STAT 532 Assignment 8 -2015
Due: Friday, Oct 30 (by 4:15 pm)

Show all work **neatly** and **in order** for full credit. Same instructions as previous homeworks...ask if you need clarification.

1. Go through the logit-normal code for the model we went over in class. Recall the Beta-Binomial model we used as an example for rejection sampling. Let's now use the data from the rejection sampling example to make inference about the rate of death from stomach cancer for at risk males between the ages 45-64 for the largest cities in Missouri (data from Tsutakawa et al. 1985). The data are provided in the rejection sampling code.

   (a) (8 pts) Fit the logit-normal model (via Gibbs sampling) to the stomach cancer data and compare the results to those obtained from the beta-binomial model (via rejection sampling). Make sure to specify your priors completely, justifying choice of hyperparameter values as much as possible. Use plots to compare "results", and briefly discuss what you find.

   (b) (8 pts) Implement a posterior predictive check to help compare the use of the two models, and briefly discuss the results and how you would use the checks.

   (c) (2 pts) Which model would you choose for inference? Why?

   (d) (6 pts) Fit the logit-normal model in JAGS or BUGS or Stan. (Include your model code in the answer to the question.)

   (e) (4 pts) Compare the canned software package results to those obtained from the Gibbs sampler "we" programmed.

   (f) (Extra credit) Fit the beta-binomial model in JAGS or BUGS or Stan. Compare to other results.

2. (10 pts) Read the Gelman's 2005 paper entitled "Analysis of Variance - Why it is more important than ever." I do not expect you grasp everything in the paper on the first time through.

   (a) (10 pts) Take some notes while you are reading it and provide at least two questions about the paper here (these should reflect your biggest questions, not just your first two questions).

   (b) (10 pts) Write R code to simulate data from a statistical model that could come from the design and type of response variable found in the consulting project Kenny and Leslie are working on this semester. (We will go over the details in class). Be thinking about "batches" as you do this.

(c) (5 pts) Explicitly write out the model you used to simulate the data (you may do this before, during, or after) part (b). Sometimes thinking about how to simulate data can help you write work through the details of writing down a model with all its needed parts.

3. Chapter 12 discusses ways to make Markov chain simulation more efficient. You should be aware of this information, though we will not go through all of the details. To make Gibbs samplers more efficient, we can use transformations, reparameterizations, incorporation of auxiliary variables (data augmentation), and parameter expansion. Here, we will look at auxiliary variables and parameter expansion (pages 293-295 in the text). Reparameterization, parameter expansion are two strategies that can help improve convergence.

   (a) (2 pts) Explicitly write out the model of interest in the Example that starts on page 293.

   (b) (4 pts) Implement the Gibbs sampler described on page 294 for the data augmentation approach for obtaining the posterior distribution of μ and σ.

   (c) (4 pts) Implement the Gibbs sampler described on page 295 for the parameter expansion approach for obtaining the posterior distribution of μ and σ.

   (d) (3 pts) Compare convergence for the two approaches and briefly discuss.

4. *Hamiltonian Monte Carlo*:

   (a) (3 pts) Read Sections 12.4 and 12.5 in the text. Tell me if you read it. You don't have to pour over all the details, but try to get the general idea of what is going on.

   (b) (8 pts) Complete the *Introduction to HMC Exercise* (Files: `Intro_HMC_Exercise.pdf` and `roll.R`). Turn in the functions you wrote. Did this help you understand the material in Sections 12.4 and 12.5 (briefly explain).

   (c) (6 pts) Look through and run the examples provided in the Introduction Michael Lerch created last year. (Files: `stan_example.R`, `sample_cens.stan`, `sample_lm.stan`, `sample_normal.stan`). Include a couple of comments/questions here.

5. (20 pts - will be a grade on its own and due by November 11) Write up an approximately one-page reference for your assigned convergence diagnostic to share with your fellow classmates as a quick reference. Most importantly, it should be understandable and usable by your peers in the future. You may also research other criteria not listed here, just be sure to let me know if you would like to do this. Use the work of previous students as a starting point and keep previous names in the author list (add yours).

   Names in bold should take the lead on the diagnostic, since names not in bold are assigned to two diagnostics.

   - PSR ($\widehat{R}$): ***Doug***, *Leslie*
   - n effective: ***Garland***, *Leslie*
   - Geweke's diagnostic: ***Matt***, *Jordan*
   - Raftery and Lewis' diagnostic: ***Claire***, *Jordan*
   - Brooks and Gelman's diagnostic (multivariate): ***Brandon***, *Wilson*
   - Heidelberg and Welchs' diagnostic: ***Allison***, *Kenny*
   - Quantile Equivalence Diagnostic (Michael Lerch's): ***Chris***, *Kenny*
   - Edward Boone's: ***Jake***, *Wilson*