

Problem 4

Bayesian analysis of a simple hierarchical model using
MCMC

Longitudinal data and random effects

Suppose that for data for n individuals are recorded over T timepoints: for $i = 1, \dots, n$ we have

- ▶ response data: $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})$
- ▶ predictors: $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})$ for $t = 1, \dots, T$

We may use the following model

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{z}_{it}v_i + \epsilon_{it}$$

where

- ▶ $\mathbf{z}_{it} = (z_{it1}, \dots, z_{itq})$ is another vector of predictors
- ▶ $v_i = (v_{i1}, \dots, v_{iq})^\top$ is a vector of **random effects**.

Longitudinal data and random effects

Here we typically assume that

- ▶ $v_1, \dots, v_n$ are iid with

$$v_i \sim \text{Normal}(0, \Sigma_v) \quad q \times 1$$

- ▶ the residual vectors $\epsilon_1, \dots, \epsilon_n$ are iid with

$$\epsilon_i \sim \text{Normal}(0, \Sigma_\epsilon) \quad T \times 1$$

to complete the model.

For identifiability, we must introduce one equality constraint; this is as the model with

$$\beta_1 \longrightarrow \beta_1 + c \quad v_i \longrightarrow v_i - c$$

for any constant c yields the same likelihood. For example, we might have

$$\sum_{i=1}^n v_i = 0 \quad \text{or} \quad v_1 \equiv 0.$$

Bayesian inference using MCMC

The parameters of interest are β , but

$$\Sigma_v \quad \Sigma_\epsilon$$

are also unknown (matrix-valued) nuisance parameters.

In addition, the random effects

$$v_1, \dots, v_n$$

can also be considered unknown (vector-valued) nuisance parameters.

Bayesian inference using MCMC

The posterior of all unknown quantities is

$$\pi(\beta, \Sigma_v, \Sigma_\epsilon, v_1, \dots, v_n | \mathbf{X}, \mathbf{Y}, \mathbf{Z})$$

where $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are the collected data values.

The posterior is proportional to the (conditional) likelihood of the observed responses given the random effects

$$\mathcal{L}(\beta, \Sigma_\epsilon; v_1, \dots, v_n) = \prod_{i=1}^n f(\mathbf{y}_i | v_i, \mathbf{x}_i, \mathbf{z}_i; \beta, \Sigma_\epsilon)$$

multiplied by the ‘prior’ for the random effects

$$\pi_0^v(v_1, \dots, v_n | \Sigma_u) = \prod_{i=1}^n \pi_0^{v_i}(v_i; \Sigma_v)$$

multiplied by the prior for the remaining unknown parameters.

Bayesian inference using MCMC

It is typical to assume a prior structure of the form

$$\pi_0(\beta, \Sigma_v, \Sigma_\epsilon) = \pi_0^{\beta, \epsilon}(\beta, \Sigma_\epsilon) \pi_0^v(\Sigma_v)$$

so that for the posterior we have

$$\begin{aligned} \pi(\beta, \Sigma_v, \Sigma_\epsilon, v_1, \dots, v_n | \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \\ \propto \mathcal{L}(\beta, \Sigma_\epsilon; v_1, \dots, v_n) \pi_0^v(v_1, \dots, v_n | \Sigma_u) \\ \pi_0^{\beta, \epsilon}(\beta, \Sigma_\epsilon) \pi_0^v(\Sigma_v) \end{aligned}$$

To perform MCMC using Metropolis-Hastings on the entire unknown parameter vector is very complicated. Therefore it is common to use a Metropolis-within-Gibbs strategy

- need to identify ‘blocks’ of parameters that can be updated jointly, conditional on other fixed values.

Blocks for the Gibbs sampler

The obvious partitioning is

- ▶ β
- ▶ Σ_ϵ
- ▶ Σ_v
- ▶ $v_1, \dots, v_n$

as there are natural conditional independencies that stem from the hierarchical structure.

Full conditionals

- β : we have

$$\pi(\underline{\beta}|\Sigma_v, \Sigma_\epsilon, v_1, \dots, v_n) \propto \mathcal{L}(\underline{\beta}, \Sigma_\epsilon; v_1, \dots, v_n) \pi_0^{\beta, \epsilon}(\underline{\beta}, \Sigma_\epsilon)$$

- Σ_ϵ : similarly

$$\pi(\underline{\Sigma}_\epsilon|\beta, \Sigma_v, v_1, \dots, v_n) \propto \mathcal{L}(\beta, \underline{\Sigma}_\epsilon; v_1, \dots, v_n) \pi_0^{\beta, \epsilon}(\beta, \underline{\Sigma}_\epsilon)$$

where the underlining indicates the argument to the function.

Full conditionals (cont.)

- Σ_v : we have

$$\pi(\underline{\Sigma_v} | \beta, \Sigma_\epsilon, v_1, \dots, v_n) \propto \pi_0^v(v_1, \dots, v_n | \underline{\Sigma_u}) \pi_0^v(\underline{\Sigma_u})$$

that is, **independent** of the response data $\mathbf{Y}$

- $v_1, \dots, v_n$

$$\pi(\underline{v_1, \dots, v_n} | \beta, \Sigma_\epsilon, \Sigma_v) \propto \mathcal{L}(\beta, \Sigma_\epsilon; \underline{v_1, \dots, v_n}) \pi_0^v(\underline{v_1, \dots, v_n} | \Sigma_u)$$

which further factorizes

$$\pi(\underline{v_1, \dots, v_n} | \beta, \Sigma_\epsilon, \Sigma_v) = \prod_{i=1}^n \pi(\underline{v_i} | \beta, \Sigma_\epsilon, \Sigma_v)$$

where

$$\pi(v_i | \beta, \Sigma_\epsilon, \Sigma_v) \propto f(\mathbf{y}_i | v_i, \mathbf{x}_i, \mathbf{z}_i; \beta, \Sigma_\epsilon) \pi_0^{v_i}(v_i | \Sigma_u)$$

Notes

1. Typically we factorize the prior for (β, Σ_ϵ) as

$$\pi_0^{\beta, \epsilon}(\beta, \Sigma_\epsilon) = \pi_0^{\beta|\epsilon}(\beta|\Sigma_\epsilon)\pi_0^\epsilon(\Sigma_\epsilon)$$

2. The covariance matrices Σ_ϵ and Σ_v are positive definite symmetric matrices; this must be respected when these parameters are sampled using MCMC.
3. The **Inverse Wishart** distribution is commonly chosen as a prior for a covariance matrix.

Example: Growth curves

Suppose that at time t the growth of the amount of bacteria in a plate is well represented using a curve of the form

$$\mathbb{E}[Y_t] = \beta_1 + \beta_2 t + \beta_3 \log t$$

However, due to experimental variation in the plate medium, growth profiles are slightly different for each plate.

We may use the statistical model

$$y_{it} = \beta_1 + \beta_2 t + \beta_3 \log t + v_i + \epsilon_{it}$$

Here we have a random effects model, with $p = 3$ and $q = 1$

$$\mathbf{x}_{it} = (1, t, \log t)$$

for each i , and $\mathbf{z}_i \equiv 1$ for all i .

Example: Growth curves (cont.)

This is termed the **random intercepts** model.

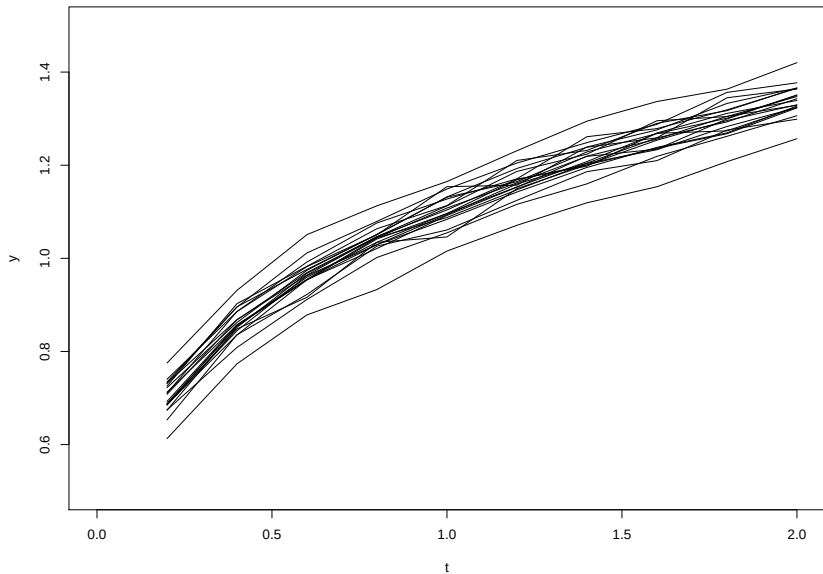
We let $\Sigma_v \equiv \sigma_v$, and suppose further that

$$\Sigma_\epsilon = \sigma_\epsilon^2 \mathbf{I}_T$$

that is, the residual errors are serially uncorrelated.

We have scalar variance parameters which makes sampling more straightforward.

Example: Growth curves (cont.)



Analysis

In our data set we have

- ▶ $n = 20$
- ▶ $T = 10$ with $t = 0.2, 0.4, \dots, 1.8, 2.0$.

For the priors, we select

- ▶ $\pi_0^{\beta|\epsilon}(\beta|\sigma_\epsilon) \equiv \text{Normal}(0, \sigma_\epsilon^2 \sigma_0^2 \mathbf{I}_p)$
- ▶ $\pi_0^\epsilon(\sigma_\epsilon^2) \equiv \text{InverseGamma}(a_\epsilon, b_\epsilon)$
- ▶ $\pi_0^v(\sigma_v^2) \equiv \text{InverseGamma}(a_v, b_v)$

We select the hyperparameters

- ▶ $\sigma_0^2 = 10^2$
- ▶ $a_\epsilon = 5, b_\epsilon = 0.05$
- ▶ $a_v = 20, b_v = 1$

Results

Run the Metropolis-within-Gibbs for 40000 iterations, after a burn-in of 100 iterations, and take every 20th sample.

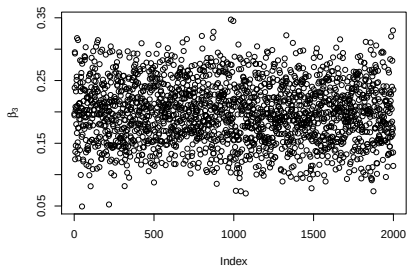
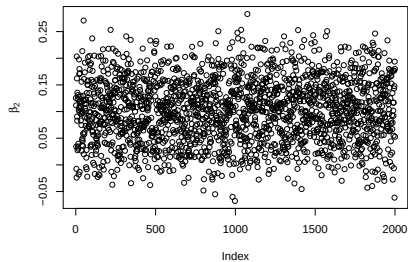
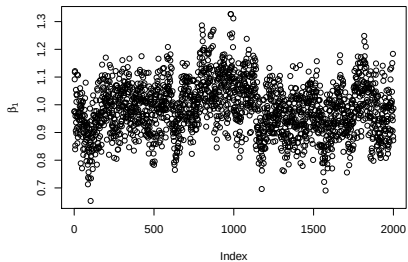
This takes roughly 6 minutes of computing time in R.

We inspect

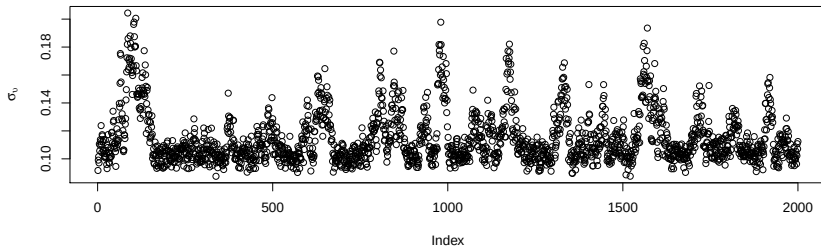
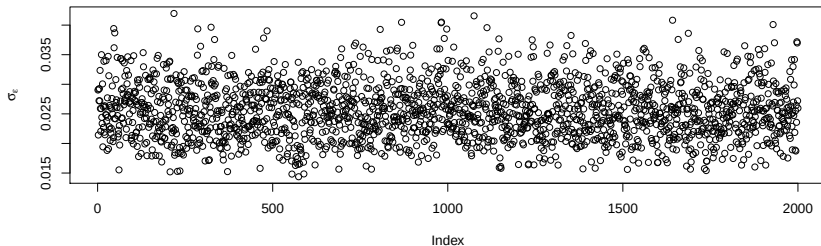
- ▶ trace plots
- ▶ ACF plots
- ▶ scatterplot matrices

for the posterior samples.

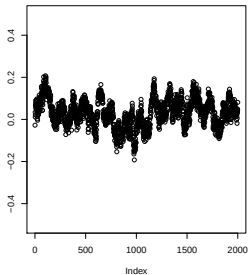
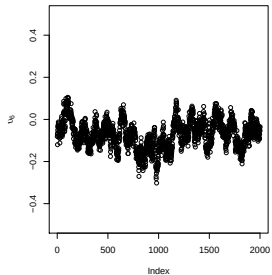
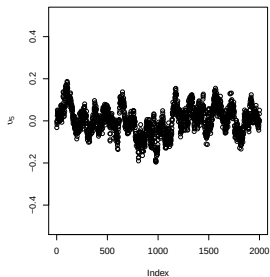
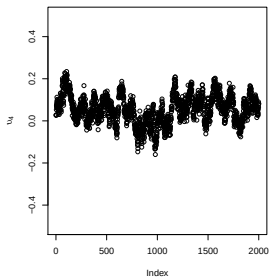
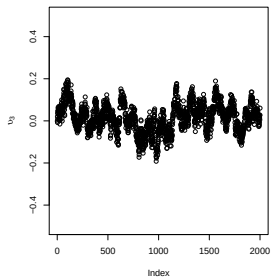
Results: posterior for β



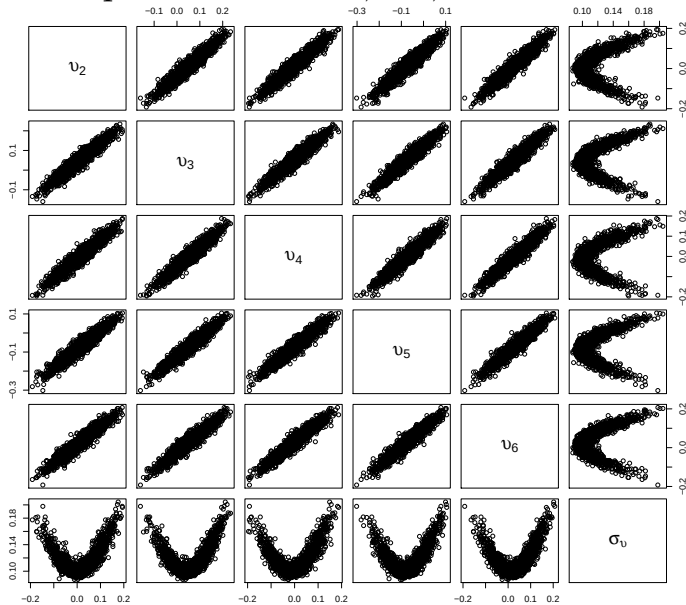
Results: posterior for σ_ϵ and σ_v



Results: posterior for v_i for $i = 1, \dots, 5$



Results: scatterplot v_i for $i = 1, \dots, 5$ and σ_v



Reparameterization

The above plots demonstrate that there are some extremely strong dependencies in the joint posterior. This is, in part, due to the parameterization used.

- ▶ by parameterizing using β_1 and the v_i , we have ensured that the v_i are ‘separated’ from the data $\mathbf{y}_i$ in the model hierarchy.
- ▶ this has the effect of inducing strong dependence across the v_i and with σ_v .

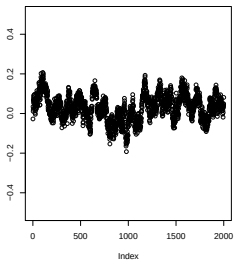
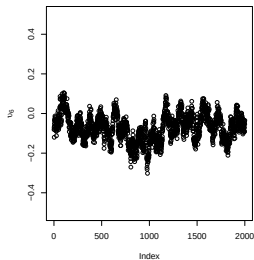
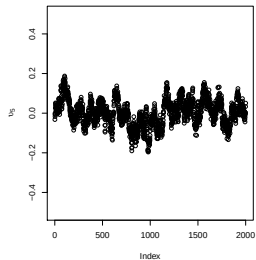
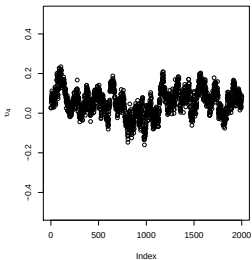
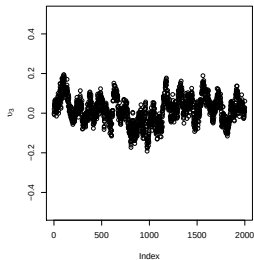
A preferable parameterization therefore is to write the model instead as

$$y_{it} = \xi_i + \beta_2 t + \beta_3 \log t + \epsilon_{it}$$

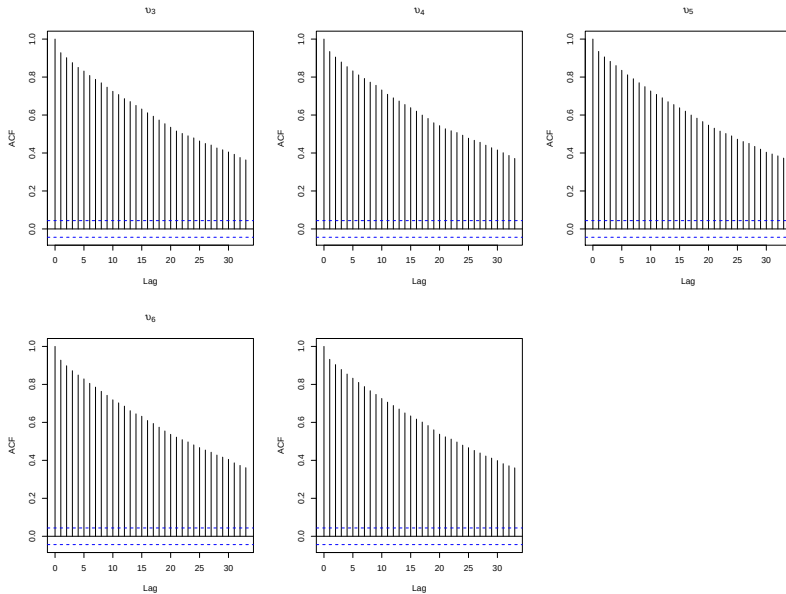
where

$$\xi_i = \beta_1 + v_i \quad i = 1, \dots, n$$

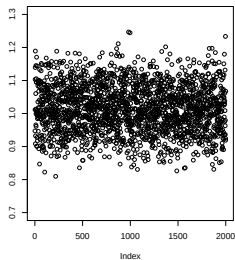
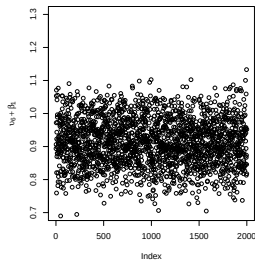
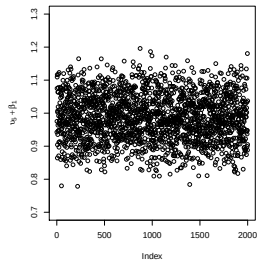
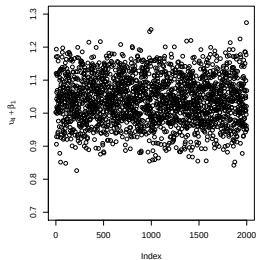
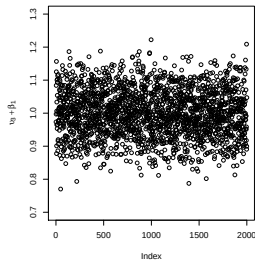
Results: posterior for v_i for $i = 1, \dots, 5$



Results: ACF for v_i for $i = 1, \dots, 5$



Results: posterior for $\xi_i = v_i + \beta_1$ for $i = 1, \dots, 5$



Results: ACF for $\xi_i = v_i + \beta_1$ for $i = 1, \dots, 5$

