

MATH41112/61112

Ergodic Theory

Charles Walkden

4th January, 2018

Contents

0	Preliminaries	2
1	An introduction to ergodic theory. Uniform distribution of real sequences	4
2	More on uniform distribution mod 1. Measure spaces.	13
3	Lebesgue integration. Invariant measures	23
4	More examples of invariant measures	38
5	Ergodic measures: definition, criteria, and basic examples	43
6	Ergodic measures: Using the Hahn-Kolmogorov Extension Theorem to prove ergodicity	53
7	Continuous transformations on compact metric spaces	62
8	Ergodic measures for continuous transformations	72
9	Recurrence	83
10	Birkhoff's Ergodic Theorem	89
11	Applications of Birkhoff's Ergodic Theorem	99
12	Solutions to the Exercises	108

0. Preliminaries

§0.1 Contact details

The lecturer is Dr Charles Walkden, Room 2.241, Tel: 0161 275 5805,
Email: charles.walkden@manchester.ac.uk.

My office hour is: Monday 2pm-3pm. If you want to see me at another time then please email me first to arrange a mutually convenient time.

§0.2 Course structure

This is a reading course, supported by one lecture per week. I have split the notes into weekly sections. You are expected to have read through the material before the lecture, and then go over it again afterwards in your own time. In the lectures I will highlight the most important parts, explain the statements of the theorems and what they mean in practice, and point out common misunderstandings. As a general rule, I will *not* normally go through the proofs in great detail (but they are examinable unless indicated otherwise). You will be expected to work through the proofs yourself in your own time. All the material in the notes is examinable, unless it says otherwise. (Note that if a proof is marked ‘not examinable’ then it means that I won’t expect you to reproduce the proof, but you will be expected to know and understand the statement of the result. If an entire section is marked ‘not examinable’ (for example, the review of Riemann integration in §1.3, and the discussions on the proofs of von Neumann’s Ergodic Theorem and Birkhoff’s Ergodic Theorem in §§9.6, 10.5, respectively) then you don’t need to know the statements of any subsidiary lemmas/propositions in those sections that are not used elsewhere but reading this material may help your understanding.)

Each section of the notes contains exercises. The exercises are a key part of the course and you are expected to attempt them. The solutions to the exercises are contained in the notes; I would strongly recommend attempting the exercises first without referring to the solutions.

Please point out any mistakes (typographical or mathematical) in the notes.

§0.3 The exam

The exam is a 3 hour written exam. There are several past exam papers on the course website. Note that some topics (for example, entropy) were covered in previous years and are not covered this year; there are also some new topics (for example, Kac’s Lemma) that were not covered in 2010 or earlier.

The format of the exam is the same as last year’s. The exam has 5 questions, of which you must do 4. If you attempt all 5 questions then you will get credit for your best 4 answers. The style of the questions is similar to last year’s exam as well as to ‘Section B’ questions from earlier years.

Each question is worth 30 marks. Thus the total number of marks available on the exam is $4 \times 30 = 120$. This will then be converted to a mark out of 100 (by multiplying by $100/120$).

There is no coursework, in-class test or mid-term for this course.

§0.4 Recommended texts

There are several suitable introductory texts on ergodic theory, including

W. Parry, Topics in Ergodic Theory

P. Walters, An Introduction to Ergodic Theory

I.P. Cornfeld, S.V. Fomin and Ya.G. Sinai, Ergodic Theory

K. Petersen, Ergodic Theory

M. Einsiedler and T. Ward, Ergodic Theory: With a View Towards Number Theory.

Parry's or Walter's books are the most suitable for this course.

1. An introduction to ergodic theory. Uniform distribution of real sequences

§1.1 Introduction

A dynamical system consists of a space X , often called a *phase space*, and a rule that determines how points in X evolve in time. Time can be either continuous (in which case a dynamical system is given by a first order autonomous differential equation) or discrete (in which case we are studying the iterates of a single map $T : X \rightarrow X$).

We will only consider the case of discrete time in this course. Thus we will be studying the iterates of a single map $T : X \rightarrow X$. We will write $T^n = T \circ \dots \circ T$ (n times) to denote the n th composition of T . If $x \in X$ then we can think of $T^n(x)$, the result of applying the map T n times to the point x , as being where x has moved to after time n .

We call the sequence $x, T(x), T^2(x), \dots, T^n(x), \dots$ the *orbit of x* . If T is invertible (and so we can iterate backwards by repeatedly applying T^{-1}) then sometimes we refer to the doubly-infinite sequence $\dots, T^{-n}(x), \dots, T^{-1}(x), x, T(x), \dots, T^n(x), \dots$ as the *orbit of x* and the sequence $x, T(x), \dots, T^n(x), \dots$ as the *forward orbit of x* .

As an example, consider the map $T : [0, 1] \rightarrow [0, 1]$ defined by

$$T(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1/2, \\ 2x - 1 & \text{if } 1/2 < x \leq 1. \end{cases}$$

We call this the *doubling map*.

Some orbits for the doubling map are *periodic*, i.e. they return to where they started after a finite number of iterations. For example, $2/5$ is periodic as

$$T(2/5) = 4/5, T(4/5) = 3/5, T(3/5) = 1/5, T(1/5) = 2/5.$$

Thus $T^4(2/5) = 2/5$. We say that $2/5$ has *period 4*.

In general, for a general dynamical system $T : X \rightarrow X$, a point $x \in X$ is a *periodic point* with *period* $n > 0$ if $T^n(x) = x$. (Note that we do not assume that n is least.) If x is a periodic point of period n then we call $\{x, T(x), \dots, T^{n-1}(x)\}$ a *periodic orbit of period n* .

Other points for the doubling map may have a dense orbit in $[0, 1]$. Recall that a set Y is dense in $[0, 1]$ if any point in $[0, 1]$ can be arbitrarily well approximated by a point in Y . Thus the orbit of x is dense in $[0, 1]$ if: for all $x' \in X$ and for all $\varepsilon > 0$ there exists $n > 0$ such that $|T^n(x) - x'| < \varepsilon$.

Consider a subinterval $[a, b] \subset [0, 1]$. How frequently does an orbit of a point under the doubling map visit the interval $[a, b]$? Define the *characteristic function* χ_B of a set B by

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

Then

$$\sum_{j=0}^{n-1} \chi_{[a,b]}(T^j(x))$$

denotes the number of the first n points in the orbit of x that lie in $[a, b]$. Hence

$$\frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(T^j(x))$$

denotes the proportion of the first n points in the orbit of x that lie in $[a, b]$. Hence

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(T^j(x))$$

denotes the frequency with which the orbit of x lies in $[a, b]$. In ergodic theory, one wants to understand when this is equal to the ‘size’ of the interval $[a, b]$ (we will make ‘size’ precise later by using measure theory; for the moment, ‘size’=‘length’). That is, when does

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(T^j(x)) = b - a \quad (1.1.1)$$

for every interval $[a, b]$? Note that if x satisfies (1.1.1) then the proportion of time that the orbit of x spends in an interval $[a, b]$ is equal to the length of that interval; i.e. the orbit of x is equidistributed in $[0, 1]$ and does not favour one region of $[0, 1]$ over another.

In general, one cannot expect (1.1.1) to hold for every point x ; indeed, if x is periodic then (1.1.1) does not hold. Even if the orbit of x is dense, then (1.1.1) may not hold. However, one might expect (1.1.1) to hold for ‘typical’ points $x \in X$ (where again we can make ‘typical’ precise using measure theory). One might also want to replace the function $\chi_{[a,b]}$ with an arbitrary function $f : X \rightarrow \mathbb{R}$. In this case one would want to ask: for the doubling map T , when is it the case that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \int_0^1 f(x) dx?$$

The goal of the course is to understand the statement, prove, and explain how to apply the following result.

Theorem 1.1.1 (Birkhoff’s Ergodic Theorem)

Let (X, \mathcal{B}, μ) be a probability space. Let $f \in L^1(X, \mathcal{B}, \mu)$ be an integrable function. Suppose that $T : X \rightarrow X$ is an ergodic measure-preserving transformation of X . Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \int f d\mu$$

for μ -a.e. point $x \in X$.

Ergodic theory has many applications to other areas of mathematics, notably hyperbolic geometry, number theory, fractal geometry, and mathematical physics. We shall see some of the (simpler) applications to number theory throughout the course.

§1.2 Uniform distribution mod 1

Let $T : X \rightarrow X$ be a dynamical system. In ergodic theory we are interested in the long-term distributional behaviour of the sequence of points $x, T(x), T^2(x), \dots$. Before studying this problem, we consider an analogous problem in the context of sequences of real numbers.

Let $x_n \in \mathbb{R}$ be a sequence of real numbers. We may decompose x_n as the sum of its integer part $[x_n] = \sup\{m \in \mathbb{Z} \mid m \leq x_n\}$ (i.e. the largest integer which is less than or equal to x_n) and its fractional part $\{x_n\} = x_n - [x_n]$. Clearly, $0 \leq \{x_n\} < 1$. The study of $x_n \bmod 1$ is the study of the sequence $\{x_n\}$ in $[0, 1]$.

Definition. We say that the sequence x_n is *uniformly distributed* mod 1 (udm1 for short) if for every a, b with $0 \leq a < b \leq 1$, we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{card} \{0 \leq j \leq n-1 \mid \{x_j\} \in [a, b]\} = b - a.$$

Remarks.

- (i) Here, card denotes the cardinality of a set.
- (ii) Thus x_n is uniformly distributed mod 1 if, given any interval $[a, b] \subset [0, 1]$, the frequency with which the fractional parts of x_n lie in the interval $[a, b]$ is equal to its length, $b - a$.
- (iii) We can replace $[a, b]$ by $[a, b)$, $(a, b]$ or (a, b) without altering the definition.

The following result gives a necessary and sufficient condition for the sequence $x_n \in \mathbb{R}$ to be uniformly distributed mod 1.

Theorem 1.2.1 (Weyl's Criterion)

The following are equivalent:

- (i) the sequence $x_n \in \mathbb{R}$ is uniformly distributed mod 1;
- (ii) for every continuous function $f : [0, 1] \rightarrow \mathbb{R}$ with $f(0) = f(1)$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\{x_j\}) = \int_0^1 f(x) dx; \quad (1.2.1)$$

- (iii) for each $\ell \in \mathbb{Z} \setminus \{0\}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} = 0.$$

Remarks.

- (i) As a grammatical point, criterion is singular (the plural is criteria). Weyl's criterion is that (i) and (iii) are equivalent. Statement (ii) has been included because it is an important intermediate step in the proof and, as we shall see, it closely resembles an ergodic theorem.
- (ii) One can replace the hypothesis that f is continuous in (1.2.1) with f is Riemann integrable.

(iii) To prove that (i) is equivalent to (iii) we work, in fact, not on the unit interval $[0, 1]$ but on the unit circle \mathbb{R}/\mathbb{Z} . To form \mathbb{R}/\mathbb{Z} , we work with real numbers modulo the integers (informally: we ignore integer parts). Note that, ignoring integer parts means that 0 and 1 in $[0, 1]$ are ‘the same’. Thus the end-points of the unit interval ‘join up’ and we see that \mathbb{R}/\mathbb{Z} is a circle. More formally, \mathbb{R} is an additive group, \mathbb{Z} is a subgroup and the quotient group \mathbb{R}/\mathbb{Z} is, topologically, a circle. Note that the requirement in (ii) that $f(0) = f(1)$ means that $f : [0, 1] \rightarrow \mathbb{R}$ is a well-defined function on the circle \mathbb{R}/\mathbb{Z} .

It is, however, the case that (i) is equivalent to (ii) without the hypothesis in (ii) that $f(0) = f(1)$.

§1.2.1 The sequence $x_n = \alpha n$

The behaviour of the sequence $x_n = \alpha n$ depends on whether α is rational or irrational.

If $\alpha \in \mathbb{Q}$ then it is easy to see that $\{\alpha n\}$ can take on only finitely many values in $[0, 1]$. Indeed, if $\alpha = p/q$ ($p \in \mathbb{Z}, q \in \mathbb{Z}, q \neq 0, \text{hcf}(p, q) = 1$) then $\{\alpha n\}$ takes the q values

$$0 = \left\{ \frac{0}{q} \right\}, \left\{ \frac{p}{q} \right\}, \left\{ \frac{2p}{q} \right\}, \dots, \left\{ \frac{(q-1)p}{q} \right\}$$

as $\{qp/q\} = 0$. In particular αn is not uniformly distributed mod 1.

If $\alpha \notin \mathbb{Q}$ then the situation is completely different. We shall show that αn is uniformly distributed mod 1 by applying Weyl’s Criterion. Let $\ell \in \mathbb{Z} \setminus \{0\}$. As $\alpha \notin \mathbb{Q}$ we have that $\ell\alpha$ is never an integer; hence $e^{2\pi i \ell \alpha} \neq 1$. Note that

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} = \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell \alpha j} = \frac{1}{n} \frac{e^{2\pi i \ell \alpha n} - 1}{e^{2\pi i \ell \alpha} - 1}$$

by summing the geometric progression. Hence

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} \right| = \frac{1}{n} \frac{|e^{2\pi i \ell \alpha n} - 1|}{|e^{2\pi i \ell \alpha} - 1|} \leq \frac{1}{n} \frac{2}{|e^{2\pi i \ell \alpha} - 1|}. \quad (1.2.2)$$

As $\alpha \notin \mathbb{Q}$, the denominator in the right-hand side of (1.2.2) is not 0. Letting $n \rightarrow \infty$ we see that

$$\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} \right| = 0.$$

Hence x_n is uniformly distributed mod 1.

Remarks.

1. More generally, we could consider the sequence $x_n = \alpha n + \beta$. It is easy to see by modifying the above argument that x_n is uniformly distributed mod 1 if and only if α is irrational. (See Exercise 1.2.)
2. Fix $\alpha > 1$ and consider the sequence $x_n = \alpha^n x$. Then it is possible to show that for (Lebesgue) almost every $x \in \mathbb{R}$, the sequence x_n is uniformly distributed mod 1. We will prove this, at least for the cases when $\alpha = 2, 3, 4, \dots$

3. Suppose we set $x = 1$ in the above remark and consider the sequence $x_n = \alpha^n$. Then one can show that x_n is uniformly distributed mod 1 for almost every $\alpha > 1$. However, not a single example of such an α is known! Indeed, it is not even known if $(3/2)^n$ is dense mod 1.

§1.2.2 Proof of Weyl's Criterion

We prove (i) implies (ii). Suppose that the sequence $x_n \in \mathbb{R}$ is uniformly distributed mod 1. If $\chi_{[a,b]}$ is the characteristic function of the interval $[a, b]$, then we may rewrite the definition of uniform distribution mod 1 as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(\{x_j\}) = \int_0^1 \chi_{[a,b]}(x) dx.$$

From this we deduce that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} g(\{x_j\}) = \int_0^1 g(x) dx$$

whenever g is a step function, i.e., when $g(x) = \sum_{k=1}^m c_k \chi_{[a_k, b_k]}(x)$ is a finite linear combination of characteristic functions of intervals.

Now let f be a continuous function on $[0, 1]$. Then, given $\varepsilon > 0$, we can find a step function g with $\|f - g\|_\infty \leq \varepsilon$. We have the estimate

$$\begin{aligned} & \left| \frac{1}{n} \sum_{j=0}^{n-1} f(\{x_j\}) - \int_0^1 f(x) dx \right| \\ & \leq \left| \frac{1}{n} \sum_{j=0}^{n-1} (f(\{x_j\}) - g(\{x_j\})) \right| + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(\{x_j\}) - \int_0^1 g(x) dx \right| \\ & \quad + \left| \int_0^1 g(x) dx - \int_0^1 f(x) dx \right| \\ & \leq \frac{1}{n} \sum_{j=0}^{n-1} |f(\{x_j\}) - g(\{x_j\})| + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(\{x_j\}) - \int_0^1 g(x) dx \right| \\ & \quad + \int_0^1 |g(x) - f(x)| dx \\ & \leq 2\varepsilon + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(\{x_j\}) - \int_0^1 g(x) dx \right|. \end{aligned}$$

Since the last term converges to zero as $n \rightarrow \infty$, we obtain

$$\limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{j=0}^{n-1} f(\{x_j\}) - \int_0^1 f(x) dx \right| \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this gives us that

$$\frac{1}{n} \sum_{j=0}^{n-1} f(\{x_j\}) \rightarrow \int_0^1 f(x) dx$$

as $n \rightarrow \infty$.

We now prove (ii) implies (iii). Suppose that $f : [0, 1] \rightarrow \mathbb{C}$ is continuous and $f(0) = f(1)$. By writing $f = \operatorname{Re}f + i\operatorname{Im}f$ and applying (ii) to the real and imaginary parts of f we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\{x_j\}) = \int_0^1 f(x) dx.$$

For $\ell \in \mathbb{Z}, \ell \neq 0$ we let $f(x) = e^{2\pi i \ell x}$. Note that, as \exp is $2\pi i$ -periodic, $f(\{x_j\}) = e^{2\pi i \ell x_j}$. Hence

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} = \int_0^1 e^{2\pi i \ell x} dx = \frac{1}{2\pi i \ell} e^{2\pi i \ell x} \Big|_0^1 = 0,$$

as $\ell \neq 0$.

We prove (iii) implies (i). Suppose that (iii) holds. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} g(\{x_j\}) = \int_0^1 g(x) dx$$

whenever $g(x) = \sum_{k=1}^m c_k e^{2\pi i \ell_k x}$ is a trigonometric polynomial, $c_k \in \mathbb{C}$, i.e. a finite linear combination of exponential functions.

Note that the space $C(X, \mathbb{C})$ is a vector space: if $f, g \in C(X, \mathbb{C})$ then $f + g \in C(X, \mathbb{C})$ and if $f \in C(X, \mathbb{C}), \lambda \in \mathbb{C}$ then $\lambda f \in C(X, \mathbb{C})$. A linear subspace $\mathcal{S} \subset C(X, \mathbb{C})$ is an *algebra* if whenever $f, g \in \mathcal{S}$ then $fg \in \mathcal{S}$. We will need the following result:

Theorem 1.2.2 (Stone-Weierstrass Theorem)

Let X be a compact metric space and let $C(X, \mathbb{C})$ denote the space of continuous functions defined on X . Suppose that $\mathcal{S} \subset C(X, \mathbb{C})$ is an algebra of continuous functions such that

- (i) if $f \in \mathcal{S}$ then $\bar{f} \in \mathcal{S}$,
- (ii) \mathcal{S} separates the points of X , i.e. for all $x, y \in X, x \neq y$, there exists $f \in \mathcal{S}$ such that $f(x) \neq f(y)$,
- (iii) for every $x \in X$ there exists $f \in \mathcal{S}$ such that $f(x) \neq 0$.

Then \mathcal{S} is uniformly dense in $C(X, \mathbb{C})$, i.e. for all $f \in C(X, \mathbb{C})$ and all $\varepsilon > 0$, there exists $g \in \mathcal{S}$ such that $\|f - g\|_\infty = \sup_{x \in X} |f(x) - g(x)| < \varepsilon$.

We shall apply the Stone-Weierstrass Theorem with \mathcal{S} given by the set of trigonometric polynomials. It is easy to see that \mathcal{S} satisfies the hypotheses of Theorem 1.2.2. Let f be any continuous function on $[0, 1]$ with $f(0) = f(1)$. Given $\varepsilon > 0$ we can find a trigonometric polynomial g such that $\|f - g\|_\infty \leq \varepsilon$. As in the first part of the proof, we can conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\{x_j\}) = \int_0^1 f(x) dx.$$

Now consider the interval $[a, b] \subset [0, 1]$. Given $\varepsilon > 0$, we can find continuous functions f_1 and f_2 (with $f_1(0) = f_1(1), f_2(0) = f_2(1)$) such that

$$f_1 \leq \chi_{[a,b]} \leq f_2$$

and

$$\int_0^1 f_2(x) - f_1(x) dx \leq \varepsilon.$$

We then have that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(\{x_j\}) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_1(\{x_j\}) = \int_0^1 f_1(x) dx \\ &\geq \int_0^1 f_2(x) dx - \varepsilon \geq \int_0^1 \chi_{[a,b]}(x) dx - \varepsilon \end{aligned}$$

and

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(\{x_j\}) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_2(\{x_j\}) = \int_0^1 f_2(x) dx \\ &\leq \int_0^1 f_1(x) dx + \varepsilon \leq \int_0^1 \chi_{[a,b]}(x) dx + \varepsilon. \end{aligned}$$

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(\{x_j\}) = \int_0^1 \chi_{[a,b]}(x) dx = b - a,$$

so that x_n is uniformly distributed mod 1. □

§1.2.3 Exercises

Exercise 1.1

Show that if x_n is uniformly distributed mod 1 then $\{x_n\}$ is dense in $[0, 1]$. (The converse is not true.)

Exercise 1.2

Let $\alpha, \beta \in \mathbb{R}$. Let $x_n = \alpha n + \beta$. Show that x_n is uniformly distributed mod 1 if and only if $\alpha \notin \mathbb{Q}$.

Exercise 1.3

(i) Prove that $\log_{10} 2$ is irrational.

(ii) The leading digit of an integer is the left-most digit of its base 10 representation. (Thus the leading digit of 32 is 3, the leading digit of 1024 is 1, etc.) Show that the frequency with which 2^n has leading digit r ($r = 1, 2, \dots, 9$) is $\log_{10}(1 + 1/r)$.

(Hint: first show that 2^n has leading digit r if and only if

$$r10^k \leq 2^n < (r+1)10^k$$

for some $k \in \mathbb{N}$.)

Exercise 1.4

Calculate the frequency with which the penultimate leading digit of 2^n is equal to r , $r = 0, 1, 2, \dots, 9$. (The penultimate leading digit is the second-to-leftmost digit in the base 10 expansion. The penultimate leading digit of 2048 is 0, etc.)

§1.3 Appendix: a recap on the Riemann Integral

(This subsection is included for general interest and to motivate the Lebesgue integral. Hence it is not examinable.)

You have probably already seen the construction of the Riemann integral. This gives a method for defining the integral of suitable functions defined on an interval $[a, b]$. In the next section we will see how the Lebesgue integral is a generalisation of the Riemann integral in the sense that it allows us to integrate functions defined on spaces more general than subintervals of \mathbb{R} . The Lebesgue integral has other nice properties, for example it is well-behaved with respect to limits. Here we give a brief exposition about some inadequacies of the Riemann integral and how they motivate the Lebesgue integral.

Let $f : [a, b] \rightarrow \mathbb{R}$ be a bounded function (for the moment we impose no other conditions on f).

A *partition* Δ of $[a, b]$ is a finite set of points $\Delta = \{x_0, x_1, x_2, \dots, x_n\}$ with

$$a = x_0 < x_1 < x_2 < \dots < x_n = b.$$

In other words, we are dividing $[a, b]$ up into subintervals.

We then form the *upper* and *lower Riemann sums*

$$U(f, \Delta) = \sum_{i=0}^{n-1} \sup_{x \in [x_i, x_{i+1}]} f(x) (x_{i+1} - x_i),$$

$$L(f, \Delta) = \sum_{i=0}^{n-1} \inf_{x \in [x_i, x_{i+1}]} f(x) (x_{i+1} - x_i).$$

The idea is then that if we make the subintervals in the partition small, these sums will be a good approximation to our intuitive notion of the integral of f over $[a, b]$ as the area bounded by the graph of f . More precisely, if

$$\inf_{\Delta} U(f, \Delta) = \sup_{\Delta} L(f, \Delta),$$

where the infimum and supremum are taken over all possible partitions of $[a, b]$, then we write

$$\int_a^b f(x) dx$$

for their common value and call it the (*Riemann*) *integral* of f between those limits. We also say that f is *Riemann integrable*.

The class of Riemann integrable functions includes continuous functions and step functions (i.e. finite linear combinations of characteristic functions of intervals).

However, there are many functions for which one wishes to define an integral but which are not Riemann integrable, making the theory rather unsatisfactory. For example, define $f : [0, 1] \rightarrow \mathbb{R}$ by

$$f(x) = \chi_{\mathbb{Q} \cap [0, 1]}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{otherwise.} \end{cases}$$

Since between any two distinct real numbers we can find both a rational number and an irrational number, given $0 \leq y < z \leq 1$, we can find $y < x < z$ with $f(x) = 1$ and $y < x' < z$

with $f(x') = 0$. Hence for any partition $\Delta = \{x_0, x_1, \dots, x_n\}$ of $[0, 1]$, we have

$$\begin{aligned} U(f, \Delta) &= \sum_{i=0}^{n-1} (x_{i+1} - x_i) = 1, \\ L(f, \Delta) &= 0. \end{aligned}$$

Taking the infimum and supremum, respectively, over all partitions Δ shows that f is *not* Riemann integrable.

Why does Riemann integration not work for the above function and how could we go about improving it? Let us look again at (and slightly rewrite) the formulæ for $U(f, \Delta)$ and $L(f, \Delta)$. We have

$$U(f, \Delta) = \sum_{i=0}^{n-1} \sup_{x \in [x_i, x_{i+1}]} f(x) \lambda([x_i, x_{i+1}])$$

and

$$L(f, \Delta) = \sum_{i=0}^{n-1} \inf_{x \in [x_i, x_{i+1}]} f(x) \lambda([x_i, x_{i+1}]),$$

where, for an interval $[y, z]$,

$$\lambda([y, z]) = z - y$$

denotes its *length*. In the example above, things did not work because dividing $[0, 1]$ into intervals (no matter how small) did not ‘separate out’ the different values that f could take. But suppose we had a notion of ‘length’ that worked for more general sets than intervals. Then we could do better by considering more complicated ‘partitions’ of $[0, 1]$, where by partition we now mean a collection of subsets $\{E_1, \dots, E_m\}$ of $[0, 1]$ such that $E_i \cap E_j = \emptyset$, if $i \neq j$, and $\bigcup_{i=1}^m E_i = [0, 1]$.

In the example, for instance, it might be reasonable to write

$$\begin{aligned} \int_0^1 f(x) dx &= 1 \times \lambda([0, 1] \cap \mathbb{Q}) + 0 \times \lambda([0, 1] \setminus \mathbb{Q}) \\ &= \lambda([0, 1] \cap \mathbb{Q}). \end{aligned}$$

Instead of using subintervals, the Lebesgue integral uses a much wider class of subsets (namely, sets in a given σ -algebra) together with a notion of ‘generalised length’ (namely, measure).

2. More on uniform distribution mod 1. Measure spaces

§2.1 Uniform distribution of sequences in \mathbb{R}^k

We shall now look at the uniform distribution of sequences in \mathbb{R}^k . We will say that a sequence $\mathbf{x}_n = (x_{n,1}, \dots, x_{n,k}) \in \mathbb{R}^k$ is uniformly distributed mod 1 if, given any k -dimensional cube, the frequency with which the fractional parts of \mathbf{x}_n lie in the cube is equal to its k -dimensional volume.

Definition. A sequence $\mathbf{x}_n = (x_{n,1}, \dots, x_{n,k}) \in \mathbb{R}^k$ is said to be *uniformly distributed mod 1* if, for each choice of k intervals $[a_1, b_1], \dots, [a_k, b_k] \subset [0, 1]$, we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \text{card}\{j \in \{0, 1, \dots, n-1\} \mid \mathbf{x}_j \in [a_1, b_1] \times \dots \times [a_k, b_k]\} = (b_1 - a_1) \cdots (b_k - a_k).$$

We have the following criterion for uniform distribution.

Theorem 2.1.1 (Multi-dimensional Weyl's Criterion)

Let $\mathbf{x}_n = (x_{n,1}, \dots, x_{n,k}) \in \mathbb{R}^k$. The following are equivalent:

- (i) the sequence $\mathbf{x}_n \in \mathbb{R}^k$ is uniformly distributed mod 1;
- (ii) for any continuous function $f : \mathbb{R}^k / \mathbb{Z}^k \rightarrow \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\{x_{j,1}\}, \dots, \{x_{j,k}\}) = \int \cdots \int f(x_1, \dots, x_k) dx_1 \dots dx_k;$$

- (iii) for all $\underline{\ell} = (\ell_1, \dots, \ell_k) \in \mathbb{Z}^k \setminus \{\mathbf{0}\}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i(\ell_1 x_{j,1} + \cdots + \ell_k x_{j,k})} = 0.$$

Remark. Here and throughout $\mathbf{0} \in \mathbb{Z}^k$ denotes the zero vector $(0, \dots, 0)$.

Remark. In §1 we commented that, topologically, the quotient group \mathbb{R}/\mathbb{Z} is a circle. More generally, the quotient group $\mathbb{R}^k/\mathbb{Z}^k$ is a k -dimensional torus.

Remark. Consider the case when $k = 2$ so that $\mathbb{R}^2/\mathbb{Z}^2$ is the 2-dimensional torus. We can regard $\mathbb{R}^2/\mathbb{Z}^2$ as the square $[0, 1] \times [0, 1]$ with the top and bottom sides identified and left and right sides identified. Thus a continuous function $f : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{R}$ has the property that $f(0, y) = f(1, y)$ and $f(x, 0) = f(x, 1)$. More generally, we can identify the k -dimensional torus $\mathbb{R}^k/\mathbb{Z}^k$ with $[0, 1]^k$ with $(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_k)$ and $(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_k)$ identified, $1 \leq i \leq k$. A continuous function $f : \mathbb{R}^k/\mathbb{Z}^k \rightarrow \mathbb{R}$ then corresponds to a continuous function $f : [0, 1]^k \rightarrow \mathbb{R}$ such that

$$f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_k) = f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_k)$$

for each i , $1 \leq i \leq k$.

Proof of Theorem 2.1.1. The proof of Theorem 2.1.1 is essentially the same as in the case $k = 1$. \square

§2.2 The sequence $\mathbf{x}_n = (\alpha_1 n, \dots, \alpha_k n)$

We shall apply Theorem 2.1.1 to the sequence $\mathbf{x}_n = (\alpha_1 n, \dots, \alpha_k n)$, for real numbers $\alpha_1, \dots, \alpha_k$.

Definition. Real numbers $\beta_1, \dots, \beta_s \in \mathbb{R}$ are said to be *rationally independent* if the only rationals $r_1, \dots, r_s \in \mathbb{Q}$ such that

$$r_1 \beta_1 + \dots + r_s \beta_s = 0$$

are $r_1 = \dots = r_s = 0$.

Proposition 2.2.1

Let $\alpha_1, \dots, \alpha_k \in \mathbb{R}$. Then the following are equivalent:

- (i) the sequence $\mathbf{x}_n = (\alpha_1 n, \dots, \alpha_k n) \in \mathbb{R}^k$ is uniformly distributed mod 1;
- (ii) $\alpha_1, \dots, \alpha_k$ and 1 are rationally independent.

Proof. The proof is similar to the discussion in §1.2.1 and we leave it as an exercise. (See Exercise 2.1.) \square

Remark. Note that in the case $k = 1$, Proposition 2.2.1 reduces to the results of §1.2.1. To see this, note that $\alpha, 1$ are rationally dependent if and only if there exist rationals r, s (not both zero) such that $r\alpha + s = 0$. This holds if and only if α is rational. Hence $\alpha, 1$ are rationally independent if and only if α is irrational.

§2.3 Weyl's Theorem on Polynomials

We have seen that $\alpha n + \beta$ is uniformly distributed mod 1 if α is irrational. Weyl's Theorem generalises this to polynomials of higher degree. Write

$$p(n) = \alpha_k n^k + \alpha_{k-1} n^{k-1} + \dots + \alpha_1 n + \alpha_0.$$

Theorem 2.3.1 (Weyl's Theorem on Polynomials)

If any one of $\alpha_1, \dots, \alpha_k$ is irrational then $p(n)$ is uniformly distributed mod 1.

To prove this theorem we shall need the following technical result.

Lemma 2.3.2 (van der Corput's Inequality)

Let $z_0, \dots, z_{n-1} \in \mathbb{C}$ and let $1 < m < n$. Then

$$m^2 \left| \sum_{j=0}^{n-1} z_j \right|^2 \leq m(n+m-1) \sum_{j=0}^{n-1} |z_j|^2 + 2(n+m-1) \operatorname{Re} \sum_{j=1}^{m-1} (m-j) \sum_{i=0}^{n-1-j} z_{i+j} \bar{z}_i.$$

Let $x_n \in \mathbb{R}$. For each $m \geq 1$ define the sequence $x_n^{(m)} = x_{n+m} - x_n$ to be the sequence of m^{th} differences. The following lemma allows us to infer the uniform distribution of the sequence x_n if we know the uniform distribution of each of the m^{th} differences of x_n .

Lemma 2.3.3

Let $x_n \in \mathbb{R}$ be a sequence. Suppose that for each $m \geq 1$ the sequence $x_n^{(m)}$ of m^{th} differences is uniformly distributed mod 1. Then x_n is uniformly distributed mod 1.

Proof. We shall apply Weyl's Criterion. We need to show that if $\ell \in \mathbb{Z} \setminus \{0\}$ then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} = 0.$$

Let $z_j = e^{2\pi i \ell x_j}$ for $j = 0, \dots, n-1$. Note that $|z_j| = 1$. Let $1 < m < n$. By van der Corput's inequality,

$$\begin{aligned} \frac{m^2}{n^2} \left| \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} \right|^2 &\leq \frac{m}{n^2} (n+m-1)n + \frac{2(n+m-1)}{n} \operatorname{Re} \sum_{j=1}^{m-1} \frac{(m-j)}{n} \sum_{i=0}^{n-1-j} e^{2\pi i \ell (x_{i+j} - x_i)} \\ &= \frac{m}{n} (m+n-1) + \frac{2(n+m-1)}{n} \operatorname{Re} \sum_{j=1}^{m-1} (m-j) A_{n,j} \end{aligned}$$

where

$$A_{n,j} = \frac{1}{n} \sum_{i=0}^{n-1-j} e^{2\pi i \ell (x_{i+j} - x_i)} = \frac{1}{n} \sum_{i=0}^{n-1-j} e^{2\pi i \ell x_i^{(j)}}.$$

As the sequence $x_i^{(j)}$ of j^{th} differences is uniformly distributed mod 1, by Weyl's criterion we have that $A_{n,j} \rightarrow 0$ for each $j = 1, \dots, m-1$. Hence for each $m \geq 1$

$$\limsup_{n \rightarrow \infty} \frac{m^2}{n^2} \left| \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} \right|^2 \leq \limsup_{n \rightarrow \infty} m \frac{(n+m-1)}{n} = m.$$

Hence, for each $m > 1$ we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \left| \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} \right| \leq \frac{1}{\sqrt{m}}.$$

As $m > 1$ is arbitrary, the result follows. \square

Proof of Weyl's Theorem. We will only prove Weyl's Theorem on Polynomials (Theorem 2.3.1) in the special case where the leading coefficient α_k of

$$p(n) = \alpha_k n^k + \dots + \alpha_1 n + \alpha_0$$

is irrational. (The general case, where α_i is irrational for some $1 \leq i \leq k$, can be deduced easily from this special case and we leave this as an exercise. See Exercise 2.2.)

We shall use induction on the degree of p . Let $\Delta(k)$ denote the statement 'for every polynomial p of degree $\leq k$, with irrational leading coefficient, the sequence $p(n)$ is uniformly distributed mod 1'. We know that $\Delta(1)$ is true; this follows immediately from Exercise 1.2.

Suppose that $\Delta(k-1)$ is true. Let $p(n) = \alpha_k n^k + \cdots + \alpha_1 n + \alpha_0$ be any polynomial of degree k with α_k irrational. Let $m \in \mathbb{N}$ and consider the sequence $p^{(m)}(n) = p(n+m) - p(n)$ of m th differences. We have that

$$\begin{aligned} p^{(m)}(n) &= p(n+m) - p(n) \\ &= \alpha_k(n+m)^k + \alpha_{k-1}(n+m)^{k-1} + \cdots + \alpha_1(n+m) + \alpha_0 \\ &\quad - \alpha_k n^k - \alpha_{k-1} n^{k-1} - \cdots - \alpha_1 n - \alpha_0 \\ &= \alpha_k n^k + \alpha_k k n^{k-1} m + \cdots + \alpha_{k-1} n^{k-1} + \alpha_{k-1}(k-1)n^{k-2} m \\ &\quad + \cdots + \alpha_1 n + \alpha_1 m + \alpha_0 - \alpha_k n^k - \alpha_{k-1} n^{k-1} - \cdots - \alpha_1 n - \alpha_0. \end{aligned}$$

After cancellation, we can see that, for each m , $p^{(m)}(n)$ is a polynomial of degree $k-1$ with irrational leading coefficient $\alpha_k k m$. Therefore, by the inductive hypothesis, $p^{(m)}(n)$ is uniformly distributed mod 1. We may now apply Lemma 2.3.3 to conclude that $p(n)$ is uniformly distributed mod 1 and so $\Delta(k)$ holds. This completes the induction. \square

§2.4 Measures and the Lebesgue integral

You may have seen the definition of Lebesgue measure, Lebesgue outer measure and the Lebesgue integral in other courses, for example in Fourier Analysis and Lebesgue Integration. The theory developed in that course is one particular example of a more general theory, which we sketch here. Measure theory is a key technical tool in ergodic theory, and so a good knowledge of measures and integration is essential for this course (although we will not need to know the (many) technical intricacies).

§2.4.1 Measure spaces

Loosely speaking, a measure is a function that, when given a subset of a space X , will say how ‘big’ that subset is. A motivating example is given by Lebesgue measure on $[0, 1]$. The Lebesgue measure of an interval $[a, b]$ is given by its length $b - a$. In defining an abstract measure space, we will be taking the properties of ‘length’ (or, in higher dimensions, ‘volume’) and abstracting them, in much the same way that a metric space abstracts the properties of ‘distance’.

It turns out that in general it is not possible to be able to define the measure of an arbitrary subset of X . Instead, we will usually have to restrict our attention to a class of subsets of X .

Definition. A collection \mathcal{B} of subsets of X is called a σ -algebra if the following properties hold:

- (i) $\emptyset \in \mathcal{B}$,
- (ii) if $E \in \mathcal{B}$ then its complement $X \setminus E \in \mathcal{B}$,
- (iii) if $E_n \in \mathcal{B}$, $n = 1, 2, 3, \dots$, is a countable sequence of sets in \mathcal{B} then their union $\bigcup_{n=1}^{\infty} E_n \in \mathcal{B}$.

Definition. If X is a set and \mathcal{B} a σ -algebra of subsets of X then we call (X, \mathcal{B}) a *measurable space*.

Examples.

1. The trivial σ -algebra is given by $\mathcal{B} = \{\emptyset, X\}$.
2. The full σ -algebra is given by $\mathcal{B} = \mathcal{P}(X)$, i.e. the collection of *all* subsets of X .

Remark. In general, the trivial σ -algebra is too small and the full σ -algebra is too big. We shall see some more interesting examples of σ -algebras later.

Here are some easy properties of σ -algebras:

Lemma 2.4.1

Let \mathcal{B} be a σ -algebra of subsets of X . Then

- (i) $X \in \mathcal{B}$;
- (ii) if $E_n \in \mathcal{B}$ then $\bigcap_{n=1}^{\infty} E_n \in \mathcal{B}$.

In the special case when X is a compact metric space there is a particularly important σ -algebra.

Definition. Let X be a compact metric space. We define the *Borel σ -algebra* $\mathcal{B}(X)$ to be the smallest σ -algebra of subsets of X which contains all the open subsets of X .

Remarks.

1. By ‘smallest’ we mean that if \mathcal{C} is another σ -algebra that contains all open subsets of X then $\mathcal{B}(X) \subset \mathcal{C}$, that is:

$$\mathcal{B}(X) = \bigcap \{ \mathcal{C} \mid \mathcal{C} \text{ is a } \sigma\text{-algebra that contains the open sets} \}.$$

2. We say that the Borel σ -algebra is generated by the open sets. We call a set in $\mathcal{B}(X)$ a *Borel set*.
3. By Definition 2.4.1(ii), the Borel σ -algebra also contains all the closed sets and is the smallest σ -algebra with this property.
4. By Lemma 2.4.1 it follows that \mathcal{B} contains all countable intersections of open sets, all countable unions of countable intersections of open sets, all countable intersections of countable unions of countable intersections of open sets, etc—and indeed many other sets.
5. There are plenty of sets that are not Borel sets, although by necessity they are rather complicated. For example, consider \mathbb{R} as an additive group and $\mathbb{Q} \subset \mathbb{R}$ as a subgroup. Form the quotient group \mathbb{R}/\mathbb{Q} and choose an element in $[0, 1]$ for each coset (this requires the Axiom of Choice.) The set E of coset representatives is a non-Borel set.
6. In the case when $X = [0, 1]$ or \mathbb{R}/\mathbb{Z} , the Borel σ -algebra is also the smallest σ -algebra that contains all sub-intervals.

Let X be a set and let \mathcal{B} be a σ -algebra of subsets of X .

Definition. A function $\mu : \mathcal{B} \rightarrow \mathbb{R}$ is called a (finite) *measure* if:

- (i) $\mu(\emptyset) = 0$;

- (ii) if E_n is a countable collection of pairwise disjoint sets in \mathcal{B} (i.e. $E_n \cap E_m = \emptyset$ for $n \neq m$) then

$$\mu \left(\bigcup_{n=1}^{\infty} E_n \right) = \sum_{n=1}^{\infty} \mu(E_n).$$

We call (X, \mathcal{B}, μ) a *measure space*.

If $\mu(X) = 1$ then we call μ a *probability* or *probability measure* and refer to (X, \mathcal{B}, μ) as a *probability space*.

Remark. Thus a measure just abstracts properties of ‘length’ or ‘volume’. Condition (i) says that the empty set has zero length, and condition (ii) says that the length of a disjoint union is the sum of the lengths of the individual sets.

Definition. We say that a property holds *almost everywhere* if the set of points on which the property fails to hold has measure zero. We will often abbreviate this to ‘ μ -a.e.’ or to ‘a.e.’ when the implied measure is clear from the context.

Example. We shall see (Exercise 2.9) that the set of rationals in $[0, 1]$ forms a Borel set with zero Lebesgue measure. Thus Lebesgue almost every point in $[0, 1]$ is irrational. (Thus, ‘typical’ (in the sense of measure theory, and with respect to Lebesgue measure) points in $[0, 1]$ are irrational.)

We will usually be interested in studying measures on the Borel σ -algebra of a compact metric space X . To define such a measure, we need to define the measure of an arbitrary Borel set. In general, the Borel σ -algebra is extremely large. We shall see that it is often unnecessary to do this and instead it is sufficient to define the measure of a certain class of subsets.

§2.4.2 The Hahn-Kolmogorov Extension Theorem

A collection \mathcal{A} of subsets of X is called an *algebra* if:

- (i) $\emptyset \in \mathcal{A}$,
- (ii) if $A_1, A_2, \dots, A_n \in \mathcal{A}$ then $\bigcup_{j=1}^n A_j \in \mathcal{A}$,
- (iii) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.

Thus an algebra is like a σ -algebra, except that it is closed under finite unions and not necessarily closed under countable unions.

Example. Take $X = [0, 1]$, and $\mathcal{A} = \{\text{all finite unions of subintervals}\}$.

Let $\mathcal{B}(\mathcal{A})$ denote the σ -algebra generated by \mathcal{A} , i.e., the smallest σ -algebra containing \mathcal{A} . More precisely:

$$\mathcal{B}(\mathcal{A}) = \bigcap \{ \mathcal{C} \mid \mathcal{C} \text{ is a } \sigma\text{-algebra, } \mathcal{C} \supset \mathcal{A} \}.$$

In the case when $X = [0, 1]$ and \mathcal{A} is the algebra of finite unions of intervals, we have that $\mathcal{B}(\mathcal{A})$ is the Borel σ -algebra. Indeed, in the special case of the Borel σ -algebra of a compact metric space X , it is usually straightforward to check whether an algebra generates the Borel σ -algebra.

Proposition 2.4.2

Let X be a compact metric space and let \mathcal{B} be the Borel σ -algebra. Let \mathcal{A} be an algebra of Borel subsets, $\mathcal{A} \subset \mathcal{B}$. Suppose that for every $x_1, x_2 \in X$, $x_1 \neq x_2$, there exist disjoint open sets $A_1, A_2 \in \mathcal{A}$ such that $x_1 \in A_1, x_2 \in A_2$. Then \mathcal{A} generates the Borel σ -algebra \mathcal{B} .

The following result says that if we have a function which looks like a measure defined on an algebra, then it extends uniquely to a measure defined on the σ -algebra generated by the algebra.

Theorem 2.4.3 (Hahn-Kolmogorov Extension Theorem)

Let \mathcal{A} be an algebra of subsets of X . Suppose that $\mu : \mathcal{A} \rightarrow [0, 1]$ satisfies:

- (i) $\mu(\emptyset) = 0$;
- (ii) if $A_n \in \mathcal{A}$, $n \geq 1$, are pairwise disjoint and if $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ then

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Then there is a unique probability measure $\mu : \mathcal{B}(\mathcal{A}) \rightarrow [0, 1]$ which is an extension of $\mu : \mathcal{A} \rightarrow [0, 1]$.

Remarks.

- (i) We will often use the Hahn-Kolmogorov Extension Theorem as follows. Take $X = [0, 1]$ and take \mathcal{A} to be the algebra consisting of all finite unions of subintervals of X . We then define the ‘measure’ μ of a subinterval in such a way as to be consistent with the hypotheses of the Hahn-Kolmogorov Extension Theorem. It then follows that μ does indeed define a measure on the Borel σ -algebra.
- (ii) Here is another way in which we shall use the Hahn-Kolmogorov Extension Theorem. Suppose we have two measures, μ and ν , and we want to see if $\mu = \nu$. A priori we would have to check that $\mu(B) = \nu(B)$ for all $B \in \mathcal{B}$. The Hahn-Kolmogorov Extension Theorem says that it is sufficient to check that $\mu(A) = \nu(A)$ for all A in an algebra \mathcal{A} that generates \mathcal{B} . For example, to show that two Borel probability measures on $[0, 1]$ are equal, it is sufficient to show that they give the same measure to each subinterval.
- (iii) There is a more general version of the Hahn-Kolmogorov Extension Theorem for the case when X does not have finite measure (indeed, this is the setting in which the Hahn-Kolmogorov Theorem is usually stated). Suppose that X is a set, \mathcal{B} is a σ -algebra of subsets of X , and \mathcal{A} is an algebra that generates \mathcal{B} . Suppose that $\mu : \mathcal{A} \rightarrow \mathbb{R} \cup \{\infty\}$ satisfies conditions (i) and (ii) of Theorem 2.4.3. Suppose in addition that there exist a countable number of sets $A_n \in \mathcal{A}$, $n = 1, 2, 3, \dots$ such that $X = \bigcup_{n=1}^{\infty} A_n$ such that $\mu(A_n) < 1$. Then there exists a unique measure $\mu : \mathcal{B}(\mathcal{A}) \rightarrow \mathbb{R} \cup \{\infty\}$ which is an extension of $\mu : \mathcal{A} \rightarrow \mathbb{R} \cup \{\infty\}$.

A consequence of the proof (which we omit) of the Hahn-Kolmogorov Extension Theorem is that sets in \mathcal{B} can be arbitrarily well approximated by sets in \mathcal{A} in the following sense. We define the *symmetric difference* between two sets A, B by

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Thus, two sets are ‘close’ if their symmetric difference is small.

Proposition 2.4.4

Suppose that \mathcal{A} is an algebra that generates the σ -algebra \mathcal{B} . Let $B \in \mathcal{B}$ and let $\varepsilon > 0$. Then there exists $A \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$.

Remark. It is straightforward to check that if $\mu(A \Delta B) < \varepsilon$ then $|\mu(A) - \mu(B)| < \varepsilon$.

§2.4.3 Examples of measure spaces

Lebesgue measure on $[0, 1]$. Take $X = [0, 1]$ and take \mathcal{A} to be the collection of all finite unions of subintervals of $[0, 1]$. For a subinterval $[a, b]$ define

$$\mu([a, b]) = b - a.$$

This satisfies the hypotheses of the Hahn-Kolmogorov Extension Theorem, and so defines a measure on the Borel σ -algebra \mathcal{B} . This is *Lebesgue measure*.

Lebesgue measure on \mathbb{R}/\mathbb{Z} . Take $X = \mathbb{R}/\mathbb{Z}$ and take \mathcal{A} to be the collection of all finite unions of subintervals of $[0, 1)$. For a subinterval $[a, b]$ define

$$\mu([a, b]) = b - a.$$

This satisfies the hypotheses of the Hahn-Kolmogorov Extension Theorem, and so defines a measure on the Borel σ -algebra \mathcal{B} . This is *Lebesgue measure on the circle*.

Lebesgue measure on the k -dimensional torus. Take $X = \mathbb{R}^k/\mathbb{Z}^k$ and take \mathcal{A} to be the collection of all finite unions of k -dimensional sub-cubes $\prod_{j=1}^k [a_j, b_j]$ of $[0, 1]^k$. For a sub-cube $\prod_{j=1}^k [a_j, b_j]$ of $[0, 1]^k$, define

$$\mu\left(\prod_{j=1}^k [a_j, b_j]\right) = \prod_{j=1}^k (b_j - a_j).$$

This satisfies the hypotheses of the Hahn-Kolmogorov Extension Theorem, and so defines a measure on the Borel σ -algebra \mathcal{B} . This is *Lebesgue measure on the torus*.

Stieltjes measures.¹ Take $X = [0, 1]$ and let $\rho : [0, 1] \rightarrow \mathbb{R}^+$ be an increasing function such that $\rho(1) - \rho(0) = 1$. Take \mathcal{A} to be the algebra of finite unions of subintervals and define

$$\mu_\rho([a, b]) = \rho(b) - \rho(a).$$

This satisfies the hypotheses of the Hahn-Kolmogorov Extension Theorem, and so defines a measure on the Borel σ -algebra \mathcal{B} . We say that μ_ρ is the measure on $[0, 1]$ with density ρ .

Dirac measures. Finally, we give an example of a class of measures that do not fall into the above categories. Let X be an arbitrary space and let \mathcal{B} be an arbitrary σ -algebra. Let $x \in X$. Define the measure δ_x by

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

Then δ_x defines a probability measure. It is called the *Dirac measure at x* .

¹An approximate pronunciation of Stieltjes is ‘Steel-tyuz’.

§2.5 Exercises

Exercise 2.1

Prove Proposition 2.2.1: let $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ and let $\mathbf{x}_n = (\alpha_1 n, \dots, \alpha_k n) \in \mathbb{R}^k$. Prove that x_n is uniformly distributed mod 1 if and only if $\alpha_1, \dots, \alpha_k, 1$ are rationally independent.

Exercise 2.2

Deduce the general case of Weyl's Theorem on Polynomials (where at least one non-constant coefficient is irrational) from the special case proved above (where the leading coefficient is irrational).

Exercise 2.3

Let α be irrational. Show that $p(n) = \alpha n^2 + n + 1$ is uniformly distributed mod 1 by using Lemma 2.3.3 and Exercise 1.2: i.e. show that, for each $m \geq 1$, the sequence $p^{(m)}(n) = p(n+m) - p(n)$ of m th differences is uniformly distributed mod 1.

Exercise 2.4

Let $p(n) = \alpha_k n^k + \alpha_{k-1} n^{k-1} + \dots + \alpha_1 n + \alpha_0$, $q(n) = \beta_k n^k + \beta_{k-1} n^{k-1} + \dots + \beta_1 n + \beta_0$. Show that $(p(n), q(n)) \in \mathbb{R}^2$ is uniformly distributed mod 1 if, for some $1 \leq i \leq k$, α_i, β_i and 1 are rationally independent.

Exercise 2.5

Prove Lemma 2.4.1.

Exercise 2.6

Let $X = [0, 1]$. Find the smallest σ -algebra that contains the sets: $[0, 1/4]$, $[1/4, 1/2]$, $[1/2, 3/4]$, and $[3/4, 1]$

Exercise 2.7

Let $X = [0, 1]$ and let \mathcal{B} denote the Borel σ -algebra. A dyadic interval is an interval of the form

$$\left[\frac{p_1}{2^k}, \frac{p_2}{2^k} \right], \quad p_1, p_2 \in \{0, 1, \dots, 2^k\}.$$

Show that the algebra formed by taking finite unions of all dyadic intervals (over all $k \in \mathbb{N}$) generates the Borel σ -algebra.

Exercise 2.8

Show that $\mathcal{A} = \{\text{all finite unions of subintervals of } [0, 1]\}$ is an algebra.

Exercise 2.9

Let μ denote Lebesgue measure on $[0, 1]$. Show that for any $x \in [0, 1]$ we have that $\mu(\{x\}) = 0$. Hence show that the Lebesgue measure of any countable set is zero.

Show that Lebesgue almost every point in $[0, 1]$ is irrational.

Exercise 2.10

Let $X = [0, 1]$. Let $\mu = \delta_{1/2}$ denote the Dirac δ -measure at $1/2$. Show that

$$\mu([0, 1/2) \cup (1/2, 1]) = 0.$$

Conclude that μ -almost every point in $[0, 1]$ is equal to $1/2$.

3. Lebesgue integration. Invariant measures

§3.1 Lebesgue integration

Let (X, \mathcal{B}, μ) be a measure space. We are interested in how to integrate functions defined on X with respect to the measure μ . In the special case when $X = [0, 1]$, \mathcal{B} is the Borel σ -algebra and μ is Lebesgue measure, this will extend the definition of the Riemann integral to a class of functions that are not Riemann integrable.

Definition. Let $f : X \rightarrow \mathbb{R}$ be a function. If $D \subset \mathbb{R}$ then we define the *pre-image of D* to be the set $f^{-1}D = \{x \in X \mid f(x) \in D\}$.

A function $f : X \rightarrow \mathbb{R}$ is *measurable* if $f^{-1}D \in \mathcal{B}$ for every Borel subset D of \mathbb{R} . One can show that this is equivalent to requiring that $f^{-1}(-\infty, c) \in \mathcal{B}$ for all $c \in \mathbb{R}$.

A function $f : X \rightarrow \mathbb{C}$ is *measurable* if both the real and imaginary parts, $\operatorname{Re}f$ and $\operatorname{Im}f$, are measurable.

Remark. In writing $f^{-1}D$, we are not assuming that f is a bijection. We are writing $f^{-1}D$ to denote the pre-image of the set D .

We define integration via simple functions.

Definition. A function $f : X \rightarrow \mathbb{R}$ is *simple* if it can be written as a linear combination of characteristic functions of sets in \mathcal{B} , i.e.:

$$f = \sum_{j=1}^r a_j \chi_{B_j},$$

for some $a_j \in \mathbb{R}$, $B_i \in \mathcal{B}$, where the B_j are pairwise disjoint.

Remarks.

- (i) Note that the sets B_j are sets in the σ -algebra \mathcal{B} ; even in the case when $X = [0, 1]$ we do not assume that the sets B_j are intervals.
- (ii) For example, $\chi_{\mathbb{Q} \cap [0, 1]}$ is a simple function. Note, however, that $\chi_{\mathbb{Q} \cap [0, 1]}$ is not Riemann integrable.

For a simple function $f : X \rightarrow \mathbb{R}$ we define

$$\int f \, d\mu = \sum_{j=1}^r a_j \mu(B_j).$$

For example, if μ denotes Lebesgue measure on $[0, 1]$ then

$$\int \chi_{\mathbb{Q} \cap [0, 1]} \, d\mu = \mu(\mathbb{Q} \cap [0, 1]) = 0,$$

as $\mathbb{Q} \cap [0, 1]$ is a countable set and so has Lebesgue measure zero.

A simple function can be written as a linear combination of characteristic functions of pairwise disjoint sets in many different ways (for example, $\chi_{[1/4, 3/4]} = \chi_{[1/4, 1/2]} + \chi_{[1/2, 3/4]}$). However, one can show that the definition of a simple function f given in (3.1.1) is independent of the choice of representation of f as a linear combination of characteristic functions. Thus for a simple function f , the integral of f can be regarded as being the area of the region in $X \times \mathbb{R}$ bounded by the graph of f .

If $f : X \rightarrow \mathbb{R}$, $f \geq 0$, is measurable then one can show that there exists an increasing sequence of simple functions f_n such that $f_n \uparrow f$ pointwise¹ as $n \rightarrow \infty$ and we define

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

This can be shown to exist (although it may be ∞) and to be independent of the choice of sequence f_n .

For an arbitrary measurable function $f : X \rightarrow \mathbb{R}$, we write $f = f^+ - f^-$, where $f^+ = \max\{f, 0\} \geq 0$ and $f^- = \max\{-f, 0\} \geq 0$ and define

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

If $\int f^+ d\mu = \infty$ and $\int f^- d\mu$ is finite then we set $\int f d\mu = \infty$. Similarly, if $\int f^+ d\mu$ is finite but $\int f^- d\mu = \infty$ then we set $\int f d\mu = -\infty$. If both $\int f^+ d\mu$ and $\int f^- d\mu$ are infinite then we leave $\int f d\mu$ undefined.

Finally, for a measurable function $f : X \rightarrow \mathbb{C}$, we define

$$\int f d\mu = \int \operatorname{Re} f d\mu + i \int \operatorname{Im} f d\mu.$$

We say that f is *integrable* if

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

(Note that, in the case of a measurable function $f : X \rightarrow \mathbb{R}$, saying that f is integrable is equivalent to saying that both $\int f^+ d\mu$ and $\int f^- d\mu$ are finite.)

Denote the space of \mathbb{C} -valued integrable functions by $L^1(X, \mathcal{B}, \mu)$. (We shall see a slightly more sophisticated definition of this space below.)

Note that when we write $\int f d\mu$ we are implicitly integrating over the whole space X . We can define integration over subsets of X as follows.

Definition. Let (X, \mathcal{B}, μ) be a probability space. Let $f \in L^1(X, \mathcal{B}, \mu)$ and let $B \in \mathcal{B}$. Then $\chi_B f \in L^1(X, \mathcal{B}, \mu)$. We define

$$\int_B f d\mu = \int \chi_B f d\mu.$$

¹ $f_n \uparrow f$ pointwise means: for every x , $f_n(x)$ is an increasing sequence of real numbers and $\lim_{n \rightarrow \infty} f_n(x) = f(x)$.

§3.1.1 Examples

Lebesgue measure. Let $X = [0, 1]$ and let μ denote Lebesgue measure on the Borel σ -algebra. If $f : [0, 1] \rightarrow \mathbb{R}$ is Riemann integrable then it is also Lebesgue integrable and the two definitions agree. However, there are plenty of examples of functions which are Lebesgue integrable but not Riemann integrable. For example, take $f(x) = \chi_{\mathbb{Q} \cap [0, 1]}(x)$ defined on $[0, 1]$ to be the characteristic function of the rationals. Then $f(x) = 0$ μ -a.e. Hence f is integrable and $\int f d\mu = 0$. However, f is not Riemann integrable.

The Stieltjes integral. Let $\rho : [0, 1] \rightarrow \mathbb{R}^+$ and suppose that ρ is differentiable. Then one can show that

$$\int f d\mu_\rho = \int f(x)\rho'(x) dx.$$

Integration with respect to Dirac measures. Let $x \in X$. Recall that we defined the Dirac measure at x by

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

If χ_B denotes the characteristic function of B then

$$\int \chi_B d\delta_x = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B. \end{cases}$$

Suppose that $f = \sum a_j \chi_{B_j}$ is a simple function and that, without loss of generality, the B_j are pairwise disjoint. Then $\int f d\delta_x = a_j$ where j is chosen so that $x \in B_j$ (and equals zero if no such B_j exists). Now let $f : X \rightarrow \mathbb{R}$. By choosing an increasing sequence of simple functions, we see that

$$\int f d\delta_x = f(x).$$

We say that two measurable functions $f, g : X \rightarrow \mathbb{C}$ are *equivalent* or *equal μ -a.e.* if $f = g$ μ -a.e., i.e. if $\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0$. The following result says that if two functions differ only on a set of measure zero then their integrals are equal.

Lemma 3.1.1

Suppose that $f, g \in L^1(X, \mathcal{B}, \mu)$ and f, g are equal μ -a.e. Then $\int f d\mu = \int g d\mu$.

Functions being equivalent is an equivalence relation. We shall write $L^1(X, \mathcal{B}, \mu)$ for the set of equivalence classes of integrable functions $f : X \rightarrow \mathbb{C}$ on (X, \mathcal{B}, μ) . We define

$$\|f\|_1 = \int |f| d\mu.$$

Then $d(f, g) = \|f - g\|_1$ is a metric on $L^1(X, \mathcal{B}, \mu)$. One can show that $L^1(X, \mathcal{B}, \mu)$ is a vector space; indeed, it is complete in the L^1 metric, and so is a Banach space.

Remark. In practice, we will often abuse notation and regard elements of $L^1(X, \mathcal{B}, \mu)$ as functions rather than equivalence classes of functions. In general, in measure theory one can often ignore sets of measure zero and treat two objects (functions, sets, etc) that differ only on a set of measure zero as ‘the same’.

More generally, for any $p \geq 1$, we can define the space $L^p(X, \mathcal{B}, \mu)$ consisting of (equivalence classes of) measurable functions $f : X \rightarrow \mathbb{C}$ such that $|f|^p$ is integrable. We can again define a metric on $L^p(X, \mathcal{B}, \mu)$ by defining $d(f, g) = \|f - g\|_p$ where

$$\|f\|_p = \left(\int |f|^p d\mu \right)^{1/p}$$

is the L^p norm.

Apart from L^1 , the most interesting L^p space is $L^2(X, \mathcal{B}, \mu)$. This is a Hilbert space² with the inner product

$$\langle f, g \rangle = \int f \bar{g} d\mu.$$

The Cauchy-Schwarz inequality holds: $|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$ for all $f, g \in L^2(X, \mathcal{B}, \mu)$.

Suppose that μ is a finite measure. It follows from the Cauchy-Schwarz inequality that $L^2(X, \mathcal{B}, \mu) \subset L^1(X, \mathcal{B}, \mu)$.

In general, the Riemann integral does not behave well with respect to limits. For example, if f_n is a sequence of Riemann integrable functions such that $f_n(x) \rightarrow f(x)$ at every point x then it does not follow that f is Riemann integrable. Even if f is Riemann integrable, it does not follow that $\int f_n(x) dx \rightarrow \int f(x) dx$. The following convergence theorems hold for the Lebesgue integral.

Theorem 3.1.2 (Monotone Convergence Theorem)

Suppose that $f_n : X \rightarrow \mathbb{R}$ is an increasing sequence of integrable functions on (X, \mathcal{B}, μ) . Suppose that $\int f_n d\mu$ is a bounded sequence of real numbers (i.e. there exists $M > 0$ such that $|\int f_n d\mu| \leq M$ for all n). Then $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ exists μ -a.e. Moreover, f is integrable and

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

Theorem 3.1.3 (Dominated Convergence Theorem)

Suppose that $g : X \rightarrow \mathbb{R}$ is integrable and that $f_n : X \rightarrow \mathbb{R}$ is a sequence of measurable functions with $|f_n(x)| \leq g(x)$ μ -a.e. and $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ μ -a.e. Then f is integrable and

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Remark. Both the Monotone Convergence Theorem and the Dominated Convergence Theorem fail for Riemann integration.

§3.2 Invariant measures

We are now in a position to study dynamical systems. Let (X, \mathcal{B}, μ) be a probability space. Let $T : X \rightarrow X$ be a dynamical system. If $B \in \mathcal{B}$ then we define

$$T^{-1}B = \{x \in X \mid T(x) \in B\},$$

that is, $T^{-1}B$ is the pre-image of B under T .

²An inner product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ on a complex vector space \mathcal{H} is a function such that: (i) $\langle v, v \rangle \geq 0$ for all $v \in \mathcal{H}$ with equality if and only if $v = 0$, (ii) $\langle u, v \rangle = \overline{\langle v, u \rangle}$, and (iii) for each $v \in \mathcal{H}$, $u \mapsto \langle u, v \rangle$ is linear. An inner product determines a norm by setting $\|v\| = (\langle v, v \rangle)^{1/2}$. A norm determines a metric by setting $d(u, v) = \|u - v\|$. We say that \mathcal{H} is a Hilbert space if the vector space \mathcal{H} is complete with respect to the metric induced from the inner product.

Remark. Note that we do not have to assume that T is a bijection for this definition to make sense. For example, let $T(x) = 2x \bmod 1$ be the doubling map on $[0, 1]$. Then T is not a bijection. One can easily check that, for example, $T^{-1}(0, 1/2) = (0, 1/4) \cup (1/2, 3/4)$.

Definition. A transformation $T : X \rightarrow X$ is said to be *measurable* if $T^{-1}B \in \mathcal{B}$ for all $B \in \mathcal{B}$.

Remark. We will often work with compact metric spaces X equipped with the Borel σ -algebra. In this setting, any continuous transformation is measurable.

Remark. Suppose that \mathcal{A} is an algebra of sets that generates the σ -algebra \mathcal{B} . One can show that if $T^{-1}A \in \mathcal{B}$ for all $A \in \mathcal{A}$ then T is measurable.

Definition. We say that T is a *measure-preserving transformation* (m.p.t. for short) or, equivalently, μ is said to be a *T -invariant measure*, if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$.

§3.3 Using the Hahn-Kolmogorov Extension Theorem to prove invariance

Recall the Hahn-Kolmogorov Extension Theorem:

Theorem 3.3.1 (Hahn-Kolmogorov Extension Theorem)

Let \mathcal{A} be an algebra of subsets of X and let $\mathcal{B}(\mathcal{A})$ denote the σ -algebra generated by \mathcal{A} . Suppose that $\mu : \mathcal{A} \rightarrow [0, 1]$ satisfies:

- (i) $\mu(\emptyset) = 0$;
- (ii) if $A_n \in \mathcal{A}$, $n \geq 1$, are pairwise disjoint and if $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Then there is a unique probability measure $\mu : \mathcal{B}(\mathcal{A}) \rightarrow [0, 1]$ which is an extension of $\mu : \mathcal{A} \rightarrow [0, 1]$.

That is, if μ looks like a measure on an algebra \mathcal{A} , then it extends *uniquely* to a measure defined on the σ -algebra $\mathcal{B}(\mathcal{A})$ generated by \mathcal{A} .

Corollary 3.3.2

Let \mathcal{A} be an algebra of subsets of X . Suppose that μ_1 and μ_2 are two measures on $\mathcal{B}(\mathcal{A})$ such that $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{A}$. Then $\mu_1 = \mu_2$ on $\mathcal{B}(\mathcal{A})$.

We shall discuss several examples of dynamical systems and prove that certain naturally occurring measures are invariant using the Hahn-Kolmogorov Extension Theorem.

Suppose that (X, \mathcal{B}, μ) is a probability space and suppose that $T : X \rightarrow X$ is measurable. We define a new measure $T_*\mu$ by

$$T_*\mu(B) = \mu(T^{-1}B) \tag{3.3.1}$$

where $B \in \mathcal{B}$. It is straightforward to check that $T_*\mu$ is a probability measure on (X, \mathcal{B}, μ) (see Exercise 3.4). Thus μ is a T -invariant measure if and only if $T_*\mu = \mu$, i.e. $T_*\mu$ and μ are the same measure. Corollary 3.3.2 says that if two measures agree on an algebra, then they agree on the σ -algebra generated by that algebra. Hence if we can show that $T_*\mu(A) = \mu(A)$ for all sets $A \in \mathcal{A}$ for some algebra \mathcal{A} that generates \mathcal{B} , then $T_*\mu = \mu$, and so μ is a T -invariant measure.

§3.3.1 The doubling map

Let $X = \mathbb{R}/\mathbb{Z}$ be the circle, \mathcal{B} be the Borel σ -algebra, and let μ denote Lebesgue measure. Define the doubling map by $T(x) = 2x \bmod 1$.

Proposition 3.3.3

Let $X = \mathbb{R}/\mathbb{Z}$ be the circle, \mathcal{B} be the Borel σ -algebra, and let μ denote Lebesgue measure. Define the doubling map by $T(x) = 2x \bmod 1$. Then Lebesgue measure μ is T -invariant.

Proof. Let \mathcal{A} denote the algebra of finite unions of intervals. For an interval $[a, b]$ we have that

$$T^{-1}[a, b] = \{x \in \mathbb{R}/\mathbb{Z} \mid T(x) \in [a, b]\} = \left[\frac{a}{2}, \frac{b}{2}\right] \cup \left[\frac{a+1}{2}, \frac{b+1}{2}\right].$$

See Figure 3.3.1.

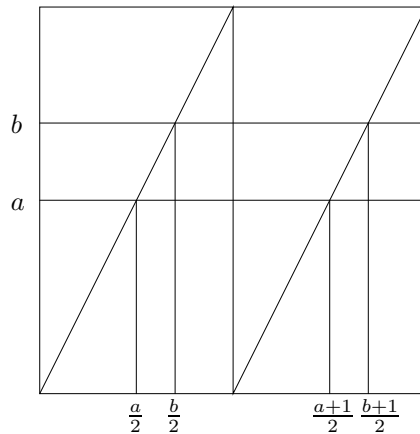


Figure 3.3.1: The pre-image of an interval under the doubling map

Hence

$$\begin{aligned} T_*\mu([a, b]) &= \mu(T^{-1}[a, b]) \\ &= \mu\left(\left[\frac{a}{2}, \frac{b}{2}\right] \cup \left[\frac{a+1}{2}, \frac{b+1}{2}\right]\right) \\ &= \frac{b}{2} - \frac{a}{2} + \frac{(b+1)}{2} - \frac{(a+1)}{2} \\ &= b - a = \mu([a, b]). \end{aligned}$$

Hence $T_*\mu = \mu$ on the algebra \mathcal{A} . As \mathcal{A} generates the Borel σ -algebra, by uniqueness in the Hahn-Kolmogorov Extension Theorem we see that $T_*\mu = \mu$. Hence Lebesgue measure is T -invariant. \square

§3.3.2 Rotations on a circle

Let $X = \mathbb{R}/\mathbb{Z}$ be the circle, let \mathcal{B} be the Borel σ -algebra and let μ be Lebesgue measure. Fix $\alpha \in \mathbb{R}$. Define $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ by $T(x) = x + \alpha \bmod 1$. We call T a *rotation through angle α* .

One can also regard \mathbb{R}/\mathbb{Z} as the unit circle $K = \{z \in \mathbb{C} \mid |z| = 1\}$ in the complex plane via the map $t \mapsto e^{2\pi it}$. In these co-ordinates, the map T becomes $T(e^{2\pi i\theta}) = e^{2\pi i\alpha}e^{2\pi i\theta}$, which is a rotation about the origin through the angle $2\pi\alpha$.

Proposition 3.3.4

Let $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$, $T(x) = x + \alpha \pmod{1}$, be a circle rotation. Then Lebesgue measure is an invariant measure.

Proof. Let $[a, b] \subset \mathbb{R}/\mathbb{Z}$ be an interval. By the Hahn-Kolmogorov Extension Theorem, if we can show that $T_*\mu([a, b]) = \mu([a, b])$ then it follows that $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$, hence μ is T -invariant.

Note that $T^{-1}[a, b] = [a - \alpha, b - \alpha]$ where we interpret the endpoints mod 1. (One needs to be careful here: if $a - \alpha < 0 < b - \alpha$ then $T^{-1}([a, b]) = [0, b - \alpha] \cup [a - \alpha + 1, 1]$, etc.) Hence

$$T_*\mu([a, b]) = \mu([a - \alpha, b - \alpha]) = (b - \alpha) - (a - \alpha) = b - a = \mu([a, b]).$$

Hence $T_*\mu = \mu$ on the algebra \mathcal{A} . As \mathcal{A} generates the Borel σ -algebra, by uniqueness in the Hahn-Kolmogorov Extension Theorem we see that $T_*\mu = \mu$. Hence Lebesgue measure is T -invariant. \square

§3.3.3 The Gauss map

Let $X = [0, 1]$ be the unit interval and let \mathcal{B} be the Borel σ -algebra. Define the Gauss map $T : X \rightarrow X$ by

$$T(x) = \begin{cases} \frac{1}{x} \pmod{1} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

See Figure 3.3.2

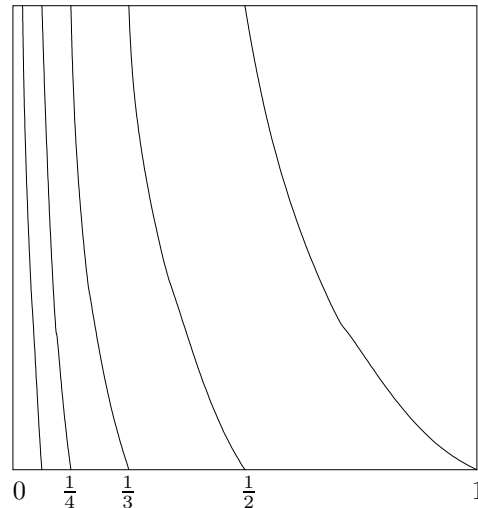


Figure 3.3.2: The graph of the Gauss map (note that there are, in fact, infinitely many branches to the graph, only the first 5 are illustrated)

The Gauss map is very closely related to continued fractions. Recall that if $x \in (0, 1)$

then x has a continued fraction expansion of the form

$$x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \dots}}} \quad (3.3.2)$$

where $x_j \in \mathbb{N}$. If x is rational then this expansion is finite. One can show that x is irrational if and only if it has an infinite continued fraction expansion. Moreover, if x is irrational then it has a unique infinite continued fraction expansion.

If x has continued fraction expansion given by (3.3.2) then

$$\frac{1}{x} = x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \dots}}}$$

Hence, taking the fractional part, we see that $T(x)$ has continued fraction expansion given by

$$T(x) = \frac{1}{x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \dots}}}$$

i.e. T acts by deleting the zeroth term in the continued fraction expansion of x and then shifting the remaining digits one place to the left.

The Gauss map does not preserve Lebesgue measure (see Exercise 3.5). However it does preserve Gauss' measure μ defined by

$$\mu(B) = \frac{1}{\log 2} \int_B \frac{dx}{1+x}$$

(here \log denotes the natural logarithm; the factor $\log 2$ is a normalising constant to make this a probability measure).

Proposition 3.3.5

Gauss' measure is an invariant measure for the Gauss map.

Proof. It is sufficient to check that $\mu([a, b]) = \mu(T^{-1}[a, b])$ for any interval $[a, b]$. First note that

$$T^{-1}[a, b] = \bigcup_{n=1}^{\infty} \left[\frac{1}{b+n}, \frac{1}{a+n} \right].$$

Thus

$$\begin{aligned} \mu(T^{-1}[a, b]) &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{1}{1+x} dx \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left[\log \left(1 + \frac{1}{a+n} \right) - \log \left(1 + \frac{1}{b+n} \right) \right] \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} [\log(a+n+1) - \log(a+n) - \log(b+n+1) + \log(b+n)] \end{aligned}$$

$$\begin{aligned}
&= \lim_{N \rightarrow \infty} \frac{1}{\log 2} \sum_{n=1}^N [\log(a+n+1) - \log(a+n) - \log(b+n+1) + \log(b+n)] \\
&= \frac{1}{\log 2} \lim_{N \rightarrow \infty} [\log(a+N+1) - \log(a+1) - \log(b+N+1) + \log(b+1)] \\
&= \frac{1}{\log 2} \left(\log(b+1) - \log(a+1) + \lim_{N \rightarrow \infty} \log \left(\frac{a+N+1}{b+N+1} \right) \right) \\
&= \frac{1}{\log 2} (\log(b+1) - \log(a+1)) \\
&= \frac{1}{\log 2} \int_a^b \frac{1}{1+x} dx = \mu([a, b]),
\end{aligned}$$

as required. □

§3.3.4 Markov shifts

Let S be a finite set, for example $S = \{1, 2, \dots, k\}$, with $k \geq 2$. Let

$$\Sigma = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \mid x_j \in S\}$$

denote the set of all infinite sequences of symbols chosen from S . Thus a point \mathbf{x} in the phase space Σ is an infinite sequence of symbols $\mathbf{x} = (x_0, x_1, x_2, \dots)$.

Define the *shift map* $\sigma : \Sigma \rightarrow \Sigma$ by

$$\sigma((x_0, x_1, x_2, \dots)) = (x_1, x_2, x_3, \dots)$$

(equivalently, $(\sigma(\mathbf{x}))_j = x_{j+1}$). Thus σ takes a sequence, deletes the zeroth term in this sequence, and then shifts the remaining terms in the sequence one place to the left.

When constructing a measure μ on the Borel σ -algebra \mathcal{B} of $[0, 1]$ we first defined μ on an algebra \mathcal{A} that generates the σ -algebra \mathcal{B} and then extended μ to \mathcal{B} using the Hahn-Kolmogorov Extension Theorem. In this case, our algebra \mathcal{A} was the collection of finite unions of intervals; thus to define μ on \mathcal{A} it was sufficient to define μ on an interval. We want to use a similar procedure to define measures on Σ . To do this, we first need to define a metric on Σ , so that it makes sense to talk about the Borel σ -algebra, and then we need an algebra of subsets that generates the Borel σ -algebra.

Let $\mathbf{x}, \mathbf{y} \in \Sigma$. Suppose that $\mathbf{x} \neq \mathbf{y}$. Define $n(\mathbf{x}, \mathbf{y}) = n$ where $x_n \neq y_n$ but $x_j = y_j$ for $0 \leq j \leq n-1$. Thus $n(\mathbf{x}, \mathbf{y})$ is the index of the first place in which the sequences \mathbf{x}, \mathbf{y} disagree. For convenience, define $n(\mathbf{x}, \mathbf{y}) = \infty$ if $\mathbf{x} = \mathbf{y}$. Define

$$d(\mathbf{x}, \mathbf{y}) = \frac{1}{2^{n(\mathbf{x}, \mathbf{y})}}.$$

Thus two sequences \mathbf{x}, \mathbf{y} are close if they agree for a large number of initial places.

One can show (see Exercise 3.10) that d is a metric on Σ and that the shift map $\sigma : \Sigma \rightarrow \Sigma$ is continuous.

Fix $i_j \in S$, $j = 0, 1, \dots, n-1$. We define the *cylinder set*

$$[i_0, i_1, \dots, i_{n-1}] = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \in \Sigma \mid x_j = i_j, j = 0, 1, \dots, n-1\}.$$

That is, the cylinder set $[i_0, i_1, \dots, i_{n-1}]$ consists of all infinite sequences of symbols from S that begin i_0, i_1, \dots, i_{n-1} . We call n the *rank* of the cylinder. Cylinder sets for shifts often

play the same role that intervals do for maps of the unit interval or circle. Let \mathcal{A} denote the algebra of all finite unions of cylinders. Then \mathcal{A} generates the Borel σ -algebra \mathcal{B} . To see this we use Proposition 2.4.2. It is sufficient to check that \mathcal{A} separates every pair of distinct points in Σ . Let $\mathbf{x} = (x_j)_{j=0}^\infty, \mathbf{y} = (y_j)_{j=0}^\infty \in \Sigma$ and suppose that $\mathbf{x} \neq \mathbf{y}$. Then there exists $n \geq 0$ such that $x_n \neq y_n$. Hence \mathbf{x}, \mathbf{y} are in different cylinders of rank $n + 1$, and the claim follows.

We will construct a family of σ -invariant measures on Σ by first constructing them on cylinders and then extending them to the Borel σ -algebra by using the Hahn-Kolmogorov Extension Theorem. A $k \times k$ matrix P is called a *stochastic matrix* if

- (i) $P(i, j) \in [0, 1]$
- (ii) each row of P sums to 1: for each i , $\sum_{j=1}^k P(i, j) = 1$.

(Here, $P(i, j)$ denotes the (i, j) th entry of the matrix P .)

We say that P is *irreducible* if: for all i, j , there exists $n > 0$ such that $P^n(i, j) > 0$. We say that P is *aperiodic* if there exists $n > 0$ such that every entry of P^n is strictly positive. Thus P is irreducible if for every (i, j) there exists an n such that the (i, j) th entry of P^n is positive, and P is aperiodic if this n can be chosen to be independent of (i, j) .

Suppose that P is irreducible. Let d be the highest common factor of $\{n > 0 \mid P^n(i, i) > 0\}$. One can show that P is aperiodic if and only if $d = 1$. We call d the *period* of P .

In general, if d is the period of an irreducible matrix P then $\{1, 2, \dots, k\}$ can be partitioned into d sets, S_0, S_1, \dots, S_{d-1} , say, such that $P(i, j) > 0$ only if $i \in S_\ell, j \in S_{\ell+1 \bmod d}$. The matrix P^d restricted to the indices that comprise each set S_j is then aperiodic.

The eigenvalues of aperiodic (or, more generally, irreducible) stochastic matrices are extremely well-behaved.

Theorem 3.3.6 (Perron-Frobenius Theorem)

Let P be an irreducible stochastic matrix with period d . Then the following statements hold:

- (i) The d th roots of unity are simple eigenvalues for P and all other eigenvalues have modulus strictly less than 1.
- (ii) Let $\mathbf{1}$ denote the column vector $(1, 1, \dots, 1)^T$. Then $P\mathbf{1} = \mathbf{1}$ so that $\mathbf{1}$ is a right eigenvector corresponding to the maximal eigenvalue 1. Moreover, there exists a corresponding left eigenvector $p = (p(1), \dots, p(k))$ for the eigenvalue 1, that is $pP = p$. The vector p has strictly positive entries $p(j) > 0$, and we can assume that p is normalised so that $\sum_{j=1}^k p(j) = 1$.
- (iii) for all $i, j \in \{1, 2, \dots, k\}$, we have that $P^{nd}(i, j) \rightarrow p(j)$ as $n \rightarrow \infty$.

Proof (not examinable). We prove only the aperiodic case. In this case, the period $d = 1$. We must show that 1 is a simple eigenvalue, construct the positive left eigenvector p , and show that $P^n(i, j) \rightarrow p(j)$ as $n \rightarrow \infty$.

First note that 1 is an eigenvalue of P as $P\mathbf{1} = \mathbf{1}$; this follows from the fact that, for a stochastic matrix, the rows sum to 1.

Suppose P has an eigenvalue λ with corresponding eigenvector v . Then $Pv = \lambda v$. Hence $P^n v = \lambda^n v$. As the entries of P are non-negative we have that

$$|\lambda^n| |v| \leq P^n |v|.$$

Note that if P is stochastic then so is P^n for any $n \geq 1$. As P^n is stochastic, the right-hand side is a bounded sequence in n . Hence P keeps an eigenvector in a bounded region of \mathbb{C}^k . If $|\lambda| > 1$ then $|\lambda^n||v| \rightarrow \infty$, a contradiction if $v \neq 0$. Hence the eigenvalues of P have modulus less than or equal to 1.

Suppose that $Pv = \lambda v$ and $|\lambda| = 1$. Then $P^n|v| \geq |v|$. As P is aperiodic, we can choose n such that $P^n(i, j) > 0$ for all i, j . Hence

$$\sum_{j=1}^k P^n(i, j)|v(j)| \geq |v(i)| \quad (3.3.3)$$

and choose i_0 such that $|v(i_0)| = \max\{|v(j)| \mid 1 \leq j \leq k\}$. Also, as P^n is stochastic and $P^n(i, j) > 0$, we must have that

$$|v(i_0)| \geq \sum_{j=1}^k P^n(i_0, j)|v(j)| \quad (3.3.4)$$

as the right-hand side of (3.3.4) is a convex combination of the $|v(j)|$. Thus $|v(j)| = |v(i_0)|$ for every j , $1 \leq j \leq k$. We can assume, by normalising, that $|v(j)| = 1$ for all $1 \leq j \leq k$.

Now $Pv = \lambda v$, i.e. $\lambda v(i) = \sum_{j=1}^k P(i, j)v(j)$, a convex combination of $v(j)$. As the $|v(j)|$ all have the same modulus, this can only happen if all of the $v(j)$ are the same. Hence v is a multiple of $\mathbf{1}$ and $\lambda = 1$. So 1 is a simple eigenvalue and there are no other eigenvalues of modulus 1.

Since 1 is a simple eigenvalue, there is a unique (up to scalar multiples) left eigenvector p such that $pP = p$. As P is non-negative, we have that $|p|P \geq |p|$, i.e.

$$\sum_{i=1}^k |p(i)|P(i, j) \geq |p(j)| \quad (3.3.5)$$

and summing over j gives

$$\sum_{i=1}^k |p(i)| \geq \sum_{j=1}^k |p(j)|$$

as P is stochastic. Hence we must have equality in (3.3.5), i.e. $|p|P = |p|$. Hence $|p|$ is a left eigenvector for P . Hence p is a scalar multiple of $|p|$, so without loss of generality we can assume that $p(i) \geq 0$ for all i .

To see that $p(i) > 0$, choose n such that all of the entries of P^n are positive. Then $pP^n = p$. Hence $p(j) = \sum_{i=1}^k p(i)P^n(i, j)$. The right-hand side of this expression is a sum of non-negative terms and can only be zero if $p(i) = 0$ for all $1 \leq i \leq k$, i.e. if $p = 0$. Hence all of the entries of p are strictly positive.

We can normalise p and assume that $\sum_{j=1}^k p(j) = 1$.

Decompose \mathbb{R}^k into the sum $V_0 + V_1$ of eigenspaces where

$$V_0 = \{v \mid \langle p, v \rangle = 0\}, \quad V_1 = \text{span}\{\mathbf{1}\}$$

so that V_1 is the eigenspace corresponding to the eigenvalue 1 and V_0 is the sum of the eigenspaces of the remaining eigenvalues. Then $P(V_1) = V_1$ and $P(V_0) \subset V_0$. Note that if $w \in V_0$ then $P^n w \rightarrow 0$ as 1 is not an eigenvalue of P when restricted to V_0 and the eigenvalues of P restricted to V_0 have modulus strictly less than 1.

Let $v \in \mathbb{R}^k$ and write $v = c\mathbf{1} + w$ where $\langle p, w \rangle = 0$. Hence $c = \langle p, v \rangle$. Then

$$P^n v = \langle p, v \rangle \mathbf{1} + P^n w.$$

Hence $P^n v \rightarrow \langle p, v \rangle \mathbf{1}$ as $n \rightarrow \infty$. Taking $v = e_j = (0, \dots, 0, 1, 0, \dots, 0)$, the standard basis vectors, we see that $P^n(i, j) \rightarrow p(j)$. \square

Given an irreducible stochastic matrix P with corresponding normalised left eigenvector p , we define a *Markov measure* μ_P on cylinders by defining

$$\mu_P([i_0, i_1, \dots, i_{n-1}]) = p(i_0)P(i_0, i_1)P(i_1, i_2) \cdots P(i_{n-2}, i_{n-1}).$$

We can then extend μ_P to a probability measure on the Borel σ -algebra of Σ .

Bernoulli measures are particular examples of Markov measures. Let $p = (p(1), \dots, p(k))$, $p(j) \in (0, 1)$, $\sum_{j=1}^k p(j) = 1$ be a probability vector. Define

$$\mu_p([i_0, i_1, \dots, i_{n-1}]) = p(i_0)p(i_1) \cdots p(i_{n-1}).$$

and then extend to the Borel σ -algebra. We call μ_p the *Bernoulli- p measure*.

We can now prove that Markov measures are invariant for shift maps.

Proposition 3.3.7

Let $\sigma : \Sigma \rightarrow \Sigma$ be a shift map on the set of k symbols S . Let P be an irreducible stochastic matrix with left eigenvector p . Then the Markov measure μ_P is a σ -invariant measure.

Proof. It is sufficient to prove that $\mu_P(\sigma^{-1}[i_0, \dots, i_{n-1}]) = \mu_P([i_0, \dots, i_{n-1}])$ for each cylinder $[i_0, \dots, i_{n-1}]$. First note that

$$\begin{aligned} \sigma^{-1}[i_0, \dots, i_{n-1}] &= \{\mathbf{x} \in \Sigma \mid \sigma(\mathbf{x}) \in [i_0, \dots, i_{n-1}]\} \\ &= \{\mathbf{x} \in \Sigma \mid \mathbf{x} = (i, i_0, \dots, i_{n-1}, \dots), i \in S\} \\ &= \bigcup_{i=1}^k [i, i_0, \dots, i_{n-1}]. \end{aligned}$$

Hence

$$\begin{aligned} \mu_P(\sigma^{-1}[i_0, \dots, i_{n-1}]) &= \mu_P\left(\bigcup_{i=1}^k [i, i_0, \dots, i_{n-1}]\right) \\ &= \sum_{i=1}^k \mu_P([i, i_0, \dots, i_{n-1}]) \text{ as this is a disjoint union} \\ &= \sum_{i=1}^k p(i)P(i, i_0)P(i_0, i_1) \cdots P(i_{n-2}, i_{n-1}) \\ &= p(i_0)P(i_0, i_1) \cdots P(i_{n-2}, i_{n-1}) \text{ as } pP = p \\ &= \mu_P([i_0, \dots, i_{n-1}]) \end{aligned}$$

where we have used the fact that $pP = p$. \square

Remark. Bernoulli measures are familiar to you from probability theory. Suppose that $S = \{H, T\}$ so that Σ denotes all infinite sequences of H s and T s. We can think of an element of Σ as the outcome of an infinite sequence of coin tosses. Suppose that $p = (p_H, p_T)$ is a probability vector with corresponding Bernoulli measure μ_p . Then, for example, the cylinder set $[H, H, T]$ denotes the set of (infinite) coin tosses that start H, H, T , and this set has measure $p_H p_H p_T$, corresponding to the probability of tossing H, H, T .

Markov measures are similar. Given a stochastic matrix $P = (P(i, j))$ and a left probability eigenvector $p = (p(1), \dots, p(k))$ we defined

$$\mu_P([i_0, i_1, \dots, i_{n-1}]) = p(i_0)P(i_0, i_1)P(i_1, i_2) \cdots P(i_{n-2}, i_{n-1}).$$

We can regard $p(i_0)$ as being the probability of outcome i_0 . Then we can regard $P(i_0, i_1)$ as being the probability of outcome i_1 , given that the previous outcome was i_0 .

§3.4 Exercises

Exercise 3.1

Show that in Weyl's Criterion (Theorem 1.2.1) one cannot replace the hypothesis in equation (1.2.1) that f is continuous with the hypothesis that $f \in L^1(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu)$ (where μ denotes Lebesgue measure).

Exercise 3.2

Let X be a compact metric space equipped with the Borel σ -algebra \mathcal{B} . Show that a continuous transformation $T : X \rightarrow X$ is measurable.

Exercise 3.3

Give an example of a sequence of functions $f_n \in L^1([0, 1], \mathcal{B}, \mu)$ ($\mu =$ Lebesgue measure) such that $f_n \rightarrow 0$ μ -a.e. but $f_n \not\rightarrow 0$ in L^1 .

Exercise 3.4

Let (X, \mathcal{B}, μ) be a probability space and suppose that $T : X \rightarrow X$ is measurable. Show that $T_*\mu$ is a probability measure on (X, \mathcal{B}, μ) .

Exercise 3.5

- (i) Show that the Gauss map does not preserve Lebesgue measure. (That is, find an example of a Borel set B such that $T^{-1}B$ and B have different Lebesgue measures.)
- (ii) Let μ denote Gauss' measure and let λ denote Lebesgue measure. Show that if $B \in \mathcal{B}$, the Borel σ -algebra of $[0, 1]$, then

$$\frac{1}{2 \log 2} \lambda(B) \leq \mu(B) \leq \frac{1}{\log 2} \lambda(B). \quad (3.4.1)$$

Conclude that a set $B \in \mathcal{B}$ has Lebesgue measure zero if and only if it has Gauss' measure zero. (Two measures with the same sets of measure zero are said to be *equivalent*.)

- (iii) Using (3.4.1), show that $f \in L^1([0, 1], \mathcal{B}, \mu)$ if and only if $f \in L^1([0, 1], \mathcal{B}, \lambda)$.

Exercise 3.6

For an integer $k \geq 2$ define $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ by $T(x) = kx \pmod{1}$. Show that T preserves Lebesgue measure.

Exercise 3.7

Let $\beta > 1$ denote the golden ratio (so that $\beta^2 = \beta + 1$). Define $T : [0, 1] \rightarrow [0, 1]$ by $T(x) = \beta x \bmod 1$. Show that T does not preserve Lebesgue measure. Define the measure μ by $\mu(B) = \int_B k(x) dx$ where

$$k(x) = \begin{cases} \frac{1}{\frac{1}{\beta} + \frac{1}{\beta^3}} & \text{on } [0, 1/\beta) \\ \frac{1}{\beta(\frac{1}{\beta} + \frac{1}{\beta^3})} & \text{on } [1/\beta, 1]. \end{cases}$$

By using the Hahn-Kolmogorov Extension Theorem, show that μ is a T -invariant measure.

Exercise 3.8

Define the logistic map $T : [0, 1] \rightarrow [0, 1]$ by $T(x) = 4x(1 - x)$. Define the measure μ by

$$\mu(B) = \frac{1}{\pi} \int_B \frac{1}{\sqrt{x(1-x)}} dx.$$

- (i) Check that μ is a probability measure.
- (ii) By using the Hahn-Kolmogorov Extension Theorem, show that μ is a T -invariant measure.

Exercise 3.9

Define $T : [0, 1] \rightarrow [0, 1]$ by

$$T(x) = \begin{cases} n(n+1)x - n & \text{if } x \in \left(\frac{1}{n+1}, \frac{1}{n}\right], n \in \mathbb{N} \\ 0 & \text{if } x = 0. \end{cases}$$

This is called the Lüroth map.

Show that

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

Show that T preserves Lebesgue measure.

Exercise 3.10

Let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \mid x_j \in \{1, 2, \dots, k\}\}$ denote the shift space on k symbols. For $\mathbf{x}, \mathbf{y} \in \Sigma$, define $n(\mathbf{x}, \mathbf{y})$ to be the index of the first place in which the two sequences \mathbf{x}, \mathbf{y} disagree, and write $n(\mathbf{x}, \mathbf{y}) = \infty$ if $\mathbf{x} = \mathbf{y}$. Define

$$d(\mathbf{x}, \mathbf{y}) = \frac{1}{2^{n(\mathbf{x}, \mathbf{y})}}.$$

- (i) Show that $d(\mathbf{x}, \mathbf{y})$ is a metric.
 - (ii) Show that the shift map σ is continuous.
 - (iii) Show that a cylinder set $[i_0, \dots, i_{n-1}]$ is both open and closed.
- (One can also prove that Σ is compact; we shall use this fact later.)

Exercise 3.11

Show that the matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

is irreducible but not aperiodic. Show that P has period 2. Show that $\{1, 2, \dots, 5\}$ can be partitioned into two sets $S_0 \cup S_1$ so that $P(i, j) > 0$ only if $i \in S_\ell$ and $j \in S_{\ell+1 \pmod 2}$. Show that P^2 , when restricted to indices in S_0 and in S_1 is aperiodic.

Determine the eigenvalues of P . Find the unique left probability eigenvector p such that $pP = p$.

Exercise 3.12

Show that Bernoulli measures are Markov measures. That is, given a probability vector $p = (p(1), \dots, p(k))$, construct a stochastic matrix P such that $pP = p$. Show that the corresponding Markov measure is the Bernoulli- p measure.

4. More examples of invariant measures

§4.1 Criteria for invariance

We shall give more examples of invariant measures. Recall that, given a measurable transformation $T : X \rightarrow X$ of a probability space (X, \mathcal{B}, μ) , we say that μ is a T -invariant measure (or, equivalently, T is a measure-preserving transformation) if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$.

We will need the following characterisations of invariance.

Lemma 4.1.1

Let $T : X \rightarrow X$ be a measurable transformation of a probability space (X, \mathcal{B}, μ) . Then the following are equivalent:

- (i) T is a measure-preserving transformation;
- (ii) for each $f \in L^1(X, \mathcal{B}, \mu)$ we have

$$\int f \, d\mu = \int f \circ T \, d\mu;$$

- (iii) for each $f \in L^2(X, \mathcal{B}, \mu)$ we have

$$\int f \, d\mu = \int f \circ T \, d\mu.$$

Proof. We will use the identity $\chi_{T^{-1}B} = \chi_B \circ T$; this is straightforward to check, see Exercise 4.1.

We prove that (i) implies (ii). Suppose that T is a measure-preserving transformation. For any characteristic function χ_B , $B \in \mathcal{B}$,

$$\int \chi_B \, d\mu = \mu(B) = \mu(T^{-1}B) = \int \chi_{T^{-1}B} \, d\mu = \int \chi_B \circ T \, d\mu$$

and so the equality holds for any simple function (a finite linear combination of characteristic functions). Given any $f \in L^1(X, \mathcal{B}, \mu)$ with $f \geq 0$, we can find an increasing sequence of simple functions f_n with $f_n \rightarrow f$ pointwise, as $n \rightarrow \infty$. For each n we have

$$\int f_n \, d\mu = \int f_n \circ T \, d\mu$$

and, applying the Monotone Convergence Theorem to both sides, we obtain

$$\int f \, d\mu = \int f \circ T \, d\mu.$$

To extend the result to a general real-valued integrable function f , consider the positive and negative parts. To extend the result to complex-valued integrable functions f , take real and imaginary parts.

That (ii) implies (iii) follows immediately, as $L^2(X, \mathcal{B}, \mu) \subset L^1(X, \mathcal{B}, \mu)$.

Finally, we prove that (iii) implies (i). Let $B \in \mathcal{B}$. Then $\chi_B \in L^2(X, \mathcal{B}, \mu)$ as $\int |\chi_B|^2 d\mu = \int \chi_B d\mu = \mu(B)$. Recalling that $\chi_B \circ T = \chi_{T^{-1}B}$ we have that

$$\mu(B) = \int \chi_B d\mu = \int \chi_B \circ T d\mu = \int \chi_{T^{-1}B} d\mu = \mu(T^{-1}B)$$

so that μ is a T -invariant probability measure. \square

§4.2 Invariant measures on periodic orbits

Recall that if $x \in X$ then we define the Dirac measure δ_x by

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

We also recall that if $f : X \rightarrow \mathbb{R}$ then $\int f d\delta_x = f(x)$.

Let $T : X \rightarrow X$ be a measurable dynamical system defined on a measurable space (X, \mathcal{B}) . Suppose that $x = T^n x$ is a periodic point with period n . Then the probability measure

$$\mu = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x}$$

is T -invariant. This is clear from Lemma 4.1.1, noting that for $f \in L^1(X, \mathcal{B}, \mu)$

$$\begin{aligned} \int f \circ T d\mu &= \frac{1}{n} (f(Tx) + \cdots + f(T^{n-1}x) + f(T^n x)) \\ &= \frac{1}{n} (f(x) + f(Tx) + \cdots + f(T^{n-1}x)) \\ &= \int f d\mu, \end{aligned}$$

using the fact that $T^n x = x$.

§4.3 The change of variables formula

The change of variables formula (equivalently, integration by substitution) for (Riemann) integration should be familiar to you. It can be stated in the following way: if $u : [a, b] \rightarrow [c, d]$ is a differentiable bijection with continuous derivative and $f : [c, d] \rightarrow \mathbb{R}$ is (Riemann) integrable then $f \circ u : [a, b] \rightarrow \mathbb{R}$ is (Riemann) integrable and

$$\int_{u(a)}^{u(b)} f(x) dx = \int_a^b f(u(x))u'(x) dx. \quad (4.3.1)$$

Allowing for the possibility that u is decreasing (so that $u(b) < u(a)$), we can rewrite (4.3.1) as

$$\int_{[c,d]} f(x) dx = \int_{[a,b]} f(u(x))|u'(x)| dx. \quad (4.3.2)$$

We would like a version of (4.3.2) that holds for (Lebesgue) integrable functions on subsets of \mathbb{R}^n , equipped with Lebesgue measure on \mathbb{R}^n .

Theorem 4.3.1 (Change of variables formula)

Let $B \subset \mathbb{R}^n$ be a Borel subset of \mathbb{R}^n and suppose that $B \subset U$ for some open subset U . Suppose that $u : U \rightarrow \mathbb{R}^n$ is a diffeomorphism onto its image (i.e. $u : U \rightarrow u(U)$ is a differentiable bijection with differentiable inverse). Then $u(B)$ is a Borel set.

Let μ denote Lebesgue measure on \mathbb{R}^n and let $f : \mathbb{R}^n \rightarrow \mathbb{C}$ be integrable. Then

$$\int_{u(B)} f \, d\mu = \int_B f \circ u |\det Du| \, d\mu$$

where Du denotes the matrix of partial derivatives of u .

There are more sophisticated versions of the change of variables formula that hold for arbitrary measures on \mathbb{R}^n .

§4.4 Rotations of a circle

We illustrate how one can use the change of variables formula for integration to prove that Lebesgue measure is an invariant measure for certain maps on the circle.

Proposition 4.4.1

Fix $\alpha \in \mathbb{R}$ and define $T(x) = x + \alpha \pmod{1}$. Then Lebesgue measure μ is T -invariant.

Proof. By Lemma 4.1.1(ii) we need to show that $\int f \circ T \, d\mu = \int f \, d\mu$ for every $f \in L^1(X, \mathcal{B}, \mu)$.

Recall that we can identify functions on \mathbb{R}/\mathbb{Z} with 1-periodic functions on \mathbb{R} . By using the substitution $u(x) = x + \alpha$ and the change of variables formula for integration we have that

$$\begin{aligned} \int f \circ T \, d\mu &= \int_0^1 f(Tx) \, dx = \int_0^1 f(x + \alpha) \, dx = \int_\alpha^{1+\alpha} f(x) \, dx \\ &= \int_\alpha^1 f(x) \, dx + \int_1^{1+\alpha} f(x) \, dx = \int_\alpha^1 f(x) \, dx + \int_0^\alpha f(x) \, dx = \int_0^1 f(x) \, dx \end{aligned}$$

where we have used the fact that $\int_1^\alpha f(x) \, dx = \int_1^{1+\alpha} f(x) \, dx$ by the periodicity of f . \square

§4.5 Toral automorphisms

Let $X = \mathbb{R}^k / \mathbb{Z}^k$ be the k -dimensional torus. Let $A = (a(i, j))$ be a $k \times k$ matrix with entries in \mathbb{Z} and with $\det A \neq 0$. We can define a linear map $\mathbb{R}^k \rightarrow \mathbb{R}^k$ by

$$\begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \mapsto A \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

For brevity, we shall often abuse this notation by writing this as $(x_1, \dots, x_k) \mapsto A(x_1, \dots, x_k)$.

Since A is an integer matrix it maps \mathbb{Z}^k to itself. We claim that A allows us to define a map

$$T = T_A : X \rightarrow X : (x_1, \dots, x_k) + \mathbb{Z}^k \mapsto A(x_1, \dots, x_k) + \mathbb{Z}^k.$$

We shall often abuse notation and write $T(x_1, \dots, x_k) = A(x_1, \dots, x_k) \pmod{1}$.

To see that this map is well defined, we need to check that if $x + \mathbb{Z}^k = y + \mathbb{Z}^k$ then $Ax + \mathbb{Z}^k = Ay + \mathbb{Z}^k$. If $x, y \in \mathbb{R}^k$ give the same point in the torus, then $x = y + n$ for some $n \in \mathbb{Z}^k$. Hence $Ax = A(y + n) = Ay + An$. As A maps \mathbb{Z}^k to itself, we see that $An \in \mathbb{Z}^k$ so that Ax, Ay determine the same point in the torus.

Definition. Let $A = (a(i, j))$ denote a $k \times k$ matrix with integer entries such that $\det A \neq 0$. Then we call the map $T_A : \mathbb{R}^k / \mathbb{Z}^k \rightarrow \mathbb{R}^k / \mathbb{Z}^k$ a *linear toral endomorphism*.

The map T is not invertible in general. However, if $\det A = \pm 1$ then A^{-1} exists and is an integer matrix. Hence we have a map T^{-1} given by

$$T^{-1}(x_1, \dots, x_k) = A^{-1}(x_1, \dots, x_k) \bmod 1.$$

One can easily check that T^{-1} is the inverse of T .

Definition. Let $A = (a(i, j))$ denote a $k \times k$ matrix with integer entries such that $\det A = \pm 1$. Then we call the map $T_A : \mathbb{R}^k / \mathbb{Z}^k \rightarrow \mathbb{R}^k / \mathbb{Z}^k$ a *linear toral automorphism*.

Remark. The reason for this nomenclature is clear. If T_A is either a linear toral endomorphism or linear toral automorphism, then it is an endomorphism or automorphism, respectively, of the torus regarded as an additive group.

Example. Take A to be the matrix

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

and define $T : \mathbb{R}^2 / \mathbb{Z}^2 \rightarrow \mathbb{R}^2 / \mathbb{Z}^2$ to be the induced map:

$$T(x_1, x_2) = (2x_1 + x_2 \bmod 1, x_1 + x_2 \bmod 1).$$

Then T is a linear toral automorphism and is called Arnold's CAT map (CAT stands for 'C'ontinuous 'A'utomorphism of the 'T'orus). See Figure 4.5.1.

Definition. Suppose that $\det A = \pm 1$. Then we call T a *hyperbolic* toral automorphism if A has no eigenvalues of modulus 1.

Proposition 4.5.1

Let T be a linear toral automorphism of the k -dimensional torus $X = \mathbb{R}^k / \mathbb{Z}^k$. Then Lebesgue measure μ is T -invariant.

Proof. By Lemma 4.1.1(ii) we need to show that $\int f \circ T d\mu = \int f d\mu$ for every $f \in L^1(X, \mathcal{B}, \mu)$.

Recall that we can identify functions $f : \mathbb{R}^k / \mathbb{Z}^k \rightarrow \mathbb{C}$ with functions $f : \mathbb{R}^k \rightarrow \mathbb{C}$ that satisfy $f(x + n) = f(x)$ for all $n \in \mathbb{Z}^k$. We apply the change of variables formula with the substitution $T(x) = Ax$. Note that $DT(x) = A$ and $|\det DT| = 1$. Hence, by the change of variables formula

$$\int f \circ T d\mu = \int_{\mathbb{R}^k / \mathbb{Z}^k} f \circ T |\det DT| d\mu = \int_{T(\mathbb{R}^k / \mathbb{Z}^k)} f d\mu = \int f d\mu.$$

□

We shall see in §5.4.3 that linear toral endomorphisms (i.e. when A is a $k \times k$ integer matrix with $\det A \neq 0$) also preserves Lebesgue measure.

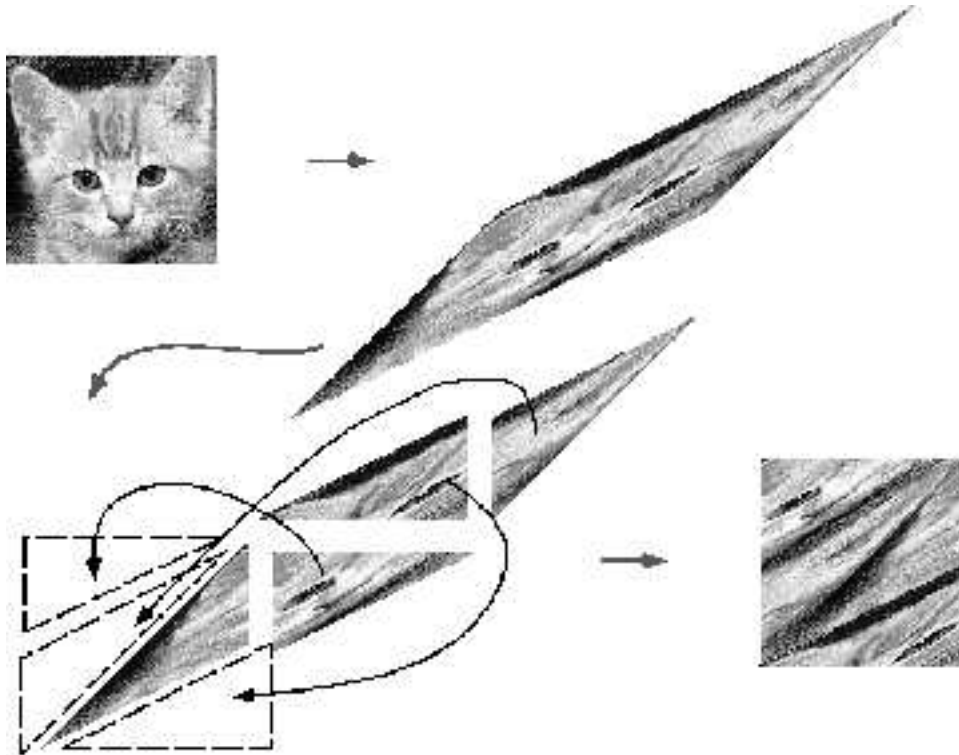


Figure 4.5.1: Arnold's CAT map

§4.6 Exercises

Exercise 4.1

Suppose that $T : X \rightarrow X$. Show that $\chi_{T^{-1}B} = \chi_B \circ T$.

Exercise 4.2

Let $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$, $T(x) = 2x \bmod 1$, denote the doubling map. Show that the periodic points for T are points of the form $p/(2^n - 1)$, $p = 0, 1, \dots, 2^n - 2$. Conclude that T has infinitely many invariant measures.

Exercise 4.3

By using the change of variables formula, prove that the doubling map $T(x) = 2x \bmod 1$ on \mathbb{R}/\mathbb{Z} preserves Lebesgue measure.

Exercise 4.4

Fix $\alpha \in \mathbb{R}$ and define the map $T : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2$ by

$$T((x, y) + \mathbb{Z}^2) = (x + \alpha, x + y) + \mathbb{Z}^2.$$

By using the change of variables formula, prove that Lebesgue measure is T -invariant.

5. Ergodic measures: definition, criteria, and basic examples

§5.1 Introduction

In section 3 we defined what is meant by an invariant measure or, equivalently, what is meant by a measure-preserving transformation. In this section, we define what is meant by an ergodic measure. The primary motivation for ergodicity is Birkhoff's Ergodic Theorem: if T is an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) then, for each $f \in L^1(X, \mathcal{B}, \mu)$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu$$

for μ -a.e. $x \in X$. Checking that a given measure-preserving transformation is ergodic is often a highly non-trivial task and we shall study some methods for proving ergodicity.

§5.2 Ergodicity

We define what it means to say that a measure-preserving transformation is ergodic.

Definition. Let (X, \mathcal{B}, μ) be a probability space and let $T : X \rightarrow X$ be a measure-preserving transformation. We say that T is an *ergodic transformation with respect to μ* (or that μ is an *ergodic measure*) if, whenever $B \in \mathcal{B}$ satisfies $T^{-1}B = B$, then we have that $\mu(B) = 0$ or 1.

Remark. Ergodicity can be viewed as an indecomposability condition. If ergodicity does not hold then we can find a set $B \in \mathcal{B}$ such that $T^{-1}B = B$ and $0 < \mu(B) < 1$. We can then split $T : X \rightarrow X$ into $T : B \rightarrow B$ and $T : X \setminus B \rightarrow X \setminus B$ with invariant probability measures $\frac{1}{\mu(B)}\mu(\cdot \cap B)$ and $\frac{1}{1-\mu(B)}\mu(\cdot \cap (X \setminus B))$, respectively.

It will sometimes be convenient for us to weaken the condition $T^{-1}B = B$ to $\mu(T^{-1}B \Delta B) = 0$, where Δ denotes the symmetric difference:

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

We will often write that $A = B$ μ -a.e. or $A = B \bmod 0$ to mean that $\mu(A \Delta B) = 0$.

Remark. It is easy to see that if $A = B$ μ -a.e. then $\mu(A) = \mu(B)$.

Lemma 5.2.1

Let T be a measure-preserving transformation of the probability space (X, \mathcal{B}, μ) .

Suppose that $B \in \mathcal{B}$ is such that $\mu(T^{-1}B \Delta B) = 0$. Then there exists $B' \in \mathcal{B}$ with $T^{-1}B' = B'$ and $\mu(B \Delta B') = 0$. (In particular, $\mu(B) = \mu(B')$.)

Proof (not examinable). For each $n \geq 0$, we have the inclusion

$$T^{-n}B\Delta B \subset \bigcup_{j=0}^{n-1} \left(T^{-(j+1)}B\Delta T^{-j}B \right) = \bigcup_{j=0}^{n-1} T^{-j}(T^{-1}B\Delta B).$$

Hence, as T preserves μ ,

$$\mu(T^{-n}B\Delta B) \leq n\mu(T^{-1}B\Delta B) = 0.$$

Let

$$B' = \bigcap_{n=0}^{\infty} \bigcup_{j=n}^{\infty} T^{-j}B.$$

We have that

$$\mu \left(B\Delta \bigcup_{j=n}^{\infty} T^{-j}B \right) \leq \sum_{j=n}^{\infty} \mu(B\Delta T^{-n}B) = 0.$$

Since the sets $\bigcup_{j=n}^{\infty} T^{-j}B$ decrease as n increases we have that $\mu(B\Delta B') = 0$. Also,

$$T^{-1}B' = \bigcap_{n=0}^{\infty} \bigcup_{j=n}^{\infty} T^{-(j+1)}B = \bigcap_{n=0}^{\infty} \bigcup_{j=n+1}^{\infty} T^{-j}B = B',$$

as required. □

Corollary 5.2.2

If T is ergodic and $\mu(T^{-1}B\Delta B) = 0$ then $\mu(B) = 0$ or 1 .

We have the following convenient characterisations of ergodicity.

Proposition 5.2.3

Let T be a measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . The following are equivalent:

- (i) T is ergodic;
- (ii) whenever $f \in L^1(X, \mathcal{B}, \mu)$ satisfies $f \circ T = f$ μ -a.e. we have that f is constant μ -a.e.
- (iii) whenever $f \in L^2(X, \mathcal{B}, \mu)$ satisfies $f \circ T = f$ μ -a.e. we have that f is constant μ -a.e.

Remark. If f is a constant function then clearly $f \circ T = f$. Proposition 5.2.3 says that, when T is ergodic, the constants are the *only* T -invariant functions (up to sets of measure zero).

Proof. We prove that (i) implies (ii). Suppose that T is ergodic. Suppose that $f \in L^1(X, \mathcal{B}, \mu)$ is such that $f \circ T = f$ μ -a.e. By taking real and imaginary parts, we can assume without loss of generality that f is real-valued. For $k \in \mathbb{Z}$ and $n \in \mathbb{N}$, define

$$X(k, n) = \left\{ x \in X \mid \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\} = f^{-1} \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right).$$

Since f is measurable, we have that $X(k, n) \in \mathcal{B}$.

We have that

$$T^{-1}X(k, n) \Delta X(k, n) \subset \{x \in X \mid f(Tx) \neq f(x)\}$$

so that

$$\mu(T^{-1}X(k, n) \Delta X(k, n)) = 0.$$

Hence, as T is ergodic, we have by Corollary 5.2.2 that $\mu(X(k, n)) = 0$ or $\mu(X(k, n)) = 1$.

As $f \in L^1(X, \mathcal{B}, \mu)$ is integrable, we have that f is finite almost everywhere. Hence, for each n ,

$$f^{-1}\mathbb{R} = f^{-1}\left(\bigcup_{k=-\infty}^{\infty} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right) = \bigcup_{k=-\infty}^{\infty} f^{-1}\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) = \bigcup_{k=-\infty}^{\infty} X(k, n)$$

is equal to X up to a set of measure zero, i.e.,

$$\mu\left(X \Delta \bigcup_{k \in \mathbb{Z}} X(k, n)\right) = 0;$$

moreover, this union is disjoint. Hence we have

$$\sum_{k \in \mathbb{Z}} \mu(X(k, n)) = \mu(X) = 1$$

and so there is a unique k_n for which $\mu(X(k_n, n)) = 1$. Let

$$Y = \bigcap_{n=1}^{\infty} X(k_n, n).$$

Then $\mu(Y) = 1$. Let $x, y \in Y$. Then for each n we have that $f(x), f(y) \in [k_n/2^n, (k_n + 1)/2^n)$. Hence for all $n \geq 1$ we have that

$$|f(x) - f(y)| \leq \frac{1}{2^n}.$$

Hence $f(x) = f(y)$. Hence f is constant on the set Y . Hence f is constant μ -a.e.

That (ii) implies (iii) is clear as if $f \in L^2(X, \mathcal{B}, \mu)$ then $f \in L^1(X, \mathcal{B}, \mu)$.

Finally, we prove that (iii) implies (i). Suppose that $B \in \mathcal{B}$ is such that $T^{-1}B = B$. Then $\chi_B \in L^2(X, \mathcal{B}, \mu)$ and $\chi_B \circ T(x) = \chi_B(x)$ for all $x \in X$. Hence χ_B is constant μ -a.e. Since χ_B only takes the values 0 and 1, we must have $\chi_B = 0$ μ -a.e. or $\chi_B = 1$ μ -a.e. Therefore

$$\mu(B) = \int_X \chi_B d\mu = \begin{cases} 0 & \text{if } \chi_B = 0 \text{ } \mu\text{-a.e.} \\ 1 & \text{if } \chi_B = 1 \text{ } \mu\text{-a.e.} \end{cases}$$

Hence T is ergodic with respect to μ . □

§5.3 Fourier series

We shall give a method for proving that certain transformations of the circle or torus are ergodic with respect to Lebesgue measure. To do this, we use Proposition 5.2.3 and Fourier series.

Let $X = \mathbb{R}/\mathbb{Z}$ denote the unit circle and let $f : X \rightarrow \mathbb{R}$. (Alternatively, we can think of f as a periodic function $\mathbb{R} \rightarrow \mathbb{R}$ with $f(x) = f(x + n)$ for all $n \in \mathbb{Z}$.) Equip X with the Borel σ -algebra, let μ denote Lebesgue measure and assume that $f \in L^2(X, \mathcal{B}, \mu)$.

We can associate to f its Fourier series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos 2\pi nx + b_n \sin 2\pi nx), \tag{5.3.1}$$

where

$$a_n = 2 \int_0^1 f(x) \cos 2\pi nx \, d\mu, \quad b_n = 2 \int_0^1 f(x) \sin 2\pi nx \, d\mu.$$

(Notice that we are not claiming that the series converges—we are just formally associating the Fourier series to f .)

We shall find it more convenient to work with a complex form of the Fourier series and rewrite (5.3.1) as

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}, \tag{5.3.2}$$

where

$$c_n = \int_0^1 f(x) e^{-2\pi i n x} \, d\mu.$$

(In particular, $c_0 = \int_0^1 f \, d\mu$.) We call c_n the n th Fourier coefficient.

Remark. That (5.3.2) and (5.3.1) are equivalent follows from the fact that

$$\cos 2\pi nx = \frac{e^{2\pi i n x} + e^{-2\pi i n x}}{2}, \quad \sin 2\pi nx = \frac{e^{2\pi i n x} - e^{-2\pi i n x}}{2i}$$

One can explain Fourier series by considering a more general construction. Recall that an inner product on a complex vector space \mathcal{H} is a function $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ such that

- (i) $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in \mathcal{H}$,
- (ii) for each $v \in \mathcal{H}$, $u \mapsto \langle u, v \rangle$ is linear,
- (iii) $\langle v, v \rangle \geq 0$ for all $v \in \mathcal{H}$, with equality if and only if $v = 0$.

Given an inner product, one can define a norm on \mathcal{H} by setting $\|v\| = \sqrt{\langle v, v \rangle}$. One can then define a metric on \mathcal{H} by setting $d_{\mathcal{H}}(u, v) = \|u - v\|$.

If \mathcal{H} is a complex vector space with an inner product $\langle \cdot, \cdot \rangle$ such that \mathcal{H} is complete with respect to the metric given by the inner product then we call \mathcal{H} a *Hilbert space*.

Recall that $L^2(X, \mathcal{B}, \mu)$ is a Hilbert space with the inner product

$$\langle f, g \rangle = \int f \bar{g} \, d\mu.$$

The metric on $L^2(X, \mathcal{B}, \mu)$ is then given by

$$d(f, g) = \left(\int |f - g|^2 \, d\mu \right)^{1/2}.$$

Let \mathcal{H} be an infinite dimensional Hilbert space. We say that $\{e_j\}_{j=0}^{\infty}$ is an *orthonormal basis* for \mathcal{H} if:

$$(i) \langle e_i, e_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

(ii) every $v \in \mathcal{H}$ can be written in the form

$$v = \sum_{j=0}^{\infty} c_j e_j. \quad (5.3.3)$$

As (5.3.3) involves an infinite sum, we need to be careful about what convergence means. To make (5.3.3) precise, let $s_n = \sum_{j=0}^n c_j e_j$ denote the n th partial sum. Then (5.3.3) means that $\|v - s_n\| \rightarrow 0$ as $n \rightarrow \infty$.

As the vectors $\{e_j\}_{j=0}^{\infty}$ are orthonormal, taking the inner product of (5.3.3) with e_i shows that

$$\langle v, e_i \rangle = \left\langle \sum_{j=0}^{\infty} c_j e_j, e_i \right\rangle = \sum_{j=0}^{\infty} c_j \langle e_j, e_i \rangle = c_i.$$

Let $X = \mathbb{R}/\mathbb{Z}$ be the circle and let \mathcal{B} be the Borel σ -algebra. Let μ denote Lebesgue measure. Let $x \in \mathbb{R}/\mathbb{Z}$. Let $e_n(x) = e^{2\pi i n x}$. Then $\{e_n\}_{n=-\infty}^{\infty}$ is an orthonormal basis for the Hilbert space $L^2(X, \mathcal{B}, \mu)$. Thus if $f \in L^2(X, \mathcal{B}, \mu)$ then we can write

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}$$

(in the sense that the sequence of partial sums L^2 -converges to f) where

$$c_n = \langle f, e_n \rangle = \int f(x) e^{-2\pi i n x} d\mu. \quad (5.3.4)$$

If we want to make the dependence of c_n on f clear, then we will sometimes write $c_n(f)$ for c_n .

We shall need the following facts about Fourier coefficients.

Proposition 5.3.1

(i) Let $f, g \in L^2(X, \mathcal{B}, \mu)$. Then $f = g$ μ -a.e. if and only if their Fourier coefficients are equal, i.e. $c_n(f) = c_n(g)$ for all $n \in \mathbb{Z}$.

(ii) Let $f \in L^2(X, \mathcal{B}, \mu)$. Then $c_n \rightarrow 0$ as $n \rightarrow \pm\infty$.

Remark. Proposition 5.3.1(ii) is better known as the Riemann-Lebesgue Lemma.

So far, we have studied Fourier series for functions defined on the circle; a similar construction works for functions defined on the k -dimensional torus. Let $X = \mathbb{R}^k/\mathbb{Z}^k$ be the k -dimensional torus equipped with the Borel σ -algebra and let μ denote Lebesgue measure on X . Then $L^2(X, \mathcal{B}, \mu)$ is a Hilbert space when equipped with the inner product

$$\langle f, g \rangle = \int f \bar{g} d\mu.$$

Let $x \in \mathbb{R}^k/\mathbb{Z}^k$. Let $\mathbf{n} = (n_1, \dots, n_k) \in \mathbb{Z}^k$ and define $e_{\mathbf{n}}(x) = e^{2\pi i \langle \mathbf{n}, x \rangle}$ where $\langle \mathbf{n}, x \rangle = n_1 x_1 + \dots + n_k x_k$. Then $\{e_{\mathbf{n}}\}_{\mathbf{n} \in \mathbb{Z}^k}$ is an orthonormal basis for $L^2(X, \mathcal{B}, \mu)$. Thus we can write $f \in L^2(X, \mathcal{B}, \mu)$ as

$$f(x) = \sum_{\mathbf{n} \in \mathbb{Z}^k} c_{\mathbf{n}} e^{2\pi i \langle \mathbf{n}, x \rangle}$$

in the sense that the sequence of partial sums s_N converges in $L^2(X, \mathcal{B}, \mu)$ where

$$s_N(x) = \sum_{\mathbf{n}=(n_1, \dots, n_k) \in \mathbb{Z}^k, |n_j| \leq N} c_{\mathbf{n}} e^{2\pi i \langle \mathbf{n}, x \rangle}.$$

The \mathbf{n} th Fourier coefficient is given by

$$c_{\mathbf{n}} = c_{\mathbf{n}}(f) = \int f(x) e^{-2\pi i \langle \mathbf{n}, x \rangle} d\mu.$$

We have the following analogue of Proposition 5.3.1:

Proposition 5.3.2

- (i) Let $f, g \in L^2(X, \mathcal{B}, \mu)$. Then $f = g$ μ -a.e. if and only if their Fourier coefficients are equal.
- (ii) Let $f \in L^2(X, \mathcal{B}, \mu)$. Let $\mathbf{n} = (n_1, \dots, n_k) \in \mathbb{Z}^k$ and define $\|\mathbf{n}\| = \max_{1 \leq j \leq k} |n_j|$. Then $c_{\mathbf{n}} \rightarrow 0$ as $\|\mathbf{n}\| \rightarrow \infty$.

Remark. We could have used any norm on \mathbb{Z}^k in (ii).

§5.4 Proving ergodicity using Fourier series

In the previous section we studied a number of examples of dynamical systems defined on the circle or the torus and we proved that Lebesgue measure is invariant. We show how Proposition 5.2.3 can be used in conjunction with Fourier series to determine whether Lebesgue measure is ergodic.

Recall that if $f \in L^2(X, \mathcal{B}, \mu)$ then we associate to f the Fourier series

$$\sum_{n=-\infty}^{\infty} c_n(f) e^{2\pi i n x}$$

where

$$c_n(f) = \int f(x) e^{-2\pi i n x} d\mu.$$

If we let $s_n(x) = \sum_{\ell=-n}^n c_{\ell}(f) e^{2\pi i \ell x}$ then $\|s_n - f\|_2 \rightarrow 0$ as $n \rightarrow \infty$.

If T is a measure-preserving transformation then it follows that

$$\begin{aligned} \|s_n \circ T - f \circ T\|_2 &= \left(\int |s_n \circ T - f \circ T|^2 d\mu \right)^{1/2} = \left(\int (|s_n - f|^2) \circ T d\mu \right)^{1/2} \\ &= \left(\int (|s_n - f|^2) d\mu \right)^{1/2} = \|s_n - f\|_2 \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$, where we have used Lemma 4.1.1. By Proposition 5.3.2(i) it follows that, if $\lim_{n \rightarrow \infty} s_n \circ T$ is a possibly infinite sum of terms of the form $e^{2\pi i n x}$, then it must be the Fourier series of $f \circ T$. In practice, this means that if we take the Fourier series for $f(x)$ and evaluate it at Tx , then we obtain the Fourier series for $f(Tx)$. If $f \circ T = f$ almost everywhere, then we can use Proposition 5.3.1(i) to compare Fourier coefficients to obtain relationships between the Fourier coefficients, and then show that f must be constant.

A similar method works for Fourier series on the torus, as we shall see.

§5.4.1 Rotations on a circle

Fix $\alpha \in \mathbb{R}$ and define $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ by $T(x) = x + \alpha \pmod{1}$. We have already seen that T preserves Lebesgue measure. The following result gives a necessary and sufficient condition for T to be ergodic.

Theorem 5.4.1

Let $T(x) = x + \alpha \pmod{1}$.

(i) If $\alpha \in \mathbb{Q}$ then T is not ergodic with respect to Lebesgue measure.

(ii) If $\alpha \notin \mathbb{Q}$ then T is ergodic with respect to Lebesgue measure.

Proof. Suppose that $\alpha \in \mathbb{Q}$ and write $\alpha = p/q$ for $p, q \in \mathbb{Z}$ with $q \neq 0$. Define

$$f(x) = e^{2\pi i q x} \in L^2(X, \mathcal{B}, \mu).$$

Then f is not constant but

$$f(Tx) = e^{2\pi i q(x+p/q)} = e^{2\pi i(qx+p)} = e^{2\pi i q x} = f(x).$$

Hence T is not ergodic.

Suppose that $\alpha \notin \mathbb{Q}$. Suppose that $f \in L^2(X, \mathcal{B}, \mu)$ is such that $f \circ T = f$ a.e. We want to prove that f is constant. Suppose that f has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}.$$

Then $f \circ T$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n \alpha} e^{2\pi i n x}.$$

Comparing Fourier coefficients we see that

$$c_n = c_n e^{2\pi i n \alpha},$$

for all $n \in \mathbb{Z}$. As $\alpha \notin \mathbb{Q}$, we see that $e^{2\pi i n \alpha} \neq 1$ unless $n = 0$. Hence $c_n = 0$ for $n \neq 0$. Hence f has Fourier series c_0 , i.e. f is constant a.e. \square

§5.4.2 The doubling map

Let $X = \mathbb{R}/\mathbb{Z}$. Recall that if $f \in L^2(X, \mathcal{B}, \mu)$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}$$

then the Riemann-Lebesgue Lemma (Proposition 5.3.1(ii)) tells us that $c_n \rightarrow 0$ as $n \rightarrow \infty$.

Proposition 5.4.2

The doubling map $T : X \rightarrow X$ defined by $T(x) = 2x \pmod{1}$ is ergodic with respect to Lebesgue measure μ .

Proof. Let $f \in L^2(X, \mathcal{B}, \mu)$ and suppose that $f \circ T = f$ μ -a.e. Let f have Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}.$$

For each $p \geq 0$, $f \circ T^p$ has Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n 2^p x}.$$

Comparing Fourier coefficients we see that

$$c_n = c_{2^p n}$$

for all $n \in \mathbb{Z}$ and each $p = 0, 1, 2, \dots$. Suppose that $n \neq 0$. Then $|2^p n| \rightarrow \infty$ as $p \rightarrow \infty$. By the Riemann-Lebesgue Lemma (Proposition 5.3.1(ii)), $c_{2^p n} \rightarrow 0$ as $p \rightarrow \infty$. As $c_{2^p n} = c_n$, we must have that $c_n = 0$ for $n \neq 0$. Thus f has Fourier series c_0 , and so must be equal to a constant a.e. Hence T is ergodic with respect to μ . \square

§5.4.3 Toral endomorphisms

The argument for the doubling map can be generalised using higher-dimensional Fourier series to study toral endomorphisms. Let $X = \mathbb{R}^k / \mathbb{Z}^k$ and let μ denote Lebesgue measure. When T is invertible (and so a linear toral automorphism) we have already seen that Lebesgue measure is an invariant measure; in §7 we shall see that Lebesgue measure is an invariant measure when T is a linear toral endomorphism.

Recall that $f \in L^2(X, \mathcal{B}, \mu)$ has Fourier series

$$\sum_{\mathbf{n} \in \mathbb{Z}^k} c_{\mathbf{n}} e^{2\pi i \langle \mathbf{n}, x \rangle},$$

where $\mathbf{n} = (n_1, \dots, n_k)$, $x = (x_1, \dots, x_k)$. Define $|\mathbf{n}| = \max_{1 \leq j \leq k} |n_j|$. Then the Riemann-Lebesgue Lemma tells us that $c_{\mathbf{n}} \rightarrow 0$ as $|\mathbf{n}| \rightarrow \infty$.

Let A be a $k \times k$ integer matrix with $\det A \neq 0$ and define $T : X \rightarrow X$ by

$$T((x_1, \dots, x_k) + \mathbb{Z}^k) = A(x_1, \dots, x_k) + \mathbb{Z}^k.$$

Proposition 5.4.3

A linear toral endomorphism T is ergodic with respect to μ if and only if no eigenvalue of A is a root of unity.

Remark. In particular, hyperbolic toral automorphisms (i.e. $\det A = \pm 1$ and A has no eigenvalues of modulus 1) are ergodic with respect to Lebesgue measure.

Proof. Suppose that T is ergodic but, for a contradiction, that A has a p th root of unity as an eigenvalue. We choose $p > 0$ to be the least such integer. Then A^p has 1 as an eigenvalue, and so $\mathbf{n}(A^p - I) = 0$ for some non-zero vector $\mathbf{n} = (n_1, \dots, n_k) \in \mathbb{R}^k$. Since A is an integer matrix, we have that $A^p - I$ is an integer matrix, and so we can in fact take $\mathbf{n} \in \mathbb{Z}^k$. Note that

$$e^{2\pi i \langle \mathbf{n}, A^p x \rangle} = e^{2\pi i \langle \mathbf{n} A^p, x \rangle} = e^{2\pi i \langle \mathbf{n}, x \rangle}.$$

This is because, writing $x = (x_1, \dots, x_k)^T$,

$$\langle \mathbf{n}, Ax \rangle = (n_1, \dots, n_k) \begin{pmatrix} a_{1,1} & \cdots & a_{1,k} \\ \vdots & & \vdots \\ a_{k,1} & \cdots & a_{k,k} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} = \langle \mathbf{n}A, x \rangle.$$

Define

$$f(x) = \sum_{j=0}^{p-1} e^{2\pi i \langle \mathbf{n}, A^j x \rangle}.$$

Then $f \in L^2(X, \mathcal{B}, \mu)$ and is T -invariant. Since T is ergodic, we must have that f is constant. But the only way in which this can happen is if $\mathbf{n} = \mathbf{0}$, a contradiction.

Conversely suppose that no eigenvalue of A is a root of unity; we prove that T is ergodic with respect to Lebesgue measure. Suppose that $f \in L^2(X, \mathcal{B}, \mu)$ is T -invariant μ -a.e. We show that f is constant μ -a.e. Associate to f its Fourier series:

$$\sum_{\mathbf{n} \in \mathbb{Z}^k} c_{\mathbf{n}} e^{2\pi i \langle \mathbf{n}, x \rangle}.$$

Since $fT^p = f$ μ -a.e., for all $p > 0$, we have that

$$\sum_{\mathbf{n} \in \mathbb{Z}^k} c_{\mathbf{n}} e^{2\pi i \langle \mathbf{n}A^p, x \rangle} = \sum_{\mathbf{n} \in \mathbb{Z}^k} c_{\mathbf{n}} e^{2\pi i \langle \mathbf{n}, x \rangle}.$$

Comparing Fourier coefficients we see that, for every $\mathbf{n} \in \mathbb{Z}^k$,

$$c_{\mathbf{n}} = c_{\mathbf{n}A} = \cdots = c_{\mathbf{n}A^p} = \cdots.$$

If $c_{\mathbf{n}} \neq 0$ then there can only be finitely many indices in the above list, for otherwise it would contradict the fact that $c_{\mathbf{n}} \rightarrow 0$ as $|\mathbf{n}| \rightarrow \infty$, by the Riemann-Lebesgue Lemma (Proposition 5.3.1(ii)). Hence there exist $q_1 > q_2 \geq 0$ such that $\mathbf{n}A^{q_1} = \mathbf{n}A^{q_2}$. Letting $p = q_1 - q_2 > 0$ we see that $\mathbf{n}A^p = \mathbf{n}$. Thus \mathbf{n} is either equal to $\mathbf{0}$ or \mathbf{n} is an eigenvector for A^p with eigenvalue 1. In the latter case, A would have a p th root of unity as an eigenvalue. Hence $\mathbf{n} = \mathbf{0}$. Hence $c_{\mathbf{n}} = 0$ unless $\mathbf{n} = \mathbf{0}$ and so f is equal to the constant c_0 μ -a.e. Thus T is ergodic. \square

§5.5 Exercises

Exercise 5.1

Suppose that $\alpha \in \mathbb{Q}$. Show directly from the definition that the rotation $T(x) = x + \alpha \pmod{1}$ is not ergodic, i.e. find an invariant set $B = T^{-1}B$, $B \in \mathcal{B}$, which has Lebesgue measure $0 < \mu(B) < 1$.

Exercise 5.2

Define $T : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2$ by

$$T((x, y) + \mathbb{Z}^2) = (x + \alpha, x + y) + \mathbb{Z}^2.$$

Suppose that $\alpha \notin \mathbb{Q}$. By using Fourier series, show that T is ergodic with respect to Lebesgue measure.

Exercise 5.3

Let $T : X \rightarrow X$ be a measurable transformation of a measurable space (X, \mathcal{B}) . Suppose that $x = T^n x$ is a periodic point with period n . Define the measure μ supported on the periodic orbit of x by

$$\mu = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x}$$

where δ_x denotes the Dirac measure at x . Show from the definition of ergodicity that μ is an ergodic measure.

Exercise 5.4

(Part (iv) of this exercise is outside the scope of the course!)

It is easy to construct lots of examples of hyperbolic toral automorphisms (i.e. no eigenvalues of modulus 1—the CAT map is such an example), which must necessarily be ergodic with respect to Lebesgue measure. It is harder to show that there are ergodic toral automorphisms with some eigenvalues of modulus 1.

- (i) Show that to have an ergodic toral automorphism of $\mathbb{R}^k/\mathbb{Z}^k$ with an eigenvalue of modulus 1, we must have $k \geq 4$.

Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 8 & -6 & 8 \end{pmatrix}.$$

- (ii) Show that A defines a linear toral automorphism T_A of the 4-dimensional torus $\mathbb{R}^4/\mathbb{Z}^4$.
- (iii) Show that A has four eigenvalues, two of which have modulus 1.
- (iv) Show that T_A is ergodic with respect to Lebesgue measure. (Hint: you have to show that the two eigenvalues of modulus 1 are not roots of unity, i.e. are not solutions to $\lambda^n - 1 = 0$ for some n . The best way to do this is to use results from Galois theory on the irreducibility of polynomials.)

6. Ergodic measures: Using the Hahn-Kolmogorov Extension Theorem to prove ergodicity

§6.1 Introduction

We illustrate a method for proving that a given transformation is ergodic using the Hahn-Kolmogorov Extension Theorem. The key observation is the following technical lemma.

Lemma 6.1.1

Let (X, \mathcal{B}, μ) be a probability space and suppose that $\mathcal{A} \subset \mathcal{B}$ is an algebra that generates \mathcal{B} . Let $B \in \mathcal{B}$. Suppose there exists $K > 0$ such that

$$\mu(B)\mu(I) \leq K\mu(B \cap I) \tag{6.1.1}$$

for all $I \in \mathcal{A}$. Then $\mu(B) = 0$ or 1 .

Proof. Let $\varepsilon > 0$. As \mathcal{A} generates \mathcal{B} there exists $I \in \mathcal{A}$ such that $\mu(B^c \Delta I) < \varepsilon$. Hence $|\mu(B^c) - \mu(I)| < \varepsilon$. Moreover, note that $B \cap I \subset B^c \Delta I$ so that $\mu(B \cap I) < \varepsilon$. Hence

$$\mu(B)\mu(B^c) \leq \mu(B)(\mu(I) + \varepsilon) \leq \mu(B)\mu(I) + \mu(B)\varepsilon \leq K\mu(B \cap I) + \varepsilon \leq (K + 1)\varepsilon.$$

As $\varepsilon > 0$ is arbitrary, it follows that $\mu(B)\mu(B^c) = 0$. Hence $\mu(B) = 0$ or 1 . □

Remark. We will often apply Lemma 6.1.1 when \mathcal{A} is an algebra of finite unions of intervals or cylinders. In this case, we need only check that there exists a constant $K > 0$ such that (6.1.1) holds for intervals or cylinders. To see this, let $I = \bigcup_{j=1}^k I_j$ be a finite union of pairwise disjoint sets in \mathcal{A} . Then if (6.1.1) holds for I_j then

$$\begin{aligned} \mu(B)\mu(I) &= \mu(B)\mu\left(\bigcup_{j=1}^k I_j\right) = \sum_{j=1}^k \mu(B)\mu(I_j) \\ &\leq K \sum_{j=1}^k \mu(B \cap I_j) = K\mu\left(B \cap \bigcup_{j=1}^k I_j\right) = K\mu(B \cap I). \end{aligned}$$

We will also use the change of variables formula for integration. Recall that if $I, J \subset \mathbb{R}$ are intervals, $u : I \rightarrow J$ is a differentiable bijection, and $f : J \rightarrow \mathbb{R}$ is integrable, then

$$\int_J f(x) dx = \int_I f(u(x))|u'(x)| dx.$$

§6.2 The doubling map

To illustrate the method, we give another proof that the doubling map is ergodic with respect to Lebesgue measure. Let $X = [0, 1]$ be the unit interval, let \mathcal{B} be the Borel σ -algebra, and let μ be Lebesgue measure.

Given $x \in [0, 1]$, we can write x as a base 2 ‘decimal’ expansion:

$$x = \cdot x_0 x_1 x_2 \dots = \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}} \quad (6.2.1)$$

where $x_j \in \{0, 1\}$. Note that

$$T(x) = 2 \sum_{j=0}^{\infty} \frac{x_j}{2^{j+1}} \bmod 1 = x_0 + \sum_{j=0}^{\infty} \frac{x_{j+1}}{2^{j+1}} \bmod 1 = \sum_{j=0}^{\infty} \frac{x_{j+1}}{2^{j+1}}.$$

Hence if x has base 2 expansion given by (6.2.1) then $T(x)$ has base 2 expansion given by

$$T(x) = \cdot x_1 x_2 x_3 \dots$$

i.e. T deletes the zeroth term in the base 2 expansion of x and shifts the remaining terms one place to the left.

We introduce *dyadic intervals* or *cylinders* to be the sets

$$I(i_0, i_1, \dots, i_{n-1}) = \{x \in [0, 1] \mid x_j = i_j, j = 0, \dots, n-1\}.$$

(So, for example, $I(0) = [0, 1/2]$, $I(1) = [1/2, 1]$, $I(0, 0) = [0, 1/4]$, $I(0, 1) = [1/4, 1/2]$, etc.) We call n the *rank* of the cylinder. A dyadic interval is an interval with end-points at $k/2^n, (k+1)/2^n$ where $n \geq 1$ and $k \in \{0, 1, \dots, 2^n\}$.

Let \mathcal{A} denote the algebra of finite unions of cylinders. Then \mathcal{A} generates the Borel σ -algebra. This follows from Proposition 2.4.2 by noting that cylinders are intervals (and so Borel) and that they separate points: if $x, y \in [0, 1]$, $x \neq y$, then they have base 2 expansions that differ at some index, say $x_n \neq y_n$. Hence x, y belong to disjoint cylinders of rank n .

Define the maps

$$\phi_0(x) = \frac{x}{2}, \quad \phi_1(x) = \frac{x+1}{2}.$$

Then $\phi_0 : [0, 1] \rightarrow I(0)$ and $\phi_1 : [0, 1] \rightarrow I(1)$ are differentiable bijections. Indeed, if $x \in [0, 1]$ has base 2 expansion

$$x = \cdot x_0 x_1 x_2 \dots$$

then $\phi_0(x)$ and $\phi_1(x)$ have base 2 expansions given by

$$\phi_0(x) = \cdot 0 x_0 x_1 x_2 \dots, \quad \phi_1(x) = \cdot 1 x_0 x_1 x_2 \dots$$

Thus ϕ_0 and ϕ_1 act on base 2 expansions as a shift to the right, inserting the digits 0 and 1 in the zeroth place, respectively. Note that $T\phi_0(x) = x$ and $T\phi_1(x) = x$ for all $x \in [0, 1]$.

Given $i_0, i_1, \dots, i_{n-1} \in \{0, 1\}$, define

$$\phi_{i_0, i_1, \dots, i_{n-1}} : [0, 1] \rightarrow I(i_0, i_1, \dots, i_{n-1})$$

by

$$\phi_{i_0, i_1, \dots, i_{n-1}} = \phi_{i_0} \phi_{i_1} \cdots \phi_{i_{n-1}}. \quad (6.2.2)$$

Thus $\phi_{i_0, i_1, \dots, i_{n-1}}$ takes the point x with base 2 expansion given by (6.2.1), shifts the digits n places to the right, and inserts the digits i_0, i_1, \dots, i_{n-1} in the first n places. Note that $T^n \phi_{i_0, i_1, \dots, i_{n-1}}(x) = x$ for all $x \in [0, 1]$.

We are now in a position to prove that T is ergodic with respect to Lebesgue measure. Let $B \in \mathcal{B}$ be such that $T^{-1}B = B$. We must show that $\mu(B) = 0$ or 1 . By Lemma 6.1.1, it is sufficient to prove that there exists $K > 0$ such that $\mu(B)\mu(I) \leq K\mu(B \cap I)$ for all intervals I ; in fact, we shall prove that $\mu(B)\mu(I) = \mu(B \cap I)$ for all dyadic intervals I .

Note that $T^{-n}B = B$. Let $I = I(i_0, i_1, \dots, i_{n-1})$ be a cylinder of rank n and let $\phi = \phi_{i_0, i_1, \dots, i_{n-1}}$. Then $T^n \phi(x) = x$. Note also that $\mu(I) = 1/2^n$. We will also need the fact that $\phi'(x) = 1/2^n$ (this follows by noting that $\phi'_0(x) = \phi'_1(x) = 1/2$ and differentiating (6.2.2) using the chain rule).

Finally, we observe that

$$\begin{aligned}
 \mu(B \cap I) &= \int \chi_{B \cap I}(x) dx \\
 &= \int \chi_B(x) \chi_I(x) dx \\
 &= \int_I \chi_B(x) dx \\
 &= \int_0^1 \chi_B(\phi(x)) \phi'(x) dx \text{ by the change of variables formula} \\
 &= \int_0^1 \chi_{T^{-n}B}(\phi(x)) \phi'(x) dx \text{ as } T^{-n}B = B \\
 &= \int_0^1 \chi_B(T^n(\phi(x))) \phi'(x) dx \text{ as } \chi_{T^{-n}B} = \chi_B \circ T^n \\
 &= \int_0^1 \chi_B(x) \phi'(x) dx \text{ as } T^n \phi(x) = x \\
 &= \frac{1}{2^n} \int_0^1 \chi_B(x) dx \text{ as } \phi'(x) = 1/2^n \\
 &= \mu(I)\mu(B) \text{ as } \mu(I) = 1/2^n.
 \end{aligned}$$

Hence $\mu(B \cap I) = \mu(B)\mu(I)$ for all sets I in the algebra of cylinders. By Lemma 6.1.1 it follows that $\mu(B) = 0$ or 1 . Hence Lebesgue measure is an ergodic measure for T .

§6.3 The Gauss map

Let $x \in [0, 1]$. If x has continued fraction expansion

$$x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \dots}}}$$

then for brevity we write $x = [x_0, x_1, x_2, \dots]$.

Let $X = [0, 1]$ and recall that the Gauss map is defined by $T(x) = 1/x \bmod 1$ (with T defined at 0 by setting $T(0) = 0$). If x has continued fraction expansion $[x_0, x_1, x_2, \dots]$ then $T(x)$ has continued fraction expansion $[x_1, x_2, \dots]$. We have already seen that T leaves

Gauss' measure μ invariant, where Gauss' measure is defined by

$$\mu(B) = \frac{1}{\log 2} \int_B \frac{1}{1+x} dx.$$

We shall find it convenient to swap between Gauss' measure μ and Lebesgue measure, which we shall denote here by λ . Recall from Exercise 3.5 that for any set $B \in \mathcal{B}$ we have

$$\frac{1}{2 \log 2} \lambda(B) \leq \mu(B) \leq \frac{1}{\log 2} \lambda(B).$$

Hence $\mu(B) = 0$ if and only if $\lambda(B) = 0$. Thus to prove ergodicity it suffices to show that any T -invariant set B has either $\lambda(B) = 0$ or $\lambda(B^c) = 0$.

We shall also need some basic facts about continued fractions. Let $x \in (0, 1)$ be irrational and have continued fraction expansion $[x_0, x_1, \dots]$. For any $t \in [0, 1]$, write

$$[x_0, x_1, \dots, x_{n-1} + t] = \frac{P_n(x_0, x_1, \dots, x_{n-1}; t)}{Q_n(x_0, x_1, \dots, x_{n-1}; t)}$$

where $P_n(x_0, x_1, \dots, x_{n-1}; t)$ and $Q_n(x_0, x_1, \dots, x_{n-1}; t)$ are polynomials in x_0, x_1, \dots, x_{n-1} and t . Let $P_n = P_n(x_0, x_1, \dots, x_{n-1}; 0)$, $Q_n = Q_n(x_0, x_1, \dots, x_{n-1}; 0)$ (we suppress the dependence of P_n and Q_n on x_0, \dots, x_{n-1} for brevity). The following lemma is easily proved using induction.

Lemma 6.3.1

(i) We have

$$P_n(x_0, x_1, \dots, x_{n-1}; t) = P_n + tP_{n-1}, \quad Q_n(x_0, x_1, \dots, x_{n-1}; t) = Q_n + tQ_{n-1}.$$

and the following recurrence relations hold:

$$P_{n+1} = x_n P_n + P_{n-1}, \quad Q_{n+1} = x_n Q_n + Q_{n-1}$$

with initial conditions $P_0 = 0, P_1 = 1, Q_0 = 1, Q_1 = x_0$.

(ii) The following identity holds:

$$Q_n P_{n-1} - Q_{n-1} P_n = (-1)^n.$$

Let $i_0, i_1, \dots, i_{n-1} \in \mathbb{N}$. Define the cylinder $I(i_0, i_1, \dots, i_{n-1})$ to be the set of all points $x \in (0, 1)$ whose continued fraction expansion starts with i_0, \dots, i_{n-1} . This is easily seen to be an interval; indeed

$$I(i_0, i_1, \dots, i_{n-1}) = \{[i_0, i_1, \dots, i_{n-1} + t] \mid t \in [0, 1)\}.$$

Let \mathcal{A} denote the algebra of finite unions of cylinders. Then \mathcal{A} generates the Borel σ -algebra. (This follows from Proposition 2.4.2: cylinders are clearly Borel sets and they separate points. To see this, note that if $x \neq y$ then they have different continued fraction expansions. Hence there exists n such that $x_n \neq y_n$. Hence x, y are in different cylinders of rank n , and these cylinders are disjoint.)

For each $i \in \mathbb{N}$ define the map $\phi_i : [0, 1) \rightarrow I(i)$ by

$$\phi_i(x) = \frac{1}{i+x}.$$

Thus if x has continued fraction expansion $[x_0, x_1, \dots]$ then $\phi_i(x)$ has continued fraction expansion $[i, x_0, x_1, \dots]$. Clearly $T(\phi_i(x)) = x$ for all $x \in [0, 1)$.

For $i_0, i_1, \dots, i_{n-1} \in \mathbb{N}$, define

$$\phi_{i_0, i_1, \dots, i_{n-1}} = \phi_{i_0} \phi_{i_1} \cdots \phi_{i_{n-1}} : [0, 1) \rightarrow I(i_0, i_1, \dots, i_{n-1}).$$

Then $\phi_{i_0, i_1, \dots, i_{n-1}}(x)$ takes the continued fraction expansion of x , shifts every digit n places to the right, and inserts the digit i_0, i_1, \dots, i_{n-1} in the first n places. Clearly

$$T^n(\phi_{i_0, i_1, \dots, i_{n-1}}(x)) = x$$

for all $x \in [0, 1)$.

We first need an estimate on the length of (i.e. the Lebesgue measure of) the cylinder $I(i_0, i_1, \dots, i_{n-1})$. Note that

$$\phi_{i_0, i_1, \dots, i_{n-1}}(t) = \frac{P_n(i_0, \dots, i_{n-1}; t)}{Q_n(i_0, \dots, i_{n-1}; t)} = \frac{P_n + tP_{n-1}}{Q_n + tQ_{n-1}}.$$

Differentiating this expression with respect to t and using Lemma 6.3.1(ii), we see that

$$|\phi'_{i_0, i_1, \dots, i_{n-1}}(t)| = \left| \frac{Q_n P_{n-1} - P_n Q_{n-1}}{(Q_n + tQ_{n-1})^2} \right| = \frac{1}{(Q_n + tQ_{n-1})^2}.$$

It follows from Lemma 6.3.1(ii) that $Q_n + Q_{n-1} \leq 2Q_n$. Hence

$$\frac{1}{4} \frac{1}{Q_n^2} \leq \frac{1}{(Q_n + Q_{n-1})^2} \leq |\phi'_{i_0, i_1, \dots, i_{n-1}}(t)| \leq \frac{1}{Q_n^2}. \quad (6.3.1)$$

Hence

$$\lambda(I(i_0, i_1, \dots, i_{n-1})) = \int \chi_{I(i_0, i_1, \dots, i_{n-1})}(t) dt = \int_{I(i_0, i_1, \dots, i_{n-1})} dt = \int_0^1 |\phi'_{i_0, i_1, \dots, i_{n-1}}(t)| dt \quad (6.3.2)$$

where we have used the change of variables formula. Combining (6.3.2) with (6.3.1) we see that

$$\frac{1}{4} \frac{1}{Q_n^2} \leq \lambda(I(i_0, i_1, \dots, i_{n-1})) \leq \frac{1}{Q_n^2}. \quad (6.3.3)$$

We can now prove that the Gauss map is ergodic with respect to Gauss' measure μ . Suppose that $T^{-1}B = B$ where $B \in \mathcal{B}$. Let $I(i_0, i_1, \dots, i_{n-1})$ be a cylinder. Then

$$\begin{aligned} & \lambda(B \cap I(i_0, i_1, \dots, i_{n-1})) \\ &= \int_{I(i_0, i_1, \dots, i_{n-1})} \chi_B(x) dx \\ &= \int_0^1 \chi_B(\phi_{i_0, i_1, \dots, i_{n-1}}(x)) |\phi'_{i_0, i_1, \dots, i_{n-1}}(x)| dx \text{ by the change of variables formula.} \\ &= \int_0^1 \chi_{T^{-n}B}(\phi_{i_0, i_1, \dots, i_{n-1}}(x)) |\phi'_{i_0, i_1, \dots, i_{n-1}}(x)| dx \text{ as } T^{-n}B = B \\ &= \int_0^1 \chi_B(T^n(\phi_{i_0, i_1, \dots, i_{n-1}}(x))) |\phi'_{i_0, i_1, \dots, i_{n-1}}(x)| dx \text{ as } \chi_{T^{-n}B} = \chi_B \circ T^n \\ &= \int_0^1 \chi_B(x) |\phi'_{i_0, i_1, \dots, i_{n-1}}(x)| dx \text{ as } T^n \phi_{i_0, i_1, \dots, i_{n-1}}(x) = x. \end{aligned}$$

By (6.3.1) and (6.3.3) it follows that

$$\lambda(B \cap I(i_0, i_1, \dots, i_{n-1})) \geq \frac{1}{4Q_n^2} \lambda(B) \geq \frac{1}{4} \lambda(B) \lambda(I(i_0, i_1, \dots, i_{n-1}))$$

so that

$$\lambda(B) \lambda(I(i_0, i_1, \dots, i_{n-1})) \leq 4 \lambda(B \cap I(i_0, i_1, \dots, i_{n-1})).$$

By Lemma 6.1.1 it follows that $\lambda(B) = 0$ or $\lambda(B^c) = 0$. Hence, as Lebesgue measure and Gauss' measure have the same sets of measure zero, it follows that either $\mu(B) = 0$ or $\mu(B^c) = 0$. Hence T is ergodic with respect to Gauss' measure.

§6.4 Bernoulli shifts

Let $S = \{1, \dots, k\}$ be a finite set of symbols and let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \mid x_j \in \{1, 2, \dots, k\}\}$ denote the shift space on k symbols. Let $\sigma : \Sigma \rightarrow \Sigma$ denote the left shift map, so that $(\sigma(\mathbf{x}))_j = x_{j+1}$.

Recall that we defined the cylinder $[i_0, \dots, i_{n-1}]$ to be the set of all sequences in Σ that start with symbols i_0, \dots, i_{n-1} , that is

$$[i_0, \dots, i_{n-1}] = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \in \Sigma \mid x_j = i_j, j = 0, 1, \dots, n-1\}.$$

Let $p = (p(1), \dots, p(k))$ be a probability vector (that is, $p(j) > 0$, $\sum_{j=1}^k p(j) = 1$). We defined the Bernoulli measure μ_p on cylinders by setting

$$\mu_p[i_0, \dots, i_{n-1}] = p(i_0)p(i_1) \cdots p(i_{n-1}).$$

We have already seen that μ_p is a σ -invariant measure.

Proposition 6.4.1

Let μ_p be a Bernoulli measure. Then μ_p is ergodic.

Proof. We first make the following observation: let $I = [i_0, \dots, i_{p-1}]$, $J = [j_0, \dots, j_{q-1}]$ be cylinders of ranks p, q , respectively. Consider $I \cap \sigma^{-n}J$ where $n \geq p$. Then

$$\begin{aligned} I \cap \sigma^{-n}J &= \{\mathbf{x} = (x_k)_{k=0}^{\infty} \in \Sigma \mid x_k = i_k \text{ for } k = 0, 1, \dots, p-1, x_{k+n} = j_k \text{ for } k = 0, 1, \dots, q-1\} \\ &= \bigcup_{x_p, \dots, x_{n-1}} [i_0, i_1, \dots, i_{p-1}, x_p, \dots, x_{n-1}, j_0, \dots, j_{q-1}], \end{aligned}$$

a disjoint union. Hence

$$\begin{aligned} \mu_p(I \cap \sigma^{-n}J) &= \sum_{x_p, \dots, x_{n-1}} \mu_p[i_0, i_1, \dots, i_{p-1}, x_p, \dots, x_{n-1}, j_0, \dots, j_{q-1}] \\ &= \sum_{x_p, \dots, x_{n-1}} p(i_0)p(i_1) \cdots p(i_{p-1})p(x_p) \cdots p(x_{n-1})p(j_0)p(j_1) \cdots p(j_{q-1}) \\ &= p(i_0)p(i_1) \cdots p(i_{p-1})p(j_0)p(j_1) \cdots p(j_{q-1}) \text{ as } \sum_{x_p=1}^k p(x_p) = \cdots = \sum_{x_{n-1}=1}^k p(x_{n-1}) = 1 \\ &= \mu_p(I)\mu_p(J). \end{aligned} \tag{6.4.1}$$

Let $B \in \mathcal{B}$ be σ -invariant. By Lemma 6.1.1 it is sufficient to prove that $\mu_p(B)\mu_p(I) \leq \mu_p(B \cap I)$ for each cylinder I . Let $\varepsilon > 0$. We first approximate the invariant set B by a finite

union of cylinders. By Proposition 2.4.4, we can find a finite disjoint union of cylinders $A = \bigcup_{j=1}^r J_j$ such that $\mu_p(B\Delta A) < \varepsilon$. Note that $|\mu_p(A) - \mu_p(B)| < \varepsilon$.

Let n be any integer greater than the rank of I . Note that $\sigma^{-n}B\Delta\sigma^{-n}A = \sigma^{-n}(B\Delta A)$. Hence

$$\mu_p(\sigma^{-n}B\Delta\sigma^{-n}A) = \mu_p(\sigma^{-n}(B\Delta A)) = \mu_p(B\Delta A) < \varepsilon,$$

where we have used the facts that $\sigma^{-n}B = B$ and that μ_p is an invariant measure.

As $A = \bigcup_{j=1}^r J_j$ is a finite union of cylinders and n is greater than the rank of I , it follows from (6.4.1) that

$$\begin{aligned} \mu_p(\sigma^{-n}A \cap I) &= \mu_p\left(\sigma^{-n}\left(\bigcup_{j=1}^r J_j\right) \cap I\right) = \sum_{j=1}^r \mu_p(\sigma^{-n}J_j \cap I) \\ &= \sum_{j=1}^r \mu_p(J_j)\mu_p(I) = \mu_p\left(\bigcup_{j=1}^r J_j\right)\mu_p(I) \\ &= \mu_p(A)\mu_p(I). \end{aligned}$$

Finally, note that $(\sigma^{-n}A \cap I)\Delta(\sigma^{-n}B \cap I) \subset (\sigma^{-n}A)\Delta(\sigma^{-n}B)$. Hence $\mu_p((\sigma^{-n}A \cap I)\Delta(\sigma^{-n}B \cap I)) < \varepsilon$ so that $\mu_p(\sigma^{-n}A \cap I) < \mu_p(\sigma^{-n}B \cap I) + \varepsilon$. Hence

$$\begin{aligned} \mu_p(B)\mu_p(I) &= \mu_p(\sigma^{-n}B)\mu_p(I) \leq \mu_p(\sigma^{-n}A)\mu_p(I) + \varepsilon = \mu_p(\sigma^{-n}A \cap I) + \varepsilon \\ &\leq \mu_p(\sigma^{-n}B \cap I) + 2\varepsilon = \mu_p(B \cap I) + 2\varepsilon. \end{aligned}$$

As $\varepsilon > 0$ is arbitrary, we have that $\mu_p(B)\mu_p(I) \leq \mu_p(B \cap I)$ for any cylinder I . By Lemma 6.1.1, it follows that $\mu_p(B) = 0$ or 1. Hence μ_p is ergodic. \square

§6.5 Markov shifts

Let P be an irreducible stochastic $k \times k$ matrix with entries $P(i, j)$. Let $p = (p(1), \dots, p(k))$ be the unique left probability eigenvector corresponding to eigenvalue 1, so that $pP = p$. Recall that the Markov measure μ_P is defined on the Borel σ -algebra by defining it on cylinders in the following way:

$$\mu_P([i_0, i_1, \dots, i_{n-1}]) = p(i_0)P(i_0, i_1)P(i_1, i_2) \cdots P(i_{n-2}, i_{n-1}).$$

We have seen that μ_P is an invariant measure for the shift map σ . We can adapt the proof of Proposition 6.4.1 to show that μ_P is ergodic.

Proposition 6.5.1

Let P be an irreducible stochastic matrix. Then the corresponding Markov measure μ_P is ergodic.

Proof (not examinable). Let d denote the period of P .

Let $I = [i_0, \dots, i_{p-1}]$, $J = [j_0, \dots, j_{q-1}]$ be cylinders of ranks p, q , respectively. Consider $I \cap \sigma^{-n}J$ where $n \geq p$. Then

$$\begin{aligned} I \cap \sigma^{-n}J &= \{\mathbf{x} \in \Sigma \mid x_j = i_j \text{ for } j = 0, 1, \dots, p-1, x_{j+n} = j_j \text{ for } j = 0, 1, \dots, q-1\} \\ &= \bigcup_{x_p, \dots, x_{n-1}} [i_0, i_1, \dots, i_{p-1}, x_p, \dots, x_{n-1}, j_0, \dots, j_{q-1}], \end{aligned}$$

a disjoint union. Hence

$$\begin{aligned}
& \mu_P(I \cap \sigma^{-n}J) \\
&= \sum_{x_p, \dots, x_{n-1}} \mu_p[i_0, i_1, \dots, i_{p-1}, x_p, \dots, x_{n-1}, j_0, \dots, j_{q-1}] \\
&= \sum_{x_p, \dots, x_{n-1}} p(i_0)P(i_0, i_1) \cdots P(i_{p-2}, i_{p-1})P(i_{p-1}, x_p)P(x_p, x_{p+1}) \cdots P(x_{n-1}, j_0) \\
&\quad \times P(j_0, j_1) \cdots P(j_{q-2}, j_{q-1}) \\
&= \mu_P(I)\mu_P(J) \frac{1}{p(j_0)} \sum_{x_p, \dots, x_{n-1}} P(i_{p-1}, x_p)P(x_p, x_{p+1}) \cdots P(x_{n-1}, j_0) \\
&= \mu_P(I)\mu_P(J) \frac{P^{n-1-p}(i_{p-1}, j_0)}{p(j_0)}
\end{aligned}$$

By the Perron-Frobenius Theorem (Theorem 3.3.6), we know that $P^{nd}(i, j) \rightarrow p(j)$ as $n \rightarrow \infty$. Hence, letting $n \rightarrow \infty$ through an appropriate subsequence, we see that

$$\mu_P(I \cap \sigma^{-n}J) \rightarrow \mu_P(I)\mu_P(J). \quad (6.5.1)$$

The remainder of the proof is almost identical to the proof of Proposition 6.4.1. Let $B \in \mathcal{B}$ be σ -invariant. By Proposition 6.1.1 it is sufficient to prove that $\mu_P(B)\mu_P(I) \leq \mu_P(B \cap I)$ for every cylinder I . Let $\varepsilon > 0$. We approximate B by a finite union of cylinders by using Proposition 2.4.4. That is, we can find a finite disjoint union of cylinders $A = \bigcup_{j=1}^r J_j$ such that $\mu_P(B \Delta A) < \varepsilon$. Note that $|\mu_P(A) - \mu_P(B)| < \varepsilon$.

Let n be any integer greater than the rank of I . Note that $\sigma^{-n}B \Delta \sigma^{-n}(A) = \sigma^{-n}(B \Delta A)$. Hence

$$\mu_P(\sigma^{-n}B \Delta \sigma^{-n}A) = \mu_P(\sigma^{-n}(B \Delta A)) = \mu_P(B \Delta A) < \varepsilon$$

where we have used the facts that $\sigma^{-n}B = B$ and that μ_P is an invariant measure.

As $A = \bigcup_{j=1}^r J_j$ is a finite union of cylinders, it follows from (6.5.1) that by choosing n sufficiently large, we have that $\mu_P(\sigma^{-n}J_j \cap I) \leq \mu_P(J_j)\mu_P(I) + \varepsilon$ for $j = 1, 2, \dots, r$. Hence

$$\begin{aligned}
\mu_P(\sigma^{-n}A)\mu_P(I) &= \mu_P\left(\sigma^{-n}\left(\bigcup_{j=1}^r J_j\right)\right)\mu_P(I) = \sum_{j=1}^r \mu_P(\sigma^{-n}J_j)\mu_P(I) \\
&= \sum_{j=1}^r \mu_P(J_j \cap I) + \varepsilon = \mu_P\left(\bigcup_{j=1}^r J_j \cap I\right) + \varepsilon \\
&= \mu_P(A \cap I) + \varepsilon.
\end{aligned}$$

Finally, note that $(\sigma^{-n}A \cap I) \Delta (\sigma^{-n}B \cap I) \subset (\sigma^{-n}A) \Delta (\sigma^{-n}B)$. Hence $\mu_P((\sigma^{-n}A \cap I) \Delta (\sigma^{-n}B \cap I)) < \varepsilon$ so that $\mu_P(\sigma^{-n}B \cap I) < \mu_P(\sigma^{-n}A \cap I) + \varepsilon$. Hence

$$\begin{aligned}
\mu_P(B)\mu_P(I) &= \mu_P(\sigma^{-n}B)\mu_P(I) \leq \mu_P(\sigma^{-n}A)\mu_P(I) + \varepsilon = \mu_P(\sigma^{-n}A \cap I) + 2\varepsilon \\
&\leq \mu_P(\sigma^{-n}B \cap I) + 3\varepsilon = \mu_P(B \cap I) + 3\varepsilon.
\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, we have that $\mu_P(B)\mu_P(I) \leq \mu_P(B \cap I)$. By Proposition 6.1.1, the result follows. \square

§6.6 Exercises

Exercise 6.1

The dynamical system $T : [0, 1] \rightarrow [0, 1]$ defined by

$$T(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1/2 \\ 2(1-x) & \text{if } 1/2 \leq x \leq 1 \end{cases}$$

is called the tent map.

- (i) Prove that T preserves Lebesgue measure.
- (ii) Prove that T is ergodic with respect to Lebesgue measure.

Exercise 6.2

Recall that the Lüroth map $T : [0, 1] \rightarrow [0, 1]$ is defined to be

$$T(x) = \begin{cases} n(n+1)x - n & \text{if } x \in \left(\frac{1}{n+1}, \frac{1}{n}\right] \\ 0 & \text{if } x = 0. \end{cases}$$

We saw in Exercise 3.9 that Lebesgue measure is a T -invariant probability measure. Prove that Lebesgue measure is ergodic.

Exercise 6.3

Prove (using induction on n) Lemma 6.3.1.

Exercise 6.4

Let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \mid x_j \in \{0, 1\}\}$ and let $\sigma : \Sigma \rightarrow \Sigma$, $(\sigma(\mathbf{x}))_j = x_{j+1}$ be the shift map on the space of infinite sequences of two symbols $\{0, 1\}$. Note that Σ supports uncountably many different σ -invariant measures (for example, the Bernoulli- $(p, 1-p)$ measures are all ergodic and all distinct for $p \in (0, 1)$). We will use this observation to prove that the doubling map has uncountably many ergodic measures.

Define $\pi : \Sigma \rightarrow \mathbb{R}/\mathbb{Z}$ by

$$\pi(\mathbf{x}) = \pi(x_0, x_1, \dots) = \frac{x_0}{2} + \frac{x_1}{2^2} + \dots + \frac{x_n}{2^{n+1}} + \dots$$

- (i) Show that π is continuous.
- (ii) Let $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ be the doubling map: $T(x) = 2x \bmod 1$. Show that $\pi \circ \sigma = T \circ \pi$.
- (iii) If μ is a σ -invariant probability measure on Σ , show that $\pi_*\mu$ (where $\pi_*\mu(B) = \mu(\pi^{-1}B)$ for a Borel subset $B \subset \mathbb{R}/\mathbb{Z}$) is a T -invariant probability measure on \mathbb{R}/\mathbb{Z} . (Lebesgue measure on \mathbb{R}/\mathbb{Z} corresponds to choosing μ to be the Bernoulli- $(1/2, 1/2)$ -measure on Σ .)
- (iv) Show that if μ is an ergodic measure for σ , then $\pi_*\mu$ is an ergodic measure for T .
- (v) Conclude that there are uncountably many different ergodic measures for the doubling map.

7. Continuous transformations on compact metric spaces

§7.1 Introduction

So far, we have been studying a measurable map T defined on a probability space (X, \mathcal{B}, μ) . We have asked whether the given measure μ is invariant or ergodic. In this section, we shift our focus slightly and consider, for a given transformation $T : X \rightarrow X$, the space $M(X, T)$ of all probability measures that are invariant under T . In order to equip $M(X, T)$ with some structure we will need to assume that the underlying space X is itself equipped with some additional structure other than merely being a measure space. Throughout this section we will work in the context of X being a compact metric space and T being a continuous transformation.

§7.2 Probability measures on compact metric spaces

Let X be a compact metric space equipped with the Borel σ -algebra \mathcal{B} . (Recall that the Borel σ -algebra is the smallest σ -algebra that contains all the open subsets of X .)

Let $C(X, \mathbb{R}) = \{f : X \rightarrow \mathbb{R} \mid f \text{ is continuous}\}$ denote the space of real-valued continuous functions defined on X . Define the *uniform norm* of $f \in C(X, \mathbb{R})$ by

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

With this norm, $C(X, \mathbb{R})$ is a Banach space.

An important property of $C(X, \mathbb{R})$ that will prove to be useful later on is that it is *separable*: $C(X, \mathbb{R})$ contains a countable dense subset. Thus we can choose a sequence $\{f_n \in C(X, \mathbb{R})\}_{n=1}^\infty$ such that, for all $f \in C(X, \mathbb{R})$ and all $\varepsilon > 0$, there exists n such that $\|f - f_n\|_\infty < \varepsilon$.

Let $M(X)$ denote the set of all Borel probability measures on (X, \mathcal{B}) .

It will be very important to have a sensible notion of convergence in $M(X)$; the appropriate notion for us is called *weak* convergence*. We say that a sequence of probability measures μ_n *weak* converges* to μ as $n \rightarrow \infty$ if, for every $f \in C(X, \mathbb{R})$,

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu.$$

If μ_n weak* converges to μ then we write $\mu_n \rightarrow \mu$ as $n \rightarrow \infty$. We can make $M(X)$ into a metric space compatible with this definition of convergence by choosing a countable dense subset $\{f_n\}_{n=1}^\infty \subset C(X, \mathbb{R})$ and, for $\mu_1, \mu_2 \in M(X)$, setting

$$d_{M(X)}(\mu_1, \mu_2) = \sum_{n=1}^{\infty} \frac{1}{2^n \|f_n\|_\infty} \left| \int f_n d\mu_1 - \int f_n d\mu_2 \right|$$

(we can assume that $f_n \not\equiv 0$ for any n). It is easy to check that $\mu_n \rightarrow \mu$ as $n \rightarrow \infty$ if and only if $d_{M(X)}(\mu_n, \mu) \rightarrow 0$ as $n \rightarrow \infty$.

However, we will not need to work with a particular metric: what will be important is the definition of convergence.

Remark. Note that with this definition it is not necessarily true that $\lim_{n \rightarrow \infty} \mu_n(B) = \mu(B)$ for $B \in \mathcal{B}$.

§7.2.1 The Riesz Representation Theorem

Let $\mu \in M(X)$ be a Borel probability measure. Then we can think of μ as a functional that acts on $C(X, \mathbb{R})$, that is we can regard μ as a map

$$\mu : C(X, \mathbb{R}) \rightarrow \mathbb{R} : f \mapsto \int f d\mu.$$

We will often write $\mu(f)$ for $\int f d\mu$.

Notice that this functional enjoys several natural properties:

(i) the functional defined by μ is linear:

$$\mu(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 \mu(f_1) + \lambda_2 \mu(f_2)$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ and $f_1, f_2 \in C(X, \mathbb{R})$.

(ii) the functional defined by μ is bounded: i.e. if $f \in C(X, \mathbb{R})$ then $|\mu(f)| \leq \|f\|_\infty$.

(iii) if $f \geq 0$ then $\mu(f) \geq 0$ (we say that the functional μ is *positive*);

(iv) consider the function $\mathbf{1}$ defined by $\mathbf{1}(x) \equiv 1$ for all x ; then $\mu(\mathbf{1}) = 1$ (we say that the functional μ is *normalised*).

The Riesz Representation Theorem says that the above properties characterise all Borel probability measures on X . That is, if we have a map $w : C(X, \mathbb{R}) \rightarrow \mathbb{R}$ that satisfies the above four properties, then w must be given by integrating with respect to a Borel probability measure. This will be a very useful method of constructing measures: we need only construct bounded positive normalised linear functionals.

Theorem 7.2.1 (Riesz Representation Theorem)

Let $w : C(X, \mathbb{R}) \rightarrow \mathbb{R}$ be a functional such that:

(i) w is linear: i.e. $w(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 w(f_1) + \lambda_2 w(f_2)$;

(ii) w is bounded: i.e. for all $f \in C(X, \mathbb{R})$ we have $|w(f)| \leq \|f\|_\infty$;

(iii) w is positive: i.e. if $f \geq 0$ then $w(f) \geq 0$;

(iv) w is normalised: i.e. $w(\mathbf{1}) = 1$.

Then there exists a Borel probability measure $\mu \in M(X)$ such that

$$w(f) = \int f d\mu.$$

Moreover, μ is unique.

Thus the Riesz Representation Theorem says that “if it looks like integration on continuous functions, then it is integration with respect to a (unique) Borel probability measure.”

§7.2.2 Properties of $M(X)$

First note that the space $M(X)$ of Borel probability measures on the compact metric space X is non-empty (provided $X \neq \emptyset$). This is because, for each $x \in X$, the Dirac measure δ_x is a Borel probability measure. Indeed, we have the following result:

Proposition 7.2.2

There is a continuous embedding of X in $M(X)$ given by the map $X \rightarrow M(X) : x \mapsto \delta_x$, i.e. if $x_n \rightarrow x$ then $\delta_{x_n} \rightarrow \delta_x$.

Proof. See Exercise 7.1. □

Recall that a subset C of a vector space is *convex* if whenever $v_1, v_2 \in C$ and $\alpha \in [0, 1]$ then $\alpha v_1 + (1 - \alpha)v_2 \in C$.

Proposition 7.2.3

The space $M(X)$ is convex.

Proof. Let $\mu_1, \mu_2 \in M(X)$, $\alpha \in [0, 1]$. Then it is easy to check that $\alpha\mu_1 + (1 - \alpha)\mu_2$, defined by

$$(\alpha\mu_1 + (1 - \alpha)\mu_2)(B) = \alpha\mu_1(B) + (1 - \alpha)\mu_2(B),$$

is a Borel probability measure. □

Finally, recall that a metric space K is said to be (sequentially) compact if every sequence of points in K has a convergent subsequence.

Proposition 7.2.4

The space $M(X)$ is weak* compact.

Proof. For convenience, we shall write $\mu(f) = \int f d\mu$.

Since $C(X, \mathbb{R})$ is separable, we can choose a countable dense subset of functions $\{f_i\}_{i=1}^\infty \subset C(X, \mathbb{R})$. Given a sequence $\mu_n \in M(X)$, we shall first consider the sequence of real numbers $\mu_n(f_1) \in \mathbb{R}$. We have that $|\mu_n(f_1)| \leq \|f_1\|_\infty$ for all n , so $\mu_n(f_1)$ is a bounded sequence of real numbers. As such, it has a convergent subsequence, $\mu_n^{(1)}(f_1)$ say.

Next we apply the sequence of measures $\mu_n^{(1)}$ to f_2 and consider the sequence $\mu_n^{(1)}(f_2) \in \mathbb{R}$. Again, this is a bounded sequence of real numbers and so it has a convergent subsequence $\mu_n^{(2)}(f_2)$.

In this way we obtain, for each $i \geq 1$, nested subsequences $\{\mu_n^{(i)}\} \subset \{\mu_n^{(i-1)}\}$ such that $\mu_n^{(i)}(f_j)$ converges for $1 \leq j \leq i$. Now consider the diagonal sequence $\mu_n^{(n)}$. Since, for $n \geq i$, $\mu_n^{(n)}$ is a subsequence of $\mu_n^{(i)}$, $\mu_n^{(n)}(f_i)$ converges for every $i \geq 1$.

We can now use the fact that $\{f_i\}$ is dense to show that $\mu_n^{(n)}(f)$ converges for all $f \in C(X, \mathbb{R})$, as follows. For any $\varepsilon > 0$, we can choose f_i such that $\|f - f_i\|_\infty \leq \varepsilon$. Since $\mu_n^{(n)}(f_i)$ converges, there exists N such that if $n, m \geq N$ then

$$|\mu_n^{(n)}(f_i) - \mu_m^{(m)}(f_i)| \leq \varepsilon.$$

Thus if $n, m \geq N$ we have

$$\begin{aligned} |\mu_n^{(n)}(f) - \mu_m^{(m)}(f)| &\leq |\mu_n^{(n)}(f) - \mu_n^{(n)}(f_i)| + |\mu_n^{(n)}(f_i) - \mu_m^{(m)}(f_i)| + |\mu_m^{(m)}(f_i) - \mu_m^{(m)}(f)| \\ &\leq 3\varepsilon, \end{aligned}$$

so $\mu_n^{(n)}(f)$ converges, as required.

To complete the proof, write $w(f) = \lim_{n \rightarrow \infty} \mu_n^{(n)}(f)$. We claim that w satisfies the hypotheses of the Riesz Representation Theorem and so corresponds to integration with respect to a probability measure.

(i) By construction, w is a linear mapping: $w(\lambda f + \mu g) = \lambda w(f) + \mu w(g)$.

(ii) As $|w(f)| \leq \|f\|_\infty$, we see that w is bounded.

(iii) If $f \geq 0$ then it is easy to check that $w(f) \geq 0$. Hence w is positive.

(iv) It is easy to check that w is normalised: $w(\mathbf{1}) = 1$.

Therefore, by the Riesz Representation Theorem, there exists $\mu \in M(X)$ such that $w(f) = \int f d\mu$. We then have that $\int f d\mu_n^{(n)} \rightarrow \int f d\mu$, as $n \rightarrow \infty$, for all $f \in C(X, \mathbb{R})$, i.e., that $\mu_n^{(n)}$ converges weak* to μ , as $n \rightarrow \infty$. \square

§7.3 Invariant measures for continuous transformations

Let X be a compact metric space equipped with the Borel σ -algebra and let $T : X \rightarrow X$ be a continuous transformation. It is clear that T is measurable.

Given a measure μ , we have already defined the measure $T_*\mu$ by $T_*\mu(B) = \mu(T^{-1}B)$. If μ is a Borel probability measure, then it is straightforward to check that $T_*\mu$ is a Borel probability measure. We can think of T_* as a transformation on $M(X)$, namely:

$$T_* : M(X) \rightarrow M(X), \quad T_*\mu = \mu \circ T^{-1}.$$

That is, if $B \in \mathcal{B}$ then $T_*\mu(B) = \mu(T^{-1}B)$.

The following result tells us how to integrate with respect to $T_*\mu$.

Lemma 7.3.1

For $f \in L^1(X, \mathcal{B}, \mu)$ we have

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

Proof. From the definition, for $B \in \mathcal{B}$,

$$\int \chi_B d(T_*\mu) = (T_*\mu)(B) = \mu(T^{-1}B) = \int \chi_{T^{-1}B} d\mu = \int \chi_B \circ T d\mu.$$

Thus the result holds for simple functions. If $f \geq 0$ is a positive measurable function then we can choose an increasing sequence of simple functions f_n increasing to f pointwise. We have

$$\int f_n d(T_*\mu) = \int f_n \circ T d\mu$$

and, applying the Monotone Convergence Theorem (Theorem 3.1.2) to each side, we obtain

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

The result extends to an arbitrary real-valued $f \in L^1(X, \mathcal{B}, \mu)$ by considering positive and negative parts and then to complex-valued integrable functions by taking real and imaginary parts. \square

Recall that a measure μ is said to be T -invariant if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$. Hence μ is T -invariant if and only if $T_*\mu = \mu$. Write

$$M(X, T) = \{\mu \in M(X) \mid T_*\mu = \mu\}$$

to denote the space of all T -invariant Borel probability measures.

The following result gives a useful criterion for checking whether a measure is T -invariant.

Lemma 7.3.2

Let $T : X \rightarrow X$ be a continuous mapping of a compact metric space. The following are equivalent:

- (i) $\mu \in M(X, T)$;
- (ii) for all $f \in C(X, \mathbb{R})$ we have that

$$\int f \circ T d\mu = \int f d\mu. \quad (7.3.1)$$

Proof. We prove (i) implies (ii). Suppose that $\mu \in M(X, T)$ so that $T_*\mu = \mu$. Let $f \in C(X, \mathbb{R})$. Then $f \in L^1(X, \mathcal{B}, \mu)$. Hence by Lemma 7.3.1, for any $f \in C(X, \mathbb{R})$ we have

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

Conversely, Lemma 7.3.1 allows us to write (7.3.1) as: $\mu(f) = (T_*\mu)(f)$ for all $f \in C(X, \mathbb{R})$. Hence μ and $T_*\mu$ determine the same linear functional on $C(X, \mathbb{R})$. By uniqueness in the Riesz Representation theorem, we have $T_*\mu = \mu$. \square

§7.4 Invariant measures for continuous maps on the torus

We can use Lemma 7.3.2 to prove that a given measure is invariant for certain dynamical systems. We first note that we need only check (7.3.1) for a dense set of continuous functions.

Lemma 7.4.1

Suppose that $\mathcal{S} \subset C(X, \mathbb{R})$ is a uniformly dense subset of functions (that is, for all $f \in C(X, \mathbb{R})$ and all $\varepsilon > 0$ there exists $g \in \mathcal{S}$ such that $\|f - g\|_\infty < \varepsilon$). Suppose that $\int g \circ T d\mu = \int g d\mu$ for all $g \in \mathcal{S}$. Then $\int f \circ T d\mu = \int f d\mu$ for all $f \in C(X, \mathbb{R})$.

Proof. Let $f \in C(X, \mathbb{R})$ and let $\varepsilon > 0$. Choose $g \in \mathcal{S}$ such that $\|f - g\|_\infty < \varepsilon$. Then

$$\begin{aligned} & \left| \int f \circ T d\mu - \int f d\mu \right| \\ & \leq \left| \int f \circ T d\mu - \int g \circ T d\mu \right| + \left| \int g \circ T d\mu - \int g d\mu \right| + \left| \int g d\mu - \int f d\mu \right| \\ & \leq \int |f \circ T - g \circ T| d\mu + \left| \int g \circ T d\mu - \int g d\mu \right| + \int |f - g| d\mu. \end{aligned}$$

Noting that, as $\|f - g\|_\infty < \varepsilon$, we have that $|f(Tx) - g(Tx)| < \varepsilon$ for all x , and that $\int g \circ T d\mu = \int g d\mu$, we have that

$$\left| \int f \circ T d\mu - \int f d\mu \right| < 2\varepsilon.$$

As ε is arbitrary, the result follows. \square

Corollary 7.4.2

Let T be a continuous transformation of a compact metric space X , equipped with the Borel σ -algebra. Let μ be a Borel probability measure on X .

Suppose that $\mathcal{S} \subset C(X, \mathbb{R})$ is a uniformly dense subset of functions such that $\int g \circ T d\mu = \int g d\mu$ for all $g \in \mathcal{S}$. Then μ is a T -invariant measure.

Proof. This follows immediately from Lemma 7.3.2 and Lemma 7.4.1. \square

We show how to use Corollary 7.4.2 by studying some of our examples.

§7.4.1 Circle rotations

Let $T(x) = x + \alpha \pmod{1}$ be a circle rotation. We show how to use Corollary 7.4.2 to prove that Lebesgue measure μ is T -invariant.

Let $\ell \in \mathbb{Z}$. We first note that if $\ell \neq 0$ then

$$\int e^{2\pi i \ell x} dx = \frac{1}{2\pi i \ell} e^{2\pi i \ell x} \Big|_0^1 = 0.$$

We also note that if $\ell = 0$ then $\int e^{2\pi i \ell x} dx = 1$.

Let \mathcal{S} denote the set of trigonometric polynomials, i.e.

$$\mathcal{S} = \left\{ \sum_{j=0}^{r-1} c_j e^{2\pi i \ell_j x} \mid c_j \in \mathbb{R}, \ell_j \in \mathbb{Z}, r \in \mathbb{N} \right\}.$$

Then \mathcal{S} is uniformly dense in $C(X, \mathbb{R})$ by the Stone-Weierstrass Theorem (Theorem 1.2.2).

Let $g \in \mathcal{S}$ be a trigonometric polynomial and write

$$g(x) = \sum_{j=0}^{r-1} c_j e^{2\pi i \ell_j x}$$

where $\ell_j = 0$ if and only if $j = 0$. Hence $\int g d\mu = c_0$.

Note that

$$g(Tx) = \sum_{j=0}^{r-1} c_j e^{2\pi i \ell_j (x+\alpha)} = \sum_{j=0}^{r-1} c_j e^{2\pi i \ell_j \alpha} e^{2\pi i \ell_j x}.$$

Hence

$$\int g \circ T d\mu = \int \sum_{j=0}^{r-1} c_j e^{2\pi i \ell_j \alpha} e^{2\pi i \ell_j x} d\mu = \sum_{j=0}^{r-1} c_j e^{2\pi i \ell_j \alpha} \int e^{2\pi i \ell_j x} d\mu$$

and the only non-zero integral occurs when $\ell_j = 0$, i.e. $j = 0$. We must therefore have that $\int g \circ T d\mu = c_0$.

Hence $\int g \circ T d\mu = \int g d\mu$ for all $g \in \mathcal{S}$. It follows from Corollary 7.4.2 that μ is T -invariant.

§7.4.2 Toral endomorphisms

Let A be a $k \times k$ integer matrix with $\det A \neq 0$. Define the linear toral endomorphism $T : \mathbb{R}^k/\mathbb{Z}^k \rightarrow \mathbb{R}^k/\mathbb{Z}^k$ by

$$T((x_1, \dots, x_k) + \mathbb{Z}^k) = A(x_1, \dots, x_k) + \mathbb{Z}^k.$$

When T is a linear toral automorphism (i.e. when $\det A = \pm 1$) we have already seen that Lebesgue measure is invariant. We use Corollary 7.4.2 to prove the Lebesgue measure μ is T -invariant when $\det A \neq 0$.

For $\mathbf{n} = (n_1, \dots, n_k) \in \mathbb{Z}^k$ and $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ define, as before, $\langle \mathbf{n}, x \rangle = n_1x_1 + \dots + n_kx_k$. Note that

$$\int e^{2\pi i \langle \mathbf{n}, x \rangle} d\mu = \int \dots \int e^{2\pi i n_1 x_1} \dots e^{2\pi i n_k x_k} dx_1 \dots dx_k.$$

Hence

$$\int e^{2\pi i \langle \mathbf{n}, x \rangle} d\mu = \begin{cases} 0 & \text{if } \mathbf{n} \neq \mathbf{0} \\ 1 & \text{if } \mathbf{n} = \mathbf{0} \end{cases}$$

where $\mathbf{0} = (0, \dots, 0) \in \mathbb{Z}^k$.

Let

$$\mathcal{S} = \left\{ \sum_{j=0}^{r-1} c_j e^{2\pi i \langle \mathbf{n}^{(j)}, x \rangle} \mid r \in \mathbb{N}, c_j \in \mathbb{R}, \mathbf{n}^{(j)} = (n_1^{(j)}, \dots, n_k^{(j)}) \in \mathbb{Z}^k \right\}.$$

By the Stone-Weierstrass Theorem (Theorem 1.2.2), we see that \mathcal{S} is uniformly dense in $C(\mathbb{R}^k/\mathbb{Z}^k, \mathbb{R})$.

Let $g \in \mathcal{S}$ and write

$$g(x) = \sum_{j=0}^{r-1} c_j e^{2\pi i \langle \mathbf{n}^{(j)}, x \rangle}$$

where $\mathbf{n}^{(j)} = \mathbf{0}$ if and only if $j = 0$. Then

$$\int g d\mu = \int \sum_{j=0}^{r-1} c_j e^{2\pi i \langle \mathbf{n}^{(j)}, x \rangle} d\mu = \sum_{j=0}^{r-1} c_j \int e^{2\pi i \langle \mathbf{n}^{(j)}, x \rangle} d\mu = c_0.$$

Note that

$$g(Tx) = \sum_{j=0}^{r-1} c_j e^{2\pi i \langle \mathbf{n}^{(j)}, Ax \rangle} = \sum_{j=0}^{r-1} c_j e^{2\pi i \langle \mathbf{n}^{(j)} A, x \rangle}.$$

Hence

$$\int g \circ T d\mu = \int \sum_{j=0}^{r-1} c_j e^{2\pi i \langle \mathbf{n}^{(j)} A, x \rangle} d\mu = \sum_{j=0}^{r-1} c_j \int e^{2\pi i \langle \mathbf{n}^{(j)} A, x \rangle} d\mu.$$

These integrals are zero unless $\mathbf{n}^{(j)} A = \mathbf{0}$. As $\det A \neq 0$ this happens only when $\mathbf{n}^{(j)} = \mathbf{0}$, i.e. when $j = 0$. Hence

$$\int g \circ T d\mu = c_0 = \int g d\mu.$$

Hence by Corollary 7.4.2, μ is a T -invariant measure.

Remark. You will notice a strong connection between the above arguments and Fourier series and you may think that we could take $g(x)$ to be the n th partial sum of the Fourier series for f . However, one needs to take care. Suppose $f \in C(\mathbb{R}^k/\mathbb{Z}^k, \mathbb{R})$ has Fourier series $\sum_{\mathbf{n}} c_{\mathbf{n}}(f)e^{2\pi i\langle \mathbf{n}, x \rangle}$. We need to be careful about what it means for this infinite series to converge. We know that the sequence of partial sums s_n converges in L^2 to f , but we do not know that the partial sums converge uniformly to f . That is, we know that $\|f - s_n\|_2 \rightarrow 0$, but not necessarily that $\|f - s_n\|_{\infty} \rightarrow 0$. In fact, in general, it is not true that $\|f - s_n\|_{\infty} \rightarrow 0$.

However, if one defines $\sigma_n = 1/n \sum_{j=0}^{n-1} s_j$ to be the average of the first n partial sums, then it is true that $\|f - \sigma_n\|_{\infty} \rightarrow 0$. (This is quite a deep result.)

§7.5 Existence of invariant measures

Given a continuous mapping $T : X \rightarrow X$ of a compact metric space, it is natural to ask whether invariant measures necessarily exist, i.e., whether $M(X, T) \neq \emptyset$. The next result shows that this is the case.

Theorem 7.5.1

Let $T : X \rightarrow X$ be a continuous mapping of a compact metric space. Then there exists at least one T -invariant probability measure.

Proof. Let $\nu \in M(X)$ be a probability measure (for example, we could take ν to be a Dirac measure). Define the sequence $\mu_n \in M(X)$ by

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T_*^j \nu,$$

so that, for $B \in \mathcal{B}$,

$$\mu_n(B) = \frac{1}{n} (\nu(B) + \nu(T^{-1}B) + \dots + \nu(T^{-(n-1)}B)).$$

Since $M(X)$ is weak* compact, some subsequence μ_{n_k} converges, as $k \rightarrow \infty$, to a measure $\mu \in M(X)$. We shall show that $\mu \in M(X, T)$. By Lemma 7.3.2, this is equivalent to showing that

$$\int f d\mu = \int f \circ T d\mu \text{ for all } f \in C(X, \mathbb{R}).$$

To see this, first note that $f \circ T - f$ is continuous. Then

$$\begin{aligned} \left| \int f \circ T d\mu - \int f d\mu \right| &= \left| \int (f \circ T - f) d\mu \right| \\ &= \lim_{k \rightarrow \infty} \left| \int (f \circ T - f) d\mu_{n_k} \right| \\ &= \lim_{k \rightarrow \infty} \left| \int (f \circ T - f) d \left(\frac{1}{n_k} \sum_{j=0}^{n_k-1} T_*^j \nu \right) \right| \\ &= \lim_{k \rightarrow \infty} \left| \frac{1}{n_k} \int \sum_{j=0}^{n_k-1} (f \circ T - f) dT_*^j \nu \right| \end{aligned}$$

$$\begin{aligned}
 &= \lim_{k \rightarrow \infty} \left| \frac{1}{n_k} \int \sum_{j=0}^{n_k-1} (f \circ T^{j+1} - f \circ T^j) d\nu \right| \\
 &= \lim_{k \rightarrow \infty} \left| \frac{1}{n_k} \int (f \circ T^{n_k} - f) d\nu \right| \\
 &\leq \lim_{k \rightarrow \infty} \frac{2\|f\|_\infty}{n_k} = 0.
 \end{aligned}$$

Therefore, $\mu \in M(X, T)$, as required. □

We will need the following additional properties of $M(X, T)$.

Theorem 7.5.2

Let $T : X \rightarrow X$ be a continuous mapping of a compact metric space. Then $M(X, T)$ is a weak* compact and convex subset of $M(X)$.

Proof. The fact that $M(X, T)$ is convex is straightforward from the definition.

To see that $M(X, T)$ is weak* compact it is sufficient to show that it is a weak* closed subset of the weak* compact $M(X)$. Suppose that $\mu_n \in M(X, T)$ is such that $\mu_n \rightharpoonup \mu \in M(X)$. We need to show that $\mu \in M(X, T)$. To see this, observe that for any $f \in C(X, \mathbb{R})$ we have that

$$\begin{aligned}
 \int f \circ T d\mu &= \lim_{n \rightarrow \infty} \int f \circ T d\mu_n \text{ as } f \circ T \text{ is continuous} \\
 &= \lim_{n \rightarrow \infty} \int f d\mu_n \text{ as } \mu_n \in M(X, T) \\
 &= \int f d\mu \text{ as } \mu_n \rightharpoonup \mu.
 \end{aligned}$$

□

§7.6 Exercises

Exercise 7.1

Prove Proposition 7.2.2: show that if $x_n, x \in X$ and $x_n \rightarrow x$ then $\delta_{x_n} \rightarrow \delta_x$.

Exercise 7.2

Prove that $T_* : M(X) \rightarrow M(X)$ is weak* continuous (i.e. if $\mu_n \rightharpoonup \mu$ then $T_*\mu_n \rightharpoonup T_*\mu$).

Exercise 7.3

Let X be a compact metric space. For $\mu \in M(X)$ define

$$\|\mu\| = \sup_{f \in C(X, \mathbb{R}), \|f\|_\infty \leq 1} \left| \int f d\mu \right|.$$

We say that μ_n converges *strongly* to μ if $\|\mu_n - \mu\| \rightarrow 0$ as $n \rightarrow \infty$. The topology this determines is called the *strong topology* (or the *operator topology*).

- (i) Show that if $\mu_n \rightarrow \mu$ strongly then $\mu_n \rightharpoonup \mu$ in the weak* topology.
- (ii) Suppose that X is infinite. Show that $X \hookrightarrow M(X) : x \mapsto \delta_x$ is *not* continuous in the strong topology.

- (iii) Prove that $\|\delta_x - \delta_y\| = 2$ if $x \neq y$. (You may use Urysohn's Lemma: Let A and B be disjoint closed subsets of a metric space X . Then there is a continuous function $f \in C(X, \mathbb{R})$ such that $0 \leq f \leq 1$ on X while $f \equiv 0$ on A and $f \equiv 1$ on B .)

Hence prove that $M(X)$ is not compact in the strong topology when X is infinite.

Exercise 7.4

Give an example of a sequence of measures μ_n and a set B such that $\mu_n \rightarrow \mu$ but $\mu_n(B) \not\rightarrow \mu(B)$.

Exercise 7.5

Prove that $M(X, T)$ is convex.

Exercise 7.6

Suppose that $\mathcal{S} \subset C(X, \mathbb{R})$ is a uniformly dense subset of functions (that is, for all $f \in C(X, \mathbb{R})$ and all $\varepsilon > 0$ there exists $g \in \mathcal{S}$ such that $\|f - g\|_\infty < \varepsilon$). Let $\mu_n, \mu \in M(X)$. Suppose that $\int f d\mu_n \rightarrow \int f d\mu$ for all $f \in \mathcal{S}$. Prove that $\mu_n \rightarrow \mu$.

Exercise 7.7

Let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^\infty \mid x_j \in \{0, 1\}\}$ denote the shift space on two symbols 0, 1. Let $\sigma : \Sigma \rightarrow \Sigma$, $(\sigma(\mathbf{x}))_j = x_{j+1}$ denote the shift map.

- (i) How many periodic points of period n are there?
 (ii) Let $\text{Per}(n)$ denote the set of periodic points of period n . Define

$$\mu_n = \frac{1}{2^n} \sum_{\mathbf{x} \in \text{Per}(n)} \delta_{\mathbf{x}}.$$

Let $i_j \in \{0, 1\}$, $0 \leq j \leq m - 1$ and define the *cylinder set*

$$[i_0, i_1, \dots, i_{m-1}] = \{\mathbf{x} = (x_j)_{j=0}^\infty \in \Sigma \mid x_j = i_j, j = 0, 1, \dots, m - 1\}.$$

Let μ denote the Bernoulli-(1/2, 1/2) measure. Prove that

$$\int \chi_{[i_0, i_1, \dots, i_{m-1}]} d\mu_n \rightarrow \int \chi_{[i_0, i_1, \dots, i_{m-1}]} d\mu \text{ as } n \rightarrow \infty.$$

- (iii) Prove that $\chi_{[i_0, i_1, \dots, i_{m-1}]}$ is a continuous function.
 (iv) Use Exercise 7.6 and the Stone-Weierstrass Theorem (Theorem 1.2.2) to show that $\mu_n \rightarrow \mu$ as $n \rightarrow \infty$.

Exercise 7.8

Let $X = \mathbb{R}^3 / \mathbb{Z}^3$ be the 3-dimensional torus. Let $\alpha \in \mathbb{R}$. Define $T : X \rightarrow X$ by

$$T \left(\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbb{Z}^3 \right) \right) = \left(\begin{pmatrix} \alpha + x \\ y + x \\ z + y \end{pmatrix} + \mathbb{Z}^3 \right).$$

Use Corollary 7.4.2 to prove that Lebesgue measure μ is a T -invariant measure.

8. Ergodic measures for continuous transformations

§8.1 Introduction

In the previous section we saw that, given a continuous transformation of a compact metric space, the set of T -invariant Borel probability measures is non-empty. One can ask a similar question: is the set of ergodic Borel probability measures non-empty? In this section we address this question. We let $E(X, T) \subset M(X, T)$ denote the set of ergodic T -invariant Borel probability measures on X .

§8.2 Radon-Nikodym derivatives

We will need the concept of Radon-Nikodym derivatives.

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

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

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

$$\nu(B) = \int_B f d\mu. \quad (8.2.1)$$

Then $\nu \ll \mu$.

As a particular example, let $X = [0, 1]$ be equipped with the Borel σ -algebra \mathcal{B} . Define $f : [0, 1] \rightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1/2 \\ 0 & \text{if } 1/2 < x \leq 1. \end{cases}$$

Let μ be Lebesgue measure and let ν be the measure given by

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

If $A \subset [1/2, 1]$ is any Borel set then $\nu(A) = 0$.

The following theorem says that, essentially, all absolutely continuous measures occur by the construction in (8.2.1).

Theorem 8.2.1 (Radon-Nikodym)

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

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

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

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

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

The following relations are all easy to prove, and indicate why the notation was chosen in this way.

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

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

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

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

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

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

§8.3 Ergodic measures as extreme points

§8.3.1 Extreme points of convex sets

A point in a convex set is called an *extreme point* if it cannot be written as a non-trivial convex combination of (other) elements of the set. More precisely, μ is an extreme point of $M(X, T)$ if, whenever

$$\mu = \alpha\mu_1 + (1 - \alpha)\mu_2,$$

with $\mu_1, \mu_2 \in M(X, T)$, $0 < \alpha < 1$, then we have $\mu_1 = \mu_2 = \mu$.

Remarks.

(i) Let Y be the unit square

$$Y = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\} \subset \mathbb{R}^2.$$

Then the extreme points of Y are the corners $(0, 0), (0, 1), (1, 0), (1, 1)$.

(ii) Let Y be the (closed) unit disc

$$Y = \{(x, y) \mid x^2 + y^2 \leq 1\} \subset \mathbb{R}^2.$$

Then the set of extreme points of Y is precisely the unit circle $\{(x, y) \mid x^2 + y^2 = 1\}$.

§8.3.2 Existence of ergodic measures

The next result will allow us to show that ergodic measures for continuous transformations on compact metric spaces always exist.

Theorem 8.3.1

Let T be a continuous transformation of a compact metric space X equipped with the Borel σ -algebra \mathcal{B} . The following are equivalent:

- (i) the T -invariant probability measure μ is ergodic;
(ii) μ is an extreme point of $M(X, T)$.

Proof. We prove (ii) implies (i). If μ is an extreme point of $M(X, T)$ then it is ergodic. In fact, we shall prove the contrapositive. Suppose that μ is not ergodic; we show that μ is not an extreme point of $M(X, T)$. As μ is not ergodic, there exists $B \in \mathcal{B}$ such that $T^{-1}B = B$ and $0 < \mu(B) < 1$.

Define probability measures μ_1 and μ_2 on X by

$$\mu_1(A) = \frac{\mu(A \cap B)}{\mu(B)}, \quad \mu_2(A) = \frac{\mu(A \cap (X \setminus B))}{\mu(X \setminus B)}.$$

(The assumption that $0 < \mu(B) < 1$ ensures that the denominators are not equal to zero.) Clearly, $\mu_1 \neq \mu_2$, since $\mu_1(B) = 1$ while $\mu_2(B) = 0$.

Since $T^{-1}B = B$, we also have $T^{-1}(X \setminus B) = X \setminus B$. Thus we have

$$\begin{aligned} \mu_1(T^{-1}A) &= \frac{\mu(T^{-1}A \cap B)}{\mu(B)} \\ &= \frac{\mu(T^{-1}A \cap T^{-1}B)}{\mu(B)} \\ &= \frac{\mu(T^{-1}(A \cap B))}{\mu(B)} \\ &= \frac{\mu(A \cap B)}{\mu(B)} \\ &= \mu_1(A) \end{aligned}$$

and (by the same argument)

$$\mu_2(T^{-1}A) = \frac{\mu(T^{-1}A \cap (X \setminus B))}{\mu(X \setminus B)} = \mu_2(A),$$

i.e., μ_1 and μ_2 are both in $M(X, T)$.

However, we may write μ as the non-trivial (since $0 < \mu(B) < 1$) convex combination

$$\mu = \mu(B)\mu_1 + (1 - \mu(B))\mu_2,$$

so that μ is not an extreme point. □

Proof (not examinable). We prove (i) implies (ii). Suppose that μ is ergodic and that $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$, with $\mu_1, \mu_2 \in M(X, T)$ and $0 < \alpha < 1$. We shall show that $\mu_1 = \mu$ (so that $\mu_2 = \mu$, also). This will show that μ is an extreme point of $M(X, T)$.

If $\mu(A) = 0$ then $\mu_1(A) = 0$, so that $\mu_1 \ll \mu$. Therefore the Radon-Nikodym derivative $d\mu_1/d\mu \geq 0$ exists. One can easily deduce from the statement of the Radon-Nikodym Theorem that $\mu_1 = \mu$ if and only if $d\mu_1/d\mu = 1$ μ -a.e. We shall show that this is indeed the case by showing that the sets where, respectively, $d\mu_1/d\mu < 1$ and $d\mu_1/d\mu > 1$ both have μ -measure zero.

Let

$$B = \left\{ x \in X \mid \frac{d\mu_1}{d\mu}(x) < 1 \right\}.$$

Now

$$\mu_1(B) = \int_B \frac{d\mu_1}{d\mu} d\mu = \int_{B \cap T^{-1}B} \frac{d\mu_1}{d\mu} d\mu + \int_{B \setminus T^{-1}B} \frac{d\mu_1}{d\mu} d\mu \quad (8.3.1)$$

and

$$\mu_1(T^{-1}B) = \int_{T^{-1}B} \frac{d\mu_1}{d\mu} d\mu = \int_{B \cap T^{-1}B} \frac{d\mu_1}{d\mu} d\mu + \int_{T^{-1}B \setminus B} \frac{d\mu_1}{d\mu} d\mu. \quad (8.3.2)$$

As $\mu_1 \in M(X, T)$, we have that $\mu_1(B) = \mu_1(T^{-1}B)$. Hence comparing the last summands in both (8.3.1) and (8.3.2) we obtain

$$\int_{B \setminus T^{-1}B} \frac{d\mu_1}{d\mu} d\mu = \int_{T^{-1}B \setminus B} \frac{d\mu_1}{d\mu} d\mu. \quad (8.3.3)$$

In fact, these integrals are taken over sets of the same μ -measure:

$$\begin{aligned} \mu(T^{-1}B \setminus B) &= \mu(T^{-1}B) - \mu(T^{-1}B \cap B) \\ &= \mu(B) - \mu(T^{-1}B \cap B) \\ &= \mu(B \setminus T^{-1}B). \end{aligned}$$

Note that on the left-hand side of (8.3.3), the integrand $d\mu_1/d\mu < 1$. However, on the right-hand side of (8.3.3), the integrand $d\mu_1/d\mu \geq 1$. Thus we must have that $\mu(B \setminus T^{-1}B) = \mu(T^{-1}B \setminus B) = 0$, which is to say that $\mu(T^{-1}B \triangle B) = 0$, i.e. $T^{-1}B = B$ μ -a.e. Therefore, since μ is ergodic, we have that $\mu(B) = 0$ or $\mu(B) = 1$.

We can rule out the possibility that $\mu(B) = 1$ by observing that if $\mu(B) = 1$ then

$$1 = \mu_1(X) = \int_X \frac{d\mu_1}{d\mu} d\mu = \int_B \frac{d\mu_1}{d\mu} d\mu < \mu(B) = 1,$$

a contradiction. Therefore $\mu(B) = 0$.

If we define

$$C = \left\{ x \in X \mid \frac{d\mu_1}{d\mu}(x) > 1 \right\}$$

then repeating essentially the same argument gives $\mu(C) = 0$.

Hence

$$\mu \left\{ x \in X \mid \frac{d\mu_1}{d\mu}(x) = 1 \right\} = \mu(X \setminus (B \cup C)) = \mu(X) - \mu(B) - \mu(C) = 1,$$

i.e., $d\mu_1/d\mu = 1$ μ -a.e. Therefore $\mu_1 = \mu$, as required. \square

We can now prove that a continuous transformation of a compact metric space always has an ergodic measure. To do this, we will show that $M(X, T)$ has an extreme point.

Theorem 8.3.2

Let $T : X \rightarrow X$ be a continuous mapping of a compact metric space. Then there exists at least one ergodic measure in $M(X, T)$.

Proof. By Theorem 8.3.1, it is equivalent to prove that $M(X, T)$ has an extreme point.

Choose a countable dense subset of $C(X, \mathbb{R})$, $\{f_i\}_{i=0}^\infty$ say. Consider the first function f_0 . Since the map

$$M(X, T) \rightarrow \mathbb{R} : \mu \mapsto \int f_0 d\mu$$

is (weak*) continuous and $M(X, T)$ is compact, there exists (at least one) $\nu \in M(X, T)$ such that

$$\int f_0 d\nu = \sup_{\mu \in M(X, T)} \int f_0 d\mu.$$

If we define

$$M_0 = \left\{ \nu \in M(X, T) \mid \int f_0 d\nu = \sup_{\mu \in M(X, T)} \int f_0 d\mu \right\}$$

then the above shows that M_0 is non-empty. Also, M_0 is closed and hence compact.

We now consider the next function f_1 and define

$$M_1 = \left\{ \nu \in M_0 \mid \int f_1 d\nu = \sup_{\mu \in M_0} \int f_1 d\mu \right\}.$$

By the same reasoning as above, M_1 is a non-empty closed subset of M_0 .

Continuing inductively, we define

$$M_j = \left\{ \nu \in M_{j-1} \mid \int f_j d\nu = \sup_{\mu \in M_{j-1}} \int f_j d\mu \right\}$$

and hence obtain a nested sequence of sets

$$M(X, T) \supset M_0 \supset M_1 \supset \cdots \supset M_j \supset \cdots$$

with each M_j non-empty and closed.

Now consider the intersection

$$M_\infty = \bigcap_{j=0}^{\infty} M_j.$$

Recall that the intersection of a decreasing sequence of non-empty compact sets is non-empty. Hence M_∞ is non-empty and we can pick $\mu_\infty \in M_\infty$. We shall show that μ_∞ is an extreme point (and hence ergodic).

Suppose that we can write $\mu_\infty = \alpha\mu_1 + (1 - \alpha)\mu_2$, $\mu_1, \mu_2 \in M(X, T)$, $0 < \alpha < 1$. We have to show that $\mu_1 = \mu_2$. Since $\{f_j\}_{j=0}^{\infty}$ is dense in $C(X, \mathbb{R})$, it suffices to show that

$$\int f_j d\mu_1 = \int f_j d\mu_2 \quad \forall j \geq 0.$$

Consider f_0 . By assumption

$$\int f_0 d\mu_\infty = \alpha \int f_0 d\mu_1 + (1 - \alpha) \int f_0 d\mu_2.$$

In particular,

$$\int f_0 d\mu_\infty \leq \max \left\{ \int f_0 d\mu_1, \int f_0 d\mu_2 \right\}.$$

However $\mu_\infty \in M_0$ and so

$$\int f_0 d\mu_\infty = \sup_{\mu \in M(X, T)} \int f_0 d\mu \geq \max \left\{ \int f_0 d\mu_1, \int f_0 d\mu_2 \right\}.$$

Therefore

$$\int f_0 d\mu_1 = \int f_0 d\mu_2 = \int f_0 d\mu_\infty.$$

Thus, the first identity we require is proved and $\mu_1, \mu_2 \in M_0$. This last fact allows us to employ the same argument on f_1 (with $M(X, T)$ replaced by M_0) and conclude that

$$\int f_1 d\mu_1 = \int f_1 d\mu_2 = \int f_1 d\mu_\infty$$

and $\mu_1, \mu_2 \in M_1$.

Continuing inductively, we show that for an arbitrary $j \geq 0$,

$$\int f_j d\mu_1 = \int f_j d\mu_2$$

and $\mu_1, \mu_2 \in M_j$. This completes the proof. □

§8.4 An example: the North-South map

For many dynamical systems there exist uncountably many different ergodic measures. This is the case for the doubling map, Markov shifts, toral automorphisms, etc. Here we give an example of a dynamical system $T : X \rightarrow X$ for which one can construct $M(X, T)$ and $E(X, T)$ explicitly.

Let $X \subset \mathbb{R}^2$ denote the circle of radius 1 centred at $(0, 1) \in \mathbb{R}^2$. Call $N = (0, 2)$ the North Pole and $S = (0, 0)$ the South Pole (S) of X . Define a map $\phi : X \setminus \{N\} \rightarrow \mathbb{R} \times \{0\}$ by drawing a straight line through N and x and denoting by $\phi(x)$ the unique point on the x -axis that this line crosses (this is just stereographic projection of the circle). Define $T : X \rightarrow X$ by

$$T(x) = \begin{cases} \phi^{-1} \left(\frac{1}{2} \phi(x) \right) & \text{if } x \in X \setminus \{N\}, \\ N & \text{if } x = N. \end{cases}$$

Hence $T(N) = N$, $T(S) = S$ and if $x \neq N, S$ then $T^n(x) \rightarrow S$ as $n \rightarrow \infty$. We call T the

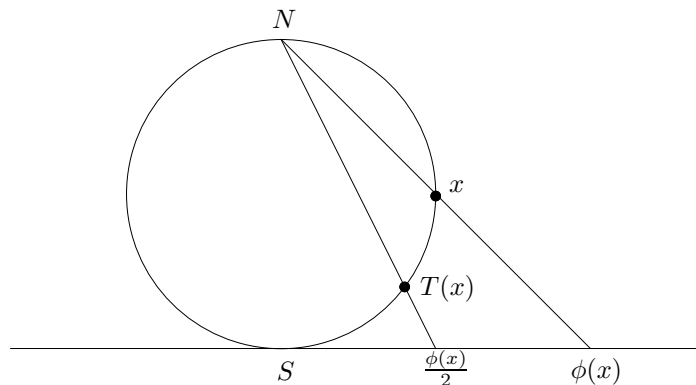


Figure 8.4.1: The North-South map

North-South map.

Clearly both N and S are fixed points for T . Hence δ_N and δ_S (the Dirac delta measures at N , S , respectively) are T -invariant. It is easy to see that both δ_N and δ_S are ergodic.

Now let $\mu \in M(X, T)$ be an invariant measure. We claim that μ assigns zero measure to the set $X \setminus \{N, S\}$. Let $x \in X$ be any point in the right semi-circle (for example, take $x = (1, 1) \in \mathbb{R}^2$) and consider the arc I of semi-circle from x to $T(x)$. Then $\bigcup_{n=-\infty}^{\infty} T^{-n}I$ is a disjoint union of arcs of semi-circle and, moreover, is equal to the entire right semi-circle. Now

$$\mu \left(\bigcup_{n=-\infty}^{\infty} T^{-n}I \right) = \sum_{n=-\infty}^{\infty} \mu(T^{-n}I) = \sum_{n=-\infty}^{\infty} \mu(I)$$

and the only way for this to be finite is if $\mu(I) = 0$. Hence μ assigns zero measure to the entire right semi-circle. Similarly, μ assigns zero measure to the left semi-circle.

Hence μ is concentrated on the two points N, S , and so must be a convex combination of the Dirac delta measures δ_N and δ_S . Hence

$$M(X, T) = \{\alpha\delta_N + (1 - \alpha)\delta_S \mid \alpha \in [0, 1]\}$$

and the ergodic measures are the extreme points of $M(X, T)$, namely δ_N, δ_S .

§8.5 Unique ergodicity

We conclude by looking at the case where $T : X \rightarrow X$ has a *unique* invariant probability measure.

Definition. Let $T : X \rightarrow X$ be a continuous transformation of a compact metric space X . If there is a *unique* T -invariant probability measure then we say that T is *uniquely ergodic*.

Remark. You might wonder why such T are not instead called ‘uniquely invariant’. Recall that the extreme points of $M(X, T)$ are precisely the ergodic measures. If $M(X, T)$ consists of just one measure then that measure is an extreme, and so must be ergodic.

Unique ergodicity implies the following strong convergence result.

Theorem 8.5.1 (Oxtoby’s Ergodic Theorem)

Let X be a compact metric space and let $T : X \rightarrow X$ be a continuous transformation. The following are equivalent:

- (i) T is uniquely ergodic;
- (ii) for each $f \in C(X, \mathbb{R})$ there exists a constant $c(f)$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \rightarrow c(f), \tag{8.5.1}$$

uniformly for $x \in X$.

Remark. The convergence in (8.5.1) means that

$$\lim_{n \rightarrow \infty} \sup_{x \in X} \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - c(f) \right| = 0.$$

Remark. If $M(X, T) = \{\mu\}$ then the constant $c(f)$ in (8.5.1) is $\int f d\mu$.

Proof. We prove (ii) implies (i). Suppose that μ, ν are T -invariant probability measures; we shall show that $\mu = \nu$. Integrating the expression in (ii), we obtain

$$\int f d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int f \circ T^j d\mu = \int \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j d\mu = \int c(f) d\mu = c(f),$$

(that the convergence in (8.5.1) is uniform allows us to interchange integration and taking limits) and, by the same argument

$$\int f d\nu = c(f).$$

Therefore

$$\int f d\mu = \int f d\nu \text{ for all } f \in C(X, \mathbb{R})$$

and so $\mu = \nu$ (by the Riesz Representation Theorem).

We prove (i) implies (ii). Let $M(X, T) = \{\mu\}$. If (ii) is true, then, by the Dominated Convergence Theorem (Theorem 3.1.3), we must necessarily have $c(f) = \int f d\mu$.

The convergence in (ii) means: $\forall f \in C(X, \mathbb{R}), \forall \varepsilon > 0, \exists N \in \mathbb{N}$ such that if $n \geq N$ then for all $x \in X$ we have

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - \int f d\mu \right| < \varepsilon.$$

Suppose that (ii) is false. Then, negating the above quantifiers, we see that there exists $f_0 \in C(X, \mathbb{R})$ and $\varepsilon > 0$ and an increasing sequence $n_k \uparrow \infty$ such that there exists x_{n_k} for which

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} f_0(T^j x_{n_k}) - \int f_0 d\mu \right| \geq \varepsilon. \quad (8.5.2)$$

Define the probability measure $\mu_k \in M(X)$ by

$$\mu_k = \frac{1}{n_k} \sum_{j=0}^{n_k-1} T_*^j \delta_{x_{n_k}},$$

so that (8.5.2) can be written as

$$\left| \int f_0 d\mu_k - \int f_0 d\mu \right| \geq \varepsilon.$$

Now $\mu_k \in M(X)$ and $M(X)$ is weak* compact. Hence there exists a weak* convergent subsequence, say with weak* limit ν . By following the proof of Theorem 7.5.1, it is easy to see that $\nu \in M(X, T)$. In particular, we have

$$\left| \int f_0 d\nu - \int f_0 d\mu \right| \geq \varepsilon.$$

Therefore, $\nu \neq \mu$, contradicting unique ergodicity. \square

§8.6 Irrational rotations

Let $X = \mathbb{R}/\mathbb{Z}$, $T : X \rightarrow X$, $T(x) = x + \alpha \pmod{1}$ where α is irrational. We have already seen that Lebesgue measure μ is an ergodic T -invariant measure. We can prove that Lebesgue measure is the only invariant measure.

Proposition 8.6.1

An irrational rotation of a circle is uniquely ergodic and the unique T -invariant measure is Lebesgue measure.

Proof. We use Oxtoby's Ergodic Theorem. To prove that T is uniquely ergodic, we must show that (8.5.1) holds for every continuous function $f \in C(X, \mathbb{R})$. Note that the convergence in (8.5.1) is uniform, i.e. we must show that

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_{\infty} \rightarrow 0 \quad (8.6.1)$$

as $n \rightarrow \infty$.

We first prove (8.6.1) in the case when $f(x) = e^{2\pi i \ell x}$, $\ell \in \mathbb{Z} \setminus \{0\}$. Note that $T^j(x) = x + j\alpha$. Hence

$$\begin{aligned} \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \right| &= \left| \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell (x+j\alpha)} \right| \\ &= \left| e^{2\pi i \ell x} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell \alpha j} \right| \\ &= \frac{1}{n} \frac{|e^{2\pi i \ell \alpha n} - 1|}{|e^{2\pi i \ell \alpha} - 1|} \\ &\leq \frac{1}{n} \frac{2}{|e^{2\pi i \ell \alpha} - 1|}. \end{aligned} \quad (8.6.2)$$

As α is irrational, the denominator in (8.6.2) is not zero. Note also that $\int e^{2\pi i \ell x} d\mu = 0$. Hence

$$\sup_{x \in X} \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right| \rightarrow 0$$

as $n \rightarrow \infty$ when $f(x) = e^{2\pi i \ell x}$, $\ell \in \mathbb{Z} \setminus \{0\}$. Clearly (8.6.1) holds when f is a constant function. By taking finite linear combinations of exponential functions we see that

$$\sup_{x \in X} \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right| \rightarrow 0$$

as $n \rightarrow \infty$ for all trigonometric polynomials g . By the Stone-Weierstrass Theorem (Theorem 1.2.2), trigonometric polynomials are uniformly dense in $C(X, \mathbb{R})$. Let $f \in C(X, \mathbb{R})$ and let $\varepsilon > 0$. Then there exists a trigonometric polynomial g such that $\|f - g\|_{\infty} < \varepsilon$.

Hence for any $x \in X$ we have

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right| \\
& \leq \left| \frac{1}{n} \sum_{j=0}^{n-1} (f(T^j(x)) - g(T^j(x))) \right| + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right| + \left| \int g - f d\mu \right| \\
& \leq \frac{1}{n} \sum_{j=0}^{n-1} |f(T^j(x)) - g(T^j(x))| + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right| + \int |g - f| d\mu \\
& \leq 2\varepsilon + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right|.
\end{aligned}$$

Hence, taking the supremum over all $x \in X$, we have

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_{\infty} < 2\varepsilon + \left\| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right\|_{\infty}.$$

Letting $n \rightarrow \infty$ we see that

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_{\infty} < 2\varepsilon.$$

As $\varepsilon > 0$ is arbitrary, it follows that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_{\infty} = 0.$$

Hence statement (ii) in Oxtoby's Ergodic Theorem holds. As (i) and (ii) in Oxtoby's Ergodic Theorem are equivalent, it follows that T is uniquely ergodic and Lebesgue measure is the unique invariant measure. \square

§8.7 Exercises

Exercise 8.1

Prove the following identities concerning Radon-Nikodym derivatives.

- (i) If $\nu \ll \mu$ and $f \in L^1(X, \mathcal{B}, \mu)$ then $f \in L^1(X, \mathcal{B}, \nu)$ and

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

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

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

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

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

Exercise 8.2

Let $X = \mathbb{R}^3/\mathbb{Z}^3$ be the 3-dimensional torus. Fix $\alpha \notin \mathbb{Q}$. Define $T : X \rightarrow X$ by

$$T \left(\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbb{Z}^3 \right) \right) = \left(\left(\begin{pmatrix} \alpha + x \\ y + x \\ z + y \end{pmatrix} + \mathbb{Z}^3 \right) \right).$$

Prove by induction that for $n \geq 3$

$$T^n \left(\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbb{Z}^3 \right) \right) = \left(\left(\begin{pmatrix} \binom{n}{1} \alpha + x \\ \binom{n}{2} \alpha + \binom{n}{1} x + y \\ \binom{n}{3} \alpha + \binom{n}{2} x + \binom{n}{1} y + z \end{pmatrix} + \mathbb{Z}^3 \right) \right)$$

(here $\binom{n}{r}$ denotes the binomial coefficient).

Let $f(x, y, z) = e^{2\pi i(kx + \ell y + mz)}$ where $k, \ell, m \in \mathbb{Z}$. Assuming Weyl's Theorem on Polynomials (Theorem 2.3.1), prove using Weyl's Criterion (Theorem 1.2.1) that

$$\sup_{x,y,z} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j((x, y, z) + \mathbb{Z}^3)) \rightarrow 0$$

as $n \rightarrow \infty$, whenever $(k, \ell, m) \in \mathbb{Z}^3 \setminus \{(0, 0, 0)\}$.

Hence, using Oxtoby's Ergodic Theorem, prove that T is uniquely ergodic and Lebesgue measure is the unique invariant measure.

Exercise 8.3

Let T be a homeomorphism of a compact metric space X . Suppose that T is uniquely ergodic with unique invariant measure μ . Prove that every orbit of T is dense if, and only if, $\mu(U) > 0$ for every non-empty open set U .

9. Recurrence

§9.1 Introduction

We can now begin to study ergodic theorems. Before we do this, we discuss a remarkable result due to Poincaré.

§9.2 Poincaré's Recurrence Theorem

Theorem 9.2.1 (Poincaré's Recurrence Theorem)

Let $T : X \rightarrow X$ be a measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . Let $B \in \mathcal{B}$ be such that $\mu(B) > 0$. Then for μ -a.e. $x \in B$, the orbit $\{T^n x\}_{n=0}^{\infty}$ returns to B infinitely often.

Proof. Let

$$E = \{x \in B \mid T^n x \in B \text{ for infinitely many } n \geq 1\},$$

then we have to show that $\mu(B \setminus E) = 0$.

If we write

$$F = \{x \in B \mid T^n x \notin B \forall n \geq 1\}$$

then we have the identity

$$B \setminus E = \bigcup_{k=0}^{\infty} (T^{-k} F \cap B).$$

Thus we have the estimate

$$\mu(B \setminus E) = \mu\left(\bigcup_{k=0}^{\infty} (T^{-k} F \cap B)\right) \leq \mu\left(\bigcup_{k=0}^{\infty} T^{-k} F\right) \leq \sum_{k=0}^{\infty} \mu(T^{-k} F).$$

Since $\mu(T^{-k} F) = \mu(F) \forall k \geq 0$ (because the measure is preserved), it suffices to show that $\mu(F) = 0$.

First suppose that $n > m$ and that $T^{-m} F \cap T^{-n} F \neq \emptyset$. If y lies in this intersection then $T^m y \in F$ and $T^{n-m}(T^m y) = T^n y \in F \subset B$, which contradicts the definition of F . Thus $T^{-m} F$ and $T^{-n} F$ are disjoint.

Since $\{T^{-k} F\}_{k=0}^{\infty}$ is a disjoint family, we have

$$\sum_{k=0}^{\infty} \mu(T^{-k} F) = \mu\left(\bigcup_{k=0}^{\infty} T^{-k} F\right) \leq \mu(X) = 1.$$

Since the terms in the summation have the constant value $\mu(F)$, we must have $\mu(F) = 0$. \square

Remark. Note that the hypotheses of the Poincaré Recurrence Theorem are very mild: all one needs is for T to be a measure-preserving transformation of a probability space. (One does not need T to be ergodic.) If you carefully look at the proof, you will see that the fact that T is measure-preserving and the fact that $\mu(X) = 1$ are used just once. The same proof continues to hold in the case when $\mu(X)$ is finite. Poincaré's Recurrence Theorem is false with either of the hypotheses that $\mu(X)$ is finite or T is measure-preserving removed.

§9.3 Ergodic Theorems

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

$$\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j \quad (9.3.1)$$

as $n \rightarrow \infty$. The precise formulation of an ergodic theorem depends on the class of function f (for example, one could assume that f is integrable, L^2 , or continuous), and the notion of convergence used (for example, the convergence could be pointwise, L^2 , or uniform). We have already studied when one has uniform convergence of (9.3.1): this is Oxtoby's Ergodic Theorem and only holds in the very special circumstances when T is uniquely ergodic. In what follows we will discuss von Neumann's (Mean) Ergodic Theorem and Birkhoff's Ergodic Theorem. Von Neumann's Ergodic Theorem is in the context of $f \in L^2(X, \mathcal{B}, \mu)$ and L^2 -convergence of the ergodic averages (9.3.1); Birkhoff's Ergodic Theorem is in the context of $f \in L^1(X, \mathcal{B}, \mu)$ and almost everywhere pointwise convergence of (9.3.1). Note that L^2 convergence neither implies nor is implied by almost everywhere pointwise convergence.

Before stating these theorems, we first need to discuss conditional expectation.

§9.4 Conditional expectation

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

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

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

$$\nu(A) = \int_A E(f | \mathcal{A}) d\mu$$

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

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

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

For a complex-valued f we split f into its real and imaginary parts and define

$$E(f | \mathcal{A}) = E(\operatorname{Re}(f) | \mathcal{A}) + iE(\operatorname{Im}(f) | \mathcal{A}).$$

Thus we can view conditional expectation as an operator

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

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

- (i) $E(f | \mathcal{A})$ is \mathcal{A} -measurable, and
- (ii) $\int_A f d\mu = \int_A E(f | \mathcal{A}) d\mu$ for all $A \in \mathcal{A}$.

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

Let T be a measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . To state von Neumann's and Birkhoff's Ergodic Theorems precisely, we will need the sub- σ -algebra \mathcal{I} of T -invariant subsets, namely:

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

It is straightforward to check that \mathcal{I} is a σ -algebra. Note that if T is ergodic then \mathcal{I} is the trivial σ -algebra consisting of all sets in \mathcal{B} of measure 0 or 1.

§9.5 Von Neumann's Ergodic Theorem

Von Neumann's Ergodic Theorem deals with the L^2 -limiting behaviour of $\frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j$ for $f \in L^2(X, \mathcal{B}, \mu)$.

Theorem 9.5.1 (von Neumann's Ergodic Theorem)

Let (X, \mathcal{B}, μ) be a probability space and let $T : X \rightarrow X$ be a measure-preserving transformation. Let \mathcal{I} denote the σ -algebra of T -invariant sets. Then for every $f \in L^2(X, \mathcal{B}, \mu)$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j = E(f | \mathcal{I})$$

where the convergence is in L^2 .

When T is ergodic with respect to μ then von Neumann's Ergodic Theorem takes a particularly simple form.

Corollary 9.5.2

Let (X, \mathcal{B}, μ) be a probability space and let $T : X \rightarrow X$ be an ergodic measure-preserving transformation. Let $f \in L^2(X, \mathcal{B}, \mu)$. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j = \int f d\mu, \tag{9.5.1}$$

where the convergence is in L^2 .

Proof. If T is ergodic then \mathcal{I} is the trivial σ -algebra \mathcal{N} consisting of sets of measure 0 and 1. If $f \in L^2(X, \mathcal{B}, \mu)$ then $E(f | \mathcal{N}) = \int f d\mu$. □

Remark. The meaning of convergence in (9.5.1) is that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j - \int f d\mu \right\|_2 = 0$$

i.e.

$$\lim_{n \rightarrow \infty} \left(\int \left(\frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - \int f d\mu \right)^2 d\mu \right)^{1/2} = 0.$$

§9.6 Proof of von Neumann’s Ergodic Theorem

None of this section is examinable—it is included for people who like hard-core functional analysis!

We prove von Neumann’s Ergodic Theorem in the case where T is invertible.

In order to prove von Neumann’s Ergodic Theorem, it is useful to recast it in terms of linear analysis.

Theorem 9.6.1 (von Neumann’s Ergodic Theorem for Operators)

Let U be a unitary operator of a complex Hilbert space H . Let $I = \{v \in H \mid Uv = v\}$ be the closed subspace of U -invariant functions and let $P_I : H \rightarrow I$ be orthogonal projection onto I . Then for all $v \in H$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} U^j v = P_I v \tag{9.6.1}$$

in the norm induced on H by the inner product.

Proof of Theorem 9.6.1. Denote the inner product and norm on \mathcal{H} by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively.

First note that if $v \in I$ then (9.6.1) holds, as

$$\frac{1}{n} \sum_{j=0}^{n-1} U^j v = v = P_I v.$$

If $v = Uw - w$ for some $w \in H$ then

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} U^j v \right\| = \frac{1}{n} \|U^n w - w\| \leq \frac{1}{n} 2\|w\| \rightarrow 0$$

as $n \rightarrow \infty$. If we let C denote the norm-closure of the subspace $\{Uw - w \mid w \in H\}$ then it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} U^j v = 0$$

for all $v \in C$, by approximation.

We claim that $H = I \oplus C$, an orthogonal decomposition. Suppose that $v \perp C$. Then $\langle v, Uw - w \rangle = 0$ for all $w \in H$. Hence $\langle U^*v, w \rangle = \langle v, w \rangle$ for all $w \in H$. Hence $U^*v = v$. As U is unitary, we have that $U^* = U^{-1}$. Hence $v = Uv$, so that $v \in I$. Reversing each implication we see that $v \in I$ implies $v \perp C$, and the claim follows. \square

Remark. Note that an isometry of a Hilbert space H is a linear operator U such that $\langle Uv, Uw \rangle = \langle v, w \rangle$ for all $v, w \in H$. We say that U is unitary if, in addition, it is invertible. Equivalently, U is unitary if the dual operator U^* is the inverse of U : $U^*U = UU^* = \text{id}$.

We can prove von Neumann's Ergodic Theorem for an invertible measure-preserving transformation T of a probability space (X, \mathcal{B}, μ) as follows. Recall that $L^2(X, \mathcal{B}, \mu)$ is a Hilbert space with respect to the inner product

$$\langle f, g \rangle = \int f \bar{g} d\mu$$

and that T induces a linear operator $U : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{B}, \mu)$ by $Uf = f \circ T$. As T is measure-preserving, we have that U is an isometry; if T is invertible then U is unitary.

Let $P_{\mathcal{I}} : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{I}, \mu)$ denote the orthogonal projection onto the subspace of T -invariant functions. One can easily check (see Exercise 9.6) that $P_{\mathcal{I}}f = E(f | \mathcal{I})$.

Hence, when T is invertible, Theorem 9.5.1 follows immediately from Theorem 9.6.1.

One can deduce from Theorem 9.6.1 that the result continues to hold when U is an isometry and is not assumed to be invertible.

§9.7 Exercises

Exercise 9.1

Construct an example to show that Poincaré's recurrence theorem does not hold on infinite measure spaces. That is, find a measure space (X, \mathcal{B}, μ) with $\mu(X) = \infty$ and a measure-preserving transformation $T : X \rightarrow X$ such that the conclusion of Poincaré's Recurrence Theorem does not hold.

Exercise 9.2

Poincaré's Recurrence Theorem says that, if we have a measure-preserving transformation T of a probability space (X, \mathcal{B}, μ) and a set $A \in \mathcal{B}$, $\mu(A) > 0$, then, if we start iterating a typical point $x \in A$ then the orbit of x will return to A infinitely often.

Construct an example to show that if we have a measure-preserving transformation T of a probability space (X, \mathcal{B}, μ) and two sets $A, B \in \mathcal{B}$, $\mu(A), \mu(B) > 0$, then, if we start iterating a typical point $x \in A$ then the orbit of x does not necessarily visit B infinitely often.

Exercise 9.3

- (i) Prove that $f \mapsto E(f | \mathcal{A})$ is linear.
- (ii) Suppose that T is a measure-preserving transformation. Show that $E(f | \mathcal{A}) \circ T = E(f \circ T | T^{-1}\mathcal{A})$.
- (iii) Show that $E(f | \mathcal{B}) = f$.
- (iv) Let \mathcal{N} denote the trivial σ -algebra consisting of all sets of measure 0 and 1. Show that a function f is \mathcal{N} -measurable if and only if it is constant a.e. Show that $E(f | \mathcal{N}) = \int f d\mu$.

Exercise 9.4

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

- (i) Let $\alpha = \{A_1, \dots, A_n\}$, $A_j \in \mathcal{B}$ be a finite partition of X . (By a partition we mean that $X = \bigcup_{j=1}^n A_j$ and $A_i \cap A_j = \emptyset$ if $i \neq j$.) Let \mathcal{A} denote the set of all finite unions of sets in α . Check that \mathcal{A} is a σ -algebra.
- (ii) Show that $g : X \rightarrow \mathbb{R}$ is \mathcal{A} -measurable if and only if g is constant on each A_j , i.e.

$$g(x) = \sum_{j=1}^n c_j \chi_{A_j}(x).$$

- (iii) Let $f \in L^1(X, \mathcal{B}, \mu)$. Show that

$$E(f | \mathcal{A})(x) = \sum_{j=1}^r \chi_{A_j}(x) \frac{\int_{A_j} f d\mu}{\mu(A_j)}.$$

Thus $E(f | \mathcal{A})$ is the best approximation to f that is constant on sets in the partition α .

Exercise 9.5

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

Exercise 9.6

Let T be a measure-preserving transformation of the probability space (X, \mathcal{B}, μ) and let \mathcal{I} denote the sub- σ -algebra of T -invariant sets. Let $P_{\mathcal{I}} : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{I}, \mu)$ denote the orthogonal projection onto the subspace of T -invariant functions. Prove that $P_{\mathcal{I}}f = E(f | \mathcal{I})$ for all $f \in L^2(X, \mathcal{B}, \mu)$.

10. Birkhoff's Ergodic Theorem

§10.1 Birkhoff's Ergodic Theorem

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

Theorem 10.1.1 (Birkhoff's Ergodic Theorem)

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

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

for μ -a.e. $x \in X$.

Corollary 10.1.2 (Birkhoff's Ergodic Theorem for an ergodic transformation)

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu$$

for μ -a.e. $x \in X$.

§10.2 Consequences of, and criteria for, ergodicity

Here we give some simple corollaries of Birkhoff's Ergodic Theorem. The first result says that, for a typical orbit of an ergodic dynamical system, 'time averages' equal 'space averages'.

Corollary 10.2.1

Let T be an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . Suppose that $B \in \mathcal{B}$. Then for μ -a.e. $x \in X$, the frequency with which the orbit of x lies in B is given by $\mu(B)$, i.e.,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{card}\{j \in \{0, 1, \dots, n-1\} \mid T^j x \in B\} = \mu(B) \quad \mu\text{-a.e.}$$

Proof. Apply the Birkhoff Ergodic Theorem with $f = \chi_B$. □

It is possible to characterise ergodicity in terms of the behaviour of iteration of pre-images of sets, rather than the iteration points, under the dynamics. The next result deals with this.

Proposition 10.2.2

Let (X, \mathcal{B}, μ) be a probability space and let $T : X \rightarrow X$ be a measure-preserving transformation. The following are equivalent:

- (i) T is ergodic;
- (ii) for all $A, B \in \mathcal{B}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \mu(A)\mu(B).$$

Proof. We prove that (i) implies (ii). Suppose that T is ergodic. Since $\chi_A \in L^1(X, \mathcal{B}, \mu)$, Birkhoff's Ergodic Theorem tells us that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j x) = \mu(A)$$

for μ -a.e. $x \in X$. Multiplying both sides by $\chi_B(x)$ gives

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j x) \chi_B(x) = \mu(A)\chi_B(x)$$

for μ -a.e. $x \in X$. Since the left-hand side is bounded (by 1), we can apply the Dominated Convergence Theorem (Theorem 3.1.3) to see that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int \chi_A \circ T^j \chi_B d\mu = \lim_{n \rightarrow \infty} \int \frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ T^j \chi_B d\mu = \mu(A)\mu(B).$$

We prove that (ii) implies (i). Now suppose that the convergence holds. Suppose that $T^{-1}B = B$ and take $A = B$. Then $\mu(T^{-j}A \cap B) = \mu(B)$ so

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(B) = \mu(B)^2.$$

This gives $\mu(B) = \mu(B)^2$. Therefore $\mu(B) = 0$ or 1 and so T is ergodic. \square

§10.3 Kac's Lemma

Poincaré's Recurrence Theorem tells us that, under a measure-preserving transformation, almost every point of a subset A of positive measure will return to A . However, it does not tell us how long we should have to wait for this to happen. One would expect that return times to sets of large measure are small, and that return times to sets of small measure are large. This is indeed the case, and forms the content of Kac's Lemma.

Let $T : X \rightarrow X$ be a measure-preserving transformation of a probability space (X, \mathcal{B}, μ) and let $A \subset X$ be a measurable subset with $\mu(A) > 0$. By Poincaré's Recurrence Theorem, the integer

$$n_A(x) = \inf\{n \geq 1 \mid T^n(x) \in A\}$$

is defined for a.e. $x \in A$.

Theorem 10.3.1 (Kac's Lemma)

Let T be an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . Let $A \in \mathcal{B}$ be such that $\mu(A) > 0$. Then

$$\int_A n_A d\mu = 1.$$

Proof. Let

$$A_n = A \cap T^{-1}A^c \cap \dots \cap T^{-(n-1)}A^c \cap T^{-n}A.$$

Then A_n consists of those points in A that return to A after exactly n iterations of T , i.e. $A_n = \{x \in A \mid n_A(x) = n\}$.

Consider the illustration in Figure 10.3. As T is ergodic, almost every point of X

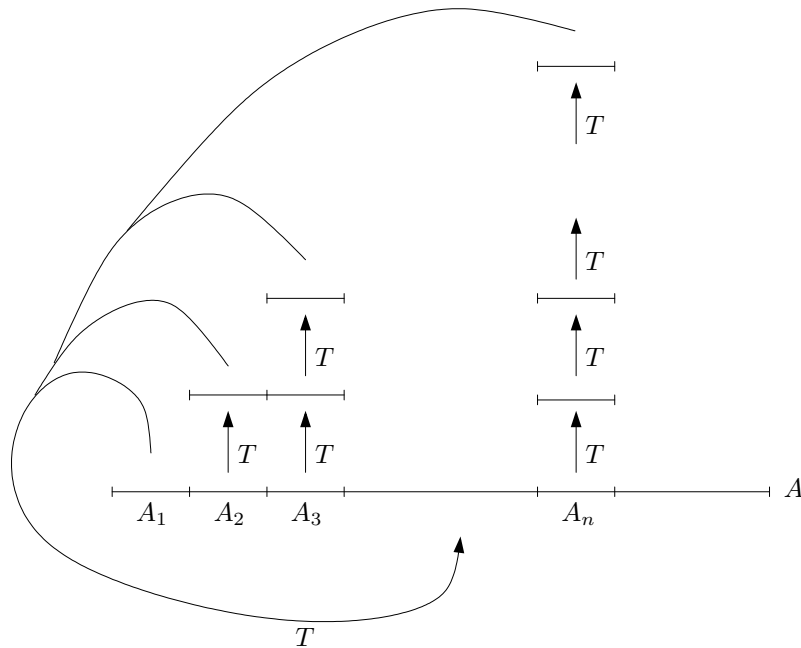


Figure 10.3.1: The return times to A

eventually enters A . Hence the diagram represents almost all of X . Note that the column above A_n in the diagram consists of n sets, $A_{n,0}, \dots, A_{n,n-1}$ say, with $A_{n,0} = A_n$. Note that $T^{-k}A_{n,k} = A_n$. As T is measure-preserving, it follows that $\mu(A_{n,k}) = \mu(A_n)$ for $k = 0, \dots, n - 1$. Hence

$$\begin{aligned} 1 &= \mu(X) = \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mu(A_{n,k}) = \sum_{n=1}^{\infty} n\mu(A_n) \\ &= \sum_{n=1}^{\infty} \int_{A_n} n_A d\mu = \int_A n_A d\mu. \end{aligned}$$

□

Remark. Let A be as in the statement of Kac's Lemma (Theorem 10.3.1). Define a probability measure μ_A on A by $\mu_A = \mu/\mu(A)$ so that $\mu_A(A) = 1$. Then Kac's Lemma says

that

$$\int_A n_A d\mu_A = \frac{1}{\mu(A)},$$

i.e. the expected return time of a point in A to the set A is $1/\mu(A)$.

§10.4 Ehrenfests' example

The following example, due to P. and T. Ehrenfest, demonstrates that the return times in Poincaré's Recurrence Theorem may be extremely large.

Consider two urns. One urn contains 100 balls, numbered 1 to 100, and the other urn is empty. We also have a random number generator: this could be a bag containing 100 slips of paper, numbered 1 to 100.

Each second, a slip of paper is drawn from the bag, the number is noted, and the slip of paper is returned to the bag. The ball bearing that number is then moved from whichever urn it is currently in to the other urn.

Naively, we would expect that the system will settle into an equilibrium state in which there are 50 balls in each urn. Of course, there will continue to be small random fluctuations about the 50-50 distribution. However, it would appear highly unlikely for the system to return to the state in which 100 balls are in the first urn. Nevertheless, the Poincaré Recurrence Theorem tells us that this situation will occur almost surely and Kac's Lemma tells us how long we should expect to wait.

To see this, we represent the system as a shift on 101 symbols with an appropriate Markov measure. Regard $x_j \in \{0, \dots, 100\}$ as being the number of balls in the first urn after j seconds. Hence a sequence $(x_j)_{j=0}^{\infty}$ records the number of balls in the first urn at each time. Let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^{\infty} \mid x_j \in \{0, 1, \dots, 100\}\}$.

Let $p(i)$ denote the probability of there being i balls in the first urn. This is equal to the number of possible ways of choosing i balls from 100, divided by the total number of ways of distributing 100 balls across the 2 urns. There are $\binom{100}{i}$ ways of choosing i balls from 100 balls. As there are 2 possible urns for each ball to be in, there are 2^{100} possible arrangements of all the balls. Hence the probability of there being i balls in the first urn is

$$p(i) = \frac{1}{2^{100}} \binom{100}{i}.$$

If we have i balls in the first urn then at the next stage we must have either $i-1$ or $i+1$ balls in the first urn. The number of balls becomes $i-1$ if the random number chosen is equal to the number of one of the balls in the first urn. As there are currently i such balls, the probability of this happening is $i/100$. Hence the probability $P(i, i-1)$ that there are $i-1$ balls remaining given that we started with i balls in the first urn is $i/100$. Similarly, the probability $P(i, i+1)$ that there are $i+1$ balls in the first urn given that we started with i balls is $(100-i)/100$. If $j \neq i-1, i+1$ then we cannot have j balls in the first urn given that we started with i balls; thus $P(i, j) = 0$. This defines a stochastic matrix:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ \frac{1}{100} & 0 & \frac{99}{100} & 0 & 0 & \cdots \\ 0 & \frac{2}{100} & 0 & \frac{98}{100} & 0 & \cdots \\ 0 & 0 & \frac{3}{100} & 0 & \frac{97}{100} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

It is straightforward to check that $pP = p$. Hence we have a Markov probability measure μ_P defined on Σ . The matrix P is irreducible (but is not aperiodic); this ensures that μ_P is ergodic.

Consider the cylinder $A = [100]$ of length 1. This represents there being 100 balls in the first urn. By Poincaré's Recurrence Theorem, if we start in A then we return to A infinitely often. Thus, with probability 1, we will return to the situation where all 100 balls have returned to the first urn—and this will happen infinitely often! We can use Kac's Lemma to calculate the expected amount of time we will have to wait until all the balls first return to the first urn. By Kac's lemma, the expected first return time to A is

$$\frac{1}{\mu_P(A)} = 2^{100} \text{ seconds,}$$

which is about 4×10^{22} years, or about 3×10^{12} times the length of time that the Universe has so far existed!

(This measure-preserving transformation system, with 4 balls rather than 100, was also studied in Exercise 3.11.)

§10.5 Proof of Birkhoff's Ergodic Theorem

None of this section is examinable—it is included for people who like hard-core ε - δ analysis!

The proof is based on the following inequality.

Theorem 10.5.1 (Maximal Inequality)

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

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

For $n \geq 1$, set $F_n(x) = \max_{0 \leq j \leq n} f_j(x)$ so that $F_n(x) \geq 0$. Then

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

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

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

and therefore

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

If $F_n(x) > 0$ then

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

so we obtain that

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

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

Hence

$$\begin{aligned}
\int_A f d\mu &\geq \int_A F_n d\mu - \int_A F_n \circ T d\mu \\
&= \int_X F_n d\mu - \int_A F_n \circ T d\mu \text{ as } F_n = 0 \text{ on } X \setminus A \\
&\geq \int_X F_n d\mu - \int_X F_n \circ T d\mu \text{ as } F_n \circ T \geq 0 \\
&= 0 \text{ as } \mu \text{ is } T\text{-invariant.}
\end{aligned}$$

□

Corollary 10.5.2

Let $g \in L^1(X, \mathcal{B}, \mu)$ and let

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

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

$$\int_{M_\alpha \cap B} g d\mu \geq \alpha \mu(M_\alpha \cap B).$$

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

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

(since $f_n(x) > 0 \Rightarrow F_n(x) > 0$ and $F_n(x) > 0 \Rightarrow f_j(x) > 0$ for some $1 \leq j \leq n$). Write $C_n = \{x \mid F_n(x) > 0\}$ and observe that $C_n \subset C_{n+1}$. Thus χ_{C_n} converges to χ_{M_α} and so $f\chi_{C_n}$ converges to $f\chi_{M_\alpha}$, as $n \rightarrow \infty$. Furthermore, $|f\chi_{C_n}| \leq |f|$. Hence, by the Dominated Convergence Theorem,

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

Applying the Maximal Inequality, we have for all $n \geq 1$ that $\int_{C_n} f d\mu \geq 0$. Therefore $\int_{M_\alpha} f d\mu \geq 0$, i.e., $\int_{M_\alpha} g d\mu \geq \alpha \mu(M_\alpha)$.

For the general case, we work with the restriction of T to B , $T|_B : B \rightarrow B$, and apply the Maximal Inequality on this subset to get

$$\int_{M_\alpha \cap B} g d\mu \geq \alpha \mu(M_\alpha \cap B),$$

as required. □

We will also need the following convergence result.

Proposition 10.5.3 (Fatou's Lemma)

Let (X, \mathcal{B}, μ) be a probability space and suppose that $f_n : X \rightarrow \mathbb{R}$ are measurable functions. Define $f(x) = \liminf_{n \rightarrow \infty} f_n(x)$. Then f is measurable and

$$\int f d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu$$

(one or both of these expressions may be infinite).

Proof of Birkhoff's Ergodic Theorem. Let

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

These exist (but may be $\pm\infty$, respectively) at all points $x \in X$. Clearly $f_*(x) \leq f^*(x)$.

Let

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

Observe that

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

As f is finite μ -a.e., we have that $f(x)/n \rightarrow 0$ μ -a.e. as $n \rightarrow \infty$. Hence, taking the lim sup and lim inf as $n \rightarrow \infty$, gives us that $f^* \circ T = f^*$ μ -a.e. and $f_* \circ T = f_*$ μ -a.e.

We have to show

- (i) $f^* = f_*$ μ -a.e
- (ii) $f^* \in L^1(X, \mathcal{B}, \mu)$
- (iii) $\int f^* d\mu = \int f d\mu$.

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

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

Note that

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

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

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

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

Applying Corollary 10.5.2 we have that

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

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

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

Therefore

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

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

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

We prove (ii). Let

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

Then $g_n \geq 0$ and

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

so we can apply Fatou's Lemma (Proposition 10.5.3) to conclude that $\lim_{n \rightarrow \infty} g_n = |f^*|$ is integrable, i.e., that $f^* \in L^1(X, \mathcal{B}, \mu)$.

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

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

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

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

Since $T^{-1}D_k^n = D_k^n$, we can apply Corollary 10.5.2 again to obtain

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

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

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

Thus

$$\int_{D_k^n} f^* d\mu \leq \frac{k+1}{n} \mu(D_k^n) \leq \frac{1}{n} \mu(D_k^n) + \int_{D_k^n} f d\mu$$

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

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

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

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

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

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

Applying the same argument to $-f$ gives

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

so that

$$\int f^* d\mu = \int f_* d\mu \geq \int f d\mu.$$

Therefore

$$\int f^* d\mu = \int f d\mu,$$

as required.

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

$$\int_I f d\mu = \int_I f^* d\mu.$$

Hence $f^* = E(f | \mathcal{I})$. □

§10.6 Exercises

Exercise 10.1

Suppose that T is an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) and suppose that $f \in L^1(X, \mathcal{B}, \mu)$. Prove that

$$\lim_{n \rightarrow \infty} \frac{f(T^n x)}{n} = 0 \text{ } \mu\text{-a.e.}$$

Exercise 10.2

Deduce from Birkhoff's Ergodic Theorem that if T is an ergodic measure-preserving transformation of a probability space (X, \mathcal{B}, μ) and $f \geq 0$ is measurable but $\int f d\mu = \infty$ then

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

(Hint: define $f_M = \min\{f, M\}$ and note that $f_M \in L^1(X, \mathcal{B}, \mu)$. Apply Birkhoff's Ergodic Theorem to each f_M .)

Exercise 10.3

Let T be a measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . Prove that the following are equivalent:

- (i) T is ergodic with respect to μ ,
- (ii) for all $f, g \in L^2(X, \mathcal{B}, \mu)$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int f(T^j x) g(x) d\mu = \int f d\mu \int g d\mu.$$

Exercise 10.4

Let X be a compact metric space equipped with the Borel σ -algebra \mathcal{B} and let $T : X \rightarrow X$ be continuous. Suppose that $\mu \in M(X)$ is an ergodic measure.

Prove that there exists a set $Y \in \mathcal{B}$ with $\mu(Y) = 1$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu$$

for all $x \in Y$ and for all $f \in C(X, \mathbb{R})$.

(Thus, in the special case of a continuous transformation of a compact metric space and continuous functions f , the set of full measure for which Corollary 10.1.2 holds can be chosen to be independent of the function f .)

Exercise 10.5

A popular illustration of recurrence concerns a monkey typing the complete works of Shakespeare on a typewriter. Here we study this from an ergodic-theoretic viewpoint.

Imagine a(n idealised) monkey typing on a typewriter. Each second he types one letter, and each letter occurs with equal probability (independently of the preceding letter). Suppose that the keyboard has 26 keys (so no space bar, carriage return, numbers, etc). Show how to model this using a shift space on 26 symbols with an appropriate Bernoulli measure. Use Birkhoff's Ergodic Theorem to show that the monkey must, with probability 1, eventually type the word 'MONKEY'. Use Kac's Lemma to calculate the expected time it would take for the monkey to first type 'MONKEY'.

11. Applications of Birkhoff's Ergodic Theorem

§11.1 Introduction

We will show how to use Birkhoff's Ergodic Theorem to prove some interesting results in number theory.

§11.2 Normal and simply normal numbers

Recall that any number $x \in [0, 1]$ can be written as a decimal

$$x = .x_0x_1x_2\dots = \sum_{j=0}^{\infty} \frac{x_j}{10^{j+1}}$$

where $x_j \in \{0, 1, \dots, 9\}$. This decimal expansion is unique unless the decimal expansion ends in either infinitely repeated 0s or infinitely repeated 9s.

More generally, given any integer base $b \geq 2$, we can write $x \in [0, 1]$ as a base b expansion:

$$x = .x_0x_1x_2\dots = \sum_{j=0}^{\infty} \frac{x_j}{b^{j+1}}$$

where $x_j \in \{0, 1, \dots, b-1\}$. This expansion is unique unless it ends in either infinitely repeated 0s or infinitely repeated $(b-1)$ s. In what follows, if x has two expansions in base b then we always choose the expansion that ends in infinitely repeated 0s; note that this is a countable set and so has Lebesgue measure zero.

Definition. A number $x \in [0, 1]$ is said to be *simply normal in base b* if for each $k = 0, 1, \dots, b-1$, the frequency with which digit k occurs in the base b expansion of x is equal to $1/b$.

Remarks.

1. Thus a number is simply normal in base b if all of the b possible digits in its base b expansion are equally likely to occur.
2. It is straightforward to construct examples of simply normal numbers in a given base. For example,

$$x = .012\dots 89012\dots 89\dots \tag{11.2.1}$$

consisting of the block of decimal digits $012\dots 89$ infinitely repeated is simply normal in base 10. If a number is simply normal in one base then it need not be simply normal in any other base. For example, x as defined in (11.2.1) is not simply normal in base 10^{10} .

Fix $b \geq 2$. Define the map $T : [0, 1] \rightarrow [0, 1]$ by $T(x) = T_b(x) = bx \bmod 1$. It is easy to see, by following any of the arguments we have seen for the doubling map, that Lebesgue measure μ on $[0, 1]$ is an ergodic invariant measure for T .

There is a close connection between the map T_b and base b expansions. Note that if $x \in [0, 1]$ has base b expansion

$$x = \sum_{j=0}^{\infty} \frac{x_j}{b^{j+1}} = \cdot x_0 x_1 x_2 \cdots$$

then

$$T_b(x) = b \sum_{j=0}^{\infty} \frac{x_j}{b^{j+1}} \bmod 1 = x_0 + \sum_{j=1}^{\infty} \frac{x_j}{b^j} \bmod 1 = \sum_{j=0}^{\infty} \frac{x_{j+1}}{b^{j+1}} = \cdot x_1 x_2 x_3 \cdots.$$

Thus T_b acts on base b expansions by deleting the zeroth term and then shifting the remaining digits one place to the left. This relationship between base b expansions and the map T_b can be used to prove the following result.

Proposition 11.2.1

Let $b \geq 2$. Then Lebesgue almost every real number in $[0, 1]$ is simply normal in base b .

Proof. Fix $k \in \{0, 1, \dots, b-1\}$. Note that $x_0 = k$ if and only if $x \in [k/b, (k+1)/b)$. Hence $x_j = k$ if and only if $T_b^j(x) \in [k/b, (k+1)/b)$. Thus

$$\frac{1}{n} \text{card}\{0 \leq j \leq n-1 \mid x_j = k\} = \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[k/b, (k+1)/b)}(T^j x). \quad (11.2.2)$$

By Birkhoff's Ergodic Theorem, for Lebesgue almost every point x the above expression converges to $\int \chi_{[k/b, (k+1)/b)}(x) dx = 1/b$. Let $X_b(k)$ denote the set of points $x \in [0, 1]$ for which (11.2.2) converges. Then $\mu(X_b(k)) = 1$ for each $k = 0, 1, \dots, b-1$.

Let $X_b = \bigcap_{k=0}^{b-1} X_b(k)$. Let μ denote Lebesgue measure. Then $\mu(X_b) = 1$. If $x \in X_b$ then the frequency with which each digit $k \in \{0, 1, \dots, b-1\}$ occurs in the base b expansion of x is equal to $1/b$, i.e. x is simply normal in base b . \square

We can consider a more general notion of normality of numbers as follows. Take $x \in [0, 1]$ and write x as a base b expansion

$$x = \cdot x_0 x_1 x_2 \dots = \sum_{j=0}^{\infty} \frac{x_j}{b^{j+1}}$$

where $x_j \in \{0, 1, \dots, b-1\}$. Fix a finite word of symbols i_0, i_1, \dots, i_{k-1} where $i_j \in \{0, 1, \dots, b-1\}$, $j = 0, \dots, k-1$. We can ask what is the frequency with which the block of symbols i_0, i_1, \dots, i_{k-1} occurs in the base b expansion of x . We call a block a symbols i_0, i_1, \dots, i_{k-1} a *word of length k* . Note that x has a base b expansion that starts $i_0 i_1 \cdots i_{k-1}$ precisely when

$$x \in \left[\sum_{j=0}^{k-1} \frac{i_j}{b^{j+1}}, \sum_{j=0}^{k-1} \frac{i_j}{b^{j+1}} + \frac{1}{b^k} \right).$$

Call this interval $I(i_0, \dots, i_{k-1})$ and note that it has Lebesgue measure $1/b^k$.

Definition. A number $x \in [0, 1]$ is *simply normal* if it is simply normal in base b for all $b \geq 2$.

Proposition 11.2.2

Lebesgue almost every real number $x \in [0, 1]$ is simply normal.

Proof. See Exercise 11.2(i). □

Definition. A number $x \in [0, 1]$ is said to be *normal in base b* if, for each $k \geq 1$ and for each word i_0, i_1, \dots, i_{k-1} of length k , the frequency with which this word occurs in the base b expansion of x is equal to $1/b^k$.

Proposition 11.2.3

Let $b \geq 2$ be an integer. Lebesgue almost every real number in $[0, 1]$ is normal in base b .

Proof. Fix a word i_0, i_1, \dots, i_{k-1} of length k and define the interval $I(i_0, \dots, i_{k-1})$ as above. Then the word i_0, i_1, \dots, i_{k-1} occurs at the j th place in the base b expansion of x if and only if $T_b^j(x) \in I(i_0, \dots, i_{k-1})$. Thus

$$\begin{aligned} & \frac{1}{n} \text{card}\{0 \leq j \leq n-1 \mid i_0, i_1, \dots, i_{k-1} \text{ occurs at the } j\text{th place in the base } b \text{ expansion of } x\} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \chi_{I(i_0, \dots, i_{k-1})}(T_b^j x). \end{aligned} \quad (11.2.3)$$

By Birkhoff's Ergodic Theorem, for Lebesgue almost every point x the above expression converges to $\int \chi_{I(i_0, \dots, i_{k-1})}(x) dx = 1/b^k$. Let $X_b(i_0, i_1, \dots, i_{k-1})$ denote the set of points $x \in [0, 1]$ for which (11.2.3) converges. Let μ denote Lebesgue measure. Then $\mu(X_b(i_0, i_1, \dots, i_{k-1})) = 1$ for each word i_0, i_1, \dots, i_{k-1} of length k .

Let

$$X_b = \bigcap_{k=1}^{\infty} \bigcap_{i_0, i_1, \dots, i_{k-1}} X_b(i_0, i_1, \dots, i_{k-1})$$

where the second intersection is taken over all words of length k . As this is a countable intersection, we have that $\mu(X_b) = 1$. If $x \in X_b$ then the frequency with which any word of length k occurs in the base b expansion of x is equal to $1/b^k$, i.e. x is normal in base b . □

We can then make the following definition.

Definition. A number $x \in [0, 1]$ is *normal* if it is normal in base b for every base $b \geq 2$.

We can now prove the following result:

Proposition 11.2.4

Lebesgue almost every number $x \in [0, 1]$ is normal.

Proof. See Exercise 11.2(ii). □

Remark. Although a 'typical' number is normal, there are no known examples of normal numbers!

§11.3 Continued fractions

We can use Birkhoff's Ergodic Theorem to study the frequency with which a given digit occurs in the continued fraction expansion of real numbers.

Proposition 11.3.1

For Lebesgue-almost every $x \in [0, 1]$, the frequency with which the natural number k occurs in the continued fraction expansion of x is

$$\frac{1}{\log 2} \log \left(\frac{(k+1)^2}{k(k+2)} \right).$$

Proof. Let λ denote Lebesgue measure and let μ denote Gauss' measure. Recall that λ and μ are equivalent, i.e. they have the same sets of measure zero. Then λ -a.e. and μ -a.e. $x \in [0, 1]$ is irrational and has an infinite continued fraction expansion

$$x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \cdots}}}. \quad (11.3.1)$$

Let T denote the continued fraction map. Then

$$T(x) = \frac{1}{x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \cdots}}}$$

so that

$$\frac{1}{T(x)} = x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \cdots}}.$$

Hence $x_1 = [1/T(x)]$, where $[x]$ denotes the integer part of x . More generally, $x_n = [1/T^n x]$.

Fix $k \in \mathbb{N}$. Note that x has a continued fraction expansion starting with digit k (i.e. $x_0 = k$) precisely when $[1/x] = k$. That is, $x_0 = k$ precisely when

$$k \leq \frac{1}{x} < k+1$$

which is equivalent to requiring

$$\frac{1}{k+1} < x \leq \frac{1}{k}$$

i.e. $x \in (1/(k+1), 1/k]$. Similarly $x_n = k$ precisely when $T^n x \in (1/(k+1), 1/k]$.

Hence

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \text{card}\{0 \leq j \leq n-1 \mid x_j = k\} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{(1/(k+1), 1/k]}(T^j x) \\ &= \int \chi_{(1/(k+1), 1/k]} d\mu \text{ for } \mu\text{-a.e. } x \\ &= \frac{1}{\log 2} \left[\log \left(1 + \frac{1}{k} \right) - \log \left(1 + \frac{1}{k+1} \right) \right] \text{ for } \mu\text{-a.e. } x \\ &= \frac{1}{\log 2} \log \frac{(k+1)^2}{k(k+2)} \text{ for } \mu\text{-a.e. } x. \end{aligned}$$

As μ and λ have the same sets of measure zero, this holds for Lebesgue almost every point. \square

We can also study the limiting arithmetic and geometric means of the digits in the continued fraction expansion of Lebesgue almost every point $x \in [0, 1]$.

Proposition 11.3.2

(i) For Lebesgue-almost every $x \in [0, 1]$, the limiting arithmetic mean of the digits in the continued fraction expansion of x is infinite. More specifically, for Lebesgue almost every $x \in [0, 1]$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} (x_0 + x_1 + \dots + x_{n-1}) = \infty$$

where x has continued fraction expansion given by (11.3.1).

(ii) For Lebesgue-almost every $x \in [0, 1]$, the limiting geometric mean of the digits in the continued fraction expansion of x is

$$\prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2 + 2k} \right)^{\log k / \log 2}.$$

More specifically, for Lebesgue almost every $x \in [0, 1]$ we have

$$\lim_{n \rightarrow \infty} (x_0 x_1 \dots x_{n-1})^{1/n} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2 + 2k} \right)^{\log k / \log 2}.$$

where x has continued fraction expansion given by (11.3.1).

Proof. Writing

$$x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \dots}}}$$

the proposition claims that

$$\lim_{n \rightarrow \infty} \frac{1}{n} (x_0 + x_1 + \dots + x_{n-1}) = \infty \tag{11.3.2}$$

for Lebesgue almost every point, and that

$$\lim_{n \rightarrow \infty} (x_0 x_1 \dots x_{n-1})^{1/n} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2 + 2k} \right)^{\log k / \log 2} \tag{11.3.3}$$

for Lebesgue almost every point.

We leave (11.3.2) as an exercise.

We prove (11.3.3). Define $f(x) = \log k$ for $x \in (1/(k+1), 1/k]$ so that $f(x) = \log k$ precisely when $x_0 = k$. Then $f(T^j x) = \log k$ precisely when $x_j = k$. By Exercise 3.5(iii),

to show $f \in L^1(X, \mathcal{B}, \mu)$ it is sufficient to show that $f \in L^1(X, \mathcal{B}, \lambda)$. Note that

$$\begin{aligned} \int f d\lambda &= \sum_{k=1}^{\infty} \log k \lambda \left(\left(\frac{1}{k+1}, \frac{1}{k} \right] \right) \\ &= \sum_{k=1}^{\infty} \frac{\log k}{k(k+1)} \\ &\leq \sum_{k=1}^{\infty} \frac{\log k}{k^2}, \end{aligned}$$

which converges. Hence $f \in L^1(X, \mathcal{B}, \mu)$.

Now

$$\begin{aligned} \lim_{n \rightarrow \infty} \log(x_0 x_1 \cdots x_{n-1})^{1/n} &= \lim_{n \rightarrow \infty} \frac{1}{n} (\log x_0 + \log x_1 + \cdots + \log x_{n-1}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \\ &= \frac{1}{\log 2} \int_0^1 \frac{f(x)}{1+x} dx \\ &= \frac{1}{\log 2} \sum_{k=1}^{\infty} \int_{1/(k+1)}^{1/k} \frac{\log k}{1+x} dx \\ &= \sum_{k=1}^{\infty} \frac{\log k}{\log 2} \log \left(1 + \frac{1}{k^2 + 2k} \right) \end{aligned}$$

as $n \rightarrow \infty$, for Gauss-almost every point $x \in [0, 1]$. As Gauss' measure and Lebesgue measure have the same sets of measure zero, this limit also exists for Lebesgue almost every point. Exponentiating both sides of the above gives the result. \square

Let $x \in [0, 1]$ be irrational and have continued fraction expansion $[x_0, x_1, \dots]$. Then $[x_0, x_1, \dots, x_{n-1}]$ is a rational number; write $[x_0, x_1, \dots, x_{n-1}] = P_n(x)/Q_n(x)$, where $P_n(x), Q_n(x)$ are co-prime integers. Then $P_n(x)/Q_n(x)$ is a 'good' rational approximation to x . We write $P_n(x), Q_n(x)$ if we wish to indicate the dependence on x . As x and $P_n(x)/Q_n(x)$ lie in the same cylinder $I(x_0, \dots, x_{n-1})$ of rank n , we must have that

$$\left| x - \frac{P_n(x)}{Q_n(x)} \right| \leq \text{diam } I(x_0, \dots, x_{n-1}) \leq \frac{1}{Q_n(x)^2},$$

where $\text{diam } I$ denotes the length of the interval I . Thus we can quantify how good a rational approximation $P_n(x)/Q_n(x)$ is to x by looking at the denominator $Q_n(x)$. Thus understanding how $Q_n(x)$ grows gives us information about x . For a typical point, $Q_n(x)$ grows exponentially fast and we can determine the exponential growth rate.

Proposition 11.3.3

For Lebesgue almost every real number $x \in [0, 1]$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log Q_n(x) = \frac{\pi^2}{12 \log 2}.$$

Remark. Thus, for a typical point $x \in [0, 1]$, we have that $Q_n(x) \sim e^{n\pi^2/12 \log 2}$.

Proof (not examinable). Let $x \in [0, 1]$ be irrational and have continued fraction expansion $[x_0, x_1, \dots]$. Write

$$[x_0, x_1, \dots, x_{n-1}] = \frac{P_n(x)}{Q_n(x)}.$$

Then

$$\frac{P_n(x)}{Q_n(x)} = \frac{1}{x_0 + [x_1, \dots, x_{n-1}]} = \frac{1}{x_0 + \frac{P_{n-1}(Tx)}{Q_{n-1}(Tx)}} = \frac{Q_{n-1}(Tx)}{P_{n-1}(Tx) + x_0 Q_{n-1}(Tx)}. \quad (11.3.4)$$

By Lemma 6.3.1(ii) and the Euclidean algorithm, we know that for all n and all x , $P_n(x)$ and $Q_n(x)$ are coprime. As $P_{n-1}(Tx)$ and $Q_{n-1}(Tx)$ are coprime, it follows that $P_{n-1}(Tx) + x_0 Q_{n-1}(Tx)$ and $Q_{n-1}(Tx)$ are coprime. Hence, comparing the numerators in (11.3.4), we see that $P_n(x) = Q_{n-1}(Tx)$. Also note that $P_1(x) = 1$. Hence

$$\frac{P_n(x)}{Q_n(x)} \frac{P_{n-1}(Tx)}{Q_{n-1}(Tx)} \dots \frac{P_1(T^{n-1}x)}{Q_1(T^{n-1}x)} = \frac{P_1(T^{n-1}x)}{Q_n(x)} = \frac{1}{Q_n(x)}.$$

Taking the logarithm and dividing by n gives that

$$-\frac{1}{n} \log Q_n(x) = \frac{1}{n} \log \prod_{j=0}^{n-1} \frac{P_{n-j}(T^j x)}{Q_{n-j}(T^j x)} = \frac{1}{n} \sum_{j=0}^{n-1} \log \frac{P_{n-j}(T^j x)}{Q_{n-j}(T^j x)}. \quad (11.3.5)$$

This resembles an ergodic sum, except that the function P_{n-j}/Q_{n-j} depends on j and so we cannot immediately apply Birkhoff's Ergodic Theorem. We will consider ergodic sums using the function $f(x) = \log x$ and show that the difference between $\frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$ and (11.3.5) is small.

Let $f(x) = \log x$. Then we can write (11.3.5) as

$$-\frac{1}{n} \log Q_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - \frac{1}{n} \sum_{j=0}^{n-1} \left(\log T^j(x) - \log \frac{P_{n-j}(T^j x)}{Q_{n-j}(T^j x)} \right) = \frac{1}{n} \Sigma_n^{(1)} - \frac{1}{n} \Sigma_n^{(2)}.$$

We evaluate $\lim_{n \rightarrow \infty} \frac{1}{n} \Sigma_n^{(1)}$. By Birkhoff's Ergodic Theorem and the fact that Gauss' measure μ and Lebesgue measure are equivalent, it follows that for Lebesgue almost every $x \in [0, 1]$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Sigma_n^{(1)} = \frac{1}{\log 2} \int \frac{f(x)}{1+x} dx = \frac{1}{\log 2} \int_0^1 \frac{\log x}{1+x} dx.$$

Integrating by parts we have that

$$\int_0^1 \frac{\log x}{1+x} dx = \log x \log(1+x) \Big|_0^1 - \int_0^1 \frac{\log(1+x)}{x} dx = - \int_0^1 \frac{\log(1+x)}{x} dx.$$

The Taylor series expansion of $\log(1+x)$ about zero is

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} x^k}{k}$$

so that

$$\frac{\log(1+x)}{x} = \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k+1}.$$

Hence for almost every x ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Sigma_n^{(1)} = -\frac{1}{\log 2} \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} \int_0^1 x^k dx = -\frac{1}{\log 2} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)^2}.$$

Note that

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} - 2 \sum_{n=1}^{\infty} \frac{1}{(2n)^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} - \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{12},$$

using the well-known fact that $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. Hence for almost every x ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Sigma_n^{(1)} = -\frac{\pi^2}{12 \log 2}.$$

It remains to show that $\frac{1}{n} \Sigma_n^{(2)} \rightarrow 0$ as $n \rightarrow \infty$. Recall that in §6.3 we introduced the cylinder set $I(x_0, x_1, \dots, x_{n-1})$ of rank n to denote the set of points x with continued fraction expansion that starts x_0, \dots, x_{n-1} . We proved in §6.3 that $I(x_0, x_1, \dots, x_{n-1})$ is an interval with length at most $1/Q_n(x)^2$. Note that both x and $P_n(x)/Q_n(x)$ lie in the same interval of rank n . Hence

$$\left| \frac{x}{P_n(x)/Q_n(x)} - 1 \right| = \frac{Q_n(x)}{P_n(x)} \left| x - \frac{P_n(x)}{Q_n(x)} \right| \leq \frac{Q_n(x)}{P_n(x)} \frac{1}{Q_n(x)^2} = \frac{1}{P_n(x)Q_n(x)}.$$

It follows from Lemma 6.3.1(i) that $P_n(x) \geq 2^{(n-2)/2}$ and $Q_n(x) \geq 2^{(n-1)/2}$. Hence

$$\left| \frac{x}{P_n(x)/Q_n(x)} - 1 \right| \leq \frac{1}{2^{n-3/2}}.$$

By the triangle inequality and the fact that $\log y \leq y - 1$ we have that

$$\begin{aligned} \left| \Sigma_2^{(n)} \right| &\leq \sum_{j=0}^{n-1} \left| \log \left(\frac{T^j(x)}{P_{n-j}(T^j(x))/Q_{n-j}(T^j(x))} \right) \right| \\ &\leq \sum_{j=0}^{n-1} \left| \frac{T^j(x)}{P_{n-j}(T^j(x))/Q_{n-j}(T^j(x))} - 1 \right| \\ &\leq \sum_{j=0}^{n-1} \frac{1}{2^{n-j-\frac{3}{2}}}. \end{aligned}$$

Note that

$$\sum_{j=0}^{n-1} \frac{1}{2^{n-j-\frac{3}{2}}} = \frac{1}{2^{-3/2}} \sum_{j=0}^{n-1} \frac{1}{2^j} \leq \frac{1}{2^{-3/2}} \sum_{j=0}^{\infty} \frac{1}{2^j} = 2^{5/2}.$$

Hence $\Sigma_n^{(2)} \leq 2^{5/2}$ for all n . Hence

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Sigma_n^{(2)} = 0$$

and the result follows. \square

§11.4 Exercises

Exercise 11.1

Let $b \geq 2$ be an integer. Prove that Lebesgue measure is an ergodic invariant measure for $T_b(x) = bx \bmod 1$ defined on the unit interval.

Exercise 11.2

(i) A number $x \in [0, 1]$ is said to be *simply normal* if it is simply normal in base b for all $b \geq 2$. Prove that Lebesgue a.e. number $x \in [0, 1]$ is simply normal.

(ii) Prove Proposition 11.2.4.

Exercise 11.3

Let $r \geq 2$ be an integer. Prove that for Lebesgue almost every $x \in [0, 1]$, the sequence $x_n = r^n x$ is uniformly distributed mod 1.

Exercise 11.4

Prove that the arithmetic mean of the digits appearing in the base 10 expansion of Lebesgue-a.e. $x \in [0, 1)$ is equal to 4.5, i.e. prove that if $x = \sum_{j=0}^{\infty} x_j/10^{j+1}$, $x_j \in \{0, 1, \dots, 9\}$ then

$$\lim_{n \rightarrow \infty} \frac{1}{n} (x_0 + x_1 + \dots + x_{n-1}) = 4.5 \text{ a.e.}$$

Exercise 11.5

Let $x \in [0, 1]$ have continued fraction expansion $x = [x_0, x_1, x_2, \dots]$.

Prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} (x_0 + x_1 + \dots + x_{n-1}) = \infty$$

for Lebesgue almost every $x \in [0, 1]$. (Hint: use Exercise 10.2.)

12. Solutions to the Exercises

Solution 1.1

Suppose that $x_n \in \mathbb{R}$ is uniformly distributed mod 1. Let $x \in [0, 1]$ and let $\varepsilon > 0$. We want to show that there exists n such that $\{x_n\} \in (x - \varepsilon, x + \varepsilon) \cap [0, 1]$ (as usual, $\{x_n\}$ denotes the fractional part of x_n).

By the definition of uniform distribution mod 1 we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{card}\{j \mid 0 \leq j \leq n-1, \{x_j\} \in (x - \varepsilon, x + \varepsilon)\} = 2\varepsilon.$$

Then there exists n_0 such that if $n \geq n_0$ then

$$\frac{1}{n} \text{card}\{j \mid 0 \leq j \leq n-1, \{x_j\} \in (x - \varepsilon, x + \varepsilon)\} > \varepsilon > 0.$$

Hence

$$\text{card}\{j \mid 0 \leq j \leq n-1, \{x_j\} \in (x - \varepsilon, x + \varepsilon)\} > 0$$

for some n , so there exists j such that $\{x_j\} \in (x - \varepsilon, x + \varepsilon)$.

Solution 1.2

We use Weyl's Criterion. Let $\ell \in \mathbb{Z} \setminus \{0\}$. Then

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} = \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell (\alpha j + \beta)} = \frac{1}{n} e^{2\pi i \ell \beta} \sum_{j=0}^{n-1} e^{2\pi i \ell \alpha j} = \frac{1}{n} e^{2\pi i \ell \beta} \left(\frac{e^{2\pi i \ell \alpha n} - 1}{e^{2\pi i \ell \alpha} - 1} \right),$$

summing the geometric progression. As $\alpha \notin \mathbb{Q}$, we have that $e^{2\pi i \ell \alpha} \neq 1$ for any $\ell \in \mathbb{Z} \setminus \{0\}$.

Hence

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} \right| \leq \frac{1}{n} \left| \frac{e^{2\pi i \ell \alpha n} - 1}{e^{2\pi i \ell \alpha} - 1} \right| \leq \frac{1}{n} \frac{2}{|e^{2\pi i \ell \alpha} - 1|} \rightarrow 0$$

as $n \rightarrow \infty$, as $|e^{2\pi i \ell \beta}| = 1$. Hence $x_n = \alpha n + \beta$ is uniformly distributed.

Solution 1.3

(i) If $\log_{10} 2 = p/q$ with p, q integers, $\text{hcf}(p, q) = 1$, then $2 = 10^{p/q}$, i.e. $2^q = 10^p = 5^p 2^p$. Comparing indices, we see that $0 = p = q$, a contradiction.

(ii) Let 2^n have leading digit r . Then

$$2^n = r \cdot 10^\ell + \text{terms involving lower powers of 10}$$

where the terms involving lower powers of 10 are integers lying in $[0, 10^\ell)$. Hence

$$\begin{aligned} 2^n \text{ has leading digit } r &\Leftrightarrow r \cdot 10^\ell \leq 2^n < (r+1) \cdot 10^\ell \\ &\Leftrightarrow \log_{10} r + \ell \leq n \log_{10} 2 < \log_{10}(r+1) + \ell \\ &\Leftrightarrow \log_{10} r \leq \{n \log_{10} 2\} < \log_{10}(r+1). \end{aligned}$$

Hence

$$\begin{aligned} & \frac{1}{n} \text{card}\{k \mid 0 \leq k \leq n-1, 2^k \text{ has leading digit } r\} \\ &= \frac{1}{n} \text{card}\{k \mid 0 \leq k \leq n-1, \{k \log_{10} 2\} \in [\log_{10} r, \log_{10}(r+1))\} \end{aligned}$$

which, by uniform distribution, converges to $\log_{10}(r+1) - \log_{10} r = \log_{10}(1 + 1/r)$ as $n \rightarrow \infty$.

Solution 1.4

The frequency with which the penultimate leading digit of 2^n is r is given by $\sum_{q=1}^9 A(q, r)$ where $A(q, r)$ is the frequency with which the leading digit is q and the penultimate leading digit is r .

Now 2^n has leading digit q and penultimate digit r precisely when

$$q \cdot 10^\ell + r \cdot 10^{\ell-1} \leq 2^n < q \cdot 10^\ell + (r+1) \cdot 10^{\ell-1}.$$

Taking logs shows that 2^n has leading digit q and penultimate leading digit r when

$$\log_{10}(10q+r) + \ell - 1 \leq n \log_{10} 2 < \log_{10}(10q+r+1) + (\ell - 1).$$

Reducing this mod 1 gives

$$\log_{10}(10q+r) - 1 \leq \{n \log_{10} 2\} < \log_{10}(10q+r+1) - 1$$

(the -1 s appear because $1 < \log_{10}(10q+r), \log_{10}(10q+r+1) < 2$). As $\{n \log_{10} 2\}$ is uniformly distributed mod 1, we see that

$$\begin{aligned} A(q, r) &= (\log_{10}(10q+r+1) - 1) - (\log_{10}(10q+r) - 1) \\ &= \log_{10} \left(1 + \frac{1}{10q+r} \right). \end{aligned}$$

Hence the frequency with which the penultimate leading digit of 2^n is r is

$$\sum_{q=1}^9 \log_{10} \left(1 + \frac{1}{10q+r} \right) = \log_{10} \prod_{q=1}^9 \left(1 + \frac{1}{10q+r} \right).$$

Solution 2.1

Suppose first that the numbers $\alpha_1, \dots, \alpha_k, 1$ are rationally independent. This means that if r_1, \dots, r_k, r are rational numbers such that

$$r_1 \alpha_1 + \dots + r_k \alpha_k + r = 0,$$

then $r_1 = \dots = r_k = r = 0$. In particular, for $\ell = (\ell_1, \dots, \ell_k) \in \mathbb{Z}^k \setminus \{\mathbf{0}\}$

$$\ell_1 \alpha_1 + \dots + \ell_k \alpha_k \notin \mathbb{Z},$$

so that

$$e^{2\pi i(\ell_1 \alpha_1 + \dots + \ell_k \alpha_k)} \neq 1.$$

By summing the geometric progression we have that

$$\begin{aligned} \left| \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i(\ell_1 j \alpha_1 + \dots + \ell_k j \alpha_k)} \right| &= \left| \frac{1}{n} \frac{e^{2\pi i n(\ell_1 \alpha_1 + \dots + \ell_k \alpha_k)} - 1}{e^{2\pi i(\ell_1 \alpha_1 + \dots + \ell_k \alpha_k)} - 1} \right| \\ &\leq \frac{1}{n} \frac{2}{|e^{2\pi i(\ell_1 \alpha_1 + \dots + \ell_k \alpha_k)} - 1|} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Therefore, by Weyl's Criterion, $(n\alpha_1, \dots, n\alpha_k)$ is uniformly distributed mod 1.

Now suppose that the numbers $\alpha_1, \dots, \alpha_k, 1$ are rationally dependent. Thus there exist rationals r_1, \dots, r_k, r (not all zero) such that $r_1\alpha_1 + \dots + r_k\alpha_k + r = 0$. By multiplying by a common denominator we can find $\ell = (\ell_1, \dots, \ell_k) \in \mathbb{Z}^k \setminus \{\mathbf{0}\}$ such that

$$\ell_1\alpha_1 + \dots + \ell_k\alpha_k \in \mathbb{Z}.$$

Thus $e^{2\pi i(\ell_1 n\alpha_1 + \dots + \ell_k n\alpha_k)} = 1$ for all $n \in \mathbb{N}$ and so

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i(\ell_1 j \alpha_1 + \dots + \ell_k j \alpha_k)} = 1 \not\rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Therefore, $(n\alpha_1, \dots, n\alpha_k)$ is *not* uniformly distributed mod 1.

Solution 2.2

Let $p(n) = \alpha_k n^k + \dots + \alpha_1 n + \alpha_0$. Suppose that $\alpha_k, \dots, \alpha_{s+1} \in \mathbb{Q}$ but $\alpha_s \notin \mathbb{Q}$. Let

$$\begin{aligned} p_1(n) &= \alpha_k n^k + \dots + \alpha_{s+1} n^{s+1} \\ p_2(n) &= \alpha_s n^s + \dots + \alpha_1 n + \alpha_0 \end{aligned}$$

so that $p(n) = p_1(n) + p_2(n)$. By choosing q to be a common denominator for $\alpha_k, \dots, \alpha_{s+1}$, we can write

$$p_1(n) = \frac{1}{q} (m_k n^k + \dots + m_{s+1} n^{s+1})$$

where $m_j \in \mathbb{Z}$.

By Weyl's Criterion, we want to show that for $\ell \in \mathbb{Z} \setminus \{0\}$ we have

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell p(j)} \rightarrow 0$$

as $n \rightarrow \infty$.

Write $j = qm + r$ where $r = 0, \dots, q-1$. Then $p_1(qm + r) = d_r \pmod{1}$ for some $d_r \in \mathbb{Q}$. Moreover, $p_2(qm + r) = p_2^{(q,r)}(m)$ is a polynomial in m with irrational leading coefficient. Now

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell p(j)} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{\lfloor \frac{n}{q} \rfloor - 1} \sum_{r=0}^{q-1} e^{2\pi i \ell d_r} e^{2\pi i \ell p_2^{(q,r)}(m)} \\ &= \lim_{n \rightarrow \infty} \frac{\lfloor \frac{n}{q} \rfloor}{n} \sum_{r=0}^{q-1} e^{2\pi i \ell d_r} \frac{1}{\lfloor \frac{n}{q} \rfloor} \sum_{m=0}^{\lfloor \frac{n}{q} \rfloor - 1} e^{2\pi i \ell p_2^{(q,r)}(m)} \\ &= 0 \end{aligned}$$

as $p_2^{(q,r)}(m)$ is uniformly distributed mod 1.

Solution 2.3

Let $p(n) = \alpha n^2 + n + 1$ where $\alpha \notin \mathbb{Q}$. Let $m \geq 1$ and consider the sequence $p^{(m)}(n) = p(n+m) - p(n)$ of m th differences. We have that

$$p^{(m)}(n) = \alpha(n+m)^2 + (n+m) + 1 - \alpha n^2 - n - 1 = 2\alpha mn + \alpha m^2 + m$$

which is a degree 1 polynomial in n with leading coefficient $2\alpha m \notin \mathbb{Q}$. Note that $2\alpha m \neq 0$, as $m \geq 1$. By Exercise 1.2 we have that $p^{(m)}(n)$ is uniformly distributed mod 1 for every $m \geq 1$. By Lemma 2.3.3, it follows that $p(n)$ is uniformly distributed mod 1.

Solution 2.4

By Weyl's Criterion, we require that for each $(\ell_1, \ell_2) \in \mathbb{Z}^2 \setminus \{(0,0)\}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \exp 2\pi i(\ell_1 p(k) + \ell_2 q(k)) = 0 \tag{12.0.1}$$

Let

$$\begin{aligned} p_{\ell_1, \ell_2}(n) &= \ell_1 p(n) + \ell_2 q(n) \\ &= (\ell_1 \alpha_k + \ell_2 \beta_k) n^k + \dots + (\ell_1 \alpha_1 + \ell_2 \beta_1) n + (\ell_1 \alpha_0 + \ell_2 \beta_0) \end{aligned}$$

This is a polynomial of degree at most k . Then (12.0.1) can be written as

$$\frac{1}{n} \sum_{k=0}^{n-1} \exp 2\pi i p_{\ell_1, \ell_2}(k).$$

By the 1-dimensional version of Weyl's criterion (using the integer $\ell = 1$), this will converge to 0 as $n \rightarrow \infty$ if $p_{\ell_1, \ell_2}(n)$ is uniformly distributed mod 1. By Weyl's Theorem on Polynomials (Theorem 2.3.1), this happens if at least one of $\ell_1 \alpha_k + \ell_2 \beta_k, \ell_1 \alpha_{k-1} + \ell_2 \beta_{k-1}, \dots, \ell_1 \alpha_1 + \ell_2 \beta_1$ is irrational. Note that $\ell_1 \alpha_i + \ell_2 \beta_i \notin \mathbb{Q}$ if and only if $\alpha_i, \beta_i, 1$ are rationally independent.

Solution 2.5

- (i) We know that $\emptyset \in \mathcal{B}$ and that if $E \in \mathcal{B}$ then $X \setminus E \in \mathcal{B}$. Hence $X = X \setminus \emptyset \in \mathcal{B}$.
- (ii) Let $E_n \in \mathcal{B}$. Then $X \setminus E_n \in \mathcal{B}$. Then $\bigcup_n (X \setminus E_n) \in \mathcal{B}$. Now $\bigcup_n (X \setminus E_n) = X \setminus \bigcap_n E_n$. Hence $\bigcap_n E_n = X \setminus (X \setminus \bigcap_n E_n) \in \mathcal{B}$.

Solution 2.6

The smallest σ -algebra containing the sets $[0, 1/4), [1/4, 1/2), [1/2, 3/4)$ and $[3/4, 1]$ is

$$\begin{aligned} \mathcal{B} = \{ & \emptyset, [0, 1/4), [1/4, 1/2), [1/2, 3/4), [3/4, 1], [0, 1/2), [0, 1/4) \cup [1/2, 3/4), [0, 1/4) \cup [3/4, 1], \\ & [1/4, 3/4), [1/4, 1/2) \cup [3/4, 1], [1/2, 1], [0, 3/4), [0, 1/2) \cup [3/4, 1], \\ & [0, 1/4) \cup [1/2, 1], [1/4, 1], [0, 1] \} \end{aligned}$$

Solution 2.7

Clearly a finite union of dyadic intervals is a Borel set.

By Proposition 2.4.2 we need to show that if $x, y \in [0, 1], x \neq y$, then there exist disjoint dyadic intervals I_1, I_2 such that $x \in I_1, y \in I_2$. Let $\varepsilon = |x - y|$ and choose n such that $1/2^n < \varepsilon/2$. Without loss of generality, assume that $x < y$. Then there exist integers $p, q, p < q$, such that

$$\frac{p-1}{2^n} \leq x < \frac{p}{2^n} < \frac{q}{2^n} < y \leq \frac{q+1}{2^n}.$$

Hence x, y belong to different dyadic intervals.

Solution 2.8

Let \mathcal{A} denote the collection of finite unions of intervals. Trivially $\emptyset \in \mathcal{A}$. If $A, B \in \mathcal{A}$ are finite unions of intervals then $A \cup B$ is a finite union of intervals. Hence \mathcal{A} is closed under taking finite unions. If $A = [a, b] \subset [0, 1]$ then $A^c = [0, a) \cup (b, 1]$ is a finite union of intervals. Hence \mathcal{A} is an algebra.

Solution 2.9

First note that if μ is a measure and $A \subset B$ then $\mu(A) \leq \mu(B)$. (To see this, note that if $A \subset B$ then $B = A \cup (B \setminus A)$ is a disjoint union. Hence $\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$.)

Let μ denote Lebesgue measure on $[0, 1]$. Let $x \in [0, 1]$. For any $\varepsilon > 0$, we have that $\{x\} \subset (x - \varepsilon, x + \varepsilon) \cap [0, 1]$. Hence $\mu(\{x\}) \leq 2\varepsilon$. As $\varepsilon > 0$ is arbitrary, it follows that $\mu(\{x\}) = 0$.

Let $E = \{x_j\}_{j=1}^{\infty}$ be a countable set. Then

$$\mu(E) = \mu\left(\bigcup_{j=1}^{\infty} \{x_j\}\right) = \sum_{j=1}^{\infty} \mu(\{x_j\}) = 0.$$

Hence any countable set has Lebesgue measure 0.

As the rational points in $[0, 1]$ are countable, it follows that $\mu(\mathbb{Q} \cap [0, 1]) = 0$. Hence Lebesgue almost every point in $[0, 1]$ is irrational.

Solution 2.10

Let $\mu = \delta_{1/2}$ be the Dirac δ -measure at $1/2$. Then, by definition, $\mu([0, 1/2) \cup (1/2, 1]) = 0$ as $1/2 \notin [0, 1/2) \cup (1/2, 1]$. Hence $\mu\{x \in [0, 1] \mid x \neq 1/2\} = 0$, so that μ -a.e. point in $[0, 1]$ is equal to $1/2$.

Solution 3.1

Let $x_n = \alpha n$ where $\alpha \in \mathbb{R}$ is irrational. Then x_n is uniformly distributed mod 1 (by the results in §1.2.1). Let $A = \{\{\alpha n\} \mid n \geq 0\} \subset [0, 1]$ denote the set of fractional parts of the sequence x_n ; note that A is a countable set. Let $f = \chi_A$. Then $f \in L^1([0, 1], \mathcal{B}, \mu)$ (where \mathcal{B} denotes the Borel σ -algebra and μ denotes Lebesgue measure on $[0, 1]$) and $f \equiv 0$ a.e. Hence $\int f d\mu = 0$. However, $f(\{x_n\}) = 1$ for each n . Hence

$$\frac{1}{n} \sum_{j=0}^{n-1} f(\{x_n\}) = 1 \not\rightarrow \int f d\mu = 0$$

as $n \rightarrow \infty$.

Solution 3.2

Let X be a compact metric space equipped with the Borel σ -algebra \mathcal{B} . Let $T : X \rightarrow X$ be continuous. Recall that \mathcal{B} is generated by the open sets. It is sufficient to check that $T^{-1}U \in \mathcal{B}$ for all open sets U . But this is clear: as T is continuous, the pre-image $T^{-1}U$ of any open set is open, hence $T^{-1}U \in \mathcal{B}$.

Solution 3.3

Define $f_n : [0, 1] \rightarrow \mathbb{R}$ by

$$f_n(x) = \begin{cases} n - n^2x & \text{if } 0 \leq x \leq 1/n \\ 0 & \text{if } 1/n \leq x \leq 1. \end{cases}$$

(Draw a picture!) Then f_n is continuous, hence $f_n \in L^1(X, \mathcal{B}, \mu)$. Moreover, $\int f_n d\mu = 1/2$ for each n . Hence $f_n \not\rightarrow 0$ in $L^1(X, \mathcal{B}, \mu)$.

However, $f_n \rightarrow 0$ μ -a.e. To see this, let $x \in [0, 1]$, $x \neq 0$. Choose n such that $1/n < x$. Then $f_n(x) = 0$ for any $n \geq N$. Hence, if $x \neq 0$, we have that $f_n(x) = 0$ for all sufficiently large n . Hence $f_n \rightarrow 0$ μ -a.e.

Solution 3.4

First note that if $B \in \mathcal{B}$ then $T^{-1}B \in \mathcal{B}$. Hence $T_*\mu(B) = \mu(T^{-1}B)$ is well-defined.

Clearly $T^{-1}(\emptyset) = \emptyset$. Hence $T_*\mu(\emptyset) = \mu(T^{-1}\emptyset) = \mu(\emptyset) = 0$.

Let $E_n \in \mathcal{B}$ be pairwise disjoint. Then $T^{-1}E_n \in \mathcal{B}$ are pairwise disjoint. (To see this, suppose that $x \in T^{-1}E_n \cap T^{-1}E_m$. Then $T(x) \in E_n$ and $T(x) \in E_m$. Hence $T(x) \in E_n \cap E_m$. As the E_n are pairwise disjoint, this implies that $n = m$. Hence $T^{-1}E_n = T^{-1}E_m$.) Hence

$$T_*\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \mu\left(T^{-1}\bigcup_{n=1}^{\infty} E_n\right) = \mu\left(\bigcup_{n=1}^{\infty} T^{-1}E_n\right) = \sum_{n=1}^{\infty} \mu(T^{-1}E_n) = \sum_{n=1}^{\infty} T_*\mu(E_n)$$

where we have used the fact that $T^{-1}\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} T^{-1}E_n$.

Hence $T_*\mu$ is a measure.

Finally, note that $T^{-1}(X) = X$. Hence $T_*\mu(X) = \mu(T^{-1}X) = \mu(X) = 1$, so that T is a probability measure.

Solution 3.5

- (i) Let λ denote Lebesgue measure on $[0, 1]$. All one needs to do is to find a set B such that $\lambda(B) \neq \lambda(T^{-1}B)$, and any (reasonable) choice of set B will work. For example, take $B = (1/2, 1)$. Then

$$T^{-1}(B) = \bigcup_{n=1}^{\infty} \left(\frac{1}{n+1}, \frac{1}{n+1/2}\right).$$

It follows that

$$\lambda(T^{-1}B) = \sum_{n=1}^{\infty} \frac{1}{(1+2n)(1+n)} = \log(4) - 1 < \frac{1}{2} = \lambda(B).$$

- (ii) Recall that

$$\mu(B) = \frac{1}{\log 2} \int_B \frac{dx}{1+x} = \frac{1}{\log 2} \int \frac{\chi_B(x)}{1+x} dx.$$

Note that $1/2 \leq 1/(1+x) \leq 1$ if $0 \leq x \leq 1$. Hence

$$\frac{1}{\log 2} \int \frac{\chi_B(x)}{2} dx \leq \mu(B) \leq \frac{1}{\log 2} \int \chi_B(x) dx$$

so that

$$\frac{1}{2\log 2} \lambda(B) = \frac{1}{2\log 2} \int \chi_B(x) dx \leq \mu(B) \leq \frac{1}{\log 2} \int \chi_B(x) dx = \frac{1}{\log 2} \lambda(B).$$

(iii) From (3.4.1) it follows that

$$\frac{1}{2\log 2} \int f d\lambda \leq \int f d\mu \leq \frac{1}{\log 2} \int f d\lambda \tag{12.0.2}$$

for all simple functions f . By taking increasing sequences of simple functions, we see that (12.0.2) continues to hold for non-negative measurable functions. Let $f \in L^1(X, \mathcal{B}, \mu)$. Then

$$\frac{1}{2\log 2} \int |f| d\lambda \leq \int |f| d\mu$$

so that $f \in L^1(X, \mathcal{B}, \lambda)$. Similarly, if $f \in L^1(X, \mathcal{B}, \lambda)$ then $f \in L^1(X, \mathcal{B}, \mu)$.

Solution 3.6

Let $[a, b] \subset [0, 1]$. Then

$$T^{-1}[a, b] = \bigcup_{j=0}^{k-1} \left[\frac{a+j}{k}, \frac{b+j}{k} \right]$$

so that

$$T_*\mu([a, b]) = \sum_{j=0}^{k-1} \frac{b+j}{k} - \frac{a+j}{k} = \sum_{j=0}^{k-1} \frac{b-a}{k} = b-a = \mu([a, b]).$$

Hence $T_*\mu$ and μ agree on intervals. Hence, by the Hahn-Kolmogorov Extension Theorem, $T_*\mu = \mu$ so that μ is a T -invariant measure.

Solution 3.7

Let $T(x) = \beta x \text{ mod } 1$. Then T has a graph as illustrated in Figure 12.1.

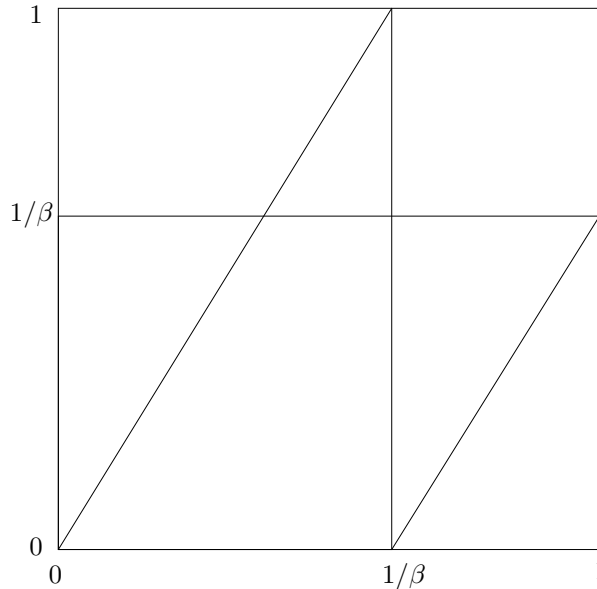


Figure 12.1: The graph of $T(x) = \beta x \text{ mod } 1$.

Let us first show that T does not preserve Lebesgue measure. For this, it is sufficient to find a set B such that B and $T^{-1}B$ do not have equal Lebesgue measure; in fact, almost any

reasonable choice of B will suffice, but here is a specific example. Let λ denote Lebesgue measure. Take $B = [1/\beta, 1]$. Then $\lambda(B) = 1 - 1/\beta = 1/\beta^2$ (as $\beta - 1 = 1/\beta$). Now $T^{-1}[1/\beta, 1] = [1/\beta^2, 1/\beta]$ so that $\lambda(T^{-1}B) = 1/\beta - 1/\beta^2 = 1/\beta^3 \neq \lambda(B)$.

We now show that T does preserve the measure μ defined as in the statement of the question. To do this we again use the Hahn-Kolmogorov Extension Theorem, which tells us that it is sufficient to prove that $\mu(T^{-1}[a, b]) = \mu[a, b]$ for all intervals $[a, b] \subset [0, 1]$.

If $[a, b] \subset [0, 1/\beta]$ then

$$T^{-1}[a, b] = \left[\frac{a}{\beta}, \frac{b}{\beta} \right] \cup \left[\frac{a+1}{\beta}, \frac{b+1}{\beta} \right],$$

a disjoint union. Hence,

$$\begin{aligned} \mu(T^{-1}[a, b]) &= \frac{1}{\frac{1}{\beta} + \frac{1}{\beta^3}} \left(\frac{b-a}{\beta} \right) + \frac{1}{\beta \left(\frac{1}{\beta} + \frac{1}{\beta^3} \right)} \left(\frac{(b+1) - (a+1)}{\beta} \right) \\ &= \frac{b-a}{\frac{1}{\beta} + \frac{1}{\beta^3}} \left(\frac{1}{\beta} + \frac{1}{\beta^2} \right) \\ &= \mu([a, b]). \end{aligned}$$

If $[a, b] \subset [1/\beta, 1]$ then $T^{-1}[a, b] = [a/\beta, b/\beta]$ and

$$\mu(T^{-1}[a, b]) = \mu \left(\left[\frac{a}{\beta}, \frac{b}{\beta} \right] \right) = \frac{1}{\frac{1}{\beta} + \frac{1}{\beta^3}} \left(\frac{b-a}{\beta} \right) = \mu([a, b]).$$

If $a < 1/\beta < b$ then we write $[a, b] = [a, 1/\beta] \cup [1/\beta, b]$. Then $T^{-1}[a, b] = T^{-1}[a, 1/\beta] \cup T^{-1}[1/\beta, b]$, a disjoint union. Hence

$$\begin{aligned} \mu(T^{-1}[a, b]) &= \mu(T^{-1}[a, 1/\beta] \cup T^{-1}[1/\beta, b]) \\ &= \mu(T^{-1}[a, 1/\beta]) + \mu(T^{-1}[1/\beta, b]) \\ &= \mu([a, 1/\beta]) + \mu([1/\beta, b]) = \mu([a, b]). \end{aligned}$$

Solution 3.8

(i) Note that

$$\mu([0, 1]) = \frac{1}{\pi} \int_0^1 \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \int_0^{\pi/2} d\theta = 1,$$

using the substitution $x = \sin^2 \theta$.

(ii) By the Hahn-Kolmogorov Extension Theorem it is sufficient to prove that $\mu([a, b]) = \mu(T^{-1}[a, b])$ for all intervals $[a, b]$.

Note that

$$T^{-1}[a, b] = \left[\frac{1 - \sqrt{1-a}}{2}, \frac{1 - \sqrt{1-b}}{2} \right] \cup \left[\frac{1 + \sqrt{1-b}}{2}, \frac{1 + \sqrt{1-a}}{2} \right]$$

(as the graph of T is decreasing on $[1/2, 1]$ the order of a, b are reversed in the second sub-interval). It is sufficient to prove that

$$\mu\left(\left(\frac{1 - \sqrt{1-a}}{2}, \frac{1 - \sqrt{1-b}}{2}\right) \cup \left(\frac{1 + \sqrt{1-b}}{2}, \frac{1 + \sqrt{1-a}}{2}\right)\right) = \frac{1}{2}\mu([a, b])$$

and

$$\mu([(1 + \sqrt{1-b})/2, (1 + \sqrt{1-a})/2]) = \frac{1}{2}\mu([a, b])$$

We prove the first equality (the second is similar). Now

$$\mu([(1 - \sqrt{1-a})/2, (1 - \sqrt{1-b})/2]) = \frac{1}{\pi} \int_{\frac{1-\sqrt{1-a}}{2}}^{\frac{1-\sqrt{1-b}}{2}} \frac{1}{\sqrt{x(1-x)}} dx. \quad (12.0.3)$$

Consider the substitution $u = 4x(1-x)$. Then $du = 4(1-2x)dx$ and as x ranges between $(1 - \sqrt{1-a})/2$ and $(1 - \sqrt{1-b})/2$, u ranges between a, b . Note also that a simple manipulation shows that $(1-2x)^2 = 1-u$. Hence the right-hand side of (12.0.3) is equal to

$$\frac{1}{2\pi} \int_a^b \frac{1}{\sqrt{u(1-u)}} du = \frac{1}{2}\mu([a, b]).$$

Similarly,

$$\mu([(1 + \sqrt{1-b})/2, (1 + \sqrt{1-a})/2]) = \frac{1}{2}\mu([a, b])$$

and the result follows.

Solution 3.9

Note that

$$X = \bigcup_{n=1}^{\infty} \left(\frac{1}{n+1}, \frac{1}{n} \right] \cup \{0\}$$

and that this is a disjoint union. Hence, denoting Lebesgue measure by μ ,

$$1 = \mu(X) = \sum_{n=1}^{\infty} \mu \left(\left(\frac{1}{n+1}, \frac{1}{n} \right] \right) = \sum_{n=1}^{\infty} \frac{1}{n} - \frac{1}{n+1} = \sum_{n=1}^{\infty} \frac{1}{n(n+1)}.$$

By the Hahn-Kolmogorov Extension Theorem, it is sufficient to check that $\mu(T^{-1}[a, b]) = \mu([a, b])$ for all intervals $[a, b]$. It is straightforward to check that

$$T^{-1}[a, b] = \bigcup_{n=1}^{\infty} \left[\frac{a+n}{n(n+1)}, \frac{b+n}{n(n+1)} \right]$$

and that this is a disjoint union. Hence

$$\begin{aligned} \mu(T^{-1}[a, b]) &= \sum_{n=1}^{\infty} \mu \left(\left[\frac{a+n}{n(n+1)}, \frac{b+n}{n(n+1)} \right] \right) \\ &= \sum_{n=1}^{\infty} \frac{b-a}{n(n+1)} \\ &= b-a = \mu([a, b]). \end{aligned}$$

Solution 3.10

- (i) Clearly $d(\mathbf{x}, \mathbf{y}) \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{y}$. It is also clear that $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$. It remains to prove the triangle inequality: $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Sigma$.

If any of $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are equal then the triangle inequality is clear, so we can assume that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are all distinct. Suppose that \mathbf{x} and \mathbf{y} agree in the first n places and that \mathbf{y} and \mathbf{z} agree in the first m places. Then \mathbf{x} and \mathbf{z} agree in at least the first $\min\{n, m\}$ places. Hence $n(\mathbf{x}, \mathbf{z}) \geq \min\{n(\mathbf{x}, \mathbf{y}), n(\mathbf{y}, \mathbf{z})\}$. Hence

$$d(\mathbf{x}, \mathbf{z}) = \frac{1}{2^{n(\mathbf{x}, \mathbf{z})}} \leq \frac{1}{2^{\min\{n(\mathbf{x}, \mathbf{y}), n(\mathbf{y}, \mathbf{z})\}}} \leq \frac{1}{2^{n(\mathbf{x}, \mathbf{y})}} + \frac{1}{2^{n(\mathbf{y}, \mathbf{z})}} = d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}).$$

(ii) Let $\varepsilon > 0$ and choose $n \geq 1$ such that $1/2^n < \varepsilon$. Choose $\delta = 1/2^{n+1}$. Suppose that $d(\mathbf{x}, \mathbf{y}) < \delta = 1/2^{n+1}$. Then $n(\mathbf{x}, \mathbf{y}) > n + 1$, i.e. \mathbf{x} and \mathbf{y} agree in at least the first $n + 1$ places. Hence $\sigma(\mathbf{x})$ and $\sigma(\mathbf{y})$ agree in at least the first n places. Hence $n(\sigma(\mathbf{x}), \sigma(\mathbf{y})) > n$. Hence $d(\sigma(\mathbf{x}), \sigma(\mathbf{y})) < 1/2^n < \varepsilon$.

(iii) We show that $[i_0, \dots, i_{n-1}]$ is open. Let $\mathbf{x} \in [i_0, \dots, i_{n-1}]$ so that $x_j = i_j$ for $j = 0, 1, \dots, n - 1$. Choose $\varepsilon = 1/2^n$. Suppose that $d(\mathbf{x}, \mathbf{y}) < \varepsilon$. Then $n(\mathbf{x}, \mathbf{y}) > n$, i.e. \mathbf{x} and \mathbf{y} agree in at least the first n places. Hence $x_j = y_j$ for $j = 0, 1, \dots, n - 1$. Hence $y_j = i_j$ for $j = 0, 1, \dots, n - 1$ so that $\mathbf{y} \in [i_0, \dots, i_{n-1}]$. Hence $[i_0, \dots, i_{n-1}]$ is open.

To see that $[i_0, \dots, i_{n-1}]$ is closed, note that

$$\Sigma \setminus [i_0, \dots, i_{n-1}] = \bigcup [i'_0, \dots, i'_{n-1}]$$

where the union is over all n -tuples $(i'_0, i'_1, \dots, i'_{n-1}) \neq (i_0, i_1, \dots, i_{n-1})$. This is a finite union of open sets, and so is open. Hence $[i_0, \dots, i_{n-1}]$, as the complement of an open set, is closed.

Solution 3.11

First note that

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, P^2 = \begin{pmatrix} 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 5/8 & 0 & 3/8 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 3/8 & 0 & 5/8 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \end{pmatrix},$$

$$P^3 = \begin{pmatrix} 0 & 5/8 & 0 & 3/8 & 0 \\ 5/32 & 0 & 3/4 & 0 & 3/32 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 3/32 & 0 & 3/4 & 0 & 5/32 \\ 0 & 3/8 & 0 & 5/8 & 0 \end{pmatrix}, P^4 = \begin{pmatrix} 5/32 & 0 & 3/4 & 0 & 3/32 \\ 0 & 17/32 & 0 & 15/32 & 0 \\ 1/8 & 0 & 3/4 & 0 & 1/8 \\ 0 & 15/32 & 0 & 17/32 & 0 \\ 3/32 & 0 & 3/4 & 0 & 5/32 \end{pmatrix}.$$

As for each i, j there exists n for which $P^n(i, j) > 0$, it follows that P is irreducible.

Recall that the period of P is the highest common factor of $\{n > 0 \mid P^n(i, i) > 0\}$. As all the diagonal entries of P^2 are positive, it follows that P has period 2.

Decompose $\{1, 2, 3, 4, 5\} = \{1, 3, 5\} \cup \{2, 4\} = S_0 \cup S_1$. If $P(i, j) > 0$ then either $i \in S_0$ and $j \in S_1$, or $i \in S_1$ and $j \in S_0$, i.e. $i \in S_\ell$ and $j \in S_{\ell+1 \pmod 2}$. When restricted to the indices $\{1, 3, 5\}$, P^2 has the form

$$\begin{pmatrix} 1/4 & 3/4 & 0 \\ 1/8 & 3/4 & 1/8 \\ 0 & 3/4 & 1/4 \end{pmatrix}$$

which is easily seen to be irreducible and aperiodic. When restricted to the indices $\{2, 4\}$, P^2 has the form

$$\begin{pmatrix} 5/8 & 3/8 \\ 3/8 & 5/8 \end{pmatrix}$$

which is clearly irreducible and aperiodic.

The eigenvalues of P are found by evaluating the determinant

$$\begin{vmatrix} -\lambda & 1 & 0 & 0 & 0 \\ 1/4 & -\lambda & 3/4 & 0 & 0 \\ 0 & 1/2 & -\lambda & 1/2 & 0 \\ 0 & 0 & 3/4 & -\lambda & 1/4 \\ 0 & 0 & 0 & 1 & -\lambda \end{vmatrix}.$$

After simplifying this expression, we obtain

$$(1 - \lambda)(1 + \lambda)\lambda \left(\lambda + \frac{1}{4} \right).$$

(Note that, as P has period 2, we expect from the Perron-Frobenius Theorem that the square roots of 1 to be the eigenvalues of modulus 1 for P .)

A left eigenvector $p = (p(1), p(2), p(3), p(4), p(5))$ for the eigenvalue 1 is determined by

$$(p(1), p(2), p(3), p(4), p(5)) \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} = (p(1), p(2), p(3), p(4), p(5))$$

which simplifies to

$$\frac{1}{4}p(2) = p(1), \quad p(1) + \frac{1}{2}p(3) = p(2), \quad \frac{3}{4}p(2) + \frac{3}{4}p(4) = p(3), \quad \frac{1}{2}p(3) + p(5) = p(4), \quad \frac{1}{4}p(4) = p(5).$$

Setting $p(1) = 1$ we obtain $(p(1), p(2), p(3), p(4), p(5)) = (1, 4, 6, 4, 1)$, and normalising this to form a probability vector we obtain

$$p = \left(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16} \right).$$

Solution 3.12

Let $p = (p(1), \dots, p(k))$ be a probability vector. Let P be the matrix

$$P = \begin{pmatrix} p(1) & p(2) & \cdots & p(k) \\ p(1) & p(2) & \cdots & p(k) \\ \vdots & & & \vdots \\ p(1) & p(2) & \cdots & p(k) \end{pmatrix}.$$

Then P is a stochastic matrix. As each $p(j) > 0$, it follows that P is aperiodic. It is straightforward to check that $pP = p$.

As $P(i, j) = p(j)$, the Markov measure determined by the matrix P is the same as Bernoulli measure determined by the probability vector p .

Solution 4.1

Note that

$$\chi_{T^{-1}B}(x) = 1 \Leftrightarrow x \in T^{-1}B \Leftrightarrow T(x) \in B \Leftrightarrow \chi_B(T(x)) = 1.$$

Hence $\chi_{T^{-1}B} = \chi_B \circ T$.

Solution 4.2

Note that $T^n(x) = x$ if and only if $2^n x = x \pmod{1}$, i.e. $2^n x = x + p$ for some integer p . Hence $x = p/(2^n - 1)$. We get distinct values of x in \mathbb{R}/\mathbb{Z} when $p = 0, 1, \dots, 2^n - 2$ (note that when $p = 2^n - 1$ then $x = 1$, which is the same as 0 in \mathbb{R}/\mathbb{Z}).

Hence there are infinitely many distinct periodic orbits for the doubling map. If $x, Tx, \dots, T^{n-1}x$ is a periodic orbit of period n then let $\delta_{(x)} = 1/n \sum_{j=0}^{n-1} \delta_{T^j x}$ denote the periodic orbit measure supported on the orbit of x . As there are infinitely many distinct periodic orbits, there are infinitely many distinct measures supported on periodic orbits.

Solution 4.3

Recall that \mathbb{R}/\mathbb{Z} can be regarded as $[0, 1]$ where 0 and 1 are identified. Suppose that $f : [0, 1] \rightarrow \mathbb{R}$ is integrable and that $f(0) = f(1)$ so that f is a well-defined function on \mathbb{R}/\mathbb{Z} . Then

$$\begin{aligned} \int f \circ T \, d\mu &= \int_0^{1/2} f \circ T \, d\mu + \int_{1/2}^1 f \circ T \, d\mu \\ &= \int_0^{1/2} f(2x) \, dx + \int_{1/2}^1 f(2x-1) \, dx \\ &= \frac{1}{2} \int_0^1 f(x) \, dx + \frac{1}{2} \int_0^1 f(x) \, dx \\ &= \int f \, d\mu \end{aligned} \tag{12.0.4}$$

where we have used the substitution $u(x) = 2x$ for the first integral and $u(x) = 2x - 1$ for the second integral in (12.0.4)

Solution 4.4

It is straightforward to check that $T : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{R}^2/\mathbb{Z}^2$ is a diffeomorphism.

Recall that we can identify functions $f : \mathbb{R}^2/\mathbb{Z}^2 \rightarrow \mathbb{C}$ with functions $f : \mathbb{R}^2 \rightarrow \mathbb{C}$ that satisfy $f(x+n) = f(x)$ for all $n \in \mathbb{Z}^2$. We apply the change of variables formula with the substitution $u(x) = T(x)$. Note that

$$DT(x) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

so that $|\det DT| = 1$. Hence, by the change of variables formula

$$\int f \circ T \, d\mu = \int_{\mathbb{R}^2/\mathbb{Z}^2} f \circ T |\det DT| \, d\mu = \int_{T(\mathbb{R}^2/\mathbb{Z}^2)} f \, d\mu = \int f \, d\mu.$$

Solution 5.1

Let $\alpha = p/q$ with $p, q \in \mathbb{Z}$, $q \neq 0$, $\text{hcf}(p, q) = 1$. Let

$$B = \bigcup_{j=0}^{q-1} \left[\frac{j}{q}, \frac{j}{q} + \frac{1}{2q} \right].$$

Then

$$T^{-1} \left[\frac{j}{q}, \frac{j}{q} + \frac{1}{2q} \right] = \left[\frac{j-p}{q}, \frac{j-p}{q} + \frac{1}{2q} \right]$$

so that $T^{-1}B = B$ (draw a picture to understand this better). However $\mu(B) = 1/2$, so that T is not ergodic with respect to Lebesgue measure.

Solution 5.2

Suppose that $f \in L^2(X, \mathcal{B}, \mu)$ has Fourier series

$$\sum_{(n,m) \in \mathbb{Z}^2} c_{(n,m)} e^{2\pi i(nx+my)}.$$

Then $f \circ T$ has Fourier series

$$\sum_{(n,m) \in \mathbb{Z}^2} c_{(n,m)} e^{2\pi i(n(x+\alpha)+m(x+y))} = \sum_{(n,m) \in \mathbb{Z}^2} c_{(n,m)} e^{2\pi i n \alpha} e^{2\pi i((n+m)x+my)}.$$

Comparing coefficients we see that

$$c_{(n+m,m)} = e^{2\pi i n \alpha} c_{(n,m)}.$$

Suppose that $m \neq 0$. Then for each $j > 0$,

$$|c_{(n+jm,m)}| = \cdots = |c_{(n+m,m)}| = |c_{(n,m)}|,$$

as $|e^{2\pi i n \alpha}| = 1$. Note that if $m \neq 0$ then $(n+jm, m) \rightarrow \infty$ as $j \rightarrow \infty$. By the Riemann-Lebesgue Lemma (Proposition 5.3.2(ii)), we must have that $c_{(n,m)} = 0$ if $m \neq 0$. Hence f has Fourier series

$$\sum_{(n,0) \in \mathbb{Z}^2} c_{(n,0)} e^{2\pi i n x}$$

and $f \circ T$ has Fourier series

$$\sum_{(n,0) \in \mathbb{Z}^2} c_{(n,0)} e^{2\pi i n \alpha} e^{2\pi i n x}.$$

Comparing Fourier coefficients we see that

$$c_{(n,0)} = c_{(n,0)} e^{2\pi i n \alpha}.$$

Suppose that $n \neq 0$. As $\alpha \notin \mathbb{Q}$, $e^{2\pi i n \alpha} \neq 1$. Hence $c_{(n,0)} = 0$ unless $n = 0$. Hence f has Fourier series $c_{(0,0)}$, i.e. f is constant a.e. Hence T is ergodic with respect to Lebesgue measure.

Solution 5.3

Suppose that $T : X \rightarrow X$ has a periodic point x with period n . Let

$$\mu = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x}.$$

Let $B \in \mathcal{B}$ and suppose that $T^{-1}B = B$. We must show that $\mu(B) = 0$ or 1 .

Suppose that $x \in B$. Then $x \in T^{-1}B$. Hence $T(x) \in B$. Continuing inductively, we see that $T^j(x) \in B$ for $j = 0, 1, \dots, n-1$. Hence

$$\mu(B) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x}(B) = \frac{1}{n} \sum_{j=0}^{n-1} 1 = 1.$$

Similarly, if $x \in X \setminus B$ then $T^j(x) \in X \setminus B$ for $j = 0, 1, \dots, n-1$ (we have used the fact that if B is T -invariant then $X \setminus B$ is T -invariant). Hence $\mu(B) = 0$.

Solution 5.4

(i) Recall that the determinant of a matrix is equal to the product of all the eigenvalues.

Let T be a linear toral automorphism with corresponding matrix A . Suppose that A has an eigenvalue of modulus 1. By considering A^2 if necessary, there is no loss in generality in assuming that $\det A = 1$.

Suppose $k = 2$. Then the matrix A has two eigenvalues, $\lambda, \bar{\lambda}$. As A has an eigenvalue of modulus 1, we must have that $\lambda = e^{2\pi i\theta}$ for some $\theta \in [0, 1)$. Then $\lambda, \bar{\lambda}$ satisfy the equation $\lambda^2 + 2\cos\theta\lambda + 1 = 0$. However the matrix $A = (a, b; c, d)$ has characteristic equation $\lambda^2 + (a+d)\lambda + 1 = 0$. Hence $2\cos\theta = a+d$, an integer. Thus $\theta = 0, \pm\pi/2, \pm\pi$. Hence $\lambda = \pm 1, \pm i$, and is a root of unity and so T cannot be ergodic.

Now suppose $k = 3$. Then, assuming that A has an eigenvalue of modulus 1, the eigenvalues must be $\lambda = e^{2\pi i\theta}, \bar{\lambda}$ and $\mu \in \mathbb{R}$. As $\det A = 1$, we must have that $\lambda\bar{\lambda}\mu = 1$. As $\lambda\bar{\lambda} = 1$, it follows that $\mu = 1$, Hence A has 1 as an eigenvalue and so T cannot be ergodic.

Thus $k \geq 4$.

(ii) A has integer entries and it is easy to see that $\det A = 1$. Hence A determines a linear toral automorphism of $\mathbb{R}^4/\mathbb{Z}^4$.

(iii) It is straightforward to calculate that the characteristic equation for A is

$$\lambda^4 - 8\lambda^3 + 6\lambda^2 - 8\lambda + 1 = 0.$$

Clearly, $\lambda \neq 0$. Dividing by λ^2 and substituting $u = \lambda + \lambda^{-1}$ we see that

$$u^2 - 8u + 4 = 0.$$

Hence

$$u = 4 \pm 2\sqrt{3}.$$

Multiplying $\lambda + \lambda^{-1} = u$ by λ we obtain a quadratic in λ with solution

$$\lambda = \frac{u \pm \sqrt{u^2 - 4}}{2}.$$

Substituting the two different values of u gives four values of λ , namely:

$$2 + \sqrt{3} \pm \sqrt{6 + 4\sqrt{3}}, \quad 2 - \sqrt{3} \pm i\sqrt{4\sqrt{3} - 6}.$$

The first two are real and not of unit modulus, whereas the second two are complex numbers of unit modulus.

- (iv) This question is *not* part of the course and is included for completeness only. The solution requires ideas from Galois theory.

We first claim that $f(\lambda) = \lambda^4 - 8\lambda^3 + 6\lambda^2 - 6\lambda + 1$ is irreducible over \mathbb{Q} . (To see this, recall that irreducibility over \mathbb{Q} is equivalent to irreducibility over \mathbb{Z} . Consider the polynomial $f(\lambda + 1) = \lambda^4 - 4\lambda^3 - 12\lambda^2 - 14\lambda - 6$ and apply Eisenstein's criterion using the prime 2 to see that $f(\lambda + 1)$ is irreducible over \mathbb{Z} . Hence $f(\lambda)$ is irreducible over \mathbb{Z} .) Hence $\lambda^4 - 8\lambda^3 + 6\lambda^2 - 6\lambda + 1$ has no common factors with $\lambda^n - 1$ for any n . Hence λ is not a root of unity.

Solution 6.1

- (i) Let μ denote Lebesgue measure. We prove that $T_*\mu = \mu$ by using the Hahn-Kolmogorov Extension Theorem. It is sufficient to prove that $T_*\mu([a, b]) = \mu([a, b])$ for all intervals $[a, b]$. Note that

$$T^{-1}[a, b] = \left[\frac{a}{2}, \frac{b}{2}\right] \cup \left[1 - \frac{b}{2}, 1 - \frac{a}{2}\right].$$

Hence

$$T_*\mu([a, b]) = \frac{b}{2} - \frac{a}{2} + \left(1 - \frac{a}{2}\right) - \left(1 - \frac{b}{2}\right) = b - a = \mu([a, b]).$$

Hence μ is a T -invariant measure.

- (ii) Define $I(0) = [0, 1/2]$, $I(1) = [1/2, 1]$ and define the maps

$$\phi_0 : [0, 1] \rightarrow I(0) : x \mapsto \frac{x}{2}, \quad \phi_1 : [0, 1] \rightarrow I(1) : x \mapsto 1 - \frac{x}{2}.$$

Then $T\phi_0(x) = x$, $T\phi_1(x) = x$.

Given $i_0, \dots, i_{n-1} \in \{0, 1\}$ define

$$\phi_{i_0, i_1, \dots, i_{n-1}} = \phi_{i_0} \phi_{i_1} \cdots \phi_{i_{n-1}}$$

and note that $T^n \phi_{i_0, i_1, \dots, i_{n-1}}(x) = x$ for all $x \in [0, 1]$.

Define

$$I(i_0, i_1, \dots, i_{n-1}) = \phi_{i_0, i_1, \dots, i_{n-1}}([0, 1])$$

and call this a cylinder of rank n . It is easy to see that cylinders of rank n are dyadic intervals (although the labelling of these cylinders is not the same as the labelling that one gets when using the doubling map: for example, for the tent map $I(1, 1) = [1/2, 3/4]$ whereas for the doubling map $I(1, 1) = [3/4, 1]$). Hence the algebra \mathcal{A} of finite unions of cylinders generates the Borel σ -algebra.

Let $B \in \mathcal{B}$ be such that $T^{-1}B = B$. Note that $T^{-n}B = B$. Let $I = I(i_0, i_1, \dots, i_{n-1})$ be a cylinder of rank n and let $\phi = \phi_{i_0, i_1, \dots, i_{n-1}}$. Then $T^n\phi(x) = x$. Note also that $\mu(I) = 1/2^n$. We will also need the fact that $|\phi'(x)| = 1/2^n$ (this follows from noting that $|\phi'_0(x)| = |\phi'_1(x)| = 1/2$ and using the chain rule).

Finally, we observe that

$$\begin{aligned} \mu(B \cap I) &= \int \chi_{B \cap I}(x) dx \\ &= \int \chi_B(x) \chi_I(x) dx \end{aligned}$$

$$\begin{aligned}
&= \int_I \chi_B(x) dx \\
&= \int_0^1 \chi_B(\phi(x)) |\phi'(x)| dx \text{ by the change of variables formula} \\
&= \int_0^1 \chi_{T^{-n}B}(\phi(x)) |\phi'(x)| dx \text{ as } T^{-n}B = B \\
&= \int_0^1 \chi_B(T^n(\phi(x))) |\phi'(x)| dx \\
&= \int_0^1 \chi_B(x) |\phi'(x)| dx \text{ as } T^n\phi(x) = x \\
&= \frac{1}{2^n} \int_0^1 \chi_B(x) dx \text{ as } |\phi'(x)| = 1/2^n \\
&= \mu(I)\mu(B) \text{ as } \mu(I) = 1/2^n.
\end{aligned}$$

Hence $\mu(B \cap I) = \mu(B)\mu(I)$ for all sets I in the algebra of cylinders. By Lemma 6.1.1 it follows that $\mu(B) = 0$ or 1 . Hence Lebesgue measure is an ergodic measure for T .

Solution 6.2

For each $n \geq 1$ define $I(n) = [1/(n+1), 1/n]$ and define the maps

$$\phi_n : [0, 1] \rightarrow I(n) : x \mapsto \frac{x - n}{n(n+1)}.$$

Note that $T\phi_n(x) = x$ for all $x \in [0, 1]$.

Given $i_0, i_1, \dots, i_{n-1} \in \mathbb{N}$ define

$$\phi_{i_0, i_1, \dots, i_{n-1}} = \phi_{i_0} \phi_{i_1} \cdots \phi_{i_{n-1}}$$

and note that $T^n \phi_{i_0, i_1, \dots, i_{n-1}}(x) = x$ for all $x \in [0, 1]$.

Define $I(i_0, i_1, \dots, i_{n-1}) = \phi_{i_0, i_1, \dots, i_{n-1}}([0, 1])$ and call this a cylinder of rank n . Note that

$$\phi'_n(x) = \frac{1}{n(n+1)} \leq \frac{1}{2}$$

so that, by the chain rule,

$$\phi'_{i_0, i_1, \dots, i_{n-1}}(x) = \prod_{j=0}^{n-1} \frac{1}{i_j(i_j+1)} \leq \frac{1}{2^n}.$$

By the Intermediate Value Theorem, $I(i_0, i_1, \dots, i_{n-1})$ is an interval of length no more than $1/2^n$. For each n , the cylinders of rank n partition $[0, 1]$. Let $x, y \in [0, 1]$ and suppose that $x \neq y$. Choose n such that $|x - y| > 1/2^n$. Then x, y must lie in different cylinders of rank n . Hence the cylinders separate the points of $[0, 1]$. By Proposition 2.4.2 it follows that the algebra \mathcal{A} of finite unions of cylinders generates the Borel σ -algebra.

Let $B \in \mathcal{B}$ be such that $T^{-1}B = B$. Note that $T^{-n}B = B$. Let $I = I(i_0, i_1, \dots, i_{n-1})$ be a cylinder of rank n and let $\phi = \phi_{i_0, i_1, \dots, i_{n-1}}$. Then $T^n\phi(x) = x$. Note that

$$\mu(I) = \prod_{j=0}^{n-1} \frac{1}{i_j(i_j+1)} = \phi'(x) \tag{12.0.5}$$

for any $x \in [0, 1]$.

Finally, we observe that

$$\begin{aligned}
 \mu(B \cap I) &= \int \chi_{B \cap I}(x) dx \\
 &= \int \chi_B(x) \chi_I(x) dx \\
 &= \int_I \chi_B(x) dx \\
 &= \int_0^1 \chi_B(\phi(x)) |\phi'(x)| dx \text{ by the change of variables formula} \\
 &= \int_0^1 \chi_{T^{-n}B}(\phi(x)) |\phi'(x)| dx \text{ as } T^{-n}B = B \\
 &= \int_0^1 \chi_B(T^n(\phi(x))) |\phi'(x)| dx \\
 &= \int_0^1 \chi_B(x) |\phi'(x)| dx \\
 &= \mu(I) \mu(B) \text{ by (12.0.5)}.
 \end{aligned}$$

Hence $\mu(B \cap I) = \mu(B)\mu(I)$ for all sets I in the algebra of cylinders. By Lemma 6.1.1 it follows that $\mu(B) = 0$ or 1 . Hence Lebesgue measure is an ergodic measure for T .

Solution 6.3

(i) First note that

$$\frac{1}{x_0} = \frac{P_1}{Q_1}, \quad \frac{1}{x_0 + \frac{1}{x_1}} = \frac{x_1}{x_0 x_1 + 1} = \frac{P_2}{Q_2}.$$

If we define $P_0 = 0, Q_0 = 1$ then we have that $P_2 = x_1 P_1 + P_0$ and $Q_2 = x_1 Q_1 + Q_0$.

Similarly,

$$\frac{1}{x_0 + t} = \frac{P_1(x_0; t)}{Q_1(x_0; t)}, \quad \frac{1}{x_0 + \frac{1}{x_1 + t}} = \frac{x_1 + t}{x_0 x_1 + 1 + t} = \frac{P_2(x_0, x_1; t)}{Q_2(x_0, x_1; t)}.$$

then

$$P_2(x_0, x_1; t) = P_2 + tP_1, \quad Q_2(x_0, x_1; t) = Q_2 + tQ_1.$$

Suppose that $P_n(x_0, \dots, x_{n-1}) = P_n + tP_{n-1}$, $Q_n(x_0, \dots, x_{n-1}) = Q_n + tQ_{n-1}$. Then

$$\begin{aligned}
 \frac{P_{n+1}(x_0, x_1, \dots, x_n; t)}{Q_{n+1}(x_0, x_1, \dots, x_n; t)} &= [x_0, \dots, x_{n-1}, x_n + t] \\
 &= \left[x_0, \dots, x_{n-1} + \frac{1}{x_n + t} \right] \\
 &= \frac{P_n(x_0, x_1, \dots, x_{n-1}; \frac{1}{x_n + t})}{Q_n(x_0, x_1, \dots, x_{n-1}; \frac{1}{x_n + t})} \\
 &= \frac{P_n + \frac{1}{x_n + t} P_{n-1}}{Q_n + \frac{1}{x_n + t} Q_{n-1}} \\
 &= \frac{x_n P_n + P_{n-1} + t P_n}{x_n Q_n + Q_{n-1} + t Q_n}.
 \end{aligned}$$

Hence

$$P_{n+1}(x_0, x_1, \dots, x_n; t) = x_n P_n + P_{n-1} + t P_n, \quad Q_{n+1}(x_0, x_1, \dots, x_n; t) = x_n Q_n + Q_{n-1} + t Q_n.$$

Putting $t = 0$ we obtain the recurrence relations

$$P_{n+1} = x_n P_n + P_{n-1}, \quad Q_{n+1} = x_n Q_n + Q_{n-1}.$$

Hence

$$P_{n+1}(x_0, x_1, \dots, x_n; t) = P_{n+1} + t P_n, \quad Q_{n+1}(x_0, x_1, \dots, x_n; t) = Q_{n+1} + t Q_n.$$

By induction, the recurrence relations hold.

(ii) Note that

$$\begin{aligned} Q_n P_{n-1} - Q_{n-1} P_n &= (x_{n-1} Q_{n-1} + Q_{n-2}) P_{n-1} - Q_{n-1} (x_{n-1} P_{n-1} + P_{n-2}) \\ &= -(Q_{n-1} P_{n-2} - Q_{n-2} P_{n-1}) = \dots = (-1)^n. \end{aligned}$$

Solution 6.4

(i) Let $\mathbf{x} = (x_0, x_1, \dots), \mathbf{y} = (y_0, y_1, \dots) \in \Sigma$. Let $d_{\mathbb{R}/\mathbb{Z}}$ and d_Σ denote the usual metrics on \mathbb{R}/\mathbb{Z} and Σ , respectively. Now

$$d_{\mathbb{R}/\mathbb{Z}}(\pi(\mathbf{x}), \pi(\mathbf{y})) \tag{12.0.6}$$

$$\begin{aligned} &\leq |\pi(x_0, x_1, \dots) - \pi(y_0, y_1, \dots)| \\ &= \left| \frac{x_0 - y_0}{2} + \frac{x_1 - y_1}{2^2} + \dots \right| \\ &\leq \frac{|x_0 - y_0|}{2} + \frac{|x_1 - y_1|}{2^2} + \dots. \end{aligned} \tag{12.0.7}$$

Now if $d_\Sigma(\mathbf{x}, \mathbf{y}) < 1/2^n$ then $x_j = y_j$ for $j = 0, \dots, n$. Hence we can bound the right-hand side of (12.0.7) by

$$\begin{aligned} \frac{|x_{n+1} - y_{n+1}|}{2^{n+2}} + \frac{|x_{n+2} - y_{n+2}|}{2^{n+3}} + \dots &\leq \frac{1}{2^{n+2}} + \frac{1}{2^{n+3}} + \dots \\ &= \frac{1}{2^{n+2}} \left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots \right) \\ &\leq \frac{1}{2^{n+1}}, \end{aligned}$$

summing the geometric progression. This implies that π is continuous. To see this, let $\varepsilon > 0$. Choose n such that $1/2^{n+1} < \varepsilon$. Choose $\delta = 1/2^n$. If $d_\Sigma(\mathbf{x}, \mathbf{y}) < \delta$ then $d_{\mathbb{R}/\mathbb{Z}}(\pi(\mathbf{x}), \pi(\mathbf{y})) < \varepsilon$.

(ii) Observe that if $\mathbf{x} = (x_j)_{j=0}^\infty \in \Sigma$ then

$$\pi(\sigma(\mathbf{x})) = \pi(\sigma(x_0, x_1, \dots)) = \pi(x_1, x_2, \dots) = \frac{x_1}{2} + \frac{x_2}{2^2} + \dots$$

and

$$\begin{aligned} T(\pi(\mathbf{x})) &= T(\pi(x_0, x_1, \dots)) \\ &= T\left(\frac{x_0}{2} + \frac{x_1}{2^2} + \dots\right) \\ &= x_0 + \frac{x_1}{2} + \frac{x_2}{2^2} + \dots \pmod{1} \\ &= \frac{x_1}{2} + \frac{x_2}{2^2} + \dots \end{aligned}$$

(iii) We must show that $T_*(\pi_*\mu) = \pi_*\mu$. To see this, observe that

$$\begin{aligned} T_*(\pi_*\mu)(B) &= \pi_*\mu(T^{-1}B) \\ &= \mu(\pi^{-1}T^{-1}B) \\ &= \mu(\sigma^{-1}\pi^{-1}B) \text{ as } \pi\sigma = T\pi \\ &= (\sigma_*\mu)(\pi^{-1}B) \\ &= \mu(\pi^{-1}B) \text{ as } \mu \text{ is } \sigma\text{-invariant} \\ &= (\pi_*\mu)(B). \end{aligned}$$

(iv) Suppose that μ is an ergodic measure for σ . We claim that $\pi_*\mu$ is an ergodic measure for T , i.e. if $B \in \mathcal{B}(\mathbb{R}/\mathbb{Z})$ is such that $T^{-1}B = B$ then $\pi_*\mu(B) = 0$ or 1 .

First observe that $\pi^{-1}(B)$ is σ -invariant. This follows as:

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

As μ is an ergodic measure for σ , we must have that $\mu(\pi^{-1}(B)) = 0$ or 1 . Hence $\pi_*\mu(B) = 0$ or 1 .

(v) There are uncountably many different Bernoulli measures μ_p for Σ given by the family of probability vectors $(p, 1 - p)$. These are ergodic for σ . To see that $\pi_*\mu_p$ are all different, notice that $\pi_*\mu_p([0, 1/2)) = \mu_p(\pi^{-1}[0, 1/2)) = \mu_p([0]) = p$, where $[0]$ denotes the cylinder consisting of all sequences that start with 0.

Solution 7.1

Suppose that $x_n \rightarrow x$. We must show that $\delta_{x_n} \rightarrow \delta_x$. Let $f \in C(X, \mathbb{R})$. Then

$$\int f d\delta_{x_n} = f(x_n) \rightarrow f(x) = \int f d\delta_x$$

as f is continuous. Hence $\delta_{x_n} \rightarrow \delta_x$.

Solution 7.2

Suppose that $\mu_n \rightarrow \mu$. We must show that $T_*\mu_n \rightarrow T_*\mu$. Let $f \in C(X, \mathbb{R})$. Then

$$\lim_{n \rightarrow \infty} \int f d(T_*\mu_n) = \lim_{n \rightarrow \infty} \int f \circ T d\mu_n = \int f \circ T d\mu = \int f d(T_*\mu)$$

as $f \circ T$ is continuous. Hence $T_*\mu_n \rightarrow T_*\mu$ as $n \rightarrow \infty$.

Solution 7.3

(i) Suppose that $\mu_n \rightarrow \mu$ as $n \rightarrow \infty$. We claim that $\mu_n \rightarrow \mu$. To show this, we have to prove that if $f \in C(X, \mathbb{R})$ then $\int f d\mu_n \rightarrow \int f d\mu$ as $n \rightarrow \infty$.

Let $f \in C(X, \mathbb{R})$. Note that $f/\|f\|_\infty \in C(X, \mathbb{R})$ and that $\|(f/\|f\|_\infty)\|_\infty = 1$. Hence

$$\begin{aligned} \left| \int f d\mu_n - \int f d\mu \right| &= \|f\|_\infty \left| \int \frac{f}{\|f\|_\infty} d\mu_n - \int \frac{f}{\|f\|_\infty} d\mu \right| \\ &\leq \sup_{g \in C(X, \mathbb{R}), \|g\|_\infty \leq 1} \left| \int g d\mu_n - \int g d\mu \right| \\ &= \|\mu_n - \mu\|, \end{aligned}$$

which tends to 0 as $n \rightarrow \infty$.

(ii) Suppose that $x_n \rightarrow x$ but that $x_n \neq x$ for all n . We claim that $\delta_{x_n} \not\rightarrow \delta_x$. Note that

$$\|\delta_{x_n} - \delta_x\| = \sup_{f \in C(X, \mathbb{R}), \|f\|_\infty \leq 1} |f(x_n) - f(x)|.$$

For each n , we can choose a continuous function $f_n \in C(X, \mathbb{R})$ such that $f_n(x) = 1$, $f_n(x_n) = 0$ and $\|f_n\|_\infty \leq 1$. Hence

$$\begin{aligned} \|\delta_{x_n} - \delta_x\| &= \sup_{f \in C(X, \mathbb{R}), \|f\|_\infty \leq 1} |f(x_n) - f(x)| \\ &\geq |f_n(x_n) - f_n(x)| \\ &= 1. \end{aligned}$$

Hence $\delta_{x_n} \not\rightarrow \delta_x$.

(iii) First note that if $f \in C(X, \mathbb{R})$ is any continuous function with $\|f\|_\infty \leq 1$, then

$$\left| \int f d\delta_x - \int f d\delta_y \right| = |f(x) - f(y)| \leq |f(x)| + |f(y)| \leq 2.$$

Hence

$$\|\delta_x - \delta_y\| = \sup_{f \in C(X, \mathbb{R}), \|f\|_\infty \leq 1} \left| \int f d\delta_x - \int f d\delta_y \right| \leq 2.$$

Conversely, by Urysohn's Lemma, there exist continuous functions g_1, g_2 such that $g_1(x) = g_2(y) = 1, g_1(y) = g_2(x) = 0$ and $0 \leq g_1, g_2 \leq 1$. Let $h = g_1 - g_2$. Then $h(x) = 1, h(y) = -1$ and $-1 \leq h \leq 1$ (so that $\|h\|_\infty = 1$). Hence

$$\begin{aligned} 2 &= |h(x) - h(y)| \\ &= \left| \int h d\delta_x - \int h d\delta_y \right| \\ &\leq \sup_{f \in C(X, \mathbb{R}), \|f\|_\infty \leq 1} \left| \int f d\delta_x - \int f d\delta_y \right| \\ &= \|\delta_x - \delta_y\|. \end{aligned}$$

Hence if $x \neq y$ then $\|\delta_x - \delta_y\| = 2$.

Suppose that X is infinite. Let $x_n \in X$ be pairwise distinct and consider the sequence $\mu_n = \delta_{x_n}$. Then $\|\mu_n - \mu_m\| = 2$ if $n \neq m$. Hence μ_n cannot have a convergent subsequence, and so $M(X)$ is not compact in the strong topology when X is infinite.

Solution 7.4

Let $x_n \in X$ be a sequence such that $x_n \rightarrow x$ and $x_n \neq x$ for all n . Let $\mu_n = \delta_{x_n}$ and $\mu = \delta_x$. Then $\mu_n \rightarrow \mu$. Take $B = \{x\}$. Then $\mu_n(B) = 0$ but $\mu(B) = 1$. Hence $\mu_n(B) \not\rightarrow \mu(B)$.

Solution 7.5

Let $\mu_1, \mu_2 \in M(X, T)$ and suppose that $\alpha \in [0, 1]$. Then $\alpha\mu_1 + (1 - \alpha)\mu_2 \in M(X)$. To check that $\alpha\mu_1 + (1 - \alpha)\mu_2 \in M(X, T)$, note that

$$\begin{aligned} (T_*(\alpha\mu_1 + (1 - \alpha)\mu_2))(B) &= (\alpha\mu_1 + (1 - \alpha)\mu_2)(T^{-1}B) \\ &= \alpha\mu_1(T^{-1}B) + (1 - \alpha)\mu_2(T^{-1}B) \\ &= \alpha\mu_1(B) + (1 - \alpha)\mu_2(B) \\ &= (\alpha\mu_1 + (1 - \alpha)\mu_2)(B). \end{aligned}$$

Solution 7.6

Let $\mathcal{S} \subset C(X, \mathbb{R})$ be uniformly dense. Let $f \in C(X, \mathbb{R})$. Let $\varepsilon > 0$. Choose $g \in \mathcal{S}$ such that $\|f - g\|_\infty < \varepsilon$. Choose N such that if $n \geq N$ then $|\int f d\mu_n - \int f d\mu| < \varepsilon$. Then

$$\begin{aligned} & \left| \int f d\mu_n - \int f d\mu \right| \\ & \leq \left| \int f d\mu_n - \int g d\mu_n \right| + \left| \int g d\mu_n - \int g d\mu \right| + \left| \int f d\mu - \int g d\mu \right| \\ & \leq \int |f - g| d\mu_n + \left| \int g d\mu_n - \int g d\mu \right| + \int |f - g| d\mu \\ & \leq 3\varepsilon. \end{aligned}$$

As $\varepsilon > 0$ is arbitrary, the result follows.

Solution 7.7

- (i) Recall that $\mathbf{x} \in \Sigma$ is a periodic point with period n if $\sigma^n(\mathbf{x}) = \mathbf{x}$. If \mathbf{x} is periodic with period n then $x_{j+n} = x_j$ for all $j = 0, 1, 2, \dots$. Hence \mathbf{x} is determined by the first n symbols, which then repeat. As there are two choices for each x_j , there are 2^n periodic points with period n .
- (ii) First note that μ_n is a Borel probability measure.

Let $[i_0, i_1, \dots, i_{m-1}]$ be a cylinder. Let $n \geq m$. Then the periodic points \mathbf{x} of period n in $[i_0, i_1, \dots, i_{m-1}]$ have the form

$$\mathbf{x} = (i_0, i_1, \dots, i_{m-1}, x_m, \dots, x_{n-1}, i_0, i_1, \dots, i_{m-1}, x_m, \dots, x_{n-1}, i_0, \dots)$$

where the finite string of symbols $i_0, i_1, \dots, i_{m-1}, x_m, \dots, x_{n-1}$ repeats. The symbols x_m, \dots, x_{n-1} can be chosen arbitrarily. Hence there are 2^{n-m} such periodic points.

Hence, if $n \geq m$,

$$\int \chi_{[i_0, i_1, \dots, i_{m-1}]} d\mu_n = \frac{1}{2^n} \times 2^{n-m} = \frac{1}{2^m} = \int \chi_{[i_0, i_1, \dots, i_{m-1}]} d\mu.$$

- (iii) To prove that $\chi_{[i_0, i_1, \dots, i_{m-1}]}$ is continuous we need to show that, if $\mathbf{x}_n \rightarrow \mathbf{x}$ then $\chi_{[i_0, i_1, \dots, i_{m-1}]}(\mathbf{x}_n) \rightarrow \chi_{[i_0, i_1, \dots, i_{m-1}]}(\mathbf{x})$.

First suppose that $\mathbf{x} \in [i_0, i_1, \dots, i_{m-1}]$. As $\mathbf{x}_n \rightarrow \mathbf{x}$, it follows from the definition of the metric on Σ that there exists $N \in \mathbb{N}$ such that if $n \geq N$ then \mathbf{x}_n and \mathbf{x} agree in the first m places. Hence if $n \geq N$ then $\mathbf{x}_n \in [i_0, i_1, \dots, i_{m-1}]$. Hence, if $n \geq N$, then $\chi_{[i_0, i_1, \dots, i_{m-1}]}(\mathbf{x}_n) = 1 = \chi_{[i_0, i_1, \dots, i_{m-1}]}(\mathbf{x})$.

Now suppose that $\mathbf{x} \notin [i_0, i_1, \dots, i_{m-1}]$. As $\mathbf{x}_n \rightarrow \mathbf{x}$, it follows from the definition of the metric on Σ that there exists $N \in \mathbb{N}$ such that if $n \geq N$ then \mathbf{x}_n and \mathbf{x} agree in the first m places. Hence if $n \geq N$, there exists $j \in \{0, 1, \dots, m-1\}$ such that $(\mathbf{x}_n)_j \neq i_j$; that is, if $n \geq N$, then $\mathbf{x}_n \notin [i_0, i_1, \dots, i_{m-1}]$. Hence, if $n \geq N$, then $\chi_{[i_0, i_1, \dots, i_{m-1}]}(\mathbf{x}_n) = 0 = \chi_{[i_0, i_1, \dots, i_{m-1}]}(\mathbf{x})$.

Hence $\chi_{[i_0, i_1, \dots, i_{m-1}]}$ is continuous.

- (iv) Let \mathcal{S} denote the set of finite linear combinations of characteristic functions of cylinders. By the Stone-Weierstrass Theorem, \mathcal{S} is uniformly dense in $C(X, \mathbb{R})$. By (ii) above, if $g \in \mathcal{S}$ then $\int g d\mu_n \rightarrow \int g d\mu$ as $n \rightarrow \infty$. Let $f \in C(X, \mathbb{R})$ and let $\varepsilon > 0$. Choose $g \in \mathcal{S}$ such that $\|f - g\|_\infty < \varepsilon$. Then a 3ε argument as in the solutions to Exercise 7.6 proves that $\limsup_{n \rightarrow \infty} |\int f d\mu_n - \int f d\mu| < 3\varepsilon$ and the result follows.

Solution 7.8

As trigonometric polynomials are uniformly dense in $C(X, \mathbb{R})$, it is sufficient to prove that $\int g \circ T d\mu = \int g d\mu$ for all trigonometric polynomials g . Let $g(x) = \sum_{j=0}^r c_j e^{2\pi i \langle \mathbf{n}^{(j)}, x \rangle}$, $c_j \in \mathbb{R}$, $\mathbf{n}^{(j)} = (n_1^{(j)}, n_2^{(j)}, n_3^{(j)}) \in \mathbb{Z}^3$ be a trigonometric polynomial. We label the coefficients so that $\mathbf{n}^{(j)} = \mathbf{0}$ if and only if $j = 0$. Then $\int g d\mu = c_0$.

Note that

$$\begin{aligned} g \circ T \left(\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbb{Z}^3 \right) \right) &= g \left(\left(\begin{pmatrix} \alpha + x \\ x + y \\ y + z \end{pmatrix} + \mathbb{Z}^3 \right) \right) \\ &= \sum_{j=0}^r c_j e^{2\pi i \langle (n_1^{(j)}, n_2^{(j)}, n_3^{(j)}), (\alpha + x, x + y, y + z) \rangle} \\ &= \sum_{j=0}^r c_j e^{2\pi i n_1^{(j)} \alpha} e^{2\pi i (n_1^{(j)} x + n_2^{(j)} (x + y) + n_3^{(j)} (y + z))} \\ &= \sum_{j=0}^r c_j e^{2\pi i n_1^{(j)} \alpha} e^{2\pi i \langle (n_1^{(j)} + n_2^{(j)}, n_2^{(j)} + n_3^{(j)}, n_3^{(j)}), (x, y, z) \rangle}. \end{aligned}$$

Hence

$$\begin{aligned} \int g \circ T d\mu &= \int \sum_{j=0}^r c_j e^{2\pi i n_1^{(j)} \alpha} e^{2\pi i \langle (n_1^{(j)} + n_2^{(j)}, n_2^{(j)} + n_3^{(j)}, n_3^{(j)}), (x, y, z) \rangle} d\mu \\ &= \sum_{j=0}^r c_j e^{2\pi i n_1^{(j)} \alpha} \int e^{2\pi i \langle (n_1^{(j)} + n_2^{(j)}, n_2^{(j)} + n_3^{(j)}, n_3^{(j)}), (x, y, z) \rangle} d\mu. \end{aligned}$$

The integral is equal to zero unless $(n_1^{(j)} + n_2^{(j)}, n_2^{(j)} + n_3^{(j)}, n_3^{(j)}) = (0, 0, 0)$, i.e. unless $n_1^{(j)} = n_2^{(j)} = n_3^{(j)} = 0$. By our choice of labelling the coefficients, this only happens if $j = 0$. Hence

$$\int g \circ T d\mu = c_0 = \int g d\mu.$$

Solution 8.1

(i) Let $B \in \mathcal{B}$ and let $f = \chi_B$. Note that

$$\int f d\nu = \int \chi_B d\nu = \int_B d\nu = \nu(B) = \int_B \frac{d\nu}{d\mu} d\mu = \int \chi_B \frac{d\nu}{d\mu} d\mu = \int f \frac{d\nu}{d\mu} d\mu.$$

Hence the result holds for characteristic functions, hence for simple functions (finite linear combinations of characteristic functions). Let $f \in L^1(X, \mathcal{B}, \mu)$ be such that $f \geq 0$. By considering an increasing sequence of simple functions, the result follows for positive L^1 functions. By splitting an arbitrary real-valued L^1 function into its positive and negative parts, and then an arbitrary $L^1(X, \mathcal{B}, \mu)$ function into its real and imaginary parts, the result holds.

(ii) Now $d\nu_1/d\mu$, $d\nu_2/d\mu$ are the unique functions such that

$$\nu_1(B) = \int_B \frac{d\nu_1}{d\mu} d\mu, \quad \nu_2(B) = \int_B \frac{d\nu_2}{d\mu} d\mu,$$

respectively. Hence

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

However,

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

Hence, by uniqueness in the Radon-Nikodym theorem, we have that

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

(iii) Suppose that $\mu(B) = 0$. As $\nu \ll \mu$ then $\nu(B) = 0$. As $\lambda \ll \nu$ then $\lambda(B) = 0$. Hence $\lambda \ll \mu$.

Now as $\lambda \ll \mu$ we have

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

As $\lambda \ll \nu$ we have

$$\lambda(B) = \int_B \frac{d\lambda}{d\nu} d\nu = \int \chi_B \frac{d\lambda}{d\nu} d\nu.$$

By part (i), using the fact that $\nu \ll \mu$, it follows that

$$\int \chi_B \frac{d\lambda}{d\nu} d\nu = \int \chi_B \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu} d\mu = \int_B \frac{d\lambda}{d\nu} \frac{d\nu}{d\mu} d\mu.$$

Hence by uniqueness in the Radon-Nikodym theorem,

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

Solution 8.2

The claimed formula is easily seen to be valid for $n = 3$. Suppose the formula is valid for n . Then

$$\begin{aligned} & T^{n+1} \left(\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbb{Z}^3 \right) \right) \\ &= TT^n \left(\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbb{Z}^3 \right) \right) \\ &= T \left(\left(\left(\begin{pmatrix} \binom{n}{1} \alpha + x \\ \binom{n}{2} \alpha + \binom{n}{1} x + y \\ \binom{n}{3} \alpha + \binom{n}{2} x + \binom{n}{1} y + z \end{pmatrix} + \mathbb{Z}^3 \right) \right) \right) \end{aligned}$$

$$\begin{aligned}
&= \left(\left(\left(\binom{n}{1} \alpha + x + \alpha \right) \right. \right. \\
&\quad \left. \left(\binom{n}{2} \alpha + \binom{n}{1} x + y + \binom{n}{1} \alpha + x \right) \right. \\
&\quad \left. \left. \left(\binom{n}{3} \alpha + \binom{n}{2} x + \binom{n}{1} y + z + \binom{n}{2} \alpha + \binom{n}{1} x + y \right) + \mathbb{Z}^3 \right) \right) \\
&= \left(\left(\left(\binom{n+1}{1} \alpha + x \right) \right. \right. \\
&\quad \left. \left(\binom{n+1}{2} \alpha + \binom{n+1}{1} x + y \right) \right. \\
&\quad \left. \left. \left(\binom{n+1}{3} \alpha + \binom{n+1}{2} x + \binom{n+1}{1} y + z \right) + \mathbb{Z}^3 \right) \right).
\end{aligned}$$

Hence the claimed formula holds by induction.

Let $f(x, y, z) = e^{2\pi i(kx + \ell y + mz)}$. Then

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j((x, y, z) + \mathbb{Z}^3)) = \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i p_{x,y,z}^{(k,\ell,m)}(j)}$$

where $p_{x,y,z}^{(k,\ell,m)}(n)$ is a polynomial. When $m \neq 0$, $p_{x,y,z}^{(k,\ell,m)}(n)$ is a degree 3 polynomial with leading coefficient $m\alpha/6 \notin \mathbb{Q}$. When $m = 0, \ell \neq 0$, $p_{x,y,z}^{(k,\ell,m)}(n)$ is a degree 2 polynomial with leading coefficient $\ell\alpha/2 \notin \mathbb{Q}$. When $m = \ell = 0, k \neq 0$, $p_{x,y,z}^{(k,\ell,m)}(n)$ is a degree 1 polynomial with leading coefficient $k\alpha \notin \mathbb{Q}$. In all three cases, $p_{x,y,z}^{(k,\ell,m)}(n)$ is uniformly distributed mod 1, by Weyl's Theorem on Polynomials (Theorem 2.3.1). Hence by Weyl's Criterion (Theorem 1.2.1) for all $(k, \ell, m) \in \mathbb{Z}^3 \setminus \{(0, 0, 0)\}$ we have

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i p_{x,y,z}^{(k,\ell,m)}(j)} \rightarrow 0$$

as $n \rightarrow \infty$. When $k = \ell = m = 0$ we trivially have that

$$\frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i p_{x,y,z}^{(k,\ell,m)}(j)} = \frac{1}{n} \sum_{j=0}^{n-1} 1 \rightarrow 1$$

as $n \rightarrow \infty$. Hence

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j((x, y, z) + \mathbb{Z}^3)) \rightarrow \int f d\mu$$

whenever $f(x, y, z) = e^{2\pi i(kx + \ell y + mz)}$.

By taking finite linear combinations of exponential functions we see that

$$\sup_{x \in X} \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right| \rightarrow 0$$

as $n \rightarrow \infty$ for all trigonometric polynomials g . By the Stone-Weierstrass Theorem (Theorem 1.2.2), trigonometric polynomials are uniformly dense in $C(X, \mathbb{R})$. Let $f \in C(X, \mathbb{R})$

and let $\varepsilon > 0$. Then there exists a trigonometric polynomial g such that $\|f - g\|_\infty < \varepsilon$. Hence for any $x \in X$ we have

$$\begin{aligned} & \left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right| \\ & \leq \left| \frac{1}{n} \sum_{j=0}^{n-1} (f(T^j(x)) - g(T^j(x))) \right| + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right| + \left| \int g - f d\mu \right| \\ & \leq \frac{1}{n} \sum_{j=0}^{n-1} |f(T^j(x)) - g(T^j(x))| + \left| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right| + \int |g - f| d\mu \\ & \leq 2\varepsilon + \left\| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right\|_\infty. \end{aligned}$$

Hence, taking the supremum over all $x \in X$, we have

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_\infty \leq 2\varepsilon + \left\| \frac{1}{n} \sum_{j=0}^{n-1} g(T^j(x)) - \int g d\mu \right\|_\infty.$$

Letting $n \rightarrow \infty$ we see that

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_\infty \leq 2\varepsilon.$$

As $\varepsilon > 0$ is arbitrary, it follows that

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) - \int f d\mu \right\|_\infty = 0.$$

Hence statement (ii) in Oxtoby's Ergodic Theorem holds. As (i) and (ii) in Oxtoby's Ergodic Theorem are equivalent, it follows that T is uniquely ergodic and Lebesgue measure is the unique invariant measure.

Solution 8.3

Let T be a uniquely ergodic homeomorphism with unique invariant measure μ .

Suppose that every orbit is dense. Let U be a non-empty open set. Then for all $x \in X$, there exists $n \in \mathbb{Z}$ such that $T^n(x) \in U$. Hence $X = \bigcup_{n=-\infty}^{\infty} T^{-n}U$. Hence

$$1 = \mu(X) = \mu \left(\bigcup_{n=-\infty}^{\infty} T^{-n}U \right) \leq \sum_{n=-\infty}^{\infty} \mu(T^{-n}U) = \sum_{n=-\infty}^{\infty} \mu(U)$$

as μ is T -invariant. Hence $\mu(U) > 0$.

Conversely, suppose that $\mu(U) > 0$ for all non-empty open sets. Suppose for a contradiction that there exists $x_0 \in X$ such that the orbit of x_0 is not dense.

Clearly $\{T^n(x_0) \mid n \in \mathbb{Z}\}$ is T -invariant. As T is continuous, the set

$$Y = \text{cl}\{T^n(x_0) \mid n \in \mathbb{Z}\}$$

is also T -invariant. As the orbit of x_0 is not dense, Y is a proper subset of X . As Y is closed and X is compact, it follows that Y is compact. By Theorem 7.5.1 there exists an invariant probability measure ν for the map $T : Y \rightarrow Y$. Extend ν to X by setting $\nu(B) = \nu(B \cap Y)$ for Borel subsets $B \subset X$. Noting that $X \setminus Y$ is also T -invariant, it follows that ν is an invariant measure for $T : X \rightarrow X$. This contradicts unique ergodicity as $\nu(X \setminus Y) = 0$ but $\mu(X \setminus Y) > 0$.

Solution 9.1

Let $X = \mathbb{R}$ equipped with the Borel σ -algebra and Lebesgue measure. Define $T(x) = x + 1$. Then Lebesgue measure is T -invariant. Take $A = [0, 1)$. Then A has positive measure, but no point of A returns to A under T .

Solution 9.2

Take $X = \{0, 1\}$ to be a set consisting of two elements. Let \mathcal{B} be the set of all subsets of X and equip X with the measure $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ that assigns measure $1/2$ to both 0 and 1 . Take $T(x) = x$ to be the identity. Then T is a measure-preserving transformation. Let $A = \{0\}$, $B = \{1\}$. Then $\mu(A) = \mu(B) = 1/2 > 0$. However, $T^j(0)$ never lands in B .

Solution 9.3

Recall that $E(f | \mathcal{A})$ is determined as being the unique \mathcal{A} -measurable function such that

$$\int_A E(f | \mathcal{A}) d\mu = \int_A f d\mu$$

for all $A \in \mathcal{A}$.

(i) We need to show that

$$E(\alpha f + \beta g | \mathcal{A}) = \alpha E(f | \mathcal{A}) + \beta E(g | \mathcal{A}).$$

Note that $\alpha E(f | \mathcal{A}) + \beta E(g | \mathcal{A})$ is \mathcal{A} -measurable. Moreover, as

$$\begin{aligned} \int_A \alpha E(f | \mathcal{A}) + \beta E(g | \mathcal{A}) d\mu &= \alpha \int_A E(f | \mathcal{A}) d\mu + \beta \int_A E(g | \mathcal{A}) d\mu \\ &= \alpha \int_A f d\mu + \beta \int_A g d\mu \\ &= \int_A \alpha f + \beta g d\mu \\ &= \int_A E(\alpha f + \beta g | \mathcal{A}) d\mu \end{aligned}$$

for all $A \in \mathcal{A}$, the claim follows.

(ii) First note that $E(f | \mathcal{A}) \circ T$ is $T^{-1}\mathcal{A}$ -measurable. To see this, note that $E(f | \mathcal{A})$ is \mathcal{A} -measurable, i.e.

$$\{x \in X \mid E(f | \mathcal{A})(x) \leq c\} \in \mathcal{A} \text{ for all } c \in \mathbb{R}.$$

Hence

$$\{x \in X \mid E(f | \mathcal{A})(Tx) \leq c\} = T^{-1}\{x \in X \mid E(f | \mathcal{A})(x) \leq c\} \in T^{-1}\mathcal{A}$$

so that $E(f | \mathcal{A}) \circ T$ is $T^{-1}\mathcal{A}$ -measurable.

Note that for any $A \in \mathcal{A}$

$$\begin{aligned} \int_{T^{-1}A} E(f | \mathcal{A}) \circ T d\mu &= \int \chi_{T^{-1}A} E(f | \mathcal{A}) \circ T d\mu \\ &= \int \chi_A \circ T \cdot E(f | \mathcal{A}) \circ T d\mu \\ &= \int \chi_A E(f | \mathcal{A}) d\mu \text{ as } \mu \text{ is } T\text{-invariant} \\ &= \int_A E(f | \mathcal{A}) d\mu \\ &= \int_A f d\mu. \end{aligned}$$

Moreover

$$\begin{aligned} \int_{T^{-1}A} E(f \circ T | T^{-1}\mathcal{A}) d\mu &= \int_{T^{-1}A} f \circ T d\mu \\ &= \int \chi_{T^{-1}A} f \circ T d\mu \\ &= \int \chi_A \circ T \cdot f \circ T d\mu \\ &= \int \chi_A f d\mu \\ &= \int_A f d\mu. \end{aligned}$$

Hence

$$\int_{T^{-1}A} E(f \circ T | T^{-1}\mathcal{A}) d\mu = \int_{T^{-1}A} E(f | \mathcal{A}) \circ T d\mu$$

for all $A \in \mathcal{A}$. By the characterisation of conditional expectation, it follows that

$$E(f \circ T | T^{-1}\mathcal{A}) = E(f | \mathcal{A}) \circ T.$$

- (iii) That $E(f | \mathcal{B}) = f$ is immediate from the above characterisation of conditional expectation.
- (iv) Recall that a function $f : X \rightarrow \mathbb{R}$ is \mathcal{A} -measurable if $f^{-1}(-\infty, c) \in \mathcal{A}$ for all $c \in \mathbb{R}$.

Suppose that f is \mathcal{N} -measurable. Let $B_c = f^{-1}(-\infty, c) \in \mathcal{N}$. Hence $\mu(B_c) = 0$ or 1 . Note that $c_1 < c_2$ implies $B_{c_1} \subset B_{c_2}$. Hence there exists c_0 such that

$$c_0 = \sup\{c \mid \mu(B_c) = 0\} = \inf\{c \mid \mu(B_c) = 1\}.$$

We claim that $f(x) = c_0$ μ -a.e. If $c < c_0$ then $\mu(\{x \in X \mid f(x) < c\}) = 0$. Hence $f(x) \geq c_0$ μ -a.e. Let $c > c_0$. Then

$$\mu(\{x \in X \mid f(x) \geq c\}) = \mu(X \setminus \{x \in X \mid f(x) < c\}) = 1 - \mu(\{x \in X \mid f(x) < c\}) = 0.$$

Hence $\mu(\{x \in X \mid f(x) > c_0\}) = 0$. Hence $f(x) = c_0$ μ -a.e.

Suppose that f is constant almost everywhere, say $f(x) = a$ μ -a.e. Then $f^{-1}(-\infty, c) = \emptyset$ μ -a.e. if $c < a$ and $f^{-1}(-\infty, c) = X$ μ -a.e. if $c > a$. Hence $\mu(f^{-1}(-\infty, c)) = 0$ or 1 . Hence $f^{-1}(-\infty, c) \in \mathcal{N}$ for all $c \in \mathbb{R}$. Hence f is \mathcal{N} -measurable.

If $N \in \mathcal{N}$ has measure 0 then

$$\int_N f d\mu = 0 = \int_N \left(\int f d\mu \right) d\mu$$

and if $N \in \mathcal{N}$ has measure 1 then

$$\int_N f d\mu = \int f d\mu = \int_N \left(\int f d\mu \right) d\mu.$$

Hence $E(f | \mathcal{N}) = \int f d\mu$.

Solution 9.4

- (i) Let $\alpha = \{A_1, \dots, A_n\}$ be a finite partition of X into sets $A_j \in \mathcal{B}$ and let \mathcal{A} be the set of all finite unions of sets in α .

Trivially $\emptyset \in \mathcal{A}$.

Let $B_j = \bigcup_{i=1}^{\ell_j} A_{i,j}$, $A_{i,j} \in \alpha$, be a countable collection of finite unions of sets in α . Then $\bigcup_j B_j$ is a union of sets in α . As there are only finitely many sets in α , we have that $\bigcup_j B_j$ is a finite union of sets in α . Hence $\bigcup_j B_j \in \mathcal{A}$.

It is clear that \mathcal{A} is closed under taking complements.

Hence \mathcal{A} is a σ -algebra.

- (ii) Recall that $g : X \rightarrow \mathbb{R}$ is \mathcal{A} -measurable if $g^{-1}(-\infty, c) \in \mathcal{A}$ for all $c \in \mathbb{R}$.

Suppose that g is constant on each $A_j \in \alpha$ and write $g(x) = \sum_j c_j \chi_{A_j}(x)$. Then $g^{-1}(-\infty, c) = \bigcup A_j$ where the union is taken over sets A_j for which $c_j < c$. Hence g is \mathcal{A} -measurable.

Conversely, suppose that g is \mathcal{A} -measurable. For each $c \in \mathbb{R}$, let $A_c = g^{-1}(-\infty, c)$. Then $A_c \in \mathcal{A}$. Moreover, $A_c \downarrow \emptyset$ as $c \rightarrow -\infty$ (in the sense that $\bigcap_{c \in \mathbb{R}} A_c = \emptyset$) and $A_c \uparrow X$ as $c \rightarrow \infty$ (in the sense that $\bigcup_c A_c = X$). Let $A \in \alpha$. Then there exists c_0 such that $A \not\subset A_c$ for $c < c_0$ and $A \subset A_c$ for $c > c_0$. Hence $g(x) = c_0$ for all $x \in A$. Hence g is constant on each element of α .

- (iii) Define g by

$$g(x) = \sum_{j=1}^n \chi_{A_j}(x) \frac{\int_{A_j} f d\mu}{\mu(A_j)}.$$

Then g is constant on each set in α , hence g is \mathcal{A} -measurable.

Let $A_i \in \alpha$. Then

$$\int_{A_i} g d\mu = \sum_{j=1}^n \int \chi_{A_i} \chi_{A_j} \frac{\int_{A_j} f d\mu}{\mu(A_j)} d\mu = \sum_{j=1}^n \int \chi_{A_i \cap A_j} \frac{\int_{A_j} f d\mu}{\mu(A_j)} d\mu = \int_{A_i} f d\mu.$$

Hence $\int_A g d\mu = \int_A f d\mu$ for all $A \in \mathcal{A}$. It follows that $g = E(f | \mathcal{A})$.

Solution 9.5

Clearly $\emptyset \in \mathcal{I}$.

Let $I \in \mathcal{I}$, so that $T^{-1}(I) = I$. Then $T^{-1}(X \setminus I) = X \setminus I$, so that the complement of I is in \mathcal{I} .

Let $I_n \in \mathcal{I}$. Then $T^{-1}(\bigcup_n I_n) = \bigcup_n T^{-1}I_n = \bigcup_n I_n$ so that $\bigcup_n I_n \in \mathcal{I}$.

Hence \mathcal{I} is a σ -algebra.

Solution 9.6

Recall that $E(f | \mathcal{I})$ is determined by the requirements that $E(f | \mathcal{I})$ is \mathcal{I} -measurable and that

$$\int_I E(f | \mathcal{I}) d\mu = \int_I f d\mu$$

for all $I \in \mathcal{I}$. Let $P_{\mathcal{I}} : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(X, \mathcal{I}, \mu)$ denote the orthogonal projection onto the subspace of \mathcal{I} -measurable functions. To show that $P_{\mathcal{I}}f = E(f | \mathcal{I})$ it is thus sufficient to check that for each $I \in \mathcal{I}$ we have

$$\int_I P_{\mathcal{I}}f d\mu = \int_I f d\mu$$

for all $f \in L^2(X, \mathcal{B}, \mu)$.

Note that $\int_I P_{\mathcal{I}}f d\mu = \int \chi_I P_{\mathcal{I}}f d\mu = \langle \chi_I, P_{\mathcal{I}}f \rangle$ and, similarly, $\int_I f d\mu = \langle \chi_I, f \rangle$, where we use $\langle \cdot, \cdot \rangle$ to denote the inner product on $L^2(X, \mathcal{B}, \mu)$. Hence it is sufficient to prove that, for all $I \in \mathcal{I}$, $\langle \chi_I, f - P_{\mathcal{I}}f \rangle = 0$.

It is proved in the proof of Theorem 9.6.1 that $L^2(X, \mathcal{B}, \mu) = L^2(X, \mathcal{I}, \mu) \oplus C$ where C denotes the norm-closure of the subspace $\{w \circ T - w \mid w \in L^2(X, \mathcal{B}, \mu)\}$. Hence it is sufficient to prove that $\langle \chi_I, g \rangle = 0$ for all $g \in C$. To see this, first note that for $w \in L^2(X, \mathcal{B}, \mu)$ we have that

$$\begin{aligned} \langle \chi_I, w \circ T - w \rangle &= \langle \chi_I, w \circ T \rangle - \langle \chi_I, w \rangle \\ &= \langle \chi_{T^{-1}I}, w \circ T \rangle - \langle \chi_I, w \rangle \\ &= \langle \chi_I \circ T, w \circ T \rangle - \langle \chi_I, w \rangle = 0, \end{aligned}$$

using the facts that $I = T^{-1}I$ a.e. and that T is measure-preserving. It follows that $\langle \chi_I, g \rangle = 0$ for all $g \in C$.

Solution 10.1

Let T be an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) and let $f \in L^1(X, \mathcal{B}, \mu)$. Let

$$S_n = \sum_{j=0}^{n-1} f(T^j x).$$

By Birkhoff's Ergodic Theorem, there exists a set N such that $\mu(N) = 0$ and if $x \notin N$ then $S_n/n \rightarrow \int f d\mu$ as $n \rightarrow \infty$. Let $x \notin N$. Note that

$$\frac{n+1}{n} \frac{S_{n+1}}{n+1} = \frac{f(T^n x)}{n} + \frac{S_n}{n}.$$

Letting $n \rightarrow \infty$ we have that $(n+1)/n \rightarrow 1$, $\frac{1}{n+1}S_{n+1} \rightarrow \int f d\mu$ and $\frac{1}{n}S_n \rightarrow \int f d\mu$ as $n \rightarrow \infty$. Hence if $x \notin N$ then $f(T^n x)/n \rightarrow 0$ as $n \rightarrow \infty$. Hence $f(T^n x)/n \rightarrow 0$ as $n \rightarrow \infty$ for μ -a.e. $x \in X$.

Solution 10.2

Let $f \geq 0$ be measurable and suppose that $\int f d\mu = \infty$. For each integer $M > 0$ define $f_M(x) = \min\{f(x), M\}$. Then $0 \leq f_M \leq M$, hence $f_M \in L^1(X, \mathcal{B}, \mu)$. Moreover $f_M(x) \uparrow f(x)$ as $M \rightarrow \infty$ for all $x \in X$. Hence by the Monotone Convergence Theorem (Theorem 3.1.2), $\int f_M d\mu \rightarrow \int f d\mu = \infty$.

By Birkhoff's Ergodic Theorem, there exists $N_M \subset X$ with $\mu(N_M) = 0$ such that for all $x \notin N_M$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_M(T^j x) = \int f_M d\mu. \quad (12.0.8)$$

Let $N = \bigcup_{M=1}^{\infty} N_M$. Then $\mu(N) = 0$. Moreover, for any $M > 0$ we have that if $x \notin N$ then (12.0.8) holds.

Let $K \geq 0$ be arbitrary. As $\int f_M d\mu \rightarrow \infty$, it follows that there exists $M > 0$ such that $\int f_M d\mu \geq K$. Hence for all $x \notin N$ we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_M(T^j x) = \int f_M d\mu \geq K.$$

As K is arbitrary, we have that for all $x \notin N$

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

Hence $\frac{1}{n} \sum_{j=0}^{\infty} f(T^j x) \rightarrow \infty$ for μ -a.e. $x \in X$.

Solution 10.3

We prove that (i) implies (ii). Suppose that T is an ergodic measure-preserving transformation of the probability space (X, \mathcal{B}, μ) . Recall from Proposition 10.2.2 that for all $A, B \in \mathcal{B}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j} A \cap B) = \mu(A)\mu(B),$$

Equivalently, for all $A, B \in \mathcal{B}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int \chi_A(T^j x) \chi_B(x) d\mu = \int \chi_A d\mu \int \chi_B d\mu. \quad (12.0.9)$$

Let $f(x) = \sum_{k=1}^r c_k \chi_{A_k}(x)$ be a simple function. Then taking linear combinations of expressions of the form (12.0.9) we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int f(T^j x) \chi_B(x) d\mu = \int f d\mu \int \chi_B d\mu.$$

If $f \geq 0$ is a positive measurable function then we can choose a sequence of simple functions $f_n \uparrow f$ that increase pointwise to f . By the Monotone Convergence Theorem (Theorem 3.1.2) we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int f(T^j x) \chi_B(x) d\mu = \int f d\mu \int \chi_B d\mu \quad (12.0.10)$$

for all positive measurable functions f . Suppose that $f \in L^1(X, \mathcal{B}, \mu)$ is real-valued. Then by writing $f = f^+ - f^-$ where f^+, f^- are positive, we have that (12.0.10) holds when f is integrable and real-valued. By taking real and imaginary parts of f , we have that (12.0.10) holds for all $f \in L^1(X, \mathcal{B}, \mu)$.

By taking finite linear combinations of characteristic functions in (12.0.10) we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int f(T^j x) g(x) d\mu = \int f d\mu \int g d\mu. \tag{12.0.11}$$

for all simple functions g . By taking an increasing sequence of simple functions and applying the Monotone Convergence Theorem as above, we have that (12.0.11) holds for all positive measurable functions g . By writing $g = g^+ - g^-$ where g^+, g^- are positive, we have that (12.0.11) holds for any real-valued integrable function g . By taking real and imaginary parts, we have that (12.0.11) holds for any $g \in L^1(X, \mathcal{B}, \mu)$.

We prove that (ii) implies (i). Suppose that for all $f, g \in L^2(X, \mathcal{B}, \mu)$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int f(T^j x) g(x) d\mu = \int f d\mu \int g d\mu.$$

Suppose that $T^{-1}B = B, B \in \mathcal{B}$. Then $\chi_B \in L^2(X, \mathcal{B}, \mu)$. Taking $f = g = \chi_B$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \int \chi_B(T^j x) \chi_B(x) d\mu \rightarrow \int \chi_B d\mu \int \chi_B d\mu = \mu(B)^2.$$

Note that $\chi_B(T^j x) \chi_B(x) = \chi_{T^{-j}B \cap B}(x) = \chi_B(x)$ as $T^{-j}B = B$. Hence

$$\frac{1}{n} \sum_{j=0}^{n-1} \int \chi_B(T^j x) \chi_B(x) d\mu = \frac{1}{n} \sum_{j=0}^{n-1} \int \chi_B d\mu = \int \chi_B d\mu = \mu(B).$$

Hence $\mu(B) = \mu(B)^2$ so that $\mu(B) = 0$ or 1 .

Solution 10.4

Choose a countable dense set of continuous functions $\{f_i\}_{i=1}^\infty \subset C(X, \mathbb{R})$. By Birkhoff's Ergodic Theorem there exists $Y_i \in \mathcal{B}$ such that $\mu(Y_i) = 1$ and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_i(T^j x) = \int f_i d\mu$$

for all $x \in Y_i$. Let $Y = \bigcap_{i=1}^\infty Y_i$. Then $Y \in \mathcal{B}$ and $\mu(Y) = 1$.

Let $f \in C(X, \mathbb{R}), \varepsilon > 0, x \in Y$. Choose i such that $\|f - f_i\|_\infty < \varepsilon$. Choose N such that if $n \geq N$ then

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} f_i(T^j x) - \int f_i d\mu \right| < \varepsilon.$$

Then

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) - \int f d\mu \right|$$

$$\begin{aligned} &\leq \left| \frac{1}{n} \sum_{j=0}^{n-1} (f(T^j x) - f_i(T^j x)) \right| + \left| \frac{1}{n} \sum_{j=0}^{n-1} f_i(T^j x) - \int f_i d\mu \right| + \left| \int f_i d\mu - \int f d\mu \right|. \\ &< 3\varepsilon. \end{aligned}$$

As $\varepsilon > 0$ is arbitrary, we have that for all $f \in C(X, \mathbb{R})$ and for all $x \in Y$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu.$$

Solution 10.5

Let $S = \{A, B, C, \dots, Z\}$ denote the finite set of letters (symbols) in the alphabet. Let $\Sigma = \{\mathbf{x} = (x_j)_{j=0}^\infty \mid x_j \in S, j = 0, 1, 2, \dots\}$ denote the space of all infinite sequences of symbols. For each $s \in S$, let $p(s) = 1/26$ denote the probability of choosing symbol s . Let \mathcal{B} denote the Borel σ -algebra on Σ and equip Σ with the Bernoulli probability measure μ defined on cylinders by

$$\mu([i_0, i_1, \dots, i_{n-1}]) = p(i_0)p(i_1) \cdots p(i_{n-1}).$$

Define $\sigma : \Sigma \rightarrow \Sigma$ by $(\sigma(\mathbf{x}))_j = x_{j+1}$.

We regard an element $\mathbf{x} \in \Sigma$ as one possible outcome of the monkey typing an infinite sequence of letters.

Let B denote the cylinder $[M, O, N, K, E, Y]$. Then $\mu(B) = 1/26^6 > 0$. By Birkhoff's Ergodic Theorem, for μ -a.e. $\mathbf{x} \in \Sigma$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_B(\sigma^j(\mathbf{x})) = \mu(B) > 0.$$

Hence, almost surely, the infinite sequence of letters \mathbf{x} will contain 'MONKEY'. Hence, with probability 1, the monkey will type the word 'MONKEY'. (Indeed, with probability one he will type 'MONKEY' infinitely often.)

By Kac's Lemma, the expected first time at which 'MONKEY' appears is $1/\mu(B) = 26^6$. If the monkey types 1 letter a second, then one would expect to wait 26^6 seconds (about 9.8 years) until 'MONKEY' first appears in a block of 6.

Solution 11.1

We first claim that for each integer $b \geq 2$, $T(x) = T_b(x) = bx \bmod 1$ is ergodic with respect to Lebesgue measure μ (we already know that Lebesgue measure is invariant by Exercise 3.6). To see this, we use Fourier series, following the argument that was used to prove that the doubling map is ergodic with respect to Lebesgue measure.

Suppose that $f \in L^2(\mathbb{R}/\mathbb{Z}, \mathcal{B}, \mu)$ is such that $f \circ T = f$ μ -a.e. Then $f \circ T^p = f$ μ -a.e. for all $p \in \mathbb{N}$. Associate to f its Fourier series $\sum_{n=-\infty}^\infty c_n e^{2\pi i n x}$. Then $f \circ T^p$ has Fourier series $\sum_{n=-\infty}^\infty c_n e^{2\pi i n b^p x}$. Comparing Fourier coefficients we see that $c_{b^p n} = c_n$. Suppose that $n \neq 0$. Then $b^p n \rightarrow \infty$ as $n \rightarrow \infty$. By the Riemann-Lebesgue Lemma (Proposition 5.3.1(ii)), $c_n = c_{b^p n} \rightarrow 0$ as $n \rightarrow \infty$. Hence $c_n = 0$ if $n \neq 0$. Hence f has Fourier series c_0 , i.e. f is constant a.e. Hence T is ergodic with respect to Lebesgue measure.

Solution 11.2

(i) Let

$$X_b = \{x \in [0, 1) \mid x \text{ is simply normal in base } b\}.$$

Then for each $b \geq 2$, X_b has Lebesgue measure $\mu(X_b) = 1$. Hence

$$X_\infty = \bigcap_{b=2}^{\infty} X_b$$

consists of all numbers that are simply normal in every base $b \geq 2$. Clearly $\mu(X_\infty) = 1$.

- (ii) Let $X(b)$ denote the set of numbers that are normal in base b , $b \geq 2$. Then $X_\infty = \bigcap_{b=2}^{\infty} X_b$ consists of all normal numbers. Clearly $\mu(X_\infty) = 1$.

Alternatively, note that $x \in [0, 1]$ is simply normal in base b^k if and only if every word of length k occurs with frequency $1/b^k$ in the base b expansion of x . Hence a number is normal in every base if and only if it is simply normal in every base.

Solution 11.3

Let $T(x) = rx \bmod 1$, $T : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$. From Exercise 11.1 we know that T is ergodic with respect to Lebesgue measure μ . Let $x \in [0, 1]$ and let $x_n = r^n x$. Then $\{x_n\}$, the fractional part of x_n , is equal to $T^n x$. Let $\ell \in \mathbb{Z} \setminus \{0\}$ and let $f_\ell(x) = e^{2\pi i \ell x}$. Then there exists $N_\ell \in \mathcal{B}$, $\mu(N_\ell) = 0$, such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_j} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f_\ell(T^j x) = \int f_\ell(x) dx = 0$$

for all $x \notin N_\ell$.

Let $N = \bigcup_{\ell \in \mathbb{Z} \setminus \{0\}} N_\ell$. As $\mu(N_\ell) = 0$ and this is a countable union, we have that $\mu(N) = 0$. Hence if $x \notin N$ we have for all $\ell \in \mathbb{Z} \setminus \{0\}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \ell x_n} = 0.$$

By Weyl's Criterion it follows that if $x \notin N$ then x_n is uniformly distributed mod 1.

(Aside: you might wonder why we had to use Weyl's Criterion and did not just use the definition of uniform distribution. Whilst it is certainly true that

$$\frac{1}{n} \text{card}\{j \in \{0, 1, \dots, n-1\} \mid \{x_j\} \in [a, b]\} = \frac{1}{n} \sum_{j=0}^{n-1} \chi_{[a,b]}(T^j x) \rightarrow \int \chi_{[a,b]} d\mu = b - a$$

for μ -a.e. $x \in X$, the set of measure zero for which this fails depends on the interval $[a, b]$. We need a set of measure zero that works for all intervals. As there are uncountably many intervals, we cannot just take the union of all the sets of measure zero as we did above. One can make an argument along these lines work, by considering intervals with rational endpoints (and so a countable collection of intervals) and then approximate an arbitrary interval.)

Solution 11.4

Let $T(x) = 10x \bmod 1$. From Exercise 11.1 we know that T is ergodic with respect to Lebesgue measure. Let $x \in [0, 1]$ have decimal expansion

$$x = \sum_{j=0}^{\infty} \frac{x_j}{10^{j+1}}$$

with $x_j \in \{0, 1, \dots, 9\}$. Let

$$f(x) = \sum_{k=0}^9 k \chi_{[k/10, (k+1)/10)}$$

so that $f(x) = k$ precisely when $x_0 = k$. Note that $f(T^j x) = k$ precisely when $x_j = k$. Then

$$\frac{1}{n}(x_0 + x_1 + \dots + x_{n-1}) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x).$$

Hence by Birkhoff's Ergodic Theorem,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n}(x_0 + x_1 + \dots + x_{n-1}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \\ &= \int f(x) dx \text{ a.e.} \\ &= \sum_{k=0}^9 \frac{k}{10} \text{ a.e.} = 4.5 \text{ a.e.} \end{aligned}$$

Solution 11.5

If $x \in [0, 1]$ then write the continued fraction expansion of x as $[x_0, x_1, \dots]$.

Define $f : [0, 1] \rightarrow \mathbb{R}$ by

$$f(x) = \sum_{k=1}^{\infty} k \chi_{(1/(k+1), 1/k]}(x)$$

if $x \in (0, 1]$ and set $f(0) = 0$. Then for $x \in (0, 1]$ we have that $f(x) = k$ precisely when $1/(k+1) < x \leq 1/k$, i.e. $f(x) = k$ when $x_0 = k$. Hence $f(T^j x) = k$ precisely when $x_j = k$.

We can write

$$\frac{1}{n}(x_0 + \dots + x_{n-1}) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x).$$

Clearly $f \geq 0$ and is measurable. However $f \notin L^1(X, \mathcal{B}, \mu)$. To see this, using Exercise 3.5(iii), it is sufficient to show that $f \notin L^1(X, \mathcal{B}, \lambda)$ where λ denotes Lebesgue measure. Note that

$$\begin{aligned} \int f d\lambda &= \sum_{k=1}^{\infty} k \lambda \left(\left(\frac{1}{k+1}, \frac{1}{k} \right] \right) \\ &= \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty. \end{aligned}$$

By the results of Exercise 10.2, it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \infty$$

for μ -a.e. $x \in X$. As Gauss' measure and Lebesgue measure have the same sets of measure zero. we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \infty$$

for Lebesgue almost every point $x \in X$.