THE BIRKHOFF ERGODIC THEOREM

ALESSIO DEL VIGNA

Ergodic theory studies the properties of dynamical systems which hold for the orbits of almost every initial conditions with respect to measures that remain invariant under time evolution. In this note we present one of the most important results in this field, the Birkhoff ergodic theorem.

1. Motivations

We first introduce some notation. Throughout the note, $(M, \mathfrak{B}, \mu)$ is a measure space and $f : M \to M$ a measurable transformation.

Definition 1. The measure μ is *invariant* under f if $\mu(E) = \mu(f^{-1}E)$ for every $E \in \mathfrak{B}$.

If μ is finite, Poincaré recurrence theorem states that almost every point in any positive-measure set E returns to E an infinite number of times. Furthermore, when the system is ergodic, Kač theorem says that the mean return time to E is inversely proportional to the measure of E. We now recall the definition of ergodicity.

Definition 2. If μ is a probability measure, we say that f is *ergodic* with respect to the measure μ if for all measurable sets $B \in \mathfrak{B}$ such that $f^{-1}B \subseteq B$ it holds that $\mu(B) = 0$ or $\mu(B) = 1$.

Given an arbitrary $x \in M$, we consider $\left\{j \geq 0 \,:\, f^j(x) \in E\right\}$, the set of iterates of x which visit the set E. Another way of stating Poincaré recurrence is that this set is infinite. We would like to give more precise quantitative information. Let

$$\tau(E, x) \coloneqq \lim_{n\to\infty} \frac{1}{n} \# \left\{0 \leq j < n \,:\, f^j(x) \in E\right\}.$$

When this limit exists we call it the *mean sojourn time* of x in the set E. A convenient way to write it is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_E(f^j(x)).$$

Birkhoff ergodic theorem says that the mean sojourn time exists for μ-a.e. initial condition $x \in M$ and, additionally, that if μ is ergodic then $\tau(E, x) = \mu(E)$. Thus the orbit of almost every point of M enters the set E with asymptotic frequency $\mu(E)$.

Example 1. In general the mean sojourn time does not exist for every initial condition. Consider $M = [0, 1]$ and $f : M \to M$ defined to be

$$f(x) = 10x - \lfloor 10x \rfloor,$$

which preserves the Lebesgue measure m on M. Let

$$E = \left[0, \frac{1}{10}\right) = \{x \in M \,:\, x = 0.0x_2x_3 \dots\}.$$

In other words, E is the set of numbers in $[0,1)$ whose decimal expansion starts with a 0. Consider the initial condition $x \in (0,1)$ whose decimal expansion $x = 0.x_1x_2x_3\dots$ is such that $x_i = 1$ if there exists $k \geq 0$ even such that $2^k \leq i < 2^{k+1}$ and $x_i = 0$ otherwise. That is

$$x = 0.100111100000000111111111111110\dots$$

where the lengths of the alternating blocks of 0's and 1's are the successive powers of 2. If $n = 2^k - 1$ with $k \geq 1$ even then $\frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_E(f^j(x)) = \frac{\sum_{j=1}^{k-1} 2^j}{2^k-1} \to \frac{2}{3}$ as $k \to \infty$. Instead if $n = 2^k - 1$ with $k \geq 1$ odd we have $\frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}_E(f^j(x)) \to \frac{1}{3}$ as $k \to \infty$. Thus the mean sojourn time of x in E does not exist.

2. The theorem

We defined the mean sojourn time of x in a positive-measure set E as the asymptotic mean of the values of the characteristic function of E along the orbit of x. More in general, we can replace $\mathbb{1}_E$ with a measurable function $\varphi \in L^1(\mu)$ and consider the asymptotic behaviour of

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)),$$

which is called the *time average* of φ along the orbit of x. Birkhoff ergodic theorem states that the time average exists for μ-a.e. initial condition and that, if f is also ergodic, it is equal to $\frac{1}{\mu(M)}\int \varphi \, \mathrm{d}\mu$, the so-called *space average* of φ.

We now turn to the theorem and its proof, following [1]. Without further specifications, here we assume the measure μ being *σ-finite*, that is there exists a countable collection $\{A_n\}_{n\geq 1}$ of measurable sets such that $\bigcup_{n=1}^{\infty} A_n = M$ and $\mu(A_n) < \infty$ for all n.

Theorem 3 (maximal ergodic theorem)**.** *Let $(M, \mathfrak{B}, \mu)$ be a measure space and $f : M \to M$ a measure-preserving transformation. Let $U : L^1(\mu) \to L^1(\mu)$ be a positive linear operator such that $\|U\| \leq 1$. Given $\psi \in L^1_{\mathbb{R}}(\mu)$ define*

$$\psi_0 = 0, \quad \psi_n = \psi + U\psi + \dots + U^{n-1}\psi \quad \text{for } n \geq 1.$$

For $N \geq 0$ integer let $\Psi_N = \max_{0\leq n\leq N} \psi_n$ and $E_N = \{x \in M : \Psi_N(x) > 0\}$. Then

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

Proof. It is clear that $\Psi_N \in L^1_{\mathbb{R}}(\mu)$. For $0 \leq n \leq N$ we have $\Psi_N \geq \psi_n$, and by the positivity of U follows

$$U\Psi_N \geq U\psi_n = U\psi + \dots + U^n\psi,$$

so that $U\Psi_N + \psi \geq \psi_{n+1}$. If $x \in E_N$ we have

$$U\Psi_N(x) + \psi(x) \geq \max_{1\leq n\leq N} \psi_n(x) \overset{\Psi_N(x)>0}{=} \max_{0\leq n\leq N} \psi_n(x) = \Psi_N(x).$$

Thus $\psi \geq \Psi_N - U\Psi_N$ on E_N, from which

$$\int_{E_N} \psi \, \mathrm{d}\mu \geq \int_{E_N} \Psi_N \, \mathrm{d}\mu - \int_{E_N} U\Psi_N \, \mathrm{d}\mu = \int_M \Psi_N \, \mathrm{d}\mu - \int_{E_N} U\Psi_N \, \mathrm{d}\mu, \tag{1}$$

where the last equality holds because $\Psi_N = 0$ on $M \setminus E_N$. Since by definition $\Psi_N \geq 0$, the positivity of U implies $U\Psi_N \geq 0$, so that $\int_{E_N} U\Psi_N \,\mathrm{d}\mu \leq \int_M U\Psi_N \,\mathrm{d}\mu$. Using (1) we have

$$\int_{E_N} \psi \,\mathrm{d}\mu \geq \int_M \Psi_N \,\mathrm{d}\mu - \int_M U\Psi_N \,\mathrm{d}\mu$$

and the right hand side is ≥ 0 because $\|U\| \leq 1$, as we wanted to prove. □

Let $(M, \mathfrak{B}, \mu)$ be a measure space, $f : M \to M$ a measure-preserving transformation. The *Koopman operator* (or composition operator) is the linear operator

$$U_f : L^1(\mu) \to L^1(\mu), \quad U_f(\varphi) \coloneqq \varphi \circ f.$$

Note that U_f is a linear isometry of $L^1(\mu)$: since μ is preserved by f we have

$$\|U_f(\varphi)\|_1 = \int |U_f(\varphi)| \,\mathrm{d}\mu = \int |\varphi| \circ f \,\mathrm{d}\mu = \int |\varphi| \,\mathrm{d}\mu = \|\varphi\|_1 .$$

Moreover, U_f is a positive linear operator, that is $U_f(\varphi) \geq 0$ μ-a.e. for every $\varphi \geq 0$. The next corollary is the application of the maximal ergodic theorem to U_f, obtaining a fundamental inequality.

Corollary 4. *Let $(M, \mathfrak{B}, \mu)$ be a measure space, $f : M \to M$ a measure-preserving transformation, and $g \in L^1_{\mathbb{R}}(\mu)$. For $\alpha \in \mathbb{R}$ set*

$$B_\alpha = \left\{ x \in M \,:\, \sup_{n\geq 1} \frac{1}{n} \sum_{j=0}^{n-1} g(f^j(x)) > \alpha \right\}.$$

Then for all invariant subset A with $\mu(A) < \infty$ we have

$$\int_{B_\alpha \cap A} g \,\mathrm{d}\mu \geq \alpha \cdot \mu(B_\alpha \cap A).$$

Proof. We first prove the result assuming that M has finite measure and $A = M$. Let $\psi = g - \alpha$. Following the notation of the maximal ergodic theorem applied to the operator U_f we have $\psi_0 = 0$ and for $n \geq 1$

$$\psi_n = \sum_{j=0}^{n-1} U_f^j \psi = \sum_{j=0}^{n-1} \psi \circ f^j,$$

and $E_n = \{x \in M \,:\, \Psi_n(x) > 0\}$. Then

$$B_\alpha = \bigcup_{n=1}^{\infty} \left\{ x \in M \,:\, \sum_{j=0}^{n-1} g(f^j(x)) > n\alpha \right\} = \bigcup_{n=1}^{\infty} \{x \in M \,:\, \psi_n(x) > 0\} = \bigcup_{n=1}^{\infty} E_n$$

since $\psi_n(x) > 0$ for some $n \geq 1$ implies $\Psi_n(x) > 0$ and $\Psi_n(x) > 0$ for some $n \geq 1$ implies $\psi_j(x) > 0$ for some $1 \leq j \leq n$. From the maximal ergodic theorem we have $\int_{E_n} \psi \,\mathrm{d}\mu \geq 0$ for all $n \geq 1$. Since $E_n \subseteq E_{n+1}$ for all $n \geq 1$, we have $\mathbb{1}_{E_n} \to \mathbb{1}_{B_\alpha}$ for $n \to \infty$. Moreover, $|\psi \mathbb{1}_{E_n}| \leq |\psi|$ and thus from the dominated convergence theorem we have

$$\int_{E_n} \psi \,\mathrm{d}\mu = \int \psi \mathbb{1}_{E_n} \,\mathrm{d}\mu \to \int \psi \mathbb{1}_{B_\alpha} \,\mathrm{d}\mu = \int_{B_\alpha} \psi \,\mathrm{d}\mu \quad \text{as } n \to \infty.$$

Therefore $\int_{B_\alpha} \psi \,\mathrm{d}\mu \geq 0$ and the thesis follows by recalling that $\psi = g - \alpha$. In the general case it suffices to apply this argument to $f|_A$ to get $\int_{B_\alpha \cap A} g \,\mathrm{d}\mu \geq \alpha \cdot \mu(B_\alpha \cap A)$. □

We are now ready to prove the Birkhoff ergodic theorem.

Theorem 5 (Birkhoff ergodic theorem)**.** *Let $(M, \mathfrak{B}, \mu)$ be a measure space, $f : M \to M$ a measure-preserving transformation, and $\varphi \in L^1(\mu)$. Then*

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x))$$

converges pointwise for μ-a.e. $x \in M$ to a function $\varphi^* \in L^1(\mu)$. Also $\varphi^* \circ f = \varphi^*$ μ-a.e. and if $\mu(M) < \infty$ then

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

Proof. We first assume that $\mu(M) < \infty$. By considering real and imaginary parts separately, we can assume $\varphi : M \to \mathbb{R}$. Let

$$\varphi^*(x) = \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \quad \text{and} \quad \varphi_*(x) = \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)),$$

which are defined for all $x \in M$ and are measurable functions. We divide the argument in the proof of the following properties:

(i) φ^* and φ_* are f-invariant;

(ii) $\varphi^* = \varphi_*$ μ-a.e.;

(iii) $\varphi^* \in L^1(\mu)$;

(iv) $\int \varphi^* \, \mathrm{d}\mu = \int \varphi \, \mathrm{d}\mu$.

For (i), we first note that the following identity holds:

$$\frac{n+1}{n}\left(\frac{1}{n+1}\sum_{j=0}^{n} \varphi(f^j(x))\right) - \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(f(x))) = \frac{1}{n}\varphi(x).$$

Then by taking the limit superior and the limit inferior we get $\varphi^* \circ f = \varphi^*$ and $\varphi_* \circ f = \varphi_*$, respectively.

We now show (ii). For real numbers α and β let $E_{\alpha,\beta} = \{x \in M : \varphi_*(x) < \beta \text{ and } \varphi^*(x) > \alpha\}$. Since

$$\{x \in M : \varphi_*(x) < \varphi^*(x)\} = \bigcup_{\substack{\beta<\alpha \\ \alpha,\beta\in\mathbb{Q}}} E_{\alpha,\beta}$$

it is suffices to prove that $\mu(E_{\alpha,\beta}) = 0$ for all $\beta < \alpha$ (note that the above union is countable). Clearly $f^{-1}E_{\alpha,\beta} = E_{\alpha,\beta}$ and if we set

$$B_\alpha = \left\{x \in M : \sup_{n\geq 1} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) > \alpha\right\}$$

we have $E_{\alpha,\beta} \subseteq B_\alpha$. From Corollary 4 we get

$$\int_{E_{\alpha,\beta}} \varphi \, \mathrm{d}\mu = \int_{B_\alpha \cap E_{\alpha,\beta}} \varphi \, \mathrm{d}\mu \geq \alpha \cdot \mu(E_{\alpha,\beta}).$$

If we replace φ, α, β with $-\varphi$, $-\beta$, $-\alpha$ respectively, since $(-\varphi)^* = -\varphi_*$ and $(-\varphi)_* = \varphi^*$ we also get the inequality $\int_{E_{\alpha,\beta}} \varphi \,\mathrm{d}\mu \leq \beta \cdot \mu(E_{\alpha,\beta})$. Therefore $\alpha \cdot \mu(E_{\alpha,\beta}) \leq \beta \cdot \mu(E_{\alpha,\beta})$ and if $\beta < \alpha$ we get $\mu(E_{\alpha,\beta}) = 0$. Note that here we also used that $\mu(M) < \infty$. This gives $\varphi^* = \varphi_*$ almost everywhere with respect to μ.

Property (iii), that is $\varphi^* \in L^1(\mu)$, is a simple consequence of the Fatou Lemma. Let $(f_n)_{n\geq 1}$ be the sequence of functions given by $f_n(x) = \left|\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x))\right|$. They are non-negative measurable functions and $\int f_n \,\mathrm{d}\mu \leq \int |\varphi| \,\mathrm{d}\mu < \infty$, so that

$$\int |\varphi^*| \,\mathrm{d}\mu = \int \lim_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu < \infty.$$

We are left to show $\int \bar{\varphi} \,\mathrm{d}\mu = \int \varphi \,\mathrm{d}\mu$, which is (iv). To this end, for k and $n \geq 1$ integers, define

$$D_{n,k} = \left\{ x \in M \,:\, \frac{k}{n} \leq \varphi^*(x) < \frac{k+1}{n} \right\}.$$

The sets $D_{n,k}$ are measurable and invariant under f, and moreover for all $\varepsilon > 0$ we have $D_{k,n} \subseteq B_{\frac{k}{n}-\varepsilon}$. From Corollary 4 we have

$$\int_{D_{n,k}} \varphi \,\mathrm{d}\mu = \int_{B_{\frac{k}{n}-\varepsilon} \cap D_{n,k}} \varphi \,\mathrm{d}\mu \geq \left(\frac{k}{n} - \varepsilon\right) \mu(D_{n,k})$$

and since ε is arbitrary it follows that $\int_{D_{n,k}} \varphi \,\mathrm{d}\mu \geq \frac{k}{n}\mu(D_{n,k})$. Then

$$\int_{D_{n,k}} \varphi^* \,\mathrm{d}\mu \leq \frac{k+1}{n}\mu(D_{n,k}) \leq \frac{1}{n}\mu(D_{n,k}) + \int_{D_{n,k}} \varphi \,\mathrm{d}\mu$$

and summing over k yields

$$\int \varphi^* \,\mathrm{d}\mu \leq \frac{1}{n}\mu(M) + \int \varphi \,\mathrm{d}\mu.$$

This holds for every $n \geq 1$, so that $\int \varphi^* \,\mathrm{d}\mu \leq \int \varphi \,\mathrm{d}\mu$. The same argument applied to $-\varphi$ gives $\int (-\varphi)^* \,\mathrm{d}\mu \leq \int -\varphi \,\mathrm{d}\mu$, which is the same as $\int \varphi_* \,\mathrm{d}\mu \geq \int \varphi \,\mathrm{d}\mu$. Since $\varphi^* = \varphi_*$ μ-a.e. we can conclude that the two integrals do coincide. This finishes the proof when $\mu(M) < \infty$.

For the case $\mu(M) = \infty$ the above proof is still valid if we prove that $\mu(E_{\alpha,\beta}) < \infty$ when $\beta < \alpha$, so that we are allowed to apply Corollary 4. Suppose $\alpha > 0$ and let $C \in \mathfrak{B}$ be such that $C \subseteq E_{\alpha,\beta}$ and $\mu(C) < \infty$. Such a set exists because we are assuming M to be σ-finite. By letting $\psi = \varphi - \alpha \mathbb{1}_C \in L^1(\mu)$, the maximal ergodic theorem yields

$$\int_{\{x\in M \,:\, \Psi_n(x)>0\}} (\varphi - \alpha \mathbb{1}_C) \,\mathrm{d}\mu \geq 0 \tag{2}$$

for all $n \geq 1$. Thus

$$\alpha \cdot \mu(C \cap \{x \in M \,:\, \Psi_n(x) > 0\}) \overset{(2)}{\leq} \int_{\{x\in M \,:\, \Psi_n(x)>0\}} \varphi \,\mathrm{d}\mu \leq \int_{\{x\in M \,:\, \Psi_n(x)>0\}} |\varphi| \,\mathrm{d}\mu \leq \int |\varphi| \,\mathrm{d}\mu$$

We claim that $E_{\alpha,\beta} \subseteq \bigcup_{n=0}^{\infty} \{x \in M \,:\, \Psi_n(x) > 0\}$. If $x \in E_{\alpha,\beta}$ then there is an n (actually infinitely many) such that $\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) > \alpha$, so that with the notation of the maximal ergodic theorem

$$\psi_n(x) = \sum_{j=0}^{n-1} \varphi(f^j(x)) - \alpha \sum_{j=0}^{n-1} \mathbb{1}_C(f^j(x)) \geq \sum_{j=0}^{n-1} \varphi(f^j(x)) - \alpha n > 0,$$

which implies $\Psi_n(x) > 0$. So we have the inclusion

$$C \subseteq E_{\alpha,\beta} \subseteq \bigcup_{n=0}^{\infty} \{x \in M \,:\, \Psi_n(x) > 0\}.$$

The sequence of sets $(C \cap \{x \in M \,:\, \Psi_n(x) > 0\})_{n\geq 1}$ is non-decreasing and the union of these sets is C. Thus $\mu(C \cap \{x \in M \,:\, \Psi_n(x) > 0\}) \to \mu(C)$ as $n \to \infty$, and hence

$$\alpha \cdot \mu(C) \leq \int |\varphi| \, \mathrm{d}\mu.$$

We thus proved that $\mu(C) \leq \frac{1}{\alpha} \int |\varphi| \, \mathrm{d}\mu$ for each $C \in \mathfrak{B}$ with $C \subseteq E_{\alpha,\beta}$ and $\mu(C) < \infty$. Since M is σ-finite, we can write $E_{\alpha,\beta}$ as the union of a non-decreasing sequence $(C_n)_{n\geq 1}$ of finite-measure sets. Since for each n we have $\mu(C_n) \leq \frac{1}{\alpha} \int |\varphi| \, \mathrm{d}\mu$, it follows that $\mu(E_{\alpha,\beta}) \leq \frac{1}{\alpha} \int |\varphi| \, \mathrm{d}\mu < \infty$. If $\alpha \leq 0$ then $\beta < 0$ and we can apply the argument to $-\varphi$ and $-\beta$ instead of φ and α to get $\mu(E_{\alpha,\beta}) < \infty$. □

Corollary 6. *Let $(M, \mathfrak{B}, \mu)$ be a finite measure space, $f : M \to M$ a measure-preserving transformation, and $\varphi \in L^1(\mu)$. The map f is ergodic if and only if for μ-a.e. $x \in M$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \frac{1}{\mu(M)} \int \varphi \, \mathrm{d}\mu.$$

Proof. ($\Rightarrow$) From the Birkhoff ergodic theorem we know that φ^* is invariant under f, that $\varphi^* \in L^1(\mu)$ and that $\int \varphi^* \, \mathrm{d}\mu = \int \varphi \, \mathrm{d}\mu$. Since f is ergodic, φ^* must be constant μ-a.e. and the thesis follows.

($\Leftarrow$) Let $B \in \mathfrak{B}$ be an invariant set and $\varphi = \mathbb{1}_B$. From the hypothesis we have that the mean sojourn time

$$\tau(B, x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_B(f^j(x))$$

is constant for μ-a.e. $x \in M$. But $\tau(B, x) = 1$ for every $x \in B$ and $\tau(B, x) = 0$ for every $x \in M \setminus B$, thus necessarily $\mu(B) = 0$ or $\mu(B) = 1$. □

References

[1] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, Springer-Verlag New York, 1982

Dipartimento di Matematica, Università di Pisa
Email address: alessio.delvigna@dm.unipi.it