Fig. 12.10 Density estimate of marginal posterior of the mean parameter μ for Madison writing example with negative binomial sampling.

```
> quantile(mu, c(0.05, 0.95))
      5%       95%
1.108965 1.371458
```

Comparing with the 90% interval estimate of μ obtained with Poisson sampling (1.119, 1.339), the negative binomial interval computed here is a bit wider. The longer interval is a consequence of the fact that, in negative binomial sampling, the overdispersion parameter a is unknown and the uncertainty in this value results in a longer interval estimate. (For a similar reason, the interval estimate for a mean from a normal population with unknown variance will be longer than the corresponding interval for a normal mean when the variance is known.)

Exercises

12.1 (Learning about students sleep habits). In Chapter 6, one is interested in the mean number of hours μ slept by students at a particular college. Suppose that a particular professor's prior beliefs about μ are represented by a normal curve with mean $\mu_0 = 6$ hours and variance $\tau_0^2 = 0.25$.

$$g(\mu) \propto \exp\left(-\frac{1}{2\tau_0^2}(\mu - \mu_0)^2\right).$$

a. Use the `curve` and `dnorm` functions to construct a plot of the prior density.
b. Use the `qnorm` function to find the quartiles of the prior density.
c. Find the probability (from the prior) that the mean amount of sleep exceeds 7 hours (Use the `pnorm` function.)

12.2 (Learning about students sleep habits, continued). Suppose the number of hours slept by a sample of n students, $y_1, ..., y_n$, represent a random sample from a normal density with unknown mean μ and known variance σ^2. The likelihood function of μ is given by

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

If we combine the normal prior density with this likelihood, it can be shown that the posterior density for μ also has the normal form with updated variance and mean parameters

$$\tau_1^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}, \mu_1 = \tau_1^2\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right).$$

a. For the sleeping data collected in Chapter 6, $n = 24$ students were sampled and the mean sleeping time was $\bar{y} = 7.688$. Assume that we know the sampling variance is given by $\sigma^2 = 2.0$. Use these values together with the prior mean and prior variance to compute the mean and variance of the posterior density of μ.
b. Using two applications of the `curve` function, plot the prior and posterior densities together using contrasting colors.
c. Use the `qnorm` function to construct a 90% posterior interval estimate of the mean sleeping time.
d. Find the posterior probability that the mean amount of sleep exceeds 7 hours.

12.3 (Waiting until a hit in baseball). In sports, fans are fascinated with the patterns of streaky behavior of athletes. In baseball, a batter wishes to get a "base hit"; otherwise he records an "out." Suppose one records the number of outs between consecutive hits (the spacing) for a particular baseball player. For example, if the player begins the season with the outcomes H, O, O, H, H, O, O, O, O, H, O, H, H, then the spacings are given by 0, 2, 4, 1, 0 (one starts counting before the first outcome). One observes the following spacings for the player Ian Kinsler for the 2008 baseball season:

```
0  2  0  4  1  0  2  0  1  0  0  1  1  3  1  0  0  0  1
6  0  9  0  4  1  9  1  0  3  4  5  5  1  0  2  4  0  4
0  3  2  1  0  1  3  7  0  3  1  2 14  4  0  1  6  1 10
1  2  0  1  0  4  5  0  7  3  1  2  1  2  1  2  2  4  3
3  1  1  2  1  2  7  0  3  1  2  2  2  2  0  3  4  1  1
0  0  1  1  1 11  2  2  1  3  1  0  1  2  1  1  1  0  0
2  0 10  1  2  2  1  1  3  1  1  0  0  1  0  1  0  1  1
```

```
 0   1   0   0   0   2   1   4   5   5   0   0   0   0   2   0   8   5   2
11   8   0   7   1   3   1
```

Let y denote the number of outs before the next base hit. A basic assumption is that y has a geometric(p) distribution with probability function

$$f(y|p) = p(1-p)^y, \; y = 0, 1, 2, \ldots$$

where p is the player's hitting probability, If we assume independence in spacings, then the likelihood function of p is given by

$$L(p) = \prod_y f(y|p) = p^n (1-p)^s,$$

where n is the sample size and $s = \sum y$ is the sum of the spacings.

a. Compute the values of n and s for Kinsler's data.
b. Use the `curve` function to graph the likelihood function for values of p between 0.2 and 0.5.
c. Based on the graph of the likelihood, which value of the hitting probability p is "most likely" given the data?

12.4 (Waiting until a hit, continued). Based on Ian Kinsler's performance in previous seasons, a baseball fan has some prior beliefs about Kinsler's hitting probability p. She believes that $P(p < 0.300) = 0.5$ and $P(p < 0.350) = 0.90$. This prior information can be matched to a beta density with shape parameters $a = 44$ and $b = 102$.

$$g(p) = \frac{1}{B(44, 102)} p^{44-1} (1-p)^{102-1}, \; 0 < p < 1.$$

If one multiplies this prior density with the likelihood function found in Exercise 12.3, the posterior density for p is given (up to a proportionality constant) by

$$g(p|\text{data}) \propto p^{n+44-1} (1-p)^{s+102-1}, \; 0 < p < 1,$$

where n is the sample size and $s = \sum y$ is the sum of the spacings. This is a beta density with shape parameters $a = n + 44$ and $b = s + 102$. (The values of n and s are found from the data in Exercise 12.3.)

a. Using the `curve` function, graph the posterior density for values of the hitting probability p between values 0.2 and 0.5.
b. Using the `qbeta` function, find the median of the posterior density of p.
c. Using the `qbeta` function, construct a 95% Bayesian interval estimate for p

12.5 (Waiting until a hit, continued). In Exercise 12.4, we saw that the posterior density for Ian Kinsler's hitting probability is a beta density with shape parameters $a = n + 44$ and $b = s + 102$. (The values of n and s are found from the data in Exercise 12.3.)

a. Using the `rbeta` function, simulate 1000 values from the posterior density of p.
b. Use the `hist` function on the simulated sample to display the posterior density.
c. Using the simulated draws, approximate the mean and standard deviation of the posterior density.
d. Using the simulated draws, construct a 95% Bayesian interval estimate. Compare the interval with the exact 95% interval estimate using the `qbeta` function.

12.6 (Waiting until a hit, continued). In Exercise 12.4, we saw that the posterior density for Ian Kinsler's hitting probability is a beta density with shape parameters $a = n + 44$ and $b = s + 102$. (The values of n and s are found from the data in Exercise 12.3.)

The function `metrop.hasting.rw` described in the chapter can be used to simulate a sample from the posterior density of the hitting probability p. The following function `betalogpost` will compute the logarithm of the beta density with shape parameters a and b.

```
betalogpost = function(p, a, b)
  dbeta(p, a, b, log=TRUE)
```

a. Use the function `metrop.hasting.rw` together with the function `betalog-post` to simulate from the posterior density using the Metropolis Hastings random walk algorithm. Use $p = 0.2$ as a starting value, take 1000 iterations, and use the scale constant $C = 0.1$. Construct a trace plot of the simulated values and find the acceptance rate of the algorithm. Compute the posterior mean of p from the simulated draws and compare the simulation estimate with the exact posterior mean $(n + 44)/(n + s + 44 + 102)$.
b. Rerun the random walk algorithm using the alternative scale constant values $C = 0.01$ and $C = 0.30$. In each case, construct a trace plot and compute the acceptance rate of the algorithm. Of the three choices for the scale constant C, are any of the values unsuitable? Explain.