Birkhoff's Ergodic Theorem

(Proof 2)

We present an alternative proof of Birkhoff's 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*

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

exists for μ-almost every x. Furthermore,

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

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

We use the notation f^* instead of $\overline{f}$ to reserve $\overline{f}$ and $\underline{f}$ for the limit superior and limit inferior of the ergodic averages.

We present the proof proposed by Katznelson and Weiss (1982). Katznelson and Weiss attributed the idea of their proof to Kamae (1982), who gave a proof of Birkhoff's theorem using nonstandard analysis. A similar proof was later given by Shields (1987) in the language of probability theory.

The proof follows the same general pattern as the proof of the ergodic theorem in the elementary case of finite-state dynamical systems (exercise) and the regeneration-based proof of the ergodic theorem of positive recurrent Markov chains. Namely, we break the orbit into segments over each of which we can control the average of f. To illustrate this idea, we first present the proof in the special case in which f is bounded, and then sketch how it can be adapted to cover arbitrary integrable function. The complete proof in the general case will be left as an exercise.

Proof of Birkhoff's theorem for bounded functions. Let $M\in\mathbb{R}_+$ be such that $f(x)\leq M$ for every $x\in\mathcal{X}$. For $x\in\mathcal{X}$, let

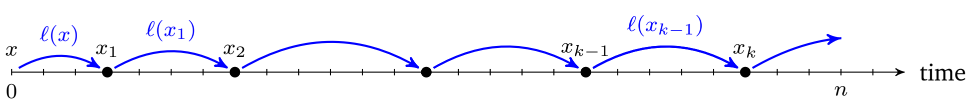
$$\overline{f}(x) \coloneqq \limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f\big(T^i(x)\big)\,, \qquad\qquad \underline{f}(x) \coloneqq \liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f\big(T^i(x)\big)\,.$$

It is easy to verify that $\overline{f}\big(T(x)\big)=\overline{f}(x)$ and $\underline{f}\big(T(x)\big)=\underline{f}(x)$ for every $x\in\mathcal{X}$ (exercise). In other words, $\overline{f}$ and $\underline{f}$ are constant over the entire orbit of x.

We estimate $\overline{f}_n(x)\coloneqq\frac{1}{n}\sum_{i=0}^{n-1} f\big(T^i(x)\big)$ in terms of $\overline{f}(x)$. Let $\varepsilon>0$. For each $x\in\mathcal{X}$, define

$$\ell(x)\coloneqq \inf\big\{\ell\in\mathbb{Z}_+ : \overline{f}_\ell(x) > \overline{f}(x)-\varepsilon\big\}\,,$$

and note that $\ell(x)<\infty$. The idea is to break the orbit of x over the time interval $[\![0,n-1]\!]$ into blocks as depicted in the following figure:[1]



Notes for Introduction to Ergodic Theory (MATH 307K), Siamak Taati

Last update: February 23, 2025

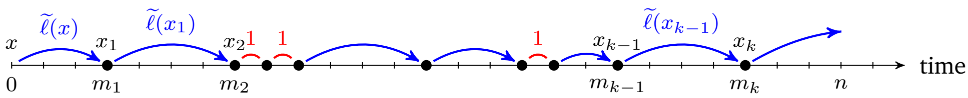
[1]We use the notation $[\![a,b]\!]\coloneqq[a,b]\cap\mathbb{Z}$ for integer intervals.

The average of f over each block, with the exception of the incomplete one at the end, will be larger than $\overline{f}(x)-\varepsilon$. We may hope that the contribution of the last incomplete block remains relatively small (i.e., of order $o(n)$), so that, in the limit, we obtain $\underline{f}(x) \geq \overline{f}(x)-\varepsilon$. However, $\ell(\cdot)$ is not bounded, hence there is no immediate reason for the contribution of the last incomplete black to remain small.

To circumvent this, let L be large and define

$$\widetilde{\ell}(x) \coloneqq \begin{cases} \ell(x) & \text{if } \ell(x) \leq L, \\ 1 & \text{if } \ell(x) > L. \end{cases}$$

Instead of $\ell(\cdot)$, we use $\widetilde{\ell}(\cdot)$ to break the orbit of x into blocks, so that the length of each block is at most L.



This ensures that the contribution of the last incomplete block remains bounded, but introduces some blocks on which the average is not necessarily larger than $\overline{f}(x)-\varepsilon$ (the red ones in the above figure). Nevertheless, we expect that, with high probability, such blocks form only a small fraction of the interval $[\![0, n-1]\!]$.

Let us make this idea more precise. Let $n \in \mathbb{Z}_+$. Given $x \in \mathcal{X}$, let $m_0 \coloneqq 0$ and $x_0 \coloneqq x$, and for $j = 0, 1, 2, \ldots$, recursively define $\ell_{j+1} \coloneqq \widetilde{\ell}(x_j)$, $m_{j+1} \coloneqq m_j + \ell_{j+1}$ and $x_{j+1} \coloneqq T^{\ell_{j+1}}(x_j)$. Let $k \coloneqq \max\{j \in \mathbb{N} : m_j \leq n\}$. The interval $[\![0, n-1]\!]$ is partitioned into k (complete) blocks $[\![m_{j-1}, m_j - 1]\!]$ with $j = 1, 2, \ldots, k$ and an incomplete block $[\![m_k, n-1]\!]$. Let

$$E_L \coloneqq \{x \in \mathcal{X} : \ell(x) > L\} \ .$$

We say that the jth block is *proper* if $x_{j-1} \notin E_L$ and *skipped* otherwise. In the above figure, the proper blocks are depicted by blue and the skipped blocks by red.

We consider the three types of blocks (proper, skipped, incomplete) separately:

- (Proper blocks) If the jth block is proper, then by definition, we have $\overline{f}_{\ell_j}(x_{j-1}) > \overline{f}(x_{j-1}) - \varepsilon$. Multiplying both sides by ℓ_j, and using the fact that $\overline{f}$ is constant over the orbit of x, we can rewrite the latter as

$$\sum_{i=m_{j-1}}^{m_j-1} f\big(T^i(x)\big) > \sum_{i=m_{j-1}}^{m_j-1} \overline{f}\big(T^i(x)\big) - \varepsilon \ell_j \ . \tag{$\clubsuit_1$}$$

- (Skipped blocks) If the jth block is skipped, then $\ell_j = 1$. In this case, we have the bound

$$f\big(T^{m_{j-1}}(x)\big) \geq \overline{f}\big(T^{m_{j-1}}(x)\big) - 2M \tag{$\clubsuit_2$}$$

  because $f(\cdot) \geq -M$ and $\overline{f}(\cdot) \leq M$.

- (Incomplete block) The length of the incomplete block is bounded by L, hence

$$\sum_{i=m_k}^{n} f\big(T^i(x)\big) \geq \sum_{i=m_k}^{n} \overline{f}\big(T^i(x)\big) - 2ML \ . \tag{$\clubsuit_3$}$$

Combining ($\clubsuit_1$), ($\clubsuit_2$), and ($\clubsuit_3$), it follows that

$$\sum_{i=0}^{n-1} f\big(T^i(x)\big) \geq \sum_{i=0}^{n-1} \overline{f}\big(T^i(x)\big) - \varepsilon n - 2M \sum_{i=0}^{n-1} \mathbb{1}_{E_L}\big(T^i(x)\big) - 2ML \ .$$

Dividing by n, and using again the T-invariance of $\overline{f}$, we obtain

$$\overline{f}_n(x) \geq \overline{f}(x) - \varepsilon - \frac{2M}{n} \sum_{i=0}^{n-1} \mathbb{1}_{E_L}\big(T^i(x)\big) - \frac{2ML}{n} \ . \tag{$\diamondsuit$}$$

As $n \to \infty$, the last term on the right-hand side goes to 0, but it is not clear how to bound the third term.

To exploit ($\diamondsuit$), we integrate both sides with respect to μ to get

$$\int \overline{f}_n(x)\,\mathrm{d}\mu(x) \geq \int \overline{f}(x)\,\mathrm{d}\mu(x) - \varepsilon - \frac{2M}{n}\sum_{i=0}^{n-1} \underbrace{\int \mathbb{1}_{E_L}\big(T^i(x)\big)\,\mathrm{d}\mu(x)}_{(T^i\mu)(E_L)} - \frac{2ML}{n}\,,$$

which, using the T-invariance of μ, simplifies to

$$\int f\,\mathrm{d}\mu \geq \int \overline{f}\,\mathrm{d}\mu - \varepsilon - 2M\mu(E_L) - \frac{2ML}{n}\,.$$

Letting first $n \to \infty$, next $L \to \infty$, and then $\varepsilon \searrow 0$ yields

$$\int f\,\mathrm{d}\mu \geq \int \overline{f}\,\mathrm{d}\mu\,.$$

By symmetry (i.e., applying the same conclusion to $-f$), we also have $\int f\,\mathrm{d}\mu \leq \int \underline{f}\,\mathrm{d}\mu$, hence

$$\int \underline{f}\,\mathrm{d}\mu \geq \int f\,\mathrm{d}\mu \geq \int \overline{f}\,\mathrm{d}\mu\,.$$

Since $\underline{f}(x) \leq \overline{f}(x)$ for every x, we conclude that $\overline{f} = \underline{f}$ and

$$\int \overline{f}\,\mathrm{d}\mu = \int \underline{f}\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu\,.$$

In other words, the limit in ($\varnothing$) exists, the function f^* is T-invariant, and we have

$$\int f^*\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$$

as claimed. $\square$

Birkhoff's theorem for general integrable functions can be proven by adapting the above argument. Using decomposition into positive and negative parts, we can assume that f is non-negative. Since the symmetry is lost, we must consider $\overline{f}$ and $\underline{f}$ separately.

- To deal with $\overline{f}$, we first truncate it as $\overline{f} \wedge M$, repeat the same argument, and then let $M \to \infty$. Similar bounds for the skipped and incomplete blocks still hold because $f \geq 0$ and $\overline{f} \wedge M \leq M$.
- To deal with $\underline{f}$, truncation is not needed because $\underline{f} \geq 0$. Here, we do not have an upper bound on f, but the integral $\int_{E_L} f\,\mathrm{d}\mu$ still goes to 0 as $L \to \infty$.

Exercise. Prove Birkhoff's theorem for general integrable functions by following the above sketch (or any other approach).

Remark. Here, and in other proofs of the ergodic theorem, assuming f to be bounded makes the proof significantly less technical. It would therefore be convenient to have a separate argument reducing the general case to the bounded case. See this MathOverflow post[2] for such a reduction attributed to Furstenberg and Weiss. $\diamond$

[2] https://mathoverflow.net/a/406846/23297