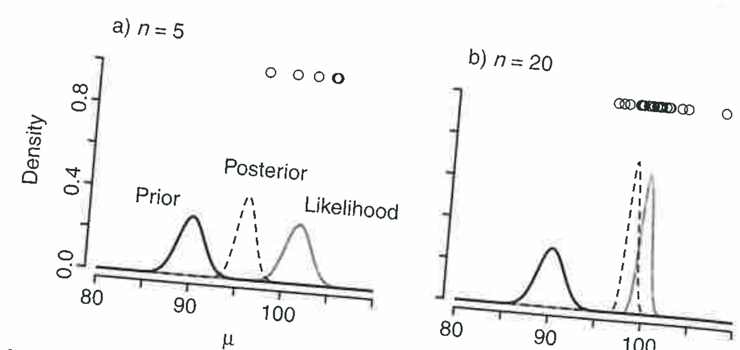

FIGURE 4.3. A Bayesian analysis of the mean of a normal distribution showing how the posterior balance the input of the prior and the likelihood. Sample values (x_i) are shown as dots above. Curves are plotted against values of parameter μ.

This equal contribution from prior and likelihood explains why the posterior in Figure 4.3a splits the difference between the two. In Figure 4.3b, the ratio is $n/\sigma^2 = 2$. This places more weight on the likelihood than on the prior, because $0.2 > 0.05$. The posterior in Figure 4.3b shows this stronger influence of data. If the sample size were $n = 100$, we would find no discernible effect of the prior.

At this point one might ask whether the density 4.11 is of much use. The parameter μ lies to the left of the bar, and y and σ^2 are to the right of the bar. In other words this is the posterior density for the mean μ conditioned on a known variance—σ^2 is written to the right of the vertical bar on the left-hand side to indicate that it is known. This is consistent with the stated objective of determining the posterior for the mean, when the variance is known. Usually σ^2 will also be unknown, in which case we need to estimate both mean and variance from the data. In other words, it would be more useful to know the joint distribution of the two unknown parameter values, $p(\mu,\sigma^2|y)$, and the marginal posterior of the mean that integrates the uncertainty in σ^2. When the density for the variance is inverse gamma (following section), this integral happens to be a Student's t density,

$$p(\mu|y) = \int p(\mu, \sigma^2|y)\,d\sigma^2$$

$$= t_{n-1}\left(\mu \middle| \bar{y}, \frac{\sigma^2}{n-1}\right)$$

Often such integrals are not available. It turns out that sampling-based approaches of Chapter 8 for simulating a posterior distribution rely on simple conditional relationships, such as that for the mean in Equation 4.11. So this is an important tool that will be especially useful when we get to computation. Before proceeding further, I consider calculations for the variance conditioned on a known mean.

4.2.4 The Variance of a Normal Distribution

The mean of a normal distribution might be known, and we wish to determine the posterior variance. A prior commonly used for the variance is inverse

gamma, because it is conjugate with the normal. This example of a prior inverse gamma density for the variance demonstrates this relationship. Here is the prior:

$$IG(\sigma^2|\alpha, \beta) \propto (\sigma^2)^{-(\alpha+1)}\exp\left[-\frac{\beta}{\sigma^2}\right]$$

Combined with the normal likelihood this yields

$$p(\sigma^2|y,\mu) \propto p(y|\mu,\sigma^2)p(\sigma^2)$$

$$= \overbrace{(\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]}^{likelihood} \times \overbrace{(\sigma^2)^{-(\alpha+1)}\exp\left[-\frac{\beta}{\sigma^2}\right]}^{prior}$$

Now collect coefficients containing σ^2 to obtain a new inverse gamma density,

$$= (\sigma^2)^{-(\alpha+n/2+1)}\exp\left[-\frac{1}{\sigma^2}\left(\beta + \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right)\right]$$

$$= IG\left(\sigma^2 \middle| \alpha + n/2, \beta + 1/2\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

Thus, conditioned on a specific value of the mean, the posterior variance is also inverse gamma. As before, the shape of the distribution depends on both the prior and the likelihood. The inverse gamma distribution is sharply peaked when its parameter values are large. Both parameter values are sums of a prior parameter and terms that increase with sample size. As seen in the previous example, the prior has impact when prior parameter values are small relative to the amount of data.

Example 4.2. Dispersal studies can involve a Gaussian dispersal kernel for the scatter of seeds about a parent plant (e.g., Clark et al. 1998). In this case, the mean of the distribution is known (at distance zero), and we seek to estimate the variance. A dispersal kernel is two-dimensional. For a normal dispersal kernel with no directional bias, we have

$$p(x,y|\sigma^2) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 + y_i^2)\right]$$

Note that the normalization constant for this bivariate density has units of distance^{-2}. To simplify, replace the Cartesian coordinates with distance from source,

$$r^2 = x^2 + y^2$$