Part III

Asymptotic Theory

1. Introduction

1. Modes of Convergence

In this section, we will study various modes of convergence for a sequence $\{X_n\}$ of random variables.

Definition 1 (a.s. convergence) *Let X_n be defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We say that $\{X_n\}$ converges almost surely to X if*

$$\mathbf{P}\left\{\omega \mid X_n(\omega) \to X(\omega)\right\} \;=\; 1$$

and write $X_n \to_{a.s.} X$.

Remark We may equivalently formulate the a.s. convergence as

$$\begin{aligned}
\mathbf{P}\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon \;\; i.o.\} &= \mathbf{P} \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{\omega : |X_k(\omega) - X(\omega)| > \varepsilon\} \\
&= \mathbf{P} \limsup_{n \to \infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} \\
&= 0
\end{aligned}$$

or similarly as

$$\begin{aligned}
\mathbf{P}\{\omega : |X_n(\omega) - X(\omega)| < \varepsilon \;\; ev.\} &= \mathbf{P} \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \{\omega : |X_k(\omega) - X(\omega)| < \varepsilon\} \\
&= \mathbf{P} \liminf_{n \to \infty} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\} \\
&= 1
\end{aligned}$$

for any $\varepsilon > 0$.

Definition 2 (convergence in probability) *Let X_n be defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We say that $\{X_n\}$ converges in probability to X if for any $\varepsilon > 0$*

$$\mathbf{P}\left\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\right\} \;\to\; 0$$

and denote by $X_n \to_p X$.

Definition 3 (convergence in L^p) *Let X_n be defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We say that $\{X_n\}$ converges in L^p to X if*

$$\mathbf{E}\,|X_n - X|^p \;\to\; 0$$

and write $X_n \to_{L^p} X$.

Remark Most commonly considered is the case $p = 2$, i.e. L^2-convergence, which we often refer to as the *mean squared error* convergence.

For a random variable X defined on $(\Omega, \mathcal{F}, \mathbf{P})$, we define the *distribution* of X to be the probability measure given by

$$P_X(B) = \mathbf{P} \circ X^{-1}(B)$$

for any Borel set $B \in \mathcal{B}(\mathbf{R})$, the Borel σ-field on $\mathbf{R}$.

Definition 4 (convergence in distribution) *We say that $\{X_n\}$ converges in distribution to X if*

$$\mathbf{E}f(X_n) \;\to\; \mathbf{E}f(X)$$

for every function f that is bounded and continuous a.s. in P_X, and write in symbols $X_n \to_d X$.

Remarks

(a) In the above definition, the function f need not be continuous at every point. It suffices that f is continuous with P_X probability 1, i.e., f may be discontinuous on a set N such that $P_X(N) = 0$.

(b) For the convergence in distribution, $\{X_n\}$ need not be defined on a common probability space. It is not a convergence of $\{X_n\}$, but that of probability measures $\{P_n\}$ induced by $\{X_n\}$. We may regard it as a convergence in the set of probability measures with some weak topology. For this reason, it is often referred to as the *weak convergence.*

(c) Let P and Q be two probability measures on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Then $P = Q$ if and only if $P(B) = Q(B)$ for all $B \in \mathcal{B}(\mathbf{R})$. We may show that the condition holds if and only if $\int f\, dP = \int f\, dQ$ for every bounded and continuous f. Our definition of convergence in distribution may similarly be motivated.

In what follows, we let F_n (and F) and φ_n (and φ), respectively, be the distribution and characteristic functions of X_n (and X).

Lemma 1 *The following are equivalent:*

(a) $X_n \to_d X$.

(b) $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$ *for every bounded and uniformly continuous* f.

(c) $F_n(t) \to F(t)$ *for every continuity point* t *of* F.

(d) $\varphi_n(t) \to \varphi(t)$ *for all* t.

Proof It is trivial that (a) $\Rightarrow$ (b). To see that (a) $\Rightarrow$ (c), consider

$$f_t(x) = \mathrm{I}\{x \leq t\}$$

Clearly, f_t is bounded, and continuous a.s. in P_X whenever t is a continuity point of F. It thus follows that $F_n(t) = \mathbf{E}f_t(X_n)$ converges to $F(t) = \mathbf{E}f_t(X)$ for such point t. For the implication (a) $\Rightarrow$ (d), look at the class of functions

$$f_t(x) = e^{itx}$$

For every t, the function f_t is bounded and continuous. Therefore, $\varphi_n(t) = \mathbf{E}f_t(X_n)$ converges to $\varphi(t) = \mathbf{E}f_t(X)$. The proofs for other implications are more involved and omitted. ■

Definition 5 (convergence of moments) *We say that* $\{X_n\}$ *converges to* X *in the* p*-th moment if*

$$\mathbf{E}\, X_n^p \;\to\; \mathbf{E}\, X^p$$

Remark Let $\hat{\theta}_n$ be an estimator of the parameter θ. For $\hat{\theta}_n$ to be a good estimator, it must be *consistent* and *asymptotically normal*, if properly standardized. The estimator

$\hat{\theta}_n$ is said to be consistent if it converges to θ. It is called *strongly* or *weakly* consistent, depending upon whether the mode of convergence is almost sure or in probability. The asymptotic normality implies

$$\sqrt{n}(\hat{\theta}_n - \theta) \;\to_d\; \mathbf{N}(0, \Sigma)$$

for some covariance matrix Σ.

2. Relationships among Modes of Convergence

In below we summarize the relationships existing among various modes of convergence introduced in the previous section.

Theorem 2

$$\begin{array}{ccccc} \to_{a.s.} & \text{(a)} & & \text{(c)} & \\ & \Longrightarrow & \to_p & \Longrightarrow & \to_d \\ \to_{L^p} & \text{(b)} & & & \end{array}$$

Proof To show (a), notice that

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}\{|X_n - X| > \varepsilon\} &\le \lim_{n\to\infty} \mathbf{P} \bigcup_{k\ge n} \{|X_k - X| > \varepsilon\} \\ &= \mathbf{P} \bigcap_{n=1}^{\infty} \bigcup_{k\ge n} \{|X_k - X| > \varepsilon\} \end{aligned}$$

Part (b) follows directly from the Chebyshev inequality

$$\mathbf{P}\{|X_n - X| > \varepsilon\} \le \frac{\mathbf{E}|X_n - X|^p}{\varepsilon^p}$$

For (c), pick an arbitrary f that is bounded and uniformly continuous. Let $M = \sup|f(x)|$, and for any $\varepsilon > 0$ choose δ such that

$$|X_n - X| \le \delta \quad \text{implies} \quad |f(X_n) - f(X)| \le \varepsilon$$

We have

$$|f(X_n) - f(X)| \le \varepsilon + 2M\,\mathrm{I}\{|X_n - X| > \delta\}$$

Then it follows that

$$\begin{aligned} |\mathbf{E}f(X_n) - \mathbf{E}f(X)| &\leq \mathbf{E}|f(X_n) - f(X)| \\ &\leq \varepsilon + 2M\,\mathbf{P}\{|X_n - X| > \delta\} \end{aligned}$$

from which the stated implication is immediate. ■

Other implications do not hold, as we will see in the following counterexamples.

Counterexamples Consider a probability space $([0,1], \mathcal{B}[0,1], \lambda)$, where λ is the Lebesgue measure and $\mathcal{B}[0,1]$ is the Borel σ-field on $[0,1]$. Define a sequence $\{X_n\}$ of random variables by

$$X_n(\omega) = n^{\frac{1}{p}}\,\mathrm{I}\left\{0 \leq \omega \leq \frac{1}{n}\right\}$$

for $p > 0$, and a sequence $\{Y_n\}$ by

$$Y_n = \mathrm{I}\left\{\frac{b-1}{a} \leq \omega \leq \frac{b}{a}\right\}$$

for $n = \dfrac{a(a-1)}{2} + b$ with $a = 1, 2, \ldots$ and $1 \leq b \leq a$.

It is not difficult to see that

$$X_n \to_{a.s.} 0 \quad \text{but not} \quad X_n \to_{L^p} 0$$

We indeed have $X_n(\omega) \to 0$ for all $\omega \in (0,1]$, but $\mathbf{E}X_n^p = 1$ for all n. On the contrary,

$$Y_n \to_{L^p} 0 \quad \text{but not} \quad Y_n \to_{a.s.} 0$$

Clearly, $\mathbf{E}Y_n^p = 1/a \to 0$, but $Y_n(\omega)$ does not converge for any $\omega \in [0,1]$.

Remarks

(a) It is also obvious from the above counter examples that $\to_p$ does not imply $\to_{a.s.}$ nor $\to_{L^p}$. If $\to_p \Rightarrow \to_{a.s.}$, for instance, it falsely follows that $\to_{L^p} \Rightarrow \to_p \Rightarrow \to_{a.s.}$, i.e., $\to_{L^p} \Rightarrow \to_{a.s.}$.

(b) It is clearly untrue that $\to_d \Rightarrow \to_p$ because the former does not even require that $\{X_n\}$ be defined on a common probability space. If, however, they are defined on a common probability space, then the implication does hold, as laid out in Theorem 3 (b) below.

Theorem 3

(a) *If* $X_n \to_p X$*, then there exists a subsequence* $\{X_{n_k}\}$ *such that* $X_{n_k} \to_{a.s.} X$*.*

(b) *Let* $\{X_n\}$ *be defined on a common probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, and* c *be a constant. If* $X_n \to_d c$*, then* $X_n \to_p c$*.*

Proof For (a), let $X_n \to_p X$. Since

$$\mathbf{P}\left\{|X_n - X| > \frac{1}{2^k}\right\} \;\to\; 0$$

for any k, we may choose n_k for each k such that

$$\mathbf{P}\left\{|X_{n_k} - X| > \frac{1}{2^k}\right\} \;\leq\; \frac{1}{2^k}$$

Now notice that

$$\sum_{k=1}^{\infty} \mathbf{P}\left\{|X_{n_k} - X| > \frac{1}{2^k}\right\} \;\leq\; \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty$$

Then it follows from Borel-Cantelli lemma that

$$\mathbf{P} \limsup_{k\to\infty} \left\{|X_{n_k} - X| > \frac{1}{2^k}\right\} = 0$$

i.e.

$$\mathbf{P}\left\{|X_{n_k} - X| > \frac{1}{2^k} \; i.o.\right\} = 0$$

as was to be shown.

To show part (b), let

$$f_c(x) = \mathrm{I}\{|x - c| > \varepsilon\}$$

for a given $\varepsilon > 0$. If $X_n \to_d c$, then $\mathbf{E} f_c(X_n) = \mathbf{P}\{|X_n - c| > \varepsilon\}$ converges to $\mathbf{E} f_c(c) = 0$, since f_c is continuous at c. The stated result thus follows easily. ■

3. Basics of Asymptotic Analysis

We introduce in this section some basic tools for asymptotic analysis.

Theorem 4 *Let f be a continuous function. Then we have*

(a) *If $X_n \to_{a.s.} X$, then $f(X_n) \to_{a.s.} f(X)$.*

(b) *If $X_n \to_p X$, then $f(X_n) \to_p f(X)$.*

(c) *If $X_n \to_d X$, then $f(X_n) \to_d f(X)$.*

Part (c), in particular, is called the *continuous mapping theorem*(CMT).

Proof Part (a) is obvious, since due to the continuity of f

$$X_n(\omega) \to X(\omega) \quad \text{implies} \quad f(X_n(\omega)) \to f(X(\omega))$$

for all ω.

For (b), let $X_n \to_p X$ and $f(X_{n_k})$ be an arbitrary subsequence of $f(X_n)$. Now it follows that

$$\begin{aligned} X_n \to_p X \quad &\Longrightarrow \quad X_{n_k} \to_p X \\ &\Longrightarrow \quad \exists \{X_{n_{k_i}}\} \text{ such that } X_{n_{k_i}} \to_{a.s.} X \\ &\Longrightarrow \quad f(X_{n_{k_i}}) \to_{a.s.} f(X) \\ &\Longrightarrow \quad f(X_{n_{k_i}}) \to_p f(X) \end{aligned}$$

We have thus shown that any subsequence $f(X_{n_k})$ of $f(X_n)$ has a subsequece $f(X_{n_{k_i}})$ such that $f(X_{n_{k_i}}) \to_p f(X)$, which implies that $f(X_n) \to_p f(X)$.

For part (c), it suffices to show that

$$\mathbf{E}g(f(X_n)) \to \mathbf{E}g(f(X))$$

for $X_n \to_d X$ and g continuous and bounded. This however follows directly from the definition of the convergence in distribution of X_n to X, since $g \circ f$ is continuous and bounded. ■

In what follows, let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. We define

Definition 6 *We write $x_n = o(a_n)$ and $y_n = O(b_n)$, respectively, when*

$$\frac{x_n}{a_n} \to 0 \quad \text{and} \quad \left|\frac{y_n}{b_n}\right| < M$$

for $M > 0$.

Remarks

(a) In particular, $x_n = o(1)$ if the sequence $\{x_n\}$ converges to zero, and $y_n = O(1)$ if the sequence $\{y_n\}$ is bounded.

(b) We may write

$$o(a_n) = a_n o(1) \quad \text{and} \quad O(b_n) = b_n O(1)$$

in general.

(c) The equality containing o and O is not really an equality. For instance, $o(1) = O(1)$, but $O(1) \neq o(1)$.

(d) For $y_n = O(1)$, it suffices to have $|y_n| < M$ for large n. If $|y_n| < M$ for all $n > N$, say, then it follows that $|y_n| < M_*$ for all n with $M_* = \max\{y_1, \ldots, y_N, M\}$.

Lemma 5

(a) $O(o(1)) = o(1)$

(b) $o(O(1)) = o(1)$

(c) $o(1)O(1) = o(1)$

Proof For part (a), let $x_n = o(1)$ and $y_n = O(x_n)$. With M such that $|y_n/x_n| < M$, we have

$$|y_n| < M|x_n| \to 0$$

which shows that $y_n \to 0$, as required.

For (b), assume $x_n = O(1)$ and $y_n = o(x_n)$. Choose M such that $|x_n| < M$. Then it follows that

$$\frac{|y_n|}{M} < \left|\frac{y_n}{x_n}\right| \to 0$$

as was to be shown.

To prove (c), let $x_n = o(1)$ and $y_n = O(1)$. Then, for M such that $|y_n| < M$,

$$|x_n y_n| < |x_n|M \to 0$$

which yields the stated result. ■

Remark In general, we have

$$O(o(a_n)) = O(a_n o(1)) = a_n O(o(1)) = a_n o(1) = o(a_n)$$

for the result comparable to part (a) of Lemma 5.

Definition 7 *We write $X_n = o_p(a_n)$ if X_n/a_n converges in probability to zero. Moreover, $Y_n = O_p(b_n)$ if for any $\varepsilon > 0$, there exists $M > 0$ such that*

$$\mathbf{P}\left\{\left|\frac{Y_n}{b_n}\right| > M\right\} < \varepsilon$$

Remarks

(a) In particular, $X_n = o_p(1)$ if $X_n \to_p 0$. When $Y_n = O_p(1)$, there exists M such that $\mathbf{P}\{|Y_n| > M\} < \varepsilon$ for any $\varepsilon > 0$. In this case, we say that Y_n is stochastically bounded.

(b) The remarks (b) - (d) for o and O also hold for o_p and O_p, with obvious modifications.

Lemma 6

(a) $O(o_p(1)) = O_p(o(1)) = O_p(o_p(1)) = o_p(1)$.

(b) $o(O_p(1)) = o_p(O(1)) = o_p(O_p(1)) = o_p(1)$.

(c) $o(1)O_p(1) = o_p(1)O(1) = o_p(1)O_p(1) = o_p(1)$.

Theorem 7 *Let $X_n \to_d X$. Then we have*

(a) $X_n = O_p(1)$.

(b) $X_n + o_p(1) \;\to_d\; X$.

Proof For part (a), notice that we may choose sufficiently large M such that

$$\mathbf{P}\{|X| > M\} < \varepsilon \quad \text{and} \quad \mathbf{P}\{|X| = M\} = 0$$

since $\{|X| > M\} \downarrow \emptyset$ as $M \uparrow \infty$. Let $f(x) = \mathrm{I}\{|x| > M\}$. Since $X_n \to_d X$ and f bounded and continuous a.s., we have $\mathbf{E}f(X_n) = \mathbf{P}\{|X_n| > M\}$ converges to $\mathbf{E}f(X) = \mathbf{P}\{|X| > M\}$. Therefore, $\mathbf{P}\{|X_n| > M\} < \varepsilon$ for large n.

For (b), let $Y_n = o_p(1)$ and assume that f is uniformly continuous and bounded. It suffices to show that

$$\begin{aligned} |\mathbf{E}f(X_n + Y_n) - \mathbf{E}f(X)| &\leq \mathbf{E}\,|f(X_n + Y_n) - f(X_n)| + |\mathbf{E}f(X_n) - \mathbf{E}f(X)| \\ &\to 0 \end{aligned}$$

The second term can be made arbitrarily small, since $X_n \to_d X$. To show that the first term is also negligible, we note that

$$|f(X_n + Y_n) - f(X_n)| \leq \varepsilon + 2M\,\mathrm{I}\{|Y_n| > \delta\}$$

where $M = \sup|f(x)|$ and ε and δ are chosen as in the proof of Theorem 2. The stated result now follows from $\mathbf{P}\{|Y_n| > \delta\} \to 0$. ■

Corollary 8 *If $X_n \to_d X$ and $Y_n \to_p c$, then $X_n Y_n \to_d cX$.*

Proof Since $X_n = O_p(1)$ and $Y_n = c + o_p(1)$, we have

$$\begin{aligned} X_n Y_n &= X_n(c + o_p(1)) \\ &= cX_n + O_p(1)o_p(1) \\ &= cX_n + o_p(1) \end{aligned}$$

Apply CMT to get the stated result. ■

Remark Let $\hat{\theta}_n$ be an estimator of the parameter θ with the true value θ_0. If $\hat{\theta}_n$ is consistent, then

$$\hat{\theta}_n = \theta_0 + o_p(1)$$

If it is asymptotically normal, then

$$\hat{\theta}_n = \theta_0 + O_p\left(\frac{1}{\sqrt{n}}\right)$$

in particular.

Now we introduce so-called Δ-*method*, which is useful in obtaining asymptotic distribution of nonlinear functions of $\hat{\theta}_n$ whose distribution is asymptotically normal. Suppose

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \ \to_d \ \mathbf{N}(0, \Sigma)$$

and f is continuous and differentiable at θ_0 with $F(\theta) = \partial f(\theta)/\partial \theta'$. It follows from the Taylor expansion of $f(\theta)$ around θ_0 that

$$\begin{aligned}
f(\hat{\theta}_n) &= f(\theta_0) + F(\theta_0)(\hat{\theta}_n - \theta_0) + o(\|\hat{\theta}_n - \theta_0\|) \\
&= f(\theta_0) + F(\theta_0)(\hat{\theta}_n - \theta_0) + o\left(O_p\left(\frac{1}{\sqrt{n}}\right)\right) \\
&= f(\theta_0) + F(\theta_0)(\hat{\theta}_n - \theta_0) + o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

which implies

$$\begin{aligned}
\sqrt{n}(f(\hat{\theta}_n) - f(\theta_0)) &= F(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \\
&\to_d \ \mathbf{N}(0, F(\theta_0)\Sigma F(\theta_0)')
\end{aligned}$$

If, for instance, we let $\theta = (\alpha, \beta)'$ and $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)'$, then the limiting distribution of $\hat{\alpha}_n/\hat{\beta}_n$ is given as above with

$$F(\theta) = \left(\frac{\partial f}{\partial \alpha}, \frac{\partial f}{\partial \beta}\right) = \left(\frac{1}{\beta}, -\frac{\alpha}{\beta^2}\right)$$