5.8.   THE CORRECTION FOR CONTINUITY

Approximating a Discrete Distribution by a Continuous Distribution

Suppose that $X_1, \ldots, X_n$ form a random sample from a discrete distribution, and let $X = X_1 + \cdots + X_n$. It was shown in the previous section that even though the distribution of X will be discrete, this distribution can be approximated by the normal distribution, which is continuous. In this section, we shall describe a standard method for improving the quality of the approximation that is obtained when a probability based on a discrete distribution is approximated by one based on a continuous distribution.

Suppose, therefore, that the random variable X has a discrete distribution with p.f. $f(x)$ and it is desired to approximate this distribution by a continuous distribution with p.d.f. $g(x)$. For simplicity, we shall consider only a discrete distribution for which all possible values of X are integers. This condition is satisfied for the binomial, hypergeometric, Poisson, and negative binomial distributions described in this chapter.

If the p.d.f. $g(x)$ provides a good approximation to the distribution of X, then for any integers a and b, we can simply approximate the probability

$$\Pr(a \leqslant X \leqslant b) = \sum_{x=a}^{b} f(x) \tag{1}$$

by the integral

$$\int_a^b g(x)\, dx. \tag{2}$$

Indeed, this approximation was used in Examples 1 and 3 of Sec. 5.7, where $g(x)$ was the appropriate normal p.d.f. specified by the central limit theorem.

This simple approximation has the following shortcoming: Although $\Pr(X \geqslant a)$ and $\Pr(X > a)$ will typically have different values for the discrete distribution, these probabilities will always be equal for the continuous distribution. Another way of expressing this shortcoming is as follows: Although $\Pr(X = x) > 0$ for any integer x that is a possible value of X, this probability is necessarily 0 under the approximating p.d.f.

Approximating a Histogram

The p.f. $f(x)$ of X can be represented by a *histogram*, or *bar chart*, as sketched in Fig. 5.4. For each integer x, the probability of x is represented by the area of a rectangle with a base that extends from $x - \frac{1}{2}$ to $x + \frac{1}{2}$ and with a height $f(x)$.
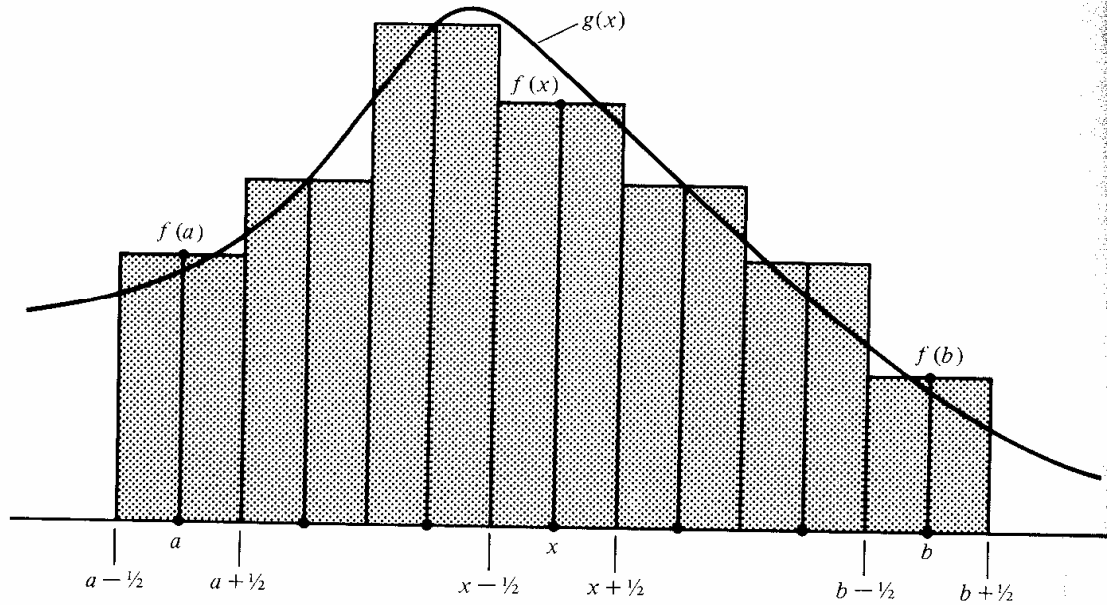
Figure 5.4 Approximating a histogram by using a p.d.f.

Thus, the area of the rectangle for which the center of the base is at the integer x is simply $f(x)$. An approximating p.d.f. $g(x)$ is also sketched in Fig. 5.4.

From this point of view it can be seen that $\Pr(a \leqslant X \leqslant b)$, as specified in Eq. (1), is the sum of the areas of the rectangles in Fig. 5.4 that are centered at $a, a + 1, \ldots, b$. It can also be seen from Fig. 5.4 that the sum of these areas is approximated by the integral

$$\int_{a-(1/2)}^{b+(1/2)} g(x)\, dx.$$

(3)

The adjustment from the integral (2) to the integral (3) is called the *correction for continuity*.

If we use the correction for continuity, we find that the probability $f(a)$ of the single integer a can be approximated as follows:

$$\Pr(X = a) = \Pr\left(a - \frac{1}{2} \leqslant X \leqslant a + \frac{1}{2}\right)$$

(4)

$$\approx \int_{a-(1/2)}^{a+(1/2)} g(x)\, dx.$$

Similarly,

$$\Pr(X > a) = \Pr(X \geqslant a + 1) = \Pr\left(X \geqslant a + \frac{1}{2}\right)$$

$$\approx \int_{a+(1/2)}^{\infty} g(x)\,dx. \tag{5}$$

Example 1: Examination Questions. To illustrate the use of the correction for continuity, we shall again consider Example 3 of Sec. 5.7. In that example, an examination contains 99 questions of varying difficulty and it is desired to determine $\Pr(X \geqslant 60)$, where X denotes the total number of questions that a particular student answers correctly. Then under the conditions of the example, it is found from the central limit theorem that the discrete distribution of X could be approximated by a normal distribution with mean 49.5 and standard deviation 4.08.

If we use the correction for continuity, we obtain

$$\Pr(X \geqslant 60) = \Pr(X \geqslant 59.5) = \Pr\left(Z \geqslant \frac{59.5 - 49.5}{4.08}\right)$$

$$\approx 1 - \Phi(2.4510) = 0.007.$$

This value can be compared with the value 0.005 that was obtained in Sec. 5.7 without the correction.   □

Example 2: Coin Tossing. Suppose that a fair coin is tossed 20 times, and that all tosses are independent. What is the probability of obtaining exactly 10 heads?

Let X denote the total number of heads obtained in the 20 tosses. According to the central limit theorem, the distribution of X will be approximately a normal distribution with mean 10 and standard deviation $[(20)(1/2)(1/2)]^{1/2} = 2.236$. If we use the correction for continuity,

$$\Pr(X = 10) = \Pr(9.5 \leqslant X \leqslant 10.5)$$

$$= \Pr\left(-\frac{0.5}{2.236} \leqslant Z \leqslant \frac{0.5}{2.236}\right)$$

$$\approx \Phi(0.2236) - \Phi(-0.2236) = 0.177.$$

The exact value of $\Pr(X = 10)$ found from the table of binomial probabilities given at the back of this book is 0.1762. Thus, the normal approximation with the correction for continuity is quite good.   □

EXERCISES

1. Let X denote the total number of successes in 15 Bernoulli trials, with probability of success $p = 0.3$ on each trial.