ity coincides with that of F, it is continuous from the right.[7] Finally, we define

$$EY = \int_{-\infty}^{\infty} h(x)\, dF(x) = \lim_{\substack{a \to -\infty \\ b \to \infty}} \int_a^b h(x)\, dF(x), \tag{3.1.2}$$

provided the limit (which may be $+\infty$ or $-\infty$) exists regardless of the way $a \to -\infty$ and $b \to \infty$.

If dF/dx exists and is equal to $f(x)$, $F(x_{i+1}) - F(x_i) = f(x_i^*)(x_{i+1} - x_i)$ for some $x_i^* \in [x_{i+1}, x_i]$ by the mean value theorem. Therefore

$$Eh(X) = \int_{-\infty}^{\infty} h(x)f(x)\, dx. \tag{3.1.3}$$

On the other hand, suppose $X = c_i$ with probability p_i, $i = 1, 2, \ldots, K$. Take $a < c_1$ and $c_K < b$; then, for sufficiently large n, each interval contains at most one of the c_i's. Then, of the n terms in the summand of (3.1.1), only K terms containing c_i's are nonzero. Therefore

$$\int_a^b h(x)\, dF(x) = \sum_{i=1}^{K} h(c_i)p_i. \tag{3.1.4}$$

3.2 Various Modes of Convergence

In this section, we shall define four modes of convergence for a sequence of random variables and shall state relationships among them in the form of several theorems.

DEFINITION 3.2.1 (convergence in probability). A sequence of random variables $\{X_n\}$ is said to converge to a random variable X *in probability* if $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$ for any $\epsilon > 0$. We write $X_n \xrightarrow{\text{P}} X$ or plim $X_n = X$.

DEFINITION 3.2.2 (convergence in mean square). *A* sequence $\{X_n\}$ is said to converge to X *in mean square* if $\lim_{n\to\infty} E(X_n - X)^2 = 0$. We write $X_n \xrightarrow{\text{M}} X$.

DEFINITION 3.2.3 (convergence in distribution). *A* sequence $\{X_n\}$ is said to converge to X *in distribution* if the distribution function F_n of X_n converges to the distribution function F of X at every continuity point of F. We write $X_n \xrightarrow{\text{d}} X$, and we call F the *limit distribution* of $\{X_n\}$. If $\{X_n\}$ and $\{Y_n\}$ have the same limit distribution, we write $X_n \overset{\text{LD}}{=} Y_n$.

The reason for adding the phrase "at every continuity point of F" can be understood by considering the following example: Consider the sequence

$F_n(\cdot)$ such that

$$\begin{aligned} F_n(x) &= 0, \qquad x < \alpha - \frac{1}{n} \qquad\qquad (3.2.1) \\ &= \frac{n}{2}\left(x - \alpha + \frac{1}{n}\right), \qquad \alpha - \frac{1}{n} \leqq x \leqq \alpha + \frac{1}{n} \\ &= 1, \qquad \alpha + \frac{1}{n} < x. \end{aligned}$$

Then $\lim F_n$ is not continuous from the left at α and therefore is not a distribution function. However, we would like to say that the random variable with the distribution (3.2.1) converges in distribution to a degenerate random variable which takes the value α with probability one. The phrase "at every continuity point of F" enables us to do so.

DEFINITION 3.2.4 (almost sure convergence). A sequence $\{X_n\}$ is said to converge to X *almost surely*[8] if

$$P\{\omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1.$$

We write $X_n \xrightarrow{\text{a.s.}} X$.

The next four theorems establish the logical relationships among the four modes of convergence, depicted in Figure 3.2.[9]

THEOREM 3.2.1 (Chebyshev). $EX_n^2 \to 0 \Rightarrow X_n \xrightarrow{\text{P}} 0$.

Proof. We have

$$EX_n^2 = \int_{-\infty}^{\infty} x^2\, dF_n(x) \geqq \epsilon^2 \int_S dF_n(x), \qquad (3.2.2)$$

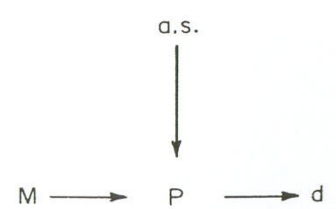

Figure 3.2 Logical relationships among four modes of convergence

where $S = \{x \mid x^2 \geqq \epsilon^2\}$. But we have

$$\begin{aligned} \int_S dF_n(x) &= \int_{-\infty}^{-\epsilon} dF_n(x) + \int_{\epsilon}^{\infty} dF_n(x) \qquad (3.2.3) \\ &= F_n(-\epsilon) + [1 - F_n(\epsilon)] \\ &= P(X_n < -\epsilon) + P(X_n \geqq \epsilon) \\ &\geqq P[X_n^2 > \epsilon^2]. \end{aligned}$$

Therefore, from (3.2.2) and (3.2.3), we obtain

$$P[X_n^2 > \epsilon^2] \leqq \frac{EX_n^2}{\epsilon^2}. \qquad (3.2.4)$$

The theorem immediately follows from (3.2.4).

The inequality (3.2.4) is called *Chebyshev's inequality*. By slightly modifying the proof, we can establish the following generalized form of Chebyshev's inequality:

$$P[g(X_n) > \epsilon^2] \leqq \frac{Eg(X_n)}{\epsilon^2}, \qquad (3.2.5)$$

where $g(\cdot)$ is any nonnegative continuous function.

Note that the statement $X_n \xrightarrow{\text{M}} X \Rightarrow X_n \xrightarrow{\text{P}} X$, where X may be either a constant or a random variable, follows from Theorem 3.2.1 if we regard $X_n - X$ as the X_n of the theorem.

We shall state the next two theorems without proof. The proof of Theorem 3.2.2 can be found in Mann and Wald (1943) or Rao (1973, p. 122). The proof of Theorem 3.2.3 is left as an exercise.

THEOREM 3.2.2. $X_n \xrightarrow{\text{P}} X \Rightarrow X_n \xrightarrow{\text{d}} X$.

THEOREM 3.2.3. $X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{\text{P}} X$.

The converse of Theorem 3.2.2 is not generally true, but it holds in the special case where X is equal to a constant α. We shall state it as a theorem, the proof of which is simple and left as an exercise.

THEOREM 3.2.4. $X_n \xrightarrow{\text{d}} \alpha \Rightarrow X_n \xrightarrow{\text{P}} \alpha$.

The converse of Theorem 3.2.3 does not hold either, as we shall show by a well-known example. Define a probability space $(\Omega, \mathcal{A}, P)$ as follows: $\Omega = [0, 1]$, $\mathcal{A}$ = Lebesgue-measurable sets in $[0, 1]$, and P = Lebesgue measure as in Example 3.1.2. Define a sequence of random variables $X_n(\omega)$ as

$$
\begin{aligned}
X_1(\omega) &= 1 && \text{for } 0 \leqq \omega \leqq 1 \\
X_2(\omega) &= 1 && \text{for } 0 \leqq \omega \leqq \frac{1}{2} \\
&= 0 && \text{elsewhere} \\
X_3(\omega) &= 1 && \text{for } \frac{1}{2} \leqq \omega \leqq \frac{1}{2} + \frac{1}{3} \\
&= 0 && \text{elsewhere} \\
X_4(\omega) &= 1 && \text{for } 0 \leqq \omega \leqq \frac{1}{12} \text{ and } \frac{1}{2} + \frac{1}{3} \leqq \omega \leqq 1 \\
&= 0 && \text{elsewhere} \\
X_5(\omega) &= 1 && \text{for } \frac{1}{12} \leqq \omega \leqq \frac{1}{12} + \frac{1}{5} \\
&= 0 && \text{elsewhere} \\
&\vdots
\end{aligned}
$$

In other words, the subset of Ω over which X_n assumes unity has the total length $1/n$ and keeps moving to the right until it reaches the right end point of [0, 1], at which point it moves back to 0 and starts again. For any $1 > \epsilon > 0$, we clearly have

$$P(|X_n| > \epsilon) = \frac{1}{n}$$

and therefore $X_n \xrightarrow{\text{P}} 0$. However, because $\Sigma_{i=1}^{\infty} i^{-1} = \infty$, there is no element in Ω for which $\lim_{n\to\infty} X_n(\omega) = 0$. Therefore $P\{\omega | \lim_{n\to\infty} X_n(\omega) = 0\} = 0$, implying that X_n does not converge to 0 almost surely.

The next three convergence theorems are extremely useful in obtaining the asymptotic properties of estimators.

THEOREM 3.2.5 (Mann and Wald). Let $\mathbf{X}_n$ and $\mathbf{X}$ be K-vectors of random variables and let $g(\cdot)$ be a function from R^K to R such that the set E of discontinuity points of $g(\cdot)$ is closed and $P(\mathbf{X} \in E) = 0$. If $\mathbf{X}_n \xrightarrow{\text{d}} \mathbf{X}$, then $g(\mathbf{X}_n) \xrightarrow{\text{d}} g(\mathbf{X})$.

A slightly more general theorem, in which a continuous function is replaced by a Borel measurable function, was proved by Mann and Wald (1943). The convergence in distribution of the individual elements of the vector $\mathbf{X}_n$ to the corresponding elements of the vector $\mathbf{X}$ is not sufficient for obtaining the above results. However, if the elements of $\mathbf{X}_n$ are independent for every n, the separate convergence is sufficient.

THEOREM 3.2.6. Let $\mathbf{X}_n$ be a vector of random variables with a fixed finite number of elements. Let g be a real-valued function continuous at a constant vector point $\boldsymbol{\alpha}$. Then $\mathbf{X}_n \xrightarrow{\text{P}} \boldsymbol{\alpha} \Rightarrow g(\mathbf{X}_n) \xrightarrow{\text{P}} g(\boldsymbol{\alpha})$.

Proof. Continuity at $\boldsymbol{\alpha}$ means that for any $\epsilon > 0$ we can find δ such that $\|\mathbf{X}_n - \boldsymbol{\alpha}\| < \delta$ implies $|g(\mathbf{X}_n) - g(\boldsymbol{\alpha})| < \epsilon$. Therefore

$$P[\|\mathbf{X}_n - \boldsymbol{\alpha}\| < \delta] \leqq P[|g(\mathbf{X}_n) - g(\boldsymbol{\alpha})| < \epsilon]. \tag{3.2.6}$$

The theorem follows because the left-hand side of (3.2.6) converges to 1 by the assumption of the theorem.

THEOREM 3.2.7 (Slutsky). If $X_n \xrightarrow{\text{d}} X$ and $Y_n \xrightarrow{\text{P}} \alpha$, then

(i) $X_n + Y_n \xrightarrow{\text{d}} X + \alpha$,

(ii) $X_n Y_n \xrightarrow{\text{d}} \alpha X$,

(iii) $(X_n/Y_n) \xrightarrow{\text{d}} X/\alpha$, provided $\alpha \neq 0$.

The proof has been given by Rao (1973, p. 122). By repeated applications of Theorem 3.2.7, we can prove the more general theorem that if g is a rational function and $\operatorname{plim} Y_{in} = \alpha_i$, $i = 1, 2, \ldots, J$, and $X_{in} \xrightarrow{\text{d}} X_i$ jointly in all $i = 1, 2, \ldots, K$, then the limit distribution of $g(X_{1n}, X_{2n}, \ldots, X_{Kn}, Y_{1n}, Y_{2n}, \ldots, Y_{Jn})$ is the same as the distribution of $g(X_1, X_2, \ldots, X_K, \alpha_1, \alpha_2, \ldots, \alpha_J)$. By using Theorem 3.2.2, this result can also be obtained from Theorem 3.2.5.

The following definition concerning the stochastic order relationship is useful (see Mann and Wald, 1943, for more details).

DEFINITION 3.2.5. Let $\{X_n\}$ be a sequence of random variables and let $\{a_n\}$ be a sequence of positive constants. Then we can write $X_n = o(a_n)$ if $\operatorname{plim}_{n\to\infty} a_n^{-1} X_n = 0$ and $X_n = O(a_n)$ if for any $\epsilon > 0$ there exists an M_ϵ such that

$$P[a_n^{-1}|X_n| \leqq M_\epsilon] \geqq 1 - \epsilon$$

for all values of n.

Sometimes these order relationships are denoted o_p and O_p respectively to distinguish them from the cases where $\{X_n\}$ are nonstochastic. However, we