Modes of Convergence

Statistics · Module 7 — Asymptotic Theory

Node 7.1

Position in Knowledge Graph

System: Statistics. Structural Domain: Module 7 — Asymptotic Theory. Node 7.1 — Modes of Convergence.

Modes of convergence form the foundational language for all large-sample results in statistics. Every asymptotic statement — consistency, asymptotic normality, convergence of test statistics — is expressed through one of the convergence modes defined here. This node depends on the measure-theoretic probability framework (Node 1.8) and moment theory (Node 1.9), and enables the consistency (Node 7.2) and asymptotic normality (Node 7.3) results that follow.

Formal Definition

Let $(T_n)_{n\geq 1}$ be a sequence of random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in $\mathbb{R}^k$.

Convergence in probability. $T_n \xrightarrow{p} T$ if for every $\varepsilon > 0$,

$$\mathbb{P}(|T_n - T| > \varepsilon) \to 0 \quad \text{as } n \to \infty.$$

Almost sure convergence. $T_n \xrightarrow{a.s.} T$ if

$$\mathbb{P}\Big(\lim_{n\to\infty} T_n = T\Big) = 1.$$

Equivalently, for every $\varepsilon > 0$, $\mathbb{P}(\sup_{m\geq n} |T_m - T| > \varepsilon) \to 0$ as $n \to \infty$.

Convergence in distribution. $T_n \xrightarrow{d} T$ if for every continuity point t of the CDF F_T,

$$F_{T_n}(t) \to F_T(t) \quad \text{as } n \to \infty.$$

Equivalently, $\mathbb{E}[f(T_n)] \to \mathbb{E}[f(T)]$ for every bounded continuous function f (the Portmanteau characterization).

L^p **convergence.** For $p \geq 1$, $T_n \xrightarrow{L^p} T$ if $\mathbb{E}[|T_n - T|^p] \to 0$. The case $p = 2$ is mean-square convergence.

Intuition

Almost sure convergence is path-level: on almost every realization ω, the sequence $T_n(\omega)$ literally converges. Convergence in probability is weaker: the probability of a large deviation vanishes, but exceptional paths may misbehave infinitely often. Convergence in distribution is weakest: it says only that the laws of T_n approach the law of T, without requiring the random variables to be defined on the same space. L^p convergence controls the average magnitude of the deviation, and its strength depends on p.

The distinction matters because different statistical guarantees require different modes. Consistency of an estimator is convergence in probability; the central limit theorem gives convergence in distribution; the strong law of large numbers gives almost sure convergence.

Structural Role

Modes of convergence are the formal language for all large-sample statements in statistics. Consistency (Node 7.2) is convergence in probability of an estimator to the true parameter. Asymptotic normality (Node 7.3) is convergence in distribution of a standardized estimator to a Gaussian. Slutsky's theorem and the continuous mapping theorem, stated below, are the primary tools for combining and transforming convergent sequences to establish distributional limits of composite statistics. The hierarchy of modes determines which conclusions transfer between settings.

Mathematical Development

Theorem (Hierarchy of modes).

$$T_n \xrightarrow{a.s.} T \implies T_n \xrightarrow{p} T \implies T_n \xrightarrow{d} T.$$

$$T_n \xrightarrow{L^p} T \implies T_n \xrightarrow{p} T.$$

Proof that a.s. implies in probability. Fix $\varepsilon > 0$. Define $A_n = \{|T_n - T| > \varepsilon\}$. Almost sure convergence means $\mathbb{P}(\limsup_n A_n) = 0$. Since $\mathbb{P}(A_n) \leq \mathbb{P}(\bigcup_{m \geq n} A_m)$ and $\mathbb{P}(\bigcup_{m \geq n} A_m) \downarrow \mathbb{P}(\limsup_n A_n) = 0$, we have $\mathbb{P}(A_n) \to 0$. $\square$

Proof that L^p implies in probability. By Markov's inequality, $\mathbb{P}(|T_n - T| > \varepsilon) \leq \varepsilon^{-p}\mathbb{E}[|T_n - T|^p] \to 0$. $\square$

Proof that in probability implies in distribution. Fix a continuity point t of F_T. For any $\delta > 0$,

$$F_{T_n}(t) = \mathbb{P}(T_n \leq t) \leq \mathbb{P}(T \leq t + \delta) + \mathbb{P}(|T_n - T| > \delta).$$

Similarly, $F_{T_n}(t) \geq \mathbb{P}(T \leq t - \delta) - \mathbb{P}(|T_n - T| > \delta)$. Taking $n \to \infty$ and then $\delta \to 0$ using continuity of F_T at t yields $F_{T_n}(t) \to F_T(t)$. $\square$

Converse: convergence in distribution to a constant. If $T_n \xrightarrow{d} c$ for a constant c, then $T_n \xrightarrow{p} c$. This follows because $F_c(t) = \mathbf{1}(t \geq c)$ has a single discontinuity, and pointwise convergence of CDFs at all continuity points forces the mass to concentrate at c.

Theorem (Continuous Mapping Theorem). If $T_n \xrightarrow{d} T$ and $g : \mathbb{R}^k \to \mathbb{R}^m$ is continuous $\mathbb{P}_T$-almost everywhere, then $g(T_n) \xrightarrow{d} g(T)$. The same holds with $\xrightarrow{p}$ and $\xrightarrow{a.s.}$ in place of $\xrightarrow{d}$.

Proof sketch for the distributional case. Let D_g be the set of discontinuity points of g. By assumption, $\mathbb{P}(T \in D_g) = 0$. For any bounded continuous f, the composition $f \circ g$ is continuous on $\mathbb{R}^k \setminus D_g$. By the Portmanteau theorem extended to functions continuous a.e., $\mathbb{E}[f(g(T_n))] \to \mathbb{E}[f(g(T))]$. $\square$

Theorem (Slutsky). If $T_n \xrightarrow{d} T$ and $S_n \xrightarrow{p} c$ for a constant c, then:

$$T_n + S_n \xrightarrow{d} T + c, \qquad T_n S_n \xrightarrow{d} cT, \qquad T_n / S_n \xrightarrow{d} T/c \;\; (c \neq 0).$$

Proof. Consider the map $g(t, s) = t + s$. The pair $(T_n, S_n) \xrightarrow{d} (T, c)$ since $S_n \xrightarrow{p} c$ (the joint convergence follows because convergence in probability to a constant makes the marginals asymptotically independent).

Applying the continuous mapping theorem to g yields $T_n + S_n \xrightarrow{d} T + c$. The product and ratio cases follow analogously. $\square$

Skorokhod representation theorem. If $T_n \xrightarrow{d} T$, there exist random variables T_n', T' on a common probability space with the same marginal distributions such that $T_n' \xrightarrow{a.s.} T'$. This converts distributional convergence to almost sure convergence on a possibly different probability space, often simplifying proofs.

Borel-Cantelli and almost sure convergence. The first Borel-Cantelli lemma states: if $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_n A_n) = 0$. Applied with $A_n = \{|T_n - T| > \varepsilon\}$: if $\sum_n \mathbb{P}(|T_n - T| > \varepsilon) < \infty$ for every $\varepsilon > 0$, then $T_n \xrightarrow{a.s.} T$. This provides a checkable sufficient condition for almost sure convergence from tail probability bounds.

Subsequence characterization. $T_n \xrightarrow{p} T$ if and only if every subsequence of (T_n) has a further subsequence that converges almost surely to T. This characterization is frequently used in proofs: to show convergence in probability, extract a subsequence via the diagonal argument and upgrade to almost sure convergence.

Dominated convergence and L^p convergence. If $T_n \xrightarrow{p} T$ and $|T_n|^p \leq Y$ a.s. for some integrable Y, then $T_n \xrightarrow{L^p} T$. This is the probabilistic version of the dominated convergence theorem, and it provides the bridge from convergence in probability back up to L^p convergence when a dominating bound exists.

Multivariate extension. All four modes extend to $\mathbb{R}^k$-valued random vectors by replacing $|T_n - T|$ with $\|T_n - T\|$ (any norm; all norms on $\mathbb{R}^k$ are equivalent). The Cramer-Wold device characterizes convergence in distribution of random vectors: $T_n \xrightarrow{d} T$ in $\mathbb{R}^k$ if and only if $a^\top T_n \xrightarrow{d} a^\top T$ for every $a \in \mathbb{R}^k$. This reduces multivariate distributional convergence to checking one-dimensional convergence along all projections.

Worked Examples

Example 1. Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Set $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. By the weak law of large numbers, $\bar{X}_n \xrightarrow{p} \mu$. By the strong law, $\bar{X}_n \xrightarrow{a.s.} \mu$. Since $\mathbb{E}[(\bar{X}_n - \mu)^2] = \sigma^2/n \to 0$, we also have $\bar{X}_n \xrightarrow{L^2} \mu$. By the CLT, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$. All four modes appear in this single example.

Example 2 (Slutsky application). Suppose $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and we estimate σ^2 by $S_n^2 = (n-1)^{-1} \sum (X_i - \bar{X}_n)^2$. Since $S_n^2 \xrightarrow{p} \sigma^2$, Slutsky gives $\sqrt{n}(\bar{X}_n - \mu)/S_n \xrightarrow{d} \mathcal{N}(0,1)$. This justifies the standard z-type confidence interval with estimated variance.

Example 3 (Continuous mapping). Let $T_n \xrightarrow{d} \mathcal{N}(0,1)$. Then $T_n^2 \xrightarrow{d} \chi_1^2$ by the continuous mapping theorem with $g(t) = t^2$. More generally, if $T_n \xrightarrow{d} \mathcal{N}(0, I_k)$, then $\|T_n\|^2 \xrightarrow{d} \chi_k^2$.

Example 4 (Convergence in distribution but not in probability). Let $X \sim \mathcal{N}(0,1)$ and define $T_n = (-1)^n X$. Then $T_n \sim \mathcal{N}(0,1)$ for every n, so $T_n \xrightarrow{d} \mathcal{N}(0,1)$. But $T_n - T_{n+1} = (-1)^n X - (-1)^{n+1} X = 2(-1)^n X$, which does not converge to 0, so T_n does not converge in probability (or almost surely). Convergence in distribution says nothing about the joint behavior of the sequence.

Example 5 (L^p convergence vs. almost sure). Let $\Omega = [0,1]$ with Lebesgue measure. Define $T_n = n \cdot \mathbf{1}_{[0,1/n]}$. Then $T_n \xrightarrow{a.s.} 0$ (for any $\omega > 0$, eventually $T_n(\omega) = 0$). However, $\mathbb{E}[|T_n|^p] = n^p/n = n^{p-1} \to \infty$ for $p > 1$, so $T_n \overset{L^p}{\not\to} 0$ for $p > 1$. Almost sure convergence does not imply L^p convergence without uniform integrability.

Theorem (Uniform integrability bridge). $T_n \xrightarrow{L^1} T$ if and only if $T_n \xrightarrow{p} T$ and $\{T_n\}$ is uniformly integrable: $\lim_{M \to \infty} \sup_n \mathbb{E}[|T_n| \mathbf{1}_{|T_n| > M}] = 0$. This provides the missing link between convergence in probability and L^p convergence.

Cross-Links

AI. Convergence in distribution underpins the theoretical analysis of stochastic gradient descent: under appropriate step-size schedules, the iterate distribution converges to a stationary distribution, and Slutsky-type arguments justify plug-in variance estimation in online learning.

Economics. Asymptotic results in econometrics (consistency and normality of GMM estimators) are stated entirely in terms of these modes; the distinction between convergence in probability and almost sure convergence determines whether path-dependent economic dynamics can be analyzed.

Linear Algebra. Convergence in L^2 is norm convergence in the Hilbert space $L^2(\Omega, \mathbb{P})$; the hierarchy of modes mirrors the hierarchy of topologies (norm, weak, weak-$*$) on Banach spaces.

Quantum Mechanics. Convergence of density matrices in trace norm is the quantum analogue of L^1 convergence; weak convergence of quantum states corresponds to convergence in distribution of measurement outcomes.

Structural Compression

Invariant. The hierarchy a.s. $\Rightarrow$ in prob. $\Rightarrow$ in dist. is strict; each mode imposes a progressively weaker condition on the probability measure. $L^p \Rightarrow$ in prob. provides an independent chain.

Mechanism. Each mode controls a different aspect of the probability mass: almost sure convergence controls path behavior; convergence in probability controls tail probabilities; convergence in distribution controls only the law. L^p convergence controls moment-weighted deviations. The continuous mapping theorem and Slutsky's theorem transfer convergence through smooth operations.

Cross-System Echo. Convergence in distribution is weak convergence of measures; in functional analysis, this is weak-$*$ convergence of probability measures as linear functionals on $C_b(\mathbb{R})$. The Skorokhod representation theorem bridges distributional and pathwise convergence by constructing a coupling.

Exercises

1. Construct a sequence T_n that converges to 0 in probability but not almost surely. Verify both claims.
2. Let $T_n \xrightarrow{d} \mathcal{N}(0, 1)$. Prove that $|T_n| \xrightarrow{d} |Z|$ where $Z \sim \mathcal{N}(0, 1)$, and identify the distribution of $|Z|$.
3. Suppose $T_n \xrightarrow{L^2} T$. Prove that $\mathbb{E}[T_n] \to \mathbb{E}[T]$ and $\text{Var}(T_n) \to \text{Var}(T)$.
4. Let $X_1, \ldots, X_n$ be i.i.d. Uniform$(0, \theta)$. Show that $X_{(n)} = \max_i X_i$ satisfies $X_{(n)} \xrightarrow{a.s.} \theta$. Find the rate: determine the limit distribution of $n(\theta - X_{(n)})$.
5. Prove that if $T_n \xrightarrow{p} T$ and $T_n \xrightarrow{p} T'$, then $T = T'$ a.s. (uniqueness of limits in probability).
6. Let $T_n \xrightarrow{d} T$ and $S_n \xrightarrow{d} S$. Give a counterexample showing that $T_n + S_n$ need not converge in distribution to $T + S$ without additional conditions.
7. Argue, structurally, why convergence in distribution is the natural mode for the central limit theorem while convergence in probability is the natural mode for the law of large numbers, relating each to the type of statistical conclusion it supports.

Statistics · Module 7 — Asymptotic Theory · Node 7.1
Concepts: convergence, probability-space
Prerequisites: 1.8, 1.9
Enables: 7.2, 7.3