# Chapter 4

# Unbiasedness, Consistency, and Limiting Distributions

In the previous chapters, we were concerned with probability models and distributions. In the next chapter, we begin discussing statistical inference which will remain our focus for the remainder of this book. In this chapter, we present some tools drawn from asymptotic theory. These are useful in statistics as well as in probability theory.

In our discussion, we use some examples from statistics so the concept of a random sample will prove helpful for this chapter. More details on sampling are given in Chapter 5. Suppose we have a random variable $X$ which has pdf, (or pmf), $f(x;\theta)$, $(p(x;\theta))$, where $\theta$ is either a real number or a vector of real numbers. Assume that $\theta \in \Omega$ which is a subset of $R^p$, for $p \geq 1$. For example, $\theta$ could be the vector $(\mu, \sigma^2)$ when $X$ has a $N(\mu, \sigma^2)$ distribution or $\theta$ could be the probability of success $p$ when $X$ has a binomial distribution. In the previous chapters, say to work a probability problem, we would know $\theta$. In statistics, though, $\theta$ is unknown. Our information about $\theta$ comes from a sample $X_1, X_2, \ldots, X_n$. We often assume that this is a **random sample** which means that the random variables $X_1, X_2, \ldots, X_n$ are independent and have the same distribution as $X$; that is, $X_1, X_2, \ldots, X_n$ are iid. A **statistic** $T$ is a function of the sample; i.e, $T = T(X_1, X_2, \ldots, X_n)$. We may use $T$ to estimate $\theta$. In which case, we would say that $T$ is a **point estimator** of $\theta$. For example, suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean $\mu$ and variance $\sigma^2$. Then the statistics $\overline{X}$ and $S^2$ are referred to as the **sample mean** and the **sample variance** of this random sample. They are point estimators of $\mu$ and $\sigma^2$, respectively.

As another illustration, consider the case where $X$ is Bernoulli with probability of success $p$; that is, $X$ assumes the values 1 or 0 with probabilities $p$ or $1-p$, respectively. Suppose we perform $n$ Bernoulli trials. Recall that Bernoulli trials are performed independently of one another and under identical conditions. Let $X_i$ be the outcome on the $ith$ trial, $i = 1, 2, \ldots, n$. Then $X_1, X_2, \ldots, X_n$ form a random sample from the distribution $X$. A statistic of interest here is $\overline{X}$, which is the proportion of successes in the sample. It is a point estimator of $p$.

## 4.1 Expectations of Functions

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ denote a random vector from some experiment. Often we are interested in a function of $\mathbf{X}$, say, $T = T(\mathbf{X})$. For example, if $\mathbf{X}$ is a sample, $T$ may be a statistic of interest. We begin by considering linear functions of $\mathbf{X}$; i.e., functions of the form

$$T = \mathbf{a}'\mathbf{X} = \sum_{i=1}^{n} a_i X_i,$$

for a specified vector $\mathbf{a} = (a_1, \ldots, a_n)'$. We will obtain the mean and variance of such random variables.

The mean of $T$ follows immediately from the linearity of the expectation operator, $E$, but for easy reference we state this as a theorem:

**Theorem 4.1.1.** *Let $T = \sum_{i=1}^{n} a_i X_i$. Provided $E[|X_i|] < \infty$, for $i = 1, \ldots, n$,*

$$E(T) = \sum_{i=1}^{n} a_i E(X_i).$$

For the variance of $T$, we first state a very general result involving covariances. Let $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ denote another random vector and let $W = \mathbf{b}'\mathbf{Y}$ for a specified vector $\mathbf{b} = (b_1, \ldots, b_m)'$.

**Theorem 4.1.2.** *Let $T = \sum_{i=1}^{n} a_i X_i$ and let $W = \sum_{i=1}^{m} b_i Y_i$. If $E[X_i^2] < \infty$, and $E[Y_j^2] < \infty$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, then*

$$cov(T, W) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j cov(X_i, Y_j).$$

*Proof:* Using the definition of the covariance and Theorem 4.1.1 we have the first equality below, while the second equality follows from the linearity of $E$

$$\begin{aligned} \text{cov}(T, W) &= E[\sum_{i=1}^{n} \sum_{j=1}^{m} (a_i X_i - a_i E(X_i))(b_j Y_j - b_j E(Y_j))] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j E[(X_i - E(X_i))(Y_j - E(Y_j))], \end{aligned}$$

which is the desired result. ■

To obtain the variance of $T$, simply replace $W$ by $T$ in Theorem 4.1.2. We state the result as a theorem:

**Corollary 4.1.1.** *Let $T = \sum_{i=1}^{n} a_i X_i$. Provided $E[X_i^2] < \infty$, for $i = 1, \ldots, n$,*

$$Var(T) = cov(T, T) = \sum_{i=1}^{n} a_i^2 \, Var(X_i) + 2 \sum_{i<j} a_i a_j cov(X_i, X_j). \tag{4.1.1}$$

Note that if $X_1, \ldots, X_n$ are independent random variables then the covariance $\text{cov}(X_i, X_j) = 0$; see Example 2.5.4. This leads to a simplification of (4.1.1) which we record in the following corollary.

**Corollary 4.1.2.** *If $X_1, \ldots, X_n$ are independent random variables with finite variances, then*

$$Var(T) = \sum_{i=1}^{n} a_i^2 \, Var(X_i). \tag{4.1.2}$$

Note that we need only $X_i$ and $X_j$ to be uncorrelated for all $i \neq j$ to obtain this result; for example, $\text{Cov}(X_i, X_j) = 0$, $i \neq j$, which is true when $X_1, \ldots, X_n$ are independent.

Let us now return to the discussion of sampling and statistics found at the beginning of this chapter. Consider the situation where we have a random variable $X$ of interest whose density is given by $f(x; \theta)$, for $\theta \in \Omega$. The parameter $\theta$ is unknown and we seek a statistic based on a sample to estimate it. Our first property of an estimator concerns its expectation.

**Definition 4.1.1.** *Let $X$ be a random variable with pdf $f(x;\theta)$ or pmf $p(x;\theta)$, $\theta \in \Omega$. Let $X_1, \ldots, X_n$ be a random sample from the distribution of $X$ and let $T$ denote a statistic. We say $T$ is an* **unbiased** *estimator of $\theta$ if*

$$E(T) = \theta, \quad \text{for all } \theta \in \Omega. \tag{4.1.3}$$

*If $T$ is not unbiased (that is, $E(T) \neq \theta$), we say that $T$ is a* **biased** *estimator of $\theta$.*

**Example 4.1.1 (Sample Mean).** Let $X_1, \ldots, X_n$ be a random sample from the distribution of a random variable $X$ which has mean $\mu$ and variance $\sigma^2$. Recall that the **sample mean** is given by $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$. This is a linear combination of the sample observations with $a_i \equiv n^{-1}$; hence, by Theorem 4.1.1 and Corollary 4.1.2 we have,

$$E(\overline{X}) = \mu \text{ and } Var(\overline{X}) = \frac{\sigma^2}{n}. \tag{4.1.4}$$

Hence, $\overline{X}$ is an unbiased estimator of $\mu$. Furthermore, the variance of $\overline{X}$ becomes small as $n$ gets large. That is, it seems in the limit that the mass of the distribution of the sample mean $\overline{X}$ is converging to $\mu$ as $n$ gets large. This is presented in the next section. ■

**Example 4.1.2 (Sample Variance).** As in the last example, let $X_1, \ldots, X_n$ be a random sample from the distribution of a random variable $X$ which has mean $\mu$ and variance $\sigma^2$. Define the **sample variance** by

$$S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = (n-1)^{-1} \left( \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right), \tag{4.1.5}$$

where the second equality follows after some algebra; see Exercise 4.1.3. Using the above theorems, the results of the last example, and the fact that $E(X^2) = \sigma^2 + \mu^2$, we have the following

$$\begin{aligned} E(S^2) &= (n-1)^{-1}\left(\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right) \\ &= (n-1)^{-1}\left\{n\sigma^2 + n\mu^2 - n[(\sigma^2/n) + \mu^2]\right\} \\ &= \sigma^2. \end{aligned} \tag{4.1.6}$$

Hence, the sample variance is an unbiased estimate of $\sigma^2$. If $V = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ then $E(V) = ((n-1)/n)\sigma^2$. That is $V$ is a biased estimator of $\sigma^2$. This is one reason for dividing by $n-1$ instead of $n$ in the definition of the sample variance. ■

**Example 4.1.3 (Maximum of a Sample from a Uniform Distribution).** Let $X_1, \ldots, X_n$ be a random sample from a uniform$(0, \theta)$ distribution. Suppose $\theta$ is unknown. An intuitive estimate of $\theta$ is the maximum of the sample. Let $Y_n = \max\{X_1, \ldots, X_n\}$. Exercise 4.1.2 shows that the cdf of $Y_n$ is

$$F_{Y_n}(t) = \begin{cases} 1 & t > \theta \\ \left(\frac{t}{\theta}\right)^n & 0 < t \le \theta \\ 0 & t \le 0. \end{cases} \tag{4.1.7}$$

Hence, the pdf of $Y_n$ is

$$f_{Y_n}(t) = \begin{cases} \frac{n}{\theta^n} t^{n-1} & 0 < t \le \theta \\ 0 & \text{elsewhere.} \end{cases} \tag{4.1.8}$$

Based on its pdf, it is easy to show that $E(Y_n) = (n/(n+1))\theta$. Thus, $Y_n$ is a biased estimator of $\theta$. Note, however, that $((n+1)/n)Y_n$ is an unbiased estimator of $\theta$. ■

**Example 4.1.4 (Sample Median).** Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution of $X$, which has pdf $f(x)$. Suppose $\mu = E(X)$ exists and, further, that the pdf $f(x)$ is symmetric about $\mu$. In Example 4.1.1, we showed that the sample mean was an unbiased estimator of $\mu$. What about the sample median, $T = T(X_1, X_2, \ldots, X_n) = \text{med}\{X_1, X_2, \ldots, X_n\}$? The sample median satisfies two properties: (1), if we increase (or decrease) the sample items by $b$ then the sample median increases (or decreases) by $b$, and (2), if we multiply each sample item by $-1$, then the median gets multiplied by $-1$. We can abbreviate these properties as:

$$T(X_1 + b, X_2 + b, \ldots, X_n + b) = T(X_1, X_2, \ldots, X_n) + b \tag{4.1.9}$$

$$T(-X_1, -X_2, \ldots, -X_n) = -T(X_1, X_2, \ldots, X_n). \tag{4.1.10}$$

As Exercise 4.1.1 shows, if $X_i$ is symmetrically distributed about $\mu$, the distribution of the random vector $(X_1 - \mu, \ldots, X_n - \mu)$ is the same as the distribution of the random vector $(-(X_1 - \mu), \ldots, -(X_n - \mu))$. In particular, expectations taken under these random vectors are the same. By this fact and (4.1.9) and (4.1.10), we have the following:

$$\begin{aligned} E[T] - \mu &= E[T(X_1, \ldots, X_n)] - \mu = E[T(X_1 - \mu, \ldots, X_n - \mu)] \\ &= E[T(-(X_1 - \mu), \ldots, -(X_n - \mu))] \\ &= -E[T(X_1 - \mu, \ldots, X_n - \mu)] \\ &= -E[T(X_1, \ldots, X_n)] + \mu = -E[T] + \mu. \end{aligned} \tag{4.1.11}$$

That is, $2E(T) = 2\mu$, so we have $E[T] = \mu$. However, the sample median satisfies (4.1.9) and (4.1.10); thus, under these conditions the sample median is an unbiased estimator of $\theta$. Which estimator, the sample mean or the sample median, is better? We will consider this question later. ■

Note that the median is transparent to the argument in the last example. That is, if $T$ is an estimator of $\mu$ which satisfies the conditions (4.1.9) and (4.1.10) and the pdf of $X$ is symmetric about $\mu$, then $T$ is an unbiased estimator of $\mu$.

## EXERCISES

**4.1.1.** Suppose $X$ has a pdf which is symmetric about $b$; i.e., $f(b+x) = f(b-x)$, for all $-\infty < x < \infty$. We say that $X$ is symmetrically distributed about $b$.

(a) Show that $Y = X - b$ is symmetrically distributed about 0.

(b) Show that $Z = -(X - b)$ has the same distribution as $Y$ in Part (a).

(c) Show that $(X_1 - \mu, \ldots, X_n - \mu)$ and $(-(X_1 - \mu), \ldots, -(X_n - \mu))$ as defined in Example 4.1.4 have the same distribution.

**4.1.2.** Derive the cdf given in expression (4.1.7).

**4.1.3.** Derive the second equality in expression (4.1.5).

**4.1.4.** Let $X_1, X_2, X_3, X_4$ be four iid random variables having the same pdf $f(x) = 2x,\ 0 < x < 1$, zero elsewhere. Find the mean and variance of the sum $Y$ of these four random variables.

**4.1.5.** Let $X_1$ and $X_2$ be two independent random variables so that the variances of $X_1$ and $X_2$ are $\sigma_1^2 = k$ and $\sigma_2^2 = 2$, respectively. Given that the variance of $Y = 3X_2 - X_1$ is 25, find $k$.

**4.1.6.** If the independent variables $X_1$ and $X_2$ have means $\mu_1,\ \mu_2$ and variances $\sigma_1^2,\ \sigma_2^2$, respectively, show that the mean and variance of the product $Y = X_1X_2$ are $\mu_1\mu_2$ and $\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2$, respectively.

**4.1.7.** Find the mean and variance of the sum $Y$ of the observations of a random sample of size 5 from the distribution having pdf $f(x) = 6x(1-x),\ 0 < x < 1$, zero elsewhere.

**4.1.8.** Determine the mean and variance of the mean $\overline{X}$ of a random sample of size 9 from a distribution having pdf $f(x) = 4x^3,\ 0 < x < 1$, zero elsewhere.

**4.1.9.** Let $X$ and $Y$ be random variables with $\mu_1 = 1,\ \mu_2 = 4,\ \sigma_1^2 = 4,\ \sigma_2^2 = 6,\ \rho = \frac{1}{2}$. Find the mean and variance of $Z = 3X - 2Y$.

**4.1.10.** Let $X$ and $Y$ be independent random variables with means $\mu_1,\ \mu_2$ and variances $\sigma_1^2,\ \sigma_2^2$. Determine the correlation coefficient of $X$ and $Z = X - Y$ in terms of $\mu_1,\ \mu_2, \sigma_1^2, \sigma_2^2$.

**4.1.11.** Let $\mu$ and $\sigma^2$ denote the mean and variance of the random variable $X$. Let $Y = c+bX$, where $b$ and $c$ are real constants. Show that the mean and the variance of $Y$ are, respectively, $c + b\mu$ and $b^2\sigma^2$.

**4.1.12.** Find the mean and the variance of $Y = X_1 - 2X_2 + 3X_3$, where $X_1, X_2, X_3$ are observations of a random sample from a chi-square distribution with 6 degrees of freedom.

**4.1.13.** Determine the correlation coefficient of the random variables $X$ and $Y$ if $\text{var}(X) = 4$, $\text{var}(Y) = 2$, and $\text{var}(X + 2Y) = 15$.

**4.1.14.** Let $X$ and $Y$ be random variables with means $\mu_1$, $\mu_2$; variances $\sigma_1^2$, $\sigma_2^2$; and correlation coefficient $\rho$. Show that the correlation coefficient of $W = aX+b$, $a > 0$, and $Z = cY + d$, $c > 0$, is $\rho$.

**4.1.15.** A person rolls a die, tosses a coin, and draws a card from an ordinary deck. He receives \$3 for each point up on the die, \$10 for a head and \$0 for a tail, and \$1 for each spot on the card (jack = 11, queen = 12, king = 13). If we assume that the three random variables involved are independent and uniformly distributed, compute the mean and variance of the amount to be received.

**4.1.16.** Let $X_1$ and $X_2$ be independent random variables with nonzero variances. Find the correlation coefficient of $Y = X_1X_2$ and $X_1$ in terms of the means and variances of $X_1$ and $X_2$.

**4.1.17.** Let $X_1$ and $X_2$ have a joint distribution with parameters $\mu_1$, $\mu_2$, $\sigma_1^2$, $\sigma_2^2$, and $\rho$. Find the correlation coefficient of the linear functions of $Y = a_1X_1 + a_2X_2$ and $Z = b_1X_1 + b_2X_2$ in terms of the real constants $a_1$, $a_2$, $b_1$, $b_2$, and the parameters of the distribution.

**4.1.18.** Let $X_1, X_2$, and $X_3$ be random variables with equal variances but with correlation coefficients $\rho_{12} = 0.3$, $\rho_{13} = 0.5$, and $\rho_{23} = 0.2$. Find the correlation coefficient of the linear functions $Y = X_1 + X_2$ and $Z = X_2 + X_3$.

**4.1.19.** Find the variance of the sum of 10 random variables if each has variance 5 and if each pair has correlation coefficient 0.5.

**4.1.20.** Let $X$ and $Y$ have the parameters $\mu_1$, $\mu_2$, $\sigma_1^2$, $\sigma_2^2$, and $\rho$. Show that the correlation coefficient of $X$ and $[Y - \rho(\sigma_2/\sigma_1)X]$ is zero.

**4.1.21.** Let $X_1$ and $X_2$ have a bivariate normal distribution with parameters $\mu_1$, $\mu_2$, $\sigma_1^2$, $\sigma_2^2$, and $\rho$. Compute the means, the variances, and the correlation coefficient of $Y_1 = \exp(X_1)$ and $Y_2 = \exp(X_2)$.
*Hint:* Various moments of $Y_1$ and $Y_2$ can be found by assigning appropriate values to $t_1$ and $t_2$ in $E[\exp(t_1X_1 + t_2X_2)]$.

**4.1.22.** Let $X$ be $N(\mu, \sigma^2)$ and consider the transformation $X = \log(Y)$ or, equivalently, $Y = e^X$.

(a) Find the mean and the variance of $Y$ by first determining $E(e^X)$ and $E[(e^X)^2]$, by using the mgf of $X$.

(b) Find the pdf of $Y$. This is the pdf of the *lognormal distribution*.

**4.1.23.** Let $X_1$ and $X_2$ have a trinomial distribution with parameters $n$, $p_1$, $p_2$.

(a) What is the distribution of $Y = X_1 + X_2$?

(b) From the equality $\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$, once again determine the correlation coefficient $\rho$ of $X_1$ and $X_2$.

**4.1.24.** Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$, where $X_1, X_2$, and $X_3$ are three independent random variables. Find the joint mgf and the correlation coefficient of $Y_1$ and $Y_2$ provided that:

(a) $X_i$ has a Poisson distribution with mean $\mu_i$, $i = 1, 2, 3$.

(b) $X_i$ is $N(\mu_i, \sigma_i^2)$, $i = 1, 2, 3$.

**4.1.25.** Let $S^2$ be the sample variance of a random sample from a distribution with variance $\sigma^2 > 0$. Since $E(S^2) = \sigma^2$, why isn't $E(S) = \sigma$?
*Hint:* Use Jensen's inequality to show that $E(S) < \sigma$.

**4.1.26.** For the last exercise, suppose that the sample is drawn from a $N(\mu, \sigma^2)$ distribution. Recall that $(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution. Use Theorem 3.3.1 to determine an unbiased estimator of $\sigma$.

**4.1.27.** Let $S^2$ be the sample variance of a random sample drawn from a $N(\mu, \sigma^2)$ distribution. Show that the constant $c = (n-1)/(n+1)$ minimizes $E[(cS^2 - \sigma^2)^2]$. Hence, the estimator $(n+1)^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$ of $\sigma^2$ minimizes the mean square error among estimators of the form $cS^2$.

## 4.2 Convergence in Probability

In this section, we formalize a way of saying that a sequence of random variables is getting "close" to another random variable. We will use this concept throughout the book.

**Definition 4.2.1.** *Let $\{X_n\}$ be a sequence of random variables and let $X$ be a random variable defined on a sample space. We say that $X_n$* **converges in probability** *to $X$ if for all $\epsilon > 0$*

$$\lim_{n\to\infty} P[|X_n - X| \geq \epsilon] = 0,$$

*or equivalently,*

$$\lim_{n\to\infty} P[|X_n - X| < \epsilon] = 1.$$

*If so, we write*

$$X_n \xrightarrow{P} X.$$

If $X_n \xrightarrow{P} X$, we often say that the mass of the difference $X_n - X$ is converging to 0. In statistics, often the limiting random variable $X$ is a constant; i.e., $X$ is a degenerate random variable with all its mass at some constant $a$. In this case, we write $X_n \xrightarrow{P} a$. Also, as Exercise 4.2.1 shows, convergence of real sequence $a_n \to a$ is equivalent to $a_n \xrightarrow{P} a$.

One way of showing convergence in probability is to use Chebyshev's Theorem (1.10.3). An illustration of this is given in the following proof. To emphasize the fact that we are working with sequences of random variables, we may place a subscript $n$ on random variables, like $\overline{X}$ to read $\overline{X}_n$.

**Theorem 4.2.1 (Weak Law of Large Numbers).** *Let $\{X_n\}$ be a sequence of iid random variables having common mean $\mu$ and variance $\sigma^2 < \infty$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then*

$$\overline{X}_n \xrightarrow{P} \mu.$$

*Proof.* Recall from Example 4.1.1 that mean and variance of $\overline{X}_n$ is $\mu$ and $\sigma^2/n$, respectively. Hence, by Chebyshev's Theorem we have for any $\epsilon > 0$,

$$P[|\overline{X}_n - \mu| \geq \epsilon] = P[|\overline{X}_n - \mu| \geq (\epsilon\sqrt{n}/\sigma)(\sigma/\sqrt{n})] \leq \frac{\sigma^2}{n\epsilon^2} \to 0. \quad \blacksquare$$

This theorem says that all the mass of the distribution of $\overline{X}_n$ is converging to $\mu$, as $n$ converges to $\infty$. In a sense, for $n$ large, $\overline{X}_n$ is close to $\mu$. But how close? For instance, if we were to estimate $\mu$ by $\overline{X}_n$, what can we say about the error of estimation? We will answer this in Section 4.3.

Actually in a more advanced course a Strong Law of Large Numbers is proven; see page 124 of Chung (1968). One result of this theorem is that we can weaken the hypothesis of Theorem 4.2.1 to the assumption that the random variables $X_i$ are independent and each has finite mean $\mu$. Thus the Strong Law of Large Numbers is a first moment theorem, while the Weak Law requires the existence of the second moment.

There are several theorems concerning convergence in probability which will be useful in the sequel. Together the next two theorems say that convergence in probability is closed under linearity.

**Theorem 4.2.2.** *Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then $X_n + Y_n \xrightarrow{P} X + Y$.*

*Proof:* Let $\epsilon > 0$ be given. Using the triangle inequality we can write

$$|X_n - X| + |Y_n - Y| \geq |(X_n + Y_n) - (X + Y)| \geq \epsilon.$$

Since $P$ is monotone relative to set containment, we have

$$\begin{aligned} P[|(X_n + Y_n) - (X + Y)| \geq \epsilon] &\leq P[|X_n - X| + |Y_n - Y| \geq \epsilon] \\ &\leq P[|X_n - X| \geq \epsilon/2] + P[|Y_n - Y| \geq \epsilon/2]. \end{aligned}$$

By the hypothesis of the theorem, the last two terms converge to 0 which gives us the desired result. $\blacksquare$

**Theorem 4.2.3.** *Suppose $X_n \xrightarrow{P} X$ and $a$ is a constant. Then $aX_n \xrightarrow{P} aX$.*

*Proof:* If $a = 0$, the result is immediate. Suppose $a \neq 0$. Let $\epsilon > 0$. The result follows from these equalities:

$$P[|aX_n - aX| \geq \epsilon] = P[|a||X_n - X| \geq \epsilon] = P[|X_n - X| \geq \epsilon/|a|],$$

and by hypotheses the last term goes to 0. ∎

**Theorem 4.2.4.** *Suppose $X_n \xrightarrow{P} a$ and the real function $g$ is continuous at $a$. Then $g(X_n) \xrightarrow{P} g(a)$.*

*Proof:* Let $\epsilon > 0$. Then since $g$ is continuous at $a$, there exists a $\delta > 0$ such that if $|x - a| < \delta$, then $|g(x) - g(a)| < \epsilon$. Thus

$$|g(x) - g(a)| \geq \epsilon \Rightarrow |x - a| \geq \delta.$$

Substituting $X_n$ for $x$ in the above implication, we obtain

$$P[|g(X_n) - g(a)| \geq \epsilon] \leq P[|X_n - a| \geq \delta].$$

By the hypothesis, the last term goes to 0 as $n \to \infty$, which gives us the result. ∎

This theorem gives us many useful results. For instance, if $X_n \xrightarrow{P} a$, then

$$\begin{aligned} X_n^2 &\xrightarrow{P} a^2, \\ 1/X_n &\xrightarrow{P} 1/a, \quad \text{provided } a \neq 0, \\ \sqrt{X_n} &\xrightarrow{P} \sqrt{a}, \quad \text{provided } a \geq 0. \end{aligned}$$

Actually, in a more advanced class, it is shown that if $X_n \xrightarrow{P} X$ and $g$ is a continuous function then $g(X_n) \xrightarrow{P} g(X)$; see page 104 of Tucker (1967). We make use of this in the next theorem.

**Theorem 4.2.5.** *Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then $X_nY_n \xrightarrow{P} XY$.*

*Proof:* Using the above results, we have

$$\begin{aligned} X_nY_n &= \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2 \\ &\xrightarrow{P} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X - Y)^2 = XY. \end{aligned}$$ ∎

Let us return to our discussion of sampling and statistics. Consider the situation where we have a random variable $X$ whose distribution has an unknown parameter $\theta \in \Omega$. We seek a statistic based on a sample to estimate $\theta$. In the last section, we introduced the property of unbiasedness for an estimator. We now introduce consistency:

**Definition 4.2.2.** *Let $X$ be a random variable with cdf $F(x,\theta)$, $\theta \in \Omega$. Let $X_1, \ldots, X_n$ be a sample from the distribution of $X$ and let $T_n$ denote a statistic. We say $T_n$ is a* **consistent** *estimator of $\theta$ if*

$$T_n \xrightarrow{P} \theta.$$

If $X_1, \ldots, X_n$ is a random sample from a distribution with finite mean $\mu$ and variance $\sigma^2$, then by the Weak Law of Large Numbers, the sample mean, $\overline{X}$, is a consistent estimator of $\mu$.

**Example 4.2.1 (Sample Variance).** Let $X_1, \ldots, X_n$ denote a random sample from a distribution with mean $\mu$ and variance $\sigma^2$. Theorem 4.2.1 showed that $\overline{X}_n \xrightarrow{P} \mu$. To show that the sample variance converges in probability to $\sigma^2$, assume further that $E[X_1^4] < \infty$, so that $\mathrm{Var}(S^2) < \infty$. Using the preceding results, we can show the following:

$$\begin{aligned} S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 &= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\right) \\ &\xrightarrow{P} 1 \cdot [E(X_1^2) - \mu^2] = \sigma^2. \end{aligned}$$

Hence, the sample variance is a consistent estimator of $\sigma^2$. ■

Unlike the last example, sometimes we can obtain the convergence by using the distribution function. We illustrate this with the following example:

**Example 4.2.2 (Maximum of a Sample from a Uniform Distribution).** Reconsider Example 4.1.3, where $X_1, \ldots, X_n$ is a random sample from a uniform$(0,\theta)$ distribution. Let $Y_n = \max\{X_1, \ldots, X_n\}$. The cdf of $Y_n$ is given by expression (4.1.7), from which it is easily seen that $Y_n \xrightarrow{P} \theta$ and the sample maximum is a consistent estimate of $\theta$. Note that the unbiased estimator, $((n+1)/n)Y_n$, is also consistent. ■

To expand on Example 4.2.2, by the Weak Law of Large Numbers, Theorem 4.2.1, it follows that $\overline{X}_n$ is a consistent estimator of $\theta/2$ so $2\overline{X}_n$ is a consistent estimator of $\theta$. Note the difference in how we showed that $Y_n$ and $2\overline{X}_n$ converge to $\theta$ in probability. For $Y_n$ we used the cdf of $Y_n$ but for $2\overline{X}_n$ we appealed to the Weak Law of Large Numbers. In fact, the cdf of $2\overline{X}_n$ is quite complicated for the uniform model. In many situations, the cdf of the statistic cannot be obtained but we can appeal to asymptotic theory to establish the result. There are other estimators of $\theta$. Which is the "best" estimator? In future chapters we will be concerned with such questions.

Consistency is a very important property for an estimator to have. It is a poor estimator that does not approach its target as the sample size gets large. Note that the same cannot be said for the property of unbiasedness. For example, instead of using the sample variance to estimate $\sigma^2$, suppose we use $V = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Then $V$ is consistent for $\sigma^2$, but it is biased, because $E(V) = (n-1)\sigma^2/n$. Thus the bias of $V$ is $\sigma^2/n$, which vanishes as $n \to \infty$.

EXERCISES

**4.2.1.** Let $\{a_n\}$ be a sequence of real numbers. Hence, we can also say that $\{a_n\}$ is a sequence of constant (degenerate) random variables. Let $a$ be a real number. Show that $a_n \to a$ is equivalent to $a_n \xrightarrow{P} a$.

**4.2.2.** Let the random variable $Y_n$ have a distribution that is $b(n, p)$.

(a) Prove that $Y_n/n$ converges in probability $p$. This result is one form of the weak law of large numbers.

(b) Prove that $1 - Y_n/n$ converges in probability to $1 - p$.

(c) Prove that $(Y_n/n)(1 - Y_n/n)$ converges in probability to $p(1 - p)$.

**4.2.3.** Let $W_n$ denote a random variable with mean $\mu$ and variance $b/n^p$, where $p > 0$, $\mu$, and $b$ are constants (not functions of $n$). Prove that $W_n$ converges in probability to $\mu$.
*Hint:* Use Chebyshev's inequality.

**4.2.4.** Let $X_1, \ldots, X_n$ be iid random variables with common pdf

$$f(x) = \begin{cases} e^{-(x-\theta)} & x > \theta \quad -\infty < \theta < \infty \\ 0 & \text{elsewhere.} \end{cases} \tag{4.2.1}$$

This pdf is called the **shifted exponential**. Let $Y_n = \min\{X_1, \ldots, X_n\}$. Prove that $Y_n \to \theta$ in probability, by obtaining the cdf and the pdf of $Y_n$.

**4.2.5.** For Exercise 4.2.4, obtain the mean of $Y_n$. Is $Y_n$ an unbiased estimator of $\theta$? Obtain an unbiased estimator of $\theta$ based on $Y_n$.

## 4.3 Convergence in Distribution

In the last section, we introduced the concept of convergence in probability. With this concept, we can formally say, for instance, that a statistic converges to a parameter and, furthermore, in many situations we can show this without having to obtain the distribution function of the statistic. But how close is the statistic to the estimator? For instance, can we obtain the error of estimation with some credence? The method of convergence discussed in this section, in conjunction with earlier results, gives us affirmative answers to these questions.

**Definition 4.3.1 (Convergence in Distribution).** *Let $\{X_n\}$ be a sequence of random variables and let $X$ be a random variable. Let $F_{X_n}$ and $F_X$ be, respectively, the cdfs of $X_n$ and $X$. Let $C(F_X)$ denote the set of all points where $F_X$ is continuous. We say that $X_n$* **converges in distribution** *to $X$ if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \textit{for all } x \in C(F_X).$$

*We denote this convergence by*

$$X_n \xrightarrow{D} X.$$

**Remark 4.3.1.** This material on convergence in probability and in distribution comes under what statisticians and probabilists refer to as *asymptotic theory.* Often, we say that the ditsribution of $X$ is the **asymptotic distribution** or the **limiting distribution** of the sequence $\{X_n\}$. We might even refer informally to the asymptotics of certain situations. Moreover, for illustration, instead of saying $X_n \xrightarrow{D} X$, where $X$ has a standard normal random, we may write

$$X_n \xrightarrow{D} N(0,1),$$

as an abbreviated way of saying the same thing. Clearly the right-hand member of this last expression is a distribution and not a random variable as it should be, but we will make use of this convention. In addition, we may say that $X_n$ has a *limiting* standard normal distribution to mean that $X_n \xrightarrow{D} X$, where $X$ has a standard normal random, or equivalently $X_n \xrightarrow{D} N(0,1)$. ■

Motivation for only considering points of continuity of $F_X$ is given by the following simple example. Let $X_n$ be a random variable with all its mass at $\frac{1}{n}$ and let $X$ be a random variable with all its mass at 0. Then as Figure 4.3.1 shows all the mass of $X_n$ is converging to 0, i.e., the distribution of $X$. At the point of discontinuity of $F_X$, $\lim F_{X_n}(0) = 0 \neq 1 = F_X(0)$; while at continuity points $x$ of $F_X$, (i.e., , $x \neq 0$), $\lim F_{X_n}(x) = F_X(x)$. Hence, according to the definition, $X_n \xrightarrow{D} X$.

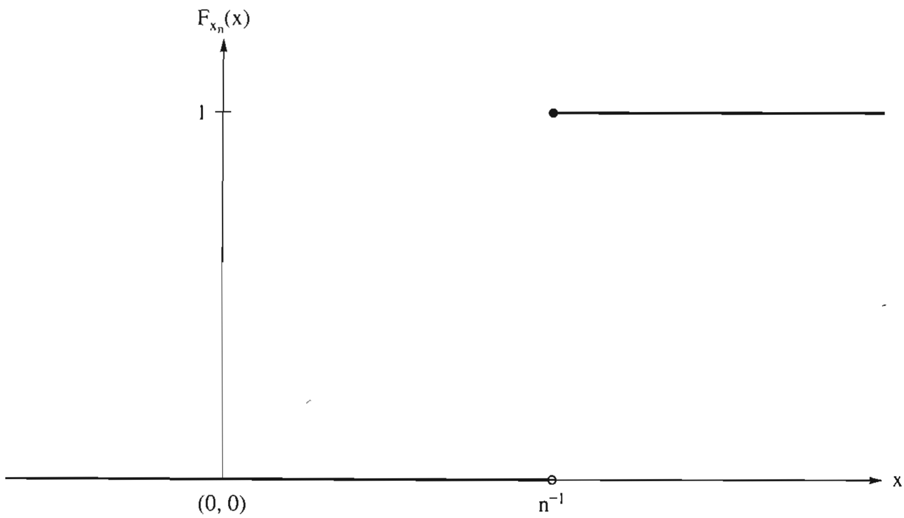


**Figure 4.3.1:** Cdf of $X_n$ which has all its mass at $n^{-1}$

Convergence in probability is a way of saying that a sequence of random variables $X_n$ is getting close to another random variable $X$. On the other hand, convergence in distribution is only concerned with the cdfs $F_{X_n}$ and $F_X$. A simple example illustrates this. Let $X$ be a continuous random variable with a pdf $f_X(x)$ which is symmetric about 0; i.e., $f_X(-x) = f_X(x)$. Then is easy to show that the density of the random variable $-X$ is also $f_X(x)$. Thus $X$ and $-X$ have the same distributions. Define the sequence of random variables $X_n$ as

$$X_n = \begin{cases} X & \text{if } n \text{ is odd} \\ -X & \text{if } n \text{ is even.} \end{cases} \tag{4.3.1}$$

Clearly $F_{X_n}(x) = F_X(x)$ for all $x$ in the support of $X$, so that $X_n \xrightarrow{D} X$. On the other hand, the sequence $X_n$ does not get close to $X$. In particular, $X_n \not\rightarrow X$ in probability.

**Example 4.3.1.** Let $\overline{X}_n$ have the cdf

$$F_n(\overline{x}) = \int_{-\infty}^{\overline{x}} \frac{1}{\sqrt{1/n}\sqrt{2\pi}} e^{-nw^2/2}\, dw.$$

If the change of variable $v = \sqrt{n}w$ is made, we have

$$F_n(\overline{x}) = \int_{-\infty}^{\sqrt{n}\overline{x}} \frac{1}{\sqrt{2\pi}} e^{-v^2/2}\, dv.$$

It is clear that

$$\lim_{n\to\infty} F_n(\overline{x}) = \begin{cases} 0 & \overline{x} < 0 \\ \frac{1}{2} & \overline{x} = 0 \\ 1 & \overline{x} > 0. \end{cases}$$

Now the function

$$F(\overline{x}) = \begin{cases} 0 & \overline{x} < 0 \\ 1 & \overline{x} \geq 0, \end{cases}$$

is a cdf and $\lim_{n\to\infty} F_n(\overline{x}) = F(\overline{x})$ at every point of continuity of $F(\overline{x})$. To be sure, $\lim_{n\to\infty} F_n(0) \neq F(0)$, but $F(\overline{x})$ is not continuous at $\overline{x} = 0$. Accordingly, the sequence $\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots$ converges in distribution to a random variable that has a degenerate distribution at $\overline{x} = 0$. ■

**Example 4.3.2.** Even if a sequence $X_1, X_2, X_3, \ldots$ converges in distribution to a random variable $X$, we cannot in general determine the distribution of $X$ by taking the limit of the pmf of $X_n$. This is illustrated by letting $X_n$ have the pmf

$$p_n(x) = \begin{cases} 1 & x = 2 + n^{-1} \\ 0 & \text{elsewhere.} \end{cases}$$

Clearly, $\lim_{n\to\infty} p_n(x) = 0$ for all values of $x$. This may suggest that $X_n$, for $n = 1, 2, 3, \ldots$, does not converge in distribution. However, the cdf of $X_n$ is

$$F_n(x) = \begin{cases} 0 & x < 2 + n^{-1} \\ 1 & x \geq 2 + n^{-1}, \end{cases}$$

and

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \leq 2 \\ 1 & x > 2. \end{cases}$$

Since

$$F(x) = \begin{cases} 0 & x < 2 \\ 1 & x \geq 2, \end{cases}$$

is a cdf, and since $\lim_{n\to\infty} F_n(x) = F(x)$ at all points of continuity of $F(x)$, the sequence $X_1, X_2, X_3, \ldots$ converges in distribution to a random variable with cdf $F(x)$. ■

The last example showed in general that we cannot determine limiting distributions by considering pmfs or pdfs. But under certain conditions we can determine convergence in distribution by considering the sequence of pdfs as the following example shows.

**Example 4.3.3.** Let $T_n$ have a $t$-distribution with $n$ degrees of freedom, $n = 1, 2, 3, \ldots$. Thus its cdf is

$$F_n(t) = \int_{-\infty}^{t} \frac{\Gamma[(n+1))/2]}{\sqrt{\pi n}\,\Gamma(n/2)} \frac{1}{(1+y^2/n)^{(n+1)/2}}\,dy,$$

where the integrand is the pdf $f_n(y)$ of $T_n$. Accordingly,

$$\lim_{n\to\infty} F_n(t) = \lim_{n\to\infty} \int_{-\infty}^{t} f_n(y)\,dy = \int_{-\infty}^{t} \lim_{n\to\infty} f_n(y)\,dy.$$

By a result in analysis, (Lebesgue Dominated Convergence Theorem), that allows us to interchange the order of the limit and integration provided $|f_n(y)|$ is dominated by a function which is integrable. This is true because

$$|f_n(y)| \leq 10 f_1(y)$$

and

$$\int_{-\infty}^{t} 10 f_1(y)\,dy = \frac{10}{\pi} \arctan t < \infty,$$

for all real $t$. Hence, we can find the limiting distribution by finding the limit of the pdf of $T_n$. It is

$$\begin{aligned} \lim_{n\to\infty} f_n(y) &= \lim_{n\to\infty} \left\{ \frac{\Gamma[(n+1)/2]}{\sqrt{n/2}\,\Gamma(n/2)} \right\} \lim_{n\to\infty} \left\{ \frac{1}{(1+y^2/n)^{1/2}} \right\} \\ &\quad \times \lim_{n\to\infty} \left\{ \frac{1}{\sqrt{2\pi}} \left[ \left(1+\frac{y^2}{n}\right) \right]^{-n/2} \right\}. \end{aligned}$$

Using the fact from elementary calculus that

$$\lim_{n\to\infty} \left(1+\frac{y^2}{n}\right)^{n} = e^{y^2},$$

the limit associated with the third factor is clearly the pdf of the standard normal distribution. The second limit obviously equals 1. By Remark 4.3.2, the first limit also equals 1. Thus we have

$$\lim_{n\to\infty} F_n(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,$$

and hence $T_n$ has a limiting standard normal distribution. ■

**Remark 4.3.2 (Stirling's Formula).** In advanced calculus the following approximation is derived,

$$\Gamma(k+1) \doteq \sqrt{2\pi} k^{k+1/2} e^{-k}. \tag{4.3.2}$$

This is known as *Stirling's formula* and it is an excellent approximation when $k$ is large. Because $\Gamma(k+1) = k!$, for $k$ an integer, this formula gives one an idea of how fast $k!$ grows. As Exercise 4.3.20 shows, this approximation can be used to show that the first limit in Example 4.3.3 is 1. ■

**Example 4.3.4 (Maximum of a Sample from a Uniform Distribution, Continued).** Recall Example 4.1.3, where $X_1, \ldots, X_n$ is a random sample from a uniform$(0, \theta)$ distribution. Again let $Y_n = \max\{X_1, \ldots, X_n\}$, but now consider the random variable $Z_n = n(\theta - Y_n)$. Let $t \in (0, n\theta)$. Then, using the cdf of $Y_n$, (4.1.7), the cdf of $Z_n$ is

$$\begin{aligned} P[Z_n \le t] &= P[Y_n \ge \theta - (t/n)] \\ &= 1 - \left(\frac{\theta - (t/n)}{\theta}\right)^n \\ &= 1 - \left(1 - \frac{t/\theta}{n}\right)^n \\ &\to 1 - e^{-t/\theta}. \end{aligned}$$

Note that the last quantity is the cdf of an exponential random variable with mean $\theta$, (3.3.2). So we would say that $Z_n \xrightarrow{D} Z$, where $Z$ is distributed $\exp(\theta)$. ■

**Remark 4.3.3.** To simplify several of the proofs of this section, we make use of the $\underline{\lim}$ and $\overline{\lim}$ of a sequence. For readers who are unfamilar with these concepts, we discuss them in Appendix A. In this brief remark, we highlight the properties needed for understanding the proofs. Let $\{a_n\}$ be a sequence of real numbers and define the two subsequences,

$$b_n = \sup\{a_n, a_{n+1}, \ldots\}, \quad n = 1, 2, 3 \ldots, \tag{4.3.3}$$

$$c_n = \inf\{a_n, a_{n+1}, \ldots\}, \quad n = 1, 2, 3 \ldots. \tag{4.3.4}$$

While $\{c_n\}$ is a nondecreasing sequence, $\{b_n\}$ is a nonincreasing sequence. Hence, their limits always exist, (may be $\pm\infty$). Denote them respectively by $\underline{\lim}_{n\to\infty} a_n$ and $\overline{\lim}_{n\to\infty} a_n$. Further, $c_n \le a_n \le b_n$, for all $n$. Hence, by the Sandwich Theorem (see Theorem A.2.1 of Appendix A), if $\underline{\lim}_{n\to\infty} a_n = \overline{\lim}_{n\to\infty} a_n$, then $\lim_{n\to\infty} a_n$ exists and is given by $\lim_{n\to\infty} a_n = \overline{\lim}_{n\to\infty} a_n$.

As discussed in the appendix, several other properties of these concepts are useful. For example, suppose $\{p_n\}$ is a sequence of probabilities and $\overline{\lim}_{n\to\infty} p_n = 0$. Then by the Sandwich Theorem, since $0 \le p_n \le \sup\{p_n, p_{n+1}, \ldots\}$, for all $n$, we have $\lim_{n\to\infty} p_n = 0$. Also, for any two sequences $\{a_n\}$ and $\{b_n\}$, it easily follows that $\overline{\lim}_{n\to\infty}(a_n + b_n) \le \overline{\lim}_{n\to\infty} a_n + \overline{\lim}_{n\to\infty} b_n$. ■

As the following theorem shows, convergence in distribution is weaker than convergence in probability. Thus convergence in distribution is often called weak convergence.

**Theorem 4.3.1.** *If $X_n$ converges to $X$ in probability, then $X_n$ converges to $X$ in distribution.*

*Proof:* Let $x$ be a point of continuity of $F_X(x)$. Let $\epsilon > 0$. We have,

$$\begin{aligned} F_{X_n}(x) &= P[X_n \le x] \\ &= P[\{X_n \le x\} \cap \{|X_n - X| < \epsilon\}] + P[\{X_n \le x\} \cap \{|X_n - X| \ge \epsilon\}] \\ &\le P[X \le x + \epsilon] + P[|X_n - X| \ge \epsilon]. \end{aligned}$$

Based on this inequality and the fact that $X_n \xrightarrow{P} X$, we see that

$$\overline{\lim_{n\to\infty}} F_{X_n}(x) \le F_X(x + \epsilon). \tag{4.3.5}$$

To get a lower bound, we proceed similarly with the complement to show that

$$P[X_n > x] \le P[X \ge x - \epsilon] + P[|X_n - X| \ge \epsilon].$$

Hence,

$$\underline{\lim_{n\to\infty}} F_{X_n}(x) \ge F_X(x - \epsilon). \tag{4.3.6}$$

Using a relationship between $\overline{\lim}$ and $\underline{\lim}$, it follows from (4.3.5) and (4.3.6) that

$$F_X(x - \epsilon) \le \underline{\lim_{n\to\infty}} F_{X_n}(x) \le \overline{\lim_{n\to\infty}} F_{X_n}(x) \le F_X(x + \epsilon).$$

Letting $\epsilon \downarrow 0$ gives us the desired result. ■

Reconsider the sequence of random variables $\{X_n\}$ defined by expression (4.3.1). Here, $X_n \xrightarrow{D} X$ but $X_n \overset{P}{\not\to} X$. So in general the converse of the above theorem is not true. However, it is true if $X$ is degenerate as shown by the following theorem.

**Theorem 4.3.2.** *If $X_n$ converges to the constant $b$ in distribution, then $X_n$ converges to $b$ in probability.*

*Proof:* Let $\epsilon > 0$ be given. Then,

$$\lim_{n\to\infty} P[|X_n - b| \le \epsilon] = \lim_{n\to\infty} F_{X_n}(b + \epsilon) - \lim_{n\to\infty} F_{X_n}(b - \epsilon - 0) = 1 - 0 = 1,$$

which is the desired result. ■

A result that will prove quite useful is the following:

**Theorem 4.3.3.** *Suppose $X_n$ converges to $X$ in distribution and $Y_n$ converges in probability to 0. Then $X_n + Y_n$ converges to $X$ in distribution.*

The proof is similar to that of Theorem 4.3.2 and is left to Exercise 4.3.12. We often use this last result as follows. Suppose it is difficult to show that $X_n$ converges to $X$ in distribution; but it is easy to show that $Y_n$ converges in distribution to $X$ and that $X_n - Y_n$ converges to 0 in probability. Hence, by this last theorem, $X_n = Y_n + (X_n - Y_n) \xrightarrow{D} X$, as desired.

The next two theorems state general results. A proof of the first result can be found in a more advanced text while the second, Slutsky's Theorem, follows similar to that of Theorem 4.3.1.

**Theorem 4.3.4.** *Suppose $X_n$ converges to $X$ in distribution and $g$ is a continuous function on the support of $X$. Then $g(X_n)$ converges to $g(X)$ in distribution.*

**Theorem 4.3.5 (Slutsky's Theorem).** *Let $X_n$, $X$, $A_n$, $B_n$ be random variables and let $a$ and $b$ be constants. If $X_n \xrightarrow{D} X$, $A_n \xrightarrow{P} a$, and $B_n \xrightarrow{P} b$, then*

$$A_n + B_n X_n \xrightarrow{D} a + bX.$$

### 4.3.1 Bounded in Probability

Another useful concept, related to convergence in distribution, is boundedness in probability of a sequence of random variables.

First consider any random variable $X$ with cdf $F_X(x)$. Then given $\epsilon > 0$, we can bound $X$ in the following way. Because the lower limit of $F_X$ is 0 and its upper limit is 1 we can find $\eta_1$ and $\eta_2$ such that

$$F_X(x) < \epsilon/2 \text{ for } x \le \eta_1 \text{ and } F_X(x) > 1 - (\epsilon/2) \text{ for } x \ge \eta_2.$$

Let $\eta = \max\{|\eta_1|, |\eta_2|\}$ then

$$P[|X| \le \eta] = F_X(\eta) - F_X(-\eta - 0) \ge 1 - (\epsilon/2) - (\epsilon/2) = 1 - \epsilon. \tag{4.3.7}$$

Thus random variables which are not bounded (e.g., $X$ is $N(0,1)$) are still bounded in the above way. This is a useful concept for sequences of random variables which we define next.

**Definition 4.3.2 (Bounded in Probability).** *We say that the sequence of random variables $\{X_n\}$ is bounded in probability, if for all $\epsilon > 0$ there exists a constant $B_\epsilon > 0$ and an integer $N_\epsilon$ such that*

$$n \ge N_\epsilon \Rightarrow P[|X_n| \le B_\epsilon] \ge 1 - \epsilon.$$

Next, consider a sequence of random variables $\{X_n\}$ which converge in distribution to a random variable $X$ which has cdf $F$. Let $\epsilon > 0$ be given and choose $\eta$ so that (4.3.7) holds for $X$. We can always choose $\eta$ so that $\eta$ and $-\eta$ are continuity points of $F$. We then have,

$$\lim_{n\to\infty} P[|X_n| \le \eta] \ge \lim_{n\to\infty} F_{X_n}(\eta) - \lim_{n\to\infty} F_{X_n}(-\eta - 0) = F_X(\eta) - F_X(-\eta) \ge 1 - \epsilon.$$

To be precise, we can then choose $N$ so large that $P[|X_n| \leq \eta] \geq 1 - \epsilon$, for $n \geq N$. We have thus proved the following theorem

**Theorem 4.3.6.** *Let $\{X_n\}$ be a sequence of random variables and let $X$ be a random variable. If $X_n \to X$ in distribution, then $\{X_n\}$ is bounded in probability.*

As the following example shows, the converse of this theorem is not true.

**Example 4.3.5.** Take $\{X_n\}$ to be the following sequence of degenerate random variables. For $n = 2m$ even, $X_{2m} = 2 + (1/(2m))$ with probability one. For $n = 2m - 1$ odd, $X_{2m-1} = 1 + (1/(2m))$ with probability one. Then the sequence $\{X_2, X_4, X_6, \ldots\}$ converges in distribution to the degenerate random variable $Y = 2$, while the sequence $\{X_1, X_3, X_5, \ldots\}$ converges in distribution to the degenerate random variable $W = 1$. Since the distributions of $Y$ and $W$ are not the same, the sequence $\{X_n\}$ does not converge in distribution. Because all of the mass of the sequence $\{X_n\}$ is in the interval $[1, 5/2]$, however, the sequence $\{X_n\}$ is bounded in probability. ■

One way of thinking of a sequence which is bounded in probability (or one which is converging to a random variable in distribution) is that the probability mass of $|X_n|$ is not escaping to $\infty$. At times we can use boundedness in probability instead of convergence in distribution. A property we will need later is given in the following theorem

**Theorem 4.3.7.** *Let $\{X_n\}$ be a sequence of random variables bounded in probability and let $\{Y_n\}$ be a sequence of random variables which converge to 0 in probability. Then*

$$X_n Y_n \xrightarrow{P} 0.$$

*Proof:* Let $\epsilon > 0$ be given. Choose $B_\epsilon > 0$ and an integer $N_\epsilon$ such that

$$n \geq N_\epsilon \Rightarrow P[|X_n| \leq B_\epsilon] \geq 1 - \epsilon.$$

Then,

$$\begin{aligned}
\overline{\lim_{n\to\infty}} P[|X_n Y_n| \geq \epsilon] &\leq \overline{\lim_{n\to\infty}} P[|X_n Y_n| \geq \epsilon, |X_n| \leq B_\epsilon] \\
&\quad + \overline{\lim_{n\to\infty}} P[|X_n Y_n| \geq \epsilon, |X_n| > B_\epsilon] \\
&\leq \overline{\lim_{n\to\infty}} P[|Y_n| \geq \epsilon/B_\epsilon] + \epsilon = \epsilon. \qquad (4.3.8)
\end{aligned}$$

From which the desired result follows. ■

## 4.3.2 Δ-Method

Recall a common problem discussed in the last three chapters is the situation where we know the distribution of a random variable, but we want to determine the distribution of a function of it. This is also true in asymptotic theory and Theorems 4.3.4 and 4.3.5 are illustrations of this. Another such result is called the Δ-method. To establish this result we need a convenient form of the mean value theorem with remainder, sometimes called Young's Theorem; see Hardy (1992) or Lehmann (1999). Suppose $g(x)$ is differentiable at $x$. Then we can write,

$$g(y) = g(x) + g'(x)(y - x) + o(|y - x|), \tag{4.3.9}$$

where the notation $o$ means

$$a = o(b) \text{ if and only if } \tfrac{a}{b} \to 0, \text{ as } b \to 0.$$

The *little o* notation is used in terms of convergence in probability, also. We often write $o_p(X_n)$, which means

$$Y_n = o_p(X_n) \text{ if and only if } \tfrac{Y_n}{X_n} \xrightarrow{P} 0, \text{ as } n \to \infty. \tag{4.3.10}$$

There is a corresponding *big O* notation which is given by

$$Y_n = O_p(X_n) \text{ if and only if } \tfrac{Y_n}{X_n} \text{ is bounded in probability as } n \to \infty. \tag{4.3.11}$$

The following theorem illustrates the little-o notation, but it is also serves as a lemma for Theorem 4.3.9.

**Theorem 4.3.8.** *Suppose $\{Y_n\}$ is a sequence of random variables which is bounded in probability. Suppose $X_n = o_p(Y_n)$. Then $X_n \xrightarrow{P} 0$, as $n \to \infty$.*

*Proof:* Let $\epsilon > 0$ be given. Because the sequence $\{Y_n\}$ is bounded in probability, there exists positive constants $N_\epsilon$ and $B_\epsilon$ such that

$$n \geq N_\epsilon \Longrightarrow P[|Y_n| \leq B_\epsilon] \geq 1 - \epsilon. \tag{4.3.12}$$

Also, because $X_n = o_p(Y_n)$, we have

$$\frac{X_n}{Y_n} \xrightarrow{P} 0, \tag{4.3.13}$$

as $n \to \infty$. We then have,

$$\begin{aligned} P[|X_n| \geq \epsilon] &= P[|X_n| \geq \epsilon, |Y_n| \leq B_\epsilon] + P[|X_n| \geq \epsilon, |Y_n| > B_\epsilon] \\ &\leq P\left[\frac{X_n}{|Y_n|} \geq \frac{\epsilon}{B_\epsilon}\right] + P\left[|Y_n| > B_\epsilon\right]. \end{aligned}$$

By (4.3.13) and (4.3.12), respectively, the first and second terms on the right-side can be made arbitrarily small by choosing $n$ sufficiently large. Hence, the result is true. ■

**Theorem 4.3.9.** *Let $\{X_n\}$ be a sequence of random variables such that*

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2). \tag{4.3.14}$$

*Suppose the function $g(x)$ is differentiable at $\theta$ and $g'(\theta) \neq 0$. Then*

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2(g'(\theta))^2). \tag{4.3.15}$$

*Proof:* Using expression ( 4.3.9), we have,

$$g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + o_p(|X_n - \theta|),$$

where $o_p$ is interpretated as in ( 4.3.10). Rearranging, we have

$$\sqrt{n}(g(X_n) - g(\theta)) = g'(\theta)\sqrt{n}(X_n - \theta) + o_p(\sqrt{n}|X_n - \theta|).$$

Because (4.3.14) holds, Theorem 4.3.6 implies that $\sqrt{n}|X_n - \theta|$ is bounded in probability. Therefore by Theorem 4.3.8 $o_p(\sqrt{n}|X_n - \theta|) \to 0$, in probability. Hence by (4.3.14) and Theorem 4.3.1 the result follows. ■

Illustrations of the Δ- method can be found in Example 4.3.8 and the exercises.

### 4.3.3 Moment Generating Function Technique

To find the limiting distribution function of a random variable $X_n$ by using the definition obviously requires that we know $F_{X_n}(x)$ for each positive integer $n$. But it is often difficult to obtain $F_{X_n}(x)$ in closed form. Fortunately, if it exists, the mgf that corresponds to the cdf $F_{X_n}(x)$ often provides a convenient method of determining the limiting cdf.

The following theorem, which is essentially Curtiss' modification of a theorem of Lévy and Cramér, explains how the mgf may be used in problems of limiting distributions. A proof of the theorem is beyond of the scope of this book. It can readily be found in more advanced books; see, for instance, page 171 of Breiman (1968).

**Theorem 4.3.10.** *Let $\{X_n\}$ be a sequence of random variables with mgf $M_{X_n}(t)$ that exists for $-h < t < h$ for all $n$. Let $X$ be a random variable with mgf $M(t)$, which exists for $|t| \leq h_1 \leq h$. If $\lim_{n\to\infty} M_{X_n}(t) = M(t)$ for $|t| \leq h_1$, then $X_n \xrightarrow{D} X$.*

In this and the subsequent sections are several illustrations of the use of Theorem 4.3.10. In some of these examples it is convenient to use a certain limit that is established in some courses in advanced calculus. We refer to a limit of the form

$$\lim_{n\to\infty}\left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn},$$

where $b$ and $c$ do not depend upon $n$ and where $\lim_{n\to\infty}\psi(n) = 0$. Then

$$\lim_{n\to\infty}\left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn} = \lim_{n\to\infty}\left(1 + \frac{b}{n}\right)^{cn} = e^{bc}. \tag{4.3.16}$$

For example,

$$\lim_{n\to\infty}\left(1 - \frac{t^2}{n} + \frac{t^3}{n^{3/2}}\right)^{-n/2} = \lim_{n\to\infty}\left(1 - \frac{t^2}{n} + \frac{t^2/\sqrt{n}}{n}\right)^{-n/2}.$$

Here $b = -t^2, c = -\frac{1}{2}$, and $\psi(n) = t^2/\sqrt{n}$. Accordingly, for every fixed value of $t$, the limit is $e^{t^2/2}$.

**Example 4.3.6.** Let $Y_n$ have a distribution that is $b(n,p)$. Suppose that the mean $\mu = np$ is the same for every $n$; that is, $p = \mu/n$, where $\mu$ is a constant. We shall find the limiting distribution of the binomial distribution, when $p = \mu/n$, by finding the limit of $M(t;n)$. Now

$$M(t;n) = E(e^{tY_n}) = [(1-p)+pe^t]^n = \left[1 + \frac{\mu(e^t-1)}{n}\right]^n$$

for all real values of $t$. Hence we have

$$\lim_{n\to\infty} M(t;n) = e^{\mu(e^t-1)}$$

for all real values of $t$. Since there exists a distribution, namely the Poisson distribution with mean $\mu$, that has mgf $e^{\mu(e^t-1)}$, then, in accordance with the theorem and under the conditions stated, it is seen that $Y_n$ has a limiting Poisson distribution with mean $\mu$.

Whenever a random variable has a limiting distribution, we may, if we wish, use the limiting distribution as an approximation to the exact distribution function. The result of this example enables us to use the Poisson distribution as an approximation to the binomial distribution when $n$ is large and $p$ is small. To illustrate the use of the approximation, let $Y$ have a binomial distribution with $n = 50$ and $p = \frac{1}{25}$. Then

$$Pr(Y \le 1) = (\tfrac{24}{25})^{50} + 50(\tfrac{1}{25})(\tfrac{24}{25})^{49} = 0.400,$$

approximately. Since $\mu = np = 2$, the Poisson approximation to this probability is

$$e^{-2} + 2e^{-2} = 0.406. \quad \blacksquare$$

**Example 4.3.7.** Let $Z_n$ be $\chi^2(n)$. Then the mgf of $Z_n$ is $(1-2t)^{-n/2}$, $t < \frac{1}{2}$. The mean and the variance of $Z_n$ are, respectively, $n$ and $2n$. The limiting distribution of the random variable $Y_n = (Z_n - n)/\sqrt{2n}$ will be investigated. Now the mgf of $Y_n$ is

$$\begin{aligned} M(t;n) &= E\left\{\exp\left[t\left(\frac{Z_n - n}{\sqrt{2n}}\right)\right]\right\} \\ &= e^{-tn/\sqrt{2n}}E(e^{tZ_n/\sqrt{2n}}) \\ &= \exp\left[-\left(t\sqrt{\frac{2}{n}}\right)\left(\frac{n}{2}\right)\right]\left(1 - 2\frac{t}{\sqrt{2n}}\right)^{-n/2}, \quad t < \frac{\sqrt{2n}}{2}. \end{aligned}$$

This may be written in the form

$$M(t;n) = \left(e^{t\sqrt{2/n}} - t\sqrt{\frac{2}{n}}e^{t\sqrt{2/n}}\right)^{-n/2}, \quad t < \sqrt{\frac{n}{2}}.$$

In accordance with Taylor's formula, there exists a number $\xi(n)$, between 0 and $t\sqrt{2/n}$, such that

$$e^{t\sqrt{2/n}} = 1 + t\sqrt{\frac{2}{n}} + \frac{1}{2}\left(t\sqrt{\frac{2}{n}}\right)^2 + \frac{e^{\xi(n)}}{6}\left(t\sqrt{\frac{2}{n}}\right)^3.$$

If this sum is substituted for $e^{t\sqrt{2/n}}$ in the last expression for $M(t;n)$, it is seen that

$$M(t;n) = \left(1 - \frac{t^2}{n} + \frac{\psi(n)}{n}\right)^{-n/2},$$

where

$$\psi(n) = \frac{\sqrt{2}t^3 e^{\xi(n)}}{3\sqrt{n}} - \frac{\sqrt{2}t^3}{\sqrt{n}} - \frac{2t^4 e^{\xi(n)}}{3n}.$$

Since $\xi(n) \to 0$ as $n \to \infty$, then $\lim \psi(n) = 0$ for every fixed value of $t$. In accordance with the limit proposition cited earlier in this section, we have

$$\lim_{n\to\infty} M(t;n) = e^{t^2/2}$$

for all real values of $t$. That is, the random variable $Y_n = (Z_n - n)/\sqrt{2n}$ has a limiting standard normal distribution. ■

**Example 4.3.8 (Example 4.3.7 Continued).** In the notation of the last example, we showed that

$$\sqrt{n}\left[\frac{1}{\sqrt{2n}}Z_n - \frac{1}{\sqrt{2}}\right] \xrightarrow{D} N(0,1). \tag{4.3.17}$$

For this situation, though, there are times when we are interested in the square-root of $Z_n$. Let $g(t) = \sqrt{t}$ and let $W_n = g(Z_n/(\sqrt{2}n)) = (Z_n/(\sqrt{2}n))^{1/2}$. Note that $g(1/\sqrt{2}) = 1/2^{1/4}$ and $g'(1/\sqrt{2}) = 2^{-3/4}$. Therefore, by the $\Delta$-method, Theorem 4.3.9, and (4.3.17), we have

$$\sqrt{n}\left[W_n - 1/2^{1/4}\right] \xrightarrow{D} N(0, 2^{-3/2}). \quad ■ \tag{4.3.18}$$

## EXERCISES

**4.3.1.** Let $\overline{X}_n$ denote the mean of a random sample of size $n$ from a distribution that is $N(\mu, \sigma^2)$. Find the limiting distribution of $\overline{X}_n$.

**4.3.2.** Let $Y_1$ denote the first order statistic of a random sample of size $n$ from a distribution that has pdf $f(x) = e^{-(x-\theta)}$, $\theta < x < \infty$, zero elsewhere. Let $Z_n = n(Y_1 - \theta)$. Investigate the limiting distribution of $Z_n$.

**4.3.3.** Let $Y_n$ denote the $n$th order statistic of a random sample from a distribution of the continuous type that has cdf $F(x)$ and pdf $f(x) = F'(x)$. Find the limiting distribution of $Z_n = n[1 - F(Y_n)]$.

**4.3.4.** Let $Y_2$ denote the second order statistic of a random sample of size $n$ from a distribution of the continuous type that has cdf $F(x)$ and pdf $f(x) = F'(x)$. Find the limiting distribution of $W_n = nF(Y_2)$.

**4.3.5.** Let the pmf of $Y_n$ be $p_n(y) = 1$, $y = n$, zero elsewhere. Show that $Y_n$ does not have a limiting distribution. (In this case, the probability has "escaped" to infinity.)

**4.3.6.** Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from a distribution that is $N(\mu, \sigma^2)$, where $\sigma^2 > 0$. Show that the sum $Z_n = \sum_1^n X_i$ does not have a limiting distribution.

**4.3.7.** Let $X_n$ have a gamma distribution with parameter $\alpha = n$ and $\beta$, where $\beta$ is not a function of $n$. Let $Y_n = X_n/n$. Find the limiting distribution of $Y_n$.

**4.3.8.** Let $Z_n$ be $\chi^2(n)$ and let $W_n = Z_n/n^2$. Find the limiting distribution of $W_n$.

**4.3.9.** Let $X$ be $\chi^2(50)$. Approximate $P(40 < X < 60)$.

**4.3.10.** Let $p = 0.95$ be the probability that a man, in a certain age group, lives at least 5 years.

**(a)** If we are to observe 60 such men and if we assume independence, find the probability that at least 56 of them live 5 or more years.

**(b)** Find an approximation to the result of part (a) by using the Poisson distribution.
*Hint:* Redefine $p$ to be 0.05 and $1 - p = 0.95$.

**4.3.11.** Let the random variable $Z_n$ have a Poisson distribution with parameter $\mu = n$. Show that the limiting distribution of the random variable $Y_n = (Z_n - n)/\sqrt{n}$ is normal with mean zero and variance 1.

**4.3.12.** Prove Theorem 4.3.3

**4.3.13.** Let $X_n$ and $Y_n$ have a bivariate normal distribution with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ (free of $n$) but $\rho = 1 - 1/n$. Consider the conditional distribution of $Y_n$, given $X_n = x$. Investigate the limit of this conditional distribution as $n \to \infty$. What is the limiting distribution if $\rho = -1 + 1/n$? Reference to these facts is made in the Remark in Section 2.4.

**4.3.14.** Let $\overline{X}_n$ denote the mean of a random sample of size $n$ from a Poisson distribution with parameter $\mu = 1$.

**(a)** Show that the mgf of $Y_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma = \sqrt{n}(\overline{X}_n - 1)$ is given by $\exp[-t\sqrt{n} + n(e^{t/\sqrt{n}} - 1)]$.

**(b)** Investigate the limiting distribution of $Y_n$ as $n \to \infty$.
*Hint:* Replace, by its MacLaurin's series, the expression $e^{t/\sqrt{n}}$, which is in the exponent of the mgf of $Y_n$.

**4.3.15.** Using Exercise 4.3.14, find the limiting distribution of $\sqrt{n}(\sqrt{\overline{X}_n} - 1)$.

**4.3.16.** Let $\overline{X}_n$ denote the mean of a random sample of size $n$ from a distribution that has pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere.

**(a)** Show that the mgf $M(t; n)$ of $Y_n = \sqrt{n}(\overline{X}_n - 1)$ is

$$M(t; n) = [e^{t/\sqrt{n}} - (t/\sqrt{n})e^{t/\sqrt{n}}]^{-n}, \quad t < \sqrt{n}.$$

(b) Find the limiting distribution of $Y_n$ as $n \to \infty$.

Exercises 4.3.14 and 4.3.16 are special instances of an important theorem that will be proved in the next section.

**4.3.17.** Using Exercise 4.3.16, find the limiting distribution of $\sqrt{n}(\sqrt{\overline{X}_n} - 1)$.

**4.3.18.** Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a distribution with pdf $f(x) = e^{-x}, 0 < x < \infty$, zero elsewhere. Determine the limiting distribution of $Z_n = (Y_n - \log n)$.

**4.3.19.** Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a distribution with pdf $f(x) = 5x^4, 0 < x < 1$, zero elsewhere. Find $p$ so that $Z_n = n^p Y_1$ converges in distribution.

**4.3.20.** Use Stirling's formula, (4.3.2), to show that the first limit in Example 4.3.3 is 1.

## 4.4 Central Limit Theorem

It was seen (Section 3.4) that, if $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$, the random variable

$$\frac{\sum_1^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$$

is, for every positive integer $n$, normally distributed with zero mean and unit variance. In probability theory there is a very elegant theorem called the *central limit theorem*. A special case of this theorem asserts the remarkable and important fact that if $X_1, X_2, \ldots, X_n$ denote the observations of a random sample of size $n$ from any distribution having finite variance $\sigma^2 > 0$ (and hence finite mean $\mu$), then the random variable $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ converges in distribution to a random variable having a standard normal distribution. Thus, whenever the conditions of the theorem are satisfied, for large $n$ the random variable $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ has an approximate normal distribution with mean zero and variance 1. It will then be possible to use this approximate normal distribution to compute approximate probabilities concerning $\overline{X}$. In the statistical problem where $\mu$ is unknown, we will use this approximate distribution of $\overline{X}_n$ to establish approximate confidence intervals for $\mu$; see Section 5.4.

We will often use the notation $Y_n$ has a limiting standard normal distribution to mean that $Y_n$ converges in distribution to a standard normal random variable; see Remark 4.3.1.

The more general form of the theorem is stated, but it is proved only in the modified case. However, this is exactly the proof of the theorem that would be given if we could use the characteristic function in place of the mgf.

**Theorem 4.4.1.** *Let $X_1, X_2, \ldots, X_n$ denote the observations of a random sample from a distribution that has mean $\mu$ and positive variance $\sigma^2$. Then the random variable $Y_n = (\sum_1^n X_i - n\mu)/\sqrt{n}\sigma = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ converges in distribution to a random variable which has a normal distribution with mean zero and variance 1.*

*Proof:* For this proof, additionaly assume that the mgf $M(t) = E(e^{tX})$ exists for $-h < t < h$. If one replaces the mgf by the characteristic function $\varphi(t) = E(e^{itX})$, which always exists, then our proof is essentially the same as the proof in a more advanced course which uses characteristic functions.

The function

$$m(t) = E[e^{t(X-\mu)}] = e^{-\mu t}M(t)$$

also exists for $-h < t < h$. Since $m(t)$ is the mgf for $X - \mu$, it must follow that $m(0) = 1$, $m'(0) = E(X - \mu) = 0$, and $m''(0) = E[(X - \mu)^2] = \sigma^2$. By Taylor's formula there exists a number $\xi$ between 0 and $t$ such that

$$\begin{aligned} m(t) &= m(0) + m'(0)t + \frac{m''(\xi)t^2}{2} \\ &= 1 + \frac{m''(\xi)t^2}{2}. \end{aligned}$$

If $\sigma^2 t^2/2$ is added and subtracted, then

$$m(t) = 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(\xi) - \sigma^2]t^2}{2} \tag{4.4.1}$$

Next consider $M(t;n)$, where

$$\begin{aligned} M(t;n) &= E\left[\exp\left(t\frac{\sum X_i - n\mu}{\sigma\sqrt{n}}\right)\right] \\ &= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\exp\left(t\frac{X_2 - \mu}{\sigma\sqrt{n}}\right)\cdots\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right] \\ &= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\right]\cdots E\left[\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right] \\ &= \left\{E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]\right\}^n \\ &= \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n, \quad -h < \frac{t}{\sigma\sqrt{n}} < h. \end{aligned}$$

In Equation 4.4.1 replace $t$ by $t/\sigma\sqrt{n}$ to obtain

$$m\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{[m''(\xi) - \sigma^2]t^2}{2n\sigma^2},$$

where now $\xi$ is between 0 and $t/\sigma\sqrt{n}$ with $-h\sigma\sqrt{n} < t < h\sigma\sqrt{n}$. Accordingly,

$$M(t;n) = \left\{1 + \frac{t^2}{2n} + \frac{[m''(\xi) - \sigma^2]t^2}{2n\sigma^2}\right\}^n.$$

Since $m''(t)$ is continuous at $t = 0$ and since $\xi \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty}[m''(\xi) - \sigma^2] = 0.$$

The limit proposition (4.3.16) cited in Section 4.3 shows that

$$\lim_{n\to\infty} M(t;n) = e^{t^2/2},$$

for all real values of $t$. This proves that the random variable $Y_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ has a limiting standard normal distribution. ■

As cited in Remark 4.3.1, we say that $Y_n$ has a limiting standard normal distribution. We interpret this theorem as saying that, when $n$ is a large, fixed positive integer, the random variable $\overline{X}$ has an approximate normal distribution with mean $\mu$ and variance $\sigma^2/n$; and in applications we use the approximate normal pdf as though it were the exact pdf of $\overline{X}$.

Some illustrative examples, here and below, will help show the importance of this version of the central limit theorem.

**Example 4.4.1.** Let $\overline{X}$ denote the mean of a random sample of size 75 from the distribution that has the pdf

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

For this situation, it can be shown that $g(\overline{x})$ has a graph when $0 < \overline{x} < 1$ that is composed of arcs of 75 different polynomials of degree 74. The computation of such a probability as $P(0.45 < \overline{X} < 0.55)$ would be extremely laborious. The conditions of the theorem are satisfied, since $M(t)$ exists for all real values of $t$. Moreover, $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$, so that we have approximately

$$\begin{aligned} P(0.45 < \overline{X} < 0.55) &= P\left[\frac{\sqrt{n}(0.45-\mu)}{\sigma} < \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} < \frac{\sqrt{n}(0.55-\mu)}{\sigma}\right] \\ &= P[-1.5 < 30(\overline{X}-0.5) < 1.5] \\ &= 0.866, \end{aligned}$$

from Table III in Appendix B. ■

**Example 4.4.2.** Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $b(1,p)$. Here $\mu = p$, $\sigma^2 = p(1-p)$, and $M(t)$ exists for all real values of $t$. If $Y_n = X_1 + \cdots + X_n$, it is known that $Y_n$ is $b(n,p)$. Calculation of probabilities for $Y_n$, when we do not use the Poisson approximation, are simplified by making use of the fact that $(Y_n - np)/\sqrt{np(1-p)} = \sqrt{n}(\overline{X}_n - p)/\sqrt{p(1-p)} = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ has a limiting distribution that is normal with mean zero and variance 1. Frequently, statisticians say that $Y_n$, or more simply $Y$, has an approximate normal distribution with mean $np$ and variance $np(1-p)$. Even with $n$ as small as 10, with $p = \frac{1}{2}$ so that the binomial distribution is symmetric about $np = 5$, we note in Figure 4.4.1 how well the normal distribution, $N(5, \frac{5}{2})$, fits the binomial distribution, $b(10, \frac{1}{2})$, where the heights of the rectangles represent the probabilities of the respective integers $0, 1, 2, \ldots, 10$. Note that the area of the rectangle whose base is $(k - 0.5, k + 0.5)$ and the area under the normal pdf between $k - 0.5$ and $k + 0.5$ are *approximately* equal for each $k = 0, 1, 2, \ldots, 10$, even with $n = 10$. This example should help the reader understand Example 4.4.3. ■

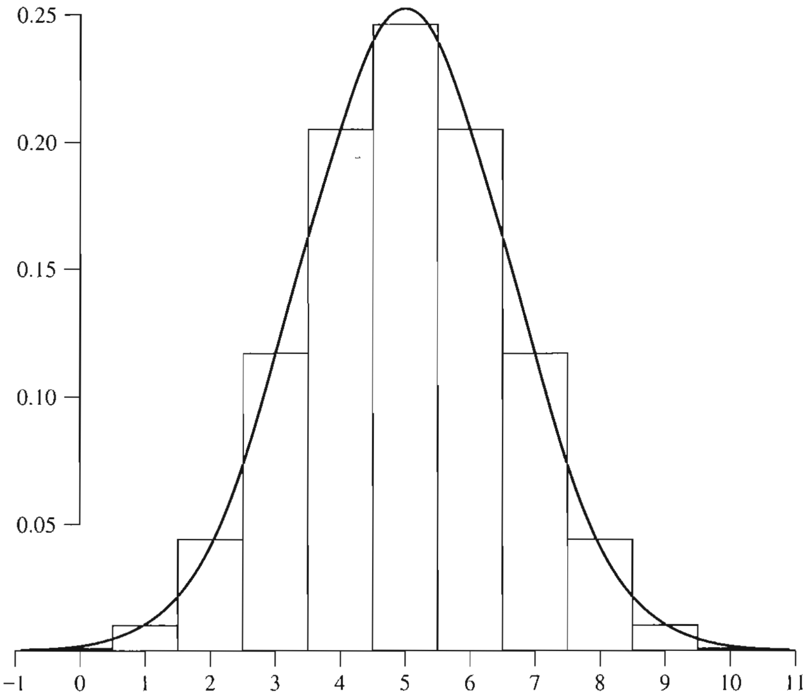


**Figure 4.4.1:** The $b\left(10, \frac{1}{2}\right)$ pmf overlaid by the $N\left(5, \frac{5}{2}\right)$ pdf

**Example 4.4.3.** With the background of Example 4.4.2, let $n = 100$ and $p = \frac{1}{2}$, and suppose that we wish to compute $P(Y = 48, 49, 50, 51, 52)$. Since $Y$ is a random variable of the discrete type, $\{Y = 48, 49, 50, 51, 52\}$ and $\{47.5 < Y < 52.5\}$ are equivalent events. That is, $P(Y = 48, 49, 50, 51, 52) = P(47.5 < Y < 52.5)$. Since $np = 50$ and $np(1-p) = 25$, the latter probability may be written

$$\begin{aligned} P(47.5 < Y < 52.5) &= P\left(\frac{47.5-50}{5} < \frac{Y-50}{5} < \frac{52.5-50}{5}\right) \\ &= P\left(-0.5 < \frac{Y-50}{5} < 0.5\right). \end{aligned}$$

Since $(Y-50)/5$ has an approximate normal distribution with mean zero and variance 1, Table III show this probability to be approximately 0.382.

The convention of selecting the event $47.5 < Y < 52.5$, instead of another event, say, $47.8 < Y < 52.3$, as the event equivalent to the event $Y = 48, 49, 50, 51, 52$ seems to have originated as: The probability, $P(Y = 48, 49, 50, 51, 52)$, can be interpreted as the sum of five rectangular areas where the rectangles have widths one but the heights are, respectively, $P(Y = 48), \ldots, P(Y = 52)$. If these rectangles are so located that the midpoints of their bases are, respectively, at the points $48, 49, \ldots, 52$ on a horizontal axis, then in approximating the sum of these areas by an area bounded by the horizontal axis, the graph of a normal pdf, and two ordinates, it seems reasonable to take the two ordinates at the points 47.5 and 52.5. This is called the **continuity correction.** ■

We know that $\overline{X}$ and $\sum_1^n X_i$ have approximate normal distributions, provided that $n$ is large enough. Later, we find that other statistics also have approximate normal distributions, and this is the reason that the normal distribution is so important to statisticians. That is, while not many underlying distributions are normal, the distributions of statistics calculated from random samples arising from these distributions are often very close to being normal.

Frequently, we are interested in functions of statistics that have approximate normal distributions. To illustrate, consider the sequence of random variable $Y_n$ of Example 4.4.2. As discussed there, $Y_n$ has an approximate $N[np, np(1-p)]$. So $np(1-p)$ is an important function of $p$ as it is the variance of $Y_n$. Thus, if $p$ is unknown, we might want to estimate the variance of $Y_n$. Since $E(Y_n/n) = p$, we might use $n(Y_n/n)(1 - Y_n/n)$ as such an estimator and would want to know something about the latter's distribution. In particular, does it also have an approximate normal distribution? If so, what are its mean and variance? To answer questions like these, we can apply the $\Delta$-method, Theorem 4.3.9.

As an illustration of the $\Delta$-method, we consider a function of the sample mean. We know that $\overline{X}_n$ converges in probability to $\mu$ and $\overline{X}_n$ is approximately $N(\mu, \sigma^2/n)$. Suppose that we are interested in a function of $\overline{X}_n$, say $u(\overline{X}_n)$, where $u$ is differentiable at $\mu$ and $u'(\mu) \neq 0$. By Theorem 4.3.9, $u(\overline{X}_n)$ is approximately distributed as $N\{u(\mu), [u'(\mu)]^2\sigma^2/n\}$. *More formally, we could say that*

$$\frac{u(\overline{X}_n) - u(\mu)}{\sqrt{[u'(\mu)]^2\sigma^2/n}}$$

has a limiting standard normal distribution.

**Example 4.4.4.** Let $Y_n$ (or $Y$ for simplicity) be $b(n, p)$. Thus $Y/n$ is approximately $N[p, p(1-p)/n]$. Statisticians often look for functions of statistics whose variances do not depend upon the parameter. Here the variance of $Y/n$ depends upon $p$. Can we find a function, say $u(Y/n)$, whose variance is essentially free of $p$? Since $Y/n$ converges in probability to $p$, we can approximate $u(Y/n)$ by the first two terms of its Taylor's expansion about $p$, namely by

$$u\left(\frac{Y}{n}\right) \doteq v\left(\frac{Y}{n}\right) = u(p) + \left(\frac{Y}{n} - p\right)u'(p).$$

Of course, $v(Y/n)$ is a linear function of $Y/n$ and thus also has an approximate normal distribution; clearly, it has mean $u(p)$ and variance

$$[u'(p)]^2\frac{p(1-p)}{n}.$$

But it is the latter that we want to be essentially free of $p$; thus we set it equal to a constant, obtaining the differential equation

$$u'(p) = \frac{c}{\sqrt{p(1-p)}}.$$

A solution of this is

$$u(p) = (2c) \arcsin \sqrt{p}.$$

If we take $c = \frac{1}{2}$, we have, since $u(Y/n)$ is approximately equal to $v(Y/n)$, that

$$u\left(\frac{Y}{n}\right) = \arcsin \sqrt{\frac{Y}{n}}$$

has an approximate normal distribution with mean $\arcsin \sqrt{p}$ and variance $1/4n$, which is free of $p$. ■

## EXERCISES

**4.4.1.** Let $\overline{X}$ denote the mean of a random sample of size 100 from a distribution that is $\chi^2(50)$. Compute an approximate value of $P(49 < \overline{X} < 51)$.

**4.4.2.** Let $\overline{X}$ denote the mean of a random sample of size 128 from a gamma distribution with $\alpha = 2$ and $\beta = 4$. Approximate $P(7 < \overline{X} < 9)$.

**4.4.3.** Let $Y$ be $b(72, \frac{1}{3})$. Approximate $P(22 \leq Y \leq 28)$.

**4.4.4.** Compute an approximate probability that the mean of a random sample of size 15 from a distribution having pdf $f(x) = 3x^2$, $0 < x < 1$, zero elsewhere, is between $\frac{3}{5}$ and $\frac{4}{5}$.

**4.4.5.** Let $Y$ denote the sum of the observations of a random sample of size 12 from a distribution having pmf $p(x) = \frac{1}{6}$, $x = 1, 2, 3, 4, 5, 6$, zero elsewhere. Compute an approximate value of $P(36 \leq Y \leq 48)$.
*Hint:* Since the event of interest is $Y = 36, 37, \ldots, 48$, rewrite the probability as $P(35.5 < Y < 48.5)$.

**4.4.6.** Let $Y$ be $b(400, \frac{1}{5})$. Compute an approximate value of $P(0.25 < Y/n)$.

**4.4.7.** If $Y$ is $b(100, \frac{1}{2})$, approximate the value of $P(Y = 50)$.

**4.4.8.** Let $Y$ be $b(n, 0.55)$. Find the smallest value of $n$ which is such that (approximately) $P(Y/n > \frac{1}{2}) \geq 0.95$.

**4.4.9.** Let $f(x) = 1/x^2$, $1 < x < \infty$, zero elsewhere, be the pdf of a random variable $X$. Consider a random sample of size 72 from the distribution having this pdf. Compute approximately the probability that more than 50 of the observations of the random sample are less than 3.

**4.4.10.** Forty-eight measurements are recorded to several decimal places. Each of these 48 numbers is rounded off to the nearest integer. The sum of the original 48 numbers is approximated by the sum of these integers. If we assume that the errors made by rounding off are iid and have a uniform distribution over the interval $(-\frac{1}{2}, \frac{1}{2})$, compute approximately the probability that the sum of the integers is within two units of the true sum.

**4.4.11.** We know that $\overline{X}$ is approximately $N(\mu, \sigma^2/n)$ for large $n$. Find the approximate distribution of $u(\overline{X}) = \overline{X}^3$.

**4.4.12.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean $\mu$. Thus $Y = \sum_{i=1}^{n} X_i$ has a Poisson distribution with mean $n\mu$. Moreover, $\overline{X} = Y/n$ is approximately $N(\mu, \mu/n)$ for large $n$. Show that $u(Y/n) = \sqrt{Y/n}$ is a function of $Y/n$ whose variance is essentially free of $\mu$.

## 4.5 *Asymptotics for Multivariate Distributions

In this section, we briefly discuss asymptotic concepts for sequences of random vectors. The concepts introduced for univariate random variables generalize in a straightforward manner to the multivariate case. Our development is brief and the interested reader can consult more advanced texts for more depth; see Serfling (1980).

We need some notation. For a vector $\mathbf{v} \in R^p$, recall that Euclidean norm of $\mathbf{v}$ is defined to be

$$\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{p} v_i^2}. \tag{4.5.1}$$

This norm satisfies the usual three properties given by

$$\begin{array}{ll} \text{(a). For all } \mathbf{v} \in R^p,\ \|\mathbf{v}\| \geq 0, \text{ and } \|\mathbf{v}\| = 0 \text{ if and only if } \mathbf{v} = \mathbf{0}. \\ \text{(b). For all } \mathbf{v} \in R^p \text{ and } a \in R,\ \|a\mathbf{v}\| = |a|\|\mathbf{v}\|. \\ \text{(c). For all } \mathbf{v}, \mathbf{u} \in R^p,\ \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|. \end{array} \tag{4.5.2}$$

Denote the standard basis of $R^p$ by the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_p$, where all the components of $\mathbf{e}_i$ are 0 except for the $ith$ component which is 1. Then we can always write any vector $\mathbf{v}' = (v_1, \ldots, v_p)$ as

$$\mathbf{v} = \sum_{i=1}^{p} v_i \mathbf{e}_i.$$

The following lemma will be useful:

**Lemma 4.5.1.** *Let $\mathbf{v}' = (v_1, \ldots, v_p)$ be any vector in $R^p$. Then,*

$$|v_j| \leq \|\mathbf{v}\| \leq \sum_{i=1}^{n} |v_i|, \quad \textit{for all } j = 1, \ldots, p\,. \tag{4.5.3}$$

*Proof*: Note that for all $j$,

$$v_j^2 \leq \sum_{i=1}^{p} v_i^2 = \|\mathbf{v}\|^2;$$

hence, taking the square root of this equality leads to the first part of the desired inequality. The second part is

$$\|\mathbf{v}\| = \|\sum_{i=1}^{p} v_i \mathbf{e}_i\| \leq \sum_{i=1}^{p} |v_i| \|\mathbf{e}_i\| = \sum_{i=1}^{p} |v_i|. \quad \blacksquare$$

Let $\{\mathbf{X}_n\}$ denote a sequence of $p$ dimensional vectors. Because the absolute value is the Euclidean norm in $R^1$, the definition of convergence in probability for random vectors is an immediate generalization:

**Definition 4.5.1.** *Let $\{\mathbf{X}_n\}$ be a sequence of $p$ dimensional vectors and let $\mathbf{X}$ be a random vector, all defined on the same sample space. We say that $\{\mathbf{X}_n\}$ converges in probability to $\mathbf{X}$ if*

$$\lim_{n\to\infty} P[\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon] = 0, \tag{4.5.4}$$

*for all $\epsilon > 0$. As in the univariate case, we write $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$.*

As the next theorem shows, convergence in probability of vectors is equivalent to componentwise convergence in probability.

**Theorem 4.5.1.** *Let $\{\mathbf{X}_n\}$ be a sequence of $p$ dimensional vectors and let $\mathbf{X}$ be a random vector, all defined on the same sample space. Then*

$$\mathbf{X}_n \xrightarrow{P} \mathbf{X} \text{ if and only if } X_{nj} \xrightarrow{P} X_j \text{ for all } j = 1, \ldots, p.$$

*Proof:* This follows immediately from Lemma 4.5.1. Suppose $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$. For any $j$, from the first part of the inequality (4.5.3), we have, for $\epsilon > 0$,

$$\epsilon \leq |X_{nj} - X_j| \leq \|\mathbf{X}_n - \mathbf{X}\|.$$

Hence,

$$\overline{\lim}_{n\to\infty} P[|X_{nj} - X_j| \geq \epsilon] \leq \overline{\lim}_{n\to\infty} P[\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon] = 0,$$

which is the desired result.

Conversely, if $X_{nj} \xrightarrow{P} X_j$ for all $j = 1, \ldots, p$, then by the second part of the inequality ( 4.5.3),

$$\epsilon \leq \|\mathbf{X}_n - \mathbf{X}\| \leq \sum_{i=1}^{p} |X_{nj} - X_j|,$$

for any $\epsilon > 0$. Hence,

$$\begin{aligned} \overline{\lim}_{n\to\infty} P[\|\mathbf{X}_n - \mathbf{X}\| \geq \epsilon] &\leq \overline{\lim}_{n\to\infty} P[\sum_{i=1}^{p} |X_{nj} - X_j| \geq \epsilon] \\ &\leq \sum_{i=1}^{p} \overline{\lim}_{n\to\infty} P[|X_{nj} - X_j| \geq \epsilon/p] = 0. \quad \blacksquare \end{aligned}$$

Based on this result many of the theorems involving convergence in probability can easily be extended to the multivariate setting. Some of these results are given in the exercises. This is true of statistical results, too. For example, in Section 4.2, we showed that if $X_1, \ldots, X_n$ is a random sample from the distribution of a random variable $X$ with mean, $\mu$, and variance, $\sigma^2$, then $\overline{X}_n$ and $S_n^2$ are consistent estimates of $\mu$ and $\sigma^2$. By the last theorem, we have that $(\overline{X}_n, S_n^2)$ is a consistent estimate of $(\mu, \sigma^2)$.

As another simple application, consider the multivariate analog of the sample mean and sample variance. Let $\{\mathbf{X}_n\}$ be a sequence of iid random vectors with common mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Denote the vector of means by

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i. \tag{4.5.5}$$

Of course, $\overline{\mathbf{X}}_n$ is just the vector of sample means, $(\overline{X}_1, \ldots, \overline{X}_p)'$. By the Weak Law of Large Numbers, Theorem 4.2.1, $\overline{X}_j \to \mu_j$, in probability, for each $j$. Hence, by Theorem 4.5.1, $\overline{\mathbf{X}}_n \to \boldsymbol{\mu}$, in probability.

How about the analog of the sample variances? Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})'$. Define the sample variances and covariances by,

$$S_{n,j}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2 \tag{4.5.6}$$

$$S_{n,jk} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k), \tag{4.5.7}$$

for $j, k = 1, \ldots, p$. Assuming finite fourth moments, the Weak Law of Large Numbers shows that all these componentwise sample variances and sample covariances converge in probability to distribution variances and covariances, respectively. If we define the $p \times p$ matrix $\mathbf{S}$ to be the matrix with the $jth$ diagonal entry $S_{n,j}^2$ and $(j,k)th$ entry $S_{n,jk}$ then $\mathbf{S} \to \boldsymbol{\Sigma}$, in probability.

The definition of convergence in distribution remains the same. We state it here in terms of vector notation.

**Definition 4.5.2.** *Let $\{\mathbf{X}_n\}$ be a sequence of random vectors with $\mathbf{X}_n$ having distribution function $F_n(\mathbf{x})$ and $\mathbf{X}$ be a random vector with distribution function $F(\mathbf{x})$. Then $\{\mathbf{X}_n\}$* **converges in distribution** *to $\mathbf{X}$ if*

$$\lim_{n\to\infty} F_n(\mathbf{x}) = F(\mathbf{x}), \tag{4.5.8}$$

*for all points $\mathbf{x}$ at which $F(\mathbf{x})$ is continuous. We write $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$.*

In the multivariate case, there are analogs to many of the theorems in Section 4.3. We state two important theorems without proof.

**Theorem 4.5.2.** *Let $\{\mathbf{X}_n\}$ be a sequence of random vectors which converge in distribution to a random vector $\mathbf{X}$ and let $g(\mathbf{x})$ be a function which is continuous on the support of $\mathbf{X}$. Then $g(\mathbf{X}_n)$ converges in distribution to $g(\mathbf{X})$.*

We can apply this theorem to show that convergence in distribution implies marginal convergence. Simply take $g(\mathbf{x}) = x_j$ where $\mathbf{X} = (x_1, \ldots, x_p)'$. Since $g$ is continuous, the desired result follows.

It is often difficult to determine convergence in distribution by using the definition. As in the univariate case, convergence in distribution is equivalent to convergence of moment generating functions, which we state in the following theorem.

**Theorem 4.5.3.** *Let $\{\mathbf{X}_n\}$ be a sequence of random vectors with $\mathbf{X}_n$ having distribution function $F_n(\mathbf{x})$ and moment generating function $M_n(\mathbf{t})$. Let $\mathbf{X}$ be a random vector with distribution function $F(\mathbf{x})$ and moment generating function $M(\mathbf{t})$. Then $\{\mathbf{X}_n\}$ converges in distribution to $\mathbf{X}$ if and only if for some $h > 0$,*

$$\lim_{n\to\infty} M_n(\mathbf{t}) = M(\mathbf{t}), \tag{4.5.9}$$

*for all $\mathbf{t}$ such that $\|\mathbf{t}\| < h$.*

The proof of this theorem can be found in more advanced books; see, for instance, Tucker (1967). Also, the usual proof is for characteristic functions instead of moment generating functions. As we mentioned previously, characteristic functions always exist, so convergence in distribution is completely characterized by convergence of corresponding characteristic functions.

*The moment generating function of $\mathbf{X}_n$ is $E[\exp\{\mathbf{t}'\mathbf{X}_n\}]$. Note that $\mathbf{t}'\mathbf{X}_n$* is a random variable. We can frequently use this and univariate theory to derive results in the multivariate case. A perfect example of this is the multivariate central limit theorem.

**Theorem 4.5.4 (Multivariate Central Limit Theorem).** *Let $\{\mathbf{X}_n\}$ be a sequence of iid random vectors with common mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ which is positive definite. Assume the common moment generating function $M(\mathbf{t})$ exists in an open neighborhood of $\mathbf{0}$. Let*

$$\mathbf{Y}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu}).$$

*Then $\mathbf{Y}_n$ converges in distribution to a $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution.*

*Proof.* Let $\mathbf{t} \in R^p$ be a vector in the stipulated neighbohood of $\mathbf{0}$. The moment generating function of $\mathbf{Y}_n$ is,

$$\begin{aligned} M_n(\mathbf{t}) &= E\left[\exp\left\{\mathbf{t}'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu})\right\}\right] \\ &= E\left[\exp\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{t}'(\mathbf{X}_i - \boldsymbol{\mu})\right\}\right] \\ &= E\left[\exp\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i\right\}\right], \end{aligned} \tag{4.5.10}$$

where $W_i = \mathbf{t}'(\mathbf{X}_i - \boldsymbol{\mu})$. Note that $W_1, \ldots, W_n$ are iid with mean 0 and variance $\mathrm{Var}(W_i) = \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}$. Hence, by the simple Central Limit Theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i \xrightarrow{D} N(0, \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}). \tag{4.5.11}$$

Expression (4.5.10), though, is the mgf of $(1/\sqrt{n})\sum_{i=1}^{n} W_i$ evaluated at 1. Therefore by (4.5.11), we must have

$$M_n(\mathbf{t}) = E\left[\exp\left\{1\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i\right\}\right] \to e^{1^2\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2} = e^{\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2}.$$

Because the last quantity is the moment generating function of a $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, we have the desired result. ■

Suppose $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ is a random sample from a distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Let $\overline{\mathbf{X}}_n$ be the vector of sample means. Then, from the Central Limit Theorem, we say that

$$\overline{\mathbf{X}}_n \text{ has an approximate } N_p\left(\boldsymbol{\mu}, \tfrac{1}{n}\boldsymbol{\Sigma}\right) \text{ distribution.} \tag{4.5.12}$$

A result that we use frequently concerns linear transformations. Its proof is obtained by using moment generating functions and is left as an exercise.

**Theorem 4.5.5.** *Let $\{\mathbf{X}_n\}$ be a sequence of $p$-dimensional random vectors. Suppose $\mathbf{X}_n \xrightarrow{D} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{A}$ be an $m \times p$ matrix of constants and let $\mathbf{b}$ be an $m$-dimensional vector of constants. Then $\mathbf{A}\mathbf{X}_n + \mathbf{b} \xrightarrow{D} N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.*

A result that will prove to be quite useful is the extension of the $\Delta$-method; see Theorem 4.3.9. A proof can be found in Chapter 3 of Serfling (1980).

**Theorem 4.5.6.** *Let $\{\mathbf{X}_n\}$ be a sequence of $p$-dimensional random vectors. Suppose*

$$\sqrt{n}(\mathbf{X}_n - \boldsymbol{\mu}_0) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}).$$

*Let $\mathbf{g}$ be a transformation $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_k(\mathbf{x}))'$ such that $1 \le k \le p$ and the $k \times p$ matrix of partial derivatives,*

$$\mathbf{B} = \left[\frac{\partial g_i}{\partial \mu_j}\right], \quad i = 1, \ldots k; \ j = 1, \ldots, p,$$

*are continuous and do not vanish in a neighborhood of $\boldsymbol{\mu}_0$. Let $\mathbf{B}_0 = \mathbf{B}$ at $\boldsymbol{\mu}_0$. Then*

$$\sqrt{n}(\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\boldsymbol{\mu}_0)) \xrightarrow{D} N_k(\mathbf{0}, \mathbf{B}_0\boldsymbol{\Sigma}\mathbf{B}_0'). \tag{4.5.13}$$

## EXERCISES

**4.5.1.** Let $\{\mathbf{X}_n\}$ be a sequence of $p$ dimensional random vectors. Show that

$$\mathbf{X}_n \xrightarrow{D} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ if and only if } \mathbf{a}'\mathbf{X}_n \xrightarrow{D} N_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}),$$

for all vectors $\mathbf{a} \in R^p$.

**4.5.2.** Let $X_1, \ldots, X_n$ be a random sample from a uniform$(a, b)$ distribution. Let $Y_1 = \min X_i$ and let $Y_2 = \max X_i$. Show that $(Y_1, Y_2)'$ converges in probability to the vector $(a, b)'$.

**4.5.3.** Let $\mathbf{X}_n$ and $\mathbf{Y}_n$ be $p$ dimensional random vectors. Show that if

$$\mathbf{X}_n - \mathbf{Y}_n \xrightarrow{P} \mathbf{0} \text{ and } \mathbf{X}_n \xrightarrow{D} \mathbf{X},$$

where $\mathbf{X}$ is a $p$ dimensional random vector, then $\mathbf{Y}_n \xrightarrow{D} \mathbf{X}$.

**4.5.4.** Let $\mathbf{X}_n$ and $\mathbf{Y}_n$ be $p$ dimensional random vectors such that $\mathbf{X}_n$ and $\mathbf{Y}_n$ are independent for each $n$ and their mgfs exist. Show that if

$$\mathbf{X}_n \xrightarrow{D} \mathbf{X} \text{ and } \mathbf{Y}_n \xrightarrow{D} \mathbf{Y},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are $p$ dimensional random vectors, then $(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{D} (\mathbf{X}, \mathbf{Y})$.

**4.5.5.** Suppose $\mathbf{X}_n$ has a $N_p(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ distribution. Show that

$$\mathbf{X}_n \xrightarrow{D} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ iff } \boldsymbol{\mu}_n \to \boldsymbol{\mu} \text{ and } \boldsymbol{\Sigma}_n \to \boldsymbol{\Sigma}.$$

**3.6.2** 1.761 .

**3.6.9** $\frac{1}{4.74}$ ; 3.33 .

## Chapter 4

**4.1.4** $\frac{8}{3}$ ; $\frac{2}{9}$ .

**4.1.5** 7 .

**4.1.7** 2.5 ; 0.25 .

**4.1.9** −5 ; 30.6 .

**4.1.10** $\frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$ .

**4.1.12** 12 ; 168 .

**4.1.13** 0.265 .

**4.1.15** 22.5 ; 65.25 .

**4.1.16** $\frac{\mu_2\sigma_1}{\sqrt{\sigma_1^2\sigma_2^2+\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}}$ .

**4.1.18** 0.801 .

**4.1.22** (a)$e^{\mu+(\sigma^2/2)}$ ; $e^{2\mu}\left(e^{2\sigma^2}-e^{\sigma^2}\right)$ .

**4.2.5** No ; $Y_n - \frac{1}{n}$ .

**4.3.1** Degenerate at $\mu$ .

**4.3.2** Gamma($\alpha = 1, \beta = 1$) .

**4.3.3** Gamma($\alpha = 1, \beta = 1$) .

**4.3.4** Gamma($\alpha = 2, \beta = 1$) .

**4.3.7** Degenerate at $\beta$ .

**4.3.9** 0.682 .

**4.3.10** (b) 0.815 .

**4.3.13** Degenerate at $\mu_2 + \frac{\sigma_2}{\sigma_1}(x - \mu_1)$ .

**4.3.14** (b) $N(0,1)$ .

**4.3.16** (b) $N(0,1)$ .

**4.3.19** $\frac{1}{5}$ .

**4.4.2** 0.954 .

**4.4.3** 0.604 .

**4.4.4** 0.840 .

**4.4.5** 0.728 .

**4.4.7** 0.08 .

**4.4.9** 0.267 .

## Chapter 5

**5.1.1** (a) $\frac{1}{m(m-1)}$ , $x_i = 1, 2, \ldots, m$ , $x_j = 1, 2, \ldots, m$ , $x_i \neq x_j$ .

**5.1.2** (a)$b(n,p)$ ; (c) $\frac{p(p-1)}{n}$ .

**5.1.3** (b)Gamma($\alpha = n, \beta = \theta/n$) ; (d) $c = 9.59$ , $d = 34.2$ .

**5.1.5** 9.5 .

**5.2.5** $1 - (1 - e^{-3})^4$ .

**5.2.6** (a) $\frac{1}{8}$ .

**5.2.10** Weibull.

**5.2.11** $\frac{5}{16}$ .

**5.2.12** pdf: $(2z_1)(4z_2^3)(6z_3^5)$ , $0 < z_i < 1$ .

**5.2.13** $\frac{7}{12}$ .

**5.2.17** (a) $48y_3^5y_4$ , $0 < y_3 < y_4 < 1$ ; (b) $\frac{6y_3^5}{y_4^6}$ , $0 < y_3 < y_4$ ; (c) $\frac{6}{7}y_4$ .

**5.2.18** $\frac{1}{4}$ .

**5.2.19** $6uv(u+v)$ , $0 < u < v < 1$ .

**5.2.24** 14 .

**5.2.25** (a) $\frac{15}{16}$ ; (b) $\frac{675}{1024}$ ; (c) $(0.8)^4$ .

**5.2.26** 0.824 .

**5.2.27** 8 .

**5.2.28** (a) $1.13\sigma$ ; (b) $0.92\sigma$ .

**5.3.2** 8 .

**5.3.5** (a) Beta($n - j + 1, j$) ; (b) Beta($n - j + i - 1, j - i + 2$) .

**5.3.6** $\frac{10!}{1!3!4!}v_1v_2^3(1 - v_1 - v_2)^4$ , $0 < v_2$ , $v_1 + v_2 < 1$ .

**5.4.1** (77.28, 85.12) .

**5.4.2** 24 or 25 .