1 Lecture 13 — The Ergodic Theorem

The strong law of large numbers proved in Lecture 12 says that, for i.i.d. summands, time averages $n^{-1}S_n$ converge almost surely to the expected value. Ergodic theory generalises this picture to any measure-preserving dynamical system: replace "i.i.d." by "measure-preserving" and "$\mathbb{E}X_1$" by a conditional expectation on the σ-field of invariant sets. The two foundational results are Birkhoff's pointwise theorem (almost-sure convergence) and von Neumann's mean ergodic theorem (L^p convergence). Specialising to the shift on a product space recovers the SLLN.

1.1 Measure-preserving maps, invariance, ergodicity

Throughout this section $(\Omega, \mathcal{F}, \mu)$ is a measure space and $T\colon \Omega \to \Omega$ a measurable map. We are interested in time averages along the orbit ω, $T\omega$, $T^2\omega, \dots$

Definition 1.1: Measure-preserving map

The map $T\colon \Omega \to \Omega$ is *measure preserving* if

$$\mu(T^{-1}(A)) \;=\; \mu(A), \qquad \text{for all } A \in \mathcal{F}.$$

Equivalently, the pushforward measure $\mu \circ T^{-1}$ coincides with μ: the dynamics does not distort the size of any measurable set.

Definition 1.2: Invariant set, invariant function

A set $A \in \mathcal{F}$ is T-*invariant* if $T^{-1}(A) = A$. The collection

$$\mathcal{F}_T \;=\; \{A \in \mathcal{F} : T^{-1}(A) = A\}$$

of all T-invariant sets is a σ-field. A measurable function $f\colon \Omega \to \mathbb{R}$ is *invariant* if $f = f \circ T$; equivalently, f is invariant if and only if it is $\mathcal{F}_T$-measurable.

Definition 1.3: Ergodic map

A measure-preserving map T is *ergodic* if every invariant set is trivial: for all $A \in \mathcal{F}_T$,

$$\mu(A) \;=\; 0 \quad \text{or} \quad \mu(A^c) \;=\; 0.$$

Equivalently, every T-invariant measurable function is constant μ-almost everywhere.

Example 1.1 (Shift mod 1 on the circle). On $((0,1], \mathcal{B}, \lambda)$ and a fixed $a \in (0,1]$, define the rotation

$$T(x) \;=\; x + a \mod 1 \;=\; \begin{cases} x + a & x + a \le 1, \\ x + a - 1 & x + a > 1. \end{cases}$$

T preserves Lebesgue measure: every half-open arc and its preimage have the same length. It is ergodic precisely when a is irrational.

■ Example 1.2 (Baker's map). On $(0,1]$ define $T(x) = 2x - \lfloor 2x \rfloor$. T is the doubling map; preimages of intervals split into two intervals of half the length, so Lebesgue measure is preserved. T is ergodic.

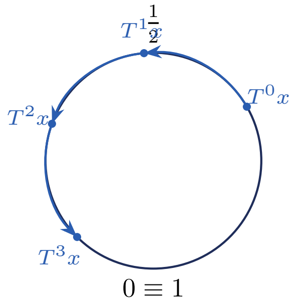

Figure 1. Orbit of a point under the rotation $T(x) = x + a \bmod 1$: for irrational a the orbit is dense, the dynamics is ergodic, and Birkhoff's theorem says time averages equal space averages.

The next two facts are the everyday tools used below; both follow directly from Results 1.1 and 1.3.

Proposition 1.4: Two basic facts

Let T be measure preserving on $(\Omega, \mathcal{F}, \mu)$.

1. If $f \in L^1(\Omega, \mathcal{F}, \mu)$ then $f \circ T \in L^1$ and
$$\int f \, d\mu \;=\; \int f \circ T \, d\mu.$$
2. If, in addition, T is ergodic and f is invariant, then $f = c$ μ-a.e. for some constant c.

1.2 Ergodic theorems

For the rest of the lecture, fix $(\Omega, \mathcal{F}, \mu)$ and a measure-preserving T. For $f \colon \Omega \to \mathbb{R}$ measurable set the *Birkhoff sums*

$$S_n \;=\; S_n(f) \;=\; f + f \circ T + f \circ T^2 + \cdots + f \circ T^{n-1}, \qquad S_0 \equiv 0.$$

Birkhoff's theorem controls the time averages $n^{-1} S_n(f)$ almost everywhere; von Neumann's controls them in L^p. Both rest on a single combinatorial estimate, the maximal ergodic lemma.

Lemma 1.5: Maximal ergodic lemma

Let $f \in L^1(\Omega, \mathcal{F}, \mu)$ and set $S^* = \sup_{n \geq 0} S_n(f)$. Then

$$\int_{\{S^* > 0\}} f \, d\mu \;\geq\; 0.$$

Theorem 1.6: Birkhoff's pointwise ergodic theorem

Let $(\Omega, \mathcal{F}, \mu)$ be σ-finite, T measure preserving, and $f \in L^1(\Omega, \mathcal{F}, \mu)$. There exists an invariant function $\bar{f} \in L^1(\Omega, \mathcal{F}, \mu)$ with

$$\int |\bar{f}|\, d\mu \;\le\; \int |f|\, d\mu \qquad \text{and} \qquad \frac{S_n(f)}{n} \;\longrightarrow\; \bar{f} \quad \mu\text{-a.e. as } n \to \infty.$$

If T is ergodic and μ is a probability, then $\bar{f} = \int f\, d\mu$ almost everywhere.

Remark 1.1. The strategy is to show that $\liminf_n n^{-1} S_n(f)$ and $\limsup_n n^{-1} S_n(f)$ are both T-invariant and equal a.e. Invariance follows from

$$n^{-1} S_n(f) \circ T \;=\; n^{-1}[S_{n+1}(f) - f] \;=\; \frac{n+1}{n} \cdot \frac{S_{n+1}(f)}{n+1} \;-\; \frac{f}{n},$$

and one isolates the bad set

$$D_{a,b} \;=\; \Big\{ \omega \in \Omega : \liminf_n n^{-1} S_n(f) < a < b < \limsup_n n^{-1} S_n(f) \Big\}$$

for rationals $a < b$. Each $D_{a,b}$ is T-invariant; an application of Result 1.5 to $g = f - b\,\mathbf{1}_B$ on a finite-measure subset $B \subseteq D_{a,b}$ yields

$$b\,\mu(D_{a,b}) \;\le\; \int_{D_{a,b}} f\, d\mu \;\le\; a\,\mu(D_{a,b}),$$

and $a < b$ forces $\mu(D_{a,b}) = 0$. Taking the countable union over rationals gives convergence in $[-\infty, \infty]$ on a full-measure set; the integrability bound $\int |\bar{f}|\, d\mu \le \int |f|\, d\mu$ falls out of Fatou's lemma applied to $n^{-1}|S_n(f)|$.

Theorem 1.7: von Neumann's mean ergodic theorem

Suppose $\mu(\Omega) < \infty$ and $p \in [1, \infty)$. For every $f \in L^p(\Omega, \mathcal{F}, \mu)$ there exists $\bar{f} \in L^p$ such that

$$\frac{S_n(f)}{n} \;\longrightarrow\; \bar{f} \qquad \text{in } L^p.$$

Remark 1.2. The argument is a three-epsilon truncation. Because T is measure-preserving, $\|f \circ T^n\|_p = \|f\|_p$, so by Minkowski $\|n^{-1} S_n(f)\|_p \le \|f\|_p$. Given $\varepsilon > 0$, choose $C > 0$ and set $g = \min\{\max\{-C, f\}, C\}$; then $\|f - g\|_p < \varepsilon/3$ and g is bounded by C, so dominated convergence upgrades the a.e. convergence $n^{-1} S_n(g) \to \bar{g}$ of **??** to L^p convergence. Fatou applied to $|n^{-1} S_n(f-g)|^p$ gives $\|\bar{f} - \bar{g}\|_p \le \|f - g\|_p$, and the triangle inequality

$$\left\| \tfrac{S_n(f)}{n} - \bar{f} \right\|_p \;\le\; \left\| \tfrac{S_n(f-g)}{n} \right\|_p + \left\| \tfrac{S_n(g)}{n} - \bar{g} \right\|_p + \|\bar{g} - \bar{f}\|_p \;<\; \varepsilon$$

finishes the proof.

1.3 Application: the strong law of large numbers, again

The two ergodic theorems give an almost free derivation of the SLLN by running the canonical i.i.d. construction through the shift map.

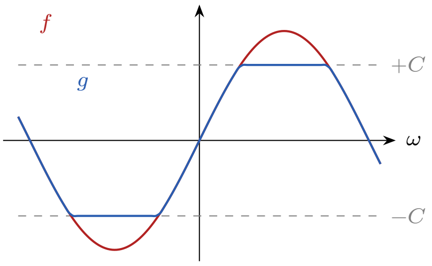

Figure 2. Truncation step in von Neumann's proof: the unbounded f (red) is clipped to a bounded $g = \min\{\max\{-C, f\}, C\}$ (blue); the tails are absorbed in $\|f - g\|_p < \varepsilon/3$, and dominated convergence handles g.

Let $(\Omega, \mathcal{F}, P)$ be a probability space carrying i.i.d. real-valued random variables $\{X_i\}_{i=1}^{\infty}$ with common distribution F. Set $(S, \mathcal{S}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{S})$ where $\mathcal{S}$ is generated by the π-system of cylinder sets

$$\mathcal{A} \;=\; \left\{\prod_{n\in\mathbb{N}} A_n \;:\; A_n \in \mathcal{B}(\mathbb{R})\ \forall n,\ A_n = \mathbb{R} \text{ eventually}\right\}.$$

The map $X\colon \Omega \to \mathbb{R}^{\mathbb{N}}$, $X(\omega) = (X_1(\omega), X_2(\omega), \dots)$, induces the product measure

$$\mu(A) \;=\; P \circ X^{-1}(A) \;=\; \prod_{n\in\mathbb{N}} dF(A_n), \qquad A = \prod A_n.$$

Definition 1.8: Shift map on $\mathbb{R}^{\mathbb{N}}$

The *shift map* $T\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ drops the first coordinate:

$$T(x_1, x_2, x_3, \dots) \;=\; (x_2, x_3, x_4, \dots).$$

Proposition 1.9: The shift is measure-preserving and ergodic

Under the i.i.d. product measure μ above, the shift map T is measure preserving and ergodic. Ergodicity follows from Kolmogorov's zero-one law: every shift-invariant cylinder event lies in the tail σ-field $\bigcap_n \sigma(X_n, X_{n+1}, \dots)$ and so has probability 0 or 1.

Theorem 1.10: Strong law of large numbers, again

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. real-valued random variables with $\mathbb{E}|X_i| < \infty$. Then

$$\frac{S_n}{n} \;=\; \frac{X_1 + \cdots + X_n}{n} \;\xrightarrow{\text{a.s.}}\; \mathbb{E}X_i.$$

Remark 1.3. Take $f\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ to be the first-coordinate projection $f(x_1, x_2, \dots) = x_1$. With T the shift, $f \circ T^k(x) = x_{k+1}$, so the Birkhoff sums recover the partial sums:

$$S_n(f) \;=\; f + f\circ T + \cdots + f \circ T^{n-1} \;=\; X_1 + \cdots + X_n.$$

?? gives an invariant $\bar{f} \in L^1$ with $n^{-1}S_n \to \bar{f}$ a.s. Since the shift is ergodic (**??**), Result 1.3 forces $\bar{f}$ to be constant a.e.; identifying that constant via **??** at $p = 1$,

$$\bar{f} \;=\; \int \bar{f}\, d\mu \;=\; \lim_{n\to\infty} \int n^{-1} S_n(f)\, d\mu \;=\; \mathbb{E}X_i,$$

which is the SLLN.