Ergodic Theory
(Lecture Notes)

Joan Andreu Lázaro Camí

Department of Mathematics
Imperial College London
January 2010

Contents

Chapter 1

Introduction

Ergodic theory lies in somewhere among measure theory, analysis, probability, dynamical systems, and differential equations and can be motivated from many different angles. We will choose one specific point of view but there are many others. Let

$$\dot{x} = f(x)$$

be an *ordinary differential equation.* The problem of studying differential equations goes back centuries and, throughout the years, many different approaches and techniques have been developed. The most classical approach is that of finding explicit analytic solutions. This approach can provide a great deal of information but is essentially only applicable to an extremely restricted class of differential equations. From the very beginning of the 20th century, there has been a great development on topological methods to obtain qualitative topological information such as the existence of periodic solutions. Again, this can be a very successful approach in certain situations but there are a lot of equations which have, for example, infinitely many periodic solutions, possibly intertwined in very complicated ways, to which these methods do not really apply. Finally there are numerical methods to approximate solutions. In the last few decades, with the increase of computing power, there has been hope that numerical methods could play an important role. Again, while this is true in some situations, there are also a lot of equations for which the numerical methods have very limited applicability because the approximation errors grow exponentially and quickly become uncontrollable. Moreover, the *sensitive dependence on initial conditions* is now understood to be an intrinsic feature of certain equations than cannot be resolved by increasing the computing power.

Example 1 Lorenz's equations were introduced by the meteorologist E.

Lorenz in 1963 as an extremely simplified model of the Navier-Stokes equations for a fluid flow:

$$\begin{cases} \dot{x}_1 = \sigma\,(x_2 - x_1) \\ \dot{x}_2 = x_1\,(\rho - x_3) - x_2 \\ \dot{x}_3 = x_1 x_2 - \beta x_3, \end{cases} \tag{1.1}$$

where σ is the ***Prandlt number*** and ρ the ***Rayleigh number***. Usually $\sigma = 10$, $\beta = 8/3$, and ρ varies. However, for $\rho = 28$, the systems exhibits a chaotic behaviour. This is a very good example of a relatively simple ODE which is quite intractable from many angles. It does not admit any explicit analytic solutions; the topology is extremely complicated with infinitely many periodic solutions which are knotted in many different ways (there are studies of the structure of the periodic solutions of Lorenz's equations from the point of view of knot theory); on the other hand, numerical integration has very limited use since nearby solutions diverge very quickly.

Using classical methods, one can prove that the solutions of Lorenz's equations, eventually, end up in some bounded region $U \subset \mathbb{R}^3$. This simplifies our approach significantly since it means that it is sufficient to concentrate on the solutions inside U. A combination of results obtained over almost 40 years by several different mathematicians can be formulated in the following theorem which can be thought of essentially as a statement in ergodic theory. We give here a precise but slightly informal statement as some of the terms will be defined more precisely later on these notes.

Theorem 2 *For every ball $B \subset \mathbb{R}^3$, there exists a "probability" $p\,(B) \in [0,1]$ such that, for "almost every" initial condition $x_0 \in \mathbb{R}^3$, we have*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{1}_B\,(x_t)\,dt = p\,(B)\,, \tag{1.2}$$

where x_t is the solution of (1.1) with initial condition x_0.

First of all, recall that $\mathbf{1}_B$ is the characteristic function of the set B defined by

$$\mathbf{1}_B\,(x) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{if } x \notin B. \end{cases}$$

The integral $\int_0^T \mathbf{1}_B\,(x_t)\,dt$ is simply the amount of time that the solution x_t spends inside the ball B between time 0 and time T, and $T^{-1}\int_0^T \mathbf{1}_B\,(x_t)\,dt$ is therefore the *proportion* of time that the solution spends in B from $t = 0$ to T. Theorem 2 makes two highly non trivial assertions:

1. that the proportion $T^{-1}\int_0^T \mathbf{1}_B(x_t)\,dt$ converges as $T\to\infty$;

2. that this limit is independent of the initial condition x_0.

There is no a priori reason why the limit (1.2) should exist. But perhaps the most remarkable fact is that the limit is the same for *almost all* initial conditions (the concept *almost all* will be made precise later). This says that the asymptotic time averages of the solution x_t with initial condition x_0 are actually independent of this initial condition. Therefore, independently of the initial condition, the proportion of time that the systems spends on B is $P(B)$. In other words, there exists a way of *measuring* the balls B such that the measure $P(B)$ gives us information on the amount of time that the system, on average, spends on B. Theorem 2 is just a particular case of the more general Birkoff's Ergodic Theorem which we will state and prove in Chapter 4.

The moral of the story is that even though Lorenz's equations are difficult to describe from an analytic, numerical, or topological point of view, they are very well behaved from a probabilistic point of view. The tools and methods of probability theory are therefore very well suited to study and understand these equations and other similar dynamical systems. This is essentially the point of view on ergodic theory that we will take in these lectures. Since this is an introductory course, we will focus on the simplest examples of dynamical systems for which there is already an extremely rich and interesting theory, which are one-dimensional maps of the interval or the circle. However, the ideas and methods which we will present often apply in much more general situations and usually form the conceptual foundation for analogous results in higher dimensions. Indeed, results about interval maps are applied directly to higher dimensional systems. For example, Lorenz's equations can be studied taking a cross section for the flow and using Poincaré's first return map, which essentially reduces the system to a one dimensional map.

Chapter 2

Measure Theory

In this chapter, we will introduce the minimal requirements of Measure Theory which will be needed later. In particular, we will review one of the pillars of measure theory, namely, the concept of integral with respect to an arbitrary measure. For a more extensive exposition, the reader is encouraged to check, for example, with [2].

2.1 Motivation: Positive measures and Cantor sets

The notion of measure is, in the first instance, a generalization of the standard idea of length, or, in general, volume. Indeed, while we know how to define the length λ of an interval $[a, b]$, namely, $\lambda([a, b]) = b - a$, we do not a priori know how to measure the size of sets which contain no intervals but which, logically, have *positive measure*. For example, let $\{r_i\}_{i=0}^{\infty}$ be a sequence of positive numbers such that $\sum_i r_i < 1$. Define a set $C \subseteq [0, 1]$ recursively removing open subintervals from $[0, 1]$ in the following way. To start with, we remove an open subinterval I_0 of length r_0 from the interior of $[0, 1]$ so that $[0, 1] \backslash I_0$ has two connected components. Then we remove intervals I_1 and I_2 of lengths r_1 and r_2 respectively from the interior of these components so that $[0, 1] \backslash (I_0 \cup I_1 \cup I_2)$ has 4 connected components. Now remove intervals $I_3, ..., I_6$ from each of the interiors of these components and continue in this way. Let

$$C = [0, 1] \backslash \bigcup_{i=0}^{\infty} I_i \tag{2.1}$$

be a ***Cantor set***. By construction, C does not contain any intervals since every interval is eventually subdivided by the removal of one of the subintervals I_k from its interior. Therefore, it seems that it does not make sense to talk about C as having any *length*. Nevertheless, the total length of the intervals removed is $\sum_{i\geq 0} r_i < 1$ so it would make sense to say that the size of C is $1 - \sum_{i\geq 0} r_i$. Measure theory formalizes this notion in a rigorous way and makes it possible to assign a size to sets such as C.

Remark 3 *If $\sum_{i\geq 0} r_i = 1$ is exactly 1, then C is an example of a non-countable set of zero Lebesgue measure.*

Non-measurable sets

The example above shows that it is desirable to generalise the notion of *length* so that we can apply it to *measure* more complicated subsets which are not intervals. In particular, we would like to say that the Cantor set defined above has positive measure. It turns out that, in general, it is not possible to define a measure in a consistent way on all possible subsets. In 1924 Banach and Tarski showed that it is possible to divide the unit ball in 3-dimensional space into 5 parts and re-assemble these parts to form two unit balls, thus apparently doubling the volume of the original set. This implies that it is impossible to consistently assign a well defined volume to any subset in an additive way. See a very interesting discussion on wikipedia on this point (Banach-Tarski paradox).

Consider the following simpler example. Let $\mathbb{S}^1$ be the unit circle and let $f_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ be an irrational circle rotation. We will see that, in this case, every orbit is dense in $\mathbb{S}^1$ (Theorem 27) Let $A \subset \mathbb{S}^1$ be a set containing exactly one point from each orbit. Suppose that we have defined a general notion of a measure m on $\mathbb{S}^1$ that generalises the notion of length of an interval so that the measure $m(A)$ has a meaning. In particular, in order to be well-defined, such a measure will be translation invariant in the sense that the measure of a set cannot be changed by simply translating this set. Therefore, since a circle rotation f_α is just a translation, we have $m(f_\alpha^n(A)) = m(A)$ for every $n \in \mathbb{Z}$, where $f_\alpha^n := f_\alpha \circ \overset{n \text{ times}}{\cdots} \circ f_\alpha$. Moreover, since A contains only one single point from each orbit and all points on a given orbit are distinct, we have $f_\alpha^n(A) \cap f_\alpha^m(A) = \emptyset$ if $n \neq m$. Consequently,

$$m\left(f_\alpha^n(A) \bigcup f_\alpha^m(A)\right) = m(f_\alpha^n(A)) + m(f_\alpha^m(A)).$$

Therefore,

$$1 = m\left(\mathbb{S}^1\right) = m\left(\bigcup_{n=-\infty}^{\infty} f^n(A)\right) = \sum_{n=-\infty}^{\infty} m(f^n(A)) = \sum_{n=-\infty}^{\infty} m(A)$$

which is clearly impossible as the right hand side is zero if $m(A) = 0$ or infinity if $m(A) > 0$. In order to overcome this difficulty, one has to restrict the family of subsets which can be assigned a length consistently. This subsets will be called *measurable sets* and the family a *σ-algebra*.

Remark 4 *The previous counterexample depends on the Axiom of Choice to ensure that the set constructed by choosing a single point from each of an uncountable family of subsets exists.*

2.2 Measures and σ-algebras

Let X be a set and $\mathcal{A}$ a collection of (not necessarily disjoint) subsets of X.

Definition 5 *We say that $\mathcal{A}$ is an **algebra** (of subsets of X) if*

1. $\emptyset \in \mathcal{A}$ *and* $X \in \mathcal{A}$,
2. $A \in \mathcal{A}$ *implies* $A^c \in \mathcal{A}$,
3. *for any finite collection* $A_1, ..., A_n$ *of subsets in* $\mathcal{A}$ *we have that* $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$.

*We say that $\mathcal{A}$ is a σ**-algebra** if, additionally,*

3'. *for any countable collection* $\{A_i\}_{i\in\mathbb{N}}$ *of subsets in* $\mathcal{A}$, *we have*

$$\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}.$$

The family of all subsets of a set X is obviously a σ-algebra. Given $\mathcal{C}$ a family of subsets of X we define the ***σ-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$*** as the smallest σ-algebra containing $\mathcal{C}$. That is, as the intersection of all the σ-algebras containing $\mathcal{C}$. This is always well defined and is in general smaller than the σ-algebra of all subsets of X.

Exercise 6 Prove that the intersection of all the σ-algebras containing $\mathcal{C}$ is indeed a σ-algebra.

If X is a topological space, the σ-algebra generated by open sets is called the ***Borel σ-algebra*** and is denoted by $\mathcal{B}(X)$. Observe that a Cantor set C introduced in (2.1) is the complement of a countable union of open intervals and, therefore, belongs to Borel σ-algebra $\mathcal{B}([0,1])$.

Definition 7 *A real-valued set function μ on a class of sets $\mathcal{C}$ is called*

1. ***additive*** *if*
$$\mu\left(\bigcup\nolimits_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu\left(A_i\right)$$
for any finite sequence $\{A_1, ..., A_n\} \subset \mathcal{C}$ of pairwise disjoint sets such that $\bigcup_{i=1}^{n} A_i \in \mathcal{C}$.

2. ***countably additive*** *(or σ-additive) if*
$$\mu\left(\bigcup\nolimits_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu\left(A_i\right)$$
for any countably collection $\{A_i\}_{i\geq 1} \subseteq \mathcal{F}$ of pairwise disjoint sets such that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition 8 *Let $\mathcal{F}$ be a σ-algebra of X. A **measure** μ is a function*
$$\mu : \mathcal{F} \longrightarrow [0, \infty]$$
which is countably additive.

This definition shows that the σ-algebra is as intrinsic to the definition of a measure as the space itself. In general, we refer to a ***measure space*** as a triple $(X, \mathcal{F}, \mu)$. The elements in the σ-algebra $\mathcal{F}$ are called ***measurable sets***. We say that μ is ***finite*** if $\mu(X) < \infty$ and that μ is a ***probability measure*** if $\mu(X) = 1$. A measure μ is called ***σ-finite*** if $X = \bigcup_{i=1}^{\infty} A_i$ such that $A_i \in \mathcal{F}$ and $\mu(A_i) < \infty$ for any i. For example, the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is σ-finite because
$$\mathbb{R} = \bigcup\nolimits_{n\in\mathbb{Z}} [n, n+1]$$
because the Lebesgue measure of an interval is its length. On the other hand, observe that if $\hat{\mu}$ is a finite measure we can easily define a probability measure μ by
$$\mu(A) = \frac{\hat{\mu}(A)}{\hat{\mu}(X)}, \quad A \in \mathcal{F}.$$

Exercise 9 Let $\mathcal{F}$ be a σ-algebra and let $\mu : \mathcal{F} \to [0, \infty]$ be an additive positive function, $\mu \neq \infty$. Then,

1. μ is σ-additive $\Longleftrightarrow$ For any increasing sequence $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ (i.e., $A_n \subseteq A_{n+1}$) we have
$$\lim_{n\to\infty} \mu(A_n) = \mu(A) \quad \text{where } A := \bigcup\nolimits_{n\in\mathbb{N}} A_n.$$

2.
   (a) μ is σ-additive $\Longrightarrow$ For any decreasing sequence $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ (i.e., $A_{n+1} \subseteq A_n$) such that $\mu(A_1) < \infty$ we have
   $$\lim_{n\to\infty} \mu(A_n) = \mu(A) \quad \text{where } A := \bigcap\nolimits_{n\in\mathbb{N}} A_n.$$
   (b) If for any decreasing sequence $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ such that $A_n \searrow \emptyset$, we have $\lim_{n\to\infty} \mu(A_n) = 0$, then μ is σ-additive.

Defining a countably additive function on σ-algebras is non-trivial. It is usually easier to define countably additive functions on algebras because the class of sequences $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ such that $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ is smaller than in σ-algebras. Observe that, unlike what happens in σ-algebras, $\bigcup_{n\in\mathbb{N}} A_n$ needs not belong to $\mathcal{A}$ if $\mathcal{A}$ is only an algebra. For example, the standard length is a countably additive function on the algebra generated by finite unions of intervals. The fact that this extends to a countably additive function on the corresponding σ-algebra (and therefore, that we can measure Cantor sets) is guaranteed by the following fundamental result.

Theorem 10 (Carathéodory's Theorem, [3, Theorem 1.5.6]) *Let $\widetilde{\mu}$ be a countably additive function defined on an algebra $\mathcal{A}$ of subsets. Then $\widetilde{\mu}$ can be extended in a unique way to a countably additive function μ on the σ-algebra $\mathcal{F} = \sigma(\mathcal{A})$.*

Remark 11 The σ-additivity of $\widetilde{\mu}$ cannot be removed as the following counter-example shows. Let $\mathcal{A}$ be the algebra of sets $A \subset \mathbb{N}$ such that either A or $\mathbb{N}\backslash A$ is finite. For finite A, let $\mu(A) = 0$, and for A with a finite complement let $\mu(A) = 1$. Then μ is an additive, but not countably additive set function.

Proof. It is clear that A is indeed an algebra. $\mu(A \cup B) = \mu(A) + \mu(B)$ is obvious for disjoint sets A and B if A is finite. Finally, A and B in $\mathcal{A}$ cannot be infinite simultaneously being disjoint. If μ was countably additive, we would have
$$\mu(\mathbb{N}) = \sum_{n=1}^{\infty} \mu(\{n\}) = 0,$$
which is clearly a contradiction. ■

Nevertheless, defining measures on $\mathbb{R}$ is easier, as the next subsection summarises.

2.2.1 Measures on $\mathbb{R}$

In this subsection, we are going to gather some definitions and results that, roughly speaking, state that a measure on $\mathbb{R}$ is completely determined by the value of that measure on intervals of the form $(a, b]$, $a \leq b$.

Definition 12 *A* ***distribution function*** $F : \mathbb{R} \to \mathbb{R}$ *is a right-continuous increasing function. That is,*

$$x \leq y \Rightarrow F(x) \leq F(y) \ \text{ and } \lim_{x \to a^+} F(x) = F(a).$$

Let now $\mathfrak{J} := \{(a, b] : a \leq b \in \mathbb{R}\}$ and let $F : \mathbb{R} \to \mathbb{R}$ be a distribution function. Define

$$\begin{array}{rcl} \mu : \mathfrak{J} & \longrightarrow & [0, \infty] \\ (a, b] & \longmapsto & F(b) - F(a). \end{array} \tag{2.2}$$

Then, one can prove that μ thus defined is a σ-additive and σ-finite function on $\mathfrak{J}$. Moreover, there exists a unique σ-additive extension of μ onto $\mathcal{B}(\mathbb{R})$. That is,

Theorem 13 *A distribution function* $F : \mathbb{R} \to \mathbb{R}$ *determines a measure* μ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *by means of the formula*

$$\mu((a, b]) = F(b) - F(a), \quad (a, b] \in \mathfrak{J}.$$

A natural question now arises. Can we obtain any measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ from a distribution function? The answer is no, but *almost* any of them. Observe that the measure defined through (2.2) is finite on any bounded interval. These are precisely the measures we can generate by means of distribution functions. They are called Lebesgue-Stieljes measures.

Definition 14 *A* ***Lebesgue-Stieljes measure*** *is a measure* μ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *such that, for any bounded* $A \in \mathcal{B}(\mathbb{R})$, $\mu(A) < \infty$.

Proposition 15 *Let* μ *be a Lebesgue-Stieljes measure. Then, there exists a distribution function* F *such that*

$$\forall a < b \in \mathbb{R}, \quad F(b) - F(a) = \mu((a, b]). \tag{2.3}$$

Proof. Define $F(0) = c \in \mathbb{R}$ any arbitrary value and

$$F(x) := \begin{cases} F(0) + \mu((0,x]) & \text{if } x > 0 \\ F(0) - \mu((x,0]) & \text{if } x < 0. \end{cases}$$

F is a distribution function. It is clearly increasing and, by definition, satisfies (2.3). In order to check the right-continuity, let $b \geq 0$. Then

$$\lim_{x \to b^+} F(x) = F(0) + \lim_{x \to b^+} \mu((0,x]) = F(0) + \mu((0,b]) = F(b),$$

where in the second equality we have used that

$$\lim_{x \to b} \mu(A_x) = \mu(A_b) \quad \text{where } A_x = (0,x],\ A_x \searrow A_b \text{ as } x \to b$$

(see Exercise 9.2(a). The case $b < 0$ is analogous. ■

2.2.2 Examples

1. **Dirac delta measures**. Dirac measures δ_a, $a \in \mathbb{R}$, are defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as

$$\delta_a(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{if } a \notin A, \end{cases} \qquad A \in \mathcal{B}(\mathbb{R}).$$

   In this case, we say that the *entire mass* is concentrated at the single point a. The distribution function of δ_a is

$$F(x) = \begin{cases} 1 & \text{if } x \geq a \\ 0 & \text{if } x < a. \end{cases}$$

   An immediate generalization is the case of a measure concentrated on finite set of points $\{a_1, ..., a_n\}$ each of which carries some proportion $\rho_1, ..., \rho_n$ of the total mass, i.e., $\mu := \sum_{i=1}^n \rho_i \delta_{a_i}$ with $\rho_1 + ... + \rho_n = 1$. Then, given $A \in \mathcal{B}(\mathbb{R})$,

$$\mu(A) = \sum_{\{i : a_i \in A\}} \rho_i$$

   is the sum of the weights carried by those points contained in A.

2. **Lebesgue measure**. Lebesgue measure is defined on $\mathcal{B}(\mathbb{R})$ and assigns to any subinterval $I \subseteq \mathbb{R}$ its length. Lebesgue measure λ is characterised by the distribution function $F(x) = x$. That is,

$$\lambda((a,b]) = b - a.$$

3. **Absolutely continuous measures**. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. For any subinterval $I \subseteq \mathbb{R}$ define
$$\mu(I) := \int_I f(y)\, dy.$$
Then μ defines a σ-finitely additive function on the algebra of finite unions of subintervals of $\mathbb{R}$ and thus extends uniquely to a measure on $\mathcal{B}(\mathbb{R})$. Indeed, a possible distribution function F associated to μ is
$$F(x) = \int_0^x f(y)\, dy.$$

4. **Normal law**. The probability measure given by the distribution function
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \mathrm{e}^{-y^2/2}\, dy$$
is called the ***standard normal law*** and is denoted by $N(0,1)$.

5. **Measures on spaces of sequences**. Let Σ_k^+ denote the set of infinite sequences of k symbols. That is, and element $a \in \Sigma_k^+$ is a sequence $a = (a_0, a_1, ...)$ with $a_i \in \{0, 1, ..., k-1\}$. For any given *finite block* $(x_0, ..., x_{n-1})$ of length n with $x_i \in \{0, 1, ..., k-1\}$, let
$$I_{x_0 \ldots x_{n-1}} := \left\{ a \in \Sigma_k^+ : a_i = x_i,\ i = 0, ..., n-1 \right\}$$
denote the set of all infinite sequences which start precisely with the prescribed finite block $(x_0, ..., x_{n-1})$. We call this a ***cylinder set*** of order n. Let
$$\mathcal{A} = \{\text{finite unions of cylinder sets}\}.$$

Exercise 16 Show that $\mathcal{A}$ is an algebra of subsets of Σ_k^+.

Fix now k numbers $\{p_0, ..., p_{k-1}\} \subset [0,1]$ such that $p_0 + ... + p_{k-1} = 1$ and define a function $\mu : \mathcal{A} \to \mathbb{R}_+$ on the algebra of cylinder sets by
$$\mu\left(I_{x_0 \ldots x_{n-1}}\right) := \prod_{i=0}^{n-1} p_{x_i}.$$

Exercise 17 Prove that μ is σ-additive.

Therefore, the function μ extends uniquely to a measure on the σ-algebra $\mathcal{F} = \sigma(\mathcal{A})$.

2.3 Integration

Integration with respect to a measure can be regarded as a powerful generalization of the standard Riemann integral. In this section, we are going to review the basics of integration with respect to an arbitrary measure. Before, we need to introduce some definitions.

Let $(X, \mathcal{F}, \mu)$ be a measure space and let $A \in \mathcal{F}$. We define the ***characteristic function*** $\mathbf{1}_A$ as

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

On the other hand, a ***simple*** or ***elementary function*** $\zeta : X \to \mathbb{R}$ is a function that can be written in the form

$$\zeta = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}$$

for some constants $c_i \in \mathbb{R}$ and some disjoint measurable sets $A_i \in \mathcal{F}$, $i = 1, ..., n$. The integral of a simple function ζ with respect to the measure μ is defined in a straightforward manner as

$$\int \zeta d\mu := \sum_{i=1}^{n} c_i \mu(A_i).$$

The idea is to extend this integral to more general functions. More concretely, we can define the integral of a measurable function. Recall that a function $f : X \to \mathbb{R}$ is ***measurable*** if $f^{-1}(I) \in \mathcal{F}$ for any $I \in \mathcal{B}(\mathbb{R})$. If $(X, \mathcal{F}, \mu)$ is a probability space (i.e., μ is a probability), measurable functions are usually called ***random variables***.

Exercise 18 *Let $f : X \to \mathbb{R}_+$ be a measurable function. Show that f is the (pointwise) limit of an increasing sequence of elementary functions.* ***Hint:*** *define, for any $n \in \mathbb{N}$,*

$$\zeta_n := \sum_{k=1}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{\left\{\frac{k}{2^n} \le f \le \frac{k+1}{2^n}\right\}} + n\mathbf{1}_{\{n \le f\}}.$$

The integral of a general, measurable, *non-negative* function $f : \mathbb{R} \to \mathbb{R}_+$ can be defined in two equivalent ways. On the one hand,

$$\int f d\mu := \sup \left\{ \int \zeta d\mu : \zeta \text{ is simple, } \zeta \le f \right\}. \tag{2.4}$$

On the other hand, one can prove that

$$\int f d\mu = \lim_{n\to\infty} \int \zeta_n d\mu$$

where $\{\zeta_n\}_{n\in\mathbb{N}}$ is an increasing sequence of elementary functions converging to f. The integral is usually called the ***Lebesgue integral*** of the function f with respect to the measure μ (even if μ is not Lebesgue measure). Observe that $\int f d\mu$ may be ∞.

Remark 19 *Note that, unlike the Riemann integral which is defined by a limiting process that may or may not converge, the supremum in (2.4) is always well defined, though it needs not be finite.*

In general, let $f : X \to \mathbb{R}$ be a measurable function and write

$$f = f^+ - f^-$$

where $f^+ = \max\{f, 0\}$ and $f^- = \min\{-f, 0\}$. It is not difficult to prove that both f^+ and f^- are measurable, non-negative functions.

Definition 20 *Let $f : X \to \mathbb{R}$ be a measurable function. If*

$$\int f^+ d\mu < \infty \quad \textit{and} \quad \int f^- d\mu < \infty$$

*we say that f is μ-**integrable** and we define*

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

The set of all μ-integrable functions is denoted by $L^1(X, \mu)$.

2.3.1 Properties of the Lebesgue integral

Let $f, g : X \to \mathbb{R}$ be two arbitrary measurable functions. The Lebesgue integral has the following properties (that, in general, are not difficult to prove):

1.
$$\int (af + bg)\, d\mu = a \int f d\mu + b \int g d\mu$$

   for any $a, b \in \mathbb{R}$.

2. If $A \in \mathcal{F}$ is such that $\mu(A) = 0$, then

$$\int_A f d\mu := \int \mathbf{1}_A f d\mu = 0.$$

That is, the integral of f over a set of measure 0 is 0. This is true even if f takes the values $\pm\infty$ on A. That is, if A contains singularities of f. Recall that we say that a point $x = a$ is a ***singularity*** if $f(a) = \pm\infty$. For example, if a non-negative function $f \geq 0$ is integrable, $\int f d\mu < \infty$, then we can say that $\mu(\{x : f(x) = \infty\}) = 0$.

3. If $f \geq 0$ and $\int f d\mu = 0$, then $\mu(\{x : f(x) > 0\}) = 0$.

4. If $f \leq g$, then

$$\int f d\mu \leq \int g d\mu.$$

5.
$$\left| \int f d\mu \right| \leq \int |f| \, d\mu. \tag{2.5}$$

Indeed, $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$ and

$$\begin{aligned} \left| \int f d\mu \right| &= \left| \int f^+ d\mu - \int f^- d\mu \right| \leq \left| \int f^+ d\mu \right| + \left| \int f^- d\mu \right| \\ &= \int f^+ d\mu + \int f^- d\mu = \int |f| \, d\mu. \end{aligned}$$

By (2.5), we can characterise $L^1(X, \mu)$ as the space of measurable functions $f : X \to \mathbb{R}$ such that

$$\int |f| \, d\mu < \infty.$$

In general, for $p \geq 1$, we introduce the spaces $L^p(X, \mu)$ as the space of measurable functions such that $\int_X |f|^p \, d\mu < \infty$, where two functions are identified if they differ, at most, on a set of zero measure. $L^p(X, \mu)$ is a *Banach space* with the norm $\|f\|_p := (\int_X |f|^p \, d\mu)^{1/p}$.

Example 21 Let $f : [0, 1] \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \\ 1 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

It is well known that this function is not Riemann integrable because the limit of the upper and lower Riemann sums do not coincide. However, as a function measurable with respect to the Lebesgue measure λ, f is simple: it values 0 on the measurable set $\mathbb{Q}$ and values 1 on the measurable set $[0,1]\backslash\mathbb{Q}$. The set of rational numbers $\mathbb{Q}$ has zero Lebesgue measure because $\mathbb{Q}$ is countable. Therefore $\lambda([0,1]\backslash\mathbb{Q}) = 1$ and

$$\int_{[0,1]} f d\lambda = \lambda\left([0,1]\backslash\mathbb{Q}\right) = 1.$$

Chapter 3

Invariant measures

3.1 Invariant measures: definitions and examples

Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be two measure spaces. A map $T : X \to Y$ is called ***measurable*** if the preimage $T^{-1}(A)$ of any measurable set $A \in \mathcal{G}$ is measurable, i.e., $T^{-1}(A) \in \mathcal{F}$. A measurable map T is ***non-singular*** if the preimage of every set of measure 0 has measure 0. The map $T : X \to Y$ is ***measure-preserving*** if $\mu\left(T^{-1}(A)\right) = \nu(A)$ for any $A \in \mathcal{G}$. A non-singular map from a measure space $(X, \mathcal{F}, \mu)$ into itself is called a ***non-singular transformation***, or simply a ***transformation***. If a transformation $T : X \to X$ preserves a measure μ, then μ is called T***-invariant***. Usually, we will deal with measurable maps between topological spaces. In that case, the σ-algebras involved will be always the corresponding Borel σ-algebras.

A set has ***full measure*** if its complement has measure 0. We say that a property holds for μ***-almost every*** x (μ-a.e.) or μ***-almost surely*** (μ-a.s.) if it holds on a subset of full μ-measure.

Let λ be the Lebesgue measure on $\mathbb{R}$. A ***flow*** $\{T^t\}_{t \in J}$ on a measurable space $(X, \mathcal{F}, \mu)$ is a family of measurable maps $T^t : X \to X$ where, usually, J equals $\mathbb{R}$ (time-continuous flows) or $\mathbb{N}$ (discrete flows). If $J = \mathbb{N}$, we say that the flow $\{T^n\}_{n \in \mathbb{N}}$ is ***measurable*** if T^n is measurable for any n. When $J = \mathbb{R}$, $\{T^t\}_{t \in \mathbb{R}}$ is ***measurable*** if the product map $T : X \times \mathbb{R} \to X$ given by $T(x, t) = T^t(x)$ is measurable with respect to the product σ-algebra $\sigma(\mathcal{F} \times \mathcal{B}(\mathbb{R}))$ on $X \times \mathbb{R}$, and $T^t : X \to X$ is a non-singular measurable transformation for any $t \in \mathbb{R}$. A measurable flow T^t is a ***measure-preserving flow*** if each T^t is a measure-preserving transformation. Discrete flows are usually built from measurable maps $T : X \to X$ as follows: for any $n \in \mathbb{N}$, we define

$$T^n(x) = T \circ \overset{n)}{\cdots} \circ T(x)$$

and $T^0 = \mathrm{Id}$, the identity on X.

3.1.1 Examples

1. **Dirac measures on fixed points**. If $T : X \to X$ is a measurable map and p a fixed point of T, $T(p) = p$, then the Dirac measure δ_p is invariant. Indeed, let $A \in \mathcal{F}$ be a arbitrary measurable set. We have to prove that

$$\delta_p\left(T^{-1}(A)\right) = \delta_p(A). \tag{3.1}$$

   We consider two cases. First of all, suppose $p \in A$ so that $\delta_p(A) = 1$. In this case $p \in T^{-1}(A)$ clearly so $\delta_p\left(T^{-1}(A)\right) = 1$ and (3.1) holds. Secondly, suppose that $p \notin A$. Then $\delta_p(A) = 0$ and we also have $p \notin T^{-1}(A)$ because if $p \in T^{-1}(A)$ then $p = T(p) \in T\left(T^{-1}(A)\right) \subseteq A$, which would be a contradiction. Therefore $p \notin T^{-1}(A)$, $\delta_p\left(T^{-1}(A)\right) = 0$, and (3.1) holds again.

2. **Dirac measures on periodic orbits**. Let $T : X \to X$ be a measurable map and let $P = \{a_1, ..., a_n\}$ be a periodic orbit with minimal period n. That is, $T(a_i) = a_{i+1}$ for $i = 1, ..., n-1$ and $T(a_n) = a_1$. Let $\rho_1, ..., \rho_n$ be constants such that $\rho_i \in (0,1)$ and $\sum_{i=1}^n \rho_i = 1$. Consider the measure

$$\delta_P(A) = \sum_{\{i : a_i \in A\}} \rho_i.$$

   Exercise 22 Show that δ_P is invariant if and only if $\rho_i = 1/n$ for every $i = 1, ..., n$.

3. **Circle rotations**.

   Proposition 23 *Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a circle rotation, $T(x) = x + \alpha$ for some $\alpha \in \mathbb{R}$. The Lebesgue measure is invariant.*

   Proof. T is just a translation and Lebesgue measure is invariant under translations. ■

   However, depending on the value of α, there may be other invariant measures. If $2\pi/\alpha$ is rational then all points $x \in \mathbb{S}^1$ are periodic of the same period and, therefore, T admits also infinitely many distinct

Dirac measures on the periodic orbits (see Example 26). If $2\pi/\alpha$ is irrational, then all orbits are dense in $\mathbb{S}^1$ (Example 26) and the Lebesgue measure is the unique invariant measure of T.

4. **Measure-preserving flows in** $\mathbb{R}^n$. Let $U \subseteq \mathbb{R}^n$ be an open set and $v : U \to \mathbb{R}^n$ a C^r vector field, $r \geq 1$. Consider the differential equation

$$\dot{x} = v(x). \tag{3.2}$$

Suppose that, for every $p \in U$, there exists a (unique) solution $x : \mathbb{R} \to U$ of (3.2) with initial condition p, which means that, $\dot{x}_t = v(x_t)$ and $x_{t=0} = p$. For any $t \in \mathbb{R}$, we define the map $\varphi_t : U \to U$ by $\varphi_t(p) = x_t$ where $x : \mathbb{R} \to U$ is the solution of (3.2) with initial condition $x_0 = p$. Basic results of ordinary differential equations show that, for every t, the map φ_t is a C^r diffeomorphism and the family of maps $\varphi_t : U \to U$ defines a one-parameter group, i.e., $\varphi_{t=0} = \mathrm{Id}$ (the identity) and $\varphi_{t+s} = \varphi_t \circ \varphi_s$ for any $t, s \in \mathbb{R}$. Moreover, by Liouville's formula,

$$\det \left| \frac{\partial \varphi_t}{\partial x^i}(p) \right| = \exp\left(\int_0^t \operatorname{div} v\left(\varphi_s(p)\right) ds \right)$$

for any $p \in U$ and t. Hence, if we assume $\operatorname{div} v = 0$, we have $\det \left| \frac{\partial \varphi_t}{\partial x^i}(p) \right| = 1$ and φ_t preserves the *n-dimensional volume* (or Lebesgue measure). Hamiltonian vector fields are examples of vector fields that satisfy $\operatorname{div} v = 0$. Recall that a vector field is called ***Hamiltonian*** if $n = 2m$ is an even number and there exists a function $H : U \to \mathbb{R}$ such that, denoting the points in $\mathbb{R}^n$ as $(q_1, ..., q_m, p_1, ..., p_m)$,

$$v = \left(\frac{\partial H}{\partial p_1}, ..., \frac{\partial H}{\partial p_m}, -\frac{\partial H}{\partial q_1}, ..., -\frac{\partial H}{\partial q_m} \right).$$

Exercise 24 Complete the proof and show that $\operatorname{div} v = 0$ implies that the flow φ_t associated to v preserves the Lebesgue measure.

3.2 Poincaré's recurrence Theorem

Invariant measures play a fundamental role in dynamics. As a first example, we state and prove the following famous result by Poincaré which implies that recurrence is a generic property of orbits of measure-preserving dynamical systems.

Theorem 25 (Poincaré's Recurrence Theorem) *Let $(X, \mathcal{F}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving map. Let $A \in \mathcal{F}$ such that $\mu(A) > 0$. Then, μ-almost every point $x \in A$, there exists some $n \in \mathbb{N}$ such that $T^n(x) \in A$. Consequently, there are infinitely many $k \in \mathbb{N}$ for which $T^k(x) \in A$; a point $x \in A$ returns to A infinitely often.*

Proof. Let

$$B := \{x \in A : T^k(x) \notin A \text{ for all } k \in \mathbb{N}\} = A \Big\backslash \bigcup_{k \in \mathbb{N}} T^{-k}(A).$$

Then $B \in \mathcal{F}$ and all the preimages $T^{-k}(B)$ are measurable, have the same measure as B, and *disjoint*. Indeed, suppose that

$$T^{-n}(B) \bigcap T^{-m}(B) \neq \emptyset, \quad n \neq m, \ n > m.$$

That is, there exists some $x \in T^{-n}(B) \bigcap T^{-m}(B)$ such that

$$\begin{aligned} T^m(x) &\in T^m\left(T^{-n}(B) \bigcap T^{-m}(B)\right) \\ &= T^m\left(T^{-n}(B)\right) \bigcap T^m\left(T^{-m}(B)\right) = T^{n-m}(B) \bigcap B. \end{aligned}$$

But this implies that $T^{n-m}(B) \bigcap B \neq \emptyset$ which contradicts the definition of B.

Now, since X has finite total measure, it follows that B has measure 0. Actually,

$$\infty > \mu(X) \geq \mu\left(\bigcup_{k \in \mathbb{N}} T^{-k}(B)\right) = \sum_{k \in \mathbb{N}} \mu\left(T^{-k}(B)\right) = \sum_{k \in \mathbb{N}} \mu(B),$$

which implies $\mu(B) = 0$. In other words, $\mu(A) = \mu(A \backslash B)$ and every point in $A \backslash B$ returns to A, which proves the first assertion.

To show that almost every point of A returns to A infinitely often let

$$\widetilde{B}_n := \{x \in A : T^n(x) \in A \text{ and } T^k(x) \notin A, \ k > n\}, \quad n \geq 1,$$

denote the set of points which return to A for the last time after exactly n iterations. We will show that $\mu(\widetilde{B}_n) = 0$ for any $n \geq 1$ so that the set

$$\widetilde{B} := \bigcup_{n \geq 1} \widetilde{B}_n \subset A$$

of the points with return to A only finitely many times has measure 0 as well. Indeed, consider the set $T^n(\widetilde{B}_n) \subset B$, which is by definition contained in A

and consists of points that never return to A. Therefore, $\mu(T^n(\widetilde{B}_n)) = 0$. But

$$\widetilde{B}_n \subseteq T^{-n}\left(T^n(\widetilde{B}_n)\right).$$

Consequently, using that μ is T-invariant we have

$$\mu(\widetilde{B}_n) \leq \mu\left(T^{-n}\left(T^n(\widetilde{B}_n)\right)\right) = \mu\left(T^n(\widetilde{B}_n)\right) = 0.$$

■

The conclusion of Poincaré's Recurrence Theorem may be useless if the preserved measure μ has no *physical* meaning. For example, if $p \in X$ is a fixed point, i.e., $T(p) = (p)$, then the Dirac measure δ_p is invariant. However, with respect to this measure, any set that does not contain p has measure zero so we cannot state anything about the recurrence properties of the systems (except for p, which is a fixed point).

On the other hand, Poincaré's Recurrence Theorem leads us to some paradoxical conclusions. For example, particle dynamics are ruled by Hamiltonian vector fields, which preserve the Lebesgue volume of the phase space. If we open a partition separating a chamber containing gas and a chamber with a vacuum, then Poincaré's Theorem implies that, after a while, the gas molecules will again collect in the first chamber. This is because there exists a set of strictly positive Lebesgue measure in $\mathbb{R}^{3N} \times \mathbb{R}^{3N}$ whose points correspond to the positions and velocities of N particles in the first chamber. The resolution of this paradox lies in the fact that *a while* may be longer than the duration of the solar's system existence. And, of course, that particle dynamics are described at a microscopic level by quantum mechanics, whose effects cannot be taken into account deterministically.

Example 26 Let $T : \mathbb{S}^1 \to \mathbb{S}^1$ be a circle rotation of angle $\alpha \in \mathbb{R}$. If $\alpha = 2\pi\frac{m}{n}$, $m, n \in \mathbb{N}$, then T^n is the identity and Theorem 25 is obvious. If α is not commensurable with 2π, then Poincaré's Recurrence Theorem gives

$$\forall\delta > 0, \; \exists n \in \mathbb{N} \text{ such that } |T^n(x) - x| < \delta.$$

It easily follows that

Theorem 27 *If $\alpha \neq 2\pi\frac{m}{n}$, then the orbit $\{T^k(x), \; k = 1, 2, ...\}$ is dense on the circle $\mathbb{S}^1$.*

Exercise 28 Prove Theorem 27.

3.3 Invariant measures for continuous maps

In this section, we show that a continuous map $T : X \to X$ of a compact metric space X into itself has at least one invariant Borel probability measure.

Theorem 29 (Krylov-Bogolubov) *Let X be a compact metric space and $T : X \to X$ a continuous map. Then there exists a T-invariant Borel probability μ on X.*

Exercise 30 The compactness condition is essential here. Consider the open interval $I = (0,1)$ and the map $T : (0,1) \to (0,1)$ given by $T(x) = x/2$. Show that T admits no invariant probabilities.

We will prove Krylov-Bogolubov's Theorem in several steps. Before proceeding, we need to introduce some definitions.

Let $\mathcal{M}$ denote the set of all Borel probability measures on a topological space $(X, \mathcal{F})$, $\mathcal{F} = \mathcal{B}(X)$. A sequence of measures $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}$ converges in the ***weak* topology*** to a measure $\mu \in \mathcal{M}$ if $\int_X f d\mu_n \to \int_X f d\mu$ for any continuous function $f \in C(X)$ as $n \to \infty$. A measurable map $T : X \to X$ induces a map $T_* : \mathcal{M} \to \mathcal{M}$ defined by $(T_*\mu)(A) := \mu\left(T^{-1}(A)\right)$, $A \in \mathcal{F}$. We call $T_*\mu$ the ***push-forward*** of μ. Similarly, we can define $(T_*^n\mu)(A) = \mu(T^{-n}(A))$. Obviously, μ is T-invariant if and only if $T_*\mu = \mu$.

This notion of convergence is called *weak-star convergence* because the space of finite Borel measures can be canonically identified with the space of linear functionals on the space of continuous functions, i.e. with the dual space of continuous functions. Actually, every finite Borel measure μ on X defines a bounded linear functional $L_\mu(f) = \int_X f d\mu$ on the space $C_c(X)$ of continuous functions on X with compact support. Furthermore, L_μ is positive in the sense that $L_\mu(f) \geq 0$ if $f \geq 0$. The Riesz Representation Theorem (see [9, Theorem 2.14]) states that the converse is also true: for every positive bounded linear functional L on $C_c(X)$, there is a finite Borel measure μ on X such that $L(f) = \int_X f d\mu$, $f \in C_c(X)$. If X is compact, then trivially $C(X) = C_c(X)$.

For an arbitrary measure $\mu_0 \in \mathcal{M}$, we define the sequence of measures

$$\mu_n := \frac{1}{n}\sum_{i=0}^{n-1} T_*^i \mu_0. \tag{3.3}$$

Our aim is to show that the sequence μ_n converges in the weak* topology to a probability measure μ and that this measure is invariant. In order to do that, we will need a couple of Lemmas.

Lemma 31 *$\mathcal{M}$ is compact in the weak* topology.*

Proof (sketch). $C(X)$ is a Banach space with the norm of the supremum

$$\|f\|_{\infty} = \sup_{x \in X} f(x), \quad f \in C(X).$$

Its dual space, $C(X)^*$ is a normed space with the norm

$$\|L\| = \sup\{|L(f)| : \|f\|_{\infty} = 1, \ f \in C(X)\}$$

such that the set of probabilities $\mathcal{M} \subset C(X)^*$ is contained in the unit ball

$$\{L \in C(X)^* : \|L\| = 1\}.$$

Indeed, if $\|f\|_{\infty} = 1$ then $f \leq 1$ a.s. so that, for any measure $\mu \in C(X)^*$,

$$L_{\mu}(f) = \int_X f d\mu \leq \int_X 1 d\mu, \ \ \|f\|_{\infty} = 1,$$

by monotonicity of the Lebesgue integral. Therefore, since the constant function 1 has norm $\|1\|_{\infty} = 1$ one,

$$\|\mu\| = \sup\left\{\left|\int_X f d\mu\right| : \|f\|_{\infty} = 1, \ f \in C(X)\right\} = \int_X 1 d\mu = \mu(X),$$

which implies that the set of probability measures $\mathcal{M}$ is contained in the unit ball of $C(X)^*$.

Moreover, $\mathcal{M}$ is closed. To prove this statement, let $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathcal{M}$ be a sequence of probabilities that, regarded as a sequence of linear functionals $\{L_{\mu_n}\}_{n \in \mathbb{N}}$, converge weakly to a linear map $L \in C^*(X)$. We need to show that $L = L_{\mu}$ for some $\mu \in \mathcal{M}$. Since L_{μ_n} are positive,

$$L_{\mu_n}(f) \geq 0 \ \ \forall n \ \ \text{if} \ \ f \geq 0,$$

then taking the limit $n \to \infty$ we conclude $L(f) \geq 0$ if $f \geq 0$, so L is positive. By the Riesz Representation Theorem, there exists a finite Borel measure μ such that $L = L_{\mu}$. But $\{\mu_n\}_{n \in \mathbb{N}}$ are probabilities, so

$$1 = L_{\mu_n}(1) = \mu_n(X) \quad \forall n.$$

Letting $n \to \infty$, we also conclude that $1 = L_{\mu}(1)$, which implies that μ is a probability as required.

Finally, by the Banach-Alaoglu Theorem, we have that the unit ball of $C^*(M)$ is compact. Since $\mathcal{M}$ is a closed set of a compact one, $\mathcal{M}$ is compact too:

Theorem 32 (Banach-Alaoglu, [6, Chapter V, §4]) *Let E be a normed space. Then, the closed balls in E^* are compact in the weak* topology.*

■

Lemma 33 *For all integrable functions $f : X \to \mathbb{R}$ we have*

$$\int_X f d(T_*\mu) = \int_X (f \circ T) d\mu. \tag{3.4}$$

Proof. We first prove that the statement holds for characteristic functions. If $f = \mathbf{1}_A$, $A \in \mathcal{F}$, then

$$\int_X \mathbf{1}_A d(T_*\mu) = (T_*\mu)(A) = \mu\left(T^{-1}(A)\right) = \int_X \mathbf{1}_{T^{-1}(A)} d\mu = \int_X \mathbf{1}_A \circ T d\mu.$$

Obviously, (3.4) also holds if f is a simple function, that is, a linear combination of characteristic functions. Now suppose that f is a non-negative integrable function. By Exercise 18, there exists an increasing sequence $\{f_n\}_{n\in\mathbb{N}}$ of simple functions converging to f. Moreover, the sequence $\{f_n \circ T\}_{n\in\mathbb{N}}$ is clearly an increasing sequence of simple functions converging to $f \circ T$. By the definition of Lebesgue integral, we have

$$\int_X f_n \circ T d\mu \underset{n\to\infty}{\longrightarrow} \int_X f \circ T d\mu \quad \text{and} \quad \int_X f_n d(T_*\mu) \underset{n\to\infty}{\longrightarrow} \int_X f d(T_*\mu).$$

Since we already proved that $\int_X f_n \circ T d\mu = \int_X f_n d(T_*\mu)$ for simple functions and the limit of a sequence is unique, we conclude that

$$\int_X f \circ T d\mu = \int_X f d(T_*\mu).$$

For a general measurable f, we repeat the same argument for the positive f^+ and negative f^- parts of f. ■

Lemma 34 *$T_* : \mathcal{M} \to \mathcal{M}$ is continuous.*

Proof. Suppose $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}$ is a sequence converging to μ, i.e., $\mu_n \to \mu$ as $n \to \infty$. Then, by the definition of convergence in the weak* topology, for any continuous function $f : X \to \mathbb{R}$ we have

$$\int_X f d(T_*\mu_n) = \int_X f \circ T d\mu_n \underset{n\to\infty}{\longrightarrow} \int_X f \circ T d\mu = \int_X f d(T_*\mu)$$

where we have used Lemma 33. In other words, $T_*\mu_n \to T_*\mu$ as $n \to \infty$, which, in turn, implies that T is continuous. ∎

Proof of Theorem 29. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be the sequence of measure defined in (3.3). The compactness of $\mathcal{M}$ implies, in particular, that the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ has a converging subsequence $\{\mu_{n_j}\}_{j\in\mathbb{N}}$. Define

$$\mu := \lim_{j\to\infty} \mu_{n_j}.$$

We will show that μ is invariant.

One the one hand, using the linearity of T_*, we have

$$\begin{aligned} T_*\mu_{n_j} &= T_*\left(\frac{1}{n_j}\sum_{i=0}^{n_j-1} T_*^i\mu_0\right) = \frac{1}{n_j}\sum_{i=0}^{n_j-1} T_*^{i+1}\mu_0 \\ &= \frac{1}{n_j}\sum_{i=0}^{n_j-1} T_*^i\mu_0 - \frac{1}{n_j}\mu_0 + \frac{1}{n_j}T_*^{n_j}\mu_0 \\ &= \mu_{n_j} - \frac{1}{n_j}\mu_0 + \frac{1}{n_j}T_*^{n_j}\mu_0. \end{aligned}$$

Since the last two terms of this expression tend to 0 as $j \to \infty$, we conclude that

$$T_*\mu_{n_j} \underset{n\to\infty}{\longrightarrow} \mu.$$

On the other hand, from the continuity of T (Lemma 34), we have $T_*\mu_{n_j} \to T_*\mu$. That is, $T_*\mu = \mu$ and μ is an invariant measure. ∎

Chapter 4

Birkhoff's Ergodic Theorem

In this chapter, we will state and prove Birkhoff's Ergodic Theorem which, in the case of ergodic measures, gives a much more sophisticated statement about recurrence than Poincaré's Recurrence Theorem. We will review the examples of the previous chapter and discuss the ergodicity of their measures. Finally, we will conclude with a section on the existence of ergodic measures in general.

4.1 Ergodic transformations.

Throughout this section, $(X, \mathcal{F}, \mu)$ will denote a measure space and $T : X \to X$ a measure-preserving transformation.

Definition 35 *We say that T or μ is **ergodic** if, for any $A \in \mathcal{F}$ such that $T^{-1}(A) = A$, then A has either measure 0 or full measure.*

Exercise 36 Show that $T^{-1}(A) = A$ implies $T(A) = A$ but that the converse is not true in general.

A set $A \in \mathcal{F}$ such that $T^{-1}(A) = A$ is called ***T-invariant***. A measurable function $f : X \to \mathbb{R}$ is called ***T-invariant*** if $f \circ T = f$ almost everywhere.

Proposition 37 *A measurable transformation is ergodic if and only if every invariant measurable function is constant a.e..*

Proof. Exercise. (**Hint:** see the proof of Proposition 38). ■

Therefore, the ergodicity a transformation $T : X \to X$ can be characterised saying that T-invariant measurable functions are constant a.e..

However, when $(X, \mathcal{F}, \mu)$ is a finite measure space, we can use this characterisation in a smaller set of functions.

Proposition 38 *Suppose that $\mu(X) < \infty$. The following properties are equivalent:*

1. *T is ergodic.*
2. *If $f \in L^p(X, \mu)$ is T-invariant, $p \geq 1$, then f is constant almost everywhere.*

Proof. $2 \Longrightarrow 1$. If $A \in \mathcal{F}$ is T-invariant, the characteristic function $\mathbf{1}_A$ is T-invariant and belongs to $L^p(X, \mu)$. Therefore, $\mathbf{1}_A$ is constant a.e.. That is, $\mu(A) = 0$ or 1.

$1 \Longrightarrow 2$. If $f \in L^p(X, \mu)$ is T-invariant, the set $A_c := \{x : f(x) \leq c\}$ is invariant for each $c \in \mathbb{R}$. Since T is ergodic, this means that $\mu(A_c)$ is either 0 or 1.

Exercise 39 *Show that this implies that f is constant almost everywhere.*

■

4.2 Conditional Expectation

Let $(X, \mathcal{F}, \mu)$ be a probability space. That is, $\mu(X) = 1$. A measurable function $f : X \to \mathbb{R}$ between the measurable spaces $(X, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called a ***random variable***. If f is a random variable, $\mathrm{E}[f]$ will denote $\int_X f d\mu$ and we will call this integral the ***expectation*** or ***mean value*** of f.

Definition 40 *Let $f : X \to \mathbb{R}$ be a real-valued random variable such that $\mathrm{E}[|f|] < \infty$ and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$, $\mathcal{G} \subseteq \mathcal{F}$. The **conditional expectation** of f with respect to $\mathcal{G}$ is a $\mathcal{G}$-measurable random variable f' such that*

$$\int_A f d\mu = \int_A f' d\mu \quad \text{for any } A \in \mathcal{G}. \tag{4.1}$$

We denote f' by $\mathrm{E}[f|\mathcal{G}]$.

The existence of $\mathrm{E}[f|\mathcal{G}]$ is not a trivial issue and it is based on the Radon-Nikodym's Theorem, one of the most important theorems in measure theory. The explicit computation of $\mathrm{E}[f|\mathcal{G}]$ can be carried out in some particular situations.

Example 41 Let $\mathcal{G}$ be the σ-algebra generated by a finite partition $A_1, ..., A_n$ of X and suppose that $\mu(A_i) > 0$ for any $i = 1, ..., n$. Then

$$\mathrm{E}[f|\,\mathcal{G}] = \sum_{i=1}^{n} \frac{\mathrm{E}[f\mathbf{1}_{A_i}]}{\mu(A_i)} \mathbf{1}_{A_i}.$$

In particular if $\mathcal{G} = \{\emptyset, X, A, A^c\}$ is the σ-algebra generated by $A \in \mathcal{F}$, then

$$\mathrm{E}[f|\,\mathcal{G}] = \frac{\mathrm{E}[f\mathbf{1}_A]}{\mu(A)} \mathbf{1}_A + \frac{\mathrm{E}[f\mathbf{1}_{A^c}]}{\mu(A^c)} \mathbf{1}_{A^c}.$$

Example 42 If $\mathcal{G}$ is the trivial σ-algebra, $\mathcal{G} = \{\emptyset, X\}$, then $\mathrm{E}[f|\,\mathcal{G}] = \mathrm{E}[f]$ for any random variable f. Indeed, on the one hand, only constants are measurable with respect to the trivial σ-algebra; on the other hand, (4.1) implies the conditional expectation to be equal to the mean value of f.

We review Radon-Nikodym's Theorem for the benefit of a clearer exposition. Recall that given two finite measures μ and ν on a measurable space $(X, \mathcal{F})$, we say that ν is ***absolutely continuous*** with respect to μ if $\mu(A) = 0$ implies $\nu(A) = 0$, where $A \in \mathcal{F}$. We will write $\nu \ll \mu$.

Theorem 43 (Radon-Nikodym's Theorem) *Let μ and ν be two finite measures on $(X, \mathcal{F})$ such that ν is absolutely continuous with respect to μ. Then, there exists an essentially unique measurable function $g : X \to \mathbb{R}$ such that*

$$\nu(A) = \int_A g\,d\mu. \tag{4.2}$$

*The density g is denoted by $\frac{d\nu}{d\mu}$ and is usually called the **Radon-Nikodym derivative**.*

Essentially unique in Radon-Nykodym's Theorem means that any two functions satisfying (4.2) may only differ on a set of μ-measure 0.

Proposition 44 *With the same notation as in Definition 40, the conditional expectation $\mathrm{E}[f|\,\mathcal{G}]$ exists and is essentially unique.*

Proof. We continue denoting by μ the restriction of μ to $\mathcal{G}$ and define the measure ν on $\mathcal{G}$ by

$$\nu(A) = \int_A f\,d\mu, \quad A \in \mathcal{G}.$$

It is clear that ν is absolutely continuous with respect to μ. Its Radon-Nikodym derivative is then the required conditional expectation. The uniqueness follows from the uniqueness statement in the Radon-Nikodym's Theorem. ■

4.2.1 Properties of the conditional expectation

1. **Linearity:** for any two random variables $f, g : X \to \mathbb{R}$ and two real numbers $a, b \in \mathbb{R}$,
$$\mathrm{E}\left[af + bg \middle| \mathcal{G}\right] = a\,\mathrm{E}\left[f \middle| \mathcal{G}\right] + b\,\mathrm{E}\left[g \middle| \mathcal{G}\right].$$
This property follows from the definition of the conditional expectation and that of the integral.

2. **monotonicity:** If $g \leq f$ then $\mathrm{E}\left[g \middle| \mathcal{G}\right] \leq \mathrm{E}\left[f \middle| \mathcal{G}\right]$.

   Exercise 45 Show the monotonicity property using that, if two $\mathcal{G}$-measurable random variables $f_1, f_2 : X \to \mathbb{R}$ satisfy $\mathrm{E}\left[f_1 \mathbf{1}_A\right] \leq \mathrm{E}\left[f_2 \mathbf{1}_A\right]$ for any $A \in \mathcal{G}$, then $f_1 \leq f_2$.

3. The mean value of a random variable is the same as that of its conditional expectation:
$$\mathrm{E}\left[\mathrm{E}\left[f \middle| \mathcal{G}\right]\right] = \mathrm{E}\left[f\right].$$
This is a consequence of (4.1) with $A = X \in \mathcal{G}$.

4. If $f : X \to \mathbb{R}$ is a $\mathcal{G}$-measurable random variable, then $\mathrm{E}\left[f \middle| \mathcal{G}\right] = f$. Indeed, f is already $\mathcal{G}$-measurable and satisfies (4.1).

5. Two elements $A, B \in \mathcal{F}$ are ***independent*** if $\mu\left(A \cap B\right) = \mu\left(A\right)\mu\left(B\right)$. Two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ are independent if, for any $A \in \mathcal{F}$ and any $B \in \mathcal{G}$, A and B are independent. We say that a random variable $f : X \to \mathbb{R}$ is independent of a σ-algebra $\mathcal{G}$ if the σ-algebra generated by f, $\mathcal{F} = \sigma\left(\left\{f^{-1}\left(A\right) : A \in \mathcal{B}(\mathbb{R})\right\}\right)$, is independent of $\mathcal{G}$. Finally, we say that two random variables are independent it the σ-algebras they generate are independent.

   If $f : X \to \mathbb{R}$ is independent of $\mathcal{G}$, then $\mathrm{E}\left[f \middle| \mathcal{G}\right] = \mathrm{E}\left[f\right]$.

   Exercise 46 Prove this statement using that, if $g, f : X \to \mathbb{R}$ are two independent random variables, $\mathrm{E}\left[fg\right] = \mathrm{E}\left[f\right]\mathrm{E}\left[g\right]$.

6. **Factorization:** If g is a bounded, $\mathcal{G}$-measurable random variable,
$$\mathrm{E}\left[gf \middle| \mathcal{G}\right] = g\,\mathrm{E}\left[f \middle| \mathcal{G}\right].$$

7. If $\mathcal{G}_i$, $i = 1, 2$, are σ-algebras such that with $\mathcal{G}_1 \subseteq \mathcal{G}_2$,
$$\mathrm{E}\left[\mathrm{E}\left[f|\,\mathcal{G}_1\right]|\,\mathcal{G}_2\right] = \mathrm{E}\left[\mathrm{E}\left[f|\,\mathcal{G}_2\right]|\,\mathcal{G}_1\right] = \mathrm{E}\left[f|\,\mathcal{G}_1\right].$$

8. Let f be a random variable independent of $\mathcal{G}$ and let g be a $\mathcal{G}$-measurable random variable. Then, for any measurable function $h : \mathbb{R}^2 \to \mathbb{R}$ such that the random variable
$$\begin{array}{rcl} h\left(f,g\right) : X & \longrightarrow & \mathbb{R} \\ x & \longmapsto & h\left(f\left(x\right), g\left(x\right)\right) \end{array}$$
is in $L^1\left(X, \mu\right)$, we have
$$\mathrm{E}\left[h\left(f,g\right)|\,\mathcal{G}\right] = \mathrm{E}\left[h\left(f,x\right)\right]|_{x=g}.$$
In this expression, $\mathrm{E}\left[h\left(f,x\right)\right]$ denotes the random variable that, to any fixed $x \in X$, associates the expectation $\mathrm{E}\left[h\left(f,x\right)\right]$. $\mathrm{E}\left[h\left(f,x\right)\right]|_{x=g}$ denotes the composition of this random variable with g.

4.3 Birkhoff's Ergodic Theorem

Let $(X, \mathcal{F}, \mu)$ be a probability space and $T : X \to X$ a probability preserving map. The results of this sections are also true if we replace the probability μ with a finite measure, $\mu\left(X\right) < \infty$. We define the σ-algebra of T-invariant sets $\mathcal{G} = \sigma\left(\left\{A \in \mathcal{F} : T^{-1}(A) = A\right\}\right)$.

Exercise 47 Show that any $\mathcal{G}$-measurable random variable $f : X \to \mathbb{R}$ is T-invariant.

Theorem 48 (Birkhoff's Ergodic Theorem) *Let $f : X \to \mathbb{R}$ be an integrable random variable (i.e., $\mathrm{E}\left[|f|\right] < \infty$). With the notation introduced so far,*
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(T^n(x)\right) = \mathrm{E}\left[f|\,\mathcal{G}\right](x) \quad a.s.. \tag{4.3}$$

Remark 49 *Birkhoff's Ergodic Theorem implies that the limit*
$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(T^n(x)\right)$$
exists a.s. and, moreover, defines a T-invariant integrable function function because $\mathrm{E}\left[f|\,\mathcal{G}\right]$ is $\mathcal{G}$-measurable (Exercise 47). In the literature, these important consequences of Birkhoff's Ergodic Theorem are sometimes explicitly stated.

Birkhoff's Ergodic Theorem has an important corollary when T is ergodic. Observe that, if T is ergodic, then any subset in the σ-algebra $\mathcal{G}$ of invariant sets has probability either 0 or 1. Roughly speaking, one might think of $\mathcal{G}$ as the trivial σ-algebra so, by Example 42, $\mathrm{E}[f|\mathcal{G}] = \mathrm{E}[f]$. We give a rigorous proof of this fact in the next corollary:

Corollary 50 *If T is ergodic with respect to μ,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(T^n(x)\right) = \mathrm{E}[f] \quad a.s..$$

Proof. Let $f' = \mathrm{E}[f|\mathcal{G}]$ and define the sets $A_+ := \{x \in X : f'(x) > \mathrm{E}[f]\}$, $A_0 := \{x \in X : f'(x) = \mathrm{E}[f]\}$, and $A_- := \{x \in X : f'(x) < \mathrm{E}[f]\}$. These three sets are T-invariant (f' is T-invariant by Exercise 47) and, therefore, belong to $\mathcal{G}$. They form a partition of X and, consequently, exactly one of them must have measure 1 and the other two probability 0. If $\mu(A_+) = 1$, then $\mathrm{E}[f'] = \int_{A_+} f' d\mu$ and, using the monotonicity of the integral,

$$\mathrm{E}\left[f'\right] = \int_{A_+} f' d\mu > \int_{A_+} \mathrm{E}[f]\, d\mu = \mathrm{E}[f]\, \mu(A_+) = \mathrm{E}[f],$$

which is clearly a contradiction because, by definition, $\mathrm{E}[f'] = \mathrm{E}[f]$. Similarly, $\mu(A_-)$ must also be 0. Consequently, $\mu(A_0) = 1$ and $f' = \mathrm{E}[f]$ a.s..

■

Corollary 50 is often referred to as Birkhoff's Theorem in the literature. Its physical interpretation is the following. An integrable function $f : X \to \mathbb{R}$ is sometimes called an *observable* since it can be thought of as the result of a *measurement* which depends on the point x of the phase space X at which f is evaluated. The integral $\int_X f d\mu$ is sometimes called the ***space average*** of f (with respect to the measure μ) whereas, for a given point $x \in X$, the averages $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x))$ are often referred to as the ***time averages*** of f along the orbit of x. Corollary 50 claims that, when μ is ergodic, *time averages converge to space averages.*

In order to prove Theorem 48, we need an auxiliary result.

Lemma 51 (Maximal Ergodic Theorem) *Define*

$$S_N(x) = \sum_{n=0}^{N-1} f\left(T^n(x)\right) \;\; \textit{and}\;\; M_N(x) := \max\{S_0(x), ..., S_N(x)\}$$

with the convention $S_0 = 0$. Then $\int_{\{M_N>0\}} f d\mu \geq 0$.

Proof. For every $0 \leq k \leq N$ and every $x \in X$, by definition, one has $M_N(T(x)) > S_k(T(x))$ and $f(x) + M_N(T(x)) \geq f(x) + S_k(T(x)) = S_{k+1}(x)$. Therefore,

$$f(x) \geq \max\{S_1(x), ..., S_N(x)\} - M_N(T(x)).$$

Furthermore, $\max\{S_1(x), ..., S_N(x)\} = M_N(x)$ on the set $\{M_N > 0\}$, so that

$$\begin{aligned} \int_{\{M_N>0\}} f d\mu &\geq \int_{\{M_N>0\}} (M_N - M_N \circ T) d\mu \\ &\geq \mathrm{E}[M_N] - \int_{\{M_N>0\}} (M_N \circ T) d\mu \end{aligned} \tag{4.4}$$

where $\int_{\{M_N>0\}} M_N d\mu \geq \mathrm{E}[M_N]$ because $M_N \geq 0$. Now,

$$\begin{aligned} \int_{\{M_N>0\}} (M_N \circ T) d\mu &= \int_X \mathbf{1}_{\{M_N>0\}} (M_N \circ T) d\mu = \int_X \mathbf{1}_{\{T(x) \mid M_N(x)>0\}} M_N d(T_* \mu) \\ &= \int_{\{T(x) \mid M_N(x)>0\}} M_N d\mu \end{aligned}$$

because T is measure-preserving. Since $M_N \geq 0$, $\int_B M_N d\mu \leq \mathrm{E}[M_N]$ for any $B \in \mathcal{F}$, so that (4.4) implies

$$\int_{\{M_N>0\}} f d\mu \geq \mathrm{E}[M_N] - \int_{\{T(x) \mid M_N(x)>0\}} (M_N \circ T) d\mu \geq 0,$$

which is the required result. ■

Proof of Theorem 48. First of all, observe that proving (4.3) is equivalent to proving that

$$0 = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) - \mathrm{E}[f|\mathcal{G}] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} (f - \mathrm{E}[f|\mathcal{G}])(T^n(x))$$

where we have used that $\mathrm{E}[f|\mathcal{G}]$ is T-invariant. Therefore, replacing f by $f - \mathrm{E}[f|\mathcal{G}]$ in the statement of Birkhoff's Ergodic Theorem, we can assume without loss of generality that $\mathrm{E}[f|\mathcal{G}] = 0$. Define $\overline{S} = \limsup_{n\to\infty} S_n/n$ and $\underline{S} = \liminf_{n\to\infty} S_n/n$. We want to show that $\overline{S} = \underline{S} = 0$. It is enough to show that $\overline{S} \leq 0$ a.s. since this implies (by considering $-f$ instead of f) that $\underline{S} \geq 0$. Therefore $0 \leq \underline{S} \leq \overline{S} \leq 0$, which means $\overline{S} = \underline{S} = 0$ a.s..

It is clear that $\overline{S}(T(x)) = \overline{S}(x)$ for every $x \in X$, so that, if $\varepsilon > 0$, one has $A^\varepsilon := \left\{x \in X : \overline{S}(x) > \varepsilon\right\} \in \mathcal{G}$. That is, A^ε belongs to the σ-algebra $\mathcal{G}$ of T-invariant sets. We want to show that $\mu(A^\varepsilon) = 0$. Define

$$f^\varepsilon := (f - \varepsilon)\,\mathbf{1}_{A^\varepsilon},$$

and S_N^ε and M_N^ε according to Lemma 51. With these definitions, we have

$$\frac{S_N^\varepsilon}{N} = \begin{cases} 0 \ \text{ if } \overline{S}(x) \le \varepsilon \\ \frac{S_N}{N} - \varepsilon \ \text{ otherwise.} \end{cases} \tag{4.5}$$

The sequence of sets $\{M_N^\varepsilon > 0\}$ increases to the set $B^\varepsilon := \{\sup_N S_N^\varepsilon > 0\} = \{\sup_N \frac{S_N^\varepsilon}{N} > 0\}$. From (4.5),

$$\sup\nolimits_N \frac{S_N^\varepsilon(x)}{N} > 0 \Longleftrightarrow \exists N \in \mathbb{N} : \frac{S_N}{N} - \varepsilon > 0 \Longleftrightarrow \overline{S}(x) > \varepsilon.$$

Therefore

$$B^\varepsilon = \left\{\sup\nolimits_N \frac{S_N}{N} > \varepsilon\right\} = \left\{\overline{S} > \varepsilon\right\} = A^\varepsilon.$$

Now, on the one hand, $\int_{\{M_N^\varepsilon > 0\}} f^\varepsilon d\mu \ge 0$ for any $N \ge 1$ from Lemma 51; on the other hand, $\mathrm{E}[|f^\varepsilon|] \le \mathrm{E}[|f|] + \varepsilon < \infty$. In this situation, the Dominated Convergence Theorem implies that

$$0 \le \lim_{N\to\infty} \int_{\{M_N^\varepsilon > 0\}} f^\varepsilon d\mu = \int_{A^\varepsilon} f^\varepsilon d\mu$$

and, therefore,

$$\begin{aligned} 0 \le \int_{A^\varepsilon} f^\varepsilon d\mu &= \int_{A^\varepsilon} (f - \varepsilon)\, d\mu = \int_{A^\varepsilon} f d\mu - \varepsilon\mu(A^\varepsilon) \\ &= \int_{A^\varepsilon} \mathrm{E}[f|\,\mathcal{G}]\, d\mu - \varepsilon\mu(A^\varepsilon) = -\varepsilon\mu(A^\varepsilon) \end{aligned}$$

because $A^\varepsilon \in \mathcal{G}$ and we assumed that $\mathrm{E}[f|\,\mathcal{G}] = 0$. In conclusion, one must have $\mu(A^\varepsilon) = 0$ for any $\varepsilon > 0$, which implies that $\overline{S} \le 0$ almost surely. ∎

Corollary 52 *Let $T : X \to X$ be a measurable transformation and μ a T-invariant ergodic probability. Then, for any $A \in \mathcal{F}$*

$$\frac{\#\left\{1 \le j \le N : T^j(x) \in A\right\}}{N} \underset{N\to\infty}{\longrightarrow} \mu(A) \quad a.s..$$

Proof. It is a straightforward consequence of Birkhoff's Ergodic Theorem applied to the characteristic function $f = \mathbf{1}_A$. ■

Example 53 Dirac measures on fixed points and periodic orbits. Let $T : X \to X$ be a measurable transformation and let $P = \{a_1, ..., a_n\}$ be a periodic orbit. Let $\delta_P = \frac{1}{n}\sum_{i=1}^{n} \delta_{a_i}(A)$ be the Dirac measure uniformly distributed on P (see Subsection 3.1.1). We already know that δ_P is T-invariant.

Proposition 54 *δ_P is ergodic.*

Proof. If $P = \{a_1\}$ is a fixed point, the statement is trivial because every measurable set $A \in \mathcal{F}$ has measure 0 or 1 with respect to δ_{a_1}. In particular, this is true for any backward invariant set. If P is a periodic orbit with $n \geq 2$ points, then every measurable set A such that $T^{-1}(A) = A$ must contain either all points of P or none of them. Therefore A has measure either 0 or 1. ■

Now, let $p \neq q$ be two fixed points for T and define the measure

$$\mu = \frac{1}{2}(\delta_p + \delta_q).$$

Proposition 55 *μ is not ergodic.*

Proof. Consider the set $A = \bigcup_{n\in\mathbb{N}} T^{-1}(p)$ of all the preimages of the point p. Then clearly $T^{-1}(A) = A$. Moreover, $q \notin A$ since q is a fixed point and therefore cannot be sent to p under a forward iteration. Therefore $\mu(A) = 1/2$ and μ is not ergodic. ■

4.4 Structure of the set of invariant measures

Let X be a topological space and $T : X \to X$ a measurable transformation. Recall that $\mathcal{M}$ denotes the space of all Borel probabilities on $(X, \mathcal{B}(X))$. The larger space of finite Borel measures is a vector space since for any two measures μ_1 and μ_2 and any two scalars $a, b \in \mathbb{R}$ we have that $a\mu_1 + b\mu_2$ also defines a finite measure. Let $\mathcal{M}_T \subset \mathcal{M}$ be the subset of all T-invariant Borel probability measures on X. A subset of a linear space is ***convex*** if, for any $t \in [0, 1]$ and every $\mu_1, \mu_2 \in \mathcal{M}$ we have $t\mu_1 + (1 - t)\mu_2 \in \mathcal{M}$.

Exercise 56 $\mathcal{M}$ and $\mathcal{M}_T$ are convex.

Moreover, $\mathcal{M}_T \subset \mathcal{M}$ is closed. If, additionally, $\mathcal{M}$ is compact (for instance if X is compact, see Lemma 31) then $\mathcal{M}_T$ is also compact in the weak* topology. We say that $\mu \in \mathcal{M}_T$ is an ***extremal point*** of $\mathcal{M}_T$ if it cannot be written as a linear combination of any two other points of $\mathcal{M}_T$, i.e., if $\mu = t\mu_1 + (1-t)\,\mu_2$ for some $\mu_1, \mu_2 \in \mathcal{M}_T$, then necessarily $t = 0$ or 1. The Krein-Milman Theorem claims that a convex set is the convex hull of its extremal points. In particular, a convex set has always extremal points. As we will see in Proposition 59, ergodic probabilities correspond to extremal points and, therefore, the existence of ergodic measures is always guaranteed provided that $\mathcal{M}_T \neq \emptyset$. For example, if X is compact and $T : X \to X$ is continuous, then $\mathcal{M}_T \neq \emptyset$ by Krylov-Bogolubov's Theorem (Theorem 29). Consequently,

Proposition 57 *If X is compact and $T : X \to X$ is continuous, then there exists one ergodic probability at least.*

We say that two measures μ and ν are ***equivalent***, and we will write $\mu = \nu$, if $\mu(A) = \nu(A)$ for any $A \in \mathcal{F}$. In particular, this means that $\int_X f d\mu = \int_X f d\nu$ for any bounded measurable function $f : X \to \mathbb{R}$. We say that two measures μ and ν are ***mutually singular*** if there exists a measurable set $A \in \mathcal{F}$ such that $\mu(A) = 1$ and $\nu(A) = 0$.

Lemma 58 *If $\mu_1, \mu_2 \in \mathcal{M}_T$ are ergodic measures such that $\mu_1 \ll \mu_2$ then $\mu_1 = \mu_2$.*

Proof. Let $f : X \to \mathbb{R}$ be an arbitrary bounded measurable function (and thus in particular integrable with respect to any invariant probability measure). Since μ_2 is ergodic, by Birkhoff's Ergodic Theorem,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(T^n(x)\right) = \int_X f d\mu_2$$

for any $x \in \Omega$ on a measurable set of full μ_2-measure, i.e., $\mu_2(\Omega) = 1$. Since $\mu_1 \ll \mu_2$ and $\mu_2(\Omega^c) = 0$, we have $\mu_1(\Omega^c) = 0$ and, consequently, $\mu_1(\Omega) = 1$ as well. Therefore,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(T^n(x)\right) = \int_X f d\mu_2 \quad \mu_1\text{-a.s.}.$$

However, applying Birkhoff's Ergodic Theorem to μ_1, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f\left(T^n(x)\right) = \int_X f d\mu_1 \quad \mu_1\text{-a.s.}.$$

In other words,

$$\int_X f d\mu_2 = \int_X f d\mu_1 \tag{4.6}$$

for any measurable bounded function f. Writing (4.6) for a characteristic function $\mathbf{1}_A$, we obtain $\mu_1(A) = \mu_2(A)$ for any measurable set $A \in \mathcal{F}$ and $\mu_1 = \mu_2$. ■

Proposition 59 *$\mu \in \mathcal{M}_T$ is ergodic if and only if it is an extremal point of $\mathcal{M}_T$.*

Proof. $\mu \in \mathcal{M}_T$ extremal point $\Longrightarrow \mu \in \mathcal{M}_T$ ergodic. Suppose that μ is not ergodic. Then there exists a T-invariant measurable subset $A \subset X$ with $0 < \mu(A) < 1$. Define the measures $\mu_A(B) = \mu(B \cap A)/\mu(A)$ and $\mu_{X\backslash A}(B) = \mu(B \cap (X\backslash A))/\mu(X\backslash A)$ where $B \in \mathcal{F}$. Then μ_A and $\mu_{X\backslash A}$ are T-invariant and

$$\mu = \mu(A)\,\mu_A + \mu(X\backslash A)\,\mu_{X\backslash A},$$

so μ is not an extreme point.

$\mu \in \mathcal{M}_T$ ergodic $\Longrightarrow \mu \in \mathcal{M}_T$ extremal point. Suppose by contradiction that μ is not extremal so that $\mu = t\mu_1 + (1-t)\mu_2$ for two invariant different probability measures $\mu_1, \mu_2 \in \mathcal{M}_T$ and some $t \in (0,1)$. Since $\mu(A) = 0$ always implies $\mu_1(A) = 0$ and $\mu_2(A) = 0$ both μ_1 and μ_2 are absolutely continuous with respect to μ and, moreover, they are ergodic. Indeed, if $\mu(A) = 1$, then necessarily $\mu_1(A) = \mu_2(A) = 1$. Therefore, by Lemma 58, we have $\mu_1 = \mu = \mu_2$ contradicting thus our assumption. ■

Ergodic measures are not only extremal points but also mutually singular each other.

Proposition 60 *Let μ_1 and μ_2 be distinct ergodic invariant measures. Then μ_1 and μ_2 are mutually singular.*

Proof. By Lemma 58, μ_1 and μ_2 cannot be absolutely continuous. Therefore, there exists a measurable set E such that $\mu_1(E) > 0$ and $\mu_2(E) = 0$. Define

$$A = \bigcap_{m=0}^{\infty} \bigcup_{j=m}^{\infty} T^{-j}(E).$$

We will show that $\mu_1(A) = 1$ and $\mu_2(A) = 0$ which will imply that μ_1 and μ_2 are mutually singular.

First, we claim that $T^{-1}(A) = A$. Indeed, if $x \in A$ then $x \in \bigcup_{j=m}^{\infty} T^{-j}(E)$ for any $m \in \mathbb{N}$. That is, $T^j(x) \in E$ for infinitely many values of $j \in \mathbb{N}$.

If x satisfies this property, then so do $T^{-1}(x)$ and $T(x)$, which implies $T^{-1}(A) = A$. Therefore, it is sufficient to show that $\mu_1(A) > 0$ to imply $\mu_1(A) = 0$ by the ergodicity of μ_1.

By the invariance of both μ_1 and μ_2 we have

$$\mu_1 \left(\bigcup_{j=0}^{\infty} T^{-j}(E) \right) \geq \mu_1(E) > 0 \tag{4.7}$$

and

$$\mu_2 \left(\bigcup_{j=0}^{\infty} T^{-j}(E) \right) = 0. \tag{4.8}$$

Observe that (4.8) implies that $\mu_2(A) = 0$. On the other hand,

$$\bigcup_{j=m}^{\infty} T^{-j}(E) = T^{-m} \left(\bigcup_{j=0}^{\infty} T^{-j}(E) \right)$$

and, consequently,

$$\mu_i \left(\bigcup_{j=m}^{\infty} T^{-j}(E) \right) = \mu_i \left(T^{-m} \left(\bigcup_{j=0}^{\infty} T^{-j}(E) \right) \right) = \mu_i \left(\bigcup_{j=0}^{\infty} T^{-j}(E) \right)$$

for $i = 1, 2$. In particular, the measure of each $\bigcup_{j=m}^{\infty} T^{-j}(E)$ is constant. Since the sets $\bigcup_{j=m}^{\infty} T^{-j}(E)$ are nested, i.e.,

$$\bigcup_{j=m+1}^{\infty} T^{-j}(E) \subseteq \bigcup_{j=m}^{\infty} T^{-j}(E),$$

A is the countable intersection of a nested sequence of sets all of them with the same measure (strictly positive by (4.7)). It follows that $\mu_1(A) > 0$ and $\mu_2(A) = 0$ as required.

Exercise 61 Prove this last sentence.

■

Chapter 5

Circle rotations

In this chapter, we are going to deal with a very important example, that of circle rotations

$$\begin{aligned} T:\mathbb{S}^1 &\longrightarrow \mathbb{S}^1 \\ x &\longmapsto x+\alpha, \quad \alpha\in\mathbb{R}. \end{aligned} \tag{5.1}$$

More concretely, we are going to prove that the *Lebesgue measure* λ on $\left(\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1)\right)$ is ergodic if and only if α is a irrational multiple of 2π, i.e., $\alpha \neq 2\pi\frac{m}{n}$, $m, n \in \mathbb{Z}$. Indeed if $\alpha = 2\pi\frac{m}{n}$ with $m, n \in \mathbb{Z}$ such that m and n have no common factors, then $T^n = \mathrm{Id}$ is the identity and, for any $x \in \mathbb{S}^1$, $O_x = \left\{x, T(x), ..., T^{n-1}(x)\right\}$ is a periodic orbit of period n. Then any set built as a family of arcs $\left\{B_\varepsilon\left(T^i(x)\right)\right\}_{i=0...n-1}$ of length $\varepsilon > 0$ centered at the points of an orbit O_x is invariant and of strictly positive Lebesgue measure. Therefore, the Lebesgue measure is not ergodic. Furthermore, according to Example 53, the Dirac measure supported on O_x

$$\delta_{O_x} = \frac{1}{n}\sum_{i=0}^{n}\delta_{T^i(x)}$$

is ergodic. Therefore, a rational rotation admits infinitely many ergodic measures.

5.1 Irrational case

We will prove that the Lebesgue measure is ergodic when α in (5.1) is an irrational multiple of 2π in two different ways. The first proof is rather easy and uses Fourier analysis. The second one is longer and requires some non-trivial results such as the Lebesgue Density Theorem. Nevertheless, this theorem is quite important in order to prove the ergodicity of some concrete maps, hence the reason why we choose to prove the ergodicity of an irrational rotation by this slightly more sophisticated way. To start with, we will give the easier proof.

Proposition 62 *The Lebesgue measure is ergodic with respect to the rotation (5.1) if and only if α is an irrational multiple of 2π.*

Proof. By Proposition 38, it is enough to prove that any T-invariant $f \in L^2\left(\mathbb{S}^1, \lambda\right)$ is constant a.e.. Identify $\mathbb{S}^1$ with the unit interval $[0,1]/_{0\sim 1}$ and think of α as an irrational number between 0 and 1. The Fourier series $\sum_{n=-\infty}^{n=\infty} a_n \, \mathrm{e}^{2n\pi i x}$ of f converges to f in the L^2 norm. The series $\sum_{n=-\infty}^{n=\infty} a_n \, \mathrm{e}^{2n\pi i(x+\alpha)}$ converges to $f \circ T$. Since $f = f \circ T$ a.e., uniqueness of Fourier coefficients implies that $a_n = a_n \, \mathrm{e}^{2n\pi i\alpha}$ for all $n \in \mathbb{Z}$. Since $\mathrm{e}^{2n\pi i\alpha} \neq 1$ for $n \neq 0$, we conclude that $a_n = 0$ for $n \neq 0$, so f is constant a.e..

The converse is immediate because if $\alpha = \frac{n}{m}$ is rational, $n, m \in \mathbb{Z}$, then we already showed that the Lebesgue measure is not ergodic. ■

For the alternative proof, we need first to define new concepts and give additional results.

Lemma 63 ([9, Lemma 7.3]) *Let $W \subset \mathbb{R}^n$ be a measurable set that is contained in a finite union of balls $B_{r_i}(x_i)$ where $x_i \in \mathbb{R}^n$ and $r_i > 0$, $i = 1, ..., N$. Then there is a set $S \subset \{1, ..., N\}$ so that*

(a) the balls $B_{r_i}(x_i)$ with $i \in S$ are disjoint,

(b) $W \subset \bigcup_{i \in S} B_{3r_i}(x_i)$, and

(c) $\lambda(W) \leq 3^n \sum_{i \in S} \lambda\left(B_{r_i}(x_i)\right)$.

Proof. Order the balls $B_i = B_{r_i}(x_i)$ so that $r_1 \geq r_2 \geq ... \geq r_N$. Put $i_1 = 1$. Discard all B_j that intersect B_{i_i}. Let B_{i_2} be the first of the remaining B_j if there are any. Discard all B_j with $j > i_2$ that intersect B_{i_2}, let B_{i_3} be the first of the remaining ones, and so on as long as possible. This process stops after a finite number of steps and gives $S = \{i_1, i_2, ...\}$. It is clear that (a) holds. Every discarded B_j is a subset of $B_{3r_i}(x_i)$ for some $i \in S$, for if $r' < r$ and $B_{r'}(x')$ intersects $B_r(x)$, then $B_{r'}(x') \subseteq B_r(x)$. This proves (b). (c) follows from (b) because, in $\mathbb{R}^n$,

$$\lambda\left(B_{3r}(x)\right) = 3^n \lambda\left(B_r(x)\right).$$

■

Corollary 64 *Let* $T : \mathbb{S}^1 \to \mathbb{S}^1$ *be an irrational circle rotation. Then, for every* $x \in \mathbb{S}^1$*, there exists a sequence of arc neighbourhoods* J_n *of* x *with* $\lambda(J_n) \to 0$ *as* $n \to \infty$ *and a sequence of finite sets* $S_n \subset \mathbb{N}$ *such that*

1. $\mathbb{S}^1 \subseteq \bigcup_{i \in S_n} T^i(J_n)$;
2. $\sum_{i \in S_n} \lambda\left(T^i(J_n)\right) \leq 3\left(1 + \frac{2}{n}\right)$.

Observe that while the first statement is relatively intuitive, the second is highly non-trivial beforehand. Nevertheless, this is a consequence of Lemma 63. The number three is not as important as the fact that there exists a bound on how much the intervals $T^i(J_n)$ can overlap, so that we can give a bound uniform in n. This will be crucially used at the end of the proof of Proposition 68.

Proof of Corollary 64. Identify $\mathbb{S}^1$ with $[0,1]/\, 0 \sim 1$ and define the projection

$$\begin{array}{rcl} \pi : \mathbb{R} & \longrightarrow & [0,1]/\, 0 \sim 1 \\ z & \longmapsto & [z] \end{array}$$

that send any real number to its equivalent class in $[0,1]/\, 0 \sim 1$. Let $z \in [0,1]$ and $n \in \mathbb{N}$. Take $J_n := \pi\left(B_{3/n}(z)\right)$ as the image by π of the open ball of radius $\frac{3}{n}$ centered at z. $\{J_n\}_{n \in \mathbb{N}}$ defines a sequence of arc neighbourhoods of $x = [z]$ such that $\lambda(J_n) = \frac{6}{n} \to 0$ as $n \to 0$.

Exercise 65 *Using that the orbit* $\{T^i(x)\}_{i \in \mathbb{N}}$ *of* x *is dense in* $\mathbb{S}^1$ *for an irrational rotation (Theorem 27), prove that there exists a finite subset* $I = \{i_1, ..., i_N\} \subset \mathbb{N}$ *such that* $[0,1] \subseteq \bigcup_{i \in I} B_{1/n}(T^i(x))$.

Now, we have

$$[0,1] \subseteq \bigcup_{i \in I} B_{1/n}\left(T^i(x)\right).$$

By Corollary 64, there exists a finite set $S \subset I$ such that

$$[0,1] \subseteq \bigcup_{i \in S} B_{3/n}\left(T^i(x)\right)$$

where the balls $B_{1/n}\left(T^i(x)\right)$ are disjoint. Moreover, $\pi\left(B_{3/n}\left(T^i(x)\right)\right) = T^i(J_n)$ and

$$\mathbb{S}^1 \subseteq \bigcup_{i \in S} \pi\left(B_{3/n}\left(T^i(x)\right)\right) = \bigcup_{i \in S} T^i(J_n).$$

On the other hand,

$$\sum_{i\in S}\lambda\left(T^{i}\left(J_{n}\right)\right)=\sum_{i\in S}\lambda\left(\pi\left(B_{3/n}\left(T^{i}\left(x\right)\right)\right)\right)$$
$$\leq\sum_{i\in S}\lambda\left(B_{3/n}\left(T^{i}\left(x\right)\right)\right)\leq 3\sum_{i\in S}\lambda\left(B_{1/n}\left(T^{i}\left(x\right)\right)\right). \quad (5.2)$$

Observe now that the union $\bigcup_{i\in S}B_{1/n}\left(T^{i}\left(x\right)\right)$ is contained in the open interval $(-\frac{1}{n},1+\frac{1}{n})$. Since the balls $B_{1/n}\left(T^{i}\left(x\right)\right)$, $i\in S$, are disjoint,

$$\sum_{i\in S}\lambda\left(B_{1/n}\left(T^{i}\left(x\right)\right)\right)=\lambda\left(\bigcup_{i\in S}B_{1/n}\left(T^{i}\left(x\right)\right)\right)\leq 1+\frac{2}{n}.$$

Therefore, (5.2) implies

$$\sum_{i\in S}\lambda\left(T^{i}\left(J_{n}\right)\right)\leq 3\left(1+\frac{2}{n}\right).$$

■

If $f\in L^{1}\left(\mathbb{R}^{n},\lambda\right)$, $n\in\mathbb{N}$, we say that $x\in\mathbb{R}^{n}$ is a ***Lebesgue point*** of f if

$$f\left(x\right)=\lim_{r\to 0}\frac{1}{B_{r}\left(x\right)}\int_{B_{r}(x)}fd\lambda.$$

It is probably far from obvious that every $f\in L^{1}\left(\mathbb{R}^{n},\lambda\right)$ has Lebesgue points. But the following remarkable theorem, which we are not going to prove, shows that they always exist. The reader is encouraged to check with [9].

Theorem 66 ([9, Theorem 7.7]) *If $f\in L^{1}\left(\mathbb{R}^{n},\lambda\right)$, then almost every $x\in\mathbb{R}^{n}$ is a Lebesgue point of f.*

One of the most important corollaries of Theorem 66 is what in the literature is sometimes referred to as *Lebesgue's Density Theorem*. This result gives us information about the *density of (Lebesgue) measure* on almost every point of a measurable set.

Corollary 67 (Lebesgue's Density Theorem) *Let $A\in\mathcal{B}\left(\mathbb{R}^{n}\right)$ be a Lebesgue measurable subset of $\mathbb{R}^{n}$ with positive measure, $\lambda\left(A\right)>0$. Then for λ-almost every point $x\in A$,*

$$\lim_{r\to 0}\frac{\lambda\left(A\cap B_{r}(x)\right)}{\lambda(B_{r}(x))}=1. \quad (5.3)$$

Proof. This result is a consequence of Theorem 66 applied to the characteristic function $f = \mathbf{1}_A$. ■

This result says that in some very subtle way. A priori, one may expect that if $\lambda(A) = 1/2$, then for any subinterval J the ratio between $A \cap J$ and J might be $1/2$, i.e., that the ratio between the measure of the whole interval and the measure of the set A is constant at every scale. This theorem shows that this is not the case. Points $x \in A$ for which (5.3) holds are called ***Lebesgue's density points***.

We are now ready to tackle the prove of ergodicity of Lebesgue measure for irrational circle rotations. Similar arguments of those used in the proof of Proposition 68 will be used later.

Proposition 68 *If $\alpha/2\pi$ is irrational then Lebesgue measure is ergodic.*

Proof. Let $A \in \mathcal{B}\left(\mathbb{S}^1\right)$ satisfy $T^{-1}(A) = A$ and $\lambda(A) > 0$. We want to show that $\lambda(A) = 1$. By Lebesgue's Density Theorem, λ-almost every point of A is a Lebesgue density point of A. Let $x \in A$ be one of such points and fix an arbitrary $\varepsilon > 0$. Choose $n_\varepsilon \in \mathbb{N}$ large enough so that

$$\lambda\left(A \cap J_{n_\varepsilon}\right) \geq (1-\varepsilon)\,\lambda\left(J_{n_\varepsilon}\right) \tag{5.4}$$

where J_{n_ε} is a sufficiently small arc neighbourhood of x as in Corollary 64. We shall make three simple statements which combined will give us the desired result. First of all, observe that (5.4) is equivalent to

$$\frac{\lambda\left(J_{n_\varepsilon} \backslash A\right)}{\lambda\left(J_{n_\varepsilon}\right)} \leq \varepsilon. \tag{5.5}$$

Secondly, since T is just a translation and Lebesgue measure is invariant by translations, we have $\lambda\left(T^i\left(J_{n_\varepsilon}\right)\right) = \lambda\left(J_{n_\varepsilon}\right)$ and $\lambda\left(T^i\left(J_{n_\varepsilon} \backslash A\right)\right) = \lambda\left(J_{n_\varepsilon} \backslash A\right)$ for any $i \in \mathbb{N}$ (these equalities stem from the fact that λ is invariant by T^{-1}, which is again a rotation of angle $-\alpha$). In particular,

$$\frac{\lambda\left(T^i\left(J_{n_\varepsilon} \backslash A\right)\right)}{\lambda\left(T^i\left(J_{n_\varepsilon}\right)\right)} = \frac{\lambda\left(J_{n_\varepsilon} \backslash A\right)}{\lambda\left(J_{n_\varepsilon}\right)}. \tag{5.6}$$

In third place, using the invariance of A ($T^{-1}(A) = A$ which, in turn, implies $T(A) = A$) and the fact that $\mathbb{S}^1 \subseteq \bigcup_{i \in S_{n_\varepsilon}} T^i\left(J_{n_\varepsilon}\right)$ we have

$$\mathbb{S}^1 \backslash A \subseteq \left(\bigcup_{i \in S_{n_\varepsilon}} T^i\left(J_{n_\varepsilon}\right)\right) \Big\backslash A = \bigcup_{i \in S_{n_\varepsilon}} \left(T^i\left(J_{n_\varepsilon}\right) \backslash A\right) = \bigcup_{i \in S_{n_\varepsilon}} T^i\left(J_{n_\varepsilon} \backslash A\right)$$

so

$$\lambda\left(\mathbb{S}^1 \setminus A\right) \le \sum_{i\in S_{n_\varepsilon}} \lambda\left(T^i\left(J_{n_\varepsilon}\setminus A\right)\right), \tag{5.7}$$

Now, from (5.5), (5.6), and (5.7)

$$\begin{aligned}\lambda\left(\mathbb{S}^1 \setminus A\right) &\le \sum_{i\in S_{n_\varepsilon}} \lambda\left(T^i\left(J_{n_\varepsilon}\setminus A\right)\right) \le \sum_{i\in S_{n_\varepsilon}} \frac{\lambda\left(J_{n_\varepsilon}\setminus A\right)}{\lambda\left(J_{n_\varepsilon}\right)}\lambda\left(T^i\left(J_{n_\varepsilon}\right)\right) \\ &= \frac{\lambda\left(J_{n_\varepsilon}\setminus A\right)}{\lambda\left(J_{n_\varepsilon}\right)} \sum_{i\in S_{n_\varepsilon}} \lambda\left(T^i\left(J_{n_\varepsilon}\right)\right) \le 3\varepsilon\left(1+\frac{2}{n_\varepsilon}\right).\end{aligned}$$

by Corollary 64. Since ε is arbitrary, this means that $\lambda\left(\mathbb{S}^1 \setminus A\right) = 0$ so $\lambda(A) = 1$. ■

Chapter 6

Central Limit Theorem

In this chapter we will state a Central Limit Theorem for the random variables $f \circ T^n$ built from an observable $f \in L^1(X, \mu)$ and an ergodic map $T : X \to X$. This Central Limit Theorem is a firs step to give confidence intervals for an estimation of $\mathrm{E}[f]$ by means of Birkhoff's Ergodic Theorem,

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n(x).$$

The Central Limit Theorem will only hold for mixing maps.

6.1 Mixing maps

Definition 69 *Let $T : X \to X$ be a measurable transformation on a measure space $(X, \mathcal{F}, \mu)$ that preserves μ. We say that T is mixing if, for any two sets $A, B \in \mathcal{F}$ such that $\mu(A) > 0$ and $\mu(B) > 0$, we have*

$$\lim_{n\to\infty} \mu\left(T^{-n}(B) \cap A\right) = \mu(B)\,\mu(A).$$

There are two natural interpretations of mixing, one geometrical and one probabilistic. From a geometrical point of view (recall that $\mu\left(T^{-1}(B)\right) = \mu(B)$) one can think of $T^{-n}(B)$ as a *redistribution of mass*. The mixing condition then says that for large n the proportion of $T^{-n}(B)$ which intersects A is just proportional to the measure of A. In other words $T^{-n}(B)$ is spreading itself uniformly with respect to the measure. A more probabilistic point of view is to think of $\mu(T^{-n}(B) \cap A)/\mu(B)$ as the conditional probability of having $x \in A$ given that $T^n(x) \in B$, i.e. the probability that the occurrence of the event B today is a consequence of the occurrence of the event A n steps in the past. The mixing condition then says that this probability converges to the probability of A, i.e., asymptotically, there is no causal relation between the two events.

A classical example by Arnold and Avez ([1]) explains what a mixing map does. Suppose a cocktail shaker X, $\mu(X) = 1$ is filled by 85% of lemon juice and 15% of vodka. Let A be the part of the cocktail shaker originally occupied by the vodka and B any part of the shaker. Let $T^{-1} : X \to X$ be the transformation of the content of the shaker made during one move by the waiter (who is shaking the cocktail repeatedly and redistributes the volume of the two liquids). Then after n moves the fraction of juice in the part B is $\mu(T^{-n}(A) \cap B) / \mu(B)$. As the waiter keeps shaking the cocktail ($n \to \infty$), the fraction of vodka in any part B approaches $\mu(A) = 0.15$, i.e. the vodka spreads uniformly in the mixture.

Proposition 70 *Let $(X, \mathcal{F}, \mu)$ be a probability space. Any mixing map $T : X \to X$ is ergodic.*

Proof. Let $A \in \mathcal{F}$ be any T-invariant measurable set. Then $T^{-n}(A) = A$ and

$$\mu(A \cap B) = \lim_{n\to\infty} \mu\left(T^{-n}(A) \cap B\right) = \mu(A)\,\mu(B).$$

In particular, for $A = B$ we have $\mu(A) = \mu(A)^2$. This means $\mu(A) = 0$ or $\mu(A) = 1$, hence T is ergodic. ■

Proposition 71 *Suppose that $T : X \to X$ is mixing. Then, for any $f, g \in L^2(X, \mu)$,*

$$\lim_{n\to\infty} \int g\,(f \circ T^n)\, d\mu = \int g d\mu \int f d\mu. \tag{6.1}$$

Proof. Equation (6.1) trivially holds for characteristic functions. Indeed, if $g = \mathbf{1}_A$ and $f = \mathbf{1}_B$ for some sets $A, B \in \mathcal{F}$, then

$$\begin{aligned}\int g\,(f \circ T^n)\, d\mu &= \int \mathbf{1}_A\,(\mathbf{1}_B \circ T^n)\, d\mu = \int \mathbf{1}_A \mathbf{1}_{T^{-n}(B)} d\mu = \int \mathbf{1}_{A\cap T^{-n}(B)} d\mu \\ &= \mu\left(A \cap T^{-n}(B)\right) \underset{n\to\infty}{\longrightarrow} \mu(A)\,\mu(B) = \int \mathbf{1}_A d\mu \int \mathbf{1}_B d\mu.\end{aligned}$$

(6.1) is obviously true for elementary functions as well. The general result follows approximating two arbitrary functions f and g by sequences of elementary functions. ■

6.2 Central Limit Theorem

In probability, the ***Strong Law of Large Numbers*** claims that if $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of independent random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ that are identically distributed, then

$$\frac{X_1 + \ldots + X_n}{n} \underset{n\to\infty}{\longrightarrow} m \text{ a.s.},$$

where $m = \mathrm{E}[X_1]$ is the common expectation. We usually say that $\{X_n\}_{n\in\mathbb{N}}$ is and ***i.i.d. sequence***. Observe that, whenever we have an ergodic map $T : X \to X$ on a probability space $(X, \mathcal{F}, \mu)$, for any $f \in L^1(X, \mu)$, the sequence of random variables $X_n := f \circ T^n$ satisfies the Strong Law of Large Numbers by Birkhoff's Ergodic Theorem,

$$\frac{1}{n}\sum_{i=0}^{n-1} f\left(T^i(x)\right) \underset{n\to\infty}{\longrightarrow} \mathrm{E}[f] \text{ a.s.}. \tag{6.2}$$

Unlike the standard result in probability, now $X_n = f \circ T^n$ are not independent in general (although they might be in some particular situations).

Let $S_n := \sum_{i=0}^{n-1} f\left(T^i(x)\right)$. Even if we know that $T : X \to X$ is an ergodic map, we would like to know how fast the convergence of the averages S_n/n is to the expected value $\mathrm{E}[f]$. Unfortunately, this convergence is in general very slow except for some concrete functions and for dynamical systems that, in a broad sense, exhibit some *strong mixing properties* (in addition to being ergodic). Since we do not know how many iterates are required in 6.2 to obtain a good approximation of $\mathrm{E}[f]$ and since the rate of convergence to that value may vary from point to point, having a *Central Limit Theorem* for the sequence $X_n = f \circ T^n$ is crucial to estimate confidence intervals for the expectation $\mathrm{E}[f]$.

Definition 72 *Given $f \in L^1(X, \mu)$, we say that the random variables $X_n = f \circ T^n$ satisfy the Central Limit Theorem if*

$$\lim_{n\to\infty} \mu\left(\left\{\frac{S_n - n\,\mathrm{E}[f]}{\sqrt{n}} \le z\right\}\right) = \frac{1}{\sqrt{2\pi}\sigma_f}\int_{-\infty}^{z} \mathrm{e}^{-\frac{s^2}{2\sigma_f^2}}\,ds$$

for some finite $\sigma_f^2 \ge 0$. That is, $\left(S_n - n\,\mathrm{E}[f]\right)/\sqrt{n}$ converges in law to $\mathcal{N}(0, \sigma_f^2)$.

For example, if $f \in L^1(X,\mu)$ is such that $f \circ T^n$ satisfy the Central Limit Theorem, that is,

$$\frac{S_n - n\,\mathrm{E}[f]}{\sigma_f \sqrt{n}} = \frac{S_n/n - \mathrm{E}[f]}{\sigma_f/\sqrt{n}} \longrightarrow \mathcal{N}(0,1) \text{ as } n \to \infty \tag{6.3}$$

and we suppose we know σ_f then, for n *large enough*,

$$\mathrm{E}[f] \in \left(\frac{S_n}{n} - 1.96\frac{\sigma_f}{\sqrt{n}}, \frac{S_n}{n} + 1.96\frac{\sigma_f}{\sqrt{n}} \right)$$

with an *approximately* 95% confidence level. In general σ_f has to be estimated as well, which means that, strictly speaking, the Gaussian distribution in (6.3) must be replaced with a different law in order to obtain confidence intervals.

In probability, the Central Limit Theorem is proved for i.i.d sequences of square integrable random variables. Observe that, again, we are in completely different context because the random variables $X_n = f \circ T^n$ need not be independent (actually they will not be in general).

From the definition of S_n, it can be argued that, provided that such a σ_f^2 exists, it must be

$$\sigma_f^2 = C_f(0) + 2\sum_{n=1}^{\infty} C_f(n), \tag{6.4}$$

where

$$C_f(n) := \mathrm{Cov}(f, f \circ T^n) = \mathrm{E}[f(f \circ T^n)] - \mathrm{E}[f]^2, \ n \in \mathbb{N},$$

is the ***autocorrelation function***. More generally, given $f, g \in L^2(X,\mu)$, we introduce the ***correlation function***

$$C_{g,f}(n) := \mathrm{Cov}(g, f \circ T^n) = \mathrm{E}[g(f \circ T^n)] - \mathrm{E}[g]\,\mathrm{E}[f].$$

Exercise 73 Without loss of generality, we can assume that $\mathrm{E}[f] = 0$. Verify the following formula

$$\mathrm{Var}[S_n] = \mathrm{E}\left[S_n^2\right] = nC_f(0) + 2\sum_{i=1}^{n-1}(n-i)\,C_f(n).$$

Observe that in order that σ_f^2 in (6.4) be finite, we must have $C_f(n) \to 0$ as $n \to \infty$. By Proposition 71, this is guaranteed if T is mixing. However, to prove the Central Limit Theorem we need a fast convergence to 0. One can prove that

$$\lim_{n\to\infty} \left(\mathrm{Var}\left[S_n\right] - n\sigma_f^2\right) = -2\sum_{n=1}^{\infty} nC_f(n)$$

which implies that

$$\lim_{n\to\infty} \mathrm{Var}\left[\frac{S_n}{\sqrt{n}}\right] = \sigma_f^2$$

provided that

$$\sum_{n=1}^{\infty} n\left|C_f(n)\right| < \infty. \tag{6.5}$$

For example, (6.5) holds if there exist constants $K \geq 0$ and $\alpha > 2$ such that $|C_f(n)| \leq Kn^{-\alpha}$ (polynomial decay of correlations) or some constant $\beta > 0$ such $|C_f(n)| \leq \mathrm{e}^{-\beta n}$ (exponential decay). One particular class of functions exhibiting exponential decay are *Hölder continuous functions*:

Definition 74 *A function $f : X \to \mathbb{R}$ defined on a metric space X is called* ***Hölder continuous*** *if there exist constants $\alpha_f \in (0,1]$ and $K_f > 0$ such that*

$$|f(x) - f(y)| \leq K_f \,\mathrm{dist}(x,y)^{\alpha_f} \quad \forall x, y \in X.$$

Theorem 75 (exponential decay of correlations) *Let $T : X \to X$ be a mixing map on a metric space. For every pair of Hölder continuous functions f and g, there exist constants $B_{f,g} > 0$ and $\theta_{f,g} < 1$ such that*

$$|C_{f,g}(n)| \leq B_{f,g}\theta_{f,g}^n, \quad n \geq 1.$$

Therefore,

Theorem 76 *Let $f : X \to \mathbb{R}$ be a Hölder continuous functions. Then $X_n = f \circ T^n$ satisfy the Central Limit Theorem with σ_f^2 as in (6.4).*

Bibliography

[1] V. I. Arnold and A. Avez. *Problèmes ergodiques de la Mécanique Classique.* Gauthier-Villars, Paris (1967).

[2] R. G. Bartle. *The Elements of Integration and Lebesgue Measure.* Wiley Classics Library. John Wiley and Sons, Inc. 1995.

[3] V. I. Bogachev. *Measure Theory* (Volume I). Springer, 2007.

[4] M. Brin and G. Stuck. *Introduction to Dynamical Systems.* Cambridge university press, 2002.

[5] N. Chernov and R. Markarian. *Introduction to the Ergodic Theory of Chaotic Billiards.* Monografías del Instituto de Matemática y Ciencias Affines, **19**. Instituto de Matemática y Ciencias Affies; Pontificia Universidad Católica del Perú, Lima, 2001. Available at `premat.fing.edu.uy/papers/2003/75.ps` (Secon Edition) and `http://www.math.uab.edu/chernov/papers/rbook.pdf`.

[6] N. Dunford and J.T. Schwartz. *Linear operators, I. General Theory.* Interscience, New York, 1958.

[7] M. Hairer. *Ergodic Properties of Markov Processes.* Lecture notes, 2006. Available at `www.hairer.org/notes/Markov.pdf`.

[8] S. Luzzatto. Itroduction to Ergodic Theory. Lecture notes, 2009. Available at `http://web.mac.com/stefanoluzzatto/Teaching/M4A36.html`.

[9] W. Rudin. *Real and Complex Analysis.* Third edition. McGraw-Hill Book Co., 1987.

[10] I. G. Sinai. *Introduction to Ergodic Theory.* Princeton University Press, 1976.