Bayesian Statistics: Assessment Exercise 3

Marks for this assignment: 11 of course total of 88 (12.5%)

See Moodle for submission and deadline

For this exercise, you need to submit a single text file or Word document that encompasses the answers to all of the questions below, with the section number (a) to (e) as a header for each part. Where JAGS code is asked for, you do NOT need to specify the R code needed to run it.

You have been approached by a biologist who wants to analyse a dataset. He has collected information on the exact proportion (variable name: Proportion) of insects in ten different fields (variable name: Field) that are exhibiting a phenotype indicative of a genetic mutation he is interested in. He has also recorded the number of insects counted in each field (variable name: Total), and previous exposure of that field to a known carcinogen in fertiliser (variable name: Exposure – coded as YES or NO). He wants you to run an MCMC model to describe his data.

(a) Conceptually, what distribution does the experiment follow? What link function would you choose for this model?
    (1 mark)

(b) What data would you need to pass to JAGS to run this model, and how would you calculate this data from the information that you have been given?
    (1 mark)

We have a population of animals some of which have a disease. The data is in the file "Disease.R". Our prior for prevalence of this disease is Beta(1,1), and have no useful individual animal level predictors. We test all 500 animals using a perfect test (i.e. sensitivity and specificity are 100%), and we get 19 positive results.

(c) Write a JAGS model and then calculate the posterior distribution of belief for the prevalence using MCMC.
    (4 marks)

(d) There are two ways of programming this in JAGS – one uses a loop and the other doesn't, and they have a slightly different format for the observed data *(hint: the Bernoulli distribution is a special case of the Binomial)*. Try both.
    (4 marks)

(e) Is there any difference between the results? There is one difference … try increasing the number of iterations to see if that makes it more obvious.
    (1 mark)