Introduction to Bayesian Statistics

Practicum 3

Using WinBugs to estimate a single proportion

**Background**

Intra-ventricular hemorrhage (IVH) is a bleeding of the brain which ranges in severity from grade I being the least severe to grade IV being the most severe. Severe IVH (grade III or IV) often results in death, long term disability, or cognitive delay. Among very preterm infants, a population with extremely sensitive and immature brains, it is estimated that 20% will have grade III or IV IVH (Stoll 2010) within the first 72 hours of life. A researcher at OUHSC wants to study the prevalence of severe IVH among the population of OUHSC inborn neonates. Over a six month period, infants born before 28 weeks of gestation were recruited for a study that involved biomarker monitoring for 72 hours concurrent with cranial ultrasound every 12 hours. 22 neonates were included in the study and the presence (1 if present, 0 otherwise) of severe (grade III or IV) IVH after 72 hours is given below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | - |
| 0 | 0 | 0 | 0 | 0 | - |

1. Using WinBUGS, estimate the prevalence of neonate with severe IVH at OUHSC along with a 95% Bayesian credible interval for the following priors.
	1. Beta(1,1) which is a diffuse or vague prior.
	2. MATCH vague: For this prior, use the MATCH software we discussed earlier together with any pertinent prior information on the prevalence of premature neonate with severe IVH at OUHSC. Since we don’t have much info on the variability, let this one represent a prior with a lot of uncertainty. Write down the distribution you identify from MATCH along with any parameter values.
	3. MATCH strong: For this prior, use the MATCH software we discussed earlier together with any pertinent prior information on the prevalence of premature neonate with severe IVH at OUHSC. Let’s assume that the expert you consult to assess this prior had very little uncertainty about its belief of how frequent this condition among premature neonates is at the OUHSC. In other words, we expect that the distribution which will represent the prior will be quite narrow. Write down the distribution you identify from MATCH along with any parameter values.

|  |  |  |
| --- | --- | --- |
| Prior Distribution (parameters) | Median Estimate | 95% Credible Interval |
| 1. Beta (1,1)
 |  |  |
| 1. MATCH Diffuse \_\_\_\_\_\_\_\_\_\_\_\_
 |  |  |
| 1. MATCH Strong \_\_\_\_\_\_\_\_\_\_\_\_
 |  |  |

Compare the median estimates and credible intervals using the three options for priors in the Table above.

1. Why are they different/ the same?
2. Which do you suppose would most closely approximates the frequentist estimate and 95% confidence interval? Explain why.

Suppose a larger study was done later over a 5 year period at OUHSC to estimate the prevalence of severe IVH in a larger cohort of premature neonates. This study enrolled an additional 253 premature neonates, 76 of whom experienced severe IVH. For convenience, whether or not the premature neonate experienced severe IVH (1 yes, 0 no) is given in the list below.

4. In light of the work you have already done in 1, describe how you might obtain a prior for this study

5. Use that prior with WinBUGS to obtain a new estimate of the prevalence of severe IVH among premature neonates along with a 95% credible interval. For convenience the data are given below.

#data

list(x=c(0,0,0,1,1,0,0,0,1,0,0,1,0,0,1,0,0,0,1,0,0,0,1,0,1,0,0,1,1,0,0,0,0,1,0,1,0,0,1,0,0,0,0,0,0,0,1,0,0,1,0,0,1,0,0,0,0,1,0,0,0,1,0,0,1,0,0,1,1,0,1,0,0,0,1,0,0,1,1,0,0,0,1,0,0,0,1,0,0,0,0,1,0,0,1,0,0,1,0,0,0,1,0,0,1,1,0,0,0,0,1,0,0,0,1,0,0,0,0,1,1,1,0,0,0,1,0,0,1,0,0,1,1,0,0,1,0,0,0,0,1,0,1,0,1,1,0,0,1,0,0,1,0,0,0,0,1,0,0,1,0,0,0,0,0,1,0,1,0,0,0,1,0,0,0,1,0,0,0,0,0,1,0,0,0,0,1,0,0,1,0,0,1,0,1,0,0,1,0,1,0,0,0,1,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,1,0,1,0,0,1,0,0,1,0,0,0,1,0,0,0,1,0,1,0,0,0))

Re-run the analyses in WinBugs with the 3 priors which you had developed in Table 1.

6. Do the priors influence the posterior the same way using this larger dataset as it did with the smaller dataset? Why or why not?

BONUS!!

As you can see with the larger dataset, the list for the data can get a little cumbersome with large sample sizes. As was mentioned earlier, there is a nice relationship between the Bernoulli distribution and the Binomial distribution that can streamline the commands we need to enter into WinBUGS as well as the amount of Bernoulli data we must enter when estimating proportions. (See “Intro to WinBUGS”). See if you can rewrite the model and data commands from question 5 to use a Binomial likelihood.

7. Use the same prior you specified in question 5 and compute an estimate of the prevalence of severe IVH along with a 95% credible interval. How do they compare to those obtained in question 5?