UNIVERSITY OF WOLLONGONG

School of Mathematics and Applied Statistics

STAT902. Advanced Data Analysis

Assignment 9

Due: 5:00pm Monday 17th May, 2010.
Local students: please put under lecturer's door.
Remote students: please e-mail or post by this date.

1. Consider the Bayesian normal random sample model:

$$[y_1, \ldots, y_n | \, \mu, \sigma^2] \stackrel{\text{ind.}}{\sim} N(\mu, \sigma^2)$$

$$[\mu] \sim N(0, \sigma_\mu^2), \qquad [\sigma^2] \sim \text{IG}(A, B),$$

where σ_μ^2, A and B are hyperparameters to be specified by the analyst. Note that the prior density on σ^2 is

$$[\sigma^2] = \frac{B^A}{\Gamma(A)} (\sigma^2)^{-A-1} e^{-B/\sigma^2}, \quad \sigma^2 > 0.$$

Derive the distributions of the full conditionals:

$$[\mu | \sigma^2, \boldsymbol{y}] \qquad \text{and} \qquad [\sigma^2 | \mu, \boldsymbol{y}]$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$.

Hint: Write the model in linear model format and use the result in the handout 'A Useful Result Involving Normal Distributions'.

2. This question involves Bayesian linear mixed model analysis of the pigweights and orthodontal data from Assignment 6.

   (a) Make sure that the files `pigweights.txt` and `summariseMCMC.txt` from previous assignments are available.

   (b) Download the files `pigwtsBayesModel.txt` and `pigwtsBayes.Rs` from the `Computer Code and Data` page of the course web-site.

   (c) Type `source("pigwtsBayes.Rs")` to fit the Bayesian normal mixed model:

$$[y_{ij} | U_i, \sigma_U^2, \sigma_\varepsilon^2] \stackrel{\text{ind.}}{\sim} N(\beta_0 + U_i + \beta_1 \, x_{ij}, \sigma_\varepsilon^2)$$

$$[U_i | \, \sigma_U^2] \stackrel{\text{ind.}}{\sim} N(0, \sigma_U^2)$$

$$\beta_0, \beta_1 \stackrel{\text{ind.}}{\sim} N(0, 10^8), \quad \sigma_\varepsilon^2, \sigma_U^2 \stackrel{\text{ind.}}{\sim} \text{IG}(0.01, 0.01)$$

   where the y variable is weight and the x variable is weeks, after undergoing standardisation.

   (d) Issue the following commands to obtain the maximum likelihood fit of the frequentist counterpart of the above model:

```
library(nlme)
gD <- groupedData(weight~sweeks|idnum,data=pigweights)
fit <- lme(weight~sweeks,random=~1,data=gD)
print(intervals(fit))
```

Write up a brief comparison of the two sets of results.

(e) Fit the above Bayesian normal mixed model to the orthodontal data from Assignment 6, with the y variable being the `distance` measurement and the x variable being age. Provide graphical and numerical summaries of the analysis.

IMPORTANT: The data frame `Orthodont` (within the package `nlme`) has the individuals in the study labelled using the `Subject` column as follows:

```
M01 M01 M01 M01 M02 M02 M02 M02 M03 M03 M03 M03 M04 M04 M04 M04 M05 M05
M05 M05 M06 M06 M06 M06 M07 M07 M07 M07 M08 M08 M08 M08 M09 M09 M09 M09
M10 M10 M10 M10 M11 M11 M11 M11 M12 M12 M12 M12 M13 M13 M13 M13 M14 M14
M14 M14 M15 M15 M15 M15 M16 M16 M16 M16 F01 F01 F01 F01 F02 F02 F02 F02
F03 F03 F03 F03 F04 F04 F04 F04 F05 F05 F05 F05 F06 F06 F06 F06 F07 F07
F07 F07 F08 F08 F08 F08 F09 F09 F09 F09 F10 F10 F10 F10 F11 F11 F11 F11
```

However, the `BUGS` code for the pig weights example assumes numerical labels. Specifically, to mimic what was done for the pig weights example, we need to work with the labels:

```
 1  1  1  1  2  2  2  2  3  3  3  3  4  4  4  4  5  5  5  5  6  6  6  6  7
 7  7  7  8  8  8  8  9  9  9  9 10 10 10 10 11 11 11 11 12 12 12 12 13 13
13 13 14 14 14 14 15 15 15 15 16 16 16 16 17 17 17 17 18 18 18 18 19 19 19
19 20 20 20 20 21 21 21 21 22 22 22 22 23 23 23 23 24 24 24 24 25 25 25 25
26 26 26 26 27 27 27 27
```

for the orthodontal data.

This can be achieved through the creation of a new identification number variable as follows:

```
idnum <- rep(1:27,each=4)
```

3. This question involves Bayesian Poisson mixed model (special case of the Generalised Linear Mixed Model (GLMM)) analysis of longitudinal data on seizure counts.

(a) Obtain the files `seizure.txt`, `seizure.Rs`, `seizureModel.txt` from the `Computer Code and Data` page of the course web-site.

(b) Type `source("seizure.Rs")` to fit the Bayesian normal mixed model:

$$[y_{ij}|U_i,\sigma_U^2] \overset{\text{ind.}}{\sim} \text{Poisson}\{\exp(\beta_0 + U_i + \beta_1\,x_{1ij} + \beta_2\,x_{2ij} + \beta_3\,x_{3ij})\}$$

$$[U_i|\,\sigma_U^2] \overset{\text{ind.}}{\sim} N(0,\sigma_U^2)$$

$$\beta_0,\beta_1 \overset{\text{ind.}}{\sim} N(0,10^8), \quad \sigma_U^2 \sim \text{IG}(0.01,0.01)$$

where the y variable is seizure count, the x_1 variable is an indicator of treatment via the drug *progabide*, the x_2 variable is a baseline measurement on seizure count and x_3 is age (the last two predictors are normalised).

(c) Provide graphical and numerical summaries of the analysis and make some brief concluding remarks about the respective impacts of the 3 predictors on the response variable.

Note: It is not clear whether the above Poisson mixed model is a good one, and diagnostic checking should still be done in practice. The main goal of this question is illustration of GLMM fitting using `BRugs`.

4. Consider the Bayesian Poisson random intercept model:

$$[\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \sigma_U^2] = \exp\{\boldsymbol{y}^T(\boldsymbol{X\beta} + \boldsymbol{Zu}) - \mathbf{1}^T \exp(\boldsymbol{X\beta} + \boldsymbol{Zu}) - \mathbf{1}^T \ln(\boldsymbol{y}!)\}$$

$$[\boldsymbol{u}|\sigma_U^2] \sim N(\mathbf{0}, \sigma_U^2 \boldsymbol{I})$$

$$[\boldsymbol{\beta}] \sim N(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}), \quad [\sigma_U^2] \sim \text{IG}(A, B)$$

where $\boldsymbol{u}$ is $q \times 1$. Obtain the full conditional distribution for σ_U^2:

$$[\sigma_U^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{y}].$$