STAT 457 - FINAL

Martha Eichlersmith

2019-12-12

Problem 1

Recall the genetic linkage model of Section 5.1.

Problem 1a

Problem 1a(1)

For the data $Y = (125, 18, 30, 34)$, implement the Gibbs sampler algorithm. Use a flat prior on θ.
Plot θ^i versus iteration i. How long a chain (or chains) did you use? Did you toss out any initial values?

Problem 1a(2)

Compute the posterior mean and posterior variance based on your chain.

Problem 1a(3)

Plot the estimated observed posterior along with the normalized likelihood.

Problem 1a(4)

Discuss the adequacy of the estimate.

Problem 1b

Repeat 1a for $Y = (14, 0, 1, 5)$.

Problem 1c

Problem 1c(1)

Run 20 chains with independent starting values. Compute the average of the θ's in each chain.

Problem 1c(2)

Calculate the standard deviation of the 20 averages. Interpret this value.

Problem 1c(3)

Compute the standard deviation of the θ's in each chain.
Divide each SD by the square root of the number of iterations. Average these "standard errors".

Problem 1c(4)

Compare the values in 1c(2) and 1c(3).
Would you expect these numbers to be similar or different?

Problem 2

Problem 2a

For the genetic linkage model applied to $Y = (125, 18, 20, 34)$, implement the Metropolis algorithm.
(use a flat prior on θ). Use one long chain and plot θ^i versus i.
Try several driver functions:

Problem 2a(1) - Uniform on (0,1)

Problem 2a(2) - Normal Centered at the Current Point of the Chain and sd = 0.01

Problem 2a(3) - Normal Centered at the Current Point of the Chain and sd = 0.1

Problem 2a(4) - Normal Centered at the Current Point of the Chain and sd = 0.5

Problem 2a(5) - Normal Centered at 0.4 and sd = 0.1

Problem 2b

Repeat 2a for $Y = (14, 0, 1, 5)$

Problem 2c

Compute both the posterior mean and standard deviation for both data sets.
Compare to results from the previous problem.

Problem 2d

Problem 2d(1)

For each of the drives in part 2a, run 20 chains with independent starting values.
Compute the averages of the θ's in each chain.

Problem 2d(2)

Calculate the standard deviation of the 20 averages. Interpret this value.

Problem 2d(3)

Compute the standard deviation of the θ's in each chain. Divide each SD by the square root of the number of iterations.
Average these "standard errors".

Problem 2d(4)

Compare the 2d(2) values to 2d(3). Would you expect these number to be similar or different?
Compare to the results of Exercise 1c.

Problem 3

Problem 3a

Consider the 1-way variance components model

$$Y_{ij} = \theta_i + \epsilon_{ij}$$

where Y_{ij} is the jth observation from the ith group, θ_i is the effect, ϵ_{ij}=error, $i = 1, ..., K$ and $j = 1, ..., J$. It is assume that $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\epsilon)$ and $\theta_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2_\theta)$. Under the prior specification $p(\sigma^2_\epsilon, \sigma^2_\theta, \mu) = p(\sigma^2_\epsilon)p(\sigma^2_\theta)p(\mu)$, with $p(\sigma^2_\theta) = \text{InverseGamma}(a_1, b_1)$, $p(\sigma^2_\epsilon) = \text{InverseGamma}(a_2, b_2)$, and $p(\mu) = \mathcal{N}(\mu_0, \sigma^2_0)$.
Let $\overline{Y}_i = \frac{1}{J}\sum_{j=1}^J Y_{ij}$ and $\theta = (theta_1, \cdots, \theta_k)$.
Show the following:

Problem 3a(1)

$$p(\mu \mid \theta, \sigma^2_\epsilon, \sigma^2_\theta, Y) = \mathcal{N}\left(\frac{\sigma^2_\theta \mu_0 + \sigma^2_0 \sum \theta_i}{\sigma^2_\theta + K\sigma^2_0},\ \frac{\sigma^2_\theta \sigma^2_0}{\sigma^2_\theta + K\sigma^2_0}\right)$$

Problem 3a(2)

$$p(\theta_i \mid \mu, \sigma_\epsilon^2, \sigma_\theta^2, Y) = \mathcal{N}\left(\frac{J\sigma_\theta^2}{J\sigma_\theta^2 + \sigma_\epsilon^2} \cdot \overline{Y}_i + \frac{\sigma_\epsilon^2}{J\sigma_\theta^2 + \sigma_\epsilon^2} \cdot \mu, \quad \frac{\sigma_\theta^2 \sigma_\epsilon^2}{J\sigma_\theta^2 + \sigma_\epsilon^2}\right)$$

Problem 3a(3)

$$p(\sigma_\epsilon^2 \mid \mu, \theta, \sigma_\theta^2, Y) = \text{InverseGamma}\left(a_2 + \frac{KJ}{2},\ b_2 + \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij} - \theta_i)^2\right)$$

Problem 3a(4)

$$p(\sigma_\theta^2 \mid \mu, \theta, \sigma_\epsilon^2, Y) = \text{InverseGamma}\left(a_1 + \frac{K}{2},\ b_1 + \frac{1}{2}\sum_{i=1}^{K}(\theta_i - \mu)^2\right)$$

Problem 3b

Run the Gibbs sampler for the data below. Use one chain of length 75,000. Take $p(\mu) = \mathcal{N}(0, 10^{12})$, $p(\sigma_\epsilon^2) = IG(0, 0)$, and $p(\sigma_\theta^2) = IG(1, 1)$. For each θ_i, for σ_ϵ, and for θ_θ, plot the simulated value at iteration j versus j. Summarize each posterior marginal.

Problem 3c

Repeat 3b using the prior specification $p(\mu) = \mathcal{N}(0, 10^{12})$, $p(\sigma_\epsilon^2) = IG(0, 0)$, and $p(\sigma_\theta^2) = IG(0, 0)$. Does this specification violate the Hobart_Casella conditions? Describe what happens to the Gibbs sampler chain in this case.

Problem 5

Suppose that X and Y have exponential conditional distributions restricted over the interval $(0, B)$, i.e. $p(x \mid y) \propto y \exp\{-yx\}$ for $0 < x < B < \infty$ and $p(y \mid x) \propto x \exp\{-xy\}$ for $0 < y < B < \infty$, where B is known constant.

Problem 5a

Take $m = 1$ and $B = 3$. Run the data augmentation algorithm using these conditionals. (Hist: Reject the exponential deviates that lie outside $(0, B)$). How did you assess convergence of this chain? Obtain the marginal for x using the mixture of conditionals $p(x \mid y)$, mixed over the simulated y deviates in your chain.

Problem 5b

Show that the marginal for x is proportional to $(1 - \exp\{-Bx\})/x$. Compare your results in 5a to this curve.

Problem 5c

Repeat 5a and 5b using $B = \infty$. Describe what happens. Is the marginal for x a proper density in this case?