6. A HIERARCHICAL MODEL

Applications of hierarchical models of the kind introduced by Lindley and Smith (1972) abound in fields as diverse as educational testing (Rubin 1981), cancer studies (DuMouchel and Harris 1983), and biological growth curves (Strenio, Weisberg, and Bryk 1983). However, both Bayesian and empirical Bayesian models are typically forced to invoke a number of approximations, whose consequences are often unclear under the multiparameter likelihoods induced by the modelling. See, for example, Morris (1983), Racine-Poon (1985), and Racine-Poon and Smith (1990) for details of some approaches to to implementing hierarchical model analysis. By contrast, a full implementation of the Bayesian approach is easily achieved using the Gibbs sampler, at least for the widely used normal hierarchical model structure.

For illustration, we focus on the following population grown problem. In a study conducted by the CIBA-GEIGY company, the weights of 30 young rats in a control group were meassured weekly for five weeks. The data are given in Table 3, with weight measurements available for all five weeks. Later we discuss the substantive problem of comparison with data from a treatment group. Initially, however, we shall focus attention on the control group in order to illustrate the Gibbs sampling methodology.

Table 3. Rat Population Growth Data: Control Group

Rat\Week	1	2	3	4	5	Rat\Week	1	2	3	4	5
1	151	199	246	283	320	16	160	207	248	288	324
2	145	199	249	293	354	17	142	187	234	280	316
3	147	214	263	312	328	18	156	203	243	283	317
4	155	200	237	272	297	19	157	212	259	307	336
5	135	188	230	280	323	20	152	203	246	286	321
6	159	210	252	298	331	21	154	205	253	298	334
7	141	189	231	275	305	22	139	190	225	267	302
8	159	201	248	297	338	23	146	191	229	272	302
9	177	236	285	340	376	24	157	211	250	285	323
10	134	182	220	260	296	25	132	185	237	286	331
11	160	208	261	313	352	26	160	207	257	303	345
12	143	188	220	273	314	27	169	216	261	295	333
13	154	200	244	289	325	28	157	205	248	289	316
14	171	221	270	326	358	29	137	180	219	258	291
15	163	216	242	281	312	30	153	200	244	286	324

NOTE: $x_{i1} = 8$, $x_{i2} = 15$, $x_{i3} = 22$, $x_{i4} = 29$, $x_{i5} = 36$ days; $i = 1, ..., 30$.

For the time period considered, it is reasonable to assume individual straight-line growth curves, although non-linear curves can be handled as well. We also assume homoscedastic normal measurement errors (nonhomogeneous varianes can be accommodated as in the previous section), so that

$$Y_{ij} \sim N(\alpha_i + \beta_i x_{ij}, \sigma_c^2),\ i = 1, ..., k;\ j = 1, ..., n_i,$$

provides the full measurement model (with $k = 30$, $n_i = 5$, and x_{ij} denoting the age in days of the ith rate when measurement j was taken). The population structure is modeled as

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim N\left\{ \begin{pmatrix} \alpha_c \\ \beta_c \end{pmatrix}, \Sigma_c \right\}, \qquad i = 1, ..., k,$$

assuming independence throughout. A full Bayesian analysis now requires the specification of a prior for σ_c^2, $\mu_c = (\alpha_c, \beta_c)^T$, and Σ_c. Typical inferences of interest in such studies include marginal posteriors for the population parameters (α_c, β_c) and predictive intervals for individual future growth given the first-week measurement. We shall see that these are easily obtained using the Gibbs sampler.

For the prior specification, we assume independence, as is customary, taking

$$[\mu_c, \Sigma_c^{-1}, \sigma_c^2] = [\mu_c][\Sigma_c^{-1}][\sigma_c^2]$$

to have a normal-Wishart-inverse-gamma form:

$$\begin{aligned} [\mu_c] &= N(\eta, C), \\ [\Sigma_c^{-1}] &= W((\rho R)^{-1}, \rho), \\ [\sigma_c^2] &= IG\left(\frac{\nu_0}{2}, \frac{\nu_0 \tau_0^2}{2}\right). \end{aligned}$$

Rewriting the measurement model for the ith individual as $Y_i \sim N(X_i\theta, \sigma_c^2 I_{n_i})$ where $\theta_i = (\alpha_i, \beta_i)^T$ and X_i denotes the appropriate design matrix and defining

$$\begin{aligned} Y &= (Y_1, ..., Y_k)^T, \quad \overline{\theta} = k^{-1} \sum_{i=1}^{k} \theta_i, \quad n = \sum_{i=1}^{k} n_i, \\ D_i &= \sigma_c^{-1} X_i^T X_i + \Sigma_c^{-1}, \\ V &= (k\Sigma_c^{-1} + C^{-1})^{-1}, \end{aligned}$$

the Gibbs sampler for $\theta = (\theta_1, ..., \theta_k)$, Σ_c, and σ_c^2 (a total of 66 parameters in the above example) is straightforwardly seen to be specified by the conditional distributions

$$\begin{aligned} [\theta_i \,|\, Y, \mu_c, \Sigma_c^{-1}, \sigma_c^2] &= N\{D_i(\sigma_c^{-2} X_i^T Y_i + \Sigma_c^{-1}\mu_c),\ D_i\}, \quad i = 1, ..., k \\ [\mu_c \,|\, Y, \{\theta\}, \Sigma_c^{-1}, \sigma_c^2] &= N\{V(k\Sigma_c^{-1}\overline{\theta} + C^{-1}\eta),\ V\}, \\ [\Sigma_c^{-1} \,|\, Y, \{\theta\}, \mu_c, \sigma_c^2] &= W\left\{ \left[\sum_i (\theta_i - \mu_c)(\theta_i - \mu_c)^T + \rho R \right]^{-1},\ k + \rho \right\}, \\ [\sigma_c^2 \,|\, Y, \{\theta\}, \mu_c, \Sigma_c^{-1}] &= IG\left(\frac{n + \nu_0}{2},\ \frac{1}{2} \left[\sum_i (Y_i - X_i\theta_i)^T (Y_i - X_i\theta_i) + \nu_0 \tau_0^2 \right] \right) \end{aligned} \tag{5}$$

For the analysis of the rat growth data given above, the hyperparameter specification was defined by

$$C^{-1} = 0, \quad \nu_0 = 0, \quad p = 2, \quad R = \begin{pmatrix} 100 & 0 \\ 0 & 0.1 \end{pmatrix},$$

[because $C^{-1} = 0$ the value of η disappears entirely from the full conditionals] reflecting rather vague initial information relative to that to be provided by the data. Simulation from the Wishart distribution for the 2×2 matrix Σ_c^{-1} is easily accomplished using the algorithm of Odell and Feiveson (1966): with $G(\cdot, \cdot)$ denoting gamma distributions, draw independently from

$$[U_1] = G\left(\frac{\nu}{2}, \frac{1}{2}\right),$$
$$[U_2] = G\left(\frac{\nu - 1}{2}, \frac{1}{2}\right),$$

and

$$[N] = N(0, 1);$$

set

$$W = \begin{bmatrix} U_1 & N\sqrt{U_1} \\ N\sqrt{U_1} & U_2 + N^2 \end{bmatrix};$$

then if $S^{-1} = (H^{1/2})^T (H^{1/2})$,

$$\Sigma_c^{-1} = (H^{1/2})^T W (H^{1/2}) \sim W(S^{-1}, \nu).$$

The iterative process was monitored by observing empirical Q–Q plots for successive samples from α_c, β_c, σ_c^2, and the eigenvalues of Σ_c^{-1}. Though the α_i and β_i are of less interest, spot checking revealed satisfactory convergence, not surprising in view of (5), which suggests that convergence for the θ_i is comparable to that of μ_c. For the data set summarized in Table 3, convergence was achieved with about 35 cycles of $m = 50$ drawings.

REFERENCES

DuMouchel, W. H., and Harris, J. E. (1983), "Bayes methods for Combining the Results of Cancer studies in Humans and Other Species" (with discussion), *Journal of the American Statistical Association*, 78, 293–305.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model" (with discussion), *Journnal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Morris, C. (1983), "Parametric Empirical bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78, 47–59.

Odell, P. L., and Feiveson, A. H. (1966), "A Numerical Procedure to Generate a Sample Covariance Matrix," *Journal of the American Statistical Association*, 61, 198–203.

Racine-Poon, A. (1985), "A Bayesian Approach to Non-linear Random Effects Models, *Biometrics*, 41, 1015–1024.

Racine-Poon, A., and Smith, A. F. M. (1990), "Population Models," in *Statistical Methodology in the Pharmaceutical Sciences*, ed. D. Berry, New York: Marcel Dekker.

Rubin, D. B. (1981), "Estimation in Parallel Randomized Experiments, *Journal of Educational Statistics*, 6, 377–401.

Strenio, J. F., Weisberg, H. I., and Bryk, A. S. (1983), "Empirical Bayes Estimation of Individual Growth-Curve Parameters and Their Relationship to Covariates, *Biometrics*, 39, 71–86.

Taken from "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling", Alan E. Gelfand, Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith, *Journal of the American Statistical Association* **85** (412) (1990), 972–985, Section 6, pp. 978–979. See also B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis (2nd edn)*, Chapman and Hall 2000, pp. 148-149.