Course EPIB-668 - Bayesian Analysis in Medicine

Assignment 4

1. The attributable risk in epidemiology is defined as the proportion of the incidence of a disease in an exposed group that is due to the exposure. In other words, it is the incidence of the disease in the exposed that would be eliminated if the exposure were eliminated. The attributable risk percent (AR%) is simply the percentage of the risk in the exposed group that is due to the exposure.

Consider the 2×2 table below:

Eat Fish	Stroke	No Stroke	Total
Never	82	1549	1631
Daily	23	779	802
Total	105	2328	2433

(a) What is the point estimate of the population attributable risk from the above data set?

(b) Assume uniform beta(1,1) priors for the probability of stroke in each of the never and daily consumption of fish groups. What is the posterior distribution for the probability of disease in each of these exposure groups?

(c) Using direct simulation programmed in R, find the mean and standard deviation, and provide a 95% credible interval from the posterior distribution of the attributable risk. Print out a histogram of your sampled values.

(d) Using direct simulation programmed in R, find the mean and standard deviation, and provide a 95% credible interval from the posterior distribution of the attributable risk percent. Print out a histogram of your sampled values.

2. A diagnostic test for detecting a certain parasite via serology is known to have good sensitivity, but poor specificity. The exact values of the sensitivity (S) and specificity (C) are unknown, but it is known that the sensitivity is

likely to be in the range from 80% to 90%, and specificity is likely to be in the range from 30% to 50%. When imperfect diagnostic tests are used, the probability of a positive test can be seen to be

$$
\begin{aligned}
P(+'ve\ test\ result) &= P(true\ positive\ result) + P(false\ positive\ result) \\
&= \pi S + (1 - \pi)(1 - C)
\end{aligned}
$$

Suppose we have a group of 100 subjects given this test, and that 40 of them test positive. The likelihood function then is (using the above probability in the binomial distribution):

$$
(\pi S + (1 - \pi)(1 - C))^{40} \left(1 - [\pi S + (1 - \pi)(1 - C)]\right)^{60}
$$

where π is the prevalence of the disease. Suppose that nothing at all is known about the prevalence of the parasite in this group of patients, such that a uniform beta(1,1) prior is appropriate. It is then rather easy to formulate the numerator (prior $\times$ likelihood) of Bayes theorem, simply by multiplying the three priors (for π, S, and C) and the likelihood function given above. However, the denominator is difficult, because we would need to do multiple integration over the expression in the numerator. Thus, this problem is a prime candidate for the SIR algorithm.

(a) Write down the full mathematical expression for the numerator of Bayes theorem, by writing out the formulae for the three beta distributions (actually, you can omit the expression for π, since it is just equal to one everywhere) times the above likelihood function.

(b) Create an R program that uses a SIR algorithm to find the posterior distribution for the prevalence, π. Remember the steps: Draw a sample from a proposal distribution (simplest to use the prior distributions for each of π, S, and C here), weight each value (if you use the prior, the weight is simply the likelihood), and then resample the original sample with the weights.

(c) Use the program you created in (b) to obtain a sample from the posterior distribution for the prevalence π. Print out a histogram for π, the mean and SD of the posterior for π, and a 95% credible interval.

Use WinBUGS to solve the rest of the questions on this assignment. You will redo three questions from a previous assignment, but now using WinBUGS. Examples similar to the WinBUGS programs you will need here are found in the course notes.

As in all WinBUGS programs, remember that it is good practice to run a number of "burn-in" iterations (to ensure convergence of the algorithm), and begin monitoring the parameters of interest after this. For the programs below, a burn-in of 500 iterations is plenty (although in other more complex problems, many thousands are often needed), and 5000 iterations after burn-in should provide excellent accuracy. These 5500 iterations should take just a few seconds to run on most computers.

In providing your answers, please include a printout of the program you used and all of the summary results from the monitored variables. Also provide on a separate sheet of paper the direct answers to the questions.

3. Redo example #1 from Assignment #2 using WinBUGS. Use the same prior distribution you used at that time. In particular, answer the following questions:

(a) 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 properties of the posterior density using BUGS. That is, state the posterior mean, SD, and provide a 95% credible interval.

(b) Suppose that instead of 9 patients, 90 patients had been studied with 30 surviving past 12 months. Determine the properties of the posterior density using BUGS. State the posterior mean, and provide a 95% credible interval.

4. Redo example #4 from Assignment #2 using WinBUGS. Again, use the same prior distribution you had there, and answer the following questions:

(a) Determine the properties of the posterior density for μ using WinBUGS. State the posterior mean, SD, and provide a 95% credible interval. Compare your posterior mean and 95% interval that you obtained previously to the mean and 95% credible set produced by WinBUGS. Remember that you did this problem with a known variance for the data of 144.

(b) Redo part (a) with an unknown mean AND unknown variance, using a non-informative prior for both parameters. In WinBUGS, one usually considers uniform priors for the SD as non-informative. Select an appropriate range for this SD, and defend your choice based on practical concerns from this problem. In other words, find a range within which the SD must be, by thinking about possible minimum and maximum values for the SD here. Using a WinBUGS program, provide posterior means, SD's and 95% intervals for both the mean and the now assumed unknown variance.

(c) Compare the estimated variance from part (b) with the fixed value of 144 used in Assignment #2. Is 144 included in the 95% interval? What effect would estimating the variance (rather than assuming a fixed value) have on the 95% credible interval for mu?

5. Redo problem #5 from Assignment #2, now using WinBUGS. Again use the same prior distribution that you used last time, calculate the posterior mean, SD and 95% credible intervals for the Poisson parameter, and compare it to what you had last time.