22S:138, Bayesian Statistics
Fall 2008, Homework 7

Due: Mon., 11/03 in class

1 Directed graphs and full conditional distributions

1. The exponential distribution is commonly used to model data on time from some starting point until an event occurs. The probability density function for a random variable Y drawn from an exponential distribution with parameter λ is

$$p(y \mid \lambda) \;=\; \lambda\, e^{-\lambda y}, \;\; 0 < y < \infty$$

Researchers wish to compare 5 medications that reduce fever. These medications are all different formulations of the same drug (capsule, liquid, tablet, etc.). The experiment involves 30 patients, each with a temperature of at least 102 degrees. Six patients are randomly assigned to medication 1, six patients to medication 2, etc. After the patient takes the medication assigned to him, his temperature is monitored continuously. The outcome variable of interest is the time in hours from when the medication is taken until the patient's temperature goes below 99.5 degrees.

Let y_{ij} denote this time for patient j who received medication i.

The researchers will carry out a Bayesian analysis. They assume that all the times for patients receiving the same medication, say i, follow an exponential distribution with the same parameter λ_i. They believe that the λs for different treatments are different, but they have no idea of which ones will be larger or smaller.

Below is WinBUGS code for fitting a 3-stage hierarchical model to their data.

```
model {
for (i in 1:N) {
     for (j in 1:M) {
          Y[i,j] ~ dexp( lambda[i])
               }
    lambda[i] ~ dgamma( alpha, beta)
}
alpha ~ dgamma( 1,1)
beta ~ dgamma(1,1)
overall <- alpha / beta
}
```

(a) Draw a directed graph for this model. You may either do this by hand or use DoodleBUGS or any graphing software that you know.

(b) Derive expressions proportional to the full conditional distributions for lambda[i], alpha, and beta. If any of them are standard distributions, identify them by family and parameters. If any of them are not standard distributions, say so.

2. Suppose another level were added to the hierarchy as follows:

```
model {
for (i in 1:N) {
     for (j in 1:M) {
          Y[i,j] ~ dexp( lambda[i])
               }
    lambda[i] ~ dgamma( alpha, beta)
}

alpha ~ dgamma( a1, b1)
beta ~ dgamma(a2, b2 )

a1 ~ dexp(2)
b1 ~ dexp(2)
a2 ~ dexp(2)
b2 ~ dexp(2)

overall <- alpha / beta
}
```

(a) Will the full conditionals for the lambda[i]s change? Why or why not?

(b) Derive expressions proportional to the full conditionals for alpha, beta, a1, and b1. If possible, identify them as standard distributions.

2 Regression

You will fit a linear regression model to the temperature/heart rate data discussed in class. This dataset is available in WinBUGS format under "Datasets" on the course web page as "normtbug.dat." You may copy it from there into a WinBUGS window. We will consider heart rate the response variable and body temperature the independent or predictor variable.

See the lab 5 handout for example code for fitting a regression model. You will need to change the variable names in the code to match the variable names in the data (i.e. "heart" instead of "Y," etc.).

You will use the noninformative priors from the example code.

1. Fit the model with centering of the predictor variable.

```
mu[i] <- alpha + beta * (temp[i]-temp.bar)
```

Run 3 separate 1000-iteration samplers. Use the following initial values. Monitor alpha, beta, and sigma.

```
list(alpha = -100, beta = 5, tau = 1)
list(alpha = 0, beta = 2.5, tau = .1)
list(alpha = 100, beta = 0, tau = 10)
```

Look at history plots of all 3 chains together, as well as autocorrelation plots. Also run the Gelman and Rubin diagnostic. (The only plot you need to print is the history plot for α.)

Write a paragraph describing the plots and commenting on whether the samplers are converging satisfactorily.

- How many initial iterations should you throw out in order that the remaining iterations appear to be draws from the target distribution?

2. Print the summary statistics for the model parameters. Have WinBUGS begin computing them starting at the first iteration that you do not throw out. Use the posterior means of α and β to hand-calculate the expected value of *heart* when $temp = 100.8$. Recall that the covariate has been centered.

3. Run the plots in "Inference/Compare." Comment on whether you see outliers, nonlinearity, or unequal variance, etc.

4. Prediction: In the data list, change the last value of "heart" to NA. Then fit the model again. Run the same number of iterations as you threw out in a previous step. Then start monitoring heart[130] as well as alpha, beta, and sigma. Obtain a 95% prediction interval for heart[130].

5. Informative priors: Suppose you had the following SAS output from a linear regression fit to a dataset of body temperatures and heart rates measured on a different random sample of 100 healthy adults. The predictor variable was centered in this analysis.

```
                     Model: MODEL1
               Dependent Variable: heart

                  Analysis of Variance

                              Sum of         Mean
Source                DF     Squares       Square  F Value  Pr > F

Model                  1   651.29310    651.29310    43.93  <.0001
Error                 98  1452.80040     14.82449
Corrected Total       99  2104.09349

        Root MSE             3.85026   R-Square     0.3095
        Dependent Mean      77.26720   Adj R-Sq     0.3025
        Coeff Var            4.98304
```



```
                         The REG Procedure
                          Model: MODEL1
                    Dependent Variable: heart

                       Parameter Estimates

                        Parameter       Standard
Variable      DF         Estimate          Error    t Value    Pr > |t|


Intercept      1         75.73899        0.60151     122.59      <.0001
temp           1          2.73008        0.41189       6.63      <.0001
```

(a) Write one choice of informative prior based on this previous analysis.

(b) Would the posterior variance of the regression slope parameter be larger or smaller if you fit a model using this prior rather than the vague prior previously used? Why? (You do not have to refit the model to answer this question.)