PubH 7440: Introduction to Bayesian Analysis
Spring 2020, Homework 2
Due: February 20, 2020

1. (Gelman, Chapter 3, Problem 3): An experiment was performed on the effects of magnetic fields on the flow of calcium out of chicken brains. Two groups of chickens were involved: a control group of 32 chickens and an exposed group of 36 chickens. One measurement was taken on each chicken, and the purpose of the experiment was to measure the average flow μ_c in untreated (control) chickens and the average flow of μ_t in treated chickens. The 32 measurements on the control group had a sample mean of 1.013 and a sample standard deviation of 0.025. The 36 measurements on the treatment group had a sample mean of 1.173 and a sample standard deviation of 0.20.

    a. Assuming the control measurements were taken at random from a normal distribution with mean μ_c and variance σ_c^2, what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. Assume that $p(\mu,\ \sigma^2) \propto \frac{1}{\sigma^2}$.

    b. What is the posterior distribution for the difference: $\mu_t - \mu_c$? To get this, you may sample from the independent posterior distributions you obtained in part (a). Plot a histogram of your samples and give an approximate 95% posterior interval for $\mu_t - \mu_c$.

2. (Gelman, Chapter 3, Problem 9): Suppose *y* is an i.i.d. sample of size *n* from the distribution $N(\mu, \sigma^2)$, where the prior distribution for (μ, σ^2) is defined as $\mu|\sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{\kappa}\right)$ and $\sigma^2 \sim InvGamma(\nu_0, \sigma_0^2)$. This is the $N - InvGamma\left(\mu, \sigma^2 \middle| \mu_0, \frac{\sigma_0^2}{\kappa};\ \nu_0, \sigma_0^2\right)$. Show that the posterior takes the same form and write it in terms of the sufficient statistics.

Table 1: Data for Intersections from streets that are bike routes (the left two columns) and not bike routes (the right two columns)

streets that are bike routes		**streets that are not bike routes**	
Intersection	**Number of Bicycles**	**Intersection**	**Number of Bicycles**
1	16	1	12
2	9	2	1
3	10	3	2
4	13	4	4
5	19	5	9
6	20	6	7
7	18	7	9
8	17	8	8
9	35		
10	55		

3. Consider the data in Table 1, which are on the number of bicycles at intersections for streets that are and are not bike routes. The goal is to analyze the mean rate of bicycles for each intersection (separately for the bike-route and non-bike-route streets) using the following model:

$$Y_i|\lambda_i \sim Poisson(\lambda_i)$$

$$\lambda_i \sim Gamma(\alpha, rate = \beta)$$

$$\alpha \sim Gamma(0.01, rate = 0.01)$$

$$\beta \sim Gamma(0.01, rate = 0.01)$$

Where:

$$p(y_i|\lambda_i) = \frac{(\lambda_i)^{y_i} e^{-\lambda_i}}{y_i!}$$

and

$$p(\lambda_i|\alpha, \beta) = \frac{\lambda_i^{\alpha-1} \beta^{\alpha} e^{-\beta\lambda_i}}{\Gamma(\alpha)}$$

a. Provide BUGS code to fit the model described above, specifying a separate hierarchical model for the bike-route and non-bike-route intersections.

b. Using the data and initial values provided on the canvas site, run three parallel chains for 4000 iterations each and provide traceplots for the four population-level parameters (i.e. α and β for bikes and non-bikes). Comment on what these plots tell you about the convergence of the chain.

c. Throw out the first 4000 iterations and run the chain for 2000 additional iterations. Provide summary statistics for the four population-level parameters, as well as the intersection-specific rate parameters.

d. Edit your BUGS code to test if the posterior mean of the population-level bike-route intersection rates is larger than the non-bike-route intersection rates.

e. Edit your BUGS code to provide the predictive distribution for the number of bikes observed at a NEW bike-route intersection and provide summary statistics and a plot of the predictive density.

4. Consider the following mixture density:

$$f(x) = \frac{1}{3}\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x+5)^2} + \frac{1}{3}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} + \frac{1}{3}\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-5)^2}$$

This distribution will have mean 0 and variance 17.67. Write code (in R or your preferred statistical programming language) to approximate the mean and variance of $f(x)$ using importance and rejection sampling, using a normal candidate density in both cases. Approximate the mean and variance using 200 draws from the target density.