BIRKHOFF'S ERGODIC THEOREM AND THE LAW OF LARGE NUMBERS

CHENJIA LIN

Contents

1. Introduction

This paper will introduce some basic ideas in ergodic theory and demonstrate that the strong law of large numbers (SLLN) follows from Birkhoff's ergodic theorem.

Theorem 1.1. *(Strong Law of Large Numbers) If $X_1, X_2, \dots$ is a family of independent, identically-distributed random variables with finite mean α, then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \alpha \quad \textit{almost surely.}$$

There are multiple "weak" laws of large numbers and they have different assumptions e.g. $X_1, X_2, \dots$ are uncorrelated with the same mean and bounded variance or i.i.d. with a condition on the tail distributions (see [2]), though their common point is convergence *in probability*. The SLLN is "strong" in the sense that we obtain a stronger almost sure convergence without any assumptions about the variance of the random variables.

Birkhoff's ergodic theorem—also known as the pointwise ergodic theorem—leads to a proof of the SLLN. We state the theorem below and unpack its components in the following section. Proofs of the theorem are given in each of the references.

Theorem 1.2 (Birkhoff's Ergodic Theorem)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $T : \Omega \to \Omega$ a measure-preserving transformation. If $f \in L^1(\Omega)$, then there exists $\widehat{f} \in L^1(\Omega)$ such that*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^{-n}(\omega) = \widehat{f}(\omega) \quad \textit{almost everywhere.}$$

In particular, if T is also ergodic, then $\widehat{f}$ is constant a.e. The constant equals 0 *if* $\mu(\Omega) = \infty$ *and*

$$\frac{1}{\mu(\Omega)} \int_\Omega f d\mu$$

if $\mu(\Omega) < \infty$. *Hence, if* μ *is a probability measure,* $\widehat{f} \equiv \mathbb{E}f$.

2. Basics of Ergodic Theory

Throughout this section, we assume that $(\Omega, \mathcal{F}, \mu)$ is a measure space. If $\mathcal{A}$ is a collection of subsets of Ω, then we let $\sigma(\mathcal{A})$ denote the σ-algebra generated by $\mathcal{A}$.

Definition 2.1. We say that a measureable function $T : \Omega \to \Omega$ is *measure-preserving* if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{F}$. If $A \in \mathcal{F}$ such that $T^{-1}(A) = A$, then we say that A is T-invariant.

If T is measure-preserving, then $\mu(T^{-n}(A)) = \mu(A)$ for all $A \in \mathcal{F}$ and $n \in \mathbb{N}$. If A is T-invariant, then $T^{-n}(A) = A$. In practice, it can be somewhat difficult to directly show that a function T is measure-preserving on $\mathcal{F}$. Fortunately, it suffices to show that T is measure-preserving on a π-system that generates $\mathcal{F}$.

Proposition 2.2. *If* $\mathcal{A} \subseteq \mathcal{F}$ *is a* π*-system such that* $\mu(T^{-1}(A)) = \mu(A)$ *for all* $A \in \mathcal{A}$*, then* T *is measure-preserving.*

Proof. Define $\mathcal{L} = \{A \in \sigma(\mathcal{A}) : \mu(T^{-1}(A)) = \mu(A)\}$ and note that it is a λ-system; if $A \in \mathcal{L}$, then

$$\begin{aligned} \mu(T^{-1}(\Omega \setminus A)) &= \mu(\Omega \setminus T^{-1}(A)) \\ &= \mu(\Omega) - \mu(T^{-1}(A)) \\ &= \mu(\Omega) - \mu(A) \\ &= \mu(\Omega \setminus A) \end{aligned}$$

implies that $\mathcal{L}$ is closed under complements. If $A_1, \mathcal{A}_2, \dots \in \mathcal{A}$ are disjoint, then

$$\begin{aligned} \mu\left(T^{-1}\left(\bigsqcup_{i=1}^{\infty} A_i\right)\right) &= \mu\left(\bigsqcup_{i=1}^{\infty} T^{-1}(A_i)\right) \\ &= \sum_{i=1}^{\infty} \mu(T^{-1}(A_i)) \\ &= \sum_{i=1}^{\infty} \mu(A_i) \\ &= \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \end{aligned}$$

shows that $\mathcal{L}$ is closed under countable disjoint unions. Notice that $\mathcal{A} \subseteq \mathcal{L}$, so Dynkin's π-λ theorem gives $\mathcal{A} \subseteq \sigma(\mathcal{A}) \subseteq \mathcal{L}$, which implies $\sigma(\mathcal{A}) = \mathcal{L}$ and T is measure-preserving on $\sigma(\mathcal{A})$. $\square$

One can verify using properties about preimages that the T-invariant subsets of Ω form a σ-algebra.

Definition 2.3. We say that a measure-preserving transformation T is *ergodic* if every T-invariant subset $A \in \mathcal{F}$ has $\mu(A) = 0$ or $\mu(\Omega \setminus A) = 0$. In particular, if μ is a probability measure, then either $\mu(A) = 0$ or 1.

Ergodicity is analogous to the notion of *(topological) transitivity* when studying dynamical systems defined on topological spaces. Recall that a homeomorphism $f : X \to X$ on a topological space X is transitive if there exists a point whose orbit is dense. A set having full measure is roughly analogous to a set being dense. If T is ergodic, then T-invariant sets are either negligible or as "big" as Ω.

The notion of topologically mixing also has a measure-theoretic counterpart, though we require μ to be a probability measure.

Definition 2.4. If $(\Omega, \mathcal{F}, \mu)$ is a probability space, then we say that a measure-preserving transformation T is *mixing* if

$$\lim_{n\to\infty} \mu\big(A \cap T^{-n}(B)\big) = \mu(A) \cdot \mu(B)$$

for any $A, B \in \mathcal{F}$.

The following result demonstrates the need for μ to be a probability measure.

Proposition 2.5. *If $T : \Omega \to \Omega$ is mixing, then T is ergodic.*

Proof. Let $A \in \mathcal{F}$ where $T^{-1}(A) = A$. If T is mixing, then

$$\lim_{n\to\infty} \mu(A \cap T^{-n}(A)) = \lim_{n\to\infty} \mu(A) = \mu(A) = \mu(A)\mu(A)$$

implies $\mu(A) = 0$ or 1. In particular, this also gives $\mu(\Omega) = 0$ or 1. □

It is generally difficult to prove that a transformation T is ergodic, though this follows immediately if we know that T is mixing. As a word of caution, ergodic transformations need not be mixing, and an example of this is the rotation $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ on $\mathbb{R}/\mathbb{Z}$ by an irrational angle α. Nevertheless, we will invoke the result above in our proof of the SLLN.

3. Law of Large Numbers

To arrive at the strong law of large numbers, we will define a Borel probability measure $\mathbb{P}$ on $\mathbb{R}^{\mathbb{N}}$ (endowed with the product topology) and show that the *leftward shift operator* $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ defined by

$$T(y_1, y_2, \dots) = (y_2, y_3, \dots)$$

is measure-preserving and ergodic.

Recall from topology that a (countable) basis for $\mathbb{R}^{\mathbb{N}}$ is the collection of all sets of the form

$$\underbrace{I_1 \times \cdots \times I_k \times \mathbb{R} \times \mathbb{R} \times \cdots}_{\text{"rational open cylinder set"}}, \tag{3.1}$$

where $k \in \mathbb{N}$ and $I_1, \dots, I_k$ are open intervals with rational endpoints. Notice that this collection is a π-system that generates the Borel σ-algebra on $\mathbb{R}^{\mathbb{N}}$. If $X_1, X_2, \dots$ are random variables on a probability space $(\Omega, \mathcal{F}, \mu)$, then they induce pushforward measures $\mu_1, \mu_2, \dots$ defined by $\mu_i(A) := \mu(X_i^{-1}(A))$ for all $A \subseteq \mathbb{R}$. If $X_1, X_2, \dots$ are also i.i.d, then let X be a random variable with the same distribution. Through a measure extension theorem, we may obtain a unique Borel probability measure $\mathbb{P}$ on $\mathbb{R}^{\mathbb{N}}$ that satisfies

$$\mathbb{P}(I_1 \times \cdots I_k \times \mathbb{R} \times \mathbb{R} \times \cdots) = \mu_X(I_1) \cdots \mu_X(I_k)$$

(Section 2.7 of [1]).

We now proceed to show that the shift $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is measure-preserving and mixing, which implies T is ergodic. The following result will allow us to approximate Borel sets on $\mathbb{R}^{\mathbb{N}}$ with the open rational cyliner sets, which we can compute the measures of.

Lemma 3.2. *Let $(\mathcal{M}, d)$ be a metric space with a countable basis $\mathcal{C}$. Define*

$$\overline{\mathcal{C}} := \{U \subseteq \mathcal{M} : U \in \mathcal{C} \text{ or } \mathcal{M} \setminus U \in \mathcal{C}\} \cup \{\mathcal{M}\}.$$

If $B \subseteq \mathcal{M}$ is a Borel set and $\varepsilon > 0$, then there exist $C_1, \dots, C_n \in \overline{\mathcal{C}}$ such that

$$\mu\left(B \triangle \left(\bigcup_{i=1}^{n} C_i\right)\right) < \varepsilon.$$

Lemma 3.3. *If A and B are measureable, then*

$$|\mu(A) - \mu(B)| \leq \mu(A \triangle B).$$

Proof. Observe that

$$\begin{aligned} |\mu(A) - \mu(B)| &= |\mu(A \setminus B) + \mu(A \cap B) - \mu(B \setminus A) - \mu(B \cap A)| \\ &= |\mu(A \setminus B) - \mu(B \setminus A)| \\ &\leq \mu(A \setminus B) + \mu(B \setminus A) \\ &= \mu(A \triangle B). \end{aligned}$$

□

Proposition 3.4. *The shift $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is measure-preserving.*

Proof. By Proposition 2.2, it suffices to show that T is measure-preserving for sets of the form (3.1) since they form a π-system that generates the Borel σ-algebra on $\mathbb{R}^{\mathbb{N}}$; we have

$$\begin{aligned} \mathbb{P}\big(T^{-1}(I_1 \times \cdots \times I_k \times \mathbb{R} \times \mathbb{R} \times \cdots)\big) &= \mathbb{P}\left(\mathbb{R} \times I_1 \times \cdots \times I_k \times \mathbb{R} \times \mathbb{R} \times \cdots\right) \\ &= \mu_X(\mathbb{R})\mu_X(I_1) \cdots \mu_X(I_k) \\ &= \mathbb{P}(I_1 \times \cdots \times I_k \times \mathbb{R} \times \mathbb{R} \times \cdots). \end{aligned}$$

□

Proposition 3.5. *The shift $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is mixing.*

Proof. If $A, B \subseteq \mathbb{R}^{\mathbb{N}}$ are Borel sets and $\varepsilon > 0$, then let $U, V \subseteq \mathbb{R}^{\mathbb{N}}$ be finite unions of rational open cylinder sets such that $\mathbb{P}(A \triangle U) < \varepsilon/4$ and $\mathbb{P}(B \triangle V) < \varepsilon/4$. In particular, we may write

$$\begin{aligned} U &= U_1 \times \cdots \times U_k \times \mathbb{R} \times \mathbb{R} \times \cdots, \\ V &= V_1 \times \cdots \times V_\ell \times \mathbb{R} \times \mathbb{R} \times \cdots \end{aligned}$$

where U_i, V_j are open sets in $\mathbb{R}$. Letting $n \geq k$, we have

$$\begin{aligned} &\mathbb{P}(U \cap T^{-n}(V)) \\ &= \mathbb{P}\left\{\omega \in \mathbb{R}^{\mathbb{N}} : (\omega_1, \dots, \omega_k) \in U_1 \times \cdots \times U_k, (\omega_{n+1}, \dots, \omega_{n+\ell}) \in V_1 \times \cdots \times V_\ell\right\} \\ &= \mu_X(U_1) \cdots \mu_X(U_k) \cdot \mu_X(V_1) \cdots \mu_X(V_\ell) \\ &= \mathbb{P}(U) \cdot \mathbb{P}(V). \end{aligned}$$

Now observe that

$$\begin{aligned}
&|\mathbb{P}(A \cap T^{-n}(B)) - \mathbb{P}(A)\mathbb{P}(B)| \\
&\quad \leq |\mathbb{P}(A \cap T^{-n}(B)) - \mathbb{P}(U \cap T^{-n}(V))| + |\mathbb{P}(U \cap T^{-n}(V)) - \mathbb{P}(U)\mathbb{P}(V)| \\
&\qquad + |\mathbb{P}(U)\mathbb{P}(V) - \mathbb{P}(U)\mathbb{P}(B)| + |\mathbb{P}(U)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)| \\
&\quad \leq |\mathbb{P}(A \cap T^{-n}(B)) - \mathbb{P}(U \cap T^{-n}(V))| + 0 \\
&\qquad + \mathbb{P}(U)|\mathbb{P}(V) - \mathbb{P}(B)| + \mathbb{P}(B)|\mathbb{P}(U) - \mathbb{P}(A)| \\
&\quad \leq |\mathbb{P}(A \cap T^{-n}(B)) - \mathbb{P}(U \cap T^{-n}(V))| + \mathbb{P}(B \triangle V) + \mathbb{P}(A \triangle U) \\
&\quad \leq \mathbb{P}\big((A \cap T^{-n}(B)) \triangle (U \cap T^{-n}(V))\big) + \varepsilon/4 + \varepsilon/4.
\end{aligned}$$

Noting that $\mathbb{P}(T^{-n}(B) \triangle T^{-n}(V)) = \mathbb{P}(B \triangle V)$ since T is measure-preserving, we have

$$\begin{aligned}
\mathbb{P}\big((A \cap T^{-n}(B)) \triangle (U \cap T^{-n}(V))\big) &\leq \mathbb{P}\big((A \triangle U) \cup (T^{-n}(B) \triangle T^{-n}(V))\big) \\
&\leq \mathbb{P}(A \triangle U) + \mathbb{P}(T^{-n}(B) \triangle T^{-n}(V)) \\
&\leq \varepsilon/4 + \varepsilon/4 \\
&= \varepsilon/2,
\end{aligned}$$

which altogether gives $|\mathbb{P}(A \cap T^{-n}(B)) - \mathbb{P}(A)\mathbb{P}(B)| < \varepsilon$ for all $\varepsilon > 0$. Since n is independent of ε, we obtain $\lim_{n\to\infty} \mathbb{P}(A \cap T^{-n}(B)) = \mathbb{P}(A)\mathbb{P}(B)$ as desired. □

Having verified that the shift operator on $\mathbb{R}^{\mathbb{N}}$ is ergodic, we conclude with a proof of the SLLN.

Proof of the SLLN. Let $Y : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be the projection onto the first coordinate i.e. $Y(\omega_1, \omega_2, \dots) = \omega_1$. Evaluating at $(X_1(\omega), X_2(\omega), \dots) \in \mathbb{R}^{\mathbb{N}}$ gives us

$$\frac{1}{n}\sum_{j=0}^{n-1} (Y \circ T^j)\big(X_1(\omega), X_2(\omega), \dots\big) = \frac{1}{n}\sum_{j=0}^{n-1} X_{j+1}(\omega) = \frac{1}{n}\sum_{j=1}^{n} X_j(\omega).$$

Then by Birkhoff's ergodic theorem, we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j = \mathbb{E}Y \quad \text{almost surely,}$$

where

$$\mathbb{E}Y = \int_{\mathbb{R}^{\mathbb{N}}} Y \, d\mathbb{P} = \int_{\Omega} X_1 \, d\mu = \mathbb{E}X. \qquad \square$$

References

[1] R. Ash. *Probability and Measure Theory.* Academic Press. 2000.
[2] R. Durrett. *Probability: Theory and Examples.* 2019.
[3] J. Hawkins. *Ergodic Dynamics: From Basic Theory to Applications.* Springer. 2021.
[4] S. Lalley. *Birkhoff's Ergodic Theorem.* 2019.
http://galton.uchicago.edu/~lalley/Courses/381/ErgodicTheorem.pdf