and must be obtained before the algorithm can be applied. Fortunately this is often possible for 'nice' models, where the full conditional densities have conjugate forms.

Example 11.22 (Random effects model) The sampling model in the simplest normal one-way layout is

$$y_{tr} = \theta_t + \varepsilon_{tr}, \quad t = 1, \ldots, T, \; r = 1, \ldots, R,$$

where $\theta_1, \ldots, \theta_T \overset{\text{iid}}{\sim} N(\mu, \sigma_\theta^2)$ and independent of this $\varepsilon_{tr} \overset{\text{iid}}{\sim} N(0, \sigma^2)$. The focus of interest is usually σ^2 and σ_θ^2.

Bayesian analysis requires prior information, which we suppose to be expressed through the conjugate densities

$$\mu \sim N(\mu_0, \tau^2), \quad \sigma^2 \sim IG(\alpha, \beta), \quad \sigma_\theta^2 \sim IG(\alpha_\theta, \beta_\theta).$$

The full posterior density is then

$$\pi(\mu, \theta, \sigma^2, \sigma_\theta^2 \mid y) \propto f(y \mid \theta, \sigma^2) f(\theta \mid \mu, \sigma_\theta^2) \pi(\mu) \pi(\sigma^2) \pi(\sigma_\theta^2). \tag{11.42}$$

We now take $(U_1, U_2, U_3, U_4) = (\sigma_\theta^2, \sigma^2, \mu, \theta)$, and calculate the full conditional densities needed for Gibbs sampling, always treating the data y as fixed. Each calculation requires integration over just one parameter. For example,

$$\begin{aligned}
\pi(\sigma_\theta^2 \mid \sigma^2, \mu, \theta, y) &= \frac{f(y \mid \theta, \sigma^2) f(\theta \mid \mu, \sigma_\theta^2) \pi(\mu) \pi(\sigma^2) \pi(\sigma_\theta^2)}{\int f(y \mid \theta, \sigma^2) f(\theta \mid \mu, \sigma_\theta^2) \pi(\mu) \pi(\sigma^2) \pi(\sigma_\theta^2) d\sigma_\theta^2} \\
&= \frac{f(\theta \mid \mu, \sigma_\theta^2) \pi(\sigma_\theta^2)}{\int f(\theta \mid \mu, \sigma_\theta^2) \pi(\sigma_\theta^2) d\sigma_\theta^2} \\
&= \pi(\sigma_\theta^2 \mid \mu, \theta).
\end{aligned}$$

Similar calculations reveal that $\pi(\theta \mid \sigma_\theta^2, \sigma^2, \mu, y)$ does not simplify, but that

$$\pi(\sigma^2 \mid \sigma_\theta^2, \mu, \theta, y) = \pi(\sigma^2 \mid \theta, y), \quad \pi(\mu \mid \sigma_\theta^2, \sigma^2, \theta, y) = \pi(\mu \mid \sigma_\theta^2, \theta). \tag{11.43}$$

Arguments paralleling those in Example 11.12 lead to

$$\sigma_\theta^2 \mid \mu, \theta \sim IG\left(\alpha_\theta + \frac{1}{2}T, \beta_\theta + \frac{1}{2}\sum_{t=1}^{T}(\theta_t - \mu)^2\right), \tag{11.44}$$

$$\sigma^2 \mid \theta, y \sim IG\left(\alpha + \frac{1}{2}TR, \beta + \frac{1}{2}\sum_{t=1}^{T}\sum_{r=1}^{R}(y_{tr} - \theta_t)^2\right), \tag{11.45}$$

$$\mu \mid \sigma_\theta^2, \theta \sim N\left(\frac{\sigma_\theta^2 \mu_0 + \tau^2 \sum_{t=1}^{T} \theta_t}{\sigma_\theta^2 + T\tau^2}, \frac{\sigma_\theta^2 \tau^2}{\sigma_\theta^2 + T\tau^2}\right). \tag{11.46}$$

The conditional density $\pi(\theta \mid \sigma_\theta^2, \sigma^2, \mu, y)$ is most readily calculated by noting that given μ, σ_θ^2 and σ^2, the statistic $\overline{y}_t$ is sufficient for θ_t, with distribution $N(\theta_t, \sigma^2/R)$, while the prior density for θ_t given σ_θ^2, σ^2, and μ is $N(\mu, \sigma_\theta^2)$. Hence the posterior density for θ_t is

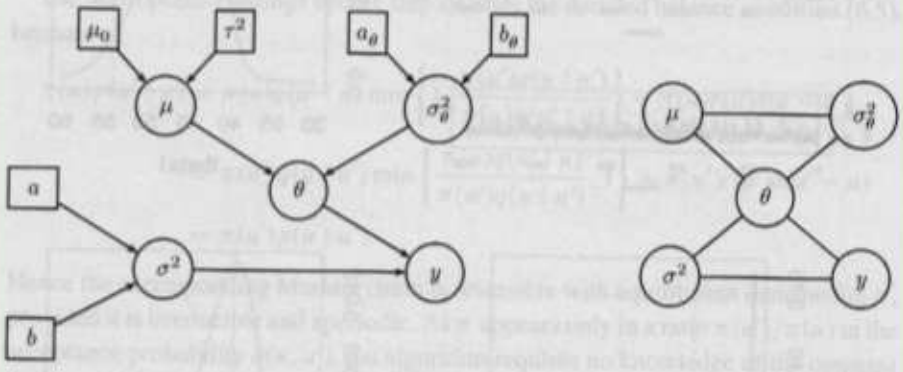
$$\theta_t \mid \sigma_\theta^2, \sigma^2, \mu, y \sim N\left(\frac{R\sigma_\theta^2 \overline{y}_t + \sigma^2 \mu}{R\sigma_\theta^2 + \sigma^2}, \frac{\sigma_\theta^2 \sigma^2}{R\sigma_\theta^2 + \sigma^2}\right), \quad t = 1, \ldots, T, \tag{11.47}$$

and the θ_t are conditionally independent.

Table 11.9 Estimated posterior means and standard deviations for the model fitted to the blood data, and simple frequentist estimates from analysis of variance.

Figure 11.7 Graphs for random effects model of Example 11.22. Left: directed acyclic graph showing dependence of random variables (circles) on themselves and on fixed quantities (rectangles). Right: conditional independence graph, formed by moralizing the directed acyclic graph, that is, joining parents and dropping arrowheads.

	σ_θ^2	σ^2	μ	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
Estimate	23.8	126.4	41.9	53.9	43.0	34.9	39.9	41.3	38.6
Posterior mean	17.1	138.0	41.9	45.8	42.3	39.6	41.2	41.7	40.8
Posterior SD	30.3	33.8	2.4	4.1	2.9	3.4	2.9	2.9	3.0



Expressions (11.44)–(11.47) give the steps required for an iteration of the Gibbs sampler. As the T updates in (11.47) are independent, they may all be performed at once, if the programming language used permits simultaneous generation of several non-identically-distributed normal variates.

Ideas from Section 6.2.2 render the structure of the full conditional densities more intelligible. Figure 11.7 shows the directed acyclic graph and the corresponding conditional independence graph for the present model. Each of μ, σ_θ^2, and σ^2 has two hyperparameters, considered fixed, and μ and σ_θ^2 are parents of $\theta_1, \ldots, \theta_T$. Each iteration of the Gibbs sampler traverses the parameter nodes in the conditional independence graph, simulating from the full conditional distribution corresponding to each node with remaining parameters set at their current values. The data y are held fixed throughout.

We applied this algorithm to the data in Table 9.22 on the stickiness of blood. For illustration we took $\alpha = \alpha_\theta = 0.5$, $\beta = \beta_\theta = 1$, $\mu_0 = 0$, and $\tau^2 = 1000$, and generated starting-values for the parameters from the uniform distribution on (0, 100). We ran 25 independent chains with $I = 1000$.

Figure 11.8 shows simulated series for three parameters and estimates of their posterior densities. The burn-in period seems to last for about $B = 100$ iterations, after which the chains seem stable. The chain for σ_θ^2 makes some large positive excursions, but the others seem fairly homogeneous, though they both show fairly strong autocorrelations. Estimated variance inflation factors are about 10 for σ_θ^2 and μ, but only 1–2.5 for the other parameters, consistent with the top left panels of the figure.

Table 11.9 shows the posterior means and standard deviations for the parameters, with their frequentist estimates. The posterior mean for μ is essentially equal to the overall average $\bar{y}$, but the posterior densities of the θ_t are strongly shrunk towards it, because there is evidence that σ_θ^2 is small; its posterior 0.1, 0.5, and 0.9 quantiles

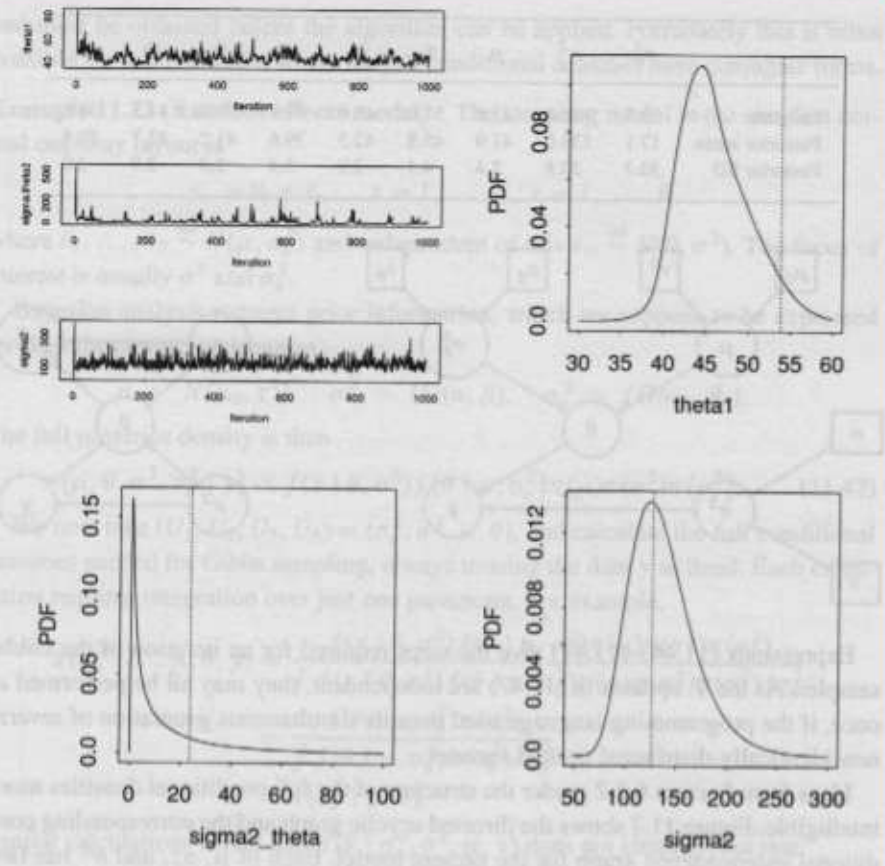


Figure 11.8 Gibbs sampler for normal components of variance model and blood data. Top left: time plots of θ_1, σ_θ^2, and σ^2. The other panels show estimated posterior densities for these parameters, based on applying analogues of (11.41) to the last 200 estimates from each of 25 parallel chains of length 1000. Frequentist estimates are shown as the dotted vertical lines.

are 0.46, 7.1, and 42.1. The variability mostly comes from measurement error, not inter-subject variation. ■

Metropolis–Hastings algorithm

The Gibbs sampler is easy to program, but if the full conditional densities it involves are unavailable or too nasty then a more general algorithm may be needed. A powerful approach known as the *Metropolis–Hastings algorithm* works as follows. In order to update the current value u of a Markov chain, a new value u' is generated using a proposal density $q(u' \mid u)$. Any density q can be used provided $q(u' \mid u) > 0$ if and only if $q(u \mid u') > 0$ and the resulting chain has the properties desired. Having generated u', a move from u to u' is accepted with probability

$$a(u, u') = \min\left\{1, \frac{\pi(u')q(u \mid u')}{\pi(u)q(u' \mid u)}\right\},$$

but otherwise the chain remains at u. Hence the probability density for a move to u', given that the chain has current value u, is

$$p(u' \mid u) = q(u' \mid u)a(u, u') + r(u)\delta(u - u'),$$

δ denotes the Dirac delta function.

Subject					
1	2	3	4	5	6
68	49	41	33	40	30
42	52	40	27	45	42
69	41	26	48	50	35
64	56	33	54	41	44
39	40	42	42	37	49
66	43	27	56	34	25
29	20	35	19	42	45

Table 9.22 Blood data: seven measurements from each of six subjects on a property related to the stickiness of their blood.

variance $\sigma_b^2/T + \sigma^2/(TR) = (\sigma^2 + R\sigma_b^2)/(TR)$, which is estimated unbiasedly by $SS_b/\{(T-1)TR\}$, independent of $\overline{y}_{..}$, and confidence intervals are based on the t_{T-1} distribution of $(\overline{y}_{..} - \mu)/[SS_b/\{(T-1)TR\}]^{1/2}$.

The assumptions of homogeneous variance across all blocks and of normality can be checked using probability plots.

Example 9.14 (Blood data) Six subjects were selected at random from a large population, and a property related to stickiness of samples of blood was measured seven times on each subject. The data are given in Table 9.22.

For these data, $SS_w = 4549.7$ and $SS_b = 1466.0$ on 36 and 5 degrees of freedom respectively. A point estimate of the variance for different measurements on the same subject is $SS_w/36 = 126.4$. and a point estimate of the variance of mean stickiness between subjects is $(SS_b/5 - SS_w/36)/7 = 23.83$. An equi-tailed 90% confidence interval for the ratio σ_b^2/σ^2 based on (9.9) is $(-0.01, 1.34)$; this overlaps the negative half-axis and would not usually be appropriate. ■

Nested variation

The previous example had two levels of nested variation, for subjects and for measurements. In practice data with several levels of variation arise. Consider for example comparison of the success of a surgical procedure, measured on a continuous scale. Data are available on patients, P of whom are treated by each surgeon and with S surgeons working at H hospitals. We suppose that surgeons at different hospitals are independent, and likewise for the patients, so patients are nested within surgeons within hospitals — there is no relation between the first patient of surgeon 1 at hospital 1 and the first patient of surgeon 2 at hospital 1, nor between surgeon 1 at hospital 1 and surgeon 1 at hospital 2. Put another way, labels for patients can be permuted independently within each surgeon without changing the data structure, and likewise for surgeons within each hospital. A simple model for the outcome y_{hsp} for the pth patient of the sth surgeon at the hth hospital is

$$y_{hsp} = \mu + b_h + e_{hs} + \varepsilon_{hsp}, \quad h = 1, \ldots, H, s = 1, \ldots, S, p = 1, \ldots, P, \tag{9.10}$$

Term

Between hospitals

Between surgeons within hospitals

Between patients within surgeons

Table 9.23 Analysis of variance table for nested model. Each sum of squares is summed over h, s and p. Mean squares are formed by dividing sums of squares by their degrees of freedom. λ_b^2 and λ_e^2 are non-centrality parameters measuring differences among the b_h and e_{hs} when they are treated as fixed.