3

Single-Parameter Models

3.1 Introduction

In this chapter, we introduce the use of R in summarizing the posterior distributions for several single-parameter models. We begin by describing Bayesian inference for a variance for a normal population and inference for a Poisson mean when informative prior information is available. For both problems, summarization of the posterior distribution is facilitated by the use of R functions to compute and simulate distributions from the exponential family. In Bayesian analyses, one may have limited beliefs about a parameter and there may be several priors that provide suitable matches to these beliefs. In estimating a normal mean, we illustrate the use of two distinct priors in modeling beliefs, and show that inferences may or may not be sensitive to the choice of prior. In this example, we illustrate the "brute force" method of summarizing a posterior where the density is computed by the "prior times likelihood" recipe over a fine grid. We conclude by describing a Bayesian test of the simple hypothesis that a coin is fair. The computation of the posterior probability of "fair coin" is facilitated using `beta` and `binom` functions in R.

3.2 Normal Distribution with Known Mean but Unknown Variance

Gelman et al (2003) consider a problem of estimating an unknown variance using American football scores. The focus is on the difference d between a game outcome (winning score minus losing score) and a published point spread. We observe $d_1, ..., d_n$, the observed differences between game outcomes and point spreads for n football games. If these differences are assumed to be a random sample from a normal distribution with mean 0 and unknown variance σ^2, the likelihood function is given by

$$L(\sigma^2) = (\sigma^2)^{-n/2} \exp\{-\sum_{i=1}^{n} d_i^2/(2\sigma^2)\}, \ \sigma^2 > 0.$$

Suppose the noninformative prior density $p(\sigma^2) = 1/\sigma^2$ is assigned to the variance. Then the posterior density of σ^2 is given, up to a proportionality constant, by

$$g(\sigma^2|\text{data}) \propto (\sigma^2)^{-n/2-1} \exp\{-v/(2\sigma^2)\},$$

where $v = \sum_{i=1}^{n} d_i^2$. If we define the precision parameter $P = 1/\sigma^2$, then it can be shown that P is distributed as U/v, where U has a chi-squared distribution with n degrees of freedom. Suppose we are interested in a point estimate and a 95% probability interval for the standard deviation σ.

In the following R output, we first read in the datafile `footballscores` that is available in the LearnBayes package. For each of 672 games, the data file contains `favorite` and `underdog`, the actual scores of the favorite and underdog teams, and `spread`, the published point spread. We compute the difference variable `d`. As in the preceding notation, `n` is the sample size and `v` is the sum of squares of the differences.

```
> data(footballscores)
> attach(footballscores)
> d = favorite - underdog - spread
> n = length(d)
> v = sum(d^2)
```

We simulate 1000 values from the posterior distribution of the standard deviation σ in two steps. First, we simulate values of the precision parameter $P = 1/\sigma^2$ from the scaled chi-square(n) distribution by the command `rchisq(1000, n)/v`. Then we perform the transformation $\sigma = \sqrt{1/P}$ to get values from the posterior distribution of the standard deviation σ. We use the `hist` command to construct a histogram of the draws of σ (see Fig. 3.1).

```
> P = rchisq(1000, n)/v
> s = sqrt(1/P)
> hist(s,main="")
```

The R `quantile` command is used to extract the 2.5%, 50%, and 97.5% percentiles of this simulated sample. A point estimate for σ is provided by the posterior median 13.85 . In addition, the extreme percentiles (13.2, 14.6) represent a 95% probability interval for σ.

```
> quantile(s, probs = c(0.025, 0.5, 0.975))

    2.5%      50%     97.5%
13.17012 13.85135 14.56599
```
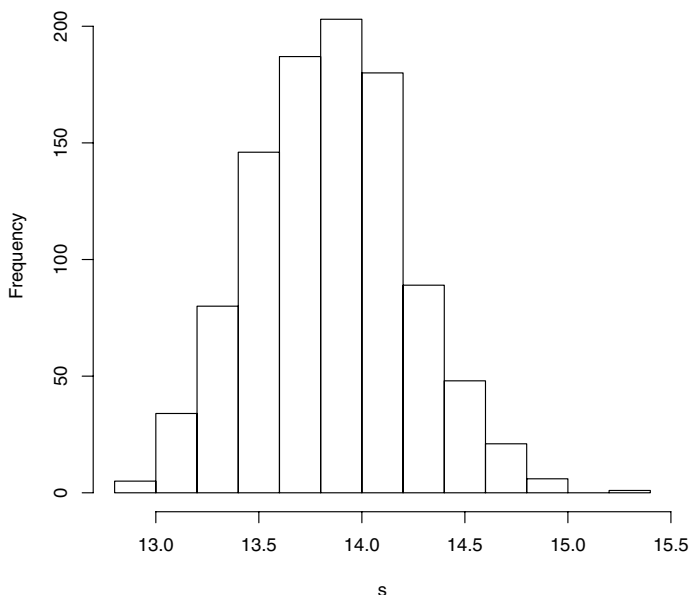
Fig. 3.1. Histogram of simulated sample of the standard deviation σ of differences between game outcomes and point spreads.

3.3 Estimating a Heart Transplant Mortality Rate

Consider the problem of learning about the rate of success of heart transplant surgery of a particular hospital in the United States. For this hospital, we observe the number of transplant surgeries n and the number of deaths within 30 days of surgery y is recorded. In addition, one can predict the probability of death for an individual patient. This prediction is based on a model that uses information such as patients' medical condition before surgery, gender, and race. Based on these predicted probabilities, one can obtain an expected number of deaths, denoted by e. A standard model assumes that the number of deaths y follows a Poisson distribution with mean $e\lambda$, and the objective is to estimate the mortality rate per unit exposure λ.

The standard estimate of λ is the maximum likelihood estimate $\hat{\lambda} = y/e$. Unfortunately, this estimate can be poor when the number of deaths y is close to zero. In this situation when small death counts are possible, it is desirable to use a Bayesian estimate that uses prior knowledge about the size of the mortality rate. A convenient choice for a prior distribution is a member of the gamma(α, β) density of the form

$$p(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda), \ \lambda > 0.$$

A convenient source of prior information is heart transplant data from a small group of hospitals that we believe has the same rate of mortality as the rate from the hospital of interest. Suppose we observe the number of deaths z_j and the exposure o_j for 10 hospitals ($j = 1, ..., 10$), where z_j is Poisson with mean $o_j\lambda$. If we assign λ the standard noninformative prior $p(\lambda) = \lambda^{-1}$, then the updated distribution for λ, given this data from the 10 hospitals, is

$$p(\lambda) \propto \lambda^{\sum_{j=1}^{10} z_j - 1} \exp\big(-(\sum_{j=1}^{10} o_j)\lambda\big).$$

Using this information, we have a gamma(α, β) prior for λ, where $\alpha = \sum_{j=1}^{10} z_j$ and $\beta = \sum_{j=1}^{10} o_j$. In this example, we have

$$\sum_{j=1}^{10} z_j = 16, \sum_{j=1}^{10} o_j = 15174,$$

and so we assign λ a gamma(16, 15174) prior.

If the observed number of deaths from surgery y_{obs} for a given hospital with exposure e is Poisson $(e\lambda)$ and λ is assigned the gamma(α, β) prior, then the posterior distribution will also have the gamma form with parameters $\alpha + y_{obs}$ and $\beta + e$. Also the (prior) predictive density of y (before any data are observed) can be computed by the formula

$$f(y) = \frac{f(y|\lambda)g(\lambda)}{g(\lambda|y)},$$

where $f(y|\lambda)$ is the Poisson$(e\lambda)$ sampling density and $g(\lambda)$ and $g(\lambda|y)$ are, respectively, the prior and posterior densities of λ.

By the model checking strategy of Box (1980), both the posterior density $g(\lambda|y)$ and the predictive density $f(y)$ play important roles in a Bayesian analysis. By use of the posterior density, one performs inference about the unknown parameter conditional on the Bayesian model that includes the assumptions of sampling density and the prior density. One can check the validity of the proposed model by inspecting the predictive density. If the observed data value y_{obs} is consistent with the predictive density $p(y)$, then the model seems reasonable. On the other hand, if y_{obs} is in the extreme tail portion of the predictive density, then this casts doubt on the validity of the Bayesian model, and perhaps the prior density or the sampling density has been misspecified.

We consider inference about the heart transplant rate for two hospitals – one that has experienced a small number of surgeries and a second that has experienced many surgeries. First consider hospital A which experienced only one death ($y_{obs} = 1$) with an exposure of $e = 66$. The standard estimate of this hospital's rate, 1/66, is suspect due to the small observed number of

deaths. The following R calculations illustrate the Bayesian calculations. After the gamma prior parameters `alpha` and `beta` and exposure `ex` are defined, the predictive density of the values $y = 0, 1, ..., 10$ are found by use of the preceding formula and the R functions `dpois` and `dgamma`. The formula for the predictive density is valid for all λ, but to ensure that there is not any underflow in the calculations, the values of $f(y)$ are computed for the prior mean value $\lambda = \alpha/\beta$. Note that practically all of the probability of the predictive density is concentrated on the two values $y = 0$ and 1. The observed number of deaths ($y_{\text{obs}} = 1$) is in the middle of this predictive distribution and so there is no reason to doubt our Bayesian model.

```
> alpha=16;beta=15174
> yobs=1; ex=66
> y=0:10
> lam=alpha/beta
> py=dpois(y, lam*ex)*dgamma(lam, shape = alpha,
+   rate = beta)/dgamma(lam, shape= alpha + y,
+   rate = beta + ex)
> cbind(y, round(py, 3))

        y
 [1,]  0 0.933
 [2,]  1 0.065
 [3,]  2 0.002
 [4,]  3 0.000
 [5,]  4 0.000
 [6,]  5 0.000
 [7,]  6 0.000
 [8,]  7 0.000
 [9,]  8 0.000
[10,]  9 0.000
[11,] 10 0.000
```

The posterior density of λ can be summarized by simulating 1000 values from the gamma density.

```
> lambdaA = rgamma(1000, shape = alpha + yobs, rate = beta + ex)
```

Let's consider the estimation of a different hospital that experiences many surgeries. Hospital B had $y_{\text{obs}} = 4$ deaths with an exposure of $e = 1767$. For this data, we again have R compute the prior predictive density and simulate 1000 draws from the posterior density using the `rgamma` command. Again we see that the observed number of deaths seems consistent with this model since $y_{\text{obs}} = 4$ is not in the extreme tails of this distribution.

```
> ex = 1767; yobs=4
> y = 0:10
```

```
> py = dpois(y, lam * ex) * dgamma(lam, shape = alpha,
+      rate = beta)/dgamma(lam, shape = alpha + y,
+      rate = beta + ex)
> cbind(y, round(py, 3))

        y
 [1,]  0 0.172
 [2,]  1 0.286
 [3,]  2 0.254
 [4,]  3 0.159
 [5,]  4 0.079
 [6,]  5 0.033
 [7,]  6 0.012
 [8,]  7 0.004
 [9,]  8 0.001
[10,]  9 0.000
[11,] 10 0.000

> lambdaB = rgamma(1000, shape = alpha + yobs, rate = beta + ex)
```

To see the impact of the prior density on the inference, it is helpful to display the prior and posterior distributions on the same graph. In Fig. 3.2, the histograms of the draws from the posterior distributions of the rates are shown for hospitals A and B. The gamma prior density is placed on top of the histogram in each case. We see that for hospital A with relatively little experience in surgeries, the prior information is significant and the posterior distribution resembles the prior distribution. In contrast, for hospital B with many surgeries, the prior information is less influential and the posterior distribution resembles the likelihood function.

```
> lambda = seq(0, max(c(lambdaA,lambdaB)), length = 500)
> par(mfrow = c(2, 1))
> hist(lambdaA, freq = FALSE, main="", ylim=c(0,1500))
> lines(lambda, dgamma(lambda, shape = alpha, rate = beta))
> hist(lambdaB, freq = FALSE, main="", ylim=c(0,1500))
> lines(lambda, dgamma(lambda, shape = alpha, rate = beta))
```

3.4 An Illustration of Bayesian Robustness

In practice, one may have incomplete prior information about a parameter in the sense that one's beliefs won't entirely define a prior density. There may be a number of different priors that match the given prior information. For example, if you believe a priori that the median of a parameter θ is 30 and its 80th percentile is 50, certainly there are many prior probability distributions that can be chosen that match these two percentiles. In this situation where
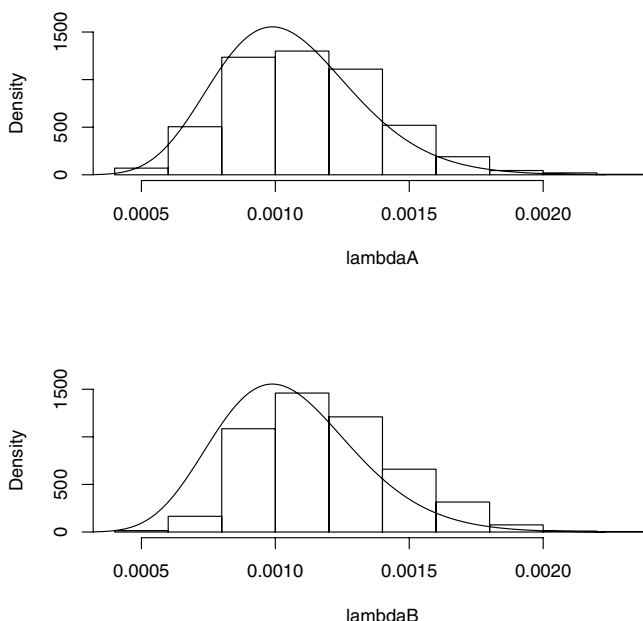
Fig. 3.2. Histograms for simulated samples from the posterior distributions for two transplant rates. The prior density for the corresponding rate is drawn in each graph.

different priors are possible, it is desirable that inferences from the posterior not be dependent on the exact functional form of the prior. A Bayesian analysis is said to be *robust* to the choice of prior if the inference is insensitive to different priors that match the user's beliefs.

   To illustrate this idea, suppose you are interested in estimating the true IQ θ for a person we'll call Joe. You believe Joe has an average intelligence and the median of your prior distribution is 100. Also you are 90% confident that Joe's IQ falls between 80 and 120. You construct a prior density by matching this information with a normal density with mean μ and standard deviation τ, or $N(\mu, \tau)$. It is straightforward to show that the parameter values that match this prior information are $\mu = 100$ and $\tau = 12.16$.

   Joe takes four IQ tests and his scores are y_1, y_2, y_3, y_4. Assuming that an individual score y is distributed $N(\theta, \sigma)$ with known standard deviation $\sigma = 15$, the observed mean score $\bar{y}$ is $N(\theta, \sigma/\sqrt{4})$.

   With the use of a normal prior in this case, the posterior density of θ will also be normal with standard deviation

$$\tau_1 = 1/(\sqrt{4/\sigma^2 + 1/\tau^2})$$

and mean

$$\mu_1 = \frac{\bar{y}(4/\sigma^2) + \mu(1/\tau^2)}{4/\sigma^2 + 1/\tau^2}.$$

We illustrate the posterior calculations for three hypothetical test results for Joe. We suppose that the observed mean test score is $\bar{y} = 110$, or $\bar{y} = 125$, or $\bar{y} = 140$. In each case we compute the posterior mean and posterior standard deviation of Joe's true IQ θ. These values are denoted by the R variables `mu1` and `tau1` in the following output.

```
> mu = 100
> tau = 12.16
> sigma = 15
> n = 4
> se = sigma/sqrt(4)
> ybar = c(110, 125, 140)
> tau1 = 1/sqrt(1/se^2 + 1/tau^2)
> mu1 = (ybar/se^2 + mu/tau^2) * tau1^2
> summ1=cbind(ybar, mu1, tau1)
> summ1

      ybar      mu1      tau1
[1,]   110 107.2442 6.383469
[2,]   125 118.1105 6.383469
[3,]   140 128.9768 6.383469
```

Let's now consider an alternative prior density to model our beliefs about Joe's true IQ. Any symmetric density instead of a normal could be used, so we use a t density with location μ, scale τ, and two degrees of freedom. Since our prior median is 100, we let the median of our t density be equal to $\mu = 100$. We find the scale parameter τ so the t density matches our prior belief that the 95th percentile of θ is equal to 120. Note that

$$P(\theta < 120) = P(T < \frac{20}{\tau}) = .95,$$

where T is a standard t variate with two degrees of freedom. It follows that

$$\tau = 20/t_2(.95),$$

where $t_v(p)$ is the pth quantile of a t random variable with v degrees of freedom. We find τ by use of the t quantile function `qt` in R.

```
> tscale = 20/qt(0.95, 2)
> tscale

[1] 6.849349
```

We display the normal and t priors in a single graph in Fig. 3.3. Although they have the same basic shape, note that the t density has significantly flatter tails – we will see that this will impact the posterior density for "extreme" test scores.

```
> theta = seq(60, 140, length = 200)
> plot(theta,1/tscale*dt((theta-mu)/tscale,2),
+   type="l",ylab="Prior Density")
> lines(theta,1/10*dnorm((theta-mu)/tau),lwd=3)
> legend(locator(1),legend=c("t density","normal density"),
+   lwd=c(1,3))
```
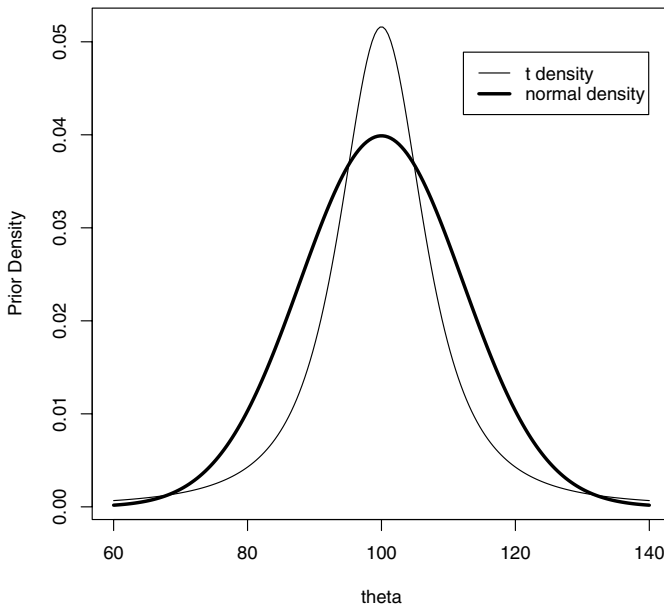

Fig. 3.3. Normal and t priors for representing prior opinion about a person's true IQ score.

We perform the posterior calculations using the t prior for each of the possible sample results. Note that the posterior density of θ is given, up to a proportionality constant, by

$$g(\theta|\text{data}) \propto \phi(\bar{y}|\theta, \sigma/\sqrt{n})g_T(\theta|v, \mu, \tau),$$

where $\phi(y|\theta, \sigma)$ is a normal density with mean θ and standard deviation σ, and $g_T(\mu|v, \mu, \tau)$ is a t density with median μ, scale parameter τ and degrees of freedom v. Since this density does not have a convenient functional form, we summarize it by a direct "prior times likelihood" approach. We construct a grid of θ values that covers the posterior density, compute the product of the normal likelihood and the t prior on the grid, and convert these products to probabilities by dividing by the sum. Essentially we are approximating the continuous posterior density by a discrete distribution on this grid. We then use this discrete distribution to compute the posterior mean and posterior standard deviation. We apply this computation algorithm for the three values of $\bar{y}$ and the posterior moments are displayed in the second and third columns of the R matrix `summ2`.

```
> summ2 = c()
> for (i in 1:3) {
+       theta = seq(60, 180, length = 500)
+       like = dnorm((theta - ybar[i])/7.5)
+       prior = dt((theta - mu)/tscale, 2)
+       post = prior * like
+       post = post/sum(post)
+       m = sum(theta * post)
+       s = sqrt(sum(theta^2 * post) - m^2)
+       summ2 = rbind(summ2, c(ybar[i], m, s))
+ }
> summ2

      [,1]      [,2]      [,3]
[1,]   110 105.2921 5.841676
[2,]   125 118.0841 7.885174
[3,]   140 135.4134 7.973498
```

Let's compare the posterior moments of θ using the two priors by combining the two R matrices `summ1` and `summ2`.

```
> cbind(summ1,summ2)

      [,1]      [,2]      [,3] [,4]      [,5]      [,6]
[1,]   110 107.2442 6.383469  110 105.2921 5.841676
[2,]   125 118.1105 6.383469  125 118.0841 7.885174
[3,]   140 128.9768 6.383469  140 135.4134 7.973498
```

When $\bar{y} = 110$, the values of the posterior mean and posterior standard deviation are similar using the normal and t priors. However, there can be substantial differences in the posterior moments using the two priors when the observed mean score is inconsistent with the prior mean. In the "extreme" case where $\bar{y} = 140$, Fig. 3.4 graphs the posterior densities for the two priors.

```
> normpost = dnorm(theta, mu1[3], tau1)
> normpost = normpost/sum(normpost)
> plot(theta,normpost,type="l",lwd=3,ylab="Posterior Density")
> lines(theta,post)
> legend(locator(1),legend=c("t prior","normal prior"),lwd=c(1,3))
```
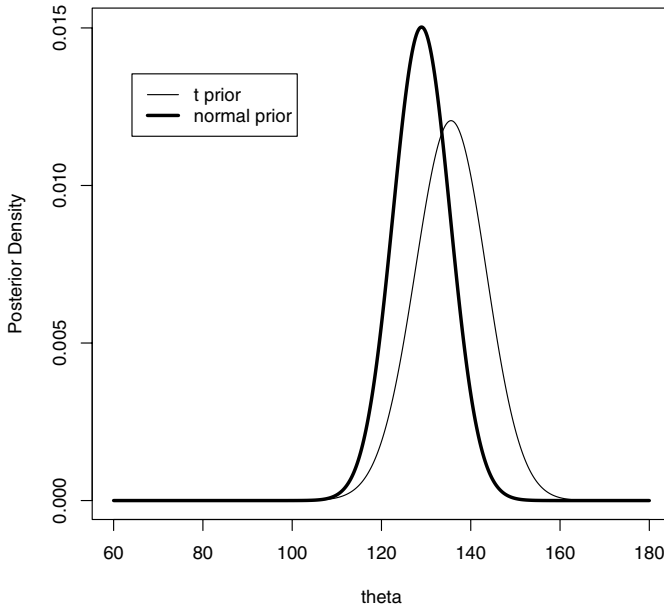

Fig. 3.4. Posterior densities for a person's true IQ using normal and t priors.

When a normal prior is used, the posterior will always be a compromise be-
tween the prior information and the observed data, even when the data result
conflicts with one's prior beliefs about the location of Joe's IQ. In contrast,
when a t prior is used, the likelihood will be in the flat-tailed portion of the
prior and the posterior will resemble the likelihood function.

In this case, the inference about the mean is robust to the choice of prior
(normal or t) when the observed mean IQ score is consistent with the prior
beliefs. But in the case when an extreme IQ score is observed, we see that the
inference is not robust to the choice of prior density.

3.5 A Bayesian Test of the Fairness of a Coin

Suppose you are interested in assessing the fairness of a coin. You observe y distributed binomial with parameters n and p and you are interested in testing the hypothesis H that $p = .5$. If y is observed, then it is usual practice to make a decision on the basis of the p-value

$$2 \times P(Y \leq y).$$

If this p-value is *small*, then you reject the hypothesis H and conclude that the coin is not fair. Suppose, for example, the coin is flipped 20 times and only 5 heads are observed. In R we compute the probability of obtaining five or fewer heads:

```
> pbinom(5, 20, 0.5)
```

```
[1] 0.02069473
```

The p-value here is $2 \times .021 = .042$. Since this value is smaller than the common significance level of .05, you would decide to reject the hypothesis H and conclude that the coin is not fair.

Let's consider this problem from a Bayesian perspective. There are two possible models here – either the coin is fair ($p = .5$) or the coin is not fair ($p \neq .5$). Suppose that you are indifferent between the two possibilities, and so you initially assign each model a probability of $1/2$. Now if you believe the coin is fair, then your entire prior distribution for p would be concentrated on the value $p = .5$. If instead the coin is unfair, you would assign a different prior distribution on $(0, 1)$, call it $g_1(p)$, that would reflect your beliefs about the unfair coin probability. Suppose you assign a beta(a, a) prior on p. This beta distribution is symmetric about .5 – it says that you believe the coin is not fair, the probability is close to $p = .5$. To summarize, your prior distribution in this testing situation can be written

$$g(p) = .5I(p = .5) + .5I(p \neq .5)g_1(p),$$

where $I(A)$ is an indicator function equal to 1 if the event A is true and otherwise equal to 0.

After observing the number of heads in n tosses, we would update our prior distribution by Bayes' rule. The posterior density for p can be written as

$$g(p|y) = \lambda(y)I(p = .5) + (1 - \lambda(y))g_1(p|y),$$

where g_1 is a beta$(a+y, a+n-y)$ density and $\lambda(y)$ is the posterior probability of the model that the coin is fair

$$\lambda(y) = \frac{.5p(y|.5)}{.5p(y|.5) + .5m_1(y)}.$$

In the expression for $\lambda(y)$, $p(y|.5)$ is the binomial density for y when $p = .5$, and $m_1(y)$ is the (prior) predictive density for y using the beta density.

In R the posterior probability of fairness $\lambda(y)$ is easily computed. The R command `dbinom` will compute the binomial probability $p(y|.5)$ and the predictive density for y can be computed using the identity

$$m_1(y) = \frac{f(y|p)g_1(p)}{g_1(p|y)}.$$

Assume first that we assign a beta(10, 10) prior for p when the coin is not fair and we observe $y = 5$ heads in $n = 20$ tosses. The posterior probability of fairness is stored in the R variable `lambda`.

```
> n = 20
> y = 5
> a = 10
> p = 0.5
> m1 = dbinom(y, n, p) * dbeta(p, a, a)/dbeta(p, a + y, a + n -
+     y)
> lambda = dbinom(y, n, p)/(dbinom(y, n, p) + m1)
> lambda

[1] 0.2802215
```

We get the surprising result that the posterior probability of the hypothesis of fairness H is .28, which is less evidence against fairness than implied by the above p-value calculation.

The function `pbetat` in the LearnBayes package performs a test of a binomial proportion. The inputs to the function are the value of p to be tested, the prior probability of that value, a vector of parameters of the beta prior when the hypothesis is not true, and a vector of numbers of successes and failures. In this example, the syntax would be

```
> pbetat(p,.5,c(a,a),c(y,n-y))

$bf
[1] 0.3893163

$post
[1] 0.2802215
```

The output variable `post` is the posterior probability that $p = .5$, which agrees with the calculation. The output variable `bf` is the Bayes factor in support of the null hypothesis which is discussed in Chapter 8.

Since the choice of the prior parameter $a = 10$ in this analysis seems arbitrary, it is natural to ask about the sensitivity of this posterior calculation to the choice of this parameter. To answer this question, we compute the posterior probability of fairness for a range of values of $\log a$. We graph the posterior probability against $\log a$ in Fig. 3.5.

```
> loga = seq(-4, 5, length = 100)
> a = exp(loga)
> m2 = dbinom(y, n, p) * dbeta(p, a, a)/dbeta(p, a + y, a + n -
+    y)
> lambda = dbinom(y, n, p)/(dbinom(y, n, p) + m2)

> plot(loga,lambda,type="l",xlab="log(a)",ylab="Prob(coin is fair)")
```
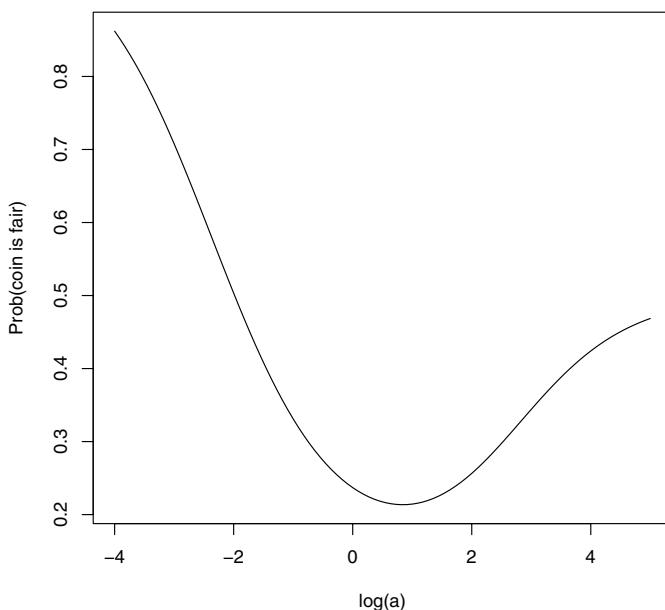

Fig. 3.5. Posterior probability that coin is fair graphed against values of the prior parameter $\log a$.

We see from this graph that the probability of fairness appears to be greater than .2 for all choices of a. Since the value of .2 is larger than the p-value calculation of .042, this suggests that the p-value is overstating the evidence against the hypothesis that the coin is fair.

Another distinction between the frequentist and Bayesian calculations is the event that led to the decision about rejecting the hypothesis that the coin was fair. The p-value calculation was based on the probability of the event "5 heads or fewer," but the Bayesian calculation was based solely on the likelihood based on the event "exactly 5 heads." That raises the question: how

would the Bayesian answers change if we observed "5 heads or fewer"? One can show that the posterior probability that the coin is fair is given by

$$\lambda(y) = \frac{.5 P_0(Y \leq 5)}{.5 P_0(Y \leq 5) + .5 P_1(Y \leq 5)}.$$

where $P_0(Y \leq 5)$ is the probability of five heads or fewer under the binomial model with $p = .5$, and $P_1(Y \leq 5)$ is the predictive probability of this event under the alternative model with a beta$(10, 10)$ prior on p. In the following R output, the cumulative probability of five heads under the binomial model is computed by the R function `pbinom`. The probability of five or fewer heads under the alternative model is computed by summing the predictive density over the six values of y.

```
> n=20
> y=5
> a=10
> p=.5
> m2=0
> for (k in 0:y)
+ {m2=m2+dbinom(k,n,p)*dbeta(p,a,a)/dbeta(p,a+k,a+n-k)}
> lambda=pbinom(y,n,p)/(pbinom(y,n,p)+m2)
> lambda

[1] 0.2184649
```

Note that the posterior probability of fairness is .218 based on the data "5 heads or fewer." This posterior probability is smaller than the value of .280 found earlier based on $y = 5$. This is a reasonable result, since observing "5 heads or fewer" is stronger evidence against fairness than the result "5 heads."

3.6 Further Reading

Carlin and Louis (2000), chapter 2, and Gelman et al (2003), chapter 2, provide general discussions of Bayesian inference for one-parameter problems. Lee (2004), Antleman (1996), and Bolstad (2004) provide extensive descriptions of inference for a variety of one-parameter models. The notion of Bayesian robustness is discussed in detail in Berger (1985). Bayesian testing for basic inference problems is outlined in Lee (2004).

3.7 Summary of R Functions

`pbetat` – Bayesian test that a proportion is equal to a specified prior using a beta prior

Usage: `pbetat(p0,prob,ab,data)`
Arguments: `p0`, the value of the proportion to be tested; `prob`, the prior probability of the hypothesis; `ab`, the vector of parameter values of the beta prior under the alternative hypothesis; `data`, vector containing the number of successes and number of failures
Value: `bf`, the Bayes factor in support of the null hypothesis; `post`, the posterior probability of the null hypothesis

3.8 Exercises

1. **Cauchy sampling model**
    Suppose one observes a random sample $y_1, ..., y_n$ from a Cauchy density with location θ and scale parameter 1. If a uniform prior is placed on θ, then the posterior density is given (up to a proportionality constant) by

    $$g(\theta|\text{data}) \propto \prod_{i=1}^{n} \frac{1}{1 + (y_i - \theta)^2}.$$

    Suppose one observes the data 0, 10, 9, 8, 11, 3, 3, 8, 8, 11.
    a) Using the R command `seq`, set up a grid of values of θ from -2 to 12 in steps of 0.1.
    b) Compute the posterior density on this grid.
    c) Plot the density and comment on its main features.
    d) Compute the posterior mean and posterior standard deviation of θ.

2. **Learning about an exponential mean**
    Suppose a random sample is taken from an exponential distribution with mean λ. If we assign the usual noninformative prior $g(\lambda) = 1/\lambda$, then the posterior density is given, up to a proportionality constant, by

    $$g(\lambda|\text{data}) = \lambda^{-n-1} \exp\{-s/\lambda\},$$

    where n is the sample size and s is the sum of the observations.
    a) Show that if we transform λ to $\theta = 1/\lambda$, then θ has a gamma density with shape parameter n and rate parameter s. (A gamma density with shape α and rate β is proportional to $h(x) = x^{\alpha-1} \exp(-\beta x)$.)
    b) In a life-testing illustration, five bulbs are tested with observed burn times (in hours) of 751, 594, 1213, 1126, and 819. Using the R function `rgamma`, simulate 1000 values from the posterior distribution of θ.
    c) By transforming these simulated draws, obtain a simulated sample from the posterior distribution of λ.
    d) Estimate the posterior probability that λ exceeds 1000 hours.

3. **Learning about the upper bound of a discrete uniform density**
    Suppose one takes independent observations $y_1, ..., y_n$ from a uniform distribution on the set $\{1, 2, ..., N\}$, where the upper bound N is unknown.

Suppose one places a uniform prior for N on the values $1, ..., B$, where B is known. Then the posterior probabilities for N are given by

$$g(N|y) \propto \frac{1}{N^n}, \ y_{(n)} \leq N \leq B,$$

where $y_{(n)}$ is the maximum observation. To illustrate this situation, suppose a tourist is waiting for a taxi in a city. During this waiting time, she observes five taxis with the numbers 43, 24, 100, 35, and 85. She assumes that taxis in this city are numbered from 1 to N, she is equally likely to observe any numbered taxi at a given time, and observations are independent. She also knows that there cannot be more than 200 taxis in the city.

a) Use R to compute the posterior probabilities of N on a grid of values.
b) Compute the posterior mean and posterior standard deviation of N.
c) Find the probability that there are more than 150 taxis in the city.

4. **Bayesian robustness**
Suppose you are about to flip a coin that you believe is fair. If p denotes the probability of flipping a head, then your "best guess" at p is .5. Moreover, you believe that it is highly likely that the coin is close to fair, which you quantify by $P(.44 < p < .56) = .9$. Consider the following two priors for p:

P1:p distributed beta(100, 100)
P2:p distributed according to the mixture prior

$$g(p) = .9 f_B(p; 500, 500) + .1 f_B(p; 1, 1),$$

where $f_B(p; a, b)$ is the beta density with parameters a and b.

a) Simulate 1000 values from each prior density P1 and P2. By summarizing the simulated samples, show that both priors match the given prior beliefs about the coin flipping probability p.
b) Suppose you flip the coin 100 times and obtain 45 heads. Simulate 1000 values from the posteriors from priors P1 and P2, and compute 90% probability intervals.
c) Suppose that you only observe 30 heads out of 100 flips. Again simulate 1000 values from the two posteriors and compute 90% probability intervals.
d) Looking at your results from (b) and (c), comment on the robustness of the inference with respect to the choice of prior density in each case.

5. **Test of a proportion**
In the standard Rhine test of extra-sensory perception (ESP), a set of cards is used where each card has a circle, a square, wavy lines, a cross, or a star. A card is selected at random from the deck and a person tries to guess the symbol on the card. This experiment is repeated 20 times and the number of correct guesses y is recorded. Let p denote the probability that the person makes a correct guess, where $p = .2$ if the person does not

have ESP and is just guessing at the card symbol. To see if the person truly has some ESP, we would like to test the hypothesis $H : p = .2$.

a) If the person identifies $y = 8$ cards correctly, compute the p-value.
b) Suppose you believe a priori that the probability that $p = .2$ is .5 and if $p \neq .2$, you assign a beta(1, 4) prior on the proportion. Using the function `pbetat`, compute the posterior probability of the hypothesis H. Compare your answer with the p-value computed in part (a).
c) The posterior probability computed in part (b) depended on your belief about plausible values of the proportion p when $p \neq .2$. For each of the following priors, compute the posterior probability of H: (1) $p \sim$ beta(.5, 2), (2) $p \sim$ beta(2, 8), (3) $p \sim$ beta(8, 32).
d) On the basis of your Bayesian computations, do you think that $y = 8$ is convincing evidence that the person really has some ESP? Explain.

6. **Learning from grouped data**

Suppose you drive on a particular interstate roadway and typically drive at a constant speed of 70 mph. One day, you pass one car and get passed by 17 cars. Suppose that the speeds are normally distributed with unknown mean μ and standard deviation $\sigma = 10$. If you pass s cars, and f cars pass you, the likelihood of μ is given by

$$L(\mu) = \Phi(70, \mu, \sigma)^s (1 - \Phi(70, \mu, \sigma))^f,$$

where $\Phi(y, \mu, \sigma)$ is the cdf of the normal distribution with mean μ and standard deviation σ. Assign the unknown mean μ a flat prior density.

a) If $s = 1$ and $f = 17$, plot the posterior density of μ.
b) Using the density found in part (a), find the posterior mean of μ.
c) Find the probability that the average speed of the cars on this interstate roadway exceeds 80 mph.