Ergodic Theory

Karma Dajani

September 20, 2006

Contents

1 Introduction and preliminaries **5**
- 1.1 What is Ergodic Theory? . . . 5
- 1.2 Measure Preserving Transformations . . . 6
- 1.3 Basic Examples . . . 9
- 1.4 Recurrence . . . 15
- 1.5 Induced and Integral Transformations . . . 15
  - 1.5.1 Induced Transformations . . . 15
  - 1.5.2 Integral Transformations . . . 18
- 1.6 Ergodicity . . . 20
- 1.7 Other Characterizations of Ergodicity . . . 22
- 1.8 Examples of Ergodic Transformations . . . 24

2 The Ergodic Theorem **29**
- 2.1 The Ergodic Theorem and its consequences . . . 29
- 2.2 Characterization of Irreducible Markov Chains . . . 41
- 2.3 Mixing . . . 45

3 Measure Preserving Isomorphisms and Factor Maps **47**
- 3.1 Measure Preserving Isomorphisms . . . 47
- 3.2 Factor Maps . . . 51
- 3.3 Natural Extensions . . . 52

4 Entropy **55**
- 4.1 Randomness and Information . . . 55
- 4.2 Definitions and Properties . . . 56
- 4.3 Calculation of Entropy and Examples . . . 63
- 4.4 The Shannon-McMillan-Breiman Theorem . . . 66
- 4.5 Lochs' Theorem . . . 72

5 Invariant Measures for Continuous Transformations **79**
5.1 Existence 79
5.2 Unique Ergodicity 86

6 Hurewicz Ergodic Theorem **91**
6.1 Equivalent measures 91
6.2 Non-singular and conservative transformations 92
6.3 Hurewicz Ergodic Theorem 95

Chapter 1

Introduction and preliminaries

1.1 What is Ergodic Theory?

Ergodic Theory is difficult to characterize, as it stands at the junction of so many fields. It uses techniques and examples from many fields such as probability theory, statistical mechanics, number theory, vector fields on manifolds, group actions of homogeneous spaces and many more.

The word *ergodic* is an amalgamation of two Greek words: *ergon* (work) and *odos* (path). The word was introduced by Boltzmann (in statistical mechanics) regarding his hypothesis: *for large systems of interacting particles in equilibrium, the time average along a single trajectory equals the space average.* The hypothesis as it was stated was false, and the investigation for the conditions under which these two quantities are equal lead to the birth of ergodic theory as is known nowadays.

A modern description of what ergodic theory is would be: it is the study of the long term average behavior of systems evolving in time. The collection of all states of the system form a space X, and the evolution is represented by either

- a transformation $T : X \to X$, where Tx is the state of the system at time $t = 1$, when the system (i.e., at time $t = 0$) was initially in state x. (This is analogous to the setup of discrete time stochastic processes).

- if the evolution is continuous or if the configurations have spacial structure, then we describe the evolution by looking at a group of transformations G (like $\mathbb{Z}^2$, $\mathbb{R}$, $\mathbb{R}^2$) acting on X, i.e., every $g \in G$ is identified with a transformation $T_g : X \to X$, and $T_{gg'} = T_g \circ T_{g'}$.

The space X usually has a special structure, and we want T to preserve the basic structure on X. For example
–if X is a measure space, then T must be measurable.
–if X is a topological space, then T must be continuous.
–if X has a differentiable structure, then T is a diffeomorphism.
In this course our space is a probability space $(X, \mathcal{B}, \mu)$, and our time is discrete. So the evolution is described by a measurable map $T : X \to X$, so that $T^{-1}A \in \mathcal{B}$ for all $A \in \mathcal{B}$. For each $x \in X$, the orbit of x is the sequence

$$x, Tx, T^2x, \ldots .$$

If T is invertible, then one speaks of the two sided orbit

$$\ldots, T^{-1}x, x, Tx, \ldots .$$

We want also that the evolution is in steady state i.e. stationary. In the language of ergodic theory, we want T to be *measure preserving.*

1.2 Measure Preserving Transformations

Definition 1.2.1 *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ measurable. The map T is said to be measure preserving with respect to μ if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$.*

This definition implies that for any measurable function $f : X \to \mathbb{R}$, the process

$$f, f \circ T, f \circ T^2, \ldots$$

is stationary. This means that for all Borel sets $B_1, \ldots, B_n$, and all integers $r_1 < r_2 < \ldots < r_n$, one has for any $k \geq 1$,

$$\mu\left(\{x : f(T^{r_1}x) \in B_1, \ldots f(T^{r_n}x) \in B_n\}\right) =$$
$$\mu\left(\{x : f(T^{r_1+k}x) \in B_1, \ldots f(T^{r_n+k}x) \in B_n\}\right).$$

In case T is invertible, then T is measure preserving if and only if $\mu(TA) = \mu(A)$ for all $A \in \mathcal{B}$. We can generalize the definition of measure preserving to the following case. Let $T : (X_1, \mathcal{B}_1, \mu_1) \to (X_2, \mathcal{B}_2, \mu_2)$ be measurable, then T is measure preserving if $\mu_1(T^{-1}A) = \mu_2(A)$ for all $A \in \mathcal{B}_2$. The following

gives a useful tool for verifying that a transformation is measure preserving. For this we need the notions of algebra and semi-algebra.
Recall that a collection $\mathcal{S}$ of subsets of X is said to be a *semi-algebra* if (i) $\emptyset \in \mathcal{S}$, (ii) $A \cap B \in \mathcal{S}$ whenever $A, B \in \mathcal{S}$, and (iii) if $A \in \mathcal{S}$, then $X \setminus A = \cup_{i=1}^{n} E_i$ is a disjoint union of elements of $\mathcal{S}$. For example if $X = [0, 1)$, and $\mathcal{S}$ is the collection of all subintervals, then $\mathcal{S}$ is a semi-algebra. Or if $X = \{0, 1\}^{\mathbb{Z}}$, then the collection of all cylinder sets $\{x : x_i = a_i, \ldots, x_j = a_j\}$ is a semi-algebra.
An *algebra* $\mathcal{A}$ is a collection of subsets of X satisfying:(i) $\emptyset \in \mathcal{A}$, (ii) if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$, and finally (iii) if $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$. Clearly an algebra is a semi-algebra. Furthermore, given a semi-algebra $\mathcal{S}$ one can form an algebra by taking all finite disjoint unions of elements of $\mathcal{S}$. We denote this algebra by $\mathcal{A}(\mathcal{S})$, and we call it the *algebra generated* by $\mathcal{S}$. It is in fact the smallest algebra containing $\mathcal{S}$. Likewise, given a semi-algebra $\mathcal{S}$ (or an algebra $\mathcal{A}$), the σ-algebra generated by $\mathcal{S}$ ($\mathcal{A}$) is denoted by $\mathcal{B}(\mathcal{S})$ ($\mathcal{B}(\mathcal{A})$), and is the smallest σ-algebra containing $\mathcal{S}$ (or $\mathcal{A}$).
A *monotone class* $\mathcal{C}$ is a collection of subsets of X with the following two properties

- if $E_1 \subseteq E_2 \subseteq \ldots$ are elements of $\mathcal{C}$, then $\cup_{i=1}^{\infty} E_i \in \mathcal{C}$,
- if $F_1 \supseteq F_2 \supseteq \ldots$ are elements of $\mathcal{C}$, then $\cap_{i=1}^{\infty} F_i \in \mathcal{C}$.

The *monotone class generated* by a collection $\mathcal{S}$ of subsets of X is the smallest monotone class containing $\mathcal{S}$.

Theorem 1.2.1 *Let $\mathcal{A}$ be an algebra of X, then the σ-algebra $\mathcal{B}(\mathcal{A})$ generated by $\mathcal{A}$ equals the monotone class generated by $\mathcal{A}$.*

Using the above Theorem, one can get an easier criterion for checking that a transformation is measure preserving.

Theorem 1.2.2 *Let $(X_i, \mathcal{B}_i, \mu_i)$ be probability spaces, $i = 1, 2$, and $T : X_1 \to X_2$ a transformation. Suppose $\mathcal{S}_2$ is a generating semi-algebra of $\mathcal{B}_2$. Then, T is measurable and measure preserving if and only if for each $A \in \mathcal{S}_2$, we have $T^{-1}A \in \mathcal{B}_1$ and $\mu_1(T^{-1}A) = \mu_2(A)$.*

Proof. Let

$$\mathcal{C} = \{B \in \mathcal{B}_2 : T^{-1}B \in \mathcal{B}_1, \text{ and } \mu_1(T^{-1}B) = \mu_2(B)\}.$$

Then, $\mathcal{S}_2 \subseteq \mathcal{C} \subseteq \mathcal{B}_2$, and hence $\mathcal{A}(\mathcal{S}_2) \subset \mathcal{C}$. We show that $\mathcal{C}$ is a monotone class. Let $E_1 \subseteq E_2 \subseteq \ldots$ be elements of $\mathcal{C}$, and let $E = \cup_{i=1}^{\infty} E_i$. Then, $T^{-1}E = \cup_{i=1}^{\infty} T^{-1}E_i \in \mathcal{B}_1$.

$$\begin{aligned}
\mu_1(T^{-1}E) = \mu_1(\cup_{n=1}^{\infty} T^{-1}E_n) &= \lim_{n\to\infty} \mu_1(T^{-1}E_n) \\
&= \lim_{n\to\infty} \mu_2(E_n) \\
&= \mu_2(\cup_{n=1}^{\infty} E_n) \\
&= \mu_2(E).
\end{aligned}$$

Thus, $E \in \mathcal{C}$. A similar proof shows that if $F_1 \supseteq F_2 \supseteq \ldots$ are elements of $\mathcal{C}$, then $\cap_{i=1}^{\infty} F_i \in \mathcal{C}$. Hence, $\mathcal{C}$ is a monotone class containing the algebra $\mathcal{A}(\mathcal{S}_2)$. By the monotone class theorem, $\mathcal{B}_2$ is the smallest monotone class containing $\mathcal{A}(\mathcal{S}_2)$, hence $\mathcal{B}_2 \subseteq \mathcal{C}$. This shows that $\mathcal{B}_2 = \mathcal{C}$, therefore T is measurable and measure preserving. □

For example if
– $X = [0, 1)$ with the Borel σ-algebra $\mathcal{B}$, and μ a probability measure on $\mathcal{B}$. Then a transformation $T : X \to X$ is measurable and measure preserving if and only if $T^{-1}[a, b) \in \mathcal{B}$ and $\mu\left(T^{-1}[a, b)\right) = \mu\left([a, b)\right)$ for any interval $[a, b)$.
– $X = \{0, 1\}^{\mathbb{N}}$ with product σ-algebra and product measure μ. A transformation $T : X \to X$ is measurable and measure preserving if and only if

$$T^{-1}\left(\{x : x_0 = a_0, \ldots, x_n = a_n\}\right) \in \mathcal{B},$$

and

$$\mu\left(T^{-1}\{x : x_0 = a_0, \ldots, x_n = a_n\}\right) = \mu\left(\{x : x_0 = a_0, \ldots, x_n = a_n\}\right)$$

for any cylinder set.

Exercise 1.2.1 Recall that if A and B are measurable sets, then

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

Show that for any measurable sets A, B, C one has

$$\mu(A\Delta B) \leq \mu(A\Delta C) + \mu(C\Delta B).$$

Another useful lemma is the following (see also ([KT]).

Lemma 1.2.1 *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $\mathcal{A}$ an algebra generating $\mathcal{B}$. Then, for any $A \in \mathcal{B}$ and any $\epsilon > 0$, there exists $C \in \mathcal{A}$ such that $\mu(A\Delta C) < \epsilon$.*

Proof. Let

$$\mathcal{D} = \{A \in \mathcal{B} : \text{ for any } \epsilon > 0, \text{ there exists } C \in \mathcal{A} \text{ such that } \mu(A\Delta C) < \epsilon\}.$$

Clearly, $\mathcal{A} \subseteq \mathcal{D} \subseteq \mathcal{B}$. By the Monotone Class Theorem (Theorem (1.2.1)), we need to show that $\mathcal{D}$ is a monotone class. To this end, let $A_1 \subseteq A_2 \subseteq \cdots$ be a sequence in $\mathcal{D}$, and let $A = \bigcup_{n=1}^{\infty} A_n$, notice that $\mu(A) = \lim_{n\to\infty} \mu(A_n)$. Let $\epsilon > 0$, there exists an N such that $\mu(A\Delta A_N) = |\mu(A) - \mu(A_N)| < \epsilon/2$. Since $A_N \in \mathcal{D}$, then there exists $C \in \mathcal{A}$ such that $\mu(A_N\Delta C) < \epsilon/2$. Then,

$$\mu(A\Delta C) \leq \mu(A\Delta A_N) + \mu(A_N\Delta C) < \epsilon.$$

Hence, $A \in \mathcal{D}$. Similarly, one can show that $\mathcal{D}$ is closed under decreasing intersections so that $\mathcal{D}$ is a monotone class containg $\mathcal{A}$, hence by the Monotone class Theorem $\mathcal{B} \subseteq \mathcal{D}$. Therefore, $\mathcal{B} = \mathcal{D}$, and the theorem is proved.

$\square$

1.3 Basic Examples

(a) *Translations* – Let $X = [0, 1)$ with the Lebesgue σ-algebra $\mathcal{B}$, and Lebesgue measure λ. Let $0 < \theta < 1$, define $T : X \to X$ by

$$Tx = x + \theta \bmod 1 = x + \theta - \lfloor x + \theta \rfloor.$$

Then, by considering intervals it is easy to see that T is measurable and measure preserving.

(b) *Multiplication by 2 modulo 1* – Let $(X, \mathcal{B}, \lambda)$ be as in example (a), and let $T : X \to X$ be given by

$$Tx = 2x \bmod 1 = \begin{cases} 2x & 0 \leq x < 1/2 \\ 2x - 1 & 1/2 \leq x < 1. \end{cases}$$

For any interval $[a, b)$,

$$T^{-1}[a, b) = [\frac{a}{2}, \frac{b}{2}) \cup [\frac{a+1}{2}, \frac{b+1}{2}),$$

and

$$\lambda\left(T^{-1}[a,b)\right) = b - a = \lambda\left([a,b)\right).$$

Although this map is very simple, it has in fact many facets. For example, iterations of this map yield the binary expansion of points in $[0,1)$ i.e., using T one can associate with each point in $[0,1)$ an infinite sequence of 0's and 1's. To do so, we define the function a_1 by

$$a_1(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1/2 \\ 1 & \text{if } 1/2 \leq x < 1, \end{cases}$$

then $Tx = 2x - a_1(x)$. Now, for $n \geq 1$ set $a_n(x) = a_1(T^{n-1}x)$. Fix $x \in X$, for simplicity, we write a_n instead of $a_n(x)$, then $Tx = 2x - a_1$. Rewriting we get $x = \frac{a_1}{2} + \frac{Tx}{2}$. Similarly, $Tx = \frac{a_2}{2} + \frac{T^2x}{2}$. Continuing in this manner, we see that for each $n \geq 1$,

$$x = \frac{a_1}{2} + \frac{a_2}{2^2} + \cdots + \frac{a_n}{2^n} + \frac{T^n x}{2^n}.$$

Since $0 < T^n x < 1$, we get

$$x - \sum_{i=1}^{n} \frac{a_i}{2^i} = \frac{T^n x}{2^n} \to 0 \text{ as } n \to \infty.$$

Thus, $x = \sum_{i=1}^{\infty} \frac{a_i}{2^i}$. We shall later see that the sequence of digits $a_1, a_2, \ldots$ forms an i.i.d. sequence of Bernoulli random variables.

(c) *Baker's Transformation* – This example is the two-dimensional version of example (b). The underlying probability space is $[0,1)^2$ with product Lebesgue σ-algebra $\mathcal{B} \times \mathcal{B}$ and product Lebesgue measure $\lambda \times \lambda$. Define $T : [0,1)^2 \to [0,1)^2$ by

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

Exercise 1.3.1 Verify that T is invertible, measurable and measure preserving.

(d) *β-transformations* – Let $X = [0,1)$ with the Lebesgue σ-algebra $\mathcal{B}$. Let $\beta = \frac{1+\sqrt{5}}{2}$, *the golden mean.* Notice that $\beta^2 = \beta + 1$. Define a transformation

$T : X \to X$ by

$$Tx = \beta x \bmod 1 = \begin{cases} \beta x & 0 \le x < 1/\beta \\ \beta x - 1 & 1/\beta \le x < 1. \end{cases}$$

Then, T is **not** measure preserving with respect to Lebesgue measure (give a counterexample), but is measure preserving with respect to the measure μ given by

$$\mu(B) = \int_B g(x)\,\mathrm{d}x,$$

where

$$g(x) = \begin{cases} \frac{5+3\sqrt{5}}{10} & 0 \le x < 1/\beta \\ \frac{5+\sqrt{5}}{10} & 1/\beta \le x < 1. \end{cases}$$

Exercise 1.3.2 Verify that T is measure preserving with respect to μ, and show that (similar to example (b)) iterations of this map generate expansions for points $x \in [0, 1)$ (known as *β-expansions)* of the form

$$x = \sum_{i=1}^{\infty} \frac{b_i}{\beta^i},$$

where $b_i \in \{0, 1\}$ and $b_i b_{i+1} = 0$ for all $i \ge 1$.

(e) *Bernoulli Shifts* – Let $X = \{0, 1, \ldots k-1\}^{\mathbb{Z}}$ (or $X = \{0, 1, \ldots k-1\}^{\mathbb{N}}$), $\mathcal{F}$ the σ-algebra generated by the cylinders. Let $p = (p_0, p_1, \ldots, p_{k-1})$ be a positive probability vector, define a measure μ on $\mathcal{F}$ by specifying it on the cylinder sets as follows

$$\mu\left(\{x : x_{-n} = a_{-n}, \ldots, x_n = a_n\}\right) = p_{a_{-n}} \ldots p_{a_n}.$$

Let $T : X \to X$ be defined by $Tx = y$, where $y_n = x_{n+1}$. The map T, called the *left shift*, is measurable and measure preserving, since

$$T^{-1}\{x : x_{-n} = a_{-n}, \ldots x_n = a_n\} = \{x : x_{-n+1} = a_{-n}, \ldots, x_{n+1} = a_n\},$$

and

$$\mu\left(\{x : x_{-n+1} = a_{-n}, \ldots, x_{n+1} = a_n\}\right) = p_{a_{-n}} \ldots p_{a_n}.$$

Notice that in case $X = \{0, 1, \dots k-1\}^{\mathbb{N}}$, then one should consider cylinder sets of the form $\{x : x_0 = a_0, \dots x_n = a_n\}$. In this case

$$T^{-1}\{x : x_0 = a_0, \dots, x_n = a_n\} = \cup_{j=0}^{k-1}\{x : x_0 = j, x_1 = a_0, \dots, x_{n+1} = a_n\},$$

and it is easy to see that T is measurable and measure preserving.

(f) *Markov Shifts* – Let $(X, \mathcal{F}, T)$ be as in example (e). We define a measure ν on $\mathcal{F}$ as follows. Let $P = (p_{ij})$ be a stochastic $k \times k$ matrix, and $q = (q_0, q_1, \dots, q_{k-1})$ a positive probability vector such that $qP = q$. Define ν on cylinders by

$$\nu\left(\{x : x_{-n} = a_{-n}, \dots x_n = a_n\}\right) = q_{a_{-n}} p_{a_{-n}a_{-n+1}} \cdots p_{a_{n-1}a_n}.$$

Just as in example (e), one sees that T is measurable and measure preserving.

(g) *Stationary Stochastic Processes*– Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and

$$\dots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \dots$$

a stationary stochastic process on Ω with values in $\mathbb{R}$. Hence, for each $k \in \mathbb{Z}$

$$\mathbb{P}\left(Y_{n_1} \in B_1, \dots, Y_{n_r} \in B_r\right) = \mathbb{P}\left(Y_{n_1+k} \in B_1, \dots, Y_{n_r+k} \in B_r\right)$$

for any $n_1 < n_2 < \cdots < n_r$ and any Lebesgue sets $B_1, \dots, B_r$. We want to see this process as coming from a measure preserving transformation.
Let $X = \mathbb{R}^{\mathbb{Z}} = \{x = (\dots, x_1, x_0, x_1, \dots) : x_i \in \mathbb{R}\}$ with the product σ-algebra (i.e. generated by the cylinder sets). Let $T : X \to X$ be the left shift i.e. $Tx = z$ where $z_n = x_{n+1}$. Define $\phi : \Omega \to X$ by

$$\phi(\omega) = (\dots, Y_{-2}(\omega), Y_{-1}(\omega), Y_0(\omega), Y_1(\omega), Y_2(\omega), \dots).$$

Then, ϕ is measurable since if $B_1, \dots, B_r$ are Lebesgue sets in $\mathbb{R}$, then

$$\phi^{-1}\left(\{x \in X : x_{n_1} \in B_1, \dots x_{n_r} \in B_r\}\right) = Y_{n_1}^{-1}(B_1) \cap \dots \cap Y_{n_r}^{-1}(B_r) \in \mathcal{F}.$$

Define a measure μ on X by

$$\mu(E) = \mathbb{P}\left(\phi^{-1}(E)\right).$$

On cylinder sets μ has the form,

$$\mu\left(\{x \in X : x_{n_1} \in B_1, \dots x_{n_r} \in B_r\}\right) = \mathbb{P}\left(Y_{n_1} \in B_1, \dots, Y_{n_r} \in B_r\right).$$

Since

$$T^{-1}\left(\{x : x_{n_1} \in B_1, \ldots x_{n_r} \in B_r\}\right) = \{x : x_{n_1+1} \in B_1, \ldots x_{n_r+1} \in B_r\},$$

stationarity of the process Y_n implies that T is measure preserving. Furthermore, if we let $\pi_i : X \to \mathbb{R}$ be the natural projection onto the i^{th} coordinate, then $Y_i(\omega) = \pi_i(\phi(\omega)) = \pi_0 \circ T^i(\phi(\omega))$.

(h) *Random Shifts* – Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ an invertible measure preserving transformation. Then, T^{-1} is measurable and measure preserving with respect to μ. Suppose now that at each moment instead of moving forward by T $(x \to Tx)$, we first flip a fair coin to decide whether we will use T or T^{-1}. We can describe this random system by means of a measure preserving transformation in the following way.
Let $\Omega = \{-1, 1\}^{\mathbb{Z}}$ with product σ-algebra $\mathcal{F}$ (i.e. the σ-algebra generated by the cylinder sets), and the uniform product measure $\mathbb{P}$ (see example (e)), and let $\sigma : \Omega \to \Omega$ be the left shift. As in example (e), the map σ is measure preserving. Now, let $Y = \Omega \times X$ with the product σ-algebra, and product measure $\mathbb{P} \times \mu$. Define $S : Y \to Y$ by

$$S(\omega, x) = (\sigma\omega, T^{\omega_0}x).$$

Then S is invertible (why?), and measure preserving with respect to $\mathbb{P} \times \mu$. To see the latter, for any set $C \in \mathcal{F}$, and any $A \in \mathcal{B}$, we have

$$\begin{aligned}
(\mathbb{P} \times \mu)\left(S^{-1}(C \times A)\right) &= (\mathbb{P} \times \mu)\left(\{(\omega, x) : S(\omega, x) \in (C \times A)\right) \\
&= (\mathbb{P} \times \mu)\left(\{(\omega, x) : \omega_0 = 1, \sigma\omega \in C, Tx \in A\right) \\
&+ (\mathbb{P} \times \mu)\left(\{(\omega, x) : \omega_0 = -1, \sigma\omega \in C, T^{-1}x \in A\right) \\
&= (\mathbb{P} \times \mu)\left(\{\omega_0 = 1\} \cap \sigma^{-1}C \times T^{-1}A\right) \\
&+ (\mathbb{P} \times \mu)\left(\{\omega_0 = -1\} \cap \sigma^{-1}C \times TA\right) \\
&= \mathbb{P}\left(\{\omega_0 = 1\} \cap \sigma^{-1}C\right)\mu\left(T^{-1}A\right) \\
&+ \mathbb{P}\left(\{\omega_0 = -1\} \cap \sigma^{-1}C\right)\mu\left(TA\right) \\
&= \mathbb{P}\left(\{\omega_0 = 1\} \cap \sigma^{-1}C\right)\mu(A) \\
&+ \mathbb{P}\left(\{\omega_0 = -1\} \cap \sigma^{-1}C\right)\mu(A) \\
&= \mathbb{P}(\sigma^{-1}C)\mu(A) = \mathbb{P}(C)\mu(A) = (\mathbb{P} \times \mu)(C \times A).
\end{aligned}$$

(h) *continued fractions* – Consider $([0, 1), \mathcal{B})$, where $\mathcal{B}$ is the Lebesgue σ-algebra. Define a transformation $T : [0, 1) \to [0, 1)$ by $T0 = 0$ and for $x \neq 0$

$$Tx = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor.$$

Exercise 1.3.3 *Show that T is* **not** *measure preserving with respect to Lebesgue measure, but is measure preserving with respect to the so called Gauss probability measure μ given by*

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

An interesting feature of this map is that its iterations generate the continued fraction expansion for points in $(0, 1)$. For if we define

$$a_1 = a_1(x) = \begin{cases} 1 & \text{if } x \in (\frac{1}{2}, 1) \\ n & \text{if } x \in (\frac{1}{n+1}, \frac{1}{n}], n \geq 2, \end{cases}$$

then, $Tx = \frac{1}{x} - a_1$ and hence $x = \dfrac{1}{a_1 + Tx}$. For $n \geq 1$, let $a_n = a_n(x) = a_1(T^{n-1}x)$. Then, after n iterations we see that

$$x = \frac{1}{a_1 + Tx} = \ldots = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + T^n x}}}.$$

In fact, if $\dfrac{p_n}{q_n} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}}$, then one can show that $\{q_n\}$ are monotonically increasing, and

$$|x - \frac{p_n}{q_n}| < \frac{1}{q_n^2} \to 0 \text{ as } n \to \infty.$$

The last statement implies that

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}.$$

1.4 Recurrence

Let T be a measure preserving transformation on a probability space $(X, \mathcal{F}, \mu)$, and let $B \in \mathcal{F}$. A point $x \in B$ is said to be *B-recurrent* if there exists $k \geq 1$ such that $T^k x \in B$.

Theorem 1.4.1 (Poincaré Recurrence Theorem) *If $\mu(B) > 0$, then a.e. $x \in B$ is B-recurrent.*

Proof Let F be the subset of B consisting of all elements that are not B-recurrent. Then,

$$F = \{x \in B : T^k x \notin B \text{ for all } k \geq 1\}.$$

We want to show that $\mu(F) = 0$. First notice that $F \cap T^{-k}F = \emptyset$ for all $k \geq 1$, hence $T^{-l}F \cap T^{-m}F = \emptyset$ for all $l \neq m$. Thus, the sets $F, T^{-1}F, \ldots$ are pairwise disjoint, and $\mu(T^{-n}F) = \mu(F)$ for all $n \geq 1$ (T is measure preserving). If $\mu(F) > 0$, then

$$1 = \mu(X) \geq \mu\left(\cup_{k\geq 0} T^{-k}F\right) = \sum_{k\geq 0} \mu(F) = \infty,$$

a contradiction. □

The proof of the above theorem implies that almost every $x \in B$ returns to B infinitely often. In other words, there exist infinitely many integers $n_1 < n_2 < \ldots$ such that $T^{n_i}x \in B$. To see this, let

$$D = \{x \in B : T^k x \in B \text{ for finitely many } k \geq 1\}.$$

Then,

$$D = \{x \in B : T^k x \in F \text{ for some } k \geq 0\} \subseteq \cup_{k=0}^{\infty} T^{-k}F.$$

Thus, $\mu(D) = 0$ since $\mu(F) = 0$ and T is measure preserving.

1.5 Induced and Integral Transformations

1.5.1 Induced Transformations

Let T be a measure preserving transformation on the probability space $(X, \mathcal{F}, \mu)$. Let $A \subset X$ with $\mu(A) > 0$. By Poincaré's Recurrence Theorem almost every $x \in A$ returns to A infinitely often under the action of T.

For $x \in A$, let $n(x) := \inf\{n \geq 1 : T^n x \in A\}$. We call $n(x)$ the *first return time* of x to A.

Exercise 1.5.1 Show that n is measurable with respect to the σ-algebra $\mathcal{F} \cap A$ on A.

By Poincaré Theorem, $n(x)$ is finite a.e. on A. In the sequel we remove from A the set of measure zero on which $n(x) = \infty$, and we denote the new set again by A. Consider the σ-algebra $\mathcal{F} \cap A$ on A, which is the restriction of $\mathcal{F}$ to A. Furthermore, let μ_A be the probability measure on A, defined by

$$\mu_A(B) = \frac{\mu(B)}{\mu(A)}, \quad \text{for } B \in \mathcal{F} \cap A,$$

so that $(A, \mathcal{F} \cap A, \mu_A)$ is a probability space. Finally, define the induced map $T_A : A \to A$ by

$$T_A x = T^{n(x)} x, \quad \text{for } x \in A.$$

From the above we see that T_A is defined on A. What kind of a transformation is T_A?

Exercise 1.5.2 Show that T_A is measurable with respect to the σ-algebra $\mathcal{F} \cap A$.

Proposition 1.5.1 *T_A is measure preserving with respect to μ_A.*

Proof For $k \geq 1$, let

$$A_k = \{x \in A : n(x) = k\}$$

$$B_k = \{x \in X \setminus A : Tx, \ldots, T^{k-1}x \notin A, T^k x \in A\}.$$

Notice that $A = \bigcup_{k=1}^{\infty} A_k$, and

$$T^{-1}A = A_1 \cup B_1 \quad \text{and} \quad T^{-1}B_n = A_{n+1} \cup B_{n+1}. \tag{1.1}$$

Let $C \in \mathcal{F} \cap A$, since T is measure preserving it follows that $\mu(C) = \mu(T^{-1}C)$. To show that $\mu_A(C) = \mu_A(T_A^{-1}C)$, we show that

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

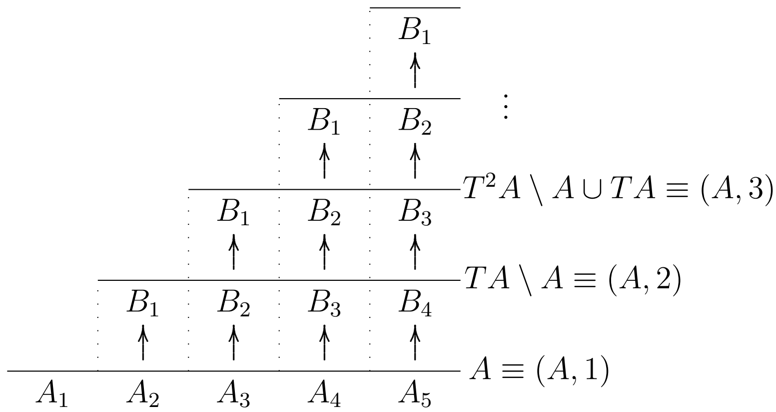

Figure 1.1: A tower.

Now,

$$T_A^{-1}(C) \;=\; \bigcup_{k=1}^{\infty} A_k \cap T_A^{-1}C \;=\; \bigcup_{k=1}^{\infty} A_k \cap T^{-k}C,$$

hence

$$\mu\left(T_A^{-1}(C)\right) \;=\; \sum_{k=1}^{\infty} \mu\left(A_k \cap T^{-k}C\right).$$

On the other hand, using repeatedly (1.1) , one gets for any $n \geq 1$,

$$\begin{aligned}
\mu\left(T^{-1}(C)\right) &= \mu(A_1 \cap T^{-1}C) + \mu(B_1 \cap T^{-1}C) \\
&= \mu(A_1 \cap T^{-1}C) + \mu(T^{-1}(B_1 \cap T^{-1}C)) \\
&= \mu(A_1 \cap T^{-1}C) + \mu(A_2 \cap T^{-2}C) + \mu(B_2 \cap T^{-2}C) \\
&\;\;\vdots \\
&= \sum_{k=1}^{n} \mu(A_k \cap T^{-k}C) + \mu(B_n \cap T^{-n}C).
\end{aligned}$$

Since

$$1 \;\geq\; \mu\left(\bigcup_{n=1}^{\infty} B_n \cap T^{-n}C\right) \;=\; \sum_{n=1}^{\infty} \mu(B_n \cap T^{-n}C),$$

it follows that

$$\lim_{n\to\infty} \mu(B_n \cap T^{-n}C) \;=\; 0.$$

Thus,

$$\mu(C) = \mu(T^{-1}C) = \sum_{k=1}^{\infty} \mu\left(A_k \cap T^{-k}C\right) = \mu(T_A^{-1}C).$$

This shows that $\mu_A(C) = \mu_A(T_A^{-1}C)$, which implies that T_A is measure preserving with respect to μ_A. □

Exercise 1.5.3 Assume T is invertible. Without using Proposition 1.5.1 show that for all $C \in \mathcal{F} \cap A$,

$$\mu_A(C) = \mu_A(T_A C).$$

Exercise 1.5.4 Let $G = \dfrac{1+\sqrt{5}}{2}$, so that $G^2 = G + 1$. Consider the set

$$X = [0, \frac{1}{G}) \times [0,1) \bigcup [\frac{1}{G}, 1) \times [0, \frac{1}{G}),$$

endowed with the product Borel σ-algebra, and the normalized Lebesgue measure $\lambda \times \lambda$. Define the transformation

$$\mathcal{T}(x,y) = \begin{cases} (Gx, \dfrac{y}{G}), & (x,y) \in [0,\frac{1}{G}) \times [0,1] \\ \\ (Gx-1, \dfrac{1+y}{G}), & (x,y) \in [\frac{1}{G}, 1) \times [0, \frac{1}{G}). \end{cases}$$

(a) Show that $\mathcal{T}$ is measure preserving with respect to $\lambda \times \lambda$.

(b) Determine explicitely the induced transformation of $\mathcal{T}$ on the set $[0,1) \times [0, \frac{1}{G})$.

1.5.2 Integral Transformations

Let S be a measure preserving transformation on a probability space $(A, \mathcal{F}, \nu)$, and let $f \in L^1(A, \nu)$ be positive and integer valued. We now construct a measure preserving transformation T on a probability space $(X, \mathcal{C}, \mu)$, such that the original transformation S can be seen as the induced transformation on X with return time f.

(1) $X = \{(y,i) : y \in A \text{ and } 1 \leq i \leq f(y),\, i \in \mathbb{N}\}$,

(2) $\mathcal{C}$ is generated by sets of the form

$$(B,i) = \{(y,i) : y \in B \text{ and } f(y) \geq i\} ,$$

where $B \subset A$, $B \in \mathcal{F}$ and $i \in \mathbb{N}$.

(3) $\mu(B,i) = \dfrac{\nu(B)}{\displaystyle\int_A f(y)\,\mathrm{d}\nu(y)}$ and then extend μ to all of X.

(4) Define $T : X \to X$ as follows:

$$T(y,i) = \begin{cases} (y, i+1), & \text{if } i+1 \leq f(y), \\ (Sy, 1), & \text{if } i+1 > f(y). \end{cases}$$

Now $(X, \mathcal{C}, \mu, T)$ is called an *integral system* of $(A, \mathcal{F}, \nu, S)$ under f. We now show that T is μ-measure preserving. In fact, it suffices to check this on the generators.

Let $B \subset A$ be $\mathcal{F}$-measurable, and let $i \geq 1$. We have to discern the following two cases:

(1) If $i > 1$, then $T^{-1}(B,i) = (B, i-1)$ and clearly

$$\mu(T^{-1}(B,i)) = \mu(B, i-1) = \mu(B,i) = \frac{\nu(B)}{\int_A f(y)\,\mathrm{d}\nu(y)} .$$

(2) If $i = 1$, we write $A_n = \{y \in A : f(y) = n\}$, and we have

$$T^{-1}(B,1) = \bigcup_{n=1}^{\infty} (A_n \cap S^{-1}B,\, n) \quad \text{(disjoint union)}.$$

Since $\bigcup_{n=1}^{\infty} A_n = A$ we therefore find that

$$\begin{aligned} \mu(T^{-1}(B,1)) &= \sum_{n=1}^{\infty} \frac{\nu(A_n \cap S^{-1}B)}{\displaystyle\int_A f(y)\,\mathrm{d}\nu(y)} = \frac{\nu(S^{-1}B)}{\displaystyle\int_A f(y)\,\mathrm{d}\nu(y)} \\ &= \frac{\nu(B)}{\displaystyle\int_A f(y)\,\mathrm{d}\nu(y)} = \mu(B,1) . \end{aligned}$$

This shows that T is measure preserving. Moreover, if we consider the induced transformation of T on the set $(A,1)$, then the first return time $n(x,1) = \inf\{k \geq 1 : T^k(x,1) \in (A,1)\}$ is given by $n(x,1) = f(x)$, and $T_{(A,1)}(x,1) = (Sx,1)$.

1.6 Ergodicity

Definition 1.6.1 *Let T be a measure preserving transformation on a probability space $(X,\mathcal{F},\mu)$. The map T is said to be ergodic if for every measurable set A satisfying $T^{-1}A = A$, we have $\mu(A) = 0$ or 1.*

Theorem 1.6.1 *Let $(X,\mathcal{F},\mu)$ be a probability space and $T : X \to X$ measure preserving. The following are equivalent:*

(i) T is ergodic.

(ii) If $B \in \mathcal{F}$ with $\mu(T^{-1}B\Delta B) = 0$, then $\mu(B) = 0$ or 1.

(iii) If $A \in \mathcal{F}$ with $\mu(A) > 0$, then $\mu\left(\cup_{n=1}^{\infty} T^{-n}A\right) = 1$.

(iv) If $A, B \in \mathcal{F}$ with $\mu(A) > 0$ and $\mu(B) > 0$, then there exists $n > 0$ such that $\mu(T^{-n}A \cap B) > 0$.

Remark 1.6.1

1. In case T is invertible, then in the above characterization one can replace T^{-n} by T^n.

2. Note that if $\mu(B\triangle T^{-1}B) = 0$, then $\mu(B \setminus T^{-1}B) = \mu(T^{-1}B \setminus B) = 0$. Since

$$B = \left(B \setminus T^{-1}B\right) \cup \left(B \cap T^{-1}B\right),$$

and

$$T^{-1}B = \left(T^{-1}B \setminus B\right) \cup \left(B \cap T^{-1}B\right),$$

we see that after removing a set of measure 0 from B and a set of measure 0 from $T^{-1}B$, the remaining parts are equal. In this case we say that B equals $T^{-1}B$ modulo sets of measure 0.
3. In words, (iii) says that if A is a set of positive measure, almost every $x \in X$ eventually (in fact infinitely often) will visit A.
4. (iv) says that elements of B will eventually enter A.

Proof of Theorem 1.6.1

(i)⇒(ii) Let $B \in \mathcal{F}$ be such that $\mu(B\Delta T^{-1}B) = 0$. We shall define a measurable set C with $C = T^{-1}C$ and $\mu(C\Delta B) = 0$. Let

$$C = \{x \in X : T^n x \in B \text{ i.o. }\} = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} T^{-k}B.$$

Then, $T^{-1}C = C$, hence by (i) $\mu(C) = 0$ or 1. Furthermore,

$$\begin{aligned}
\mu(C\Delta B) &= \mu\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} T^{-k}B \cap B^c\right) + \mu\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} T^{-k}B^c \cap B\right)\\
&\leq \mu\left(\bigcup_{k=1}^{\infty} T^{-k}B \cap B^c\right) + \mu\left(\bigcup_{k=1}^{\infty} T^{-k}B^c \cap B\right)\\
&\leq \sum_{k=1}^{\infty}\mu\left(T^{-k}B\Delta B\right).
\end{aligned}$$

Using induction (and the fact that $\mu(E\Delta F) \leq \mu(E\Delta G) + \mu(G\Delta F)$), one can show that for each $k \geq 1$ one has $\mu\left(T^{-k}B\Delta B\right) = 0$. Hence, $\mu(C\Delta B) = 0$ which implies that $\mu(C) = \mu(B)$. Therefore, $\mu(B) = 0$ or 1.

(ii)⇒(iii) Let $\mu(A) > 0$ and let $B = \bigcup_{n=1}^{\infty} T^{-n}A$. Then $T^{-1}B \subset B$. Since T is measure preserving, then $\mu(B) > 0$ and

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

Thus, by (ii) $\mu(B) = 1$.

(iii)⇒(iv) Suppose $\mu(A)\mu(B) > 0$. By (iii)

$$\mu(B) = \mu\left(B \cap \bigcup_{n=1}^{\infty} T^{-n}A\right) = \mu\left(\bigcup_{n=1}^{\infty}(B \cap T^{-n}A)\right) > 0.$$

Hence, there exists $k \geq 1$ such that $\mu(B \cap T^{-k}A) > 0$.

(iv)⇒(i) Suppose $T^{-1}A = A$ with $\mu(A) > 0$. If $\mu(A^c) > 0$, then by (iv) there exists $k \geq 1$ such that $\mu(A^c \cap T^{-k}A) > 0$. Since $T^{-k}A = A$, it follows that $\mu(A^c \cap A) > 0$, a contradiction. Hence, $\mu(A) = 1$ and T is ergodic. $\square$

1.7 Other Characterizations of Ergodicity

We denote by $L^0(X,\mathcal{F},\mu)$ the space of all complex valued measurable functions on the probability space $(X,\mathcal{F},\mu)$. Let

$$L^p(X,\mathcal{F},\mu)=\{f\in L^0(X,\mathcal{F},\mu):\int_X |f|^p\,\mathrm{d}\mu(x)<\infty\}.$$

We use the subscript $\mathbb{R}$ whenever we are dealing only with real-valued functions.

Let $(X_i,\mathcal{F}_i,\mu_i)$, $i=1,2$ be two probability spaces, and $T:X_1\to X_2$ a measure preserving transformation i.e., $\mu_2(A)=\mu_1(T^{-1}A)$. Define the *induced* operator $U_T:L^0(X_2,\mathcal{F}_2,\mu_2)\to L^0(X_1,\mathcal{F}_1,\mu_1)$ by

$$U_T f = f\circ T.$$

The following properties of U_T are easy to prove.

Proposition 1.7.1 *The operator U_T has the following properties:*

(i) U_T *is linear*

(ii) $U_T(fg)=U_T(f)U_T(g)$

(iii) $U_T c = c$ *for any constant* c.

(iv) U_T *is a positive linear operator*

(v) $U_T 1_B = 1_B\circ T = 1_{T^{-1}B}$ *for all* $B\in\mathcal{F}_2$.

(vi) $\int_{X_1} U_T f\,\mathrm{d}\mu_1=\int_{X_2} f\,\mathrm{d}\mu_2$ *for all* $f\in L^0(X_2,\mathcal{F}_2,\mu_2)$, *(where if one side doesn't exist or is infinite, then the other side has the same property).*

(vii) Let $p\geq 1$. *Then,* $U_T L^p(X_2,\mathcal{F}_2,\mu_2)\subset L^p(X_1,\mathcal{F}_1,\mu_1)$, *and* $||U_T f||_p=||f||_p$ *for all* $f\in L^p(X_2,\mathcal{F}_2,\mu_2)$.

Exercise 1.7.1 Prove Proposition 1.7.1

Using the above properties, we can give the following characterization of ergodicity

Theorem 1.7.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ measure preserving. The following are equivalent:*

(i) *T is ergodic.*

(ii) *If $f \in L^0(X, \mathcal{F}, \mu)$, with $f(Tx) = f(x)$ for all x, then f is a constant a.e.*

(iii) *If $f \in L^0(X, \mathcal{F}, \mu)$, with $f(Tx) = f(x)$ for a.e. x, then f is a constant a.e.*

(iv) *If $f \in L^2(X, \mathcal{F}, \mu)$, with $f(Tx) = f(x)$ for all x, then f is a constant a.e.*

(v) *If $f \in L^2(X, \mathcal{F}, \mu)$, with $f(Tx) = f(x)$ for a.e. x, then f is a constant a.e.*

Proof

The implications (iii)⇒(ii), (ii)⇒(iv), (v)⇒(iv), and (iii)⇒(v) are all clear. It remains to show (i)⇒(iii) and (iv)⇒(i).

(i)⇒(iii) Suppose $f(Tx) = f(x)$ a.e. and assume without any loss of generality that f is real (otherwise we consider separately the real and imaginary parts of f). For each $n \geq 1$ and $k \in \mathbb{Z}$, let

$$X(k, n) = \{x \in X : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\}.$$

Then, $T^{-1}X(k, n)\Delta X(k, n) \subseteq \{x : f(Tx) \neq f(x)\}$ which implies that

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

By ergodicity of T, $\mu(X(k, n)) = 0$ or 1, for each $k \in \mathbb{Z}$. On the other hand, for each $n \geq 1$, we have

$$X = \bigcup_{k \in \mathbb{Z}} X(k, n) \text{ (disjoint union)}.$$

Hence, for each $n \geq 1$, there exists a unique integer k_n such that $\mu\left(X(k_n, n)\right) = 1$. In fact, $X(k_1, 1) \supseteq X(k_2, 2) \supseteq \ldots$, and $\{\frac{k_n}{2^n}\}$ is a bounded increasing sequence, hence $\lim_{n\to\infty} \frac{k_n}{2^n}$ exists. Let $Y = \bigcap_{n\geq 1} X(k_n, n)$, then $\mu(Y) = 1$.

Now, if $x \in Y$, then $0 \leq |f(x) - k_n/2^n| < 1/2^n$ for all n. Hence, $f(x) = \lim_{n\to\infty} \frac{k_n}{2^n}$, and f is a constant on Y.

(iv)⇒(i) Suppose $T^{-1}A = A$ and $\mu(A) > 0$. We want to show that $\mu(A) = 1$. Consider 1_A, the indicator function of A. We have $1_A \in L^2(X, \mathcal{F}, \mu)$, and $1_A \circ T = 1_{T^{-1}A} = 1_A$. Hence, by (iv), 1_A is a constant a.e., hence $1_A = 1$ a.e. and therefore $\mu(A) = 1$. □

1.8 Examples of Ergodic Transformations

Example 1–Irrational Rotations. Consider $([0,1), \mathcal{B}, \lambda)$, where $\mathcal{B}$ is the Lebesgue σ-algebra, and λ Lebesgue measure. For $\theta \in (0,1)$, consider the transformation $T_\theta : [0,1) \to [0,1)$ defined by $T_\theta x = x + \theta \pmod 1$. We have seen in example (a) that T_θ is measure preserving with respect λ. When is T_θ ergodic?
If θ is rational, then T_θ is not ergodic. Consider for example $\theta = 1/4$, then the set

$$A = [0, 1/8) \cup [1/4, 3/8) \cup [1/2, 5/8) \cup [3/4, 7/8)$$

is T_θ-invariant but $\mu(A) = 1/2$.

Exercise 1.8.1 Suppose $\theta = \dfrac{p}{q}$ with *gcd*$(p,q) = 1$. Find a non-trivial T_θ-*invariant* set. Conclude that T_θ is not ergodic if θ is a rational.

Claim. T_θ is ergodic if and only if θ is irrational.

Proof of Claim.

(⇒) The contrapositive statement is given in Exercise 1.8.1 i.e. if θ is rational, then T_θ is not ergodic.

(⇐) Suppose θ is irrational, and let $f \in L^2(X, \mathcal{B}, \lambda)$ be T_θ-invariant. Write f in its Fourier series

$$f(x) = \sum_{n \in Z} a_n e^{2\pi i n x}.$$

Since $f(T_\theta x) = f(x)$, then

$$\begin{aligned} f(T_\theta x) &= \sum_{n\in Z} a_n e^{2\pi in(x+\theta)} = \sum_{n\in Z} a_n e^{2\pi in\theta} e^{2\pi inx} \\ &= f(x) = \sum_{n\in Z} a_n e^{2\pi inx}. \end{aligned}$$

Hence, $\sum_{n\in Z} a_n(1-e^{2\pi in\theta})e^{2\pi inx} = 0$. By the uniqueness of the Fourier coefficients, we have $a_n(1-e^{2\pi in\theta}) = 0$ for all $n \in \mathbb{Z}$. If $n \neq 0$, since θ is irrational we have $1 - e^{2\pi in\theta} \neq 0$. Thus, $a_n = 0$ for all $n \neq 0$, and therefore $f(x) = a_0$ is a constant. By Theorem 1.7.1, T_θ is ergodic.

Exercise 1.8.2 Consider the probability space $([0,1), \mathcal{B}\times\mathcal{B}, \lambda\times\lambda)$, where as above $\mathcal{B}$ is the Lebesgue σ-algebra on $[0,1)$, and λ normalized Lebesgue measure. Suppose $\theta \in (0,1)$ is irrational, and define $T_\theta \times T_\theta : [0,1)\times[0,1) \to [0,1)\times[0,1)$ by

$$T_\theta \times T_\theta(x,y) = (x+\theta \ mod \ (1) \ , y+\theta \ mod \ (1) \ ).$$

Show that $T_\theta \times T_\theta$ is measure preserving, but is **not** ergodic.

Example 2–One (or Two) sided shift. Let $X = \{0,1,\dots k-1\}^{\mathbb{N}}$, $\mathcal{F}$ the σ-algebra generated by the cylinders, and μ the product measure defined on cylinder sets by

$$\mu\left(\{x : x_0 = a_0, \dots x_n = a_n\}\right) = p_{a_0} \dots p_{a_n},$$

where $p = (p_0, p_1, \dots, p_{k-1})$ is a positive probability vector. Consider the left shift T defined on X by $Tx = y$, where $y_n = x_{n+1}$ (See Example (e) in Subsection 1.3). We show that T is ergodic. Let E be a measurable subset of X which is T-invariant i.e., $T^{-1}E = E$. For any $\epsilon > 0$, by Lemma 1.2.1 (see subsection 1.2), there exists $A \in \mathcal{F}$ which is a finite disjoint union of cylinders such that $\mu(E\Delta A) < \epsilon$. Then

$$\begin{aligned} |\mu(E) - \mu(A)| &= |\mu(E\setminus A) - \mu(A\setminus E)| \\ &\leq \mu(E\setminus A) + \mu(A\setminus E) = \mu(E\Delta A) < \epsilon. \end{aligned}$$

Since A depends on finitely many coordinates only, there exists $n_0 > 0$ such that $T^{-n_0}A$ depends on different coordinates than A. Since μ is a product measure, we have

$$\mu(A \cap T^{-n_0}A) = \mu(A)\mu(T^{-n_0}A) = \mu(A)^2.$$

Further,

$$\mu(E\Delta T^{-n_0}A) = \mu(T^{-n_0}E\Delta T^{-n_0}A) = \mu(E\Delta A) < \epsilon,$$

and

$$\mu\left(E\Delta(A\cap T^{-n_0}A)\right) \le \mu(E\Delta A) + \mu(E\Delta T^{-n_0}A) < 2\epsilon.$$

Hence,

$$|\mu(E) - \mu((A\cap T^{-n_0}A))| \le \mu\left(E\Delta(A\cap T^{-n_0}A)\right) < 2\epsilon.$$

Thus,

$$\begin{aligned} |\mu(E)-\mu(E)^2| &\le |\mu(E)-\mu(A)^2| + |\mu(A)^2-\mu(E)^2| \\ &= |\mu(E)-\mu((A\cap T^{-n_0}A))| + (\mu(A)+\mu(E))|\mu(A)-\mu(E)| \\ &< 4\epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that $\mu(E) = \mu(E)^2$, hence $\mu(E) = 0$ or 1. Therefore, T is ergodic.

The following lemma provides, in some cases, a useful tool to verify that a measure preserving transformation defined on $([0,1),\mathcal{B},\mu)$ is ergodic, where $\mathcal{B}$ is the Lebesgue σ-algebra, and μ is a probability measure equivalent to Lebesgue measure λ (i.e., $\mu(A) = 0$ if and only if $\lambda(A) = 0$).

Lemma 1.8.1 *(Knopp's Lemma) . If B is a Lebesgue set and $\mathcal{C}$ is a class of subintervals of $[0,1)$ satisfying*

(a) *every open subinterval of $[0,1)$ is at most a countable union of disjoint elements from $\mathcal{C}$,*

(b) $\forall A\in\mathcal{C}$, $\lambda(A\cap B)\ge\gamma\lambda(A)$, *where $\gamma>0$ is independent of A,*

then $\lambda(B) = 1$.

Proof The proof is done by contradiction. Suppose $\lambda(B^c) > 0$. Given $\varepsilon > 0$ there exists by Lemma 1.2.1 a set E_ε that is a finite disjoint union of open intervals such that $\lambda(B^c\triangle E_\varepsilon) < \varepsilon$. Now by conditions (a) and (b) (that is, writing E_ε as a countable union of disjoint elements of $\mathcal{C}$) one gets that $\lambda(B\cap E_\varepsilon)\ge\gamma\lambda(E_\varepsilon)$.

Also from our choice of E_ε and the fact that

$$\lambda(B^c \triangle E_\varepsilon) \geq \lambda(B \cap E_\varepsilon) \geq \gamma\lambda(E_\varepsilon) \geq \gamma\lambda(B^c \cap E_\varepsilon) > \gamma(\lambda(B^c) - \varepsilon),$$

we have that

$$\gamma(\lambda(B^c) - \varepsilon) \; < \; \lambda(B^c \triangle E_\varepsilon) \; < \; \varepsilon \; .$$

Hence $\gamma\lambda(B^c) < \varepsilon + \gamma\varepsilon$, and since $\varepsilon > 0$ is arbitrary, we get a contradiction. □

Example 3–Multiplication by 2 modulo 1–Consider $([0,1), \mathcal{B}, \lambda)$ be as in Example (1) above, and let $T : X \to X$ be given by

$$Tx = 2x \bmod 1 = \begin{cases} 2x & 0 \leq x < 1/2 \\ 2x - 1 & 1/2 \leq x < 1, \end{cases}$$

(see Example (b), subsection 1.3). We have seen that T is measure preserving. We will use Lemma 1.8.1 to show that T is ergodic. Let $\mathcal{C}$ be the collection of all intervals of the form $[k/2^n, (k+1)/2^n)$ with $n \geq 1$ and $0 \leq k \leq 2^n - 1$. Notice that the the set $\{k/2^n : n \geq 1, \, 0 \leq k < 2^n - 1\}$ of dyadic rationals is dense in $[0,1)$, hence each open interval is at most a countable union of disjoint elements of $\mathcal{C}$. Hence, $\mathcal{C}$ satisfies the first hypothesis of Knopp's Lemma. Now, T^n maps each dyadic interval of the form $[k/2^n, (k+1)/2^n)$ linearly onto $[0,1)$, (we call such an interval dyadic of order n); in fact, $T^n x = 2^n x \bmod(1)$. Let $B \in \mathcal{B}$ be T-invariant, and assume $\lambda(B) > 0$. Let $A \in \mathcal{C}$, and assume that A is dyadic of order n. Then, $T^n A = [0,1)$ and

$$\begin{aligned} \lambda(A \cap B) &= \lambda(A \cap T^{-n}B) = \frac{1}{\lambda(A)}\lambda(T^n A \cap B) \\ &= \frac{1}{2^n}\lambda(B) = \lambda(A)\lambda(B). \end{aligned}$$

Thus, the second hypothesis of Knopp's Lemma is satisfied with $\gamma = \lambda(B) > 0$. Hence, $\lambda(B) = 1$. Therefore T is ergodic.

Exercise 1.8.3 Let $\beta > 1$ be a non-integer, and consider the transformation $T_\beta : [0,1) \to [0,1)$ given by $T_\beta x = \beta x \bmod(1) \; = \beta x - \lfloor \beta x \rfloor$. Use Lemma 1.8.1 to show that T_β is ergodic with respect to Lebesgue measure λ, i.e. if $T_\beta^{-1} A = A$, then $\lambda(A) = 0$ or 1.

Example 4–Induced transformations of ergodic transformations– Let T be an ergodic measure preserving transformation on the probability space $(X, \mathcal{F}, \mu)$, and $A \in \mathcal{F}$ with $\mu(A) > 0$. Consider the induced transformation T_A on $(A, \mathcal{F} \cap A, \mu_A)$ of T (see subsection 1.5). Recall that $T_A x = T^{n(x)}x$, where $n(x) := \inf\{n \geq 1 : T^n x \in A\}$. Let (as before)

$$A_k = \{x \in A : n(x) = k\}$$

$$B_k = \{x \in X \setminus A : Tx, \ldots, T^{k-1}x \notin A, T^k x \in A\}.$$

Proposition 1.8.1 *If T is ergodic on $(X, \mathcal{F}, \mu)$, then T_A is ergodic on $(A, \mathcal{F} \cap A, \mu_A)$.*

Proof Let $C \in \mathcal{F} \cap A$ be such that $T_A^{-1}C = C$. We want to show that $\mu_A(C) = 0$ or 1; equivalently, $\mu(C) = 0$ or $\mu(C) = \mu(A)$. Since $A = \bigcup_{k\geq 1} A_k$, we have $C = T_A^{-1}C = \bigcup_{k\geq 1} A_k \cap T^{-k}C$. Let $E = \bigcup_{k\geq 1} B_k \cap T^{-k}C$, and $F = E \cup C$ (disjoint union). Recall that (see subsection 1.5) $T^{-1}A = A_1 \cup B_1$, and $T^{-1}B_k = A_{k+1} \cup B_{k+1}$. Hence,

$$\begin{aligned}
T^{-1}F &= T^{-1}E \cup T^{-1}C \\
&= \bigcup_{k\geq 1} \left[(A_{k+1} \cup B_{k+1}) \cap T^{-(k+1)}C\right] \cup \left[(A_1 \cup B_1) \cap T^{-1}C\right] \\
&= \bigcup_{k\geq 1} (A_k \cap T^{-k}C) \cup \bigcup_{k\geq 1} (B_k \cap T^{-k}C) \\
&= C \cup E = F.
\end{aligned}$$

Hence, F is T-invariant, and by ergodicity of T we have $\mu(F) = 0$ or 1.
–If $\mu(F) = 0$, then $\mu(C) = 0$, and hence $\mu_A(C) = 0$.
–If $\mu(F) = 1$, then $\mu(X \setminus F) = 0$. Since

$$X \setminus F = (A \setminus C) \cup ((X \setminus A) \setminus E) \supseteq A \setminus C,$$

it follows that

$$\mu(A \setminus C) \leq \mu(X \setminus F) = 0.$$

Since $\mu(A \setminus C) = \mu(A) - \mu(C)$, we have $\mu(A) = \mu(C)$, i.e., $\mu_A(C) = 1$. $\square$

Exercise 1.8.4 Show that if T_A is ergodic and $\mu\left(\bigcup_{k\geq 1} T^{-k}A\right) = 1$, then, T is ergodic.

Chapter 2

The Ergodic Theorem

2.1 The Ergodic Theorem and its consequences

The Ergodic Theorem is also known as Birkhoff's Ergodic Theorem or the Individual Ergodic Theorem (1931). This theorem is in fact a generalization of the Strong Law of Large Numbers (SLLN) which states that for a sequence $Y_1, Y_2, \ldots$ of i.i.d. random variables on a probability space $(X, \mathcal{F}, \mu)$, with $\mathrm{E}|Y_i| < \infty$; one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_i \;=\; \mathrm{E}Y_1 \text{ (a.e.)}.$$

For example consider $X = \{0,1\}^{\mathbb{N}}$, $\mathcal{F}$ the σ-algebra generated by the cylinder sets, and μ the uniform product measure, i.e.,

$$\mu\left(\{x : x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n\}\right) = 1/2^n.$$

Suppose one is interested in finding the frequency of the digit 1. More precisely, for a.e. x we would like to find

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \le i \le n : x_i = 1\}.$$

Using the Strong Law of Large Numbers one can answer this question easily. Define

$$Y_i(x) \;:=\; \begin{cases} 1, & \text{if } x_i = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Since μ is product measure, it is easy to see that $Y_1, Y_2, \ldots$ form an i.i.d. Bernoulli process, and $EY_i = E|Y_i| = 1/2$. Further, $\#\{1 \leq i \leq n : x_i = 1\} = \sum_{i=1}^n Y_i(x)$. Hence, by SLLN one has

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \leq i \leq n : x_i = 1\} = \frac{1}{2}.$$

Suppose now we are interested in the frequency of the block 011, i.e., we would like to find

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \leq i \leq n : x_i = 0, x_{i+1} = 1, x_{i+2} = 1\}.$$

We can start as above by defining random variables

$$Z_i(x) := \begin{cases} 1, & \text{if } x_i = 0, x_{i+1} = 1, x_{i+2} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\frac{1}{n}\#\{1 \leq i \leq n : x_i = 0, x_{i+1} = 1, x_{i+2} = 1\} = \frac{1}{n}\sum_{i=1}^{n} Z_i(x).$$

It is not hard to see that this sequence is stationary but not independent. So one cannot directly apply the strong law of large numbers. Notice that if T is the left shift on X, then $Y_n = Y_1 \circ T^{n-1}$ and $Z_n = Z_1 \circ T^{n-1}$.
In general, suppose $(X, \mathcal{F}, \mu)$ is a probability space and $T : X \to X$ a measure preserving transformation. For $f \in L^1(X, \mathcal{F}, \mu)$, we would like to know under what conditions does the limit $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)$ exist a.e. If it does exist what is its value? This is answered by the Ergodic Theorem which was originally proved by G.D. Birkhoff in 1931. Since then, several proofs of this important theorem have been obtained; here we present a recent proof given by T. Kamae and M.S. Keane in [KK].

Theorem 2.1.1 (The Ergodic Theorem) *Let $(X, \mathcal{F}, \mu)$ be a probability space and $T : X \to X$ a measure preserving transformation. Then, for any f in $L^1(\mu)$,*

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

exists a.e., is T-invariant and $\int_X f \,\mathrm{d}\mu = \int_X f^* \,\mathrm{d}\mu$. If moreover T is ergodic, then f^* is a constant a.e. and $f^* = \int_X f \,\mathrm{d}\mu$.

For the proof of the above theorem, we need the following simple lemma.

Lemma 2.1.1 *Let $M > 0$ be an integer, and suppose $\{a_n\}_{n\geq 0}$, $\{b_n\}_{n\geq 0}$ are sequences of non-negative real numbers such that for each $n = 0, 1, 2, \ldots$ there exists an integer $1 \leq m \leq M$ with*

$$a_n + \cdots + a_{n+m-1} \geq b_n + \cdots + b_{n+m-1}.$$

Then, for each positive integer $N > M$, one has

$$a_0 + \cdots + a_{N-1} \geq b_0 + \cdots + b_{N-M-1}.$$

Proof of Lemma 2.1.1 Using the hypothesis we recursively find integers $m_0 < m_1 < \ldots < m_k < N$ with the following properties

$$m_0 \leq M,\ m_{i+1} - m_i \leq M \text{ for } i = 0, \ldots, k-1, \text{ and } N - m_k < M,$$
$$a_0 + \ldots + a_{m_0-1} \geq b_0 + \ldots + b_{m_0-1},$$
$$a_{m_0} + \ldots + a_{m_1-1} \geq b_{m_0} + \ldots + b_{m_1-1},$$
$$\vdots$$
$$a_{m_{k-1}} + \ldots + a_{m_k-1} \geq b_{m_{k-1}} + \ldots + b_{m_k-1}.$$

Then,

$$\begin{aligned} a_0 + \ldots + a_{N-1} &\geq a_0 + \ldots + a_{m_k-1} \\ &\geq b_0 + \ldots + b_{m_k-1} \geq b_0 + \ldots b_{N-M-1}. \end{aligned}$$

$\square$

Proof of Theorem 2.1.1 Assume with no loss of generality that $f \geq 0$ (otherwise we write $f = f^+ - f^-$, and we consider each part separately). Let $f_n(x) = f(x) + \ldots + f(T^{n-1}x)$, $\overline{f}(x) = \limsup_{n\to\infty} \dfrac{f_n(x)}{n}$, and $\underline{f}(x) = \liminf_{n\to\infty} \dfrac{f_n(x)}{n}$. Then $\overline{f}$ and $\underline{f}$ are T-invariant. This follows from

$$\begin{aligned} \overline{f}(Tx) &= \limsup_{n\to\infty} \frac{f_n(Tx)}{n} \\ &= \limsup_{n\to\infty} \left[\frac{f_{n+1}(x)}{n+1} \cdot \frac{n+1}{n} - \frac{f(x)}{n} \right] \\ &= \limsup_{n\to\infty} \frac{f_{n+1}(x)}{n+1} = \overline{f}(x). \end{aligned}$$

(Similarly $\underline{f}$ is T-invariant). Now, to prove that f^* exists, is integrable and T-invariant, it is enough to show that

$$\int_X \underline{f}\, \mathrm{d}\mu \geq \int_X f\, \mathrm{d}\mu \geq \int_X \overline{f}\, \mathrm{d}\mu.$$

For since $\overline{f} - \underline{f} \geq 0$, this would imply that $\overline{f} = \underline{f} = f^*$. a.e.
We first prove that $\int_X \overline{f} d\mu \leq \int_X f\, \mathrm{d}\mu$. Fix any $0 < \epsilon < 1$, and let $L > 0$ be any real number. By definition of $\overline{f}$, for any $x \in X$, there exists an integer $m > 0$ such that

$$\frac{f_m(x)}{m} \geq \min(\overline{f}(x), L)(1-\epsilon).$$

Now, for any $\delta > 0$ there exists an integer $M > 0$ such that the set

$$X_0 = \{x \in X : \exists\ 1 \leq m \leq M \text{ with } f_m(x) \geq m\ \min(\overline{f}(x), L)(1-\epsilon)\}$$

has measure at least $1 - \delta$. Define F on X by

$$F(x) = \begin{cases} f(x) & x \in X_0 \\ L & x \notin X_0. \end{cases}$$

Notice that $f \leq F$ (why?). For any $x \in X$, let $a_n = a_n(x) = F(T^n x)$, and $b_n = b_n(x) = \min(\overline{f}(x), L)(1-\epsilon)$ (so b_n is independent of n).We now show that $\{a_n\}$ and $\{b_n\}$ satisfy the hypothesis of Lemma 2.1.1 with $M > 0$ as above. For any $n = 0, 1, 2, \ldots$

–if $T^n x \in X_0$, then there exists $1 \leq m \leq M$ such that

$$\begin{aligned} f_m(T^n x) &\geq m\ \min(\overline{f}(T^n x), L)(1-\epsilon) \\ &= m\ \min(\overline{f}(x), L)(1-\epsilon) \\ &= b_n + \ldots + b_{n+m-1}. \end{aligned}$$

Hence,

$$\begin{aligned} a_n + \ldots + a_{n+m-1} &= F(T^n x) + \ldots + F(T^{n+m-1}x) \\ &\geq f(T^n x) + \ldots + f(T^{n+m-1}x) = f_m(T^n x) \\ &\geq b_n + \ldots + b_{n+m-1}. \end{aligned}$$

–If $T^n x \notin X_0$, then take $m = 1$ since

$$a_n = F(T^n x) = L \geq \min(\overline{f}(x), L)(1-\epsilon) = b_n.$$

Hence by Lemma 2.1.1 for all integers $N > M$ one has

$$F(x) + \ldots + F(T^{N-1}x) \geq (N - M) \min(\overline{f}(x), L)(1 - \epsilon).$$

Integrating both sides, and using the fact that T is measure preserving one gets

$$N \int_X F(x) \, \mathrm{d}\mu(x) \geq (N - M) \int_X \min(\overline{f}(x), L)(1 - \epsilon) \, \mathrm{d}\mu(x).$$

Since

$$\int_X F(x) \, \mathrm{d}\mu(x) = \int_{X_0} f(x) \, \mathrm{d}\mu(x) + L\mu(X \setminus X_0),$$

one has

$$\begin{aligned}
\int_X f(x) \, \mathrm{d}\mu(x) &\geq \int_{X_0} f(x) \, \mathrm{d}\mu(x) \\
&= \int_X F(x) \, \mathrm{d}\mu(x) - L\mu(X \setminus X_0) \\
&\geq \frac{(N - M)}{N} \int_X \min(\overline{f}(x), L)(1 - \epsilon) \, \mathrm{d}\mu(x) - L\delta.
\end{aligned}$$

Now letting first $N \to \infty$, then $\delta \to 0$, then $\epsilon \to 0$, and lastly $L \to \infty$ one gets together with the monotone convergence theorem that $\overline{f}$ is integrable, and

$$\int_X f(x) \, \mathrm{d}\mu(x) \geq \int_X \overline{f}(x) \, \mathrm{d}\mu(x).$$

We now prove that

$$\int_X f(x) \, \mathrm{d}\mu(x) \leq \int_X \underline{f}(x) \, \mathrm{d}\mu(x).$$

Fix $\epsilon > 0$, for any $x \in X$ there exists an integer m such that

$$\frac{f_m(x)}{m} \leq (\underline{f}(x) + \epsilon).$$

For any $\delta > 0$ there exists an integer $M > 0$ such that the set

$$Y_0 = \{x \in X : \exists\ 1 \leq m \leq M \text{ with } f_m(x) \leq m\,(\underline{f}(x) + \epsilon)\}$$

has measure at least $1-\delta$. Define G on X by

$$G(x) = \begin{cases} f(x) & x \in Y_0 \\ 0 & x \notin Y_0. \end{cases}$$

Notice that $G \le f$. Let $b_n = G(T^n x)$, and $a_n = \underline{f}(x)+\epsilon$ (so a_n is independent of n). One can easily check that the sequences $\{a_n\}$ and $\{b_n\}$ satisfy the hypothesis of Lemma 2.1.1 with $M > 0$ as above. hence

$$G(x) + \ldots + G(T^{N-M-1}x) \le N(\underline{f}(x) + \epsilon).$$

Integrating both sides yields

$$(N-M)\int_X G(x)d\mu(x) \le N(\int_X \underline{f}(x)d\mu(x) + \epsilon).$$

Since $f \ge 0$, the measure ν defined by $\nu(A) = \int_A f(x)\,\mathrm{d}\mu(x)$ is absolutely continuous with respect to the measure μ. Hence, there exists $\delta_0 > 0$ such that if $\mu(A) < \delta$, then $\nu(A) < \delta_0$. Since $\mu(X \setminus Y_0) < \delta$, then $\nu(X \setminus Y_0) = \int_{X\setminus Y_0} f(x)d\mu(x) < \delta_0$. Hence,

$$\begin{aligned} \int_X f(x)\,\mathrm{d}\mu(x) &= \int_X G(x)\,\mathrm{d}\mu(x) + \int_{X\setminus Y_0} f(x)\,\mathrm{d}\mu(x) \\ &\le \frac{N}{N-M}\int_X (\underline{f}(x)+\epsilon)\,\mathrm{d}\mu(x) + \delta_0. \end{aligned}$$

Now, let first $N \to \infty$, then $\delta \to 0$ (and hence $\delta_0 \to 0$), and finally $\epsilon \to 0$, one gets

$$\int_X f(x)\,\mathrm{d}\mu(x) \le \int_X \underline{f}(x)\,\mathrm{d}\mu(x).$$

This shows that

$$\int_X \underline{f}\,\mathrm{d}\mu \ge \int_X f\,\mathrm{d}\mu \ge \int_X \overline{f}\,\mathrm{d}\mu,$$

hence, $\overline{f} = \underline{f} = f^*$ a.e., and f^* is T-invariant. In case T is ergodic, then the T-invariance of f^* implies that f^* is a constant a.e. Therefore,

$$f^*(x) = \int_X f^*(y)d\mu(y) = \int_X f(y)\,\mathrm{d}\mu(y).$$

$\square$

Remarks

(1) Let us study further the limit f^* in the case that T is not ergodic. Let $\mathcal{I}$ be the sub-σ-algebra of $\mathcal{F}$ consisting of all T-invariant subsets $A \in \mathcal{F}$. Notice that if $f \in L^1(\mu)$, then the *conditional expectation* of f given $\mathcal{I}$ (denoted by $E_\mu(f|\mathcal{I})$), is the unique a.e. $\mathcal{I}$-measurable $L^1(\mu)$ function with the property that

$$\int_A f(x)\, \mathrm{d}\mu(x) = \int_A E_\mu(f|\mathcal{I})(x)\, \mathrm{d}\mu(x)$$

for all $A \in \mathcal{I}$ i.e., $T^{-1}A = A$. We claim that $f^* = E_\mu(f|\mathcal{I})$. Since the limit function f^* is T-invariant, it follows that f^* is $\mathcal{I}$-measurable. Furthermore, for any $A \in \mathcal{I}$, by the ergodic theorem and the T-invariance of 1_A,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (f1_A)(T^i x) = 1_A(x) \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = 1_A(x) f^*(x) \ a.e.$$

and

$$\int_X f1_A(x)\, \mathrm{d}\mu(x) = \int_X f^* 1_A(x)\, \mathrm{d}\mu(x).$$

This shows that $f^* = E_\mu(f|\mathcal{I})$.

(2) Suppose T is ergodic and measure preserving with respect to μ, and let ν be a probability measure which is equivalent to μ (i.e. μ and ν have the same sets of measure zero so $\mu(A) = 0$ if and only if $\nu(A) = 0$), then for every $f \in L^1(\mu)$ one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \ = \ \int_X f\, \mathrm{d}\mu$$

ν a.e.

Exercise 2.1.1 (*Kac's Lemma*) Let T be an invertible, measure preserving and ergodic transformation on a probability space $(X, \mathcal{F}, \mu)$. Let A be a measurable subset of X of positive μ measure, and denote by n the first return time map and let T_A be the induced transformation of T on A (see section 1.5). Prove that

$$\int_A n(x)\, \mathrm{d}\mu = 1.$$

Conclude that $n(x) \in L^1(A, \mu_A)$, and that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} n(T_A^i(x)) = \frac{1}{\mu(A)},$$

almost everywhere on A.

Exercise 2.1.2 Let $\beta = \dfrac{1+\sqrt{5}}{2}$, and consider the transformation $T_\beta : [0,1) \to [0,1)$ given by $T_\beta x = \beta x \bmod(1) = \beta x - \lfloor \beta x \rfloor$. Define b_1 on $[0,1)$ by

$$b_1(x) = \begin{cases} 0 & \text{if } 0 \le x < 1/\beta \\ 1 & \text{if } 1/\beta \le x < 1, \end{cases}$$

Fix $k \ge 0$. Find the a.e. value (with respect to Lebesgue measure) of the following limit

$$\lim_{n\to\infty} \frac{1}{n}\#\{1 \le i \le n : b_i = 0, b_{i+1} = 0, \ldots, b_{i+k} = 0\}.$$

Using the Ergodic Theorem, one can give yet another characterization of ergodicity.

Corollary 2.1.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Then, T is ergodic if and only if for all $A, B \in \mathcal{F}$, one has*

$$\lim_{n\to\infty} \frac{1}{n-1}\sum_{i=0}^{n} \mu(T^{-i}A \cap B) = \mu(A)\mu(B). \tag{2.1}$$

Proof Suppose T is ergodic, and let $A, B \in \mathcal{F}$. Since the indicator function $1_A \in L^1(X, \mathcal{F}, \mu)$, by the ergodic theorem one has

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} 1_A(T^i x) = \int_X 1_A(x)\,\mathrm{d}\mu(x) = \mu(A) \text{ a.e.}$$

Then,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_{T^{-i}A\cap B}(x) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_{T^{-i}A}(x)1_B(x)\\ &= 1_B(x)\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_A(T^ix)\\ &= 1_B(x)\mu(A) \text{ a.e.}\end{aligned}$$

Since for each n, the function $\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_{T^{-i}A\cap B}$ is dominated by the constant function 1, it follows by the dominated convergence theorem that

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n}\mu(T^{-i}A\cap B) &= \int_X \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}1_{T^{-i}A\cap B}(x)\,\mathrm{d}\mu(x)\\ &= \int_X 1_B\mu(A)\,\mathrm{d}\mu(x) = \mu(A)\mu(B).\end{aligned}$$

Conversely, suppose (2.1) holds for every $A, B\in\mathcal{F}$. Let $E\in\mathcal{F}$ be such that $T^{-1}E = E$ and $\mu(E) > 0$. By invariance of E, we have $\mu(T^{-i}E\cap E) = \mu(E)$, hence

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}E\cap E) = \mu(E).$$

On the other hand, by (2.1)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}E\cap E) = \mu(E)^2.$$

Hence, $\mu(E) = \mu(E)^2$. Since $\mu(E) > 0$, this implies $\mu(E) = 1$. Therefore, T is ergodic. □

To show ergodicity one needs to verify equation (2.1) for sets A and B belonging to a generating semi-algebra only as the next proposition shows.

Proposition 2.1.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $\mathcal{S}$ a generating semi-algebra of $\mathcal{F}$. Let $T : X \to X$ be a measure preserving transformation. Then, T is ergodic if and only if for all $A, B \in \mathcal{S}$, one has*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}A\cap B) = \mu(A)\mu(B). \tag{2.2}$$

Proof We only need to show that if (2.2) holds for all $A, B \in \mathcal{S}$, then it holds for all $A, B \in \mathcal{F}$. Let $\epsilon > 0$, and $A, B \in \mathcal{F}$. Then, by Lemma 1.2.1 (subsection 1.2) there exist sets A_0, B_0 each of which is a finite disjoint union of elements of $\mathcal{S}$ such that

$$\mu(A\Delta A_0) < \epsilon, \text{ and } \mu(B\Delta B_0) < \epsilon.$$

Since,

$$(T^{-i}A \cap B)\Delta(T^{-i}A_0 \cap B_0) \subseteq (T^{-i}A\Delta T^{-i}A_0) \cup (B\Delta B_0),$$

it follows that

$$\begin{aligned} |\mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0)| &\leq \mu\left[(T^{-i}A \cap B)\Delta(T^{-i}A_0 \cap B_0)\right] \\ &\leq \mu(T^{-i}A\Delta T^{-i}A_0) + \mu(B\Delta B_0) \\ &< 2\epsilon. \end{aligned}$$

Further,

$$\begin{aligned} |\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| &\leq \mu(A)|\mu(B) - \mu(B_0)| + \mu(B_0)|\mu(A) - \mu(A_0)| \\ &\leq |\mu(B) - \mu(B_0)| + |\mu(A) - \mu(A_0)| \\ &\leq \mu(B\Delta B_0) + \mu(A\Delta A_0) \\ &< 2\epsilon. \end{aligned}$$

Hence,

$$\begin{aligned} &\left| \left(\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) \right) - \left(\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}A_0 \cap B_0) - \mu(A_0)\mu(B_0) \right) \right| \\ \leq\ & \frac{1}{n}\sum_{i=0}^{n-1} \left|\mu(T^{-i}A \cap B) + \mu(T^{-i}A_0 \cap B_0)\right| - |\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| \\ <\ & 4\epsilon. \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \left[\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) \right] = 0.$$

$\square$

Theorem 2.1.2 *Suppose μ_1 and μ_2 are probability measures on $(X, \mathcal{F})$, and $T : X \to X$ is measurable and measure preserving with respect to μ_1 and μ_2. Then,*

(i) if T is ergodic with respect to μ_1, and μ_2 is absolutely continuous with respect to μ_1, then $\mu_1 = \mu_2$,

(ii) if T is ergodic with respect to μ_1 and μ_2, then either $\mu_1 = \mu_2$ or μ_1 and μ_2 are singular with respect to each other.

Proof (i) Suppose T is ergodic with respect to μ_1 and μ_2 is absolutely continuous with respect to μ_1. For any $A \in \mathcal{F}$, by the ergodic theorem for a.e. x one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \mu_1(A).$$

Let

$$C_A = \{x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \mu_1(A)\},$$

then $\mu_1(C_A) = 1$, and by absolute continuity of μ_2 one has $\mu_2(C_A) = 1$. Since T is measure preserving with respect to μ_2, for each $n \geq 1$ one has

$$\frac{1}{n} \sum_{i=0}^{n-1} \int_X 1_A(T^i x)\, \mathrm{d}\mu_2(x) = \mu_2(A).$$

On the other hand, by the dominated convergence theorem one has

$$\lim_{n\to\infty} \int_X \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) d\mu_2(x) = \int_X \mu_1(A)\, \mathrm{d}\mu_2(x).$$

This implies that $\mu_1(A) = \mu_2(A)$. Since $A \in \mathcal{F}$ is arbitrary, we have $\mu_1 = \mu_2$.

(ii) Suppose T is ergodic with respect to μ_1 and μ_2. Assume that $\mu_1 \neq \mu_2$. Then, there exists a set $A \in \mathcal{F}$ such that $\mu_1(A) \neq \mu_2(A)$. For $i = 1, 2$ let

$$C_i = \{x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} 1_A(T^j x) = \mu_i(A)\}.$$

By the ergodic theorem $\mu_i(C_i) = 1$ for $i = 1, 2$. Since $\mu_1(A) \neq \mu_2(A)$, then $C_1 \cap C_2 = \emptyset$. Thus μ_1 and μ_2 are supported on disjoint sets, and hence μ_1 and μ_2 are mutually singular. $\square$

We end this subsection with a short discussion that the assumption of ergodicity is not very restrictive. Let T be a transformation on the probability space $(X, \mathcal{F}, \mu)$, and suppose T is measure preserving but not necessarily ergodic. We assume that X is a complete separable metric space, and $\mathcal{F}$ the corresponding Borel σ-algebra (in order to make sure that the conditional expectation is well-defined a.e.). Let $\mathcal{I}$ be the sub-σ-algebra of T-invariant measurable sets. We can decompose μ into T-invariant ergodic components in the following way. For $x \in X$, define a measure μ_x on $\mathcal{F}$ by

$$\mu_x(A) = E_\mu(1_A|\mathcal{I})(x).$$

Then, for any $f \in L^1(X, \mathcal{F}, \mu)$,

$$\int_X f(y)\, \mathrm{d}\mu_x(y) = E_\mu(f|\mathcal{I})(x).$$

Note that

$$\mu(A) = \int_X E_\mu(1_A|\mathcal{I})(x)\, \mathrm{d}\mu(x) = \int_X \mu_x(A)\, \mathrm{d}\mu(x),$$

and that $E_\mu(1_A|\mathcal{I})(x)$ is T-invariant. We show that μ_x is T-invariant and ergodic for a.e. $x \in X$. So let $A \in \mathcal{F}$, then for a.e. x

$$\mu_x(T^{-1}A) = E_\mu(1_A \circ T|\mathcal{I})(x) = E_\mu(I_A|\mathcal{I})(Tx) = E_\mu(I_A|\mathcal{I})(x) = \mu_x(A).$$

Now, let $A \in \mathcal{F}$ be such that $T^{-1}A = A$. Then, 1_A is T-invariant, and hence $\mathcal{I}$-measurable. Then,

$$\mu_x(A) = E_\mu(1_A|\mathcal{I})(x) = 1_A(x) \ \ a.e.$$

Hence, for a.e. x and for any $B \in \mathcal{F}$,

$$\mu_x(A \cap B) = E_\mu(1_A 1_B|\mathcal{I})(x) = 1_A(x) E_\mu(1_B|\mathcal{I})(x) = \mu_x(A)\mu_x(B).$$

In particular, if $A = B$, then the latter equality yields $\mu_x(A) = \mu_x(A)^2$ which implies that for a.e. x, $\mu_x(A) = 0$ or 1. Therefore, μ_x is ergodic. (One in fact needs to work a little harder to show that one can find a set N of μ-measure zero, such that for any $x \in X \setminus N$, and any T-invariant set A, one has $\mu_x(A) = 0$ or 1. In the above analysis the a.e. set depended on the choice of A. Hence, the above analysis is just a rough sketch of the proof of what is called *the ergodic decomposition* of measure preserving transformations.)

2.2 Characterization of Irreducible Markov Chains

Consider the Markov Chain in Example(f) subsection 1.3. That is $X = \{0, 1, \dots N-1\}^{\mathbb{Z}}$, $\mathcal{F}$ the σ-algebra generated by the cylinders, $T : X \to X$ the left shift, and μ the Markov measure defined by the stochastic $N \times N$ matrix $P = (p_{ij})$, and the positive probability vector $\pi = (\pi_0, \pi_1, \dots, \pi_{N-1})$ satisfying $\pi P = \pi$. That is

$$\mu(\{x : x_0 = i_0, x_1 = i_1, \dots x_n = i_n\}) = \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}.$$

We want to find necessary and sufficient conditions for T to be ergodic. To achieve this, we first set

$$Q = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k,$$

where $P^k = (p_{ij}^{(k)})$ is the k^{th} power of the matrix P, and P^0 is the $k \times k$ identity matrix. More precisely, $Q = (q_{ij})$, where

$$q_{ij} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}.$$

Lemma 2.2.1 *For each $i, j \in \{0, 1, \dots N-1\}$, the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$ exists, i.e., q_{ij} is well-defined.*

Proof For each n,

$$\frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} = \frac{1}{\pi_i} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X : x_0 = i, x_k = j\}).$$

Since T is measure preserving, by the ergodic theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{x:x_k=j\}}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{x:x_0=j\}}(T^k x) = f^*(x),$$

where f^* is T-invariant and integrable. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{x:x_0=i,x_k=j\}}(x) = 1_{\{x:x_0=i\}}(x) \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{x:x_0=j\}}(T^k x) = f^*(x) 1_{\{x:x_0=i\}}(x).$$

Since $\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{x:x_0=i,x_k=j\}}(x) \leq 1$ for all n, by the dominated convergence theorem,

$$\begin{aligned}
q_{ij} &= \frac{1}{\pi_i}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(\{x\in X : x_0=i, x_k=j\}) \\
&= \frac{1}{\pi_i}\int_X \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{x:x_0=i,x_k=j\}}(x)\,\mathrm{d}\mu(x) \\
&= \frac{1}{\pi_i}\int_X f^*(x)1_{\{x:x_0=i\}}(x)\,\mathrm{d}\mu(x) \\
&= \frac{1}{\pi_i}\int_{\{x:x_0=i\}} f^*(x)\,\mathrm{d}\mu(x)
\end{aligned}$$

which is finite since f^* is integrable. Hence q_{ij} exists. □

Exercise 2.2.1 Show that the matrix Q has the following properties*:*
(a) Q is stochastic.
(b) $Q = QP = PQ = Q^2$.
(c) $\pi Q = \pi$.

We now give a characterization for the ergodicity of T. Recall that the matrix P is said to be irreducible if for every $i, j \in \{0, 1, \ldots N-1\}$, there exists $n \geq 1$ such that $p_{ij}^{(n)} > 0$.

Theorem 2.2.1 *The following are equivalent,*

(i) T *is ergodic.*

(ii) All rows of Q *are identical.*

(iii) $q_{ij} > 0$ *for all* i, j.

(iv) P *is irreducible.*

(v) 1 *is a simple eigenvalue of* P.

Proof
(i) ⇒ (ii) By the ergodic theorem for each i, j,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{x:x_0=i,x_k=j\}}(x) = 1_{\{x:x_0=i\}}(x)\pi_j.$$

By the dominated convergence theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X : x_0 = i, x_k = j\}) = \pi_i \pi_j.$$

Hence,

$$q_{ij} = \frac{1}{\pi_i} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X : x_0 = i, x_k = j\}) = \pi_j,$$

i.e., q_{ij} is independent of i. Therefore, all rows of Q are identical.

(ii) $\Rightarrow$ (iii) If al the rows of Q are identical, then all the columns of Q are constants. Thus, for each j there exists a constant c_j such that $q_{ij} = c_j$ for all i. Since $\pi Q = \pi$, it follows that $q_{ij} = c_j = \pi_j > 0$ for all i, j.

(iii) $\Rightarrow$ (iv) For any i, j

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} \;=\; q_{ij} > 0.$$

Hence, there exists n such that $p_{ij}^{(n)} > 0$, therefore P is irreducible.

(iv) $\Rightarrow$ (iii) Suppose P is irreducible. For any state $i \in \{0, 1, \ldots, N-1\}$, let $S_i = \{j : q_{ij} > 0\}$. Since Q is a stochastic matrix, it follows that $S_i \neq \emptyset$. Let $l \in S_i$, then $q_{il} > 0$. Since $Q = QP = QP^n$ for all n, then for any state j

$$q_{ij} = \sum_{m=0}^{N-1} q_{im} p_{mj}^{(n)} \geq q_{il} p_{lj}^{(n)}$$

for any n. Since P is irreducible, there exists n such that $p_{lj}^{(n)} > 0$. Hence, $q_{ij} > 0$ for all i, j.

(iii) $\Rightarrow$ (ii) Suppose $q_{ij} > 0$ for all $j = 0, 1, \ldots, N-1$. Fix any state j, and let $q_j = \max_{0 \leq i \leq N-1} q_{ij}$. Suppose that not all the q_{ij}'s are the same. Then there exists $k \in \{0, 1, \ldots, N-1\}$ such that $q_{kj} < q_j$. Since Q is stochastic and $Q^2 = Q$, then for any $i \in \{0, 1, \ldots, N-1\}$ we have,

$$q_{ij} = \sum_{l=0}^{N-1} q_{il} q_{lj} < \sum_{l=0}^{N-1} q_{il} q_j = q_j.$$

This implies that $q_j = \max_{0\leq i\leq N-1} q_{ij} < q_j$, a contradiction. Hence, the columns of Q are constants, or all the rows are identical.

(ii) $\Rightarrow$ (i) Suppose all the rows of Q are identical. We have shown above that this implies $q_{ij} = \pi_j$ for all $i,j \in \{0,1,\ldots,N-1\}$. Hence $\pi_j = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} p_{ij}^{(k)}$.

Let

$$A = \{x : x_r = i_0, \ldots, x_{r+l} = i_l\}, \text{ and } B = \{x : x_s = j_0, \ldots, x_{s+m} = j_m\}$$

be any two cylinder sets of X. By Proposition 2.1.1 in Section 2, we must show that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}A\cap B) = \mu(A)\mu(B).$$

Since T is the left shift, for all n sufficiently large, the cylinders $T^{-n}A$ and B depend on different coordinates. Hence, for n sufficiently large,

$$\mu(T^{-n}A\cap B) = \pi_{j_0}p_{j_0j_1}\ldots p_{j_{m-1}j_m}p_{j_mi_0}^{(n+r-s-m)}p_{i_0i_1}\ldots p_{i_{l-1}i_l}.$$

Thus,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}A\cap B)\\
=\;& \pi_{j_0}p_{j_0j_1}\ldots p_{j_{m-1}j_m}p_{i_0i_1}\ldots p_{i_{l-1}i_l}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}p_{j_mi_0}^{(k)}\\
=\;& (\pi_{j_0}p_{j_0j_1}\ldots p_{j_{m-1}j_m})(\pi_{i_0}p_{i_0i_1}\ldots p_{i_{l-1}i_l})\\
=\;& \mu(B)\mu(A).
\end{aligned}$$

Therefore, T is ergodic.

(ii) $\Rightarrow$ (v) If all the rows of Q are identical, then $q_{ij} = \pi_j$ for all i,j. If $vP = v$, then $vQ = v$. This implies that for all j, $v_j = (\sum_{i=0}^{N-1} v_i)\pi_j$. Thus, v is a multiple of π. Therefore, 1 is a simple eigenvalue.

(v) $\Rightarrow$ (ii) Suppose 1 is a simple eigenvalue. For any i, let $q_i^* = (q_{i0},\ldots,q_{i(N-1)})$ denote the i^{th} row of Q then, q_i^* is a probability vector. From $Q = QP$, we get $q_i^* = q_i^*P$. By hypothesis π is the only probability vector satisfying $\pi P = P$, hence $\pi = q_i^*$, and all the rows of Q are identical. $\square$

2.3 Mixing

As a corollary to the ergodic theorem we found a new definition of ergodicity; namely, asymptotic average independence. Based on the same idea, we now define other notions of *weak* independence that are stronger than ergodicity.

Definition 2.3.1 *Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Then,*

(i) T is weakly mixing if for all $A, B \in \mathcal{F}$, one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(T^{-i}A \cap B) - \mu(A)\mu(B)\right| = 0. \tag{2.3}$$

(ii) T is strongly mixing if for all $A, B \in \mathcal{F}$, one has

$$\lim_{n\to\infty} \mu(T^{-i}A \cap B) = \mu(A)\mu(B). \tag{2.4}$$

Notice that strongly mixing implies weakly mixing, and weakly mixing implies ergodicity. This follows from the simple fact that if $\{a_n\}$ is a sequence of real numbers such that $\lim_{n\to\infty} a_n = 0$, then $\lim_{n\to\infty} \dfrac{1}{n} \displaystyle\sum_{i=0}^{n-1} |a_i| = 0$, and hence $\lim_{n\to\infty} \dfrac{1}{n} \displaystyle\sum_{i=0}^{n-1} a_i = 0$. Furthermore, if $\{a_n\}$ is a bounded sequence, then the following are equivalent:

(i) $\lim_{n\to\infty} \dfrac{1}{n} \displaystyle\sum_{i=0}^{n-1} |a_i| = 0$

(ii) $\lim_{n\to\infty} \dfrac{1}{n} \displaystyle\sum_{i=0}^{n-1} |a_i|^2 = 0$

(iii) there exists a subset J of the integers of density zero, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \# \left(\{0, 1, \ldots, n-1\} \cap J\right) = 0,$$

such that $\lim_{n\to\infty, n\notin J} a_n = 0$.

Using this one can give three equivalent characterizations of weakly mixing transformations, can you state them?

Exercise 2.3.1 Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Let $\mathcal{S}$ be a generating semi-algebra of $\mathcal{F}$.

(a) Show that if equation (2.3) holds for all $A, B \in \mathcal{S}$, then T is weakly mixing.

(b) Show that if equation (2.4) holds for all $A, B \in \mathcal{S}$, then T is strongly mixing.

Exercise 2.3.2 Consider the one or two-sided Bernoulli shift T as given in Example (e) in subsection 1.3, and Example (2) in subsection 1.8. Show that T is strongly mixing.

Exercise 2.3.3 Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ a measure preserving transformation. Consider the transformation $T \times T$ defined on $(X \times X, \mathcal{F} \times \mathcal{F}, \mu \times \mu)$ by $T \times T(x, y) = (Tx, Ty)$.

(a) Show that $T \times T$ is measure preserving with respect to $\mu \times \mu$.

(b) Show that $T \times T$ is ergodic, if and only if T *is* weakly mixing.

Chapter 3

Measure Preserving Isomorphisms and Factor Maps

3.1 Measure Preserving Isomorphisms

Given a measure preserving transformation T on a probability space $(X, \mathcal{F}, \mu)$, we call the quadruple $(X, \mathcal{F}, \mu, T)$ *a dynamical system* . Now, given two dynamical systems $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$, what should we mean by: *these systems are the same*? On each space there are two important structures:

(1) The measure structure given by the σ-algebra and the probability measure. Note, that in this context, sets of measure zero can be ignored.

(2) The dynamical structure, given by a measure preserving transformation.

So our notion of *being the same* must mean that we have a map

$$\psi : (X, \mathcal{F}, \mu, T) \to (Y, \mathcal{C}, \nu, S)$$

satisfying

(i) ψ is one-to-one and onto a.e. By this we mean, that if we remove a (suitable) set N_X of measure 0 in X, and a (suitable) set N_Y of measure 0 in Y, the map $\psi : X \setminus N_X \to Y \setminus N_Y$ is a bijection.

(ii) ψ is measurable, i.e., $\psi^{-1}(C) \in \mathcal{F}$, for all $C \in \mathcal{C}$.

(iii) ψ preserves the measures: $\nu = \mu \circ \psi^{-1}$, i.e., $\nu(C) = \mu\left(\psi^{-1}(C)\right)$ for all $C \in \mathcal{C}$.

Finally, we should have that

(iv) ψ preserves the dynamics of T and S, i.e., $\psi \circ T = S \circ \psi$, which is the same as saying that the following diagram commutes.

$$\begin{tikzcd} N \arrow[r, "T"] \arrow[d, "\psi"'] & N \arrow[d, "\psi"] \\ N' \arrow[r, "S"] & N' \end{tikzcd}$$

This means that T-orbits are mapped to S-orbits:

$$\begin{array}{ccccc} x \rightarrow & Tx \rightarrow & T^2x \rightarrow & \cdots \rightarrow & T^nx \rightarrow \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \psi(x) \rightarrow & S\left(\psi(x)\right) \rightarrow & S^2\left(\psi(x)\right) \rightarrow & \cdots \rightarrow & S^n\left(\psi(x)\right) \rightarrow \end{array}$$

Definition 3.1.1 *Two dynamical systems* $(X, \mathcal{F}, \mu, T)$ *and* $(Y, \mathcal{C}, \nu, S)$ *are isomorphic if there exist measurable sets* $N \subset X$ *and* $M \subset Y$ *with* $\mu(X \backslash N) = \nu(Y \setminus M) = 0$ *and* $T(N) \subset N$, $S(M) \subset M$, *and finally if there exists a measurable map* $\psi : N \rightarrow M$ *such that* (i)–(iv) *are satisfied.*

Exercise 3.1.1 Suppose $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ are two isomorphic dynamical systems. Show that

(a) T is ergodic if and only if S is ergodic.

(b) T is weakly mixing if and only if S is weakly mixing.

(c) T is strongly mixing if and only if S is strongly mixing.

Examples

(1) Let $K = \{z \in \mathbb{C} : |z| = 1\}$ be equipped with the Borel σ-algebra $\mathcal{B}$ on K, and Haar measure (i.e., normalized Lebeque measure on the unit circle).

Define $S : K \to K$ by $Sz = z^2$; equivalently $Se^{2\pi i\theta} = e^{2\pi i(2\theta)}$.. One can easily check that S is measure preserving. In fact, the map S is isomorphic to the map T on $([0,1), \mathcal{B}, \lambda)$ given by $Tx = 2x \pmod 1$ (see Example (b) in subsection 1.3, and Example (3) in subsection 1.8). Define a map $\phi : [0,1) \to K$ by $\phi(x) = e^{2\pi i x}$. We leave it to the reader to check that ϕ is a measurable isomorphism, i.e., ϕ is a measurable and measure preserving bijection such that $S\phi(x) = \phi(Tx)$ for all $x \in [0,1)$.

(2) Consider $([0,1), \mathcal{B}, \lambda)$, the unit interval with the Lebesgue σ-algebra, and Lebesgue measure. Let $T : [0,1) \to [0,1)$ be given by $Tx = Nx - \lfloor Nx \rfloor$. Iterations of T generate the N-adic expansion of points in the unit interval. Let $Y := \{0, 1, \ldots, N-1\}^{\mathbb{N}}$, the set of all sequences $(y_n)_{n\geq 1}$, with $y_n \in \{0, 1, \ldots, N-1\}$ for $n \geq 1$. We now construct an isomorphism between $([0,1), \mathcal{B}, \lambda, T)$ and $(Y, \mathcal{F}, \mu, S)$, where $\mathcal{F}$ is the σ-algebra generated by the cylinders, and μ the uniform product measure defined on cylinders by

$$\mu(\{(y_i)_{i\geq 1} \in Y : \, y_1 = a_1, y_2 = a_2, \ldots, y_n = a_n\}) \;=\; \frac{1}{N^n} \,,$$

for any $(a_1, a_2, a_3, \ldots) \in Y$, and where S is the left shift.

Define $\psi : [0,1) \to Y = \{0, 1, \ldots, N-1\}^{\mathbb{N}}$ by

$$\psi : \, x \;=\; \sum_{k=1}^{\infty} \frac{a_k}{N^k} \;\mapsto\; (a_k)_{k\geq 1} \,,$$

where $\sum_{k=1}^{\infty} a_k/N^k$ is the N-adic expansion of x (for example if $N = 2$ we get the binary expansion, and if $N = 10$ we get the decimal expansion). Let

$$C(i_1, \ldots, i_n) \;=\; \{(y_i)_{i\geq 1} \in Y : \, y_1 = i_1, \ldots, y_n = i_n\} \;.$$

In order to see that ψ is an isomorphism one needs to verify measurability and measure preservingness on cylinders:

$$\psi^{-1}\left(C(i_1, \ldots, i_n)\right) \;=\; \Big[\frac{i_1}{N} + \frac{i_2}{N^2} + \cdots + \frac{i_n}{N^n}, \frac{i_1}{N} + \frac{i_2}{N^2} + \cdots + \frac{i_n + 1}{N^n}\Big)$$

and

$$\lambda\left(\psi^{-1}(C(i_1, \ldots, i_n))\right) \;=\; \frac{1}{N^n} \;=\; \mu\left(C(i_1, \ldots, i_n)\right)) \;.$$

Note that

$$\mathcal{N} \;=\; \{(y_i)_{i\geq 1} \in Y : \; \text{there exists a } k \geq 1 \text{ such that } y_i = N-1 \text{ for all } i \geq k\}$$

is a subset of Y of measure 0. Setting $\tilde{Y} = Y \setminus \mathcal{N}$, then $\psi : [0,1) \to \tilde{Y}$ is a bijection, since every $x \in [0,1)$ has a unique N-adic expansion **generated** by T. Finally, it is easy to see that $\psi \circ T = S \circ \psi$.

Exercise 3.1.2 Consider $([0,1)^2, \mathcal{B} \times \mathcal{B}, \lambda \times \lambda)$, where $\mathcal{B} \times \mathcal{B}$ is the product Lebesgue σ-algebra, and $\lambda \times \lambda$ is the product Lebesgue measure Let $T : [0,1)^2 \to [0,1)^2$ be given by

$$T(x,y) = \begin{cases} (2x, \dfrac{1}{2}y), & 0 \le x < \frac{1}{2} \\ \\ (2x-1, \dfrac{1}{2}(y+1)), & \frac{1}{2} \le x < 1. \end{cases}$$

Show that T is isomorphic to the two-sided Bernoulli shift S on $\big(\{0,1\}^{\mathbb{Z}}, \mathcal{F}, \mu\big)$, where $\mathcal{F}$ is the σ-algebra generated by cylinders of the form

$$\Delta = \{x_{-k} = a_{-k}, \ldots, x_\ell = a_\ell \,:\, a_i \in \{0,1\},\ i = -k, \ldots, \ell\}, \quad k, \ell \ge 0,$$

and μ the product measure with weights $(\frac{1}{2}, \frac{1}{2})$ (so $\mu(\Delta) = (\frac{1}{2})^{k+\ell+1}$).

Exercise 3.1.3 Let $G = \dfrac{1+\sqrt{5}}{2}$, so that $G^2 = G + 1$. Consider the set

$$X = [0, \frac{1}{G}) \times [0,1) \bigcup [\frac{1}{G}, 1) \times [0, \frac{1}{G}),$$

endowed with the product Borel σ-algebra. Define the transformation

$$\mathcal{T}(x,y) = \begin{cases} (Gx, \dfrac{y}{G}), & (x,y) \in [0, \frac{1}{G}) \times [0,1] \\ \\ (Gx-1, \dfrac{1+y}{G}), & (x,y) \in [\frac{1}{G}, 1) \times [0, \frac{1}{G}). \end{cases}$$

(a) Show that $\mathcal{T}$ is measure preserving with respect to normalized Lebesgue measure on X.

(b) Now let $\mathcal{S} : [0,1) \times [0,1) \to [0,1) \times [0,1)$ be given by

$$\mathcal{S}(x,y) = \begin{cases} (Gx, \dfrac{y}{G}), & (x,y) \in [0, \frac{1}{G}) \times [0,1] \\ \\ (G^2x - G, \dfrac{G+y}{G^2}), & (x,y) \in [\frac{1}{G}, 1) \times [0,1). \end{cases}$$

Show that $\mathcal{S}$ is measure preserving with respect to normalized Lebesgue measure on $[0,1) \times [0,1)$.

(c) Let $Y = [0,1) \times [0, \frac{1}{G})$, and let U be the induced transformation of $\mathcal{T}$ on Y, i.e., for $(x,y) \in Y$, $U(x,y) = \mathcal{T}^{n(x,y)}$, where $n(x,y) = \inf\{n \geq 1 : \mathcal{T}^n(x,y) \in Y\}$. Show that the map $\phi : [0,1) \times [0,1) \to Y$ given by

$$\phi(x,y) = (x, \frac{y}{G})$$

defines an isomorphism from $\mathcal{S}$ to U, where Y has the induced measure structure (see Section 1.5).

3.2 Factor Maps

In the above section, we discussed the notion of isomorphism which describes when two dynamical systems are considered the same. Now, we give a precise definition of what it means for a dynamical system to be a subsystem of another one.

Definition 3.2.1 *Let $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ be two dynamical systems. We say that S is a factor of T if there exist measurable sets $M_1 \in \mathcal{F}$ and $M_2 \in \mathcal{C}$, such that $\mu(M_1) = \nu(M_2) = 1$ and $T(M_1) \subset M_1$, $S(M_2) \subset M_2$, and finally if there exists a measurable and measure preserving map $\psi : M_1 \to M_2$ which is surjective, and satisfies $\psi(T(x)) = S(\psi(x))$ for all $x \in M_1$. We call ψ a factor map.*

Remark Notice that if ψ is a factor map, then $\mathcal{G} = \psi^{-1}\mathcal{C}$ is a T-invariant sub-σ-algebra of $\mathcal{F}$, since

$$T^{-1}\mathcal{G} = T^{-1}\psi^{-1}\mathcal{C} = \psi^{-1}S^{-1}\mathcal{C} \subseteq \psi^{-1}\mathcal{C} = \mathcal{G}.$$

Examples Let T be the Baker's transformation on $([0,1)^2, \mathcal{B} \times \mathcal{B}, \lambda \times \lambda)$, given by

$$T(x,y) = \begin{cases} (2x, \frac{1}{2}y), & 0 \leq x < \frac{1}{2} \\ \\ (2x-1, \frac{1}{2}(y+1)), & \frac{1}{2} \leq x < 1, \end{cases}$$

and let S be the left shift on $X = \{0,1\}^{\mathbb{N}}$ with the σ-algebra $\mathcal{F}$ generated by the cylinders, and the uniform product measure μ. Define $\psi : [0,1)\times[0,1) \to X$ by

$$\psi(x,y) = (a_1, a_2, \ldots),$$

where $x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}$ is the binary expansion of x. It is easy to check that ψ is a factor map.

Exercise 3.2.1 Let T be the left shift on $X = \{0,1,2\}^{\mathbb{N}}$ which is endowed with the σ-algebra $\mathcal{F}$, generated by the cylinder sets, and the uniform product measure μ giving each symbol probability 1/3, i.e.,

$$\mu\left(\{x \in X : x_1 = i_1, x_2 = i_2, \ldots, x_n = i_n\}\right) = \frac{1}{3^n},$$

where $i_1, i_2, \ldots, i_n \in \{0,1,2\}$.
Let S be the left shift on $Y = \{0,1\}^{\mathbb{N}}$ which is endowed with the σ-algebra $\mathcal{G}$, generated by the cylinder sets, and the product measure ν giving the symbol 0 probability 1/3 and the symbol 1 probability 2/3, i.e.,

$$\mu\left(\{y \in Y : y_1 = j_1, y_2 = j_2, \ldots, y_n = j_n\}\right) = (\frac{2}{3})^{j_1+j_2+\ldots+j_n}(\frac{1}{3})^{n-(j_1+j_2+\ldots+j_n)},$$

where $j_1, j_2, \ldots, j_n \in \{0,1\}$. Show that S is a factor of T.

Exercise 3.2.2 Show that a factor of an ergodic (weakly mixing/strongly mixing) transformation is also ergodic (weakly mixing/strongly mixing).

3.3 Natural Extensions

Suppose $(Y, \mathcal{G}, \nu, S)$ is a *non-invertible* measure-preserving dynamical system. An invertible measure-preserving dynamical system $(X, \mathcal{F}, \mu, T)$ is called a *natural extension* of $(Y, \mathcal{G}, \nu, S)$ if S is a factor of T and the factor map ψ satisfies $\vee_{m=0}^{\infty} T^m \psi^{-1}\mathcal{G} = \mathcal{F}$, where

$$\bigvee_{k=0}^{\infty} T^k \psi^{-1}\mathcal{G}$$

is the smallest σ-algebra containing the σ-algebras $T^k\psi^{-1}\mathcal{G}$ for all $k \geq 0$.

Example Let T on $\left(\{0,1\}^{\mathbb{Z}}, \mathcal{F}, \mu\right)$ be the two-sided Bernoulli shift, and S on $\left(\{0,1\}^{\mathbb{N}\cup\{0\}}, \mathcal{G}, \nu\right)$ be the one-sided Bernoulli shift, both spaces are endowed with the uniform product measure. Notice that T is invertible, while S is not. Set $X = \{0,1\}^{\mathbb{Z}}$, $Y = \{0,1\}^{\mathbb{N}\cup\{0\}}$, and define $\psi : X \to Y$ by

$$\psi\left(\ldots, x_{-1}, x_0, x_1, \ldots\right) = \left(x_0, x_1, \ldots\right).$$

Then, ψ is a factor map. We claim that

$$\bigvee_{k=0}^{\infty} T^k \psi^{-1} \mathcal{G} = \mathcal{F}.$$

To prove this, we show that $\bigvee_{k=0}^{\infty} T^k \psi^{-1} \mathcal{G}$ contains all cylinders generating $\mathcal{F}$.

Let $\Delta = \{x \in X : x_{-k} = a_{-k}, \ldots, x_\ell = a_\ell\}$ be an arbitrary cylinder in $\mathcal{F}$, and let $D = \{y \in Y : y_0 = a_{-k}, \ldots, y_{k+\ell} = a_\ell\}$ which is a cylinder in $\mathcal{G}$. Then,

$$\psi^{-1} D = \{x \in X : x_0 = a_{-k}, \ldots, x_{k+\ell} = a_\ell\} \quad \text{and} \quad T^k \psi^{-1} D = \Delta.$$

This shows that

$$\bigvee_{k=0}^{\infty} T^k \psi^{-1} \mathcal{G} = \mathcal{F}.$$

Thus, T is the natural extension of S.

Chapter 4

Entropy

4.1 Randomness and Information

Given a measure preserving transformation T on a probability space $(X, \mathcal{F}, \mu)$, we want to define a nonnegative quantity $h(T)$ which measures the average uncertainty about where T moves the points of X. That is, the value of $h(T)$ reflects the amount of 'randomness' generated by T. We want to define $h(T)$ in such a way, that (i) the amount of information gained by an application of T is proportional to the amount of uncertainty removed, and (ii) that $h(T)$ is isomorphism invariant, so that isomorphic transformations have equal entropy.

The connection between entropy (that is randomness, uncertainty) and the transmission of information was first studied by Claude Shannon in 1948. As a motivation let us look at the following simple example. Consider a source (for example a ticker-tape) that produces a string of symbols $\cdots x_{-1}x_0x_1\cdots$ from the alphabet $\{a_1, a_2, \ldots, a_n\}$. Suppose that the probability of receiving symbol a_i at any given time is p_i, and that each symbol is transmitted independently of what has been transmitted earlier. Of course we must have here that each $p_i \geq 0$ and that $\sum_i p_i = 1$. In ergodic theory we view this process as the dynamical system $(X, \mathcal{F}, \mathcal{B}, \mu, T)$, where $X = \{a_1, a_2, \ldots, a_n\}^{\mathbb{N}}$, $\mathcal{B}$ the σ-algebra generated by cylinder sets of the form

$$\Delta_n(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) := \{x \in X : x_{i_1} = a_{i_1}, \ldots, x_{i_n} = a_{i_n}\}$$

μ the product measure assigning to each coordinate probability p_i of seeing the symbol a_i, and T the left shift. We define the entropy of this system by

$$H(p_1, \ldots, p_n) \;=\; h(T) \;:=\; -\sum_{i=1}^{n} p_i \log_2 p_i \,. \tag{4.1}$$

If we define $\log p_i$ as the amount of uncertainty in transmitting the symbol a_i, then H is the average amount of information gained (or uncertainty removed) per symbol (notice that H is in fact an expected value). To see why this is an appropriate definition, notice that if the source is degenerate, that is, $p_i = 1$ for some i (i.e., the source only transmits the symbol a_i), then $H = 0$. In this case we indeed have no randomness. Another reason to see why this definition is appropriate, is that H is maximal if $p_i = \frac{1}{n}$ for all i, and this agrees with the fact that the source is most random when all the symbols are equiprobable. To see this maximum, consider the function $f : [0,1] \to \mathbb{R}_+$ defined by

$$f(t) = \begin{cases} 0 & \text{if } t = 0, \\ -t \log_2 t & \text{if } 0 < t \le 1. \end{cases}$$

Then f is continuous and concave downward, and Jensen's Inequality implies that for any $p_1, \ldots, p_n$ with $p_i \ge 0$ and $p_1 + \ldots + p_n = 1$,

$$\frac{1}{n} H(p_1, \ldots, p_n) \;=\; \frac{1}{n} \sum_{i=1}^{n} f(p_i) \;\le\; f\big(\frac{1}{n} \sum_{i=1}^{n} p_i\big) \;=\; f\big(\frac{1}{n}\big) \;=\; \frac{1}{n} \log_2 n \,,$$

so $H(p_1, \ldots, p_n) \le \log_2 n$ for all probability vectors $(p_1, \ldots, p_n)$. But

$$H\big(\frac{1}{n}, \ldots, \frac{1}{n}\big) \;=\; \log_2 n,$$

so the maximum value is attained at $(\frac{1}{n}, \ldots, \frac{1}{n})$.

4.2 Definitions and Properties

So far H is defined as the average information per symbol. The above definition can be extended to define the information transmitted by the occurrence of an event E as $-\log_2 P(E)$. This definition has the property that the information transmitted by $E \cap F$ for independent events E and F is the sum of the information transmitted by each one individually, i.e.,

$$-\log_2 P(E \cap F) \;=\; -\log_2 P(E) - \log_2 P(F).$$

The only function with this property is the logarithm function to any base. We choose base 2 because information is usually measured in bits.

In the above example of the ticker-tape the symbols were transmitted independently. In general, the symbol generated might depend on what has been received before. In fact these dependencies are often 'built-in' to be able to check the transmitted sequence of symbols on errors (think here of the Morse sequence, sequences on compact discs etc.). Such dependencies must be taken into consideration in the calculation of the average information per symbol. This can be achieved if one replaces the symbols a_i by blocks of symbols of particular size. More precisely, for every n, let $\mathcal{C}_n$ be the collection of all possible n-blocks (or cylinder sets) of length n, and define

$$H_n := -\sum_{C\in\mathcal{C}_n} P(C)\log P(C)\,.$$

Then $\frac{1}{n}H_n$ can be seen as the average information per symbol when a block of length n is transmitted. The entropy of the source is now defined by

$$h := \lim_{n\to\infty}\frac{H_n}{n}\,. \tag{4.2}$$

The existence of the limit in (4.2) follows from the fact that H_n is a *subadditive sequence*, i.e., $H_{n+m} \leq H_n + H_m$, and proposition (4.2.2) (see proposition (4.2.3) below).

Now replace the source by a measure preserving system $(X,\mathcal{B},\mu,T)$. How can one define the entropy of this system similar to the case of a source? The symbols $\{a_1,a_2,\ldots,a_n\}$ can now be viewed as a partition $\mathcal{A} = \{A_1,A_2,\ldots,A_n\}$ of X, so that X is the disjoint union (up to sets of measure zero) of $A_1,A_2,\ldots,A_n$. The source can be seen as follows: with each point $x\in X$, we associate an infinite sequence $\cdots x_{-1},x_0,x_1,\cdots$, where x_i is a_j if and only if $T^ix\in A_j$. We define the *entropy of the partition* α by

$$H(\alpha) = H_\mu(\alpha) := -\sum_{i=1}^{n}\mu(A_i)\log\mu(A_i)\,.$$

Our aim is to define the entropy of the transformation T which is independent of the partition we choose. In fact $h(T)$ must be the maximal entropy over all possible finite partitions. But first we need few facts about partitions.

Exercise 4.2.1 Let $\alpha = \{A_1, \ldots, A_n\}$ and $\beta = \{B_1, \ldots, B_m\}$ be two partitions of X. Show that

$$T^{-1}\alpha := \{T^{-1}A_1, \ldots, T^{-1}A_n\}$$

and

$$\alpha \vee \beta := \{A_i \cap B_j : A_i \in \alpha,\, B_j \in \beta\}$$

are both partitions of X.

The members of a partition are called the *atoms* of the partition. We say that the partition $\beta = \{B_1, \ldots, B_m\}$ is a *refinement* of the partition $\alpha = \{A_1, \ldots, A_n\}$, and write $\alpha \leq \beta$, if for every $1 \leq j \leq m$ there exists an $1 \leq i \leq n$ such that $B_j \subset A_i$ (up to sets of measure zero). The partition $\alpha \vee \beta$ is called the *common refinement* of α and β.

Exercise 4.2.2 Show that if β is a refinement of α, each atom of α is a finite (disjoint) union of atoms of β.

Given two partitions $\alpha = \{A_1, \ldots A_n\}$ and $\beta = \{B_1, \ldots, B_m\}$ of X, we define the *conditional entropy of α given β* by

$$H(\alpha|\beta) := -\sum_{A\in\alpha}\sum_{B\in\beta} \log\left(\frac{\mu(A\cap B)}{\mu(B)}\right)\mu(A\cap B)\,.$$

(Under the convention that $0\log 0 := 0$.)

The above quantity $H(\alpha|\beta)$ is interpreted as the average uncertainty about which element of the partition α the point x will enter (under T) if we already know which element of β the point x will enter.

Proposition 4.2.1 *Let α, β and γ be partitions of X. Then,*

(a) $H(T^{-1}\alpha) = H(\alpha)$ *;*

(b) $H(\alpha \vee \beta) = H(\alpha) + H(\beta|\alpha)$*;*

(c) $H(\beta|\alpha) \leq H(\beta)$*;*

(d) $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$*;*

(e) If $\alpha \leq \beta$*, then* $H(\alpha) \leq H(\beta)$*;*

(f) $H(\alpha \vee \beta | \gamma) = H(\alpha|\gamma) + H(\beta|\alpha \vee \gamma)$;

(g) If $\beta \leq \alpha$, *then* $H(\gamma|\alpha) \leq H(\gamma|\beta)$;

(h) If $\beta \leq \alpha$, *then* $H(\beta|\alpha) = 0$.

(i) We call two partitions α *and* β independent *if*

$$\mu(A \cap B) = \mu(A)\mu(B) \quad \text{for all} \quad A \in \alpha,\ B \in \beta\,.$$

If α *and* β *are independent partitions, one has that*

$$H(\alpha \vee \beta) = H(\alpha) + H(\beta)\,.$$

Proof We prove properties (b) and (c), the rest are left as an exercise.

$$\begin{aligned}
H(\alpha \vee \beta) &= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log \mu(A\cap B) \\
&= -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log \frac{\mu(A\cap B)}{\mu(A)} \\
&+ -\sum_{A\in\alpha}\sum_{B\in\beta} \mu(A\cap B)\log \mu(A) \\
&= H(\beta|\alpha) + H(\alpha).
\end{aligned}$$

We now show that $H(\beta|\alpha) \leq H(\beta)$. Recall that the function $f(t) = -t\log t$ for $0 < t \leq 1$ is concave down. Thus,

$$\begin{aligned}
H(\beta|\alpha) &= -\sum_{B\in\beta}\sum_{A\in\alpha} \mu(A\cap B)\log \frac{\mu(A\cap B)}{\mu(A)} \\
&= -\sum_{B\in\beta}\sum_{A\in\alpha} \mu(A)\frac{\mu(A\cap B)}{\mu(A)}\log \frac{\mu(A\cap B)}{\mu(A)} \\
&= \sum_{B\in\beta}\sum_{A\in\alpha} \mu(A) f\left(\frac{\mu(A\cap B)}{\mu(A)}\right) \\
&\leq \sum_{B\in\beta} f\left(\sum_{A\in\alpha} \mu(A)\frac{\mu(A\cap B)}{\mu(A)}\right) \\
&= \sum_{B\in\beta} f(\mu(B)) = H(\beta).
\end{aligned}$$

□

Exercise 4.2.3 Prove the rest of the properties of Proposition 4.2.1

Now consider the partition $\bigvee_{i=0}^{n-1} T^{-i}\alpha$, whose atoms are of the form $A_{i_0} \cap T^{-1}A_{i_1} \cap \ldots \cap T^{-(n-1)}A_{i_{n-1}}$, consisting of all points $x \in X$ with the property that $x \in A_{i_0}$, $Tx \in A_{i_1}$, ... , $T^{n-1}x \in A_{i_{n-1}}$.

Exercise 4.2.4 Show that if α is a finite partition of $(X, \mathcal{F}, \mu, T)$, then

$$H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = H(\alpha) + \sum_{j=1}^{n-1} H(\alpha | \bigvee_{i=1}^{j} T^{-i}\alpha).$$

To define the notion of the entropy of a transformation with respect to a partition, we need the following two propositions.

Proposition 4.2.2 *If $\{a_n\}$ is a subadditive sequence of real numbers i.e., $a_{n+p} \leq a_n + a_p$ for all n, p, then*

$$\lim_{n\to\infty} \frac{a_n}{n}$$

exists.

Proof Fix any $m > 0$. For any $n \geq 0$ one has $n = km + i$ for some i between $0 \leq i \leq m - 1$. By subadditivity it follows that

$$\frac{a_n}{n} = \frac{a_{km+i}}{km+i} \leq \frac{a_{km}}{km} + \frac{a_i}{km} \leq k\frac{a_m}{km} + \frac{a_i}{km}.$$

Note that if $n \to \infty$, $k \to \infty$ and so $\limsup_{n\to\infty} a_n/n \leq a_m/m$. Since m is arbitrary one has

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \inf \frac{a_m}{m} \leq \liminf_{n\to\infty} \frac{a_n}{n}.$$

Therefore $\lim_{n\to\infty} a_n/n$ exists, and equals $\inf a_n/n$. □

Proposition 4.2.3 *Let α be a finite partitions of $(X, \mathcal{B}, \mu, T)$, where T is a measure preserving transformation. Then, $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ exists.*

Proof Let $a_n = H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \geq 0$. Then, by Proposition 4.2.1, we have

$$\begin{aligned}
a_{n+p} &= H(\bigvee_{i=0}^{n+p-1} T^{-i}\alpha) \\
&\leq H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) + H(\bigvee_{i=n}^{n+p-1} T^{-i}\alpha) \\
&= a_n + H(\bigvee_{i=0}^{p-1} T^{-i}\alpha) \\
&= a_n + a_p.
\end{aligned}$$

Hence, by Proposition 4.2.2

$$\lim_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$$

exists. □

We are now in position to give the definition of the entropy of the transformation T.

Definition 4.2.1 *The entropy of the measure preserving transformation T with respect to the partition α is given by*

$$h(\alpha, T) = h_\mu(\alpha, T) := \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)\,,$$

where

$$H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = -\sum_{D\in\bigvee_{i=0}^{n-1} T^{-i}\alpha} \mu(D)\log(\mu(D))\,.$$

Finally, the entropy of the transformation T is given by

$$h(T) = h_\mu(T) := \sup_\alpha h(\alpha, T)\,.$$

The following theorem gives an equivalent definition of entropy..

Theorem 4.2.1 *The entropy of the measure preserving transformation T with respect to the partition α is also given by*

$$h(\alpha, T) = \lim_{n\to\infty} H(\alpha | \bigvee_{i=1}^{n-1} T^{-i}\alpha)\,.$$

Proof Notice that the sequence $\{H(\alpha|\bigvee_{i=1}^{n} T^{-i}\alpha)$ is bounded from below, and is non-increasing, hence $\lim_{n\to\infty} H(\alpha|\bigvee_{i=1}^{n} T^{-i}\alpha)$ exists. Furthermore,

$$\lim_{n\to\infty} H(\alpha|\bigvee_{i=1}^{n} T^{-i}\alpha) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} H(\alpha|\bigvee_{i=1}^{j} T^{-i}\alpha).$$

From exercise 4.2.4, we have

$$H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = H(\alpha) + \sum_{j=1}^{n-1} H(\alpha|\bigvee_{i=1}^{j} T^{-i}\alpha).$$

Now, dividing by n, and taking the limit as $n \to \infty$, one gets the desired result $\square$

Theorem 4.2.2 *Entropy is an isomorphism invariant.*

Proof Let $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ be two isomorphic measure preserving systems (see Definition 1.2.3, for a definition), with $\psi : X \to Y$ the corresponding isomorphism. We need to show that $h_\mu(T) = h_\nu(S)$.

Let $\beta = \{B_1, \ldots, B_n\}$ be any partition of Y, then $\psi^{-1}\beta = \{\psi^{-1}B_1, \ldots, \psi^{-1}B_n\}$ is a partition of X. Set $A_i = \psi^{-1}B_i$, for $1 \le i \le n$. Since $\psi : X \to Y$ is an isomorphism, we have that $\nu = \mu\psi^{-1}$ and $\psi T = S\psi$, so that for any $n \ge 0$ and $B_{i_0}, \ldots, B_{i_{n-1}} \in \beta$

$$\begin{aligned}
& \nu\left(B_{i_0} \cap S^{-1}B_{i_1} \cap \ldots \cap S^{-(n-1)}B_{i_{n-1}}\right) \\
= \;& \mu\left(\psi^{-1}B_{i_0} \cap \psi^{-1}S^{-1}B_{i_1} \cap \ldots \cap \psi^{-1}S^{-(n-1)}B_{i_{n-1}}\right) \\
= \;& \mu\left(\psi^{-1}B_{i_0} \cap T^{-1}\psi^{-1}B_{i_1} \cap \ldots \cap T^{-(n-1)}\psi^{-1}B_{i_{n-1}}\right) \\
= \;& \mu\left(A_{i_0} \cap T^{-1}A_{i_1} \cap \ldots \cap T^{-(n-1)}A_{i_{n-1}}\right)\,.
\end{aligned}$$

Setting

$$A(n) = A_{i_0} \cap \ldots \cap T^{-(n-1)}A_{i_{n-1}} \;\text{ and }\; B(n) = B_{i_0} \cap \ldots \cap S^{-(n-1)}B_{i_{n-1}}\,,$$

we thus find that

$$\begin{aligned}
h_\nu(S) &= \sup_\beta h_\nu(\beta, S) = \sup_\beta \lim_{n\to\infty} \frac{1}{n} H_\nu(\bigvee_{i=0}^{n-1} S^{-i}\beta) \\
&= \sup_\beta \lim_{n\to\infty} -\frac{1}{n} \sum_{B(n)\in\bigvee_{i=0}^{n-1} S^{-i}\beta} \nu(B(n)) \log \nu(B(n)) \\
&= \sup_{\psi^{-1}\beta} \lim_{n\to\infty} -\frac{1}{n} \sum_{A(n)\in\bigvee_{i=0}^{n-1} T^{-i}\psi^{-1}\beta} \mu(A(n)) \log \mu(A(n)) \\
&= \sup_{\psi^{-1}\beta} h_\mu(\psi^{-1}\beta, T) \\
&\le \sup_\alpha h_\mu(\alpha, T) = h_\mu(T)\,,
\end{aligned}$$

where in the last inequality the supremum is taken over all possible finite partitions α of X. Thus $h_\nu(S) \le h_\mu(T)$. The proof of $h_\mu(T) \le h_\nu(S)$ is done similarly. Therefore $h_\nu(S) = h_\mu(T)$, and the proof is complete. □

4.3 Calculation of Entropy and Examples

Calculating the entropy of a transformation directly from the definition does not seem feasible, for one needs to take the supremum over **all** finite partitions, which is practically impossible. However, the entropy of a partition is relatively easier to calculate if one has full information about the partition under consideration. So the question is whether it is possible to find a partition α of X where $h(\alpha, T) = h(T)$. Naturally, such a partition contains all the information 'transmitted' by T. To answer this question we need some notations and definitions.

For $\alpha = \{A_1, \ldots, A_N\}$ and all m, $n \ge 0$, let

$$\sigma\left(\bigvee_{i=n}^{m} T^{-i}\alpha\right) \text{ and } \sigma\left(\bigvee_{i=-m}^{-n} T^{-i}\alpha\right)$$

be the smallest σ-algebras containing the partitions $\bigvee_{i=n}^{m} T^{-i}\alpha$ and $\bigvee_{i=-m}^{-n} T^{-i}\alpha$ respectively. Furthermore, let $\sigma\left(\bigvee_{i=-\infty}^{-\infty} T^{-i}\alpha\right)$ be the smallest σ-algebra containing all the partitions $\bigvee_{i=n}^{m} T^{-i}\alpha$ and $\bigvee_{i=-m}^{-n} T^{-i}\alpha$ for all n and m. We call a partition α a *generator* with respect to T if $\sigma\left(\bigvee_{i=-\infty}^{\infty} T^{-i}\alpha\right) = \mathcal{F}$, where $\mathcal{F}$ is the σ-algebra on X. If T is non-invertible, then α is said to be a generator if $\sigma\left(\bigvee_{i=0}^{\infty} T^{-i}\alpha\right) = \mathcal{F}$. Naturally, this equality is modulo sets of measure zero. One has also the following characterization of generators, saying basically, that each measurable set in X can be approximated by a finite disjoint union of cylinder sets. See also [W] for more details and proofs.

Proposition 4.3.1 *The partition α is a generator of $\mathcal{F}$ if for each $A \in \mathcal{F}$ and for each $\varepsilon > 0$ there exists a finite disjoint union C of elements of $\{\alpha_n^m\}$, such that $\mu(A \triangle C) < \varepsilon$.*

We now state (without proofs) two famous theorems known as *Kolmogorov-Sinai's Theorem* and *Krieger's Generator Theorem.* For the proofs, we refer the interested reader to the book of Karl Petersen or Peter Walter.

Theorem 4.3.1 (Kolmogorov and Sinai, 1958) *If α is a generator with respect to T and $H(\alpha) < \infty$, then $h(T) = h(\alpha, T)$.*

Theorem 4.3.2 (Krieger, 1970) *If T is an ergodic measure preserving transformation with $h(T) < \infty$, then T has a finite generator.*

We will use these two theorems to calculate the entropy of a Bernoulli shift.

Example (Entropy of a Bernoulli shift)–Let T be the left shift on $X = \{1, 2, \cdots, n\}^{\mathbb{Z}}$ endowed with the σ-algebra $\mathcal{F}$ generated by the cylinder sets, and product measure μ giving symbol i probability p_i, where $p_1 + p_2 + \ldots + p_n = 1$. Our aim is to calculate $h(T)$. To this end we need to find a partition α which generates the σ-algebra $\mathcal{F}$ under the action of T. The natural choice of α is what is known as the *time-zero partition* $\alpha = \{A_1, \ldots, A_n\}$, where

$$A_i := \{x \in X : x_0 = i\}, \ i = 1, \ldots, n.$$

Notice that for all $m \in \mathbb{Z}$,

$$T^{-m}A_i = \{x \in X : x_m = i\},$$

and

$$A_{i_0} \cap T^{-1}A_{i_1} \cap \cdots \cap T^{-m}A_{i_m} = \{x \in X : x_0 = i_0, \ldots, x_m = i_m\}.$$

In other words, $\bigvee_{i=0}^{m} T^{-i}\alpha$ is precisely the collection of cylinders of length m (i.e., the collection of all m-blocks), and these by definition generate $\mathcal{F}$. Hence, α is a generating partition, so that

$$h(T) \;=\; h(\alpha, T) \;=\; \lim_{m\to\infty} \frac{1}{m} H\left(\bigvee_{i=0}^{m-1} T^{-i}\alpha\right) .$$

First notice that $-$ since μ is product measure here $-$ the partitions

$$\alpha,\, T^{-1}\alpha,\, \cdots,\, T^{-(m-1)}\alpha$$

are all independent since each specifies a different coordinate, and so

$$\begin{aligned} & H(\alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-(m-1)}\alpha) \\ = \;& H(\alpha) + H(T^{-1}\alpha) + \cdots + H(T^{-(m-1)}\alpha) \\ = \;& mH(\alpha) \;=\; -m \sum_{i=1}^{n} p_i \log p_i . \end{aligned}$$

Thus,

$$h(T) \;=\; \lim_{m\to\infty} \frac{1}{m}(-m) \sum_{i=1}^{n} p_i \log p_i \;=\; -\sum_{i=1}^{n} p_i \log p_i \,.$$

Exercise 4.3.1 Let T be the left shift on $X = \{1, 2, \cdots, n\}^{\mathbb{Z}}$ endowed with the σ-algebra $\mathcal{F}$ generated by the cylinder sets, and the Markov measure μ given by the stochastic matrix $P = (p_{ij})$, and the probability vector $\pi = (\pi_1, \ldots, \pi_n)$ with $\pi P = \pi$. Show that

$$h(T) = -\sum_{j=1}^{n}\sum_{i=1}^{n} \pi_i p_{ij} \log p_{ij}$$

Exercise 4.3.2 *Suppose* $(X_1, \mathcal{B}_1, \mu_1, T_1)$ *and* $(X_2, \mathcal{B}_2, \mu_2, T_2)$ *are two dynamical systems. Show that*

$$h_{\mu_1\times\mu_2}(T_1 \times T_2) = h_{\mu_1}(T_1) + h_{\mu_2}(T_2).$$

4.4 The Shannon-McMillan-Breiman Theorem

In the previous sections we have considered only finite partitions on X, however all the definitions and results hold if we were to consider countable partitions of finite entropy. Before we state and prove the Shannon-McMillan-Breiman Theorem, we need to introduce the information function associated with a partition.

Let $(X, \mathcal{F}, \mu)$ be a probability space, and $\alpha = \{A_1, A_2, \ldots\}$ be a finite or a countable partition of X into measurable sets. For each $x \in X$, let $\alpha(x)$ be the element of α to which x belongs. Then, the *information function* associated to α is defined to be

$$I_\alpha(x) = -\log \mu(\alpha(x)) = -\sum_{A \in \alpha} 1_A(x) \log \mu(A).$$

For two finite or countable partitions α and β of X, we define the *conditional information function* of α given β by

$$I_{\alpha|\beta}(x) = -\sum_{B \in \beta} \sum_{A \in \alpha} 1_{(A \cap B)}(x) \log \left(\frac{\mu(A \cap B)}{\mu(B)} \right).$$

We claim that

$$I_{\alpha|\beta}(x) = -\log E_\mu(1_{\alpha(x)}|\sigma(\beta)) = -\sum_{A \in \alpha} 1_A(x) \log E(1_A|\sigma(\beta)), \tag{4.3}$$

where $\sigma(\beta)$ is the σ-algebra generated by the finite or countable partition β, (see the remark following the proof of Theorem (2.1.1)). This follows from the fact (which is easy to prove using the definition of conditional expectations) that if β is finite or countable, then for any $f \in L^1(\mu)$, one has

$$E_\mu(f|\sigma(\beta)) = \sum_{B \in \beta} 1_B \frac{1}{\mu(B)} \int_B f \, d\mu.$$

Clearly, $H(\alpha|\beta) = \int_X I_{\alpha|\beta}(x) \, \mathrm{d}\mu(x)$.

Exercise 4.4.1 Let α and β be finite or countable partitions of X. Show that

$$I_{\alpha \bigvee \beta} = I_\alpha + I_{\beta|\alpha}.$$

Now suppose $T : X \to X$ is a measure preserving transformation on $(X, \mathcal{F}, \mu)$, and let $\alpha = \{A_1, A_2, \ldots\}$ be any countable partition. Then $T^{-1} = \{T^{-1}A_1, T^{-1}A_2, \ldots\}$ is also a countable partition. Since T is measure preserving one has,

$$I_{T^{-1}\alpha}(x) = -\sum_{A_i \in \alpha} 1_{T^{-1}A_i}(x) \log \mu(T^{-1}A_i) = -\sum_{A_i \in \alpha} 1_{A_i}(Tx) \log \mu(A_i) = I_\alpha(Tx).$$

Furthermore,

$$\lim_{n\to\infty} \frac{1}{n+1} H(\bigvee_{i=0}^{n} T^{-i}\alpha) = \lim_{n\to\infty} \frac{1}{n+1} \int_X I_{\bigvee_{i=0}^{n} T^{-i}\alpha}(x)\, \mathrm{d}\mu(x) = h(\alpha, T).$$

The Shannon-McMillan-Breiman theorem says if T is ergodic and if α has finite entropy, then in fact the integrand $\dfrac{1}{n+1} I_{\bigvee_{i=0}^{n} T^{-i}\alpha}(x)$ converges a.e. to $h(\alpha, T)$. Notice that the integrand can be written as

$$\frac{1}{n+1} I_{\bigvee_{i=0}^{n} T^{-i}\alpha}(x) = -\frac{1}{n+1} \log \mu \left((\bigvee_{i=0}^{n} T^{-i}\alpha)(x) \right),$$

where $(\bigvee_{i=0}^{n} T^{-i}\alpha)(x)$ is the element of $\bigvee_{i=0}^{n} T^{-i}\alpha$ containing x (often referred to as the α-cylinder of order n containing x). Before we proceed we need the following proposition.

Proposition 4.4.1 *Let $\alpha = \{A_1, A_2, \ldots\}$ be a countable partition with finite entropy. For each $n = 1, 2, 3, \ldots$, let $f_n(x) = I_{\alpha|\bigvee_{i=1}^{n} T^{-i}\alpha}(x)$, and let $f^* = \sup_{n\geq 1} f_n$. Then, for each $\lambda \geq 0$ and for each $A \in \alpha$,*

$$\mu\left(\{x \in A : f^*(x) > \lambda\}\right) \leq 2^{-\lambda}.$$

Furthermore, $f^* \in L^1(X, \mathcal{F}, \mu)$.

Proof Let $t \geq 0$ and $A \in \alpha$. For $n \geq 1$, let

$$f_n^A(x) = -\log E_\mu \left(1_A | \bigvee_{i=1}^{n} T^{-i}\alpha \right)(x),$$

and

$$B_n = \{x \in X : f_1^A(x) \leq t, \ldots, f_{n-1}^A(x) \leq t, f_n^A(x) > t\}.$$

Notice that for $x \in A$ one has $f_n(x) = f_n^A(x)$, and for $x \in B_n$ one has $E_\mu\left(1_A | \bigvee_{i=1}^n T^{-i}\alpha\right)(x) < 2^{-t}$. Since $B_n \in \sigma\left(\bigvee_{i=1}^n T^{-i}\alpha\right)$, then

$$\begin{aligned}
\mu(B_n \cap A) &= \int_{B_n} 1_A(x)\,\mathrm{d}\mu(x) \\
&= \int_{B_n} E_\mu\left(1_A | \bigvee_{i=1}^n T^{-i}\alpha\right)(x)\,\mathrm{d}\mu(x) \\
&\leq \int_{B_n} 2^{-t}\,\mathrm{d}\mu(x) = 2^{-t}\mu(B_n).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mu\left(\{x \in A : f^*(x) > t\}\right) &= \mu\left(\{x \in A : f_n(x) > t, \text{ for some } n\}\right) \\
&= \mu\left(\{x \in A : f_n^A(x) > t, \text{ for some } n\}\right) \\
&= \mu\left(\cup_{n=1}^\infty A \cap B_n\right) \\
&= \sum_{n=1}^\infty \mu\left(A \cap B_n\right) \\
&\leq 2^{-t}\sum_{n=1}^\infty \mu(B_n) \leq 2^{-t}.
\end{aligned}$$

We now show that $f^* \in L^1(X, \mathcal{F}, \mu)$. First notice that

$$\mu\left(\{x \in A : f^*(x) > t\}\right) \leq \mu(A),$$

hence,

$$\mu\left(\{x \in A : f^*(x) > t\}\right) \leq \min(\mu(A), 2^{-t}).$$

Using Fubini's Theorem, and the fact that $f^* \geq 0$ one has

$$\begin{aligned}
\int_X f^*(x)\,\mathrm{d}\mu(x) &= \int_0^\infty \mu\left(\{x \in X : f^*(x) > t\}\right)\,\mathrm{d}t \\
&= \int_0^\infty \sum_{A\in\alpha} \mu\left(\{x \in A : f^*(x) > t\}\right)\,\mathrm{d}t \\
&= \sum_{A\in\alpha} \int_0^\infty \mu\left(\{x \in A : f^*(x) > t\}\right)\,\mathrm{d}t \\
&\leq \sum_{A\in\alpha} \int_0^\infty \min(\mu(A), 2^{-t})\,\mathrm{d}t \\
&= \sum_{A\in\alpha} \int_0^{-\log\mu(A)} \mu(A)\,\mathrm{d}t + \sum_{A\in\alpha} \int_{-\log\mu(A)}^\infty 2^{-t}\,\mathrm{d}t \\
&= -\sum_{A\in\alpha} \mu(A)\log\mu(A) + \sum_{A\in\alpha} \frac{\mu(A)}{\log_e 2} \\
&= H_\mu(\alpha) + \frac{1}{\log_e 2} < \infty.
\end{aligned}$$

□

So far we have defined the notion of the conditional entropy $I_{\alpha|\beta}$ when α and β are countable partitions. We can generalize the definition to the case α is a countable partition and $\mathcal{G}$ is a σ-algebra as follows (see equation (4.3)),

$$I_{\alpha|\mathcal{G}}(x) = -\log E_\mu(1_{\alpha(x)}|\mathcal{G}).$$

If we denote by $\bigvee_{i=1}^\infty T^{-i}\alpha = \sigma(\cup_n \bigvee_{i=1}^n T^{-i}\alpha)$, then

$$I_{\alpha|\bigvee_{i=1}^\infty T^{-i}\alpha}(x) = \lim_{n\to\infty} I_{\alpha|\bigvee_{i=1}^n T^{-i}\alpha}(x). \tag{4.4}$$

Exercise 4.4.2 Give a proof of equation (4.4) using the following important theorem, known as the Martingale Convergence Theorem (and is stated to our setting)

Theorem 4.4.1 *(Martingale Convergence Theorem) Let $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \cdots$ be a sequence of increasing σalgebras, and let $\mathcal{C} = \sigma(\cup_n \mathcal{C}_n)$. If $f \in L^1(\mu)$, then*

$$E_\mu(f|\mathcal{C}) = \lim_{n\to\infty} E_\mu(f|\mathcal{C}_n)$$

μ a.e. and in $L^1(\mu)$.

Exercise 4.4.3 Show that if T is measure preserving on the probability space $(X, \mathcal{F}, \mu)$ and $f \in L^1(\mu)$, then

$$\lim_{n\to\infty} \frac{f(T^n x)}{n} = 0, \quad \mu \ a.e.$$

Theorem 4.4.2 *(The Shannon-McMillan-Breiman Theorem) Suppose T is an ergodic measure preserving transformation on a probability space $(X, \mathcal{F}, \mu)$, and let α be a countable partition with $H(\alpha) < \infty$. Then,*

$$\lim_{n\to\infty} \frac{1}{n+1} I_{\bigvee_{i=0}^{n} T^{-i}\alpha}(x) = h(\alpha, T) \ a.e.$$

Proof For each $n = 1, 2, 3, \ldots$, let $f_n(x) = I_{\alpha | \bigvee_{i=1}^{n} T^{-i}\alpha}(x)$. Then,

$$\begin{aligned}
I_{\bigvee_{i=0}^{n} T^{-i}\alpha}(x) &= I_{\bigvee_{i=1}^{n} T^{-i}\alpha}(x) + I_{\alpha | \bigvee_{i=1}^{n} T^{-i}\alpha}(x) \\
&= I_{\bigvee_{i=0}^{n-1} T^{-i}\alpha}(Tx) + f_n(x) \\
&= I_{\bigvee_{i=1}^{n-1} T^{-i}\alpha}(Tx) + I_{\alpha | \bigvee_{i=1}^{n-1} T^{-i}\alpha}(Tx) + f_n(x) \\
&= I_{\bigvee_{i=0}^{n-2} T^{-i}\alpha}(T^2 x) + f_{n-1}(Tx) + f_n(x) \\
&\ \ \vdots \\
&= I_{\alpha}(T^n x) + f_1(T^{n-1}x) + \ldots + f_{n-1}(Tx) + f_n(x).
\end{aligned}$$

Let $f(x) = I_{\alpha | \bigvee_{i=1}^{\infty} T^{-i}\alpha}(x) = \lim_{n\to\infty} f_n(x)$. Notice that $f \in L^1(X, \mathcal{F}, \mu)$ since $\int_X f(x)\, \mathrm{d}\mu(x) = h(\alpha, T)$. Now letting $f_0 = I_\alpha$, we have

$$\begin{aligned}
\frac{1}{n+1} I_{\bigvee_{i=0}^{n} T^{-i}\alpha}(x) &= \frac{1}{n+1} \sum_{k=0}^{n} f_{n-k}(T^k x) \\
&= \frac{1}{n+1} \sum_{k=0}^{n} f(T^k x) + \frac{1}{n+1} \sum_{k=0}^{n} (f_{n-k} - f)(T^k x).
\end{aligned}$$

By the ergodic theorem,

$$\lim_{n\to\infty} \sum_{k=0}^{n} f(T^k x) = \int_X f(x)\, \mathrm{d}\mu(x) = h(\alpha, T) \ a.e.$$

We now study the sequence $\{\frac{1}{n+1}\sum_{k=0}^{n}(f_{n-k}-f)(T^kx)\}$. Let

$$F_N = \sup_{k\geq N}|f_k - f|, \text{ and } f^* = \sup_{n\geq 1} f_n.$$

Notice that $0 \leq F_N \leq f^*+f$, hence $F_N \in L^1(X,\mathcal{F},\mu)$ and $\lim_{N\to\infty} F_N(x) = 0$ a.e. Also for any k, $|f_{n-k}-f| \leq f^*+f$, so that $|f_{n-k}-f| \in L^1(X,\mathcal{F},\mu)$ and $\lim n \to \infty |f_{n-k}-f| = 0$ a.e.
For any $N \leq n$,

$$\begin{aligned}
\frac{1}{n+1}\sum_{k=0}^{n}|f_{n-k}-f|(T^kx) &= \frac{1}{n+1}\sum_{k=0}^{n-N}|f_{n-k}-f|(T^kx) \\
&+ \frac{1}{n+1}\sum_{k=n-N+1}^{n}|f_{n-k}-f|(T^kx) \\
&\leq \frac{1}{n+1}\sum_{k=0}^{n-N}F_N(T^kx) \\
&+ \frac{1}{n+1}\sum_{k=0}^{N-1}|f_k-f|(T^{n-k}x).
\end{aligned}$$

If we take the limit as $n\to\infty$, then by exercise (4.4.3) the second term tends to 0 a.e., and by the ergodic theorem as well as the dominated convergence theorem, the first term tends to zero a.e. Hence,

$$\lim_{n\to\infty}\frac{1}{n+1}I_{\bigvee_{i=0}^{n}T^{-i}\alpha}(x) = h(\alpha,T) \ a.e.$$

□

The above theorem can be interpreted as providing an estimate of the size of the atoms of $\bigvee_{i=0}^{n}T^{-i}\alpha$. For n sufficiently large, a typical element $A \in \bigvee_{i=0}^{n}T^{-i}\alpha$ satisfies

$$-\frac{1}{n+1}\log\mu(A) \approx h(\alpha,T)$$

or

$$\mu(A_n) \approx 2^{-(n+1)h(\alpha,T)}.$$

Furthermore, if α is a generating partition (i.e. $\bigvee_{i=0}^{\infty}T^{-i}\alpha = \mathcal{F}$, then in the conclusion of Shannon-McMillan-Breiman Theorem one can replace $h(\alpha,T)$ by $h(T)$.

4.5 Lochs' Theorem

In 1964, G. Lochs compared the decimal and the continued fraction expansions. Let $x \in [0,1)$ be an irrational number, and suppose $x = .d_1 d_2 \cdots$ is the decimal expansion of x (which is generated by iterating the map $Sx = 10x \pmod 1$). Suppose further that

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}} = [0; a_1, a_2, \cdots] \tag{4.5}$$

is its regular continued fraction (RCF) expansion (generated by the map $Tx = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$). Let $y = .d_1 d_2 \cdots d_n$ be the rational number determined by the first n decimal digits of x, and let $z = y + 10^{-n}$. Then, $[y, z)$ is the decimal cylinder of order n containing x, which we also denote by $B_n(x)$. Now let

$$y = \cfrac{1}{b_1 + \cfrac{1}{b_2 + \ddots + \cfrac{1}{b_l}}}$$

and

$$z = \cfrac{1}{c_1 + \cfrac{1}{c_2 + \ddots + \cfrac{1}{c_k}}}$$

be the continued fraction expansion of y and z. Let

$$m\,(n, x) = \max \left\{ i \le \min\,(l, k) : \text{ for all } j \le i,\ b_j = c_j \right\}. \tag{4.6}$$

In other words, if $B_n(x)$ denotes the decimal cylinder consisting of all points y in $[0,1)$ such that the first n decimal digits of y agree with those of x, and if $C_j(x)$ denotes the continued fraction cylinder of order j containing x, i.e., $C_j(x)$ is the set of all points in $[0,1)$ such that the first j digits in their continued fraction expansion is the same as that of x, then $m(n,x)$ is the largest integer such that $B_n(x) \subset C_{m(n,x)}(x)$. Lochs proved the following theorem:

Theorem 4.5.1 *Let λ denote Lebesgue measure on $[0,1)$. Then for a.e. $x \in [0,1)$*

$$\lim_{n\to\infty} \frac{m(n,x)}{n} = \frac{6\log 2 \log 10}{\pi^2}.$$

In this section, we will prove a generalization of Lochs' theorem that allows one to compare any two known expansions of numbers. We show that Lochs' theorem is true for any two sequences of interval partitions on $[0,1)$ satisfying the conclusion of Shannon-McMillan-Breiman theorem. We begin with few definitions that will be used in the arguments to follow.

Definition 4.5.1 *By an interval partition we mean a finite or countable partition of $[0,1)$ into subintervals. If P is an interval partition and $x \in [0,1)$, we let $P(x)$ denote the interval of P containing x.*

Let $\mathcal{P} = \{P_n\}_{n=1}^{\infty}$ be a sequence of interval partitions. Let λ denote Lebesgue probability measure on $[0,1)$.

Definition 4.5.2 *Let $c \geq 0$. We say that $\mathcal{P}$ has entropy c a.e. with respect to λ if*

$$-\frac{\log \lambda\left(P_n(x)\right)}{n} \to c \text{ a.e.}$$

Note that we do not assume that each P_n is refined by P_{n+1}.

Suppose that $\mathcal{P} = \{P_n\}_{n=1}^{\infty}$ and $\mathcal{Q} = \{Q_n\}_{n=1}^{\infty}$ are sequences of interval partitions. For each $n \in \mathbb{N}$ and $x \in [0,1)$, define

$$m_{\mathcal{P},\mathcal{Q}}(n,x) = \sup\left\{m \mid P_n(x) \subset Q_m(x)\right\}.$$

Theorem 4.5.2 *Let $\mathcal{P} = \{P_n\}_{n=1}^{\infty}$ and $\mathcal{Q} = \{Q_n\}_{n=1}^{\infty}$ be sequences of interval partitions and λ Lebesgue probability measure on $[0,1)$. Suppose that for some constants $c > 0$ and $d > 0$, $\mathcal{P}$ has entropy c a.e with respect to λ and $\mathcal{Q}$ has entropy d a.e. with respect to λ. Then*

$$\frac{m_{\mathcal{P},\mathcal{Q}}(n,x)}{n} \to \frac{c}{d} \text{ a.e.}$$

Proof First we show that

$$\limsup_{n\to\infty} \frac{m_{\mathcal{P},\mathcal{Q}}(n,x)}{n} \leq \frac{c}{d} \text{ a.e.}$$

Fix $\varepsilon > 0$. Let $x \in [0,1)$ be a point at which the convergence conditions of the hypotheses are met. Fix $\eta > 0$ so that $\dfrac{c+\eta}{c - \frac{c}{d}\eta} < 1+\varepsilon$. Choose N so that for all $n \geq N$

$$\lambda\left(P_n\left(x\right)\right) > 2^{-n(c+\eta)}$$

and

$$\lambda\left(Q_n\left(x\right)\right) < 2^{-n(d-\eta)}.$$

Fix n so that $\min\left\{n, \frac{c}{d}n\right\} \geq N$, and let m' denote any integer greater than $(1+\varepsilon)\,\frac{c}{d}n$. By the choice of η,

$$\lambda\left(P_n\left(x\right)\right) > \lambda\left(Q_{m'}\left(x\right)\right)$$

so that $P_n(x)$ is not contained in $Q_{m'}(x)$. Therefore

$$m_{\mathcal{P},\mathcal{Q}}\left(n,x\right) \leq \left(1+\varepsilon\right)\frac{c}{d}n$$

and so

$$\limsup_{n\to\infty}\frac{m_{\mathcal{P},\mathcal{Q}}\left(n,x\right)}{n} \leq \left(1+\varepsilon\right)\frac{c}{d} \text{ a.e.}$$

Since $\varepsilon > 0$ was arbitrary, we have the desired result.

Now we show that

$$\liminf_{n\to\infty}\frac{m_{\mathcal{P},\mathcal{Q}}\left(n,x\right)}{n} \geq \frac{c}{d} \text{ a.e.}$$

Fix $\varepsilon \in (0,1)$. Choose $\eta > 0$ so that $\zeta =: \varepsilon c - \eta\left(1 + (1-\varepsilon)\,\frac{c}{d}\right) > 0$. For each $n \in \mathbb{N}$ let $\bar{m}(n) = \left\lfloor (1-\varepsilon)\,\frac{c}{d}n\right\rfloor$. For brevity, for each $n \in \mathbb{N}$ we call an element of P_n (respectively Q_n) $(n,\eta)-$good if

$$\lambda\left(P_n\left(x\right)\right) < 2^{-n(c-\eta)}$$

(respectively

$$\lambda\left(Q_n\left(x\right)\right) > 2^{-n(d+\eta)}.)$$

For each $n \in \mathbb{N}$, let

$$D_n\left(\eta\right) = \left\{x : \begin{array}{c} P_n\left(x\right) \text{ is } \left(n,\eta\right)-\text{good and } Q_{\bar{m}(n)}\left(x\right) \text{ is } \left(\bar{m}\left(n\right),\eta\right)- \text{ good} \\ \text{and } P_n\left(x\right) \nsubseteq Q_{\bar{m}(n)}\left(x\right) \end{array}\right\}.$$

If $x \in D_n(\eta)$, then $P_n(x)$ contains an endpoint of the $(\bar{m}(n), \eta)$ −good interval $Q_{\bar{m}(n)}(x)$. By the definition of $D_n(\eta)$ and $\bar{m}(n)$,

$$\frac{\lambda(P_n(x))}{\lambda\left(Q_{\bar{m}(n)}(x)\right)} < 2^{-n\zeta}.$$

Since no more than one atom of P_n can contain a particular endpoint of an atom of $Q_{\bar{m}(n)}$, we see that $\lambda(D_n(\eta)) < 2 \cdot 2^{-n\zeta}$ and so

$$\sum_{n=1}^{\infty} \lambda(D_n(\eta)) < \infty,$$

which implies that

$$\lambda\{x \mid x \in D_n(\eta) \text{ i.o.}\} = 0.$$

Since $\bar{m}(n)$ goes to infinity as n does, we have shown that for almost every $x \in [0,1)$, there exists $N \in \mathbb{N}$, so that for all $n \geq N$, $P_n(x)$ is (n, η) −good and $Q_{\bar{m}(n)}(x)$ is $(\bar{m}(n), \eta)$ −good and $x \notin D_n(\eta)$. In other words, for almost every $x \in [0,1)$, there exists $N \in \mathbb{N}$, so that for all $n \geq N$, $P_n(x)$ is (n, η) −good and $Q_{\bar{m}(n)}(x)$ is $(\bar{m}(n), \eta)$ −good and $P_n(x) \subset Q_{\bar{m}(n)}(x)$. Thus, for almost every $x \in [0,1)$, there exists $N \in \mathbb{N}$, so that for all $n \geq N$, $m_{\mathcal{P},\mathcal{Q}}(n,x) \geq \bar{m}(n)$, so that

$$\frac{m_{\mathcal{P},\mathcal{Q}}(n,x)}{n} \geq \lfloor (1-\varepsilon)\frac{c}{d} \rfloor.$$

This proves that

$$\liminf_{n\to\infty} \frac{m_{\mathcal{P},\mathcal{Q}}(n,x)}{n} \geq (1-\varepsilon)\frac{c}{d} \text{ a.e.}$$

Since $\varepsilon > 0$ was arbitrary, we have established the theorem. □

The above result allows us to compare any two well-known expansions of numbers. Since the *commonly used* expansions are usually performed for points in the unit interval, our underlying space will be $([0,1), \mathcal{B}, \lambda)$, where $\mathcal{B}$ is the Lebesgue σ-algebra, and λ the Lebesgue measure. The expansions we have in mind share the following two properties.

Definition 4.5.3 *A surjective map $T : [0,1) \to [0,1)$ is called a number theoretic fibered map (NTFM) if it satisfies the following conditions:*

(a) *there exists a finite or countable partition of intervals $\alpha = \{A_i;\, i \in D\}$ such that T restricted to each atom of α (cylinder set of order 0) is monotone, continuous and injective. Furthermore, α is a generating partition.*

(b) *T is ergodic with respect to Lebesgue measure λ, and there exists a T invariant probability measure μ equivalent to λ with bounded density. (Both $\dfrac{d\mu}{d\lambda}$ and $\dfrac{d\lambda}{d\mu}$ are bounded, and $\mu(A) = 0$ if and only if $\lambda(A) = 0$ for all Lebesgue sets A.).*

Let T be an NTFM with corresponding partition α, and T-invariant measure μ equivalent to λ. Let $L, M > 0$ be such that

$$L\lambda(A) \le \mu(A) < M\lambda(A)$$

for all Lebesgue sets A (property (b)). For $n \ge 1$, let $P_n = \bigvee_{i=0}^{n-1} T^{-i}\alpha$, then by property (a), P_n is an interval partition. If $H_\mu(\alpha) < \infty$, then Shannon-McMillan-Breiman Theorem gives

$$\lim_{n\to\infty} -\frac{\log \mu(P_n(x))}{n} = h_\mu(T) \text{ a.e. with respect to } \mu.$$

Exercise 4.5.1 Show that the conclusion of the Shannon-McMillan-Breiman Theorem holds if we replace μ by λ, i.e.

$$\lim_{n\to\infty} -\frac{\log \lambda(P_n(x))}{n} = h_\mu(T) \text{ a.e. with respect to } \lambda.$$

Iterations of T generate expansions of points $x \in [0,1)$ with digits in D. We refer to the resulting expansion as the *T-expansion of x*.

Almost all known expansions on $[0,1)$ are generated by a NTFM. Among them are the n-adic expansions ($Tx = nx \pmod 1$, where n is a positive integer), β expansions ($Tx = \beta x \pmod 1$, where $\beta > 1$ is a real number), continued fraction expansions ($Tx = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$), and many others (see the book Ergodic Theory of Numbers).

Exercise 4.5.2 Prove Theorem (4.5.1) using Theorem (4.5.2). Use the fact that the continued fraction map T is ergodic with respect to Gauss measure μ, given by

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

and has entropy equal to $h_\mu(T) = \dfrac{\pi^2}{6\log 2}$.

Exercise 4.5.3 Reformulate and prove Lochs' Theorem for any two NTFM maps S and T on $[0,1)$.

Chapter 5

Invariant Measures for Continuous Transformations

5.1 Existence

Suppose X is a compact metric space, and let $\mathcal{B}$ be the Borel σ-algebra i.e., the σ-algebra generated by the open sets. Let $M(X)$ be the collection of all Borel probability measures on X. There is natural embedding of the space X in $M(X)$ via the map $x \to \delta_x$, where δ_X is the Dirac measure concentrated at x ($\delta_x(A) = 1$ if $x \in A$, and is zero otherwise). Furthermore, $M(X)$ is a convex set, i.e., $p\mu + (1-p)\nu \in M(X)$ whenever $\mu, \nu \in M(X)$ and $0 \le p \le 1$. Theorem 5.1.2 below shows that a member of $M(X)$ is determined by how it integrates continuous functions. We denote by $C(X)$ the Banach space of all complex valued continuous functions on X under the supremum norm.

Theorem 5.1.1 *Every member of $M(X)$ is regular, i.e., for all $B \in \mathcal{B}$ and every $\epsilon > 0$ there exist an open set U_ϵ and a closed sed C_ϵ such that $C_\epsilon \subseteq B \subseteq U_\epsilon$ such that $\mu(U_\epsilon \setminus C_\epsilon) < \epsilon$.*

Idea of proof Call a set $B \in \mathcal{B}$ with the above property *a regular set*. Let $\mathcal{R} = \{B \in \mathcal{B} : B \text{ is regular }\}$. Show that $\mathcal{R}$ is a σ-algebra containing all the closed sets. $\square$

Corollary 5.1.1 *For any $B \in \mathcal{B}$, and any $\mu \in M(X)$,*

$$\mu(B) = \sup_{C \subseteq B : C \text{ closed}} \mu(C) = \inf_{B \subseteq U : U \text{ open}} \mu(U).$$

Theorem 5.1.2 *Let $\mu, m \in M(X)$. If*

$$\int_X f \, \mathrm{d}\mu = \int_X f \, \mathrm{d}m$$

for all $f \in C(X)$, then $\mu = m$.

Proof From the above corollary, it suffices to show that $\mu(C) = m(C)$ for all closed subsets C of X. Let $\epsilon > 0$, by regularity of the measure m there exists an open set U_ϵ such that $C \subseteq U_\epsilon$ and $m(U_\epsilon \setminus C) < \epsilon$. Define $f \in C(X)$ as follows

$$f(x) = \begin{cases} 0 & x \notin U_\epsilon \\ \frac{d(x, X \setminus U_\epsilon)}{d(x, X \setminus U_\epsilon) + d(x, C)} & x \in U_\epsilon. \end{cases}$$

Notice that $1_C \leq f \leq 1_{U_\epsilon}$, thus

$$\mu(C) \leq \int_X f \, \mathrm{d}\mu = \int_X f \, \mathrm{d}m \leq m(U_\epsilon) \leq m(C) + \epsilon.$$

Using a similar argument, one can show that $m(C) \leq \mu(C) + \epsilon$. Therefore, $\mu(C) = m(C)$ for all closed sets, and hence for all Borel sets. □

This allows us to define a metric structure on $M(X)$ as follows. A sequence $\{\mu_n\}$ in $M(X)$ converges to $\mu \in M(X)$ if and only if

$$\lim_{n \to \infty} \int_X f(x) \, \mathrm{d}\mu_n(x) = \int_X f(x) \, \mathrm{d}\mu(x)$$

for all $f \in C(X)$. We will show that under this notion of convergence the space $M(X)$ is compact, but first we need The Riesz Representation Representation Theorem.

Theorem 5.1.3 *(The Riesz Representation Theorem) Let X be a compact metric space and $J : C(X) \to \mathbb{C}$ a continuous linear map such that J is a positive operator and $J(1) = 1$. Then there exists a $\mu \in M(X)$ such that $J(f) = \int_X f(x) \, \mathrm{d}\mu(x)$.*

Theorem 5.1.4 *The space $M(X)$ is compact.*

Idea of proof Let $\{\mu_n\}$ be a sequence in $M(X)$. Choose a countable dense subset of $\{f_n\}$ of $C(X)$. The sequence $\{\int_X f_1\,\mathrm{d}\mu_n\}$ is a bounded sequence of complex numbers, hence has a convergent subsequence $\{\int_X f_1\,\mathrm{d}\mu_n^{(1)}\}$. Now, the sequence $\{\int_X f_2\,\mathrm{d}\mu_n^{(1)}\}$ is bounded, and hence has a convergent subsequence $\{\int_X f_2\,\mathrm{d}\mu_n^{(2)}\}$. Notice that $\{\int_X f_1\,\mathrm{d}\mu_n^{(2)}\}$ is also convergent. We continue in this manner, to get for each i a subsequence $\{\mu_n^{(i)}\}$ of $\{\mu_n\}$ such that for all $j \leq i$, $\{\mu_n^{(i)}\}$ is a subsequence of $\{\mu_n^{(j)}\}$ and $\{\int_X f_j\,\mathrm{d}\mu_n^{(i)}\}$ converges. Consider the diagonal sequence $\{\mu_n^{(n)}\}$, then $\{\int_X f_j\,\mathrm{d}\mu_n^{(n)}\}$ converges for all j, and hence $\{\int_X f\,\mathrm{d}\mu_n^{(n)}\}$ converges for all $f \in C(X)$. Now define $J : C(X) \to \mathbb{C}$ by $J(f) = \lim_{n\to\infty}\{\int_X f\,\mathrm{d}\mu_n^{(n)}\}$. Then, J is linear, continuous ($|J(f)| \leq \sup_{x\in X}|f(x)|$), positive and $J(1) = 1$. Thus, by Riesz Representation Theorem, there exists a $\mu \in M(X)$ such that $J(f) = \lim_{n\to\infty}\{\int_X f\,\mathrm{d}\mu_n^{(n)}\} = \int_X f\,\mathrm{d}\mu$. Therefore, $\lim_{n\to\infty}\mu_n^{(n)} = \mu$, and $M(X)$ is compact. $\square$

Let $T : X \to X$ be a continuous transformation. Since $\mathcal{B}$ is generated by the open sets, then T is measurable with respect to $\mathcal{B}$. Furthermore, T induces in a natural way, an operator $\overline{T} : M(X) \to M(X)$ given by

$$(\overline{T}\mu)(A) = \mu(T^{-1}A)$$

for all $A \in \mathcal{B}$. Then $\overline{T}^i\mu(A) = \mu(T^{-i}A)$. Using a standard argument, one can easily show that

$$\int_X f(x)\,\mathrm{d}(\overline{T}\mu)(x) = \int_X f(Tx)d\mu(x)$$

for all continuous functions f on X. Note that T is measure preserving with respect to $\mu \in M(X)$ if and only if $\overline{T}\mu = \mu$. Equivalently, μ is measure preserving if and only if

$$\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f(Tx)\,\mathrm{d}\mu(x)$$

for all continuous functions f on X. Let

$$M(X,T) = \{\mu \in M(X) : \overline{T}\mu = \mu\}$$

be the collection of all probability measures under which T is measure preserving.

Theorem 5.1.5 *Let $T : X \to X$ be continuous, and $\{\sigma_n\}$ a sequence in $M(X)$. Define a sequence $\{\mu_n\}$ in $M(X)$ by*

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} \overline{T}^i \sigma_n.$$

Then, any limit point μ of $\{\mu_n\}$ is a member of $M(X,T)$.

Proof We need to show that for any continuous function f on X, one has $\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f(Tx)\,\mathrm{d}\mu$. Since $M(X)$ is compact there exists a $\mu \in M(X)$ and a subsequence $\{\mu_{n_i}\}$ such that $\mu_{n_i} \to \mu$ in $M(X)$. Now for any f continuous, we have

$$\begin{aligned}
\left|\int_X f(Tx)\,\mathrm{d}\mu(x) - \int_X f(x)\,\mathrm{d}\mu(x)\right| &= \lim_{j\to\infty}\left|\int_X f(Tx)\,\mathrm{d}\mu_{n_j}(x) - \int_X f(x)\,\mathrm{d}\mu_{n_j}(x)\right| \\
&= \lim_{j\to\infty}\left|\frac{1}{n_j}\int_X \sum_{i=0}^{n_j-1}\left(f(T^{i+1}x) - f(T^i x)\right)\,\mathrm{d}\sigma_{n_j}(x)\right| \\
&= \lim_{j\to\infty}\left|\frac{1}{n_j}\int_X \left(f(T^{n_j}x) - f(x)\right)\,\mathrm{d}\sigma_{n_j}(x)\right| \\
&\le \lim_{j\to\infty}\frac{2\sup_{x\in X}|f(x)|}{n_j} = 0.
\end{aligned}$$

Therefore $\mu \in M(X,T)$. □

Theorem 5.1.6 *Let T be a continuous transformation on a compact metric space. Then,*

(i) $M(X,T)$ is a compact convex subset of $M(X)$.

(ii) $\mu \in M(X,T)$ is an extreme point (i.e. μ cannot be written in a non-trivial way as a convex combination of elements of $M(X,T)$) if and only if T is ergodic with respect to μ.

Proof (i) Clearly $M(X,T)$ is convex. Now let $\{\mu_n\}$ be a sequence in $M(X,T)$ converging to μ in $M(X)$. We need to show that $\mu \in M(X,T)$.

Since T is continuous, then for any continuous function f on X, the function $f \circ T$ is also continuous. Hence,

$$\begin{aligned}\int_X f(Tx)\,\mathrm{d}\mu(x) &= \lim_{n\to\infty}\int_X f(Tx)\,\mathrm{d}\mu_n(x)\\ &= \lim_{n\to\infty}\int_X f(x)\,\mathrm{d}\mu_n(x)\\ &= \int_X f(x)\,\mathrm{d}\mu(x).\end{aligned}$$

Therefore, T is measure preserving with respect to μ, and $\mu \in M(X,T)$.
(ii) Suppose T is ergodic with respect to μ, and assume that

$$\mu = p\mu_1 + (1-p)\mu_2$$

for some $\mu_1, \mu_2 \in M(X,T)$, and some $0 < p \le 1$. We will show that $\mu = \mu_1$. Notice that the measure μ_1 is absolutely continuous with respect to μ, and T is ergodic with respect to μ, hence by Theorem (2.1.2) we have $\mu_1 = \mu$.

Conversely, (we prove the contrapositive) suppose that T is not ergodic with respect to μ. Then there exists a measurable set E such that $T^{-1}E = E$, and $0 < \mu(E) < 1$. Define measures μ_1, μ_2 on X by

$$\mu_1(B) = \frac{\mu(B\cap E)}{\mu(E)} \text{ and } \mu_1(B) = \frac{\mu\left(B\cap(X\setminus E)\right)}{\mu(X\setminus E)}.$$

Since E and $X \setminus E$ are T-invariant sets, then $\mu_1, \mu_2 \in M(X,T)$, and $\mu_1 \neq \mu_2$ since $\mu_1(E) = 1$ while $\mu_2(E) = 0$. Furthermore, for any measurable set B,

$$\mu(B) = \mu(E)\mu_1(B) + (1-\mu(E))\mu_2(B),$$

i.e. μ_1 is a non-trivial convex combination of elements of $M(X,T)$. Thus, μ is not an extreme point of $M(X,T)$. $\square$

Since the Banach space $C(X)$ of all continuous functions on X (under the sup norm) is separable i.e. $C(X)$ has a countable dense subset, one gets the following strengthening of the Ergodic Theorem.

Theorem 5.1.7 *If $T : X \to X$ is continuous and $\mu \in M(X,T)$ is ergodic, then there exists a measurable set Y such that $\mu(Y) = 1$, and*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \int_X f(x)\,\mathrm{d}\mu(x)$$

for all $x \in Y$, and $f \in C(X)$.

Proof Choose a countable dense subset $\{f_k\}$ in $C(X)$. By the Ergodic Theorem, for each k there exists a subset X_k with $\mu(X_k) = 1$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_k(T^i x) = \int_X f_k(x)\,\mathrm{d}\mu(x)$$

for all $x \in X_k$. Let $Y = \bigcap_{k=1}^{\infty} X_k$, then $\mu(Y) = 1$, and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_k(T^i x) = \int_X f_k(x) d\mu(x)$$

for all k and all $x \in Y$. Now, let $f \in C(X)$, then there exists a subsequence $\{f_{k_j}\}$ converging to f in the supremum norm, and hence is uniformly convergent. For any $x \in Y$, using uniform convergence and the dominated convergence theorem, one gets

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) &= \lim_{n\to\infty} \lim_{j\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_{k_j}(T^i x) \\
&= \lim_{j\to\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f_{k_j}(T^i x) \\
&= \lim_{j\to\infty} \int_X f_{k_j}\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu.
\end{aligned}$$

□

Theorem 5.1.8 *Let $T : X \to X$ be continuous, and $\mu \in M(X,T)$. Then T is ergodic with respect to μ if and only if*

$$\frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x} \to \mu \;\; a.e.$$

Proof Suppose T is ergodic with respect to μ. Notice that for any $f \in C(X)$,

$$\int_X f(y)\,\mathrm{d}(\delta_{T^i x})(y) = f(T^i x),$$

Hence by theorem 5.1.7, there exists a measurable set Y with $\mu(Y) = 1$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X f(y)\,\mathrm{d}(\delta_{T^i x})(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int_X f(y)\,\mathrm{d}\mu(y)$$

for all $x \in Y$, and $f \in C(X)$. Thus, $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i x} \to \mu$ for all $x \in Y$.

Conversely, suppose $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i x} \to \mu$ for all $x \in Y$, where $\mu(Y) = 1$. Then for any $f \in C(X)$ and any $g \in L^1(X, \mathcal{B}, \mu)$ one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) g(x) = g(x) \int_X f(y)\,\mathrm{d}\mu(y).$$

By the dominated convergence theorem

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X f(T^i x) g(x)\,\mathrm{d}\mu(x) = \int_X g(x) d\mu(x) \int_X f(y)\,\mathrm{d}\mu(y).$$

Now, let $F, G \in L^2(X, \mathcal{B}, \mu)$. Then, $G \in L^1(X, \mathcal{B}, \mu)$ so that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X f(T^i x) G(x)\,\mathrm{d}\mu(x) = \int_X G(x) d\mu(x) \int_X f(y) d\mu(y)$$

for all $f \in C(X)$. Let $\epsilon > 0$, there exists $f \in C(X)$ such that $||F - f||_2 < \epsilon$ which implies that $|\int F\,\mathrm{d}\mu - \int f d\mu| < \epsilon$. Furthermore, there exists N so that for $n \geq N$ one has

$$\left| \int_X \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) G(x)\,\mathrm{d}\mu(x) - \int_X G d\mu \int_X f\,\mathrm{d}\mu \right| < \epsilon.$$

Thus, for $n \geq N$ one has

$$\begin{aligned}
& \left| \int_X \frac{1}{n} \sum_{i=0}^{n-1} F(T^i x) G(x) \, \mathrm{d}\mu(x) - \int_X G d\mu \int_X F \, \mathrm{d}\mu \right| \\
\leq \; & \int_X \frac{1}{n} \sum_{i=0}^{n-1} \left| F(T^i x) - f(T^i x) \right| |G(x)| \, \mathrm{d}\mu(x) \\
+ \; & \left| \int_X \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) G(x) d\mu(x) - \int_X G d\mu \int_X f d\mu \right| \\
+ \; & \left| \int_X f d\mu \int_X G d\mu - \int_X F \, \mathrm{d}\mu \int_X G \, \mathrm{d}\mu \right| \\
< \; & \epsilon ||G||_2 + \epsilon + \epsilon ||G||_2 .
\end{aligned}$$

Thus,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X F(T^i x) G(x) \, \mathrm{d}\mu(x) = \int_X G(x) \, \mathrm{d}\mu(x) \int_X F(y) \, \mathrm{d}\mu(y)$$

for all $F, G \in L^2(X, \mathcal{B}, \mu)$ and $x \in Y$. Taking F and G to be indicator functions, one gets that T is ergodic.

□

Exercise 5.1.1 Let X be a compact metric space and $T : X \to X$ be a continuous homeomorphism. Let $x \in X$ be periodic point of T of period n, i.e. $T^n x = x$ and $T^j x \neq x$ for $j < i$. Show that if $\mu \in M(X, T)$ is ergodic and $\mu(\{x\}) > 0$, then $\mu = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x}$.

5.2 Unique Ergodicity

A continuous transformation $T : X \to X$ on a compact metric space is *uniquely ergodic* if there is only one T-invariant probabilty measure μ on X. In this case, $M(X, T) = \{\mu\}$, and μ is necessarily ergodic, since μ is an extreme point of $M(X, T)$. Recall that if $\nu \in M(X, T)$ is ergodic, then there exists a measurable subset Y such that $\nu(Y) = 1$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int_X f(y) \, \mathrm{d}\nu(y)$$

for all $x \in Y$ and all $f \in C(X)$. When T is uniquely ergodic we will see that we have a much stronger result.

Theorem 5.2.1 *Let $T : X \to X$ be a continuous transformation on a compact metric space X. Then the following are equivalent:*

(i) For every $f \in C(X)$, the sequence $\{\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)\}$ converges uniformly to a constant.

(ii) For every $f \in C(X)$, the sequence $\{\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)\}$ converges pointwise to a constant.

(iii) There exists a $\mu \in M(X,T)$ such that for every $f \in C(X)$ and all $x \in X$.

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \int_X f(y)\, \mathrm{d}\mu(y).$$

(iv) T is uniquely ergodic.

Proof (i) $\Rightarrow$ (ii) immediate.
(ii) $\Rightarrow$ (iii) Define $L : C(X) \to \mathbb{C}$ by

$$L(f) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x).$$

By assumption $L(f)$ is independent of x, hence L is well defined. It is easy to see that L is linear, continuous ($|L(f)| \le \sup_{x\in X}|f(x)|$), positive and $L(1) = 1$. Thus, by Riesz Representation Theorem there exists a probability measure $\mu \in M(X)$ such that

$$L(f) = \int_X f(x)\, \mathrm{d}\mu(x)$$

for all $f \in C(x)$. But

$$L(f \circ T) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^{i+1} x) = L(f).$$

Hence,

$$\int_X f(Tx)\,\mathrm{d}\mu(x) = \int_X f(x)\,\mathrm{d}\mu(x).$$

Thus, $\mu \in M(X,T)$, and for every $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \int_X f(x)\,\mathrm{d}\mu(x)$$

for all $x \in X$.

(iii) $\Rightarrow$ (iv) Suppose $\mu \in M(X,T)$ is such that for every $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = \int_X f(x)\,\mathrm{d}\mu(x)$$

for all $x \in X$. Assume $\nu \in M(X,T)$, we will show that $\mu = \nu$. For any $f \in C(X)$, since the sequence $\{\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)\}$ converges pointwise to the constant function $\int_X f(x)\,\mathrm{d}\mu(x)$, and since each term of the sequence is bounded in absolute value by the constant $\sup_{x\in X}|f(x)|$, it follows by the Dominated Convergence Theorem that

$$\lim_{n\to\infty} \int_X \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)\,\mathrm{d}\nu(x) = \int_X \int_X f(x)\,\mathrm{d}\mu(x) d\nu(y) = \int_X f(x)\,\mathrm{d}\mu(x).$$

But for each n,

$$\int_X \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)\,\mathrm{d}\nu(x) = \int_X f(x)\,\mathrm{d}\nu(x).$$

Thus, $\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f(x)\,\mathrm{d}\nu(x)$, and $\mu = \nu$.

(iv) $\Rightarrow$ (i) The proof is done by contradiction. Assume $M(X,T) = \{\mu\}$ and suppose $g \in C(X)$ is such that the sequence $\{\frac{1}{n}\sum_{j=0}^{n-1} g \circ T^j\}$ does not converge uniformly on X. Then there exists $\epsilon > 0$ such that for each N there exists $n > N$ and there exists $x_n \in X$ such that

$$\Big|\frac{1}{n}\sum_{j=0}^{n-1} g(T^j x_n) - \int_X g\,\mathrm{d}\mu\Big| \geq \epsilon.$$

Let

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} \delta_{T^j x_n} = \frac{1}{n}\sum_{j=0}^{n-1} \overline{T}^j \delta_{x_n}.$$

Then,

$$|\int_X g d\mu_n - \int_X g\,\mathrm{d}\mu| \geq \epsilon.$$

Since $M(X)$ is compact. there exists a subsequence μ_{n_i} converging to $\nu \in M(X)$. Hence,

$$|\int_X g\,\mathrm{d}\nu - \int_X g\,\mathrm{d}\mu| \geq \epsilon.$$

By Theorem (5.1.5), $\nu \in M(X,T)$ and by unique ergodicity $\mu = \nu$, which is a contradiction. □

Example If T_θ is an irrational rotation, then T_θ is uniquely ergodic. This is a consequence of the above theorem and Weyl's Theorem on uniform distribution: for any Riemann integrable function f on $[0,1)$, and any $x \in [0,1)$, one has

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(x + i\theta - \lfloor x + i\theta \rfloor) = \int_X f(y)\,\mathrm{d}y.$$

As an application of this, let us consider the following question. Consider the sequence of *first digits*

$$\{1, 2, 4, 8, 1, 3, 6, 1, \ldots\}$$

obtained by writing the first decimal digit of each term in the sequence

$$\{2^n : n \geq 0\} = \{1, 2, 4, 8, 16, 32, 64, 128, \ldots\}.$$

For each $k = 1, 2, \ldots, 9$, let $p_k(n)$ be the number of times the digit k appears in the first n terms of the *first digit* sequence. The asymptotic relative frequency of the digit k is then $\lim_{n\to\infty} \dfrac{p_k(n)}{n}$. We want to identify this limit for each $k \in \{1, 2, \ldots 9\}$. To do this, let $\theta = \log_{10} 2$, then θ is irrational. For $k = 1, 2, \ldots, 9$, let $J_k = [\log_{10} k, \log_{10}(k+1))$. By unique ergodicity of T_θ, we have for each $k = 1, 2, \ldots, 9$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} 1_{J_k}(T_\theta^j(0)) = \lambda(J_k) = \log_{10}\left(\frac{k+1}{k}\right).$$

Returning to our original problem, notice that the first digit of 2^i is k if and only if

$$k \cdot 10^r \leq 2^i < (k+1) \cdot 10^r$$

for some $r \geq 0$. In this case,

$$r + \log_{10} k \leq i \log_{10} 2 = i\theta < r + \log_{10}(k+1).$$

This shows that $r = \lfloor i\theta \rfloor$, and

$$\log_{10} k \leq i\theta - \lfloor i\theta \rfloor < \log_{10}(k+1).$$

But $T_\theta^i(0) = i\theta - \lfloor i\theta \rfloor$, so that $T_\theta^i(0) \in J_k$. Summarizing, we see that the first digit of 2^i is k if and only if $T_\theta^i(0) \in J_k$. Thus,

$$\lim_{n\to\infty} \frac{p_k(n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{J_k}(T_\theta^i(0)) = \log_{10}\left(\frac{k+1}{k}\right).$$

Chapter 6

Hurewicz Ergodic Theorem

In this section we consider a class of non-measure preserving transformations. In particular, we study invertible, non-singular and conservative transformations on a probability space. We first start with a quick review of equivalent measures, we then define non-singular and conservative transformations, and state some of their properties. We end this section by giving a new proof of Hurewicz Ergodic Theorem, which is a generalization of Birkhoff Ergodic Theorem to non-singular conservative transformations.

6.1 Equivalent measures

Recall that two measures μ and ν on a measure space $(Y, \mathcal{F})$ are equivalent if μ and ν have the same null-sets, i.e.,

$$\mu(A) = 0 \quad \Leftrightarrow \quad \nu(A) = 0, \qquad A \in \mathcal{F}.$$

The theorem of Radon-Nikodym says that if μ, ν are σ-finite and equivalent, then there exist measurable functions $f, g \geq 0$, such that

$$\mu(A) = \int_A f \, \mathrm{d}\nu \quad \text{and} \quad \nu(A) = \int_A g \, \mathrm{d}\mu.$$

Furthermore, for all $h \in L^1(\mu)$ (or $L^1(\nu)$),

$$\int h \, \mathrm{d}\mu = \int hf \, \mathrm{d}\nu \text{ and } \int h \, \mathrm{d}\nu = \int hg \, \mathrm{d}\mu.$$

Usually the function f is denoted by $\dfrac{d\mu}{d\nu}$ and the function g by $\dfrac{d\nu}{d\mu}$.

Now suppose that $(X, \mathcal{B}, \mu)$ is a probability space, and $T : X \to X$ a measurable transformation. One can define a new measure $\mu \circ T^{-1}$ on $(X, \mathcal{B})$ by $\mu \circ T^{-1}(A) = \mu(T^{-1}A)$ for $A \in \mathcal{B}$. It is not hard to prove that for $f \in L^1(\mu)$,

$$\int f \,\mathrm{d}(\mu \circ T^{-1}) = \int f \circ T \,\mathrm{d}\mu \tag{6.1}$$

Exercise 6.1.1 Starting with indicator functions, give a proof of (6.1).

Note that if T is invertible, then one has that

$$\int f \,\mathrm{d}(\mu \circ T) = \int f \circ T^{-1} \,\mathrm{d}\mu \tag{6.2}$$

6.2 Non-singular and conservative transformations

Definition 6.2.1 *Let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \to X$ an invertible measurable function. T is said to be non-singular if for any $A \in \mathcal{B}$,*

$$\mu(A) = 0 \ \text{ if and only if } \ \mu(T^{-1}A) = 0.$$

Since T is invertible, non-singularity implies that

$$\mu(A) = 0 \ \text{ if and only if } \ \mu(T^n A) = 0, \ n \neq 0.$$

This implies that the measures $\mu \circ T^n$ defined by $\mu \circ T^n(A) = \mu(T^n A)$ is equivalent to μ (and hence equivalent to each other). By the theorem of Radon-Nikodym, there exists for each $n \neq 0$, a non-negative measurable function $\omega_n(x) = \dfrac{d\mu \circ T^n}{d\mu}(x)$ such that

$$\mu(T^n A) = \int_A \omega_n(x) \,\mathrm{d}\mu(x).$$

We have the following propositions.

Proposition 6.2.1 *Suppose $(X, \mathcal{B}, \mu)$ is a probability space, and $T : X \to X$ is invertible and non-singular. Then for every $f \in L^1(\mu)$,*

$$\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f(Tx)\omega_1(x)\,\mathrm{d}\mu(x) = \int_X f(T^n x)\omega_n(x)\,\mathrm{d}\mu(x).$$

Proof We show the result for indicator functions only, the rest of the proof is left to the reader.

$$\begin{aligned}\int_X 1_A(x)\,\mathrm{d}\mu(x) &= \mu(A) = \mu(T(T^{-1}A)) \\ &= \int_{T^{-1}A} \omega_1(x)\,\mathrm{d}\mu(x) \\ &= \int_X 1_A(Tx)\omega_1(x)\,\mathrm{d}\mu(x).\end{aligned}$$

□

Proposition 6.2.2 *Under the assumptions of Proposition 6.2.1, one has for all $n, m \geq 1$, that*

$$\omega_{n+m}(x) = \omega_n(x)\omega_m(T^n x), \quad \mu \text{ a.e.}$$

Proof For any $A \in \mathcal{B}$,

$$\begin{aligned}\int_A \omega_n(x)\omega_m(T^n x)\mathrm{d}\mu(x) &= \int_X 1_A(x)\omega_m(T^n x)\mathrm{d}(\mu \circ T^n)(x) \\ &= \int_X 1_A(T^{-n}x)\omega_m(x)\mathrm{d}\mu(x) \\ &= \int_X 1_{T^n A}(x)\mathrm{d}(\mu \circ T^m)(x) \\ &= \mu \circ T^m(T^n A) = \mu(T^{m+n}A) = \int_A \omega_{n+m}(x)\mathrm{d}\mu(x).\end{aligned}$$

Hence, $\omega_{n+m}(x) = \omega_n(x)\omega_m(T^n x), \quad \mu$ a.e.

Exercise 6.2.1 Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ an invertible non-singular transformation. For any measurable function f, set $f_n(x) = \sum_{i=0}^{n-1} f(T^i x)\omega_i(x)$, $n \geq 1$, where $\omega_0(x) = 1$. Show that for all $n, m \geq 1$,

$$f_{n+m}(x) = f_n(x) + \omega_n(x) f_m(T^n x).$$

Definition 6.2.2 *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ a measurable transformation. We say that T is conservative if for any $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists an $n \geq 1$ such that $\mu(A \cap T^{-n}A) > 0$.*

Note that if T is invertible, non-singular and conservative , then T^{-1} is also non-singular and conservative. In this case, for any $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists an $n \neq 0$ such that $\mu(A \cap T^n A) > 0$.

Proposition 6.2.3 *Suppose T is invertible, non-singular and conservative on the probability space $(X, \mathcal{B}, \mu)$, and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for μ a.e. $x \in A$ there exist infinitely many positive and negative integers n, such that $T^n x \in A$.*

Proof Let $B = \{x \in A : T^n x \notin A \text{ for all } n \geq 1\}$. Note that for any $n \geq 1$, $B \cap T^{-n}B = \emptyset$. If $\mu(B) > 0$, then by conservativity there exists an $n \geq 1$, such that $\mu(B \cap T^{-n}B)$ is positive, which is a contradiction. Hence, $\mu(B) = 0$, and by non-singularity we have $\mu(T^{-n}B) = 0$ for all $n \geq 1$.

Now, let $C = \{x \in A; T^n x \in A \text{ for only finitely many } n \geq 1\}$, then $C \subset \bigcup_{n=1}^{\infty} T^{-n}B$, implying that

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

Therefore, for almost every $x \in A$ there exist infinitely many $n \geq 1$ such that $T^n x \in A$. Replacing T by T^{-1} yields the result for $n \leq -1$. $\square$

Proposition 6.2.4 *Suppose T is invertible, non-singular and conservative, then*

$$\sum_{n=1}^{\infty} \omega_n(x) = \infty, \quad \mu\, a.e.$$

Proof Let $A = \{x \in X : \sum_{n=1}^{\infty} \omega_n(x) < \infty\}$. Note that

$$A = \bigcup_{M=1}^{\infty} \{x \in X : \sum_{n=1}^{\infty} \omega_n(x) < M\}.$$

If $\mu(A) > 0$, then there exists an $M \geq 1$ such that the set

$$B = \{x \in X : \sum_{n=1}^{\infty} \omega_n(x) < M\}$$

has positive measure. Then, $\int_B \sum_{n=1}^{\infty} \omega_n(x)\, \mathrm{d}\mu(x) < M\mu(B) < \infty$. However,

$$\begin{aligned}
\int_B \sum_{n=1}^{\infty} \omega_n(x)\, \mathrm{d}\mu(x) &= \sum_{n=1}^{\infty} \int_B \omega_n(x)\, \mathrm{d}\mu(x) \\
&= \sum_{n=1}^{\infty} \mu(T^n B) \\
&= \sum_{n=1}^{\infty} \int_X 1_{T^n B}(x)\, \mathrm{d}\mu(x) \\
&= \int_X \sum_{n=1}^{\infty} 1_B(T^{-n}x)\, \mathrm{d}\mu(x).
\end{aligned}$$

Hence, $\int_X \sum_{n=1}^{\infty} 1_B(T^{-n}x)\, \mathrm{d}\mu(x) < \infty$, which implies that

$$\sum_{n=1}^{\infty} 1_B(T^{-n}x) < \infty \quad \mu\, a.e.$$

Therefore, for μ a.e. x one has $T^{-n}x \in B$ for only finitely many $n \geq 1$, contradicting Proposition 6.2.3. Thus $\mu(A) = 0$, and

$$\sum_{n=1}^{\infty} \omega_n(x) = \infty, \quad \mu\, a.e.$$

□

6.3 Hurewicz Ergodic Theorem

The following theorem by Hurewicz is a generalization of Birkhoff's Ergodic Theorem to our setting; see also Hurewicz' original paper [H]. We give a new prove, similar to the proof for Birkhoff's Theorem; see Section 2.1 and [KK].

Theorem 6.3.1 *Let $(X, \mathcal{B}, \mu)$ be a probability space, and $T : X \to X$ an invertible, non-singular and conservative transformation. If $f \in L^1(\mu)$, then*

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} f(T^i x)\omega_i(x)}{\sum_{i=0}^{n-1} \omega_i(x)} = f_*(x)$$

exists μ a.e. Furthermore, f_* is T-invariant and

$$\int_X f(x)\,\mathrm{d}\mu(x) = \int_X f_*(x)\,\mathrm{d}\mu(x).$$

Proof Assume with no loss of generality that $f \geq 0$ (otherwise we write $f = f^+ - f^-$, and we consider each part separately). Let

$$f_n(x) = f(x) + f(Tx)\omega_1(x) + \cdots + f(T^{n-1}x)\omega_{n-1}(x),$$

$$g_n(x) = \omega_0(x) + \omega_1(x) + \cdots + \omega_{n-1}(x), \quad \omega_0(x) = g_0(x) = 1,$$

$$\overline{f}(x) = \limsup_{n\to\infty} \frac{f_n(x)}{\sum_{i=0}^{n-1} \omega_i(x)} = \limsup_{n\to\infty} \frac{f_n(x)}{g_n(x)},$$

and

$$\underline{f}(x) = \liminf_{n\to\infty} \frac{f_n(x)}{\sum_{i=0}^{n-1} \omega_i(x)} = \liminf_{n\to\infty} \frac{f_n(x)}{g_n(x)}.$$

By Proposition (6.2.2), one has $g_{n+m}(x) = g_n(x) + g_m(T^n x)$. Using Exercise (6.2.1) and Proposition (6.2.4), we will show that $\overline{f}$ and $\underline{f}$ are T-invariant. To this end,

$$\begin{aligned}
\overline{f}(Tx) &= \limsup_{n\to\infty} \frac{f_n(Tx)}{g_n(T^n x)} \\
&= \limsup_{n\to\infty} \frac{\dfrac{f_{n+1}(x) - f(x)}{\omega_1(x)}}{\dfrac{g_{n+1}(x) - g(x)}{\omega_1(x)}} \\
&= \limsup_{n\to\infty} \frac{f_{n+1}(x) - f(x)}{g_{n+1}(x) - g(x)} \\
&= \limsup_{n\to\infty} \left[\frac{f_{n+1}(x)}{g_{n+1}(x)} \cdot \frac{g_{n+1}(x)}{g_{n+1}(x) - g(x)} - \frac{f(x)}{g_{n+1}(x) - g(x)}\right] \\
&= \limsup_{n\to\infty} \frac{f_{n+1}(x)}{g_{n+1}(x)} \\
&= \overline{f}(x).
\end{aligned}$$

(Similarly $\underline{f}$ is T-invariant).

Now, to prove that f_* exists, is integrable and T-invariant, it is enough to show that
$$\int_X \underline{f}\, \mathrm{d}\mu \geq \int_X f\, \mathrm{d}\mu \geq \int_X \overline{f}\, \mathrm{d}\mu.$$
For since $\overline{f} - \underline{f} \geq 0$, this would imply that $\overline{f} = \underline{f} = f_*$. a.e.

We first prove that $\int_X \overline{f} d\mu \leq \int_X f\, \mathrm{d}\mu$. Fix any $0 < \epsilon < 1$, and let $L > 0$ be any real number. By definition of $\overline{f}$, for any $x \in X$, there exists an integer $m > 0$ such that
$$\frac{f_m(x)}{g_m(x)} \geq \min(\overline{f}(x), L)(1-\epsilon).$$
Now, for any $\delta > 0$ there exists an integer $M > 0$ such that the set
$$X_0 = \{x \in X : \exists\ 1 \leq m \leq M \text{ with } f_m(x) \geq g_m(x)\ \min(\overline{f}(x), L)(1-\epsilon)\}$$
has measure at least $1-\delta$. Define F on X by
$$F(x) = \begin{cases} f(x) & x \in X_0 \\ L & x \notin X_0. \end{cases}$$
Notice that $f \leq F$ (why?). For any $x \in X$, let $a_n = a_n(x) = F(T^n x)\omega_n(x)$, and $b_n = b_n(x) = \min(\overline{f}(x), L)(1-\epsilon)\omega_n(x)$. We now show that $\{a_n\}$ and $\{b_n\}$ satisfy the hypothesis of Lemma 2.1.1 with $M > 0$ as above. For any $n = 0, 1, 2, \ldots$

–if $T^n x \in X_0$, then there exists $1 \leq m \leq M$ such that
$$f_m(T^n x) \geq \min(\overline{f}(x), L)(1-\epsilon) g_m(T^n x).$$
Hence,
$$\omega_n(x) f_m(T^n x) \geq \min(\overline{f}(x), L)(1-\epsilon) g_m(T^n x)\omega_n(x).$$
Now,
$$\begin{aligned} b_n + \ldots + b_{n+m-1} &= \min(\overline{f}(x), L)(1-\epsilon) g_m(T^n x)\omega_n(x) \\ &\leq \omega_n(x) f_m(T^n x) \\ &= f(T^n x)\omega_n(x) + f(T^{n+1}x)\omega_{n+1}(x) + \cdots + f(T^{n+m-1}x)\omega_{n+m-1}(x) \\ &\leq F(T^n x)\omega_n(x) + F(T^{n+1}x)\omega_{n+1}(x) + \cdots + F(T^{n+m-1}x)\omega_{n+m-1}(x) \\ &= a_n + a_{n+1} + \cdots + a_{n+m-1}. \end{aligned}$$

–If $T^n x \notin X_0$, then take $m = 1$ since

$$a_n = F(T^n x)\omega_n(x) = L\omega_n(x) \geq \min(\overline{f}(x), L)(1-\epsilon)\omega_n(x) = b_n.$$

Hence by T-invariance of $\overline{f}$, and Lemma 2.1.1 for all integers $N > M$ one has

$$F(x)+F(Tx)+\omega_1(x)+\cdots+\omega_{N-1}(x)F(T^{N-1}x) \geq \min(\overline{f}(x), L)(1-\epsilon)g_{N-M}(x).$$

Integrating both sides, and using Proposition (6.2.1) together with the T-invariance of $\overline{f}$ one gets

$$\begin{aligned} N\int_X F(x)\,\mathrm{d}\mu(x) &\geq \int_X \min(\overline{f}(x), L)(1-\epsilon)g_{N-M}(x)\,\mathrm{d}\mu(x) \\ &= (N-M)\int_X \min(\overline{f}(x), L)(1-\epsilon)\,\mathrm{d}\mu(x). \end{aligned}$$

Since

$$\int_X F(x)\,\mathrm{d}\mu(x) = \int_{X_0} f(x)\,\mathrm{d}\mu(x) + L\mu(X\setminus X_0),$$

one has

$$\begin{aligned} \int_X f(x)\,\mathrm{d}\mu(x) &\geq \int_{X_0} f(x)\,\mathrm{d}\mu(x) \\ &= \int_X F(x)\,\mathrm{d}\mu(x) - L\mu(X\setminus X_0) \\ &\geq \frac{(N-M)}{N}\int_X \min(\overline{f}(x), L)(1-\epsilon)\,\mathrm{d}\mu(x) - L\delta. \end{aligned}$$

Now letting first $N \to \infty$, then $\delta \to 0$, then $\epsilon \to 0$, and lastly $L \to \infty$ one gets together with the monotone convergence theorem that $\overline{f}$ is integrable, and

$$\int_X f(x)\,\mathrm{d}\mu(x) \geq \int_X \overline{f}(x)\,\mathrm{d}\mu(x).$$

We now prove that

$$\int_X f(x)\,\mathrm{d}\mu(x) \leq \int_X \underline{f}(x)\,\mathrm{d}\mu(x).$$

Fix $\epsilon > 0$, for any $x \in X$ there exists an integer m such that

$$\frac{f_m(x)}{g_m(x)} \leq (\underline{f}(x) + \epsilon).$$

For any $\delta > 0$ there exists an integer $M > 0$ such that the set

$$Y_0 = \{x \in X : \exists\, 1 \leq m \leq M \text{ with } f_m(x) \leq\ (\underline{f}(x) + \epsilon) g_m(x)\}$$

has measure at least $1 - \delta$. Define G on X by

$$G(x) = \begin{cases} f(x) & x \in Y_0 \\ 0 & x \notin Y_0. \end{cases}$$

Notice that $G \leq f$. Let $b_n = G(T^n x)\omega_n(x)$, and $a_n = (\underline{f}(x) + \epsilon)\omega_n(x)$. We now check that the sequences $\{a_n\}$ and $\{b_n\}$ satisfy the hypothesis of Lemma 2.1.1 with $M > 0$ as above.

–if $T^n x \in Y_0$, then there exists $1 \leq m \leq M$ such that

$$f_m(T^n x) \leq\ (\underline{f}(x) + \epsilon) g_m(T^n x).$$

Hence,

$$\omega_n(x) f_m(T^n x) \leq\ (\underline{f}(x)+\epsilon) g_m(T^n x)\omega_n(x) = (\underline{f}(x)+\epsilon)(\omega_n(x)+\cdots+\omega_{n+m-1}(x).$$

By Proposition (6.2.2), and the fact that $f \geq G$, one gets

$$\begin{aligned} b_n + \ldots + b_{n+m-1} &= G(T^n x)\omega_n(x) + \cdots + G(T^{n+m-1}x)\omega_{n+m-1}(x) \\ &\leq f(T^n x)\omega_n(x) + \cdots + f(T^{n+m-1}x)\omega_{n+m-1}(x) \\ &= \omega_n(x) f_m(T^n x) \\ &\leq (\underline{f}(x) + \epsilon)(\omega_n(x) + \cdots + \omega_{n+m+1}(x)) \\ &= a_n + \cdots a_{n+m-1}. \end{aligned}$$

–If $T^n x \notin Y_0$, then take $m = 1$ since

$$b_n = G(T^n x)\omega_n(x) = 0 \leq (\underline{f}(x) + \epsilon)(\omega_n(x)) = a_n.$$

Hence by Lemma 2.1.1 one has for all integers $N > M$

$$G(x) + G(Tx)\omega_1(x) + \ldots + G(T^{N-M-1}x)\omega_{N-M-1}(x) \leq (\underline{f}(x) + \epsilon) g_N(x).$$

Integrating both sides yields

$$(N-M)\int_X G(x)d\mu(x) \le N(\int_X \underline{f}(x)d\mu(x)+\epsilon).$$

Since $f \ge 0$, the measure ν defined by $\nu(A)=\int_A f(x)\,\mathrm{d}\mu(x)$ is absolutely continuous with respect to the measure μ. Hence, there exists $\delta_0 > 0$ such that if $\mu(A) < \delta$, then $\nu(A) < \delta_0$. Since $\mu(X\setminus Y_0) < \delta$, then $\nu(X\setminus Y_0) = \int_{X\setminus Y_0} f(x)d\mu(x) < \delta_0$. Hence,

$$\begin{aligned}\int_X f(x)\,\mathrm{d}\mu(x) &= \int_X G(x)\,\mathrm{d}\mu(x) + \int_{X\setminus Y_0} f(x)\,\mathrm{d}\mu(x)\\ &\le \frac{N}{N-M}\int_X (\underline{f}(x)+\epsilon)\,\mathrm{d}\mu(x)+\delta_0.\end{aligned}$$

Now, let first $N\to\infty$, then $\delta\to 0$ (and hence $\delta_0\to 0$), and finally $\epsilon\to 0$, one gets

$$\int_X f(x)\,\mathrm{d}\mu(x) \le \int_X \underline{f}(x)\,\mathrm{d}\mu(x).$$

This shows that

$$\int_X \underline{f}\,\mathrm{d}\mu \ge \int_X f\,\mathrm{d}\mu \ge \int_X \overline{f}\,\mathrm{d}\mu,$$

hence, $\overline{f}=\underline{f}=f_*$ a.e., and f_* is T-invariant. □

Remark We can extend the notion of ergodicity to our setting. If T is non-singular and conservative, we say that T is ergodic if for any measurable set A satisfying $\mu(A\Delta T^{-1}A)=0$, one has $\mu(A)=0$ or 1. It is easy to check that the proof of Proposition (1.7.1) holds in this case, so that T ergodic implies that each T-invariant function is a constant μ a.e. Hence, if T is invertible, non-singular, conservative and ergodic, then by Hurewicz Ergodic Theorem one has for any $f\in L^1(\mu)$,

$$\lim_{n\to\infty}\frac{\sum_{i=0}^{n-1} f(T^i x)\omega_i(x)}{\sum_{i=0}^{n-1}\omega_i(x)} = \int_X f d\mu \quad \mu \ a.e.$$

Bibliography

[B] Michael Brin and Garrett Stuck, *Introduction to Dynamical Systems*, Cambridge University Press, 2002.

[D] K. Dajani and C. Kraaikamp, *Ergodic theory of numbers*, Carus Mathematical Monographs, 29. Mathematical Association of America, Washington, DC 2002.

[H] W. Hurewicz, *Ergodic theorem withour invariant measure*. Annals of Math., **45** (1944), 192–206.

[KK] Kamae, Teturo; Keane, Michael, *A simple proof of the ratio ergodic theorem*, Osaka J. Math. 34 (1997), no. 3, 653–657.

[KT] Kingman and Taylor, *Introduction to measure and probability*, Cambridge Press, 1966.

[Pa] William Parry, *Topics in Ergodic Theory*, Reprint of the 1981 original. Cambridge Tracts in Mathematics, 75. Cambridge University Press, Cambridge, 2004.

[P] Karl Petersen, *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, 2. Cambridge University Press, Cambridge, 1989.

[W] Peter Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982.

Index

β-transformations, 10

Stationary Stochastic Processes, 11

algebra, 7
algebra generated, 7
atoms of a partition, 58

Baker's Transformation, 10
Bernoulli Shifts, 11
binary expansion, 9
Birkhoff's Ergodic Theorem, 29

common refinement, 58
conditional expectation, 35
conditional information function, 66
conservative, 94
Continued Fractions, 13

dynamical system, 47

entropy of the partition, 57
entropy of the transformation, 61
equivalent measures, 91
ergodic, 19
ergodic decomposition, 40
extreme point, 82

factor map, 51
first return time, 15

generator, 64

Hurewicz Ergodic Theorem, 95

induced map, 15
induced operator, 21
information function, 66
integral system, 18
irreducible Markov chain, 42
isomorphic, 47

Kac's Lemma, 35
Knopp's Lemma, 25

Lochs' Theorem, 72

Markov measure, 41
Markov Shifts, 11
measure preserving, 6
monotone class, 7

natural extension, 52
non-singular transformation, 92

Poincaré Recurrence Theorem, 14

Radon-Nikodym Theorem, 91
random shifts, 12
regular measure, 79
Riesz Representation Representation Theorem, 80

semi-algebra, 7
Shannon-McMillan-Breiman Theorem, 70
strongly mixing, 45
subadditive sequence, 60

translations, 9

uniquely ergodic, 86

weakly mixing, 45