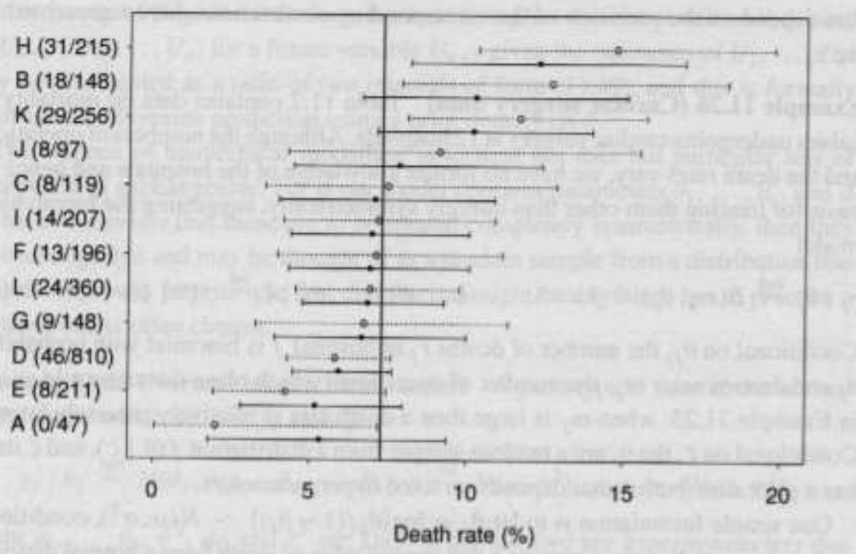


**Figure 11.11** Posterior summaries for mortality rates for cardiac surgery data. Posterior means and 0.95 equitailed credible intervals for separate analyses for each hospital are shown by hollow circles and dotted lines, while blobs and solid lines show the corresponding quantities for a hierarchical model. Note the shrinkage of the estimates for the hierarchical model towards the overall posterior mean rate, shown as the solid vertical line; the hierarchical intervals are slightly shorter than those for the simpler model.

**Example 11.27 (Spring barley data)** Table 10.21 contains data on a field trial intended to compare the yields of 75 varieties of spring barley allocated randomly to plots in three long narrow blocks. The data were analysed in Example 10.35 using a generalized additive model to accommodate the strong fertility trends over the blocks. In the absence of detailed knowledge about the varieties it seems natural to treat them as exchangeable, and we outline a Bayesian hierarchical approach. We also show how the fertility patterns may be modelled using a simple Markov random field.

Let $y = (y_1, \dots, y_n)^T$ denote the yields in the $n = 225$ plots and let $\psi_j$ denote the unknown fertility of plot $j$. Let $X$ denote the $n \times p$ design matrix that shows which of the $p = 75$ variety parameters $\beta = (\beta_1, \dots, \beta_p)^T$ have been allocated to the plots. Then a normal linear model for the yields is

$$y \mid \beta, \psi, \lambda_y \sim N_n(\psi + X\beta, I_n/\lambda_y), \tag{11.51}$$

where $\psi$ is the $n \times 1$ vector containing the fertilities and $\lambda_y$ is the unknown precision of the $y$s.

We take the prior density of $\lambda_y$ to be gamma with shape and scale parameters $a$ and $b$, $G(a, b)$, so that its prior mean and variance are $a/b$ and $a/b^2$, where $a$ and $b$ are specified. As there is no special treatment structure, we take for the $\beta_r$ the exchangeable prior $\beta \sim N_p(0, I_p/\lambda_\beta^{-1})$, with $\lambda_\beta \sim G(c, d)$ and $c$, $d$ specified. For the fertilities we take the normal Markov chain of Example 6.13, for which

$$\pi(\psi \mid \lambda_\psi) \propto \lambda_\psi^{n/2} \exp\left\{-\frac{1}{2}\lambda_\psi \sum_{i \sim j}(\psi_i - \psi_j)^2\right\}, \quad \lambda_\psi > 0, \tag{11.52}$$

the summation being over pairs of neighbouring plots and $\lambda_\psi^{-1}$ being the variance of differences between fertilities. Each $\psi_j$ occurs in $n_j$ terms, where $n_j = 1$ or 2 is the number of plots adjacent to plot $j$. The sum in (11.52) equals $\psi^{\mathrm{T}} W \psi$, where $W$ is the $n \times n$ tridiagonal matrix with elements

$$w_{ij} = \begin{cases} n_i, & i = j, \\ -1, & i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $W$ is block diagonal, with three blocks like the matrix $V$ in Example 6.13 with $\tau = 0$, corresponding to the three physical blocks of the experiment. We take $\lambda_\psi \sim G(g, h)$, with $g$ and $h$ specified.

With these conjugate prior densities, the joint posterior density is

$$\begin{aligned}\pi(\beta, \psi, \lambda) \propto \lambda_y^{n/2} \exp\left\{-\frac{1}{2}\lambda_y (y - \psi - X\beta)^{\mathrm{T}}(y - \psi - X\beta)\right\} \\ \times \lambda_\beta^{p/2} \exp\left(-\frac{1}{2}\lambda_\beta \beta^{\mathrm{T}}\beta\right) \times \lambda_\psi^{p/2} \exp\left(-\frac{1}{2}\lambda_\psi \psi^{\mathrm{T}} W \psi\right) \\ \times \lambda_y^{a-1} \exp(-b\lambda_y) \times \lambda_\beta^{c-1} \exp(-c\lambda_\beta) \times \lambda_\psi^{g-1} \exp(-h\lambda_\psi),\end{aligned}$$

where $\lambda = (\lambda_y, \lambda_\beta, \lambda_\psi)^{\mathrm{T}}$. The full conditional densities turn out to be

$$\beta \mid \psi, \lambda, y \sim N\{\lambda_y Q_\beta^{-1} X^{\mathrm{T}}(y - \psi), Q_\beta^{-1}\}, \tag{11.53}$$

$$\psi \mid \beta, \lambda, y \sim N\{\lambda_y Q_\psi^{-1}(y - X\beta), Q_\psi^{-1}\}, \tag{11.54}$$

$$\lambda_y \mid \psi, \beta, y \sim G\{a + n/2, b + (y - X\beta - \psi)^{\mathrm{T}}(y - X\beta - \psi)/2\}, \tag{11.55}$$

$$\lambda_\beta \mid \psi, \beta, y \sim G(c + p/2, d + \beta^{\mathrm{T}}\beta/2), \tag{11.56}$$

$$\lambda_\psi \mid \psi, \beta, y \sim G(g + n/2, h + \psi^{\mathrm{T}} W \psi/2), \tag{11.57}$$

where

$$Q_\beta = \lambda_y X^{\mathrm{T}} X + \lambda_\beta I_p, \quad Q_\psi = \lambda_y I_n + \lambda_\psi W.$$

The elements of $\lambda$ are independent conditional on the remaining variables. The relatively simple form of the densities in (11.53)–(11.57) suggests using a time-reversible Gibbs sampler, in which $\beta$, $\psi$, and $\lambda$ are updated in a random order at each iteration. The most direct approach to simulation in (11.53) and (11.54) is through Cholesky decomposition of $Q_\beta$ and $Q_\psi$: in (11.53), for example, we find the lower triangular matrix $L$ such that $LL^{\mathrm{T}} = Q_\beta^{-1}$, generate $\varepsilon \sim N_p(0, I_p)$, and let $\beta = \lambda_y Q_\beta^{-1} X^{\mathrm{T}}(y - \psi) + L\varepsilon$. The block diagonal structure of $W$ means that the $\psi$s for different blocks can be updated separately, so the largest Cholesky decomposition needed is that of a $75 \times 75$ matrix. An alternative is to update individual $\psi_j$s in a random order, but although the computational burden is smaller, the algorithm then converges more slowly than with direct use of (11.54).

Note the strong resemblance of (11.53) and (11.54) to the steps of the backfitting algorithm for the corresponding generalized additive model.

The missing response in block 3 is simply a further unknown whose value may be simulated using the relevant marginal density of (11.51). This adds a fourth component to the simulation in random order of $\beta$, $\psi$, and $\lambda$ at each iteration; there are no other changes to the algorithm.

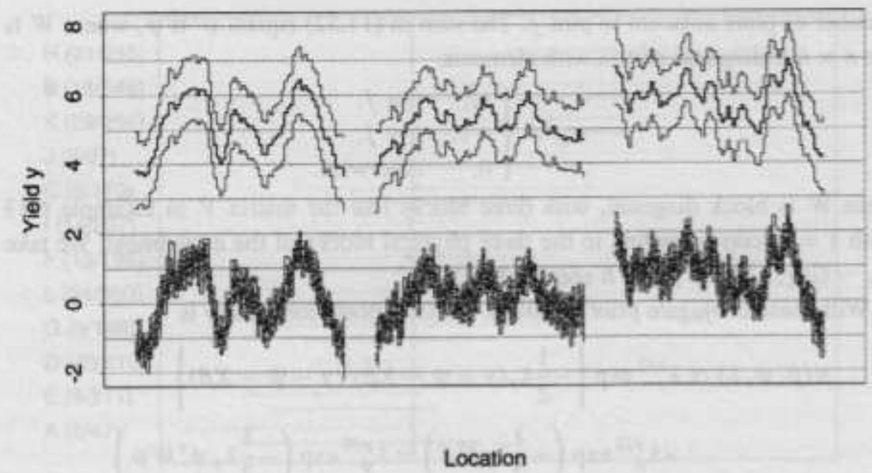


Figure 11.12 Posterior summaries for fertility trend $\psi$ for the three blocks of spring barley data, shown from left to right. Above: median trend (heavy) and overall 0.9 posterior credible bands. Below: 20 simulated trends from Gibbs sampler output.

If the matrix $X^{\mathrm{T}}X$ is diagonal, then the full conditional density for the $r$th variety effect has form

$$\beta_r \mid \psi, \lambda, y \sim N\left(\frac{\lambda_y m_r \bar{z}_r}{\lambda_\beta + \lambda_y m_r}, \frac{1}{\lambda_\beta + \lambda_y m_r}\right),$$

where $\bar{z}_r$ is the current average of $y_j - \psi_j$ for the $m_r$ plots receiving variety $r$. Thus the $\beta_r$ are shrunk towards zero by an amount that depends on the ratio $\lambda_\beta/\lambda_y$; with $\lambda_\beta = 0$ the mean for $\beta$ in (11.53) is the least squares estimate computed by regressing $y - \psi$ on the columns of $X$. Unlike in Example 11.25, however, the normal distributions of the $\beta_r$ are here averaged over the posterior densities of $\psi$, $\lambda_y$ and $\lambda_\beta$.

The algorithm described above was run with random initial values for 10,500 iterations. Time series plots of the parameters and log likelihood suggested that it had converged after 500 iterations, and inferences below are based on the final 10,000 iterations. The variance inflation factors $\widehat{\tau}$ were less than 4 for $\psi$ and $\beta$, about 44, 6 and 30 for $\lambda_y$, $\lambda_\tau$ and $\lambda_\psi$, and about 6 for $y_{187}$. Thus estimation for $\lambda_y$ is least reliable, being based on a sample equivalent to about 220 independent observations. A longer run of the algorithm would seem wise in practice. Based on this run, the posterior 0.9 credible intervals for $\lambda_y$, $\lambda_\psi$ and $\lambda_\beta$ were (5.2, 12.4), (5.0, 11.5) and (2.7, 5.7) respectively, and differences of two variety effects have posterior densities very close to normal with typical standard deviation of 0.35. The corresponding standard error for the generalized additive model was 0.41, so use of a hierarchical model and injection of prior information has increased the precision of these comparisons.

Figure 11.12 shows some simulated values of $\psi$ and pointwise 0.90 credible envelopes for the true $\psi$. These envelopes are constructed by joining the 0.05 quantiles of the fertilities simulated from the posterior density, for each location, and likewise with the 0.95 quantiles. By contrast with the analysis in Example 10.35, the effective degrees of freedom for $\psi$, controlled by $\lambda_\psi$, are here equal for each block, leading to apparent overfitting of the fertilities for block 2 compared to the generalized additive model. A difference between the models is that the current model corresponds

Table 11.12 Posterior probabilities that a variety is ranked among the best $r$ varieties, estimated from [illegible] iterations of Gibbs sampler.

| $r$ | Variety 56 | 35 | 72 | 31 | 55 | 47 | 54 | 18 | 38 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.327 | 0.182 | 0.149 | 0.129 | 0.075 | 0.055 | 0.019 | 0.015 | 0.012 | 0.006 |
| 2 | 0.518 | 0.357 | 0.299 | 0.270 | 0.174 | 0.136 | 0.050 | 0.042 | 0.035 | 0.020 |
| 5 | 0.814 | 0.690 | 0.643 | 0.621 | 0.486 | 0.416 | 0.234 | 0.183 | 0.153 | 0.106 |
| 10 | 0.959 | 0.908 | 0.887 | 0.871 | 0.795 | 0.743 | 0.560 | 0.497 | 0.429 | 0.344 |

to first differences of $\psi$ being a normal random sample, while in the earlier model the second differences are a normal random sample, giving a smoother fit.

The posterior probabilities that certain varieties rank among the $r$ best are given in Table 11.12. The ordering is somewhat different from that in Example 10.35, perhaps due to the slightly different treatment of fertility effects. As mentioned previously, no single variety strongly outperforms the rest, and future field experiments would have to include several of those included in this trial. This type of information is difficult to obtain using frequentist procedures, but is readily found by manipulating the output of the simulation algorithm described above.

This analysis is relatively easily modified when elements of the model are changed. Indeed the priors and other components chosen largely for convenience should be varied in order to assess the sensitivity of the conclusions to them; see Exercise 11.3.6. Metropolis–Hastings steps would then typically replace the Gibbs updates in the algorithm. ■

As mentioned above, more complicated hierarchies involve several layers of nested variation. Such models are widely used in certain applications, but their assessment and comparison can be difficult. For instance, shrinkage makes it unclear just how many parameters a hierarchical model has. Hierarchical modelling is an active area of current research.

*Justification of (11.49)*

To establish (11.49), suppose that $r$ lies in $0, \ldots, n$ and that $m > n$. Then exchangeability of $U_1, \ldots, U_m$ implies that the conditional probability

$$\Pr(U_1 + \cdots + U_n = r \mid U_1 + \cdots + U_m = s)$$

equals the probability of seeing $r$ 1's in $n$ draws without replacement from an urn containing $s$ 1's and $m - s$ 0's, which is $\binom{m}{n}^{-1}\binom{s}{r}\binom{m-s}{n-r}$ for $s = r, \ldots, m - (n - r)$ and zero otherwise. Hence

$$\begin{aligned}\Pr(U_1 + \cdots + U_n = r) &= \sum_{s=r}^{m-(n-r)} \binom{m}{n}^{-1} \binom{s}{r}\binom{m-s}{n-r} \Pr(U_1 + \cdots + U_m = s)\\ &= \binom{n}{r} \sum_{s=r}^{m-(n-r)} \frac{s^{(r)}(m-s)^{(n-r)}}{m^{(n)}} \Pr(U_1 + \cdots + U_m = s),\end{aligned}$$

**Table 10.21** Spring barley data (Besag et al., 1995). Spatial layout and plot yield at harvest $y$ (standardized to have unit crude variance) in a final assessment trial of 75 varieties of spring barley. The varieties are sown in three blocks, with each variety replicated thrice in the design. The yield for variety 27 is missing in the third block.

| | Block 1 | | Block 2 | | Block 3 | |
|---|---|---|---|---|---|---|
| Location $t$ | Variety | Yield $y$ | Variety | Yield $y$ | Variety | Yield $y$ |
| 1 | 57 | 9.29 | 49 | 7.99 | 63 | 11.77 |
| 2 | 39 | 8.16 | 18 | 9.56 | 38 | 12.05 |
| 3 | 3 | 8.97 | 8 | 9.02 | 14 | 12.25 |
| 4 | 48 | 8.33 | 69 | 8.91 | 71 | 10.96 |
| 5 | 75 | 8.66 | 29 | 9.17 | 22 | 9.94 |
| 6 | 21 | 9.05 | 59 | 9.49 | 46 | 9.27 |
| 7 | 66 | 9.01 | 19 | 9.73 | 6 | 11.05 |
| 8 | 12 | 9.40 | 39 | 9.38 | 30 | 11.40 |
| 9 | 30 | 10.16 | 67 | 8.80 | 16 | 10.78 |
| 10 | 32 | 10.30 | 57 | 9.72 | 24 | 10.30 |
| 11 | 59 | 10.73 | 37 | 10.24 | 40 | 11.27 |
| 12 | 50 | 9.69 | 26 | 10.85 | 64 | 11.13 |
| 13 | 5 | 11.49 | 16 | 9.67 | 8 | 10.55 |
| 14 | 23 | 10.73 | 6 | 10.17 | 56 | 12.82 |
| 15 | 14 | 10.71 | 47 | 11.46 | 32 | 10.95 |
| 16 | 68 | 10.21 | 36 | 10.05 | 48 | 10.92 |
| 17 | 41 | 10.52 | 64 | 11.47 | 54 | 10.77 |
| 18 | 1 | 11.09 | 63 | 10.63 | 37 | 11.08 |
| 19 | 64 | 11.39 | 33 | 11.03 | 21 | 10.22 |
| 20 | 28 | 11.24 | 74 | 10.85 | 29 | 10.59 |
| 21 | 46 | 10.65 | 13 | 11.35 | 62 | 11.35 |
| 22 | 73 | 10.77 | 43 | 10.25 | 5 | 11.39 |
| 23 | 37 | 10.92 | 3 | 10.08 | 70 | 10.59 |
| 24 | 55 | 12.07 | 53 | 10.25 | 13 | 11.26 |
| 25 | 19 | 11.03 | 23 | 9.57 | 11 | 11.79 |
| 26 | 10 | 11.64 | 62 | 11.34 | 44 | 12.25 |
| 27 | 35 | 11.37 | 52 | 10.19 | 36 | 12.23 |
| 28 | 26 | 10.34 | 12 | 10.80 | 52 | 10.84 |
| 29 | 17 | 9.52 | 2 | 10.04 | 60 | 10.92 |
| 30 | 71 | 8.99 | 32 | 9.69 | 68 | 10.41 |
| 31 | 8 | 8.34 | 22 | 9.36 | 3 | 10.96 |
| 32 | 62 | 9.25 | 42 | 9.43 | 19 | 9.94 |
| 33 | 44 | 9.86 | 72 | 11.46 | 67 | 11.27 |
| 34 | 53 | 9.90 | 73 | 9.29 | 59 | 11.79 |
| 35 | 74 | 11.04 | 25 | 10.10 | 2 | 11.51 |
| 36 | 20 | 10.30 | 45 | 9.53 | 75 | 11.64 |
| 37 | 56 | 11.56 | 15 | 10.55 | 27 | — |
| 38 | 29 | 9.69 | 35 | 11.34 | 43 | 9.78 |
| 39 | 2 | 10.68 | 66 | 11.36 | 51 | 8.86 |
| 40 | 47 | 10.91 | 5 | 10.88 | 10 | 10.28 |
| 41 | 11 | 10.05 | 56 | 11.61 | 35 | 12.15 |
| 42 | 38 | 10.80 | 46 | 10.33 | 74 | 10.36 |
| 43 | 65 | 10.06 | 71 | 10.53 | 66 | 9.59 |
| 44 | 13 | 10.04 | 51 | 8.67 | 34 | 10.53 |
| 45 | 31 | 10.50 | 21 | 9.56 | 18 | 11.26 |
| 46 | 40 | 9.51 | 1 | 9.95 | 50 | 10.37 |
| 47 | 4 | 9.20 | 31 | 11.10 | 42 | 10.10 |
| 48 | 67 | 9.74 | 11 | 10.11 | 1 | 9.95 |
| 49 | 22 | 8.84 | 41 | 9.36 | 58 | 9.80 |
| 50 | 49 | 9.33 | 61 | 10.23 | 26 | 10.58 |
| 51 | 58 | 9.51 | 55 | 11.38 | 41 | 9.31 |
| 52 | 43 | 9.35 | 14 | 11.30 | 25 | 9.29 |

Table 10.21 (cont.)

| Location $t$ | Block 1 | | Block 2 | | Block 3 | |
|---|---|---|---|---|---|---|
| | Variety | Yield $y$ | Variety | Yield $y$ | Variety | Yield $y$ |
| 53 | 7 | 9.01 | 44 | 10.90 | 33 | 10.03 |
| 54 | 25 | 10.58 | 34 | 10.97 | 9 | 9.49 |
| 55 | 61 | 11.03 | 54 | 12.22 | 17 | 11.52 |
| 56 | 16 | 9.89 | 24 | 10.10 | 57 | 12.24 |
| 57 | 52 | 11.39 | 4 | 11.22 | 65 | 11.64 |
| 58 | 70 | 11.24 | 65 | 10.01 | 49 | 10.74 |
| 59 | 34 | 12.18 | 75 | 10.29 | 73 | 10.29 |
| 60 | 42 | 10.21 | 38 | 10.95 | 7 | 10.25 |
| 61 | 24 | 11.08 | 17 | 9.66 | 23 | 11.39 |
| 62 | 33 | 11.05 | 68 | 9.31 | 72 | 13.34 |
| 63 | 51 | 10.29 | 7 | 8.84 | 55 | 12.73 |
| 64 | 60 | 10.57 | 27 | 10.64 | 31 | 12.62 |
| 65 | 69 | 10.42 | 58 | 9.45 | 39 | 10.19 |
| 66 | 15 | 10.49 | 48 | 9.66 | 47 | 11.61 |
| 67 | 6 | 10.00 | 28 | 9.85 | 15 | 10.52 |
| 68 | 63 | 9.23 | 60 | 9.24 | 20 | 9.07 |
| 69 | 54 | 10.57 | 30 | 10.11 | 61 | 10.76 |
| 70 | 18 | 10.27 | 70 | 9.63 | 28 | 9.91 |
| 71 | 45 | 8.86 | 20 | 9.04 | 53 | 10.17 |
| 72 | 72 | 9.45 | 9 | 8.43 | 69 | 8.68 |
| 73 | 9 | 8.03 | 40 | 10.97 | 45 | 8.74 |
| 74 | 36 | 9.22 | 50 | 8.98 | 12 | 9.15 |
| 75 | 27 | 8.70 | 10 | 9.88 | 4 | 9.39 |

### 10.7.3 More general models

We now consider how the discussion above should be modified when there are explanatory variables as well as a smooth variable, treating certain covariates nonparametrically and others not, and allowing the response to have a density other than the normal.

Let the data consist of independent triples $(x_1, t_1, y_1), \dots, (x_n, t_n, y_n)$, with $j$th log likelihood contribution $\ell_j(\eta_j, \kappa)$, where $\eta_j = x_j^{\mathrm{T}}\beta + g(t_j)$; for now we suppress dependence on $\kappa$. Then the analogue of (10.47) is the *penalized log likelihood*

$$\ell_\lambda(\beta, g) = \sum_{j=1}^{n} \ell_j(\eta_j) - \frac{1}{2}\lambda \int_a^b \{g''(t)\}^2\, dt, \quad \lambda > 0, \qquad (10.49)$$

where $a$ and $b$ are chosen so that $a < t_1, \dots, t_n < b$. If all the $t_j$ are distinct and $\lambda = 0$, the maximum is obtained by choosing $g_j = g(t_j)$ to maximise the $j$th log likelihood contribution, but this is not useful because the resulting model has $n$ parameters and is too rough. The integral in (10.49) penalizes roughness of $g(t)$, so $\lambda$ has the same interpretation as before.

If the ordered distinct values of $t_1, \dots, t_n$ are $s_1 < \cdots < s_q$ and if $g(t)$ is a natural cubic spline with knots at the $s_i$, then the integral in (10.49) may be written $g^{\mathrm{T}}Kg$, where the $q \times 1$ vector $g$ has $i$th element $g_i = g(s_i)$. Given a value of $\lambda$, our aim