Birkhoff's Ergodic Theorem

Theorem (Birkhoff's Ergodic Theorem; 1931)**.** *Let $(\mathcal{X}, \mathscr{F}, \mu)$ be a probability space and $T\colon \mathcal{X} \to \mathcal{X}$ a measurable map that preserves μ. For every $f \in L^1_\mu(\mathcal{X})$, the limit*

$$\overline{f}(x) \coloneqq \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f\big(T^k(x)\big)$$

exists for μ-almost every x. Furthermore,

(i) $\overline{f}\big(T(x)\big) = \overline{f}(x)$ *for μ-a.e. x.*

(ii) $\displaystyle\int \overline{f}\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$.

Several proofs of Birkhoff's theorem are available. We follow the one presented in Parry's book "Topics in ergodic theory" (1981). This proof is not the shortest but has the advantage that it follows the same pattern as the proof of von Neumann's theorem. Parry attributes this proof to Garsia (1970).

There are two main ingredients in the proof: one is a decomposition lemma similar to the one used for von Neumann's theorem; the other is a general inequality known as the maximal ergodic inequality.

Lemma (Decomposition lemma for L^1)**.** *The two linear subspace*

$$I_1 \coloneqq \big\{g \in L^1_\mu(\mathcal{X}) : g\circ T = g \ \mu\text{-a.e.}\big\}\ , \qquad\qquad B_1 \coloneqq \big\langle h\circ T - h : h \in L^1_\mu(\mathcal{X})\big\rangle\ ,$$

satisfy

$$L^1_\mu(\mathcal{X}) = \overline{B}_1 + I_1 \qquad \textit{and} \qquad \overline{B}_1 \cap I_1 = \{0\}\ .$$

In other words, every $f \in L^1_\mu(\mathcal{X})$ has a unique decomposition $f = f^* + f_0$ where $f^* \in I_1$ and $f_0 \in \overline{B}_1$.

Lemma (Maximal Ergodic Inequality: variant I)**.** *For every $f \in L^1_\mu(\mathcal{X})$ and $\lambda > 0$, we have*

$$\begin{aligned}&\mu\bigg(\Big\{x : \limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k(x)\Big| > \lambda\|f\|_1\Big\}\bigg) \\ &\qquad \leq \mu\bigg(\Big\{x : \sup_{n\geq 0}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k(x)\Big| > \lambda\|f\|_1\Big\}\bigg) \leq \frac{1}{\lambda}\ .\end{aligned}$$

We first present the proof of Birkhoff's theorem using these two lemmas and then come back to prove the lemmas.

Exercise (Induced isometries)**.** Verify that for every $1 \leq p \leq \infty$, the map $f \mapsto f\circ T$ is a positive linear isometry of $L^p_\mu(\mathcal{X})$.

Exercise (L^1 ergodic theorem)**.** Use the L^1 decomposition lemma to give a proof of the L^1 ergodic theorem. *Hint:* Mimic the proof of von Neumann's theorem.

Proof of Birkhoff's theorem. We first focus on the proof of the existence of the limit $\overline{f}$, although the argument will also imply (i). Proof of (ii) will be left as an exercise.

As in the proof of von Neumann's theorem, we start with some special cases:

Notes for Introduction to Ergodic Theory (MATH 307K), Siamak Taati
Last update: March 7, 2022

- If $f \in I_1$, then $\overline{f}$ trivially exists and $\overline{f} = f$.
- If $f = h \circ T - h$ for some $h \in L^\infty_\mu(\mathcal{X})$, then $\displaystyle\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to 0$ (uniformly) μ-a.e., hence $\overline{f} = 0$.

Argument. As in the L^2 case, we have

$$\Big\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\Big\|_\infty = \frac{1}{n}\|h \circ T^n - h\|_\infty \leq \frac{1}{n}\big(\|h \circ T^n\|_\infty + \|h\|_\infty\big) = \frac{2}{n}\|h\|_\infty \ .$$

It follows that $\displaystyle\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to 0$ in $L^\infty_\mu(\mathcal{X})$, and hence uniformly on a set $E \subseteq \mathcal{X}$ with $\mu(E) = 1$.

- If $f = c_1 f_1 + c_2 f_2$ for some $f_1, f_2 \in L^1_\mu(\mathcal{X})$ and the limits $\overline{f}_1$ and $\overline{f}_2$ exist μ-a.e., then $\overline{f}$ also exists μ-a.e. with $\overline{f} = c_1\overline{f}_1 + c_2\overline{f}_2$.
- If $f_m \in B_\infty$ and $f_m \to f$ in $L^1_\mu(\mathcal{X})$, then $\displaystyle\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to 0$ μ-a.e., hence $\overline{f} = 0$.

Argument. Without loss of generality, we may assume that $\|f - f_m\|_1 \to 0$ monotonically as $m \to \infty$.
For each m we have

$$\limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\Big| \leq \underbrace{\limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f_m \circ T^k\Big|}_{=\,0\ \mu\text{-a.e.}} + \limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} (f - f_m) \circ T^k\Big|$$

Applying the maximal inequality on $f - f_m$, we have

$$\mu\bigg(\Big\{x : \limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} (f - f_m) \circ T^k(x)\Big| > \lambda\|f - f_m\|_1\Big\}\bigg) \leq \frac{1}{\lambda} \ ,$$

from which it follows

$$\mu\bigg(\Big\{x : \limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x)\Big| > \lambda\|f - f_m\|_1\Big\}\bigg) \leq \frac{1}{\lambda} \ .$$

Letting $m \to \infty$ and using monotone continuity, we obtain

$$\mu\bigg(\Big\{x : \limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x)\Big| > 0\Big\}\bigg) \leq \frac{1}{\lambda} \ .$$

Letting $\lambda \to \infty$, we find

$$\mu\bigg(\Big\{x : \limsup_{n\to\infty}\Big|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x)\Big| = 0\Big\}\bigg) = 1 \ ,$$

as claimed.

Let

$$B_\infty := \big\langle h \circ T - h : h \in L^\infty_\mu(\mathcal{X})\big\rangle$$

and observe that B_∞ and B_1 have the same closures in $L^1_\mu(\mathcal{X})$. Hence, we also have the following:

- If $f \in \overline{B}_1 = \overline{B}_\infty$, then $\displaystyle\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to 0$ μ-a.e., hence $\overline{f} = 0$.

Now, by the L^1 decomposition lemma, every $f \in L^1_\mu(\mathcal{X})$ can be written as $f = f^* + f_0$ where $f^* \in I_1$ and $f_0 \in \overline{B}_1$. It follows that

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = \underbrace{\frac{1}{n}\sum_{k=0}^{n-1} f^* \circ T^k}_{f^*} + \underbrace{\frac{1}{n}\sum_{k=0}^{n-1} f_0 \circ T^k}_{\to 0} \to f^* \qquad \mu\text{-a.e. as } n \to \infty,$$

hence $\overline{f}$ exists μ-a.e. and satisfies $\overline{f} = f^*$. This also shows that (i) holds.

Exercise. Show that if $f = f^* + f_0$ with $f^* \in I_1$ and $f_0 \in \overline{B}_1$, then $\int f^* \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$.

This concludes the proof of Birkhoff's theorem. $\square$

Maximal ergodic inequality

The maximal ergodic inequality mentioned earlier is due to Wiener (1939) and Yosida and Kakutani (1939). It has the following variant due to Hopf (1954):

Lemma (Maximal Ergodic Inequality: variant II). *Let $f \in L^1_\mu(\mathcal{X})$. For $n, N \in \mathbb{N}$, define*

$$f_n := \begin{cases} f + f\circ T + \cdots + f \circ T^{n-1} & \text{if } n > 0, \\ 0 & \text{if } n = 0, \end{cases}$$
$$E_N := \left\{ x \in \mathcal{X} : \max_{0\leq n\leq N} f_n(x) > 0 \right\} .$$

Then, $\int_{E_N} f \,\mathrm{d}\mu \geq 0$.

Interpretation. Think of $(f_n)_{n\in\mathbb{N}}$ as a "random walk" on $\mathbb{R}$ starting from position 0, where the jump at time step k is given by the value of $f \circ T^{k-1}$. Let $F_N(x) := \max_{0\leq n\leq N} f_n(x)$ be the largest value reached within the first N steps. The latter lemma states that conditioned on $F_N > 0$, the expected value of f is non-negative.

Let us first show that the former variant of the maximal ergodic inequality follows from the latter.

Proof of variant I using variant II. The first inequality is obvious. To prove the second inequality, note that:

- Without loss of generality, we may assume that $\|f\|_1 = 1$. (If $\|f\|_1 = 0$, the claim holds trivially. Otherwise, we can scale f so as to get a function with norm 1.)
- Without loss of generality, we may assume that $f \geq 0$. (Otherwise, we can replace f with $|f|$, and this would not increase the left-hand side.)

Set $g := f - \lambda$. We apply variant II on g. Note that $g_n = f_n - n\lambda$, hence

$$E_N := \left\{ x \in \mathcal{X} : \max_{0\leq n\leq N} g_n(x) > 0 \right\} = \left\{ x \in \mathcal{X} : \max_{0\leq n\leq N} \frac{1}{n} f_n(x) > \lambda \right\} .$$

The inequality $\int_{E_N} g \,\mathrm{d}\mu \geq 0$ can be rewritten as

$$\lambda\mu(E_N) \leq \int_{E_N} f \,\mathrm{d}\mu . \qquad (✐)$$

Now, observe that $E_0 \subseteq E_1 \subseteq E_2 \subseteq \cdots$. Furthermore,

$$E := \bigcup_{N\geq 0} E_N = \left\{ x : \sup_{n\geq 0} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k > \lambda \right\} .$$

Letting $N \to \infty$ in (✐), and using monotone continuity and the facts that $f \geq 0$ and $\|f\|_1 = 1$, we get

$$\lambda\mu(E) \leq \int_E f \,\mathrm{d}\mu \leq \int f \,\mathrm{d}\mu = 1 .$$

Therefore, $\mu(E) \leq 1/\lambda$, as claimed. $\square$

Proof of variant II (Garsia, 1965). Let $F_N(x) := \max_{0\leq n\leq N} f_n(x)$. Note that $F_N \geq 0$ because $f_0 = 0$. Hence, $F_N(x) = 0$ for $x \in \mathcal{X} \setminus E_N$.

Claim. $f \geq F_N - F_N \circ T$ on E_N.

Argument. Note that

$$F_{N+1} = \max\{F_N, f_{N+1}\} = \max\{0, f + F_N \circ T\} \tag{☼}$$

Indeed, the first identity is clear. To see the latter, observe that

$$\begin{aligned} f + F_N \circ T &= f + \max_{0 \leq n \leq N} (f + f \circ T + \cdots f \circ T^{n-1}) \circ T \\ &= f + \max_{0 \leq n \leq N} (f \circ T + f \circ T^2 + \cdots f \circ T^n) = \max_{1 \leq m \leq N+1} f_m \end{aligned}$$

which implies $F_{N+1} = \max\{f_0, f + F_N \circ T\} = \max\{0, f + F_N \circ T\}$.

From (☼) it follows that either $0 \geq F_N$ or $f + F_N \circ T \geq F_N$. Hence, $f \geq F_N - F_N \circ T$ on E_N.

Integrating on E_N, we get

$$\begin{aligned} \int_{E_N} f \,\mathrm{d}\mu &\geq \int_{E_N} F_N \,\mathrm{d}\mu - \int_{E_N} F_N \circ T \,\mathrm{d}\mu \\ &\geq \int F_N \,\mathrm{d}\mu - \int F_N \circ T \,\mathrm{d}\mu && \text{(because } F_N = 0 \text{ outside } E_N\text{, and } F_N \circ T \geq 0) \\ &= 0 && \text{(because } T \text{ preserves } \mu) \end{aligned}$$

which concludes the proof of the lemma. □

Decomposition lemma

Recall that in the proof of the L^2 decomposition lemma, the component of an L^2 function f in the subspace I_2 of invariant functions was simply the orthogonal projection of f on I_2. For the decomposition of L^1, the role of the orthogonal projection is played by a conditional expectation with respect to the σ-algebra of invariant sets.

More specifically, let

$$\mathscr{I}_\mu := \big\{E \in \mathscr{F} : \mu\big(E \triangle T^{-1}E\big) = 0\big\}$$

be the family of all measurable sets that are invariant modulo μ.

Exercise. Verify that $\mathscr{I}_\mu$ is a sub-σ-algebra of $\mathscr{F}$.

Recall the linear subspaces

$$I_1 := \big\{g \in L^1_\mu(\mathcal{X}) : g \circ T = g \ \mu\text{-a.e.}\big\}\,, \qquad B_1 := \big\langle h \circ T - h : h \in L^1_\mu(\mathcal{X})\big\rangle\,.$$

The L^1 decomposition lemma follows from the following characterizations of I_1 and $\overline{B}_1$.

Lemma (Characterizations of I_1 and $\overline{B}_1$)**.** *Let $f \in L^1_\mu(\mathcal{X})$.*

(i) *$f \in I_1$ if and only if $\mu(f \,|\, \mathscr{I}_\mu) = f$ μ-a.e.*

(ii) *$f \in \overline{B}_1$ if and only if $\mu(f \,|\, \mathscr{I}_\mu) = 0$ μ-a.e.*

Proof of the L^1 decomposition lemma. Let $f \in L^1_\mu(\mathcal{X})$. Define $f^* := \mu(f \,|\, \mathscr{I}_\mu)$ and $f_0 := f - f^*$, so that $f = f^* + f_0$. Note that $f^*, f_0 \in L^1_\mu(\mathcal{X})$. Using the above characterizations, it is easy to see that $f^* \in I_1$ and $f_0 \in \overline{B}_1$. Furthermore, $f = \widehat{f}^* + \widehat{f}_0$ is another decomposition with $\widehat{f}^* \in I_1$ and $\widehat{f}_0 \in \overline{B}_1$, then $\widehat{f}^* = f^*$ and $\widehat{f}_0 = f_0$ μ-a.e. □

Exercise. Prove the above characterization of I_1.

Proof of the characterization of $\overline{B}_1$. (See the book of Parry for an alternative proof.)
[$\Longrightarrow$] Let us first prove that $\mu(f \,|\, \mathscr{I}_\mu) = 0$ μ-a.e. for every $f \in \overline{B}_1$.

- If $f = h \circ T - h$ for some $h \in L^1_\mu(\mathcal{X})$, then $\mu(f \,|\, \mathscr{I}_\mu) = \mu(h \circ T \,|\, \mathscr{I}_\mu) - \mu(h \,|\, \mathscr{I}_\mu) = 0$.

Argument. Clearly, $\mu(h \circ T \mid \mathscr{I}_\mu)$ is $\mathscr{I}_\mu$-measurable. Furthermore, for every $E \in \mathscr{I}_\mu$,

$$\begin{aligned}\int_E \mu(h \circ T \mid \mathscr{I}_\mu)\,\mathrm{d}\mu &= \int_E h \circ T\,\mathrm{d}\mu = \int \mathbb{1}_E \cdot (h \circ T)\,\mathrm{d}\mu \\ &= \int (\mathbb{1}_E \circ T) \cdot (h \circ T)\,\mathrm{d}\mu = \int \mathbb{1}_E h\,\mathrm{d}(T\mu) = \int_E h\,\mathrm{d}\mu\end{aligned}$$

hence $\mu(h \circ T \mid \mathscr{I}_\mu) = \mu(h \mid \mathscr{I}_\mu)$ by the uniqueness of the conditional expectation.

- The map $f \mapsto \mu(f \mid \mathscr{I}_\mu)$ is linear, hence the set of functions f with $\mu(f \mid \mathscr{I}_\mu) = 0$ (i.e., the kernel of $\mu(\cdot \mid \mathscr{I}_\mu)$) is a linear subspace of $L^1_\mu(\mathcal{X})$.

It follows that $\mu(f \mid \mathscr{I}_\mu) = 0$ μ-a.e. for every $f \in B_1$. Furthermore:

- If $f_m \in B_1$ and $f_m \to f$ in $L^1_\mu(\mathcal{X})$, then $\mu(f \mid \mathscr{I}_\mu) = 0$ μ-a.e.

Argument. Given $E \in \mathscr{I}_\mu$, we have

$$\int_E \mu(f \mid \mathscr{I}_\mu)\,\mathrm{d}\mu = \int_E f\,\mathrm{d}\mu = \overbrace{\int_E f_m\,\mathrm{d}\mu}^{=0} + \int_E (f - f_m)\,\mathrm{d}\mu$$

Hence,

$$\begin{aligned}\left|\int_E \mu(f \mid \mathscr{I}_\mu)\,\mathrm{d}\mu\right| &= \left|\int_E (f - f_m)\,\mathrm{d}\mu\right| \\ &\leq \int_E |f - f_m|\,\mathrm{d}\mu \leq \int |f - f_m|\,\mathrm{d}\mu = \|f - f_m\|_1\ .\end{aligned}$$

Since $\|f - f_m\|_1 \to 0$ as $m \to \infty$, we find that $\int_E \mu(f \mid \mathscr{I}_\mu) = 0$. It follows from the uniqueness of conditional expectation that $\mu(f \mid \mathscr{I}_\mu) = 0$ μ-a.e.

We conclude that $\mu(f \mid \mathscr{I}_\mu) = 0$ μ-a.e. for every $f \in \overline{B}_1$ as claimed.

[$\Longleftarrow$] Next, we prove that if $f \in L^1_\mu(\mathcal{X})$ is such that $\mu(f \mid \mathscr{I}_\mu) = 0$ μ-a.e., then $f \in \overline{B}_1$.

We prove the contrapositive. Let $f \in L^1_\mu(\mathcal{X}) \setminus \overline{B}_1$. Then, every $u \in \langle \overline{B}_1, f\rangle$ has a unique representation $u = u_0 + \alpha f$ where $u_0 \in \overline{B}_1$ and $\alpha \in \mathbb{R}$. Define a linear functional

$$\begin{aligned}J\colon \langle \overline{B}_1, f\rangle &\longrightarrow \mathbb{R} \\ u = u_0 + \alpha f &\longmapsto \alpha\ .\end{aligned}$$

Note that J is continuous.

Argument. Let $\delta \coloneqq d(\overline{B}_1, f)$ be the distance between f and $\overline{B}_1$. Since $\overline{B}_1$ is closed, $\delta > 0$. Therefore, for every $u = u_0 + \alpha f \in \langle \overline{B}_1, f\rangle$,

$$\|u\|_1 = \|u_0 + \alpha f\|_1 = |\alpha| \cdot \|f - (-1/\alpha)u\|_1 \geq |\alpha|\delta = |J(u)|\delta\ .$$

It follows that J is continuous.

Therefore, by the Hahn–Banach theorem, J can be extended to a continuous linear functional $\widehat{J}\colon L^1_\mu(\mathcal{X}) \to \mathbb{R}$. Since $L^\infty_\mu(\mathcal{X})$ is the dual of $L^1_\mu(\mathcal{X})$, there exists a $g \in L^\infty_\mu(\mathcal{X})$ such that

$$J(u) = \int ug\,\mathrm{d}\mu$$

for every $u \in L^1_\mu(\mathcal{X})$.

Now, note that

- $g \circ T = g$ μ-a.e.

Argument. For every $h \in L^1_\mu(\mathcal{X})$, we have $0 = J(h \circ T - h) = \int (h \circ T - h) g \, \mathrm{d}\mu$ hence

$$\int (h \circ T) g \, \mathrm{d}\mu = \int hg \, \mathrm{d}\mu \, . \tag{$\circledast$}$$

In particular,

$$\int (g \circ T - g)^2 \, \mathrm{d}\mu = \int (g \circ T)^2 \, \mathrm{d}\mu - 2 \int (g \circ T) g \, \mathrm{d}\mu + \int g^2 \, \mathrm{d}\mu = 0$$

(using ($\circledast$) and the invariance of μ), which implies $g \circ T = g$ μ-a.e.

- $\int fg \, \mathrm{d}\mu = J(f) = 1$.

Therefore, $\mu(f \mid \mathscr{I}_\mu) \neq 0$.

Argument. Since $g \circ T = g$, the function g is $\mathscr{I}_\mu$-measurable. Therefore,

$$\int \mu(f \mid \mathscr{I}_\mu) g \, \mathrm{d}\mu = \int \mu(fg \mid \mathscr{I}_\mu) \, \mathrm{d}\mu = \int fg \, \mathrm{d}\mu = 1 \, .$$

This concludes the proof. □

Back to the ergodic theorem

The above proof allows us to rephrase Birkhoff's theorem in the following more concise form:

Theorem (Birkhoff's Ergodic Theorem; rephrased)**.** *Let $(\mathcal{X}, \mathscr{F}, \mu)$ be a probability space and $T \colon \mathcal{X} \to \mathcal{X}$ a measurable map that preserves μ. For every $f \in L^1_\mu(\mathcal{X})$,*

$$\frac{1}{n} \sum_{k=0}^{n-1} f\big(T^k(x)\big) \to \mu(f \mid \mathscr{I}_\mu) \qquad \mu\text{-a.e. as } n \to \infty.$$

Exercise (Characterization of $\overline{f}$)**.** Without relying on the decomposition lemma, verify that the function $\overline{f}$ in the original statement of Birkhoff's theorem is $\mu(f \mid \mathscr{I}_\mu)$.