4 Birkhoff's Ergodic Theorem

Ergodic theorems, roughly speaking, are concerned with the question: When do averages of quantities, generated in a somehow 'stationary' manner, converge? A thorough treatment can be found in Krengel, Ergodic Thoerems, de Gruyter, 1985. We present one particularly important basic result, Birkhoff's ergodic theorem, and some special cases.

In this section, we consider the complete normed space $L^1 = L^1(\Omega, \mathfrak{A}, P)$. For a continuous linear mapping $T : L^1 \to L^1$, its *norm* is defined as

$$\|T\| := \sup_{\substack{X \in L^1 \\ X \neq 0}} \frac{\|T(X)\|_{L^1}}{\|X\|_{L^1}} \ ;$$

this number is always finite; further,

$$\|T(X)\|_{L^1} \leq \|T\| \cdot \|X\|_{L^1} \ , \qquad \forall X \in L^1 \ . \tag{1}$$

For two operators $S, T : L^1 \to L^1$, we set in short ST for the composition. It is not hard to see that $\|ST\| \leq \|S\|\|T\|$. In particular, we define powers $T^i := TT^{i-1}$ for $i > 1$; then $\|T^i\| \leq \|T\|^i$.

An linear operator T is called a *contraction* iff $\|T\| \leq 1$; this is equivalent to the assumption that that $\|T(X) - T(X)\|_{L^1} \leq \|g - h\|_{L^1}$ for all $g, h \in L^1$.

For $X, Y \in L^1$, we write in short $X \leq Y$ iff $X \leq Y$ μ–a.e.. We say that a continuous linear operator $T : L^1 \to L^1$ is *positive* iff $X \geq 0$ implies $T(X) \geq 0$. This is equivalent to the assumption that $X \leq Y$ implies $T(X) \leq T(Y)$.

Example 1. Let $\tau : \Omega \to \Omega$ be an *endomorphism*, that is, τ is measurable and $P \circ \tau^{-1} = P$. Then this induces a positive contraction

$$T_\tau : L_2 \to L_2, \qquad T_\tau(f) := f \circ \tau \ .$$

Examples of endomorphisms τ are

- om $\Omega = \mathbb{R}$, translations $\tau(\omega) = \omega + x$;
- on a product space with product measure, $(\bigotimes_{i \in \mathbb{N}} \Omega, \bigotimes_{i \in \mathbb{N}} \mathfrak{A}, \bigotimes_{i \in \mathbb{N}} \mu)$, the shift operator $\tau((\omega_1, \omega_2, \ldots) = (\omega_2, \omega_3, \ldots)$;
- for a random walk $S_n = \sum_{i \leq n} X_i$ with X_i i.i.d. and integrable, we have $(S_n)_{n \in \mathfrak{N}} \stackrel{d}{=} (S_n - X_1)_{n \geq 2}$. This leads to an endomorphism as follows: Set $S = (S_n)_{n \in \mathfrak{N}}$ and consider the product space

  $$(\bigotimes_{i \in \mathbb{N}} \mathbb{R}, \bigotimes_{i \in \mathbb{N}} \mathfrak{B}, P_S) \ .$$

  Then the shift and reset operator

  $$\tau((\omega_1, \omega_2, \ldots,)) := (\omega_2 - \omega_1, \omega_3 - \omega_1, \ldots)$$

  is an endomorphism.

If T, T' are positive linear operators (plops) and $\alpha, \beta \geq 0$, then $T \circ T'$ and $\alpha T + \beta T'$ are plops as well. In particular, for a plop T and $n \in \mathfrak{N}$, the operators

$$S_n := \sum_{i=0}^{n-1} T^i \,, \qquad A_n := \frac{1}{n} S_n \tag{2}$$

are plops. Further, if T is a contraction, A_n is a contraction as well. For two random variables X, Y in L^1, their pointwise maximum $\max\{X, Y\}$ is in L^1 as well. Further, for a plop $T : L^1 \to L^1$ and A_n,S_n defined in (2) as in the previous subsection are plops as well, and for $X \in L^1$ we may define

$$M_n^A(X) := \max\{A_1(X), \ldots, A_n(X)\}, \qquad M_n^S(X) := \max\{S_1(X), \ldots, S_n(X)\} \; ; \tag{3}$$

then $M_n^{A/S}(X) \in L^1$ and $X \leq Y \Rightarrow M_n^{A/S}(X) \leq M_n^{A/S}(Y)$.

Definition 1. Let $\tau : \Omega \to \Omega$ be an endormorphism. $A \in \mathfrak{A}$ is called τ–invariant iff $A = \tau^{-1}(A)$. The class of τ–invariant sets is denoted by $\mathcal{I}_\tau$. A r.v. X is called τ–invariant if $X = X \circ \tau$.

It is clear that $\mathcal{I}_\tau$ forms a σ–algebra; further, it is easy to see that a r.v. X is τ–invariant iff for all $\alpha \in \mathbb{R}$ the level sets $\{X \geq \alpha\}$ are τ–invariant; hence X is τ–invariant iff X is $\mathcal{I}_\tau$–measurable. A particularly nice case occurs if $\mathcal{I}_\tau$ consists only of 0–1–sets (i.e., no 'really random' event is τ–invariant); in this case, the conditional expectation above equals the expectation. This case is so important that it is given a name:

Definition 2. An endomorphism $\tau : \Omega \to \Omega$ is called *ergodic* iff for all $A \in \mathcal{I}_\mu$, $\mu(A) \in \big\{0, 1\big\}$.

Example 2. Example 2 continued: It is trivial to see that the endomorphism in (i) is ergodic. By Kolmogorov's 0–1–Law, it is easy to see that the endomorphism given in (ii) is ergodic. Also, in Example (iii), any measurable set A can be written as $\{S \in B\}$; if A is τ–invariant, it follows for any k that $A = \{(S_{n+k} - S_k)_{n \in \mathbb{N}} \in B\}$; hence, by induction one derives that B is independent of all X_i and similarly as in the proof of Kolmogorov's 0–1–Law, this entails $P(A) \in \{0, 1\}$. Thus, also Part (iii) is an example of an ergodic τ.

Lemma 1. Let $Y \in L^1$ with $Y \geq 0$, and let $T : L^1 \to L^1$ be a plop as well as a contraction. Then

$$\int T(Y) dP \leq \int Y \, dP \,.$$

Proof. Since $Y \geq 0$, $T(Y) \geq 0$, and

$$\int T(Y) dP = \|T(Y)\|_{L^1} \leq \|Y\|_{L^1} = \int Y dP \,.$$

□

Theorem 1 (Hopf, 1954). Let T be a positive contraction in L^1, and $X \in L^1$. For $n \in \mathbb{N}$ denote $E_n := \big\{M_n^A(X) \geq 0\big\} = \big\{M_n^S(X) \geq 0\big\} \in \mathfrak{A}$, and set $E_\infty = \bigcup_n E_n$. Then

$$\int_{E_n} X \, dP \geq 0 \qquad \text{and} \qquad \int_{E_\infty} X \, dP \geq 0 \,.$$

A very rough and naïve interpretation of this theorem is that, whenever at least one $1/k \sum_{i<k} T^i(X)$ is ≥ 0, X might be negative 'sometimes', but not 'most of the time'; even more, it is positive 'on the average'.

Proof (A.M. Garsia, 1965). The second inequality follows form the first by monotone convergence. For a r.v. Y, we set as usual $Y^+ = \max\{Y, 0\}$. Now let n be fixed and $k \leq n$; then

$$(M_n^S(X))^+ \geq M_n^S(X) \geq S_k(X) \ .$$

Since T is positive, this implies $T[(M_n^S(X))^+] \geq T(S_k(X))$ and thus also

$$X + T[(M_n^S(X))^+] \geq X + T(S_k(X)) = S_{k+1}(X) \ .$$

This entails

$$X \geq S_{k+1}(X) - T[(M_n^S(X))^+] \ ;$$

this equality is valid for $k = 0$ as well, and by taking the maximum over $k = 0, \ldots, n-1$ on the right hand side it follows that

$$X \geq M_n^S(X) - T[(M_n^S(X))^+] \ .$$

From this we infer

$$\begin{aligned}
\int_{E_n} X \quad & \geq \quad \int_{E_n} M_n^S(X) - \int_{E_n} T[(M_n^S(X))^+] \, dP \\
& = \quad \int_{\Omega} (M_n^S(X))^+ \, dP - \int_{E_n} T[(M_n^S(X))^+] \, dP \\
& \geq \quad \int_{\Omega} (M_n^S(X))^+ \, dP - \int_{E_n} T[(M_n^S(X))^+] \, dP \\
& \overset{T[(M_n^S(X))^+] \geq 0}{\geq} \quad \int_{\Omega} (M_n^S(X))^+ \, dP - \int_{\Omega} T[(M_n^S(X))^+] \, dP \\
& \overset{Lem.1}{\geq} \quad 0 \ .
\end{aligned}$$

□

Corollary 1 (Wiener (1939)). Let $\tau : \Omega \to \Omega$ be an endomorphism. Then for any $X \in L^1$ we have

$$P\Big(\max_{k \leq n} \frac{1}{k} \sum_{i \leq k} (X \circ \tau^i) \ \geq \ \alpha \Big) \ \leq \ \alpha^{-1} E|X| \ .$$

This is a remarkable strengthening of the Chebyshev–Markov inequality III.2.1. Compare also with the Kolmogorov inequality IV.2.1.

Proof. Set

$$E_n = \Big\{ \max_{k \leq n} \frac{1}{k} \sum_{i \leq k} (X \circ \tau^i) \ \geq \ \alpha \Big\} = \Big\{ \max_{k \leq n} \frac{1}{k} \sum_{i \leq k} (T^i(\tilde{X})) \ \geq \ 0 \Big\}$$

with $\tilde{X} = X - \alpha$ and $T(\tilde{X}) := \tilde{X} \circ \tau$. Then Theorem 1 yields $\int_{E_n} \tilde{X} \geq 0$, hence

$$\alpha P(E_n) \leq \int_{E_n} X dP \leq \int_\Omega |X| dP = E|X| \ .$$

□

Theorem 2 (Birkhoff (1931)). Let $\tau : \Omega \to \Omega$ be an endomorphism and $X \in L^1$. Then

$$\frac{1}{n} \sum_{i<n} (X \circ \tau^i) \ \to \ E(X \,|\, \mathcal{I}_\tau) \qquad\qquad P - a.s..$$

Proof. We set $T(X) = X \circ \tau$ and adopt and use the above facts and notations. We will first show that the left hand side tends to a τ–invariant $\overline{X}$ a.s.; then we show that $\overline{X}$ is indeed the conditional expectation. We start with

$$A_{n+1}(X) = \frac{1}{n+1} S_{n+1}(X) = \frac{1}{n+1} X + \frac{n}{n+1} A_n(X) \circ \tau \ .$$

Thus, the random variables

$$X^{u/l} := \overline{\lim_n} \, / \, \underline{\lim_n} \, A_n(X)$$

are τ–invariant: Indeed,

$$X^u \circ \tau = \big(\overline{\lim_n} \, A_n(X)\big) \circ \tau = \overline{\lim_n} (A_n(X) \circ \tau) = \overline{\lim_n} \, A_{n+1}(X) = X^u \ .$$

We will now proceed in several steps.
Step 1: We prove that $P(X^u = \infty \vee X^l = -\infty) = 0$.
To this end, note that $X^u > \beta$ implies that there is n such that $M_n^A(X) > \beta$. Hence

$$P(X^u > \beta) \leq \sup_n P(M_n^A(X) > \beta) \leq \beta^{-1} E|X|$$

by Corollary 1. It follows that $P(X^u > \beta) \to 0$ for $\beta \to \infty$. On the other hand, $X^l = -(-X)^u$, and thus $P(X^l < -\alpha) \leq \alpha^{-1} E|X|$ for $\alpha > 0$.
Step 2: We prove that $X^u = X^l$ a.s..
To this end, it suffices to show that for $\alpha < \beta$, $P(X^l < \alpha < \beta < X^u) = 0$; set $B = \{X^l < \alpha < \beta < X^u\}$. Since X^l, X^u are τ–invariant, B is a τ–invariant event. If we define $\tilde{X} := (X - \beta)\mathbf{1}_B$, then

$$\tilde{X} \circ \tau^k = (X \circ \tau^k - \beta)\mathbf{1}_{\tau^{-k}(B)} = (X \circ \tau^k - \beta)\mathbf{1}_B \ ;$$

consequently, $A_n(\tilde{X}) = (A_n(X) - \beta)\mathbf{1}_B$. Recall the notion $E_n := \{M_n^A(\tilde{X}) \geq 0\}$ and $E_\infty = \bigcup_n E_n$. If $\omega \in B$, then $X^u(\omega) - \beta > 0$, thus there is n such that $M_n^A(\tilde{X}(\omega)) > 0$; hence, $B \subseteq E_\infty$. On the other hand, $\tilde{X}$ equals 0 outside B. This implies that

$$\int_B \tilde{X} dP = \int_{E_\infty} \tilde{X} dP \geq 0$$

by Theorem 1. It follows that $\beta\mu(B) \leq \int_B X dP$. The same arguments, applied to $\hat{X} = (\alpha - X)\mathbf{1}_B$, yield $\alpha\mu(B) \geq \int_B X dP$. But since $\alpha < \beta$, this can only be true if

$P(B) = 0$.
Step 3: $A_n(X)$ converges a.s..
Indeed, this follows immediately from Step 2. Let $\overline{X} = \lim_n A_n(X)$.
Step 3: $A_n(X)$ is uniformly integrable.
Indeed, if $A \in \mathfrak{A}$ is arbitrary, then $P(\tau^{-i}(A)) = P(A)$ for all $i \geq 0$. Let $\varepsilon > 0$ be given and choose $\delta > 0$ such that for $P(B) < \delta$ we have $\int_B |X| dP < \varepsilon$ (this is possible by Lemma III.4.1); then we have for A with $P(A) < \delta$ that $P(\tau^{-i}(A)) < \delta$ and hence

$$\int_A |A_n(X)| dP \leq \int_A A_n(|X|) = \frac{1}{n} \sum_{i<n} \int_{\tau^{-i}(A)} |X| dP < \varepsilon \; .$$

This proves by Lemma III.4.1 that $A_n(X)$ is uniformly integrable.
Step 4: If A is τ–invariant, then $\int_A \overline{X} dP = \int_A X dP$.
Indeed, for A τ–invariant, we have for $k \geq 0$,

$$\int_A X \circ \tau^k dP = \int_{\tau^{-k}(A)} X dP \circ \tau^{-k} = \int_A X dP \; ;$$

it follows that $\int_A A_n(X) dP = \int_A X dP$. However, by uniform integrability and a.s. convergence of $A_n(X)$,

$$\int_A \overline{X} dP = \lim_n \int_A A_n(X) dP = \int_A X dP \; .$$

In summary, $A_n(X) \to \overline{X}$ a.s.; $\overline{X}$ is τ–invariant, hence $\mathcal{I}_\tau$–measurable, and from Step 4 it follows that $\overline{X} = E(X|\mathcal{I}_\tau)$. □

Corollary 2. Let $\tau : \Omega \to \Omega$ be ergodic and $X \in L_1$. Then

$$\frac{1}{n} \sum_{i=0}^{n-1} (X \circ \tau^i) \;\to\; E(X) \qquad\qquad P - a.s. \, .$$

Proof. Since τ is ergodic, $\mathcal{I}_\tau$ consists only of 0–1–sets; hence by Remark 1.1(i) $\mathrm{E}(X \mid \mathcal{I}_\tau) = \mathrm{E}(X)$. □

From part (ii) of Example 2 we infer in particular with $X(\omega) = f(\omega_1, \ldots, \omega_d)$:

Corollary 3. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. in L^1; further, let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be measurable. Then

$$\frac{1}{n} \sum_{i<n} \varphi(X_{i+1}, \ldots, X_{i+d}) \overset{P\text{-a.s.}}{\longrightarrow} E(\varphi(X_1, \ldots, X_d)) \; .$$

This is a considerable generalization of the i.i.d.-SLLN IV.2.4.

Definition 3. A sequence $(X_n)_{n\in\mathbb{N}}$ of r.v. is called *ergodic* with respect to X iff

$$\frac{1}{n} \sum_{i\leq n} X_i \overset{P\text{-a.s.}}{\longrightarrow} \mathfrak{E}X \; .$$

...