Bayesian Statistics in R: Assignment 3

(40 points)

1. Consider the code in section 3.3.3. Which is the correct interpretation of totch.y1. Is it (a) the number of subjects with y=1, or (b) the number of subjects with actual response y=1 whose predicted response matches the actual response, or (c) the number of subjects with actual response y=0 whose predicted response matches the actual response. (2 points)

2. Suggest how the sensitivity in the code in section 3.3.3 could be calculated from totch.y1. Dividing totch.y1 by (a) the grand total of subjects, or (b) the number of subjects with actual response y=1, or (c) the number of subjects with actual response y=0. (2 points)

3. On the basis of your answer in Q2, include extra commands as necessary in the code to calculate the sensitivity. Is the posterior mean sensitivity (a) under 0.25, or (b) over 0.25. (2 points)

4. The existing code in section 3.3.3 compares predicted responses under two scenarios regarding subject background and treatment. P[1] is for subjects at the upper age quartile of 37, median Beck score of 17, recent IV drug use, lower ndrugtx quartile of 1, and with longer duration treatment. P[2] is for subjects at the lower age quartile of 27, median Beck score of 17, recent IV drug use, upper ndrugtx quartile of 6, and with shorter duration treatment. Include a command to calculate the ratio of P[1] to P[2], and then monitor it. Is the posterior mean for this ratio (a) over 2.5 or (b) under 2.5? (2 points)

5. Change the code in section 3.3.3 so that P[1] refers to subjects never making IV drug use. So P[1] is now to be defined for subjects at the upper age quartile of 37, median Beck score of 17, never IV drug use, lower ndrugtx quartile of 1, and with longer duration treatment. Is the posterior mean for the ratio of P[1] to P[2] now (a) over 4.5 or (b) under 4.5? (2 points)

6. Add a command to the code in section 3.3.3 to calculate standardized residuals,

$$r_i = (y_i-\pi_i)/[\pi_i(1-\pi_i)]^{0.5}$$

It is suggested to use the jagsresults command to extract the residuals. Which cases have the highest and lowest posterior mean standardized residuals?
(3 points)

7. Monitor the predictive checks in the code in 3.4.2, which derive

$$Pr(y_{rep,i} > y_i|y)+0.5Pr(y_{rep,i} = y_i|y)$$

For which observation does the check raise most concerns? Does the model tend to (a) overpredict or (b) underpredict for this observation. (3 points)

8. Include an extra line in the observation loop to obtain the elements needed to derive the scaled deviance, something like

dv[i] <- y[i]*log(…)

and an extra statement outside the loop to obtain the total scaled deviance, something like

ScD <- 2*sum(dv[]).

Is the posterior mean deviance (a) above 5, or (b) below 5. (5 points)

9. Modify the code in section 3.4.2 to include a second regression of hypertension rates against snoring levels. The code could look something like

```
model { for (i in 1:4) {y[i] ~ dbin(p1[i],n[i])
                z[i] ~ dbin(p2[i],n[i])
logit(p1[i]) <- beta[1]+beta[2]*score[i]
logit(p2[i]) <- gam[1]+gam[2]*score[i]}
…}
```

Find the probability that the trend slope for hypertension exceeds that for CHD. Is this probability (a) over 0.75, or (b) under 0.75. (5 points)

10. Include commands in the code in section 3.6.5 to calculate the scaled Poisson deviance ScD under the Poisson regression option in JAGS (see section 3.5.4). Is the posterior mean for ScD (a) under 24000 or (b) over 24000.

(6 points)

11. Include a command in the code in section 3.6.5 to obtain standardized residuals under the negative binomial option, namely

$$r_i=(y_i-\mu_i)/V_i^{0.5}$$

where $V_i=V(y_i|X_i)=\mu_i+\mu_i^2/\theta$. Remember that in the code the response is denoted ofp[i]. Which two cases have the highest posterior mean standardized residuals? Note it is best when extracting large arrays to first ensure convergence in the main model parameters (e.g. regression coefficients). Then use a subsidiary coda.samples command to monitor the large array, and also use jagsresults, as in

```
R1 <- coda.samples(M,c("r"),n.iter=500)
r.R1 <- jagsresults(R1, c("r"))
```

Note that the summary(R1) and gelman.diag(R1) commands need not be done, as they are likely to be time-consuming. (6 points)

12. Modify the prior on θ in the NB regression code to be

$$\theta \sim U(0,10).$$

Run the model for 5000 iterations. Is the posterior mean for θ now (a) over 1.25, or (b) under 1.25. (2 points)