STAT 532 Midterm II - Fall 2015
Due: Thursday, November 12 by 10:00 a.m.

Work independently using only your brain, the Gelman et al. text, *your* notes/handouts, *your* homeworks, and R. If you have questions about resources please ask *before* using them. **Show all work neatly** (and in order). Plots should appear with the corresponding exercise, and only include computer code/output necessary to completely answer a question (i.e. before including R-code ask yourself - would I want to see this if I was grading this?). Other well organized code and output can be included in an appendix as usual, but is not required.

1. Data are collected from five study units. The number of "successes" is counted for each study unit, but the number of Bernoulli trials associated with each count is unknown. Researchers would like to make inference about the probability of success *and* the number of trials using the Binomial(m, π) distribution. (Each observation is the number of successes observed from an unknown number of bernoulli trials, though we assume the number of Bernoulli trials going into each count is the same and the probability of success is the same.)

   (a) (2 pts) Make up a research scenario to go with this set-up. Describe it in a couple of sentences to give yourself a context for the problem.

   (b) (4 pts) Suppose researchers observed the following data: $\mathbf{y}_1 = (16, 18, 22, 25, 27)$. Plot the likelihood function (For convenience, for all parts of this problem you may use a contour plot even though m is technically discrete). From your plot, approximate where the MLE is for (m, π). You may actually find the MLE analytically or computationally, but it is not required as long as your approximation via the plot is not too far off. Add a point to the plot showing your approximation of the MLE.

   (c) (2 pts) The researchers published results for m and then years later discovered a mistake in data entry so that the data should have actually been $\mathbf{y}_2 = (16, 18, 22, 25, 28)$. Again, plot the likelihood function and approximate the MLE for (m, π) using the updated data set.

   (d) (3 pts) Compare the results and discuss this in the context of stability of MLEs for m.

   (e) (3 pts) Now, implement a Bayesian analysis with the following prior: $p(m, \pi) = p(m)p(\pi)$ where $\pi \sim Beta(1, 1)$ and $m \sim Poisson(\lambda = 100)$.

      i. Display the joint posterior distribution of (m, π) and briefly compare to the likelihood functions from above.

      ii. Show work for deriving complete conditionals to be used to program your own Gibbs sampler

      iii. Implement the Gibbs Sampler (show organized code)

   (f) (4 pts) Display the marginal posterior distributions for m and π and compare results to the likelihood analysis. Find the posterior probability that $m > 100$ and $\pi < 0.3$ ($Pr(m > 100, \pi < 0.3|y)$)

   (g) (3 pts) What are the disadvantages of the prior specification in the above model?

   (h) (3 pts) Write out a hierarchical version of the model (i.e., put priors on pi and λ with unknown hyperparameters). Use hyper-prior distributions meant to reflect no prior knowledge, and explain your particular choice of prior.

   (i) (6 pts) Obtain draws from the joint posterior distribution using your method of choice (program your own algorithm or use available software). Show your work.

(j) (3 pts) Compare the marginal posterior distributions for p and m obtained from the hierarchical version to those displayed in part (f). Also, compare the posterior distributions of the hyperparameters to the fixed values used in (e).

(k) Raftery (1988) considered a hierarchical approach assigning $m \sim Poisson(\mu)$ where μ is unknown. To define a joint prior distribution, he defines $\mu = \lambda\pi$ and specifies a prior distribution on (μ, π). because "'it would seem easier to formulate prior information about μ, the unconditional expectation of the observations, than about μ, the mean of the unobserved m. He used $p(\mu, p) \propto 1/\mu$.

  i. (2 pts) Speculate on his reasons for suggesting this prior.

  ii. (3 pts) Transform to look at the prior induced on $p(m, \pi)$ and compare it to the prior you used above.

2. Suppose researchers would like to study the dose - mortality relationship for a specific toxin and a species of frog that is difficult and expensive to get for laboratory studies because of where it lives and the fact it is a threatened species. Researchers managed to obtain 20 frogs. They know the toxin is lethal to the frogs, but do not have a good understanding of a finer relationship between dose and the probability of death.

   They randomly assigned one of 4 doses to each frog, so that 5 frogs were in each dose. The 5 frogs in each treatment were held in two tanks, with 2 frogs in one tank and 3 frogs in the other tank. Dose is measured in grams per ml (g/ml), they were shooting for 0.40, 0.75, 1.0, and 2.0, but after taking measurements after applying the dose, they corrected the doses to better reflect what was actually applied (averaged multiple measurements from the two tanks per treatment). Data are in the file `FrogDoseData.csv`.

   (a) (4 pts) Write out a simple logistic regression model to estimate the dose-death relationship. Use glm() to estimate the relationship between dose and the probability of death using traditional logistic regression. Graphically display the fitted model and estimate the dose at which half of animals are expected to die (LD50).

   (b) (3 pts) Write out a model to do a similar analysis in a Bayesian framework, using the prior we used for the bioassay experiment example.

   (c) (8 pts) Fit your model in Stan. Provide model code here (and other code if needed to see all information related to your choice of prior). This should match the model you wrote out in part (b). See the Stan *Modeling Language User's Guide and Reference Manual* posted at `http://mc-stan.org/documentation/`. In particular, Section 5.4 starting on page 49 should be helpful.

   (d) (3 pts) Compare results to those obtained from glm() and briefly discuss.

   (e) (4 pts) Compare the results between the Bayesian model and the traditional logistic regression model. Discuss and attempt to explain any differences.

   (f) (4 pts) To help in specifying "automatic"' priors with minimal influence on the posterior, people suggest standardizing the inputs to the regression. Does standardizing your input variable appear to change inference or efficiency of the sampler?

   (g) (3 pts) Cauchy priors for the regression coefficients have been suggested for use as "automatic" weakly informative priors. The suggested parameterization is a Cauchy(mean=0, scale=10) for the intercept and a Cauchy(mean=0, scale=2.5) on regression coefficients for explanatory variables (that are centered and scaled to have a standard deviation of 0.5). Briefly outline why these might be considered weakly informative priors.

(h) (3 pts) Re-fit the model with these Cauchy priors. Display results and discuss the results relative to previous results. To display the posterior uncertainty, plot the regression lines from many posterior draws in light gray. Include a vertical line at the posterior mean for LD50.

(i) (3 pts) Re-fit the model one more time with a more informative prior. Display the results and briefly discuss in the context of data such as found for this problem.

(j) (2 pts) Which of the three analyses do you think is most appropriate? Briefly justify your answer.

(k) (8 pts) Discuss the assumption of independence (i.e. speculate on possible sources of dependence in the design). And, **clearly** describe how you could implement a posterior predictive check to assess a possible violation of independence due to the design. (You do not have to do it for this problem because these data make it particularly hard. It might help to think about having the following data following order in actual data set: c(1,0,0,0,0, 0,0,0,1,1, 1,1,1,0,0, 1,1,1,1,1)).

(l) (4 pts) Clearly write out a full Bayesian model you might use to attempt to account for the source of dependence you identified in the previous part.

3. It is quite common for researchers to want to specify "non-informative" or "uninformative" priors. This is typically done use "vague" or "diffuse" prior distributions.

   (a) (2 pts) Theses distributions are, however, providing information about the relative probabilities or densities of parameter values *a priori*, so in what sense are they supposed to be *non-* or *un*-informative? (Max 1/4 page)

   (b) (3 pts) Suppose you are reviewing a paper and the authors state "We chose common non-informative priors for all our variance parameters and therefore did not need to check sensitivity of the posterior to choice of prior." Formally respond to the authors regarding the statement as if you are the reviewer (in a respectful and professional way). (Max 1/4 page)

4. (3 pts) To specify proposal distributions to use within rejection or Metropolis-Hastings algorithms, we have used information from the likelihood $p(y|\theta)$. Briefly describe why is this okay to do for a proposal distribution and not okay for a prior distribution. (Max 1/4 page)

5. (3 pts) Today, the buzz is all about "big data", but there will always be situations that will arise where data are so expensive that "small data" is still a pertinent issue. Briefly explain to another graduate student (not taking this class) why and how Bayesian analysis can be helpful in "small data" situations. (Max 1/4 page)