Course EPIB-668 - Bayesian Analysis in Medicine

Assignment 2

1. Researchers seek to estimate the efficacy of a novel form of neutron therapy in treating bladder cancer using survival at 12 months post treatment as the outcome of interest. The probability of survival to twelve months for such patients with pelvic cancer treated with conventional therapy has averaged approximately 20%. Although the researchers do not hold strong opinions, they believe that the new form of neutron therapy may double the probability of survival to 0.4, and that the chances of the success rate being above that value are about as likely as it being below. They also feel that values closer to 0.4 are more likely than those further from 0.4.

(a) Given the above, indicate what seems to be a relevant conjugate prior for the probability of survival past twelve months.

(b) Using your prior distribution and the fact that that a small sample of 9 patients has now been treated with neutron therapy and 3 of those patients survived past 12 months, determine the posterior density.

(c) Provide the mean and variance for your posterior distribution from part (b).

2. We will continue discussing the situation of the first question, but now with a larger data set.

(a) Suppose that instead of 9 patients, 90 patients had been studied, with 30 surviving past 12 months. What would the posterior distribution be?

(b) For large parameter values, the beta density can be reasonably approximated by a normal density with matching mean and variance. Using a normal approximation to the posterior density of part (a), find the probability that the true rate is greater than 25%. Show all of your calculations.

3. In this question, we will investigate *predictive distributions*, which are used to make predictions about values of future data.

(a) Suppose you are responsible for allocating hospital beds, and are trying to figure out how many beds you will need for surgical complications. You know that 50 operations are carried out in any given day. Suppose you assume the rate of complications is exactly 10%. What is the distribution of beds you will need? Provide the mean value and variance of this distribution.

(b) In part (a), we assumed that the exact rate of complications is known to be 10%, but there may be only less precise prior information about that rate. Suppose the director is about 95% certain that the true rate should be somewhere between 5% and 15%. Find a reasonable prior distribution that closely approximates this prior information. [Hint: You can use a beta density, with prior mean of 10% (i.e., 0.1), and prior SD of 2.5% (i.e., 0.025). This should come quite close to matching the available prior informaiton.]

(c) We can now redo part (a), but using the prior information of part (b) to make predictions, rather than the exact value of 10%. We can think of this as a two-step process: first, a rate of complications is selected from the prior distribution, and then, assuming for the moment that this rate is correct, there is a distribution for the number of beds needed. If we do this calculation many times, we get a *mixture* of binomial distributions, over the prior distribution of the rate. This mixture of distributions is called the *predictive distribution*. In mathematical terms, it can be written as:

$$\text{predictive distribution } = \int_0^1 binomial(50, \theta) f(\theta) d\theta$$

where θ is the rate, and $f(\theta)$ is the prior distribution on that rate. Note that if $f(\theta)$ is simply a constant at 10%, then we are back to the case of a standard (non-mixture) binomial distribution, as in part (a).

Using R, simulate a sample from the predictive distribution for the number of beds needed using the prior distribution you derived in part (b). Provide the mean and variance of this distribution, as well as a 95% interval. [Hints: First generate a sample (say of size 10,000) from the beta density, and then feed these values into the binomial density, using the beta values as the probability of success for the binomial. To get a 95% interval, take the resulting 10,000 random draws from the binomial, and use the quantile command with c(0.025, 0.975) to define the desired quantiles.]

(d) Compare the distribution of part (a) to that in part (c). How do the means compare? Which distribution is more dispersed? Which do you think is more realistic?

4. (a) Construct your normal prior distribution for the average age at which men get their first myocardial infarction (MI, i.e., a heart attack), among all men who do have MI's. There is no "correct" answer here, but you should justify your choice; this distribution should correctly represent your prior knowledge about ages at which men have heart attacks.

(b) Update your prior to a posterior distribution using the data given below. Assume that the variance of the age at first MI is known to be 144. Show all of your calculations.

68 52 62 67 68 56 53 81 36 77

(c) Find prior and posterior 95% credible intervals for the mean age.

5. Suppose there are 1000 cases of cancer observed in Quebec this year. Use a non-informative Gamma prior distribution combined with a Poisson likelihood function to derive the posterior distribution for these data.