STA 365 Applied Bayesian Statistics

Homework 1
Tuesday January 15, 2013
DUE: THURSDAY January 31, 2013

For the below homework, please show your work. You should be showing at least some examples of your computer code (R code for example) if you use the computer to compute some things. When you are asked to explains something, you should not be afraid write.

1. Suppose there is a box with 6 coins in it. The coins have different probabilities of heads. They are 0/5, 1/5, 2/5, 3/5, 4/5, and 5/5.
   (a) Suppose that one of the coins is draw (with equal probability), the coin is flipped and a heads is observed. What is the posterior probability of each coin that it was that coin that was drawn.
   (b) Suppose that the same coin is flipped 3 more times. The results of the 4 flips is 3 heads and one tail. Now what are the probabilities of each coin being the coin that was selected.
2. Please note the following definitions: The prevalence of a disease is the proportion of individuals in a population or subpopulation who have a disease. The results for a diagnostic test for a disease will be called positive if the test classifies the subject as having the disease and negative if it does not. Please note that the test can be wrong and it can classify a subject as having the disease when the subject does not and vice a versa. The sensitivity of a diagnostic test is the probability of the test result will come up positive (for the present of a disease) if the person does indeed have the disease. (This could be signified by $P(T^{+}|D^{+})$ or called the "true positive" rate.). The specificity of a disease is the probability that the test will give a negative result if the subject does not have the disease. (This could be signified by $P(T^{-}|D^{-})$ or called the "true negative" rate.). Answer the following questions:
   (a) A doctor is concerned that a patient has Disease X. Since the patient has several symptoms, the doctor knows from the literature that there is a 25% chance that the patient has Disease X. So the doctor orders Test A. This test has a sensitivity of 90% and a specificity of 80%. If the patient has a positive result from this test, what is the probability new probability of the patient having Disease X?
   (b) Now, with the Test A positive, in the above question you know what the present probability of the patient having Disease X. However, the doctor was not satisfied with this probability and would like to be more certain. So, the doctor order another test, Test B which has a sensitivity of 70% and specificity of 90%. If this test also comes back positive, what is the probability that the patient has Disease X?
   (c) The above sequence of test is fairly common. The first test of this nature is often called the screening test and a second test like the above test is used to, hopefully, confirm the diagnosis. Please discuss what properties might make a good screening test and what might make a good confirmatory test and why.
   (d) suppose there is a family practice physician who decides to test all his patients for Disease X. For males in the general population, the population prevalence of Disease X is about 1 in 25 000. If a subject had a positive Test A for Disease X, what is his probability of actually having Disease X?
3. In this question, we will consider developing priors and using these priors to calculate posterior distributions. The main subject of this investigation is the birth weight of babies. There has been a

lot of work looking at birth weight as a health indicator. For simplicity, assume that the distribution of birth weights is approximately normally distributed with mean μ and precision τ. Therefore, the purpose of this question is to learn about the values of μ and τ.

Do the following:

(a) (Priors) Write down two sets of priors. For the first prior, choose a very non-informative prior. For the second prior, consider a somewhat "honest vague" prior. In considering this second prior, please consider information about what you know about birth weights. That is, consider the value of birth weights that you know about. Perhaps what the value was for yourself and other members of your family. (If you don't know anyone, some weights of people that I can remember are 9, 8.5, 6.5, and 7, all in pounds). Also, you might want to know that 1 pound = 16 ounces = 453.6 grams.

When developing your priors, please explain your reasoning for the priors that you come up. Note that there are many different "correct" answers, but you need to be clear in your explanation.

(b) Assume that for a sample of 30 birth weights (in grams), we have $\bar{X} = 2750.63$, $\sum(X_i - \bar{X})^2 = 421721$. Give the marginal posterior distribution of the parameters of the model for birth weight given these sampled values. Use the two different priors that you specified above. When specifying these distributions, calculate the posterior mean, standard deviation, and a 95% credible region for these parameters under each of the two priors. (Note: the standard precision of the t-distribution is not 1/variance.)

(c) Generate a histogram of values for the parameters of your model from the marginal posterior distribution. (That is, generate samples from the joint posterior distribution of the parameter and then provide the histogram for each of the parameters separately.)

(d) Generate 400 new sampled values of the predictive distribution. Give a histogram of these values and simple summary statistics of these 400 sampled values (such as mean and standard deviation.) (Hint: First generate a parameter value from your posterior, then generate a new sampled value from your probability model with the new parameter value which was generated. When providing your answer, you should be providing at least a sketch of the computer program used if not the actual R code.)

4. Preamble: You have been hired to predict the results of an election. The election is for town selectperson. There are three selectpersons who make up the the town council and who run the town. There is a division between the purple party and the brown party. If one side has a majority, they will control the issues in the town. Each selectperson is elected in one of three town districts and both parties have candidates running in each district. If a candidate wins a majority of votes in a district, that candidate will be the selectperson for that district. Also, assume that there will be 5001 citizens voting in each district.

(a) Prior distributions. Create two different analyses. One based on a "non-informative prior" and the other on an informative prior. Justify your answer. For the informative prior consider the following information. In the past, the two parties have been quite balanced. In the past, each party would get between 40% to 60% of the votes. Use this information to create an informative prior.

(b) Posterior distribution for the voting percentage for each district. For each district, a simple random sample of the citizen of each district is asked whom they plan to vote for. In district one: 85 said purple and 65 said brown; in district two: 70 said purple and 80 said brown; and in district three: 50 said purple and 100 said brown. Calculate the posterior probability for the percent who will vote for purple in each district. In reporting your results, provide the distribution of the posterior and the value of the parameters of the parameters and also report appropriate statistics for these distribution (which includes the posterior mean, standard deviation, and some kind of 95% interval). (Note: do this for each of the two priors specified in the first part.)

(c) Given the above information, provide the probability that the purple party will have a majority in the town council. Also, provide the probability that purple will win each of the districts. Do this by simulation. That is, simulate 10000 elections. For each simulated election, generate a probabilty, θ_i, which is the probability of a citizen voting for the purple party in district i. Then, generate the number of citizens voting for purple in each district. From there, one can elect either the purple candidate or the brown candidate for each district for each simulated election. Finally, one can see which party had a majority in the town council in the simulated election. (Note: do this for each of the two priors specified in the first part.)