Hierarchical linear models: a growth curve analysis using the Gibbs sampler


On the course web page:
- rats.dat,
- BUGS programs:
  1. Rats: a Normal hierarchical model
  2. Birats: a bivariate Normal hierarchical model

in BUGS Examples Volume I and II
http://www.mrc-bsu.cam.ac.uk/bugs/documentation/contents.html
**Gibbs sampling**

With 66 unobservable variables, Bayesian computation can be conceivably performed only by Gibbs sampling.

Gibbs sampling is indeed relatively straightforward to implement thanks to the

- **conditional independence structure** characterizing a hierarchical Bayesian model;
- **conditionally conjugate priors** at each stage in the hierarchy.

In the rats example, e.g., the prior of each parameter was chosen to be conjugate with the corresponding likelihood term that preceded it in the hierarchy.

The **conditionally conjugate priors** enables all of the full conditional distributions, necessary for implementing the Gibbs sampler, to arise in closed form.

**Computing the full conditionals from the DAG**

The full conditional for each variable can be inferred from the links and the conditional distributions in the **Directed Acyclic Graph**

The graphical model depicts the causal structure in the problem (see the DAG model for the rats example):

- Directed edges (stochastic or logical dependences) running into nodes (stochastic or deterministic variables and constants) represent direct influence of the antecedent on the descendants.
- Missing links represent irrelevance between nodes.

**A hierarchical (Normal) linear model**

or what a classical statistician might refer to as a random coefficient linear growth curve model:

**stage 1**

\[ Y_{ij} \sim \mathcal{N}(\alpha_i + \beta_i x_{ij}, \sigma^2) \]

**stage 2**

\[ \theta_i \equiv \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N}(\theta_0 \equiv \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix}, \Sigma), \]

\[ i = 1, \ldots, 30, j = 1, \ldots, 5. \]

**To perform a (fully) Bayesian analysis**, we must add at this stage \( \sigma^{-2} \sim \mathcal{G}(a, b) \), and at stage 3

\[ \theta_0 \sim \mathcal{N}(\eta, C) \text{ and } \Sigma^{-1} \sim \mathcal{W}((\rho R)^{-1}, \rho), \]

i.e. the priors for the founder (stochastic) nodes of the DAG.

All the priors for the founder (stochastic) nodes are conjugate. Their hyperparameters are assumed known.

- \( \sigma^{-2} \sim \mathcal{G}(\epsilon, \epsilon) \),
- \( \theta_0 \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0E-6 & 0 \\ 0 & 1.0E-6 \end{pmatrix} \right) \),
- \( \Sigma^{-1} \sim \mathcal{W} \left( \begin{pmatrix} 200 & 0 \\ 0 & 0.2 \end{pmatrix}, 2 \right) \),

where \( df \rho = 2 = rank(\Sigma) \), i.e. we choose the df as small as possible to leave the prior for the precision matrix \( \Sigma^{-1} \) vague. The scale matrix (divided per df) is our prior guess at the order of magnitude of the covariance matrix \( \Sigma \).
To add the probability structure to the model, we denote $V = \{v\}$ the collection of nodes and $p(v|pa(v))$ the conditional distribution of $v$ given its parent nodes (direct antecedents).

Then, the probability structure is defined by

$$p(V) = \prod_{v \in V} p(v|pa(v)),$$

assuming conditional independence wherever there is irrelevance between nodes.

Thus, it consists in the product of the priors for the nodes having no parents and the conditional distributions for the remaining children given their parents.

The marginal distribution of any node (given any observed data) nodes can be found via Bayes’ rule.

The full conditional for $v$ is computed as

$$p(v|V - v) \propto p(v, V - v) = p(V).$$

Thus it is proportional to the product of the terms in $p(V)$ containing $v$, i.e.,

$$p(v|V - v) \propto p(v|pa(v)) \prod_{w \in pa(w)} p(w|pa(w)),$$

where $\prod_{w \in pa(w)} p(w|pa(w))$ is the product of distributions in which $v$ is itself a parent.

Then, exploiting the DAG and conjugacy, the full conditionals in the Normal hierarchical model are:

- $\theta_i | \theta_0, \Sigma, y, \sigma \propto N(\theta_0, \Sigma) \cdot \prod_j N(\alpha_i + \beta_i x_{ij}, \sigma^2) \sim N(\ldots)$,
- $\theta_0 | \{\theta_i\}, \Sigma^{-1} \propto \theta_0 \sim N(\eta, C) \cdot \prod_i N(\theta_0, \Sigma) \sim N(\ldots)$,
- $\Sigma^{-1} | \{\theta_i\}, \theta_0 \propto W((\rho R)^{-1}, \rho) \cdot \prod_i N(\theta_0, \Sigma) \sim W(\ldots)$,
- $\sigma^{-2} | y, \{\theta_i\} \propto G(a, b) \cdot \prod_{i,j} N(\alpha_i + \beta_i x_{ij}, \sigma^2) \sim G(\ldots)$

Simplifying and improving Gibbs sampling computation:

**Variable transformation**: centering the covariates around the grand mean.

Variable transformation, here, is strategically used for improving MCMC. In fact, cross-correlation between intercept and slope is reduced with covariates centering.

The univariate version of the Normal hierarchical model is:

stage 1 $Y_{ij} \overset{\text{ind}}{\sim} N(\alpha_i + \beta_i(x_{ij} - \bar{x}), \sigma^2)$, so that $\alpha_i$ and $\beta_i$ are independent at priori (full balanced data).

stage 2 $\alpha_i \overset{\text{i.i.d.}}{\sim} N(\alpha_0, \sigma_{\alpha}^2)$, $\beta_i \overset{\text{i.i.d.}}{\sim} N(\beta_0, \sigma_{\beta}^2)$,

stage 3 $\alpha_0 \sim N(0, .00001)$, $\beta_0 \sim N(0, .00001)$, $\sigma^2 \sim G(\epsilon, \epsilon)$, $\sigma_{\beta}^{-2} \sim G(\epsilon, \epsilon)$
Diagnostic strategy

Using WinBUGS and *Rats: a Normal hierarchical model* program (in BUGS Examples Volume I) we run a few Gibbs sampling chains according to the following diagnostic strategy:

- Run 3 or 5 parallel chains, with starting points drawn from a distribution believed to be overdispersed w.r.t. the stationary distribution;
- Visually inspect these chains by overlaying them;
- Gelman-Rubin statistic;
- check autocorrelation;
- investigate crosscorrelation.

CODA

We summarize the Gibbs samples generated by BUGS using CODA:

*Convergence Diagnostics and Output Analysis*

It is available for R the R-CODA version:

Suite of S-functions to analyse output from BUGS and any other MCMC algorithm.