# Statistical Modelling: Practical 3

A. C. Davison and J. J. Forster

April 10, 2008

The aim of this practical session is to carry out a Bayesian analysis of the hip-replacement data previously analysed in Practical 2. We will focus on the linear mixed model: $y_{ij} \overset{\text{ind}}{\sim} N(\mu_{ij}, \sigma^2)$ where

$$\mu_{ij} = x_{ij}^{\mathrm{T}}\beta + b_i + \epsilon_{ij}, \quad b_i \overset{\text{ind}}{\sim} N(0, \sigma_b^2). \tag{1}$$

with `1`, `age`, `sex` and `I(time=2)` and `I(time=3)` in $x_{ij}$. Hence our unknown parameters are the $p = 5$ components of $\beta$ and the precision components $\sigma^{-2}$ and $\sigma_b^{-2}$. We denote the number of clusters by $k$ and the total number of observations by $n$. We will assume the conditionally conjugate prior distributions

$$\beta \sim N_p(\mu_\beta, \Sigma_\beta), \quad \sigma^{-2} \sim \text{gamma}(h, c) \quad \sigma_b^{-2} \sim \text{gamma}(h_b, c_b)$$

Then the joint posterior density for $(\beta, b, \sigma^{-2}, \sigma_b^{-2})$ is

$$\begin{aligned} f(y, b, \beta, \sigma^{-2}, \sigma_b^{-2}) &\propto f(y, b, \beta, \sigma^{-2}, \sigma_b^{-2}) \\ &\propto f(y|b, \beta, \sigma^{-2})\, f(b|\sigma_b^{-2})\, f(\beta)\, f(\sigma^{-2})\, f(\sigma_b^{-2}) \\ &\propto \phi_n(y|X\beta + Zb, \sigma^2 I_n)\, \phi_k(b; 0, \sigma_b^2 I_k)\, \phi_p(\beta; \mu_\beta, \Sigma_\beta) \times \\ &\qquad (\sigma^{-2})^{h-1} \exp(-c\sigma^{-2})\, (\sigma_b^{-2})^{h_b-1} \exp(-c_b\sigma^{-2}) \end{aligned} \tag{2}$$

where $\phi_p(y; \mu, \Sigma)$ denotes the density of the $p$-dimensional multivariate normal distribution with mean $\mu$ and variance $\Sigma$, evaluated at $y$.

From (2) some relatively straightforward manipulation gives the posterior conditional distributions for $b, \beta, \sigma^{-2}, \sigma_b^{-2}$ as

$$b|y, \text{rest} \sim N\left(\Sigma_b^* Z^T(y - X\beta)\sigma^{-2}, \Sigma_b^*\right) \tag{3}$$

$$\beta|y, \text{rest} \sim N\left(\Sigma^*\left[X^T(y - Zb)\sigma^{-2} + \Sigma_\beta^{-1}\mu_\beta\right], \Sigma^*\right) \tag{4}$$

$$\sigma^{-2}|y, \text{rest} \sim \text{gamma}\left(h + n/2, c + (y - X\beta - Zb)^T(y - X\beta - Zb)/2\right) \tag{5}$$

$$\sigma_b^{-2}|y, \text{rest} \sim \text{gamma}\left(h_b + k/2, c_b + b^T b/2\right) \tag{6}$$

where $\Sigma^* = (\Sigma_\beta^{-1} + \sigma^{-2}X^TX)^{-1}$ and $\Sigma_b^* = (\sigma_b^{-2}I_k + \sigma^{-2}Z^TZ)^{-1}$ and 'rest' indicates the other components of $(\beta, b, \sigma^{-2}, \sigma_b^{-2})$. If you have time, you might verify (3)-(6).

A Bayesian analysis of model (1) can now be carried out by sampling from the posterior ditribution (2) using a Gibbs sampler. This involves iteratively updating components of $(b, \beta, \sigma^{-2}, \sigma_b^{-2})$ by sampling from the conditional distributions (3)-(6).

If you are reasonably confident at programming in R, and fancy a challenge, then you could try and write a Gibbs sampler programme in R for model (1). It should be able to take arguments representing the number of observations required to be generated and the prior parameters $(\mu_\beta, \Sigma_\beta, c, h, c_b, h_b)$ – you might find it more convenient to specify the prior precision $\Sigma_\beta^{-1}$ rather than $\Sigma_\beta$. Other arguments required are starting values for $(\beta, \sigma^{-2}, \sigma_b^{-2})$.

Alternatively (and probably recommended) you might prefer to spend the time running the Gibbs sampler programme available in file `Prac3.txt` on the APTS web site. This requires the function `mvrnorm` for generating the multivariate normal conditional distributions (3)-(4) so you will need to load `library(MASS)`.

[A possible programme of analysis is described below. There is more here than can reasonably be carried out in a single session, so feel free to pick and choose the parts that interest you.]

(a) Generate 10 000 observations from your Gibbs sampler for the diffuse (relatively uninformative) prior distributions $\mu_\beta = 0,\ \Sigma_\beta = 10^6 I_5,\ c = h = h_b = 0.001,\ c_b = 1$ which are the default values in the R function provided.

(b) For a few chosen parameters, produce time series plots of your sample. You will see that the sample takes a short time to 'converge' from the initial starting value to a representative value from the posterior distribution. This initial segment of the sample is called the *burn-in* and should be discarded prior to the sample being used for inference.

(c) Having discarded the burn-in, calculate posterior means and standard deviations for your model parameters. Compare these with the estimates and standard errors you obtained in Practical 2 using likelihood-based methods. Plot estimated posterior densities using the kernel density estimation function `density`.

(d) Prediction is particularly straightforward in a Bayesian analysis. Write a function to calculate, based on your MCMC sample, the mean and variance of the predictive distribution for the missing values in the data set (subject 8, time 3 and subject 15, time 1). Write a function to calculate the mean and variance of the predictive distribution for a new subject, aged 70, male or female, at each of the three time points. Extend your function to provide predictive density estimates.

[Hint: For predicting the observations for existing clusters, you need to generate a sample of error terms $\epsilon$ from normal distributions with mean 0 and variances given by your generated Gibbs sampler values for $\sigma^2$. For a new cluster, you will have to generate, additionally, a sample of random effects $b$ for the new cluster from normal distributions with mean 0 and variances given by your generated Gibbs sampler values for $\sigma_b^2$

(e) Care needs to be taken with a Bayesian analysis involving variance components, that the prior distribution of the random effects variances is not too diffuse, as this can cause problems, not least with convergence of the Gibbs sampler. Investigate this behaviour by changing the value of $c_b$ to 0.001.

(f) The partial autocorrelation function `acf(sample ,type="partial")` gives an indication of how well the sampler is mixing (exploring the posterior distribution). High autocorrelations correspond to poor mixing. Assuming that the sample path of a given parameter $\theta$ can be approximated by an AR(1) process with lag 1 autocorrelation $\rho$ (in which case the partial autocorrelation dies off quickly after lag 1) the simulation standard error involved in etsimating the posterior mean of $\theta$ using $m$ sample observations is approximately $Var(\theta)/m^{1/2}$ (the independent sample standard error) multiplied by a Markov chain inflation factor $[(1+\rho)/(1-\rho)]^{1/2}$. Obtain approximate simulation standard errors for your posterior mean estimates in (c).