21. Birkhoff's Ergodic Theorem

§21.1 Introduction

An ergodic theorem is a result that describes the limiting behaviour of the sequence

$$\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j \tag{21.1}$$

as $n \to \infty$. The precise formulation of an ergodic theorem depends on the class of function f (for example, one could assume that f is integrable, L^2, or continuous), and the notion of convergence that we use (for example, we could study pointwise convergence, L^2 convergence, or uniform convergence). The result that we are interested here—Birkhoff's Ergodic Theorem—deals with pointwise convergence of (21.1) for an integrable function f.

§21.2 Conditional expectation

We will need the concepts of Radon-Nikodym derivates and conditional expectation.

Definition. Let μ be a measure on $(X, \mathcal{B})$. We say that a measure ν is *absolutely continuous with respect to* μ and write $\nu \ll \mu$ if $\nu(B) = 0$ whenever $\mu(B) = 0$, $B \in \mathcal{B}$.

Remark. Thus ν is absolutely continuous with respect to μ if sets of μ-measure zero also have ν-measure zero (but there may be more sets of ν-measure zero).

For example, let $f \in L^1(X, \mathcal{B}, \mu)$ be non-negative and define a measure ν by

$$\nu(B) = \int_B f \, d\mu.$$

Then $\nu \ll \mu$.

The following theorem says that, essentially, all absolutely continuous measures occur in this way.

Theorem 21.1 (Radon-Nikodym)
Let $(X, \mathcal{B}, \mu)$ be a probability space. Let ν be a measure defined on $\mathcal{B}$ and suppose that $\nu \ll \mu$. Then there is a non-negative measurable function f such that

$$\nu(B) = \int_B f \, d\mu, \quad \text{for all } B \in \mathcal{B}.$$

Moreover, f is unique in the sense that if g is a measurable function with the same property then $f = g$ μ-a.e.

Exercise 21.1
If $\nu \ll \mu$ then it is customary to write $d\nu/d\mu$ for the function given by the Radon-Nikodym theorem, that is

$$\nu(B) = \int_B \frac{d\nu}{d\mu}\, d\mu.$$

Prove the following relations:

(i) *If $\nu \ll \mu$ and f is a μ-integrable function then*

$$\int f\, d\nu = \int f \frac{d\nu}{d\mu}\, d\mu.$$

(ii) *If $\nu_1, \nu_2 \ll \mu$ then*

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu}.$$

(iii) *If $\lambda \ll \nu \ll \mu$ then*

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu}\frac{d\nu}{d\mu}.$$

Let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Note that μ defines a measure on $\mathcal{A}$ by restriction. Let $f \in L^1(X, \mathcal{B}, \mu)$. Then we can define a measure ν on $\mathcal{A}$ by setting

$$\nu(A) = \int_A f\, d\mu.$$

Note that $\nu \ll \mu|_{\mathcal{A}}$. Hence by the Radon-Nikodym theorem, there is a unique $\mathcal{A}$-measurable function $E(f \mid \mathcal{A})$ such that

$$\nu(A) = \int E(f \mid \mathcal{A})\, d\mu.$$

We call $E(f \mid \mathcal{A})$ the *conditional expectation* of f with respect to the σ-algebra $\mathcal{A}$.

So far, we have only defined $E(f \mid \mathcal{A})$ for non-negative f. To define $E(f \mid \mathcal{A})$ for an arbitrary f, we split f into positive and negative parts $f = f_+ - f_-$ where $f_+, f_- \geq 0$ and define

$$E(f \mid \mathcal{A}) = E(f_+ \mid \mathcal{A}) - E(f_- \mid \mathcal{A}).$$

Thus we can view conditional expectation as an operator

$$E(\cdot \mid \mathcal{A}) : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu).$$

Note that $E(f \mid \mathcal{A})$ is uniquely determined by the two requirements that

(i) $E(f \mid \mathcal{A})$ is $\mathcal{A}$-measurable, and

(ii) $\int_A f \, d\mu = \int_A E(f \mid \mathcal{A}) \, d\mu$ for all $A \in \mathcal{A}$.

Intuitively, one can think of $E(f \mid \mathcal{A})$ as the best approximation to f in the smaller space of all $\mathcal{A}$-measurable functions.

Exercise 21.2

(i) Prove that $f \mapsto E(f \mid \mathcal{A})$ is linear.

(ii) Suppose that g is $\mathcal{A}$-measurable and $|g| < \infty$ μ-a.e. Show that $E(fg \mid \mathcal{A}) = gE(f \mid \mathcal{A})$.

(iii) Suppose that T is a measure-preserving transformation. Show that $E(f \mid \mathcal{A}) \circ T = E(f \circ T \mid T^{-1}\mathcal{A})$.

(iv) Show that $E(f \mid \mathcal{B}) = f$.

(v) Let $\mathcal{N}$ denote the trivial σ-algebra consisting of all sets of measure 0 and 1. Show that $E(f \mid \mathcal{N}) = \int f \, d\mu$.

To state Birkhoff's Ergodic Theorem precisely, we will need the sub-σ-algebra $\mathcal{I}$ of T-invariant subsets, namely:

$$\mathcal{I} = \{B \in \mathcal{B} \mid T^{-1}B = B \text{ a.e.}\}.$$

Exercise 21.3

Prove that $\mathcal{I}$ is a σ-algebra.

§21.3 Birkhoff's Pointwise Ergodic Theorem

Birkhoff's Ergodic Theorem deals with the behaviour of $\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$ for μ-a.e. $x \in X$, and for $f \in L^1(X, \mathcal{B}, \mu)$.

Theorem 21.2 (Birkhoff's Ergodic Theorem)

Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving transformation. Let $\mathcal{I}$ denote the σ-algebra of T-invariant sets. Then for every $f \in L^1(X, \mathcal{B}, \mu)$, we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \to E(f \mid \mathcal{I})$$

for μ-a.e. $x \in X$.

Corollary 21.3
Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be an ergodic measure-preserving transformation. Let $f \in L^1(X, \mathcal{B}, \mu)$. Then

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x) \to \int f \, d\mu, \qquad \text{as } n \to \infty,$$

for μ-a.e. $x \in X$.

Proof. If T is ergodic then $\mathcal{I}$ is the trivial σ-algebra $\mathcal{N}$ consisting of sets of measure 0 and 1. If $f \in L^1(X, \mathcal{B}, \mu)$ then $E(f \mid \mathcal{N}) = \int f \, d\mu$. The result follows from the general version of Birkhoff's ergodic theorem. □

§21.4 Appendix: The proof of Birkhoff's Ergodic Theorem

The proof is something of a tour de force of hard analysis. It is based on the following inequality.

Theorem 21.4 (Maximal Inequality)
Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T : X \to X$ be a measure-preserving transformation and let $f \in L^1(X, \mathcal{B}, \mu)$. Define $f_0 = 0$ and, for $n \geq 1$,

$$f_n = f + f \circ T + \cdots + f \circ T^{n-1}.$$

For $n \geq 1$, set

$$F_n = \max_{0 \leq j \leq n} f_j$$

(so that $F_n \geq 0$). Then

$$\int_{\{x \mid F_n(x) > 0\}} f \, d\mu \geq 0.$$

Proof. Clearly $F_n \in L^1(X, \mathcal{B}, \mu)$. For $0 \leq j \leq n$, we have $F_n \geq f_j$, so $F_n \circ T \geq f_j \circ T$. Hence

$$F_n \circ T + f \geq f_j \circ T + f = f_{j+1}$$

and therefore

$$F_n \circ T(x) + f(x) \geq \max_{1 \leq j \leq n} f_j(x).$$

If $F_n(x) > 0$ then

$$\max_{1 \leq j \leq n} f_j(x) = \max_{0 \leq j \leq n} f_j(x) = F_n(x),$$

so we obtain that

$$f \geq F_n - F_n \circ T$$

on the set $A = \{x \mid F_n(x) > 0\}$.

Hence

$$\begin{aligned}\int_A f\,d\mu &\geq \int_A F_n\,d\mu - \int_A F_n \circ T\,d\mu \\ &= \int_X F_n\,d\mu - \int_A F_n \circ T\,d\mu \\ &\geq \int_X F_n\,d\mu - \int_X F_n \circ T\,d\mu \\ &= 0\end{aligned}$$

where we have used

(i) $F_n = 0$ on $X \setminus A$

(ii) $F_n \circ T \geq 0$

(iii) μ is T-invariant.

□

Corollary 21.5
If $g \in L^1(X, \mathcal{B}, \mu)$ and if

$$B_\alpha = \left\{ x \in X \mid \sup_{n \geq 1} \frac{1}{n} \sum_{j=0}^{n-1} g(T^j x) > \alpha \right\}$$

then for all $A \in \mathcal{B}$ with $T^{-1}A = A$ we have that

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

Proof. Suppose first that $A = X$. Let $f = g - \alpha$, then

$$\begin{aligned}B_\alpha &= \bigcup_{n=1}^{\infty} \left\{ x \mid \sum_{j=0}^{n-1} g(T^j x) > n\alpha \right\} \\ &= \bigcup_{n=1}^{\infty} \{x \mid f_n(x) > 0\} \\ &= \bigcup_{n=1}^{\infty} \{x \mid F_n(x) > 0\}\end{aligned}$$

(since $f_n(x) > 0 \Rightarrow F_n(x) > 0$ and $F_n(x) > 0 \Rightarrow f_j(x) > 0$ for some $1 \leq j \leq n$). Write $C_n = \{x \mid F_n(x) > 0\}$ and observe that $C_n \subset C_{n+1}$. Thus χ_{C_n}

converges to χ_{B_α} and so $f\chi_{C_n}$ converges to $f\chi_{B_\alpha}$, as $n \to \infty$. Furthermore, $|f\chi_{C_n}| \leq |f|$. Hence, by the Dominated Convergence Theorem,

$$\int_{C_n} f \, d\mu = \int_X f\chi_{C_n} \, d\mu \to \int_X f\chi_{B_\alpha} \, d\mu = \int_{B_\alpha} f \, d\mu, \quad \text{as } n \to \infty.$$

Applying the maximal inequality, we have, for all $n \geq 1$,

$$\int_{C_n} f \, d\mu \geq 0.$$

Therefore

$$\int_{B_\alpha} f \, d\mu \geq 0,$$

i.e.,

$$\int_{B_\alpha} g \, d\mu \geq \alpha\mu(B_\alpha).$$

For the general case, we work with the restriction of T to A, $T : A \to A$, and apply the maximal inequality on this subset to get

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

as required. □

Proof of Birkhoff's Ergodic Theorem. Let

$$f^*(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$$

and

$$f_*(x) = \liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x).$$

Writing

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

observe that

$$\frac{n+1}{n} a_{n+1}(x) = a_n(Tx) + \frac{1}{n} f(x).$$

Taking the lim sup and lim inf as $n \to \infty$ gives us that $f^* \circ T = f^*$ and $f_* \circ T = f_*$.

We have to show

(i) $f^* = f_*$ μ-a.e

(ii) $f^* \in L^1(X, \mathcal{B}, \mu)$

(iii) $\int f^* \, d\mu = \int f \, d\mu$.

We prove (i). For $\alpha, \beta \in \mathbb{R}$, define

$$E_{\alpha,\beta} = \{x \in X \mid f_*(x) < \beta \text{ and } f^*(x) > \alpha\}.$$

Note that

$$\{x \in X \mid f_*(x) < f^*(x)\} = \bigcup_{\beta<\alpha,\ \alpha,\beta\in\mathbb{Q}} E_{\alpha,\beta}$$

(a countable union). Thus, to show that $f^* = f_*$ μ-a.e., it suffices to show that $\mu(E_{\alpha,\beta}) = 0$ whenever $\beta < \alpha$. Since $f_* \circ T = f_*$ and $f^* \circ T = f^*$, we see that $T^{-1}E_{\alpha,\beta} = E_{\alpha,\beta}$. If we write

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

then $E_{\alpha,\beta} \cap B_\alpha = E_{\alpha,\beta}$.

Applying Corollary 21.5 we have that

$$\begin{aligned} \int_{E_{\alpha,\beta}} f \, d\mu &= \int_{E_{\alpha,\beta}\cap B_\alpha} f \, d\mu \\ &\geq \alpha\mu(E_{\alpha,\beta} \cap B_\alpha) = \alpha\mu(E_{\alpha,\beta}). \end{aligned}$$

Replacing f, α and β by $-f$, $-\beta$ and $-\alpha$ and using the fact that $(-f)^* = -f_*$ and $(-f)_* = -f^*$, we also get

$$\int_{E_{\alpha,\beta}} f \, d\mu \leq \beta\mu(E_{\alpha,\beta}).$$

Therefore

$$\alpha\mu(E_{\alpha,\beta}) \leq \beta\mu(E_{\alpha,\beta})$$

and since $\beta < \alpha$ this shows that $\mu(E_{\alpha,\beta}) = 0$. Thus $f^* = f_*$ μ-a.e. and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = f^*(x) \qquad \mu\text{-a.e.}$$

We prove (ii). Let

$$g_n = \left| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \right|.$$

Then $g_n \geq 0$ and

$$\int g_n \, d\mu \leq \int |f| \, d\mu$$

so we can apply Fatou's Lemma to conclude that $\lim_{n\to\infty} g_n = |f^*|$ is integrable, i.e., that $f^* \in L^1(X, \mathcal{B}, \mu)$.

We prove (iii). For $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, define

$$D_k^n = \left\{ x \in X \mid \frac{k}{n} \leq f^*(x) < \frac{k+1}{n} \right\}.$$

For every $\varepsilon > 0$, we have that

$$D_k^n \cap B_{\frac{k}{n}-\varepsilon} = D_k^n.$$

Since $T^{-1}D_k^n = D_k^n$, we can apply Corollary 22.4 again to obtain

$$\int_{D_k^n} f \, d\mu \geq \left(\frac{k}{n} - \varepsilon \right) \mu(D_k^n).$$

Since $\varepsilon > 0$ is arbitrary, we have

$$\int_{D_k^n} f \ d\mu \geq \frac{k}{n} \ \mu(D_k^n).$$

Thus

$$\begin{aligned} \int_{D_k^n} f^* \, d\mu &\leq \frac{k+1}{n} \ \mu(D_k^n) \\ &\leq \frac{1}{n} \ \mu(D_k^n) + \int_{D_k^n} f \, d\mu \end{aligned}$$

(where the first inequality follows from the definition of D_k^n). Since

$$X = \bigcup_{k\in\mathbb{Z}} D_k^n$$

(a disjoint union), summing over $k \in \mathbb{Z}$ gives

$$\begin{aligned} \int_X f^* \, d\mu &\leq \frac{1}{n} \ \mu(X) + \int_X f \, d\mu \\ &= \frac{1}{n} + \int_X f \, d\mu. \end{aligned}$$

Since this holds for all $n \geq 1$, we obtain

$$\int_X f^* \, d\mu \leq \int_X f \, d\mu.$$

Applying the same argument to $-f$ gives

$$\int (-f)^* \, d\mu \leq \int -f \, d\mu$$

so that
$$\int f^* \, d\mu = \int f_* \, d\mu \geq \int f \, d\mu.$$
Therefore
$$\int f^* \, d\mu = \int f \, d\mu,$$
as required.

Finally, we prove that $f^* = E(f \mid \mathcal{I})$. First note that as f^* is T-invariant, it is measurable with respect to $\mathcal{I}$. Moreover, if I is any T-invariant set then
$$\int_I f, d\mu = \int_I f^* \, d\mu.$$
Hence $f^* = E(f \mid \mathcal{I})$. □