20. It is said that a random variable X has an *increasing failure rate* if the failure rate $h(x)$ defined in Exercise 18 is an increasing function of x for $x > 0$; and it is said that X has a *decreasing failure rate* if $h(x)$ is a decreasing function of x for $x > 0$. Suppose that X has a Weibull distribution with parameters a and b, as defined in Exercise 19. Show that X has an increasing failure rate if $b > 1$, and X has a decreasing failure rate if $b < 1$.

21. Let X have a gamma distribution with parameters $\alpha > 2$ and $\beta > 0$.
   a. Prove that the mean of $1/X$ is $\beta/(\alpha - 1)$.
   b. Prove that the variance of $1/X$ is $\beta^2/[(\alpha - 1)^2(\alpha - 2)]$.

22. Consider the Poisson process of radioactive particle hits in Example 5.9.1. Suppose that the rate λ of the Poisson process is unknown and has a gamma distribution with parameters α and β. Let X be the number of particles that strike the target during t time units. Prove that the conditional distribution of λ given $X = x$ is a gamma distribution, and find the parameters of that gamma distribution.

23. Let F be a continuous d.f. satisfying $F(0) = 0$, and suppose that the distribution with d.f. F has the memoryless property. Define $\ell(x) = \log[1 - F(x)]$.
   a. Show that for all $t, h > 0$,
$$1 - F(h) = \frac{1 - F(t + h)}{1 - F(t)}.$$
   b. Prove that $\ell(t + h) = \ell(t) + \ell(h)$ for all $t, h > 0$.
   c. Prove that for all $t > 0$ and all positive integers k and m, $\ell(kt/m) = (k/m)\ell(t)$.
   d. Prove that for all $t, c > 0$, $\ell(ct) = c\ell(t)$.
   e. Prove that $g(t) = \ell(t)/t$ is constant for $t > 0$.
   f. Prove that F must be the d.f. of an exponential distribution.

24. Review the derivation of the Black-Scholes formula (5.6.12). For this exercise, assume that our stock price at time u in the future is $S_0 e^{\mu u + W_u}$, where W_u has a gamma distribution with parameters αu and β with $\beta > 1$. Let r be the risk-free interest rate.
   a. Prove that $e^{-ru}E(S_u) = S_0$ if and only if $\mu = r - \alpha \log(\beta/[\beta - 1])$.
   b. Assume that $\mu = r - \alpha \log(\beta/[\beta - 1])$. Let R be one minus the d.f. of the gamma distribution with parameters αu and 1. Prove that the risk-neutral price for the option to buy one share of the stock for the price q at time u is $S_0 R(c[\beta - 1]) - qe^{-ru}R(c\beta)$, where
$$c = \log\left(\frac{q}{S_0}\right) + \alpha u \log\left(\frac{\beta}{\beta - 1}\right) - ru.$$
   c. Find the price for the option being considered when $u = 1$, $q = S_0$, $r = 0.06$, $\alpha = 1$, and $\beta = 10$.

5.10 The Beta Distribution

The family of beta distributions is a popular model for random variables that are known to take values in the interval $[0, 1]$. *One common example of such a random variable is the unknown probability of success in a sequence of Bernoulli trials.*

Definition of the Beta Distribution

It is said that a random variable X has a *beta distribution with parameters* α *and* β ($\alpha > 0$ and $\beta > 0$) if X has a continuous distribution for which the p.d.f. $f(x|\alpha, \beta)$ is as follows:

$$f(x|\alpha,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases} \quad \textbf{(5.10.1)}$$

In order to verify that the integral of this p.d.f. over the real line has the value 1, we must show that for $\alpha > 0$ and $\beta > 0$,

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \quad \textbf{(5.10.2)}$$

From the definition of the gamma function, it follows that

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^\infty u^{\alpha-1}e^{-u}\,du \int_0^\infty v^{\beta-1}e^{-v}\,dv$$
$$= \int_0^\infty \int_0^\infty u^{\alpha-1}v^{\beta-1}e^{-(u+v)}\,du\,dv.$$

Now we shall let

$$x = \frac{u}{u+v} \quad \text{and} \quad y = u + v.$$

Then $u = xy$ and $v = (1-x)y$, and it can be found that the value of the Jacobian of this inverse transformation is y. Furthermore, as u and v vary over all positive values, x will vary over the interval $(0, 1)$, and y will vary over all positive values. From Eq. (5.10.2), we now obtain the relation

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^1 \int_0^\infty x^{\alpha-1}(1-x)^{\beta-1}y^{\alpha+\beta-1}e^{-y}\,dy\,dx$$
$$= \Gamma(\alpha+\beta)\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

Therefore, Eq. (5.10.2) has been established.

It can be seen from Eq. (5.10.1) that the beta distribution with parameters $\alpha = 1$ and $\beta = 1$ is simply the uniform distribution on the interval $[0, 1]$.

Example 5.10.1 **Castaneda v. Partida.** In Example 5.2.1 on page 250, 220 grand jurors were chosen from a population that is 79.1 percent Mexican-American, but only 100 grand jurors were Mexican-American. The expected value of a binomial random variable X with parameters 220 and 0.791 is $E(X) = 220 \times 0.791 = 174.02$. This is much larger than the observed value of $X = 100$. Of course, such a discrepancy could occur by chance. After all, there is positive probability of $X = x$ for all $x = 0, \dots, 220$. Since the court assumed that X had a binomial distribution with parameters $n = 220$ and P, we should be interested in whether P is substantially less than the value 0.791, which represents impartial juror choice. For example, suppose that we define discrimination to mean that $P \le 0.8 \times 0.791 = 0.6328$. We would like to compute the conditional probability of $P \le 0.6328$ given $X = 100$.

Suppose that the distribution of P prior to observing X was a beta distribution with parameters α and β. Then the p.d.f. of P is

$$f_2(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}, \quad \text{for } 0 < p < 1.$$

The conditional p.f. of X given $P = p$ is the binomial p.f.

$$g_1(x|p) = \binom{220}{x}p^x(1-p)^{220-x}, \quad \text{for } x = 0, \dots, 220.$$

We can now apply Bayes' theorem for random variables (3.6.13) to obtain the conditional p.d.f. of P given $X = 100$:

$$g_2(p|100) = \frac{\binom{220}{100} p^{100}(1-p)^{120} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}}{f_1(100)}, \quad (5.10.3)$$

where $f_1(100)$ is the marginal p.f. of X at 100 obtained by the law of total probability for random variables (3.6.11):

$$\begin{aligned} f_1(100) &= \int_0^1 \binom{220}{100} p^{100}(1-p)^{120} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp \\ &= \binom{220}{100} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{\alpha+100-1}(1-p)^{\beta+120-1}\, dp \\ &= \binom{220}{100} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+100)\Gamma(\beta+120)}{\Gamma(\alpha+\beta+220)}, \end{aligned}$$

where the last equation follows from Eq. (5.10.2). Substituting this into (5.10.3) yields

$$g_2(p|100) = \frac{\Gamma(\alpha+\beta+220)}{\Gamma(\alpha+100)\Gamma(\beta+120)} p^{\alpha+100-1}(1-p)^{\beta+120-1}, \text{ for } 0 < p < 1. \quad (5.10.4)$$

This is easily recognized as the p.d.f. of the beta distribution with parameters $\alpha + 100$ and $\beta + 120$. After choosing values of α and β, we could compute $\Pr(P \le 0.6328|X = 100)$ and decide how likely it is that there was discrimination. We will see how to choose α and β after we learn how to compute the expected value of a beta random variable. ◀

NOTE:* *Conditional Distribution of P after Observing X with Binomial Distribution. The calculation of the conditional distribution of P given $X = 100$ in Example 5.10.1 is a special case of a useful general result. Suppose that P has a beta distribution with parameters α and β, and the conditional distribution of X given $P = p$ is a binomial distribution with parameters n and p. Then the same calculation that led to Eq. (5.10.4) shows that the conditional distribution of P given $X = x$ is a beta distribution with parameters $\alpha + x$ and $\beta + n - x$.

Moments of the Beta Distribution

When the p.d.f. of a random variable X is given by Eq. (5.10.1), the moments of X are easily calculated. For $k = 1, 2, \ldots,$

$$\begin{aligned} E(X^k) &= \int_0^1 x^k f(x|\alpha,\beta)\, dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+k-1}(1-x)^{\beta-1}\, dx. \end{aligned}$$

Therefore, by Eq. (5.10.2),

$$E(X^k) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+k)\Gamma(\beta)}{\Gamma(\alpha+k+\beta)}$$
$$= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+k-1)}.$$

It follows that

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

and

$$\mathrm{Var}(X) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2$$
$$= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Example 5.10.2 **Castaneda v. Partida.** Continuing Example 5.10.1, we are now prepared to see why, for every reasonable choice one makes for α and β, the probability of discrimination in Castaneda v. Partida is quite large. To avoid bias either for or against the defendant, we shall suppose that, before learning X, the probability that a Mexican-American juror would be selected would be 0.791. Let $Y = 1$ if a Mexican-American juror is selected and let $Y = 0$ if not. Then Y has a Bernoulli distribution with parameter p given $P = p$ and $E(Y|p) = p$. So the law of total probability for expectations, Theorem 4.7.1, says that

$$\Pr(Y = 1) = E(Y) = E[E(Y|P)] = E(P).$$

This means that we should choose α and β so that $E(P) = 0.791$. Because $E(P) = \alpha/(\alpha+\beta)$, this means that $\alpha = 3.785\beta$. Because the conditional distribution of P given $X = 100$ is a beta distribution with parameters $\alpha + 100$ and $\beta + 120$, we can compute $\Pr(P \le 0.6328|X = 100)$ for each value of β with $\alpha = 3.785\beta$ and see if it is ever small. A plot of $\Pr(P \le 0.6328|X = 100)$ for various values of β is given in Fig. 5.7. In order for $\Pr(P \le 0.6328|X = 100) < 0.5$ we need $\beta \ge 51.5$. This makes $\alpha \ge 194.9$. We claim that the beta distribution with parameters 194.9 and 51.5 as well as all others that make $\Pr(P \le 0.6328|X = 100) < 0.5$ are unreasonable because they are incredibly prejudiced about the possibility of discrimination. For example, every such prior distribution for P satisfies $\Pr(P \le 0.6328) \le 3.28 \times 10^{-8}$, which is essentially 0. That is, anyone who thought the probability of discrimination was greater than 3.28×10^{-8} before learning $X = 100$ would believe that the probability of discrimination is at least 0.5 after learning $X = 100$. This is then fairly convincing evidence that there was discrimination in this case. ◀

Example 5.10.3 **A Clinical Trial.** Consider the clinical trial described in Example 2.1.3. Let P be the probability that a patient in the imipramine treatment group has no relapse (called success). A popular model for P is that P has a beta distribution with parameters α and β. Choosing α and β can be done based on expert opinion about the chance of success

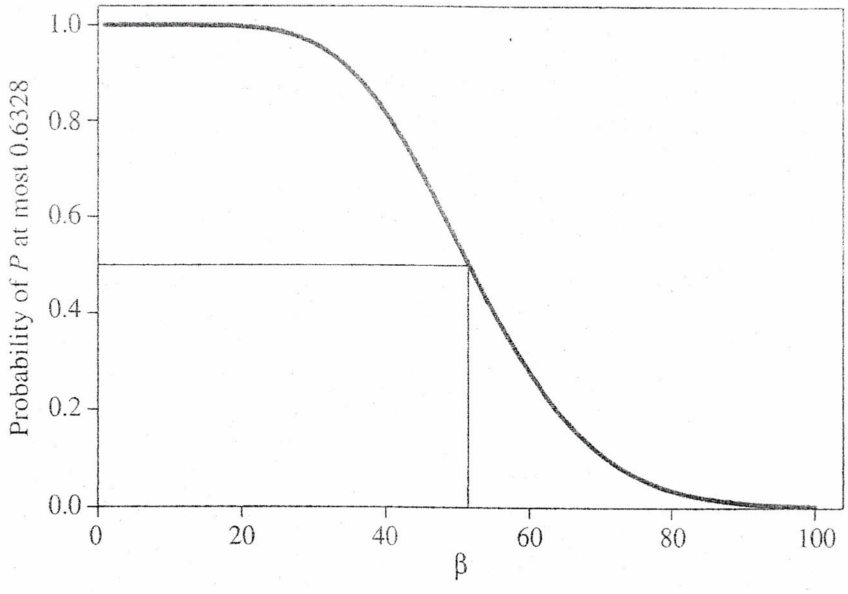

Figure 5.7 Probability of discrimination as a function of β.

and on the effect that data should have on the distribution of P after observing the data. For example, suppose that the doctors running the clinical trial think that the probability of success should be around 1/3. Let $X_i = 1$ if the ith patient is a success and $X_i = 0$ if not. We are supposing that $E(X_i|p) = \Pr(X_i = 1|p) = p$, so the law of total probability for expectations (Theorem 4.7.1) says that

$$\Pr(X_i = 1) = E(X_i) = E[E(X_i|P)] = E(P) = \frac{\alpha}{\alpha + \beta}.$$

If we want $\Pr(X_i = 1) = 1/3$, we need $\alpha/(\alpha + \beta) = 1/3$, so $\beta = 2\alpha$. Of course, the doctors will revise the probability of success after observing patients from the study. The doctors can choose β based on how that revision will occur.

Assume that the random variables $X_1, X_2, \ldots$ (the indicators of success) are conditionally independent given $P = p$. Let $X = X_1 + \cdots + X_n$ be the number of patients out of the first n who are successes. The conditional distribution of X given $P = p$ is a binomial distribution with parameters n and p, and the marginal distribution of P is a beta distribution with parameters α and β. The note following Example 5.10.1 tells us that the conditional distribution of P given $X = x$ is a beta distribution with parameters $\alpha + x$ and $\beta + n - x$. Suppose that a sequence of 20 patients, all of whom are successes, would raise the doctors' probability of success from 1/3 up to 0.9. Then

$$0.9 = E(P|X = 20) = \frac{\alpha + 20}{\alpha + \beta + 20}.$$

This equation implies that $\alpha + 20 = 9\beta$. Combining this with $\beta = 2\alpha$, we get $\alpha = 1.18$ and $\beta = 2.35$.

Finally, we can ask, what will be the distribution of P after observing some patients in the study? Suppose that 40 patients are actually observed, and 22 of them recover (as in Table 2.1). Then the conditional distribution of P given this observation is a beta distribution with parameters $1.18 + 22 = 23.18$ and $2.35 + 18 = 20.35$. It follows that

$$E(P|X = 22) = \frac{23.18}{23.18 + 20.35} = 0.5325.$$

Notice how much closer this is to the proportion of successes (0.55) than was $E(P) = 1/3$. ◀

Summary

The family of beta distributions is a popular model for random variables that lie in the interval (0, 1), such as unknown probabilities of success for sequences of Bernoulli trials. The mean of the beta distribution with parameters α and β is $\alpha/(\alpha + \beta)$. If X has a binomial distribution with parameters n and p conditional on $P = p$, and if P has a beta distribution with parameters α and β, then, conditional on $X = x$, the distribution of P is a beta distribution with parameters $\alpha + x$ and $\beta + n - x$.

EXERCISES

1. Compute the quantile function of the beta distribution with parameters $\alpha > 0$ and $\beta = 1$.

2. Determine the mode of the beta distribution with parameters α and β, assuming that $\alpha > 1$ and $\beta > 1$.

3. Sketch the p.d.f. of the beta distribution for each of the following pairs of values of the parameters:

a. $\alpha = 1/2$ and $\beta = 1/2$, **b.** $\alpha = 1/2$ and $\beta = 1$,

c. $\alpha = 1/2$ and $\beta = 2$, **d.** $\alpha = 1$ and $\beta = 1$,

e. $\alpha = 1$ and $\beta = 2$, **f.** $\alpha = 2$ and $\beta = 2$,

g. $\alpha = 25$ and $\beta = 100$, **h.** $\alpha = 100$ and $\beta = 25$.

4. Suppose that X has a beta distribution with parameters α and β. Show that $1 - X$ has a beta distribution with parameters β and α.

5. Suppose that X has a beta distribution with parameters α and β, and let r and s be given positive integers. Determine the value of $E[X^r(1 - X)^s]$.

6. Suppose that X and Y are independent random variables, X has a gamma distribution with parameters α_1 and β, and Y has a gamma distribution with parameters α_2 and β. Let $U = X/(X + Y)$ and $V = X + Y$. Show that **(a)** U has a beta distribution with parameters α_1 and α_2, and **(b)** U and V are independent.

7. Suppose that X_1 and X_2 form a random sample of two observed values from an exponential distribution with parameter β. Show that $X_1/(X_1 + X_2)$ has a uniform distribution on the interval [0, 1].

8. Suppose that the proportion X of defective items in a large lot is unknown, and X has a beta distribution with parameters α and β.

a. If one item is selected at random from the lot, what is the probability that it will be defective?

b. If two items are selected at random from the lot, what is the probability that both will be defective?

9. A manufacturer believes that defective products are produced with unknown probability P, which will be modeled as having a beta distribution. The manufacturer thinks that P should be around 0.05, but if the first 10 observed products were all defective, the mean of P would rise from 0.05 to 0.9. Find the beta distribution that has these properties.

10. A marketer is interested in how many customers are likely to buy a particular product in a particular store. Let P be the probability that a customer in that store buys the product. Let the distribution of P be uniform on the interval [0, 1] before observing any data. The marketer then observes 25 customers and only six buy the product. If the customers were conditionally independent given P, find the conditional distribution of P given the observed customers.