Name: ----------

Bayesian Statistics, 22S:138
Final exam, 2009

1. The University of Iowa requires all teaching faculty and TAs to take training in recognizing and reporting sexual harrassment. A sociology professor suspects that less than half of all public universities require such training of their teaching personnel. He conducts a Bayesian test of the following hypotheses:

$$\begin{aligned} H_0 : &\quad p \geq .5 \\ H_A : &\quad p < .5 \end{aligned}$$

where p is the proportion of all public universities that require sexual harrassment training for their instructional personnel. He commits to presenting his findings on a particular date.

Before gathering any data, he specifies a vague prior on p:

$$p \sim Beta(1, 1)$$

He then takes a simple random sample of size 30 from among all public universities. He calls an administrator at each of these institutions and learns that 18 of the 30 schools require sexual harrassment training for their teaching personnel, and 12 do not.

Here is some R output that may be useful in answering the questions that follow.

```
> pbeta(0.5, 1,1)
[1] 0.5
> pbeta(0.5, 19, 13)
[1] 0.1405208
```

(a) The hypotheses that the sociologist wishes to test are (circle one):

i. one-sided `this one is correct`
ii. two-sided
iii. simple-simple
iv. none of the above

(b) The prior odds in favor of the null hypothesis are: (numeric answer; show your work)

```
Under the Beta(1,1) prior on p, the prior probabilities on the null and
alternative hypotheses are both 0.5. Thus the prior odds are
    0.5
    ---   = 1
    0.5
```

(c) The Bayes factor in favor of the null hypothesis is: (numeric answer; show your work)

The posterior density p(p|y) is Beta(19,13). The R output provided shows that the posterior probability of the alternative hypothesis is

Pr(p < 0.5| y) = 0.1405

Thus the posterior odds in favor of the null hypothesis are:

$$\frac{1-0.1405}{0.1405} = 6.117$$

The Bayes factor is posterior odds over prior odds:

$$\frac{6.117}{1} = 6.117$$

(d) Does the data provide strong evidence either for or against the null hypothesis? Explain.

According to Kass and Raftery, a Bayes factor between 3.2 and 10 provides "substantial" but not "strong" evidence.

(e) What is the mean of the posterior distribution for p obtained by the sociologist (numeric answer; show your work).

$$\frac{19}{19+13} = 0.5938$$

(f) Suppose that the sociologist had originally intended to survey 50 universities instead of 30. He had drawn a simple random sample of 50 universities, and had contacted administrators from the first 30 on the list when he got H1N1 flu. By the time he was well enough to get back to work, he did not have time before his presentation to call the rest of the schools. Therefore, he performed his analysis using only the 30 responses he had already obtained.

Should the sociologist's inference about p be affected by the fact that he obtained fewer observations than he originally intended? Briefly justify your answer using a principle that is fundamental in the Bayesian framework.

This question was not worded well. The intended comparison was between two studies, both of which observed 18 successes out of a sample size of 30. The only difference was that in the first case, the sample size of 30 was planned whereas in the second case the planned sample size was 50 but only 30 observations were actually made. For this situation, the Likelihood Principle would say that the inference should not be affected. The likelihoods will be the same in both cases, so the contain the same information about the unknown parameter.

2. You wish to investigate how to use adults' self-reported weights to predict their actual (measured) weights. You plan to recruit a sample of 200 healthy adults and to ask each of them how much he or she weighs. You will then weigh each of them on a very accurate scale. You will convert both weights to kilograms.

   You plan to use a Bayesian linear regression model to analyze your data. You will center the predictor variable (self-reported weight).

   A similar study was previously done by C. Davis, Departments of Physical Education and Psychology, York University. A frequentist linear regression model was fit to his data, also with the predictor variable centered, producing the following output:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	29738	29738	419.59	<.0001
Error	181	12828	70.87307		
Corrected Total	182	42566			

Root MSE	8.41861	R-Square	0.6986
Dependent Mean	66.22404	Adj R-Sq	0.6970
Coeff Var	12.71232		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	66.22409	0.62232	106.41	<.0001
rwgtc	1	0.92784	0.04530	20.48	<.0001

Based on Davis's results, write an appropriate informative prior for each of the following three parameters, in the way that you would code it in WinBUGS.

(a) β_0, the intercept

```
Common practice is to specify normal priors on regression coefficients.
    beta0 ~ dnorm( 66.2, 1/0.622^2)
```

(b) β_1, the slope of actual weight on self-reported weight

```
    beta1 ~ dnorm( 0.978, 1/0.0.0453^2)
```

(c) σ^2, the variance of true values around the true regression line

```
Here we want an inverse gamma prior with n (or n-2) as the equivalent prior
sampe size, and n (or n-2) times the mean squared error over 2 as the
beta parameter.  The equivalent for WinBUGS is a gamma prior on the
precision.

    tausq ~ gamma ( 181/2, (181 * 70.82)/2 )
```

3. The table on the last page of this exam contains measurements of the blood concentration of epinephrine in ten dogs under 3 kinds of anesthesia. Each row corresponds to a single dog. The columns show epinephrine measurements under Isofluorane, Halothane, and Cyclopropane respectively. High levels of epinephrine during anesthesia are not good. The data come from a study by Perry et al. (1974) as reported in Rice (1995), and are reproduced in Table 10.2, page 340, of the online textbook *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians* by Ron Christensen, Wes Johnson, and Adam Branscum (`http://www.math.unm.edu/~fletcher`).

   Two models are fitted to the data (last 3 pages of the exam). In Model 2, the data values are assumed to be normally distributed around the population means of epinephrine for each anesthetic; the only parameters are the means for each anesthetic and the precision. Model 1 attempts to account for correlations among measurements on the same dog. It includes a parameter for each dog to represent its own typical ephinephrine level; since the dogs are assumed to be randomly selected from the population of all dogs, these parameters are random effects.

   Refer to the WinBUGS code and output on the last 3 pages of the exam in answering the following questions.

   (a) Draw a directed graph of Model 1. (Use the back of the last page if there isn't enough room here.)

(b) Copy the line or lines of code that specify the second stage of Model 1.

```
b[i] ~ dnorm( 0, taub )
```

(c) Write the line or lines of code that you would need to add in order to draw replicate datasets from the posterior predictive distribution.

```
yrep[i,j] ~ dnorm( mu[i,j], tau )
```

(d) For each model, I ran 3 parallel chains for 1000 iterations. Graphics and diagnostics suggested that convergence had occurred. I set the DIC and ran the chains for an additional 5000 iterations. The reported DIC values and the tables of node statistics are based on the post burn-in samples from all 3 chains.

Does the output indicate that the 5000 post-burn-in iterations were enough to provide acceptable precision in parameter estimation? Explain.

`Yes. All the MC errors are less than 0.05 times the corresponding standard deviations.`

(e) Give the posterior mean and 95% credible set for the population mean ephinephrine level of dogs anesthetized with Isofluorane (numeric answers taken from WinBUGS output)

i. according to Model 1: `0.435 (0.2173, 0.6577)`

ii. according to Model 2: `0.434 (0.2264, 0.6448)`

(f) Do the two models give substantially different inference regarding the research question of interest (population means of epinephrine under the three different anesthetics)? Briefly explain.

`No. The posterior means and credible sets for each beta are almost the same under both models.`

(g) How many total parameters are there in Model 1 (just count them and give a numeric answer)?

`3 betas + 10 bs + 1 taub + 1 tau = 15 parameters`

(h) How many effective parameters are reported along with the DIC for Model 1? Why is this answer equal, or not equal, to the answer you gave to part (g)?

`5.8 (roughly 6). This is smaller than 15 because the b's are correlated and so the DIC calculation does not count each b as a full parameter.`

(i) Is the added complexity of Model 1 needed for this particular dataset? Justify your answer using at least two different parts of the WinBUGS output for both models.

`No. The DIC is smaller for model 2, suggesting that it is preferred. The inference on the betas is the same under both models.`

```
# Model 1

model
{
    for(i in 1:N) {
        for(j in 1:3) {
            y[i,j] ~ dnorm( mu[i,j], tau )
            mu[i,j] <- beta[j] + b[i]
        }
        b[i] ~ dnorm( 0, taub )
    }

    sigmab ~ dunif( 0, 0.5 )
    taub <- 1 / (sigmab * sigmab)

    tau ~ dgamma(0.01, 0.01)

    for(j in 1:3) {
        beta[j] ~ dnorm( 0, 0.0001 )
    }
}

# inits
list( sigmab = 0.45, tau = 10, beta = c( .5, .5, .5) )
list( sigmab = .25, tau = 1, beta = c(0,1,2) )
list( sigmab = .05, tau = .1, beta = c(2,1,0)  )

Node statistics
node     mean     sd     MC error     2.5%     median     97.5%     start     sample
beta[1]  0.435   0.1102  0.001059    0.2173    0.434     0.6577     1001      15000
beta[2]  0.4688  0.1093  0.001034    0.2547    0.4681    0.6874     1001      15000
beta[3]  0.8535  0.1103  0.00105     0.6354    0.8524    1.071      1001      15000
sigmab   0.0788  0.06583 0.002488    0.00243   0.06396   0.2407     1001      15000
tau      9.778   2.736   0.02761     5.21      9.519    15.84       1001      15000

DIC
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
   Dbar    Dhat    pD    DIC
y  17.645  11.827  5.818 23.463
```

```
# Model 2

model
{
    for(i in 1:N) {
        for(j in 1:3) {
            y[i,j] ~ dnorm( beta[j], tau )
        }
    }
    tau ~ dgamma(0.01, 0.01)
    for(j in 1:3) {
        beta[j] ~ dnorm( 0, 0.0001 )
    }
}

# inits
list( tau = .1, beta = c( .5, .5, .5) )
list(  tau = 1, beta = c(0,1,2) )
list(  tau = 10, beta = c(2,1,0)  )

Node statistics
node    mean    sd     MC error  2.5%    median  97.5%   start   sample
beta[1] 0.4338  0.105  8.479E-4  0.2264  0.4327  0.6448  1001    15000
beta[2] 0.4663  0.1048 9.068E-4  0.2587  0.4667  0.6728  1001    15000
beta[3] 0.8528  0.1047 8.305E-4  0.6465  0.8519  1.061   1001    15000
tau     9.837   2.683  0.02307   5.262   9.586  15.81    1001    15000

DIC
Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes
Dbar Dhat pD DIC
y 17.485 13.356 4.129 21.614
```

Data

```
y[,1]  y[,2]  y[,3]
0.28  0.30  1.07
0.51  0.39  1.35
1.00  0.63  0.69
0.39  0.68  0.28
0.29  0.38  1.24
0.36  0.21  1.53
0.32  0.88  0.49
0.69  0.39  0.56
0.17  0.51  1.02
0.33  0.32  0.30
END
```