Ergodic theorems

Alex Fu

2023-01-17

This note is adapted from Chapter 6 of *Probability: Theory and Examples* by Rick Durrett, 5th edition.

We will not delve too deep into the intuition and motivations for **ergodicity** and mixing; we give just one example in ergodic Markov chains. An irreducible chain has its state space strongly connected by transitions; the state space cannot be partitioned into smaller irreducible components, or decomposed into "disjoint," non-communicating regions such that the flow of probability mass only circulates within each "closed" area. An aperiodic chain is not forced by its structure to lock in to a given pattern of cyclically alternating distributions. A positive recurrent chain does not face the issue of mass escaping off to infinity or never returning in transient chains, or the issue of mass being spread so thin as to be "infinitesimal" in null recurrent chains: it revisits its states frequently enough to generate probability mass in the long-run proportion.

A positive recurrent chain supports a stationary distribution, an "eigendistribution" invariant under any transition of the model, while an irreducible chain has at most one stationary distribution, and an aperiodic chain is free enough to eventually converge to a stationary or limiting distribution. A *regular* chain is both irreducible and aperiodic: there admits some k for which $P^k > 0$, so everything is reachable from everything else every k steps, and strictly positive transition probabilities give a maximal chance at dispersion and settling into a stable long-term distribution, as we might imagine. An *ergodic* chain is positive recurrent on top of regularity: regardless of its initial distribution, even if it starts from the most concentrated point mass measure, its dispersion will converge with probability 1 to the unique stationary distribution. The trajectory of any particle eventually stabilizes and behaves in accordance with the limiting distribution almost surely, and transitions preserve the invariant distribution.

After some long rambling, let us return to the content at hand.

Definition 1 (Stationary sequence)**.**

A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is a *stationary discrete-time random process* or a **stationary sequence** if its finite-dimensional distributions are invariant under any translations or shifts in time. That is, for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$,

$$(X_0, \ldots, X_n) \overset{\mathsf{d}}{=} (X_k, \ldots, X_{n+k}).$$

Examples of stationary sequences:

- $X_0, X_1, \ldots$ i.i.d.
- $(X_n)_{n\in\mathbb{N}}$ a Markov chain with stationary distribution and $X_0 \sim \pi$.
- Let $\Omega = [0, 1)$ with $\mathbb{P}$ the Lebesgue measure. Pick $0 < \theta < 1$ and let $X_n(\omega) = (\omega + n\theta) \bmod 1$. We may also

formulate this as a Markov chain with transition probabilities $p(x, (x+\theta) \bmod 1) = 1$. Under the inclusion $x \mapsto e^{2\pi i x}$, this process is simply rotation by θ at each time step, with $X_0 \sim$ Uniform.

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A measurable function $\varphi \colon \Omega \to \Omega$ is **measure-preserving** if $\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A)$ for all $A \in \mathcal{F}$. If X is some random variable on Ω (i.e. $\mathcal{F}$-measurable), then

$$X_n(\omega) = X(\varphi^n \omega), \quad n \in \mathbb{N}$$

forms a stationary sequence. Note that $\varphi^0 = \text{id}$. To check stationarity, let $B \subseteq \mathbb{R}^{n+1}$ be a Borel set, and let $A = \{\omega : (X_0(\omega), \ldots, X_n(\omega)) \in B\}$ be its preimage under the finite-dimensional distribution. Then

$$\mathbb{P}((X_k, \ldots, X_{n+k}) \in B) = \mathbb{P}(\varphi^k \omega \in A) = \mathbb{P}(\omega \in A) = \mathbb{P}((X_0, \ldots, X_n) \in B)$$

by the fact that φ, and by induction φ^n for any $n \in \mathbb{N}$, is measure-preserving. We suppress the parentheses for $\varphi^k \omega$ as we think of φ as a transformation *operator*, like a linear map T on a vector space.

The last example is in fact the *only* example. If $(Y_n)_{n\in\mathbb{N}}$ is any stationary sequence, taking values in a sufficiently nice space S, then by Kolmogorov's extension theorem, there exists a probability measure $\mathbb{P}$ on the sequence space $(S^{\mathbb{N}}, \Sigma^{\mathbb{N}})$ such that $X_n(\omega) = \omega_n$ has the same distribution as $(Y_n)_{n\in\mathbb{N}}$. If φ is the **shift operator**

$$\varphi(\omega_0, \omega_1, \omega_2, \ldots) \coloneqq (\omega_1, \omega_2, \ldots)$$

and $X(\omega) \coloneqq \omega_0$, then φ is measure-preserving and $X_n(\omega) = X(\varphi^n \omega)$.

From here on out, we will assume that φ is a measure-preserving transformation on Ω.

Definition 2 (Invariant event)**.**

An event $A \in \mathcal{F}$ is (φ-)**invariant** if $\varphi^{-1}(A) = A$ almost surely. We say that two events A, B are equal ($\mathbb{P}$-)*almost surely* if $\mathbb{1}_A = \mathbb{1}_B$ a.s., or equivalently $\mathbb{P}(A \triangle B) = 0$, where $A \triangle B = (A \setminus B) \sqcup (B \setminus A) = (A \cup B) \setminus (A \cap B)$ denotes the symmetric set difference.

We write $\mathcal{I}$ for the collection of φ-invariant events in $\mathcal{F}$, which we could also denote $\mathcal{F}^{\varphi}$ or $\mathcal{F}_{\varphi}$, taking inspiration from Galois theory, but we will avoid confusion with the completion $\bar{\mathcal{F}}^{\mu}$.

Proof of equivalence of definitions of almost surely equal events. If $\mathbb{1}_A = \mathbb{1}_B$ a.s., then

$$\mathbb{E}(|\mathbb{1}_A - \mathbb{1}_B|) = \mathbb{E}(\mathbb{1}_{A\setminus B}) + \mathbb{E}(\mathbb{1}_{B\setminus A}) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) = \mathbb{P}(A \triangle B) = 0.$$

Conversely, if $|\mathbb{1}_A - \mathbb{1}_B| > 0$ with some positive probability, then $\mathbb{P}(A \triangle B) = \mathbb{E}(|\mathbb{1}_A - \mathbb{1}_B|) > 0$. □

Proposition 1 (The invariant σ-algebra)**.**

The collection of invariant events $\mathcal{I}$ forms a σ-algebra.

Proof. To write it out, $\mathcal{I} = \{A \in \mathcal{F} : \varphi^{-1}A = A \text{ a.s.}\}$. Note that φ is measure-preserving: $\mathbb{P}(\varphi^{-1}A) = \mathbb{P}(A)$.

1. *Nonempty.* $\varphi^{-1}(\varnothing) = \varnothing$.
2. *Closure under complements.* If $\varphi^{-1}A = A$ a.s., then $1 - \mathbb{1}_{\varphi^{-1}A} = 1 - \mathbb{1}_A$ a.s., or $(\varphi^{-1}A)^c = A^c$ a.s.

3. *Closure under countable unions.* If $A_1, A_2, \ldots \in \mathcal{I}$, let $N_1, N_2, \ldots$ be the corresponding null sets on which $\mathbb{1}_{\varphi^{-1}A_n} \neq \mathbb{1}_{A_n}$. The countable union of null sets is null, so

$$\varphi^{-1} \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \varphi^{-1} A_n = \bigcup_{n=1}^{\infty} A_n$$

almost surely (outside of the null set $\bigcup_{n=1}^{\infty} N_n$).

□

Definition 3 (Ergodic transformation)**.**

A measure-preserving transformation φ on $(\Omega, \mathcal{F}, \mathbb{P})$ is **ergodic** if $\mathcal{I}$ is trivial, i.e. $\mathbb{P}(A) = 0$ or 1 for every $A \in \mathcal{I}$, or equivalently if $\mathcal{I}$ is independent of itself (or $\mathcal{F}$).

If φ is not ergodic, then the space Ω can be partitioned into two events A and A^c, each with positive measure, such that $\varphi(A) = A$ and $\varphi(A^c) = A^c$: that is, φ is not irreducible.

Let us check the ergodicity of our previous examples, using the fact that every measure-preserving transformation is in some sense the shift operator on the sequence space.

- If $\Omega = \mathbb{R}^{\mathbb{N}}$ and φ is the shift operator, for $A \in \mathcal{I}$, $\{\omega : \omega \in A\} = \{\omega : \varphi^n \omega \in A\} \in \sigma(X_n, X_{n+1}, \ldots)$. Thus

  $$A \in \bigcap_{n=0}^{\infty} \sigma(X_n, X_{n+1}, \ldots) = \mathcal{T},$$

  the tail σ-algebra. If $(X_n)_{n \in \mathbb{N}}$ are i.i.d., by Kolmogorov's 0-1 law, $\mathcal{T}$ is trivial, so $\mathcal{I} \subseteq \mathcal{T}$ is trivial as well, i.e. the sequence is ergodic. (That is, when we equip the corresponding measure on $\Omega = \mathbb{R}^{\mathbb{N}}$, φ is ergodic.)

- Suppose that the state space S is countable and the stationary distribution π is strictly positive. Then every state is positive recurrent, and S has a partition into closed irreducible classes R_i. If $X_0 \in R_i$, then $X_n \in R_i$ for all $n \in \mathbb{N}$ a.s., so $\{\omega : X_0(\omega) \in R_i\} \in \mathcal{I}$. Thus if this chain is reducible, it is not ergodic. (By assumption, it must start with positive mass in each R_i.)

  Conversely, let θ be the shift operator, and observe that $\mathbb{1}_A \circ \theta^n = \mathbb{1}_A$ for $A \in \mathcal{I}$. Let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. By the shift invariance of $\mathbb{1}_A$ and the Markov property, the conditional stationary probability of A is

  $$\mathbb{E}_\pi(\mathbb{1}_A \mid \mathcal{F}_n) = \mathbb{E}_\pi(\mathbb{1}_A \circ \theta^n \mid \mathcal{F}_n) = \mathbb{E}_{X_n}(\mathbb{1}_A).$$

  By Lévy's 0-1 law, the left-hand side converges to $\mathbb{1}_A$ as $n \to \infty$. Now denote $h(x) := \mathbb{E}_x(\mathbb{1}_A)$. If the chain is irreducible and (positive) recurrent, then for any $y \in S$, the right-hand side $h(X_n) = h(y)$ infinitely often (as X_n revisits y i.o.). As $h(X_n)$ converges to an indicator, we must actually have $h \equiv 0$ or $h \equiv 1$ on S, and $\mathbb{E}_\pi(\mathbb{1}_A) = \mathbb{P}_\pi(A) = 0$ or 1. Thus irreducibility and positive recurrence imply ergodicity.

  (I amend my previous remarks, or acknowledge that Durrett may be using a different convention.) This example also shows that $\mathcal{I}$ and $\mathcal{T}$ may differ. $\mathcal{I}$ is trivial above, but if the chain is periodic with period $d > 1$, then $\mathcal{T} = \sigma(\{X_0 \in S_r\} : 0 \leq r < d)$, where $S_0, \ldots, S_{d-1}$ is the cyclic decomposition of S.

- The deterministic or almost-sure rotation by θ is not ergodic if $\theta \in \mathbb{Q}$. Let $\theta = m/n$ in reduced form, $m < n$ positive integers. Then if $B \subseteq [0, 1/n)$ is a Borel subset, the collection of its rotations $A = \bigcup_{k=0}^{n-1} \left(B + \frac{k}{n}\right)$ is an invariant set.

Conversely, if θ is irrational, then φ is ergodic. One proof borrows from Fourier analysis: if $f\colon [0,1) \to \mathbb{R}$ is a measurable function with $\int_{[0,1)} f^2 \, \mathrm{d}x < \infty$, then $f(x) = \sum_{k\in\mathbb{Z}} c_k e^{2\pi i k x}$, with equality in the sense of

$$\sum_{k=-M}^{M} c_k e^{2\pi i k x} \to f(x) \text{ in } L^2([0,1)) \text{ as } M \to \infty.$$

Moreover, this representation is unique: the coefficients are determined by $c_k = \int_{[0,1)} f(x) e^{-2\pi i k x} \, \mathrm{d}x$. Thus

$$f(\varphi x) = \sum_{k\in\mathbb{Z}} (c_k e^{2\pi i k \theta}) e^{2\pi i k x},$$

where $e^{2\pi i y} = e^{2\pi i (y \bmod 1)}$. By uniqueness, $f \circ \varphi = f$ iff $c_k(e^{2\pi i k\theta} - 1) = 0$. But as θ is irrational, this implies $c_k = 0$ for k nonzero, and $f\colon [0,1) \to \mathbb{R}$ is constant. If $f = \mathbb{1}_A$ for $A \in \mathcal{I}$, then $A = \varnothing$ or $[0,1)$ almost surely.

Let us also consider a more direct proof of ergodicity when θ is irrational.

i. The "wrapped-around" lattice $x_n = n\theta \bmod 1$ is dense in $[0,1)$. By irrationality, all of the x_n are distinct (if $x_m = x_n$, then $(n-m)\theta \in \mathbb{Z}$), so for any $N < \infty$, there exist $m, n \le N$ such that $|x_n - x_m| \le 1/N$. Now let (a,b) be any "circular" open interval in $[0,1)$ with length $\ell > 0$, and take N such that $1/N < \ell$. Then there exist $x_m, x_n, x_{n+(n-m)}, x_{n+2(n-m)}, \ldots$, spaced out with distance $\le 1/N$. It is impossible to have a contiguous interval of length ℓ lie on $[0,1)$ without intersecting any of these points, which shows that $\{x_n\}$ is dense in $[0,1)$.

ii. For any Borel set $B \subseteq [0,1)$ with $\mu(B) > 0$ and for any $\delta > 0$, there exists an interval $I = [a,b)$ such that $\mu(B \cap I) > (1-\delta)\mu(I)$. By lemma A.2.1, we have $A = \bigsqcup_{i=1}^{n} [a_i, b_i)$ such that $\mu(B \triangle A) < \delta\mu(A)$. Suppose that $\mu(I_i) - \mu(I_i \cap B) \ge \delta\mu(I_i)$ for all $1 \le i \le n$. But then

$$\mu(B \triangle A) \ge \mu(A \setminus B) = \sum_{i=1}^{n} \mu(I_i \setminus B) = \sum_{i=1}^{n} (\mu(I_i) - \mu(I_i \cap B)) \ge \delta \sum_{i=1}^{n} \mu(I_i) = \delta\mu(A).$$

So there exists some $I_i = [a_i, b_i)$ satisfying the claim, and take $I = I_i$.

iii. For $A \in \mathcal{I}$ with positive measure, by part ii (Lebesgue's density theorem in $[0,1)$), we find that almost every $a \in A$ has density 1. No point in A^c has density 1 for A^c by part i, as any interval about x contains some $n\theta$-rotation of $(a-\varepsilon, a+\varepsilon)$ for $a \in A$ by the denseness of $n\theta \bmod 1$. Therefore $\mu(A^c) = 0$, i.e. $\mathbb{P}(A) = 1$, and we are done.

Now, a few more results about invariant events and random variables:

Proposition 2 (Invariant random variables).

X is $\mathcal{I}$-measurable iff X is *invariant*, i.e. $X \circ \varphi = X$ almost surely.

Proof. Recall that every $\mathcal{I}$-measurable X is the limit of simple $\mathcal{I}$-measurable functions. It suffices to follow a standard monotone class argument:

i. If $X = \mathbb{1}_A$ for $A \in \mathcal{I}$, then $\mathbb{P}(\varphi^{-1}(A) \triangle A) = \mathbb{E}(|\mathbb{1}_A \circ \varphi - \mathbb{1}_A|) = 0$ shows that $X \circ \varphi = X$ a.s.

ii. If X is simple, then it is clear by linearity that $X \circ \varphi = X$ a.s.

iii. If $X = \lim_{n\to\infty} X_n$ for $X_1, X_2, \ldots$ simple, let N_n be the null set on which $X_n \circ \varphi \neq X_n$ for each n. Then outside of the null set $\bigcup_{n=1}^{\infty} N_n$, i.e. almost surely, we see that $X \circ \varphi = X$.

Conversely, suppose that $X \circ \varphi = X$ a.s., outside of the null set N. Then for any $A \in \sigma(X)$, where $A = X^{-1}(B)$,

$$\mathbb{P}(\varphi^{-1}(A) \triangle A) = \mathbb{P}((X \circ \varphi)^{-1}(B) \triangle X^{-1}(B)) \leq \mathbb{P}(N) = 0.$$

Thus $\sigma(X) \subseteq \mathcal{I}$, i.e. X is $\mathcal{I}$-measurable.

□

Proposition 3 (Almost and strict invariance)**.**

An event A is *strictly invariant* if $A = \varphi^{-1}(A)$ and *almost invariant* if $\mathbb{P}(A \triangle \varphi^{-1}(A)) = 0$.

a. If A is any set and $B = \bigcup_{n=0}^{\infty} \varphi^{-n}(A)$, then $\varphi^{-1}(B) \subseteq B$.

b. If B is any set with $\varphi^{-1}(B) \subseteq B$ and $C = \bigcap_{n=0}^{\infty} \varphi^{-n}(B)$, then $\varphi^{-1}(C) = C$.

c. A is almost invariant iff there is C strictly invariant such that $\mathbb{P}(A \triangle C) = 0$.

Proof. As we consider events up to almost sure equality, we simply call almost invariant sets *invariant*.

a. $\varphi^{-1}(\bigcup_{n=0}^{\infty} \varphi^{-n}(A)) = \bigcup_{n=0}^{\infty} \varphi^{-1}(\varphi^{-n}(A)) = \bigcup_{n=1}^{\infty} \varphi^{-n}(A) \subseteq \bigcup_{n=0}^{\infty} \varphi^{-n}(A)$.

b. $\varphi^{-1}(\bigcap_{n=0}^{\infty} \varphi^{-n}(B)) = \bigcap_{n=1}^{\infty} \varphi^{-n}(B) \supseteq \bigcap_{n=0}^{\infty} \varphi^{-n}(B)$, and as $B = \varphi^{0}(B) \supseteq \varphi^{-1}(B)$, we have equality.

c. If A has $C = \varphi^{-1}(C)$ such that $\mathbb{P}(A \triangle C) = 0$, then

$$\mathbb{P}(\varphi^{-1}(A) \triangle C) = \mathbb{P}(\varphi^{-1}(A) \triangle \varphi^{-1}(C)) = \mathbb{P}(\varphi^{-1}(A \triangle C)) = \mathbb{P}(A \triangle C) = 0.$$

That is, $\mathbb{1}_{\varphi^{-1}(A)} = \mathbb{1}_C = \mathbb{1}_A$ a.s. Conversely, let A be almost invariant and define $C = \limsup_{n\to\infty} \varphi^{-n}(A)$ as above. Then $A \setminus C = \liminf(A \setminus \varphi^{-n}(A))$ and $C \setminus A = \limsup(\varphi^{-n}(A) \setminus A)$ both have measure zero:

$$\mathbb{P}(C \setminus A) \leq \mathbb{P}\Big(\limsup_{n\to\infty}(\varphi^{-n} A \triangle A)\Big) \leq \sum_{n=1}^{\infty} \mathbb{P}(\varphi^{-n} A \triangle A) \leq \sum_{n=1}^{\infty}\sum_{k=1}^{n} \mathbb{P}(\varphi^{-k} A \triangle \varphi^{-(k-1)} A) = 0.$$

□

If the trajectory of the average particle is "space-filling," then we might expect the time average, the average of the states sampled by the particle at times $0, 1, \ldots, n$, to the space average, the true expectation taken over all the states. This idea generalizes the SLLN, taking as stationary sequence a collection of i.i.d. integrable random variables. In this vein, the following result is also known as the pointwise or individual ergodic theorem:

Theorem 1 (Birkhoff's ergodic theorem)**.**

For any $X \in L^1(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{R})$, φ measure-preserving, and $X_n(\omega) = X(\varphi^n \omega)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k(\omega) \to \mathbb{E}(X \mid \mathcal{I}) \text{ almost surely and in } L^1.$$

As a consequence, when the sequence is ergodic, $\mathcal{I}$ is trivial, and $\mathbb{E}(X \mid \mathcal{I}) = \mathbb{E}(X)$. In particular, if $X = \mathbb{1}_A$, then the asymptotic fraction of time in which $\varphi^n \in A$ is the "spatial" average $\mathbb{P}(A)$.

We will need a slightly odd integration inequality:

Lemma 1 (Maximal ergodic lemma)**.**

Let $X_i(\omega) = X(\varphi^i \omega)$, let $S_n \coloneqq \sum_{i=0}^{n-1} X_i$, $S_0 = 0$, and let $M_n = \max_{0 \le k \le n} S_k$. Then $\mathbb{E}(X \mid M_n > 0) \ge 0$.

Proof. Note that $M_n(\varphi\omega) \ge S_k(\varphi\omega)$ for $k \le n$, so $X(\omega) + M_n(\varphi\omega) \ge X(\omega) + S_k(\varphi\omega) = S_{k+1}(\omega)$. Rearranging,

$$X(\omega) \ge S_{k+1}(\omega) - M_n(\varphi\omega) \text{ for } k = 1, \ldots, n.$$

The above holds for $k = 0$ as well, since $S_1(\omega) = X(\omega) \ge S_1(\omega) - M_n(\varphi\omega)$ by $M_n \ge 0$. Taking the maximum of the right-hand side over $k = 0, \ldots, n-1$, by the monotonicity of expectation,

$$\begin{aligned}
\mathbb{E}(X \mid M_n > 0) &\ge \int_{\{M_n > 0\}} \max_{1 \le k \le n} S_k(\omega) - M_n(\varphi\omega) \, \mathrm{d}\mathbb{P} \\
&= \int_{\{M_n > 0\}} M_n(\omega) - M_n(\varphi\omega) \, \mathrm{d}\mathbb{P} \\
&\ge \int_{\Omega} M_n(\omega) - M_n(\varphi\omega) \, \mathrm{d}\mathbb{P} \\
&= \mathbb{E}(M_n - (M_n \circ \varphi)).
\end{aligned}$$

The second inequality comes from $M_n(\omega) - M_n(\varphi\omega) = 0 - M_n(\varphi\omega) \le 0$ on the event $\{M_n > 0\}^c = \{M_n = 0\}$. Then, as φ is measure-preserving, $\mathbb{E}(M_n - (M_n \circ \varphi)) = 0$. □

Proof of Theorem 1. The $\mathcal{I}$-measurable $\mathbb{E}(X \mid \mathcal{I})$ is φ-invariant by Proposition 2, so taking $X' = X - \mathbb{E}(X \mid \mathcal{I})$, we may assume $\mathbb{E}(X \mid \mathcal{I}) = 0$ without loss of generality. Let $\varepsilon > 0$, let $\bar{X} = \limsup_{n\to\infty} S_n/n$, and let

$$\Delta \coloneqq \left\{\omega : \bar{X}(\omega) > \varepsilon\right\}.$$

We wish to show that $\mathbb{P}(\Delta) = 0$, which gives $\limsup_{n\to\infty} S_n/n \le 0$ a.s., and by symmetry we will have $S_n/n \to 0$ a.s. Matching the notation of Lemma 1, we write

$$\begin{aligned}
X^*(\omega) &\coloneqq (X(\omega) - \varepsilon) \cdot \mathbb{1}_\Delta(\omega) \\
S_n^*(\omega) &\coloneqq \sum_{i=0}^{n-1} X^*(\varphi^{i-1}\omega) \\
M_n^*(\omega) &\coloneqq \max_{0 \le k \le n} S_k^*(\omega) = \max\{0, S_1^*(\omega), \ldots, S_n^*(\omega)\} \\
A_n &\coloneqq \{M_n^* > 0\} \\
A &\coloneqq \bigcup_{n=1}^{\infty} A_n = \left\{\sup_{n \ge 1} \frac{S_n^*}{n} > 0\right\}.
\end{aligned}$$

Now, invoking Lemma 1, $\mathbb{E}(X^*; A_n) \ge 0$. By dominated convergence, where $\mathbb{E}(|X^*|) \le \mathbb{E}(|X|) + \varepsilon < \infty$, we see that $\mathbb{E}(X^*; A) \ge 0$ as well. Observe that as $X^* = X - \varepsilon$ on Δ per definition,

$$A = \left\{\sup_{n\ge 1} \frac{S_n}{n} > \varepsilon\right\} \cap \Delta = \left\{\limsup_{n\to\infty} \frac{S_n}{n} > \varepsilon\right\} = \Delta.$$

Moreover, $\Delta \in \mathcal{I}$ as $\bar{X}(\varphi\omega) = \bar{X}(\omega)$. Thus

$$\mathbb{E}(X^*; A) = \mathbb{E}(X^*; \Delta) = \mathbb{E}(X - \varepsilon; \Delta) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{I}); \Delta) - \varepsilon\,\mathbb{P}(\Delta) = -\varepsilon\,\mathbb{P}(\Delta) \geq 0,$$

where $\mathbb{E}(X \mid \mathcal{I}) = 0$, which shows that $\mathbb{P}(\Delta) = 0$. Now, we perform a routine upgrade from a.s. to L^1 convergence using truncation. Let $X'_M = X\mathbb{1}_{|X|\leq M}$ and $Y_M = X - X'_M$. We know that

$$\frac{1}{n}\sum_{k=0}^{n-1} X'_M(\varphi^k\omega) \to \mathbb{E}(X'_M \mid \mathcal{I}) \text{ a.s.}$$

by the proof above, and we have L^1 convergence by the bounded convergence theorem. For Y_M, observe that by the triangle inequality or Jensen's inequality,

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} Y_M(\varphi^k\omega)\right| \leq \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}\left|Y_M(\varphi^k\omega)\right| = \mathbb{E}\left|Y_M\right|$$

and $\mathbb{E}\left|\mathbb{E}(Y_M \mid \mathcal{I})\right| \leq \mathbb{E}\,\mathbb{E}(|Y_M| \mid \mathcal{I}) = \mathbb{E}\left|Y_M\right|$ by the tower property. Putting these two bounds together,

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} Y_M(\varphi^k\omega) - \mathbb{E}(Y_M \mid \mathcal{I})\right| \leq 2\,\mathbb{E}\left|Y_M\right|.$$

Now, given the L^1 convergence for X'_M shown above, we find that

$$\limsup_{n\to\infty}\mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} X(\varphi^k\omega) - \mathbb{E}(X \mid \mathcal{I})\right| \leq 2\,\mathbb{E}\left|Y_M\right|,$$

which tends to 0 as $M \to \infty$ by dominated convergence, and we are done. □

Note that Birkhoff's ergodic theorem implies the strong law of large numbers. The L^1 convergence follows without needing Theorem 1: decomposing $X = X^+ - X^-$, note that

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k^+ \to X^+ \text{ a.s.} \quad \text{and} \quad \mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} X_k^+\right| = \mathbb{E}\left|X^+\right|.$$

By Vitali's convergence theorem (4.6.3 in Durrett), we find $\frac{1}{n}\sum_{k=0}^{n-1} X_k^+ \to X^+$ in L^1, and likewise for the negative part. By the triangle inequality,

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} X_k - X\right| \leq \mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} X_k^+ - X^+\right| + \mathbb{E}\left|\frac{1}{n}\sum_{k=0}^{n-1} X_k^- - X^-\right| \to 0.$$

We have a few more examples of applications of the ergodic theorem:

Proposition 4 (Markov reward process).

Let $(X_n)_{n\in\mathbb{N}}$ be an irreducible Markov chain on a countable state space S with stationary distribution π, and let $f\colon S \to \mathbb{R}$ be a so-called *reward* function with

$$\mathbb{E}_\pi(|f|) = \sum_{x\in S}|f(x)|\,\pi(x) < \infty.$$

By ergodicity, $\mathcal{I}$ is trivial, so applying Theorem 1 to $f(X_0)$, we find that the "average reward" collected by a particle walking on the Markov chain converges to

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \mathbb{E}_\pi(f) \text{ almost surely and in } L^1.$$

Proposition 5 (Weyl's equidistribution theorem)**.**

For the example of rotation $\varphi(\omega) = (\omega + \theta) \bmod 1$ by irrational θ, let $X = \mathbb{1}_A$, $A \subseteq [0, 1)$ Borel. By Theorem 1,

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\varphi^k \omega \in A} \to \mathbb{P}(A) \text{ a.s.}$$

The case of $\omega = 0$ is usually known as *Weyl's equidistribution theorem*.

To be continued.

■