Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
−0.86	5	0
−0.30	5	1
−0.05	5	3
0.73	5	5

Table 3.1: *Bioassay data from Racine et al. (1986).*

Scaled inverse-Wishart model

When modeling covariance matrices it can help to extend the inverse-Wishart model by multiplying by a set of scale parameters that can be modeled separately. This gives flexibility in modeling and allows one to set up a uniform or weak prior distribution on correlations without overly constraining the variance parameters. The *scaled inverse-Wishart* model for Σ has the form,

$$\Sigma = \text{Diag}(\xi)\Sigma_\eta \text{Diag}(\xi),$$

where Σ_η is given an inverse-Wishart prior distribution (one choice is Inv-Wishart$_{d+1}(I)$, so that the marginal distributions of the correlations are uniform) and then the scale parameters ξ can be given weakly informative priors themselves. We discuss further in Section 15.4 in the context of varying-intercept, varying-slope hierarchical regression models.

3.7 Example: analysis of a bioassay experiment

Beyond the normal distribution, few multiparameter sampling models allow simple explicit calculation of posterior distributions. Data analysis for such models is possible using the computational methods described in Part III of this book. Here we present an example of a nonconjugate model for a bioassay experiment, drawn from the literature on applied Bayesian statistics. The model is a two-parameter example from the broad class of generalized linear models to be considered more thoroughly in Chapter 16. We use a particularly simple simulation approach, approximating the posterior distribution by a discrete distribution supported on a two-dimensional grid of points, that provides sufficiently accurate inferences for this two-parameter example.

The scientific problem and the data

In the development of drugs and other chemical compounds, acute toxicity tests or bioassay experiments are commonly performed on animals. Such experiments proceed by administering various dose levels of the compound to batches of animals. The animals' responses are typically characterized by a dichotomous outcome: for example, alive or dead, tumor or no tumor. An experiment of this kind gives rise to data of the form

$$(x_i, n_i, y_i);\ i = 1, \ldots, k,$$

where x_i represents the ith of k dose levels (often measured on a logarithmic scale) given to n_i animals, of which y_i subsequently respond with positive outcome. An example of real data from such an experiment is shown in Table 3.1: twenty animals were tested, five at each of four dose levels.

Modeling the dose–response relation

Given what we have seen so far, we must model the outcomes of the five animals *within each group* i as exchangeable, and it seems reasonable to model them as independent with

equal probabilities, which implies that the data points y_i are binomially distributed:

$$y_i|\theta_i \sim \text{Bin}(n_i, \theta_i),$$

where θ_i is the probability of death for animals given dose x_i. (An example of a situation in which independence and the binomial model would *not* be appropriate is if the deaths were caused by a contagious disease.) For this experiment, it is also reasonable to treat the outcomes in the four groups as independent of each other, given the parameters $\theta_1, \ldots, \theta_4$.

The simplest analysis would treat the four parameters θ_i as exchangeable in their prior distribution, perhaps using a noninformative density such as $p(\theta_1, \ldots, \theta_4) \propto 1$, in which case the parameters θ_i would have independent beta posterior distributions. The exchangeable prior model for the θ_i parameters has a serious flaw, however; we know the dose level x_i for each group i, and one would expect the probability of death to vary systematically as a function of dose.

The simplest model of the *dose–response relation*—that is, the relation of θ_i to x_i—is linear: $\theta_i = \alpha + \beta x_i$. Unfortunately, this model has the flaw that at low or high doses, x_i approaches $\pm\infty$ (recall that the dose is measured on the log scale), whereas θ_i, being a probability, must be constrained to lie between 0 and 1. The standard solution is to use a transformation of the θ's, such as the logistic, in the dose–response relation:

$$\text{logit}(\theta_i) = \alpha + \beta x_i, \tag{3.15}$$

where $\text{logit}(\theta_i) = \log(\theta_i/(1-\theta_i))$ as defined in (1.10). This is called a *logistic regression* model.

The likelihood

Under the model (3.15), we can write the sampling distribution, or likelihood, for each group i in terms of the parameters α and β as

$$p(y_i|\alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i}[1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}.$$

The model is characterized by the parameters α and β, whose joint posterior distribution is

$$\begin{aligned} p(\alpha, \beta|y, n, x) &\propto p(\alpha, \beta|n, x)p(y|\alpha, \beta, n, x) \\ &\propto p(\alpha, \beta)\prod_{i=1}^{k} p(y_i|\alpha, \beta, n_i, x_i). \end{aligned} \tag{3.16}$$

We consider the sample sizes n_i and dose levels x_i as fixed for this analysis and suppress the conditioning on (n, x) in subsequent notation.

The prior distribution

We present an analysis based on a prior distribution for (α, β) that is independent and locally uniform in the two parameters; that is, $p(\alpha, \beta) \propto 1$. In practice, we might use a uniform prior distribution if we really have no prior knowledge about the parameters, or if we want to present a simple analysis of this experiment alone. If the analysis using the noninformative prior distribution is insufficiently precise, we may consider using other sources of substantive information (for example, from other bioassay experiments) to construct an informative prior distribution.

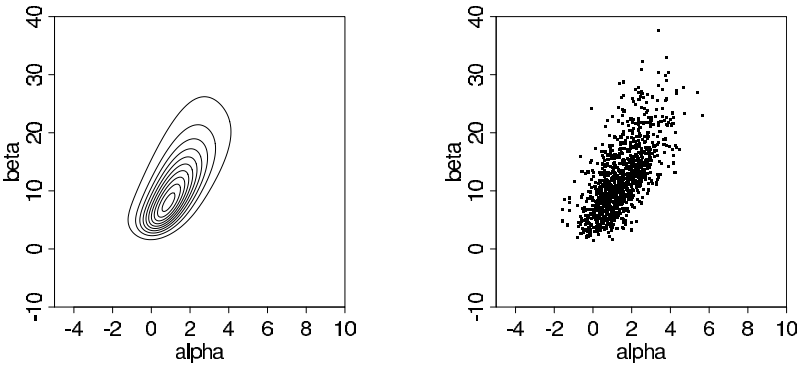

Figure 3.3 *(a) Contour plot for the posterior density of the parameters in the bioassay example. Contour lines are at* $0.05, 0.15, \ldots, 0.95$ *times the density at the mode. (b) Scatterplot of 1000 draws from the posterior distribution.*

A rough estimate of the parameters

We will compute the joint posterior distribution (3.16) at a grid of points (α, β), but before doing so, it is a good idea to get a rough estimate of (α, β) so we know where to look. To obtain the rough estimate, we use existing software to perform a logistic regression; that is, finding the maximum likelihood estimate of (α, β) in (3.16) for the four data points in Table 3.1. The estimate is $(\hat{\alpha}, \hat{\beta}) = (0.8, 7.7)$, with standard errors of 1.0 and 4.9 for α and β, respectively.

Obtaining a contour plot of the joint posterior density

We are now ready to compute the posterior density at a grid of points (α, β). After some experimentation, we use the range $(\alpha, \beta) \in [-5, 10] \times [-10, 40]$, which captures almost all the mass of the posterior distribution. The resulting contour plot appears in Figure 3.3a; a general justification for setting the lowest contour level at 0.05 for two-dimensional plots appears on page 85 in Section 4.1.

Sampling from the joint posterior distribution

Having computed the unnormalized posterior density at a grid of values that cover the effective range of (α, β), we can normalize by approximating the distribution as a step function over the grid and setting the total probability in the grid to 1. We sample 1000 random draws (α^s, β^s) from the posterior distribution using the following procedure.

1. Compute the marginal posterior distribution of α by numerically summing over β in the discrete distribution computed on the grid of Figure 3.3a.
2. For $s = 1, \ldots, 1000$:
   (a) Draw α^s from the discretely computed $p(\alpha|y)$; this can be viewed as a discrete version of the inverse cdf method described in Section 1.9.
   (b) Draw β^s from the discrete conditional distribution, $p(\beta|\alpha, y)$, given the just-sampled value of α.
   (c) For each of the sampled α and β, add a uniform random jitter centered at zero with a width equal to the spacing of the sampling grid. This gives the simulation draws a continuous distribution.

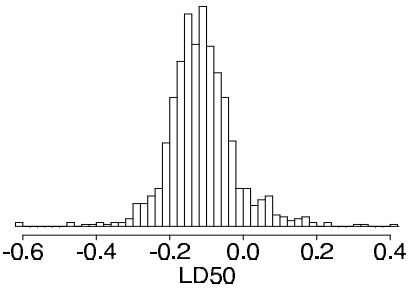

Figure 3.4 *Histogram of the draws from the posterior distribution of the LD50 (on the scale of log dose in g/ml) in the bioassay example, conditional on the parameter β being positive.*

The 1000 draws (α^s, β^s) are displayed on a scatterplot in Figure 3.3b. The scale of the plot, which is the same as the scale of Figure 3.3a, has been set large enough that all the 1000 draws would fit on the graph.

There are a number of practical considerations when applying this two-dimensional grid approximation. There can be difficulty finding the correct location and scale for the grid points. A grid that is defined on too small an area may miss important features of the posterior distribution that fall outside the grid. A grid defined on a large area with wide intervals between points can miss important features that fall between the grid points. It is also important to avoid overflow and underflow operations when computing the posterior distribution. It is usually a good idea to compute the logarithm of the unnormalized posterior distribution and subtract off the maximum value before exponentiating. This creates an unnormalized discrete approximation with maximum value 1, which can then be normalized (by setting the total probability in the grid to 1).

The posterior distribution of the LD50

A parameter of common interest in bioassay studies is the LD50—the dose level at which the probability of death is 50%. In our logistic model, a 50% survival rate means

$$\text{LD50:} \quad \text{E}\left(\frac{y_i}{n_i}\right) = \text{logit}^{-1}(\alpha + \beta x_i) = 0.5;$$

thus, $\alpha + \beta x_i = \text{logit}(0.5) = 0$, and the LD50 is $x_i = -\alpha/\beta$. Computing the posterior distribution of any summaries in the Bayesian approach is straightforward, as discussed at the end of Section 1.9. Given what we have done so far, simulating the posterior distribution of the LD50 is trivial: we just compute $-\alpha/\beta$ for the 1000 draws of (α, β) pictured in Figure 3.3b.

Difficulties with the LD50 parameterization if the drug is beneficial. In the context of this example, LD50 is a meaningless concept if $\beta \leq 0$, in which case increasing the dose does not cause the probability of death to increase. If we were certain that the drug could *not* cause the tumor rate to decrease, we should constrain the parameter space to exclude values of β less than 0. However, it seems more reasonable here to allow the possibility of $\beta \leq 0$ and just note that LD50 is hard to interpret in this case.

We summarize the inference on the LD50 scale by reporting two results: (1) the posterior probability that $\beta > 0$—that is, that the drug is harmful—and (2) the posterior distribution for the LD50 conditional on $\beta > 0$. All of the 1000 simulation draws had positive values of β, so the posterior probability that $\beta > 0$ is roughly estimated to exceed 0.999. We compute the LD50 for the simulation draws with positive values of β (which happen to be all 1000

draws for this example); a histogram is displayed in Figure 3.4. This example illustrates that the marginal posterior mean is not always a good summary of inference about a parameter. We are *not*, in general, interested in the posterior mean of the LD50, because the posterior mean includes the cases in which the dose–response relation is negative.

3.8 Summary of elementary modeling and computation

The lack of multiparameter models permitting easy calculation of posterior distributions is not a major practical handicap for three main reasons. First, when there are few parameters, posterior inference in nonconjugate multiparameter models can be obtained by simple simulation methods, as we have seen in the bioassay example. Second, sophisticated models can often be represented in a hierarchical or conditional manner, as we shall see in Chapter 5, for which effective computational strategies are available (as we discuss in general in Part III). Finally, as we discuss in Chapter 4, we can often apply a normal approximation to the posterior distribution, and therefore the conjugate structure of the normal model can play an important role in practice, well beyond its application to explicitly normal sampling models.

Our successful analysis of the bioassay example suggests the following strategy for computation of simple Bayesian posterior distributions. What follows is not truly a general approach, but it summarizes what we have done so far and foreshadows the general methods—based on successive approximations—presented in Part III.

1. Write the likelihood part of the model, $p(y|\theta)$, ignoring any factors that are free of θ.
2. Write the posterior density, $p(\theta|y) \propto p(\theta)p(y|\theta)$. If prior information is well-formulated, include it in $p(\theta)$. Otherwise use a weakly informative prior distribution or temporarily set $p(\theta) \propto$ constant, with the understanding that the prior density can be altered later to include additional information or structure.
3. Create a crude estimate of the parameters, θ, for use as a starting point and a comparison to the computation in the next step.
4. Draw simulations $\theta^1, \ldots, \theta^S$, from the posterior distribution. Use the sample draws to compute the posterior density of any functions of θ that may be of interest.
5. If any predictive quantities, $\tilde{y}$, are of interest, simulate $\tilde{y}^1, \ldots, \tilde{y}^S$ by drawing each $\tilde{y}^s$ from the sampling distribution conditional on the drawn value θ^s, $p(\tilde{y}|\theta^s)$. In Chapter 6, we discuss how to use posterior simulations of θ and $\tilde{y}$ to check the fit of the model to data and substantive knowledge.

For nonconjugate models, step 4 above can be difficult. Various methods have been developed to draw posterior simulations in complicated models, as we discuss in Part III. Occasionally, high-dimensional problems can be solved by combining analytical and numerical simulation methods. If θ has only one or two components, it is possible to draw simulations by computing on a grid, as we illustrated in the previous section for the bioassay example.

3.9 Bibliographic note

Chapter 2 of Box and Tiao (1973) thoroughly treats the univariate and multivariate normal distribution problems and also some related problems such as estimating the difference between two means and the ratio between two variances. At the time that book was written, computer simulation methods were much less convenient than they are now, and so Box and Tiao, and other Bayesian authors of the period, restricted their attention to conjugate families and devoted much effort to deriving analytic forms of marginal posterior densities.