Introduction to Bayesian Thinking

Presented by Geoffrey S. Hubona

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Proportion of Heavy Sleepers

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

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

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

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.

Using a Discrete Prior

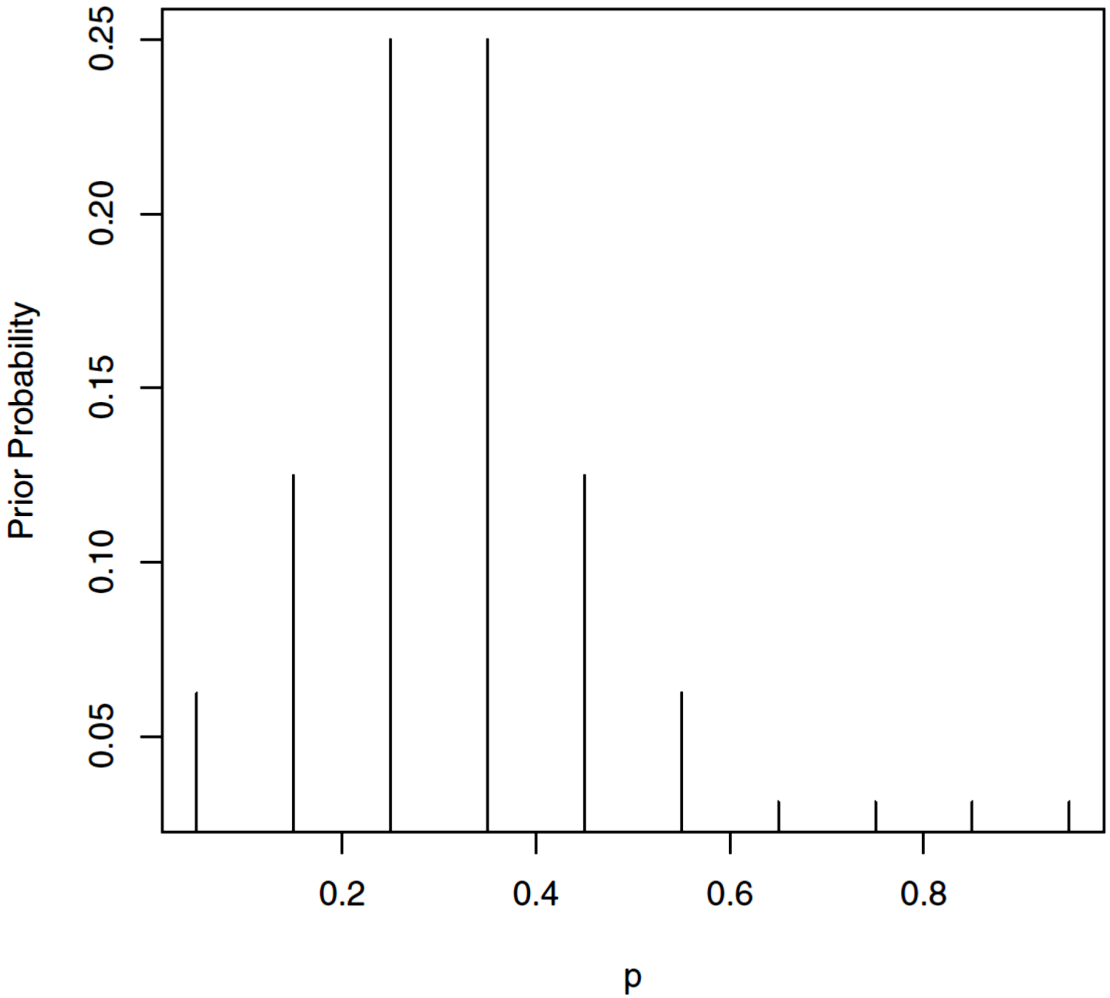

Fig. 2.1. A discrete prior distribution for a proportion p.

Using a Discrete Prior

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

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

(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.

Using a Discrete Prior

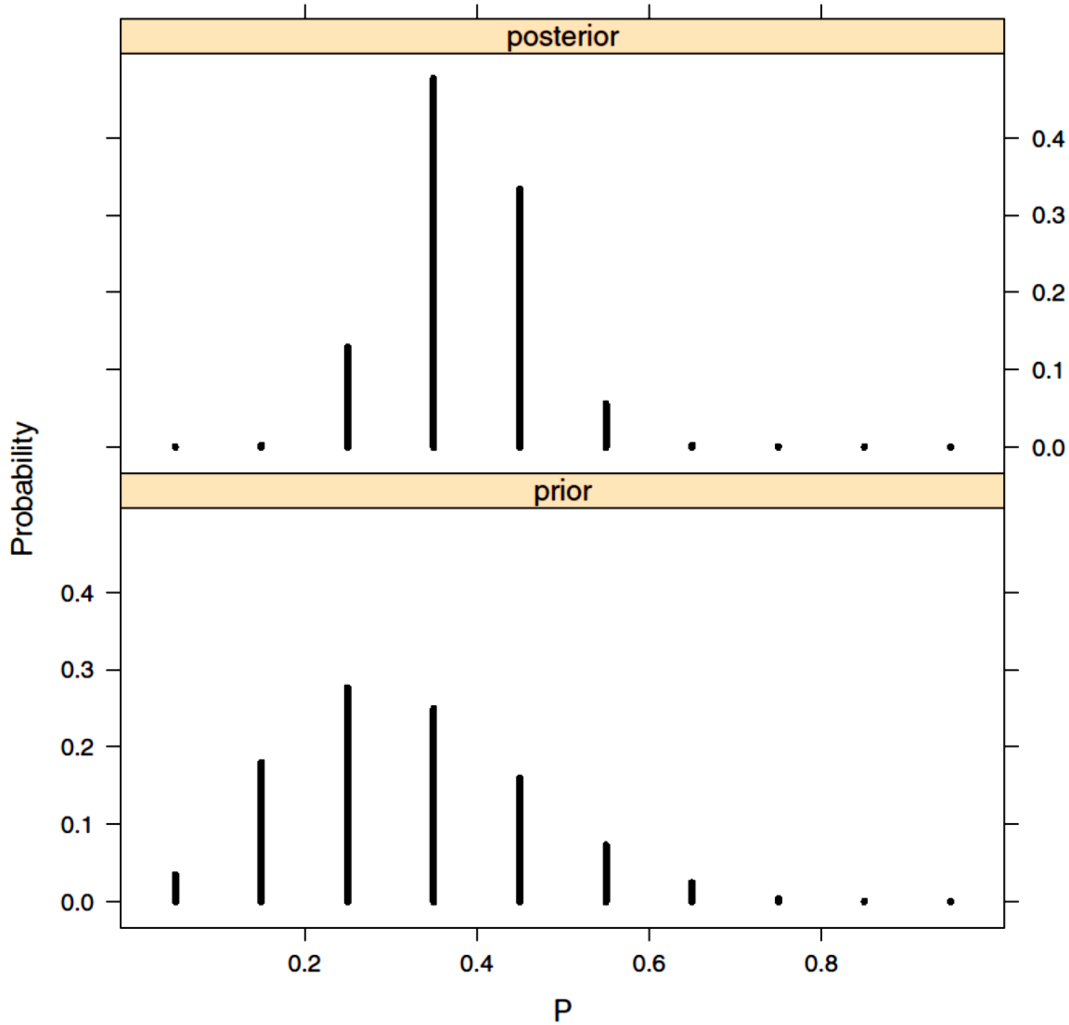

Fig. 2.2. Prior and posterior distributions for a proportion p using a discrete prior.

Using a Beta Prior

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

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

Using a Beta Prior

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.

Using a Beta Prior

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 $b + f$.

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

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.

Using a Beta Prior

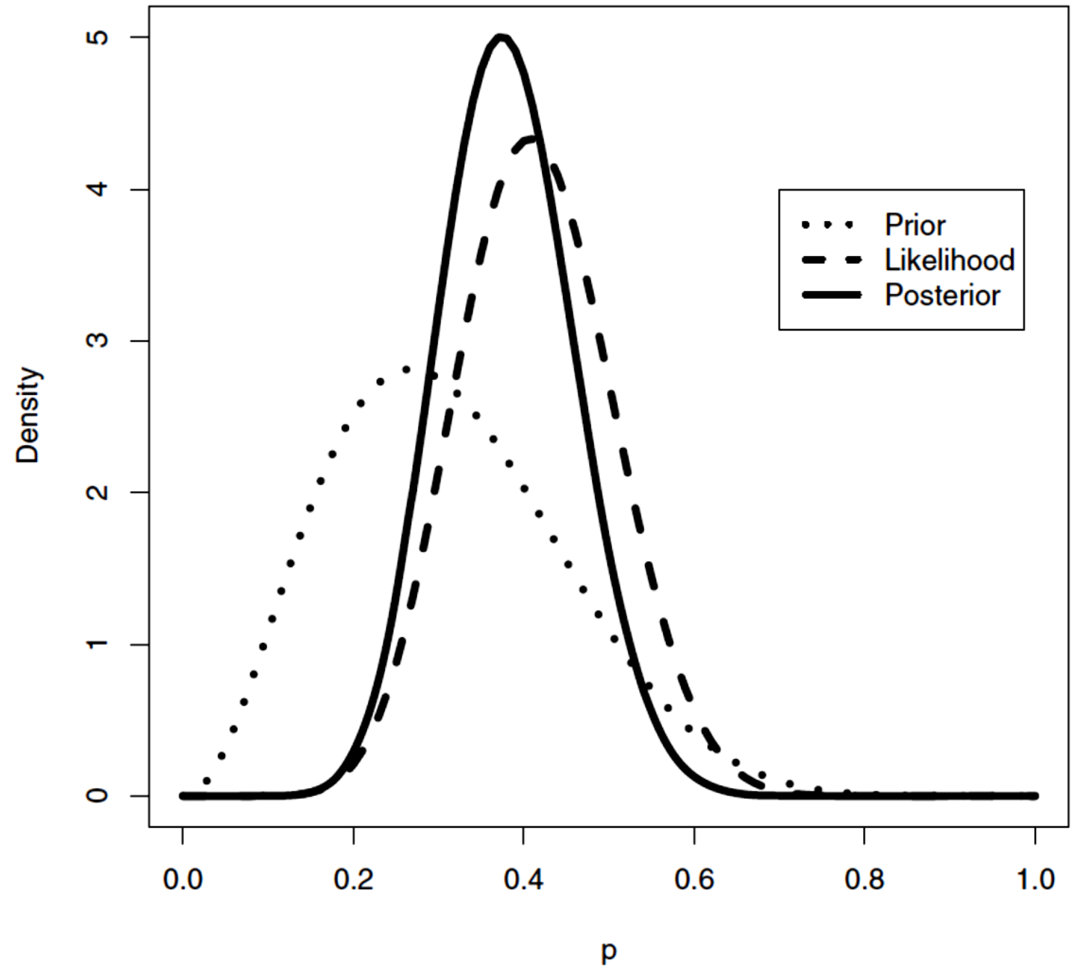

Fig. 2.3. The prior density $g(p)$, the likelihood function $L(p)$, and the posterior density $g(p|\text{data})$ for learning about a proportion p.

Using a Beta Prior

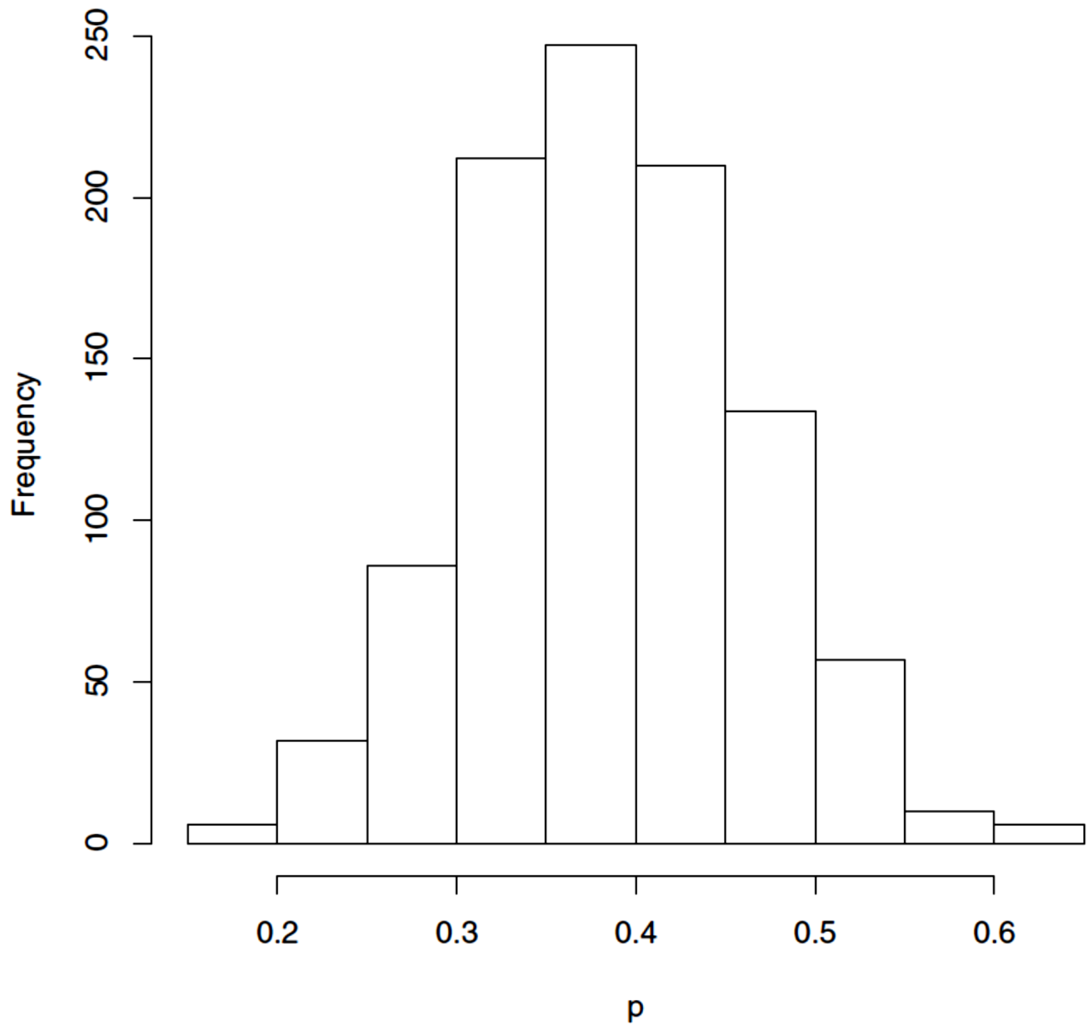

Fig. 2.4. A simulated sample from the beta posterior distribution of p.

Using a Histogram 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.

Using a Histogram Prior

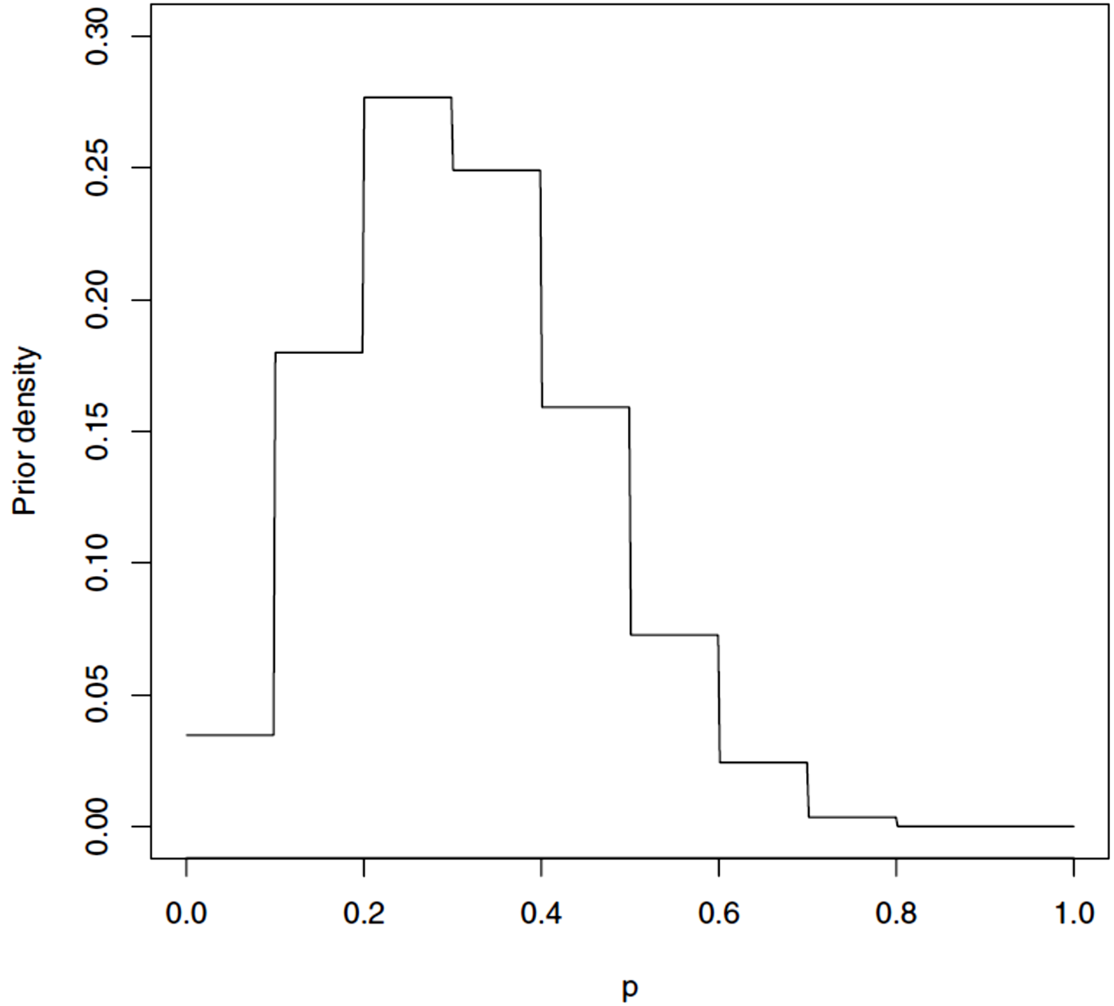

Fig. 2.5. A histogram prior for a proportion p.

Using a Histogram Prior

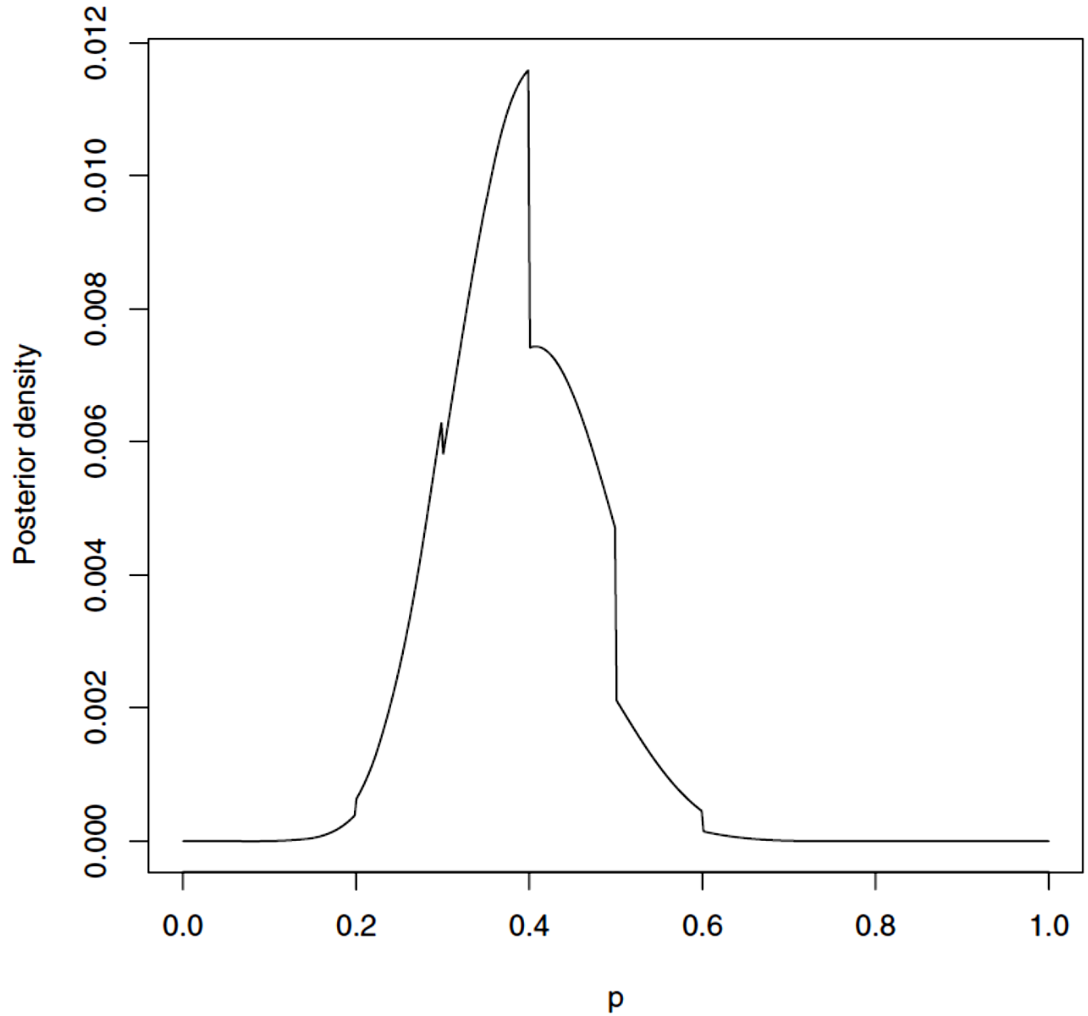

Fig. 2.6. The posterior density for a proportion using a histogram prior

Using a Histogram Prior

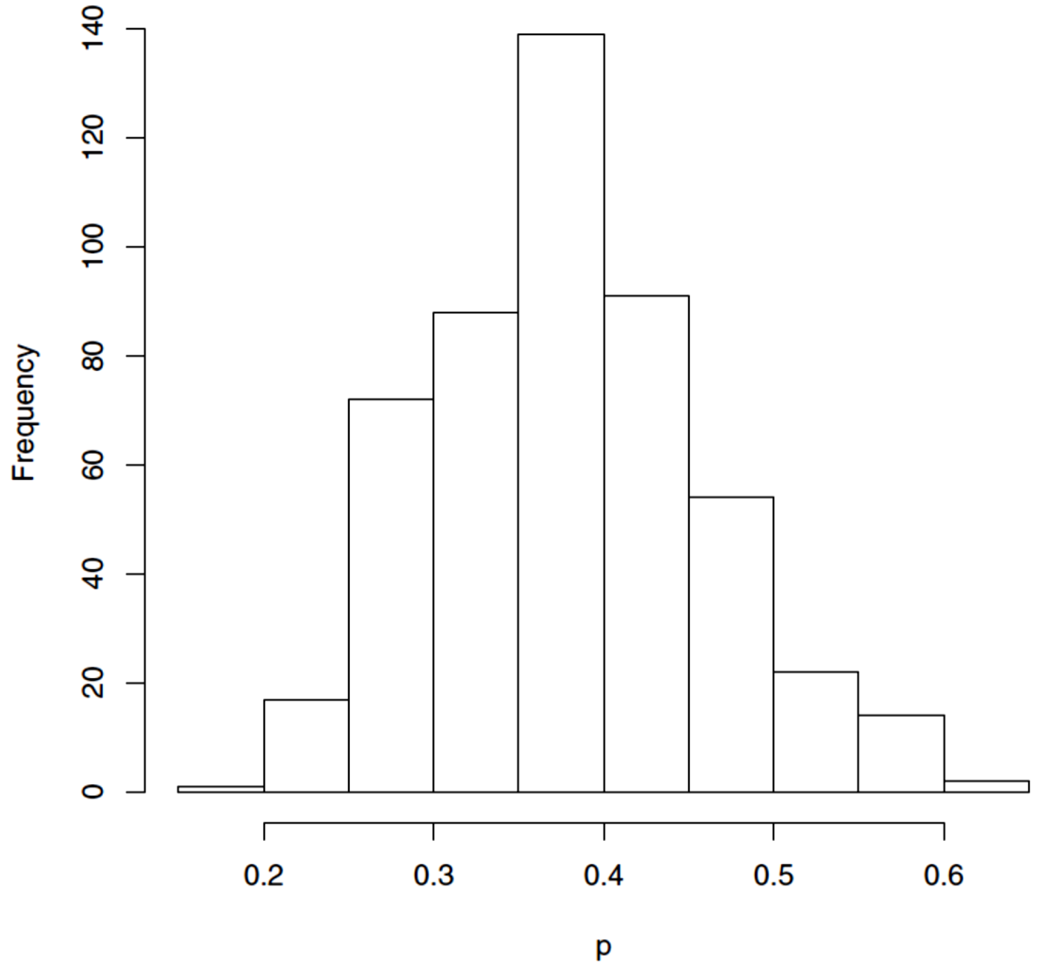

Fig. 2.7. A histogram of simulated draws from the posterior distribution of p with the use of a histogram prior.

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 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.

Prediction

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:

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

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

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

Prediction

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

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

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 beta(3.26, 7.19) prior used to reflect the person's beliefs about the proportion of heavy sleepers at the school.

Prediction

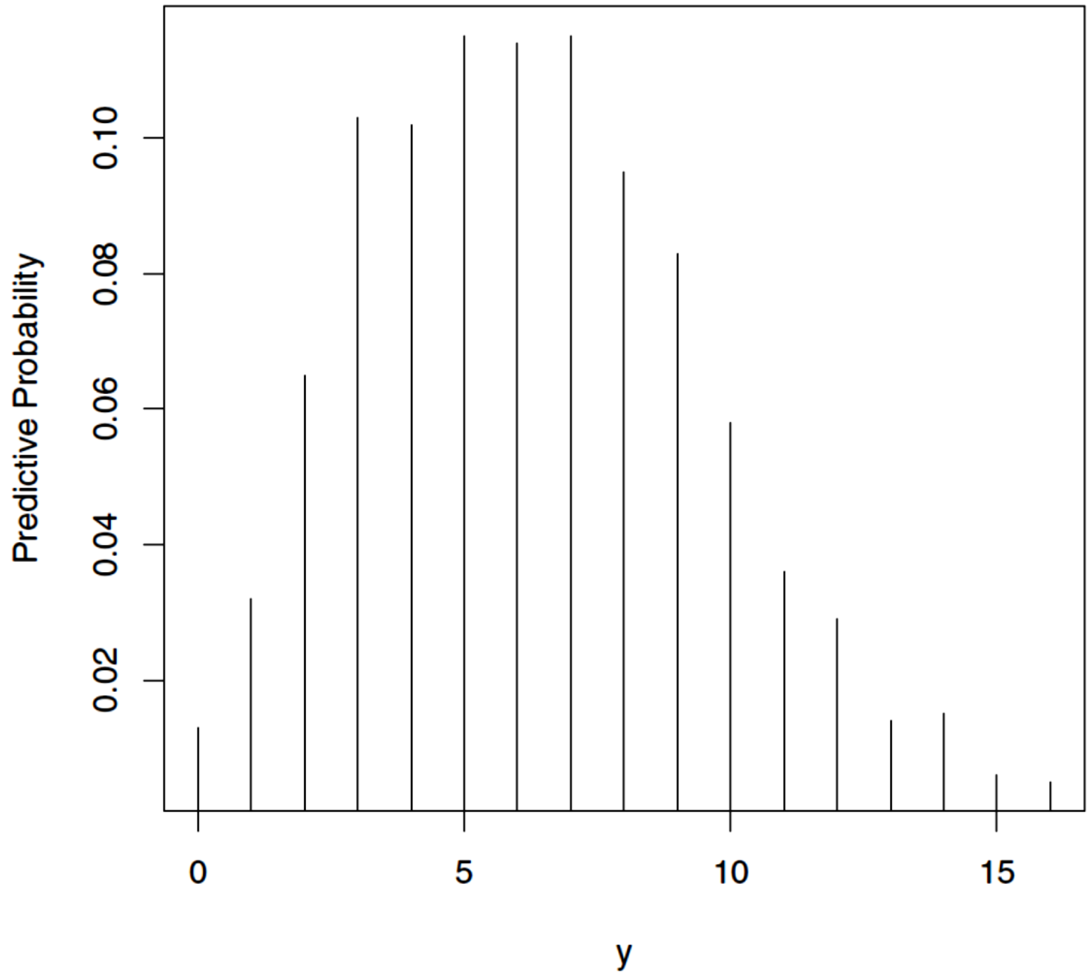

Fig. 2.8. A graph of the predictive probabilities of the number of sleepers $\tilde{y}$ in a future sample of size 20 when the proportion is assigned a beta(3.26, 7.19) prior.