Exam 2

Stat 451

3/10/2022

This is a take home-exam. The exam is due at the beginning of class on Tuesday, March 15^{th}. Please do not consult with any living person about this exam.

The first part of the exam uses the data file **mortality.dat**, and will be worth 75% of the grade.
In the data set, the first column is an identifier for the hospital. The second column is the number of operations taking place in the neonatal unit over a one year period. The third column is the number of operations that ended with the death of the patient. The idea behind the study is to rank the performances of the hospitals involved. I assume you will do the problem using JAGS. Please include relevant code with your answers.

1. Presumably, an appropriate way to rank the hospitals would be to model the proportion of operations that end in the death of the patient (p_i, where i indicates the i^{th} hospital). The maximum likelihood estimates for those proportions would simply be the number of deaths divided by the number of operations. Compute the MLE for each hospital.
2. Now, perform a Bayesian analysis where the proportion of deaths is modeled as a fixed effect. That is, you should not use a hierarchical model. Use a beta(1,1) prior for the p_i's. The answer for this problem is just the standard JAGS output.
3. Demonstrate that the parameters you are estimating in this model have converged. Show effective sample sizes, autocorrelation, and the Raftery-Lewis diagnostic.
4. Now perform a Bayesian analysis where the proportion of deaths in modeled as a random effect. That is, now use a hierarchical model for the p_i. Use a beta(α,β) prior for the p_i's and gamma(1.1,1.1) for α and β. The answer for this problem is just the standard JAGS output.
5. Demonstrate that the parameters you are estimating in this model have converged. Show effective sample sizes, autocorrelation, and the Raftery-Lewis diagnostic.
6. Produce a table with the MLE's, the point estimates from the fixed effects model, and the point estimates from the random effects model side-by-side.
7. Using the point estimates, which hospital seems to be the worst?
8. Using the point estimates, which hospital seems to be the second worst?
9. Using the random effects model, would you say those two hospitals have significantly different proportions or neonatal mortalities at the 0.05 level?
10. Using the random effects model, which of these two hospitals would you prefer for your own family for neonatal surgery?
11. (Bonus) Using the random effects model, rank the hospitals by proportion of deaths (lowest proportion has rank 1) at each draw. Then produce a mean rank and standard deviation of that rank for each hospital.

For the last set of problems, you have been brought data by an anthropologist. She has found five adult skeletons of ancient humanoids. She has been studying the ratio of the length of humerus (upper arm bone) to the length of the femur (thigh bone) in primates. For monkeys she knows that ratio is about 0.95. For

modern man, the ratio is about 0.72. The ratios for the five adult skeletons she has found are as follows: 0.857, 0.824, 0.820, 0.875, 0.844. She is interested in the probability that the population mean ratio for the group of people whose skeletons she has found is between 0.80 and 0.90. For your likelihood, you should be aware that the appropriate support for these data is between 0 and 1. So we will use a beta likelihood. You should also know that these ratios range from about 0.70 to 0.98. You can use that information to inform your prior. The parameters of a beta must be positive, so we will use gamma priors for the parameters. If the beta is parameterized with (α,β), then use a gamma(shape=1,rate=.2) for α, and a gamma(shape=1.5,rate=1) for β.

12. Since you know the likelihood and the prior distributions for the parameters, you can draw values from the prior predictive. Draw 1000 values from the prior predictive and plot a histogram of the values drawn from the prior predictive. Remember that the values must be between 0 and 1.

13. Write code in Stan to address the problem. Use 20000 burnin and 100000 iterations, thin by 10 and produce 5 chains. This will give you 40000 MCMC draws of the posterior. Print the summary of the simulation. Also include your Stan code in the R script as comments. This will make it easier for the grader to identify any mistakes and give feedback.

14. Verify that the chains you have produced have converged appropriately and have enough information to use to make inference by examining (and reporting) the effective sample size. Effective sample sizes should exceed 5000.

15. Verify that the chains you have produced have converged appropriately and have enough information to use to make inference by examining (and reporting) the Raftery-Louis diagnostic. Raftery-Louis diagnostics should be less than 3.

16. Plot the posterior density of the mean ratio which would be computed as $\alpha/(\alpha+\beta)$.

17. What is the probability the ratio for the population mean given the data from these five skeletons is between 0.80 and 0.90?