_j|\mu, \sqrt{\sigma_j^2 + \tau^2}\right),$$

  where $\phi(y|\mu, \sigma)$ denotes the normal density with mean μ and standard deviation σ.

(c) (6 marks) Let $\lambda = (\lambda_1, \lambda_2)$, where $\lambda_1 = \mu$ and $\lambda_2 = \log \tau$. Write an R function to compute the log posterior density of (λ_1, λ_2). (Remember to include the Jacobian term in the transformation.) Draw a contour plot of the log posterior density function, using the range $\lambda_1 \in [-18, 37]$ and $\lambda_2 \in [-6, 4.1]$. Find the normal approximation to the posterior of (λ_1, λ_2).

(d) (6 marks) Use the random walk Metropolis algorithm (function `rwmetrop`) to draw 10^4 samples from the posterior of (λ_1, λ_2). Drop the first 5000 draws as burn-in. Overlay the last 5000 MCMC draws on the contour plot of the log posterior density. Report the acceptance rate and assess the convergence of the Markov chain using traceplots and autocorrelation plots.

(Hint: You do not need to write down the steps of the random walk Metropolis algorithm in your report. You only need to write the R codes to implement the algorithm. Remember to set the random seed to be your student number. The same applies to all the following questions.)

(e) (6 marks) Using the simulated sample from the marginal posterior of (λ_1, λ_2), simulate 5000 draws from the joint posterior density of $\theta_1, \ldots, \theta_J$, from the conditional posterior of θ_j's given (μ, τ) in Part (b). Summarize the posterior distribution of each θ_j by reporting the posterior mean and standard deviation.

(f) (6 marks) From the conditional posteriors in Part (b), we already know that the posterior mean of θ_j, conditional on (μ, τ), can be written as

$$\mathrm{E}(\theta_j|\mu, \tau, \boldsymbol{y}, \boldsymbol{\sigma}) = (1 - B_j)y_j + B_j\mu,$$

where $B_j = \tau^{-2}/(\tau^{-2} + \sigma_j^{-2})$ is the size of the shrinkage of y_j towards μ. For all observations: (i) Compute the shrinkage size $\mathrm{E}(B_j|y)$ from the simulated draws of $(\mu, \log\tau)$ in Part (d); (ii) Rank the eight schools from the largest shrinkage to the smallest shrinkage and explain why there are differences.

(g) (2 marks) School A had the largest observed coaching effect, 28. Using the simulated draws from the joint distribution of $\theta_1, \ldots, \theta_J$, compute the posterior probability $P(\theta_1 > \theta_j)$ for $j = 2, \ldots, J$.

Question 2. (35 marks) The files `school1.txt` through `school8.txt` give weekly hours spent on homework for students sampled from eight different schools. The numbers of observations across eight schools are unequal. You can read the data in the eight text files in R by the following codes:

```
## set your working directory to the folder where the text files are located
setwd(".../school_data")   # replace ... with your directory name
y <- list()
for(i in 1:8) {
  y[[i]] <- as.vector(as.matrix(read.table(paste0("school",i,".txt"))))
}
```

Since the eight datasets have unequal lengths, we define `y` as a `list` object.

To model this dataset, we use the hierarchical normal model described in Section 8.3 of Peter Hoff's book. We use $y_{i,j}$ to denote the ith observation in the jth school, where $i = 1, \ldots, n_j$, $j = 1, \ldots, J$. $J = 8$ for this dataset and n_j is the sample size for school j. We assume that all $y_{i,j}$'s are normally distributed with mean θ_j and variance σ^2, independently across all $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$. The mean θ_j is normally distributed with mean μ and variance τ^2, independently across all $j = 1, \ldots, J$. The parameters μ, τ^2, σ^2 are assigned independent prior distributions.

The hierarchical normal model can be described as follows:

$$
\begin{aligned}
&y_{i,j}|\theta_j, \sigma^2 \overset{\text{indep}}{\sim} \mathrm{N}\left(\theta_j, \sigma^2\right) && \text{(within-group model)} \\
&\theta_j|\mu, \tau^2 \overset{\text{indep}}{\sim} \mathrm{N}(\mu, \tau^2) && \text{(between-group model)} \\
&\mu \sim \mathrm{N}(\mu_0, \sigma_0^2) && \text{(prior for global mean)} \\
&\tau^2 \sim \text{Inv-Gamma}\left(\frac{a_1}{2}, \frac{b_1}{2}\right) && \text{(prior for between-group variance)} \\
&\sigma^2 \sim \text{Inv-Gamma}\left(\frac{a_2}{2}, \frac{b_2}{2}\right) && \text{(prior for within-group variance)}
\end{aligned}
$$

Here, Inv-Gamma(α, β) denotes the inverse gamma distribution with density $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$ for $x > 0$. The hyperparameters $\mu_0 \in \mathbb{R}, \sigma_0^2 > 0, a_1 > 0, b_1 > 0, a_2 > 0, b_2 > 0$ are all fixed constants.

Let $\boldsymbol{y}_j = (y_{1,j}, \ldots, y_{n_j,j})$, $\bar{y}_j = n_j^{-1} \sum_{i=1}^{n_j} y_{i,j}$, for $j = 1, \ldots, J$, and $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J)$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$, and $\bar{\theta} = J^{-1} \sum_{j=1}^{J} \theta_j$.

Analyze this dataset by answering the following questions:

(a) (3 marks) Based on the description above, write down an expression for the unnormalized full posterior density $p(\boldsymbol{\theta}, \mu, \tau^2, \sigma^2 | \boldsymbol{y})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$.

(b) (8 marks) Write down complete derivations for the following four posterior distributions. You need to find all the missing parameters in "?".

- Conditional on (μ, τ^2, σ^2), θ_j's $(j = 1, \ldots, J)$ have independent posterior distributions

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

- Conditional on $(\boldsymbol{\theta}, \tau^2, \sigma^2)$, μ has the posterior distribution

$$\mu | \boldsymbol{\theta}, \tau^2, \sigma^2, \boldsymbol{y} \sim \mathrm{N}(?, \ ?).$$

- Conditional on $(\boldsymbol{\theta}, \mu, \sigma^2)$, τ^2 has the posterior distribution

$$\tau^2 | \boldsymbol{\theta}, \mu, \sigma^2, \boldsymbol{y} \sim \text{Inv-Gamma}(?, \ ?).$$

- Conditional on $(\boldsymbol{\theta}, \mu, \tau^2)$, σ^2 has the posterior distribution

$$\sigma^2 | \boldsymbol{\theta}, \mu, \tau^2, \boldsymbol{y} \sim \text{Inv-Gamma}(?, \ ?).$$

(c) (9 marks) Suppose that the hyperparameters have the following values

$$\mu_0 = 7, \ \sigma_0^2 = 5, \ a_1 = 10, \ b_1 = 20, \ a_2 = 15, \ b_2 = 30.$$

Write down a Gibbs sampler in R to draw $T = 10^4$ samples from the posterior $p(\boldsymbol{\theta}, \mu, \tau^2, \sigma^2 | \boldsymbol{y})$. Discard the first 5000 draws as burn-in and keep the last 5000 draws. Assess the convergence of the Markov chain by checking the traceplots and the autocorrelation plots.

(Hint: You do not need to write the steps of the Gibbs sampler in your report. You only need to write R codes to implement the algorithm. You should **not** use `gibbs` in `LearnBayes`.)

(d) (8 marks) Summarize the marginal posteriors of μ, τ^2, σ^2 as follows: (i) Report the posterior means and 95% HPD intervals for each parameter of μ, τ^2, σ^2; (ii) For each of μ, τ^2, σ^2, plot their prior density and their marginal posterior density. Discuss what is learned from the data.

(Hint: You can use `plot(density(...))` or `lines(density(...))` to plot the marginal posterior densities with kernel density estimation. You can plot the prior and the posterior densities either separately or on the same plot.)

(e) (5 marks) For the ratio $R = \frac{\tau^2}{\tau^2+\sigma^2}$, use Monte Carlo approximation with 5000 draws to plot its prior density, and then plot its posterior density. Use these plots to describe the evidence for between-school variation.

(f) (2 marks) Plot the sample averages $\bar{y}_1, \ldots, \bar{y}_J$ against the posterior expectations of $\theta_1, \ldots, \theta_J$, and describe the relationship.