MATH 538: Bayesian Data Analysis (Fall 2024)
Assignment #4

Due date: Wednesday, December 11th

Book Problems: Chapter 11: 2; Chapter 5: 4, 11, 13 - for this problem you use M-H instead of grid sampling for drawing from the marginal posterior $p(\alpha, \beta | data)$

Additional Problem:

1. Replicate our analysis of the SAT data under the normal hierarchical model in STAN. Describe/discuss your findings.

4. Exchangeable prior distributions: suppose it is known *a priori* that the $2J$ parameters $\theta_1, \ldots, \theta_{2J}$ are clustered into two groups, with exactly half being drawn from a $\mathrm{N}(1, 1)$ distribution, and the other half being drawn from a $\mathrm{N}(-1, 1)$ distribution, but we have not observed which parameters come from which distribution.

   (a) Are $\theta_1, \ldots, \theta_{2J}$ exchangeable under this prior distribution?

   (b) Show that this distribution cannot be written as a mixture of independent and identically distributed components.

   (c) Why can we not simply take the limit as $J \to \infty$ and get a counterexample to de Finetti's theorem?

   See Exercise 8.10 for a related problem.

11. Nonconjugate hierarchical models: suppose that in the rat tumor example, we wish to use a normal population distribution on the log-odds scale: $\text{logit}(\theta_j) \sim \text{N}(\mu, \tau^2)$, for $j = 1, \ldots, J$. As in Section 5.3, you will assign a noninformative prior distribution to the hyperparameters and perform a full Bayesian analysis.

$$p(\phi|y) = \int p(\theta, \phi|y) d\theta.$$

(a) Write the joint posterior density, $p(\theta, \mu, \tau|y)$.

(b) Show that the integral (5.4) has no closed-form expression.

$$p(\phi|y) = \frac{p(\theta, \phi|y)}{p(\theta|\phi, y)}$$

(c) Why is expression (5.5) no help for this problem?

In practice, we can solve this problem by normal approximation, importance sampling, and Markov chain simulation, as described in Part III.

13. Hierarchical binomial model: Exercise 3.8 described a survey of bicycle traffic in Berkeley, California, with data displayed in Table 3.3. For this problem, restrict your attention to the first two rows of the table: residential streets labeled as 'bike routes,' which we will use to illustrate this computational exercise.

    (a) Set up a model for the data in Table 3.3 so that, for $j = 1, \ldots, 10$, the observed number of bicycles at location j is binomial with unknown probability θ_j and sample size equal to the total number of vehicles (bicycles included) in that block. The parameter θ_j can be interpreted as the underlying or 'true' proportion of traffic at location j that is bicycles. (See Exercise 3.8.) Assign a beta population distribution for the parameters θ_j and a noninformative hyperprior distribution as in the rat tumor example of Section 5.3. 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, as in Section 5.3.

    (c) Compare the posterior distributions of the parameters θ_j to the raw proportions, (number of bicycles / total number of vehicles) in location j. How do the inferences from the posterior distribution differ from the raw proportions?

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

    (e) A new city block is sampled at random and is a residential street with a bike route. In an hour of observation, 100 vehicles of all kinds go by. Give a 95% posterior interval for the number of those vehicles that are bicycles. Discuss how much you trust this interval in application.

    (f) Was the beta distribution for the θ_j's reasonable?

scratch