

Birkhorff Ergodic Theorem

1. Theory of Measures: a Brief Summary

A. σ -algebra and measure space

(A.1) (σ -algebra) Let X be a given set and \mathcal{A} be a collection of subsets of X . We say that \mathcal{A} is an algebra if (i) X is in \mathcal{A} ; (ii) if A is in \mathcal{A} then $X \setminus A$ is in \mathcal{A} ; and (iii) let $\{A_m, m \in \mathbb{Z}^+\}$ be a countable sub-collections of \mathcal{A} , then $\cup_m A_m \in \mathcal{A}$.

(A.2) (Measure space) Let \mathcal{A} be a σ -algebra for X . We call a function μ from \mathcal{A} to \mathbb{R}^+ a *measure* on (X, \mathcal{A}) if (i) $\mu(\emptyset) = 0$, and (ii) For mutually disjoint measurable sets $A_i \in \mathcal{A}, i = 1, \dots$, we have $\mu(\cup_i A_i) = \sum \mu(A_i)$.

We call (X, \mathcal{A}, μ) , with \mathcal{A} a σ -algebra, μ a measure on (X, \mathcal{A}) , a *measurable space*.

Ex1: (Lebesgue measure) Let $X = \mathbb{R}$, \mathcal{B} be the smallest σ -algebra containing all intervals in \mathbb{R} . Then there exists a unique $\mu : \mathcal{B} \rightarrow [0, 1]$ such that $(\mathbb{R}, \mathcal{B}, \mu)$ is a measure space satisfying $\mu([a, b]) = |b - a| \forall [a, b] \in \mathbb{R}$.

Ex2: (Boreal measure) Let (X, d) be a metric space so the collection of open sets in X is well-defined. Let \mathcal{B} be the smallest σ -algebra containing all open sets. We call \mathcal{B} the Boreal algebra associated with metric d . We call a measure $\mu : \mathcal{B} \rightarrow \mathbb{R}^+$ a Boreal measure. Note that the Lebesgue measure on \mathbb{R} is a Boreal measure associated with $d(x, y) = |x - y|$.

Ex3: Let $M \in \mathbb{R}^n$ be smooth surface embedded in \mathbb{R}^n (such as $S^1 \in \mathbb{R}^1$ or $S^2, \mathbb{T}^2, \in \mathbb{R}^3$). We regard $B(x, \varepsilon) \cap M$ as an ε -ball around x in M where $B(x, \varepsilon)$ is the ε -ball centered at x in \mathbb{R}^n . This way the Boreal algebra on M is well-defined. Then there exists a unique measure μ such that if $B \in \mathcal{B}$ is a smooth piece of surface, then $\mu(B) =$ the surface area of B . (X, \mathcal{B}, μ) is a Boreal measure.

Ex4: (Cases of triviality) Let X be any given set. Let $\mathcal{B} = \{X, \emptyset\}$ and μ is defined by $\mu(X) = 1, \mu(\emptyset) = 0$. (X, \mathcal{B}, μ) so defined is a measure space.

Even if we have a really complicated σ -algebra such as the Boreal algebra on \mathbb{R} , A Boreal measure is not necessarily complicated. For instance, let \mathcal{B} be the Boreal algebra on real line. We can define a Boreal measure $\mu : \mathcal{B} \rightarrow \mathbb{R}^+$ by letting $\mu(B) = 1$ for $B \in \mathcal{B}$ such that $0 \in B$, and $\mu(B) = 0$ otherwise.

B. Measurable functions and integration

Let (X, \mathcal{B}, μ) be a measure space

(B.1) (Measurable functions) For a function $f : X \rightarrow \mathbb{R}$ and $A \subset \mathbb{R}$, let $f^{-1}(S) = \{x \in X, f(x) \in S\}$. We say that f is a measurable function if $f^{-1}(S) \in \mathcal{B} \forall S$ that is a Boreal set.

(B.2) (Integration) The integral for a measurable function is defined as follows

– For $A \in \mathcal{B}$, let $\mathcal{X}_A(x) = 1$ for $x \in A$ and $\mathcal{X}_A(x) = 0$ otherwise. $\mathcal{X}_A : X \rightarrow \mathbb{R}$ is a measurable function, which we call a characteristic function for A .

– Let $A_i, i = 1, \dots$ are mutually disjoint sets in \mathcal{B} , and $c_i > 0, i = 1, \dots$ be a sequence of numbers. We call

$$h(x) = \sum c_i \mathcal{X}_{A_i}(x)$$

a simple function.

– Let us define, for a simple function $f(x)$

$$\int_X h(x) d\mu := \sum c_i \mu(A_i).$$

– For $n \in \mathbb{Z}^+$, denote $\mathbb{R} = \cup_{p \geq 0} S_p$ where $S_p = [\frac{p}{n}, \frac{p+1}{n}]$. For a measurable function $f > 0$, let

$$A_i = f^{-1}(S_p); \quad h_n(x) = \sum_{p \geq 0} \frac{p}{n} \mathcal{X}_{A_i}(x).$$

The integral for f is defined by

$$\int_X f(x) d\mu = \lim_{n \rightarrow \infty} \int_X h_n(x) d\mu.$$

– For an arbitrary measurable function f , let $A^+ = \{x \in X, f(x) \geq 0\}$, $A^- = \{x \in X, f(x) \leq 0\}$ and write

$$f^+(x) = f(x) \cdot \mathcal{X}_{A^+}(x), \quad f^-(x) = -f(x) \cdot \mathcal{X}_{A^-}(x).$$

Integration for f is defined as

$$\int_X f(x) d\mu = \int_X f^+(x) d\mu - \int_X f^-(x) d\mu.$$

C. Classical space of measurable functions

– (L^p -space) For $p > 0$ given, let

$$L^p(X) = \{f : \int_X |f|^p < \infty\}.$$

– $L^p(X)$ is a linear space $((c_1f + c_2g)(x) := c_1f(x) + c_2g(x))$.

– For $f \in L^p(X)$, let

$$\|f\| = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}}.$$

$\|\cdot\|$ is a well-defined norm. $L^p(X)$ is in fact a complete metric space (Banach Space).

– $L^2(X)$ is also a Hilbert space: for $f, g \in L^2$
let

$$\langle f, g \rangle = \int_X f(x) \cdot g(x) d\mu.$$

then $\|f\|^2 = \langle f, f \rangle$.

2. Measure preserving transforms and ergodicity

A. Measure preserving transforms (m.p.t.)

[A.1] (Measurable transforms) Let T be from X to X where (X, \mathcal{B}, μ) is a measure space. For $B \in \mathcal{B}$, let $T^{-1}B = \{x \in X, T(x) \in B\}$. We say that T is measurable if $T^{-1}B \in \mathcal{B}$ for all $B \in \mathcal{B}$.

[A.2] (m.p.t.) A measurable transform $T : X \rightarrow X$ is measure preserving if for all $B \in \mathcal{B}$,

$$\mu(T^{-1}B) = \mu(B).$$

Ex1. $T(x) = x + \alpha$ is a measure preserving transform on $S^1 = \frac{\mathbb{R}}{\mathbb{Z}}$.

Also,

$$T(x, y) = (x + \alpha, y + \beta)$$

is a measure preserving transform on $\mathbb{T}^2 = \frac{\mathbb{R}^2}{\mathbb{Z}^2}$.

Ex2. Let $m \geq 1$ be an integer, then $T(x) \rightarrow mx$ is a measure preserving transform on S^1 .

Let $m, n \geq 1$ be integers, then

$$T(x, y) = (mx, ny)$$

is a measure preserving transform on \mathbb{T}^2 .

Note that in this example, T is not invertible (Homework).

Ex3. The time-one mapping induced by the solutions of a Hamiltonian system is area preserving.

B. Ergodicity

Let (X, \mathcal{B}, μ) be a measure space such that $\mu(X) < \infty$, and $T : X \rightarrow X$ is measurable.

[B.1] We say that $A \in \mathcal{B}$ is an invariant subset if $T^{-1}A = A$.

[B.2] We say that $f \in L^2(X)$ is T -invariant if $f(x) = f(Tx)$ for almost all $x \in X$ (*almost all* means a possibility of discounting a set of measure zero).

[B.2] We say that T is ergodic if $\forall A \in \mathcal{B}$, A is invariant implies that either $\mu(A) = \mu(X)$ or $\mu(A) = 0$.

Claim: T is ergodic if and only if the only invariant functions $f \in L^2(X)$ under T are the constant functions.

Proof: First we prove that if constant functions are the only ones in $L^2(X)$ that is T -invariant, then T is ergodic. If T is not ergodic, then there exists $A \in \mathcal{B}$ such that A and $X \setminus A$ are both invariant under T and $\mu(A), \mu(X \setminus A) >$

0. It is clear that $\chi_A(x) \in L^2(X)$ is T -invariant but not a constant function.

For the other direction let us assume that there exists $f \in L^2(X)$ that is not constant. Then there exists c such that $\mu(f \geq c) > 0$, $\mu(f < c) > 0$ (The existence of c is left as an exercise). But both of these two are T -invariant because f is T -invariant. \square .

Ex1. The rigid rotation $T : \theta \rightarrow \theta + \alpha$ is ergodic if and only if $\frac{\alpha}{2\pi}$ is irrational.

Proof: Let us expand $f \in L^2(S^1)$ in Fourier series, i.e., we write

$$f(x) = \sum_n e^{nxi}$$

Then

$$f(T(x)) = \sum_n e^{n\alpha} e^{nxi}$$

$f(T(x)) = f(x)$ leads to

$$a_n(1 - e^{n\alpha}) = 0.$$

If $\frac{\alpha}{2\pi}$ is irrational then we must have $a_n = 0$ for all $n \neq 0$.

If $\frac{\alpha}{2\pi} = \frac{p}{q}$, simply take

$$f(x) = \sin qx$$

We have $f(Tx) = \sin(qx + 2\pi p) = f(x)$. Note that $f(x)$ is clearly not a constant function. So T is not ergodic.

Ex2. $T : \mathbb{T}^n \rightarrow \mathbb{T}^n$ defined by

$$T(x_1, \dots, x_n) = (x_1 + \alpha_1, \dots, x_n + \alpha_n)$$

is ergodic if and only if there exists $k_1, \dots, k_n \in \mathbb{Z}$ such that

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n = 0 \pmod{2\pi}.$$

The proof is a simple generalization of Ex1. We leave it as a homework.

Ex3. $T(x) = mx$, $m > 1$ defined on S^1 is ergodic.

Proof: For $f \in L^2(S^1)$, again write

$$f(x) = \sum_n a_n e^{nxi}.$$

Then $f(x) = f(Tx)$ leads to

$$a_1 = a_m = a_{m^2} = \dots ; a_2 = a_{2m} = a_{2m^2} = \dots .$$

So if, for $k \neq 0$, $a_k \neq 0$, then

$$a_k = a_{km} = a_{km^2} = \dots .$$

This is impossible because the magnitude of a_n must decline exponentially with respect to n . □

Ex4. Let $A = (a_{ij})$ be a matrix of integer entries. The $T : \mathbb{T}^n \rightarrow \mathbb{T}^n$ defined by $z_1 = Az$ for $z = (x_1, x_2, \dots, x_n) \in \frac{\mathbb{R}^n}{\mathbb{Z}^n}$ is ergodic if and only if the matrix A does not have a root of unity as its eigenvalue.

The Proof is a simple generalization of Ex3.
Another homework.

C. Birkhoff's Ergodic Theorem

Let (X, \mathcal{B}, μ) be a measure space such that $\mu(X) < \infty$.

Assume that $T : X \rightarrow X$ is ergodic, then $\forall U \in \mathcal{B}$ satisfying $\mu(U) > 0$, $U \cap \mathbf{O}(x)$ for almost every $x \in X$ where $\mathbf{O}(x) = \{T^n x\}_{n=0}^{\infty}$ is the orbit of x .

We now move forward to consider the frequencies of visitations of $\mathbf{O}(x)$ to $U \in \mathcal{B}$. For $x \in X$, let

$$I_n(x) = \frac{1}{n} \chi_U(x) + \chi_U(Tx) + \cdots + \chi_U(T^n x).$$

$I(x) = \lim_{n \rightarrow \infty} I_n(x)$, is the frequency the orbit $\mathbf{O}(x)$ visited U .

We note that there is no reason *a priori* for I_x to be well-defined in general. For measure preserving transforms that is ergodic, however, we are in a much better situation.

Theorem (Birkhoff's Ergodic theorem) Let $T : X \rightarrow X$ be a measure preserving transform on a measure space (X, \mathcal{B}, μ) and $U \in \mathcal{B}$ be such that $\mu(U) > 0$. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n \chi_U(T^i x) = \frac{\mu(U)}{\mu(X)}.$$

for almost every $x \in X$.

Proof: Let us first assume that $I(x)$ exists for almost all $x \in X$. Then it is (as a limit of a sequence of functions $I_n(x)$ in $L^2(X)$) in $L^2(x)$. $I(x)$ is also T -invariant because

$$I(Tx) - I(x) = \lim_{n \rightarrow \infty} \frac{1}{n} (\chi_U(T^{n+1}(x)) - \chi_U(x)) = 0.$$

Since T is ergodic we have $I(x) = C$ for almost all $x \in X$. C is then determined by integrating

both side of

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n \chi_U(T^i x)$$

over X , from which we obtain

$$C\mu(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n \int_X \chi_U(T^i(x)) d\mu.$$

Observe that

$$\int_X \chi_U(T^i x) d\mu = \mu(T^{-i}U) = \mu(U).$$

We finally obtain

$$C = \frac{\mu(U)}{\mu(X)}$$

as claimed.

We now prove that $I(x)$ exists for almost all $x \in X$. Let $\varepsilon > 0$ be fixed and

$$f(x) = \chi_U - \frac{\mu(U)}{\mu(X)} - \varepsilon.$$

We also let

$$\phi_k(x) = f(x) + f(Tx) + \cdots + f(T^k(x));$$

$$F_n(x) = \max_{0 \leq k \leq n} \phi_k(x);$$

and

$$A := \{x \in X, \limsup_{n \rightarrow \infty} F_n(x) \rightarrow +\infty\}.$$

A is obviously T -invariant. Since T is ergodic, either $\mu(A) = 0$ or $\mu(A) = \mu(X)$

We now prove

$$\mu(A) = 0. \tag{1}$$

Observe that

$$F_{n+1}(x) - F_n(Tx) = f - \min(0, F_n(Tx)).$$

(If $F_n(Tx) < 0$ then $F_{n+1} = f + F_n(Tx)$. If $F_n(Tx) \geq 0$ then $F_{n+1} = f + F_n(Tx)$.)

So for $x \in A$,

$$\lim_{n \rightarrow \infty} (F_{n+1}(x) - F_n(x)) = f$$

and

$$\lim_{n \rightarrow \infty} \int_A (F_{n+1}(x) - F_n(Tx)) d\mu = \int_A f(x) d\mu := -\varepsilon \mu(A) < 0. \quad (2)$$

Here we used $\mu(A) = \mu(X)$. Observe that we also have

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_A F_n(x) d\mu &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \int_A (F_{k+1}(x) - F_k(x)) d\mu \\ &= \int_A (F_{k+1}(x) - F_k(Tx)) d\mu. \end{aligned}$$

(To obtain the last inequality we used

$$\int_A F_k(x) d\mu = \int_A F_k(Tx) d\mu.$$

This is because (a) $\mu(A) = \mu(X)$ and (b) T is measure preserving.

From the definition of A , the first quantity is $+\infty$. But from (2), the last quantity is $< +\infty$.

So (1) hold. It follows that for almost every $x \in X$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} F_n(x) d\mu \leq 0. \quad (3)$$

From (3) we have that for almost all $x \in X$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \chi_U(T^k x) \leq \frac{\mu(U)}{\mu(X)} + \varepsilon.$$

To finish the proof of this theorem we also need to prove

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \chi_U(T^k x) \geq \frac{\mu(U)}{\mu(X)} - \varepsilon. \quad (4)$$

This is obtained by re-defining f in the above as

$$f(x) = -\chi_U(x) + \frac{\mu(U)}{\mu(X)} - \varepsilon$$

and repeat the same argument. (4) follows again from (3). \square

Remarks: Let us re-write (??) as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n \mathcal{X}_U(T^i x) = \frac{1}{\mu(X)} \int_X \mathcal{X}_U(x) d\mu.$$

The quantity on the left is the average value of the function \mathcal{X}_U along a single orbit $\mathbf{O}(x)$ (*a time average*). The quantity on the right is the average value of the function $\mathcal{X}_U(x)$ over the entire space X (*the space average*). A slightly general form of the Birkhoff's ergodic theory is to replace $\mathcal{X}_U(x)$ by any given $\phi(x) \in L^1(X)$.

Claim: Assume that $T : X \rightarrow X$ is an ergodic, measure preserving transform. Let $\phi \in L^1$. Then for almost every $x \in X$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n \phi(T^k x) = \frac{1}{\mu(X)} \int_X \phi(x) d\mu.$$

This claim is what is commonly referred to as the Birkhoff's ergodic theorem, the proof of which is essentially the same as the one detailed above.