1 Lecture 11 — Law of Large Numbers

The Borel–Cantelli machinery of Lecture 10 finally pays off. We fix a sequence $\{X_i\}_{i=1}^{\infty}$ of random variables on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, valued in $(\mathbb{R}, \mathcal{B})$, and ask in what sense the sample averages $n^{-1}S_n = n^{-1}\sum_{i=1}^{n} X_i$ approach the common mean. Two answers — one in probability under uncorrelation plus a second moment, one almost sure under independence and only a first moment — are the content of this lecture.

1.1 Setup: independence and identical distribution

Throughout, $X\colon \Omega \to \mathbb{R}$ is a random variable with law $\mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$ for $A \in \mathcal{B}$, expectation $\mathbb{E}X = \int X(\omega)\, d\mathbb{P}$, and partial sums $S_n = \sum_{i=1}^{n} X_i$.

Definition 1.1: Independence of random variables

Random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$, valued in measurable spaces $(\mathcal{X}, \mathscr{X})$ and $(\mathcal{Y}, \mathscr{Y})$ respectively, are *independent* if

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) \;=\; \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B) \qquad \text{for all } A \in \mathscr{X},\ B \in \mathscr{Y}.$$

The definition extends to a finite collection $\{X_i\}_{i=1}^{n}$ by requiring $\mathbb{P}(\bigcap_{i=1}^{n}\{X_i \in A_i\}) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$. An infinite collection $\{X_i\}_{i=1}^{\infty}$ is independent if every finite sub-collection is.

Remark 1.1. Since $\{X \in A\} = X^{-1}(A)$, independence of the random variables X and Y is equivalent to independence of the generated σ-fields $\sigma(X)$ and $\sigma(Y)$ in the sense of **??**.

Definition 1.2: Identically distributed; i.i.d.

Random variables X and Y are *identically distributed* if the pushforward laws $\mathbb{P} \circ X^{-1}$ and $\mathbb{P} \circ Y^{-1}$ coincide on $\mathcal{B}$. A sequence $\{X_i\}_{i=1}^{\infty}$ is i.i.d. (*independent and identically distributed*) if it is independent and the X_i share a common law.

1.2 Weak law of large numbers

The weak law trades a strong moment hypothesis for a very mild dependence hypothesis: not full independence, only pairwise uncorrelation (a strictly weaker condition).

Theorem 1.3: Weak law of large numbers

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_i\}_{i=1}^{\infty}$ random variables with

$$\mathbb{E}X_i \;=\; c \in \mathbb{R}, \qquad \mathbb{E}X_i^2 \;=\; 1 \quad \text{for all } i,$$

and $\mathbb{E}\big[(X_i - c)(X_j - c)\big] = 0$ for all $i \neq j$. Then

$$\frac{S_n}{n} \;\xrightarrow{\mathbb{P}}\; c,$$

i.e. for every $\varepsilon > 0$, $\mathbb{P}(|n^{-1}S_n - c| \geq \varepsilon) \to 0$ as $n \to \infty$.

Remark 1.2. The proof reduces to $c = 0$ by replacing X_i with $X_i - c$, then applies Chebyshev: for any $t > 0$,

$$\mathbb{P}\left(\frac{|S_n|}{n} \geq t\right) \leq \frac{\mathbb{E}S_n^2}{t^2n^2} = \frac{1}{t^2n^2}\sum_{i,j=1}^{n}\mathbb{E}[X_iX_j] = \frac{1}{nt^2} \to 0,$$

where the cross terms vanish by uncorrelation and the diagonal sums to n by the unit second moment.

Remark 1.3. Uncorrelation is genuinely weaker than independence: independence of (X, Y) implies independence of $(f(X), g(Y))$ for any measurable f, g, hence $\mathrm{Cov}(f(X), g(Y)) = 0$ for every choice; uncorrelation asks this only for $f = g = \mathrm{id}$.

1.3 Strong law of large numbers

The strong law promotes "in probability" to "almost surely", removes the second moment hypothesis, but pays for it with full independence and identical distribution. Recall the variance

$$\mathrm{Var}(X) = \int (X - \mathbb{E}X)^2 \, d\mathbb{P}(\omega).$$

Theorem 1.4: Strong law of large numbers

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B})$. Then:

1. If $\mathbb{E}|X_1| < \infty$, then $\frac{S_n}{n} \xrightarrow{\text{a.s.}} c$ where $c = \mathbb{E}X_1$.
2. If $\mathbb{E}|X_1| = \infty$, then S_n/n does not converge to any finite limit (almost surely).

Remark 1.4. Compared with Result 1.3, no second-moment assumption is made on the X_i; only L^1 is needed. The trade-off is full independence (not just uncorrelation) and identical distribution. The "a.s." qualifier means convergence holds outside a $\mathbb{P}$-null set $N \subset \Omega$.

Remark 1.5. The divergence half (part 2) is the easier direction. The heuristic: if $\mathbb{E}|X_1| = \infty$ then $\sum_n \mathbb{P}(|X_n| > n) = \infty$, so by the second Borel–Cantelli lemma $|X_n| > n$ infinitely often. But on $\{n^{-1}S_n \to c\}$ one has $n^{-1}X_n = n^{-1}(S_n - S_{n-1}) \to 0$, contradicting $|X_n|/n > 1$ i.o.

Remark 1.6. The forward direction (part 1) is far more delicate. The standard route: reduce to $X_i \geq 0$ by writing $X_i = X_i^+ - X_i^-$ (independence of X, Y passes to X^+, Y^+), truncate $Y_i = X_i \mathbf{1}_{\{X_i \leq i\}}$ so that variances are finite, control the truncated partial sums $T_n = \sum_{i=1}^{n} Y_i$ along a geometric subsequence $k_n = \lfloor \delta^n \rfloor$ using Chebyshev plus the first Borel–Cantelli lemma, then sandwich the full sums S_i for $k_n \leq i \leq k_{n+1}$ and let $\delta \downarrow 1$.

■ **Example 1.1 (i.i.d. Bernoulli sample mean).** Let $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bernoulli}(p)$, so $\mathbb{E}X_i = p$ and $\mathrm{Var}(X_i) = p(1-p)$. Both moment hypotheses of the weak and strong laws are satisfied, so $n^{-1}S_n \to p$ both in probability (by Result 1.3) and almost surely (by Result 1.4). In particular, the empirical frequency of successes in n Bernoulli trials converges almost surely to the true success probability p — the formal statement behind the everyday claim “the average converges to the mean”.