Course EPIB-675 - Bayesian Analysis in Medicine

Assignment 3

1. The course web page provides a function for doing Bayesian analysis of a normal mean. Using similar ideas, here you will create your own R function that does Bayesian analysis for beta prior binomial likelihood functions. I suggest that you proceed along the steps outlined below, but you just need to hand in a printout of your final program and the answer to part (e).

(a) Start by creating an R function that takes as input the information needed to provide the posterior distribution given beta prior parameters and number of successes and trial of the binomial data. Use parameter names that will help the user of your program understand the inputs required. The output should be the parameters of the beta posterior distribution.

(b) Add some graphics to your function, in particular the beta prior curve and the beta posterior curve. Here are some hints:

- Note the difference between the plot command (to start a new graph) versus points (to add points to graph). Note also the use of the lty option for making different line types, and the type option for making continuous lines (e.g., lty="l").
- To avoid cutting off the tops of the graphs, think about which order you want to make plots. Starting with the tallest plot (i.e., the posterior) may be a good idea.
- Use the range of (0,1) for your x-axis in all graphs. [A more sophisticated function might first examine all curves in the tri-plot, and automatically decide on a range based on this information . . . I will leave this as an optional exercise.]

(c) Add a 95% credible interval to your output. (Hint: Investigate the use of the qbeta function, which gives quantiles of the beta density. For example perhaps using qbeta(0.025, alpha, beta) is useful for the lower limit, etc.

(d) Test your function on a trial data set and trial prior distribution.

(e) Run your program with a beta(5, 25) prior, and a data set with $x = 10$ successes in $n = 70$ trials. Print out all of the outputs (posterior, plot, and 95% interval).

2. Suppose we have a parameter θ that has a posterior distribution that is beta(20, 100). What is the probability that θ is between 0.10 and 0.20? The answer is of course given by the definite integral

$$Pr\{0.10 < \theta < 0.20\} = \int_{0.10}^{0.20} \frac{1}{B(20, 100)} \theta^{20-1}(1-\theta)^{100-1} d\theta$$

We will investigate two different ways to solve this problem.

(a) Use the R function "integrate", to directly integrate the above function, which is a dbeta(20,100) function. [Hint: The help on integrate (type ?integrate from the R command line) shows examples from the normal density, follow the example but change the density to a beta.]

(b) The R function pbeta gives you the probability of being less than any given value for any beta density. Use the difference between two pbeta's to solve the integral. Check that your answers in (a) and (b) are very similar.

3. In this problem we will use Monte Carlo integration to find the variance of a beta density. Suppose we have a beta(50, 30) density. What is its variance? We will do this in three ways:

(a) First, calculate the variance "by hand", using the formula for the variance of a beta density,

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Of course, you can do this calculation in R, or any hand calculator, etc.

(b) Next, remember that the variance is just an integral, so the R command integrate can be used, similar to problem 2 above, but with a different definite integral, defined as a beta density (using dbeta as before), but with an extra term in front, $(\theta - 0.625)^2$, where 0.625 is the mean of this beta density.

(c) Finally, this problem can be solved using Monte Carlo integration. Here, the R command rbeta can be used, and see the class notes for hints on how to calculate variances using Monte Carlo integration.

(d) Compare your answers from (a), (b), and (c) to ensure that they are all the same (or at least, very close).

4. You have just gone shopping, and received a quarter in change from the cashier.

(a) Assume that your prior probability that the coin will come up heads in any given toss can be expressed by a beta distribution with appropriately chosen α and β parameters. State your prior distribution. (Note: There is no "correct" answer, since each individual will have their own prior distribution. However, you should justify your answer in terms of your prior mean and variance (or standard deviation), that is, check to ensure that the values of α and β give reasonable means and variances. You may wish to imagine a 95% probability interval, and consider that the mean is in the center of that interval, and that four times the standard deviation will equal the length of that interval.)

(b) Suppose that the coin is now tossed 5 times, and there are no heads. Use your program from question 1 of this assignment to calculate the posterior probability for the probability of heads for that coin and provide a 95% credible interval.

(c) If you were using a frequentist approach to analyse the same data (i.e., five tails in a row), what would the 95% confidence interval be? [Note: the normal approximation does not work well here, so you may consider using an "exact" approach. There is an R program for exact binomial inferences on the course web page.]

(d) Provide interpretations of the intervals you calculated in parts (b) and (c). Which of the intervals given in (c) or (d) do you prefer? Why?

5. Two researchers are looking at preliminary results for a new surgical technique. One researcher was very optimistic about the new technique, and had a prior distribution for the success rate of the technique of beta(40,10). The second researcher was less confident in the new technique, but also less sure of his opinion, and so had a beta(5,5) prior. In the first ten trials of the use of this technique, the surgery was successful 7 times.

(a) What are the two researchers posterior distributions?

(b) Plot the two posterior distributions on the same graph in R.

(c) Using Monte Carlo integration, compare the two posterior distributions by calculating the probability that the rate from the first researcher is greater than the rate of the second researcher.