STA402L: HOMEWORK 5

DUE: 11:59 PM ON FRIDAY, MARCH 27

Instructions. Solutions must be submitted to Gradescope as a single PDF. Programming exercises must be completed in R, should be clearly presented, and include all R code. Lab questions are restated here for convenience, but you should refer to the lab itself for details.

Total points. Book exercises: 20; Lab exercises 10; Overall: 30.

BOOK EXERCISES

B1. (3 points) Hoff 10.1.

B2. (3 points) Hoff 10.4.

B3. (8 points) Consider a Markov chain on the state space $\{1, 2, 3, 4\}$ with transition kernel

$$K = \begin{pmatrix} 1/2 & 1/3 & 0 & 0 \\ 1/2 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/2 \\ 0 & 0 & 1/3 & 1/2 \end{pmatrix},$$

where K_{ij} denotes the probability of transitioning from state j to state i.

(a) (2 points) Draw a diagram of the Markov chain, labeling each edge with its transition probability.

(b) (2 points) A Markov chain is *irreducible* if for all i, j there exists $n \geq 1$ such that $(K^n)_{ij} > 0$, and *aperiodic* if the greatest common divisor of return times to each state is 1. Show that K is irreducible by arguing that any state can reach any other in a finite number of steps. Aperiodicity follows from the fact that all diagonal entries of K are positive.

(c) (2 points) A canonical result states that any irreducible and aperiodic Markov chain on a finite state space has a unique invariant measure π^*. Find the π^* that corresponds to K.

(d) (2 points) A key practical question in MCMC is how long the chain must run before it is close to π^*. Simulate the Markov chain associated to K for $S = 1000$ steps starting from $\theta^{(0)} = 1$, and at each step s and for each $j \in \{1, 2, 3, 4\}$ compute the empirical distribution

$$\hat{\pi}_j^{(s)} = \frac{1}{s} \sum_{i=1}^{s} \mathbb{1}\left(\theta^{(i)} = j\right)$$

of the jth entry of your approximation $\hat{\pi}^{(s)}$ of π^*, where $\mathbb{1}(\cdot)$ denotes the indicator function. Plot $\hat{\pi}_j^{(s)}$ as a function of s for each j, overlaying the true value π_j^* as a horizontal dashed line. At roughly what step does the chain appear to have converged?

B4. (6 points) *Gibbs sampling as a special case of Metropolis–Hastings.* Let $p(\theta \mid y)$ be our posterior or interest, where y is observed data and $\theta = (\theta_1, \ldots, \theta_d)$ be a d-dimensional parameter. Consider the following coordinate-wise Metropolis–Hastings algorithm.

- Set an initial value $\theta^{(0)}$
- For $i = 1, \ldots, S$:
  - For $j = 1, \ldots, d$:
    - Draw $\theta_j^* \sim J_j\left(\theta_j \mid \theta_1^{(i)}, \ldots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \ldots, \theta_d^{(i-1)}\right)$.
    - Compute the associated acceptance ratio, r_j.
    - Draw $u \sim \text{Uniform}(0, 1)$.
    - Set $\theta_j^{(i)} = \theta_j^*$ if $r_j > u$ or to $\theta_j^{(i-1)}$ otherwise.

(a) (2 points) Write an expression for the acceptance ratio r_j in the above algorithm.

(b) (4 points) Suppose that J_j is the full conditional distribution of θ_j, that is,

$$J_j(\theta_j \mid \theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_d) = p(\theta_j \mid \theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_d, y).$$

Prove that $r_j = 1$. Conclude that Gibbs sampling is a special case of Metropolis-Hastings.

Lab exercises

L1. (2 points) Given the multivariate normal distribution above, what are the posterior complete conditionals for X, Y , and Z? That is, derive $\pi(X|Y, Z)$, $\pi(Y|X, Z)$, and $\pi(Z|X, Y)$. Note that you should have three univariate normal distributions.

L2. (2 points) Write a Gibbs sampler that alternates updating each of the variables. You can set the initial values for all three variables to 0 and the number of mcmc samples to 1,000. Provide a trace plot and an autocorrelation plot of the draws for either X or Y. Comment on the plots.

L3. (2 points) One option for dealing with this high correlation is doing ¡font color=”green”¿**block updates**¡/font¿, where multiple variables are updated at once. Give the conditional distributions for $(X, Y)|Z$ and $Z|(X, Y)$. Note that you should have one bivariate normal distribution and one univariate normal distribution.

L4. (2 points) Write a Gibbs sampler using the conditional distributions in Exercise 3 above, where X and Y are updated together (using a random draw from a bivariate normal), alternating with Z being updated. You can once again set the initial values for all three variables to 0 and the number of mcmc samples to 1,000. Provide a trace plot and an autocorrelation plot of the draws for either X or Y, and comment on the plots.

L5. (2 points) Comment on the difference between the performance of the two Gibbs samplers. Why is the second more efficient?