

Introduction to Bayesian Thinking

2.1 Introduction

In this chapter, the basic elements of the Bayesian inferential approach are introduced through the basic problem of learning about a population proportion. Before taking data, one has beliefs about the value of the proportion and one models his or her beliefs in terms of a prior distribution. We will illustrate the use of different functional forms for this prior. After data have been observed, one updates one's beliefs about the proportion by computing the posterior distribution. One summarizes this probability distribution to perform inferences. Also, one may be interested in predicting the likely outcomes of a new sample taken from the population.

Many of the commands in the R base package can be used in this setting. The probability distribution commands such as `dbinom` and `dbeta`, and simulation commands, such as `rbeta`, `rbinom`, and `sample`, are helpful in simulating draws from the posterior and predictive distributions. Also we illustrate the special R commands `pdisc`, `histprior`, and `discint` in the `LearnBayes` package, which are helpful in constructing priors and computing and summarizing a posterior.

2.2 Learning About the Proportion of Heavy Sleepers

Suppose a person is interested in learning about the sleeping habits of American college students. She hears that doctors recommend eight hours of sleep for an average adult. What proportion of college students get at least eight hours of sleep?

Here we think of a population consisting of all American college students and let p represent the proportion of this population who sleep (on a typical night during the week) at least eight hours. We are interested in learning about the location of p .

The value of the proportion p is unknown. In the Bayesian viewpoint, a person's beliefs about the uncertainty in this proportion are represented by a probability distribution placed on this parameter. This distribution reflects the person's subjective prior opinion about plausible values of p .

A random sample of students from a particular university will be taken to learn about this proportion. But first the researcher does some initial research to learn about the sleeping habits of college students. This research will help her in constructing a prior distribution.

In the Internet article "College Students Don't Get Enough Sleep" in *The Gamecock*, the student newspaper of the University of South Carolina (April 20, 2004), the person doing the study reads that a sample survey reports that most students spend only six hours per day sleeping. She reads a second article "Sleep on It: Implementing a Relaxation Program into the College Curriculum" in *Fresh Writing*, a 2003 publication of the University of Notre Dame. Based on a sample of 100 students, "approximately 70% reported receiving only five to six hours of sleep on the weekdays, 28% receiving seven to eight, and only 2% receiving the healthy nine hours for teenagers."

Based on this information, the person doing the study believes that college students generally get less than eight hours of sleep and so p (the proportion that sleep at least eight hours) is likely smaller than .5. After some reflection, her best guess at the value of p is .3. But it is very plausible that this proportion could be any value in the interval from 0 to .5.

A sample of 27 students is taken – in this group, 11 record that they had at least eight hours of sleep the previous night. Based on the prior information and these observed data, the researcher is interested in estimating the proportion p . In addition, she is interested in predicting the number of students that get at least eight hours of sleep if a new sample of 20 students is taken.

Suppose that our prior density for p is denoted by $g(p)$. If we regard a "success" as sleeping at least eight hours and we take a random sample with s successes and f failures, then the likelihood function is given by

$$L(p) \propto p^s (1 - p)^f, \quad 0 < p < 1.$$

The posterior density for p , by Bayes' rule, is obtained, up to a proportionality constant, by multiplying the prior density by the likelihood:

$$g(p|\text{data}) \propto g(p)L(p).$$

We demonstrate posterior distribution calculations using three different choices of the prior density g corresponding to three methods for representing the researcher's prior knowledge about the proportion.

2.3 Using a Discrete Prior

A simple approach for assessing a prior for p is to write down a list of plausible proportion values and then assign weights to these values. The person in our

example believes that

.05, .15, .25, .35, .45, .55, .65, .75, .85, .95

are possible values for p . Based on her beliefs, she assigns these values the corresponding weights

1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0,

which can be converted to prior probabilities by dividing each weight by the sum. In R, we define `p` to be the vector of proportion values and `prior` the corresponding weights that we normalize to probabilities. The `plot` command is used with the “histogram” type option to graph the prior distribution, and Figure 2.1 displays the graph.

```
> p = seq(0.05, 0.95, by = 0.1)
> prior = c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
> prior = prior/sum(prior)
> plot(p, prior, type = "h", ylab="Prior Probability")
```

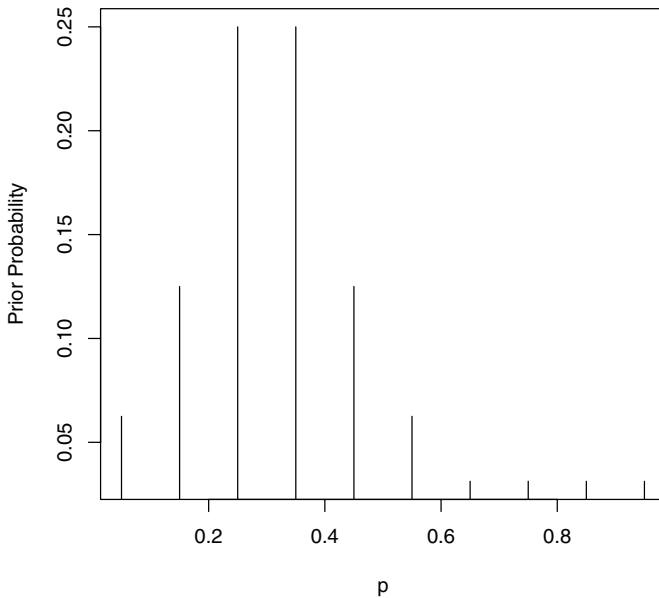


Fig. 2.1. A discrete prior distribution for a proportion p .

In our example, 11 of 27 students sleep a sufficient number of hours, so $s = 11$ and $f = 16$, and the likelihood function is

$$L(p) \propto p^{11}(1-p)^{16}, \quad 0 < p < 1.$$

(Note that the likelihood is a beta density with parameters $s + 1 = 12$ and $f + 1 = 17$.) The R function `pdisc` in the package `LearnBayes` computes the posterior probabilities. To use `pdisc`, one inputs the vector of proportion values `p`, the vector of prior probabilities `prior`, and a data vector `data` consisting of s and f . The output of `pdisc` is a vector of posterior probabilities. The `cbind` command is used to display a table of the prior and posterior probabilities. The `xyplot` function in the `lattice` package is used to construct comparative line graphs of the prior and posterior distributions in Figure 2.2.

```
> data = c(11, 16)
> post = pdisc(p, prior, data)
> round(cbind(p, prior, post), 2)

      p prior post
[1,] 0.05 0.03 0.00
[2,] 0.15 0.18 0.00
[3,] 0.25 0.28 0.13
[4,] 0.35 0.25 0.48
[5,] 0.45 0.16 0.33
[6,] 0.55 0.07 0.06
[7,] 0.65 0.02 0.00
[8,] 0.75 0.00 0.00
[9,] 0.85 0.00 0.00
[10,] 0.95 0.00 0.00

> library(lattice)
> PRIOR=data.frame("prior",p,prior)
> POST=data.frame("posterior",p,post)
> names(PRIOR)=c("Type", "P", "Probability")
> names(POST)=c("Type", "P", "Probability")
> data=rbind(PRIOR,POST)
> xyplot(Probability~P|Type,data=data,layout=c(1,2),
+       type="h",lwd=3,col="black")
```

Here we note that most of the posterior probability is concentrated on the values $p = .35$ and $p = .45$. If we combine the probabilities for the three most likely values, we can say the posterior probability that p falls in the set $\{.25, .35, .45\}$ is equal to $.940$.

2.4 Using a Beta Prior

Since the proportion is a continuous parameter, an alternative approach is to construct a density $g(p)$ on the interval $(0, 1)$ that represents the person's

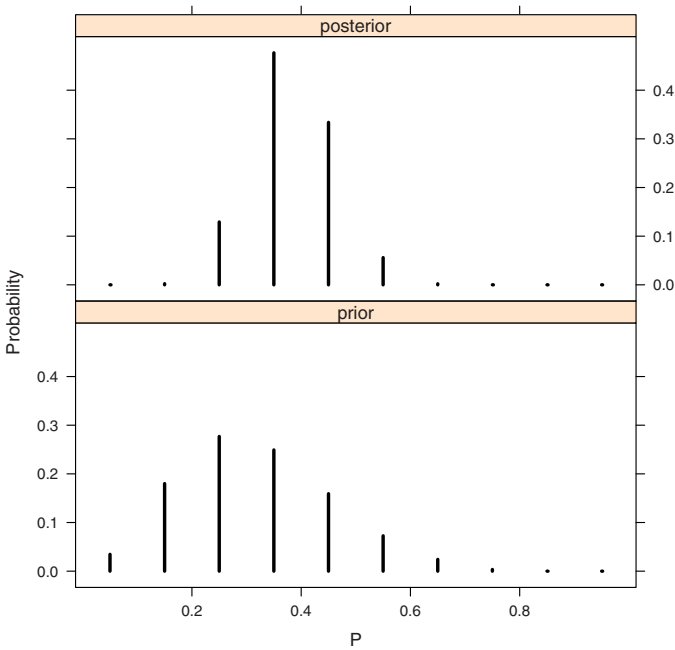


Fig. 2.2. Prior and posterior distributions for a proportion p using a discrete prior.

initial beliefs. Suppose she believes that the proportion is equally likely to be smaller or larger than $p = .3$. Moreover, she is 90% confident that p is less than $.5$. A convenient family of densities for a proportion is the beta with kernel proportional to

$$g(p) \propto p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1,$$

where the hyperparameters a and b are chosen to reflect the user's prior beliefs about p . The mean of a beta prior is $m = a/(a+b)$ and the variance of the prior is $v = m(1-m)/(a+b+1)$, but it is difficult in practice for a user to assess values of m and v to obtain values of the beta parameters a and b . It is easier to obtain a and b indirectly through statements about the percentiles of the distribution. Here the person believes that the median and 90th percentiles of the proportion are given, respectively, by $.3$ and $.5$. The function `beta.select` in the `LearnBayes` package is useful for finding the shape parameters of the beta density that match this prior knowledge. The inputs to `beta.select` are two lists, `quantile1` and `quantile2`, that define these two prior percentiles, and the function returns the values of the matching beta parameters.

```
> quantile2=list(p=.9,x=.5)
> quantile1=list(p=.5,x=.3)
> beta.select(quantile1,quantile2)
```

```
[1] 3.26 7.19
```

We see that this prior information is matched with a beta density with $a = 3.26$ and $b = 7.19$. Combining this beta prior with the likelihood function, one can show that the posterior density is also of the beta form with updated parameters $a + s$ and $b + f$.

$$g(p|\text{data}) \propto p^{a+s-1}(1-p)^{b+f-1}, \quad 0 < p < 1,$$

where $a+s = 3.26+11$ and $b+f = 7.19+16$. (This is an example of a conjugate analysis, where the prior and posterior densities have the same functional form.) Since the prior, likelihood, and posterior are all in the beta family, we can use the R command `dbeta` to compute the values of the prior, likelihood, and posterior. These three densities are displayed using three applications of the R `curve` command in the same graph in Figure 2.3. This figure helps show that the posterior density in this case compromises between the initial prior beliefs and the information in the data.

```
> a = 3.26
> b = 7.19
> s = 11
> f = 16
> curve(dbeta(x,a+s,b+f), from=0, to=1,
+       xlab="p",ylab="Density",lty=1,lwd=4)
> curve(dbeta(x,s+1,f+1),add=TRUE,lty=2,lwd=4)
> curve(dbeta(x,a,b),add=TRUE,lty=3,lwd=4)
> legend(.7,4,c("Prior","Likelihood","Posterior"),
+       lty=c(3,2,1),lwd=c(3,3,3))
```

We illustrate different ways of summarizing the beta posterior distribution to make inferences about the proportion of heavy sleepers p . The beta cdf and inverse cdf functions `pbeta` and `qbeta` are useful in computing probabilities and constructing interval estimates for p . Is it likely that the proportion of heavy sleepers is greater than .5? This is answered by computing the posterior probability $P(p \geq .5|\text{data})$, which is given by the R command

```
> 1 - pbeta(0.5, a + s, b + f)
```

```
[1] 0.0684257
```

This probability is small, so it is unlikely that more than half of the students are heavy sleepers. A 90% interval estimate for p is found by computing the 5th and 95th percentiles of the beta density:

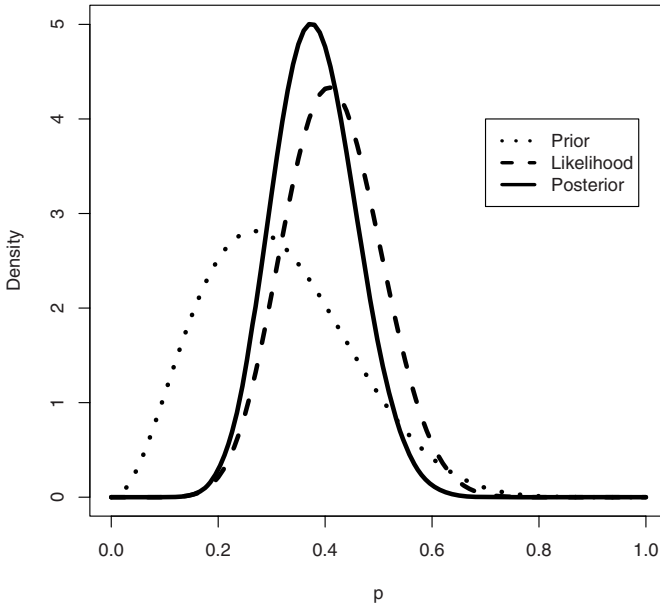


Fig. 2.3. The prior density $g(p)$, the likelihood function $L(p)$, and the posterior density $g(p|\text{data})$ for learning about a proportion p .

```
> qbeta(c(0.05, 0.95), a + s, b + f)
```

```
[1] 0.2562364 0.5129274
```

A 90% posterior credible interval for the proportion is (0.256, 0.513).

These summaries are exact because they are based on R functions for the beta posterior density. An alternative method of summarization of a posterior density is based on simulation. In this case, we can simulate a large number of values from the beta posterior density and summarize the simulated output. Using the random beta command `rbeta`, we simulate 1000 random proportion values from the $\text{beta}(a + s, b + f)$ posterior by using the command

```
> ps = rbeta(1000, a + s, b + f)
```

and display the posterior as a histogram of the simulated values in Figure 2.4.

```
> hist(ps, xlab="p", main="")
```

The probability that the proportion is larger than .5 is estimated using the proportion of simulated values in this range.

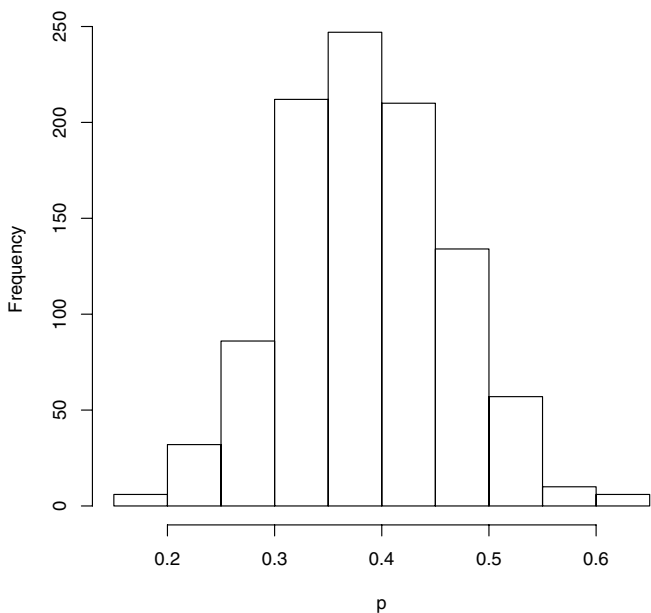


Fig. 2.4. A simulated sample from the beta posterior distribution of p .

```
> sum(ps >= 0.5)/1000
```

```
[1] 0.075
```

A 90% interval estimate can be estimated by the 5th and 95th sample quantiles of the simulated sample.

```
> quantile(ps, c(0.05, 0.95))
```

```
      5%      95%
0.2599039 0.5172406
```

Note that these summaries of the posterior density for p based on simulation are approximately equal to the exact values based on calculations from the beta distribution.

2.5 Using a Histogram Prior

Although there are computational advantages to using a beta prior, it is straightforward to perform posterior computations for any choice of prior.

We outline a “brute-force” method of summarizing posterior computations for an arbitrary prior density $g(p)$.

- Choose a grid of values of p over an interval that covers the posterior density.
- Compute the product of the likelihood $L(p)$ and the prior $g(p)$ on the grid.
- Normalize by dividing each product by the sum of the products. In this step, we are approximating the posterior density by a discrete probability distribution on the grid.
- Using the R command `sample`, take a random sample with replacement from the discrete distribution.

The resulting simulated draws are an approximate sample from the posterior distribution.

We illustrate this “brute-force” algorithm for a “histogram” prior that may better reflect the person’s prior opinion about the proportion p . Suppose it is convenient for our person to state her prior beliefs about the proportion of heavy sleepers by dividing the range of p into ten subintervals $(0, .1)$, $(.1, .2)$, \dots , $(.9, 1)$, and then assigning probabilities to the intervals. The person in our example assigns the weights 1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0 to these intervals – this can be viewed as a continuous version of the discrete prior used earlier.

In R, we represent this histogram prior with the vector `midpt`, which contains the midpoints of the intervals, and the vector `prior`, which contains the associated prior weights. We convert the prior weights to probabilities by dividing each weight by the sum. We graph this prior in Figure 2.5 using the R functions `curve` and `histprior` in the `LearnBayes` package.

```
> midpt = seq(0.05, 0.95, by = 0.1)
> prior = c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
> prior = prior/sum(prior)

> curve(histprior(x,midpt,prior), from=0, to=1,
+       ylab="Prior density",ylim=c(0,.3))
```

We compute the posterior density by multiplying the histogram prior by the likelihood function. (Recall that the likelihood function for a binomial density is given by a $\text{beta}(s+1, f+1)$ density; this function is available using the `dbeta` function.) In Figure 2.6, the posterior density is displayed using the `curve` function.

```
> curve(histprior(x,midpt,prior) * dbeta(x,s+1,f+1),
+       from=0, to=1, ylab="Posterior density")
```

To obtain a simulated sample from the posterior density by our algorithm, we first construct an equally spaced grid of values of the proportion p and compute the product of the prior and likelihood on this grid. Then we convert the products on the grid to probabilities.

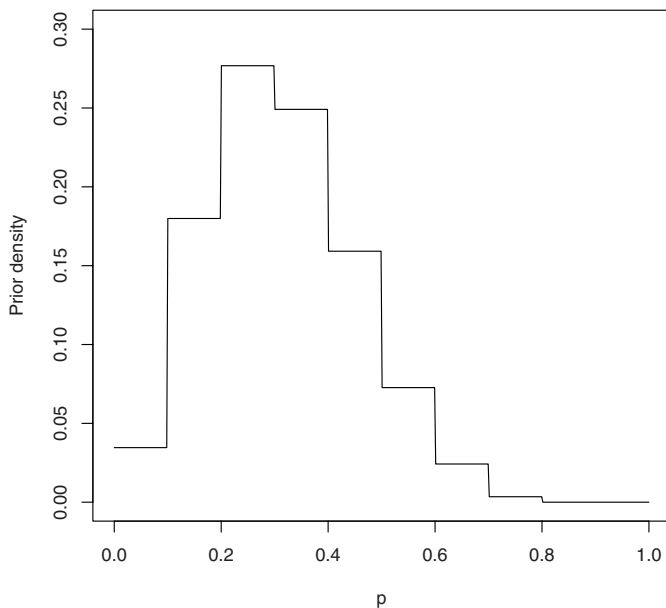


Fig. 2.5. A histogram prior for a proportion p .

```
> p = seq(0, 1, length=500)
> post = histprior(p, midpt, prior) *
+       dbeta(p, s+1, f+1)
> post = post/sum(post)
```

Last, we take a sample with replacement from the grid using the R function `sample`.

```
> ps = sample(p, replace = TRUE, prob = post)
```

Figure 2.7 shows a histogram of the simulated values.

```
> hist(ps, xlab="p", main="")
```

The simulated draws can be used as before to summarize any feature of the posterior distribution of interest.

2.6 Prediction

We have focused on learning about the population proportion of heavy sleepers p . Suppose our person is also interested in predicting the number of heavy

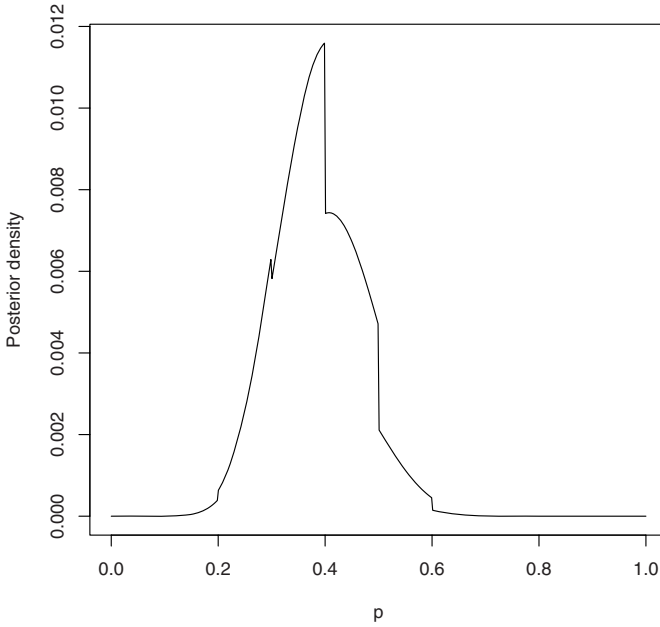


Fig. 2.6. The posterior density for a proportion using a histogram prior

sleepers \tilde{y} in a future sample of $m = 20$ students. If the current beliefs about p are contained in the density $g(p)$, then the predictive density of \tilde{y} is given by

$$f(\tilde{y}) = \int f(\tilde{y}|p)g(p)dp.$$

If g is a prior density, then we refer to this as the *prior* predictive density, and if g is a posterior, then f is a *posterior* predictive density.

We illustrate the computation of the predictive density using the different forms of prior density described in this chapter. Suppose we use a discrete prior where $\{p_i\}$ represent the possible values of the proportion with respective probabilities $\{g(p_i)\}$. Let $f_B(y|n, p)$ denote the binomial sampling density given values of the sample size n and proportion p :

$$f_B(y|n, p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, \dots, n.$$

Then the predictive probability of \tilde{y} successes in a future sample of size m is given by

$$f(\tilde{y}) = \sum f_B(\tilde{y}|m, p_i)g(p_i).$$

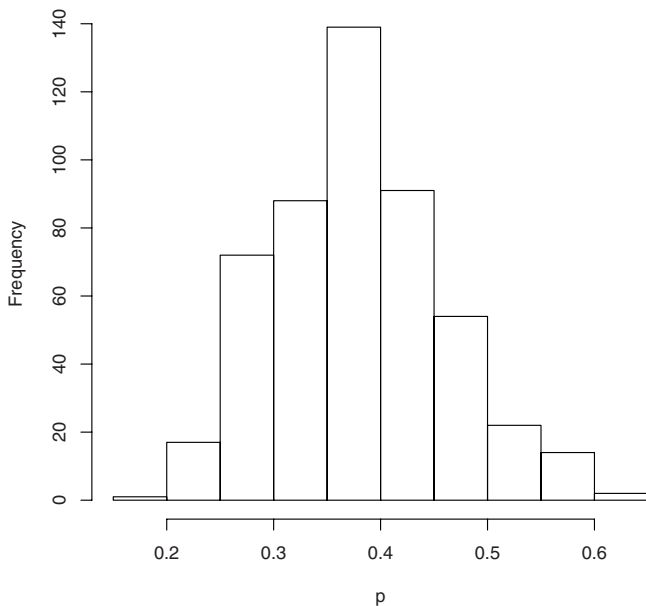


Fig. 2.7. A histogram of simulated draws from the posterior distribution of p with the use of a histogram prior.

The function `pdiscp` in the `LearnBayes` package can be used to compute the predictive probabilities when p is given a discrete distribution. As before, `p` is a vector of proportion values and `prior` a vector of current probabilities. The remaining arguments are the future sample size `m` and a vector `ys` of numbers of successes of interest. The output is a vector of the corresponding predictive probabilities.

```
> p=seq(0.05, 0.95, by=.1)
> prior = c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
> prior=prior/sum(prior)
> m=20; ys=0:20
> pred=pdiscp(p, prior, m, ys)
> round(cbind(0:20,pred),3)
```

```
      pred
[1,] 0 0.020
[2,] 1 0.044
[3,] 2 0.069
[4,] 3 0.092
[5,] 4 0.106
```

```

[6,] 5 0.112
[7,] 6 0.110
[8,] 7 0.102
[9,] 8 0.089
[10,] 9 0.074
[11,] 10 0.059
[12,] 11 0.044
[13,] 12 0.031
[14,] 13 0.021
[15,] 14 0.013
[16,] 15 0.007
[17,] 16 0.004
[18,] 17 0.002
[19,] 18 0.001
[20,] 19 0.000
[21,] 20 0.000

```

We see from the output that the most likely numbers of successes in this future sample are $\tilde{y} = 5$ and $\tilde{y} = 6$.

Suppose instead that we model our beliefs about p using a $\text{beta}(a, b)$ prior. In this case, we can analytically integrate out p to get a closed-form expression for the predictive density,

$$f(\tilde{y}) = \int f_B(\tilde{y}|m, p)g(p)dp \\ = \binom{m}{\tilde{y}} \frac{B(a + \tilde{y}, b + m - \tilde{y})}{B(a, b)}, \quad \tilde{y} = 0, \dots, m,$$

where $B(a, b)$ is the beta function. The predictive probabilities using the beta density are computed using the function `pbetap`. The inputs to this function are the vector `ab` of beta parameters a and b , the size of the future sample `m`, and the vector of numbers of successes `y`. The output is a vector of predictive probabilities corresponding to the values in `y`. We illustrate this computation using the $\text{beta}(3.26, 7.19)$ prior used to reflect the person's beliefs about the proportion of heavy sleepers at the school.

```

> ab=c(3.26, 7.19)
> m=20; ys=0:20
> pred=pbetap(ab, m, ys)

```

We have illustrated the computation of the predictive density for two choices of prior densities. One convenient way of computing a predictive density for *any* prior is by simulation. To obtain \tilde{y} , we first simulate, say, p^* from $g(p)$, and then simulate \tilde{y} from the binomial distribution $f_B(\tilde{y}|p^*)$.

We demonstrate this simulation approach for the $\text{beta}(3.26, 7.19)$ prior. We first simulate 1000 draws from the prior and store the simulated values in `p`:

```
> p=rbeta(1000, 3.26, 7.19)
```

Then we simulate values of \tilde{y} for these random ps using the `rbinom` function.

```
> y = rbinom(1000, 20, p)
```

To summarize the simulated draws of \tilde{y} , we first use the `table` command to tabulate the distinct values.

```
> table(y)
```

```
y
 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14
13 32 65 103 102 115 114 115 95 83 58 36 29 14 15
15 16
 6  5
```

We save the frequencies of \tilde{y} in a vector `freq`. Then we convert the frequencies to probabilities by dividing each frequency by the sum and use the `plot` command to graph the predictive distribution (see Figure 2.8).

```
> freq=table(y)
> ys=as.integer(names(freq))
> predprob=freq/sum(freq)
> plot(ys,predprob,type="h",xlab="y",
+      ylab="Predictive Probability")
```

Suppose we wish to summarize this discrete predictive distribution by an interval that covers at least 90% of the probability. The R function `discint` in the `LearnBayes` package is useful for this purpose. In the output, the vector `ys` contains the values of \tilde{y} and `predprob` contains the associated probabilities found from the table output. The matrix `dist` contains the probability distribution with the columns `ys` and `predprob`. The function `discint` has two inputs: the matrix `dist` and a given coverage probability `covprob`. The output is a list where the component `set` gives the credible set and `prob` gives the exact coverage probability.

```
> dist=cbind(ys,predprob)
```

```
> dist
      ys predprob
[1,]  0  0.013
[2,]  1  0.032
[3,]  2  0.065
[4,]  3  0.103
[5,]  4  0.102
[6,]  5  0.115
[7,]  6  0.114
[8,]  7  0.115
```

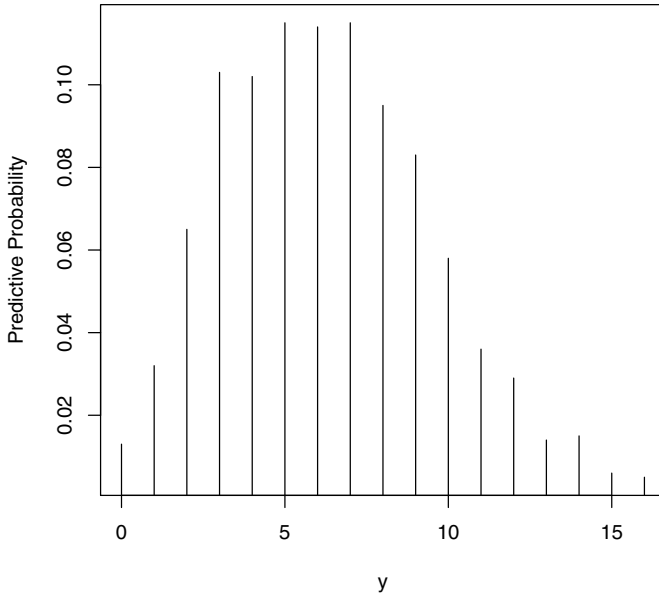


Fig. 2.8. A graph of the predictive probabilities of the number of sleepers \tilde{y} in a future sample of size 20 when the proportion is assigned a beta(3.26, 7.19) prior.

```
[9,] 8 0.095
[10,] 9 0.083
[11,] 10 0.058
[12,] 11 0.036
[13,] 12 0.029
[14,] 13 0.014
[15,] 14 0.015
[16,] 15 0.006
[17,] 16 0.005
```

```
> covprob=.9
> discint(dist,covprob)
```

```
$prob
[1] 0.918
```

```
$set
[1] 1 2 3 4 5 6 7 8 9 10 11
```

We see that the probability that \tilde{y} falls in the interval $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ is 91.8%. To say it in a different way, let $\tilde{y}/20$ denote the proportion of sleepers in the future sample. The probability that this sample proportion falls in the interval $[1/20, 11/20]$ is 91.8%. As expected, this interval is much wider than a 91.8% probability interval for the population proportion p . In predicting a future sample proportion, there are two sources of uncertainty, the uncertainty in the value of p and the binomial uncertainty in the value of \tilde{y} , and the predictive interval is relatively long since it incorporates both types of uncertainty.

2.7 Further Reading

A number of books are available that describe the basic tenets of Bayesian thinking. Berry (1996) and Albert and Rossman (2001) describe the Bayesian approach for proportions at an introductory statistics level. Albert (1996) describes Bayesian computational algorithms for proportions using the statistics package Minitab. Antleman (1996) and Bolstad (2004) provide elementary descriptions of Bayesian thinking suitable for undergraduate statistics classes.

2.8 Summary of R Functions

beta.select – finds the shape parameters of a beta density that matches knowledge of two quantiles of the distribution

Usage: `beta.select(quantile1,quantile2)`

Arguments: **quantile1**, list with components **p**, the value of the first probability, and **x**, the value of the first quantile; **quantile2**, list with components **p**, the value of the second probability, and **x**, the value of the second quantile

Value: vector of shape parameters of the matching beta distribution

discint – computes a highest probability interval for a discrete distribution

Usage: `discint(dist,prob)`

Arguments: **dist**, a probability distribution written as a matrix, where the first column contains the values and the second column contains the probabilities; **prob**, the probability content of interest

Value: **prob**, the exact probability content of the interval, and **set**, the set of values of the probability interval

histprior – computes the density of a probability distribution defined on a set of equal-width intervals

Usage: `histprior(p,midpts,prob)`

Arguments: **p**, the vector of values for which the density is to be computed; **midpts**, the vector of midpoints of the intervals; **prob**, the vector of probabilities of the intervals

Value: vector of values of the probability density

pbetap – computes the predictive distribution for the number of successes of a future binomial experiment with a beta distribution for the proportion

Usage: `pbetap(ab, n, s)`

Arguments: `ab`, the vector of parameters of the beta prior; `n`, the size of the future binomial sample; `s`, the vector of the numbers of successes for a future binomial experiment

Value: the vector of predictive probabilities for the values in the vector `s`

pdisc – computes the posterior distribution for a proportion for a discrete prior distribution

Usage: `pdisc(p, prior, data)`

Arguments: `p`, a vector of proportion values; `prior`, a vector of prior probabilities; `data`, a vector consisting of the number of successes and number of failures

Value: the vector of posterior probabilities

pdiscp – computes the predictive distribution for the number of successes of a future binomial experiment with a discrete distribution for the proportion

Usage: `pdiscp(p, probs, n, s)`

Arguments: `p`, the vector of proportion values; `probs`, the vector of probabilities; `n`, the size of the future binomial sample; `s`, the vector of the numbers of successes for a future binomial experiment

Value: the vector of predictive probabilities for the values in the vector `s`

2.9 Exercises

1. Estimating a proportion with a discrete prior

Bob claims to have ESP. To test this claim, you propose the following experiment. You will select one card from four large cards with different geometric figures, and Bob will try to identify it. Let p denote the probability that Bob is correct in identifying the figure for a single card. You believe that Bob has no ESP ability ($p = .25$), but there is a small chance that p is either larger or smaller than $.25$. After some thought, you place the following prior distribution on p :

p	0	.125	.250	.375	.500	.625	.750	.875	1
$g(p)$.001	.001	.950	.008	.008	.008	.008	.008	.008

Suppose that the experiment is repeated ten times and Bob is correct six times and incorrect four times. Using the function `pdisc`, find the posterior probabilities of these values of p . What is your posterior probability that Bob has no ESP ability?

2. Estimating a proportion with a histogram prior

Consider the following experiment. Hold a penny on edge on a flat hard surface, and spin it with your fingers. Let p denote the probability that it lands heads. To estimate this probability, we will use a histogram to model

our prior beliefs about p . Divide the interval $[0,1]$ into the ten subintervals $[0, .1]$, $[.1, .2]$, ..., $[.9, 1]$, and specify probabilities that p is in each interval. Next spin the penny 20 times and count the number of successes (heads) and failures (tails). Simulate from the posterior distribution by (1) computing the posterior density of p on a grid of values on $(0, 1)$ and (2) taking a simulated sample with replacement from the grid. (The functions `histprior` and `sample` are helpful in this computation.) How have the interval probabilities changed on the basis of your data?

3. Estimating a proportion and prediction of a future sample

A study reported on the long-term effects of exposure to low levels of lead in childhood. Researchers analyzed children's shed primary teeth for lead content. Of the children whose teeth had a lead content of more than 22.22 parts per million (ppm), 22 eventually graduated from high school and 7 did not. Suppose your prior density for p , the proportion of all such children who will graduate from high school, is `beta(1, 1)`, and so your posterior density is `beta(23, 8)`.

- Use the function `qbeta` to find a 90% interval estimate for p .
- Use the function `pbeta` to find the probability that p exceeds .6.
- Use the function `rbeta` to take a simulated sample of size 1000 from the posterior distribution of p .
- Suppose you find ten more children who have a lead content of more than 22.22 ppm. Find the predictive probability that nine or ten of them will graduate from high school. (Use your simulated sample from part (c) and the `rbinom` function to take a simulated sample from the predictive distribution.)

4. Contrasting predictions using two different priors

Suppose two persons are interested in estimating the proportion p of students at a college who commute to school. Suppose Joe uses a discrete prior given in the following table:

p	0.1	0.2	0.3	0.4	0.5
$g(p)$	0.5	0.2	0.2	0.05	0.05

Sam decides instead to use a `beta(3, 12)` prior for the proportion p .

- Use R to compute the mean and standard deviation of p for Joe's prior and for Sam's prior. Based on this computation, do Joe and Sam have similar prior beliefs about the location of p ?
 - Suppose one is interested in predicting the number of commuters y in a future sample of size 12. Use the functions `pdiscp` and `pbetap` to compute the predictive probabilities of y using both Joe's prior and Sam's prior. Do the two people have similar beliefs about the outcomes of a future sample?
- ### 5. Estimating a normal mean with a discrete prior
- Suppose you are interested in estimating the average total snowfall per year μ (in inches) for a large city on the East Coast of the United States.

Assume individual yearly snow totals y_1, \dots, y_n are collected from a population that is assumed to be normally distributed with mean μ and known standard deviation $\sigma = 10$ inches.

- a) Before collecting data, suppose you believe that the mean snowfall μ can be the values 20, 30, 40, 50, 60, and 70 inches with the following probabilities:

μ	20	30	40	50	60	70
$g(\mu)$.1	.15	.25	.25	.15	.1

Place the values of μ in the vector `mu` and the associated prior probabilities in the vector `prior`.

- b) Suppose you observe the yearly snowfall totals 38.6, 42.4, 57.5, 40.5, 51.7, 67.1, 33.4, 60.9, 64.1, 40.1, 40.7, and 6.4 inches. Enter these data into a vector `y` and compute the sample mean `ybar`.
- c) In this problem, the likelihood function is given by

$$L(\mu) \propto \exp\left(-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right),$$

where \bar{y} is the sample mean. Compute the likelihood on the list of values in `mu` and place the likelihood values in the vector `like`.

- d) One can compute the posterior probabilities for μ using the formula `post=prior*like/sum(prior*like)`

Compute the posterior probabilities of μ for this example.

- e) Using the function `discint`, find an 80% probability interval for μ .

6. Estimating a Poisson mean using a discrete prior (from Antleman (1996))

Suppose you own a trucking company with a large fleet of trucks. Breakdowns occur randomly in time and the number of breakdowns during an interval of t days is assumed to be Poisson distributed with mean $t\lambda$. The parameter λ is the daily breakdown rate. The possible values for λ are .5, 1, 1.5, 2, 2.5, and 3 with respective probabilities .1, .2, .3, .2, .15, and .05. If one observes y breakdowns, then the posterior probability of λ is proportional to

$$g(\lambda) \exp(-t\lambda)(t\lambda)^y,$$

where g is the prior probability.

- a) If 12 trucks break down in a six-day period, find the posterior probabilities for the different rate values.
- b) Find the probability that there are no breakdowns during the next week. Hint: If the rate is λ , the conditional probability of no breakdowns during a seven-day period is given by $\exp\{-7\lambda\}$. One can compute this predictive probability by multiplying a list of conditional probabilities by the posterior probabilities of λ and finding the sum of the products.