Multiparameter Models

Normal Data with Both Parameters Unknown

A standard inference problem is to learn about a normal population where both the mean and variance are unknown. To illustrate Bayesian computation for this problem, suppose we are interested in learning about the distribution of completion times for men between ages 20 and 29 who are running the New York Marathon. We observe the times $y_1, \ldots, y_{20}$ in minutes for 20 runners , and we assume they represent a random sample from an $N(\mu, \sigma)$ distribution. If we assume the standard noninformative prior $g(\mu, \sigma^2) \propto 1/\sigma^2$, then the posterior density of the mean and variance is given by

$$g(\mu, \sigma^2 | y) \propto \frac{1}{(\sigma^2)^{n/2+1}} \exp\left(-\frac{1}{2\sigma^2}(S + n(\mu - \bar{y})^2)\right),$$

where n is the sample size, $\bar{y}$ is the sample mean, and $S = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

Normal Data with Both Parameters Unknown

This joint posterior has the familiar normal/inverse chi-square form where

- the posterior of μ conditional on σ^2 is distributed as $N(\bar{y}, \sigma/\sqrt{n})$
- the marginal posterior of σ^2 is distributed as $S\chi^{-2}_{n-1}$, where χ^{-2}_{ν} denotes an inverse chi-square distribution with ν degrees of freedom

Normal Data with Both Parameters Unknown

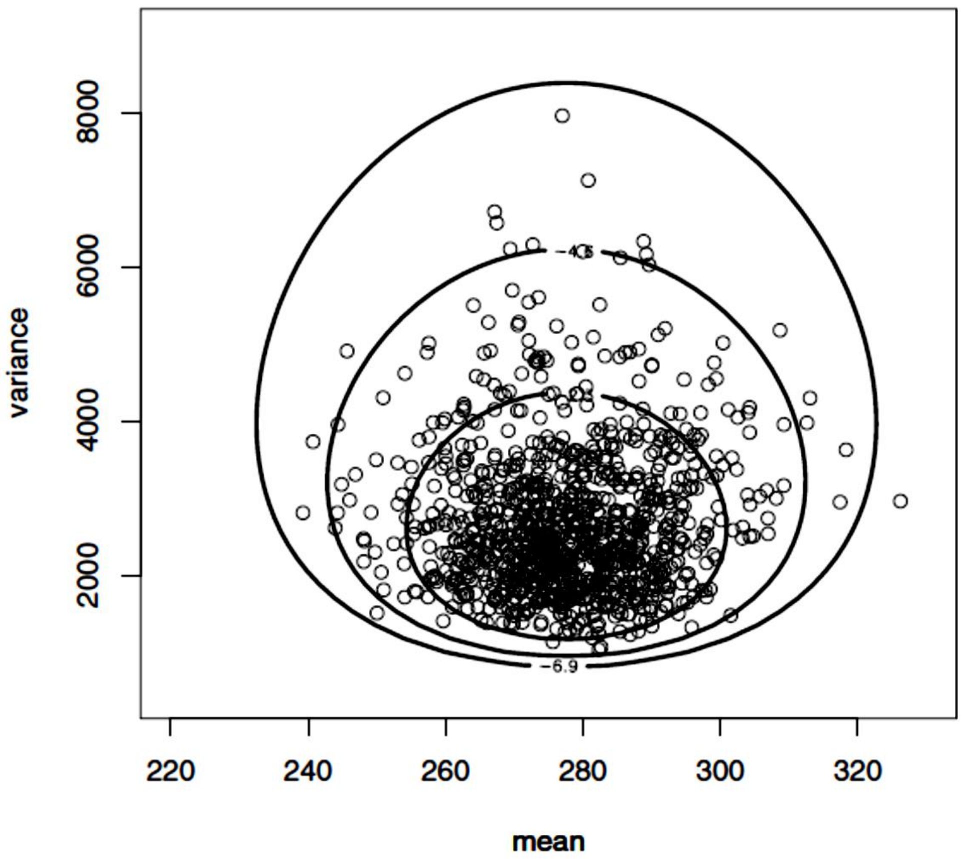

Fig. 4.1. Contour plot of the joint posterior distribution of (μ, σ^2) for a normal sampling model. The points represent a simulated random sample from this distribution.

Multinomial Model

Gelman et al. (2003) describe a sample survey conducted by CBS News before the 1988 presidential election. A total of 1447 adults were polled to indicate their preference; $y_1 = 727$ supported George Bush, $y_2 = 583$ supported Michael Dukakis, and $y_3 = 137$ supported other candidates or expressed no opinion. The counts y_1, y_2, and y_3 are assumed to have a multinomial distribution with sample size n and respective probabilities θ_1, θ_2, and θ_3. If a uniform prior distribution is assigned to the multinomial vector $\theta = (\theta_1, \theta_2, \theta_3)$, then the posterior distribution of θ is proportional to

$$g(\theta) = \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3},$$

which is recognized as a Dirichlet distribution with parameters $(y_1 + 1, y_2 + 1, y_3 + 1)$. The focus is to compare the proportions of voters for Bush and Dukakis by considering the difference $\theta_1 - \theta_2$.

Multinomial Model

The summarization of the Dirichlet posterior distribution is again conveniently done by simulation. Although the base R package does not have a function to simulate Dirichlet variates, it is easy to write a function to simulate this distribution based on the fact that if W_1, W_2, W_3 are independentlt distributed from gamma$(\alpha_1, 1)$, gamma$(\alpha_2, 1)$, gamma$(\alpha_3, 1)$ distributions and $T = W_1 + W_2 + W_3$, then the distribution of the proportions $(W_1/T, W_2/T, W_3/T)$ has a Dirichlet$(\alpha_1, \alpha_2, \alpha_3)$ distribution. The R function **`rdirichlet`** in the package **`LearnBayes`** uses this transformation of random variates to simulate draws of a Dirichlet distribution. One thousand vectors θ are simulated and stored in the matrix **`theta`**.

Multinomial Model

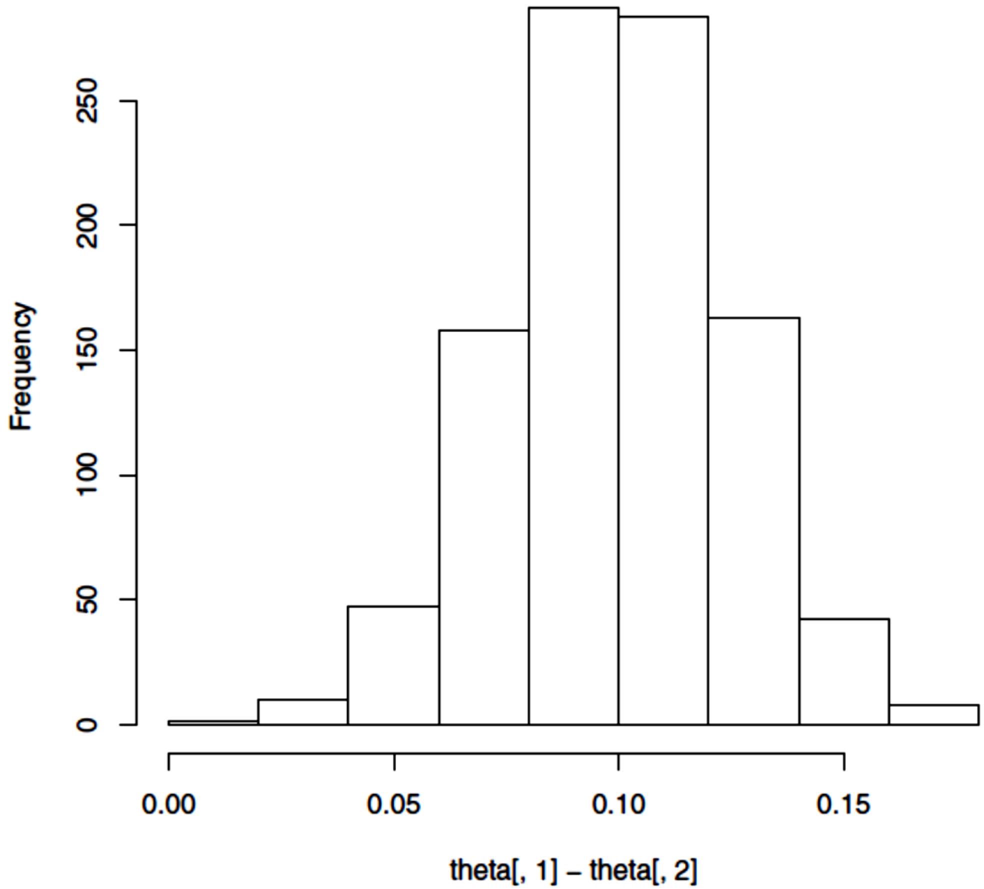

Fig. 4.2. Histogram of simulated sample of the marginal posterior distribution of $\theta_1 - \theta_2$ for the multinomial sampling example.

Multinomial Model

In the United States presidential election, there are 50 states plus the District of Columbia, and each has an assigned number of electoral votes. The candidate receiving the largest number of votes in a particular state receives the corresponding number of electoral votes, and for a candidate to be elected, he or she must receive a majority of the total number (538) of electoral votes. In the 2008 election between Barack Obama and John McCain, suppose we wish to predict the total number of electoral votes EV_O obtained by Obama. Let θ_{Oj} and θ_{Mj} denote the proportion of voters respectively for Obama and McCain in the jth state. One can express the number of electoral votes for Obama as

$$EV_O = \sum_{j=1}^{51} EV_j I(\theta_{Oj} > \theta_{Mj}),$$

where EV_j is the number of electoral votes in the jth state and $I()$ is the indicator function, which is equal to 1 if the argument is true and 0 otherwise.

Multinomial Model

On the Sunday before Election Day, the website www.cnn.com gives the results of the most recent poll in each state. Let q_{Oj} and q_{Mj} denote the sample proportions of voters for Obama and McCain in the ith state. We make the conservative assumption that each poll is based on a sample of 500 voters. Assuming a uniform prior on the vector of proportions, the vectors $(\theta_{O1}, \theta_{M1}), ..., (\theta_{O51}, \theta_{M51})$ have independent posterior distributions, where the proportions favoring the candidates in the ith state, $(\theta_{Oi}, \theta_{Mi}, 1 - \theta_{Oi}, \theta_{Mi})$, have a Dirichlet distribution with parameters $(500q_{Oj} + 1, 500q_{Mj} + 1, 500(1 - q_{Oj} - q_{Mj}) + 1)$.

Multinomial Model

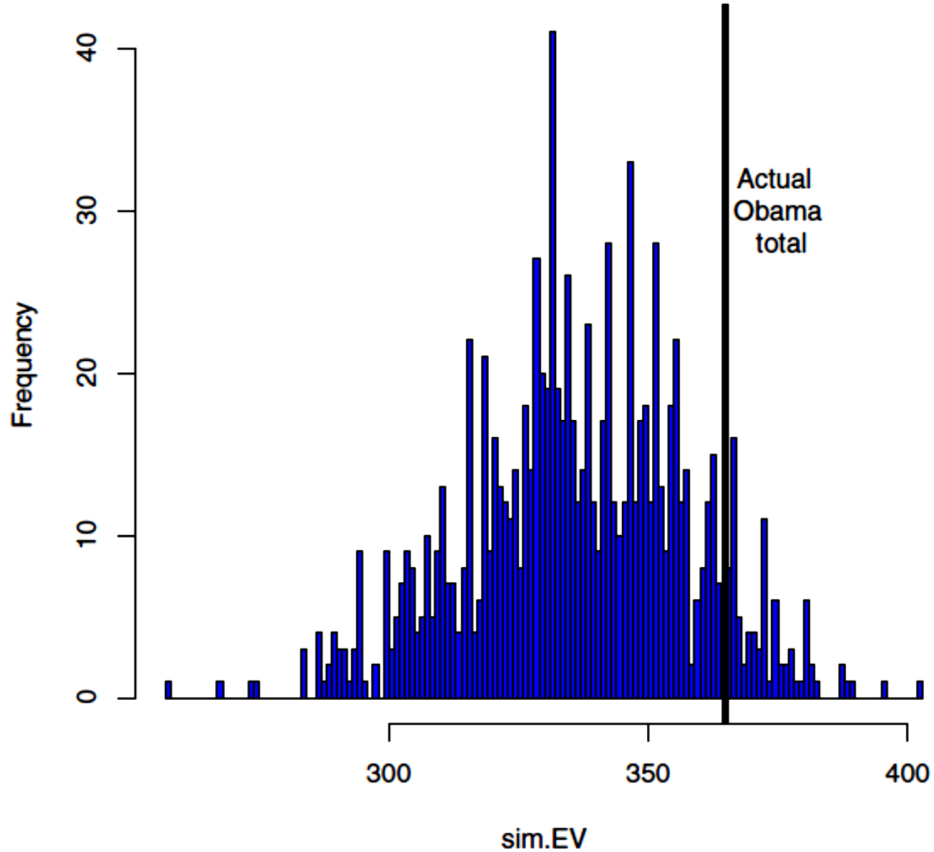

Fig. 4.3. Histogram of 1000 simulated draws of the total electoral vote for Barack Obama in the 2008 U.S. presidential election. The actual electoral vote of 365 is indicated by a vertical line.

A Bioassay Experiment

In the development of drugs, bioassay experiments are often performed on animals. In a typical experiment, various dose levels of a compound are administered to batches of animals and a binary outcome (positive or negative) is recorded for each animal. We consider data from Gelman et al. (2003), where one observes a dose level (in log g/ml), the number of animals, and the number of deaths for each of four groups. The data are displayed in Table 4.1.

Table 4.1. Data from the bioassay experiment.

Dose	Deaths	Sample Size
–0.86	0	5
–0.30	1	5
–0.05	3	5
0.73	5	5

A Bioassay Experiment

Let y_i denote the number of deaths observed out of n_i with dose level x_i. We assume y_i is binomial(n_i, p_i), where the probability p_i follows the logistic model

$$\log(p_i/(1-p_i)) = \beta_0 + \beta_1 x_i.$$

The likelihood function of the unknown regression parameters β_0 and β_1 is given by

$$L(\beta_0, \beta_1) \propto \prod_{i=1}^{4} p_i^{y_i}(1-p_i)^{n_i - y_i},$$

where $p_i = \exp(\beta_0 + \beta_1 x_i)/(1 + \exp(\beta_0 + \beta_1 x_i))$.

A Bioassay Experiment

Suppose that the beliefs about the probability p_L are independent of the beliefs about p_H. Then the joint prior of (p_L, p_H) is given by

$$g(p_L, p_H) \propto p_L^{1.12-1}(1-p_L)^{3.56-1} p_H^{2.10-1}(1-p_H)^{0.74-1}.$$

Figure 4.4 displays the conditional means prior by using error bars placed on the probability of death for two dose levels. As will be explained shortly, the smooth curve is the fitted probability curve using this prior information.

A Bioassay Experiment

Table 4.2. Prior information in the bioassay experiment.

Dose	Deaths	Sample Size
−0.7	1.12	4.68
0.6	2.10	2.84

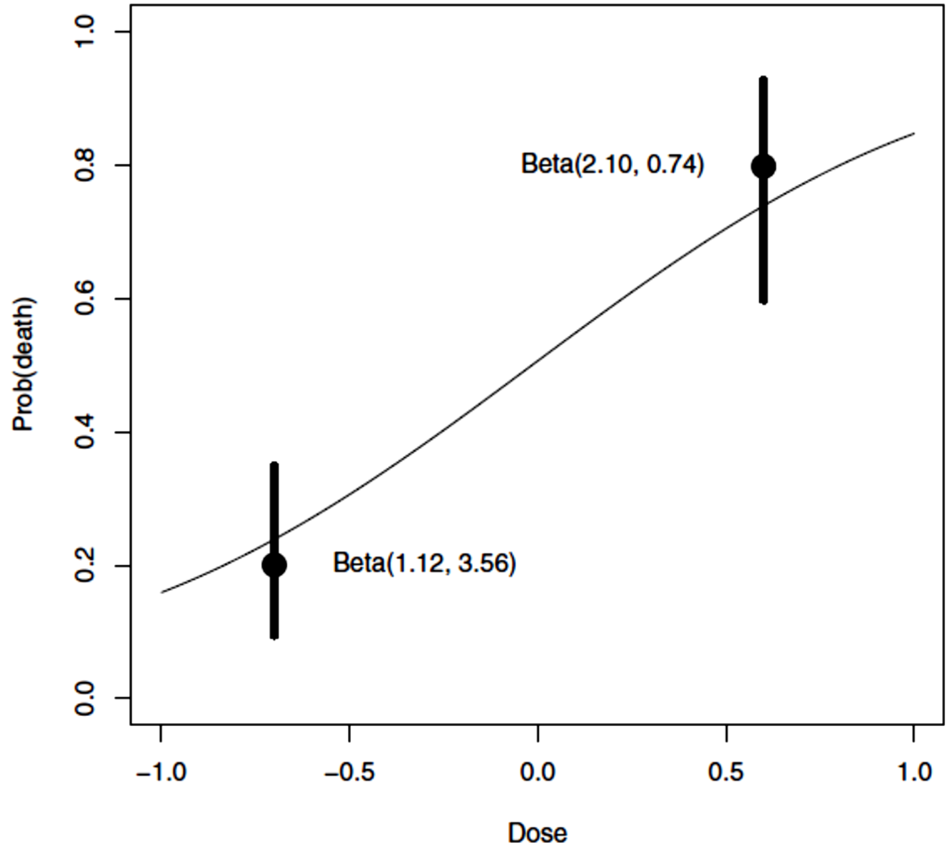

Fig. 4.4. Illustration of conditional means prior for the bioassay example. In each bar, the point corresponds to the median and the endpoints correspond to the quartiles of the prior distribution for each beta distribution.

A Bioassay Experiment

If this prior on (p_L, p_H) is transformed to the regression vector (β_0, β_1) through the transformation

$$p_L = \frac{\exp(\beta_0 + \beta_1 x_L)}{1 + \exp(\beta_0 + \beta_1 x_L)},\ p_H = \frac{\exp(\beta_0 + \beta_1 x_H)}{1 + \exp(\beta_0 + \beta_1 x_H)},$$

one can show that the induced prior is

$$g(\beta_0, \beta_1) \propto p_L^{1.12}(1 - p_L)^{3.56} p_H^{2.10}(1 - p_H)^{0.74}.$$

A Bioassay Experiment

Note that this prior has the same functional form as the likelihood, where the beta parameters can be viewed as the numbers of deaths and survivals in a prior experiment performed at two dose levels (see Table 4.2). If we combine these "prior data" with the observed data, we see that the posterior density is given by

$$g(\beta_0, \beta_1 | y) \propto \prod_{i=1}^{6} p_i^{y_i} (1 - p_i)^{n_i - y_i},$$

where $(x_j, n_j, y_j), j = 5, 6$, represent the dose, number of deaths, and sample size in the prior experiment.

A Bioassay Experiment

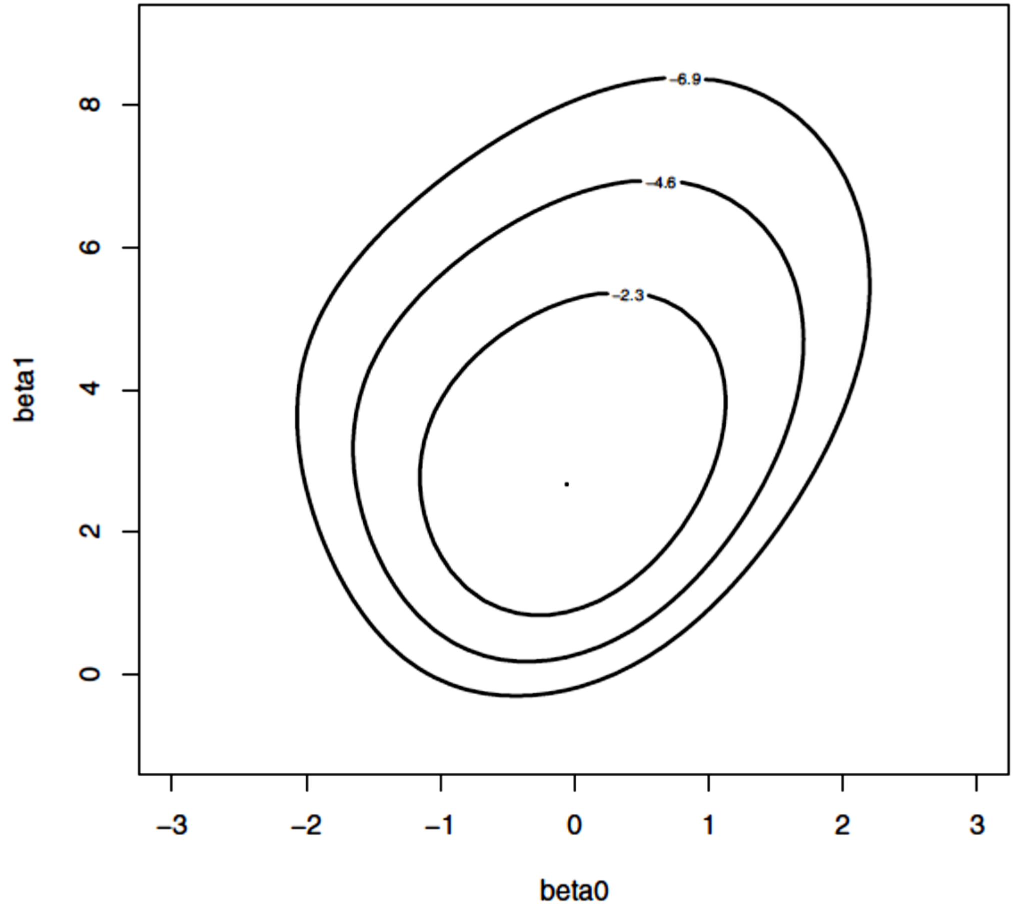

Fig. 4.5. Contour plot of the posterior distribution of (β_0, β_1) for the bioassay example. The contour lines are drawn at 10%, 1%, and .1% of the model height.

A Bioassay Experiment

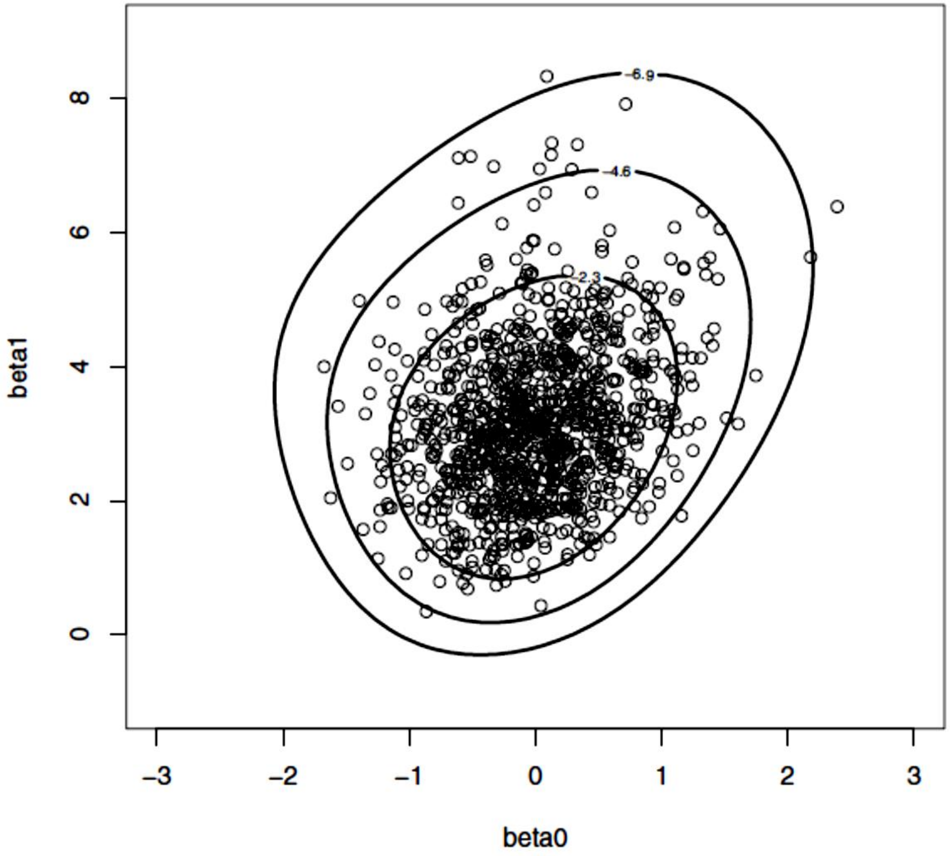

Fig. 4.6. Contour plot of the posterior distribution of (β_0, β_1) for the bioassay example. A simulated random sample from this distribution is shown on top of the contour plot.

A Bioassay Experiment

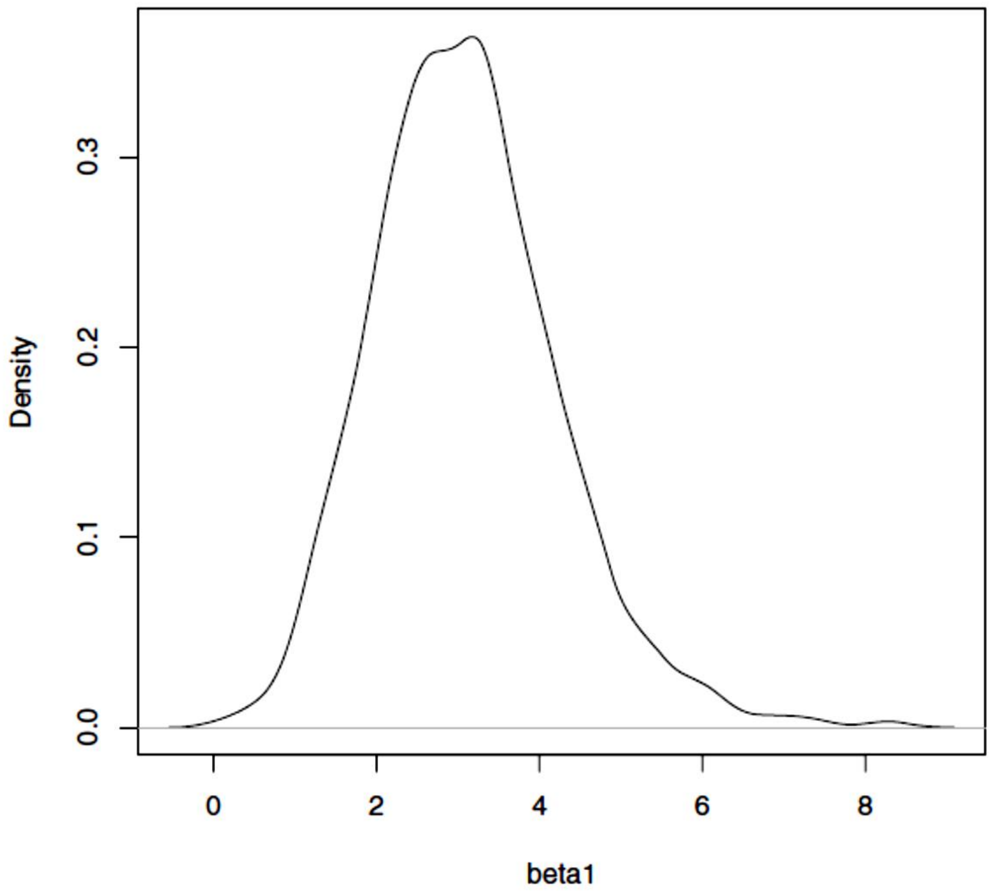

Fig. 4.7. Density of simulated values from the posterior of the slope parameter β_1 in the bioassay example.

A Bioassay Experiment

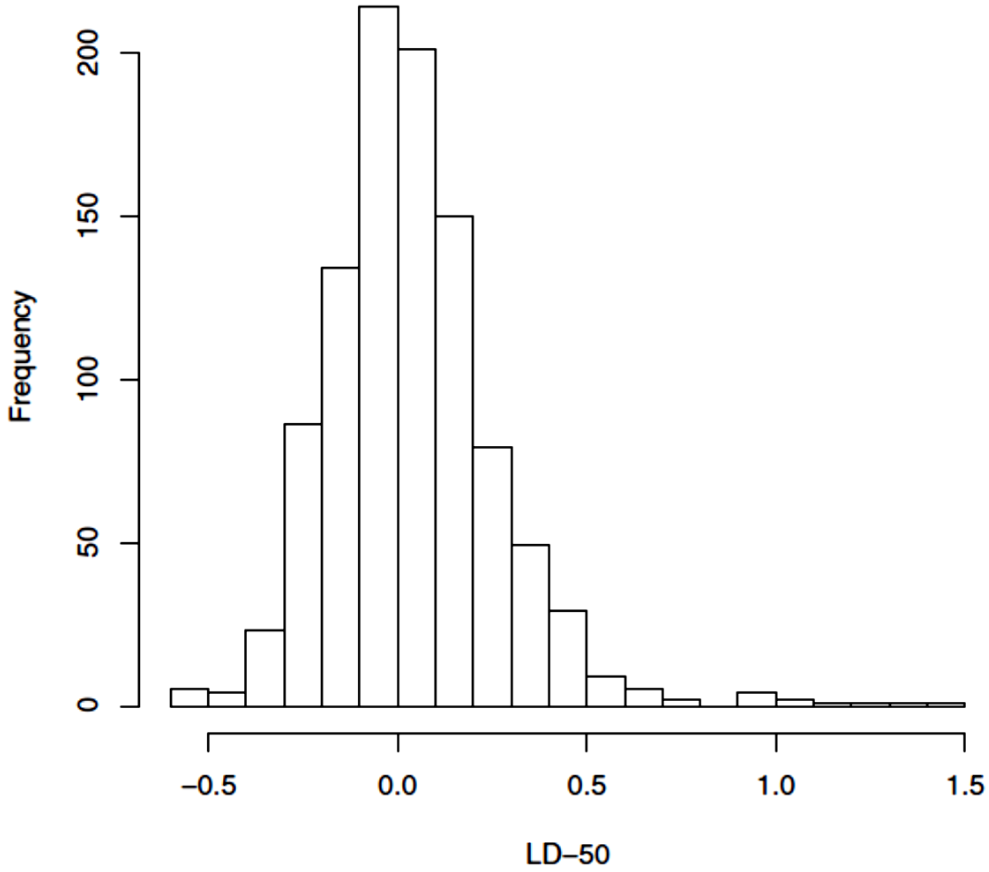

Fig. 4.8. Histogram of simulated values of the LD-50 parameter $-\beta_0/\beta_1$ in the bioassay example.

Comparing Two Proportions

Howard's special form of dependent prior is expressed as follows. First the proportions are transformed into the real-valued logit parameters

$$\theta_1 = \log \frac{p_1}{1-p_1}, \theta_2 = \log \frac{p_2}{1-p_2}.$$

Then suppose that given a value of θ_1, the logit θ_2 is assumed to be normally distributed with mean θ_1 and standard deviation σ. By generalizing this idea, Howard proposes the dependent prior of the general form

$$g(p_1, p_2) \propto e^{-(1/2)u^2} p_1^{\alpha-1}(1-p_1)^{\beta-1} p_2^{\gamma-1}(1-p_2)^{\delta-1}, 0 < p_1, p_2 < 1,$$

where

$$u = \frac{1}{\sigma}(\theta_1 - \theta_2).$$

This class of dependent priors is indexed by the parameters $(\alpha, \beta, \gamma, \delta, \sigma)$. The first four parameters reflect one's beliefs about the locations of p_1 and p_2, and the parameter σ indicates one's prior belief in the dependence between the two proportions.

Comparing Two Proportions

Suppose we observe counts y_1, y_2 from the two binomial samples. The likelihood function is given by

$$L(p_1, p_2) \propto p_1^{y_1}(1-p_1)^{n_1-y_1} p_2^{y_2}(1-p_2)^{n_2-y_2}, 0 < p_1, p_2 < 1.$$

Combining the likelihood with the prior, one sees that the posterior density has the same functional "dependent" form with updated parameters

$$(\alpha + y_1, \beta + n_1 - y_1, \gamma + y_2, \delta + n_2 - y_2, \sigma).$$

We illustrate testing the hypotheses using a dataset discussed by Pearson (1947), shown in Table 4.3.

Table 4.3. Pearson's example.

	Successes	Failures	Total
Sample 1	3	15	18
Sample 2	7	5	12
Totals	10	20	30

Comparing Two Proportions

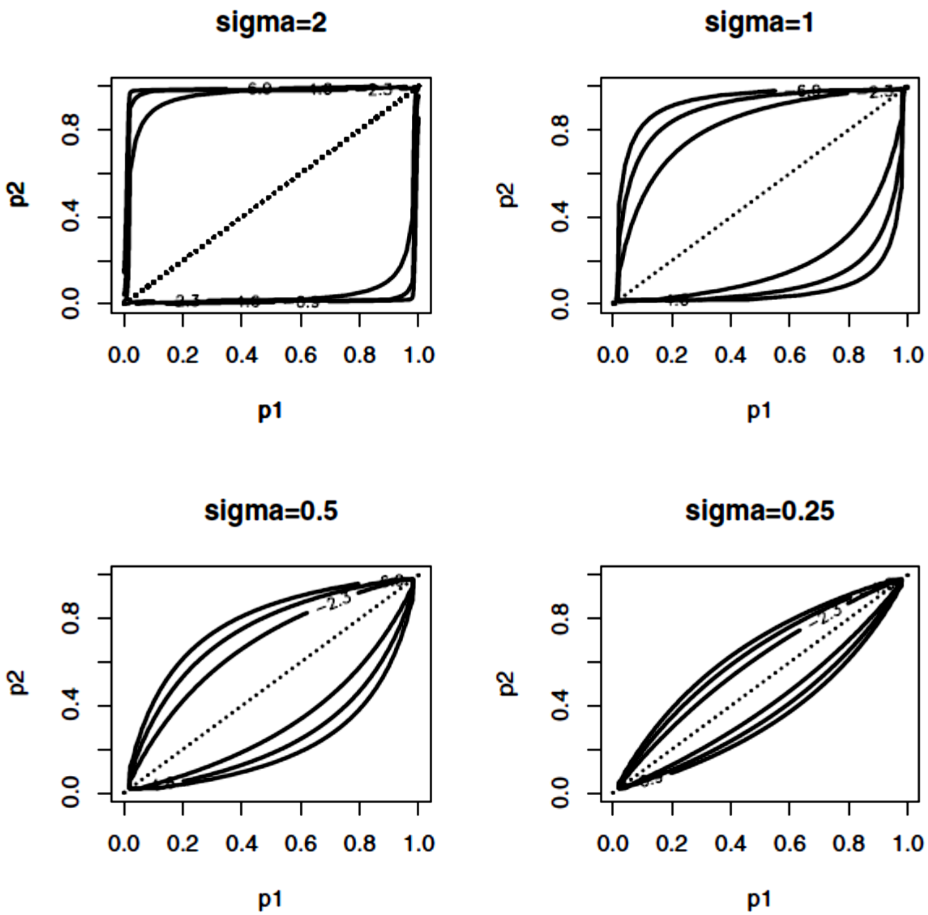

Fig. 4.9. Contour graphs of Howard's dependent prior for values of the association parameter $\sigma = 2, 1, .5$, and $.25$.

Comparing Two Proportions

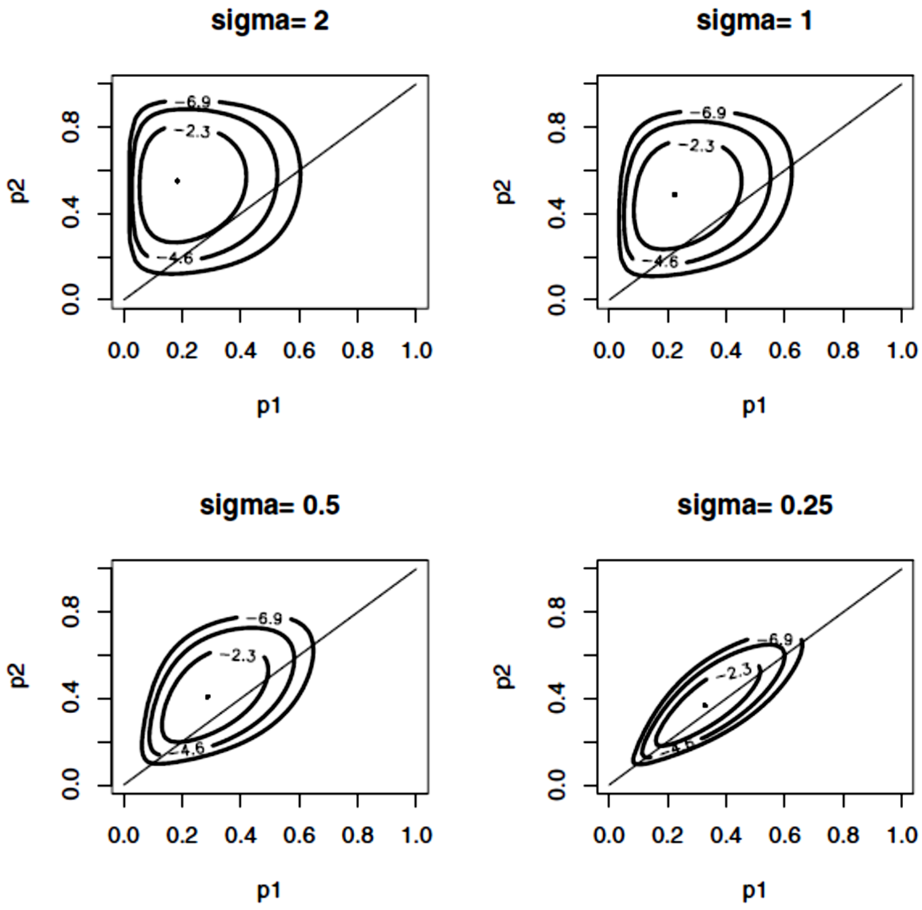

Fig. 4.10. Contour graphs of the posterior for Howard's dependent prior for values of the association parameter $\sigma = 2, 1, .5$, and $.25$.

Comparing Two Proportions

Table 4.4 displays the posterior probability that p_1 exceeds p_2 for four choices of the dependent prior parameter σ. Note that this posterior probability is sensitive to the prior belief about the dependence between the two proportions.

Table 4.4. Posterior probabilities of the hypothesis.

Dependent Parameter σ	$P(p_1 > p_2)$
2	0.012
1	0.035
.5	0.102
.25	0.201