Math 475 Homework 6

Due Monday 21st November at start of lecture

Reminder: submit all of your `R` code. I should be able to run your code exactly and get the answers you have reported. Setting the seed is a good idea.

1. You may find it helpful to read (see Blackboard)

   - Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**(4), 327–335.

   (a) Provide your own `R` code to perform random walk Metropolis-Hastings sampling from a Student's t distribution with $\nu = 5$ degrees of freedom. Use a standard normal proposal density.

   (b) Throw away the first 5,000 sampled values. Use the next 20,000 values to provide a Monte Carlo estimate of the mean of this Student's t distribution.

   (c) Compute the acceptance rate.

   (d) Give a plot of the sample autocorrelation function for the first 50 lags.

   (e) Give a Q-Q plot of the MCMC sample quantiles against the theoretical quantiles of the Student's t distribution with $\nu = 5$ degrees of freedom; use `R`'s built-in tool for computing such values, if you want.

2. Repeat Question 1, parts (b), (c), (d), and (e), but using a normal proposal density with mean 0 and variance 0.6.

3. Repeat Question 1, parts (b), (c), (d), and (e), but using a normal proposal density with mean 0 and variance 0.2.

4. Repeat Question 1, parts (b), (c), (d), and (e), but using a uniform proposal density on the interval $(-1, 1)$.

5. If you were to use the normal proposal, what value of the proposal variance gives you an acceptance rate close to the theoretically optimal 0.234? (as we learned in lecture, this should not be taken too seriously, but give an answer anyway to practice 'tuning' the algorithm).

6. For whichever of the previous problems had the greatest problem with high autocorrelation, repeat the sampling procedure but play around with a longer burn-in period and using 'thinning' of the chain in order to reduce the autocorrelation. What seems to have an effect? Remember, if you thin by keeping only every 10th value, then you need a chain that is 10 times as long to get the same number of samples. You should do so to avoid reducing the precision in your sample autocorrelation estimates.

7. Consider the following bivariate density,

$$f(x,y) \propto \binom{n}{x} y^{x+a-1}(1-y)^{n-x+b-1}, \quad x = 0, 1, \ldots, n, \;\; 0 \le y \le 1.$$

It can be shown (not required) that for fixed values of a, b, n, the conditional distributions are Binomial(n, y), and Beta$(x+a, n-x+b)$. You may find it helpful to read (see Blackboard)

- Casella, G. and George, E.I. (1992). *The American Statistician* **46**(3), 167–174.

(a) Provide your own `R` code to perform Gibbs sampling with $n = 16$, $a = 2$, and $b = 4$.

(b) Take a sample of size 20,000 from the joint density and provide an estimate of the mean and variance for each of the marginal densities.

(c) The marginal density of x can actually be found analytically (exactly), and is a beta-binomial distribution, with density

$$f(x) = \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(a+b+n)}, \quad x = 0, 1, \ldots, n.$$

This is a compound distribution. You can sample from this distribution in a straightforward way. Explain how you could do so by first sampling from a beta distribution and then sampling from a binomial distribution. Write some `R` code to do this and generate 20,000 samples from the beta-binomial distribution with $n = 16$, $a = 2$, and $b = 4$.

(d) Give a Q-Q plot of the sample quantiles for x against the sample quantiles in part (c) found by generating samples directly from this beta-binomial distribution.