

Birkhoff ergodic theorem

Assume $(\Omega, \mathcal{B}, \mathbb{P})$ is a probability space, $T: \Omega \rightarrow \Omega$ is a measurable map.

We say that

- $\mathbb{P}$ is T-invariant if $T_*\mathbb{P} = \mathbb{P}$, that is,
 $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A) \quad \forall A \in \mathcal{B}$;
- a subset $A \subset \Omega$ is T-invariant if
 $T^{-1}(A) = A$;
- a function $f: \Omega \rightarrow \mathbb{R}$ is T-invariant if
 $f \circ T = f$.

Denote by $\mathcal{B}^T \subset \mathcal{B}$ the collection of all measurable T-invariant subsets $A \subset \Omega$. It is a σ -subalgebra.

Theorem (Birkhoff)

Assume $(\Omega, \mathcal{B}, \mathbb{P})$ is a probability space, $T: \Omega \rightarrow \Omega$ is a measurable map, $\mathbb{P}$ is T-invariant. Then, for any $f \in L^1(\Omega, d\mathbb{P})$ we have

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \rightarrow \mathbb{E}(f | \mathcal{B}^T)$$

a.e. and in L^1 .

For the proof let us introduce the following notation:

Assume f is real-valued.

$$S_n = \sum_{k=0}^{n-1} f \circ T^k \quad (n \geq 1), \quad S_0 = 0,$$

$$M_n = \max \{ S_m : 0 \leq m \leq n \} \quad (n \geq 0)$$

Lemma ("maximal ergodic inequality")

Define $X_n = \{x \in \Omega : M_n(x) > 0\}$. Then

$$\int_{X_n} f \, dP \geq 0.$$

Proof

We claim that $f \geq M_n - M_n \circ T$ on X_n .

For $0 \leq m \leq n-1$, we have

$$S_{m+1} = f + S_m \circ T \leq f + M_n \circ T \quad (*)$$

Observe that for every $x \in X_n$ we have $S_m(x) > 0$ for some $1 \leq m \leq n$, hence on X_n we have

$$M_n = \max \{ S_{m+1} : 0 \leq m \leq n-1 \}.$$

By (*) we then get (on X_n) that

$$M_n \leq f + M_n \circ T,$$

proving our claim.

Now we set

$$\int_{X_n} f \, dP \geq \int_{X_n} (M_n - M_n \circ T) \, dP$$

$$= \int_{\Omega} (M_n - M_n \circ T) \, dP - \int_{X_n^c} (M_n - M_n \circ T) \, dP$$

$$= - \int_{X_n^c} (M_n - M_n \circ T) \, dP$$

(by T -invariance of P)

$$= \int_{X_n^c} M_n \circ T \, dP \geq 0.$$

$\uparrow$ X_n^c $\uparrow$
 $M_n = 0$ on X_n^c $M_n \geq 0$

Proof of Birkhoff's theorem

It suffices to consider real-valued f .

Replacing f by $f - \mathbb{E}(f | \mathcal{B}^T)$ we may also assume that $\mathbb{E}(f | \mathcal{B}^T) = 0$.

(Note that a $\mathcal{B}$ -measurable function g is $\mathcal{B}^T$ -measurable if and only if it is T -invariant.)

We want to show that $\frac{S_n}{n} \rightarrow 0$ a.e.

Define $\tilde{f} = \limsup_{n \rightarrow \infty} \frac{S_n}{n}$, $\tilde{f} : \mathcal{X} \rightarrow [-\infty, \infty]$.

It suffices to show that $\tilde{f} \leq 0$ (a.e.)

Indeed, the replacing f by $-f$ we also

get $\limsup_{n \rightarrow \infty} (-\frac{S_n}{n}) \leq 0$, that is,

$\liminf_{n \rightarrow \infty} \frac{S_n}{n} \geq 0$. Hence $\frac{S_n}{n} \rightarrow 0$ (a.e.)

Take $\varepsilon > 0$ and consider $A = \{x : \tilde{f}(x) \geq \varepsilon\}$.

We want to show that $P(A) = 0$.

Note that $\tilde{f}$ is T -invariant, since

$$\frac{1}{n} S_n \circ T = \frac{1}{n} \sum_{k=1}^n f \circ T^k = \frac{n+1}{n} S_n - \frac{1}{n} f.$$

Hence $A \in \mathcal{B}^T$.

Consider the function $f^* = (f - \varepsilon)\mathbb{1}_A$.

Put $S_n^* = \sum_{k=0}^{n-1} f^* \circ T^k$, $S_0^* = 0$ and

$M_n^* = \max \{S_m^* : 0 \leq m \leq n\}$. By T -invariance of A we have

$$S_n^* = S_n \mathbb{1}_A - n\varepsilon \mathbb{1}_A. \quad (*)$$

By the previous lemma applied to f^* we have $\int f^* dP \geq 0$. Consider the set $\{M_n^* > 0\}$.

Set $B = \bigcup_{n=1}^{\infty} \{M_n^* > 0\}$. Since the sequence of sets $\{M_n^* > 0\}$ is increasing by the dominated convergence theorem we get

$$\int_B f^* dP = \lim_{n \rightarrow \infty} \int_{\{M_n^* > 0\}} f^* dP \geq 0.$$

Now, by definition, if $x \in A$, then $\frac{S_n(x)}{n} > \varepsilon$ for infinitely many n 's and hence $\frac{S_n^*(x)}{n} > 0$ by (*). Thus, $A \subset B$.

Then

$$0 \leq \int_B f^* dP = \int_A (f - \varepsilon) dP$$

$$= \int_A \mathbb{E}(f - \varepsilon | \mathcal{B}^T) dP \quad (\text{as } A \in \mathcal{B}^T)$$

$$= -\varepsilon \int_A \mathbb{E}(1 | \mathcal{B}^T) dP = -\varepsilon P(A),$$

so $P(A) = 0$. This finishes the proof of convergence a.e.

If $f \in L^\infty(\Omega, \mathcal{B}, P) \subset L^1(\Omega, dP)$, then a.e. convergence implies that $\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \rightarrow E(f | \mathcal{B}^T)$ in L^1 by the dominated convergence theorem. Since $L^\infty(\Omega, \mathcal{B})$ is dense in $L^1(\Omega, dP)$ in the L^1 -norm and the operators $f \mapsto \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k$ and $f \mapsto E(f | \mathcal{B}^T)$ are contraction in this norm, we then get the required convergence for all $f \in L^1(\Omega, dP)$. □

Let $(\Omega, \mathcal{B}, P)$ and T be as before.

Def

If $\mathcal{B}^T$ consists only of sets A s.t. either $P(A)=0$ or $P(A)=1$, then T is called ergodic.

In this case $E(f | \mathcal{B}^T) = \int_{\Omega} f dP$, so we get the following result.

Corollary

In the setting of the theorem assume that T is ergodic. Then, for any $f \in L^1(\Omega, \mathcal{B}, dP)$, we have $\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \rightarrow \int_{\Omega} f dP$ a.e. and in L^1 .

Example (foundation of statistical mechanics)

Ω is the state space of a physical system (e.g., gas), P is the probability for the system to be in a certain state, $T: \Omega \rightarrow \Omega$ is the evolution of the system in a unit of time (e.g., second). A physical observable is a function $f: \Omega \rightarrow \mathbb{R}$. One argues is that if the system starts in a state x , then $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$ is the value of f that we can realistically measure in experiments. A key idea of statistical mechanics is that this quantity can also be computed as $\int_{\Omega} f dP$. We see know that this is indeed the case for P -a.e. x if T is ergodic.

Def

A transformation $T: \Omega \rightarrow \Omega$ is called mixing if $P(T^{-n}(A) \cap B) \xrightarrow{n \rightarrow \infty} P(A)P(B) \quad \forall A, B \in \mathcal{B}$.

Lemma

Any mixing transformation is ergodic.

Proof

If $A \in \mathcal{B}^T$, then $P(A) = P(T^{-n}(A) \cap A) \xrightarrow{n \rightarrow \infty} P(A)^2$,
hence either $P(A) = 0$ or $P(A) = 1$. □