Bayesian Inference PSTAT 215A
Homework 4 (of 4) RBG/Fall 2009

Assigned 10 Nov 2009; due 24 Nov 2009. *You may study the questions/answers with others, but what you hand in for credit must be your own work.*

1. For the multivariate normal model, Jeffreys' rule for generating a prior distribution on (μ, Σ) gives $p_J(\mu, \Sigma) \propto |\Sigma|^{-(p+2)/2}$.

   (a) Explain why the function p_J cannot actually be a probability density for (μ, Σ)

   (b) Find the form of the posterior density $p_J(\mu, \Sigma | y_1, \ldots, y_n)$ under the Jeffreys' prior and argue that it is proper under a condition that you should describe.

   (c) Obtain the closed form of $p_J(\mu | \Sigma, y_1, \ldots, y_n)$ and $p_J(\Sigma | \mu, y_1, \ldots, y_n)$.

   (d) Obtain the closed form of the posterior marginal $p_J(\Sigma | y_1, \ldots, y_n)$.

2. Marriage data: A file linked from the course web page contains data on the ages of 100 married couples sampled from the US population.

   (a) Before you look at the data, use your own knowledge to formulate a semi-conjugate prior distribution for $\mu = (\mu_h, \mu_w)^\top$ and Σ, where μ_h, μ_w are the mean husband and wife ages, respectively, and Σ is the covariance matrix. Confirm that the prior parameterization is an accurate representation of your prior knowledge by generating *prior predictive datasets* and describe how their characteristics reflect your knowledge.

   (b) Using your prior distribution and the 100 values in the dataset, obtain a MCMC approximation to $p(\mu, \Sigma | y_1, \ldots, y_{100})$. Plot the joint posterior density of μ_h and μ_w, and also the marginal posterior density of the correlation between Y_h and Y_w. Obtain 95% CIs for μ_h, μ_w and the correlation coefficient.

   (c) Perform the same analysis as in part (b) using the Jeffreys' prior and posterior calculations from Question 1. Compare the two sets of CIs and comment on the relative merits or faults of your prior versus the Jeffreys' prior for this data. What about if the sample size were much smaller, say $n = 25$?

3. A file linked from the course web page contains data on the amount of time, in seconds, that it takes each of four high school swimmers to swim 50 yards. Each swimmer has six times, taken on a bi-weekly basis.

   (a) Perform the following data analysis for each swimmer separately.

      i. Fit a Bayesian linear regression model of swimming time as the response and week as the explanatory variable. To formulate your prior, use the information that competitive times for this age group generally range from 22 to 24 seconds.

ii. For each swimmer, j, obtain a posterior predictive distribution for Y_j^*, their time if they were to swim two weeks from the last recorded time (x^*).

(b) The coach of the team must decide which of the four swimmers will compete in a swimming meet in two weeks. Using your predictive distributions, compute $P(Y_j^* = \max\{Y_1^*, \ldots, Y_4^*\}|y)$ for each swimmer j, and based on this make a recommendation to the coach.

(c) Comment on anything that might concern you about this type of *extrapolation* beyond the range of the data with the linear model.

4. A file linked from the course web page contains data on the 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: $\psi_0 = 7,\ \gamma_0 = 5,\ \tau_0^2 = 10,\ \eta_0 = 2,\ \sigma_0^2 = 15,\ \nu_0 = 2$.

   (a) Run a GS algorithm to approximate the posterior distribution of $\{\mu, \sigma^2, \psi, \tau^2\}$. Assess the convergence of the Markov chain, and find the ESS for $\{\sigma^2, \psi, \tau^2\}$. Run the chain long enough so that the ESSs are all above 1,000.

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

   (c) 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.

   (d) Obtain the posterior probability that μ_7 is smaller than μ_6, as well as the posterior probability that μ_7 is the smallest of all of the μ's.

   (e) Plot the sample averages $\bar{y}_1, \ldots, \bar{y}_8$ against the posterior expectations of $\mu_1, \ldots, \mu_8$. Also compare the sample mean of all observations to the posterior mean of ψ.

5. A file linked from the course web page has counts of the number of bicycles and other vehicles for 10 randomly selected residential streets with bike routes in Berkeley, CA.

   (a) Set up a hierarchical model for the observed number of bicycles on streets $j = 1, \ldots, 10$ that is binomial with unknown probability θ_j and sample size equal to the total number of vehicles (bicycles included) in that block. Assign a beta population distribution for the parameter θ_j and a non-informative hyperprior distribution as described in lecture. Write down the joint posterior distribution.

   (b) Compute the marginal posterior density of the hyperparameters and draw simulations from the joint posterior distribution of the parameters and hyperparameters.

   (c) Compare the posterior distributions of the parameters θ_j to the raw proportions in location j.

   (d) Give a 95% CI for the average underlying proportion of traffic that is bicycles.

   (e) A location on a new residential street with a bicycle route is sampled at random during which time 100 vehicles of all kinds go by. Give a 95% CI for the number of those vehicles that are bicycles.