STAT 5443: Homework 3

Due: 17th March 2017

1. This homework has only one problem worth 30 points and a (harder) *bonus problem* for **5 points extra credit** to your total score. Obviously, answering the bonus problem is completely optional.

2. Hand in your HW (including print outs of your source code) at the beginning of the class on 15th March, 2017. Additionally source code (if any) should be emailed to `stat5443.fall@gmail.com` **before** the assignments are submitted in the class.

3. Any attempt at academic dishonesty will automatically result in 0 points.

4. **No late submissions will be accepted unless you have prior permission from the instructor.**

PBT Example Recall the Proportion of Bicycle Traffic Example in class. In Spring 1993, a survey was done on bicycle and other traffic near UC Berkeley (Gelman, 1995). 10 city blocks on residential streets with bicycle routes were chosen and on one of those blocks n vehicles were observed on a Tuesday afternoon 3-4 PM, and S of them were bicycles. Our parameter of interest is the proportion of bicycle traffic (PBT) θ on a similar block at a similar time. One way of fitting a Bayesian model to this data would be to model:

$$S \mid \theta \sim \text{Binomial}(n, \theta), \text{ and } \theta \sim \text{Beta}(\alpha_0, \beta_0) \tag{1}$$

We can choose the hyper-parameters α_0, β_0 from our previous knowledge and set them to be $\alpha_0 = 2.0$, $\beta_0 = 6.4$. For a given n and s, we can easily calculate the posterior as:

$$p(\theta \mid s) \propto \theta^{\alpha_0+s-1}(1-\theta)^{\beta_0+n-s-1} \tag{2}$$

$$\theta \mid s \sim \text{Beta}(\alpha_0+s, \beta_0+n-s) \tag{3}$$

We know how to sample from the posterior since it is a known density, but the marginal distribution is not a common distribution. We can derive the marginal distribution of S to be:

$$f(s) = P(S=s) = \binom{n}{s} \frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \frac{\Gamma(\alpha_0+s)\Gamma(\beta_0+n-s)}{\Gamma(\alpha_0+\beta_0+n)}, s=0,\ldots,n. \tag{4}$$

This is called the Binomial-Beta distribution and our goal is to estimate the pmf of S. Assume that we have observed the data: $n = 74, s = 16$. Now answer the following questions:

Problem 1 (30 **pts)** Implement the Gibbs sampler for the model in (1) for generating bivariate samples from the joint density of (s, θ). The steps are given below:

1. Set initial values $(s^{(0)}, \theta^{(0)})$. One possibility is $s^{(0)} = s, \theta^{(0)} = \hat{\theta}_{mle} = s/n$.
2. For $i = 1, 2, \ldots, M$ (M large integer).
   (a) Sample $s^{(i)} \sim f(s \mid \theta^{(i-1)}) = \text{Binomial}(n, \theta^{(i-1)})$.
   (b) Sample $\theta^{(i)} \sim f(\theta \mid s^{(i)}) = \text{Beta}(\alpha_0 + s^{(i)}, \beta_0 + n - s^{(i)})$
3. This will produce a bivariate sample of size M of (s, θ). Now, approximate the pmf of s with the values $s^{(i)}$, e.g.

$$f(s) = P(S=s) \approx \frac{\{\#s^{(i)} = s\}}{M}$$

Write down a full computer program for implementing the Gibbs sampler described above and plot the following three things: 1) trace plot for the $s^{(i)}$ samples, 2) trace plot for the $\theta^{(i)}$ samples and 3) the histogram for the $s^{(i)}$ samples. Also, estimate the posterior median of θ based on the samples drawn - is it close to the maximum likelihood estimate s/n? How sensitive is the posterior median to the choice of initial values?

Bonus Problem: Extra Credit 5 points Implement the Gibbs sampler like the one above but treat n as an unknown parameter with $\pi(n) = \text{Poisson}(\lambda)$ (Assume $\lambda = 64$). Draw the histogram of $s^{(i)}$ samples and compare your new posterior median with the one you obtained in problem 1. Do you get similar results as before? Derive the three full conditional densities needed for the Gibbs sampler.

Hint: The conditional distribution of $(n \mid \theta, s)$ is also a Poisson.