The Chinese University of Hong Kong, Shenzhen
SDS · School of Data Science

· Fall Semester 2022

DDA 4010 – Bayesian Statistics

Exercise Sheet 5

This exercise is due on **Dec. 11st, 5:00 pm.**

Assignment A5.1 (8.1 **in Textbook):**
Components of variance: Consider the hierarchical model where

$$\theta_1, \ldots, \theta_m \mid \mu, \tau^2 \sim \text{ i.i.d. normal} \left(\mu, \tau^2\right)$$
$$y_{1,j}, \ldots, y_{n_j,j} \mid \theta_j, \sigma^2 \sim \text{ i.i.d. normal} \left(\theta_j, \sigma^2\right).$$

For this problem, we will eventually compute the following: $\operatorname{Var}\left[y_{i,j} \mid \theta_i, \sigma^2\right], \operatorname{Var}\left[\bar{y}\cdot, j\theta_i, \sigma^2\right]$, $\operatorname{Cov}\left[y_{i_1,j}, y_{i_2,j} \mid \theta_j, \sigma^2\right]$ $\operatorname{Var}\left[y_{i,j} \mid \mu, \tau^2\right], \operatorname{Var}\left[\bar{y}, j \mid \mu, \tau^2\right], \operatorname{Cov}\left[y_{i_1,j}, y_{i_2,j} \mid \mu, \tau^2\right]$
First, lets use our intuition to guess at the answers:

- Which do you think is bigger, $\operatorname{Var}\left[y_{i,j} \mid \theta_i, \sigma^2\right]$ or $\operatorname{Var}\left[y_{i,j} \mid \mu, \tau^2\right]$? 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.
- Do you think $\operatorname{Cov}\left[y_{i_1,j}, y_{i_2,j} \mid \theta_j, \sigma^2\right]$ is negative, positive, or zero? Answer the same for $\operatorname{Cov}\left[y_{i_1,j}, y_{i_2,j} \mid \mu, \tau\right]$. 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.
- Now compute each of the six quantities above and compare to your answers in a) and b).
- Now assume we have a prior $p(\mu)$ for μ. Using Bayes' rule, show that

  $$p\left(\mu \mid \theta_1, \ldots, \theta_m, \sigma^2, \tau^2, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\right) = p\left(\mu \mid \theta_1, \ldots, \theta_m, \tau^2\right)$$

  Interpret in words what this means.

Assignment A5.2 (8.3 **in Textbook):**
Hierarchical modeling: The files `school1.dat` through `school8.dat` give weekly hours spent on homework for students sampled from eight different schools. Obtain posterior distributions for the true means for the eight different schools using a hierarchical normal model with the following prior parameters:

$$\mu_0 = 7, \gamma_0^2 = 5, \quad \tau_0^2 = 10, \eta_0 = 2, \quad \sigma_0^2 = 15, \nu_0 = 2$$

- 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 .
- 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.

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

- 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.

- Plot the sample averages $\bar{y}_1, \ldots, \bar{y}_8$ against the posterior expectations of $\theta_1, \ldots, \theta_8$, and describe the relationship. Also compute the sample mean of all observations and compare it to the posterior mean of μ.

Assignment A5.3 (9.3 **in Textbook):**
Crime: The file `crime.dat` contains crime rates and data on 15 explanatory variables for 47 U.S. states, in which both the crime rates and the explanatory variables have been centered and scaled to have variance 1. A description of the variables can be obtained by typing `library(MASS);?UScrime` in R.

- Fit a regression model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ using the g-prior with $g = n, \nu_0 = 2$ and $\sigma_0^2 = 1$. Obtain marginal posterior means and 95% confidence intervals for $\boldsymbol{\beta}$, and compare to the least squares estimates. Describe the relationships between crime and the explanatory variables. Which variables seem strongly predictive of crime rates?

- Lets see how well regression models can predict crime rates based on the $\mathbf{X}$-variables. Randomly divide the crime roughly in half, into a training set $\{\boldsymbol{y}_{\text{tr}}, \mathbf{X}_{\text{tr}}\}$ and a test set $\{\boldsymbol{y}_{\text{te}}, \mathbf{X}_{\text{te}}\}$
  - Using only the training set, obtain least squares regression coefficients $\hat{\boldsymbol{\beta}}_{\text{ols}}$. Obtain predicted values for the test data by computing $\hat{\boldsymbol{y}}_{\text{ols}} = \mathbf{X}_{\text{te}}\hat{\boldsymbol{\beta}}_{\text{ols}}$. Plot $\hat{\boldsymbol{y}}_{\text{ols}}$ versus $\boldsymbol{y}_{\text{te}}$ and compute the prediction error $\frac{1}{n_{\text{te}}} \sum \left(y_{i,\text{ te}} - \hat{y}_{i,\text{ols}}\right)^2$.
  - Now obtain the posterior mean $\hat{\boldsymbol{\beta}}_{\text{Bayes}} = \mathrm{E}\left[\boldsymbol{\beta} \mid \boldsymbol{y}_{\text{tr}}\right]$ using the g-prior described above and the training data only. Obtain predictions for the test set $\hat{\boldsymbol{y}}_{\text{Bayes}} = \mathbf{X}_{\text{test}}\,\hat{\boldsymbol{\beta}}_{\text{Bayes}}$. Plot versus the test data, compute the prediction error, and compare to the OLS prediction error. Explain the results.

- Repeat the procedures in b) many times with different randomly generated test and training sets. Compute the average prediction error for both the OLS and Bayesian methods.

Assignment A5.4 (10.2 **in Textbook):**
esting success: Younger male sparrows may or may not nest during a mating season, perhaps depending on their physical characteristics. Researchers have recorded the nesting success of 43 young male sparrows of the same age, as well as their wingspan, and the data appear in the file `msparrownest.dat`. Let Y_i be the binary indicator that sparrow i successfully nests, and let x_i denote their wingspan. Our model for Y_i is $\operatorname{logit} \Pr\left(Y_i = 1 \mid \alpha, \beta, x_i\right) = \alpha + \beta x_i$, where the logit function is given by $\operatorname{logit} \theta = \log[\theta/(1-\theta)]$.

- Write out the joint sampling distribution $\prod_{i=1}^{n} p\left(y_i \mid \alpha, \beta, x_i\right)$ and simplify as much as possible.

- Formulate a prior probability distribution over α and β by considering the range of $\Pr(Y = 1 \mid \alpha, \beta, x)$ as x ranges over 10 to 15 , the approximate range of the observed wingspans.

- Implement a Metropolis algorithm that approximates $p(\alpha, \beta \mid \boldsymbol{y}, \boldsymbol{x})$. Adjust the proposal distribution to achieve a reasonable acceptance rate, and run the algorithm long enough so that the effective sample size is at least 1,000 for each parameter.
- Compare the posterior densities of α and β to their prior densities.
- Using output from the Metropolis algorithm, come up with a way to make a confidence band for the following function $f_{\alpha\beta}(x)$ of wingspan:

$$f_{\alpha\beta}(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}},$$

  where α and β are the parameters in your sampling model. Make a plot of such a band.

Sheet 5 is due on **Dec. 11st**. Submit your solutions before **Dec. 11st**, **5:00 pm**.