22S:138, Bayesian Statistics
Fall 2007, Homework 10

Does not have to be turned in. Problems like this may appear on the final.

1. This question is a continuation of your first homework for this class.

   Refer to section 1.4 (pp. 9-11) in Gelman *et al.* The question of interest is whether or not the woman is a carrier of the hemophilia gene. (If you no longer have the photocopy of this, see me.)

   Suppose the woman has 1 son, whose hemophilia status is $y = 0$. (This is different from the outcomes discussed in the book.)

   Compute:

   (a) prior odds in favor of $\theta = 0$ vs. $\theta = 1$

   (b) Bayes factor in favor of $\theta = 0$ vs. $\theta = 1$

   (c) posterior odds in favor of $\theta = 0$ vs. $\theta = 1$

2. Attached is WinBUGS code and output for the Dyes example, which you know from a previous homework. I have added 4 lines to the program. Refer to the code and output to answer the following questions (just a sentence or two for each).

   (a) Explain the meaning of the following new nodes:

      i. ypred

      ii. resid

      iii. presid

      iv. pppv

   (b) What could you learn by monitoring "pppv"?

   (c) Does the output suggest any problems with model fit?

```
model
{
for( i in 1 : batches ) {
   m[i] ~ dnorm(theta, tau.btw)
   for( j in 1 : samples ) {
     y[i , j] ~ dnorm(m[i], tau.with)
     ypred[i,j] ~ dnorm(m[i], tau.with)
     resid[i,j] <- abs(y[i,j] - m[i])
     presid[i,j] <- abs(ypred[i,j] - m[i])
   }
    large[i] <- ranked(resid[i,], samples)
    largepred[i] <- ranked(presid[i,], samples)
    pppv[i] <- step(large[i] - largepred[i])
}
```



```
sigma2.with <- 1 / tau.with
sigma2.btw <- 1 / tau.btw
tau.with ~ dgamma(0.001, 0.001)
tau.btw ~ dgamma(0.001, 0.001)
theta ~ dnorm(0.0, 1.0E-10)
}
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
m[1]	1515.0	20.31	0.4391	1472.0	1517.0	1550.0	10001	40000
m[2]	1528.0	18.62	0.2291	1490.0	1527.0	1566.0	10001	40000
m[3]	1548.0	22.12	0.5962	1512.0	1547.0	1594.0	10001	40000
m[4]	1511.0	21.42	0.5458	1466.0	1513.0	1547.0	10001	40000
m[5]	1569.0	30.38	1.143	1517.0	1572.0	1624.0	10001	40000
m[6]	1495.0	27.22	0.953	1443.0	1494.0	1544.0	10001	40000
pppv[1]	0.561	0.4963	0.005088	0.0	1.0	1.0	10001	40000
pppv[2]	0.1466	0.3537	0.003352	0.0	0.0	1.0	10001	40000
pppv[3]	0.3195	0.4663	0.003411	0.0	0.0	1.0	10001	40000
pppv[4]	0.6383	0.4805	0.006751	0.0	1.0	1.0	10001	40000
pppv[5]	0.5282	0.4992	0.003993	0.0	1.0	1.0	10001	40000
pppv[6]	0.2517	0.434	0.004228	0.0	0.0	1.0	10001	40000
sigma2.btw	2125.0	3716.0	65.35	0.004705	1237.0	9942.0	10001	40000
sigma2.with	3083.0	1125.0	31.27	1571.0	2855.0	5842.0	10001	40000

3. Use the DIC to compare the fits of two different models for the data involving growth of baby rats. First use the model, data, and initial values exactly as given in the "Rats" example in Volume 1 of examples. Then use the model, data, and initial values as given in the "Birats example in Volume 2. In both cases, run at least 1000 burn-in iterations before you Set the DIC. Then use output for the DIC based on at least 10000 additional iterations. Turn in the tables of DIC results for both models, and answer the following questions:

   (a) What is the estimated number of free parameters in the "Rats" model? In the "Birats" model? What could explain the difference between the two estimates?

   (b) Is one model strongly preferred over the other after the penalty for model complexity is taken into account? Justify your answer.