



Chapter

LARGE RANDOM SAMPLES

6

- 6.1 Introduction
- 6.2 The Law of Large Numbers
- 6.3 The Central Limit Theorem

- 6.4 The Correction for Continuity
- 6.5 Supplementary Exercises

6.1 Introduction

In this chapter, we introduce a number of approximation results that simplify the analysis of large random samples. In the first section, we give two examples to illustrate the types of analyses that we might wish to perform and how additional tools may be needed to be able to perform them.

Example
6.1.1

Proportion of Heads. If you draw a coin from your pocket, you might feel confident that it is essentially fair. That is, the probability that it will land with head up when flipped is 1/2. However, if you were to flip the coin 10 times, you would not expect to see exactly 5 heads. If you were to flip it 100 times, you would be even less likely to see exactly 50 heads. Indeed, we can calculate the probabilities of each of these two results using the fact that the number of heads in n independent flips of a fair coin has the binomial distribution with parameters n and 1/2. So, if X is the number of heads in 10 independent flips, we know that

$$\Pr(X = 5) = \binom{10}{5} \left(\frac{1}{2}\right)^5 \left(1 - \frac{1}{2}\right)^5 = 0.2461.$$

If Y is the number of heads in 100 independent flips, we have

$$\Pr(Y = 50) = \binom{100}{50} \left(\frac{1}{2}\right)^{50} \left(1 - \frac{1}{2}\right)^{50} = 0.0796.$$

Even though the probability of exactly $n/2$ heads in n flips is quite small, especially for large n , you still expect the proportion of heads to be close to 1/2 if n is large. For example, if $n = 100$, the proportion of heads is $Y/100$. In this case, the probability that the proportion is within 0.1 of 1/2 is

$$\Pr\left(0.4 \leq \frac{Y}{100} \leq 0.6\right) = \Pr(40 \leq Y \leq 60) = \sum_{i=40}^{60} \binom{100}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{100-i} = 0.9648.$$

A similar calculation with $n = 10$ yields

$$\Pr\left(0.4 \leq \frac{X}{10} \leq 0.6\right) = \Pr(4 \leq Y \leq 6) = \sum_{i=4}^6 \binom{10}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{10-i} = 0.6563.$$

Notice that the probability that the proportion of heads in n tosses is close to 1/2 is larger for $n = 100$ than for $n = 10$ in this example. This is due in part to the fact that

we have defined “close to 1/2” to be the same for both cases, namely, between 0.4 and 0.6. \blacktriangleleft

The calculations performed in Example 6.1.1 were simple enough because we have a formula for the probability function of the number of heads in any number of flips. For more complicated random variables, the situation is not so simple.

**Example
6.1.2**

Average Waiting Time. A queue is serving customers, and the i th customer waits a random time X_i to be served. Suppose that X_1, X_2, \dots are i.i.d. random variables having the uniform distribution on the interval $[0, 1]$. The mean waiting time is 0.5. Intuition suggests that the average of a large number of waiting times should be close to the mean waiting time. But the distribution of the average of X_1, \dots, X_n is rather complicated for every $n > 1$. It may not be possible to calculate precisely the probability that the sample average is close to 0.5 for large samples. \blacktriangleleft

The law of large numbers (Theorem 6.2.4) will give a mathematical foundation to the intuition that the average of a large sample of i.i.d. random variables, such as the waiting times in Example 6.1.2, should be close to their mean. The central limit theorem (Theorem 6.3.1) will give us a way to approximate the probability that the sample average is close to the mean.

Exercises

1. The solution to Exercise 1 of Sec. 3.9 is the p.d.f. of $X_1 + X_2$ in Example 6.1.2. Find the p.d.f. of $\bar{X}_2 = (X_1 + X_2)/2$. Compare the probabilities that \bar{X}_2 and X_1 are close to 0.5. In particular, compute $\Pr(|\bar{X}_2 - 0.5| < 0.1)$ and $\Pr(|X_1 - 0.5| < 0.1)$. What feature of the p.d.f. of \bar{X}_2 makes it clear that the distribution is more concentrated near the mean?

2. Let X_1, X_2, \dots be a sequence of i.i.d. random variables having the normal distribution with mean μ and variance σ^2 . Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ be the sample mean of the first n random variables in the sequence. Show that

$\Pr(|\bar{X}_n - \mu| \leq c)$ converges to 1 as $n \rightarrow \infty$. Hint: Write the probability in terms of the standard normal c.d.f. Φ and use what you know about this c.d.f.

3. This problem requires a computer program because the calculation is too tedious to do by hand. Extend the calculation in Example 6.1.1 to the case of $n = 200$ flips. That is, let W be the number of heads in 200 flips of a fair coin, and compute $\Pr(0.4 \leq \frac{W}{200} \leq 0.6)$. What do you think is the continuation of the pattern of these probabilities as the number of flips n increases without bound?

6.2 The Law of Large Numbers

The average of a random sample of i.i.d. random variables is called their sample mean. The sample mean is useful for summarizing the information in a random sample in much the same way that the mean of a probability distribution summarizes the information in the distribution. In this section, we present some results that illustrate the connection between the sample mean and the expected value of the individual random variables that comprise the random sample.

The Markov and Chebyshev Inequalities

We shall begin this section by presenting two simple and general results, known as the Markov inequality and the Chebyshev inequality. We shall then apply these inequalities to random samples.

The Markov inequality is related to the claim made on page 211 about how the mean of a distribution can be affected by moving a small amount of probability to an arbitrarily large value. The Markov inequality puts a bound on how much probability can be at arbitrarily large values once the mean is specified.

**Theorem
6.2.1**

Markov Inequality. Suppose that X is a random variable such that $\Pr(X \geq 0) = 1$. Then for every real number $t > 0$,

$$\Pr(X \geq t) \leq \frac{E(X)}{t}. \quad (6.2.1)$$

Proof For convenience, we shall assume that X has a discrete distribution for which the p.f. is f . The proof for a continuous distribution or a more general type of distribution is similar. For a discrete distribution,

$$E(X) = \sum_x xf(x) = \sum_{x < t} xf(x) + \sum_{x \geq t} xf(x).$$

Since X can have only nonnegative values, all the terms in the summations are nonnegative. Therefore,

$$E(X) \geq \sum_{x \geq t} xf(x) \geq \sum_{x \geq t} tf(x) = t \Pr(X \geq t). \quad (6.2.2)$$

Divide the extreme ends of (6.2.2) by $t > 0$ to obtain (6.2.1). ■

The Markov inequality is primarily of interest for large values of t . In fact, when $t \leq E(X)$, the inequality is of no interest whatsoever, since it is known that $\Pr(X \leq t) \leq 1$. However, it is found from the Markov inequality that for every nonnegative random variable X whose mean is 1, the maximum possible value of $\Pr(X \geq 100)$ is 0.01. Furthermore, it can be verified that this maximum value is actually attained by every random variable X for which $\Pr(X = 0) = 0.99$ and $\Pr(X = 100) = 0.01$.

The Chebyshev inequality is related to the idea that the variance of a random variable is a measure of how spread out its distribution is. The inequality says that the probability that X is far away from its mean is bounded by a quantity that increases as $\text{Var}(X)$ increases.

**Theorem
6.2.2**

Chebyshev Inequality. Let X be a random variable for which $\text{Var}(X)$ exists. Then for every number $t > 0$,

$$\Pr(|X - E(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}. \quad (6.2.3)$$

Proof Let $Y = [X - E(X)]^2$. Then $\Pr(Y \geq 0) = 1$ and $E(Y) = \text{Var}(X)$. By applying the Markov inequality to Y , we obtain the following result:

$$\Pr(|X - E(X)| \geq t) = \Pr(Y \geq t^2) \leq \frac{\text{Var}(X)}{t^2}. \quad \blacksquare$$

It can be seen from this proof that the Chebyshev inequality is simply a special case of the Markov inequality. Therefore, the comments that were given following the proof of the Markov inequality can be applied as well to the Chebyshev inequality. Because of their generality, these inequalities are very useful. For example, if $\text{Var}(X) = \sigma^2$ and we let $t = 3\sigma$, then the Chebyshev inequality yields the result that

$$\Pr(|X - E(X)| \geq 3\sigma) \leq \frac{1}{9}.$$

In words, the probability that any given random variable will differ from its mean by more than 3 standard deviations *cannot* exceed 1/9. This probability will actually be much smaller than 1/9 for many of the random variables and distributions that will be discussed in this book. The Chebyshev inequality is useful because of the fact that this probability must be 1/9 or less for *every* distribution. It can also be shown (see Exercise 4 at the end of this section) that the upper bound in (6.2.3) is sharp in the sense that it cannot be made any smaller and still hold for *all* distributions.

Properties of the Sample Mean

In Definition 5.6.3, we defined the *sample mean* of n random variables X_1, \dots, X_n to be their average,

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

The mean and the variance of \bar{X}_n are easily computed.

Theorem

6.2.3

Mean and Variance of the Sample Mean. Let X_1, \dots, X_n be a random sample from a distribution with mean μ and variance σ^2 . Let \bar{X}_n be the sample mean. Then $E(\bar{X}_n) = \mu$ and $\text{Var}(\bar{X}_n) = \sigma^2/n$.

Proof It follows from Theorems 4.2.1 and 4.2.4 that

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n\mu = \mu.$$

Furthermore, since X_1, \dots, X_n are independent, Theorems 4.3.4 and 4.3.5 say that

$$\begin{aligned} \text{Var}(\bar{X}_n) &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$
■

In words, the mean of \bar{X}_n is equal to the mean of the distribution from which the random sample was drawn, but the variance of \bar{X}_n is only $1/n$ times the variance of that distribution. It follows that the probability distribution of \bar{X}_n will be more concentrated around the mean value μ than was the original distribution. In other words, the sample mean \bar{X}_n is more likely to be close to μ than is the value of just a single observation X_i from the given distribution.

These statements can be made more precise by applying the Chebyshev inequality to \bar{X}_n . Since $E(\bar{X}_n) = \mu$ and $\text{Var}(\bar{X}_n) = \sigma^2/n$, it follows from the relation (6.2.3) that for every number $t > 0$,

$$\Pr(|\bar{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}. \quad (6.2.4)$$

Example

6.2.1

Determining the Required Number of Observations. Suppose that a random sample is to be taken from a distribution for which the value of the mean μ is not known, but for which it is known that the standard deviation σ is 2 units or less. We shall determine how large the sample size must be in order to make the probability at least 0.99 that $|\bar{X}_n - \mu|$ will be less than 1 unit.

Since $\sigma^2 \leq 2^2 = 4$, it follows from the relation (6.2.4) that for every sample size n ,

$$\Pr(|\bar{X}_n - \mu| \geq 1) \leq \frac{\sigma^2}{n} \leq \frac{4}{n}.$$

Since n must be chosen so that $\Pr(|\bar{X}_n - \mu| < 1) \geq 0.99$, it follows that n must be chosen so that $4/n \leq 0.01$. Hence, it is required that $n \geq 400$. \blacktriangleleft

**Example
6.2.2**

A Simulation. An environmental engineer believes that there are two contaminants in a water supply, arsenic and lead. The actual concentrations of the two contaminants are independent random variables X and Y , measured in the same units. The engineer is interested in what proportion of the contamination is lead on average. That is, the engineer wants to know the mean of $R = Y/(X + Y)$. We suppose that it is a simple matter to generate as many independent pseudo-random numbers with the distributions of X and Y as we desire. A common way to obtain an approximation to $E[Y/(X + Y)]$ would be the following: If we sample n pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ and compute $R_i = Y_i/(X_i + Y_i)$ for $i = 1, \dots, n$, then $\bar{R}_n = \frac{1}{n} \sum_{i=1}^n R_i$ is a sensible approximation to $E(R)$. To decide how large n should be, we can argue as in Example 6.2.1. Since it is known that $|R_i| \leq 1$, it must be that $\text{Var}(R_i) \leq 1$. (Actually, $\text{Var}(R_i) \leq 1/4$, but this is harder to prove. See Exercise 14 in this section for a way to prove it in the discrete case.) According to Chebyshev's inequality, for each $\epsilon > 0$,

$$\Pr(|\bar{R}_n - E(R)| \geq \epsilon) \leq \frac{1}{n\epsilon^2}.$$

So, if we want $|\bar{R}_n - E(R)| \leq 0.005$ with probability 0.98 or more, then we should use $n > 1/[0.2 \times 0.005^2] = 2,000,000$. \blacktriangleleft

It should be emphasized that the use of the Chebyshev inequality in Example 6.2.1 guarantees that a sample for which $n = 400$ will be large enough to meet the specified probability requirements, regardless of the particular type of distribution from which the sample is to be taken. If further information about this distribution is available, then it can often be shown that a smaller value for n will be sufficient. This property is illustrated in the next example.

**Example
6.2.3**

Tossing a Coin. Suppose that a fair coin is to be tossed n times independently. For $i = 1, \dots, n$, let $X_i = 1$ if a head is obtained on the i th toss, and let $X_i = 0$ if a tail is obtained on the i th toss. Then the sample mean \bar{X}_n will simply be equal to the proportion of heads that are obtained on the n tosses. We shall determine the number of times the coin must be tossed in order to make $\Pr(0.4 \leq \bar{X}_n \leq 0.6) \geq 0.7$. We shall determine this number in two ways: first, by using the Chebyshev inequality; second, by using the exact probabilities for the binomial distribution of the total number of heads.

Let $T = \sum_{i=1}^n X_i$ denote the total number of heads that are obtained when n tosses are made. Then T has the binomial distribution with parameters n and $p = 1/2$. Therefore, it follows from Eq. (4.2.5) on page 221 that $E(T) = n/2$, and it follows from Eq. (4.3.3) on page 232 that $\text{Var}(T) = n/4$. Because $\bar{X}_n = T/n$, we can obtain

the following relation from the Chebyshev inequality:

$$\begin{aligned}\Pr(0.4 \leq \bar{X}_n \leq 0.6) &= \Pr(0.4n \leq T \leq 0.6n) \\ &= \Pr\left(\left|T - \frac{n}{2}\right| \leq 0.1n\right) \\ &\geq 1 - \frac{n}{4(0.1n)^2} = 1 - \frac{25}{n}.\end{aligned}$$

Hence, if $n \geq 84$, this probability will be at least 0.7, as required.

However, from the table of binomial distributions given at the end of this book, it is found that for $n = 15$,

$$\Pr(0.4 \leq \bar{X}_n \leq 0.6) = \Pr(6 \leq T \leq 9) = 0.70.$$

Hence, 15 tosses would actually be sufficient to satisfy the specified probability requirement. \blacktriangleleft

The Law of Large Numbers

The discussion in Example 6.2.3 indicates that the Chebyshev inequality may not be a practical tool for determining the appropriate sample size in a particular problem, because it may specify a much greater sample size than is actually needed for the particular distribution from which the sample is being taken. However, the Chebyshev inequality is a valuable theoretical tool, and it will be used here to prove an important result known as the *law of large numbers*.

Suppose that Z_1, Z_2, \dots is a sequence of random variables. Roughly speaking, it is said that this sequence converges to a given number b if the probability distribution of Z_n becomes more and more concentrated around b as $n \rightarrow \infty$. To be more precise, we give the following definition.

**Definition
6.2.1**

Convergence in Probability. A sequence Z_1, Z_2, \dots of random variables *converges to* b in probability if for every number $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr(|Z_n - b| < \varepsilon) = 1.$$

This property is denoted by

$$Z_n \xrightarrow{P} b,$$

and is sometimes stated simply as Z_n converges to b in probability.

In other words, Z_n converges to b in probability if the probability that Z_n lies in each given interval around b , no matter how small this interval may be, approaches 1 as $n \rightarrow \infty$.

We shall now show that the sample mean of a random sample with finite variance always converges in probability to the mean of the distribution from which the random sample was taken.

**Theorem
6.2.4**

Law of Large Numbers. Suppose that X_1, \dots, X_n form a random sample from a distribution for which the mean is μ and for which the variance is finite. Let \bar{X}_n denote the sample mean. Then

$$\bar{X}_n \xrightarrow{P} \mu. \quad (6.2.5)$$

Proof Let the variance of each X_i be σ^2 . It then follows from the Chebyshev inequality that for every number $\varepsilon > 0$,

$$\Pr(|\bar{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Hence,

$$\lim_{n \rightarrow \infty} \Pr(|\bar{X}_n - \mu| < \varepsilon) = 1,$$

which means that $\bar{X}_n \xrightarrow{P} \mu$. ■

It can also be shown that Eq. (6.2.5) is satisfied if the distribution from which the random sample is taken has a finite mean μ but an infinite variance. However, the proof for this case is beyond the scope of this book.

Since \bar{X}_n converges to μ in probability, it follows that there is high probability that \bar{X}_n will be close to μ if the sample size n is large. Hence, if a large random sample is taken from a distribution for which the mean is unknown, then the arithmetic average of the values in the sample will usually be a close estimate of the unknown mean. This topic will be discussed again in Sec. 6.3, where we introduce the central limit theorem. It will then be possible to present a more precise probability distribution for the difference between \bar{X}_n and μ .

The following result can be useful if we observe random variables with mean μ but are interested in μ^2 or $\log(\mu)$ or some other continuous function of μ . The proof is left for the reader (Exercise 15).

Theorem 6.2.5 **Continuous Functions of Random Variables.** If $Z_n \xrightarrow{P} b$, and if $g(z)$ is a function that is continuous at $z = b$, then $g(Z_n) \xrightarrow{P} g(b)$. ■

Similarly, it is almost as easy to show that if $Z_n \xrightarrow{P} b$ and $Y_n \xrightarrow{P} c$, and if $g(z, y)$ is continuous at $(z, y) = (b, c)$, then $g(Z_n, Y_n) \xrightarrow{P} g(b, c)$ (Exercise 16). Indeed, Theorem 6.2.5 extends to any finite number k of sequences that converge in probability and a continuous function of k variables.

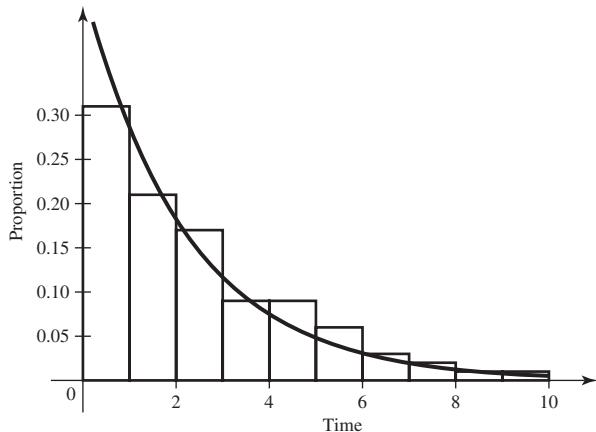
The law of large numbers helps to explain why a histogram (Definition 3.7.9) can be used as an approximation to a p.d.f.

Theorem 6.2.6 **Histograms.** Let X_1, X_2, \dots be a sequence of i.i.d. random variables. Let $c_1 < c_2$ be two constants. Define $Y_i = 1$ if $c_1 \leq X_i < c_2$ and $Y_i = 0$ if not. Then $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$ is the proportion of X_1, \dots, X_n that lie in the interval $[c_1, c_2]$, and $\bar{Y}_n \xrightarrow{P} \Pr(c_1 \leq X_1 < c_2)$. ■

Proof By construction, Y_1, Y_2, \dots are i.i.d. Bernoulli random variables with parameter $p = \Pr(c_1 \leq X_1 < c_2)$. Theorem 6.2.4 says that $\bar{Y}_n \xrightarrow{P} p$. ■

In words, Theorem 6.2.6 says the following: If we draw a histogram with the area of the bar over each subinterval being the proportion of a random sample that lies in the corresponding subinterval, then the area of each bar converges in probability to the probability that a random variable from the sequence lies in the subinterval. If the sample is large, we would then expect the area of each bar to be close to the probability. The same idea applies to a conditionally i.i.d. (given $Z = z$) sample, with $\Pr(c_1 \leq X_1 < c_2)$ replaced by $\Pr(c_1 \leq X_1 < c_2 | Z = z)$.

Figure 6.1 Histogram of service times for Example 6.2.4 together with graph of the conditional p.d.f. from which the service times were simulated.



Example 6.2.4

Rate of Service. In Example 3.7.20, we drew a histogram of an observed sample of $n = 100$ service times. The service times were actually simulated as an i.i.d. sample from the exponential distribution with parameter 0.446. Figure 6.1 reproduces the histogram overlayed with the graph of $g(x|z_0)$ where $z_0 = 0.446$. Because the width of each bar is 1, the area of each bar equals the proportion of the sample that lies in the corresponding interval. The area under the curve $g(x|z_0)$ is $\Pr(c_1 \leq X_1 < c_2 | Z = z_0)$ for each interval $[c_1, c_2]$. Notice how closely the area under the conditional p.d.f. matches the area of each bar. ◀

The reason that the p.d.f. and the heights of the bars in the histogram in Fig. 6.1 match so closely is that the area of each bar is converging in probability to the area under the graph of the p.d.f. The sum of the areas of the bars is 1, which is the same as the area under the graph of the p.d.f. If we had chosen the heights of the bars in the histogram to represent counts, then the sum of the areas of the bars would have been $n = 100$, and the bars would have been about 100 times as high as the p.d.f.

We could choose a different width for the subintervals in the histogram and still keep the areas equal to the proportions in the subintervals.

Example 6.2.5

Rate of Service. In Example 6.2.4, we can choose 20 bars of width 0.5 instead of 10 bars of width 1. To make the area of each bar represent the proportion in the subinterval, the height of each bar should equal the proportion divided by 0.5. The probability of an observation being in each interval $[c_1, c_2]$ would be

$$\begin{aligned} \Pr(c_1 \leq X_1 < c_2 | Z = x) &= \int_{c_1}^{c_2} g(x|z) dx \approx (c_2 - c_1)g([c_1 + c_2]/2|z) \\ &= 0.5 * g([c_1 + c_2]/2|z). \end{aligned} \quad (6.2.6)$$

Recall that the probability in (6.2.6) should be close to the proportion of the sample in the interval. If we divide both the probability and the proportion by 0.5, we see that the height of the histogram bar should be close to $g([c_1 + c_2]/2)$. Hence, the graph of the p.d.f. should still be close to the heights of the histogram bars. What we are doing here is choosing $r = n(b - a)/k$ in Definition 3.7.9. Figure 6.2 shows the histogram with 20 intervals of length 0.5 together with the same p.d.f. from Fig. 6.1. The bar heights are still similar to the p.d.f., but they are much more variable in

Figure 6.2 Modified histogram of service times from Example 6.2.4 together with graph of the conditional p.d.f. This time, the width of each interval is 0.5.

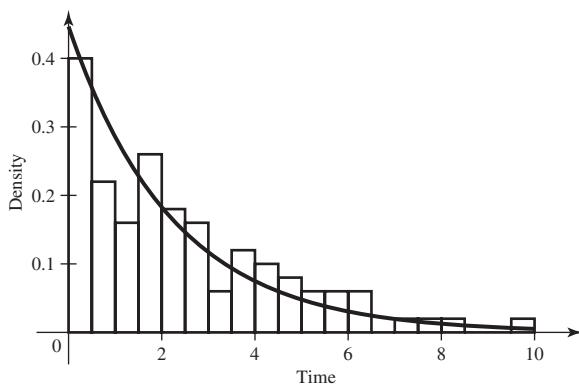


Fig. 6.2 compared to Fig. 6.1. Exercise 17 helps to explain why the bar heights are more variable in this example. ◀

The reasoning used to construct Figures 6.1 and 6.2 applies even when the subintervals used to construct the histogram have different widths. In this case, each bar should have height equal to the raw count divided by both n (the sample size) and the width of the corresponding subinterval.



Weak Laws and Strong Laws

There are other concepts of the convergence of a sequence of random variables, in addition to the concept of convergence in probability that has been presented above. For example, it is said that a sequence Z_1, Z_2, \dots converges to a constant b with probability 1 if

$$\Pr\left(\lim_{n \rightarrow \infty} Z_n = b\right) = 1.$$

A careful investigation of the concept of convergence with probability 1 is beyond the scope of this book. It can be shown that if a sequence Z_1, Z_2, \dots converges to b with probability 1, then the sequence will also converge to b in probability. For this reason, convergence with probability 1 is often called *strong convergence*, whereas convergence in probability is called *weak convergence*. In order to emphasize the distinction between these two concepts of convergence, the result that here has been called simply the law of large numbers is often called the *weak law of large numbers*. The *strong law of large numbers* can then be stated as follows: If \bar{X}_n is the sample mean of a random sample of size n from a distribution with mean μ , then

$$\Pr\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1.$$

The proof of this result will not be given here. There are examples of sequences of random variables that converge in probability but that do not converge with probability 1. Exercise 22 is one such example. Another type of converges is *convergence in quadratic mean*, which is introduced in Exercises 10–13.





Chernoff Bounds

One way to think of the Chebyshev inequality is as an application of the Markov inequality to the random variable $(X - \mu)^2$. This idea generalizes to other functions and leads to a sharper bound on the probability in the tail of a distribution when the bound applies. Before giving the general result, we give a simple example to illustrate the potential improvement that it can provide.

**Example
6.2.6**

Binomial Random Variable. Suppose that X has the binomial distribution with parameters n and $1/2$. We would like a bound to the probability that X/n is far from its mean $1/2$. To be specific, suppose that we would like a bound for

$$\Pr\left(\left|\frac{X}{n} - \frac{1}{2}\right| \geq \frac{1}{10}\right). \quad (6.2.7)$$

The Chebyshev inequality gives the bound $\text{Var}(X/n)/(1/10)^2$, which equals $25/n$.

Instead of applying the Chebyshev inequality, define $Y = X - n/2$ and rewrite the probability in (6.2.7) as the sum of the following two probabilities:

$$\begin{aligned} \Pr\left(\frac{X}{n} \geq \frac{1}{2} + \frac{1}{10}\right) &= \Pr\left(Y \geq \frac{n}{10}\right), \quad \text{and} \\ \Pr\left(\frac{X}{n} \leq \frac{1}{2} - \frac{1}{10}\right) &= \Pr\left(-Y \geq \frac{n}{10}\right). \end{aligned} \quad (6.2.8)$$

For each $s > 0$, rewrite the first of the probabilities in (6.2.8) as

$$\begin{aligned} \Pr\left(Y \geq \frac{n}{10}\right) &= \Pr\left[\exp(sY) \geq \exp\left(\frac{ns}{10}\right)\right] \\ &\leq \frac{E[\exp(sY)]}{\exp(ns/10)}, \end{aligned}$$

where the inequality follows from the Markov inequality. This equation involves the moment generating function of Y , $\psi(s) = E[\exp(sY)]$. The m.g.f. of Y can be found by applying Theorem 4.4.3 with $p = 1/2$, $a = 1$, and $b = -n/2$ together with Equation (5.2.4). The result is

$$\psi(s) = \left(\frac{1}{2} [\exp(s) + 1] \exp(-s/2)\right)^n, \quad (6.2.9)$$

for all s . Let $s = 1/2$ in (6.2.9) to obtain the bound

$$\begin{aligned} \Pr\left(Y \geq \frac{n}{10}\right) &\leq \psi(1/2) \exp(-n/20) \\ &= \exp(-n/20) \left(\frac{1}{2} [\exp(1/2) + 1] \exp(-1/4)\right)^n = 0.9811^n. \end{aligned}$$

Similarly, we can write the second probability in (6.2.8) as

$$\Pr\left(-Y \geq \frac{n}{10}\right) = \Pr\left[\exp(-sY) \geq \exp\left(\frac{ns}{10}\right)\right], \quad (6.2.10)$$

where $s > 0$. The m.g.f. of $-Y$ is $\psi(-s)$. Let $s = 1/2$ in (6.2.10) and apply the Markov inequality to obtain the bound

$$\begin{aligned}\Pr\left(-Y \geq \frac{n}{10}\right) &\leq \psi(-1/2) \exp(-n/20) \\ &= \exp(-n/20) \left(\frac{1}{2} [\exp(-1/2) + 1] \exp(1/4)\right)^n = 0.9811^n.\end{aligned}$$

Hence, we obtain the bound

$$\Pr\left(\left|\frac{X}{n} - \frac{1}{2}\right| \geq \frac{1}{10}\right) \leq 2(0.9811)^n. \quad (6.2.11)$$

The bound in (6.2.11) decreases exponentially fast as n increases, while the Chebychev bound $25/n$ decreases proportionally to $1/n$. For example, with $n = 100, 200, 300$, the Chebychev bounds are 0.25, 0.125, and 0.0833. The corresponding bounds from (6.2.11) are 0.2967, 0.0440, and 0.0065. \blacktriangleleft

The choice of $s = 1/2$ in Example 6.2.6 was arbitrary. Theorem 6.2.7 says that we can replace this arbitrary choice with the choice that leads to the smallest possible bound. The proof of Theorem 6.2.7 is a straightforward application of the Markov inequality. (See Exercise 18 in this section.)

**Theorem
6.2.7**

Chernoff Bounds. Let X be a random variable with moment generating function ψ . Then, for every real t ,

$$\Pr(X \geq t) \leq \min_{s>0} \exp(-st)\psi(s). \quad \blacksquare$$

Theorem 6.2.7 is most useful when X is the sum of n i.i.d. random variables each with finite m.g.f. and when $t = nu$ for a large value of n and some fixed u . This was the case in Example 6.2.6.

**Example
6.2.7**

Average of Geometric Random Sample. Suppose that X_1, X_2, \dots are i.i.d. geometric random variables with parameter p . We would like a bound to the probability that \bar{X}_n is far from the mean $(1-p)/p$. To be specific, for each fixed $u > 0$, we would like a bound for

$$\Pr\left(\left|\bar{X}_n - \frac{1-p}{p}\right| \geq u\right). \quad (6.2.12)$$

Let $X = \sum_{i=1}^n X_i - n(1-p)/p$. For each $u > 0$, Theorem 6.2.7 can be used to bound both

$$\begin{aligned}\Pr\left(\bar{X}_n \geq \frac{1-p}{p} + u\right) &= \Pr(X \geq nu), \quad \text{and} \\ \Pr\left(\bar{X}_n \leq \frac{1-p}{p} - u\right) &= \Pr(-X \geq nu).\end{aligned}$$

Since (6.2.12) equals $\Pr(X \geq nu) + \Pr(-X \geq nu)$, the bound we seek is the sum of the two bounds that we get for $\Pr(X \geq nu)$ and $\Pr(-X \geq nu)$.

The m.g.f. of X can be found by applying Theorem 4.4.3 with $a = 1$ and $b = -n(1-p)/p$ together with Theorem 5.5.3. The result is

$$\psi(s) = \left(\frac{p \exp[-s(1-p)/p]}{1 - (1-p) \exp(s)}\right)^n. \quad (6.2.13)$$

The m.g.f. of $-X$ is $\psi(-s)$. According to Theorem 6.2.7,

$$\Pr(X \geq nu) \leq \min_{s>0} \psi(s) \exp(-snu). \quad (6.2.14)$$

We find the minimum of $\psi(s) \exp(-snu)$ by finding the minimum of its logarithm. Using (6.2.13), we get that

$$\log[\psi(s) \exp(-snu)] = n \left\{ \log(p) - s \frac{1-p}{p} - \log[1 - (1-p) \exp(s)] - su \right\}.$$

The derivative of this expression with respect to s equals 0 at

$$s = -\log \left[\frac{(1+u)p + 1-p}{up + 1-p} (1-p) \right], \quad (6.2.15)$$

and the second derivative is positive. If $u > 0$, then the value of s in (6.2.15) is positive and $\psi(s)$ is finite. Hence, the value of s in (6.2.15) provides the minimum in (6.2.14). That minimum can be expressed as q^n where

$$q = [p(1+u) + 1-p] \left[\frac{(1+u)p + 1-p}{up + 1-p} (1-p) \right]^{u+(1-p)/p} \quad (6.2.16)$$

and $0 < q < 1$. (See Exercise 19 for a proof.) Hence, $\Pr(X \geq nu) \leq q^n$.

For $\Pr(-X \geq nu)$, we notice first that $\Pr(-X \geq nu) = 0$ if $u \geq (1-p)/p$ because $\sum_{i=1}^n X_i \geq 0$. If $u \geq (1-p)/p$, then the overall bound on (6.2.12) is q^n . For $0 < u < (1-p)/p$, the value of s that minimizes $\psi(-s) \exp(-snu)$ is

$$s = -\log \left[\frac{(1-u)p + 1-p}{1-p-up} (1-p) \right],$$

which is positive when $0 < u < (1-p)/p$. The value of $\min_{s>0} \psi(-s) \exp(-snu)$ is r^n , where

$$r = [p(1-u) + 1-p] \left[\frac{(1-u)p + 1-p}{1-p-up} (1-p) \right]^{-u+(1-p)/p}$$

and $0 < r < 1$. Hence, the Chernoff bound is q^n if $u \geq (1-p)/p$ and is $q^n + r^n$ if $0 < u < (1-p)/p$. As such, the bound decreases exponentially fast as n increases. This is a marked improvement over the Chebyshev bound, which decreases like a constant over n . ◀



Summary

The law of large numbers says that the sample mean of a random sample converges in probability to the mean μ of the individual random variables, if the variance exists. This means that the sample mean will be close to μ if the size of the random sample is sufficiently large. The Chebyshev inequality provides a (crude) bound on how high the probability is that the sample mean will be close to μ . Chernoff bounds can be sharper, but are harder to compute.

Exercises

1. For each integer n , let X_n be a nonnegative random variable with finite mean μ_n . Prove that if $\lim_{n \rightarrow \infty} \mu_n = 0$, then $X_n \xrightarrow{P} 0$.
2. Suppose that X is a random variable for which $\Pr(X \geq 0) = 1$ and $\Pr(X \geq 10) = 1/5$. Prove that $E(X) \geq 2$.
3. Suppose that X is a random variable for which $E(X) = 10$, $\Pr(X \leq 7) = 0.2$, and $\Pr(X \geq 13) = 0.3$. Prove that $\text{Var}(X) \geq 9/2$.
4. Let X be a random variable for which $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Construct a probability distribution for X such that $\Pr(|X - \mu| \geq 3\sigma) = 1/9$.

5. How large a random sample must be taken from a given distribution in order for the probability to be at least 0.99 that the sample mean will be within 2 standard deviations of the mean of the distribution?

6. Suppose that X_1, \dots, X_n form a random sample of size n from a distribution for which the mean is 6.5 and the variance is 4. Determine how large the value of n must be in order for the following relation to be satisfied:

$$\Pr(6 \leq \bar{X}_n \leq 7) \geq 0.8.$$

7. Suppose that X is a random variable for which $E(X) = \mu$ and $E[(X - \mu)^4] = \beta_4$. Prove that

$$\Pr(|X - \mu| \geq t) \leq \frac{\beta_4}{t^4}.$$

8. Suppose that 30 percent of the items in a large manufactured lot are of poor quality. Suppose also that a random sample of n items is to be taken from the lot, and let Q_n denote the proportion of the items in the sample that are of poor quality. Find a value of n such that $\Pr(0.2 \leq Q_n \leq 0.4) \geq 0.75$ by using **(a)** the Chebyshev inequality and **(b)** the tables of the binomial distribution at the end of this book.

9. Let Z_1, Z_2, \dots be a sequence of random variables, and suppose that, for $n = 1, 2, \dots$, the distribution of Z_n is as follows:

$$\Pr(Z_n = n^2) = \frac{1}{n} \quad \text{and} \quad \Pr(Z_n = 0) = 1 - \frac{1}{n}.$$

Show that

$$\lim_{n \rightarrow \infty} E(Z_n) = \infty \quad \text{but} \quad Z_n \xrightarrow{P} 0.$$

10. It is said that a sequence of random variables Z_1, Z_2, \dots converges to a constant b in quadratic mean if

$$\lim_{n \rightarrow \infty} E[(Z_n - b)^2] = 0. \quad (6.2.17)$$

Show that Eq. (6.2.17) is satisfied if and only if

$$\lim_{n \rightarrow \infty} E(Z_n) = b \quad \text{and} \quad \lim_{n \rightarrow \infty} \text{Var}(Z_n) = 0.$$

Hint: Use Exercise 5 of Sec. 4.3.

11. Prove that if a sequence Z_1, Z_2, \dots converges to a constant b in quadratic mean, then the sequence also converges to b in probability.

12. Let \bar{X}_n be the sample mean of a random sample of size n from a distribution for which the mean is μ and the variance is σ^2 , where $\sigma^2 < \infty$. Show that \bar{X}_n converges to μ in quadratic mean as $n \rightarrow \infty$.

13. Let Z_1, Z_2, \dots be a sequence of random variables, and suppose that for $n = 2, 3, \dots$, the distribution of Z_n is as follows:

$$\Pr\left(Z_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2} \quad \text{and} \quad \Pr(Z_n = n) = \frac{1}{n^2}.$$

a. Does there exist a constant c to which the sequence converges in probability?

b. Does there exist a constant c to which the sequence converges in quadratic mean?

14. Let f be a p.f. for a discrete distribution. Suppose that $f(x) = 0$ for $x \notin [0, 1]$. Prove that the variance of this distribution is at most 1/4. *Hint:* Prove that there is a distribution supported on just the two points $\{0, 1\}$ that has variance at least as large as f does and then prove that the variance of a distribution supported on $\{0, 1\}$ is at most 1/4.

15. Prove Theorem 6.2.5.

16. Suppose that $Z_n \xrightarrow{P} b$, $Y_n \xrightarrow{P} c$, and $g(z, y)$ is a function that is continuous at $(z, y) = (b, c)$. Prove that $g(Z_n, Y_n)$ converges in probability to $g(b, c)$.

17. Let X have the binomial distribution with parameters n and p . Let Y have the binomial distribution with parameters n and p/k with $k > 1$. Let $Z = kY$.

- a.** Show that X and Z have the same mean.
- b.** Find the variances of X and Z . Show that, if p is small, then the variance of Z is approximately k times as large as the variance of X .
- c.** Show why the results above explain the higher variability in the bar heights in Fig. 6.2 compared to Fig. 6.1.

18. Prove Theorem 6.2.7.

19. Return to Example 6.2.7.

- a.** Prove that the $\min_{s > 0} \psi(s) \exp(-snu)$ equals q^n , where q is given in (6.2.16).
- b.** Prove that $0 < q < 1$. *Hint:* First, show that $0 < q < 1$ if $u = 0$. Next, let $x = up + 1 - p$ and show that $\log(q)$ is a decreasing function of x .

20. Return to Example 6.2.6. Find the Chernoff bound for the probability in (6.2.7).

21. Let X_1, X_2, \dots be a sequence of i.i.d. random variables having the exponential distribution with parameter 1. Let $Y_n = \sum_{i=1}^n X_i$ for each $n = 1, 2, \dots$

- a.** For each $u > 1$, compute the Chernoff bound on $\Pr(Y_n > nu)$.
- b.** What goes wrong if we try to compute the Chernoff bound when $u < 1$?

22. In this exercise, we construct an example of a sequence of random variables Z_n such that $Z_n \xrightarrow{P} 0$ but

$$\Pr\left(\lim_{n \rightarrow \infty} Z_n = 0\right) = 0. \quad (6.2.18)$$

That is, Z_n converges in probability to 0, but Z_n does not converge to 0 with probability 1. Indeed, Z_n converges to 0 with probability 0.

Let X be a random variable having the uniform distribution on the interval $[0, 1]$. We will construct a sequence of functions $h_n(x)$ for $n = 1, 2, \dots$ and define $Z_n = h_n(X)$. Each function h_n will take only two values, 0 and 1. The set of x where $h_n(x) = 1$ is determined by dividing the interval $[0, 1]$ into k nonoverlapping subintervals of length $1/k$ for $k = 1, 2, \dots$, arranging these intervals in sequence, and letting $h_n(x) = 1$ on the n th interval in the sequence for $n = 1, 2, \dots$. For each k , there are k nonoverlapping subintervals, so the number of subintervals with lengths $1, 1/2, 1/3, \dots, 1/k$ is

$$1 + 2 + 3 + \dots + k = \frac{k(k+1)}{2}.$$

The remainder of the construction is based on this formula. The first interval in the sequence has length 1, the next two have length $1/2$, the next three have length $1/3$, etc.

- a. For each $n = 1, 2, \dots$, prove that there is a unique positive integer k_n such that

$$\frac{(k_n - 1)k_n}{2} < n \leq \frac{k_n(k_n + 1)}{2}.$$

- b. For each $n = 1, 2, \dots$, let $j_n = n - (k_n - 1)k_n/2$. Show that j_n takes the values $1, \dots, k_n$ as n runs through $1 + (k_n - 1)k_n/2, \dots, k_n(k_n + 1)/2$.

- c. Define

$$h_n(x) = \begin{cases} 1 & \text{if } (j_n - 1)/k_n \leq x < j_n/k_n, \\ 0 & \text{if not.} \end{cases}$$

Show that, for every $x \in [0, 1]$, $h_n(x) = 1$ for one and only one n among $1 + (k_n - 1)k_n/2, \dots, k_n(k_n + 1)/2$.

- d. Show that $Z_n = h_n(X)$ takes the value 1 infinitely often with probability 1.
e. Show that (6.2.18) holds.
f. Show that $\Pr(Z_n = 0) = 1 - 1/k_n$ and $\lim_{n \rightarrow \infty} k_n = \infty$.
g. Show that $Z_n \xrightarrow{P} 0$.

23. Prove that the sequence of random variables Z_n in Exercise 22 converges in quadratic mean (definition in Exercise 10) to 0.

24. In this exercise, we construct an example of a sequence of random variables Z_n such that Z_n converges to 0 with probability 1, but Z_n fails to converge to 0 in quadratic mean. Let X be a random variable having the uniform distribution on the interval $[0, 1]$. Define the sequence Z_n by $Z_n = n^2$ if $0 < X < 1/n$ and $Z_n = 0$ otherwise.

- a. Prove that Z_n converges to 0 with probability 1.
b. Prove that Z_n does not converge to 0 in quadratic mean.

6.3 The Central Limit Theorem

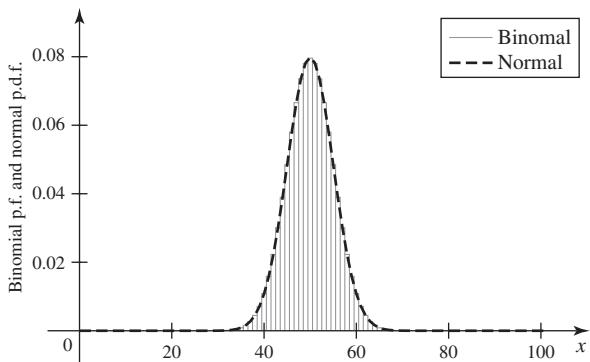
The sample mean of a large random sample of random variables with mean μ and finite variance σ^2 has approximately the normal distribution with mean μ and variance σ^2/n . This result helps to justify the use of the normal distribution as a model for many random variables that can be thought of as being made up of many independent parts. Another version of the central limit theorem is given that applies to independent random variables that are not identically distributed. We also introduce the delta method, which allows us to compute approximate distributions for functions of random variables.

Statement of the Theorem

Example 6.3.1

A Large Sample. A clinical trial has 100 patients who will receive a treatment. Patients who don't receive the treatment survive for 18 months with probability 0.5 each. We assume that all patients are independent. The trial is to see whether the new treatment can increase the probability of survival significantly. Let X be the number of patients out of the 100 who survive for 18 months. If the probability of success were 0.5 for the patients on the treatment (the same as without the treatment), then X would have the binomial distribution with parameters $n = 100$ and $p = 0.5$. The p.f. of X is graphed as a bar chart with the solid line in Fig. 6.3. The shape of the bar chart is reminiscent of a bell-shaped curve. The normal p.d.f. with the same mean $\mu = 50$ and variance $\sigma^2 = 25$ as the binomial distribution is also graphed with the dotted line. ◀

Figure 6.3 Comparison of the binomial p.f. with parameters 100 and 0.5 to the normal p.d.f. with mean 50 and variance 25.



In Examples 5.4.1 and 5.4.2, we illustrated how the Poisson distribution provides a good approximation to a binomial distribution with a large n and small p . Example 6.3.1 shows how a normal distribution can be a good approximation to a binomial distribution with a large n and not so small p . The central limit theorem (Theorem 6.3.1) is a formal statement of how normal distributions can approximate distributions of general sums or averages of i.i.d. random variables.

In Corollary 5.6.2, we saw that if a random sample of size n is taken from the normal distribution with mean μ and variance σ^2 , then the sample average \bar{X}_n has the normal distribution with mean μ and variance σ^2/n . The simple version of the central limit theorem that we give in this section says that whenever a random sample of size n is taken from *any* distribution with mean μ and variance σ^2 , the sample average \bar{X}_n will have a distribution that is *approximately* normal with mean μ and variance σ^2/n .

This result was established for a random sample from a Bernoulli distribution by A. de Moivre in the early part of the eighteenth century. The proof for a random sample from an arbitrary distribution was given independently by J. W. Lindeberg and P. Lévy in the early 1920s. A precise statement of their theorem will be given now, and an outline of the proof of that theorem will be given later in this section. We shall also state another central limit theorem pertaining to the sum of independent random variables that are not necessarily identically distributed and shall present some examples illustrating both theorems.

Theorem

6.3.1

Central Limit Theorem (Lindeberg and Lévy). If the random variables X_1, \dots, X_n form a random sample of size n from a given distribution with mean μ and variance σ^2 ($0 < \sigma^2 < \infty$), then for each fixed number x ,

$$\lim_{n \rightarrow \infty} \Pr \left[\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right] = \Phi(x), \quad (6.3.1)$$

where Φ denotes the c.d.f. of the standard normal distribution. ■

The interpretation of Eq. (6.3.1) is as follows: If a large random sample is taken from any distribution with mean μ and variance σ^2 , regardless of whether this distribution is discrete or continuous, then the distribution of the random variable $n^{1/2}(\bar{X}_n - \mu)/\sigma$ will be approximately the standard normal distribution. Therefore, the distribution of \bar{X}_n will be approximately the normal distribution with mean μ and variance σ^2/n , or, equivalently, the distribution of the sum $\sum_{i=1}^n X_i$ will be

approximately the normal distribution with mean $n\mu$ and variance $n\sigma^2$. It is in this last form that the central limit theorem was illustrated in Example 6.3.1.

**Example
6.3.2**

Tossing a Coin. Suppose that a fair coin is tossed 900 times. We shall approximate the probability of obtaining more than 495 heads.

For $i = 1, \dots, 900$, let $X_i = 1$ if a head is obtained on the i th toss and let $X_i = 0$ otherwise. Then $E(X_i) = 1/2$ and $\text{Var}(X_i) = 1/4$. Therefore, the values X_1, \dots, X_{900} form a random sample of size $n = 900$ from a distribution with mean $1/2$ and variance $1/4$. It follows from the central limit theorem that the distribution of the total number of heads $H = \sum_{i=1}^{900} X_i$ will be approximately the normal distribution for which the mean is $(900)(1/2) = 450$, the variance is $(900)(1/4) = 225$, and the standard deviation is $(225)^{1/2} = 15$. Therefore, the variable $Z = (H - 450)/15$ will have approximately the standard normal distribution. Thus,

$$\begin{aligned}\Pr(H > 495) &= \Pr\left(\frac{H - 450}{15} > \frac{495 - 450}{15}\right) \\ &= \Pr(Z > 3) \approx 1 - \Phi(3) = 0.0013.\end{aligned}$$

◀

The exact probability 0.0012 to four decimal places.

**Example
6.3.3**

Sampling from a Uniform Distribution. Suppose that a random sample of size $n = 12$ is taken from the uniform distribution on the interval $[0, 1]$. We shall approximate the value of $\Pr\left(\left|\bar{X}_n - \frac{1}{2}\right| \leq 0.1\right)$.

The mean of the uniform distribution on the interval $[0, 1]$ is $1/2$, and the variance is $1/12$ (see Exercise 3 of Sec. 4.3). Since $n = 12$ in this example, it follows from the central limit theorem that the distribution of \bar{X}_n will be approximately the normal distribution with mean $1/2$ and variance $1/144$. Therefore, the distribution of the variable $Z = 12\left(\bar{X}_n - \frac{1}{2}\right)$ will be approximately the standard normal distribution. Hence,

$$\begin{aligned}\Pr\left(\left|\bar{X}_n - \frac{1}{2}\right| \leq 0.1\right) &= \Pr\left[12\left|\bar{X}_n - \frac{1}{2}\right| \leq 1.2\right] \\ &= \Pr(|Z| \leq 1.2) \approx 2\Phi(1.2) - 1 = 0.7698.\end{aligned}$$

For the special case of $n = 12$, the random variable Z has the form $Z = \sum_{i=1}^{12} X_i - 6$. At one time, some computers produced standard normal pseudo-random numbers by adding 12 uniform pseudo-random numbers and subtracting 6. ▶

**Example
6.3.4**

Poisson Random Variables. Suppose that X_1, \dots, X_n form a random sample from the Poisson distribution with mean θ . Let \bar{X}_n be the average. Then $\mu = \theta$ and $\sigma^2 = \theta$. The central limit theorem says that $n^{1/2}(\bar{X}_n - \theta)/\theta^{1/2}$ has approximately the standard normal distribution. In particular, the central limit theorem says that \bar{X}_n should be close to μ with high probability. The probability that $|\bar{X}_n - \theta|$ is less than some small number c could be approximated using the standard normal c.d.f.:

$$\Pr(|\bar{X}_n - \theta| < c) \approx 2\Phi\left(cn^{1/2}\theta^{-1/2}\right) - 1. \quad (6.3.2)$$

◀

The type of convergence that appears in the central limit theorem, specifically, Eq. (6.3.1), arises in other contexts and has a special name.

**Definition
6.3.1**

Convergence in Distribution/Asymptotic Distribution. Let X_1, X_2, \dots be a sequence of random variables, and for $n = 1, 2, \dots$, let F_n denote the c.d.f. of X_n . Also, let F^* be a c.d.f. Then it is said that the sequence X_1, X_2, \dots converges in distribution to F^* if

$$\lim_{n \rightarrow \infty} F_n(x) = F^*(x), \quad (6.3.3)$$

for all x at which $F^*(x)$ is continuous. Sometimes, it is simply said that X_n converges in distribution to F^* , and F^* is called the *asymptotic distribution* of X_n . If F^* has a name, then we say that X_n converges in distribution to that name.

Thus, according to Theorem 6.3.1, as indicated in Eq. (6.3.1), the random variable $n^{1/2}(\bar{X}_n - \mu)/\sigma$ converges in distribution to the standard normal distribution, or, equivalently, the asymptotic distribution of $n^{1/2}(\bar{X}_n - \mu)/\sigma$ is the standard normal distribution.

Effect of the Central Limit Theorem The central limit theorem provides a plausible explanation for the fact that the distributions of many random variables studied in physical experiments are approximately normal. For example, a person's height is influenced by many random factors. If the height of each person is determined by adding the values of these individual factors, then the distribution of the heights of a large number of persons will be approximately normal. In general, the central limit theorem indicates that the distribution of the sum of many random variables can be approximately normal, even though the distribution of each random variable in the sum differs from the normal.

**Example
6.3.5**

Determining a Simulation Size. In Example 6.2.2 on page 351, an environmental engineer wanted to determine the size of a simulation to estimate the mean proportion of water contaminant that was lead. Use of the Chebyshev inequality in that example suggested that a simulation of size 2,000,000 will guarantee that the estimate will be less than 0.005 away from the true mean proportion with probability at least 0.98. In this example, we shall use the central limit theorem to determine a much smaller simulation size that should still provide the same accuracy bound. The estimate of the mean proportion will be the average \bar{R}_n of all of the simulated proportions R_1, \dots, R_n from the n simulations that will be run. As we noted in Example 6.2.2, the variance of each R_i is $\sigma^2 \leq 1$, and hence the central limit theorem says that \bar{R}_n has approximately the normal distribution with mean equal to the true mean proportion $E(R_i)$ and variance at most $1/n$. Since the probability of being close to the mean decreases as the variance increases, we see that

$$\begin{aligned} \Pr(|\bar{R}_n - E(R_i)| < 0.005) &\approx \Phi\left(\frac{0.005}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-0.005}{\sigma/\sqrt{n}}\right) \\ &\geq \Phi\left(\frac{0.005}{1/\sqrt{n}}\right) - \Phi\left(\frac{-0.005}{1/\sqrt{n}}\right) \\ &= 2\Phi(0.005\sqrt{n}) - 1. \end{aligned}$$

If we set $2\Phi(0.005\sqrt{n}) - 1 = 0.98$, we obtain

$$n = \frac{1}{0.005^2} \Phi^{-1}(0.99)^2 = 40,000 \times 2.326^2 = 216,411.$$

That is, we only need a little more than 10 percent of the simulation size that the Chebyshev inequality suggested. (Since σ^2 is actually no more than $1/4$, we really only need $n = 54,103$. See Exercise 14 in Sec. 6.2 for a proof that a discrete distribution on

the interval $[0, 1]$ can have variance at most $1/4$. The continuous case is slightly more complicated, but also true.) \blacktriangleleft

Other Examples of Convergence in Distribution In Chapter 5, we saw three examples of limit theorems involving discrete distributions. Theorems 5.3.4, 5.4.5, and 5.4.6 all showed that a sequence of p.f.'s converged to some other p.f. In Exercise 7 in Sec. 6.5, you can prove a general result that implies that the three theorems just mentioned are examples of convergence in distribution.

The Delta Method

Example
6.3.6

Rate of Service. Customers arrive at a queue for service, and the i th customer is served in some time X_i after reaching the head of the queue. If we assume that X_1, \dots, X_n form a random sample of service times with mean μ and finite variance σ^2 , we might be interested in using $1/\bar{X}_n$ to estimate the rate of service. The central limit theorem tells us something about the approximate distribution of \bar{X}_n if n is large, but what can we say about the distribution of $1/\bar{X}_n$? \blacktriangleleft

Suppose that X_1, \dots, X_n form a random sample from a distribution that has finite mean μ and finite variance σ^2 . The central limit theorem says that $n^{1/2}(\bar{X}_n - \mu)/\sigma$ has approximately the standard normal distribution. Now suppose that we are interested in the distribution of some function α of \bar{X}_n . We shall assume that α is a differentiable function whose derivative is nonzero at μ . We shall approximate the distribution of $\alpha(\bar{X}_n)$ by a method known in statistics as the *delta method*.

Theorem
6.3.2

Delta Method. Let Y_1, Y_2, \dots be a sequence of random variables, and let F^* be a continuous c.d.f. Let θ be a real number, and let a_1, a_2, \dots be a sequence of positive numbers that increase to ∞ . Suppose that $a_n(Y_n - \theta)$ converges in distribution to F^* . Let α be a function with continuous derivative such that $\alpha'(\theta) \neq 0$. Then $a_n[\alpha(Y_n) - \alpha(\theta)]/\alpha'(\theta)$ converges in distribution to F^* .

Proof We shall give only an outline of the proof. Because $a_n \rightarrow \infty$, Y_n must get close to θ with high probability as $n \rightarrow \infty$. If not, $|a_n(Y_n - \theta)|$ would go to ∞ with nonzero probability and then the c.d.f. of $a_n(Y_n - \theta)$ would not converge to a c.d.f. Because α is continuous, $\alpha(Y_n)$ must also be close to $\alpha(\theta)$ with high probability. Therefore, we shall use a Taylor series expansion of $\alpha(Y_n)$ around θ ,

$$\alpha(Y_n) \approx \alpha(\theta) + \alpha'(\theta)(Y_n - \theta), \quad (6.3.4)$$

where we have ignored all terms involving $(Y_n - \theta)^2$ and higher powers. Subtract $\alpha(\theta)$ from both sides of Eq. (6.3.4), and then multiply both sides by $a_n/\alpha'(\theta)$ to get

$$\frac{a_n}{\alpha'(\theta)}(Y_n - \theta) \approx a_n(Y_n - \theta). \quad (6.3.5)$$

We then conclude that the distribution of the left side of Eq. (6.3.5) will be approximately the same as the distribution of the right side of the equation, which is approximately F^* . \blacksquare

The most common application of Theorem 6.3.2 occurs when Y_n is the average of a random sample from a distribution with finite variance. We state that case in the following corollary.

Corollary
6.3.1

Delta Method for Average of a Random Sample. Let X_1, X_2, \dots be a sequence of i.i.d. random variables from a distribution with mean μ and finite variance σ^2 . Let α

be a function with continuous derivative such that $\alpha'(\mu) \neq 0$. Then the asymptotic distribution of

$$\frac{n^{1/2}}{\sigma\alpha'(\mu)}[\alpha(\bar{X}_n) - \alpha(\mu)]$$

is the standard normal distribution.

Proof Apply Theorem 6.3.2 with $Y_n = \bar{X}_n$, $a_n = n^{1/2}/\sigma$, $\theta = \mu$, and F^* being the standard normal c.d.f. \blacksquare

A common way to report the result in Corollary 6.3.1 is to say that the distribution of $\alpha(\bar{X}_n)$ is approximately the normal distribution with mean $\alpha(\mu)$ and variance $\sigma^2[\alpha'(\mu)]^2/n$.

**Example
6.3.7**

Rate of Service. In Example 6.3.6, we are interested in the distribution of $\alpha(\bar{X}_n)$ where $\alpha(x) = 1/x$ for $x > 0$. We can apply the delta method by finding $\alpha'(x) = -1/x^2$. It follows that the asymptotic distribution of

$$-\frac{n^{1/2}\mu^2}{\sigma} \left(\frac{1}{\bar{X}_n} - \frac{1}{\mu} \right)$$

is the standard normal distribution. Alternatively, we might say that $1/\bar{X}_n$ has approximately the normal distribution with mean $1/\mu$ and variance $\sigma^2/[n\mu^4]$. \blacktriangleleft

Variance Stabilizing Transformations If we were to observe a random sample of Poisson random variables as in Example 6.3.4, we would assume that θ is unknown. In such a case we cannot compute the probability in Eq. (6.3.2), because the approximate variance of \bar{X}_n depends on θ . For this reason, it is sometimes desirable to transform \bar{X}_n by a function α so that the approximate distribution of $\alpha(\bar{X}_n)$ has a variance that is a known value. Such a function is called a *variance stabilizing transformation*. We can often find a variance stabilizing transformation by running the delta method in reverse. In general, we note that the approximate distribution of $\alpha(\bar{X}_n)$ has variance $\alpha'(\mu)^2\sigma^2/n$. In order to make this variance constant, we need $\alpha'(\mu)$ to be a constant times $1/\sigma$. If σ^2 is a function $g(\mu)$, then we achieve this goal by letting

$$\alpha(\mu) = \int_a^\mu \frac{dx}{g(x)^{1/2}}, \quad (6.3.6)$$

where a is an arbitrary constant that makes the integral finite.

**Example
6.3.8**

Poisson Random Variables. In Example 6.3.4, we have $\sigma^2 = \theta = \mu$, so that $g(\mu) = \mu$. According to Eq. (6.3.6), we should let

$$\alpha(\mu) = \int_0^\mu \frac{dx}{x^{1/2}} = 2\mu^{1/2}.$$

It follows that $2\bar{X}_n^{1/2}$ has approximately the normal distribution with mean $2\theta^{1/2}$ and variance $1/n$. For each number $c > 0$, we have

$$\Pr(|2\bar{X}_n^{1/2} - 2\theta^{1/2}| < c) \approx 2\Phi(cn^{1/2}) - 1. \quad (6.3.7)$$

In Chapter 8, we shall see how to use Eq (6.3.7) to estimate θ when we assume that θ is unknown. \blacktriangleleft



The Central Limit Theorem (Liapounov) for the Sum of Independent Random Variables

We shall now state a central limit theorem that applies to a sequence of random variables X_1, X_2, \dots that are independent but not necessarily identically distributed. This theorem was first proved by A. Liapounov in 1901. We shall assume that $E(X_i) = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$ for $i = 1, \dots, n$. Also, we shall let

$$Y_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}}. \quad (6.3.8)$$

Then $E(Y_n) = 0$ and $\text{Var}(Y_n) = 1$. The theorem that is stated next gives a sufficient condition for the distribution of this random variable Y_n to be approximately the standard normal distribution.

Theorem 6.3.3 Suppose that the random variables X_1, X_2, \dots are independent and that $E(|X_i - \mu_i|^3) < \infty$ for $i = 1, 2, \dots$ Also, suppose that

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E(|X_i - \mu_i|^3)}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0. \quad (6.3.9)$$

Finally, let the random variable Y_n be as defined in Eq. (6.3.8). Then, for each fixed number x ,

$$\lim_{n \rightarrow \infty} \Pr(Y_n \leq x) = \Phi(x). \quad (6.3.10)$$

■

The interpretation of this theorem is as follows: If Eq. (6.3.9) is satisfied, then for every large value of n , the distribution of $\sum_{i=1}^n X_i$ will be approximately the normal distribution with mean $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2$. It should be noted that when the random variables X_1, X_2, \dots are identically distributed and the third moments of the variables exist, Eq. (6.3.9) will automatically be satisfied and Eq. (6.3.10) then reduces to Eq. (6.3.1).

The distinction between the theorem of Lindeberg and Lévy and the theorem of Liapounov should be emphasized. The theorem of Lindeberg and Lévy applies to a sequence of i.i.d. random variables. In order for this theorem to be applicable, it is sufficient to assume only that the variance of each random variable is finite. The theorem of Liapounov applies to a sequence of independent random variables that are not necessarily identically distributed. In order for this theorem to be applicable, it must be assumed that the third moment of each random variable is finite and satisfies Eq. (6.3.9).

The Central Limit Theorem for Bernoulli Random Variables By applying the theorem of Liapounov, we can establish the following result.

Theorem 6.3.4 Suppose that the random variables X_1, \dots, X_n are independent and X_i has the Bernoulli distribution with parameter p_i ($i = 1, 2, \dots$). Suppose also that the infinite series $\sum_{i=1}^{\infty} p_i(1 - p_i)$ is divergent, and let

$$Y_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n p_i}{\left(\sum_{i=1}^n p_i(1 - p_i)\right)^{1/2}}. \quad (6.3.11)$$

Then for every fixed number x ,

$$\lim_{n \rightarrow \infty} \Pr(Y_n \leq x) = \Phi(x). \quad (6.3.12)$$

Proof Here $\Pr(X_i = 1) = p_i$ and $\Pr(X_i = 0) = 1 - p_i$. Therefore,

$$\begin{aligned} E(X_i) &= p_i, \quad \text{Var}(X_i) = p_i(1 - p_i), \\ E(|X_i - p_i|^3) &= p_i(1 - p_i)^3 + (1 - p_i)p_i^3 = p_i(1 - p_i)(p_i^2 + (1 - p_i)^2) \\ &\leq p_i(1 - p_i), \end{aligned} \quad (6.3.13)$$

It follows that

$$\frac{\sum_{i=1}^n E(|X_i - p_i|^3)}{(\sum_{i=1}^n p_i(1 - p_i))^{3/2}} \leq \frac{1}{(\sum_{i=1}^n p_i(1 - p_i))^{1/2}}. \quad (6.3.14)$$

Since the infinite series $\sum_{i=1}^{\infty} p_i(1 - p_i)$ is divergent, then $\sum_{i=1}^n p_i(1 - p_i) \rightarrow \infty$ as $n \rightarrow \infty$, and it can be seen from the relation (6.3.14) that Eq. (6.3.9) will be satisfied. In turn, it follows from Theorem 6.3.3 that Eq. (6.3.10) will be satisfied. Since Eq. (6.3.12) is simply a restatement of Eq. (6.3.10) for the particular random variables being considered here, the proof of the theorem is complete. ■

Theorem 6.3.4 implies that if the infinite series $\sum_{i=1}^{\infty} p_i(1 - p_i)$ is divergent, then the distribution of the sum $\sum_{i=1}^n X_i$ of a large number of independent Bernoulli random variables will be approximately the normal distribution with mean $\sum_{i=1}^n p_i$ and variance $\sum_{i=1}^n p_i(1 - p_i)$. It should be kept in mind, however, that a typical practical problem will involve only a finite number of random variables X_1, \dots, X_n , rather than an infinite sequence of random variables. In such a problem, it is not meaningful to consider whether or not the infinite series $\sum_{i=1}^{\infty} p_i(1 - p_i)$ is divergent, because only a finite number of values p_1, \dots, p_n will be specified in the problem. In a certain sense, therefore, the distribution of the sum $\sum_{i=1}^n X_i$ can *always* be approximated by a normal distribution. The critical question is whether or not this normal distribution provides a *good* approximation to the actual distribution of $\sum_{i=1}^n X_i$. The answer depends, of course, on the values of p_1, \dots, p_n .

Since the normal distribution will be attained more and more closely as $\sum_{i=1}^n p_i(1 - p_i) \rightarrow \infty$, the normal distribution provides a good approximation when the value of $\sum_{i=1}^n p_i(1 - p_i)$ is large. Furthermore, since the value of each term $p_i(1 - p_i)$ is a maximum when $p_i = 1/2$, the approximation will be best when n is large and the values of p_1, \dots, p_n are close to 1/2.

Example 6.3.9

Examination Questions. Suppose that an examination contains 99 questions arranged in a sequence from the easiest to the most difficult. Suppose that the probability that a particular student will answer the first question correctly is 0.99, the probability that he will answer the second question correctly is 0.98, and, in general, the probability that he will answer the i th question correctly is $1 - i/100$ for $i = 1, \dots, 99$. It is assumed that all questions will be answered independently and that the student must answer at least 60 questions correctly to pass the examination. We shall determine the probability that the student will pass.

Let $X_i = 1$ if the i th question is answered correctly and $X_i = 0$ otherwise. Then $E(X_i) = p_i = 1 - (i/100)$ and $\text{Var}(X_i) = p_i(1 - p_i) = (i/100)[1 - (i/100)]$. Also,

$$\sum_{i=1}^{99} p_i = 99 - \frac{1}{100} \sum_{i=1}^{99} i = 99 - \frac{1}{100} \cdot \frac{(99)(100)}{2} = 49.5$$

and

$$\begin{aligned}\sum_{i=1}^{99} p_i(1-p_i) &= \frac{1}{100} \sum_{i=1}^{99} i - \frac{1}{(100)^2} \sum_{i=1}^{99} i^2 \\ &= 49.5 - \frac{1}{(100)^2} \cdot \frac{(99)(100)(199)}{6} = 16.665.\end{aligned}$$

It follows from the central limit theorem that the distribution of the total number of questions that are answered correctly, which is $\sum_{i=1}^{99} X_i$, will be approximately the normal distribution with mean 49.5 and standard deviation $(16.665)^{1/2} = 4.08$. Therefore, the distribution of the variable

$$Z = \frac{\sum_{i=1}^n X_i - 49.5}{4.08}$$

will be approximately the standard normal distribution. It follows that

$$\Pr\left(\sum_{i=1}^n X_i \geq 60\right) = \Pr(Z \geq 2.5735) \simeq 1 - \Phi(2.5735) = 0.0050.$$

◀



❖ Outline of Proof of Central Limit Theorem

Convergence of the Moment Generating Functions Moment generating functions are important in the study of convergence in distribution because of the following theorem, the proof of which is too advanced to be presented here.

Theorem 6.3.5 Let X_1, X_2, \dots be a sequence of random variables. For $n = 1, 2, \dots$, let F_n denote the c.d.f. of X_n , and let ψ_n denote the m.g.f. of X_n .

Also, let X^* denote another random variable with c.d.f. F^* and m.g.f. ψ^* . Suppose that the m.g.f.'s ψ_n and ψ^* exist ($n = 1, 2, \dots$). If $\lim_{n \rightarrow \infty} \psi_n(t) = \psi^*(t)$ for all values of t in some interval around the point $t = 0$, then the sequence X_1, X_2, \dots converges in distribution to X^* . ■

In other words, the sequence of c.d.f.'s F_1, F_2, \dots must converge to the c.d.f. F^* if the corresponding sequence of m.g.f.'s ψ_1, ψ_2, \dots converges to the m.g.f. ψ^* .

Outline of the Proof of Theorem 5.7.1 We are now ready to outline a proof of Theorem 6.3.1, which is the central limit theorem of Lindeberg and Lévy. We shall assume that the variables X_1, \dots, X_n form a random sample of size n from a distribution with mean μ and variance σ^2 . We shall also assume, for convenience, that the m.g.f. of this distribution exists, although the central limit theorem is true even without this assumption.

For $i = 1, \dots, n$, let $Y_i = (X_i - \mu)/\sigma$. Then the random variables Y_1, \dots, Y_n are i.i.d., and each has mean 0 and variance 1. Furthermore, let

$$Z_n = \frac{n^{1/2}(\bar{X}_n - \mu)}{\sigma} = \frac{1}{n^{1/2}} \sum_{i=1}^n Y_i.$$

We shall show that Z_n converges in distribution to a random variable having the standard normal distribution, as indicated in Eq. (6.3.1), by showing that the m.g.f. of Z_n converges to the m.g.f. of the standard normal distribution.

If $\psi(t)$ denotes the m.g.f. of each random variable Y_i ($i = 1, \dots, n$), then it follows from Theorem 4.4.4 that the m.g.f. of the sum $\sum_{i=1}^n Y_i$ will be $[\psi(t)]^n$. Also, it follows from Theorem 4.4.3 that the m.g.f. $\zeta_n(t)$ of Z_n will be

$$\zeta_n(t) = \left[\psi\left(\frac{t}{n^{1/2}}\right) \right]^n.$$

In this problem, $\psi'(0) = E(Y_i) = 0$ and $\psi''(0) = E(Y_i^2) = 1$. Therefore, the Taylor series expansion of $\psi(t)$ about the point $t = 0$ has the following form:

$$\begin{aligned}\psi(t) &= \psi(0) + t\psi'(0) + \frac{t^2}{2!}\psi''(0) + \frac{t^3}{3!}\psi'''(0) + \dots \\ &= 1 + \frac{t^2}{2} + \frac{t^3}{3!}\psi'''(0) + \dots.\end{aligned}$$

Also,

$$\zeta_n(t) = \left[1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}}\psi'''(0) + \dots \right]^n. \quad (6.3.15)$$

Apply Theorem 5.3.3 with $1 + a_n/n$ equal to the expression inside brackets in (6.3.15) and $c_n = n$. Since

$$\lim_{n \rightarrow \infty} \left[\frac{t^2}{2} + \frac{t^3}{3!n^{1/2}}\psi'''(0) + \dots \right] = \frac{t^2}{2}.$$

it follows that

$$\lim_{n \rightarrow \infty} \zeta_n(t) = \exp\left(\frac{1}{2}t^2\right). \quad (6.3.16)$$

Since the right side of Eq. (6.3.16) is the m.g.f. of the standard normal distribution, it follows from Theorem 6.3.5 that the asymptotic distribution of Z_n must be the standard normal distribution.

An outline of the proof of the central limit theorem of Liapounov can also be given by proceeding along similar lines, but we shall not consider this problem further here.



Summary

Two versions of the central limit theorem were given. They conclude that the distribution of the average of a large number of independent random variables is close to a normal distribution. One theorem requires that the random variables all have the same distribution with finite variance. The other theorem does not require that the random variables be identically distributed, but instead requires that their third moments exist and satisfy condition (6.3.9). The delta method lets us find the approximate distribution of a smooth function of a sample average.

Exercises

- 1.** Each minute a machine produces a length of rope with mean of 4 feet and standard deviation of 5 inches. Assuming that the amounts produced in different minutes are independent and identically distributed, approximate the probability that the machine will produce at least 250 feet in one hour.
- 2.** Suppose that 75 percent of the people in a certain metropolitan area live in the city and 25 percent of the people live in the suburbs. If 1200 people attending a certain concert represent a random sample from the metropolitan area, what is the probability that the number of people from the suburbs attending the concert will be fewer than 270?
- 3.** Suppose that the distribution of the number of defects on any given bolt of cloth is the Poisson distribution with mean 5, and the number of defects on each bolt is counted for a random sample of 125 bolts. Determine the probability that the average number of defects per bolt in the sample will be less than 5.5.
- 4.** Suppose that a random sample of size n is to be taken from a distribution for which the mean is μ and the standard deviation is 3. Use the central limit theorem to determine approximately the smallest value of n for which the following relation will be satisfied:

$$\Pr(|\bar{X}_n - \mu| < 0.3) \geq 0.95.$$
- 5.** Suppose that the proportion of defective items in a large manufactured lot is 0.1. What is the smallest random sample of items that must be taken from the lot in order for the probability to be at least 0.99 that the proportion of defective items in the sample will be less than 0.13?
- 6.** Suppose that three girls A , B , and C throw snowballs at a target. Suppose also that girl A throws 10 times, and the probability that she will hit the target on any given throw is 0.3; girl B throws 15 times, and the probability that she will hit the target on any given throw is 0.2; and girl C throws 20 times, and the probability that she will hit the target on any given throw is 0.1. Determine the probability that the target will be hit at least 12 times.
- 7.** Suppose that 16 digits are chosen at random with replacement from the set $\{0, \dots, 9\}$. What is the probability that their average will lie between 4 and 6?
- 8.** Suppose that people attending a party pour drinks from a bottle containing 63 ounces of a certain liquid. Suppose also that the expected size of each drink is 2 ounces, that the standard deviation of each drink is 1/2 ounce, and that all drinks are poured independently. Determine the probability that the bottle will not be empty after 36 drinks have been poured.
- 9.** A physicist makes 25 independent measurements of the specific gravity of a certain body. He knows that the limitations of his equipment are such that the standard deviation of each measurement is σ units.
 - By using the Chebyshev inequality, find a lower bound for the probability that the average of his measurements will differ from the actual specific gravity of the body by less than $\sigma/4$ units.
 - By using the central limit theorem, find an approximate value for the probability in part (a).
- 10.** A random sample of n items is to be taken from a distribution with mean μ and standard deviation σ .
 - Use the Chebyshev inequality to determine the smallest number of items n that must be taken in order to satisfy the following relation:

$$\Pr\left(|\bar{X}_n - \mu| \leq \frac{\sigma}{4}\right) \geq 0.99.$$
 - Use the central limit theorem to determine the smallest number of items n that must be taken in order to satisfy the relation in part (a) approximately.
- 11.** Suppose that, on the average, 1/3 of the graduating seniors at a certain college have two parents attend the graduation ceremony, another third of these seniors have one parent attend the ceremony, and the remaining third of these seniors have no parents attend. If there are 600 graduating seniors in a particular class, what is the probability that not more than 650 parents will attend the graduation ceremony?
- 12.** Let X_n be a random variable having the binomial distribution with parameters n and p_n . Assume that $\lim_{n \rightarrow \infty} np_n = \lambda$. Prove that the m.g.f. of X_n converges to the m.g.f. of the Poisson distribution with mean λ .
- 13.** Suppose that X_1, \dots, X_n form a random sample from a normal distribution with unknown mean θ and variance σ^2 . Assuming that $\theta \neq 0$, determine the asymptotic distribution of \bar{X}_n^3 .
- 14.** Suppose that X_1, \dots, X_n form a random sample from a normal distribution with mean 0 and unknown variance σ^2 .
 - Determine the asymptotic distribution of the statistic $\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right)^{-1}$.
 - Find a variance stabilizing transformation for the statistic $\frac{1}{n} \sum_{i=1}^n X_i^2$.
- 15.** Let X_1, X_2, \dots be a sequence of i.i.d. random variables each having the uniform distribution on the interval $[0, \theta]$ for some real number $\theta > 0$. For each n , define Y_n to be the maximum of X_1, \dots, X_n .

- a. Show that the c.d.f. of Y_n is

$$F_n(y) = \begin{cases} 0 & \text{if } y \leq 0, \\ (y/\theta)^n & \text{if } 0 < y < \theta, \\ 1 & \text{if } y \geq \theta. \end{cases}$$

Hint: Read Example 3.9.6.

- b. Show that $Z_n = n(Y_n - \theta)$ converges in distribution to the distribution with c.d.f.

$$F^*(z) = \begin{cases} \exp(z/\theta) & \text{if } z < 0, \\ 1 & \text{if } z \geq 0. \end{cases}$$

Hint: Apply Theorem 5.3.3 after finding the c.d.f. of Z_n .

- c. Use Theorem 6.3.2 to find the approximate distribution of Y_n^2 when n is large.

6.4 The Correction for Continuity

Some applications of the central limit theorem allow us to approximate the probability that a discrete random variable X lies in an interval $[a, b]$ by the probability that a normal random variable lies in that interval. The approximation can be improved slightly by being careful about how we approximate $\Pr(X = a)$ and $\Pr(X = b)$.

Approximating a Discrete Distribution by a Continuous Distribution

**Example
6.4.1**

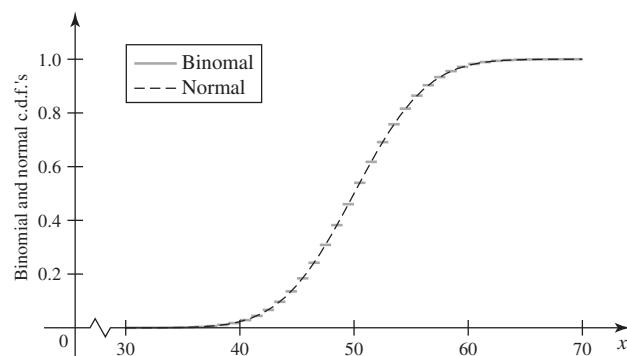
A Large Sample. In Example 6.3.1, we illustrated how the normal distribution with mean 50 and variance 25 could approximate the distribution of a random variable X that has the binomial distribution with parameters 100 and 0.5. In particular, if Y has the normal distribution with mean 50 and variance 25, we know that $\Pr(Y \leq x)$ is close to $\Pr(X \leq x)$ for all x . But the approximation has some systematic errors. Figure 6.4 shows the two c.d.f.'s over the range $30 \leq x < 70$. The two c.d.f.'s are very close at $x = n + 0.5$ for each integer n . But for each integer n , $\Pr(Y \leq x) < \Pr(X \leq x)$ for x a little above n and $\Pr(Y \leq x) > \Pr(X \leq x)$ for x a little below n . We ought to be able to make use of these systematic discrepancies in order to improve the approximation.



Suppose that X has a discrete distribution that can be approximated by a normal distribution, such as in Example 6.4.1. In this section, we shall describe a standard method for improving the quality of such an approximation based on the systematic discrepancies that were noted at the end of Example 6.4.1.

Let $f(x)$ be the p.f. of the discrete random variable X , and suppose that we wish to approximate the distribution of X by a continuous distribution with p.d.f. $g(x)$. To

Figure 6.4 Comparison of binomial and normal c.d.f.'s.



aid the discussion, let Y be a random variable with p.d.f. g . Also, for simplicity, we shall assume that all of the possible values of X are integers. This condition is satisfied for the binomial, hypergeometric, Poisson, and negative binomial distributions described in this text.

If the distribution of Y provides a good approximation to the distribution of X , then for all integers a and b , we can approximate the discrete probability

$$\Pr(a \leq X \leq b) = \sum_{x=a}^b f(x) \quad (6.4.1)$$

by the continuous probability

$$\Pr(a \leq Y \leq b) = \int_a^b g(x) dx. \quad (6.4.2)$$

Indeed, this approximation was used in Examples 6.3.2 and 6.3.9, where $g(x)$ was the appropriate normal p.d.f. derived from the central limit theorem.

This simple approximation has the following shortcoming: Although $\Pr(X \geq a)$ and $\Pr(X > a)$ will typically have different values for the discrete distribution of X , $\Pr(Y \geq a) = \Pr(Y > a)$ because Y has a continuous distribution. Another way of expressing this shortcoming is as follows: Although $\Pr(X = x) > 0$ for each integer x that is a possible value of X , $\Pr(Y = x) = 0$ for all x .

Approximating a Bar Chart

The p.f. $f(x)$ of a discrete random variable X can be represented by a *bar chart*, as sketched in Fig. 6.5. For each integer x , the probability of $\{X = x\}$ 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 $f(x)$. 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. 6.5. A bar chart with areas of bars proportional to probabilities is analogous to a histogram (see page 165) with areas of bars proportional to proportions of a sample.

From this point of view, it can be seen that $\Pr(a \leq X \leq b)$, as specified in Eq. (6.4.1), is the sum of the areas of the rectangles in Fig. 6.5 that are centered at $a, a + 1, \dots, b$. It can also be seen from Fig. 6.5 that the sum of these areas is

Figure 6.5 Approximating a bar chart by using a p.d.f.

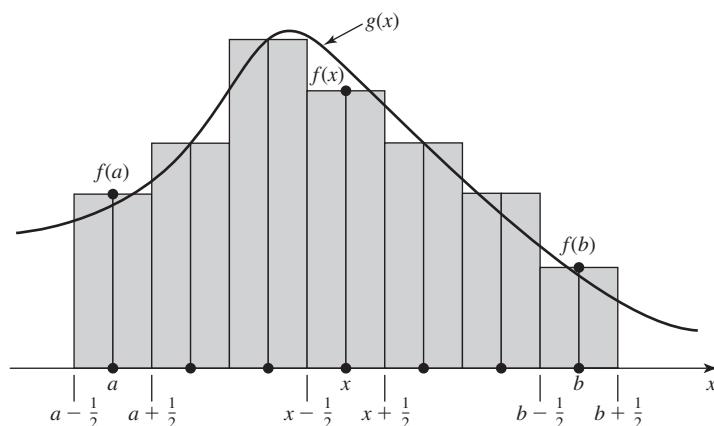
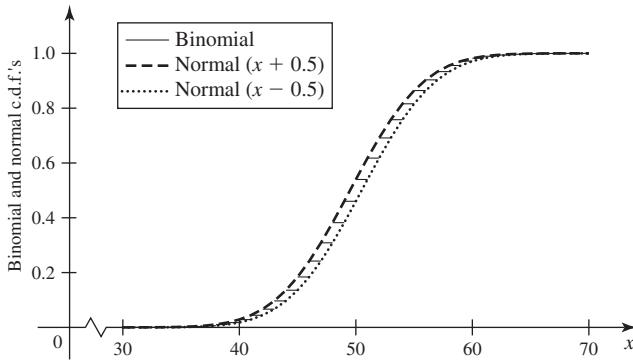


Figure 6.6 Comparison of binomial c.d.f. with normal c.d.f. shifted to the right and to the left by 0.5.



approximated by the integral

$$\Pr(a - 1/2 < Y < b + 1/2) = \int_{a-(1/2)}^{b+(1/2)} g(x) dx. \quad (6.4.3)$$

The adjustment from the integral in (6.4.2) to the integral in (6.4.3) is called the *correction for continuity*.

**Example
6.4.2**

A Large Sample. At the end of Example 6.4.1, we found that when x was a little above an integer, the approximating probability $\Pr(Y \leq x)$ is a bit smaller than the actual probability $\Pr(X \leq x)$. The correction for continuity shifts the c.d.f. of Y to the left by 0.5 when we want to compute $\Pr(Y \leq x)$ for x a little above an integer. This shift replaces $\Pr(Y \leq x)$ by $\Pr(Y \leq x + 0.5)$, which is larger and usually closer to $\Pr(X \leq x)$. Similarly, when we want to compute $\Pr(Y \leq x)$ when x is a little below an integer, the correction for continuity shifts the c.d.f. of Y to the right by 0.5 which replaces $\Pr(Y \leq x)$ by $\Pr(Y \leq x - 0.5)$. Figure 6.6 illustrates both of these shifts and shows how they each approximate the actual binomial c.d.f. better than the unshifted normal c.d.f. in Fig. 6.4. ◀

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

$$\begin{aligned} \Pr(X = a) &= \Pr\left(a - \frac{1}{2} \leq X \leq a + \frac{1}{2}\right) \\ &\approx \int_{a-(1/2)}^{a+(1/2)} g(x) dx. \end{aligned} \quad (6.4.4)$$

Similarly,

$$\begin{aligned} \Pr(X > a) &= \Pr(X \geq a + 1) = \Pr\left(X \geq a + \frac{1}{2}\right) \\ &\approx \int_{a+(1/2)}^{\infty} g(x) dx. \end{aligned} \quad (6.4.5)$$

**Example
6.4.3**

Examination Questions. To illustrate the use of the correction for continuity, we shall again consider Example 6.3.9. In that example, an examination contains 99 questions of varying difficulty and it is desired to determine $\Pr(X \geq 60)$, 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 the normal distribution with mean 49.5 and standard deviation 4.08. Let $Z = (X - 49.5)/4.08$.

If we use the correction for continuity, we obtain

$$\begin{aligned}\Pr(X \geq 60) &= \Pr(X \geq 59.5) = \Pr\left(Z \geq \frac{59.5 - 49.5}{4.08}\right) \\ &\approx 1 - \Phi(2.4510) = 0.007.\end{aligned}$$

This value is somewhat larger than the value 0.005, which was obtained in Sec. 6.3, without the correction. \blacktriangleleft

**Example
6.4.4**

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 the normal distribution with mean 10 and standard deviation $[(20)(1/2)(1/2)]^{1/2} = 2.236$. If we use the correction for continuity,

$$\begin{aligned}\Pr(X = 10) &= \Pr(9.5 \leq X \leq 10.5) \\ &= \Pr\left(-\frac{0.5}{2.236} \leq Z \leq \frac{0.5}{2.236}\right) \\ &\approx \Phi(0.2236) - \Phi(-0.2236) = 0.177.\end{aligned}$$

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. \blacktriangleleft

Summary

Let X be a random variable that takes only integer values. Suppose that X has approximately the normal distribution with mean μ and variance σ^2 . Let a and b be integers, and suppose that we wish to approximate $\Pr(a \leq X \leq b)$. The correction to the normal distribution approximation for continuity is to use $\Phi([b + 1/2 - \mu]/\sigma) - \Phi([a - 1/2 - \mu]/\sigma)$ rather than $\Phi([b - \mu]/\sigma) - \Phi([a - \mu]/\sigma)$ as the approximation.

Exercises

1. Let X_1, \dots, X_{30} be independent random variables each having a discrete distribution with p.f.

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2, \\ 1/2 & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Use the central limit theorem and the correction for continuity to approximate the probability that $X_1 + \dots + X_{30}$ is at most 33.

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

- a. Determine approximately the value of $\Pr(X = 4)$ by using the central limit theorem with the correction for continuity.
- b. Compare the answer obtained in part (a) with the exact value of this probability.
3. Using the correction for continuity, determine the probability required in Example 6.3.2.
4. Using the correction for continuity, determine the probability required in Exercise 2 of Sec. 6.3.
5. Using the correction for continuity, determine the probability required in Exercise 3 of Sec. 6.3.

6. Using the correction for continuity, determine the probability required in Exercise 6 of Sec. 6.3.

7. Using the correction for continuity, determine the probability required in Exercise 7 of Sec. 6.3.

6.5 Supplementary Exercises

1. Suppose that a pair of balanced dice are rolled 120 times, and let X denote the number of rolls on which the sum of the two numbers is 7. Use the central limit theorem to determine a value of k such that $\Pr(|X - 20| \leq k)$ is approximately 0.95.

2. Suppose that X has a Poisson distribution with a very large mean λ . Explain why the distribution of X can be approximated by the normal distribution with mean λ and variance λ . In other words, explain why $(X - \lambda)/\lambda^{1/2}$ converges in distribution, as $\lambda \rightarrow \infty$, to a random variable having the standard normal distribution.

3. Suppose that X has the Poisson distribution with mean 10. Use the central limit theorem, both without and with the correction for continuity, to determine an approximate value for $\Pr(8 \leq X \leq 12)$. Use the table of Poisson probabilities given in the back of this book to assess the quality of these approximations.

4. Suppose that X is a random variable such that $E(X^k)$ exists and $\Pr(X \geq 0) = 1$. Prove that for $k > 0$ and $t > 0$,

$$\Pr(X \geq t) \leq \frac{E(X^k)}{t^k}.$$

5. Suppose that X_1, \dots, X_n form a random sample from the Bernoulli distribution with parameter p . Let \bar{X}_n be the sample average. Find a variance stabilizing transformation for \bar{X}_n . Hint: When trying to find the integral of $(p[1-p])^{-1/2}$, make the substitution $z = \sqrt{p}$ and then think about arcsin, the inverse of the sin function.

6. Suppose that X_1, \dots, X_n form a random sample from the exponential distribution with mean θ . Let \bar{X}_n be the sample average. Find a variance stabilizing transformation for \bar{X}_n .

7. Suppose that X_1, X_2, \dots is a sequence of positive integer-valued random variables. Suppose that there is a function f such that for every $m = 1, 2, \dots$, $\lim_{n \rightarrow \infty} \Pr(X_n = m) = f(m)$, $\sum_{m=1}^{\infty} f(m) = 1$, and $f(x) = 0$ for every x that is not a positive integer. Let F be the discrete c.d.f. whose p.f. is f . Prove that X_n converges in distribution to F .

8. Let $\{p_n\}_{n=1}^{\infty}$ be a sequence of numbers such that $0 < p_n < 1$ for all n . Assume that $\lim_{n \rightarrow \infty} p_n = p$ with $0 < p < 1$. Let X_n have the binomial distribution with parameters k and p_n for some positive integer k . Prove that X_n converges in distribution to the binomial distribution with parameters k and p .

9. Suppose that the number of minutes required to serve a customer at the checkout counter of a supermarket has an exponential distribution for which the mean is 3. Using the central limit theorem, approximate the probability that the total time required to serve a random sample of 16 customers will exceed one hour.

10. Suppose that we model the occurrence of defects on a fabric manufacturing line as a Poisson process with rate 0.01 per square foot. Use the central limit theorem (both with and without the correction for continuity) to approximate the probability that one would find at least 15 defects in 2000 square feet of fabric.

11. Let X have the gamma distribution with parameters n and 3, where n is a large integer.

a. Explain why one can use the central limit theorem to approximate the distribution of X by a normal distribution.

b. Which normal distribution approximates the distribution of X ?

12. Let X have the negative binomial distribution with parameters n and 0.2, where n is a large integer.

a. Explain why one can use the central limit theorem to approximate the distribution of X by a normal distribution.

b. Which normal distribution approximates the distribution of X ?