ST4234: Bayesian Statistics

Tutorial 9, AY 19/20

Instructions: For each group who were scheduled to present on Tuesday, 7 April, please upload **one** copy of your solutions to the folder "Tutorial submission" by **6pm Tuesday, 7 April**.

1. (Continued from Tutorial 8 Question 2) Consider the genetic linkage model in Chapter 5. n ($n = y_1 + y_2 + y_3 + y_4$) animals are distributed into four categories as follows (Rao 1997):

$$\boldsymbol{y} = (y_1, y_2, y_3, y_4)$$

with cell probabilities

$$\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right),$$

where $\theta \in [0, 1]$ is the parameter. Suppose that θ is assigned a uniform prior on $[0, 1]$.

We consider using a Gibbs sampler to find the posterior $p(\theta|\boldsymbol{y})$. We augment the data into 5 categories, with probabilities

$$\left(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right).$$

We split the first category into two new categories with probabilities $1/2$ and $\theta/4$. Let z and $y_1 - z$ be the number of animals in these two new categories, where z is an integer between 0 and y_1. Clearly, z is unobserved. We use Gibbs sampler to draw (θ, z) together.

(a) Show that the joint posterior of (z, θ) is given by

$$p(\theta, z|\boldsymbol{y}) \propto \frac{1}{z!(y_1 - z)!} 2^z \theta^{y_1+y_4-z} (1-\theta)^{y_2+y_3}.$$

(b) Find the conditional posteriors $p(\theta|z, \boldsymbol{y})$ and $p(z|\theta, \boldsymbol{y})$. Based on these, write down a Gibbs sampler to sample $T = 10^4$ draws from the joint posterior $p(\theta, z|\boldsymbol{y})$, for $\boldsymbol{y} = (125, 18, 20, 34)$ and $\boldsymbol{y} = (14, 0, 1, 5)$, respectively. Compare the posterior means of θ from this Gibbs sampler with the true posterior means in Chapter 5 notes.

2. Suppose a company obtains boxes of electronic parts from a particular supplier. It is known that 80% of the lots are acceptable and the lifetimes of the "acceptable" parts follow an exponential distribution with mean λ_A. Unfortunately, 20% of the lots are unacceptable and the lifetimes of the "bad" parts are exponential with mean λ_B, where $\lambda_A > \lambda_B$. Suppose $\boldsymbol{y} = (y_1, \ldots, y_n)$ are the lifetimes of n inspected parts that can come from either acceptable or unacceptable lots. The y_is are a random sample from the mixture distribution

$$p(y|\lambda_A, \lambda_B) = 0.8\frac{\exp(-y/\lambda_A)}{\lambda_A} + 0.2\frac{\exp(-y/\lambda_B)}{\lambda_B}$$

Suppose (λ_A, λ_B) are assigned the noninformative prior proportional to $1/(\lambda_A\lambda_B)$.

(a) Show that the log posterior density of the transformed parameters $\theta = (\theta_A, \theta_B) = (\log\lambda_A, \log\lambda_B)$ is given by

$$\log p(\theta|\boldsymbol{y}) = \sum_{i=1}^{n} \log\left[0.8\frac{\exp(-y_i/\lambda_A)}{\lambda_A} + 0.2\frac{\exp(-y_i/\lambda_B)}{\lambda_B}\right] + C,$$

where C is an additive constant not depending on θ.

The following lifetimes are observed from a sample of 30 parts:

9.3, 4.9, 3.5, 26.0, 0.6, 1.0, 3.5, 26.9, 2.6, 20.4, 1.0, 10.0, 1.7, 11.3, 7.7,
14.1, 24.8, 3.8, 8.4, 1.1, 24.5, 90.7, 16.4, 30.7, 8.5, 5.9, 14.7, 0.5, 99.5, 35.2

(b) Construct a contour plot of (θ_A, θ_B) over the grid $1 \le \theta_A \le 4$ and $-2 \le \theta_B \le 8$. (Hint: You are going to observe two modes.)

(c) Using `optim`, search for the posterior mode with a starting guess of $(\theta_A, \theta_B) = (3, 0)$.

(d) Search for the posterior mode with a starting guess of $(\theta_A, \theta_B) = (2, 4)$.

(e) Explain why you obtain different estimates of the posterior mode in (c) and (d).

(f) Use a normal approximation of the posterior to construct a random walk Metropolis chain for sampling from the posterior of θ. Run the chain for 10,000 iterations, and construct density estimates for θ_A and θ_B. (Hint: You can use the R function `plot(density(...))` to plot the kernel density estimates, for the posterior marginal densities of θ_A and θ_B, respectively.)

(g) Construct a Metropolis within Gibbs sampler using the function `gibbs` in the `LearnBayes` package. Also run the chain for 10,000 iterations and construct density estimates for θ_A and θ_B. (Hint: Please check out Chapter 8 notes pages 38–41 for the usage of `gibbs`.)

(h) By looking at diagnostic plots and acceptance rates, compare the efficiency and accuracy of the two samplers in estimating θ_A and θ_B.