CHAPTER 16

Introduction to Bayesian Methods for Inference

16.1 Introduction

We begin this chapter with an example that illustrates the concepts and an application of the Bayesian approach to inference making. Suppose that we are interested in estimating the proportion of responders to a new therapy for treating a disease that is serious and difficult to cure (such a disease is said to be virulent). If p denotes the probability that any single person with the disease responds to the treatment, the number of responders Y in a sample of size n might reasonably be assumed to have a binomial distribution with parameter p. In previous chapters, we have viewed the parameter p as having a fixed but unknown value and have discussed point estimators, interval estimators, and tests of hypotheses for this parameter. Before we even collect any data, our knowledge that the disease is virulent might lead us to believe that the value of p is likely to be relatively small, perhaps in the neighborhood of .25. How can we use this information in the process of making inferences about p?

One way to use this prior information about p is to utilize a Bayesian approach. In this approach, we model the *conditional* distribution of Y given p, $Y \mid p$, as binomial:

$$p(y \mid p) = \binom{n}{y} p^y q^{n-y}, \qquad y = 0, 1, 2, \ldots, n.$$

Uncertainty about the parameter p is handled by treating it as a random variable and, before observing any data, assigning a *prior* distribution to p. Because we know that $0 < p < 1$ and the beta density function has the interval (0, 1) as support, it is convenient to use a beta distribution as a prior for p. But which beta distribution should we use? Since the mean of a beta-distributed random variable with parameters α and β is $\mu = \alpha/(\alpha+\beta)$ and we thought p might be in the neighborhood of .25, we might choose to use a beta distribution with $\alpha = 1$ and $\beta = 3$ (and $\mu = .25$) as the prior for p. Thus, the density assigned to p is

$$g(p) = \frac{1}{3}(1-p)^2, \qquad 0 < p < 1.$$

Since we have specified the conditional distribution of $Y \mid p$ and the distribution of p, we have also specified the joint distribution of (Y, p) and can determine the marginal distribution of Y and the conditional distribution of $p \mid Y$. *After* observing $Y = y$, the *posterior* density of p given $Y = y$, $g^\star(p \mid y)$, can be determined. In the next section, we derive a general result that, in our virulent-disease example, implies that the posterior density of p given $Y = y$ is

$$g^\star(p \mid y) = \frac{\Gamma(n+4)}{\Gamma(y+1)\Gamma(n-y+3)} p^y (1-p)^{n-y+2}, \qquad 0 < p < 1.$$

Notice that the posterior density for $p \mid y$ is a beta density with $\alpha = y + 1$ and $\beta = n - y + 3$. This posterior density is the "updated" (by the data) density of p and is the basis for all Bayesian inferences regarding p. In the following sections, we describe the general Bayesian approach and specify how to use the posterior density to obtain estimates, credible intervals, and hypothesis tests for p and for parameters associated with other distributions.

16.2 Bayesian Priors, Posteriors, and Estimators

If $Y_1, Y_2, \ldots, Y_n$ denote the random variables associated with a sample of size n, we previously used the notation $L(y_1, y_2, \ldots, y_n \mid \theta)$ to denote the likelihood of the sample. In the discrete case, this function is defined to be the joint probability $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$, and in the continuous case, it is the joint density of $Y_1, Y_2, \ldots, Y_n$ evaluated at $y_1, y_2, \ldots, y_n$. The parameter θ is included among the arguments of $L(y_1, y_2, \ldots, y_n \mid \theta)$ to denote that this function depends explicitly on the value of some parameter θ. In the Bayesian approach, the unknown parameter θ is viewed to be a random variable with a probability distribution, called the *prior distribution* of θ. This prior distribution is specified before any data are collected and provides a theoretical description of information about θ that was available before any data were obtained. In our initial discussion, we will assume that the parameter θ has a continuous distribution with density $g(\theta)$ that has no unknown parameters.

Using the likelihood of the data and the prior on θ, it follows that the joint likelihood of $Y_1, Y_2, \ldots, Y_n, \theta$ is

$$f(y_1, y_2, \ldots, y_n, \theta) = L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)$$

and that the marginal density or mass function of $Y_1, Y_2, \ldots, Y_n$ is

$$m(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta.$$

Finally, the posterior density of $\theta \mid y_1, y_2, \ldots, y_n$ is

$$g^\star(\theta \mid y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta}.$$

The posterior density summarizes all of the pertinent information about the parameter θ by making use of the information contained in the prior for θ *and* the information in the data.

EXAMPLE **16.1** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Bernoulli distribution where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$ and assume that the prior distribution for p is beta (α, β). Find the posterior distribution for p.

Solution Since the Bernoulli probability function can be written as

$$p(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}, \qquad y_i = 0, 1,$$

the likelihood $L(y_1, y_2, \ldots, y_n \mid p)$ is

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid p) &= p(y_1, y_2, \ldots, y_n \mid p) \\ &= p^{y_1}(1-p)^{1-y_1} \times p^{y_2}(1-p)^{1-y_2} \times \cdots \times p^{y_n}(1-p)^{1-y_n} \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i}, \qquad y_i = 0, 1 \text{ and } 0 < p < 1. \end{aligned}$$

Thus,

$$\begin{aligned} f(y_1, y_2, \ldots, y_n, p) &= L(y_1, y_2, \ldots, y_n \mid p) \times g(p) \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1} \end{aligned}$$

and

$$\begin{aligned} m(y_1, y_2, \ldots, y_n) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}\, dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma\left(\sum y_i + \alpha\right)\Gamma\left(n - \sum y_i + \beta\right)}{\Gamma(n+\alpha+\beta)}. \end{aligned}$$

Finally, the posterior density of p is

$$\begin{aligned} g^\star(p \mid y_1, y_2, \ldots, y_n) &= \frac{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}}{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \dfrac{\Gamma\left(\sum y_i + \alpha\right)\Gamma\left(n - \sum y_i + \beta\right)}{\Gamma(n+\alpha+\beta)}}, \qquad 0 < p < 1 \\ &= \frac{\Gamma(n+\alpha+\beta)}{\Gamma\left(\sum y_i + \alpha\right)\Gamma\left(n - \sum y_i + \beta\right)} \times \\ &\qquad p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}, \qquad 0 < p < 1, \end{aligned}$$

a beta density with parameters $\alpha^\star = \sum y_i + \alpha$ and $\beta^\star = n - \sum y_i + \beta$. ■

Before we proceed, let's look at some of the implications of the result in Example 16.1. In the following example, we'll compare the prior and posterior distributions for some (for now) arbitrary choices of the parameters of the prior and the results of the experiment.

EXAMPLE **16.2** Consider the virulent-disease scenario and the results of Example 16.1. Compare the prior and posterior distributions of the Bernoulli parameter p (the proportion of responders to the new therapy) if we chose the values for α and β and observed the hypothetical data given below:

a $\alpha = 1\,,\ \beta = 3\,,\ n = 5\,,\ \sum y_i = 2.$
b $\alpha = 1\,,\ \beta = 3\,,\ n = 25\,,\ \sum y_i = 10.$
c $\alpha = 10\,,\ \beta = 30\,,\ n = 5\,,\ \sum y_i = 2.$
d $\alpha = 10\,,\ \beta = 30\,,\ n = 25\,,\ \sum y_i = 10.$

Solution Before we proceed, notice that both beta priors have mean

$$\mu = \frac{\alpha}{\alpha + \beta} = \frac{1}{1+3} = \frac{10}{10+30} = .25$$

and that both hypothetical samples result in the same value of the maximum likelihood estimates (MLEs) for p:

$$\hat{p} = \frac{1}{n}\sum y_i = \frac{2}{5} = \frac{10}{25} = .40.$$

As derived in Example 16.1, if $y_1,\ y_2, \ldots,\ y_n$ denote the values in a random sample from a Bernoulli distribution, where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$, and the prior distribution for p is beta (α, β), the posterior distribution for p is beta $(\alpha^* = \sum y_i + \alpha, \beta^* = n - \sum y_i + \beta)$. Therefore, for the choices in this example,

a when the prior is beta $(1, 3)$, $n = 5$, $\sum y_i = 2$, the posterior is beta with

$$\alpha^* = \sum y_i + \alpha = 2 + 1 = 3 \quad \text{and} \quad \beta^* = n - \sum y_i + \beta = 5 - 2 + 3 = 6.$$

b when the prior is beta $(1, 3)$, $n = 25$, $\sum y_i = 10$, the posterior is beta with

$$\alpha^* = 10 + 1 = 11 \quad \text{and} \quad \beta^* = 25 - 10 + 3 = 18.$$

c when the prior is beta $(10, 30)$, $n = 5$, $\sum y_i = 2$, the posterior is beta with

$$\alpha^* = 2 + 10 = 12 \quad \text{and} \quad \beta^* = 5 - 2 + 30 = 33.$$

d when the prior is beta $(10, 30)$, $n = 25$, $\sum y_i = 10$, the posterior is beta with

$$\alpha^* = 20 \quad \text{and} \quad \beta^* = 45.$$

Recall that the mean and variance of a beta $(\alpha,\ \beta)$ distributed random variable are

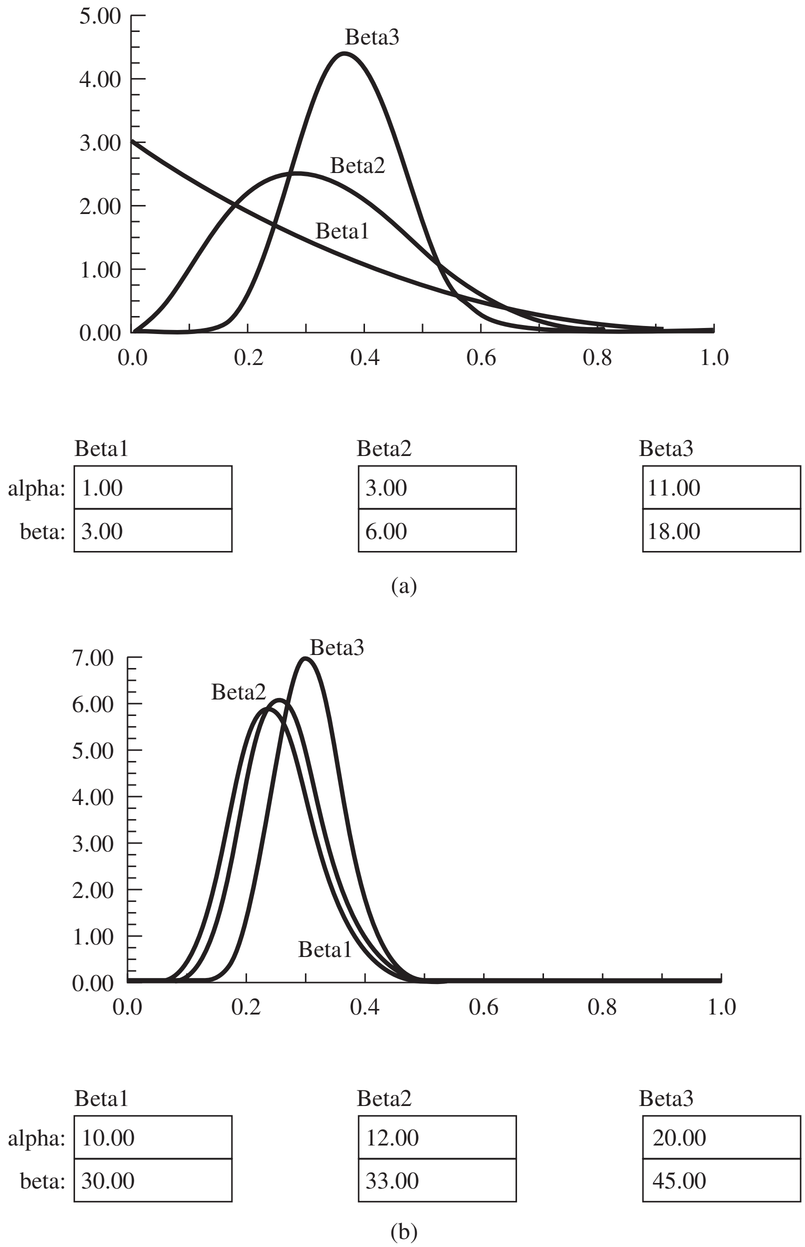
$$\mu = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

The parameters of the previous beta priors and posteriors, along with their means and variances are summarized Table 16.1. Figure 16.1(a) contains graphs of the beta distributions (priors and posteriors) associated with the beta prior with parameters

Table 16.1 Beta priors and posteriors for Example 16.2

Distribution	n	$\sum y_i$	Parameters of Beta Distribution	Mean	Variance
Prior	—	—	$\alpha = 1, \quad \beta = 3$	.2500	.0375
Posterior	5	2	$\alpha^* = 3, \quad \beta^* = 6$	.3333	.0222
Posterior	25	10	$\alpha^* = 11, \beta^* = 18$	.4074	.0078
Prior	—	—	$\alpha = 10, \quad \beta = 30$	.2500	.0046
Posterior	5	2	$\alpha^* = 12, \beta^* = 33$	.2667	.0043
Posterior	25	10	$\alpha^* = 20, \beta^* = 45$	.3077	.0032

FIGURE **16.1** Graphs of beta priors and posteriors in Example 16.2



$\alpha = 1$, $\beta = 3$. Graphs of the beta distributions associated with the beta (10, 30) prior are given in Figure 16.1(b). ■

In Examples 16.1 and 16.2, we obtained posterior densities that, like the prior, are beta densities but with altered (by the data) parameter values.

DEFINITION 16.1

Prior distributions that result in posterior distributions that are of the same functional form as the prior but with altered parameter values are called *conjugate priors*.

Any beta distribution is a conjugate prior distribution for a Bernoulli (or a binomial) distribution. When the prior is updated (using the data), the result is a beta posterior with altered parameter values. This is computationally convenient since we can determine the exact formula for the posterior and thereafter use previously developed properties of a familiar distribution. For the distributions that we use in this chapter, there are conjugate priors associated with the relevant parameters. These families of conjugate priors are often viewed to be broad enough to handle most practical situations. As a result, conjugate priors are often used in practice.

Since the posterior is a bona fide probability density function, some summary characteristic of this density provides an estimate for θ. For example, we could use the mean, the median, or the mode of the posterior density as our estimator. If we are interested in estimating some function of θ—say, $t(\theta)$—we will use the posterior expected value of $t(\theta)$ as our estimator for this function of θ.

DEFINITION 16.2

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with likelihood function $L(y_1, y_2, \ldots, y_n \mid \theta)$, and let θ have prior density $g(\theta)$. The posterior Bayes estimator for $t(\theta)$ is given by

$$\widehat{t(\theta)}_B = E(t(\theta) \mid Y_1, Y_2, \ldots, Y_n).$$

EXAMPLE 16.3 In Example 16.1, we found the posterior distribution of the Bernoulli parameter p based on a beta prior with parameters (α, β). Find the Bayes estimators for p and $p(1-p)$. [Recall that $p(1-p)$ is the variance of a Bernoulli random variable with parameter p].

Solution In Example 16.1, we found the posterior density of p to be a beta density with parameters $\alpha^\star = \sum y_i + \alpha$ and $\beta^\star = n - \sum y_i + \beta$:

$$g^\star(p \mid y_1, y_2, \ldots, y_n) = \frac{\Gamma(\alpha^\star + \beta^\star)}{\Gamma(\alpha^\star)\Gamma(\beta^\star)} p^{\alpha^\star - 1}(1-p)^{\beta^\star - 1}, \qquad 0 < p < 1.$$

The estimate for p is the posterior mean of p. From our previous study of the beta distribution, we know that

$$\begin{aligned}
\hat{p}_B &= E(p \mid y_1, y_2, \dots, y_n) \\
&= \frac{\alpha^\star}{\alpha^\star + \beta^\star} \\
&= \frac{\sum y_i + \alpha}{\sum y_i + \alpha + n - \sum y_i + \beta} = \frac{\sum y_i + \alpha}{n + \alpha + \beta}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
[p(\widehat{1-p})]_B &= E(p(1-p) \mid y_1, y_2, \dots, y_n) \\
&= \int_0^1 p(1-p)\frac{\Gamma(\alpha^\star + \beta^\star)}{\Gamma(\alpha^\star)\Gamma(\beta^\star)} p^{\alpha^\star - 1}(1-p)^{\beta^\star - 1}\, dp \\
&= \int_0^1 \frac{\Gamma(\alpha^\star + \beta^\star)}{\Gamma(\alpha^\star)\Gamma(\beta^\star)} p^{\alpha^\star}(1-p)^{\beta^\star}\, dp \\
&= \frac{\Gamma(\alpha^\star + \beta^\star)}{\Gamma(\alpha^\star)\Gamma(\beta^\star)} \times \frac{\Gamma(\alpha^\star + 1)\Gamma(\beta^\star + 1)}{\Gamma(\alpha^\star + \beta^\star + 2)} \\
&= \frac{\Gamma(\alpha^\star + \beta^\star)}{\Gamma(\alpha^\star)\Gamma(\beta^\star)} \times \frac{\alpha^\star\Gamma(\alpha^\star)\beta^\star\Gamma(\beta^\star)}{(\alpha^\star + \beta^\star + 1)(\alpha^\star + \beta^\star)\Gamma(\alpha^\star + \beta^\star)} \\
&= \frac{\alpha^\star\beta^\star}{(\alpha^\star + \beta^\star + 1)(\alpha^\star + \beta^\star)} \\
&= \frac{\left(\sum y_i + \alpha\right)\left(n - \sum y_i + \beta\right)}{(n + \alpha + \beta + 1)(n + \alpha + \beta)}.
\end{aligned}$$

So, the Bayes estimators for p and $p(1-p)$ are

$$\hat{p}_B = \frac{\sum Y_i + \alpha}{n + \alpha + \beta} \quad \text{and} \quad [p(\widehat{1-p})]_B = \frac{\left(\sum Y_i + \alpha\right)\left(n - \sum Y_i + \beta\right)}{(n + \alpha + \beta + 1)(n + \alpha + \beta)}.$$ ■

Further examination of the Bayes estimator for p given in Example 16.3 yields

$$\begin{aligned}
\hat{p}_B &= \frac{\sum Y_i + \alpha}{n + \alpha + \beta} \\
&= \left(\frac{n}{n + \alpha + \beta}\right)\left(\frac{\sum Y_i}{n}\right) + \left(\frac{\alpha + \beta}{n + \alpha + \beta}\right)\left(\frac{\alpha}{\alpha + \beta}\right) \\
&= \left(\frac{n}{n + \alpha + \beta}\right)\overline{Y} + \left(\frac{\alpha + \beta}{n + \alpha + \beta}\right)\left(\frac{\alpha}{\alpha + \beta}\right).
\end{aligned}$$

Thus, we see that the Bayes estimator for p is a weighted average of the sample mean, $\overline{Y}$ (the MLE for p) and the mean of the beta prior assigned to p. Notice that the prior mean of p is given less weight for larger sample sizes whereas the weight given to the sample mean increases for larger sample sizes. Also, since $E(\overline{Y}) = p$, it is easy to see that the Bayes estimator for p is *not* an unbiased estimator. Generally speaking, Bayes estimators are not unbiased.

Notice that the estimators obtained in Example 16.3 are both functions of the sufficient statistic $\sum Y_i$. This is no coincidence since a Bayes estimator is always a function of a sufficient statistic, a result that follows from the factorization criterion (see Theorem 9.4).

If U is a sufficient statistic for the parameter θ based on a random sample Y_1, $Y_2, \ldots, Y_n$, then

$$L(y_1, y_2, \ldots, y_n \mid \theta) = k(u, \theta) \times h(y_1, y_2, \ldots, y_n),$$

where $k(u, \theta)$ is a function only of u and θ and $h(y_1, y_2, \ldots, y_n)$ is not a function of θ. In addition (see Hogg, McKean, and Craig, 2005), the function $k(u, \theta)$ can (but need not) be chosen to be the probability mass or density function of the statistic U. In accord with the notation in this chapter, we write the conditional density of $U \mid \theta$ as $k(u \mid \theta)$. Then, because $h(y_1, y_2, \ldots, y_n)$ is not a function of θ,

$$\begin{aligned}
g^\star(\theta \mid y_1, y_2, \ldots, y_n) &= \frac{L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta} \\
&= \frac{k(u \mid \theta) \times h(y_1, y_2, \ldots, y_n) \times g(\theta)}{\int_{-\infty}^{\infty} k(u \mid \theta) \times h(y_1, y_2, \ldots, y_n) \times g(\theta)\, d\theta} \\
&= \frac{k(u \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} k(u \mid \theta)) \times g(\theta)\, d\theta}.
\end{aligned}$$

Therefore, in cases where the distribution of a sufficient statistic U is known, the posterior can be determined by using the conditional density of $U \mid \theta$. We illustrate with the following example.

EXAMPLE **16.4** Let $Y_1,\ Y_2, \ldots, Y_n$ denote a random sample from a normal population with unknown mean μ and *known* variance σ_o^2. The conjugate prior distribution for μ is a normal distribution with *known* mean η and *known* variance δ^2. Find the posterior distribution and the Bayes estimator for μ.

Solution Since $U = \sum Y_i$ is a sufficient statistic for μ and is known to have a normal distribution with mean $n\mu$ and variance $n\sigma_o^2$,

$$L(u \mid \mu) = \frac{1}{\sqrt{2\pi n\sigma_o^2}} \exp\left[\frac{1}{2n\sigma_o^2}(u - n\mu)^2\right], \qquad -\infty < u < \infty$$

and the joint density of U and μ is

$$\begin{aligned}
f(u, \mu) &= L(u \mid \mu) \times g(\mu) \\
&= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}} \exp\left[-\frac{1}{2n\sigma_o^2}(u - n\mu)^2 - \frac{1}{2\delta^2}(\mu - \eta)^2\right], \\
&\qquad -\infty < u < \infty,\ -\infty < \mu < \infty.
\end{aligned}$$

Let us look at the quantity in the above exponent:

$$
\begin{aligned}
&-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2\\
&\quad=-\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2(u-n\mu)^2+n\sigma_o^2(\mu-\eta)^2\right]\\
&\quad=-\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2u^2-2\delta^2un\mu+\delta^2n^2\mu^2+n\sigma_o^2\mu^2-2n\sigma_o^2\mu\eta+n\sigma_o^2\eta^2\right]\\
&\quad=-\frac{1}{2n\sigma_o^2\delta^2}\left[(n^2\delta^2+n\sigma_o^2)\mu^2-2(n\delta^2u+n\sigma_o^2\eta)\mu+\delta^2u^2+n\sigma_o^2\eta^2\right]\\
&\quad=-\frac{1}{2\sigma_o^2\delta^2}\left[(n\delta^2+\sigma_o^2)\mu^2-2(\delta^2u+\sigma_o^2\eta)\mu\right]-\frac{1}{2n\sigma_o^2\delta^2}(\delta^2u^2+n\sigma_o^2\eta^2)\\
&\quad=-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left[\mu^2-2\left(\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)\mu+\left(\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]\\
&\qquad\qquad-\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2u^2+n\sigma_o^2\eta^2-\frac{n(\delta^2u+\sigma_o^2\eta)^2}{n\delta^2+\sigma_o^2}\right].
\end{aligned}
$$

Finally, we obtain:

$$
\begin{aligned}
-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2 &= -\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\\
&\quad-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f(u,\mu) &= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2\right]\\
&= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]\\
&\qquad\times\exp\left[-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&m(u)=\\
&\frac{\exp\left[-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2\right]}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]d\mu\\
&=\frac{\exp\left[-\frac{1}{2(n^2\delta^2+n\sigma_o^2)}(u-n\eta)^2\right]}{\sqrt{2\pi n(n\delta^2+\sigma_o^2)}}\int_{-\infty}^{\infty}\frac{\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]}{\sqrt{\frac{2\pi\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}\,d\mu.
\end{aligned}
$$

Recognizing the above integral as that of a normal density function and hence equal to 1, we obtain that the marginal density function for U is normal with mean $n\eta$ and variance $(n^2\delta^2 + n\sigma_o^2)$. Further, the posterior density of μ given $U = u$ is

$$g^\star(\mu \mid u) = \frac{f(u, \mu)}{m(u)} = \frac{1}{\sqrt{\frac{2\pi\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}\exp\left[-\frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{\delta^2 u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right)^2\right],$$

$$-\infty < \mu < \infty,$$

a normal density with mean

$$\eta^\star = \left(\frac{\delta^2 u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right) \quad \text{and} \quad \text{variance } \delta^{\star 2} = \left(\frac{\sigma_o^2\delta^2}{n\delta^2 + \sigma_o^2}\right).$$

It follows that the Bayes estimator for μ is

$$\hat{\mu}_B = \frac{\delta^2 U + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2} = \frac{n\delta^2}{n\delta^2 + \sigma_o^2}\overline{Y} + \frac{\sigma_o^2}{n\delta^2 + \sigma_o^2}\eta.$$

Again, this Bayes estimator is a weighted average of the MLE, $\overline{Y}$, the sample mean, and the mean of the prior η. As the size of the sample n increases, the weight assigned to the sample mean $\overline{Y}$ increases whereas the weight assigned to the prior mean η decreases. ■

Exercises

16.1 Refer to the results of Example 16.2 given in Table 16.1.

- **a** Which of the two priors has the smaller variance?
- **b** Compare the means and variances of the two posteriors associated with the beta (1, 3) prior. Which of the posteriors has mean and variance that differ more from the mean and variance of the beta (1, 3) prior?
- **c** Answer the questions in parts (a) and (b) for the beta (10, 30) prior.
- **d** Are your answers to parts (a)–(c) supported by the graphs presented in Figure 16.1(a) and (b)?
- **e** Compare the posteriors based on $n = 5$ for the two priors. Which of the two posteriors has mean and variance that differs more from the mean and variance of the corresponding priors?

16.2 Define each of the following:

- **a** Prior distribution for a parameter θ
- **b** Posterior distribution for a parameter θ
- **c** Conjugate prior distribution
- **d** Bayes estimator for a function of θ, $t(\theta)$

16.3 **Applet Exercise** The applet *Binomial Revision* can be used to explore the impact of data and the prior on the posterior distribution of the Bernoulli parameter p. The demonstration at the top of the screen uses the beta prior with $\alpha = \beta = 1$.

a Click the button "Next Trial" to observe the result of taking a sample of size $n = 1$ from a Bernoulli population with $p = .4$. Did you observe a success or a failure? Does the posterior look different than the prior? Are the parameters of the posterior what you expected based on the theoretical results of Example 16.1?

b Click the button "Next Trial" once again to observe the result of taking a sample of total size $n = 2$ from a Bernoulli population with $p = .4$. How many successes and failures have you observed so far ? Does the posterior look different than the posterior that you obtained in part (a)? Are the parameters of the posterior what you expected based on the theoretical results of Example 16.1?

c Click the button "Next Trial" several times to observe the result of taking samples of larger sizes from a Bernoulli population with $p = .4$. Pay attention to the mean and variance of the posterior distributions that you obtain by taking successively larger samples. What do you observe about the values of the means of the posteriors? What do you observe about the standard deviations of posteriors based on larger sample sizes?

d On the initial demonstration on the applet, you were told that the true value of the Bernoulli parameter is $p = .4$. The mean of the beta prior with $\alpha = \beta = 1$ is .5. How many trials are necessary to obtain a posterior with mean close to .4, the true value of the Bernoulli parameter?

e Click on the button "50 Trials" to see the effect of the results of an additional 50 trials on the posterior. What do you observe about the shape of the posterior distributions based on a large number of trials?

16.4 **Applet Exercise** Scroll down to the section "Applet with Controls" on the applet *Binomial Revision*. Here, you can set the true value of the Bernoulli parameter p to any value $0 < p < 1$ (any value of "real" interest) and you can also choose any $\alpha > 0$ and $\beta > 0$ as the values of the parameters of the conjugate beta prior. What will happen if the true value of $p = .1$ and you choose a beta prior with mean 1/4? In Example 16.1, one such sets of values for α and β was illustrated: $\alpha = 1$, $\beta = 3$. Set up the applet to simulate sampling from a Bernoulli distribution with $p = .1$ and use the beta (1, 3) prior. (Be sure to press Enter after entering the appropriate values in the boxes.)

a Click the button "Next Trial" to observe the result of taking a sample of size $n = 1$ from a Bernoulli population with $p = .1$. Did you observe a success or a failure? Does the posterior look different than the prior?

b Click the button "Next Trial" once again to observe the result of taking a sample of total size $n = 2$ from a Bernoulli population with $p = .1$. How many successes and failures have you observed so far? Does the posterior look different than the posterior you obtained in part (a)?

c If you observed a success on either of the first two trials, click the "Reset" button and start over. Next, click the button "Next Trial" until you observe the first success. What happens to the shape of the posterior upon observation of the first success?

d In this demonstration, we assumed that the true value of the Bernoulli parameter is $p = .1$. The mean of the beta prior with $\alpha = 1$, $\beta = 3$ is .25. Click the button "Next Trial" until you obtain a posterior that has mean close to .1. How many trials are necessary?

16.5 Repeat the directions in Exercise 16.4, using a beta prior with $\alpha = 10$, $\beta = 30$. How does the number of trials necessary to obtain a posterior with mean close to .1 compare to the number you found in Exercise 16.4(d)?

16.6 Suppose that Y is a binomial random variable based on n trials and success probability p (this is the case for the virulent-disease example in Section 16.1). Use the conjugate beta prior with parameters α and β to derive the posterior distribution of $p \mid y$. Compare this posterior with that found in Example 16.1.

16.7 In Section 16.1 and Exercise 16.6, we considered an example where the number of responders to a treatment for a virulent disease in a sample of size n had a binomial distribution with parameter p and used a beta prior for p with parameters $\alpha = 1$ and $\beta = 3$.

a Find the Bayes estimator for $p =$ the proportion of those with the virulent disease who respond to the therapy.

b Derive the mean and variance of the Bayes estimator found in part (a).

16.8 Refer to Exercise 16.6. If Y is a binomial random variable based on n trials and success probability p and p has the conjugate beta prior with parameters $\alpha = 1$ and $\beta = 1$,

a determine the Bayes estimator for p, $\hat{p}_B$.

b what is another name for the beta distribution with $\alpha = 1$ and $\beta = 1$?

c find the mean square for error (MSE) of the Bayes estimator found in part (a). [*Hint*: Recall Exercise 8.17].

d For what values of p is the MSE of the Bayes estimator smaller than that of the unbiased estimator $\hat{p} = Y/n$?

16.9 Suppose that we conduct independent Bernoulli trials and record Y, the number of the trial on which the first success occurs. As discussed in Section 3.5, the random variable Y has a geometric distribution with success probability p. A beta distribution is again a conjugate prior for p.

a If we choose a beta prior with parameters α and β, show that the posterior distribution of $p \mid y$ is beta with parameters $\alpha^\star = \alpha + 1$ and $\beta^\star = \beta + y - 1$.

b Find the Bayes estimators for p and $p(1 - p)$.

16.10 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from an exponentially distributed population with density $f(y \mid \theta) = \theta e^{-\theta y}, 0 < y$. (*Note*: the mean of this population is $\mu = 1/\theta$.) Use the conjugate gamma (α, β) prior for θ to do the following.

a Show that the joint density of $Y_1, Y_2, \ldots, Y_n, \theta$ is

$$f(y_1, y_2, \ldots, y_n, \theta) = \frac{\theta^{n+\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp\left[-\theta \bigg/ \left(\frac{\beta}{\beta \sum y_i + 1}\right)\right].$$

b Show that the marginal density of $Y_1, Y_2, \ldots, Y_n$ is

$$m(y_1, y_2, \ldots, y_n) = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\beta^\alpha} \left(\frac{\beta}{\beta \sum y_i + 1}\right)^{\alpha+n}.$$

c Show that the posterior density for $\theta \mid (y_1, y_2, \ldots, y_n)$ is a gamma density with parameters $\alpha^\star = n + \alpha$ and $\beta^\star = \beta/\left(\beta \sum y_i + 1\right)$.

d Show that the Bayes estimator for $\mu = 1/\theta$ is

$$\hat{\mu}_B = \frac{\sum Y_i}{n + \alpha - 1} + \frac{1}{\beta(n + \alpha - 1)}.$$

[*Hint*: Recall Exercise 4.111(e).]

e Show that the Bayes estimator in part (d) can be written as a weighted average of $\overline{Y}$ and the prior mean for $1/\theta$. [*Hint*: Recall Exercise 4.111(e).]

f Show that the Bayes estimator in part (d) is a biased but consistent estimator for $\mu = 1/\theta$.

16.11 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Poisson-distributed population with mean λ. In this case, $U = \sum Y_i$ is a sufficient statistic for λ, and U has a Poisson distribution with mean $n\lambda$. Use the conjugate gamma (α, β) prior for λ to do the following.

a Show that the joint likelihood of U, λ is

$$L(u, \lambda) = \frac{n^u}{u!\beta^\alpha \Gamma(\alpha)} \lambda^{u+\alpha-1} \exp\left[-\lambda \Big/ \left(\frac{\beta}{n\beta + 1}\right)\right].$$

b Show that the marginal mass function of U is

$$m(u) = \frac{n^u \Gamma(u + \alpha)}{u!\beta^\alpha \Gamma(\alpha)} \left(\frac{\beta}{n\beta + 1}\right)^{u+\alpha}.$$

c Show that the posterior density for $\lambda \mid u$ is a gamma density with parameters $\alpha^\star = u + \alpha$ and $\beta^\star = \beta/(n\beta + 1)$.

d Show that the Bayes estimator for λ is

$$\hat{\lambda}_B = \frac{\left(\sum Y_i + \alpha\right)\beta}{n\beta + 1}.$$

e Show that the Bayes estimator in part (d) can be written as a weighted average of $\overline{Y}$ and the prior mean for λ.

f Show that the Bayes estimator in part (d) is a biased but consistent estimator for λ.

16.12 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a normal population with *known* mean μ_o and unknown variance $1/v$. In this case, $U = \sum (Y_i - \mu_o)^2$ is a sufficient statistic for v, and $W = vU$ has a χ^2 distribution with n degrees of freedom. Use the conjugate gamma (α, β) prior for v to do the following.

a Show that the joint density of U, v is

$$f(u, v) = \frac{u^{(n/2)-1} v^{(n/2)+\alpha-1}}{\Gamma(\alpha)\Gamma(n/2)\, \beta^\alpha 2^{(n/2)}} \exp\left[-v \Big/ \left(\frac{2\beta}{u\beta + 2}\right)\right].$$

b Show that the marginal density of U is

$$m(u) = \frac{u^{(n/2)-1}}{\Gamma(\alpha)\Gamma(n/2)\, \beta^\alpha 2^{(n/2)}} \left(\frac{2\beta}{u\beta + 2}\right)^{(n/2)+\alpha} \Gamma\left(\frac{n}{2} + \alpha\right).$$

c Show that the posterior density for $v \mid u$ is a gamma density with parameters $\alpha^\star = (n/2) + \alpha$ and $\beta^\star = 2\beta/(u\beta + 2)$.

d Show that the Bayes estimator for $\sigma^2 = 1/v$ is $\hat{\sigma}_B^2 = (U\beta + 2)/[\beta(n + 2\alpha - 2)]$. [*Hint*: Recall Exercise 4.111(e).]

e The MLE for σ^2 is U/n. Show that the Bayes estimator in part (d) can be written as a weighted average of the MLE and the prior mean of $1/v$. [*Hint*: Recall Exercise 4.111(e).]

16.3 Bayesian Credible Intervals

In previous sections, we have determined how to derive classical confidence intervals for various parameters of interest. In our previous approach, the parameter of interest θ had a *fixed but unknown value*. We constructed intervals by finding two *random variables* $\hat{\theta}_L$ and $\hat{\theta}_U$, the lower and upper confidence limits, such that $\hat{\theta}_L < \hat{\theta}_U$ and so that the probability that the *random interval* $\left(\hat{\theta}_L, \hat{\theta}_U\right)$ enclosed the *fixed* value θ

was equal to the prescribed confidence coefficient $1 - \alpha$. We also considered how to form one-sided confidence regions. The key realization in our pre-Bayesian work was that the *interval* was random and the parameter was fixed. In Example 8.11, we constructed a confidence interval for the mean of a normally distributed population with unknown variance using the formula

$$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) = \left\{\overline{Y} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right),\ \overline{Y} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right\}.$$

In this case, the upper and lower endpoints of the interval are clearly random variables. Upon obtaining data, calculating the *realized* values of the *sample mean* $\overline{y} = 2959$ and the sample variance $s = 39.1$ and using $n = 8$ and $t_{.025} = 2.365$, we determined that our *realized* confidence interval for the mean muzzle velocity for shells of the type considered is $(2926.3, 2991.7)$. This is a fixed interval that either contains the true mean muzzle velocity or does not. We say that the interval is a 95% confidence interval because *if independent and separate samples, each of size* $n = 8$ *were taken* and the resulting (different) intervals were determined, in the long run, 95% *of the intervals would contain the true mean*. The parameter is fixed, the endpoints of the interval are random, and different samples will yield different realized intervals.

In the Bayesian context, the parameter θ is a *random variable* with posterior density function $g^\star(\theta)$. If we consider the interval (a, b), the posterior probability that the random variable θ is in this interval is

$$P^\star(a \le \theta \le b) = \int_a^b g^\star(\theta)\, d\theta.$$

If the posterior probability $P^\star(a \le \theta \le b) = .90$, we say that (a, b) is a 90% *credible interval* for θ.

EXAMPLE **16.5** In Example 8.11, it was reasonable to assume that muzzle velocities were normally distributed with unknown mean μ. In that example, we assumed that the variance of muzzle velocities σ^2 was unknown. Assume now that we are interested in forming a Bayesian credible interval for μ and believe that there is a high probability that the muzzle velocities will be within 30 feet per second of their mean μ. Because a normally distributed population is such that approximately 95% of its values are within 2 standard deviations of its mean, it might be reasonable to assume that the underlying distribution of muzzle velocities is normally distributed with mean μ and variance σ_o^2 such that $2\sigma_o = 30$, that is with $\sigma_o^2 = 225$.

If, prior to observing any data, we believed that there was a high probability that μ was between 2700 and 2900, we might choose to use a conjugate normal prior for μ with mean η and variance δ^2 chosen such that $\eta - 2\delta = 2700$ and $\eta + 2\delta = 2900$, or $\eta = 2800$ and $\delta^2 = 50^2 = 2500$. Note that we have assumed considerably more knowledge of muzzle velocities than we did in Example 8.11 where we assumed only that muzzle velocities were normally distributed (with unknown variance). If we are comfortable with this additional structure, we now take our sample of size $n = 8$ and obtain the muzzle velocities given below:

3005	2925	2935	2965
2995	3005	2937	2905

Use the general form for the posterior density for $\mu \mid u$ developed in Example 16.4 to give a 95% credible interval for μ.

Solution This scenario is a special case of that dealt with in Example 16.4. In this application of that general result,

$$n = 8, \qquad u = \sum y_i = 23{,}672, \qquad \sigma_o^2 = 225, \qquad \eta = 2800, \qquad \delta^2 = 2500.$$

In Example 16.4, we determined that the posterior density of $\mu \mid u$ is a normal density with mean $\eta^\star$ and variance $\delta^{\star 2}$ given by

$$\eta^\star = \frac{\delta^2 u + \sigma_o^2 \eta}{n\delta^2 + \sigma_o^2} = \frac{(2500)(23672) + (225)(2800)}{8(2500) + 225} = 2957.23,$$

$$\delta^{\star 2} = \frac{\sigma_o^2 \delta^2}{n\delta^2 + \sigma_o^2} = \frac{(225)(2500)}{8(2500) + 225} = 27.81.$$

Finally, recall that any normally distributed random variable W with mean μ_W and variance σ_W^2 is such that

$$P(\mu_W - 1.96\,\sigma_W \le W \le \mu_W + 1.96\,\sigma_W) = .95.$$

It follows that a 95% credible interval for μ is

$$(\eta^\star - 1.96\,\delta^\star, \eta^\star + 1.96\,\delta^\star) = (2957.23 - 1.96\sqrt{27.81},\ 2957.23 + 1.96\sqrt{27.81})$$
$$= (2946.89,\ 2967.57).$$ ■

It is important to note that different individuals constructing credible intervals for μ using the data in Example 16.5 will obtain different intervals if they choose different values for any of the parameters η, δ^2, and σ_o^2. Nevertheless, for the choices used in Example 16.5, upon combining her prior knowledge with the information in the data, the analyst can say that the posterior probability is .95 that the (random) μ is in the (fixed) interval (2946.89, 2967.57).

EXAMPLE **16.6** In Exercise 16.10, it was stated that if $Y_1, Y_2, \ldots, Y_n$ denote a random sample from an exponentially distributed population with density $f(y \mid \theta) = \theta e^{-\theta y}$, $0 < y$, and the conjugate gamma prior (with parameters α and β) for θ was employed, then the posterior density for θ is a gamma density with parameters $\alpha^\star = n + \alpha$ and $\beta^\star = \beta/(\beta \sum y_i + 1)$. Assume that an analyst chose $\alpha = 3$ and $\beta = 5$ as appropriate parameter values for the prior and that a sample of size $n = 10$ yielded that $\sum y_i = 1.26$. Construct 90% credible intervals for θ and the mean of the population, $\mu = 1/\theta$.

Solution In this application of the general result given in Exercise 16.10,

$$n = 10, \qquad u = \sum y_i = 1.26, \qquad \alpha = 3, \qquad \beta = 5.$$

The resulting posterior density of θ is a gamma density with $\alpha^\star$ and $\beta^\star$ given by

$$\alpha^\star = n + \alpha = 10 + 3 = 13,$$

$$\beta^\star = \frac{\beta}{\beta \sum y_i + 1} = \frac{5}{5(1.26) + 1} = .685.$$

To complete our calculations, we need to find two values a and b such that

$$P^\star(a \leq \theta \leq b) = .90.$$

If we do so, a 90% credible interval for θ is (a, b). Further, because

$$a \leq \theta \leq b \quad \text{if and only if} \quad 1/b \leq 1/\theta \leq 1/a,$$

it follows that a 90% credible interval for $\mu = 1/\theta$ is $(1/b, 1/a)$.

Although we do not have a table giving probabilities associated with gamma-distributed random variables with different parameter values, such probabilities can be found using one of the applets accessible at www.thomsonedu.com/statistics/wackerly. *R*, *S*-Plus, and other statistical software can also be used to compute probabilities associated with gamma-distributed variables. Even so, there will be infinitely many choices for a and b such that $P^\star(a \leq \theta \leq b) = .90$. If we find values a and b such that

$$P^\star(\theta \geq a) = .95 \quad \text{and} \quad P^\star(\theta \geq b) = .05,$$

these values necessarily satisfy our initial requirement that $P^\star(a \leq \theta \leq b) = .90$.

In our present application, we determined that θ has a gamma posterior with parameters $\alpha^\star = 13$ and $\beta^\star = .685$. Using the applet *Gamma Probabilities and Quantiles* on the Thomson website, we determine that

$$P^\star(\theta \geq 5.2674) = .95 \quad \text{and} \quad P^\star(\theta \geq 13.3182) = .05.$$

Thus, for the data observed and the prior that we selected, (5.2674, 13.3182) is a 90% credible interval for θ whereas $[1/(13.3182), (1/5.2674)] = (.0751, .1898)$ is a 90% credible interval for $\mu = 1/\theta$.

The *R* (or *S*-Plus) command `qgamma(.05,13,1/.685)` also yields the value $a = 5.2674$ given above, whereas `qgamma(.95,13,1/.685)` gives $b = 13.3182$. ■

Exercises

16.13 **Applet Exercise** Activate the applet *Binomial Revision* and scroll down to the section labeled "Credible Interval." Change the value of the Bernoulli proportion to 0.45 and the parameters of the beta prior to $\alpha = 3$ and $\beta = 5$ and press Enter on your computer.

a What is the data-free credible interval for p based on the beta (3, 5) prior?

b Use the applet *Beta Probabilities and Quantiles* (accessible at the www.thomsonedu.com/statistics/wackerly) to calculate the prior probability that p is larger than the upper endpoint of the interval that you obtained in part (a). Also calculate the probability that p is smaller than the lower endpoint of the interval that you obtained in part (a).

c Based on your answers to part (b), what is the prior probability that p is in the interval that you obtained in part (a)? Do you agree that the interval obtained in part (a) is a 95% credible interval for p based on the beta (3, 5) prior?

d Click the button "Next Trial" once. Is the posterior based on the sample of size 1 different than the prior? How does the posterior differ from the prior?

e What is a 95% credible interval based on the prior and the result of your sample of size 1? Is it longer or shorter than the interval obtained (with no data) in part (a)?

f Click the button "Next Trial" once again. Compare the length of this interval (based on the results of a sample of size 2) to the intervals obtained in parts (a) and (e).

g Use the applet *Beta Probabilities and Quantiles* to calculate the posterior probability that p is larger than the upper endpoint of the interval that you obtained in part (f). Does the value of this posterior probability surprise you?

h Click the button "Next Trial" several times. Describe how the posterior is changed by additional data. What do you observe about the lengths of the credible intervals obtained using posteriors based on larger sample sizes?

16.14 **Applet Exercise** Refer to Exercise 16.13. Select a value for the true value of the Bernoulli proportion p and values for the parameters of the conjugate beta prior.

a Repeat Exercise 16.13(a)–(h), using the values you selected.

b Also click the button "50 Trials" a few times. Observe the values of the successive posterior standard deviations and the lengths of the successive credible intervals.

i What do you observe about the standard deviations of the successive posterior distributions?

ii Based on your answer to part (i), what effect do you expect to observe about the lengths of successive credible intervals?

iii Did the lengths of the successive credible intervals behave as you anticipated in part (ii)?

16.15 **Applet Exercise** In Exercise 16.7, we reconsidered our introductory example where the number of responders to a treatment for a virulent disease in a sample of size n had a binomial distribution with parameter p and used a beta prior for p with parameters $\alpha = 1$ and $\beta = 3$. We subsequently found that, upon observing $Y = y$ responders, the posterior density function for $p \mid y$ is a beta density with parameters $\alpha^\star = y + \alpha = y + 1$ and $\beta^\star = n - y + \beta = n - y + 3$. If we obtained a sample of size $n = 25$ that contained 4 people who responded to the new treatment, find a 95% credible interval for p. [Use the applet *Beta Probabilities and Quantiles* at www.thomsonedu.com/statistics/wackerly. Alternatively, if W is a beta-distributed random variable with parameters α and β, the R (or S-Plus) command `qbeta(p,α,β)` gives the value w such that $P(W \le w) = p$.]

16.16 **Applet Exercise** Repeat the instructions for Exercise 16.15, assuming a beta prior with parameters $\alpha = 1$ and $\beta = 1$ [a prior that is uniform on the interval (0, 1)]. (See the result of Exercise 16.8.) Compare this interval with the one obtained in Exercise 16.15.

16.17 **Applet Exercise** In Exercise 16.9, we used a beta prior with parameters α and β and found the posterior density for the parameter p associated with a geometric distribution. We determined that the posterior distribution of $p \mid y$ is beta with parameters $\alpha^\star = \alpha + 1$ and $\beta^\star = \beta + y - 1$. Suppose we used $\alpha = 10$ and $\beta = 5$ in our beta prior and observed the first success on trial 6. Determine an 80% credible interval for p.

16.18 **Applet Exercise** In Exercise 16.10, we found the posterior density for θ based on a sample of size n from an exponentially distributed population with mean $1/\theta$. Specifically, using the gamma density with parameters α and β as the prior for θ, we found that the posterior density for $\theta \mid (y_1, y_2, \ldots, y_n)$ is a gamma density with parameters $\alpha^\star = n + \alpha$ and $\beta^\star = \beta/\left(\beta \sum y_i + 1\right)$. Assuming that a sample of size $n = 15$ produced a sample such that $\sum y_i = 30.27$ and that the parameters of the gamma prior are $\alpha = 2.3$ and $\beta = 0.4$, use the applet *Gamma*

Probabilities and Quantiles to find 80% credible intervals for θ and $1/\theta$, the mean of the exponential population.

16.19 **Applet Exercise** In Exercise 16.11, we found the posterior density for λ, the mean of a Poisson-distributed population. Assuming a sample of size n and a conjugate gamma (α, β) prior for λ, we showed that the posterior density of $\lambda \mid \sum y_i$ is gamma with parameters $\alpha^\star = \sum y_i + \alpha$ and $\beta^\star = \beta/(n\beta + 1)$. If a sample of size $n = 25$ is such that $\sum y_i = 174$ and the prior parameters were $(\alpha = 2, \beta = 3)$, use the applet *Gamma Probabilities and Quantiles* to find a 95% credible interval for λ.

16.20 **Applet Exercise** In Exercise 16.12, we used a gamma (α, β) prior for v and a sample of size n from a normal population with known mean μ_o and variance $1/v$ to derive the posterior for v. Specifically, if $u = \sum(y_i - \mu_o)^2$, we determined the posterior of $v \mid u$ to be gamma with parameters $\alpha^\star = (n/2) + \alpha$ and $\beta^\star = 2\beta/(u\beta + 2)$. If we choose the parameters of the prior to be $(\alpha = 5, \beta = 2)$ and a sample of size $n = 8$ yields the value $u = .8579$, use the applet *Gamma Probabilities and Quantiles* to determine 90% credible intervals for v and $1/v$, the variance of the population from which the sample was obtained.

16.4 Bayesian Tests of Hypotheses

Tests of hypotheses can also be approached from a Bayesian perspective. As we have seen in previous sections, the Bayesian approach uses prior information about a parameter *and* information in the data about that parameter to obtain the posterior distribution. If, as in Section 10.11 where likelihood ratio tests were considered, we are interested in testing that the parameter θ lies in one of two sets of values, Ω_0 and Ω_a, we can use the posterior distribution of θ to calculate the posterior probability that θ is in each of these sets of values. When testing $H_0 : \theta \in \Omega_0$ versus $H_a : \theta \in \Omega_a$, one often-used approach is to compute the posterior probabilities $P^\star(\theta \in \Omega_0)$ and $P^\star(\theta \in \Omega_a)$ and accept the hypothesis with the higher posterior probability. That is, for testing $H_0 : \theta \in \Omega_0$ versus $H_a : \theta \in \Omega_a$,

$$\text{accept } H_0 \text{ if } P^\star(\theta \in \Omega_0) > P^\star(\theta \in \Omega_a),$$
$$\text{accept } H_a \text{ if } P^\star(\theta \in \Omega_a) > P^\star(\theta \in \Omega_0).$$

EXAMPLE **16.7** In Example 16.5, we obtained a 95% credible interval for the mean muzzle velocity associated with shells prepared with a reformulated gunpowder. We assumed that the associated muzzle velocities are normally distributed with mean μ and variance $\sigma_o^2 = 225$ and that a reasonable prior density for μ is normal with mean $\eta = 2800$ and variance $\delta^2 = 2500$. We then used the data

3005	2925	2935	2965
2995	3005	2937	2905

to obtain that the posterior density for μ is normal with mean $\eta^\star = 2957.23$ and standard deviation $\delta^\star = 5.274$. Conduct the Bayesian test for

$$H_0 : \mu \le 2950 \quad \text{versus} \quad H_a : \mu > 2950.$$

Solution In this case, if Z has a standard normal distribution,

$$\begin{aligned}P^\star(\theta \in \Omega_0) &= P^\star(\mu \le 2950)\\ &= P\left(Z \le \frac{2950 - \eta^\star}{\delta^\star}\right) = P\left(Z \le \frac{2950 - 2957.23}{5.274}\right)\\ &= P(Z \le -1.37) = .0951,\end{aligned}$$

and $P^\star(\theta \in \Omega_a) = P^\star(\mu > 2950) = 1 - P^\star(\mu \le 2950) = .9049$. Thus, we see that the posterior probability of H_a is much larger than the posterior probability of H_0 and our decision is to accept $H_a : \mu > 2950$. ■

Again, we note that if a different analyst uses the same data to conduct a Bayesian test for the same hypotheses but different values for any of η, δ^2, and σ_o^2, she will obtain posterior probabilities of the hypotheses that are different than those obtained in Example 16.7. Thus, different analysts with different choices of values for the prior parameters might reach different conclusions.

In the frequentist settings discussed in the previous chapters, the parameter θ has a fixed but unknown value, and any hypothesis is either true or false. If $\theta \in \Omega_0$, then the null hypothesis is certainly true (with probability 1), and the alternative is certainly false. If $\theta \in \Omega_a$, then the alternative hypothesis is certainly true (with probability 1), and the null is certainly false. The only way we could *know* whether or not $\theta \in \Omega_0$ is if *we knew the true value of* θ. If this were the case, conducting a test of hypotheses would be superfluous. For this reason, the frequentist makes no reference to the probabilities of the hypotheses but focuses on the probability of a type I error, α, and the power of the test, power $(\theta) = 1 - \beta(\theta)$. Conversely, the frequentist concepts of size and power are not of concern to an analyst using a Bayesian test.

EXAMPLE **16.8** In Example 16.6, we used a result given in Exercise 16.7 to obtain credible intervals for θ and the population mean μ based on $Y_1, Y_2, \dots, Y_n$, a random sample from an exponentially distributed population with density $f(y \mid \theta) = \theta e^{-\theta y}$, $0 < y$. Using a conjugate gamma prior for θ with parameters $\alpha = 3$ and $\beta = 5$, we obtained that the posterior density for θ is a gamma density with parameters $\alpha^\star = 13$ and $\beta^\star = .685$. Conduct the Bayesian test for

$$H_0 : \mu > .12 \quad \text{versus} \quad H_a : \mu \le .12.$$

Solution Since the mean of the exponential distribution is $\mu = 1/\theta$, the hypotheses are equivalent to

$$H_0 : \theta < 1/(.12) = 8.333 \quad \text{versus} \quad H_a : \theta \ge 8.333.$$

Because the posterior density for θ is a gamma density with parameters $\alpha^\star = 13$ and $\beta^\star = .685$,

$$P^\star(\theta \in \Omega_0) = P^\star(\theta < 8.333) \quad \text{and} \quad P^\star(\theta \in \Omega_a) = P^\star(\theta \ge 8.333).$$

In our present application, we determined that θ has a gamma posterior with parameters $\alpha^\star = 13$ and $\beta^\star = .685$. Using the applet *Gamma Probabilities and Quantiles*,

$$P^\star(\theta \in \Omega_a) = P^\star(\theta \geq 8.333) = 0.5570,$$

and

$$P^\star(\theta \in \Omega_0) = P^\star(\theta < 8.333) = 1 - P^\star(\theta \geq 8.333) = 0.4430.$$

In this case, the posterior probability of H_a is somewhat larger than the posterior probability of H_0. It is up to the analyst to decide whether the probabilities are sufficiently different to merit the decision to accept $H_a : \mu \leq .12$.

If you prefer to use R or S-Plus to compute the posterior probabilities of the hypotheses, `pgamma(8.333,13,1/.685)` yields $P^\star(\theta \in \Omega_0) = P^\star(\theta < 8.333)$ and $P^\star(\theta \in \Omega_a) = P^\star(\theta \geq 8.333) = 1 - P^\star(\theta \in \Omega_0)$. ■

Exercises

16.21 **Applet Exercise** In Exercise 16.15, we determined that the posterior density for p, the proportion of responders to the new treatment for a virulent disease, is a beta density with parameters $\alpha^\star = 5$ and $\beta^\star = 24$. What is the conclusion of a Bayesian test for $H_0 : p < .3$ versus $H_a : p \geq .3$? [Use the applet *Beta Probabilities and Quantiles* at www.thomsonedu.com/statistics/wackerly. Alternatively, if W is a beta-distributed random variable with parameters α and β, the R or S-Plus command `pbeta(w,`α`,`β`)` gives $P(W \leq w)$.]

16.22 **Applet Exercise** Exercise 16.16 used different prior parameters but the same data to determine that the posterior density for p, the proportion of responders to the new treatment for a virulent disease, is a beta density with parameters $\alpha^\star = 5$ and $\beta^\star = 22$. What is the conclusion of a Bayesian test for $H_0 : p < .3$ versus $H_a : p \geq .3$? Compare your conclusion to the one obtained in Exercise 16.21.

16.23 **Applet Exercise** In Exercise 16.17, we obtained a beta posterior with parameters $\alpha^\star = 11$ and $\beta^\star = 10$ for the parameter p associated with a geometric distribution. What is the conclusion of a Bayesian test for $H_0 : p < .4$ versus $H_a : p \geq .4$?

16.24 **Applet Exercise** In Exercise 16.18, we found the posterior density for θ to be a gamma density with parameters $\alpha^\star = 17.3$ and $\beta^\star = .0305$. Because the mean of the underlying exponential population is $\mu = 1/\theta$, testing the hypotheses $H_0 : \mu < 2$ versus $H_a : \mu \geq 2$ is equivalent to testing $H_0 : \theta > .5$ versus $H_a : \theta \leq .5$. What is the conclusion of a Bayesian test for these hypotheses?

16.25 **Applet Exercise** In Exercise 16.19, we found the posterior density for λ, the mean of a Poisson-distributed population, to be a gamma density with parameters $\alpha^\star = 176$ and $\beta^\star = .0395$. What is the conclusion of a Bayesian test for $H_0 : \lambda > 6$ versus $H_a : \lambda \leq 6$?

16.26 **Applet Exercise** In Exercise 16.20, we determined the posterior of $v \mid u$ to be a gamma density with parameters $\alpha^\star = 9$ and $\beta^\star = 1.0765$. Recall that $v = 1/\sigma^2$, where σ^2 is the variance of the underlying population that is normally distributed with known mean μ_o. Testing the hypotheses $H_0 : \sigma^2 > 0.1$ versus $H_a : \sigma^2 \leq 0.1$ is equivalent to testing $H_0 : v < 10$ versus $H_a : v \geq 10$. What is the conclusion of a Bayesian test for these hypotheses?

16.5 Summary and Additional Comments

As we have seen in the previous sections, the key to Bayesian inferential methods (finding estimators, credible intervals, or implementing tests of hypotheses) is finding the posterior distribution of the parameter θ. Especially when there are little data, this posterior is heavily dependent on the prior and the underlying distribution of the population from which the sample is taken. We have focused on the use of conjugate priors because of the resulting simplicity of finding the requisite posterior distribution of the parameter of interest. Of course, conjugate priors are not the only priors that can be used, but they do have the advantage of resulting in easy computations. This does not mean that a conjugate prior is necessarily the correct choice for the prior. Even if we correctly select the family from which the prior is taken (we have made repeated use of beta and gamma priors), there remains the difficulty of selecting the appropriate values associated with the parameters of the prior. We have seen, however, that the choice of the parameter values for the prior has decreasing impact for larger sample sizes.

It is probably appropriate to make a few more comments about selecting values of the parameters of the prior density. If we use a normal prior with mean ν and variance δ^2 and think that the population parameter is likely (unlikely) to be close to ν, we would use a relatively small (large) value for δ^2. When using a beta prior with parameters α and β for a parameter that we thought had value close to c, we might select α and β such that the mean of the prior, $\alpha/(\alpha+\beta)$, equals c and the variance of the prior, $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, is small. In the introductory example, we used a beta prior with $\alpha = 1$ and $\beta = 3$ because we thought that about 25% of those given the new treatment would favorably respond. The mean and standard deviation of the posterior are, respectively, .25 and .1936. Note that these are not the only choices for α and β that give .25 as the mean of the prior. In general, if $\alpha/(\alpha+\beta) = c$, then for any $k > 0$, $\alpha' = k\alpha$ and $\beta' = k\beta$ also satisfy $\alpha'/(\alpha'+\beta') = c$. However, for a beta density with parameters $\alpha' = k\alpha$ and $\beta' = k\beta$, the variance of the prior is $\alpha'\beta'[(\alpha'+\beta')^2(\alpha'+\beta'+1)] = \alpha\beta/[(\alpha+\beta)^2(k\alpha+k\beta+1)]$. Therefore, if our initial choice of α and β give an appropriate value for the mean of the prior but we prefer a smaller variance, we can achieve this by selecting some $k > 1$ and using $\alpha' = k\alpha$ and $\beta' = k\beta$ as the prior parameters. Conversely, choosing some $k < 1$ and using $\alpha' = k\alpha$ and $\beta' = k\beta$ as the prior parameters gives the same prior mean but larger prior variance. Hence, a more vague prior results from choosing small values of α and β that are such that $\alpha/(\alpha+\beta) = c$, the desired prior mean.

One of the steps in determining the prior is to determine the marginal distribution of the data. For continuous priors, this is accomplished by integrating the joint likelihood of the data and the parameter over the region of support for the prior. In our previous work, we denoted the resulting marginal mass or density function for the random variables $Y_1, Y_2, \ldots, Y_n$ in a sample of size n as $m(y_1, y_2, \ldots, y_n)$ or as $m(u)$ if U is a sufficient statistic for θ. This marginal mass or density function is called the *predictive* mass or density function of the data. We have explicitly given these predictive distributions in all of our applications. This is because, to paraphrase Berger (1985, p. 95), interest in the predictive distribution centers on the fact that this is the distribution according to which the data will actually occur. As discussed in Box

(1980, pp. 385–386), potential evidence of inappropriate model selection is provided by the predictive distribution of the data, not the posterior distribution for the parameter. Some expert Bayesian analysts choose to model the predictive distribution directly and select the prior that leads to the requisite predictive distribution. The Reverend Thomas Bayes (1784) used a uniform (0, 1) prior for the Bernoulli (or binomial) parameter *p because* this prior leads to the predictive distribution that he thought to be most appropriate. Additional comments relevant to the choice of some prior parameters can be found in Kepner and Wackerly (2002).

The preceding paragraph notwithstanding, it is true that there is a shortcut to finding the all-important posterior density for θ. As previously indicated, if $L(y_1, y_2, \ldots, y_n \mid \theta)$ is the conditional likelihood of the data and θ has continuous prior density $g(\theta)$, then the posterior density of θ is

$$g^\star(\theta \mid y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta}.$$

Notice that the denominator on the right hand side of the expression depends on $y_1, y_2, \ldots, y_n$, but *does not* depend on θ. (Definite integration with respect to θ produces a result that is free of θ.) Realizing that, with respect to θ, the denominator is a constant, we can write

$$g^\star(\theta \mid y_1, y_2, \ldots, y_n) = c(y_1, y_2, \ldots, y_n) L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta),$$

where

$$c(y_1, y_2, \ldots, y_n) = \frac{1}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta}$$

does not depend on θ. Further, notice that, because the posterior density is a bona fide density function, the quantity $c(y_1, y_2, \ldots, y_n)$ *must be* such that

$$\int_{-\infty}^{\infty} g^\star(\theta \mid y_1, y_2, \ldots, y_n)\, d\theta$$
$$= c(y_1, y_2, \ldots, y_n) \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta)\, d\theta = 1.$$

Finally, we see that the posterior density is *proportional to* the product of the conditional likelihood of the data and the prior density for θ:

$$g^\star(\theta \mid y_1, y_2, \ldots, y_n) \propto L(y_1, y_2, \ldots, y_n \mid \theta) \times g(\theta),$$

where the proportionally constant is chosen so that the integral of the posterior density function is 1. We illustrate by reconsidering Example 16.1.

EXAMPLE **16.9** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Bernoulli distribution where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$ and assume that the prior distribution for p is beta (α, β). Find the posterior distribution for p.

Solution As before,

$$
\begin{aligned}
L(y_1,\ y_2, \ldots, y_n \mid p) g(p) &= p(y_1,\ y_2, \ldots, y_n \mid p) g(p) \\
&= p^{\sum y_i}(1-p)^{n-\sum y_i}\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\right], \\
g^*(p \mid y_1,\ y_2, \ldots, y_n, p) &\propto p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}.
\end{aligned}
$$

From the above, we recognize that the resultant posterior for p must be beta with parameters $\alpha^\star = \sum y_i + \alpha$ and $\beta^\star = n - \sum y_i + \beta$. ■

What was the advantage of finding the previous posterior using this "proportionality" argument? Considerably less work! Disadvantage? We never exhibited the predictive mass function for the data and lost the opportunity to critique the Bayesian model.

Priors other than conjugate priors could well be more appropriate in specific applications. The posterior is found using the same procedure given in Section 16.2, but we might obtain a posterior distribution with which we are unfamiliar. Finding the mean of the posterior, credible intervals, and the probabilities of relevant hypotheses could be more problematic. For the examples in the previous sections, we obtained posteriors with which we were well acquainted. Posterior means were easy to find because we had already determined properties of normal, beta- and gamma-distributed random variables. Additionally, tables for these posteriors were readily available (in the appendix or easily accessed with many software packages). There is an ever-emerging set of computer procedures in which the posterior is determined based on user input of the likelihood function for the data and the prior for the parameter. Once the posterior is obtained via use of the software, this posterior is used exactly as previously described.

Bayes estimators can be evaluated using classical frequentist criteria. We have already seen that Bayes estimators are biased. However, they are usually consistent and, depending on the criteria used, can be superior to the corresponding frequentist estimators. In Exercise 16.8, you determined that the MSE of the Bayes estimator was sometimes smaller than the MSE of the unbiased MLE. Further, the influence of the choice of the prior parameter values decreases as the size of the sample increases.

In Example 8.11, we determined that the *realized* frequentist confidence interval for the mean of a normally distributed population was (2926.3, 2991.7). Using the frequentist perspective, the true population mean is fixed but unknown. As a result, this *realized* interval either captures the true value of μ or it does not. We said that this interval was a 95% confidence interval because the procedure (formula) used to produce it yields intervals that *do capture* the fixed mean about 95% of the time if samples of size 8 are *repeatedly and independently taken* and used to construct many intervals. If 100 samples of size 8 are taken and used to produce (different) realized confidence intervals, we expect approximately 95 of them to capture the parameter. We do not know which of the 100 intervals capture the unknown fixed mean. The *same data* was used in Example 16.5 to obtain (2946.89, 2967.57) as a 95% *credible* interval for μ, now viewed as a random variable. From the Bayesian perspective, it makes full sense to state that the posterior probability is .95 that the (random) mean is in this (fixed) interval.

The goodness of classical hypothesis tests is measured by α and β, the probabilities of type I and type II errors, respectively. If tests with $\alpha = .05$ are repeatedly (using different, independently selected samples) implemented, then when H_0 is true, H_0 is rejected 5% of the time. If H_0 is really true and 100 samples of the same size are independently taken, we expect to reject the (true) null hypothesis about five times. It makes no sense to even try to compute the probabilities of the hypotheses. From the Bayesian perspective, the parameter of interest is a random variable with posterior distribution derived by the analyst. Computing the posterior probabilities for each of the hypotheses is completely appropriate and is the basis for the decision in a Bayesian test.

Which is the better approach, Bayesian or frequentist? It is impossible to provide a universal answer to this question. In some applications, the Bayesian approach will be superior; in others, the frequentist approach is better.

References and Further Readings

Bayes, T. 1764. "An Essay Towards Solving a Problem in the Doctrine of Chances," *Phil. Trans. Roy. Soc.* 53, 370–418.

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis,* 2d ed. New York: Springer-Verlag.

Box, G. E. P. 1980. "Sampling and Bayes' Inference in Scientific Modeling and Robustness," *J. of the Royal Statistical Society, Series A* 143, 383–430.

Box, G. E. P., and G. C. Tiao. 1992. *Bayesian Inference in Statistical Analysis.* New York: Wiley Classics.

Casella, G., and R. L. Berger. 2002. *Statistical Inference,* 2d ed. Pacific Grove, Calif.: Duxbury.

Hogg, R. V., J. W. McKean, and A. T. Craig. 2005. *Introduction to Mathematical Statistics,* 6th ed. Upper Saddle River, N.J.: Prentice Hall.

Kepner, J., and D. Wackerly. 2002. "Observations on the Effect of the Prior Distribution on the Predictive Distribution in Bayesian Inferences," *Journal of Applied Statistics* 29(5): 761–769.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics,* 3d ed. New York: McGraw-Hill.

Rice, J. A. 1995. *Mathematical Statistics and Data Analysis,* 2d ed. Belmont, Calif.: Duxbury.