...ce, we often set $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}) = q(\boldsymbol{\theta}^*)$, i.e., we use a proposal density that ignores the current value of the variable. This algorithm is sometimes referred to as a *Hastings independence chain*, so named because the proposals (though not the final $\boldsymbol{\theta}^{(t)}$ values) form an independent sequence. Here, using equation (3.15) we obtain the acceptance ratios $r = w(\boldsymbol{\theta}^*)/w(\boldsymbol{\theta}^{(t-1)})$ where $w(\boldsymbol{\theta}) = h(\boldsymbol{\theta})/q(\boldsymbol{\theta})$, the usual weight function used in importance sampling. This suggests using the same guidelines for choosing q as for choosing a good importance sampling density (i.e., making q a good match for h), but perhaps with heavier tails to obtain an acceptance rate closer to 0.5, as encouraged by Gelman, Roberts, and Gilks (1996). While easy to implement, this algorithm can be difficult to tune since it will converge slowly unless the chosen q is rather close to the true posterior (which is of course unknown in advance).

Example 3.7 We illustrate the Metropolis-Hastings algorithm using the data in Table 3.3, which are taken from Bliss (1935). These data record the number of adult flour beetles killed after five hours of exposure to various levels of gaseous carbon disulphide (CS_2). Since the variability in these data cannot be explained by the standard logistic regression model, we attempt to fit the generalized logit model suggested by Prentice (1976),

$$P(\text{death}|w) \equiv g(w) = \{\exp(x)/(1+\exp(x))\}^{m_1} .$$

Here, w is the predictor variable (dose), and $x = (w - \mu)/\sigma$ where $\mu \in \Re$ and σ^2, $m_1 > 0$. Suppose there are y_i flour beetles dying out of n_i exposed at level w_i, $i = 1, \ldots, N$.

Dosage w_i	# Killed y_i	# Exposed n_i
1.6907	6	59
1.7242	13	60
1.7552	18	62
1.7842	28	56
1.8113	52	63
1.8369	53	59
1.8610	61	62
1.8839	60	60

Table 3.3 *Flour beetle mortality data (from Bliss, 1935).*

For our prior distributions, we assume that $m_1 \sim \text{Gamma}(a_0, b_0)$, $\mu \sim \text{Normal}(c_0, d_0^2)$, and $\sigma^2 \sim \text{Inverse Gamma}(e_0, f_0)$, where a_0, b_0, c_0, d_0, e_0, and f_0 are known, and m_1, μ, and σ^2 are independent. While these common families may appear to have been chosen to preserve some sort of conjugate structure, this is not the case: there is no closed form available for any of the three full conditional distributions needed to implement the Gibbs sampler. Thus we instead resort to the Metropolis algorithm. Our likelihood-prior specification implies the joint posterior distribution

$$\begin{aligned} p(\mu, \sigma^2, m_1|\mathbf{y}) &\propto f(\mathbf{y}|\mu, \sigma^2, m_1)\pi(\mu, \sigma^2, m_1) \\ &\propto \left\{\prod_{i=1}^{N}[g(w_i)]^{y_i}[1-g(w_i)]^{n_i-y_i}\right\} \frac{m_1^{a_0-1}}{\sigma^{2(e_0+1)}} \\ &\quad \times \exp\left[-\frac{1}{2}\left(\frac{\mu-c_0}{d_0}\right)^2 - \frac{m_1}{b_0} - \frac{1}{f_0\sigma^2}\right] . \end{aligned}$$

We begin by making a change of variables from (μ, σ^2, m_1) to $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\mu, \frac{1}{2}\log\sigma^2, \log m_1)$. This transforms the parameter space to $\Re^3$ (necessary if we wish to work with Gaussian proposal densities), and also helps to symmetrize the posterior distribution. Accounting for the Jacobian of this transformation, our target density is now

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) \propto h(\boldsymbol{\theta}) &= \left\{\prod_{i=1}^{N}[g(w_i)]^{y_i}[1-g(w_i)]^{n_i-y_i}\right\} \exp(a_0\theta_3 - 2e_0\theta_2) \\ &\quad \times \exp\left[-\frac{1}{2}\left(\frac{\theta_1-c_0}{d_0}\right)^2 - \frac{\exp(\theta_3)}{b_0} - \frac{\exp(-2\theta_2)}{f_0}\right] . \end{aligned}$$

As mentioned above, numerical stability is improved by working on the log scale, i.e., by computing the acceptance ratio as $r = \exp[\log h(\boldsymbol{\theta}^*) - \log h(\boldsymbol{\theta}^{(t-1)})]$.

We complete the prior specification by choosing the same hyperparameter values as in Carlin and Gelfand (1991b). That is, we take $a_0 = .25$ and $b_0 = 4$, so that m_1 has prior mean 1 (corresponding to the standard logit model) and prior standard deviation 2. We then specify rather vague priors for μ and σ^2 by setting $c_0 = 2$, $d_0 = 10$, $e_0 = 2.000004$, and $f_0 = 1000$; the latter two choices imply a prior mean of .001 and a prior standard deviation of .5 for σ^2. Figure 3.4 shows the output from three parallel Metropolis sampling chains, each run for 10,000 iterations using a $N_3(\boldsymbol{\theta}^{(t-1)}, \widetilde{\Sigma})$ proposal density where

$$\widetilde{\Sigma} = D = Diag(.00012, .033, .10) . \tag{3.16}$$

The histograms in the figure include all 24,000 samples obtained after a burn-in period of 2000 iterations (reasonable given the appearance of the monitoring plots). The chains mix very slowly, as can be seen from the extremely high lag 1 sample autocorrelations, estimated from the chain #2 output and printed above the monitoring plots. The reason for this slow convergence is the high correlations among the three parameters, estimated

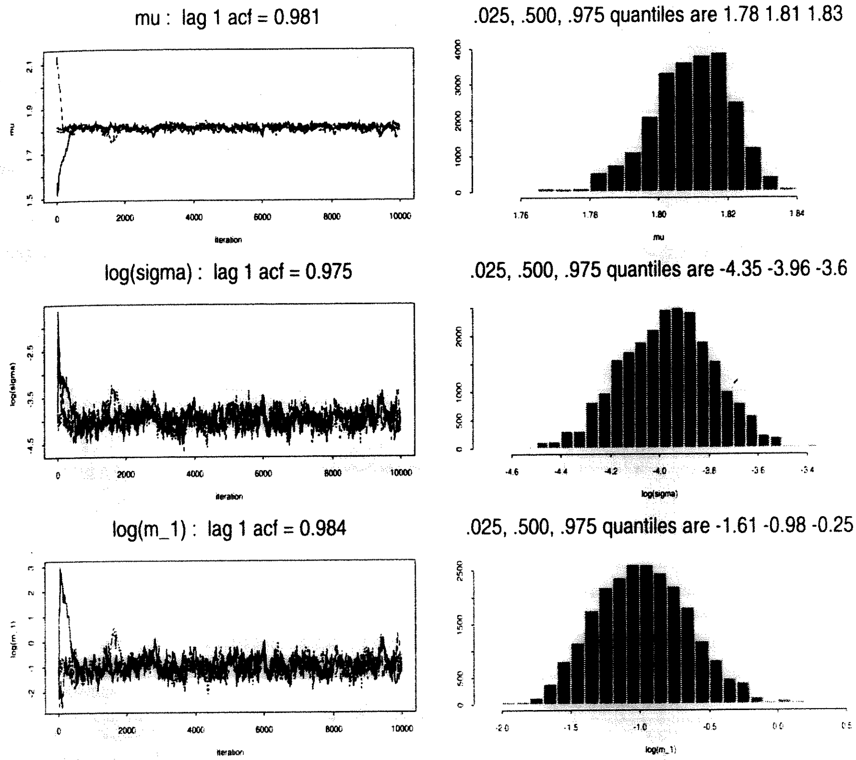

Figure 3.4 *Metropolis analysis of the flour beetle mortality data using a Gaussian proposal density with a diagonal $\widetilde{\Sigma}$ matrix. Monitoring plots use three parallel chains, and histograms use all samples following iteration 2000. Overall Metropolis acceptance rate: 13.5%.*

as $\widehat{Corr}(\theta_1, \theta_2) = -0.78$, $\widehat{Corr}(\theta_1, \theta_3) = -0.94$, and $\widehat{Corr}(\theta_2, \theta_3) = 0.89$ from the chain #2 output. As a result, the proposal acceptance rate is low (13.5%) and convergence is slow, despite the considerable experimentation in arriving at the three variances in our proposal density.

We can accelerate convergence by using a nondiagonal proposal covariance matrix designed to better mimic the posterior surface itself. From the output of the first algorithm, we can obtain an estimate of the posterior covariance matrix in the usual way as $\widehat{\Sigma} = \frac{1}{G}\sum_{g=1}^{G}(\boldsymbol{\theta}_g - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_g - \bar{\boldsymbol{\theta}})'$, where $g = 1, \ldots, G$ indexes the post-convergence Monte Carlo samples. We then reran the algorithm with proposal variance matrix

$$\widetilde{\Sigma} = 2\widehat{\Sigma} = \begin{pmatrix} 0.000292 & -0.003546 & -0.007856 \\ -0.003546 & 0.074733 & 0.117809 \\ -0.007856 & 0.117809 & 0.241551 \end{pmatrix} . \qquad (3.17)$$

The results, shown in Figure 3.5, indicate improved convergence, with lower

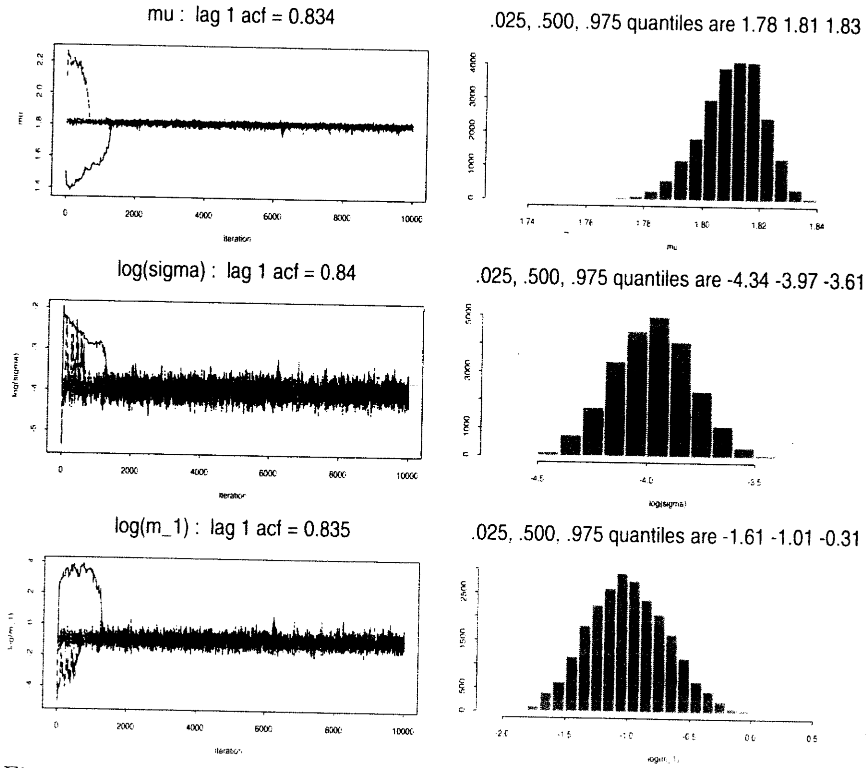

Figure 3.5 *Metropolis analysis of the flour beetle mortality data using a Gaussian proposal density with a nondiagonal $\widetilde{\Sigma}$ matrix. Monitoring plots use three parallel chains, and histograms use all samples following iteration 2000. Overall Metropolis acceptance rate: 27.3%.*

observed autocorrelations and a higher Metropolis acceptance rate (27.3%). This rate is close to the optimal rate derived by Gelman, Roberts, and Gilks (1996) of 31.6%, though this rate applies to target densities having three *independent* components. The marginal posterior results in both figures are consistent with the results obtained for this prior-likelihood combination by Carlin and Gelfand (1991b) and Müller (1991). ■

Langevin-Hastings algorithm

As mentioned above, the Metropolis algorithm is typically implemented with a multivariate normal proposal density given in (3.14), i.e., the candidate value $\boldsymbol{\theta}^*$ is drawn from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = N(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}, \widetilde{\Sigma})$. While these random walk proposals (centered at the current value of the chain, $\boldsymbol{\theta}^{(t-1)}$) can be very effective, in many situations we can do even better by introducing a systematic "drift" in the mean of this density, in order to nudge