1 Lecture 10 — Hierarchy of Convergence; Borel–Cantelli; Prohorov

Lecture 9 set up the four modes of convergence for random variables and the equivalent formulations of weak convergence (Portmanteau). The job now is twofold: assemble these modes into a single hierarchy of implications, and develop the Borel–Cantelli lemmas — the standard tool for promoting summable bounds on $\mu(A_i)$ into almost-sure statements about whether A_i occurs only finitely often. We close with a brief look at Prohorov's theorem, which gives a compactness criterion for sequences of probability measures and underlies the classical proof of the central limit theorem.

1.1 Stronger metrics on the space of probability measures

Weak convergence $\mathbb{P}_i \Rightarrow \mathbb{P}$ (**??**) is the weakest of a family of distance-like notions on probability measures, all of the form $\sup_{f\in\mathcal{F}} |\int f\,d\mathbb{P}_i - \int f\,d\mathbb{P}|$ for some test class $\mathcal{F}$. Enlarging the test class $\mathcal{F}$ yields a finer notion of closeness:

- $\mathcal{F} = C_b(S)$ (continuous bounded): weak convergence.
- $\mathcal{F} = \{f : S \to [-1, 1]$ continuous$\}$: the *Radon metric*.
- $\mathcal{F} = \{f : S \to [-1, 1]$ measurable$\}$: the *total variation* distance.
- $\mathcal{F} = \{f : S \to \mathbb{R}$ Lipschitz with constant $1\}$: the *1-Wasserstein distance*, central to optimal transport and machine learning, which quantifies *how quickly* the convergence occurs rather than only that it does.

Remark 1.1. All four are equivalent on a finite S, and all imply weak convergence on any metric space; the converse fails in general. We work almost exclusively with weak convergence below.

1.2 Hierarchy of modes of convergence

Recall from Lecture 9 the four modes for random variables $X_i, X\colon (\Omega, \mathcal{F}, \mu) \to (S, \rho)$: convergence in distribution $X_i \xrightarrow{d} X$, in probability $X_i \xrightarrow{\mathbb{P}} X$ (**??**), almost surely $X_i \xrightarrow{a.s.} X$ (Result 1.5), and in L^p $X_i \xrightarrow{L^p} X$ (**??**). The implications between them form a strict hierarchy.

Theorem 1.1: Hierarchy of convergence

Let X_i, X be random variables on $(\Omega, \mathcal{F}, \mu)$ with values in a metric space (S, ρ). Then

$$X_i \xrightarrow{a.s.} X \implies X_i \xrightarrow{\mathbb{P}} X \implies X_i \xrightarrow{d} X,$$
$$X_i \xrightarrow{L^p} X \implies X_i \xrightarrow{\mathbb{P}} X \quad \text{for every } p \in [1, \infty].$$

None of the reverse implications holds, and *a.s.* and L^p convergence are not comparable: neither implies the other without an additional uniform integrability or domination hypothesis.

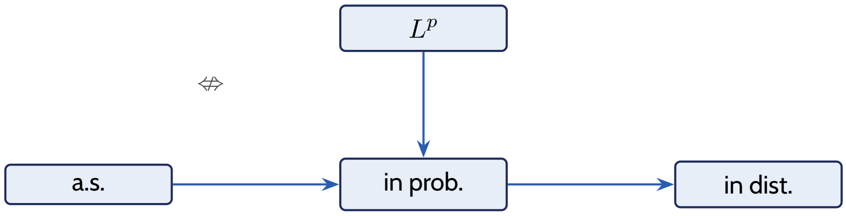

Figure 1. The four modes of convergence and the implications between them. Almost-sure and L^p convergence are not comparable.

1.3 Limsup, liminf, and the Borel–Cantelli setup

Fix a probability space $(\Omega, \mathcal{F}, \mu)$ and a sequence $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$. The set-theoretic limsup and liminf give a precise meaning to "A_i happens infinitely often" and "A_i happens eventually".

Definition 1.2: $\limsup$ and $\liminf$ of events

For $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$,

$$\limsup_i A_i = \bigcap_{i=1}^{\infty} \bigcup_{j \geq i} A_j, \qquad \liminf_i A_i = \bigcup_{i=1}^{\infty} \bigcap_{j \geq i} A_j.$$

We say A_i *infinitely often* (A_i i.o.) for $\limsup_i A_i$: $\omega \in \limsup_i A_i$ iff for every $N \in \mathbb{N}$ there exists $n \geq N$ with $\omega \in A_n$. We say A_i *eventually* (A_i ev.) for $\liminf_i A_i$: $\omega \in \liminf_i A_i$ iff there exists $N \in \mathbb{N}$ such that $\omega \in A_n$ for all $n \geq N$.

Remark 1.2. The two are dual in the sense $(\limsup_i A_i)^c = \liminf_i A_i^c$ and conversely, by De Morgan.

1.4 The Borel–Cantelli lemmas

The two lemmas are a one-sided pair: summability of $\mu(A_i)$ forces A_i to occur only finitely often almost surely; under the extra hypothesis of independence, divergence of the same series forces A_i to occur infinitely often almost surely.

Lemma 1.3: First Borel–Cantelli

Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$. If $\sum_{i=1}^{\infty} \mu(A_i) < \infty$, then

$$\mu\left(\limsup_i A_i\right) = 0.$$

Equivalently, with probability one only finitely many A_i occur. The proof is a one-line consequence of monotonicity and countable subadditivity: for every i,

$$\mu\left(\limsup_j A_j\right) \leq \mu\left(\bigcup_{j \geq i} A_j\right) \leq \sum_{j \geq i} \mu(A_j) \xrightarrow{i \to \infty} 0,$$

where the right-hand tail vanishes because the full series converges.

Lemma 1.4: Second Borel–Cantelli

Let $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ be *independent*. If $\sum_{i=1}^{\infty} \mu(A_i) = \infty$, then

$$\mu\left(\limsup_i A_i\right) = 1.$$

The argument is a complement-and-exponentiate trick. Independence of $\{A_i\}$ implies independence of $\{A_i^c\}$. For any $i \in \mathbb{N}$ and $k \geq i$,

$$\mu\left(\bigcap_{j=i}^{k} A_j^c\right) = \prod_{j=i}^{k}[1-\mu(A_j)] \leq \exp\left[-\sum_{j=i}^{k}\mu(A_j)\right],$$

using the elementary bound $1-t \leq e^{-t}$ valid for all $t \in \mathbb{R}$. Letting $k \to \infty$ makes the right-hand side vanish, so $\mu(\bigcap_{j\geq i} A_j^c) = 0$ for every i; De Morgan then gives $\mu(\limsup_i A_i) = 1$.

Remark 1.3. Independence cannot be dropped from the second lemma: if $A_1 = A_2 = \cdots = A$ with $\mu(A) = \frac{1}{2}$, then $\sum \mu(A_i) = \infty$ but $\limsup_i A_i = A$ has probability $\frac{1}{2}$, not 1.

Example 1.1 (Coin tosses produce every finite pattern). Toss a fair coin independently and let A_i be the event that positions $i, i+1, \dots, i+k-1$ spell out a fixed pattern of length k. Then $\mu(A_i) = 2^{-k}$, and the events $A_1, A_{k+1}, A_{2k+1}, \dots$ are independent with $\sum_n \mu(A_{nk+1}) = \infty$. The second Borel-Cantelli lemma gives $\mu(A_{nk+1}\text{ i.o.}) = 1$: every finite pattern appears infinitely often almost surely.

1.5 Prohorov's theorem

The final piece of the convergence apparatus is a compactness criterion for sequences of probability measures: it is to weak convergence what Bolzano–Weierstrass is to bounded sequences in $\mathbb{R}^d$. The right notion of "boundedness" is *tightness*, capturing that no mass escapes to infinity.

Definition 1.5: Uniform tightness

A collection $\{\mu_i\}_{i\in I}$ of probability measures on a metric space (S, ρ) is *uniformly tight* if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq S$ with

$$\mu_i(K_\varepsilon) > 1-\varepsilon \qquad \text{for every } i \in I.$$

Theorem 1.6: Prohorov

Let $\{\mu_i\}_{i=1}^{\infty}$ be a sequence of probability measures on a metric space S. If $\{\mu_i\}$ is uniformly tight, then it is *relatively sequentially compact* for weak convergence: every subsequence μ_{i_k} admits a further subsequence $\mu_{i_{k_r}} \Rightarrow \mu$ for some probability measure μ (depending on the subsequence).

Remark 1.4. A useful subsubsequence corollary: if every subsequence of $\{\mu_i\}$ admits a further subsequence converging weakly to the *same* limit μ, then $\mu_i \Rightarrow \mu$. This is the classical route to the central limit theorem — one shows tightness, extracts a weak limit along a subsequence, and identifies the limit as the standard normal via characteristic functions.