CHAPTER 6

Limit Theorems

6.1 INTRODUCTION

In this chapter we investigate convergence properties of sequences of random variables. The three limit results proved here, namely, the two laws of large numbers and the central limit theorem, are of considerable importance in the study of probability and statistics. Just as in analysis, we distinguish among several types of convergence. The various modes of convergence are introduced in Section 6.2. Sections 6.3 and 6.4 deal with the laws of large numbers, and the central limit theorem is proved in Section 6.6.

The reader may find some parts of this chapter difficult, at least on first reading. These have been identified with a dagger (†) and include the concept of almost sure convergence (Section 6.2) and the strong law of large numbers (Section 6.4). Since the central limit result is basic and will be used repeatedly in the rest of the book, it is important for readers to familiarize themselves with this result and its application and to understand its significance. Similarly, on the first reading it will suffice to know the strong law of large numbers and to understand its significance.

6.2 MODES OF CONVERGENCE

In this section we consider several modes of convergence and investigate their interrelationships. We begin with the weakest mode.

Definition 1. Let $\{F_n\}$ be a sequence of distribution functions. If there exists a DF F such that as $n \to \infty$,

$$F_n(x) \to F(x) \tag{1}$$

at every point x at which F is continuous, we say that F_n *converges in law* (or, *weakly*), to F, and we write $F_n \xrightarrow{w} F$.

If $\{X_n\}$ is a sequence of RVs and $\{F_n\}$ is the corresponding sequence of DFs, we say that X_n *converges in distribution* (or *law*) to X if there exists an RV X with DF F such that $F_n \xrightarrow{w} F$. We write $X_n \xrightarrow{L} X$.

It must be remembered that it is quite possible for a given sequence of DFs to converge to a function that is not a DF.

Example 1. Consider the sequence of DFs

$$F_n(x) = \begin{cases} 0, & x < n, \\ 1, & x \geq n. \end{cases}$$

Here $F_n(x)$ is the DF of the RV X_n degenerate at $x = n$. We see that $F_n(x)$ converges to a function F that is identically equal to 0, and hence is not a DF.

Example 2. Let $X_1, X_2, \ldots, X_n$ be iid RVs with common density function,

$$f(x) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \theta, \quad (0 < \theta < \infty), \\ 0, & \text{otherwise}. \end{cases}$$

Let $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$. Then the density function of $X_{(n)}$ is

$$f_n(x) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n}, & 0 < x < \theta, \\ 0, & \text{otherwise}, \end{cases}$$

and the DF of $X_{(n)}$ is

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

We see that as $n \to \infty$,

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

which is a DF. Thus $F_n \xrightarrow{w} F$.

The following example shows that convergence in distribution does not imply convergence of moments.

Example 3. Let F_n be a sequence of DFs defined by

$$F_n(x) = \begin{cases} 0, & x < 0, \\ 1 - \dfrac{1}{n}, & 0 \leq x < n, \\ 1, & n \leq x. \end{cases}$$

Clearly, $F_n \xrightarrow{w} F$, where F is the DF given by

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

Note that F_n is the DF of the RV X_n with PMF

$$P\{X_n = 0\} = 1 - \frac{1}{n}, \qquad P\{X_n = n\} = \frac{1}{n},$$

and F is the DF of the RV X degenerate at 0. We have

$$EX_n^k = n^k \left(\frac{1}{n}\right) = n^{k-1},$$

where k is a positive integer. Also, $EX^k = 0$, so that

$$EX_n^k \nrightarrow EX^k \qquad \text{for any } k \geq 1.$$

We next give an example to show that weak convergence of distribution functions does not imply the convergence of corresponding PMFs or PDFs.

Example 4. Let $\{X_n\}$ be a sequence of RVs with PMF

$$f_n(x) = P\{X_n = x\} = \begin{cases} 1, & \text{if } x = 2 + \dfrac{1}{n}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that none of the f_n's assigns any probability to the point $x = 2$. It follows that

$$f_n(x) \to f(x) \qquad \text{as} \quad n \to \infty,$$

where $f(x) = 0$ for all x. However, the sequence of DFs $\{F_n\}$ of RVs X_n converges to the function

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

at all continuity points of F. Since F is the DF of the RV degenerate at $x = 2$, $F_n \xrightarrow{w} F$.

The following result is easy to prove.

Theorem 1. Let X_n be a sequence of integer-valued RVs. Also, let $f_n(k) = P\{X_n = k\}$, $k = 0, 1, 2, \ldots$, be the PMF of X_n, $n = 1, 2, \ldots$, and $f(k) = P\{X = k\}$ be the PMF of X. Then

$$f_n(x) \to f(x) \qquad \text{for all } x \Leftrightarrow X_n \xrightarrow{L} X.$$

In the continuous case we state the following result of Scheffé [98] without proof.

Theorem 2. Let X_n, $n = 1, 2, \ldots$, and X be continuous RVs such that

$$f_n(x) \to f(x) \qquad \text{for (almost) all } x \text{ as } n \to \infty.$$

Here, f_n and f are the PDFs of X_n and X, respectively. Then $X_n \xrightarrow{L} X$.

The following result is easy to establish.

Theorem 3. Let $\{X_n\}$ be a sequence of RVs such that $X_n \xrightarrow{L} X$, and let c be a constant. Then

(a) $X_n + c \xrightarrow{L} X + c$, and
(b) $cX_n \xrightarrow{L} cX$, $c \neq 0$.

A slightly stronger concept of convergence is defined by convergence in probability.

Definition 2. Let $\{X_n\}$ be a sequence of RVs defined on some probability space $(\Omega, \mathcal{S}, P)$. We say that the sequence $\{X_n\}$ *converges in probability* to the RV X if for every $\varepsilon > 0$,

$$P\{|X_n - X| > \varepsilon\} \to 0 \qquad \text{as } n \to \infty. \tag{2}$$

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

Remark 1. We emphasize that the definition says nothing about the convergence of the RVs X_n to the RV X in the sense in which it is understood in real analysis. Thus $X_n \xrightarrow{P} X$ does not imply that given $\varepsilon > 0$, we can find an N such that $|X_n - X| < \varepsilon$ for $n \geq N$. Definition 2 speaks only of the convergence of the sequence of probabilities $P\{|X_n - X| > \varepsilon\}$ to 0.

Example 5. Let $\{X_n\}$ be a sequence of RVs with PMF

$$P\{X_n = 1\} = \frac{1}{n}, \quad \text{and} \quad P\{X_n = 0\} = 1 - \frac{1}{n}.$$

Then

$$P\{|X_n| > \varepsilon\} = \begin{cases} P\{X_n = 1\} = \dfrac{1}{n} & \text{if } 0 < \varepsilon < 1, \\ 0 & \text{if } \varepsilon \geq 1. \end{cases}$$

It follows that $P\{|X_n| > \varepsilon\} \to 0$ as $n \to \infty$, and we conclude that $X_n \xrightarrow{P} 0$.

The truth of the following statements can easily be verified.

1. $X_n \xrightarrow{P} X \Leftrightarrow X_n - X \xrightarrow{P} 0$.
2. $X_n \xrightarrow{P} X, X_n \xrightarrow{P} Y \Rightarrow P\{X = Y\} = 1$ for $P\{|X-Y| > c\} \le P\{|X_n - X| > c/2\} + P\{|X_n - Y| > c/2\}$, and it follows that $P\{|X - Y| > c\} = 0$ for every $c > 0$.
3. $X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0$ as $n, m \to \infty$ for
$$P\{|X_n - X_m| > \varepsilon\} \le P\left\{|X_n - X| > \frac{\varepsilon}{2}\right\} + P\left\{|X_m - X| > \frac{\varepsilon}{2}\right\}.$$
4. $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$.
5. $X_n \xrightarrow{P} X$, k constant, $\Rightarrow kX_n \xrightarrow{P} kX$.
6. $X_n \xrightarrow{P} k \Rightarrow X_n^2 \xrightarrow{P} k^2$.
7. $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, a, b constants $\Rightarrow X_n Y_n \xrightarrow{P} ab$, for
$$X_n Y_n = \frac{(X_n + Y_n)^2 - (X_n - Y_n)^2}{4} \xrightarrow{P} \frac{(a+b)^2 - (a-b)^2}{4} = ab.$$
8. $X_n \xrightarrow{P} 1 \Rightarrow X_n^{-1} \xrightarrow{P} 1$ for
$$\begin{aligned} P\left\{\left|\frac{1}{X_n} - 1\right| \ge \varepsilon\right\} &= P\left\{\frac{1}{X_n} \ge 1 + \varepsilon\right\} + P\left\{\frac{1}{X_n} \le 1 - \varepsilon\right\} \\ &= P\left\{\frac{1}{X_n} \ge 1 + \varepsilon\right\} + P\left\{\frac{1}{X_n} \le 0\right\} \\ &\quad + P\left\{0 < \frac{1}{X_n} \le 1 - \varepsilon\right\}, \end{aligned}$$
and each of the three terms on the right goes to 0 as $n \to \infty$.
9. $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, a, b constants, $b \ne 0 \Rightarrow X_n Y_n^{-1} \xrightarrow{P} ab^{-1}$.
10. $X_n \xrightarrow{P} X$, and Y an RV $\Rightarrow X_n Y \xrightarrow{P} XY$. Note that Y is an RV, so that given $\delta > 0$, there exists a $k > 0$ such that $P\{|Y| > k\} < \delta/2$. Thus
$$\begin{aligned} P\{|X_n Y - XY| > \varepsilon\} &= P\{|X_n - X||Y| > \varepsilon, |Y| > k\} \\ &\quad + P\{|X_n - X||Y| > \varepsilon, |Y| \le k\} \\ &< \frac{\delta}{2} + P\left(|X_n - X| > \frac{\varepsilon}{k}\right). \end{aligned}$$
11. $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y \Rightarrow X_n Y_n \xrightarrow{P} XY$, for
$$(X_n - X)(Y_n - Y) \xrightarrow{P} 0.$$

The result now follows on multiplication, using result 10. It also follows that $X_n \xrightarrow{P} X \Rightarrow X_n^2 \xrightarrow{P} X^2$.

Theorem 4. Let $X_n \xrightarrow{P} X$, and g be a continuous function defined on $\mathcal{R}$. Then $g(X_n) \xrightarrow{P} g(X)$ as $n \to \infty$.

Proof. Since X is an RV, we can, given $\varepsilon > 0$, find a constant $k = k(\varepsilon)$ such that

$$P\{|X| > k\} < \frac{\varepsilon}{2}.$$

Also, g is continuous on $\mathcal{R}$, so that g is uniformly continuous on $[-k, k]$. It follows that there exists a $\delta = \delta(\varepsilon, k)$ such that

$$|g(x_n) - g(x)| < \varepsilon$$

whenever $|x| \le k$ and $|x_n - x| < \delta$. Let

$$A = \{|X| \le k\}, \qquad B = \{|X_n - X| < \delta\}, \qquad C = \{|g(X_n) - g(X)| < \varepsilon\}.$$

Then $\omega \in A \cap B \Rightarrow \omega \in C$, so that

$$A \cap B \subseteq C.$$

It follows that

$$P\{C^c\} \le P\{A^c\} + P\{B^c\},$$

that is,

$$P\{|g(X_n) - g(X)| \ge \varepsilon\} \le P\{|X_n - X| \ge \delta\} + P\{|X| > k\} < \varepsilon$$

for $n \ge N(\varepsilon, \delta, k)$, where $N(\varepsilon, \delta, k)$ is chosen so that

$$P\{|X_n - X| \ge \delta\} < \frac{\varepsilon}{2} \qquad \text{for } n \ge N(\varepsilon, \delta, k).$$

Corollary 1. $X_n \xrightarrow{P} c$, where c is a constant $\Rightarrow g(X_n) \xrightarrow{P} g(c)$, g being a continuous function.

We remark that a more general result than Theorem 4 is true and state it without proof (see Rao [86, p. 124]): $X_n \xrightarrow{L} X$, and g continuous on $\mathcal{R} \Rightarrow g(X_n) \xrightarrow{L} g(X)$.

The following two theorems explain the relationship between weak convergence and convergence in probability.

Theorem 5. $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$.

Proof. Let F_n and F, respectively, be the DFs of X_n and X. We have

$$\begin{aligned}\{\omega\colon X(\omega) \le x'\} &= \{\omega\colon X_n(\omega) \le x, X(\omega) \le x'\} \cup \{\omega\colon X_n(\omega) > x, X(\omega) \le x'\}\\ &\subseteq \{X_n \le x\} \cup \{X_n > x, X \le x'\}.\end{aligned}$$

It follows that

$$F(x') \le F_n(x) + P\{X_n > x, X \le x'\}.$$

Since $X_n - X \xrightarrow{P} 0$, we have for $x' < x$,

$$P\{X_n > x, X \le x'\} \le P\{|X_n - X| > x - x'\} \to 0 \qquad \text{as} \quad n \to \infty.$$

Therefore,

$$F(x') \le \varliminf_{n\to\infty} F_n(x), \qquad x' < x.$$

Similarly, by interchanging X and X_n, and x and x', we get

$$\varlimsup_{n\to\infty} F_n(x) \le F(x''), \qquad x < x''.$$

Thus, for $x' < x < x''$, we have

$$F(x') \le \varliminf F_n(x) \le \varlimsup F_n(x) \le F(x'').$$

Since F has only a countable number of discontinuity points, we choose x to be a point of continuity of F, and letting $x'' \downarrow x$ and $x' \uparrow x$, we have

$$F(x) = \lim_{n\to\infty} F_n(x)$$

at all points of continuity of F.

Theorem 6. Let k be a constant. Then

$$X_n \xrightarrow{L} k \Rightarrow X_n \xrightarrow{P} k.$$

The proof is left as an exercise.

Corollary. Let k be a constant. Then

$$X_n \xrightarrow{L} k \Leftrightarrow X_n \xrightarrow{P} k.$$

Remark 2. We emphasize that we cannot improve the result above by replacing k by an RV; that is, $X_n \xrightarrow{L} X$, in general, does not imply $X_n \xrightarrow{P} X$, for let $X, X_1, X_2, \ldots$ be identically distributed RVs, and let the joint distribution of (X_n, X) be as follows:

$X \backslash X_n$	0	1	
0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	0	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	1

Clearly, $X_n \xrightarrow{L} X$. But

$$\begin{aligned} P\left\{|X_n - X| > \tfrac{1}{2}\right\} &= P\{|X_n - X| = 1\} \\ &= P\{X_n = 0, X = 1\} + P\{X_n = 1, X = 0\} \\ &= 1 \nrightarrow 0. \end{aligned}$$

Hence, $X_n \stackrel{P}{\nrightarrow} X$, but $X_n \xrightarrow{L} X$.

Remark 3. Example 3 shows that $X_n \xrightarrow{P} X$ does not imply that $EX_n^k \to EX^k$ for any $k > 0$, k integral.

Definition 3. Let $\{X_n\}$ be a sequence of RVs such that $E|X_n|^r < \infty$ for some $r > 0$. We say that X_n *converges in the* rth *mean* to an RV X if $E|X|^r < \infty$ and

$$E|X_n - X|^r \to 0 \qquad \text{as} \quad n \to \infty, \tag{3}$$

and we write $X_n \xrightarrow{r} X$.

Example 6. Let $\{X_n\}$ be a sequence of RVs defined by

$$P\{X_n = 0\} = 1 - \frac{1}{n}, \qquad P\{X_n = 1\} = \frac{1}{n}, \qquad n = 1, 2, \ldots .$$

Then

$$E|X_n|^2 = \frac{1}{n} \to 0 \qquad \text{as} \quad n \to \infty,$$

and we see that $X_n \xrightarrow{2} X$, where RV X is degenerate at 0.

Theorem 7. Let $X_n \xrightarrow{r} X$ for some $r > 0$. Then $X_n \xrightarrow{P} X$.

The proof is left as an exercise.

Example 7. Let $\{X_n\}$ be a sequence of RVs defined by

$$P\{X_n = 0\} = 1 - \frac{1}{n^r}, \quad \text{and} \quad P\{X_n = n\} = \frac{1}{n^r}, \qquad r > 0, \quad n = 1, 2, \dots .$$

Then $E|X_n|^r = 1$, so that $X_n \stackrel{r}{\nrightarrow} 0$. We show that $X_n \xrightarrow{P} 0$.

$$P\{|X_n| > \varepsilon\} = \left\{\begin{matrix} P\{X_n = n\} & \text{if} \quad \varepsilon < n \\ 0 & \text{if} \quad \varepsilon > n \end{matrix}\right\} \to 0 \qquad \text{as } n \to \infty.$$

Theorem 8. Let $\{X_n\}$ be a sequence of RVs such that $X_n \xrightarrow{2} X$. Then $EX_n \to EX$ and $EX_n^2 \to EX^2$ as $n \to \infty$.

Proof. We have

$$|E(X_n - X)| \le E|X_n - X| \le E^{1/2}|X_n - X|^2 \to 0 \qquad \text{as } n \to \infty.$$

To see that $EX_n^2 \to EX^2$ (see also Theorem 9), we write

$$EX_n^2 = E(X_n - X)^2 + EX^2 + 2\,E\{X(X_n - X)\}$$

and note that

$$|E\{X(X_n - X)\}| \le \sqrt{EX^2 E(X_n - X)^2}$$

by the Cauchy–Schwarz inequality. The result follows on passing to the limits.

We get, in addition, that $X_n \xrightarrow{2} X$ implies that $\operatorname{var}(X_n) \to \operatorname{var}(X)$.

Corollary. Let $\{X_m\}$, $\{Y_n\}$ be two sequences of RVs such that $X_m \xrightarrow{2} X$, $Y_n \xrightarrow{2} Y$. Then $E(X_m Y_n) \to E(XY)$ as $m, n \to \infty$.

The proof is left to the reader.

As a simple consequence of Theorem 8 and its corollary we see that $X_m \xrightarrow{2} X$, $Y_n \xrightarrow{2} Y$ together imply that $\operatorname{cov}(X_m, Y_n) \to \operatorname{cov}(X, Y)$.

Theorem 9. If $X_n \xrightarrow{r} X$, then $E|X_n|^r \to E|X|^r$.

Proof. Let $0 < r \le 1$. Then

$$E|X_n|^r = E|X_n - X + X|^r$$

so that

$$E|X_n|^r - E|X|^r \le E|X_n - X|^r.$$

Interchanging X_n and X, we get

$$E|X|^r - E|X_n|^r \le E|X_n - X|^r.$$

It follows that

$$|E|X|^r - E|X_n|^r| \le E|X_n - X|^r \to 0 \qquad \text{as} \quad n \to \infty.$$

For $r > 1$, we use Minkowski's inequality and obtain

$$[E|X_n|^r]^{1/r} \le [E|X_n - X|^r]^{1/r} + [E|X|^r]^{1/r}$$

and

$$[E|X|^r]^{1/r} \le [E|X_n - X|^r]^{1/r} + [E|X_n|^r]^{1/r}.$$

It follows that

$$|E^{1/r}|X_n|^r - E^{1/r}|X|^r| \le E^{1/r}|X_n - X|^r \to 0 \qquad \text{as} \quad n \to \infty.$$

This completes the proof.

Theorem 10. Let $r > s$. Then $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{s} X$.

Proof. From Theorem 3.4.3 it follows that for $s < r$,

$$E|X_n - X|^s \le [E|X_n - X|^r]^{s/r} \to 0 \qquad \text{as } n \to \infty$$

since $X_n \xrightarrow{r} X$.

Remark 4. Clearly, the converse to Theorem 10 cannot hold, since $E|X|^s < \infty$ for $s < r$ does not imply that $E|X|^r < \infty$.

Remark 5. In view of Theorem 9, it follows that $X_n \xrightarrow{r} X \Rightarrow E|X_n|^s \to E|X|^s$ for $s \le r$.

Definition 4.[†] Let $\{X_n\}$ be a sequence of RVs. We say that X_n *converges almost surely* (a.s.) to an RV X if and only if

(4) $$P\{\omega\colon X_n(\omega) \to X(\omega) \text{ as } n \to \infty\} = 1,$$

and we write $X_n \xrightarrow{\text{a.s.}} X$ or $X_n \to X$ with probability 1.

[†]May be omitted on the first reading.

The following result elucidates Definition 4.

Theorem 11. $X_n \xrightarrow{\text{a.s.}} X$ if and only if $\lim_{n\to\infty} P\{\sup_{m\geq n} |X_m - X| > \varepsilon\} = 0$ for all $\varepsilon > 0$.

Proof. Since $X_n \xrightarrow{\text{a.s.}} X$, $X_n - X \xrightarrow{\text{a.s.}} 0$, and it will be sufficient to show the equivalence of

(a) $X_n \xrightarrow{\text{a.s.}} 0$ and
(b) $\lim_{n\to\infty} P\{\sup_{m\geq n} |X_m| > \varepsilon\} = 0$.

Let us suppose that (a) holds. Let $\varepsilon > 0$, and write

$$A_n(\varepsilon) = \left\{\sup_{m\geq n} |X_m| > \varepsilon\right\} \quad \text{and} \quad C = \left\{\lim_{n\to\infty} X_n = 0\right\}.$$

Also write $B_n(\varepsilon) = C \cap A_n(\varepsilon)$, and note that $B_{n+1}(\varepsilon) \subset B_n(\varepsilon)$, and the limit set $\bigcap_{n=1}^{\infty} B_n(\varepsilon) = \emptyset$. It follows that

$$\lim_{n\to\infty} PB_n(\varepsilon) = P\left\{\bigcap_{n=1}^{\infty} B_n(\varepsilon)\right\} = 0.$$

Since $PC = 1$, $PC^c = 0$, and we have

$$\begin{aligned} PB_n(\varepsilon) = P(A_n \cap C) &= 1 - P(C^c \cup A_n^c) \\ &= 1 - PC^c - PA_n^c + P(C^c \cap A_n^c) \\ &= PA_n + P(C^c \cap A_n^c) \\ &= PA_n. \end{aligned}$$

It follows that (b) holds.

Conversely, let $\lim_{n\to\infty} PA_n(\varepsilon) = 0$, and write

$$D(\varepsilon) = \left\{\overline{\lim_{n\to\infty}} |X_n| > \varepsilon > 0\right\}.$$

Since $D(\varepsilon) \subset A_n(\varepsilon)$ for $n = 1, 2, \ldots$, it follows that $PD(\varepsilon) = 0$. Also,

$$C^c = \left\{\lim_{n\to\infty} X_n \neq 0\right\} \subset \bigcup_{k=1}^{\infty} \left\{\overline{\lim} |X_n| > \frac{1}{k}\right\},$$

so that

$$1 - PC \leq \sum_{k=1}^{\infty} PD\left(\frac{1}{k}\right) = 0,$$

and (a) holds.

Remark 6. Thus $X_n \xrightarrow{\text{a.s.}} 0$ means that for $\varepsilon > 0$, $\eta > 0$ arbitrary, we can find an n_0 such that

$$P\left\{\sup_{n \geq n_0} |X_n| > \varepsilon\right\} < \eta. \tag{5}$$

Indeed, we can write, equivalently, that

$$\lim_{n_0 \to \infty} P\left[\bigcup_{n \geq n_0} \{|X_n| > \varepsilon\}\right] = 0. \tag{6}$$

Theorem 12. $X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{P} X$.

Proof. By Remark 6, $X_n \xrightarrow{\text{a.s.}} X$ implies that for arbitrary $\varepsilon > 0$, $\eta > 0$, we can choose an $n_0 = n_0(\varepsilon, \eta)$ such that

$$P\left[\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \varepsilon\}\right] \geq 1 - \eta.$$

Clearly,

$$\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \varepsilon\} \subset \{|X_n - X| \leq \varepsilon\} \qquad \text{for} \quad n \geq n_0.$$

It follows that for $n \geq n_0$,

$$P\{|X_n - X| \leq \varepsilon\} \geq P\left[\bigcap_{n=n_0}^{\infty} \{|X_n - X| \leq \varepsilon\}\right] \geq 1 - \eta,$$

that is,

$$P\{|X_n - X| > \varepsilon\} < \eta \qquad \text{for} \quad n \geq n_0,$$

which is the same as saying that $X_n \xrightarrow{P} X$.

That the converse of Theorem 12 does not hold is shown in the following example.

Example 8. For each positive integer n there exist integers m and k (uniquely determined) such that

$$n = 2^k + m, \qquad 0 \leq m < 2^k, \quad k = 0, 1, 2, \ldots .$$

Thus, for $n = 1$, $k = 0$ and $m = 0$; for $n = 5$, $k = 2$ and $m = 1$; and so on. Define RVs X_n for $n = 1, 2, \ldots$ on $\Omega = [0, 1]$ by

$$X_n(\omega) = \begin{cases} 2^k, & \dfrac{m}{2^k} \le \omega < \dfrac{m+1}{2^k}, \\ 0, & \text{otherwise.} \end{cases}$$

Let the probability distribution of X_n be given by $P\{I\}$ = length of the interval $I \subseteq \Omega$. Thus

$$P\{X_n = 2^k\} = \frac{1}{2^k} \quad \text{and} \quad P\{X_n = 0\} = 1 - \frac{1}{2^k}.$$

The limit $\lim_{n\to\infty} X_n(\omega)$ does not exist for any $\omega \in \Omega$, so that X_n does not converge almost surely. But

$$P\{X_n| > \varepsilon\} = P\{X_n > \varepsilon\} = \begin{cases} 0 & \text{if} \quad \varepsilon \ge 2^k, \\ \dfrac{1}{2^k} & \text{if} \quad 0 < \varepsilon < 2^k, \end{cases}$$

and we see that

$$P\{|X_n| > \varepsilon\} \to 0 \qquad \text{as } n \text{ (and hence } k) \to \infty.$$

Theorem 13. Let $\{X_n\}$ be a strictly decreasing sequence of positive RVs, and suppose that $X_n \xrightarrow{P} 0$. Then $X_n \xrightarrow{\text{a.s.}} 0$.

The proof is left as an exercise.

Example 9. Let $\{X_n\}$ be a sequence of independent RVs defined by

$$P\{X_n = 0\} = 1 - \frac{1}{n}, \quad \text{and} \quad P\{X_n = 1\} = \frac{1}{n}, \qquad n = 1, 2, \dots.$$

Then

$$E|X_n - 0|^2 = E|X_n|^2 = \frac{1}{n} \to 0 \qquad \text{as } n \to \infty,$$

so that $X_n \xrightarrow{2} 0$. Also,

$$\begin{aligned} &P\{X_n = 0 \quad \text{for every } m \le n \le n_0\} \\ &= \prod_{n=m}^{n_0} \left(1 - \frac{1}{n}\right) = \frac{m-1}{n_0}, \end{aligned}$$

which diverges to zero as $n_0 \to \infty$ for all values of m. Thus X_n does not converge to 0 with probability 1.

Example 10. Let $\{X_n\}$ be independent, defined by

$$P\{X_n = 0\} = 1 - \frac{1}{n^r} \quad \text{and} \quad P\{X_n = n\} = \frac{1}{n^r}, \qquad r \geq 2, \quad n = 1, 2, \ldots.$$

Then

$$P\{X_n = 0 \quad \text{for } m \leq n \leq n_0\} = \prod_{n=m}^{n_0} \left(1 - \frac{1}{n^r}\right).$$

As $n_0 \to \infty$, the infinite product converges to some nonzero quantity, which itself converges to 1 as $m \to \infty$. Thus $X_n \xrightarrow{\text{a.s.}} 0$. However, $E|X_n|^r = 1$, and $X_n \overset{r}{\nrightarrow} 0$ as $n \to \infty$.

Example 11. Let $\{X_n\}$ be a sequence of RVs with $P\{X_n = \pm 1/n\} = \frac{1}{2}$. Then $E|X_n|^r = 1/n^r \to 0$ as $n \to \infty$, and $X_n \xrightarrow{r} 0$. For $j < k$, $|X_j| > |X_k|$, so that $\{|X_k| > \varepsilon\} \subset \{|X_j| > \varepsilon\}$. It follows that

$$\bigcup_{j=n}^{\infty} \{|X_j| > \varepsilon\} = \{|X_n| > \varepsilon\}.$$

Choosing $n > 1/\varepsilon$, we see that

$$P\left[\bigcup_{j=n}^{\infty} \{|X_j| > \varepsilon\}\right] = P\{|X_n| > \varepsilon\} \leq P\left\{|X_n| > \frac{1}{n}\right\} = 0,$$

and (6) implies that $X_n \xrightarrow{\text{a.s.}} 0$.

Remark 7. In Theorem 6.4.3 we prove a result that is sometimes useful in proving a.s. convergence of a sequence of RVs.

Theorem 14. Let $\{X_n, Y_n\}$, $n = 1, 2, \ldots$, be a sequence of RVs. Then

$$|X_n - Y_n| \xrightarrow{P} 0 \quad \text{and} \quad Y_n \xrightarrow{L} Y \Rightarrow X_n \xrightarrow{L} Y.$$

Proof. Let x be a point of continuity of the DF of Y and $\varepsilon > 0$. Then

$$\begin{aligned} P\{X_n \leq x\} &= P\{Y_n \leq x + Y_n - X_n\} \\ &= P\{Y_n \leq x + Y_n - X_n; Y_n - X_n \leq \varepsilon\} \\ &\quad + P\{Y_n \leq x + Y_n - X_n; Y_n - X_n > \varepsilon\} \\ &\leq P\{Y_n \leq x + \varepsilon\} + P\{Y_n - X_n > \varepsilon\}. \end{aligned}$$

It follows that

$$\overline{\lim_{n\to\infty}} P\{X_n \le x\} \le \underline{\lim}_{n\to\infty} P\{Y_n \le x + \varepsilon\}.$$

Similarly,

$$\underline{\lim}_{n\to\infty} P\{X_n \le x\} \ge \overline{\lim_{n\to\infty}} P\{Y_n \le x - \varepsilon\}.$$

Since $\varepsilon > 0$ is arbitrary and x is a continuity point of $P\{Y \le x\}$, we get the result by letting $\varepsilon \to 0$.

Corollary. $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$.

Theorem 15 (Slutsky's Theorem). Let $\{X_n, Y_n\}$, $n = 1, 2, \ldots$, be a sequence of pairs of RVs, and let c be a constant. Then

(a) $X_n \xrightarrow{L} X, Y_n \xrightarrow{P} c \Rightarrow X_n + Y_n \xrightarrow{L} X + c$;

(b) $X_n \xrightarrow{L} X$,

$$Y_n \xrightarrow{P} c \Rightarrow \begin{cases} X_nY_n \xrightarrow{L} cX & \text{if } c \ne 0, \\ X_nY_n \xrightarrow{P} 0 & \text{if } c = 0; \end{cases}$$

(c) $X_n \xrightarrow{L} X, Y_n \xrightarrow{P} c \Rightarrow X_n/Y_n \xrightarrow{L} X/c$ if $c \ne 0$.

Proof. (a) $X_n \xrightarrow{L} X \Rightarrow X_n + c \xrightarrow{L} X + c$ (Theorem 3). Also, $Y_n - c = (Y_n + X_n) - (X_n + c) \xrightarrow{P} 0$. A simple use of Theorem 14 shows that

$$X_n + Y_n \xrightarrow{L} X + c.$$

(b) We first consider the case where $c = 0$. We have for any fixed number $k > 0$,

$$\begin{aligned} P\{|X_nY_n| > \varepsilon\} &= P\left\{|X_nY_n| > \varepsilon, |Y_n| \le \frac{\varepsilon}{k}\right\} + P\left\{|X_nY_n| > \varepsilon, |Y_n| > \frac{\varepsilon}{k}\right\} \\ &\le P\{|X_n| > k\} + P\left\{|Y_n| > \frac{\varepsilon}{k}\right\}. \end{aligned}$$

Since $Y_n \xrightarrow{P} 0$ and $X_n \xrightarrow{L} X$, it follows that for any fixed $k > 0$,

$$\overline{\lim_{n\to\infty}} P\{|X_nY_n| > \varepsilon\} \le P\{|X| > k\}.$$

Since k is arbitrary, we can make $P\{|X| > k\}$ as small as we please by choosing k large. It follows that

$$X_nY_n \xrightarrow{P} 0.$$

Now, let $c \neq 0$. Then

$$X_nY_n - cX_n = X_n(Y_n - c),$$

and since $X_n \xrightarrow{L} X$, $Y_n \xrightarrow{P} c$, $X_n(Y_n - c) \xrightarrow{P} 0$. Using Theorem 14, we get the result that

$$X_nY_n \xrightarrow{L} cX.$$

(c) $Y_n \xrightarrow{P} c$, and $c \neq 0 \Rightarrow Y_n^{-1} \xrightarrow{P} c^{-1}$. It follows that $X_n \xrightarrow{L} X$, $Y_n \xrightarrow{P} c \Rightarrow X_nY_n^{-1} \xrightarrow{L} c^{-1}X$, and the proof of the theorem is complete.

As an application of Theorem 15, we present the following example. Many more examples appear in Chapter 7.

Example 12. Let $X_1, X_2, \dots,$ be iid RVs with common law $\mathcal{N}(0, 1)$. We shall determine the limiting distribution of the RV

$$W_n = \sqrt{n}\frac{X_1 + X_2 + \cdots + X_n}{X_1^2 + X_2^2 + \cdots + X_n^2}.$$

Let us write

$$U_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n) \quad \text{and} \quad V_n = \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}.$$

Then

$$W_n = \frac{U_n}{V_n}.$$

For the MGF of U_n we have

$$\begin{aligned} M_{U_n}(t) &= \prod_{i=1}^{n} Ee^{tX_i/\sqrt{n}} = \prod_{i=1}^{n} e^{t^2/2n} \\ &= e^{t^2/2}, \end{aligned}$$

so that U_n is an $\mathcal{N}(0, 1)$ variate (see also Corollary 2 to Theorem 5.3.22). It follows that $U_n \xrightarrow{L} Z$, where Z is an $\mathcal{N}(0, 1)$ RV. As for V_n, we note that each X_i^2 is a chi-square variate with 1 d.f. Thus

$$\begin{aligned} M_{V_n}(t) &= \prod_{i=1}^{n}\left(\frac{1}{1 - 2t/n}\right)^{1/2}, \quad t < \frac{n}{2}, \\ &= \left(1 - \frac{2t}{n}\right)^{-n/2}, \quad t < \frac{n}{2}, \end{aligned}$$

which is the MGF of a gamma variate with parameters $\alpha = n/2$ and $\beta = 2/n$. Thus the density function of V_n is given by

$$f_{V_n}(x) = \begin{cases} \dfrac{1}{\Gamma(n/2)} \dfrac{1}{(2/n)^{n/2}} x^{n/2-1} e^{-nx/2}, & 0 < x < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

We will show that $V_n \xrightarrow{P} 1$. We have for any $\varepsilon > 0$,

$$P\{|V_n - 1| > \varepsilon\} \le \frac{\operatorname{var}(V_n)}{\varepsilon^2} = \left(\frac{n}{2}\right)\left(\frac{2}{n}\right)^2 \frac{1}{\varepsilon^2} \to 0 \qquad \text{as} \quad n \to \infty.$$

We have thus shown that

$$U_n \xrightarrow{L} Z \quad \text{and} \quad V_n \xrightarrow{P} 1.$$

It follows by Theorem 15(c) that $W_n = U_n / V_n \xrightarrow{L} Z$, where Z is an $\mathcal{N}(0, 1)$ RV.

Later we will see that the condition that the X_i's be $\mathcal{N}(0, 1)$ is not needed. All we need is that $E|X_i|^2 < \infty$.

PROBLEMS 6.2

1. Let $X_1, X_2, \ldots$ be a sequence of RVs with corresponding DFs given by $F_n(x) = 0$ if $x < -n$, $= (x+n)/2n$ if $-n \le x < n$, and $= 1$ if $x \ge n$. Does F_n converge to a DF?

2. Let $X_1, X_2 \ldots$ be iid $\mathcal{N}(0, 1)$ RVs. Consider the sequence of RVs $\{\overline{X}_n\}$, where $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$. Let F_n be the DF of $\overline{X}_n$, $n = 1, 2, \ldots$. Find $\lim_{n\to\infty} F_n(x)$. Is this limit a DF?

3. Let $X_1, X_2, \ldots$ be iid $U(0, \theta)$ RVs. Let $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$, and consider the sequence $Y_n = nX_{(1)}$. Does Y_n converge in distribution to some RV Y? If so, find the DF of RV Y.

4. Let $X_1, X_2, \ldots$ be iid RVs with common absolutely continuous DF F. Let $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$, and consider the sequence of RVs $Y_n = n[1 - F(X_{(n)})]$. Find the limiting DF of Y_n.

5. Let $X_1, X_2, \ldots$ be a sequence of iid RVs with common PDF $f(x) = e^{-x+\theta}$ if $x \ge \theta$, and $= 0$ if $x < \theta$. Write $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$.
   (a) Show that $\overline{X}_n \xrightarrow{P} 1 + \theta$.
   (b) Show that $\min\{X_1, X_2, \ldots, X_n\} \xrightarrow{P} \theta$.

6. Let $X_1, X_2, \ldots$ be iid $U[0, \theta]$ RVs. Show that $\max\{X_1, X_2, \ldots, X_n\} \xrightarrow{P} \theta$.

7. Let $\{X_n\}$ be a sequence of RVs such that $X_n \xrightarrow{L} X$. Let a_n be a sequence of positive constants such that $a_n \to \infty$ as $n \to \infty$. Show that $a_n^{-1} X_n \xrightarrow{P} 0$.

8. Let $\{X_n\}$ be a sequence of RVs such that $P\{|X_n| \leq k\} = 1$ for all n and some constant $k > 0$. Suppose that $X_n \xrightarrow{P} X$. Show that $X_n \xrightarrow{r} X$ for any $r > 0$.

9. Let $X_1, X_2, \ldots, X_{2n}$ be iid $\mathcal{N}(0, 1)$ RVs. Define

$$U_n = \left(\frac{X_1}{X_2} + \frac{X_3}{X_4} + \cdots + \frac{X_{2n-1}}{X_{2n}}\right),$$

$$V_n = X_1^2 + X_2^2 + \cdots + X_n^2, \quad \text{and} \quad Z_n = \frac{U_n}{V_n}.$$

Find the limiting distribution of Z_n.

10. Let $\{X_n\}$ be a sequence of geometric RVs with parameter $\lambda/n, n > \lambda > 0$. Also, let $Z_n = X_n/n$. Show that $Z_n \xrightarrow{L} G(1, 1/\lambda)$ as $n \to \infty$. (Prochaska [80])

11. Let X_n be a sequence of RVs such that $X_n \xrightarrow{\text{a.s.}} 0$, and let c_n be a sequence of real numbers such that $c_n \to 0$ as $n \to \infty$. Show that $X_n + c_n \xrightarrow{\text{a.s.}} 0$.

12. Does convergence almost surely imply convergence of moments?

13. Let $X_1, X_2, \ldots,$ be a sequence of iid RVs with common DF F, and write $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}, n = 1, 2, \ldots$.

    (a) For $\alpha > 0$, $\lim_{x\to\infty} x^\alpha P\{X_1 > x\} = b > 0$. Find the limiting distribution of $(bn)^{-1/\alpha} X_{(n)}$. Also, find the PDF corresponding to the limiting DF and compute its moments.

    (b) If F satisfies

    $$\lim_{x\to\infty} e^x [1 - F(x)] = b > 0,$$

    find the limiting DF of $X_{(n)} - \log(bn)$ and compute the corresponding PDF and the MGF.

    (c) If X_i is bounded above by x_0 with probability 1, and for some $\alpha > 0$

    $$\lim_{x\to x_0-} (x_0 - x)^{-\alpha}[1 - F(x)] = b > 0,$$

    find the limiting distribution of $(bn)^{1/\alpha}\{X_{(n)} - x_0\}$, the corresponding PDF, and the moments of the limiting distribution.

    (*The remarkable result above, due to Gnedenko* [33], *exhausts all limiting distributions of $X_{(n)}$ with suitable norming and centering.*)

14. Let $\{F_n\}$ be a sequence of DFs that converges weakly to a DF F that is continuous everywhere. Show that $F_n(x)$ converges to $F(x)$ uniformly.

15. Prove Theorem 1.

16. Prove Theorem 6.

17. Prove Theorem 13.

18. Prove Corollary 1 to Theorem 8.

19. Let V be the class of all random variables defined on a probability space with finite expectations, and for $X \in V$ define

$$\rho(X) = E\left\{\frac{|X|}{1+|X|}\right\}.$$

Show the following:

(a) $\rho(X+Y) \le \rho(X) + \rho(Y)$; $\rho(\sigma X) \le \max(|\sigma|, 1)\rho(X)$.

(b) $d(X, Y) = \rho(X - Y)$ is a distance function on V (assuming that we identify RVs that are a.s. equal).

(c) $\lim_{n\to\infty} d(X_n, X) = 0 \Leftrightarrow X_n \xrightarrow{P} X$.

20. For the following sequences of RVs $\{X_n\}$, investigate convergence in probability and convergence in rth mean.

(a) $X_n \sim C(1/n, 0)$.

(b) $P(X_n = e^n) = 1/n^2, \quad P(X_n = 0) = 1 - 1/n^2$.

6.3 WEAK LAW OF LARGE NUMBERS

Let $\{X_n\}$ be a sequence of RVs. Write $S_n = \sum_{k=1}^n X_k$, $n = 1, 2, \ldots$. In this section we answer the following question in the affirmative: Do there exist sequences of constants A_n and $B_n > 0$, $B_n \to \infty$ as $n \to \infty$, such that the sequence of RVs $B_n^{-1}(S_n - A_n)$ converges in probability to 0 as $n \to \infty$?

Definition 1. Let $\{X_n\}$ be a sequence of RVs, and let $S_n = \sum_{k=1}^n X_k$, $n = 1, 2, \ldots$. We say that $\{X_n\}$ obeys the *weak law of large numbers* (WLLN) with respect to the sequence of constants $\{B_n\}$, $B_n > 0$, $B_n \uparrow \infty$, if there exists a sequence of real constants A_n such that $B_n^{-1}(S_n - A_n) \xrightarrow{P} 0$ as $n \to \infty$. A_n are called *centering constants*, and B_n, *norming constants*.

Theorem 1. Let $\{X_n\}$ be a sequence of pairwise uncorrelated RVs with $EX_i = \mu_i$ and $\operatorname{var}(X_i) = \sigma_i^2$, $i = 1, 2, \ldots$. If $\sum_{i=1}^n \sigma_i^2 \to \infty$ as $n \to \infty$, we can choose $A_n = \sum_{k=1}^n \mu_k$ and $B_n = \sum_{i=1}^n \sigma_i^2$, that is,

$$\sum_{i=1}^n \frac{X_i - \mu_i}{\sum_{i=1}^n \sigma_i^2} \xrightarrow{P} 0 \qquad \text{as } n \to \infty.$$

Proof. We have, by Chebychev's inequality,

$$P\left\{\left|S_n - \sum_{k=1}^{n} \mu_k\right| > \varepsilon \sum_{i=1}^{n} \sigma_i^2\right\} \le \frac{E\left[\sum_{i=1}^{n}(X_i - \mu_i)\right]^2}{\varepsilon^2 \left(\sum_{i=1}^{n} \sigma_i^2\right)^2}$$
$$= \frac{1}{\varepsilon^2 \sum_{i=1}^{n} \sigma_i^2} \to 0 \qquad \text{as } n \to \infty.$$

Corollary 1. If the X_n's are identically distributed and pairwise uncorrelated with $EX_i = \mu$ and $\operatorname{var}(X_i) = \sigma^2 < \infty$, we can choose $A_n = n\mu$ and $B_n = n\sigma^2$.

Corollary 2. In Theorem 1 we can choose $B_n = n$, provided that $n^{-2} \sum_{i=1}^{n} \sigma_i^2 \to 0$ as $n \to \infty$.

Corollary 3. In Corollary 1 we can take $A_n = n\mu$ and $B_n = n$, since $n\sigma^2/n^2 \to 0$ as $n \to \infty$. Thus, if $\{X_n\}$ are pairwise-uncorrelated identically distributed RVs with finite variance, $S_n/n \xrightarrow{P} \mu$.

Example 1. Let $X_1, X_2, \ldots$ be iid RVs with common law $b(1, p)$. Then $EX_i = p$, $\operatorname{var}(X_i) = p(1-p)$, and we have

$$\frac{S_n}{n} \xrightarrow{P} p \qquad \text{as } n \to \infty.$$

Note that S_n/n is the proportion of successes in n trials.

Hereafter, we shall be interested mainly in the case where $B_n = n$. When we say that $\{X_n\}$ obeys the WLLN, this is so with respect to the sequence $\{n\}$.

Theorem 2. Let $\{X_n\}$ be any sequence of RVs. Write $Y_n = n^{-1} \sum_{k=1}^{n} X_k$. A necessary and sufficient condition for the sequence $\{X_n\}$ to satisfy the weak law of large numbers is that

$$E\left\{\frac{Y_n^2}{1+Y_n^2}\right\} \to 0 \qquad \text{as } n \to \infty. \tag{1}$$

Proof. For any two positive numbers $a, b, a \ge b > 0$, we have

$$\frac{a}{1+a} \frac{1+b}{b} \ge 1. \tag{2}$$

Let $A = \{|Y_n| \ge \varepsilon\}$. Then $\omega \in A \Rightarrow |Y_n|^2 \ge \varepsilon^2 > 0$. Using (2), we see that $\omega \in A$ implies that

$$\frac{Y_n^2}{1+Y_n^2} \frac{1+\varepsilon^2}{\varepsilon^2} \ge 1.$$

It follows that

$$\begin{aligned} PA &\le P\left\{\frac{Y_n^2}{1+Y_n^2} \ge \frac{\varepsilon^2}{1+\varepsilon^2}\right\} \\ &\le E\frac{|Y_n^2/(1+Y_n^2)|}{\varepsilon^2/(1+\varepsilon^2)} \qquad \text{by Markov's inequality} \\ &\to 0 \qquad \text{as } n \to \infty. \end{aligned}$$

That is,

$$Y_n \xrightarrow{P} 0 \qquad \text{as } n \to \infty.$$

Conversely, we will show that for every $\varepsilon > 0$,

$$P\{|Y_n| \ge \varepsilon\} \ge E\left\{\frac{Y_n^2}{1+Y_n^2}\right\} - \varepsilon^2. \tag{3}$$

We will prove (3) for the case in which Y_n is of the continuous type. The discrete case being similar, we ask the reader to complete the proof. If Y_n has PDF $f_n(y)$, then

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{y^2}{1+y^2} f_n(y)\,dy &= \left(\int_{|y|>\varepsilon} + \int_{|y|\le\varepsilon}\right) \frac{y^2}{1+y^2} f_n(y)\,dy \\ &\le P\{|Y_n| > \varepsilon\} + \int_{-\varepsilon}^{\varepsilon} \left(1 - \frac{1}{1+y^2}\right) f_n(y)\,dy \\ &\le P\{|Y_n| > \varepsilon\} + \frac{\varepsilon^2}{1+\varepsilon^2} \le P\{|Y_n| > \varepsilon\} + \varepsilon^2, \end{aligned}$$

which is (3).

Remark 1. Since condition (1) applies not to the individual variables but to their sum, Theorem 2 is of limited use. We note, however, that all weak laws of large numbers obtained as corollaries to Theorem 1 follow easily from Theorem 2 (Problem 6).

Example 2. Let $(X_1, X_2, \ldots, X_n)$ be jointly normal with $EX_i = 0$, $EX_i^2 = 1$ for all i, and $\text{cov}(X_i, X_j) = \rho$ if $|j-i| = 1$, and $= 0$ otherwise. Then $S_n = \sum_{k=1}^{n} X_k$ is $\mathcal{N}(0, \sigma^2)$, where

$$\sigma^2 = \text{var}(S_n) = n + 2(n-1)\rho,$$

$$
\begin{aligned}
E\left\{\frac{Y_n^2}{1+Y_n^2}\right\} &= E\left\{\frac{S_n^2}{n^2+S_n^2}\right\} \\
&= \frac{2}{\sigma\sqrt{2\pi}}\int_0^\infty \frac{x^2}{n^2+x^2}e^{-x^2/2\sigma^2}\,dx \\
&= \frac{2}{\sqrt{2\pi}}\int_0^\infty \frac{y^2[n+2(n-1)\rho]}{n^2+y^2[n+2(n-1)\rho]}e^{-y^2/2}\,dy \\
&\le \frac{n+2(n-1)\rho}{n^2}\int_0^\infty \frac{2}{\sqrt{2\pi}}y^2e^{-y^2/2}\,dy \to 0 \qquad \text{as } n\to\infty.
\end{aligned}
$$

It follows from Theorem 2 that $n^{-1}S_n \xrightarrow{P} 0$. We invite the reader to compare this result to that of Problem 6.5.6.

Example 3. Let $X_1, X_2, \ldots$ be iid $C(1,0)$ RVs. We have seen (corollary to Theorem 5.3.18) that $n^{-1}S_n \sim C(1,0)$, so that $n^{-1}S_n$ does not converge in probability to 0. It follows that the WLLN does not hold (see also Problem 10).

Let $X_1, X_2, \ldots$ be an arbitrary sequence of RVs, and let $S_n = \sum_{k=1}^n X_k$, $n = 1, 2, \ldots$. Let us truncate each X_i at $c > 0$, that is, let

$$
X_i^c = \begin{cases} X_i & \text{if } |X_i| \le c \\ 0 & \text{if } |X_i| > c \end{cases}, \qquad i = 1, 2, \ldots, n.
$$

Write

$$
S_n^c = \sum_{i=1}^n X_i^c, \quad \text{and} \quad m_n = \sum_{i=1}^n EX_i^c.
$$

Lemma 1. For any $\varepsilon > 0$,

$$
P\{|S_n - m_n| > \varepsilon\} \le P\{|S_n^c - m_n| > \varepsilon\} + \sum_{k=1}^n P\{|X_k| > c\}. \tag{4}
$$

Proof. We have

$$
\begin{aligned}
P\{|S_n - m_n| > \varepsilon\} &= P\{|S_n - m_n| > \varepsilon \text{ and } |X_k| \le c \quad \text{for } k = 1,2,\ldots,n\} \\
&\quad + P\{|S_n - m_n| > \varepsilon \text{ and } |X_k| > c \quad \text{for at least one } k,\ k = 1,2,\ldots,n\} \\
&\le P\{|S_n^c - m_n| > \varepsilon\} + P\{|X_k| > c \quad \text{for at least one } k,\ 1 \le k \le n\} \\
&\le P\{|S_n^c - m_n| > \varepsilon\} + \sum_{k=1}^n P\{|X_k| > c\}.
\end{aligned}
$$

Corollary. If $X_1, X_2, \dots, X_n$ are exchangeable, then

$$P\{|S_n - m_n| > \varepsilon\} \le P\{|S_n^c - m_n| > \varepsilon\} + nP\{|X_1| > c\}. \tag{5}$$

If, in addition, the RVs $X_1, X_2, \dots, X_n$ are independent, then

$$P\{|S_n - m_n| > \varepsilon\} \le \frac{nE(X_1^c)^2}{\varepsilon^2} + nP\{|X_1| > c\}. \tag{6}$$

Inequality (6) yields the following important theorem.

Theorem 3. Let $\{X_n\}$ be a sequence of iid RVs with common finite mean $\mu = EX_1$. Then

$$n^{-1}S_n \xrightarrow{P} \mu \qquad \text{as } n \to \infty.$$

Proof. Let us take $c = n$ in (6) and replace ε by $n\varepsilon$; then we have

$$P\{|S_n - m_n| > n\varepsilon\} \le \frac{1}{n\varepsilon^2}E(X_1^n)^2 + nP\{|X_1| > n\},$$

where X_1^n is X_1 truncated at n.

First note that $E|X_1| < \infty \Rightarrow nP\{|X_1| > n\} \to 0$ as $n \to \infty$. Now (see remarks following Lemma 3.2.1)

$$\begin{aligned} E(X_1^n)^2 &= 2\int_0^n xP\{|X_1| > x\}\,dx \\ &= 2\left(\int_0^A + \int_A^n\right) xP\{|X_1| > x\}\,dx, \end{aligned}$$

where A is chosen sufficiently large that

$$xP\{|X_1| > x\} < \frac{\delta}{2} \qquad \text{for all } x \ge A, \delta > 0 \text{ arbitrary.}$$

Thus

$$E(X_1^n)^2 \le c + \delta\int_A^n dx \le c + n\delta,$$

where c is a constant. It follows that

$$\frac{1}{n\varepsilon^2}E(X_1^n)^2 \le \frac{c}{n\varepsilon^2} + \frac{\delta}{\varepsilon^2},$$

and since δ is arbitrary, $(1/n\varepsilon^2)E(X_1^n)^2$ can be made arbitrarily small for sufficiently large n. The proof is now completed by the simple observation that since $EX_j = \mu$,

$$\frac{m_n}{n} \to \mu \qquad \text{as } n \to \infty.$$

We emphasize that in Theorem 3 we require only that $E|X_1| < \infty$; nothing is said about the variance. Theorem 3 is due to Khintchine.

Example 4. Let $X_1, X_2, \ldots$ be iid RVs with $E|X_1|^k < \infty$ for some positive integer k. Then

$$\sum_{j=1}^{n} \frac{X_j^k}{n} \xrightarrow{P} EX_1^k \qquad \text{as } n \to \infty.$$

Thus, if $EX_1^2 < \infty$, then $\sum_1^n X_j^2/n \xrightarrow{P} EX_1^2$; and since $(\sum_{j=1}^n X_j/n)^2 \xrightarrow{P} (EX_1)^2$, it follows that

$$\frac{\Sigma X_j^2}{n} - \left(\frac{\Sigma X_j}{n}\right)^2 \xrightarrow{P} \operatorname{var}(X_1).$$

Example 5. Let $X_1, X_2, \ldots$ be iid RVs with common PDF

$$f(x) = \begin{cases} \dfrac{1+\delta}{x^{2+\delta}}, & x \geq 1 \\ 0, & x < 1 \end{cases}, \qquad \delta > 0.$$

Then

$$\begin{aligned} E|X| &= (1+\delta)\int_1^\infty \frac{1}{x^{1+\delta}}\,dx \\ &= \frac{1+\delta}{\delta} < \infty, \end{aligned}$$

and the law of large numbers holds, that is,

$$n^{-1}S_n \xrightarrow{P} \frac{1+\delta}{\delta} \qquad \text{as } n \to \infty.$$

PROBLEMS 6.3

1. Let $X_1, X_2, \ldots$ be a sequence of iid RVs with common uniform distribution on $[0, 1]$. Also, let $Z_n = (\prod_{i=1}^n X_i)^{1/n}$ be the geometric mean of $X_1, X_2, \ldots, X_n$, $n = 1, 2, \ldots$. Show that $Z_n \xrightarrow{P} c$, where c is a constant. Find c.
2. Let $X_1, X_2, \ldots$ be iid RVs with finite second moment. Let

$$Y_n = \frac{2}{n(n+1)} \sum_{i=1}^{n} i X_i.$$

Show that $Y_n \xrightarrow{P} EX_1$.

3. Let $X_1, X_2, \ldots$ be a sequence of iid RVs with $EX_i = \mu$ and $\text{var}(X_i) = \sigma^2$. Let $S_k = \sum_{j=1}^{k} X_j$. Does the sequence S_k obey the WLLN in the sense of Definition 1? If so, find the centering and the norming constants.

4. Let $\{X_n\}$ be a sequence of RVs for which $\text{var}(X_n) \le C$ for all n and $\rho_{ij} = \text{cov}(X_i, X_j) \to 0$ as $|i - j| \to \infty$. Show that the WLLN holds.

5. For the following sequences of independent RVs, does the WLLN hold?

(a) $P\{X_k = \pm 2^k\} = \frac{1}{2}$.

(b) $P\{X_k = \pm k\} = 1/2\sqrt{k}$, $P\{X_k = 0\} = 1 - (1/\sqrt{k})$.

(c) $P\{X_k = \pm 2^k\} = 1/2^{2k+1}$, $P\{X_k = 0\} = 1 - (1/2^{2k})$.

(d) $P\{X_k = \pm 1/k\} = \frac{1}{2}$.

(e) $P\{X_k = \pm\sqrt{k}\} = \frac{1}{2}$.

6. Let $X_1, X_2, \ldots$ be a sequence of independent RVs such that $\text{var}(X_k) < \infty$ for $k = 1, 2, \ldots$, and $(1/n^2) \sum_{k=1}^{n} \text{var}(X_k) \to 0$ as $n \to \infty$. Prove the WLLN, using Theorem 2.

7. Let X_n be a sequence of RVs with common finite variance σ^2. Suppose that the correlation coefficient between X_i and X_j is < 0 for all $i \ne j$. Show that the WLLN holds for the sequence $\{X_n\}$.

8. Let $\{X_n\}$ be a sequence of RVs such that X_k is independent of X_j for $j \ne k+1$ or $j \ne k-1$. If $\text{var}(X_k) < C$ for all k, where C is a constant, the WLLN holds for $\{X_k\}$.

9. For any sequence of RVs $\{X_n\}$, show that

$$\max_{1 \le k \le n} |X_k| \xrightarrow{P} 0 \Rightarrow n^{-1} S_n \xrightarrow{P} 0.$$

10. Let $X_1, X_2, \ldots$ be iid $C(1, 0)$ RVs. Use Theorem 2 to show that the weak law of large numbers does not hold. That is, show that

$$E \frac{S_n^2}{n^2 + S_n^2} \nrightarrow 0 \qquad \text{as } n \to \infty, \qquad \text{where} \quad S_n = \sum_{k=1}^{n} X_k, \quad n = 1, 2, \ldots.$$

11. Let $\{X_n\}$ be a sequence of iid RVs with $P\{X_n \ge 0\} = 1$. Let $S_n = \sum_{j=1}^{n} X_j$, $n = 1, 2, \ldots$. Suppose that $\{a_n\}$ is a sequence of constants such that $a_n^{-1} S_n \xrightarrow{P} 1$. Show that (a) $a_n \to \infty$ as $n \to \infty$, and (b) $a_{n+1}/a_n \to 1$.

6.4 STRONG LAW OF LARGE NUMBERS†

In this section we obtain a stronger form of the law of large numbers discussed in Section 6.3. Let $X_1, X_2, \ldots$ be a sequence of RVs defined on a probability space $(\Omega, \mathcal{S}, P)$.

Definition 1. We say that the sequence $\{X_n\}$ obeys the *strong law of large numbers* (SLLN) with respect to the norming constants $\{B_n\}$ if there exists a sequence of (centering) constants $\{A_n\}$ such that

$$B_n^{-1}(S_n - A_n) \xrightarrow{\text{a.s.}} 0 \qquad \text{as } n \to \infty. \tag{1}$$

Here $B_n > 0$ and $B_n \to \infty$ as $n \to \infty$.

We will obtain sufficient conditions for a sequence $\{X_n\}$ to obey the SLLN. In what follows we will be interested mainly in the case $B_n = n$. Indeed, when we speak of the SLLN we will assume that we are speaking of the norming constants $B_n = n$, unless specified otherwise.

We start with the *Borel–Cantelli lemma*. Let $\{A_j\}$ be any sequence of events in $\mathcal{S}$. We recall that

$$\overline{\lim_{n\to\infty}} A_n = \lim_{n\to\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \tag{2}$$

We will write $A = \overline{\lim}_{n\to\infty} A_n$. Note that A is the event that *infinitely many* of the A_n occur. We will sometimes write

$$PA = P(\overline{\lim_{n\to\infty}} A_n) = P(A_n \text{ i.o.}),$$

where "i.o." stands for "infinitely often." In view of Theorem 6.2.11 and Remark 6.2.6 we have $X_n \xrightarrow{\text{a.s.}} 0$ if and only if $P\{|X_n| > \varepsilon \text{ i.o.}\} = 0$ for all $\varepsilon > 0$.

Theorem 1 (Borel–Cantelli Lemma)

(a) Let $\{A_n\}$ be a sequence of events such that $\sum_{n=1}^{\infty} PA_n < \infty$. Then $PA = 0$.
(b) If $\{A_n\}$ is an independent sequence of events such that $\sum_{n=1}^{\infty} PA_n = \infty$, then $PA = 1$.

Proof.
(a) $PA = P(\lim_{n\to\infty} \bigcup_{k=n}^{\infty} A_k) = \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty} A_k) \le \lim_{n\to\infty} \sum_{k=n}^{\infty} PA_k = 0$.
(b) We have $A^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$, so that

†This section may be omitted on first reading

$$PA^c = P\left(\lim_{n\to\infty}\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right).$$

For $n_0 > n$, we see that $\bigcap_{k=n}^{\infty} A_k^c \subset \bigcap_{k=n}^{n_0} A_k^c$, so that

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \le \lim_{n_0\to\infty} P\left(\bigcap_{k=n}^{n_0} A_k^c\right) = \lim_{n_0\to\infty}\prod_{k=n}^{n_0}(1 - PA_k),$$

because $\{A_n\}$ is an independent sequence of events. Now we use the elementary inequality

$$1 - \exp\left(-\sum_{j=n}^{n_0}\alpha_j\right) \le 1 - \prod_{j=n}^{n_0}(1-\alpha_j) \le \sum_{j=n}^{n_0}\alpha_j, \qquad n_0 > n, \quad 1 \ge \alpha_j \ge 0,$$

to conclude that

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \le \lim_{n_0\to\infty} \exp\left(-\sum_{k=n}^{n_0} PA_k\right).$$

Since the series $\sum_{n=1}^{\infty} PA_n$ diverges, it follows that $PA^c = 0$ or $PA = 1$.

Corollary. Let $\{A_n\}$ be a sequence of independent events. Then PA is either 0 or 1.

The corollary follows since $\sum_{n=1}^{\infty} PA_n$ either converges or diverges.

As a simple application of the Borel–Cantelli lemma, we obtain a version of the SLLN.

Theorem 2. If $X_1, X_2, \ldots$ are iid RVs with common mean μ and finite fourth moment, then

$$P\left\{\lim_{n\to\infty}\frac{S_n}{n} = \mu\right\} = 1.$$

Proof. We have

$$E\{\Sigma(X_i - \mu)\}^4 = nE(X_1 - \mu)^4 + 6\binom{n}{2}\sigma^4 \le Cn^2.$$

By Markov's inequality,

$$P\left\{\left|\sum_1^n (X_i - \mu)\right| > n\varepsilon\right\} \le \frac{E\left[\sum_1^n (X_i - \mu)\right]^4}{(n\varepsilon)^4} \le \frac{Cn^2}{(n\varepsilon)^4} = \frac{C'}{n^2}.$$

Therefore,

$$\sum_{n=1}^{\infty} P\{|S_n - \mu n| > n\varepsilon\} < \infty,$$

and it follows by the Borel–Cantelli lemma that with probability 1 only finitely many of the events $\{\omega\colon |(S_n/n) - \mu| > \varepsilon\}$ occur, that is, $PA_\varepsilon = 0$, where

$$A_\varepsilon = \limsup_{n\to\infty} \left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\}.$$

The sets A_ε increase, as $\varepsilon \to 0$, to the ω set on which $S_n/n \not\to \mu$. Letting $\varepsilon \to 0$ through a countable set of values, we have

$$P\left\{\frac{S_n}{n} - \mu \not\to 0\right\} = P\left(\bigcup_k A_{1/k}\right) = 0.$$

Corollary. If $X_1, X_2, \ldots$ are iid RVs such that $P\{|X_n| < K\} = 1$ for all n, where K is a positive constant, then $n^{-1}S_n \xrightarrow{\text{a.s.}} \mu$.

Theorem 3. Let $X_1, X_2, \ldots$ be a sequence of independent RVs. Then

$$X_n \xrightarrow{\text{a.s.}} 0 \Leftrightarrow \sum_{n=1}^{\infty} P\{|X_n| > \varepsilon\} < \infty \qquad \text{for all } \varepsilon > 0.$$

Proof. Writing $A_n = \{|X_n| > \varepsilon\}$, we see that $\{A_n\}$ is a sequence of independent events. Since $X_n \xrightarrow{\text{a.s.}} 0$, $X_n \to 0$ on a set E^c with $PE = 0$. A point $\omega \in E^c$ belongs only to a finite number of A_n. It follows that

$$\limsup_{n\to\infty} A_n \subset E,$$

hence $P(A_n \text{ i.o.}) = 0$. By the Borel–Cantelli lemma [Theorem 1(b)] we must have $\sum_{n=1}^{\infty} PA_n < \infty$. [Otherwise, $\sum_{n=1}^{\infty} PA_n = \infty$, and then $P(A_n \text{ i.o.}) = 1$.]

In the other direction, let

$$A_{1/k} = \limsup_{n\to\infty} \left\{|X_n| > \frac{1}{k}\right\},$$

and use the argument in the proof of Theorem 2.

Example 1. We take an application of the Borel–Cantelli lemma to prove a.s. convergence.

Let $\{X_n\}$ have PMF

$$P(X_n = 0) = 1 - \frac{1}{n^\alpha}, \quad \text{and} \quad P(X_n = \pm n) = \frac{1}{2n^\alpha}.$$

Then $P(|X_n| > \varepsilon) = 1/n^\alpha$ and it follows that

$$\sum_{n=1}^{\infty} P(|X_n| > \varepsilon) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty \qquad \text{for } \alpha > 1.$$

Thus from Borel–Cantelli Lemma $P(A_n \text{ i.o.}) = 0$, where $A_n = \{|X_n| > \varepsilon\}$. Now using the argument in the proof of Theorem 2, we can show that $P(X_n \not\to 0\} = 0$.

We next prove some important lemmas that we will need subsequently.

Lemma 1 (Kolmogorov's Inequality). Let $X_1, X_2, \ldots, X_n$ be independent RVs with common mean 0 and variances σ_k^2, $k = 1, 2, \ldots, n$, respectively. Then for any $\varepsilon > 0$,

$$P\left\{\max_{1\le k\le n} |S_k| > \varepsilon\right\} \le \sum_1^n \frac{\sigma_i^2}{\varepsilon^2}. \tag{3}$$

Proof. Let $A_0 = \Omega$,

$$A_k = \left\{\max_{1\le j\le k} |S_j| \le \varepsilon\right\}, \qquad k = 1, 2, \ldots, n$$

and

$$\begin{aligned} B_k &= A_{k-1} \cap A_k^c \\ &= \{|S_1| \le \varepsilon, \ldots, |S_{k-1}| \le \varepsilon\} \cap \{\text{at least one of } |S_1|, \ldots, |S_k| \text{ is } > \varepsilon\} \\ &= \{|S_1| \le \varepsilon, \ldots, |S_{k-1}| \le \varepsilon, |S_k| > \varepsilon\}. \end{aligned}$$

It follows that

$$A_n^c = \sum_{k=1}^{n} B_k$$

and

$$B_k \subset \{|S_{k-1}| \le \varepsilon, |S_k| > \varepsilon\}.$$

As usual, let us write I_{B_k}, for the indicator function of the event B_k. Then

$$\begin{aligned} E(S_n I_{B_k})^2 &= E\{(S_n - S_k) I_{B_k} + S_k I_{B_k}\}^2, \\ &= E\{(S_n - S_k)^2 I_{B_k} + S_k^2 I_{B_k} + 2S_k(S_n - S_k) I_{B_k}\}. \end{aligned}$$

Since $S_n - S_k = X_{k+1} + \cdots + X_n$ and $S_k I_{B_k}$ are independent, and $EX_k = 0$ for all k, it follows that

$$E(S_n I_{B_k})^2 = E\{(S_n - S_k)I_{B_k}\}^2 + E(S_k I_{B_k})^2$$
$$\geq E(S_k I_{B_k})^2 \geq \varepsilon^2 P B_k.$$

The last inequality follows from the fact that in B_k, $|S_k| > \varepsilon$. Moreover,

$$\sum_{k=1}^{n} E(S_n I_{B_k})^2 = E(S_n^2 I_{A_n^c}) \leq E(S_n^2) = \sum_{1}^{n} \sigma_k^2,$$

so that

$$\sum_{1}^{n} \sigma_k^2 \geq \varepsilon^2 \sum_{1}^{n} P B_k = \varepsilon^2 P(A_n^c),$$

as asserted.

Corollary. Take $n = 1$; then

$$P\{|X_1| > \varepsilon\} \leq \frac{\sigma_1^2}{\varepsilon^2},$$

which is Chebychev's inequality.

Lemma 2 (Kronecker Lemma). If $\sum_{n=1}^{\infty} x_n$ converges to s (finite) and $b_n \uparrow \infty$, then

$$b_n^{-1} \sum_{k=1}^{n} b_k x_k \to 0.$$

Proof. Writing $b_0 = 0$, $a_k = b_k - b_{k-1}$, and $s_{n+1} = \sum_{k=1}^{n} x_k$, we have

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = \frac{1}{b_n} \sum_{k=1}^{n} b_k (s_{k+1} - s_k)$$
$$= \frac{1}{b_n} \left(b_n s_{n+1} + \sum_{1}^{n} b_{k-1} s_k \right) - \frac{1}{b_n} \sum_{k=1}^{n} b_k s_k$$
$$= s_{n+1} - \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) s_k$$
$$= s_{n+1} - \frac{1}{b_n} \sum_{k=1}^{n} a_k s_k.$$

It therefore suffices to show that $b_n^{-1} \sum_{k=1}^{n} a_k s_k \to s$. Since $s_n \to s$, there exists an $n_0 = n_0(\varepsilon)$ such that

$$|s_n - s| < \frac{\varepsilon}{2} \qquad \text{for } n > n_0.$$

Since $b_n \uparrow \infty$, let n_1 be an integer $> n_0$ such that

$$b_n^{-1} \left| \sum_1^{n_0} (b_k - b_{k-1})(s_k - s) \right| < \frac{\varepsilon}{2} \qquad \text{for } n > n_1.$$

Writing

$$r_n = b_n^{-1} \sum_{k=1}^{n} (b_k - b_{k-1}) s_k,$$

we see that

$$|r_n - s| = \frac{1}{b_n} \left| \sum_{k=1}^{n} (b_k - b_{k-1})(s_k - s) \right|;$$

and choosing $n > n_1$, we have

$$|r_n - s| \leq \left| \frac{1}{b_n} \sum_{k=1}^{n_0} (b_k - b_{k-1})(s_k - s) \right| + \frac{1}{b_n} \left| \sum_{k=n_0+1}^{n} (b_k - b_{k-1}) \frac{\varepsilon}{2} \right| < \varepsilon.$$

This completes the proof.

Theorem 4. If $\sum_{n=1}^{\infty} \operatorname{var}(X_n) < \infty$, then $\sum_{n=1}^{\infty} (X_n - EX_n)$ converges almost surely.

Proof. Without loss of generality, assume that $EX_n = 0$. By Kolmogorov's inequality,

$$P\left\{ \max_{1 \leq k \leq n} |S_{m+k} - S_m| \geq \varepsilon \right\} \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{n} \operatorname{var}(X_{m+k}).$$

Letting $n \to \infty$, we have

$$\begin{aligned} P\left\{ \max_{k \geq 1} |S_{m+k} - S_m| \geq \varepsilon \right\} &= P\left\{ \max_{k \geq m+1} |S_k - S_m| \geq \varepsilon \right\} \\ &\leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \operatorname{var}(X_k). \end{aligned}$$

It follows that

$$\lim_{m\to\infty} P\left\{\max_{k>m} |S_k - S_m| < \varepsilon\right\} = 1,$$

and since $\varepsilon > 0$ is arbitrary, we have

$$P\left\{\lim_{m\to\infty}\left|\sum_{j=m}^{\infty} X_j\right| = 0\right\} = 1.$$

Consequently, $\sum_{j=1}^{\infty} X_j$ converges a.s.

As a corollary we get a version of the SLLN for nonidentically distributed RVs which subsumes Theorem 2.

Corollary 1. Let $\{X_n\}$ be independent RVs. If

$$\sum_{k=1}^{\infty} \frac{\text{var}(X_k)}{B_k^2} < \infty, \qquad B_n \uparrow \infty,$$

then

$$\frac{S_n - ES_n}{B_n} \xrightarrow{\text{a.s.}} 0.$$

The corollary follows from Theorem 4 and the Kronecker lemma.

Corollary 2. Every sequence $\{X_n\}$ of independent RVs with uniformly bounded variances obeys the SLLN.

If $\text{var}(X_k) \le A$ for all k, and $B_k = k$, then

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \le A \sum_{1}^{\infty} \frac{1}{k^2} < \infty,$$

and it follows that

$$\frac{S_n - ES_n}{n} \xrightarrow{\text{a.s.}} 0.$$

Corollary 3 (Borel's Strong Law of Large Numbers). For a sequence of Bernoulli trials with (constant) probability p of success, the SLLN holds (with $B_n = n$ and $A_n = np$).

Since

$$EX_k = p, \qquad \text{var}(X_k) = p(1-p) \le \tfrac{1}{4}, \qquad 0 < p < 1,$$

the result follows from Corollary 2.

Corollary 4. Let $\{X_n\}$ be iid RVs with common mean μ and finite variance σ^2. Then

$$P\left\{\lim_{n\to\infty} \frac{S_n}{n} = \mu\right\} = 1.$$

Remark 1. Kolmogorov's SLLN is much stronger than Corollaries 1 and 4 of Theorem 4. It states that if $\{X_n\}$ is a sequence of iid RVs, then

$$n^{-1}S_n \xrightarrow{a.s.} \mu \iff E|X_1| < \infty,$$

and then $\mu = EX_1$. The proof requires more work and will not be given here. We refer the reader to Billingsley [5], Chung [14], Feller [23], or Laha and Rohatgi [56].

PROBLEMS 6.4

1. For the following sequences of independent RVs does the SLLN hold?
   (a) $P\{X_k = \pm 2^k\} = \frac{1}{2}$.
   (b) $P\{X_k = \pm k\} = 1/2\sqrt{k}$, $P\{X_k = 0\} = 1 - (1/\sqrt{k})$.
   (c) $P\{X_k = \pm 2^k\} = 1/2^{2k+1}$, $P\{X_k = 0\} = 1 - (1/2^{2k})$.
2. Let $X_1, X_2, \ldots$ be a sequence of independent RVs with $\sum_{k=1}^{\infty} \operatorname{var}(X_k)/k^2 < \infty$. Show that
$$\frac{1}{n^2}\sum_{k=1}^{n} \operatorname{var}(X_k) \to 0 \qquad \text{as } n \to \infty.$$
Does the converse also hold?
3. For what values of α does the SLLN hold for the sequence
$$P\{X_k = \pm k^{\alpha}\} = \tfrac{1}{2}\,?$$
4. Let $\{\sigma_k^2\}$ be a sequence of real numbers such that $\sum_{k=1}^{\infty} \sigma_k^2/k^2 = \infty$. Show that there exists a sequence of independent RVs $\{X_k\}$ with $\operatorname{var}(X_k) = \sigma_k^2$, $k = 1, 2, \ldots$, such that $n^{-1}\sum_{k=1}^{n}(X_k - EX_k)$ does not converge to 0 almost surely. [*Hint:* Let $P\{X_k = \pm k\} = \sigma_k^2/2k^2$, $P\{X_k = 0\} = 1 - (\sigma_k^2/k^2)$ if $\sigma_k/k \le 1$, and $P\{X_k = \pm\sigma_k\} = \frac{1}{2}$ if $\sigma_k/k > 1$. Apply the Borel–Cantelli lemma to $\{|X_n| > n\}$.]
5. Let X_n be a sequence of iid RVs with $E|X_n| = +\infty$. Show that for every positive number A, $P\{|X_n| > nA \text{ i.o.}\} = 1$ and $P\{|S_n| < nA \text{ i.o.}\} = 1$.
6. Construct an example to show that the converse of Theorem 1(a) does not hold.
7. Investigate a.s. convergence of $\{X_n\}$ to 0 in each case. (X_n's are independent in each case.)

(a) $P(X_n = e^n) = 1/n^2,\ P(X_n = 0) = 1 - 1/n^2.$

(b) $P(X_n = 0) = 1 - 1/n,\ P(X_n = \pm 1) = 1/(2n).$

6.5 LIMITING MOMENT GENERATING FUNCTIONS

Let $X_1, X_2, \ldots$ be a sequence of RVs. Let F_n be the DF of X_n, $n = 1, 2, \ldots$, and suppose that the MGF $M_n(t)$ of F_n exists. What happens to $M_n(t)$ as $n \to \infty$? If it converges, does it always converge to an MGF?

Example 1. Let $\{X_n\}$ be a sequence of RVs with PMF $P\{X_n = -n\} = 1$, $n = 1, 2, \ldots$. We have

$$M_n(t) = Ee^{tX_n} = e^{-tn} \to 0 \qquad \text{as } n \to \infty \quad \text{for all } t > 0,$$

and

$$M_n(t) \to +\infty \quad \text{for all } t < 0, \quad \text{and} \quad M_n(t) \to 1 \quad \text{at } t = 0.$$

Thus

$$M_n(t) \to M(t) = \begin{cases} 0, & t > 0 \\ 1, & t = 0 \\ \infty, & t < 0 \end{cases} \qquad \text{as } n \to \infty.$$

But $M(t)$ is not an MGF. Note that if F_n is the DF of X_n then

$$F_n(x) = \begin{cases} 0 & \text{if } x < -n \\ 1 & \text{if } x \geq -n \end{cases} \to F(x) = 1 \qquad \text{for all } x,$$

and F is not a DF.

Next suppose that X_n has MGF M_n and $X_n \xrightarrow{L} X$, where X is an RV with MGF M. Does $M_n(t) \to M(t)$ as $n \to \infty$? The answer to this question is in the negative.

Example 2 (Curtiss [18]). Consider the DF

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

where $c_n = 1/[2\tan^{-1}(n^2)]$. Clearly, as $n \to \infty$,

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

at all points of continuity of the DF F. The MGF associated with F_n is

$$M_n(t) = \int_{-n}^{n} c_n e^{tx} \frac{n}{1+n^2x^2}\,dx,$$

which exists for all t. The MGF corresponding to F is $M(t) = 1$ for all t. But $M_n(t) \nrightarrow M(t)$, since $M_n(t) \to \infty$ if $t \neq 0$. Indeed,

$$M_n(t) > \int_{0}^{n} c_n \frac{|t|^3x^3}{6} \frac{n}{1+n^2x^2}\,dx.$$

The following result is a weaker version of the continuity theorem due to Lévy and Cramér. We refer the reader to Lukacs [68, p. 47], or Curtiss [18], for details of the proof.

Theorem 1 (Continuity Theorem). Let $\{F_n\}$ be a sequence of DFs with corresponding MGFs $\{M_n\}$, and suppose that $M_n(t)$ exists for $|t| \le t_0$ for every n. If there exists a DF F with corresponding MGF M which exists for $|t| \le t_1 < t_0$, such that $M_n(t) \to M(t)$ as $n \to \infty$ for every $t \in [-t_1, t_1]$, then $F_n \xrightarrow{w} F$.

Example 3. Let X_n be an RV with PMF

$$P\{X_n = 1\} = \frac{1}{n}, \qquad P\{X_n = 0\} = 1 - \frac{1}{n}.$$

Then $M_n(t) = (1/n)e^t + [1-(1/n)]$ exists for all $t \in \mathcal{R}$, and $M_n(t) \to 1$ as $n \to \infty$ for all t. Here $M(t) = 1$ is the MGF of an RV X degenerate at 0. Thus $X_n \xrightarrow{L} X$.

Remark 1. The following notation on orders of magnitude is quite useful. We write $x_n = o(r_n)$ if given $\varepsilon > 0$, there exists an N such that $|x_n/r_n| < \varepsilon$ for all $n \ge N$, and $x_n = O(r_n)$ if there exists an N and a constant $c > 0$, such that $|x_n/r_n| < c$ for all $n \ge N$. We write $x_n = O(1)$ to express the fact that x_n is bounded for large n, and $x_n = o(1)$ to mean that $x_n \to 0$ as $n \to \infty$.

This notation is extended to RVs in an obvious manner. Thus $X_n = o_p(r_n)$ if, for every $\varepsilon > 0$ and $\delta > 0$, there exists an N such that $P(|X_n/r_n| \le \delta) \ge 1 - \varepsilon$ for $n \ge N$, and $X_n = O_p(r_n)$ if for $\varepsilon > 0$, there exists a $c > 0$ and an N such that $P(|X_n/r_n| \le c) \ge 1 - \varepsilon$. We write $X_n = o_p(1)$ to mean $X_n \xrightarrow{P} 0$.

The following lemma is quite useful in applications of Theorem 1.

Lemma 1. Let us write $f(x) = o(x)$, if $f(x)/x \to 0$ as $x \to 0$. We have

$$\lim_{n\to\infty} \left[1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right]^n = e^a \qquad \text{for every real } a.$$

Proof. By Taylor's expansion we have

$$\begin{aligned} f(x) &= f(0) + xf'(\theta x) \\ &= f(0) + xf'(0) + \{f'(\theta x) - f'(0)\}x, \qquad 0 < \theta < 1. \end{aligned}$$

If $f'(x)$ is continuous at $x = 0$, then as $x \to 0$,

$$f(x) = f(0) + xf'(0) + o(x).$$

Taking $f(x) = \log(1 + x)$, we have $f'(x) = (1 + x)^{-1}$, which is continuous at $x = 0$, so that

$$\log(1 + x) = x + o(x).$$

Then for sufficiently large n,

$$\begin{aligned} n\log\left[1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right] &= n\left\{\frac{a}{n} + o\left(\frac{1}{n}\right) + o\left[\frac{a}{n} + o\left(\frac{1}{n}\right)\right]\right\} \\ &= a + n\,o\left(\frac{1}{n}\right) \\ &= a + o(1). \end{aligned}$$

It follows that

$$\left[1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right]^n = e^{a+o(1)},$$

as asserted.

Example 4. Let $X_1, X_2, \ldots$ be iid $b(1, p)$ RVs. Also, let $S_n = \sum_1^n X_k$, and let $M_n(t)$ be the MGF of S_n. Then

$$M_n(t) = (q + pe^t)^n \qquad \text{for all } t,$$

where $q = 1 - p$. If we let $n \to \infty$ in such a way that np remains constant at λ, say, then, by Lemma 1,

$$M_n(t) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^t\right)^n = \left[1 + \frac{\lambda}{n}(e^t - 1)\right]^n \to \exp[\lambda(e^t - 1)] \qquad \text{for all } t,$$

which is the MGF of a $P(\lambda)$ RV. Thus, the binomial distribution function approaches the Poisson DF, provided that $n \to \infty$ in such a way that $np = \lambda > 0$.

Example 5. Let $X \sim P(\lambda)$. The MGF of X is given by

$$M(t) = \exp[\lambda(e^t - 1)] \qquad \text{for all } t.$$

Let $Y = (X - \lambda)/\sqrt{\lambda}$. Then the MGF of Y is given by

$$M_Y(t) = e^{-t\sqrt{\lambda}} M\left(\frac{t}{\sqrt{\lambda}}\right).$$

Also,

$$\begin{aligned}
\log M_Y(t) &= -t\sqrt{\lambda} + \log M\left(\frac{t}{\sqrt{\lambda}}\right) \\
&= -t\sqrt{\lambda} + \lambda(e^{t/\sqrt{\lambda}} - 1) \\
&= -t\sqrt{\lambda} + \lambda\left(\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{3!\,\lambda^{1/2}} + \cdots\right) \\
&= \frac{t^2}{2} + \frac{t^3}{3!\,\lambda^{3/2}} + \cdots.
\end{aligned}$$

It follows that

$$\log M_Y(t) \to \frac{t^2}{2} \qquad \text{as } \lambda \to \infty,$$

so that $M_Y(t) \to e^{t^2/2}$ as $\lambda \to \infty$, which is the MGF of an $\mathcal{N}(0, 1)$ RV.

For more examples, see Section 6.6.

Remark 2. As pointed out earlier, working with MGFs has the disadvantage that the existence of MGFs is a very strong condition. Working with CFs which always exist, on the other hand, permits a much wider application of the continuity theorem. Let ϕ_n be the CF of F_n. Then $F_n \xrightarrow{w} F$ if and only if $\phi_n \to \phi$ as $n \to \infty$ on $\mathcal{R}$, where ϕ is continuous at $t = 0$. In this case ϕ, the limit function, is the CF of the limit DF F.

Example 6. Let X be a $\mathcal{C}(0, 1)$ RV. Then its CF is given by

$$\begin{aligned}
E\exp(itX) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\cos tx}{1 + x^2}\,dx + i\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin tx}{1 + x^2}\,dx \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\cos tx}{1 + x^2}\,dx = e^{-|t|}
\end{aligned}$$

since the second integral on the right side vanishes.

Let $\{X_n\}$ be iid RVs with common law $\mathcal{L}(X)$, and set $Y_n = \sum_{j=1}^{n} X_j/n$. Then the CF of Y_n is given by

$$\varphi_n(t) = E\exp\left(it\sum_{j=1}^{n}\frac{X_j}{n}\right) = \prod_{j=1}^{n}\exp\left(-\frac{|t|}{n}\right)$$
$$= \exp(-|t|)$$

for all n. It follows φ_n is the CF of a $C(1, 0)$ RV. We could not have derived this result using MGFs. Also, if $U_n = \sum_{j=1}^{n} X_j/n^\alpha$ for $\alpha > 1$, then

$$\varphi_{U_n}(t) = \exp\left(-\frac{|t|}{n^{\alpha-1}}\right) \to 1$$

as $n \to \infty$ for all t. Since $\varphi(t) = 1$ is continuous at $t = 0$, φ is the CF of the limit DF F. Clearly, F is the DF of an RV degenerate at 0. Thus $\sum_{j=1}^{n} X_j/n^\alpha \xrightarrow{L,P} U$, where $P(U = 0) = 1$.

PROBLEMS 6.5

1. Let $X \sim NB(r; p)$. Show that

$$2pX \xrightarrow{L} Y \qquad \text{as } p \to 0,$$

where $Y \sim \chi^2(2r)$.

2. Let $X_n \sim NB(r_n; 1 - p_n)$, $n = 1, 2, \ldots$. Show that $X_n \xrightarrow{L} X$ as $r_n \to \infty$, $p_n \to 0$, in such a way that $r_n p_n \to \lambda$, where $X \sim P(\lambda)$.

3. Let $X_1, X_2, \ldots$ be independent RVs with PMF given by $P\{X_n = \pm 1\} = \frac{1}{2}$, $n = 1, 2, \ldots$. Let $Z_n = \sum_{j=1}^{n} X_j/2^j$. Show that $Z_n \xrightarrow{L} Z$, where $Z \sim U[-1, 1]$.

4. Let $\{X_n\}$ be a sequence of RVs with $X_n \sim G(n, \beta)$ where $\beta > 0$ is a constant (independent of n). Find the limiting distribution of X_n/n.

5. Let $X_n \sim \chi^2(n)$, $n = 1, 2, \ldots$. Find the limiting distribution of X_n/n^2.

6. Let $X_1, X_2, \ldots, X_n$ be jointly normal with $EX_i = 0$, $EX_i^2 = 1$ for all i and $\operatorname{cov}(X_i, X_j) = \rho$, $i, j = 1, 2, \ldots$ $(i \neq j)$. What is the limiting distribution of $n^{-1}S_n$, where $S_n = \sum_{k=1}^{n} X_k$?

6.6 CENTRAL LIMIT THEOREM

Let $X_1, X_2, \ldots$ be a sequence of RVs, and let $S_n = \sum_{k=1}^{n} X_k$, $n = 1, 2, \ldots$. In Sections 6.3 and 6.4 we investigated the convergence of the sequence of RVs $B_n^{-1}(S_n - A_n)$ to the degenerate RV. In this section we examine the convergence of $B_n^{-1}(S_n - A_n)$ to a nondegenerate RV. Suppose that for a suitable choice of constants

A_n and $B_n > 0$, the RVs $B_n^{-1}(S_n - A_n) \xrightarrow{L} Y$. What are the properties of this limit RV Y? The question as posed is far too general and is not of much interest unless the RVs X_i are suitably restricted. For example, if we take X_1 with DF F and $X_2, X_3, \ldots$ to be 0 with probability 1, choosing $A_n = 0$ and $B_n = 1$ leads to F as the limit DF.

We recall (Example 6.5.6) that if $X_1, X_2, \ldots, X_n$ are iid RVs with common law $\mathcal{C}(1, 0)$, then $n^{-1}S_n$ is also $\mathcal{C}(1, 0)$. Again, if $X_1, X_2, \ldots, X_n$ are iid $\mathcal{N}(0, 1)$ RVs then $n^{-1/2}S_n$ is also $\mathcal{N}(0, 1)$ (Corollary 2 to Theorem 5.3.22). We note thus that for certain sequences of RVs there exist sequences A_n and $B_n > 0$, $B_n \to \infty$, such that $B_n^{-1}(S_n - A_n) \xrightarrow{L} Y$. In the Cauchy case $B_n = n$, $A_n = 0$, and in the normal case $B_n = n^{1/2}$, $A_n = 0$. Moreover, we see that Cauchy and normal distributions appear as limiting distributions—in these two cases, because of the reproductive nature of the distributions. Cauchy and normal distributions are examples of stable distributions.

Definition 1. Let X_1, X_2 be iid nondegenerate RVs with common DF F. Let a_1, a_2 be any positive constants. We say that F is *stable* if there exist constants A and B (depending on a_1, a_2) such that the RV $B^{-1}(a_1X_1 + a_2X_2 - A)$ also has the DF F.

Let $X_1, X_2, \ldots$ be iid RVs with common DF F. We remark without proof (see Loève [64, p. 339]) that only stable distributions occur as limits. To make this statement more precise, we make the following definition.

Definition 2. Let $X_1, X_2, \ldots$ be iid RVs with common DF F. We say that F *belongs to the domain of attraction* of a distribution V if there exist norming constants $B_n > 0$ and centering constants A_n such that as $n \to \infty$,

$$P\{B_n^{-1}(S_n - A_n) \le x\} \to V(x) \tag{1}$$

at all continuity points x of V.

In view of the statement after Definition 1, we see that only stable distributions possess domains of attraction. From Definition 1 we also note that each stable law belongs to its own domain of attraction. The study of stable distributions is beyond the scope of this book. We restrict ourselves to seeking conditions under which the limit law V is the normal distribution. The importance of the normal distribution in statistics is due largely to the fact that a wide class of distributions F belongs to the domain of attraction of the normal law. Let us consider some examples.

Example 1. Let $X_1, X_2, \ldots, X_n$ be iid $b(1, p)$ RVs. Let

$$S_n = \sum_{k=1}^{n} X_k, \quad \text{and} \quad A_n = ES_n = np, \qquad B_n = \sqrt{\operatorname{var}(S_n)} = \sqrt{np(1-p)}.$$

Then

$$\begin{aligned}
M_n(t) &= E\exp\left[\frac{S_n - np}{\sqrt{np(1-p)}}t\right] \\
&= \prod_{i=1}^{n} E\exp\left[\frac{X_i - p}{\sqrt{np(1-p)}}t\right] \\
&= \exp\left[-\frac{npt}{\sqrt{np(1-p)}}\right]\left\{q + p\exp\left[\frac{t}{\sqrt{np(1-p)}}\right]\right\}^n, \qquad q = 1-p, \\
&= \left[q\exp\left(-\frac{pt}{\sqrt{npq}}\right) + p\exp\left(\frac{qt}{\sqrt{npq}}\right)\right]^n \\
&= \left[1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n.
\end{aligned}$$

It follows from Lemma 6.5.1 that

$$M_n(t) \to e^{t^2/2} \qquad \text{as } n \to \infty,$$

and since $e^{t^2/2}$ is the MGF of an $\mathcal{N}(0, 1)$ RV, we have by the continuity theorem

$$P\left\{\frac{S_n - np}{\sqrt{npq}} \le x\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{t^2/2}\,dt \qquad \text{for all } x \in \mathcal{R}.$$

Example 2. Let $X_1, X_2, \ldots, X_n$ be iid $\chi^2(1)$ RVs. Then $S_n \sim \chi^2(n)$, $ES_n = n$, and $\text{var}(S_n) = 2n$. Also let $Z_n = (S_n - n)/\sqrt{2n}$; then

$$\begin{aligned}
M_n(t) &= Ee^{tZ_n} \\
&= \exp\left(-t\sqrt{\frac{n}{2}}\right)\left(1 - \frac{2t}{\sqrt{2n}}\right)^{-n/2}, \qquad 2t < \sqrt{2n}, \\
&= \left[\exp\left(t\sqrt{\frac{2}{n}}\right) - t\sqrt{\frac{2}{n}}\exp\left(t\sqrt{\frac{2}{n}}\right)\right]^{-n/2}, \qquad t < \sqrt{\frac{n}{2}}.
\end{aligned}$$

Using Taylor's approximation, we get

$$\exp\left(t\sqrt{\frac{2}{n}}\right) = 1 + t\sqrt{\frac{2}{n}} + \frac{t^2}{2}\left(\sqrt{\frac{2}{n}}\right)^2 + \frac{1}{6}\exp(\theta_n)\left(t\sqrt{\frac{2}{n}}\right)^3,$$

where $0 < \theta_n < t\sqrt{2/n}$. It follows that

$$M_n(t) = \left(1 - \frac{t^2}{n} + \frac{\zeta(n)}{n}\right)^{-n/2},$$

where

$$\zeta(n) = -\sqrt{\frac{2}{n}}t^3 + \left(\frac{t^3}{3}\sqrt{\frac{2}{n}} - \frac{2t^4}{3n}\right)\exp(\theta_n) \to 0 \qquad \text{as } n \to \infty,$$

for every fixed t. We have from Lemma 6.5.1 that $M_n(t) \to e^{t^2/2}$ as $n \to \infty$ for all real t, and it follows that $Z_n \xrightarrow{L} Z$, where Z is $\mathcal{N}(0, 1)$.

These examples suggest that if we take iid RVs with finite variance, and take $A_n = ES_n$, $B_n = \sqrt{\operatorname{var}(S_n)}$, then $B_n^{-1}(S_n - A_n) \xrightarrow{L} Z$, where Z is $\mathcal{N}(0, 1)$. This is the central limit result, which we now prove. The reader should note that in both Examples 1 and 2, we used more than just the existence of $E|X|^2$. Indeed, the MGF exists and hence moments of all order exist. The existence of MGF is not a necessary condition.

Theorem 1 (Lindeberg–Lévy Central Limit Theorem). Let $\{X_n\}$ be a sequence of iid RVs with $0 < \operatorname{var}(X_n) = \sigma^2 < \infty$ and common mean μ. Let $S_n = \sum_{j=1}^n X_j$, $n = 1, 2, \ldots$. Then for every $x \in \mathcal{R}$,

$$\lim_{n\to\infty} P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right\} = \lim_{n\to\infty} P\left\{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,du.$$

Proof. The proof we give here assumes that the MGF of X_n exists. Without loss of generality, we also assume that $EX_n = 0$ and $\operatorname{var}(X_n) = 1$. Let M be the MGF of X_n. Then the MGF of $S_n/\sqrt{n}$ is given by

$$M_n(t) = E\exp\left(\frac{tS_n}{\sqrt{n}}\right) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

and

$$\ln M_n(t) = n\,\ln M(t/\sqrt{n}) = \frac{\ln M(t/\sqrt{n})}{1/n}$$
$$= \frac{L(t/\sqrt{n})}{1/n},$$

where $L(t/\sqrt{n}) = \ln M(t/\sqrt{n})$. Clearly, $L(0) = \ln(1) = 0$, so that as $n \to \infty$, the conditions for L'Hospital's rule are satisfied. It follows that

$$\lim_{n\to\infty} \ln M_n(t) = \lim_{n\to\infty} \frac{L'(t/\sqrt{n})t}{2/\sqrt{n}},$$

and since $L'(0) = EX = 0$, we can use L'Hospital's rule once again, to get

$$\lim_{n\to\infty} \ln M_n(t) = \lim_{n\to\infty} \frac{L''(t/\sqrt{n})t^2}{2} = \frac{t^2}{2}$$

using $L''(0) = \text{var}(X) = 1$. Thus

$$M_n(t) \longrightarrow \exp\left(\frac{t^2}{2}\right) = M(t)$$

where $M(t)$ is the MGF of a $\mathcal{N}(0, 1)$ RV.

Remark 1. In the proof above we could have used the Taylor series expansion of M to arrive at the same result.

Remark 2. Even though we proved Theorem 1 for the case when the MGF of X_n's exists, we will use the result whenever $0 < EX_n^2 = \sigma^2 < \infty$. The use of CFs would have provided a complete proof of Theorem 1. Let ϕ be the CF of X_n. Assuming again, without loss of generality, that $EX_n = 0$, $\text{var}(X_n) = 1$, we can write

$$\phi(t) = 1 - \tfrac{1}{2}t^2 + t^2 o(1).$$

Thus the CF of $S_n/\sqrt{n}$ is

$$\left[\phi\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 - \frac{1}{2n}t^2 + \frac{t^2}{n}o(1)\right]^n$$

which converges to $\exp(-t^2/2)$, which is the CF of a $\mathcal{N}(0, 1)$ RV. The devil is in the details of the proof.

The following converse to Theorem 1 holds.

Theorem 2. Let $X_1, X_2, \ldots, X_n$ be iid RVs such that $n^{-1/2}S_n$ has the same distribution for every $n = 1, 2, \ldots$. Then, if $EX_i = 0$, $\text{var}(X_i) = 1$, the distribution of X_i must be $\mathcal{N}(0, 1)$.

Proof. Let F be the DF of $n^{-1/2}S_n$. By the central limit theorem,

$$\lim_{n\to\infty} P\{n^{-1/2}S_n \le x\} = \Phi(x).$$

Also, $P\{n^{-1/2}S_n \le x\} = F(x)$ for each n. It follows that we must have $F(x) = \Phi(x)$.

Example 3. Let $X_1, X_2, \ldots$ be iid RVs with common PMF

$$P\{X = k\} = p(1-p)^k, \qquad k = 0, 1, 2, \ldots, \quad 0 < p < 1, \quad q = 1 - p.$$

Then $EX = q/p$, $\text{var}(X) = q/p^2$. By Theorem 1 we see that

$$P\left\{\frac{S_n - n(q/p)}{\sqrt{nq}}p \le x\right\} \to \Phi(x) \qquad \text{as } n \to \infty \text{ for all } x \in \mathcal{R}.$$

Example 4. Let $X_1, X_2, \ldots$ be iid RVs with common $B(\alpha, \beta)$ distribution. Then

$$EX = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \text{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

By the corollary to Theorem 1, it follows that

$$\frac{S_n - n[\alpha/(\alpha+\beta)]}{\sqrt{\alpha\beta n/[(\alpha+\beta+1)(\alpha+\beta)^2]}} \xrightarrow{L} Z,$$

where Z is $\mathcal{N}(0, 1)$.

For nonidentically distributed RVs we state, without proof, the following result due to Lindeberg.

Theorem 3. Let $X_1, X_2, \ldots$ be independent RVs with DFs $F_1, F_2, \ldots$, respectively. Let $EX_k = \mu_k$ and $\text{var}(X_k) = \sigma_k^2$, and write

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2.$$

If the F_k's are absolutely continuous with PDF f_k, assume that the relation

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\varepsilon s_n} (x-\mu_k)^2 f_k(x)\,dx = 0 \tag{2}$$

holds for all $\varepsilon > 0$. (A similar condition can be stated for the discrete case.) Then

$$S_n^* = \frac{\sum_{j=1}^{n} X_j - \sum_{j=1}^{n} \mu_j}{s_n} \xrightarrow{L} Z \sim \mathcal{N}(0, 1). \tag{3}$$

Condition (2) is known as the *Lindeberg condition.*

Feller [21] has shown that condition (2) is necessary as well in the following sense. For independent RVs $\{X_k\}$ for which (3) holds and

$$P\left\{\max_{1\le k\le n} \left|X_k - EX_k\right| > \varepsilon\sqrt{\text{var}(S_n)}\right\} \to 0,$$

(2) holds for every $\varepsilon > 0$.

Example 5. Let $X_1, X_2, \ldots$ be independent RVs such that X_k is $U(-a_k, a_k)$. Then $EX_k = 0$, $\text{var}(X_k) = (1/3)a_k^2$. Suppose that $|a_k| < a$ and $\sum_1^n a_k^2 \to \infty$ as $n \to \infty$. Then

$$\begin{aligned}\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|>\varepsilon s_n} x^2 f_k(x)\,dx &\le \frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|>\varepsilon s_n} a^2 \frac{1}{2a_k}\,dx \\ &\le \frac{a^2}{s_n^2}\sum_{k=1}^{n} P\{|X_k| > \varepsilon s_n\} \le \frac{a^2}{s_n^2}\sum_{k=1}^{n}\frac{\text{var}(X_k)}{\varepsilon^2 s_n^2} \\ &= \frac{a^2}{\varepsilon^2 s_n^2} \to 0 \qquad \text{as } n \to \infty.\end{aligned}$$

If $\sum_1^\infty a_k^2 < \infty$, then $s_n^2 \uparrow A^2$, say, as $n \to \infty$. For fixed k, we can find ε_k such that $\varepsilon_k A < a_k$ and then $P\{|X_k| > \varepsilon_k s_n\} \ge P\{|X_k| > \varepsilon_k A\} > 0$. For $n \ge k$, we have

$$\begin{aligned}\frac{1}{s_n^2}\sum_{j=1}^{n}\int_{|x|>\varepsilon_k s_n} x^2 f_j(x)\,dx &\ge \frac{s_n^2\varepsilon_k^2}{s_n^2}\sum_{j=1}^{n} P\{|X_j| > \varepsilon_k s_n\} \\ &\ge \varepsilon_k^2 P\{|X_k| > \varepsilon_k s_n\} \\ &> 0,\end{aligned}$$

so that the Lindeberg condition does not hold. Indeed, if $X_1, X_2, \ldots$ are independent RVs such that there exists a constant A with $P\{|X_n| \le A\} = 1$ for all n, the Lindeberg condition (2) is satisfied if $s_n^2 \to \infty$ as $n \to \infty$. To see this, suppose that $s_n^2 \to \infty$. Since the X_k's are uniformly bounded, so are the RVs $X_k - EX_k$. It follows that for every $\varepsilon > 0$ we can find an N_ε such that for $n \ge N_\varepsilon$, $P\{|X_k - EX_k| < \varepsilon s_n$, $k = 1, 2, \ldots, n\} = 1$. The Lindeberg condition follows immediately. The converse also holds, for if $\lim_{n\to\infty} s_n^2 < \infty$ and the Lindeberg condition holds, there exists a constant $A < \infty$ such that $s_n^2 \to A^2$. For any fixed j, we can find an $\varepsilon > 0$ such that $P\{|X_j - \mu_j| > \varepsilon A\} > 0$. Then, for $n \ge j$,

$$\begin{aligned}\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\varepsilon s_n} (x-\mu_k)^2 f_k(x)\,dx &\ge \varepsilon^2 \sum_{k=1}^{n} P\{|X_k - \mu_k| > \varepsilon s_n\} \\ &\ge \varepsilon^2 P\{|X_j - \mu_j| > \varepsilon A\} \\ &> 0,\end{aligned}$$

and the Lindeberg condition does not hold. This contradiction shows that $s_n^2 \to \infty$ is also a necessary condition; that is, for a sequence of uniformly bounded independent RVs, a necessary and sufficient condition for the central limit theorem to hold is $s_n^2 \to \infty$ as $n \to \infty$.

Example 6. Let $X_1, X_2, \ldots$ be independent RVs such that $\alpha_k = E|X_k|^{2+\delta} < \infty$ for some $\delta > 0$ and $\alpha_1 + \alpha_2 + \cdots + \alpha_n = o(s_n^{2+\delta})$. Then the Lindeberg condition is satisfied, and the central limit theorem holds. This result is due to Lyapunov. We have

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|>\varepsilon s_n} x^2 f_k(x)\,dx \le \frac{1}{\varepsilon^\delta s_n^{2+\delta}}\sum_{k=1}^{n}\int_{-\infty}^{\infty}|x|^{2+\delta} f_k(x)\,dx$$

$$= \frac{\sum_{k=1}^{n}\alpha_k}{\varepsilon^\delta s_n^{2+\delta}} \to 0 \qquad \text{as } n \to \infty.$$

A similar argument applies in the discrete case.

Remark 3. Both the central limit theorem (CLT) and the (weak) law of large numbers (WLLN) hold for a large class of sequences of RVs $\{X_n\}$. If the $\{X_n\}$ are independent uniformly bounded RVs, that is, if $P\{|X_n| \le M\} = 1$, the WLLN (Theorem 6.3.1) holds; the CLT holds provided that $s_n^2 \to \infty$ (Example 5).

If the RVs $\{X_n\}$ are iid, then the CLT is a stronger result than the WLLN in that the former provides an estimate of the probability $P\{|S_n - n\mu|/n \ge \varepsilon\}$. Indeed,

$$P\{|S_n - n\mu| > n\varepsilon\} = P\left\{\frac{|S_n - n\mu|}{\sigma\sqrt{n}} > \frac{\varepsilon}{\sigma}\sqrt{n}\right\}$$

$$\approx 1 - P\left\{|Z| \le \frac{\varepsilon}{\sigma}\sqrt{n}\right\},$$

where Z is $\mathcal{N}(0, 1)$, and the law of large number follows. On the other hand, we note that the WLLN does not require the existence of a second moment.

Remark 4. If $\{X_n\}$ are independent RVs, it is quite possible that the CLT may apply to the X_n's, but not the WLLN.

Example 7 (Feller [22, p. 255]). Let $\{X_k\}$ be independent RVs with PMF

$$P\{X_k = k^\lambda\} = P\{X_k = -k^\lambda\} = \tfrac{1}{2}, \qquad k = 1, 2, \ldots.$$

Then $EX_k = 0$, $\text{var}(X_k) = k^{2\lambda}$. Also let $\lambda > 0$; then

$$s_n^2 = \sum_{k=1}^{n} k^{2\lambda} \le \int_0^{n+1} x^{2\lambda}\,dx = \frac{(n+1)^{2\lambda+1}}{2\lambda+1}.$$

It follows that if $0 < \lambda < \frac{1}{2}$, $s_n/n \to 0$, and by Corollary 2 to Theorem 6.3.1, the WLLN holds. Now $k^\lambda < n^\lambda$, so that the sum $\sum_{k=1}^{n}\sum_{|x_{kl}|>\varepsilon s_n} x_{kl}^2 p_{kl}$ will be nonzero if $n^\lambda > \varepsilon s_n \approx \varepsilon[n^{\lambda+1/2}/\sqrt{(2\lambda+1)}]$. It follows that as long as $n > (2\lambda+1)\varepsilon^{-2}$,

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\sum_{|x_{kl}|>\varepsilon s_n} x_{kl}^2 p_{kl} = 0$$

and the Lindeberg condition holds. Thus the CLT holds for $\lambda > 0$. This means that

$$P\left\{a < \sqrt{\frac{2\lambda+1}{n^{2\lambda+1}}}S_n < b\right\} \to \int_a^b \frac{e^{-t^2/2}\,dt}{\sqrt{2\pi}}.$$

Thus

$$P\left\{\frac{an^{\lambda+1/2-1}}{\sqrt{2\lambda+1}} < \frac{S_n}{n} < \frac{bn^{\lambda+1/2-1}}{\sqrt{2\lambda+1}}\right\} \to \int_a^b \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$$

and the WLLN cannot hold for $\lambda \geq \frac{1}{2}$.

We conclude this section with some remarks concerning the application of the CLT. Let $X_1, X_2, \ldots$ be iid RVs with common mean μ and variance σ^2. Let us write

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

and let z_1, z_2 be two arbitrary real numbers with $z_1 < z_2$. If F_n is the DF of Z_n, then

$$\begin{aligned}\lim_{n\to\infty} P\{z_1 < Z_n \leq z_2\} &= \lim_{n\to\infty}[F_n(z_2) - F_n(z_1)]\\ &= \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-t^2/2}\,dt,\end{aligned}$$

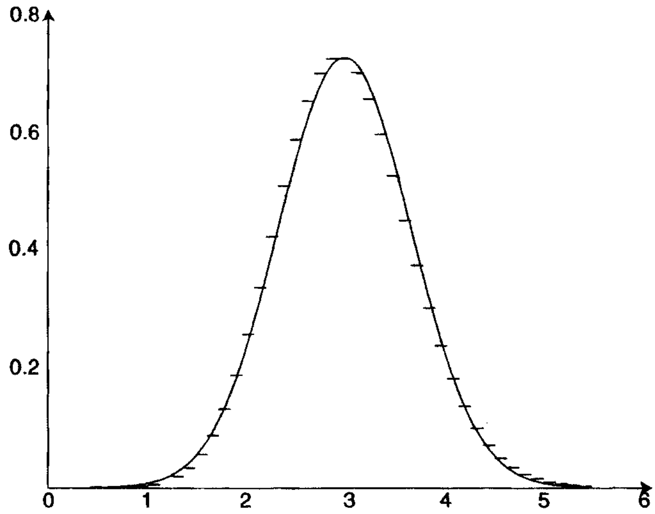
that is,

$$\lim_{n\to\infty} P\{z_1\sigma\sqrt{n} + n\mu < S_n \leq z_2\sigma\sqrt{n} + n\mu\} = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-t^2/2}\,dt. \tag{4}$$

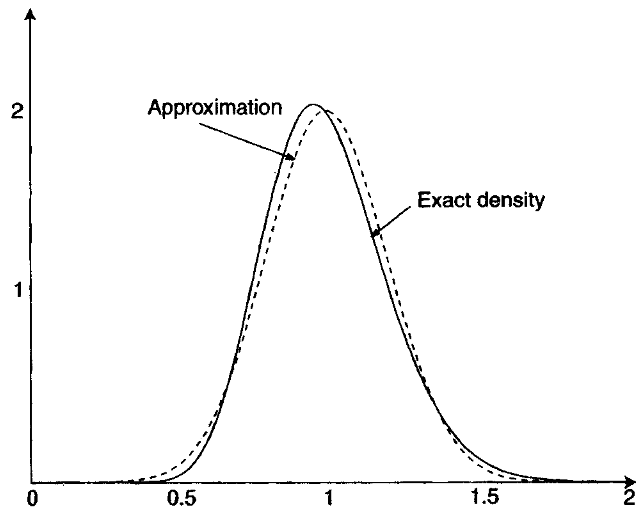
It follows that the RV $S_n = \sum_{k=1}^{n} X_k$ is asymptotically normally distributed (see Section 7.5) with mean $n\mu$ and variance $n\sigma^2$. Equivalently, the RV $n^{-1}S_n$ is asymptotically $\mathcal{N}(\mu, \sigma^2/n)$. This result is of great importance in statistics.

In Fig. 1 we show the distribution of $\overline{X}$ in sampling from $P(\lambda)$ and $G(1, 1)$. We have also superimposed, in each case, the graph of the corresponding normal approximation.

How large should n be before we apply approximation (4)? Unfortunately, the answer is not simple. Much depends on the underlying distribution, the corresponding speed of convergence and the accuracy one desires. There is a vast amount of literature on the speed of convergence and error bounds. We will content ourselves with some examples. The reader is referred to Rohatgi [88] for a detailed discussion.



(a)



(b)

Fig. 1. *(a)* Distribution of $\overline{X}$ for Poisson RV with mean 3 and normal approximation; *(b)* distribution of $\overline{X}$ for exponential RV with mean 1 and normal approximation.

In the discrete case when the underlying distribution is integer valued, approximation (4) is improved by applying the *continuity correction.* If X is integer valued, then for integers x_1, x_2

$$P\{x_1 \le X \le x_2\} = P\{x_1 - \tfrac{1}{2} < X < x_2 + \tfrac{1}{2}\}$$

which amounts to making the discrete space of values of X continuous by considering intervals of length 1 with midpoints at integers.

Example 8. Let $X_1, X_2, \ldots, X_n$ be iid $b(1, p)$ RVs. Then $ES_n = np$, and $\text{var}(S_n) = np(1-p)$, so $(S_n - np)/\sqrt{np(1-p)}$ is approximately $\mathcal{N}(0, 1)$.

Suppose that $n = 10$, $p = \frac{1}{2}$. Then from binomial tables, $P(X \le 4) = 0.3770$. Using normal approximation without continuity correction,

$$P(X \le 4) \approx P\left(Z \le \frac{4-5}{\sqrt{2.5}}\right) = P(Z \le -0.63) = 0.2643.$$

Applying continuity correction,

$$P(X \le 4) = P(X < 4.5) \approx P(Z \le -0.32) = 0.3745.$$

Next suppose that $n = 100$, $p = 0.1$. Then from binomial tables $P(X = 7) = 0.0889$. Using normal approximation, without continuity correction,

$$\begin{aligned} P(X = 7) &= P(6.0 < X < 8.0) \approx P(-1.33 < Z < -0.67) \\ &= 0.1596 \end{aligned}$$

and with continuity correction

$$\begin{aligned} P(X = 7) &= P(6.5 < X < 7.5) \approx P(-1.17 < Z < -0.83) \\ &= 0.0823 \end{aligned}$$

The rule of thumb is to use continuity correction, and normal approximation whenever $np(1-p) > 10$, and Poisson approximation with $\lambda = np$ for $p < 0.1, \lambda \le 10$.

Example 9. Let $X_1, X_2, \ldots$ be iid $P(\lambda)$ RVs. Then S_n has approximately an $\mathcal{N}(n\lambda, n\lambda)$ distribution for large n. Let $n = 64$, $\lambda = 0.125$. Then $S_n \sim P(8)$, and from Poisson distribution tables, $P(S_n = 10) = 0.099$. Using normal approximation,

$$\begin{aligned} P(S_n = 10) &= P(9.5 < S_n < 10.5) \approx P(0.53 < Z < 0.88) \\ &= 0.1087. \end{aligned}$$

If $n = 96$, $\lambda = 0.125$, then $S_n \sim P(12)$ and

$$\begin{aligned} P(S_n = 10) &= 0.105, && \text{exact,} \\ P(S_n = 10) &\approx 0.1009, && \text{normal approximation.} \end{aligned}$$

PROBLEMS 6.6

1. Let $\{X_n\}$ be a sequence of independent RVs with the following distributions. In each case, does the Lindeberg condition hold?

(a) $P\{X_n = \pm(1/2^n)\} = \frac{1}{2}$.

(b) $P\{X_n = \pm 2^{n+1}\} = 1/2^{n+3}$, $P\{X_n = 0\} = 1 - (1/2^{n+2})$.

(c) $P\{X_n = \pm 1\} = (1 - 2^{-n})/2$, $P\{X_n = \pm 2^{-n}\} = 1/2^{n+1}$.

(d) $\{X_n\}$ is a sequence of independent Poisson RVs with parameter λ_n, $n = 1, 2, \ldots$, such that $\sum_{k=1}^{n} \lambda_k \to \infty$.

(e) $P\{X_n = \pm 2^n\} = \frac{1}{2}$.

2. Let $X_1, X_2, \ldots$ be iid RVs with mean 0, variance 1, and $EX_i^4 < \infty$. Find the limiting distribution of

$$Z_n = \sqrt{n}\frac{X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n}}{X_1^2 + X_2^2 + \cdots + X_{2n}^2}.$$

3. Let $X_1, X_2, \ldots$ be iid RVs with mean α and variance σ^2, and let $Y_1, Y_2, \ldots$ be iid RVs with mean β ($\neq 0$) and variance τ^2. Find the limiting distribution of $Z_n = \sqrt{n}(\overline{X}_n - \alpha)/\overline{Y}_n$, where $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ and $\overline{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$.

4. Let $X \sim b(n, \theta)$. Use the CLT to find n such that $P_\theta\{X > n/2\} \geq 1 - \alpha$. In particular, let $\alpha = 0.10$ and $\theta = 0.45$. Calculate n, satisfying $P\{X > n/2\} \geq 0.90$.

5. Let $X_1, X_2, \ldots$ be a sequence of iid RVs with common mean μ and variance σ^2. Also, let $\overline{X} = n^{-1}\sum_{k=1}^{n} X_k$ and $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Show that $\sqrt{n}(\overline{X} - \mu)/S \xrightarrow{L} Z$, where $Z \sim \mathcal{N}(0, 1)$.

6. Let $X_1, X_2, \ldots, X_{100}$ be iid RVs with mean 75 and variance 225. Use Chebychev's inequality to calculate the probability that the sample mean will not differ from the population mean by more than 6. Then use the CLT to calculate the same probability, and compare your results.

7. Let $X_1, X_2, \ldots, X_{100}$ be iid $P(\lambda)$ RVs, where $\lambda = 0.02$. Let $S = S_{100} = \sum_{i=1}^{100} X_i$. Use the central limit result to evaluate $P\{S \geq 3\}$, and compare your result to the exact probability of the event $S \geq 3$.

8. Let $X_1, X_2, \ldots, X_{81}$ be iid RVs with mean 54 and variance 225. Use Chebychev's inequality to find the possible difference between the sample mean and the population mean with a probability of at least 0.75. Also use the CLT to do the same.

9. Use the CLT applied to a Poisson RV to show that $\lim_{n\to\infty} e^{-nt}\sum_{k=1}^{n-1}(nt)^k/k! = 1$ for $0 < t < 1$, $= \frac{1}{2}$ if $t = 1$, and 0 if $t > 1$.

10. Let $X_1, X_2, \ldots$ be a sequence of iid RVs with mean μ and variance σ^2, and assume that $EX_1^4 < \infty$. Write $V_n = \sum_{k=1}^{n}(X_k - \mu)^2$. Find the centering and norming constants A_n and B_n such that $B_n^{-1}(V_n - A_n) \xrightarrow{L} Z$, where Z is $\mathcal{N}(0, 1)$.

11. From an urn containing 10 identical balls numbered 0 through 9, n balls are drawn with replacement.

(a) What does the law of large numbers tell you about the appearance of 0's in the n drawings?

(b) How many drawings must be made in order that with probability at least 0.95, the relative frequency of the occurrence of 0's will be between 0.09 and 0.11?

(c) Use the CLT to find the probability that among the n numbers thus chosen, the number 5 will appear between $(n - 3\sqrt{n})/10$ and $(n + 3\sqrt{n})/10$ times (inclusive) if (i) $n = 25$, and (ii) $n = 100$.

12. Let $X_1, X_2, \dots, X_n$ be iid RVs with $EX_1 = 0$ and $EX_1^2 = \sigma^2 < \infty$. Let $\overline{X} = \sum_{i=1}^n X_i/n$, and for any positive real number ε, let $P_{n,\varepsilon} = P\{\overline{X} \geq \varepsilon\}$. Show that

$$P_{n,\varepsilon} \approx \frac{\sigma}{\varepsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-n\varepsilon^2/2\sigma^2} \qquad \text{as } n \to \infty.$$

[*Hint:* Use (5.3.61).]