

---

## Multiparameter Models

### 4.1 Introduction

In this chapter, we describe the use of R to summarize Bayesian models with several unknown parameters. In learning about parameters of a normal population or multinomial parameters, posterior inference is accomplished by simulating from distributions of standard forms. Once a simulated sample is obtained from the joint posterior, it is straightforward to perform transformations on these simulated draws to learn about any function of the parameters. We next consider estimating the parameters of a simple logistic regression model. Although the posterior distribution does not have a simple functional form, it can be summarized by computing the density on a fine grid of points. A common inference problem is to compare two proportions in a  $2 \times 2$  contingency table. We illustrate the computation of the posterior probability that one proportion exceeds the second proportion in the situation in which one believes a priori that the proportions are dependent.

### 4.2 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  $y_1, \dots, y_{20}$  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  $S = \sum_{i=1}^n (y_i - \bar{y})^2$ .

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

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

We first use R to construct a contour plot of the joint posterior density for this example. We read in the data `marathontimes`; when we `attach` this dataset, we can use the variable `time` that contains the vector of running times. The R function `normchi2post` in the `LearnBayes` package computes the logarithm of the joint posterior density of  $(\mu, \sigma^2)$ . We also use a function `mycontour` in the `LearnBayes` package that facilitates the use of the R `contour` command. There are four inputs to `mycontour`: the name of the function that defines the log density, a vector with the values (xlo, xhi, ylo, and yhi) that define the rectangle where the density is to be graphed, the data used in the function for the log density, and any optional parameters used with `contour`. The function produces a contour graph, shown in Figure 4.1, where the contour lines are drawn at 10%, 1%, and .1% of the maximum value of the posterior density over the grid.

```
> data(marathontimes)
> attach(marathontimes)
> d = mycontour(normchi2post, c(220, 330, 500, 9000), time,
+   xlab="mean",ylab="variance")
```

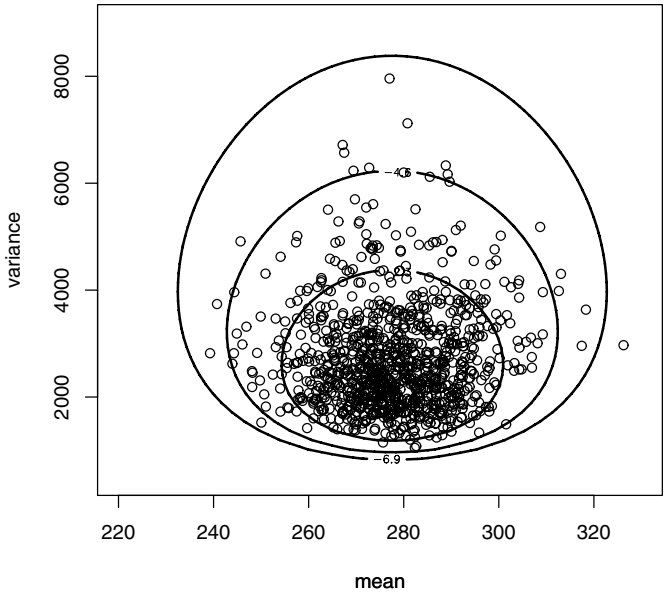
It is convenient to summarize this posterior distribution by simulation. One can simulate a value of  $(\mu, \sigma^2)$  from the joint posterior by first simulating  $\sigma^2$  from an  $S\chi_{n-1}^{-2}$  distribution and then simulating  $\mu$  from the  $N(\bar{y}, \sigma/\sqrt{n})$  distribution. In the following R output, we first simulate a sample of size 1000 from the chi-square distribution using the function `rchisq`. Then simulated draws of the “scale times inverse chi-square” distribution of the variance  $\sigma^2$  are obtained by transforming the chi-square draws. Finally, simulated draws of the mean  $\mu$  are obtained using the function `rnorm`.

```
> S = sum((time - mean(time))^2)
> n = length(time)
> sigma2 = S/rchisq(1000, n - 1)
> mu = rnorm(1000, mean = mean(time), sd = sqrt(sigma2)/sqrt(n))
```

The function `normpostsim` in the `LearnBayes` package implements this simulation algorithm. We display the simulated sampled values of  $(\mu, \sigma^2)$  on top of the contour plot of the distribution in Figure 4.1.

```
> points(mu, sigma2)
```

Inferences about the parameters or functions of the parameters are available from the simulated sample. To construct a 95% interval estimate for the mean  $\mu$ , we use the R `quantile` function to find percentiles of the simulated sample of  $\mu$ .



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

```
> quantile(mu, c(0.025, 0.975))
      2.5%    97.5%
254.0937 301.7137
```

A 95% credible interval for the mean completion time is (254.1, 301.7) minutes.

Suppose we are interested in learning about the standard deviation  $\sigma$  that describes the spread of the population of marathon running times. To obtain a sample of the posterior of  $\sigma$ , we take square roots of the simulated draws of  $\sigma^2$ . From the output, we see that an approximate 95% probability interval for  $\sigma$  is (37.5, 70.9) minutes.

```
> quantile(sqrt(sigma2), c(0.025, 0.975))
      2.5%    97.5%
37.48217 70.89521
```

### 4.3 A 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  $\theta_1, \theta_2$ , and  $\theta_3$ . If a uniform prior distribution is assigned to the multinomial vector  $\theta = (\theta_1, \theta_2, \theta_3)$ , then the posterior distribution of  $\theta$  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  $\theta_1 - \theta_2$ .

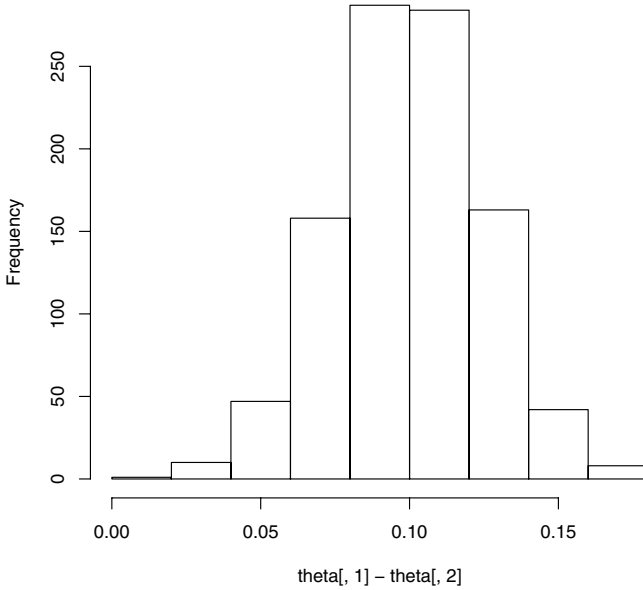
The summarization of the Dirichlet posterior distribution is again conveniently done by simulation. Although the base R package does not have a function to simulate Dirichlet variates, it is easy to write a function to simulate this distribution based on the fact that if  $W_1, W_2, W_3$  are independentlt distributed from  $\text{gamma}(\alpha_1, 1)$ ,  $\text{gamma}(\alpha_2, 1)$ ,  $\text{gamma}(\alpha_3, 1)$  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  $\theta$  are simulated and stored in the matrix `theta`.

```
> alpha = c(728, 584, 138)
> theta = rdirichlet(1000, alpha)
```

Since we are interested in comparing the proportions for Bush and Dukakis, we focus on the difference  $\theta_1 - \theta_2$ . A histogram of the simulated draws of this difference is displayed in Figure 4.2. Note that all of the mass of this distribution is on positive values, indicating that there is strong evidence that the proportion of voters for Bush exceeds the proportion for Dukakis.

```
> hist(theta[, 1] - theta[, 2], main="")
```

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  $\theta_{Oj}$  and  $\theta_{Mj}$  denote the proportion of voters respectively for Obama and



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

McCain in the  $j$ th 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  $j$ th state and  $I(\cdot)$  is the indicator function, which is equal to 1 if the argument is true and 0 otherwise.

On the Sunday before Election Day, the website [www.cnn.com](http://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  $i$ th 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  $(\theta_{O1}, \theta_{M1}), \dots, (\theta_{O51}, \theta_{M51})$  have independent posterior distributions, where the proportions favoring the candidates in the  $i$ th state,  $(\theta_{O_i}, \theta_{M_i}, 1 - \theta_{O_i}, \theta_{M_i})$ , have a Dirichlet distribution with parameters  $(500q_{Oj} + 1, 500q_{Mj} + 1, 500(1 - q_{Oj} - q_{Mj}) + 1)$ .

Based on the posterior distribution of the state proportions, one can simulate from the posterior distribution of the electoral votes for Obama. The

dataset `election.2008` in the `LearnBayes` package contains for each state the percentage of voters in the poll for McCain `M.pct`, the percentage of voters in the poll for Obama `O.pct`, and the number of electoral votes `EV`.

```
> library(LearnBayes)
> data(election.2008)
> attach(data)
```

We write a short function `prob.Obama` that will use simulation from the Dirichlet distributions to compute the posterior probability that  $\theta_{O_j}$  exceeds  $\theta_{M_j}$  in the  $j$ th state.

```
> prob.Obama=function(j)
+ {
+   p=rdirichlet(5000,
+     500*c(M.pct[j],O.pct[j],100-M.pct[j]-O.pct[j])/100+1)
+   mean(p[,2]>p[,1])
+ }
```

We compute this Obama win probability for all states by using the `sapply` function.

```
> Obama.win.probs=sapply(1:51,prob.Obama)
```

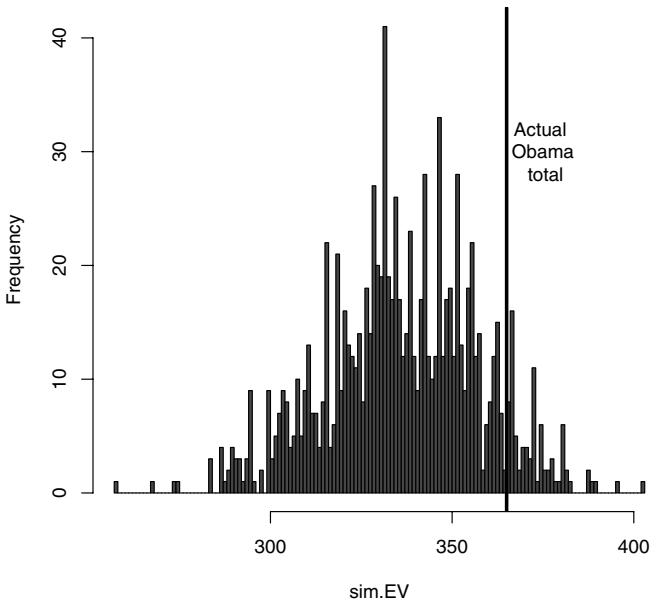
Now that we have the win probabilities, we can simulate from the posterior distribution of the Obama electoral votes by flipping a set of 51 biased coins, where the coin probabilities correspond to the Obama state win probabilities. Then we compute the number of Obama electoral votes based on the results of the coin flips. We implement one simulation using the function `sim.election` and repeat this simulation 1000 times using the `replicate` function. The vector `sim.EV` contains the number of electoral votes in the simulations.

```
> sim.election=function()
+ {
+   winner=rbinom(51,1,Obama.win.probs)
+   sum(EV*winner)
+ }
> sim.EV=replicate(1000,sim.election())
```

We construct a histogram of the posterior of  $EV_O$ , which is displayed in Figure 4.3.

```
> hist(sim.EV,min(sim.EV):max(sim.EV),col="blue")
> abline(v=365,lwd=3) # Obama received 365 votes
> text(375,30,"Actual \n Obama \n total")
```

The actual Obama electoral vote total of 365 is displayed on the graph. It would have been possible to improve our prediction by using more data than just the results of a single poll in each state. But the actual electoral vote total did fall within the 90% equal-tail prediction interval.



**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.

## 4.4 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

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

We begin in R by defining the covariate vector  $\mathbf{x}$  and the vectors of sample sizes and observed success counts  $\mathbf{n}$  and  $\mathbf{y}$ .

```
> x = c(-0.86, -0.3, -0.05, 0.73)
> n = c(5, 5, 5, 5)
> y = c(0, 1, 3, 5)
> data = cbind(x, n, y)
```

A standard classical analysis fits the model by maximum likelihood. The R function `glm` is used to do this fitting, and the summary output presents the estimates and the associated standard errors.

```
> response = cbind(y, n - y)
> results = glm(response ~ x, family = binomial)
> summary(results)
```

Call:

```
glm(formula = glmdata ~ x, family = binomial)
```

Deviance Residuals:

1	2	3	4
-0.17236	0.08133	-0.05869	0.12237

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.8466	1.0191	0.831	0.406
x	7.7488	4.8728	1.590	0.112

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 15.791412 on 3 degrees of freedom
Residual deviance: 0.054742 on 2 degrees of freedom
AIC: 7.9648
```

```
Number of Fisher Scoring iterations: 7
```



Suppose that the user has prior beliefs about the regression parameters that she inputs through the following conditional means prior. This prior is constructed by thinking about the probability of death at two different dose levels,  $x_L$  and  $x_H$ . When the dose level is  $x_L = -0.7$ , the median and 90th percentile of the probability of death  $p_L$  are respectively 0.2 and 0.5. One matches this information with a beta prior using the `beta.select` function.

```
> beta.select(list(p=.5,x=.2),list(p=.9,x=.5))
```

```
[1] 1.12 3.56
```

We see that this prior information is matched with a beta(1.12, 3.56) distribution for  $p_L$ . When the dose level is  $x_H = 0.6$ , the user believes that the median and 90th percentile of the probability of death  $p_H$  are given respectively by 0.8 and 0.98. Again using the `beta.select` function, this information is matched with a beta(2.10, 0.74) prior.

```
> beta.select(list(p=.5,x=.8),list(p=.9,x=.98))
```

```
[1] 2.10 0.74
```

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.

If this prior on  $(p_L, p_H)$  is transformed to the regression vector  $(\beta_0, \beta_1)$  through the transformation

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

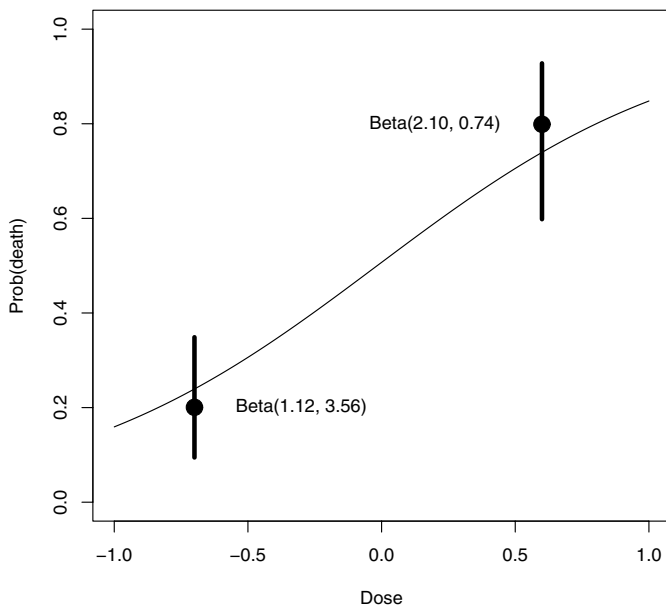
one can show that the induced prior is

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

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.



**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.

**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

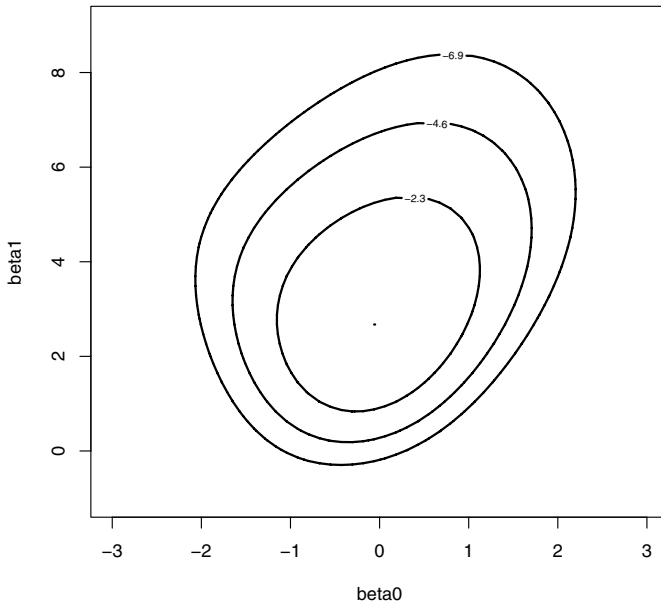
The log posterior density for  $(\beta_0, \beta_1)$  in this logistic model is contained in the R function `logisticpost`, where the `data` argument is a matrix with columns dose, number of successes, and sample size. We first combine the data (contained in the matrix `data`) with the prior data and place them in the matrix `data.new`.

```
> prior=rbind(c(-0.7, 4.68, 1.12),
+             c(0.6, 2.10, 0.74))
> data.new=rbind(data, prior)
```

To summarize the posterior distribution, we first find a rectangle that covers essentially all of the posterior probability. The maximum likelihood fit is helpful in giving a first guess at the location of this rectangle. As shown in

the contour plot displayed in Figure 4.5, we see that the rectangle  $-3 \leq \beta_0 \leq 3, -1 \leq \beta_1 \leq 9$  contains the contours that are greater than .1% of the modal value.

```
> mycontour(logisticpost,c(-3,3,-1,9),data.new,
+ xlab="beta0", ylab="beta1")
```

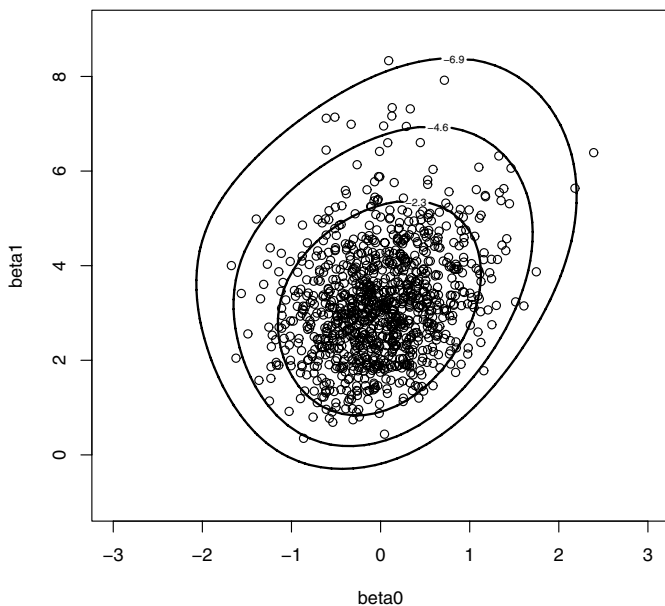


**Fig. 4.5.** Contour plot of the posterior distribution of  $(\beta_0, \beta_1)$  for the bioassay example. The contour lines are drawn at 10%, 1%, and .1% of the model height.

Now that we have found the posterior distribution, we use the function `simcontour` to simulate pairs of  $(\beta_0, \beta_1)$  from the posterior density computed on this rectangular grid. We display the contour plot with the points superimposed in Figure 4.6 to confirm that we are sampling from the posterior distribution.

```
> s=simcontour(logisticpost,c(-2,3,-1,11),data.new,1000)
> points(s)
```

We illustrate several types of inferences for this problem. Figure 4.7 displays a density estimate of the simulated values (using the R function `density`) of the slope parameter  $\beta_1$ . All of the mass of the density of  $\beta_1$  is



**Fig. 4.6.** Contour plot of the posterior distribution of  $(\beta_0, \beta_1)$  for the bioassay example. A simulated random sample from this distribution is shown on top of the contour plot.

on positive values, indicating that there is significant evidence that increasing the level of the dose does increase the probability of death.

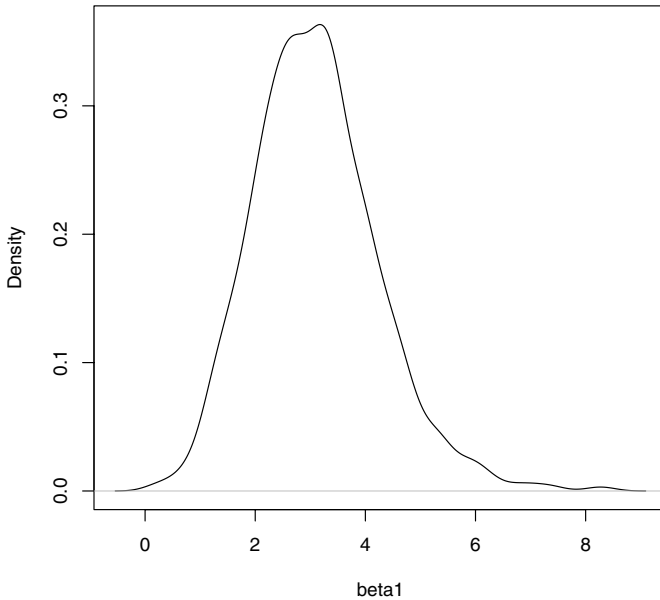
```
> plot(density(s$y), xlab="beta1")
```

In this setting, one parameter of interest is the LD-50, the value of the dose  $x$  such that the probability of death is equal to one-half. It is straightforward to show that the LD-50 is equal to  $\theta = -\beta_0/\beta_1$ . One can obtain a simulated sample from the marginal posterior density of  $\theta$  by computing a value of  $\theta$  from each simulated pair  $(\beta_0, \beta_1)$ . A histogram of the LD-50 is shown in Figure 4.8.

```
> theta=-s$x/s$y
> hist(theta, xlab="LD-50", breaks=20)
```

In contrast to the histogram of  $\beta_1$ , the LD-50 is more difficult to estimate and the posterior density of this parameter is relatively wide. We compute a 95% credible interval from the simulated draws of  $\theta$ .

```
> quantile(theta, c(.025, .975))
```



**Fig. 4.7.** Density of simulated values from the posterior of the slope parameter  $\beta_1$  in the bioassay example.

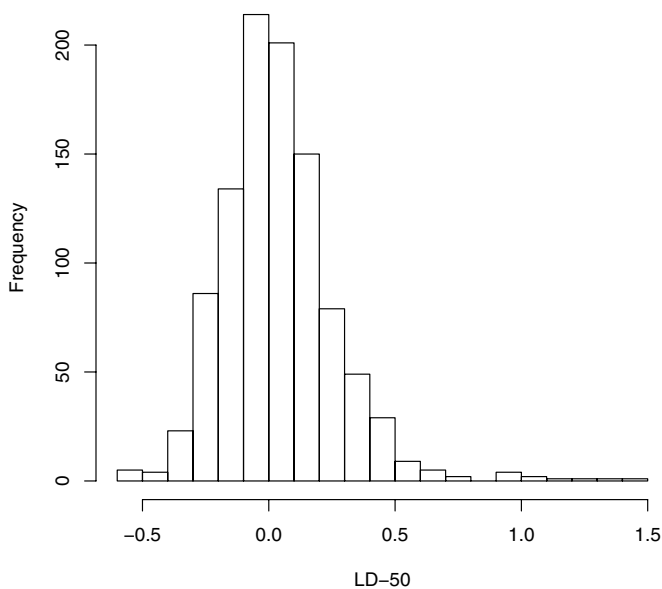
2.5%	97.5%
-0.3542899	0.5061084

The probability that  $\theta$  is contained in the interval  $(-.354, .506)$  is .95.

## 4.5 Comparing Two Proportions

Howard (1998) considers the general problem of comparing the proportions from two independent binomial distributions. Suppose we observe  $y_1$  distributed as  $\text{binomial}(n_1, p_1)$ , and  $y_2$  distributed as  $\text{binomial}(n_2, p_2)$ . One wants to know if the data favor the hypothesis  $H_1 : p_1 > p_2$  or the hypothesis  $H_2 : p_1 < p_2$  and wants a measure of the strength of the evidence in support of one hypothesis. Howard gives a broad survey of frequentist and Bayesian approaches for comparing two proportions.

From a Bayesian viewpoint, the important task is the construction of an appropriate prior distribution. In Exercise 3, we explore the assumption that  $p_1$  and  $p_2$  are independent, where each proportion is assigned a beta prior. In this case,  $p_1$  and  $p_2$  have independent beta posterior distributions and it is



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

straightforward to compute the probability of the hypotheses. However, the assumption of independence of the proportions is questionable, and we consider instead Howard's "dependent prior" that he recommends for this particular testing problem.

Suppose that one is given the information that one proportion is equal to a particular value, say  $p_1 = .8$ . This knowledge can influence a user's prior beliefs about the location of the second proportion  $p_2$ . Specifically, if the user is given that  $p_1 = .8$ , she may also believe that the value of  $p_2$  is also close to .8. This belief implies the use of dependent priors for  $p_1$  and  $p_2$ .

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  $\theta_1$ , the logit  $\theta_2$  is assumed to be normally distributed with mean  $\theta_1$  and standard deviation  $\sigma$ . 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  $\sigma$  indicates one's prior belief in the dependence between the two proportions.

Suppose that  $\alpha = \beta = \gamma = \delta = 1$ , reflecting vague prior beliefs about each individual parameter. The logarithm of the dependent prior is defined in the R function `howardprior`. Using the function `mycontour`, Figure 4.9 shows contour plots of the dependent prior for values of the association parameter  $\sigma = 2, 1, .5$ , and  $.25$ . Note that as the value of  $\sigma$  goes to zero, the prior is placing more of its mass along the line where the two proportions are equal.

```
> sigma=c(2,1,.5,.25)
> plo=.0001;phi=.9999
> par(mfrow=c(2,2))
> for (i in 1:4)
+   mycontour(howardprior,c(plo,phi,plo,phi),c(1,1,1,1,sigma[i]),
+     main=paste("sigma=",as.character(sigma[i])),
+     xlab="p1",ylab="p2")
```

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

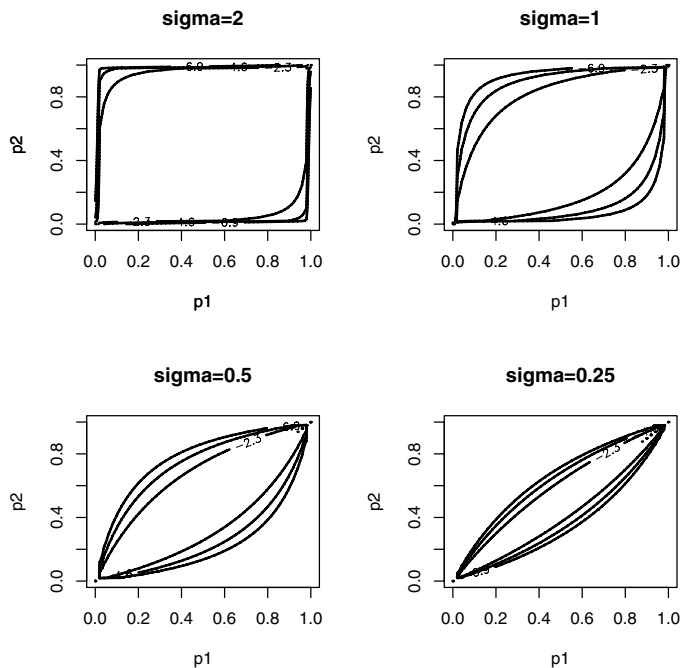
$$(\alpha + y_1, \beta + n_1 - y_1, \gamma + y_2, \delta + n_2 - y_2, \sigma).$$

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

Since the posterior distribution is of the same functional form as the prior, we can use the same `howardprior` function for the posterior calculations. In Figure 4.10, contour plots of the posterior are shown for the four values of the association parameter  $\sigma$ .



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

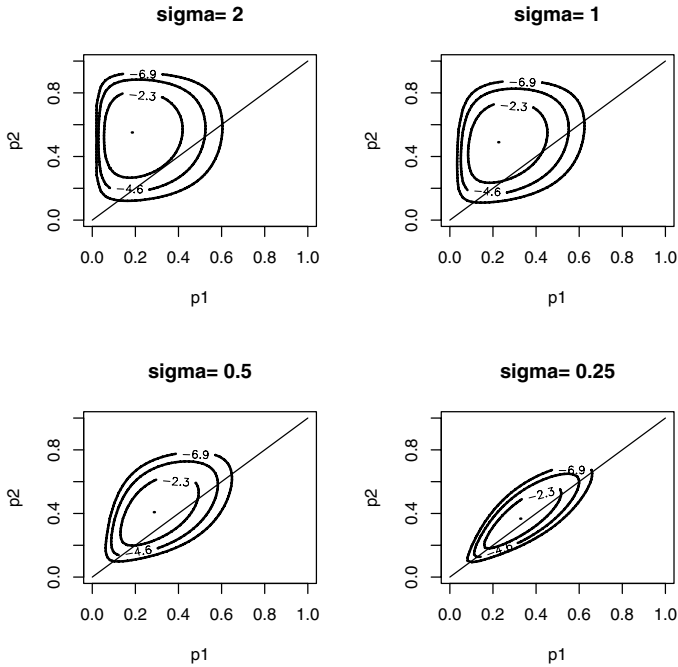
```

> sigma=c(2,1,.5,.25)
> par(mfrow=c(2,2))
> for (i in 1:4)
+{
+ mycontour(howardprior,c(plo,phi,plo,phi),
+   c(1+3,1+15,1+7,1+5,sigma[i]),
+   main=paste("sigma=",as.character(sigma[i])),
+   xlab="p1",ylab="p2")
+ lines(c(0,1),c(0,1))
+ }

```

We can test the hypothesis  $H_1 : p_1 > p_2$  simply by computing the posterior probability of this region of the parameter space. We first produce, using the function `simcontour`, a simulated sample from the posterior distribution of  $(p_1, p_2)$ , and then find the proportion of simulated pairs where  $p_1 > p_2$ . For example, we display the R commands for the computation of the posterior probability for  $\sigma = 2$ .





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

```
> s=simcontour(howardprior,c(plo,phi,plo,phi),
+ c(1+3,1+15,1+7,1+5,2),1000)
> sum(s$x>s$y)/1000
[1] 0.012
```

Table 4.4 displays the posterior probability that  $p_1$  exceeds  $p_2$  for four choices of the dependent prior parameter  $\sigma$ . 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 $\sigma$	$P(p_1 > p_2)$
2	0.012
1	0.035
.5	0.102
.25	0.201

## 4.6 Further Reading

Chapter 3 of Gelman et al. (2003) describes the normal sampling problem and other multiparameter problems from a Bayesian perspective. In particular, Gelman et al. (2003) illustrate the use of simulation when the posterior has been computed on a grid. Chapter 2 of Carlin and Louis (2009) and Lee (2004) illustrate Bayesian inference for some basic two-parameter problems. Bedrick et al. (1996) describe the use of conditional means priors for regression models. Howard (1998) gives a general discussion of inference for the two-by-two contingency table, contrasting frequentist and Bayesian approaches.

## 4.7 Summary of R Functions

**howardprior** – computes the logarithm of a dependent prior on two proportions proposed by Howard in a *Statistical Science* paper in 1998

*Usage:* `howardprior(xy, par)`

*Arguments:* `xy`, a matrix of parameter values where each row represents a value of the proportions ( $p_1$ ,  $p_2$ ); `par`, a vector containing parameter values  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\sigma$

*Value:* vector of values of the log posterior where each value corresponds to each row of the parameters in `xy`

**logisticpost** – computes the log posterior density of  $(\beta_0, \beta_1)$  when  $y_i$  are independent  $\text{binomial}(n_i, \pi_i)$  and  $\text{logit}(\pi_i) = \beta_0 + \beta_1 x_i$

*Usage:* `logisticpost(beta, data)`

*Arguments:* `beta`, a matrix of parameter values where each row represents a value of  $(\beta_0, \beta_1)$ ; `data`, a matrix of columns of covariate values  $x$ , sample sizes  $n$ , and number of successes  $y$

*Value:* vector of values of the log posterior where each value corresponds to each row of the parameters in `beta`

**mycontour** – for a general two parameter density, draws a contour graph where the contour lines are drawn at 10%, 1%, and .1% of the height at the mode

*Usage:* `mycontour(logf, limits, data, ...)`

*Arguments:* `logf`, a function that defines the logarithm of the density; `limits`, a vector of limits  $(x_{lo}, x_{hi}, y_{lo}, y_{hi})$  where the graph is to be drawn; `data`, a vector or list of parameters associated with the function `logpost`; `...`, further arguments to pass to `contour`

*Value:* a contour graph of the density is drawn

**normchi2post** – computes the log of the posterior density of a mean  $M$  and a variance  $S^2$  when a sample is taken from a normal density and a standard noninformative prior is used

*Usage:* `normchi2post(theta, data)`

*Arguments:* `theta`, a matrix of parameter values where each row is a value of  $(M, S^2)$ ; `data`, a vector containing the sample observations

*Value:* a vector of values of the log posterior where the values correspond to the rows in theta

`normpostsim` – gives a simulated sample from the joint posterior distribution of the mean and variance for a normal sampling prior with a noninformative prior

*Usage:* `normpostsim(data,m)`

*Arguments:* `data`, a vector containing the sample observations; `m`, number of simulations desired

*Value:* `mu`, vector of simulated draws of normal mean; `sigma2`, vector of simulated draws of normal variance

`rdirichlet` – simulates values from a Dirichlet distribution

*Usage:* `rdirichlet(n,par)`

*Arguments:* `n`, the number of simulations required; `par`, the vector of parameters of the Dirichlet distribution

*Value:* a matrix of simulated draws, where a row contains one simulated Dirichlet draw

`simcontour` – for a general two-parameter density defined on a grid, simulates a random sample

*Usage:* `simcontour(logf,limits,data,m)`

*Arguments:* `logf`, a function that defines the logarithm of the density; `limits`, a vector of limits (`xlo`, `xhi`, `ylo`, `yhi`) that cover the joint probability density; `data`, a vector or list of parameters associated with the function `logpost`; `m`, the size of the simulated sample

*Value:* `x`, the vector of simulated draws of the first parameter; `y`, the vector of simulated draws of the second parameter

## 4.8 Exercises

### 1. Inference about a normal population

Suppose we are interested in learning about the sleeping habits of students at a particular college. We collect  $y_1, \dots, y_{20}$ , the sleeping times (in hours) for 20 randomly selected students in a statistics course. Here are the observations:

```
9.0  8.5  7.0  8.5  6.0 12.5  6.0  9.0  8.5  7.5
8.0  6.0  9.0  8.0  7.0 10.0  9.0  7.5  5.0  6.5
```

- Assuming that the observations represent a random sample from a normal population with mean  $\mu$  and variance  $\sigma^2$  and the usual noninformative prior is placed on  $(\mu, \sigma^2)$ , simulate a sample of 1000 draws from the joint posterior distribution.
- Use the simulated sample to find 90% interval estimates for the mean  $\mu$  and the standard deviation  $\sigma$ .

- c) Suppose one is interested in estimating the upper quartile  $p_{75}$  of the normal population. Using the fact that  $p_{75} = \mu + 0.674\sigma$ , find the posterior mean and posterior standard deviation of  $p_{75}$ .

**2. The Behrens-Fisher problem**

Suppose that we observe two independent normal samples, the first distributed according to an  $N(\mu_1, \sigma_1)$  distribution, and the second according to an  $N(\mu_2, \sigma_2)$  distribution. Denote the first sample by  $x_1, \dots, x_m$  and the second sample by  $y_1, \dots, y_n$ . Suppose also that the parameters  $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$  are assigned the vague prior

$$g(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \propto \frac{1}{\sigma_1^2 \sigma_2^2}.$$

- a) Find the posterior density. Show that the vectors  $(\mu_1, \sigma_1^2)$  and  $(\mu_2, \sigma_2^2)$  have independent posterior distributions.
- b) Describe how to simulate from the joint posterior density of  $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ .
- c) The following data give the mandible lengths in millimeters for 10 male and ten female golden jackals in the collection of the British Museum. Using simulation, find the posterior density of the difference in mean mandible length between the sexes. Is there sufficient evidence to conclude that the males have a larger average?

Males

120 107 110 116 114 111 113 117 114 112

Females

110 111 107 108 110 105 107 106 111 111

**3. Comparing two proportions**

The following table gives the records of accidents in 1998 compiled by the Department of Highway Safety and Motor Vehicles in Florida.

	Injury	
	Fatal	Nonfatal
Safety Equipment in Use		
None	1601	162,527
Seat belt	510	412,368

Denote the number of accidents and fatalities when no safety equipment was in use by  $n_N$  and  $y_N$ , respectively. Similarly, let  $n_S$  and  $y_S$  denote the number of accidents and fatalities when a seat belt was in use. Assume that  $y_N$  and  $y_S$  are independent with  $y_N$  distributed as binomial( $n_N, p_N$ ) and  $y_S$  distributed as binomial( $n_S, p_S$ ). Assume a uniform prior is placed on the vector of probabilities  $(p_N, p_S)$ .

- a) Show that  $p_N$  and  $p_S$  have independent beta posterior distributions.
- b) Use the function `rbeta` to simulate 1000 values from the joint posterior distribution of  $(p_N, p_S)$ .
- c) Using your sample, construct a histogram of the relative risk  $p_N/p_S$ . Find a 95% interval estimate of this relative risk.
- d) Construct a histogram of the difference in risks  $p_N - p_S$ .

- e) Compute the posterior probability that the difference in risks exceeds 0.

#### 4. Learning from rounded data

It is a common problem for measurements to be observed in rounded form. Suppose we weigh an object five times and measure weights rounded to the nearest pound of 10, 11, 12, 11, and 9. Assume that the unrounded measurements are normally distributed with a noninformative prior distribution on the mean  $\mu$  and variance  $\sigma^2$ .

- Pretend that the observations are exact unrounded measurements. Simulate a sample of 1000 draws from the joint posterior distribution by using the algorithm described in Section 4.2.
  - Write down the correct posterior distributions for  $(\mu, \sigma^2)$ , treating the measurements as rounded.
  - By computing the correct posterior distribution on a grid of points (as in Section 4.4), simulate a sample from this distribution.
  - How do the incorrect and correct posterior distributions for  $\mu$  compare? Answer this question by comparing posterior means and variances from the two simulated samples.
5. **Estimating the parameters of a Poisson/gamma density**

Suppose that  $y_1, \dots, y_n$  are a random sample from the Poisson/gamma density

$$f(y|a, b) = \frac{\Gamma(y+a)}{\Gamma(a)y!} \frac{b^a}{(b+1)^{y+a}},$$

where  $a > 0$  and  $b > 0$ . This density is an appropriate model for observed counts that show more dispersion than predicted under a Poisson model. Suppose that  $(a, b)$  are assigned the noninformative prior proportional to  $1/(ab)^2$ . If we transform to the real-valued parameters  $\theta_1 = \log a$  and  $\theta_2 = \log b$ , the posterior density is proportional to

$$g(\theta_1, \theta_2 | \text{data}) \propto \frac{1}{ab} \prod_{i=1}^n \frac{\Gamma(y_i + a)}{\Gamma(a)y_i!} \frac{b^a}{(b+1)^{y_i+a}},$$

where  $a = \exp\{\theta_1\}$  and  $b = \exp\{\theta_2\}$ . Use this framework to model data collected by Gilchrist (1984), in which a series of 33 insect traps were set across sand dunes and the numbers of different insects caught over a fixed time were recorded. The number of insects of the taxa *Staphylinioidea* caught in the traps is shown here.

2 5 0 2 3 1 3 4 3 0 3  
2 1 1 0 6 0 0 3 0 1 1  
5 0 1 2 0 0 2 1 1 1 0

By computing the posterior density on a grid, simulate 1000 draws from the joint posterior density of  $(\theta_1, \theta_2)$ . From the simulated sample, find 90% interval estimates for the parameters  $a$  and  $b$ .

### 6. Comparison of two Poisson rates (from Antleman (1996))

A seller receives 800-number telephone orders from a first geographic area at a rate of  $\lambda_1$  per week and from a second geographic area at a rate of  $\lambda_2$  per week. Assume that incoming orders behave as if generated by a Poisson distribution; if the rate is  $\lambda$ , then the number of orders  $y$  in  $t$  weeks is distributed as  $\text{Poisson}(t\lambda)$ . Suppose a series of newspaper ads is run in the two areas for a period of four weeks, and sales for these four weeks are 260 units in area 1 and 165 units in area 2. The seller is interested in the effectiveness of these ads. One measure of this would be the probability that the sales rate in area 1 is greater than 1.5 times the sales rate in area 2:

$$P(\lambda_1 > 1.5\lambda_2).$$

Before the ads run, the seller has assessed the prior distribution for  $\lambda_1$  to be gamma with parameters 144 and 2.4 and the prior for  $\lambda_2$  to be gamma (100, 2.5).

- Show that  $\lambda_1$  and  $\lambda_2$  have independent gamma posterior distributions.
- Using the R function `rgamma`, simulate 1000 draws from the joint posterior distribution of  $(\lambda_1, \lambda_2)$ .
- Compute the posterior probability that the sales rate in area 1 is greater than 1.5 times the sales rate in area 2.

### 7. Fitting a gamma density

Suppose we observe a random sample  $y_1, \dots, y_n$  from a gamma density with shape parameter  $\alpha$  and scale parameter  $\lambda$  with density  $f(y|\alpha, \lambda) = \frac{y^{\alpha-1} \exp(-y/\lambda)}{\lambda^\alpha \Gamma(\alpha)}$ . If we place a uniform prior on  $\theta = (\alpha, \lambda)$ , then the posterior density of  $\theta$  is given by

$$g(\theta|y) \propto \prod_{i=1}^n f(y_i|\alpha, \lambda).$$

The following function `gamma.sampling.post` computes the logarithm of the posterior density:

```
gamma.sampling.post=function(theta,y)
  sum(dgamma(y,shape=theta[1],scale=theta[2],log=TRUE))
```

Suppose we use this model to fit the durations (in minutes) of the following sample of cell phone calls.

12.2, .9, .8, 5.3, 2, 1.2, 1.2, 1, .3, 1.8, 3.1, 2.8

- Compute the joint density of  $\theta$  over a suitable grid using the function `mycontour`. By simulating from the grid using the function `simcontour`, construct a 90% interval estimate for the mean parameter  $\mu = \alpha\lambda$ .

- b) Instead suppose one parameterizes the model by using the shape parameter  $\alpha$  and the rate parameter  $\beta = 1/\lambda$ . Write a function to compute the posterior density of  $(\alpha, \beta)$  (don't forget the Jacobian term) and simulate from the posterior to construct a 90% interval estimate for  $\mu$ .
- c) Instead suppose one parameterizes the model by using the shape parameter  $\alpha$  and the mean parameter  $\mu = \alpha\lambda$ . Write a function to compute the posterior density of  $(\alpha, \mu)$  (again don't forget the Jacobian term) and simulate from the posterior to construct a 90% interval estimate for  $\mu$ .
- d) Compare your three computational methods. Which is the best method for computing the interval estimate for  $\mu$ ?

### 8. Logistic modeling

A math department is interested in exploring the relationship between students' scores on the ACT test, a standard college entrance exam, and their success (getting an A or a B) in a business calculus class. Data were obtained for a sample of students; the following table gives the sample size and number of successful students for each of seven ACT scores.

ACT Score	No. of Students	No. Receiving A's and B's
16	2	0
18	7	0
20	14	6
22	26	12
24	13	7
26	14	9
28	3	3

Let  $y_i$  denote the number of successful students out of  $n_i$  with ACT score  $x_i$ . We assume that  $y_i$  is binomial( $n_i, p_i$ ), where the success probabilities follow the logistic model

$$\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_i.$$

- a) Suppose the department has some prior information that they would like to input using a conditional means prior. When the ACT score is 18, they believe that the quartiles for the success probability are 0.15 and 0.35, and when the ACT score is 26, they believe the quartiles for the success probability are 0.75 and 0.95. Using the `beta.select` function, determine the parameters for the beta distributions that match this prior information.
- b) Use the `mycontour` function together with the `logisticpost` function to find a region that contains the posterior density of  $(\beta_0, \beta_1)$ .
- c) Use the `simcontour` function to simulate 1000 draws from the posterior distribution.

- d) Use the simulated draws to find a 90% interval estimate for the probability of succeeding in the course for an ACT score equal to 20.





<http://www.springer.com/978-0-387-92297-3>

Bayesian Computation with R

Albert, J.

2009, XII, 300 p., Softcover

ISBN: 978-0-387-92297-3