Ergodic Theory and Entropy

Ashish Varghese George
MS15144

A dissertation submitted for the partial fulfilment of
BS-MS dual degree in Science

Indian Institute of Science Education and Research Mohali
June 2020

Certificate of Examination

This is to certify that the dissertation titled "Ergodic Theory and Entropy" submitted by Mr. Ashish Varghese George (Reg. No. MS15144) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr Pranab Sardar (Supervisor)	Dr Soma Maity	Dr Lingaraj Sahu

Dated: June 22, 2020

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Pranab Sardar at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Ashish Varghese George
(Candidate)
June 22, 2020

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Pranab Sardar
(Supervisor)

Acknowledgements

I'd like to extend my sincere thanks to my advisor, Dr Pranab Sardar, for his guidance, support and patience. I am genuinely grateful for his constant mentoring and efforts to motivate me to complete this work.

I'd like to express my gratitude towards my thesis committee members: Dr Soma Maity and Dr Lingaraj Sahu for their valuable inputs.

Words cannot describe how grateful I am to George Shaji, Gautham Neelakandan, and Nilangshu Bhattacharya for their support and interest in helping me build my thesis. Attending the weekly lectures without any hesitation has always boosted my confidence to explore further into this topic.

I sincerely thank Dr Krishnendu Gangopadhyay for teaching me not to give up on anything and for showing me that there is always hope.

To my dearest friends: Sunandini, for your support and keeping my sanity in check. Also, for helping me understand how to use Latex. Afham, Asif, Athul, Alan, Nitheesh and Sidharth, a.k.a. The Poker Kings, for your presence and making the last year of my campus life fun and memorable. Lopamudra Das, for your emotional support.

I'd like to thank my parents for their support and prayers. I am ever grateful for helping me achieve my goals. I am indebted to them for their continuous encouragement throughout my years of study. A big thanks to them for believing in me.

— Ashish Varghese George

Notations

$\emptyset$	Empty set
$\mathbb{1}_A$	Indicator or characteristic function of a set A
$\mathcal{F}$	σ-algebra
$\mathscr{B}$	Borel σ-algebra
$C(X)$	Set of continuous functions from a space X to $\mathbb{R}$

Contents

Introduction

The ergodic theory deals with the study of dynamical systems and their statistical properties. The inception of this branch of mathematics is from the problems that rose from the early developmental stages of statistical mechanics.

We would begin introducing some elementary notions of ergodicity illustrated with simple examples. Chapter 1 will lay foundations to understand Birkhoff's ergodic theorem in Chapter 2. After proving the theorem, Chapter 2 will prove the Von Neumann's version of ergodic theorem which is also called the Mean Ergodic Theorem. Chapter 3 will define how dynamical systems can to be isomorphic to each other.

Chapter 4 will introduce the notions of mixing in dynamical systems. After discussing several forms of mixing, I will show how they relate to ergodicity.

Chapter 5 will define entropy. This chapter can be seen as three sections. The first section will discuss the measure-theoretic entropy. The next section will explore the topological entropy. The final section will give a relation between the measure-theoretic and topological entropies. This relation is also known as the variation principle.

Remark on References:
The measure theory discussed in Chapter 0 are based on Walter Rudin's Real and Complex Analysis. Ergodic transformations in Chapter 0, Chapter 1 till Chapter 4 are based on C E Silva's Invitation to Ergodic Theory. The final Chapter that discusses entropy is based on Peter Walters' Introduction to Ergodic Theory.

Chapter 0

A Brief Review of Measure Theory

0.1 Preliminaries

In this chapter, we will give a review of some measure theory concepts from Walter Rudin's book, Real and Complex Analysis.

Definition 1.1.1 A σ-algebra in a set X is a collection of subsets of X, denoted $\mathcal{F}$, satisfying the following properties:

1. $X \in \mathcal{F}$
2. For $A \subseteq X,\ A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$
3. If $A_n \ \in \mathcal{F}$ for n = 1,2,3,... and set $A \ = \cup_{n=1}^{\infty}$ then $A_n \ \in \mathcal{F}$.

Definition 1.1.2: Suppose $\mathcal{F}$ is a σ-algebra in X. Then $(X, \mathcal{F})$ is called a **measurable space** and we call the elements of $\mathcal{F}$ as **measurable sets** in X.

Definition 0.1.3: Let $(X, \mathcal{F})$ be a measurable space. Then the function $\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ is called a **measure** on $(X, \mathcal{F})$ if μ satisfies the following,

1. For all $A \in \mathcal{F}$, $\mu(A) \leq 0$.
2. $\mu(\emptyset) = 0$.
3. For countable collections of pairwise-disjoint sets $\{A_n\}_{n=0}^{\infty} \in \mathcal{F}$,

$$\mu(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mu(A_n)$$

With μ as a measure over a measurable space $(X, \mathcal{F})$, we can call $(X, \mathcal{F}, \mu)$ a **measure space**. A measure space is called a **probability space** if $\mu(X) = 1$. $(X, \mathcal{F}, \mu)$ is a **finite** measure space if the measure μ takes only finite values.
Let $A = (a, b)$ or, $[a, b] \in \mathbb{R}$. Let l(A) = b-a. Then the Lebesgue outer measure, $\mu^*(A)$ is defined as

$$\mu^*(A) = \left\{ \sum_{n=0}^{\infty} l(I_n) : A \subseteq \bigcup_{n=0}^{\infty} I_n \text{ where for each} n \in \mathbb{Z}, I_n \text{ is a open interval} \right\}$$

A set $A \in \mathcal{F}$ is called a **null-set** if its Lebesgue outer measure is zero.

Definition 0.1.4: A measure space $(X, \mathcal{F}, \mu)$ is said to have σ-finite-measure space if $(X, \mathcal{F}, \mu)$ is a finite measure space and X is a countable union of measurable sets A_n with finite measure, i.e.,

$$X = \bigcup_{n=0}^{\infty} A_n$$

Definition 0.1.5: A **semi-ring** is a collection $\mathcal{A}$ of subsets on a non-empty set X such that

1. $\mathcal{A}$ is non-empty
2. If $U, V \in \mathcal{A}$ then, $U \cap V \in \mathcal{A}$
3. If $U, V \in \mathcal{A}$ then, $U \setminus V = \bigsqcup_{j=1}^{n} I_j$, where $I_j \in \mathcal{A}$ are disjoint.

Definition 0.1.6: We say that a property on X holds **almost everywhere** if it holds except on a set of measure zero.

Definition 0.1.7: A semi-ring $\mathcal{A}$ of measurable sets of X with finite measure is said to be a **sufficient semi-ring** for the measure space $(X, \mathcal{F}, \mu)$ if for every $U \subset \mathcal{F}$,

$$\mu(U) = \inf \left\{ \sum_{j=1}^{\infty} \mu(I_j) : U \subset \cup_{j=1}^{\infty} I_j \text{ and } I_j \in \mathcal{A} \text{ for } j \geq 1 \right\}$$

We can see that the dyadic intervals in $[0, 1]$ forms a sufficient semi-ring of $[0, 1]$.

Lemma 0.1.8: Let $\mathcal{A}$ be a sufficient semi-ring for a non-atomic measure space $(X, \mathcal{F}, \mu)$. If $U \in \mathcal{F}$ with finite measure, then for any $\delta > 0$, there exists a set $I \,\epsilon\, \mathcal{A}$ such that I is $(1 - \delta)$-full of U.

Proof. First we try to show the existence of set $H^* := \sqcup_{j=1}^{N} I_j$, where for each $j \in \mathbb{N}, I_j \in \mathcal{A}$ and are pair-wise disjoint, such that $\mu(U \triangle H^*) < \epsilon$, for any arbitrary $\epsilon > 0$.

Choose $I_j \,\epsilon\, \mathcal{A}$ such that $H_\epsilon := \cup_{j=1}^{\infty} I_j \supset U$ and $\mu(H_\epsilon) < \mu(U) + \frac{\epsilon}{2}$. Then $\mu(H_\epsilon \setminus U) < \frac{\epsilon}{2}$. There exists an $n > 1$ such that $0 \leq \mu(H_\epsilon) - \mu(\cup_{j=1}^{N} I_j) < \frac{\epsilon}{2}$. Call this union of N sets as $H^* := \cup_{j=1}^{N} I_j$. Then,

$$\begin{aligned}
\mu(H^* \triangle U) &= \mu(H^* \setminus U) + \mu(U \setminus H^*) \\
&\leq \mu(H_\epsilon \setminus U) + \mu(H^\epsilon \setminus H^*) \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

But I_j's are disjoint. So this proves the existence of such an H^*.

Now choose $\epsilon > 0$ such that,

$$\mu(H^* \triangle U) < \epsilon \ \mu(U)$$
$$\text{then, } \epsilon \ \mu(U) > \mu(U) - \mu(U \cap H^*)$$

$$\mu(U \cap H^*) > (1 - \epsilon) \ \mu(U) \tag{1}$$

also,

$$\begin{aligned}
\mu(U) - \epsilon \ \mu(U) < \mu(H^*) &< \mu(U) + \epsilon \ \mu(U) \\
&< (1 + \epsilon) \ \mu(U)
\end{aligned}$$

Assume for all $j \in \{1, 2, ..., N\}$, $\mu(U \cap I_j) \leq (1 - \delta) \ \mu(I_j)$. Then,

$$\mu(U \cap H^* < (1 - \delta) \ \mu(H^*)$$

Therefore,

$$\mu(U \cap H^* < (1 - \delta) \ (1 + \epsilon) \ \mu(U)$$

But if ϵ is chosen such that $\epsilon = \frac{\delta}{2 - \delta}$ then,

$$\begin{aligned}
(1 - \delta) \ (1 + \epsilon) &= (1 - \delta) \ (1 - \frac{\delta}{2 - \delta}) \\
&= (1 - \delta) \ \frac{2}{2 - \delta} \\
&= 1 - \frac{\delta}{2 - \delta} \\
&= 1 - \epsilon.
\end{aligned}$$

$$\implies \mu(U \cap H^*) < (1-\epsilon)\mu(U)$$

But from eqn.**(1)**, this is a contradiction. So there must exist some $j \in \{1, 2, ..., N\}$ such that $\mu(U \cap I_j) > (1-\delta)\ \mu(I_j)$. $\square$

Definition 0.1.9: A function $f : X \to Y$, where X is a measurable space and Y a topological space, is said to be measurable if $f^{-1}(V)$ is a measurable set in X for every open set V in Y.

From definitions above we can observe the following:

1. $\emptyset \in \mathcal{F}$, as $X \ \in \mathcal{F}$ implies $\emptyset = X^c \in \mathcal{F}$.
2. For any finite collection in $\mathcal{F}$, say A_i for $1 \leq i \leq n$, we have $\bigcup_{i=1}^n A_i \in \mathcal{F}$.
3. $\mathcal{F}$ is also closed under countable (or finite) intersection of measurable sets.
4. For $A,\ B \in \mathcal{F},\ A - B$ and $B - A \in \mathcal{F}$

Example 0.1.10: Consider a monotonic function $f : [a, b] \to \mathbb{R}$ where $a, b \in \mathbb{R}$. Then f is differentiable almost everywhere.

Theorem0.1.11: (Fatou's Lemma) Suppose $(X, \mathcal{F})$ be a measurable space. Let $\{f_n : X \to \mathbb{R}_{\geq 0}\}$ be a collection of measurable non-negative functions, then

$$\int \liminf_{n\to\infty} f_n d\mu \ \leq \ \liminf_{n\to\infty} \int f_n d\mu$$

Theorem 0.1.12: (Lebesgue's Monotone Convergence Theorem) Let $\{f_n\}$ be a sequence of functions such that $f_n(x)$ is monotone increasing for each x. Let $f = \lim_{n\to\infty}\ f_n$. Then,

$$\int f\ dx = \lim_{n\to\infty} \int f_n\ dx$$

Theorem 0.1.13: (Lebesgue's Dominated Convergence Theorem) Let $f_n : X \to \mathbb{R}$ be a sequence of measurable functions over a measurable space $(X, \mathcal{F})$ such that $|f_n| \leq g$ for all n = 1,2,3...,where g is integrable and let $\lim_{n\to\infty}\ f_n = f\ a.e.$ Then,

- f is integrable
- $\lim_{n\to\infty} \int f_n\ dx = \int f\ dx$

Definition 0.1.14: Let μ_1 and μ_2 be two probability measures on a measurable space $(X, \mathcal{F})$. We say that μ_1 is absolutely continuous with respect to μ_2 (denoted $\mu_1 << \mu_2$) if for $A \in \mathcal{F}$, $\mu_1(A) = 0$ whenever $\mu_2(A) = 0$.

Theorem 0.1.15: (Radon-Nikodym Theorem) Let μ_1 and μ_2 be two probability measures on a measurable space $(X, \mathcal{F})$. Then $\mu_1 << \mu_2$ if and only if there exists $f \in L^2(\mu_2)$ with $f \geq 0$ and $\int f\ d\mu = 1$ such that $\mu_1(A) = \int_A f\ d\mu_2,\ \forall A \in \mathcal{F}$. And any other function with these properties must be equal to $f, a.e.$.

This function f is called the **Radon-Nikodym derivative**. It is denoted by $\frac{d\mu_1}{d\mu_2}$.

0.2 Transformations

Definition 0.2.1: A function which has the same range and domain is called a **transformation**.

Definition 0.2.2: For any set X, if $T: \ X \to X$ is a transformation on X, the nth iterate of $x \in X$ with respect to the transformation T is denoted as $T^n(x)$, where

$$T^0(x) = x$$
$$T^{n+1}(x) = T \circ T^n \text{ for n} \geq 0$$

We can see that if T is bijective on X, then T is an invertible transformation.

Example 0.2.3: Let $X = [0, 1)$. A rotational transformation of X can be constructed using modulus 1. For any $\alpha \in \mathbb{R}$, rotation by α can be defined as

$$\begin{aligned} R_\alpha : [0,1) &\to [0,1) \\ x \ &\to x + \alpha \mod 1 \end{aligned}$$

Note: One can see that R_α is invertible and R_α^{-1} is $R_{-\alpha}$.

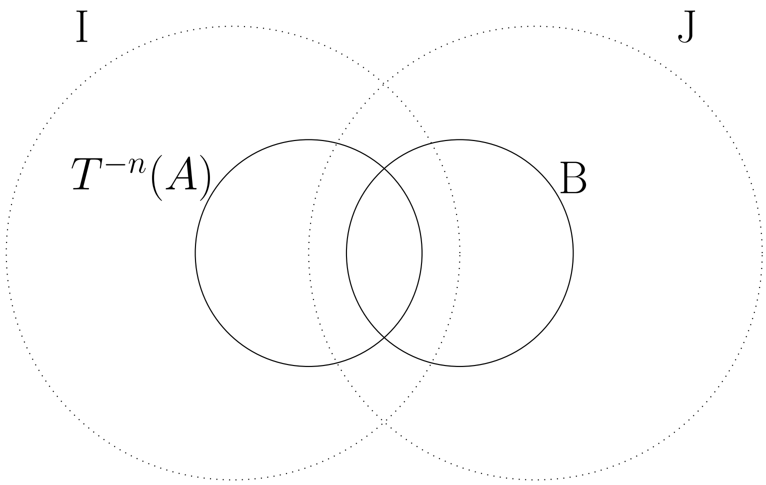
Definition 0.2.4: Let $(X, \mathcal{F})$ be a measurable space and $T : X \to X$ is a transformation such that for all $A \in \mathcal{F}$, $T^{-1}(A) \in \mathcal{F}$. Then T is called a **measurable transformation**.

Definition 0.2.5: Let (X, F, μ) be a measure space. A transformation T is **measure-preserving** if $\mu(T^{-1}(A)) = \mu(A)$, for all sets $A \in \mathcal{F}$. The measure μ will be an invariant measure for T.

Lemma 0.2.6: Let T be a measure-preserving transformation on a measure space $(X, \mathcal{F}, \mu)$. Let $A,\ B,\ I,\ \&\ J$ be measurable sets in X such that $A \subset I$ and $B \subset J$. Then for all integers $n > 0$,

$$\mu(T^{-n}(A) \cap B) \geq \mu(T^{-n}(I) \cap J) - \mu(I \setminus A) - \mu(J \setminus B).$$

Proof. The proof becomes simpler if we look at the following diagram.



Here,

$$(T^{-n}(I) \cap J) \subset (T^{-n}(A) \cap B) \cup (T^{-n}(I) \setminus T^{-n}(A)) \cup (J \setminus B)$$

With T being measure-preserving, we get,

$$\mu(T^{-n}(I) \cap J) \leq \mu(T^{-n}(A) \cap B) + \mu(I \setminus A) + \mu(J \setminus B)$$

On re-arranging the order,

$$\mu(T^{-n}(A) \cap B) \geq \mu(T^{-n}(I) \cap J) - \mu(I \setminus A) - \mu(J \setminus B).$$

□

Example 0.2.7: Define a transformation on [0,1) as follows,

$$T : [0, 1) \to [0, 1)$$

$$x \to 2x \mod 1 = \begin{cases} 2x, & \text{if } 0 \le x \le \frac{1}{2} \\ 2x - 1, & \text{if } \frac{1}{2} \le x \le 1 \end{cases}$$

Such a map is called a **Doubling Map**. To show this map T is measure-preserving on the measure space $([0, 1), \mathcal{F}, \lambda)$, where λ is the Lebesgue measure, define,

$$S_1 : [0, 1) \to [0, \frac{1}{2})$$
$$S_1(y) = \frac{y}{2}$$
$$S_2 : [0, 1) \to [\frac{1}{2}, 1)$$
$$S_2(y) = \frac{y}{2} + \frac{1}{2}$$

For any Measurable set A in $[0, 1)$, $\frac{A}{2}$ and $\frac{A}{2} + \frac{1}{2}$ are measurable. Also, $\lambda(\frac{A}{2}) = \frac{1}{2}\lambda(A) = \frac{1}{2}\lambda(A + \frac{1}{2}) = \lambda(\frac{A}{2} + \frac{1}{2})$. Then with $T^{-1}(A) = S_1(A) \sqcup S_2(A)$,

$$\begin{aligned} \lambda(T^{-1}) &= \lambda(S_1(A)) + \lambda(S_2(A)) \\ &= \lambda(\frac{A}{2}) + \lambda(\frac{A}{2} + \frac{1}{2}) \\ &= 2\lambda(\frac{A}{2}) \\ &= \lambda(A) \end{aligned}$$

Recall Example 1.2.3, on how a rotational transformation was defined. In order to prove Kronecker's Theorem for Irrational Rotations (also known as the Kronecker's Approximation Theorem), we would define a metric d on [0,1) as,

$$d(x, y) = min\{|x - y|, 1 - |x - y|\} \text{ , for } x, y \,\epsilon\, [0, 1)$$

Note that for a rotational transformation R_α, $d(R_\alpha(x), R_\alpha(y)) = d(x, y)$ for any $x, y \in [0, 1)$.

Theorem 0.2.8 (Kronecker's Theorem) For any $x \in [0, 1)$ and a rotational transformation R_α with α being irrational , the sequence $\{R^n_\alpha(x)\}_{n\ge0}$ in $[0, 1)$ is dense.

Proof. First we will show for any $m, n \in \mathbb{N}$,if $R^m_\alpha(x) = R^n_\alpha(x)$ then $m = n$. For this, if $R^m_\alpha(x) = R^n_\alpha(x)$ then,

$$x + m \mod 1 = x + n \mod 1$$

which implies,

$$0 \equiv (m - n)\alpha \mod 1$$

We get $(m - n)\alpha \in \mathbb{Z}$. But α is irrational, so the only possibility is to have $n = m$. Therefore, for each n, the values of $R^n_\alpha(x)$ are unique.
As the sequence $\{R^n_\alpha(x)\}$ fall in the interval $[0, 1)$, and $[0, 1)$ being bounded, we can invoke the Bolzano-Weierstrass Theorem here. The theorem says that there exists a converging subsequence in $\{R^n_\alpha(x)\}$. Then for an arbitrary $\epsilon > 0$, there exists $p, q \,\epsilon\, \mathbb{N}$ such that $R^p_\alpha(x)$ and $R^q_\alpha(x)$ belong to this converging subsequence and $d(R^p_\alpha(x), R^q_\alpha(x)) < \epsilon$. But $R^p_\alpha(x) = x + p\alpha \mod 1$ so,

$$\begin{aligned} |x + p\alpha - (x + q\alpha) \mod 1| &< \epsilon \\ |x + (p - q)\alpha - x \mod 1| &< \epsilon \\ d(R^{p-q}_\alpha(x), x) &< \epsilon \end{aligned}$$

Set $\delta = d(R^{p-q}_\alpha(x), x)$. Consider the sequence $\{R^{n(p-q)}_\alpha(x)\}$. Here taking any consecutive terms, say $R^{(n+1)(p-q)}_\alpha(x)$ and $R^{n(p-q)}_\alpha(x)$, they differ by δ. Thus the sequence $\{R^{n(p-q)}_\alpha(x)\}$ subdivides $[0, 1)$ into intervals of the length less than ϵ. This shows $\{R^n_\alpha(x)\}_{n\ge0}$ is dense in $[0, 1)$. □

Lemma 0.2.9: A measurable transformation T is measure preserving if and only if for all integrable non-negative function $f : X \to [0, \infty)$,

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

Proof. Let T be a measurable transformation. Suppose for all non-negative integrable function f satisfy, $\int f \circ T d\mu = \int f d\mu$. Let $f = \mathbb{1}_A$. f is clearly a non-negative integrable function. Then any measurable set A of finite measure, $f \circ T = \mathbb{1}_{T^{-1}(A)}$.
Note that $\int f \circ T d\mu = \mu(T^{-1}(A))$ and $\int_A f d\mu = \mu(A)$. But we have $\int f \circ T d\mu = \int f d\mu$. Therefore,

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

If A has infinite measure, let A_n be point-wise disjoint and $\mu(A_n) < \infty$ for each n such that, $A = \sqcup_{n=1}^{\infty} A_n$. Then,

$$\begin{aligned}
\mu(T^{-1}(A)) &= \mu(T^{-1}(\sqcup_{n=1}^{\infty} A_n)) \\
&= \sum_{n=1}^{\infty} \mu(T^{-1}(A_n)) \\
&= \sum_{n=1}^{\infty} \mu(A_n) \\
&= \mu(\sqcup_{n=1}^{\infty} A_n) \\
&= \mu(A)
\end{aligned}$$

Thus we can say T is a measure preserving transformation.

To show the other way, suppose T is measure preserving. Then $f = \mathbb{1}_A$ gives $\int f \circ T d\mu = \int f d\mu$. Let f be any non-integrable function. Then there exists a sequence of simple functions $\{S_i\}_{i=1}^{\infty}$ such that $S_1 \leq S_2 \leq S_3 \leq ... \leq f$ and $\lim_{n\to\infty} S_n = f$, for all $x \in X$. Recall that a simple function S taking distinct values $a_1, a_2, a_3, ...a_n \in \mathbb{R}$, is of the form, $S(x) = \sum_{j=1}^{N} a_i \mathbb{1}_{E_j}(x)$, where $E_j = \{x \in X : S(x) = a_j\}$ and $\sqcup_{j=1}^{N} E_j = X$.Thus each S_n's are integrable and are finite sums of indicator functions. So for each $n \leq 1$,

$$\int S_n \circ T d\mu = \int S_n d\mu$$

Applying the monotone convergence theorem here, $\lim_{n\to\infty} S_n(x) = f(x)$ gives,

$$\begin{aligned}
\int f \circ T d\mu &= \int \lim_{n\to\infty} S_n \circ T d\mu \\
&= \lim_{n\to\infty} \int S_n \circ T d\mu \\
&= \lim_{n\to\infty} \int S_n d\mu \\
&= \int \lim_{n\to\infty} S_n d\mu \\
&= \int f d\mu
\end{aligned}$$

$\square$

Chapter 1

Introduction to Ergodic Theory

1.1 Recurrence

Definition 1.1.1: Let T be a transformation on a measure space $(X, \mathcal{F}, \mu)$. If for every measurable set $A \subseteq X$ with $\mu(A) > 0$, there is a null set N in A (i.e., $\mu(N) = 0$) such that for all $x \in A/N$, and some integer n (depending on x) with $T^n(x) \in A$, then T is said to be **recurrent**.

Definition 1.1.2: A transformation T on X is said to be **conservative** if for every measurable set A with positive measure, there exists an integer n such that $\mu(A \cap T^{-n}(A)) > 0$.

Lemma 1.1.3: A transformation T on a measure space $(X, \mathcal{F}, \mu)$ is conservative if and only if T is recurrent.

Proof. Assume T is conservative and let A be a measurable set in X with positive measure. Define a set $S = A \setminus \cup_{n=1}^{\infty} T^{-n}(A)$. If this set S has positive measure, T being conservative will imply there is some in integer n such that $\mu(S \cap T^{-n}(S)) > 0$. This means there is some $x \in S$ such that $T^n(x) \in S$.But the way we defined S will discard all such x. Hence we get a contradiction to the assumption that $\mu(S) > 0$. Therefore $S = A \setminus \cup_{n=1}^{\infty} T^{-n}(A) = 0$
This result shows that the set of all points in A which after some iterations fails to be in A is a set of measure zero. So from definition 1.3.1, it can be seen that T is recurrent. In fact we can observe that, T is recurrent if and only if for all measurable sets A of positive measure, $\mu(A \setminus \cup_{n=1}^{\infty} T^{-n}(A)) = 0$.
This will make the converse easier to prove. Suppose T is recurrent and A be a measurable set of positive measure. Then,

$$\begin{aligned} \mu(A \setminus (\cup_{n=1}^{\infty} T^{-n}(A) \cap A)) &= \mu(A \setminus \cup_{n=1}^{\infty} T^{-n}(A)) \\ &= 0 \end{aligned}$$

But $\mu(A) > 0$, so there is some integer n with $\mu(T^{-n}(A) \cap A) > 0$. Therefore T is conservative. $\square$

Before heading towards the Birkhoff Ergodic Theorem which would be stated in the next chapter, we need to prove the Poincaré's Recurrence Theorem. The Ergodic Theorem can be seen as an improvement on Poincaré's theorem. This theorem, proposed in 1890 by Henri Poincaré and later proved in 1919 by Constantin Carathéodory , had a prominent role in the development of statistical mechanics.

Theorem 1.1.4: (Poincaré's Recurrence Theorem) Let T be a measure-preserving transformation on a finite measure space $(X, \mathcal{F}, \mu)$ then T is recurrent.

Proof. From Lemma 1.1.3, it is sufficient to show that T is conservative. Suppose T is not conservative. Then, for any measurable set A of positive measure, $\mu(A \cap T^{-n}(A)) = 0$, $\forall\ n > 0$. Choose two integers i and j such that $i \neq j$. Let $j = m + i$ for some integer $m > 0$. Then using the fact that T is measure-preserving,

$$\begin{aligned}\mu(T^{-j}(A)\cap T^{-i}(A)) &= \mu(T^{-(m+i)}(A)\cap T^{-i}(A))\\ &= \mu(T^{-i}(T^{-m}(A)\cap A))\\ &= \mu(T^{-m}(A)\cap A)\\ &= 0\end{aligned}$$

We Observe that $\{T^{-n}(A)\}_{n\geq 0}$ are almost pairwise disjoint. So,

$$\begin{aligned}\mu(\cup_{n=0}^{\infty}T^{-n}(A)) &= \sum_{n=0}^{\infty}\mu(T^{-n}(A))\\ &= \sum_{n=o}^{\infty}\mu(A)\\ &= \infty\end{aligned}$$

But $(X,\mathcal{F},\mu)$ is a finite measure space and thus we get a contradiction. Hence T is conservative and from Lemma 1.3.3, T is also recurrent ☐

Definition 1.1.5: Let T be a transformation on a measure space X.

- A subset $A\subseteq X$is said to be **positively T-invariant** if $x\in A \implies T(x)\in A$
- A subset $A\subseteq X$ is **strictly T-invariant** if $T^{-1}(A)=A$ or in other words, $x\in A \iff T(x)\in A$.

Observe that if a subset$A\subseteq X$ is positively invariant then$A\subseteq T^{-1}(A)$. Also if we restrict T to a subset A and still be able to define a transformation then A is positively invariant.

Example 1.1.6: Let $X=[0,\infty)$ and $T:X\to X$ such that $T(x)=x+1$. Clearly T is a measure-preserving transformation. Let $A=[1,\infty)$. $A\subseteq X$ and A is positively T-invariant. But A is not strictly T-invariant.

Definition 1.1.7: a set Y is said to be **Strictly invariant mod** μ if $Y=\mu(T^{-1}(Y)) \mod \mu$, where we say $A=B \mod \mu$ if $\mu(A\triangle B)=0$, with A and B being measurable sets.

1.2 Ergodic Transformations

Definition 1.2.1: A measure-preserving transformation T on a measure space $(X,\mathcal{F},\mu)$ is said to be **ergodic** if for every strictly T-invariant set $A\subseteq X$, either $\mu(A)=0,$ or $\mu(A^c)=0$.

Lemma 1.2.2: Let T be a measure-preserving transformation on $(X,\mathcal{F},\mu)$ which is a σ-measure space, Then the following are equivalent:

1. T is ergodic and recurrent.
2. $\mu(X\setminus\cup_{i=1}^{\infty}T^{-i}(A))=0$, for every $A\subseteq X$ of positive measure.
3. For every measurable set A of positive measure and for a.e. $x\in X$ $\exists$ an integer $n\geq 0$ such that $T^n(x)\in A$.
4. If A and B are sets of positive measure then $\exists$ an integer $n\geq 0$ such that $T^{-1}(A)\cap B\neq\emptyset$.
5. If A and B are of positive measure then $\exists$ an integer $n\geq 0$ such that $\mu(T^{-n}(A)\cap B)>0$.

Proof. First we will prove (5) $\implies$ (1). Let $A\subset X$ be a strictly invariant set of positive measure, which means $T^{-1}(A)=A$ for all integers $n>0$. Consider the complement of A, i.e., A^c. If A^c has positive measure, then (5) says that $\exists$ an integer $n>0$ such that,

$$\mu(T^{-1}(A)\cap A^c)>0$$

implies,

$$\mu(A \cap A^c) > 0$$

But this is absurd. So $\mu(A^c) = 0$ as long as A has positive measure. This shows that T is ergodic.

To show recurrence, take the set B as A and (5) gives the existence of an integer $n > 0$ such that $\mu(T^{-1}(A) \cap A) > 0$ which in fact the definition of a recurrent transformation.

For (1) $\implies$ (2), let A be a measurable set of positive measure. Let $S = \cup_{n=1}^{\infty}(A)$. Note that $T(S) = \cup_{n=1}^{\infty} T^{-(n+1)}(A) = \cup_{n=2}^{\infty} T^{-n}(A)$. So $T^{-1}(S) \subset S$. With T being recurrent, $\mu(S \setminus T^{-1}(S) = 0$. This shows, $S = T^{-1}(S)$ *mod* μ (i.e., S is strictly invariant mod μ). But S has positive measure. Thus,

$$\mu(S^c) = 0$$

implies,

$$X = S \mod \mu$$

For (2) $\implies$ (3), let $x \in A$. Then,

$$\begin{aligned} \mu(X \setminus \cup_{n=1}^{\infty} T^{-n}(A)) = 0 &\iff x \in T^{-n}(A) \text{ for some } n > 0 \\ &\iff T^{-n}(x) \in A. \end{aligned}$$

For (2) $\implies$ (4), with $\mu(X \setminus \cup_{n=1}^{\infty} T^{-n}(A)) = 0$, $\cup_{n=1}^{\infty} T^{-n}(A) = X \mod \mu$. If $A_0 \subset X$ with positive measure, then there should exist some integer $n > 0$ such that $\mu(T^{-n}(A) \cap A_0) > 0$. This would imply $T^{-n}(A) \cap A \neq \emptyset$.

For (4) $\implies$ (5), we will prove the contrapositive of the same. Let A and B be two measurable sets with positive measure in X. Suppose for all integers $n > 0$, $T^{-n}(A) \cap B = \emptyset$. This tells that there is no such $x \in B$ such that after some n^{th} iteration $T^{-n}(x) \in A$. Therefore, $\mu(T^{-n}(A) \cap B) = 0$ for all $n > 0$. □

With the help of above lemmas, we could now show that irrational rotations are ergodic. Given a measurable set, A and I (which could be an interval or a dyadic interval, but in general sense I should be from a semi-ring), we say I is $(1-\delta)$-full of A, where $0 < \delta < 1$, if $\lambda(A \cap I) > (1-\delta)\lambda(I)$.

Theorem 1.2.3: Irrational rotations are ergodic.

Proof. Let R_α be the rotational transformation by an irrational number α on $X = [0, 1)$. Let λ denote the Lebesgue measure. Let U_0 and V_0 be any sets of positive measure. Then invoking Lemma 0.1.7, there exists dyadic intervals I and J such that $\frac{3}{4}$-full of U_0 and J $\frac{3}{4}$-full of V_0, i.e.,

$$\begin{aligned} \lambda(U_0 \cap I) &\geq \frac{3}{4}\, \lambda(I) \\ \text{and,}\ \ \lambda(V_0 \cap J) &\geq \frac{3}{4}\, \lambda(J) \end{aligned}$$

If either I or J is bigger compared to the other (say J is found to be here), then take one of its half which is $\frac{3}{4}$-full of V_0. Compare this new set to I and if they do not have the same measure then repeat this halving process till the measures are of the same. Rename the set that matches the measure of I as J.

Let $U = U_0 \cap I$ and $V = V_0 \cap J$. Suppose $I = [a, b)$ and $J = [c, d)$ such that $a \leq d$. Note that both I and J are contained in $[0, 1)$.

Looking at the orbit of b, i.e., $\left\{R_\alpha^n(b)\right\}_{n\geq 0}^{\infty}$, we know it is dense in $[0, 1)$. So there exist an $n \geq 0$ such that ,

$$d - \frac{d-c}{4} < R^n(b) < d$$

Therefore,

$$\lambda(R^n(J \cap J) > \frac{3}{4}\ \lambda(J).$$

And from Lemma 0.2.6,

$$\begin{aligned}
\lambda(R^n(U) \cap V) &\geq \lambda(R^n(I) \cap J) - \lambda(I \setminus U) - \lambda(J \setminus V) \\
&\geq \frac{3}{4}\ \lambda(J) - \frac{1}{4}\ \lambda(I) - \frac{1}{4}\ \lambda(I) \\
&\geq \frac{1}{4}\ \lambda(J) > 0,\ \text{using the fact that } \lambda(I) = \lambda(J) \\
&\geq \frac{1}{4}(J) > 0.
\end{aligned}$$

Therefore, R_α is ergodic.

□

Definition 1.2.4: For any measure-preserving transformation T, if for all integers $n \geq 0$, T^{-n} is ergodic then T is a **totally ergodic** transformation.

Proposition 1.2.5: Let T be an invertible measure-preserving transformation on a non-atomic σ-finite measure space $(X, \mathcal{F}, \mu)$. If T is ergodic then T is recurrent.

Proof. Suppose T is not recurrent. Then there exists a set U of positive measure such that

$$\mu(T^{-n}(U) \cap U) = 0\ \forall\ n > 0.$$

$$\implies \mu(\cup_{n=-\infty, n\neq 0}^{\infty}(T^{-n}(U) \cap U) = 0.$$

Define,

$$\begin{aligned}
W &:= U \setminus \cup_{n=-\infty, n\neq 0}^{\infty}(T^{-n}(U) \cap U) \\
\implies \mu(W) &= \mu(U) > 0.
\end{aligned}$$

As a consequence of how we defined W, $\forall\ m \neq n, T^m(W) \cap T^n(W) = \emptyset$. As X is non-atomic, we can always find a subset $B \subset W$ such that $0 < \mu(B) < \mu(W)$. Define $B^* := \cup_{n=-\infty}^{\infty} T^{-n}(B)$. Note that B is T-invariant but also have $\mu(B^*) > 0$ and $\mu((B^*)^c) > 0$. This contradicts T being ergodic. Therefore T is recurrent. □

1.3 Eigen Values and Eigen Functions

Definition 1.3.1: Let $T : X \to X$ be a measure-preserving transformation for a probability space $(X, \mathcal{F}, \mu)$. An **eigen value** of T is the number $\lambda \in \mathbb{C}$ if there is a non-zero almost everywhere function $f \in L^2(X, \mu\mathbb{C})$ such that

$$f(T(x)) = \lambda f(x),\ a.e.$$

Such a function f is called an eigen function or eigen vector corresponding to λ.

Lemma 1.3.2: The eigen values lie in the unit circle of $\mathbb{C}$, in other words,$|\lambda| = 1$.

Proof. Let λ be an eigen value. Then λ must satisfy $f(T(x)) = \lambda f(x) f$, a.e., and the eigen function $f \in L^2(X, \mu, \mathbb{C})$. Using the fact that T is measure-preserving,

$$\begin{aligned}
\int |f \circ T|^2\ d\mu &= \int |\lambda|^2 |f|^2\ d\mu \\
\int |f|^2 \circ T\ d\mu &= |\lambda|^2 \int |f|^2 d\mu \\
\int |f|^2\ d\mu &= |\lambda|^2 \int |f|^2\ d\mu
\end{aligned}$$

With $\int |f|^2 \ d\mu \neq 0$,

$$\implies |\lambda|^2 = 1$$

$\square$

Lemma 1.3.3: Let $T : X \to X$ be a measure-preserving transformation for a space $(X, \mathcal{F}, \mu)$. T is ergodic if and only if for all measurable functions $f : X \to \mathbb{R}$, if f is T-invariant (i.e., $f(x) = f(T(x))$, a.e.) then f is constant a.e.

Proof. Suppose T is ergodic but the T-invariant function f is not constant a.e. As f is not constant there is some number α such that the sets $\{x : f(x) < \alpha\}$ and $\{x : f(x) > \alpha\}$ have positive measures. But T-invariance gives,

$$T^{-1}(\{x : f(x) < \alpha\}) = \{x : f(x) < \alpha\}$$

and

$$T^{-1}(\{x : f(x) > \alpha\}) = \{x : f(x) > \alpha\}$$

Also note that these two sets are disjoint. But this leads to a contradiction as T being ergodic, both T-invariant sets cannot have positive measures. There f must be constant a.e.

For the converse, for any T-invariant set $A \in \mathcal{F}$, the characteristic function $\mathbb{1}_A$ is T-invariant as,

$$\mathbb{1}_A(T(x)) = \mathbb{1}_{T^{-1}(A)}(x) = \mathbb{1}_A(x)$$

Recall that $\mathbb{1}_A \in L^p(X, \mu)$. So $\mathbb{1}_A$ is constant. This implies either,

$$\mu(A) = 0 \text{ or } \ \mu(A^c) = 0$$

So T is ergodic.

$\square$

Theorem 1.3.4: Let $T : X \to X$ be a measure-preserving transformation for a probability space $(X, \mathcal{F}, \mu)$. If T is ergodic and f is an eigen function, then $|f|$ is a constant a.e.

Proof. Let λ be the eigen value such that f is its eigen function. Then,

$$f(T(x)) = \lambda f(x), \text{ a.e.}$$

From Lemma 1.3.2, $|\lambda| = 1$. So,

$$\begin{aligned} |f| \circ T = |f \circ T| &= |\lambda \ f| \\ &= |\lambda| \ |f| \\ &= |f| \end{aligned}$$

So from Lemma 1.3.3, $|f|$ must be a constant a.e. $\square$

Chapter 2

The Ergodic Theorem

In statistical mechanics, upon its inception, an important question for scientists were to find a relation between the space-average and time-average of a dynamical system. The solution to this question was the Ergodic Theorem, proved in 1931 by G D Birkhoff. Around the same time, J Von Neumann also proved the theorem which is known as the mean ergodic theorem. We will prove it after showing Birkhoff's theorem.

In its simplest form, for an ergodic measure-preserving transformation, T on a probability space $(X, \mathcal{F}, \mu)$,

$$\lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N-1} \mathbb{1}_A(T^n(x)) = \mu(A)$$

for all measurable sets A in X and for each $x \in X$ outside a measure zero set in X.
The time-average of x of a dynamical system $(X, \mathcal{F}, \mu, T)$ is the average number of times the images of x under T falls in A. The space-average is obviously the measure of A. And here, the Birkhoff's theorem says that both are the same.

2.1 Birkhoff's Ergodic Theorem

Theorem 2.1.1: (Birkhoff's Ergodic Theorem) Let T be a measure-preserving transformation on a probability space $(X, \mathcal{F}, \mu)$. If $f : X \to \mathbb{R}$ is an integrable function then,

1. $\tilde{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$ exists for all $x \in X \setminus N$ where N is some null set depending on f.

2. $\tilde{f}(T(x)) = \tilde{f}(x)$, a.e.

3. For any T-invariant measurable set A,

$$\int_A f d\mu = \int_A \tilde{f} d\mu$$

   and if T is ergodic then,

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

The proof of the theorem is long and involved. Proving the existence of the limit is the hardest out of the 3 above statements. So we will prove it at the end of this section. For now let us assume that the limit,$\tilde{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$ exists for all $x \in X \setminus N$.

For the proof, we will define the following notations,

$$f_n(x) := \sum_{i=0}^{n-1} f(T^i(x)) \quad ;\text{ n} \geq 1$$

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

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

Lemma 2.1.2: $f_*(x)$ and $f^*(x)$ are T-invariant.

Proof. Observe that

$$\begin{aligned} f_n(T(x) &= \sum_{i=0}^{n-1} f(T^i(T(x))) \\ &= \sum_{i=0}^{n-1} f(T^{i+1}(x)) \\ &= f_{n+1}(x) - f(x) \end{aligned}$$

So with,

$$\frac{1}{n} f_n(T(x)) = \frac{n+1}{n} . \frac{1}{n+1} f_{n+1}(x) - \frac{1}{n} f(x) \tag{2.1}$$

taking lim inf on both sides ,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} f_n(T(x)) &= \liminf_{n\to\infty} \frac{n+1}{n} . \liminf_{n\to\infty} \frac{1}{n+1} f_{n+1}(x) - f(x) . \liminf_{n\to\infty} \frac{1}{n} \\ &= 1 \ . \liminf_{n\to\infty} \frac{1}{n+1} f_{n+1}(x) - f(x) \ . \ 0 \\ &= \liminf_{n\to\infty} \frac{1}{n+1} f_{n+1}(x) \\ &= \liminf_{n\to\infty} \frac{1}{n} f_n(x) \end{aligned}$$

which means $f_*(T(x)) = f_*(x)$. Similarly if we take lim sup on the both sides of **eqn. 2.1**, $f^*(T(x)) = f^*(x)$. □

Thus this lemma becomes the proof to the second claim of the theorem 2.1.1, which says about the T-invariant of $\tilde{f}$ (Note that we have assumed that $\tilde{f}(x)$ exists).
Now we wish to prove the claims of the Theorem 2.1.1 where the transformation T is assumed to be ergodic. For this the following two lemmas are needed.

Lemma 2.1.3: Let $g : X \to \mathbb{R}$ be a measurable function and $p \in \mathbb{Z}$ with $p \geq 1$. Suppose there is a measurable function, $\tau : X \to \{1, 2, 3, ..., p\}$ such that,

$$g_{\tau(x)} = \sum_{i=1}^{\tau(x)-1} g(T^{-i}(x)) \leq 0$$

then for all $n > p$,

$$g_n(x) \leq \sum_{i=n-p}^{n-1} |g(T^i(x))|$$

Proof. Looking at the orbit of x from x to $\tau(x)-1$, their sum of the images under g, i.e., $\sum_{i=o}^{\tau(x)-1} g(T^i(x)) \leq 0$. Hence, $g_n(x) \leq \sum_{i=\tau(x)}^{n-1} g(T^i(x))$. Also with, $y = T^{\tau(x)}(x)$ to $T^{\tau(y)-1}(y)$, we have $\sum_{i=\tau(x)}^{\tau(y)-1} g(T^i(x)) \leq 0$, as a consequence of the assumption. Similarly we can find finite sets of consecutive terms such that those finite sums are less than or equal to zero. Then after removing such sums from $g_n(x)$, the number of terms remaining is at most p (as $\tau(x)$ is bounded by p). Thus the function $g_n(x)$ is bounded by the absolute value of the finite sums of the last remaining terms. We get,

$$g_n(x) \leq \sum_{i=\tau(x)}^{n-1} |g(T^i(x))|$$

$\square$

Lemma 2.1.4: Let $p \geq 1$ be an integer and $f : X \to \mathbb{R}$ be a measurable function.

1. Let $E_p := \{x \in X | f_n(x) \geq 0$, for all n,$1 \leq n \leq p\}$. Then for all integers $n \geq p$ and for a.e., $x \in X$,

$$f_n(x) \leq \sum_{i=0}^{n-1} f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x)) + \sum_{i=n-p}^{n-1} |f(T^i(x))|$$

2. For each real number r define,

$$E_p^r = \{x \in X : \frac{1}{n}\sum_{i=o}^{n-1} f(T^i(x)) \geq r \text{ for all} 1 \leq n \leq p\}$$

then,

$$\int f d\mu \leq \int_{E_p^n} f d\mu + r(1 - \mu(E_p^r))$$

Proof. For the first part of the lemma, we wish to invoke the previous lemma over to the function, $g(x) = f(x) - f(x) \ . \ \mathbb{1}_{E_p}(x)$. For this we need to construct a function $\tau(x)$ that would satisfy the hypothesis of Lemma 2.1.3 for g. Let,

$$\tau(x) := \begin{cases} 1 & \text{, if } x \in E_p \\ \min\{1 \leq k \leq p : f_k(x) < 0\} & \text{, if } x \notin E_p \end{cases}$$

Suppose $x \in E_p$, then $\tau(x) = 1$ and $g_{\tau(x)} = g(x)$. And note that $f(x) \ . \ \mathbb{1}_{E_p}(x) = f(x)$. We get $g_\tau(x) = 0$. Suppose $x \notin E_p$, then $f_\tau(x) < 0$. Note that if for an $i \in \mathbb{N}, T^i(x) \notin E_p$, the term $f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x)) = 0$. And if $T^i(x) \in E_p$ then $f(T^i(x)) = f_1(T^i(x)) \geq 0$. Hence $f_{\tau(x)}(x) \leq \sum_{i=0}^{\tau(x)-1} f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x))$, a.e. With $\tau(x)$ satisfying the hypothesis of Lemma 2.1.3, we can apply the result of the same here, which gives, $g_n(x) \leq \sum_{i=o}^{n-1} |g(T^i(x))|$. From how g was defined,

$$f_n(x) - \sum_{i=0}^{n-1} f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x)) \leq \sum_{i=n-p}^{n-1} |f(T^i(x)) - f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x))|$$

But,

$$\sum_{i=0}^{n-1} |f(T^i(x)) - f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x))| \leq \sum_{i=n-p}^{n-1} |f(T^i(x))|$$

Therefore,

$$f_n(x) \leq \sum_{i=0}^{n-1} f(T^i(x)) \ . \ \mathbb{1}_{E_p}(T^i(x) + \sum_{i=n-p}^{n-1} |f(T^i(x)|$$

This thus proves the first part of the lemma. For the other part, we defined, $E_p^r := \{x : (f-r)_n(x) \geq 0$, for all$1 \leq n \leq p\}$. Applying the first part of the lemma to $f - r$,

$$(f-r)_n(x) \leq \sum_{i=1}^{n-1} (f-r)(T^i(x)) \ . \ \mathbb{1}_{E_p^r}(T^i(x)) + \sum_{i=n-p}^{n-1} |(f-r)(T^i(x))|$$

T is measure-preserving, so integrating on both sides of the above inequality will give,

$$\int (f-r)_n(x)d\mu \leq \int \left[\sum_{i=0}^{n-1}(f-r)(T^i(x)) \ . \ \mathbb{1}_{E_p^r}(T^i(x)) + \sum_{i=n-p}^{n-1} |(f-r)(T^i(x))|\right]$$

$$\int \sum_{i=0}^{n-1} f(T^i(x)) - nr \leq \int_{E_p^r} \left[\sum_{i=0}^{n-1} f(T^i(x)) - nr\right] d\mu + \int \sum_{i=n-p}^{n-1} |(f - r_(T^i(x))|d\mu$$

$$n\int (f-r)d\mu \leq n\int_{E_p^r}(f-r)d\mu + \sum_{i=n-p}^{n-1}\int |f-r|d\mu$$

taking m $\to \infty$

,

$$\int (f-r)d\mu \leq \int_{E_p^r}(f-r)d\mu + \lim_{n\to\infty}\frac{1}{n}\sum_{i=n-p}^{n-1}\int |f-r|d\mu$$

We get,

$$\int (f-r)d\mu \leq \int_{E_p^r}(f-r)d\mu$$

On rearrangement,

$$\int f d\mu \leq \int_{E_p} f d\mu + (1-\mu(E_p^r))^r.$$

□

With these lemmas, we can now proceed to prove the Birkhoff's theorem given that the transformation T is ergodic.

Proof. **(of Theorem 2.1.1, given that T is ergodic)** We begin with proving the following claim,

$$\int f d\mu \leq f_*(x) \text{ , a.e.} \tag{2.2}$$

Let $A := \{x : f_*(x) < \int f d\mu\}$. It is enough to show that $\mu(A) = 0$ to prove the claim. Suppose $\mu(A) > 0$, then we can write,

$$A = \bigcup_{r\in\mathbb{Q}}\{x : f_*(x) < r < \int f d\mu\}$$

Let $C_r = \{x : f_*(x) < r < \int f d\mu\}$. As $\mu(A) > 0$, $\exists$ $r \in \mathbb{Q}$ such that $\mu(C_r) > 0$. Note that C_r is T-Invariant because f_* is T-invariant. Hence with T being ergodic here, $\mu(C_r) = 1$. Recall that we have defined, $E_p^r = \{x : \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x) \geq r$, for all $1 \leq n \leq p\}$.

$$\mu(C_r) = 1 \implies \mu(\cap_{p=1}^{\infty} E_p^r) = 0$$

So,

$$\lim_{p\to\infty}\mu(E_p^r) = 0$$

But,

$$\int f d\mu \leq \int_{E_p^r} f d\mu + r(1-\mu(E_p^r)) \implies \int f d\mu \leq r.$$

This contradicts the choice of r. Therefore $\mu(A) = 0$.
With eqn.2.2 proved, applying the same on the function $' - f'$,

$$\int -f d\mu \leq \liminf_{n\to\infty} \frac{1}{n} \sum_{i=o}^{n-1} -f(T^i(x)) \text{ , a.e.}$$

or,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \leq \int f d\mu \text{ , a.e.} \tag{2.3}$$

Thus from equations 2.2 and 2.3,

$$\implies \int f d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \text{ , a.e.}$$

□

Theorem 2.1.5: Let T be a finite measure-preserving transformation on a probability space $(X, \mathcal{F}, \mu)$. T is ergodic if and only if for all measurable sets A,B,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cup B) = \mu(A)\mu(B) \tag{2.4}$$

Proof. Suppose the transformation T is ergodic, and let A and B be two measurable sets in X. We know that $\mathbb{1}_A$ is integrable. By applying the Birkhoff's Ergodic theorem here,

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

Then,

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

But,

$$|\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(T^i(x)) \ . \ \mathbb{1}_B(x)| \leq 1 \text{ , a.e.}$$

So we can invoke the dominated convergence theorem, which would give,

$$\begin{aligned}
\lim_{n\to\infty} \int \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(T^i(x)) \ . \ \mathbb{1}_B(x) d\mu &= \int \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(T^i(x)) \ . \ \mathbb{1}_B(x) d\mu(x) \\
&= \int \mu(A) \ . \ \mathbb{1}_B(x) d\mu \\
&= \mu(A)\mu(B)
\end{aligned}$$

But note that,

$$\begin{aligned}
\int \frac{1}{n} \sum_{i=0}^{n-1} (T^i(x)) \ . \ \mathbb{1}_B(x) d\mu &= \frac{1}{n} \sum_{i=0}^{n-1} \int \mathbb{1}_{T^{-i}(A) \cap B}(x) d\mu \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B)
\end{aligned}$$

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

For the converse, consider an T-invariant set, A. Then clearly $\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}(A)\cup A) = \mu(A)$. On replacing B with this set, A into the eqn.2.4,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(T^{-i}(A)\cup A) = \mu(A)\mu(A)$$
$$\implies \mu(A) = \mu(A)\mu(A)$$
$$\implies \mu(A) = 1 \text{ or } 0$$

Therefore, T is ergodic. □

To prove the Birkhoff's theorem for measure-preserving transformations, we would need the Maximal Ergodic Theorem. The following lemma will help us in proving this theorem.

Lemma 2.1.6: Let f be a measurable function and let $p \geq 1$ be an integer. Define,

$$G_p = \{x \in X : f_m(x) > 0 \text{ , for some } m, 1 \leq m \leq p\}$$

Then for all integers $n \geq p$ and for a.e, $x \in X$,

$$(f^+)_n(x) \leq \sum_{i=0}^{n-1} f(T^i(x)) \ . \ \mathbb{1}_{G_p}(T^i(x)) + \sum_{i=n-p}^{n-1} |f(T^i(x))|$$

Proof. Recall that $f^+(x) := \max\{0, f(x)\}$. The strategy to prove this lemma is by using the previous Lemma 2.1.3 over to the function, $g(x) = f^+(x) - f(x) \ . \ \mathbb{1}_{G_p}(x)$ and define the τ function as,

$$\tau(x) = \begin{cases} 1 & \text{, if } x \notin G_p \\ \min\{1 \leq k \leq p : f_k > 0\} & \text{, if } x \in G_p \end{cases}$$

But the hypothesis in Lemma 2.1.3 is needed to shown with the g and τ functions we defined, which is same as showing $f^+_{\tau(x)}(x) \leq \sum_{i=0}^{\tau-1} f(T^i(x)) \ . \ \mathbb{1}_{G_p}(T^i(x))$, a.e. For this, we consider two cases.

Case 1, $x \notin G_p$**:** Then,

$$\tau(x) = 1$$
$$f^+_\tau = f^+$$
$$\text{and, } f_n \leq 0, \ \forall n \leq p$$

Therefore, $f^+_n \leq 0$, by the definition of G_p.

Case 2, $x \in G_p$**:** Then, $\exists\ 1 \leq m \leq p$ such that, $\tau(x) = m$. So $f^+_m(x) > 0$. It can be observed that with $f^+_{\tau(x)} = f^+_m(x) = \sum_{i=0}^{m-1} f^+(T^i(x))$ then $f^+_{m-1}(T(x)) = \sum_{i=0}^{m-2} f(T^{i+1}(x)) > 0$.

$$\implies T(x) \in G_p$$

Similarly, $T^i(x) \in G_p$, $\forall\, i \leq m-1$. Therefore, $f^+_m(x) = \sum_{i=0}^{m-1} f^+(T^i(x)) = \sum_{i=0}^{m-1} f(T^i(x)) \ . \ \mathbb{1}_{G_p}(T^i(x))$. Now we can invoke Lemma 2.1.5 and we are done. □

Lemma 2.1.7 (Maximal Ergodic Theorem): Let $f : X \to \mathbb{R}$ be an integrable function and define,

$$G(f) = \{x \in X : f_n(x) > 0 \text{ , for some } n > 0\}$$

Then,

$$\int_{G(f)} f \geq 0$$

Proof. Let $p \geq 1$ be an integer and define,

$$G_p = \{x \in X : f_n(x) > 0 \text{ , for some } n, 1 \leq n \leq p\}$$

On integrating following the inequality,

$$f_n^+(x) \leq \sum_{i=0}^{n-1} f(T^i(x)) \ . \ \mathbb{1}_{G_p}(T^i(x)) + \sum_{i=n-p}^{n-1} |f(T^i(x))|$$

and using the fact that T is measure-preserving shows that,

$$n \int f^+ d\mu \leq n \int_{G_p} f d\mu + \sum_{i=n-p}^{n-1} \int |f| d\mu$$

On dividing by n and taking $n \to \infty$,

$$\int f^+ d\mu \leq \int_{G_p} f d\mu$$

$$\text{or, } \int_{G_p} f d\mu \geq 0 \text{ , for each } p \geq 1$$

It is clear from the definition of $G(f)$ that,

$$G(f) = \bigcup_{p>0} G_p$$

Then applying the dominated convergence theorem,$\int_G f d\mu < \infty$ and $\int_G f d\mu \geq \int_{G_m} f d\mu$, for some $m > 0$. Therefore, $\int_G f d\mu \geq 0$. □

Proof of the existence of $\tilde{f}$ in Birkhoff's theorem (2.1.1):
Choose r and $s \in \mathbb{R}$. Define,

$$A(r, s) := \{x \in X : f(_*(x) < r < s < f^*(x)\}$$

If we can show that the measure of this set, i.e., $\mu(A(r, s)) = 0$, for all r, s then $f_* = f^*$, a.e. This would mean the limit exists. It is already been shown that f^* and f_* are T-invariant, so in the similar way, the sets $A(r, s)$ are too T-invariant. Now if we restrict the transformation T to a set $A(r, s)$, looking at those $x \in A(r, s)$ that fall in $G(f - s) = \{x \in A(r, s) : f_m(x) > 0 \text{ , for some } m > 0\}$ are essentially all the elements in $A(r, s)$. This means we can apply the Maximal Ergodic theorem here over $A(r, s)$, i.e.

$$\int_{A(r,s)} (f - s) \ d\mu \geq 0$$

This implies

$$\int_{A(r,s)} f d\mu \geq s\mu(A(r, s)) \tag{2.5}$$

Applying the same theorem with $G(r - f)$,

$$r\mu(A(r, s)) \geq \int_{A(r,s)} f d\mu \tag{2.6}$$

From eqn. 2.5 and 2.6

$$r > s$$

But this contradicts $r < s$. Hence, $\mu(A(r,s)) = 0$.

$$\therefore f_* = f^* \text{ , a.e.}$$

Now that we have shown the existence of the limit, we can then proceed to prove the third statement of the theorem for a measure-preserving transformation T. For any function f, we can decompose it as $f = f^+ - f^-$. So for the further part of the proof it is enough to prove for functions f which are non-negative.
Let f be a non-negative bounded function, a.e. Let A be a T-invariant set. Then with dominated convergence theorem,

$$\begin{aligned}
\int_A \tilde{f} d\mu &= \int_A \lim_{n\to\infty} \sum_{i=0}^{n-1} f(T^i(x)) d\mu \\
&= \lim_{n\to\infty} \sum_{i=0}^{n-1} \int_A f(T^i(x)) d\mu \\
&= \lim_{n\to\infty} \sum_{i=0}^{n-1} \int_A f d\mu \\
&= \int_A f d\mu
\end{aligned}$$

Suppose f is now not a bounded function. Note that for any non-negative $f \in L^1$, using Fatou's lemma,

$$\begin{aligned}
||\tilde{f}||_1 &= \int |\lim_{n\to\infty} \frac{1}{n} f_n(x)| d\mu \\
&\leq \liminf \int |\frac{1}{n} f_n| d\mu \\
&= \int |f| d\mu = ||f||_1
\end{aligned}$$

Recall that for any integrable function f, $p \geq 1$, and for any $\epsilon > 0$, there exists a bounded function g such that $||f - g||_p < \epsilon$. Let f be approximated by such a bounded function g. Then,

$$\begin{aligned}
|\int_A f d\mu - \int_A \tilde{f} d\mu| &\leq |\int_A f d\mu - \int_A g d\mu| + |\int_A g d\mu - \int_A \tilde{f} d\mu| \\
&\leq \int_A |f - g| d\mu + |\int_A g d\mu - \int_A \tilde{g} d\mu + \int_A \tilde{g} d\mu - \int_A \tilde{f} d\mu| \\
&\leq \int_A |f - g| d\mu + |\int_A g d\mu - \int_A \tilde{g} d\mu| + |\int_A \tilde{g} d\mu - \int_A \tilde{f} d\mu| \\
\because g \text{ is bounded,} & \\
&\leq ||f - g||_1 + 0 + ||f - g||_1 \\
&\leq \epsilon + 0 + \epsilon \\
&= 2\epsilon
\end{aligned}$$

Hence for any T-invariant measurable set A,

$$\int_A f d\mu = \int_A \tilde{f} d\mu$$

This concludes the proof. □

2.2 Mean Ergodic Theorem

Let us recall the following lemma from functional analysis,

Lemma 2.2.1: Let S be a closed linear subspace of L^2. Then for each element f in L^2 there exists a unique element f_0 in S that is closest to f, that is ,

$$\inf \{||f-g||_2 : g \in L^2\} = ||f-f_0||_2$$

And if $P : L^2 \to S$ is defined by $P(f) = f_0$, then P is a transformation called the projection on S and every element f of L^2 can be written uniquely as,

$$f = P(f) + h,$$

where h is called the orthogonal complement of S.

Theorem 2.2.2: (Mean Ergodic Theorem) Suppose T is a measure-preserving transformation on a probability space $(X, \mathcal{F}, m)$. Let $I = \{g \in L^2 : g \circ T = g, a.e.\}$ be the subspace of the T-invariant functions and let P be the projection from L^2 to I. Then for any $f \in L^2$, the sequence, $\frac{1}{n}\sum_{i=o}^{n-1} f \circ T^i$ converges in L^2 to $P(f)$.

Proof. We would desire to decompose the L^2 space into two closed spaces of which one is the subspace I as defined in the theorem. It can be observed that for any $f \in I, f \circ T^n = f$, a.e., so,

$$\frac{1}{n}\sum_{i=o}^{n-1} f \circ T^i = f = P(f)$$

Let $B = \{f : f = g \circ T - g, g \in L^2\}$. Observe that for $f = g \circ T - g$,

$$\begin{aligned}\sum_{i=o}^{n-1} f \circ T^i &= \sum_{i=o}^{n-1} g \circ T^{i+1} - g \circ T^i \\ &= g \circ T^n - g\end{aligned}$$

With this, now it can be shown that the sequence, $\frac{1}{n}\sum_{i=o}^{n-1} f \circ T^i$ converges to zero for any function $f \in \overline{B}$. For this, suppose $\{f_i\}$ be a sequence of functions in B converging to a function $f \in L^2$. Then,

$$\begin{aligned}||\frac{1}{n}\sum_{i=o}^{n-1} f \circ T^i||_2 &= ||\frac{1}{n}\sum_{i=o}^{n-1}(f - f_j + f_j) \circ T^i||_2 \\ &\leq \frac{1}{n}\sum_{i=o}^{n-1} ||(f-f_j) \circ T^i||_2 + \frac{1}{n}\sum_{i=o}^{n-1} ||f_j \circ T^i||_2 \\ &= \frac{1}{n}\sum_{i=o}^{n-1} ||(f-f_j)||_2 + \frac{1}{n}\sum_{i=o}^{n-1} ||f_j \circ T^i||_2 \\ &= 0 + \frac{1}{n}\sum_{i=o}^{n-1}(\int |f_j \circ T^i|^2)^{\frac{1}{2}} \\ &= \frac{1}{n}\sum_{i=o}^{n-1}(\int |g_j \circ T^{i+1} - gT^i|^2)^{\frac{1}{2}} \\ &= \frac{1}{n}\sum_{i=o}^{n-1}(\int (g-g)^2)^{\frac{1}{2}} \\ &= \frac{1}{n}\sum_{i=o}^{n-1}(0)^{\frac{1}{2}} = 0\end{aligned}$$

Now to show the orthogonal complement of $\overline{B}$ is I, take $f \in I$ and $h \in B$. Then it can be seen that $h = g \circ T - g$ for $g \in L^2$. And the inner product of h and g is,

$$\begin{aligned}(f, h) = (f, g \circ T - g) &= (f, g \circ T) - (f, g) \\ &= (f \circ T, g \circ T) - (f, g) \\ &= (f, g) - (f, g) \\ &= 0\end{aligned}$$

Let $h \in \overline{B}$. Let h_k be a sequence in B converging to h in norm, i.e., $\lim_{n\to\infty}(h, h_k) = 0$. Also note that $\forall k > 0, (f, h_k) = 0$. Therefore with $(f, h) = 0$, for all $h \in \overline{B}$, I is orthogonal to $\overline{B}$.
Let $f \in \overline{B}^{\perp}$. Then for every $g \in L^2$,

$$(f, g \circ T - g) = 0$$

So,

$$\begin{aligned}(f, g) &= (f, g \circ T) \\ &= \int f \, . \, \bar{g} \circ T d\mu \\ &= \int f \circ T^{-1} \, . \, \bar{g} d\mu \\ &= (f \circ T^{-1}, g)\end{aligned}$$

As $(f - f \circ T^{-1}, g) = 0$ for all $g \in L^2$,

$$f = f \circ T^{-1}, \text{ a.e., or } \ f \circ T = f \ \text{ , a.e.}$$

$$\implies I = \overline{B}^{\perp}$$

Let P being the projection map onto the subspace I, using the above Lemma 2.2.1, every element $f \in L^2$ can be uniquely written as $f = P(f) + h$, where $h \in I^{\perp} = \overline{B}$. The existence of the limit of the sequence formed by f is ensured by the Ergodic Theorem and thus it converges in L^2 to $P(f)$.

□

Lemma 2.2.3 (Scheffé's Lemma): Let f_n and f be non-negative functions in L^1, where $n \geq 1$, such that $f_n \to f$ a.e. Then f_n converges to f in L^1 if and only if $\int f_n d\mu$ converges to $\int f d\mu$.

Proof. If f_n converges to f in L^1, it means $||f_n - f||_1 \to 0$. This clearly implies that $\int f_n d\mu$ converges to $\int f d\mu$. For the converse, With f_n and f being non-negative, $(f_n - f^-)^- \leq f$. Then applying the dominated convergence theorem here,

$$\begin{aligned}\lim_{n\to\infty} \int (f_n - f)^- d\mu &= \int \lim_{n\to\infty} (f_n - f)^- d\mu \\ &= 0, \text{ as } f_n \to f \text{ , a.e}\end{aligned}$$

Let $A_n = \{x \in X : f_n(x) \geq f(x)\}$. Then,

$$\begin{aligned}\int (f_n - f^+) d\mu &= \int_{A_n} (f_n - f) d\mu \\ &= \int_X (f_n - f) d\mu - \int_{A_n^c} (f_n - f) d\mu\end{aligned}$$

On A_n^c,

$$\begin{aligned}(f_n - f)^- &= -\min\{f_n - f, 0\} \\ &= -(f_n - f)\end{aligned}$$

So,

$$\lim_{n\to\infty} |\int_{A_n^c} f_n - f d\mu| \leq \lim_{n\to\infty} \int_X (f_n - f)^- d\mu$$

This implies,

$$\begin{aligned}\int (f_n - f)\ d\mu &= \int_X (f_n - f) d\mu + 0 \\ &= 0\end{aligned}$$

Therefore,

$$\begin{aligned}\int |f_n - f| d\mu &= \int (f_n - f)^+ + (f_n + f)^- d\mu \\ &= 0\end{aligned}$$

This shows that, f_n converges to f in L^1. □

The following theorem is also known as the Mean Ergodic Theorem in L^2.

Theorem 2.2.4: Suppose T is a measure-preserving transformation on a probability space $(X, \mathcal{F}, \mu)$. Then for any $f \in L^1$, the sequence, $f_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i$ converges in L^1 norm to a function $\tilde{f} \in L^1$.

Proof. Consider a non-negative function $f \in L^1$. Then from the way f_n is defined, $\frac{1}{n} f_n(x)$ are non-negative. Clearly $\frac{1}{n} f_n$ converges to $\tilde{f} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x)$ as $n \to \infty$, due to Birkhoff's ergodic theorem. Also,

$$\begin{aligned}\int \frac{1}{n} f_n d\mu &= \frac{1}{n}\sum_{i=0}^{n-1}\int f \circ T^i(x) d\mu \\ &= \frac{1}{n}\ .\ n \int f d\mu \\ &= \int f d\mu\end{aligned}$$

Hence from Scheffé's lemma, $\frac{1}{n} f_n$ converges to $\tilde{f}$ in L^1.

□

2.3 Examples

We can now see two examples where Birkhoff's theorem is applied. The first one is with the irrational rotation map.

Theorem 2.3.1 Let α be an irrational number and $R = R_\alpha$ be an irrational rotation map as defined in Example 1.2.3. Then for every interval $I \subset [0, 1)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_I(R^i(x)) = \lambda(I) \quad \text{for all } x \in [0, 1)$$

Proof. We have shown that with α being irrational, R is ergodic. Hence by the ergodic theorem, for any interval I, there is a null set N(I) so that for all $x \in [0, 1) \setminus N(I)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_I(R^i(x)) = \lambda(I) \tag{2.7}$$

With the collection of dyadic intervals being a countable collection of sets, let N be the union of all null sets. Then there exists a point $x \in [0, 1)$ so that eqn.2.7 holds for all such dyadic intervals.
Let I be an arbitrary interval. For any $\epsilon > 0$, there exists dyadic intervals, K, J such that,

$$J \subset I \subset K \text{ and } \lambda(K \setminus J) < \epsilon$$

From the nature of the intervals,

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_J(R^i(x)) \leq \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_I(R^i(x)) \leq \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_K(R^i(x))$$

Then,

$$\begin{aligned}
\lambda(J) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_J(R^i(x)) \\
&\leq \lim_{n\to\infty}\inf\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_I(R^i(x)) \\
&\leq \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_K(R^i(x)) \\
&= \lambda(J)+\epsilon
\end{aligned}$$

Similarly,

$$\begin{aligned}
\lambda(J) &\leq \lim_{n\to\infty}\sup\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_I(R^i(x)) \\
&\leq \lambda(J)+\epsilon
\end{aligned}$$

Hence for x,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}_I(R^i(x)) = \lambda(I)$$

To show this holds true for any $x \in [0,1)$, observe that with any interval I, we can always find an interval I' of same measure such that $R^i(0) \in I$ if and only if $R^i(x) \in I'$.

□

The second example is about normal numbers and Borel's theorem on normal numbers. We know for any $x \in [0,1)$, the corresponding binary expansion is of the form,

$$x = \sum_{i=0}^{\infty}\frac{a_i}{2_i}$$

where $a_i \in \{0,1\}$ are the digits of the binary expansion of x. Note that this expansion is unique for almost everywhere $x \in X$.

Definition 2.3.2: A number $x \in [0,1)$ is said to be normal in base 2 if the frequency of occurrence of the digit 0 and the frequency of occurrence of the digit 1 in the binary expansion equals $\frac{1}{2}$ each.

Theorem 2.3.3(Borel's Theorem on Normal Numbers): Almost everywhere $x \in [0,1)$ is normal in base 2.

Proof. Recall the doubling map, $T(x) = 2x \mod 1$ from Example 1.2.6. We have shown that T is measure-preserving (with the probability measure) and ergodic. Note that for $x = \sum_{i=0}^{\infty}\frac{a_i}{2^i}$,

$$\begin{aligned}
T^k(x) &= \sum_{i=0}^{\infty}2^k\frac{a_i}{2_i} \mod 1 \\
&= \sum_{i=0}^{\infty}\frac{a_{k+i}}{2^i}
\end{aligned}$$

Also see that,

$$a_1 = 0 \text{ if and only if } \ x \in [0, \frac{1}{2}) \text{ , and}$$
$$a_1 = 1 \text{ if and only if } \ x \in [\frac{1}{2}, 1)$$

$$\implies a_k + 1 = \begin{cases} 0 \text{ , if and only if } T^k(x) \in [0, \frac{1}{2}) \\ 1 \text{ , if and only if } T^k(x) \in [\frac{1}{2}, 1) \end{cases}$$

So,

$$\begin{aligned} \frac{1}{n} \text{ Cardinality } \Big\{0 \leq k \leq n : a_k = 0\Big\} &= \frac{1}{n} \text{ Cardinality } \Big\{1 \leq k \leq n : a_{k+1} = 0\Big\} \\ &= \frac{1}{n} \text{ Cardinality } \Big\{1 \leq k \leq n : T^k(x) \in [0, \frac{1}{2})\Big\} \end{aligned}$$

as $n \to \infty$,

$$\begin{aligned} &= \lim_{n \to \infty} \frac{1}{n} \sum_{k=o}^{n-1} \mathbb{1}_{[0, \frac{1}{2})}(T^k(x)) \\ &= \lambda([0, \frac{1}{2})) \\ &= \frac{1}{2} \end{aligned}$$

Similarly for $a_k = 1$ case, $\lambda([\frac{1}{2}, 1) = \frac{1}{2}$. $\square$

The same theorem can be shown for any base $k \in \mathbb{N}$ and here the frequency any $r \in \{1, 2, ..., k\}$ is $\frac{1}{k}$.

Chapter 3

Factors & Isomorphisms of Dynamical Systems

Formally a dynamical system consists of a measure space $(X, \mathcal{F}, \mu)$ and a measure preserving transformation T on it, denoted as $(X, \mathcal{F}, \mu, T)$. One can try to say that two dynamical systems $(X, \mathcal{F}, \mu, T)$ and $(X', \mathcal{F}', \mu', T')$ are isomorphic if they possess the same dynamical properties, such as if one is ergodic, then the other system isomorphic to this must be ergodic too. With this agenda, we can formally define isomorphism between dynamical systems. But before going into that, first we will define isomorphism between two measure spaces.

3.1 Factors & Isomorphisms

Definition 3.1.1: Consider two measure spaces, $(X, \mathcal{F}, \mu)$ and $(X', \mathcal{F}', \mu')$. They are said to be **isomorphic** if there exists measurable sets $X_0 \subset X$ and $X_0' \subset X'$ of full measure (which means $\mu(X \setminus X_0) = 0$ and $\mu(X' \setminus X_0') = 0$) and a map, $\phi : X_0 \to X_0'$ that is one-one and onto such that,

1. $A \in \mathcal{F}'(X_0)'$ if and only if $\phi^{-1}(A) \in \mathcal{F}(X_0)$
2. $\mu(\phi^{-1}(A)) = \mu'(A) \forall A \in S'(X_0)$

This isomorphism amongst measure spaces is also called measure-theoretically isomorphic or isomorphic mod 0.

Example 3.1.2: Let $(X, \mathcal{F})$ be a measurable space, where $X = [-1, 1]$ and $\mathcal{F}$ be the Lebesgue σ-algebra.Define a measure μ on $\mathcal{F}$ by $\mu(A) = \frac{1}{2}\lambda(A)$, where λ is the Lebesgue measure. It can be shown that the measure spaces, $(X, \mathcal{F}, \mu)$ and $(X, \mathcal{F}, \lambda)$ are isomorphic. For this, define,

$$\begin{aligned} \phi : [0, 1] &\to X \\ \phi(x) &= 2x - 1 \end{aligned}$$

Clearly ϕ is one-one and onto. and $\phi^{-1}(y) = \frac{y+1}{2}$. Also ϕ and ϕ^{-1} are measurable. Since ϕ is a composition of translation and dilation, it is measure-preserving, i.e.,

$$\begin{aligned} \lambda(\phi^{-1}(A)) &= \lambda(\frac{A+1}{2}) \\ &= \frac{1}{2}\lambda(A) \\ &= \mu(A) \text{ , for all } A \in \mathcal{F} \end{aligned}$$

Definition 3.1.3: Consider two finite measure-preserving dynamical systems, $(X, \mathcal{F}, \mu, T)$ and $(X', \mathcal{F}', \mu', T')$. They are said to be **isomorphic** if there exists measurable sets $X_0 \subset X$ and $X_0' \subset X'$ of full measure and a map, $\phi : X_0 \to X_0'$ that is one-one and onto such that $\forall A \in \mathcal{F}'(X_0')$,

1. $\phi^{-1}(A) \in \mathcal{F}(X_0)$
2. $\mu(\phi^{-1}(A)) = \mu'(A)$
3. $\phi(T(x)) = T'(\phi(x)), \ \forall x \in X_0$

The third property is also called the equivariance. It is illustrated as,

$$\begin{tikzcd} X_0 \arrow[r, "T"] \arrow[d, "\phi"] & X_0 \arrow[d, "\phi"] \\ X_0' \arrow[r, "T'"] & X_0' \end{tikzcd}$$

If a property is found to be invariant under isomorphism, then such a property is called a **dynamical property**.

Definition 3.1.3: Consider two finite measure-preserving dynamical systems, $(X, \mathcal{F}, \mu, T)$ and $(X', \mathcal{F}', \mu', T')$. $(X', \mathcal{F}', \mu', T')$ is said to be a **factor** of $(X, \mathcal{F}, \mu, T)$ if there exists measurable sets $X_0 \subset X$ and $X_0' \subset X'$ of full measure with,

$$T(X_0) \subset X_0, T'(X_0') \subset X_0'$$

and a map, $\phi : X_0 \to X_0'$ that is just onto such that $\forall A \in \mathcal{F}'(X_0')$,

1. $\phi^{-1}(A) \in \mathcal{F}(X_0)$
2. $\mu(\phi^{-1}(A)) = \mu'(A)$

Theorem 3.1.4: Let S be a factor of T. Then if T is ergodic then S is ergodic.

Proof. Consider two transformations, T and S on measure spaces, $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{F}, \nu)$. Let $\phi : X \to Y$ be the factor map. Choose a strictly S-invariant set $A \in Y$. See that

$$\begin{aligned} T^{-1}(\phi^{-1}(A)) &= \phi^{-1}(S^{-1}(A)) \\ &= \phi^{-1}(A) \end{aligned}$$

So $\phi^{-1}(A)$ is strictly T-invariant. That means $\mu(\phi^{-1}(A)) = 0$, or, $\mu((\phi^{-1}(A))^c) = 0$, as T is ergodic. But,

$$\mu(\phi^{-1}(A)) = \nu(A)$$

So,

$$\nu(A) = 0 \text{ or, } \ \nu(A^c) = 0$$

□

3.2 Induced Transformations

Definition 3.2.1: Let $(X, \mathcal{F}, \mu)$ be a σ-finite measure space and let T be a recurrent measure-preserving transformation. Then for every measurable set A of positive measure, there is a null set $N \subset A$ such that for all $x \in A \setminus N$, there is an integer $n = n(x) > 0$ with $T^n(x) \in A$. We call the smallest such n, the first return to A, defined by

$$n_A(x) = \min\{n > 0 : T^n(x) \in A\}$$

For all A of positive measure, $n_A(x)$ is defined, a.e., since T is recurrent. So we can define the induced transformation T_A as,

$$T_A(x) = T^{n_A(x)}(x) \ \text{ for a.e. } \ x \in A.$$

Preposition 3.2.2: Let $(X, \mathcal{F}, \mu)$ be a finite measure space with an invertible measure-preserving transformation T. If T is recurrent and $A \in X$ be a set of positive measure, then the induced transformation T_A is measure-preserving on A.

Proof. Define the set

$$A_i = \{x \in A : n_A(x) = i, i \geq 1\}$$

These sets are clearly disjoint and thus,

$$A = \bigsqcup_{i=1}^{\infty} A_i \mod \mu$$

Take a measurable set, $B \in A$. Then,

$$\begin{aligned}
\mu(T_A(B)) &= \mu(T_A(\bigsqcup_{i=1}^{\infty} B \cap A_i)) \\
&= \mu(\bigsqcup_{i=1}^{\infty} T_A(B \cap A_i)) \\
&= \sum_{i=1}^{\infty} \mu(T^i(B \cap A_i)) \\
&= \sum_{i=1}^{\infty} \mu(B \cap A_i) \\
&= \mu(\bigsqcup_{i=1}^{\infty} B \cap A_i) \\
&= \mu(B)
\end{aligned}$$

Therefore, T_A is measure-preserving on A.

□

Theorem 3.2.3: Suppose T is an invertible, recurrent, finite measure-preserving transformation over a measure space $(X, \mathcal{F}, \mu)$. Then for a measurable set $A \subset X$, the induced transformation T_A on A is ergodic if T is ergodic.

Proof. Consider two sets of positive measures in A, E and F. With T being ergodic and recurrent,

$$\mu(T^n(E) \cap F) > 0 \text{ for some integer } n > 0.$$

This means there is a point $x \in E$ such that for some $n \in \mathbb{Z}^+, T^n(x) \in F$. Let

$$\begin{aligned}
n_1 &= n_A(x) \\
n_2 &= n_A(T^{n_1}(x)) \\
n_3 &= n_A(T^{n_2}(x)) \\
&\vdots \\
n_k &= n_A(T^{n_{k-1}}(x))
\end{aligned}$$

where n_k is the first integer such that $T^{n_k}(x) \in F$. Recall that $T_A(x) := T^{n_A(x)}(x)$. Then with $T_A^n(x) = T^{n_k}(x)$,

$$\begin{aligned}
&\implies T_A^k(E) \cap F \neq \phi \\
&\implies T_A \text{ is ergodic.}
\end{aligned}$$

□

Chapter 4

Mixing

Just like mixing sugar into some water, or any two solutions in a container, the notion of mixing can be applied as an abstract concept over transformations. We would define this term to those under a certain transformation, the elements of the set will eventually be homogeneously distributed.

4.1 Defining Mixing

Definition 4.1.1: Let $(X, \mathcal{F}, \mu)$ be a probability space with a measure-preserving transformation, T. We say that T is mixing if for all measurable sets $A, B \in X$,

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B). \tag{4.1}$$

We know $\mu(X) = 1$. So re-writing eqn.4.1

$$\lim_{n\to\infty} \frac{\mu(T^{-n}(A) \cap B)}{\mu(B)} = \frac{\mu(A)}{\mu(X)}$$

This gives a clearer understanding to the definition of mixing. It says that after some $'n'$ transformations under T, the proportion of elements from B falling in A, is exactly the same as the proportion of A in the entire set, X.

For the remaining part, some notion of convergence, especially the Cesaro convergence, is needed to be recalled.Let $a_i(A, B) = \mu(T^{-i}(A) \cap B)$. Then the definition for mixing turns out to be,

$$\lim_{n\to\infty} a_i(A, B) = \mu(A)\mu(B)$$

Recall that for a bounded sequence, $\{a_i\}_{i>0}$, convergence of the same to a number can be shown by,

1. Convergence of sequences:
$$\lim_{i\to\infty} a_i - a = 0$$
2. Strong Cesaro convergence of sequences:
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i - a| = 0$$
3. Cesaro convergence of sequences:
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (a_i - a) = 0$$

Note that for a set of non-negative sequences, Cesaro and Strong Cesaro convergence are equivalent.

Lemma 4.1.2 Consider a bounded sequence $\{a_i\}_{i>0}$. Then,

1. Convergence of $\{a_i\}_{i>0}$ implies strong Cesaro convergence of $\{a_i\}_{i>0}$.

2. Strong Cesaro convergence of $\{a_i\}_{i>0}$ implies Cesaro convergence of $\{a_i\}_{i>0}$.

Proof. For (1), choose an arbitrary $\epsilon > 0$. Then with the sequence being convergent in the usual sense, $\exists\ K_1 > 0$ such that, for every $i > K_1$,

$$|a_i - a| < \frac{\epsilon}{2}$$

With the sequence $\{a_i\}_{i>0}$ being bounded, $\exists\ K_2 > 0$ such that, for every $n > K_2$,

$$\frac{1}{n}\sum_{i=o}^{K_1-1} |a_i - a| < \frac{\epsilon}{2}$$

Set $K = max\{K_1, K_2\}$. With $n > K$,

$$\begin{aligned}\frac{1}{n}\sum_{i=o}^{n-1} |a_i - a| &= \frac{1}{n}\sum_{i=o}^{K_1-1} |a_i - a| + \frac{1}{n}\sum_{i=K_1}^{K-1} |a_i - a| \\ &< \frac{\epsilon}{2} + (\frac{n-K_1}{n})\frac{\epsilon}{2} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon\end{aligned}$$

To prove (2), we just need to apply the triangle inequality property,

$$\sum_{i=0}^{n-1}(a_i - a) \leq \sum_{i=0}^{n-1} |a_i - a|$$

So,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1}(a_i - a) &\leq \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |a_i - a| \\ &= 0\end{aligned}$$

□

Definition 4.1.3: A set D of non-negative integer is said to be of **zero density** if,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=o}^{n-1} \mathbb{1}_D(i) = 0$$

And the set D is said to have a positive density, if $\limsup \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_D(i) > 0$. For a sequence of non-negative sequences $\{a_i\}_{i\geq 0}$ is said to have zero density if the set $\{a_i : i \geq 0\}$ has zero density. The following are some examples of sets with zero density.

1. Any finite set

2. $\{a_i = 2^i\}_{i\geq 0}$

3. $\{i \in \mathbb{N} : i \text{ is prime}\}$; this is a consequence of the Prime Number theorem which states that for a large number N, the probability that a random integer not greater than N is prime, is close to $\frac{1}{\log(N)}$.

Definition 4.1.4: For a set B of non-negative numbers if its complement B^c has zero density then B is said to be of **density one**.

Definition 4.1.5: A sequence $\{a_i\}$ of real numbers converges in density to a point if there exists a zero density set $D \in \mathbb{N}$ such that for all $\epsilon > 0$, there is an integer N and for every $i > N,\ i \in D$,

$$|a_i - a| < \epsilon$$

It can be denoted as,

$$\lim_{i\to\infty,\ i\notin D} a_i = a$$

Proposition 4.1.6: For a bounded sequence of non-negative real numbers, $\{b_i\}_{i\geq 0}$, it converges to 0 if and only if it converges in density to 0.

Proof. Suppose that the bounded sequence $\{b_i\}_{i\geq 0}$ converges in density to 0, outside a zero density set D. Let the sequence be bounded by M. That means,

$$\lim_{i\to\infty, i\notin D} b_i = 0$$

Then,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} b_i &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=0, i\in D}^{n-1} b_i + \lim_{n\to\infty} \frac{1}{n}\sum_{i=0, i\notin D}^{n-1} b_i \\
&\leq \lim_{n\to\infty} \frac{1}{n}\sum_{i=0, i\in D}^{n-1} M + \lim_{n\to\infty} \frac{1}{n}\sum_{i=0, i\notin D}^{n-1} b_i \\
&\leq M \lim_{n\to\infty} \frac{1}{n}\sum_{i=0, i\in D}^{n-1} \mathbb{1}_D(i) + \lim_{n\to\infty} \frac{1}{n}\sum_{i=0, i\notin D}^{n-1} b_i \\
&= 0,
\end{aligned}$$

Since D is a set with zero density and $\{b_i\}_{i\geq 0}$ converges in density.

For the converse, let $\{b_i\}_{i\geq 0}$ converges to 0 in D, i.e.,

$$\frac{1}{n}\sum_{i=0}^{n-1} b_i = 0$$

Suppose if there is some $\epsilon > 0$ such that $D_\epsilon := \{i \in \mathbb{N} : |b_i| > \epsilon\}$ has a positive density,say ω. Then the Cesaro limit,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} b_i &= \limsup \frac{1}{n}\sum_{i=0}^{n-1} b_i \\
&> \omega\ \epsilon \\
&> 0
\end{aligned}$$

This contradicts to the Cesaro convergence of $\{b_i\}_{i\geq 0}$ to zero. So all such set D_ϵ for an arbitrary ϵ must have zero density. And clearly the sequence converges outside such a set, which can be written as,

$$\lim_{i\to\infty, i\notin D_\epsilon} b_i = 0$$

Thus the sequence,$\{b_i\}_{i\geq 0}$, converges in density.

□

4.2 Weak Mixing

Definition 4.2.1: A measure-preserving transformation T on a probability space is weakly mixing if,

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

Looking at the definition of strong Cesaro convergence, weak-mixing can be seen as the sequence $\{\alpha_i(A,B) := \mu(T^i(A)\cap B)\}_{i\geq 0}$ converging to $\mu(A)\mu(B)$ in strong Cesaro convergence.

Lemma 4.2.2: With a measure-preserving transformation T over a probability space, if,

1. T is weakly mixing then T is ergodic.
2. T is mixing then T is weakly mixing.

Proof. For (1), consider the sequence, $\{\alpha_i(A,B) := \mu(T^i(A) \cap B)\}_{i\geq 0}$. Suppose A be T-invariant. Then looking at the sequence, $\{\alpha_i(A,A^c) = \mu(T^i(A) \cap A^c)\}_{i\geq 0}$,

$$\mu(T^i(A)) = \mu(A), \text{ for all } i \geq 0$$

Therefore,

$$\begin{aligned}\mu(T^i(A) \cap A^c) &= \mu(A \cap A^c)\\ &= 0\end{aligned}$$

In short, $\alpha_i(A,A^c) = 0, \forall i \geq 0$. With T being weakly mixing,

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

Then,

$$\mu(A)\ \mu(A^c) = 0$$

This implies,

$$\mu(A) = 0 \text{ or, } \mu(A^c) = 0$$

Thus T is ergodic.

For (2), T being mixing tells that the sequence $\{\alpha_i(A,B)\}$ converges to $\mu(A)\mu(B)$. And from Lemma 4.1.2, we have shown that convergence implies strong Cesaro convergence. This proves that T is weakly mixing. □

Proposition 4.2.3: For a measure-preserving transformation T on a probability space the following are equivalent;

1. T is weakly mixing
2. For every measurable sets A, B, there is a set $D = D(A,B)$ with zero density such that,
$$\lim_{i\to\infty, i\notin D} \mu(T^i(A) \cap B) = \mu(A)\mu(B)$$
3. $\lim_{n\to\infty} \frac{1}{n} \sum_{i=o}^{n-1} (\mu(T^i(A) \cap B) - \mu(A)\mu(B))^2 = 0$, for all measurable sets A, B.

Proof. To show (1) $\iff$ (2), take a sequence $\{\alpha_i(A,B) := \mu(T^i(A) \cap B) - \mu(A)\mu(B)\}$. Then apply Proposition 4.1.6.
To show (2) $\iff$ (3), from (2),

$$\lim_{i\to\infty, i\notin D} \mu(T^i(A) \cap B) - \mu(A)\mu(B) = 0$$

Then,

$$\lim_{i\to\infty, i\notin D} (\mu(T^i(A) \cap B) - \mu(A)\mu(B))^2 = 0$$

Applying Proposition 4.1.6 here,

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

The converse is just taking the above steps in the reverse order as the Proposition 4.1.6 holds both ways. □

Definition 4.2.4: If for all measurable sets A, B of a measure space $(X, \mathcal{F}, \mu)$ with a measure-preserving transformation T there is an integer $n > 0$ such that

$$\mu(T^{-n}(A) \cap A) > 0 \text{ and } \mu(T^{-n}(A) \cap B) > 0,$$

then T is said to be **doubly ergodic**.

Proposition 4.2.5: For a probability space $(X, \mathcal{F}, \mu)$, let T be a measure-preserving transformation over it. If T is weakly mixing then T is doubly ergodic.

Proof. Consider two measurable sets of positive measure, say, $A, B \in \mathcal{F}$. Let T be weakly mixing. Then we can find sets $D_1(A, B)$ and $D_2(A, A)$, of zero density such that,

$$\lim_{i \to \infty, i \notin D_1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B)$$
$$\text{and}$$
$$\lim_{i \to \infty, i \notin D_2} \mu(T^{-i}(A) \cap A) = \mu(A)\mu(A)$$

Then we can find an i such that,

$$\mu(T^{-i}(A) \cap A) > 0$$
$$\text{and}$$
$$\mu(T^{-i}(A) \cap B) > 0,$$

Thus T is **doubly ergodic**.

□

Example 4.2.6: Now we can looking to how rotation transformations are not weakly mixing. To show this, first recall how we defined R in Example 1.2.3. Take two intervals, say, $I = [0, \frac{1}{4})$ and $J = [\frac{3}{4}, 1)$. The observe that for any $x \in I$ and $y \in J$, $|T^i(x) - T^j(y)|$ remains constant for every pair (x, y) we choose. This implies that there is no integer n, such that

$$\mu(T^{-n}(I) \cap I) > 0 \text{ and } \mu(T^{-n}(I) \cap J) > 0$$

So R is not doubly ergodic. Therefore from Proposition 4.2.5, R is not weakly mixing.

4.3 Approximations to determine Dynamical Properties

Recall how semi-rings and sufficient semi-rings are defined earlier in Chapter 0. With this we would show that it is enough to verify with all the elements of a sufficient semi-ring to determine dynamical properties such as ergodicity, mixing and weak-mixing. To show this, we would first need to prove some lemmas which would act as tools to prove the approximation of sufficient semi-rings.

Lemma 4.3.1: Let $A, B, E,$ and F be measurable sets of a probability space $(X, \mathcal{F}, \mu)$. Let T be a measure-preserving transformation on the same space. Then for any $n \in \mathbb{Z}$,

$$\mu\Big((T^{-n}(A) \cap B)\Delta(T^{-n}(E) \cap F)\Big) \leq \mu(A\Delta E) + \mu(B\Delta F)$$

Proof. The triangle inequality property gives,

$$\mu\Big((T^{-n}(A)\cap B)\Delta(T^{-n}(E)\cap F)\Big) \leq \mu\Big((T^{-n}(A)\cap B)\Delta(T^{-n}(A)\cap F)\Big) + \mu\Big((T^{-n}(A)\cap F)\Delta(T^{-n}(E)\cap F)\Big)$$

Using $(A \cap E)\Delta(B \cap E) = E \cap (A\Delta B)$,

$$\begin{aligned}
\mu\Big((T^{-n}(A)\cap)\ B)\Delta(T^{-n}(E) \cap F)\Big) &\leq \mu\Big((T^{-n}(A) \cap (B\Delta F)\Big) + \mu\Big(F \cap T^{-n}(A\Delta E)\Big) \\
&\leq \mu(B\Delta F) + \mu(T^{-n}(A \cap E)) \\
&= \mu(B\Delta F) + \mu(A \cap E)
\end{aligned}$$

□

Lemma 4.3.2: For a measure-preserving transformation T over a probability space, $(X, \mathcal{F}, \mu)$, suppose that for an arbitrary $\epsilon > 0$ and $A, B \in \mathcal{F}$, there exists $E, F \in \mathcal{F}$ such that, $\mu(A\Delta E) < \frac{\epsilon}{5}$ and $\mu(B\Delta F) < \frac{\epsilon}{5}$. If for $n > 0$,

$$\left|\mu(T^{-n}(E) \cap F) - \mu(E)\mu(F)\right| < \frac{\epsilon}{5}$$

then,

$$\left|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)\right| < \epsilon.$$

Proof. From Lemma 4.3.1,

$$\mu(T^{-n}(A) \cap B)\Delta\mu(T^{-n}(E) \cap F) \leq \mu(A\Delta E) + \mu(B\Delta F) < \frac{2\epsilon}{5}$$

Then,

$$\begin{aligned}\left|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)\right| &\leq \left|\mu(T^{-n}(A) \cap B) - \mu(T^{-n}(E) \cap F)\right| + \left|\mu(T^{-n}(E) \cap F) - \mu(E)\mu(F)\right| \\ &\quad + \left|\mu(E)\mu(F) - \mu(A)\mu(F)\right| + \left|\mu(A)\mu(F) - \mu(A)\mu(B)\right| \\ &\leq \frac{2\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5}\,\mu(F) + \frac{\epsilon}{5}\mu(A) \\ &\leq \epsilon\end{aligned}$$

□

Lemma 4.3.3: For a measure-preserving transformation T on a probability space, $(X, \mathcal{F}, \mu)$, suppose that for an arbitrary $\epsilon > 0$ and $A, B \in \mathcal{F}$, there exists $E, F \in \mathcal{F}$ such that, $\mu(A\Delta E) < \frac{\epsilon}{5}$ & $\mu(B\Delta F) < \frac{\epsilon}{5}$. If for $n > 0$,

1.
$$\left|\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(E) \cap F) - \mu(E)\mu(F)\right| < \frac{\epsilon}{5}$$

   then,

$$\left|\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right| < \epsilon$$

2.
$$\frac{1}{n}\sum_{i=0}^{n-1} \left|\mu(T^{-i}(E) \cap F) - \mu(E)\mu(F)\right| < \frac{\epsilon}{5}$$

   then,

$$\frac{1}{n}\sum_{i=0}^{n-1} \left|\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right| < \epsilon$$

Proof. To show this we just need to invoke Lemma 4.3.1 and proceed with the same steps followed in the proof of Lemma 4.3.2. □

Theorem 4.3.4: For a measure-preserving transformation T over a probability space, $(X, \mathcal{F}, \mu)$, with a sufficient semi-ring, $\mathbf{C}$,

1. For every $I, J \in \mathbf{C}$,

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

   then, T is ergodic.

2. For every $I, J \in \mathbf{C}$,
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(T^{-i}(I) \cap J) - \mu(I)\mu(J)\right| = 0,$$
then, T is weakly-mixing.

3. For every $I, J \in \mathbf{C}$,
$$\lim_{n\to\infty} \mu(T^{-n}(I) \cap J) = \mu(I)\mu(J),$$
then, T is mixing.

Proof. All the statements are proved with the same procedure. Hence it's enough to show for one statement. For the ease of writing the proof, let us choose to prove (3) here.
Take two sets of positive measure, $A, B \in \mathcal{F}$. For any arbitrary $\epsilon > 0$, we know that we can find sets $E = \bigsqcup_{j=0}^{p} I_j$ and, $F = \bigsqcup_{k=0}^{q} J_k$ where each $I_j, J_k \in \mathbf{C}, j = 1, 2, ..., p$ and $k = 1, 2, ..., q$, such that,
$$\mu(A\Delta E) < \epsilon \text{ and, } \mu(B\Delta F) < \epsilon$$
See that for every $J_k \in \{1, 2, ...q\}$,
$$\begin{aligned}
\lim_{n\to\infty} \mu(T^{-n}(E) \cap J_k) &= \lim_{n\to\infty} \mu(T^{-n}(\bigsqcup_{j=1}^{p} I_j) \cap J_k) \\
&= \lim_{n\to\infty} \sum_{j=1}^{p} \mu(T^{-n}(I_j) \cap J_k) \\
&= \sum_{j=1}^{p} \lim_{n\to\infty} \mu(T^{-n}(I_j) \cap J_k)
\end{aligned}$$
From the hypothesis of the statement (3),
$$\begin{aligned}
&= \sum_{j=1}^{p} \mu(I_j \cap J_k) \\
&= \mu((\bigsqcup_{j=1}^{p} I_j)\ \mu(J_k) \\
&= \mu(E)\ \mu(J_k)
\end{aligned}$$
Using this,
$$\begin{aligned}
\lim_{n\to\infty} \mu(T^{-n}(E) \cap F) &= \lim_{n\to\infty} \mu(T^{-n}(E) \cap (\bigsqcup_{k=1}^{q} J_k)) \\
&= \lim_{n\to\infty} \sum_{k=1}^{q} \mu(T^{-n}(E) \cap J_k) \\
&= \sum_{k=1}^{q} \lim_{n\to\infty} \mu(T^{-n}(E) \cap J_k)
\end{aligned}$$
From what we saw above,
$$\begin{aligned}
&= \sum_{j=1}^{p} \mu(E)\ \mu(J_k) \\
&= \mu(E)\ \mu(\bigsqcup_{k=1}^{q} J_k) \\
&= \mu(E)\ \mu(F)
\end{aligned}$$

Hence there exists a number $N > 0$ such that for all $n > N$ and an arbitrary $\epsilon > 0$

$$\left|\mu(T^{-n}(E) \cap F) - \mu(E)\mu(F)\right| < \frac{\epsilon}{5}$$

then from Lemma 4.3.2

$$\left|\mu(T^{-n}(A) \cap B) - \mu(A)\mu(B)\right| < \epsilon$$

Since this holds for any two measurable sets with positive measure, T has to be mixing. $\square$

4.4 Mixing Sequences

Over a probability space $(X, \mathcal{F}, \mu)$, let T be a measure-preserving transformation. An infinite sequence, a_k can be defined outside a zero density set D for measurable sets A and B such that $\lim_{n\to\infty, n\notin D} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$. For this let $\mathbf{C}$ be a countable sufficient semi-ring over a probability space $(X, \mathcal{F}, \mu)$. Let $B_i, B_j \in \mathbf{C}$, then there exists a zero density set $D(B_i, B_j)$ such that $\lim_{n\to\infty, n\notin D(B_i,B_j)} \mu(T^{-n}(B_i) \cap B_j) = \mu(B_i)\mu(B_j)$. This says that for any pair of i, j we choose, there is always an infinite sequence $n_\alpha(i, j)$ such that

$$\lim_{\alpha\to\infty} \mu(T^{-n_\alpha(i,j)}(B_i) \cap B_j) = \mu(B_i)\ \mu(B_j) \tag{4.2}$$

With $\mathbf{C}$ being countable invoking Cantor's diagonalization argument will deliver a sequence $\{n_\alpha\}$ such that for all $i, j \in \mathbb{Z}$

$$\lim_{\alpha\to\infty} \mu(T^{-n_\alpha}(B_i) \cap B_j) = \mu(B_i)\ \mu(B_j)$$

Now the following lemma will show that we can find infinite sequences like above for all measurable sets of $\mathcal{F}$, which is just an application of the approximation lemmas we have proved over to eqn.4.2.

Lemma 4.4.1: Suppose there exist a countable sufficient semi-ring $\mathbf{C}$ and an infinite sequence $\{n_\alpha\}$ for a probability space $(X, \mathcal{F}, \mu)$ such that for all $I, J \in \mathbf{C}$

$$\lim_{\alpha\to\infty} \mu(T^{-n_\alpha}(I) \cap J) = \mu(I)\ \mu(J)$$

Then for all measurable sets $A, B \in \mathcal{F}$,

$$\lim_{\alpha\to\infty} \mu(T^{-n_\alpha}(A) \cap B) = \mu(A)\ \mu(B). \tag{4.3}$$

Any such sequences $\{n_\alpha\}$ that holds eqn.4.3 is called a **mixing sequence**.

4.5 Relations Associated with Weak-Mixing

Proposition 4.5.1: Over a probability space $(X, \mathcal{F}, \mu)$, let T be a measure-preserving transformation. Then the following are equivalent;

1. T is weakly-mixing
2. $T \times T$ is weakly-mixing
3. $T \times T$ is ergodic

Proof. For (1) $\implies$ (2) , consider two probability measure spaces $(X, \mathcal{F}(X), \mu_1)$ and $(Y, \mathcal{F}(Y), \mu_2)$. We will now show that the collection $\mathcal{F}(X) \times \mathcal{F}(Y)$ is a semi-ring on $X \times Y$.

Clearly $X \times Y \in \mathcal{F}(X) \times \mathcal{F}(Y)$. For $A_1 \times B_1$ and $A_2 \times B_2 \in \mathcal{F}(X) \times \mathcal{F}(Y)$, let $A_3 = A_1 \cap A_2$ and $B_3 = B_1 \cap B_2$. Then $(A_1 \times B_1) \cap (A_2 \times B_2) = A_3 \times B_3 \in \mathcal{F}(X) \times \mathcal{F}(Y)$.

Suppose $(a, b) \in (A_1 \times B_1) \setminus (A_2 \times B_2)$. If $a \in A_1 \setminus A_2$ then either $b \in B_1 \setminus B_2$ or $b \in B_1 \cap B_2$. And if $a \in A_1 \cap A_2$ then $b \in B_1 \setminus B_2$. This shows that

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \setminus A_2) \times (B_1 \setminus B_2) \bigcup (A_1 \setminus A_2) \times (B_1 \cap B_2) \bigcup (A_1 \cap A_2) \times (B_1 \setminus B_2)$$

which is a finite union of elements from $\mathcal{F}(X) \times \mathcal{F}(Y)$. Therefore $\mathcal{F}(X) \times \mathcal{F}(Y)$ forms a semi-ring.

Take measurable sets A, B, C, D in X. Then there are zero-density sets $D_1 = D(A, B)$ and $D_2 = D(C, D)$ with every $n \notin D_1 \cup D_2$ that gives

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$$
$$\lim_{n\to\infty} \mu(T^{-n}(C) \cap D) = \mu(C)\mu(D).$$

Let $\mu^2 = \mu \times \mu$ denote the product measure on $X \times X$. Then with μ^2 and every $n \notin D_1 \cup D_2$

$$\lim_{n\to\infty} \mu^2\Big((T \times T)^{-n}(A \times C) \cap (B \times D)\Big) = \mu^2(A \times C)\mu^2(B \times D). \tag{4.4}$$

With all the measurable sets $A, B, C, D \in \mathcal{F}(X) \times \mathcal{F}(X)$ and $\mathcal{F}(X) \times \mathcal{F}(X)$ being a semi-ring for $X \times X$, we can invoke Theorem 4.3.4 here. Then $T \times T$ is weakly mixing.

Applying Lemma 4.2.2 will give (2) $\implies$ (3).

To show (3) $\implies$ (1), let $A \times B$ be $T \times T$-invariant. Then with $T \times T$ being ergodic,

$$\mu(A \times B) = 0 \text{ or } \mu(A^c \times B^c) = 0$$

This implies,

$$\mu(A)\ \mu(B) = 0 \text{ or } \mu(A^c)\ \mu(B^c) = 0.$$

Hence T is ergodic. Using this fact

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

for sets $A \times A$ and $B \times B$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu^2\Big((T \times T)^{-i}(A \times A) \cap (B \times B)\Big) = \mu^2(A \times A)\ \mu^2(B \times B)$$

This gives,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Big[\mu(T^{-i}(A) \cap B)\Big]^2 = \mu^2(A)\ \mu^2(B)$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \Big[\mu(T^{-i}(A) \cap B) - \mu(A)\ \mu(B)\Big]^2 = 0$$

Then using Proposition 4.2.3, T is weakly mixing. $\square$

Definition 4.5.2: A measure-preserving transformation T is said to have a continuous spectrum if $\lambda = 1$ is its only eigen value and it is simple i.e., the set $E(\lambda) := \{f \in L^2 : f \circ T = \lambda\ f \text{ a.e.}\}$ is of dimension one.

Theorem 4.5.3: Over a probability space $(X, \mathcal{F}, \mu)$, let T be a measure-preserving transformation. Then the following are equivalent;

1. T is weakly-mixing

2. T is doubly ergodic

3. T has continuous spectrum

4. $T \times S$ is ergodic for any ergodic transformation S on a finite measure space.

Proof. For (1) $\implies$ (2), it can be shown using Proposition 4.2.5.

For (2) $\implies$ (3), with T being doubly ergodic T is ergodic. Suppose we have multiple eigen values with each of them lies on the unit circle, i.e., $|\lambda| = 1$. Then λ can be written as, $\lambda = e^{2\pi i\alpha}$ where $\alpha \in [0, 1)$. Suppose that the eigen function is of the form $f(x) = e^{2\pi i g(x)}$ where $g(x)$ is a measurable function from X to $[0, 1)$. Let R be a transformation on $[0, 1)$ defined as, $R(z) = z + \alpha$. Then it can be seen that $g \circ T = R \circ g$, i.e.,

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[d, "g"] & X \arrow[d, "g"] \\ {[0,1)} \arrow[r, "R"] & {[0,1)} \end{tikzcd}$$

Over the space $[0, 1)$, we can define a measure for any $U \in [0, 1)$

$$\nu(U) = \mu(g^{-1}(U))$$

Clearly $\nu([0, 1)) = \mu(X)$. This would make the function g a factor map from T to R. Using Theorem 3.1.4, the map R is ergodic. But from the definition of R, it is merely a rotation map. Here α can be either rational or irrational. If α is rational, then ν is atomic and is concentrated in finite number of atoms. This would mean R is not doubly ergodic. And if α is assumed to be irrational, the proof of Lemma 4.2.6 gives that R is not doubly ergodic. This shows that (2) $\implies$ (3).

For (3) $\implies$ (4), suppose $T \times S$ is not ergodic. Then there would exist a function h which is $T \times S$-invariant and non-constant. Let $U_T : L^2(\mu) \to L^2(\mu)$ and $U_S : L^2(\mu) \to L^2(\mu)$ be the unitary operators given by $U_T f = f \circ T$ and $U_S f = f \circ S$. Then decomposing the function h gives the form $f \otimes g$ where f and g are eigen functions of U_T and U_S respectively. These functions f and g must have eigen vales λ and λ' respectively satisfying $\lambda\lambda' = 1$. But eigen values in T and S are closed under complex conjugation. This would give U_T and U_S to have a common eigen value other than 1, implying T does not have continuous spectrum (neither do S).

To show (4) $\implies$ (1), since (4) holds for any ergodic measure-preserving transformation S, take $S = T$ and then apply, Proposition 4.5.1. $\square$

4.6 Rigidity & Mild Mixing

Definition 4.6.1: A measure-preserving transformation T is called **rigid** if for every measurable set A and any arbitrary $\epsilon > 0$ there exists an integer $n > 0$ such that $\mu(T^{-n}(A)\Delta A) < \epsilon$.

Clearly from the definition it can be seen that the identity map is a rigid transformation.

Theorem 4.6.2: Rotational transformations are rigid.

Proof. Consider a rotation transformation $R = R_\alpha$ on $[0, 1)$ where $\alpha \in \mathbb{R}$.
If $\alpha \in \mathbb{Q}$ then there exists p such that $R^p(A) = A$ for all sets A. This makes R a rigid transformation. So take $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. Assume the set A is an interval over $[0, 1)$, say $A = (a, b)$ where $a, b \in [0, 1)$ and $a < b$. Choose an arbitrary $\epsilon_1 > 0$. With α being irrational, $\{R^k(a)\}_{k\geq 0}$ and $\{R^k(b)\}_{k\geq 0}$ are dense in $[0, 1)$.This would mean there is some integer n satisfying

$$a < R^n(a) < a + \frac{\epsilon_1}{2}$$

and,

$$b < R^n(b) < b + \frac{\epsilon_1}{2}$$

this gives,

$$\mu(R^n((a,b))\Delta(a,b)) < \epsilon_1.$$

Now take the set A to be an arbitrary set of positive measure. Choose an arbitrary $\epsilon > 0$. We know that any set of positive measure there exists a set G^* which is a finite union of disjoint intervals and satisfies, $\mu(G^*\Delta A) < \frac{\epsilon}{3}$. Suppose this set is formed by the union of l disjoint intervals,i.e.,

$$G^* = \bigsqcup_{j=1}^{l} I_j$$

If we set $\epsilon_1 = \frac{\epsilon}{3}$, and look at one of those disjoint intervals, say, I_1 then there exists an integer k such that

$$\mu(R^j(I_1)\Delta I_1) < \epsilon_1 = \frac{\epsilon}{3}$$

this would imply for all $j = 1, 2, ..., l$

$$\mu(R^j(I_j)\Delta I_j) < \epsilon_1 = \frac{\epsilon}{3}$$

Then by triangle inequality,

$$\begin{aligned}\mu(R^j(A)\Delta A) &< \mu(R^j(A)\Delta R^j(G^*)) + \mu(R^j(G^*)\Delta G^*) + \mu(R^j(G^*)\Delta G^*)\\ &\leq \frac{\epsilon}{3} + l\ \frac{\epsilon}{3l} + \frac{\epsilon}{3}\\ &= \epsilon.\end{aligned}$$

□

Definition 4.6.3: A measure-preserving transformation T is said to be **partially rigid** if there exist a constant $\alpha > 0$ and an increasing sequence r_n such that for all sets A of finite measure,

$$\lim_{n\to\infty} \inf\ \mu(T^{-r_n}(A) \cap A) \geq \alpha\ \mu(A)$$

We would denote this constant α as the rigidity constant.

Lemma 4.6.4: Suppose $(X, \mathcal{F}, \mu)$ be an atomic probability space with a transformation T. If T is partially rigid then T is not mixing.

Proof. With a rigidity constant of α, there is some increasing sequence $\{n_i\}$ that makes T partially-rigid. Take a measurable set A such that $0 < \mu(A) < \frac{\alpha}{2}$. If T is assumed to be mixing then

$$\lim_{n_i\to\infty} \mu(T^{-n_i}(A) \cap A) = \mu^2(A).$$

But $\mu^2(A) < \frac{\alpha}{2}\ \mu(A)$ and for infinitely many $i > 0$

$$\mu(T^{-n_i}(A) \cap A) > \alpha\mu(A) \geq 2\ \mu^2(A).$$

From this contradiction, we can say T is not mixing. □

Definition 6.4.5: Over a probability space $(X, \mathcal{F}, \mu)$ a measure-preserving transformation T is said to be **mildly mixing** if

$$\liminf_{n\to\infty}\ \mu(A\Delta T^{-n}(A)) > 0$$

for all sets A with $0 < \mu(A) < 1$.

Theorem 4.6.6: Over a probability space $(X, \mathcal{F}, \mu)$ a measure-preserving transformation T is mildly mixing if and only if

$$\liminf_{n\to\infty}\ \mu(A^c \cap T^{-n}(A)) > 0$$

for all sets A with $0 < \mu(A) < 1$.

Proof. With T being mildly mixing, suppose there is an increasing sequence $\{n_i\}$ such that

$$0 = \lim_{n\to\infty} \mu(A^c \cap T^{-n_i}(A)) \tag{4.5}$$

With T being measure-preserving

$$\begin{aligned}\mu(A) &= \mu(T^{-n_i}(A)) \\ &= \mu(A^c \cap T^{-n_i}(A)) + \mu(A \cap T^{-n_i}(A)) \\ &= \lim_{i\to\infty} \mu(A \cap T^{-n_i}(A)) \end{aligned} \tag{4.6}$$

Also

$$\mu(A) = \mu\bigg(A \cap T^{-n_i}(A)\bigg) + \mu\bigg(A \cap (T^{-n_i}(A))^c\bigg) \tag{4.7}$$

Then from eqn.4.6 and eqn.4.7

$$0 = \lim_{i\to\infty} \mu\bigg(A \cap (T^{-n_i}(A))^c\bigg) \tag{4.8}$$

Then from eqn.4.5 and eqn.4.8

$$\begin{aligned}\lim_{i\to\infty} \mu\bigg(A\Delta T^{-n_i}(A)\bigg) &= \lim_{i\to\infty} \mu\bigg(A^c \cap T^{-n_i}(A)\bigg) + \lim_{i\to\infty} \mu\bigg(A \cap (T^{-n_i}(A))^c\bigg) \\ &= 0\end{aligned}$$

Thus we get a contradiction.

For the converse we know

$$\bigg(A^c \cap T^{-n}(A)\bigg) \bigcup \bigg(A \cap T^{-n}(A))^c\bigg) = A\Delta T^{-n}(A)$$

So if

$$\mu(A^c \cap T^{-n}(A)) > 0$$

then

$$\liminf_{n\to\infty} \; \mu(A\Delta T^{-n}(A)) > 0$$

Hence T is mildly-mixing. □

4.7 Examples

Now we can look into two examples where one has the dynamical property of mixing and the other do not.

4.7.1 Doubling Map

Theorem 4.7.1: Let the space be $X = [0, 1]$ and the transformation $T(x) = 2x \mod 1$. Then T is mixing on $(X, \mathcal{F}, \mu)$.

Proof. Recall that a doubling map is defined as

$$\begin{aligned} T : [0, 1] &\to [0, 1] \\ x &\mapsto 2x \mod 1 \end{aligned}$$

Any interval, $[a,b] \in [0,1]$

$$T^{-1}\Big([a,b]\Big) = \left[\frac{a}{2}, \frac{b}{2}\right] \bigcup \left[\frac{a}{2}+\frac{1}{2}, \frac{b}{2}+\frac{1}{2}\right]$$

If we can show that for all measurable sets A, B,

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B) \tag{4.9}$$

then T is mixing. But it is enough show that for all dyadic intervals eqn.4.9 is satisfied. Then Theorem 4.3.4 will give eqn.4.9. So assume the sets A, B are of the form

$$A = \left[\frac{k}{2^i}, \frac{k+1}{2^i}\right] \qquad \text{and} \qquad B = \left[\frac{l}{2^j}, \frac{l+1}{2^j}\right]$$

Then applying T over A

$$\begin{aligned} T^{-1}(A) &= \left[\frac{k}{2^{i+1}}, \frac{k+1}{2^{i+1}}\right] \bigcup \left[\frac{\frac{k}{2^i}+1}{2}, \frac{\frac{k+1}{2^i}+1}{2}\right] \\ &= \left[\frac{k}{2^{i+1}}, \frac{k+1}{2^{i+1}}\right] \bigcup \left[\frac{k}{2^{i+1}}+\frac{1}{2}, \frac{k+1}{2^{i+1}}+\frac{1}{2}\right] \end{aligned}$$

So the pre-image of A by T, i.e., $T^{-1}(A)$ gives two intervals of length $\frac{1}{2^{i+1}}$ and are spaced by $\frac{1}{2}$ among themselves. Then by induction one could see that $T^{-n}(A)$ would give 2^n intervals of length $\frac{1}{2^{i+n}}$ and spaced by $\frac{1}{2^n}$.

Clearly with $n > j$, $T^{-n}(A)$ will intersect B. $T^{-n}(A) \cap B$ will contain all the intervals of $T^{-n}(A)$ in B. Then the number of such intervals in B is given by

$$\begin{aligned} \frac{\text{Length of } B}{\text{Spacing}} &= \frac{\mu(B)}{\frac{1}{2^n}} \\ &= \frac{\frac{1}{2^j}}{\frac{1}{2^n}} \\ &= \frac{1}{2^{j-n}} = 2^{n-j} \end{aligned}$$

With each intervals of $T^{-n}(A)$ having a length of $\frac{1}{2^{i+n}}$,

$$\begin{aligned} \mu(T^{-n}(A) \cap B) &= \frac{1}{2^{i+n}} \cdot 2^{n-j} \\ &= \frac{1}{2^i} \cdot \frac{1}{2^j} \\ &= \mu(A)\ \mu(B). \end{aligned}$$

□

4.7.2 Chacón's Transformation

This transformation is defined on a system developed by R. V. Chacón. The purpose of the construction was to show an example where a measure-preserving transformation is not mixing but weakly mixing.

But before defining what a Chacón transformation is, explaining the construction of a dyadic odometer will simplify the process of understanding the construction of a Chacón's transformation.

Construction of a dyadic odometer: The construction comprises of cutting and stacking of intervals inductively from $[0,1)$.

The first step would be building the the column C_0, which would just be the whole space $[0,1)$. We say that this column has a height $h_0 = 1$. The associated transformation T_{C_0} would be identity.

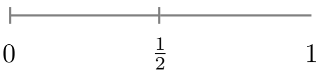


Figure 4.1: First column, C_0.

Construction of the second column C_1 is by cutting the space into two intervals, $[0, \frac{1}{2})$ & $[\frac{1}{2}, 1)$ then stacking the latter over the former. So the column C_1 has two levels and hence have a height $h_1 = 2$. Here the transformation T_{C_1} is defined from $[0, \frac{1}{2}) \rightarrow [\frac{1}{2}, 1)$. Note that T is not defined for $[\frac{1}{2}, 1)$.

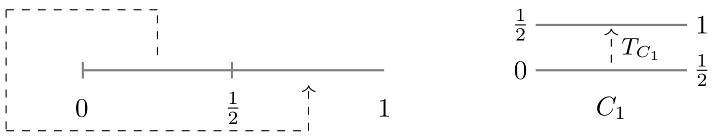


Figure 4.2: Constructing C_1 from C_0.

For the construction of the column C_2 we would cut C_1 into two columns. We then get two columns which comprises of two levels. So C_2 has four levels and height $h_2 = 4$. The transformation T_{C_2} takes $[0, \frac{1}{4}) \rightarrow [\frac{1}{2}, \frac{3}{4}) \rightarrow [\frac{1}{4}, \frac{1}{2}) \rightarrow [\frac{3}{4}, 1)$. Note that T_{C_2} is not defined for $[\frac{3}{4}, 1)$.

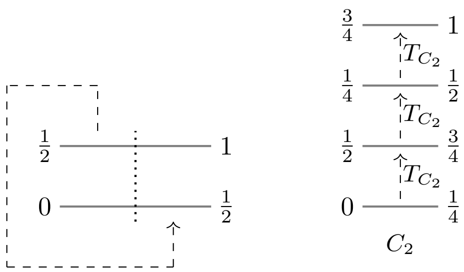


Figure 4.3: Constructing C_2 from C_1.

Now it is possible to construct a general column C_n. Take C_{n-1} and cut the same into two equal columns with each of them having $n-1$ levels. Stack the second column over the first and construct a transformation T_{C_n} which would take a point from a level l to a level $l+1$ that is right above it. From induction we can see the column C_n has 2^n levels thus height, $h_n = 2^n$. Here too T_{C_n} is not defined for the $(2^n)^{\text{th}}$ level.

Definition 4.7.2: Let T be a transformation on $[0, 1)$ defined as

$$T(x) = \lim_{n \to \infty} T_{C_n}(x)$$

Then T is called the **Dyadic odometer**.

Note that $T^{-1}(0)$ is not defined. So if we choose $X_0 = X \setminus \bigcup_{n=0}^{\infty} \{T^n(0)\}$, then X_0 has full-measure and $T : X_0 \rightarrow X_0$ is bijective.

Theorem 4.7.3: The dyadic odometer is measure-preserving and ergodic.

Proof. The dyadic intervals form a semi-ring here. $T^{-1}(I)$ is measurable for any I in the semi-ring. Also for all such I, $\mu(I) = \mu(T^{-1}(I)$. This is sufficient to show that T is measure preserving.

To show T is ergodic, take A_0 and B_0 from $[0, 1)$ and have positive measures. There exists dyadic intervals I and J which are $\frac{1}{2}$-full of A_0 and B_0 respectively. We can always choose I and J to have the same measure. With these two having the same measure, it is possible for both to exist as two levels of some column C_n. This means there exists an integer n such that

$$T^n(I) = J$$

Let $A = A_0 \cap I$ and $B = B_0 \cap J$. Then using Lemma 0.2.6,

$$\begin{aligned}\mu(T^n(A_0) \cap B_0) &\geq \mu(T^n(A) \cap B) \\ &\geq \mu(T^n(I) \cap J) - \mu(I \setminus J) - \mu(J \setminus B) \\ &> \mu(J) - \frac{1}{2}\,\mu(I) - \frac{1}{2}\,\mu(J) \\ &= 0\end{aligned}$$

Thus T is ergodic.

$\square$

Now we can define a **Chacón's transformation**. It too comprises of cutting and stacking of intervals inductively from the space $[0, 1)$.

For the first column C_0 we take the interval $[0, \frac{2}{3})$. Here the height $h_0 = 1$.

Figure 4.4: First column, C_0.

For the next column C_1 instead of splitting the whole space into two equal parts, we split C_0 into 3 equal part/columns and a 'spacer' which is an interval extended from the end point of the last level of the column to the point on which the spacer has the length same as all the levels in the column. In this case, after splitting C_0 into three parts, the new levels have a length of $\frac{2}{9}$. Thus the spacer associated with C_1 would be the interval, $[\frac{2}{3}, \frac{8}{9})$. To stack them, the second column would stack over the first but before stacking the third column the spacer is placed between the last level of the second column and the first level of the third column. So the height of C_1, which is the number of levels after stacking, $h_1 = 3 + 1$.

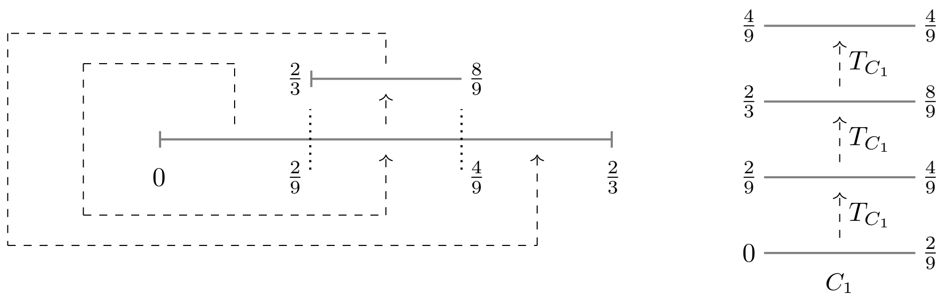


Figure 4.5: Construction of C_1 from C_0.

Therefore, inductively we could calculate the height of a general column C_n, which is given as

$$h_n = 3\ h_{n-1} + 1.$$

And in C_n, each level will have a length of $(\frac{1}{3})^n$. Note that any level I in C_n will have three sub-intervals in C_{n+1}.

Each column C_n has an associated transformation T_{C_n} which takes values from a level to adjacent upper level. Note that for each T_{C_n}, the image of the uppermost level is not defined. We would define,

$$T(x) = \lim_{n \to \infty} T_{C_n}(x)$$

as the **Chacón's transformation**. The measure of all spacers added is,

$$\frac{2}{9} + \frac{1}{3} \cdot \frac{2}{9} + \left(\frac{1}{3}\right)^2 \cdot \frac{2}{9} + ... = \frac{2}{9}\left(\frac{1}{1-\frac{1}{3}}\right)$$
$$= \frac{1}{3}$$

Thus $T\bigg([0,1)\bigg) = 1$.

Lemma 4.7.4: Suppose $n > 0$ and I, J be levels in a column C_n. Then for $k \geq n$,

$$\mu(T^{h_k}(I) \cap I) \geq \frac{1}{3}\ \mu(I).$$

Proof. Let $k = n$. For any level I in C_n, there are three sub-intervals of I. Denote these intervals as, $I[0], I[1]$, & $I[2]$. Then looking at how C_n is constructed, it can be seen that,

$$T^{h_n}(I[0]) = k[1]$$
$$T^{h_n}(I[1]) = T^{-1}(I[2])$$

This tells that $T^{h_n}(I)$ intersects both I and $T^{-1}(I)$ in measure at least $\frac{1}{3}$ times the measure of I, i.e.,

$$\mu(T^{h_n}(I) \cap I) \geq \frac{1}{3}\ \mu(I)$$
$$\mu(T^{h_n}(I) \cap T^{-1}(I)) \geq \frac{1}{3}\mu(I)$$

Now suppose $k = n + l$, for some $l > 0$. Then for a level I in C_n will have 3^l sub-intervals of I in C_{n+l}. If I'_i be one such sub-interval which lies in one of the levels of C_{n+l}. Then applying the same strategy used in the '$k = n$' case here,

$$\mu(T^{h_k}(I'_i) \cap I'_i) \geq \frac{1}{3}\ \mu(I'_i)$$

Combining the above result with the Chacón's Transformation T being measure-preserving gives,

$$\mu(T^{h_k}(I) \cap I) \geq \frac{1}{3}\bigg(\mu(I'_1) + \mu(I'_2) + ... + \mu(I'_{3^l})\bigg)$$
$$\geq \frac{1}{3}\ \mu(I)$$

□

Theorem 4.7.5: Chacón's transformation is not mixing.

Proof. Choose $n > 0$. Suppose I is a level on some column C_n. It is obvious that

$$\mu(I) < \frac{1}{3}$$

Then for all $k \geq n$

$$\mu(T^{h_k}(I) \cap I) \geq \frac{1}{3}\ \mu(I)$$
$$> \mu(I)\ \mu(I)$$

Therefore T is not mixing. □

It can be shown that a Chacón's transformation is doubly ergodic. If this is so then the transformation is weakly mixing.

Chapter 5

Entropy

Entropy as we know from Information Theory and Statistical Mechanics, it measures the uncertainty associated with a system. From the notions of entropy in thermodynamics which was used to find the energy loss of a heat-engine, an electrical engineer Claude Shannon in 1948, developed the probabilistic notions of entropy. In the mid 1950's, Shannon's theory was adapted to the theory of dynamical system by Andrei Kolmogorov. It was found that entropy is useful in determining two dynamic systems isomorphic or not. Kolmogorov's entropy was called the metric entropy, which is an invariant of measure theoretical dynamical systems. In 1961, an invariant for topological systems, called the topological entropy was developed by Roy Adler.

This chapter can be seen as three sections. First we would see Kolmogorov's metric entropy, then Adler's topological entropy and finally, we would try to bridge both these notions.

To begin with Kolmogrov's entropy, some definitions are required to be made. And to define an entropy for a transformation, it would take a three step process which will be stated after the defining some basic terminologies.

5.1 Preliminaries

Definition 5.1.1: A partition of $(X, \mathcal{F}, \mu)$ is a collection $\{A_n\} \subset$ such that $\forall i \neq j, A_i \cap A_j \neq \emptyset$ and $\bigcup A_i = X$.

A partition is said to be finite if the collection of disjoint elements from the partition are finite in number.

Notations 5.1.2:

1. Suppose there is a finite partition of $(X, \mathcal{F}, \mu)$, say ξ, then the collection of elements in $\mathcal{F}$ which are formed by union of elements in ξ is a sub-σ-algebra of $\mathcal{F}$. We would denote this collection of elements as $\mathcal{A}(\xi)$. Often $\mathcal{A}(\xi)$ is called the sub-σ-algebra generated by the partition ξ.
2. A converse of the above can also be shown. Take $\mathcal{C}$ be a finite sub-σ-algebra of $\mathcal{F}$. Then the non-empty sets of the form, $A_1 \cap A_2 \cap ... \cap A_m$, where $A_i = C_i$ or $A_i = X \setminus C_i$, with $C_i \in \mathcal{C}$ forms a partition of $(X, \mathcal{F}, \mu)$. The smallest such partition will be denoted as $\xi(\mathcal{C})$.
3. From above two points, it can be seen that

$$\mathcal{A}(\xi(\mathcal{C})) = \mathcal{C}$$

   And with η being a finite partition

$$\xi(\mathcal{A}(\eta)) = \eta$$

Definition 5.1.3: Let ξ and η be two finite partitions of $(X, \mathcal{F}, \mu)$. η is said to be a refinement of ξ if every element of ξ is a union of elements from η. This will be denoted by $\xi \leq \eta$.

Note that, with partitions ξ and η,

$$\mathcal{A}(\xi) \subset \mathcal{A}(\eta) \iff \xi \leq \eta.$$

Definition 5.1.4: Consider two partitions, $\xi = \{A_1, A_2, ..., A_m\}$ and $\eta = \{B_1, B_2, .., B_n\}$ for the space, $(X.\mathcal{F}, \mu)$. Then the join of the partition is defined as,

$$\xi \vee \eta := \{A_i \cap B_j : 1 \leq i \leq m, 1 \leq j \leq n\}.$$

It can be observed that if $\mathcal{A}$ and $\mathcal{C}$ are two sub-σ-algebra of $\mathcal{F}$, then

$$\xi(\mathcal{A} \vee \mathcal{C}) = \xi(\mathcal{A}) \vee \xi(\mathcal{C})$$

and

$$\mathcal{A}(\xi \vee \eta) = \mathcal{A}(\xi) \vee \mathcal{A}(\eta).$$

If T is a measure-preserving transformation and $\xi = \{A_1, A_2, ..., A_m\}$ is a partition, then

$$T^{-n}(\xi) = \{T^{-n}(A_1), T^{-n}(A_2), ..., T^{-n}(A_m)\}.$$

This shows that even if a transformation is applied, the above mentioned properties would still hold.

Definition 5.1.5: For sub-σ-algebras $\mathcal{C}$ and $\mathcal{D}$ of $\mathcal{F}$, we write $\mathcal{C} \dot{\subseteq} \mathcal{D}$ if for every $C \in \mathcal{C}$ $\exists$ $D \in \mathcal{D}$ with $\mu(D \Delta C) = 0$. We write $\mathcal{C} \doteq \mathcal{D}$ if $\mathcal{C} \dot{\subseteq} \mathcal{D}$ and $\mathcal{D} \dot{\subseteq} \mathcal{C}$.

We say for two partitions ξ and η, $\xi \doteq \eta$ if $\mathcal{A}(\xi) \doteq \mathcal{A}(\eta)$.

With these definitions and notations, we can proceed to define Kolmogorov's metric Entropy. For that, there would be three stage process as mentioned before. Those steps are,

1. Define entropy for a finite sub-σ-algebra
2. Define entropy for a measure-preserving transformation T relative to a sub-σ-algebra.
3. And finally, define entropy of the transformation, T.

5.2 Entropy of a partition

Suppose a finite partition, $\xi = \{A_1, A_2, ..., A_k\}$ of the probability space $(X, \mathcal{F}, \mu)$. If we see the partition as listing the possible outcomes (or events) of some experiment, then the probability associated with A_i is $\mu(A_i)$.

We desire to have a function $H(\xi)$ which would measure the uncertainty associated with performing the experiment. We wish to make $H(\xi)$ depend only on $\{\mu(A_1), \mu(A_2), ..., \mu(A_k)\}$. Then $H(\xi)$ can be denoted as $H\Bigg((\mu(A_1), \mu(A_2), ..., \mu(A_k)\Bigg)$.

Given two partitions, $\xi = \{A_1, A_2, ..., A_m\}$ and $\eta = \{B_1, B_2, .., B_n\}$, the function H must be able to calculate the uncertainty over the outcome of ξ if the outcome of η is given. To obtain an expression of this, first assume that an event B_j of η has occurred then the uncertainty about the outcome of ξ given B_j is

$$H(\xi) = H\Bigg(\frac{\mu(A_1 \cap B_j)}{\mu(B_j)}, \frac{\mu(A_2 \cap B_j)}{\mu(B_j)}, ..., \frac{\mu(A_m \cap B_j)}{\mu(B_j)}\Bigg)$$

Then if η has been told to have occurred,

$$H(\xi|\eta) = \sum_{j=1}^{n} \mu(B_j)\ H\Bigg(\frac{\mu(A_1 \cap B_j)}{\mu(B_j)}, \frac{\mu(A_2 \cap B_j)}{\mu(B_j)}, ..., \frac{\mu(A_m \cap B_j)}{\mu(B_j)}\Bigg).$$

These are some of the properties we wish to see in entropy of a partition. But there are some more desirable properties. The following theorem states all the desirable properties and finally gives the form of the function H that satisfies all of them.

Theorem 5.2.1: Let $\Delta_K := \{(p_1, p_2, ..., p_k) \in \mathbb{R}^k \ : \ p_i \geq 0, \sum_{i=0}^{k} p_i = 1\}$. Suppose a function $H : \bigcup_{k=1}^{\infty} \Delta_k \to \mathbb{R}$ has the following properties;

1. $H(p_1, p_2, ..., p_k) \geq 0$. Equality holds if and only if $p_i = 1$ for some i.
2. $H|_{\Delta_k}$ is continuous, for each $k \geq 1$.
3. $H|_{\Delta_k}$ is symmetric, for each $k \geq 1$.
4. $H|_{\Delta_k}$ has its maxima at $(\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k})$, for each $k \geq 1$.
5. $H(\xi \vee \eta) = H(\xi) + H(\eta|\xi)$.
6. $H\Big((p_1, p_2, ..., p_k, 0)\Big) = H\Big((p_1, p_2, ..., p_k)\Big)$

Then there exists a number $\lambda > 0$ such that,

$$H\Big((p_1, p_2, ..., p_k)\Big) = -\lambda \sum_{i=1}^{k} p_i \log p_i.$$

Before proving this, let us see the new properties added to the function H and some more remarks.

Property 1 says that H has a value zero (that is when no information gain is observed) occurs when there is only one possible outcome.

Property 4 shows that the maximum uncertainty is obtained when all the events have equal probability to occur.

Property 5 says the the information gained from performing ξ and η is the same as the sum of the information gained by performing ξ and the performing η given that ξ has been performed.

Definition 5.2.2: Let $\mathcal{A}$ be a finite sub-algebra of $\mathcal{F}$ with the partition generated by the same is given by $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_k\}$. Then the entropy of $\mathcal{A}$ is given as

$$H(\mathcal{A}) = H\big(\xi(\mathcal{A})\big) = -\sum_{i=1}^{k} \mu(A_i) \log \mu(A_i).$$

With this definition and applying the properties of H over a sub-algebra $\mathcal{A}$, the following can be observed,

- If $\mathcal{A} = \{X, \phi\}$,then $H(\mathcal{A}) = 0$.
- If $\xi(\mathcal{A}) = \{A_1, A_2, .., A_m\}$ such that for all A_i, $\mu(A_i) = \frac{1}{m}$ then

$$\begin{aligned} H(\mathcal{A}) &= -\sum_{i=1}^{m} \frac{1}{m} \log \frac{1}{m} \\ &= \log m \end{aligned}$$

- $H(\mathcal{A}) \geq 0$.
- $H(\mathcal{A}) = H(\mathcal{C})$ if $\mathcal{A} \doteq \mathcal{C}$.
- For a measure preserving transformation $T : X \to X$,

$$H(T^{-1}\mathcal{A}) = H(\mathbf{A}).$$

Proof of Theorem 5.2.1:
Define,

$$L(n) = H(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$$

We have to show $L(n) = \lambda \log n$ where λ is some constant. From the properties (4) and (6) stated in the theorem,

$$L(n) = H(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}, 0) \leq H(\frac{1}{n+1}, \frac{1}{n+1}, ..., \frac{1}{n+1}) = L(n+1)$$

Hence it can be seen that $L(n)$ is an increasing function of n.
Choose positive integers m and r, Let $\xi_1, \xi_2, ..., \xi_m$ be mutually independent finite partitions with each of them having r equally likely events. Then

$$\begin{aligned} H(\xi_k) &= H(\frac{1}{r}, \frac{1}{r}, ..., \frac{1}{r}) \text{ , where } 1 \leq k \leq m \\ &= L(r) \end{aligned}$$

from property (5)

$$\begin{aligned} H(\xi_1, \xi_2, ..., \xi_m) &= \sum_{k=1}^{m} H(\xi_k) \\ &= m \ . \ L(r) \end{aligned} \tag{5.1}$$

But the join of these partitions will give rise to r^m equally likely events. So,

$$H(\xi_1, \xi_2, ..., \xi_m) = L(r^m) \tag{5.2}$$

From the eqn. 5.1 and 5.2

$$L(r^m) = m \ L(r)$$

Similarly for an arbitrary s and n

$$L(s^n) = n \ L(s)$$

Arbitrarily choose r, s and n. And then choose a number m which satisfies

$$r^m \leq s^n \leq r^{m+1}$$

Then

$$m \log r \leq n \log s \leq (m+1) \log r$$

This gives

$$\frac{m}{n} = \frac{\log s}{\log r} = \frac{m}{n} + \frac{1}{n}$$

So we can write

$$\left| \frac{\log s}{\log r} - \frac{m}{n} \right| \leq \frac{1}{n}$$

In a similar way

$$\begin{aligned} L(r^m) \leq L(s^n) \leq L(r^{m+1}) \text{ or,} \\ m \ L(r) \leq n \ L(s) \leq (m+1) \ L(r) \end{aligned}$$

This implies

$$\left|\frac{L(s)}{L(r)} - \frac{m}{n}\right| \leq \frac{1}{n}$$

But we choose arbitrary r, s and n. So

$$\frac{L(s)}{L(r)} = \frac{\log s}{\log r}$$

Which means

$$L(n) = \lambda \log n.$$

Now we can look into the general case where the outcome of events are not equally likely to occur. For this let $\xi = \{A_1, A_2, ..., A_m\}$ be a partition such that for all $A_i \in \xi$, $\mu(A_i) = p_i$, where each $p_i \in \mathbb{Q}$ with $0 \leq p_i \leq 1$. Then let

$$p_i = \frac{g_i}{g} \text{ where } g = \sum_{i=1}^{m} g_i \text{ and each } g_i \in \mathbb{Q}$$

Let η be another partition which is dependent on ξ and contains g events. So let $\eta = \{B_1, B_2, ..., B_g\}$. Divide η into g groups containing $g_1, g_2, ..., g_m$ events respectively. By grouping the events of η, if an event of ξ, say, A_i is occurred then all the g_i events of the i^{th} group will occur with all of them having the same probability of $\frac{1}{g_i}$ and those events not in this group will have a probability of zero. In short, we grouped η in a way that if an event A_i of ξ occurs, then η reduces to a partition of g_i equally likely events. Thus we get,

$$\begin{aligned} H(\eta|A_i) &= H(\frac{1}{g_i}, \frac{1}{g_i}, ..., \frac{1}{g_i}) \\ &= L(g_i) \\ &= \lambda \log g_k \end{aligned}$$

So if $\xi = \{A_1, A_2, ..., A_m\}$ is performed

$$\begin{aligned} H(\eta|\xi) &= \sum_{i=0}^{m} p_i H(\eta|A_i) \\ &= \sum_{i=0}^{m} p_i \ \lambda \log g_i \end{aligned}$$

With $\frac{p_i}{g_i} = g$ and $\sum_{i=0}^{m} p_i = 1$

$$= \lambda \sum_{i=0}^{m} p_i \log p_i + \lambda \log g.$$

Looking at $\xi \vee \eta$, the event $A_i B_j$ occurs only when B_j belongs to the i^{th} group. This gives the total number of events in $\xi \vee \eta$ which is $\sum_{i=1}^{m} g_i = g$ events.The probability of those events is

$$p_i \ . \ \frac{1}{g_i} = \frac{1}{g}$$

Hence,

$$H(\xi \vee \eta) = L(g) = \lambda \log g$$

But,

$$H(\xi \vee \eta) = H(\xi) + H(\eta|\xi)$$
$$\lambda \log g = H(\xi) + \lambda \sum_{i=0}^{m} p_i \log p_i + \lambda \log g$$

or,

$$H(\xi) = -\lambda \sum_{i=0}^{m} p_i \log p_i.$$

With H being continuous, $H(\xi)$ holds for all real values $p_i \in [0,1]$. This concludes the proof. □

The following theorem is a simplified form of Jensen's Inequality theorem.

Corollary 5.2.3: For a partition $\xi = \{A_1, A_2, ..., A_m\}$, $H(\xi) \leq \log m$. And $H(\xi) = \log m$ only when $\mu(A_i) = \frac{1}{m}$ for all i.

Theorem 5.2.4 Consider a function,

$$\phi : [0, \infty) \to \mathbb{R}$$
$$\phi(x) = \begin{cases} 0, \text{ if } x = 0 \\ x \log x, \text{ if } x \neq 0 \end{cases}$$

is strictly convex i.e., $\phi(\alpha x + \beta y) \leq \alpha\phi(x) + \beta\phi(y)$ for $x, y \in [0, \infty); \alpha, \beta \geq 0$ and $\alpha + \beta = 1$. Equality is obtained only when $x = y$ or $\alpha = 0$ or $\beta = 0$.

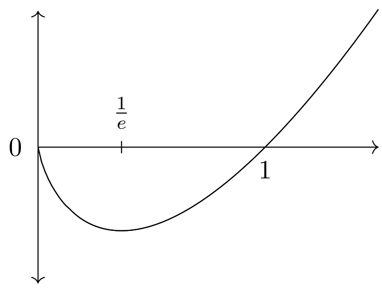


Figure 5.1: Graph of $\phi(x)$

Via induction, if $x \in [0, \infty), \alpha \geq 0$ and $\sum_{i=1}^{k} \alpha_i = 1$,

$$\phi(\sum_{i=1}^{k} \alpha_i x_i) \leq \sum_{i=1}^{k} \alpha_i \phi(x) \tag{5.3}$$

Proof. On differentiating $\phi(x)$,

$$\phi'(x) = 1 + \log x$$
$$\phi''(x) = \frac{1}{x}, \text{ on } [0, \infty)$$

Now fix α and β. Applying the Mean Value Theorem

$$\phi(y) - \phi(\alpha x + \beta y) = \phi'(z)\Big((1-\beta)\ y - \alpha\ x\Big)$$
$$= \phi'(z)\ \alpha\ (y - z)$$

and

$$\phi(\alpha x+\beta y)-\phi(x)=\phi'(w)\beta(y-x)$$

But $\phi'(x)$ is a non-decreasing function. So

$$\begin{aligned}\beta\Big(\phi(y)-\phi(\alpha x-\beta y)\Big)&=\phi'(z)\alpha\beta(y-x)\\&\geq\phi'(w)\alpha\beta(y-x)\\&=\alpha\Big(\phi(\alpha x+\beta y)-\phi x\Big)\end{aligned}$$

This implies

$$\alpha\phi(x)+\beta\phi(y)\geq\phi(\alpha x+\beta y)$$

Eqn.5.3 can be obtained using the same steps as done above with the induction method. □

5.3 Conditional Entropy

Definition 5.3.1: Given two sub-algebra $\mathcal{A}$ and $\mathcal{C}$ of $\mathcal{F}$, the entropy of $\mathcal{A}$ given $\mathcal{C}$ is

$$\begin{aligned}H(\mathcal{A}|\mathcal{C})&=H\Big(\xi(\mathcal{A})|\xi(\mathcal{C})\Big)\\&=-\sum_{j=1}^{p}\mu(C_j)\sum_{i=1}^{k}\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\log\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\\&=-\sum_{i,j}\mu(A_i\cap C_j)\log\frac{\mu(A_i\cap C_j)}{\mu(C_j)}.\end{aligned}$$

Notes:

- $H(\mathcal{A}|\mathcal{C})\geq 0$.
- $H(\mathcal{A}|\mathcal{C})=H(\mathcal{D}|\mathcal{C})$ if $\mathcal{A}\doteq\mathcal{D}$.
- $H(\mathcal{A}|\mathcal{C})=H(\mathcal{A}|\mathcal{D})$ if $\mathcal{C}\doteq\mathcal{D}$.

Theorem 5.3.2: For a probability space $(X,\mathcal{F},\mu)$, if $\mathcal{A},\mathcal{C}$, and $\mathcal{D}$ are finite sub-algebra of $\mathcal{F}$ then

1. $H(\mathcal{A}\vee\mathcal{C}\ |\mathcal{D})=H(\mathcal{A}|\mathcal{D})+H(\mathcal{C}|\ \mathcal{A}\vee\mathcal{D})$
2. $H(\mathcal{A}\vee\mathcal{C})=H(\mathcal{A})+H(\mathcal{C}|\mathcal{A})$
3. $H(\mathcal{A}|\mathcal{D})\leq H(\mathcal{C}|\mathcal{D})$, if $\mathcal{A}\dot{\subseteq}\mathcal{C}$.
4. $H(\mathcal{A})\leq H(\mathcal{C})$, if $\mathcal{A}\dot{\subseteq}\mathcal{C}$.
5. $H(\mathcal{A}|\mathcal{C})\geq H(\mathcal{A}|\mathcal{D})$ if $\mathcal{C}\dot{\subseteq}\mathcal{D}$.
6. $H(\mathcal{A})\geq H(\mathcal{A}|\mathcal{D})$
7. $H(\mathcal{A}\vee\mathcal{C}\ |\mathcal{D})\leq H(\mathcal{A}|\mathcal{D})+H(\mathcal{C}|\mathcal{D})$.
8. $H(\mathcal{A}\vee\mathcal{C})\leq H(\mathcal{A})+H(\mathcal{C})$.
9. For a measure-preserving transformation T, $H(T^{-1}\mathcal{A}\ |T^{-1}\mathcal{C})=H(\mathcal{A}|\mathcal{C})$
10. $H(T^{-1}\mathcal{A})=H(\mathcal{A})$.

Proof. For (1),

$$H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) = -\sum_{i,j,k} (A_i \cap C_j \cap D_k) \log \frac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)}$$

But,

$$\frac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)} = \frac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)} \cdot \frac{\mu(A_i \cap_k)}{\mu(D_k)}$$

Hence by splitting the log terms

$$\begin{aligned} H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) &= -\sum_{i,j,k} (A_i \cap C_j \cap D_k) \Bigg(\log \frac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)} + \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \Bigg) \\ &= H(\mathcal{C} \ |\mathcal{A} \vee \mathcal{D}) + H(\mathcal{A}|\mathcal{D}) \end{aligned}$$

For (2), plug in $\mathcal{N} = \{\emptyset, X\}$ into (1), i.e.,

$$\begin{aligned} H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{N}) &= H(\mathcal{A}|\mathcal{N}) + H(\mathcal{C}| \ \mathcal{A} \vee \mathcal{N}) \\ &= H(\mathcal{A}) + H(\mathcal{C}|\mathcal{A}) \end{aligned}$$

For (3), from (1)

$$\begin{aligned} H(\mathcal{C}|\mathcal{D} &= H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) \\ &= H(\mathcal{A}|\mathcal{D}) + H(\mathcal{C}| \ \mathcal{A} \vee \mathcal{D}) \\ &\geq H(\mathcal{A}|\mathcal{D}). \end{aligned}$$

For (4), Plug in $\mathcal{D} = \{\emptyset, X\}$ in (3). So, $H(\mathcal{C}) \geq H(\mathcal{A})$.

For (5), fix i, j and let,

$$\alpha_k = \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \text{ and } X_k = \frac{\mu(A_i \cap D_k)}{\mu(D_k)}$$

Then from Theorem 5.2.4

$$\phi\Bigg(\sum_k \frac{\mu(D_k \cap C_j) \ \mu(A_i \cap D_k)}{\mu(C_j) \ \mu(D_k)} \Bigg) \geq \sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \phi\Bigg(\frac{\mu(A_i \cap D_k)}{\mu(D_k)} \Bigg)$$

But with $\mathcal{C} \dot{\subseteq} \mathcal{D}$, the left hand side of the above equation becomes

$$\phi\Bigg(\frac{\mu(A_i \cap C_j)}{\mu(C_j)} \Bigg) = \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

Multiplying both sides with $\mu(C_j)$ and then summing over i and j gives

$$\begin{aligned} \sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} &\leq \sum_{i,j,k} \mu(D_k \cap C_j) \ \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \ \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \\ &= \sum_{i,j,k} \mu(D_k) \ \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \ \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \end{aligned}$$

or

$$-H(\mathcal{A}|\mathcal{C}) \leq -H(\mathcal{A}|\mathcal{D})$$

This implies

$$H(\mathcal{A}|\mathcal{D}) \leq H(\mathcal{A}|\mathcal{C}).$$

For (6), Plug in $\mathcal{C} = \{\emptyset, X\}$ into (5). This would give the result $H(\mathcal{A}|\mathcal{D}) \leq H(\mathcal{A})$.

For (7), Use (1) in (5), i.e.,

From (1)

$$H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) = H(\mathcal{A}|\mathcal{D}) + H(\mathcal{C}| \ \mathcal{A} \vee \mathcal{D})$$

and from (5), with $\mathcal{A} \dot{\subseteq} \mathcal{A} \vee \mathcal{D}$

$$H(\mathcal{C}|\mathcal{D}) \geq H(\mathcal{C}| \ \mathcal{A} \vee \mathcal{D})$$

So

$$H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) \leq H(\mathcal{A}|\mathcal{D}) + H(\mathcal{C}|\mathcal{D}).$$

For (8), using $\mathcal{D} = \{\emptyset, X\}$ into (7) will give the required result.

For (9) and (10), it is obvious from their definitions. □

Theorem 5.3.3: Let $\mathcal{A}$ and $\mathcal{C}$ be two sub-algebras of $\mathcal{F}$. Then,

1. $H(\mathcal{A}|\mathcal{C}) = 0$ (or $H(\mathcal{A} \vee \mathcal{C}) \ = H(\mathcal{C})$) if and only if $\mathcal{A} \dot{\subseteq} \mathcal{C}$.
2. $H(\mathcal{A}|\mathcal{C}) = H(\mathcal{A})$, (or $H(\mathcal{A} \vee \mathcal{C}) = H(\mathcal{A}) + H(\mathcal{C})$) if and only if $\mathcal{A}$ and $\mathcal{C}$ are independent i.e., $\mu(A \cap C) = \mu(A) \ \mu(C)$ for all $A \in \mathcal{A}$ and $C \in \mathcal{C}$.

Proof. Let $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_k\}$ and $\xi(\mathcal{C}) = \{C_1, C_2, ..., C_p\}$. Without loss of generality, assume each of these events to have a non-zero measure.

For proving (1), suppose $H(\mathcal{A}|\mathcal{C}) = 0$, then,

$$-\sum_{i=1}^{k} \sum_{j=1}^{p} \mu(A_j \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} = 0$$

With each term,

$$-\mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} = 0$$

implies,

$$\mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} = 0$$

So for all i, j, either,

$$\mu(A_i \cap C_j) = 0 \qquad \text{or} \qquad \mu(A_i \cap C_j) = \mu(C_j)$$

Therefore, $\mathcal{A} \dot{\subseteq} \mathcal{C}$. For the converse of (1), Using $\mathcal{A} \dot{\subseteq} \mathcal{C}$, for each i and each j, either,

$$\mu(A_i \cap C_j) = 0 \qquad \text{or} \qquad \mu(A_i \cap C_j) = \mu(C_j)$$

Then, $H(\mathcal{A}|\mathcal{C}) = 0$.
For proving (2), suppose $H(\mathcal{A}|\mathcal{C}) = H(\mathcal{A})$ then,

$$-\sum_{i=1}^{k} \sum_{j=1}^{p} \mu(A_j \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} = -\sum_{j=1}^{k} \mu(A_i) \log \mu(A_i) \tag{5.4}$$

Fixing i and applying Theorem 5.3.4 with,

$$\alpha_j = \mu(C_j)$$
$$x_j = \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

gives,

$$\phi(\sum \alpha_i x_i) \leq \sum \alpha_i \phi(x_i)$$

which is same as,

$$\phi\left(\sum_{j=1}^{p} \mu(C_j) \frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right) \leq \sum_{j=1}^{p} \mu(C_j) \phi\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right)$$
$$\phi(\mu(A_i)) \leq \sum_{j=1}^{p} \mu(C_j) \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

This implies,

$$\mu(A_i) \log \mu(A_i) \leq \sum_{j-1}^{p} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

Multiplying -1 to both sides,

$$- \leq \sum_{j-1}^{p} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \leq -\mu(A_i) \log \mu(A_i) \tag{5.5}$$

Eqn.5.5 has an equality only when $\frac{\mu(A_i \cap C_j)}{\mu(C_j)}$ is a constant for all i, j. But as we have fixed i, the equality will hold here due to eqn.5.4.

Let $a_i = \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$. Then summing $\mu(A_i \cap C_j) = a_i \mu(C_j)$ over j, we get,

$$a_i = \mu(A_i) \implies \mu(A_i \cap C_j) = \mu(A_i)\ \mu(C_j)$$

This holds for all i, j. Therefore $\mathcal{A}$ and $\mathcal{C}$ are independent.

For the converse, with $\mathcal{A}$ and $\mathcal{C}$ being independent gives,

$$H(\mathcal{A}|\mathcal{C}) = H(\mathcal{A})$$

So,

$$H(\mathcal{A} \vee \mathcal{C}) = H(\mathcal{C}) + H(\mathcal{A}|\mathcal{C})$$
$$= H(\mathcal{C}) + H(\mathcal{A})$$

□

Theorem 5.3.4: Let ∇ the space of all finite sub-algebras of $\mathcal{F}$ where two such sub-algebras $\mathcal{A}$ and $\mathcal{C}$ are identified if $\mathcal{A} \doteq \mathcal{C}$. Then,

$$d(\mathcal{A}, \mathcal{C}) = H(\mathcal{A}|\mathcal{C}) + H(\mathcal{C}|\mathcal{A})$$

is a metric on ∇.

Proof. Clearly, $d(\mathcal{A}, \mathcal{C}) \geq 0$.And when $\mathcal{A} \doteq \mathcal{C}$, $(H(\mathcal{A}|\mathcal{C}) = 0$ and $H(\mathcal{C}|\mathcal{A}) = 0$.

$$\therefore d(\mathcal{A}, \mathcal{C}) = 0$$

To show the triangle inequality property,

$$\begin{aligned}H(\mathcal{A}|\mathcal{D}) &\leq H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) \\ &= H(\mathcal{C}|\mathcal{D}) + H(\mathcal{A}| \ \mathcal{C} \vee \mathcal{D}) \\ &\leq H(\mathcal{C}|\mathcal{D}) + H(\mathcal{A}|\mathcal{C})\end{aligned}$$

Similarly,

$$H(\mathcal{D}|\mathcal{A}) \leq H(\mathcal{C}|\mathcal{A}) + H(\mathcal{D}|\mathcal{C})$$

So,

$$\begin{aligned}d(\mathcal{A}|\mathcal{D}) &= H(\mathcal{A}|\mathcal{D}) + H(\mathcal{D}|\mathcal{A}) \\ &\leq H(\mathcal{C}|\mathcal{A}) + H(\mathcal{A}|\mathcal{C}) + H(\mathcal{D}|\mathcal{C}) + H(\mathcal{C}|\mathcal{D}) \\ &\leq d(\mathcal{A}, \mathcal{C}) + d(\mathcal{C}, \mathcal{D})\end{aligned}$$

Therefore d is a metric on ∇. □

With a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$ and let $\mathscr{F}$ be an arbitrary sub-σ-algebra of $\mathcal{F}$, we can define the conditional entropy $H(\mathcal{A}|\mathscr{F})$. But for this we would require the conditional expectation map which is defined as follows.

$$E(\cdot|\mathscr{F}) : L^1(X, \mathcal{F}, \mu) \to L^1(X, \mathscr{F}, \mu)$$

For a finite sub-σ-algebra $\mathcal{C}$ of $\mathcal{F}$ with $\xi(\mathcal{C}) = \{C_1, C-2, ..., C_p\}$, the conditional expectation map is,

$$E(f|\mathcal{C})(x) = \sum_{j=1}^{p} \chi_{C_j}(x) \ . \ \frac{1}{\mu(C_j)} \int_{C_j} f d\mu$$

If $\mathcal{A}$ is a finite sub-σ-algebra with $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_m\}$,

$$\begin{aligned}E(\chi_{A_i}|C_j)(x) &= \chi_{C_j}(x) \ . \ \frac{1}{\mu(C_j)} \int_{C_j} \chi_{A_i} d\mu \\ &= \frac{\mu(A_i \cap C_j)}{\mu(C_j)}\end{aligned}$$

Then for,

$$\begin{aligned}H(\mathcal{A}|\mathcal{C}) &= -\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \\ &= -\sum_{i=1}^{m} \sum_{j=1}^{p} \mu(C_j) E(\chi_{A_i}|C_j) \log E(\chi_{A_i}|C_j) \\ &= -\sum_{i=0}^{m} \int E(\chi_{A_i}|\mathcal{C}) \log E(\chi_{A_i}|\mathcal{C}) d\mu \\ &= -\int \sum_{i=0}^{m} E(\chi_{A_i}|\mathcal{C}) \log E(\chi_{A_i}|\mathcal{C}) d\mu\end{aligned}$$

This helps us define the conditional entropy of a sub-σ-algebra given an arbitrary sub-σ-algebra.

Definition 5.3.5: For a probability space $(X, \mathcal{F}, \mu)$, if $\mathcal{A}$ is a finite sub-σ-algebra of $\mathcal{F}$ with $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_m\}$ and $\mathscr{F}$ is an arbitrary sub-σ-algebra of $\mathcal{F}$, Then the **entropy of $\mathcal{A}$ given $\mathscr{F}$** is,

$$H(\mathcal{A}|\mathscr{F}) = -\int \sum_{i=0}^{m} E(\chi_{A_i}|\mathscr{F}) \log E(\chi_{A_i}|\mathscr{F}) d\mu$$

The existence of $E(\chi_{A_i}|\mathscr{F})$ is proved by seeing that, for $B \in \mathcal{F}, \mu^{A_i} : B \to \int_B \chi_{A_i} \ d\mu$ is a finite measure on$(X, \mathcal{F})$ which is absolutely continuous with respect to μ. So if there is an inclusion map

$g : \mathscr{F} \to \mathcal{F}$, then $\mu_{A_i}|_{\mathscr{F}}$ is the restriction of μ^{A_i} to $\mathscr{F}$. Similarly $\mu|_{\mathscr{F}}$ is the restriction of μ to $\mathscr{F}$. Clearly $\mu^{A_i}|_{\mathscr{F}} << \mu|_{\mathscr{F}}$. Then by invoking the Radon-Nikodym theorem, $E(\chi_{A_i}|\mathscr{F})$ is the Radon-Nikodym derivative i.e.,

$$E(\chi_{A_i}|\mathscr{F}) = \frac{\mu^{A_i}|_{\mathscr{F}}}{\mu|_{\mathscr{F}}}$$

For further studies of the expectation map, the reader can refer to Billingsley's book, Probability and Measure[10]. Note that, $E(\cdot|\mathscr{F})$ is a positive linear operator and $\sum_{i=0}^{m} \chi_{A_i} = 1$. It can be shown that $H(\mathcal{A}|\mathscr{F})$ is finite. For this, we have,

$$0 \leq E(\chi_{A_i}|\mathscr{F})(x) \leq 1, \text{ a.e.}$$

Therefore,

$$\sum_{i=0}^{m} E(\chi_{A_i}|\mathscr{F}) \log E(\chi_{A_i}|\mathscr{F}) d\mu \leq k \max_{t\in[0,1]}(-t\log t) \leq \frac{k}{e}$$

Hence $H(\mathcal{A}|\mathscr{F})$ is finite.

For a family of sub-σ-algebra of $\{\mathscr{F}_n\}_{n=1}^{\infty}$, let $\bigvee_{n=1}^{\infty} \mathscr{F}$ denote the smallest sub-σ-algebra containing all the $\mathscr{F}_n$.

Lemma 5.3.6: Over a probability space $(X, \mathcal{F}, \mu)$, let $\{\mathscr{F}_n\}_{n=1}^{\infty}$ be an increasing sequence of sub-σ-algebra of $\mathcal{F}$. Let $\bigvee_{n=1}^{\infty} \mathscr{F}$ be denoted as $\mathscr{F}$. Then for each $f \in L^2(X, \mathcal{F}, \mu)$,

$$||E(f|\mathscr{F}_n) - E(f|\mathscr{F})||_2 \to 0$$

Proof. It can be seen that $E(f|\mathscr{F}_n)$ is an orthogonal projection of $L^2(X, \mathcal{F}, \mu)$ onto $L^2(X, \mathscr{F}_n, \mu)$. Suppose $B \in \mathscr{F}$. Choose $B_n \in \mathscr{F}_n$ such that $\mu(B_n \Delta B) \to 0$ as $n \to \infty$.The member of $L^2(X, \mathcal{F}, \mu)$ closest to χ_B is $E(\chi_B|\mathscr{F}_n)$. So,

$$\begin{aligned} ||E(\chi_B|\mathscr{F}_n) - \chi_B||_2^2 &\leq ||\chi_{B_n} - \chi_B||_2^2 \\ &= \mu(B_n \Delta B) \to 0 \end{aligned}$$

Recall that in $L^2(X, \mathcal{F}, \mu)$, finite linear combinations of characteristic functions are dense. So for all $f \in L^2(X, \mathcal{F}, \mu)$,

$$||E(f|\mathscr{F}_n) - f||_2 \to 0$$

Also,

$$E(E(f|\mathscr{F}_n)|\mathscr{F}_n) = E(f|\mathscr{F}_n)$$

Hence for all $f \in L^2(X, \mathcal{F}, \mu)$,

$$||E(f|\mathscr{F}_n) - E(f|\mathscr{F})||_2 \to 0$$

□

The same result holds for a decreasing sequence $\{\mathscr{F}_n\}_{n=1}^{\infty}$, of sub-$\sigma$-algebra with $\cap_{n=1}^{\infty} \mathscr{F}_n = \mathscr{F}$.

Theorem 5.3.7: Over a probability space $(X, \mathcal{F}, \mu)$, if $\mathcal{A}$ is a finite sub-σ-algebra of $\mathcal{F}$ and $\{\mathscr{F}_n\}_{n=1}^{\infty}$, is an increasing sequence of sub-σ-algebra of $\mathcal{F}$. Then,

$$H(\mathcal{A}|\mathscr{F}_n) \to H(\mathcal{A}|\mathscr{F})$$

Proof. Let $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_m\}$. From Lemma 5.3.6, for each i,

$$||E(\chi_{A_i}|\mathscr{F}_n) - E(\chi_{A_i}|\mathscr{F})||_2 \to 0$$

Which means $E(\chi_{A_i}|\mathscr{F}_n)$ converges in measure to $E(\chi_{A_i}|\mathscr{F})$. Thus, $-\sum_{i=1}^{m} E(\chi_{A_i}|\mathscr{F}_n) \log E(\chi_{A_i}|\mathscr{F}_n)$ converges in measure to $-\sum_{i=1}^{m} E(\chi_{A_i}|\mathscr{F}) \log E(\chi_{A_i}|\mathscr{F})$.

Also all the functions are bounded by $\frac{m}{e}$. That will give the convergence in $L^1(\mu)$.

$$\therefore H(\mathcal{A}|\mathscr{F}_n) \to H(\mathcal{A}|\mathscr{F})$$

□

The same theorem will hold for a decreasing sequence $\{\mathscr{F}_n\}_{n=1}^{\infty}$, of sub-$\sigma$-algebra with $\mathscr{F} = \cap_{n=1}^{\infty}\mathscr{F}_n$.

Theorem 5.3.8: Over a probability space $(X, \mathcal{F}, \mu)$, let $\mathcal{A}, \mathscr{F}$ be two sub-σ-algebra of $\mathcal{F}$ with $\mathcal{A}$ being finite. Then,

1. $H(\mathcal{A}|\mathscr{F}) = 0$, if and only if $\mathcal{A}\dot{\subseteq}\mathscr{F}$.
2. $H(\mathcal{A}|\mathscr{F}) = H(\mathcal{A})$ if and only if $\mathcal{A}$ and $\mathscr{F}$ are independent.

Proof. Take $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_m\}$.
To prove (1), if $\mathcal{A}\dot{\subseteq}\mathscr{F}$ and knowing that $E(\chi_{A_i}|\mathscr{F})$ takes values either 0 or 1,

$$H(\mathcal{A}|\mathscr{F}) = 0$$

For the converse,

$$0 = H(\mathcal{A}|\mathscr{F}) = -\int \sum_{i=0}^{m} E(\chi_{A_i}|\mathscr{F}) \log E(\chi_{A_i}|\mathscr{F}) d\mu$$

But for each i,

$$-E(\chi_{A_i}|\mathscr{F}) \log E(\chi_{A_i}|\mathscr{F}) \geq 0$$

Therefore,

$$E(\chi_{A_i}|\mathscr{F}) \log E(\chi_{A_i}|\mathscr{F}) = 0$$

Then for all i,

$$E(\chi_{A_i}|\mathscr{F}) = 0 \qquad \text{or} \qquad E(\chi_{A_i}|\mathscr{F}) = 1$$
$$\implies \mathcal{A}\dot{\subseteq}\mathscr{F}$$

To prove (2), suppose $B \in \mathscr{F}$, and $\mathcal{D}$ be a finite sub-σ-algebra consisting of the sets $\{\phi, B, X \setminus B, X\}$. Clearly $\mathcal{D}\dot{\subseteq}\mathscr{F}$. And,

$$H(\mathcal{A}) \geq H(\mathcal{A}|\mathcal{D}) \geq H(\mathcal{A}|\mathscr{F}) = H(\mathcal{A})$$

Hence,

$$H(\mathcal{A}|\mathcal{D}) = H(\mathcal{A})$$

So from theorem 5.3.3, $\mathcal{A}$ and $\mathcal{D}$ are independent and for all $A \in \mathcal{A}, \mu(A \cap B) = \mu(A)\ \mu(B)$. Since this would hold for all $B \in \mathscr{F}$, $\mathcal{A}$ and $\mathscr{F}$ are independent.
For the converse, suppose $\mathcal{A}$ and $\mathscr{F}$ are independent. Then for each $A_i \in \mathcal{A}$,

$$E(\chi_{A_i}|\mathscr{F}) = \mu(A_i)$$
$$\implies H(\mathcal{A}|\mathscr{F}) = H(\mathcal{A})$$

□

With this we can proceed to the next stage; defining entropy for a measure-preserving transformation with respect to a finite sub-σ-algebra .

5.4 Entropy of a Measure-Preserving Transformation

With a measure-preserving transformation T, the elements of the partition,

$$\xi(\bigvee_{i=o}^{n-1} T^{-1}\mathcal{A}) = \bigvee_{i=0}^{n-1} T^{-i}\xi(\mathcal{A})$$

are of the form,

$$\bigcap_{i=0}^{n-1} T^{-i}\mathcal{A}_{j_i}$$

. **Definition 5.4.1:** Let T be a measure-preserving transformation of the probability space $(X, \mathcal{F}, \mu)$. If $\mathcal{A}$ is a finite sub-σ-algebra of $\mathcal{F}$, then

$$h(T, \xi(\mathcal{A})) = h(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=o}^{n-1} T^{-1}\mathcal{A}) \tag{5.6}$$

is called the **entropy of T with respect to a finite sub-σ-algebra** $\mathcal{A}$.

The existence of the limit will be shown later. With the first two stages defined, we can now head into defining the entropy of T.

Definition 5.4.2: Let T be a measure-preserving transformation over a probability space $(X, \mathcal{F}, \mu)$. The **entropy of T** is,

$$h(T) = \sup h(T, \mathcal{A})$$

where the supremum is taken over all finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$.

Equivalently, $h(T) = \sup h(T, \xi)$ where supremum is taken over all finite partitions of $(X, \mathcal{F}, \mu)$. The following properties can be observed for $h(T)$.

- $h(T) \geq 0$
- $h(id_X) = 0$; where id_X is the identity map. This shows that if $h(T) = 0$ then $h(T, \mathcal{A}) = 0$ for every finite sub-σ-algebra $\mathcal{A}$. That means $\bigvee_{i=o}^{n-1} T^{-1}\mathcal{A})$ does not change as $n \to \infty$.

To show the existence of the limit in eqn.5.6 we require to prove the following lemma.

Lemma 5.4.3: Let $\{a_n\}_{n\geq 1}$ be a sequence of real numbers such $a_{n+p} \leq a_n + a_p$ for every n, p. Then $\lim_{n\to\infty} \frac{a_n}{n}$ exists and equals $\inf_n \frac{a_n}{n}$.

Proof. Fix $p > 0$. Suppose for each $n > 0$, write $n = k\ p + i$ where $0 \leq i \leq p$. Then,

$$\begin{aligned}
\frac{a_n}{n} &= \frac{a_{kp+i}}{kp+i} \leq \frac{a_i}{kp+i} + \frac{a_{kp}}{kp+i} \\
&\leq \frac{a_i}{kp} + \frac{a_{kp}}{kp} \\
&\leq \frac{a_i}{kp} + k\ \frac{a_p}{kp} \\
&= \frac{a_i}{kp} + \frac{a_p}{p}
\end{aligned}$$

Then,

$$\limsup \frac{a_n}{n} \leq \frac{a_i}{kp} + k\ \frac{a_p}{p}$$

or,

$$\limsup \frac{a_n}{n} \leq \frac{a_p}{p}$$

So,

$$\begin{aligned}
\limsup \frac{an}{n} &\leq \frac{a_p}{p} \\
\therefore \limsup \frac{a_n}{n} &\leq \inf_p \frac{a_p}{p}
\end{aligned}$$

But,

$$\inf_p \frac{a_p}{p} \leq \liminf_n \frac{a_n}{n}$$

Hence we get,

$$\limsup \frac{a_n}{n} \leq \inf_n \frac{a_n}{n} < \liminf \frac{a_n}{n}$$

Therefore the limit exists. ☐

Theorem 5.4.4: For a measure-preserving transformation T and a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$, $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=o}^{n-1} T^{-1}\mathcal{A})$ exists.

Proof. Set $a_n = H(\bigvee_{i=o}^{n-1} T^{-1}\mathcal{A}) \leq 0$. Then,

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

Then by the above Lemma 5.4.3, the limit must exist. ☐

Theorem 5.4.5: Entropy is an isomorphism invariant.

Proof. Consider two measure-preserving transformations, $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$. Let $\tilde{A}$ denote the equivalent class of sets under the relation $A \sim B$ if $\mu(A\Delta B) = 0$.

Let $\phi : (\tilde{\mathcal{F}_2}, \tilde{\mu_2}) \to (\tilde{\mathcal{F}_1}, \tilde{\mu_1})$ be an isomorphism of measure algebras such that,

$$\phi \circ \tilde{T_2}^{-1} = \tilde{T_1}^{-1} \circ \phi$$

Suppose $\mathcal{A}$ is finite where $\mathcal{A} \dot{\subseteq} \mathcal{F}_2$. Let the partition generated by the same be, $\xi(\mathcal{A}) = \{A_1, A_2, ..., A_r\}$. Choose $B_i \in \mathcal{F}_1$ such that $\tilde{\tilde{\mathcal{F}_i}} = \phi(\tilde{A_i})$ and $\eta = \{B_1, B_2, ..., B_r\}$ forms a partition of $(X_1, \mathcal{F}_1, \mu_1)$. Set $\mathcal{A}_1 = \mathcal{A}(\eta)$ so that $\bigcap_{i=0}^{n-1} T_1^{-1} B_{q_i}$ (where $q_i \in \{1, 2, ..., r\}$) has the same measure as $\bigcap_{i=0}^{n-1} T_2^{-1} A_{q_i}$ because,

$$\begin{aligned} \phi\bigg(\bigcap_{i=0}^{n-1} (T_2^{-i} A_{q_i})^{\sim}\bigg) &= \phi(\bigcap_{i=0}^{n-1} \tilde{T_2}^{-i} \tilde{A_{q_i}}) \\ &= \bigcap_{i=0}^{n-1} \tilde{T_1}^{-i} \phi(\tilde{A_{q_i}}) \\ &= \bigcap_{i=0}^{n-1} \tilde{T_1}^{-i} \tilde{B_{q_i}} \\ &= \bigcap_{i=0}^{n-1} (T_1^{-i} B_{q_i})^{\sim} \\ \therefore H(\bigvee_{i=o}^{n-1} T_1^{-i}\mathcal{A}_1) &= H(\bigvee_{i=o}^{n-1} T_2^{-i}\mathcal{A}_2) \end{aligned}$$

Since this holds with every n,

$$\begin{aligned} &H(T_1, \mathcal{A}_1) = H(T_2, \mathcal{A}_2) \\ \implies &h(T_1) \geq h(T_2) \end{aligned}$$

And by symmetry we get,

$$h(T_1) = h(T_2)$$

☐

This says that if T_2 is a factor of T_1 then $h(T_2) \leq h(T_1)$.
One can observe some properties associated with $h(T, \mathcal{A})$ and $h(T)$. The following set of theorems will talk about those.

Theorem 5.4.6: Let T be a measure-preserving transformation over a probability space $(X, \mathcal{F}, \mu)$. Let $\mathcal{A}$ and $\mathcal{C}$ be two finite sub-algebras of $\mathcal{F}$. Then

1. $h(T, \mathcal{A}) \leq H(\mathcal{A})$
2. $h(T, \mathcal{A} \vee \mathcal{C}) \leq h(T, \mathcal{A}) + h(T, \mathcal{C})$
3. $h(T, \mathcal{A}) \leq h(T, \mathcal{C})$ if $\mathcal{A} \dot{\subseteq} \mathcal{C}$.
4. $h(T, \mathcal{A}) \leq h(T, \mathcal{C}) + H(\mathcal{A}|\mathcal{C})$.
5. $h(T, T^{-1}\mathcal{A}) = h(T, \mathcal{A})$
6. For $k \geq 1, h(T, \mathcal{A}) = h(T, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})$.
7. If T is invertible and $k \geq 1$ then,

$$h(T, \mathcal{A}) = h(T, \bigvee_{i=-k}^{k} T^{-i}\mathcal{A})$$

Proof. For (1), using the property, $H(\mathcal{A} \vee \mathcal{C}) \leq H(\mathcal{A}) + H(\mathcal{C})$,

$$\frac{1}{n} H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) \leq \frac{1}{n} \sum_{i=0}^{n-1} H(T^{-i}\mathcal{A})$$

Using the property, $H(T^{-1}\mathcal{A}) = H(\mathcal{A})$,

$$\begin{aligned} &= \frac{1}{n} \sum_{i=0}^{n-1} H(\mathcal{A}) \\ &= H(\mathcal{A}) \end{aligned}$$

For (2),

$$\begin{aligned} H\Big(T^{-1}(\mathcal{A} \vee \mathcal{C})\Big) &= H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A} \vee \bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}) \\ &\leq H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) + H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}) \\ &\leq h(T, \mathcal{A}) + h(T, \mathcal{C}) \end{aligned}$$

For (3), With $\mathcal{A} \dot{\subseteq} \mathcal{C}$,

$$\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A} \dot{\subseteq} \bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}$$

Then using the property $H(\mathcal{A}) \leq H(\mathcal{C})$, if $\mathcal{A} \dot{\subseteq} \mathcal{C}$,

$$h(T, \mathcal{A}) \leq h(T, \mathcal{C})$$

For (4),

$$H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) \leq H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A} \vee \bigvee_{i=o}^{n-1} T^{-i}\mathcal{C})$$

With the property $H(\mathcal{A} \vee \mathcal{C}) = H(\mathcal{A}) + H(\mathcal{C}|\mathcal{A})$,

$$= H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}) + H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A})| \bigvee_{i=o}^{n-1} T^{-i}\mathcal{C})$$

But from the property $H(\mathcal{A} \vee \mathcal{C} \ |\mathcal{D}) \leq H(\mathcal{A}|\mathcal{D}) + H(\mathcal{C}|\mathcal{D})$,

$$\begin{aligned} H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}| \bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}) &\leq \sum_{i=0}^{n-1} H(T^{-i}\mathcal{A}| \bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}) \\ &\leq \sum_{i=0}^{n-1} H(T^{-i}\mathcal{A}| \ T^{-i}\mathcal{C}) \end{aligned}$$

And from the property $H(\mathcal{A}|\mathcal{C}) \geq H(\mathcal{A}|\mathcal{D})$ if $\mathcal{C}\dot{\subseteq}\mathcal{D}$,

$$= n \ H(\mathcal{A}|\mathcal{C})$$

Thus,

$$H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) \leq H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{C}) + n \ H(\mathcal{A}|\mathcal{C})$$

Taking $n \to \infty$,

$$h(T, \mathcal{A}) \leq h(T, \mathcal{C}) + H(\mathcal{A}|\mathcal{C})$$

For (5),

$$H(\bigvee_{i=1}^{n} T^{-i}\mathcal{A}) = H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A})$$

So,

$$h(T, T^{-1}\mathcal{A}) = h(T, \mathcal{A})$$

For (6),

$$\begin{aligned} h(T, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{j=0}^{n} T^{-j}(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})) \\ &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{k-1+n-1} T^{-i}\mathcal{A}) \\ &= \lim_{n\to\infty} \frac{k+n-2}{k+n-2} \cdot \frac{1}{n} H(\bigvee_{i=0}^{k+n-2} T^{-i}\mathcal{A}) \\ &= \lim_{n\to\infty} \frac{1}{k+n-2} H(\bigvee_{i=0}^{k+n-2} T^{-i}\mathcal{A}) \\ &= h(T, \mathcal{A}) \end{aligned}$$

For (7), from (5),

$$h(T, T^{-1}\mathcal{A}) = h(T, \mathcal{A})$$

then,

$$h(T, \bigvee_{i=-k}^{k} T^{-i}\mathcal{A}) = h(T, \bigvee_{i=0}^{2k} T^{-i}\mathcal{A})$$

from (6),

$$= h(T, \mathcal{A})$$

$\square$

Theorem 5.4.7: Let $\mathcal{A}, \mathcal{C}$ be finite sub-algebra of $\mathcal{F}$. Then $|h(T,\mathcal{A}) - h(T,\mathcal{C})| \leq d(\mathcal{A},\mathcal{C})$. Thus $h(T,\cdot)$ is a continuous function on the metric space (∇, d).

Proof. From Theorem 5.4.6, (4) says that $h(T,\mathcal{A}) \leq h(T,\mathcal{C}) + H(\mathcal{A}|\mathcal{C})$. So,

$$\begin{aligned}
|h(T,\mathcal{A}) - h(T,\mathcal{C})| &\leq \max\Big(H(\mathcal{A}|\mathcal{C}), H(\mathcal{C}|\mathcal{A})\Big) \\
&\leq H(\mathcal{A}|\mathcal{C}) + H(\mathcal{C}|\mathcal{A}) \\
&= d(\mathcal{A},\mathcal{C})
\end{aligned}$$

$\square$

Theorem 5.4.8: Let T be a measure preserving transformation over a probability space $(X, \mathcal{F}, \mu)$.

1. For $k > 0$,
$$h(T^k) = k\ h(T)$$

2. With T being invertible, then for all $k \in \mathbb{Z}$,
$$h(T^k) = |k|\ h(T)$$

Proof. Looking at,

$$\begin{aligned}
h(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{j=0}^{k-1} T^{-kj}(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})) \\
&= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{A}) \\
&= \lim_{n\to\infty} k\ .\frac{1}{kn} H(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{A}) \\
&= k\ h(T,\mathcal{A})
\end{aligned}$$

So,

$$\begin{aligned}
k\ h(T,\mathcal{A}) &= k\ \sup_{\mathcal{A}} h(T,\mathcal{A}) \\
&= \sup_{\mathcal{A}} h(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) \\
&\leq \sup_{\mathcal{C}} h(T^k, \mathcal{C}) \\
&= h(T^k)
\end{aligned}$$

And from Theorem 5.4.6 (3),

$$\begin{aligned}
h(T^k, \mathcal{A}) &\leq h(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}) \\
&= k\ h(T,\mathcal{A}) \\
\therefore h(T^k) &= kh(T)
\end{aligned}$$

For (2), it is sufficient enough to show that $h(T^{-1}) = h(T)$. Using the property that $H(T^{-1}\mathcal{A}) = H(\mathcal{A})$,

$$H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) = H(T^{-(n-1)} \bigvee_{i=o}^{n-1} T^{-i}\mathcal{A})$$

Then for all finite $\mathcal{A}$,

$$h(T^{-1}, \mathcal{A}) = h(T, \mathcal{A})$$

So if $k \in \mathbb{Z}^-$ then,

$$h(T^k) = h(T^{|k|}) = |k| \ h(T)$$

□

The following results will show when $h(T, \mathcal{A})$ goes to zero.

Theorem 5.4.9: For a measure-preserving transformation T over a probability space $(X, \mathcal{F}, \mu)$and a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$,

$$\begin{aligned} h(T, \mathcal{A}) &= \lim_{n\to\infty} H(\mathcal{A}| \bigvee_{i=0}^{n} T^{-i}\mathcal{A}) \\ &= H(\mathcal{A}| \bigvee_{i=0}^{\infty} T^{-i}\mathcal{A}) \end{aligned}$$

Proof. Note that the term with the limit on the right hand side is not increasing, hence the limit will exist due to Theorem 5.3.2 (5). We will first show via induction that,

$$H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) = H(\mathcal{A}) + \sum_{j=1}^{n-1} H(\mathcal{A}| \bigvee_{i=1}^{j} T^{-i}\mathcal{A})$$

Clearly when $n = 1$ it holds true. Suppose this is true for $n = p$. Then for $n = p + 1$,

$$H(\bigvee_{i=o}^{p} T^{-i}\mathcal{A}) = H(\bigvee_{i=1}^{p} T^{-i}\mathcal{A} \vee \mathcal{A})$$

from Theorem 5.3.2 (2),

$$= H(\bigvee_{i=1}^{p} T^{-i}\mathcal{A}) + H(\mathcal{A}| \bigvee_{i=1}^{p} T^{-i}\mathcal{A})$$

from Theorem 5.3.2 (10),

$$= H(\bigvee_{i=0}^{p-1} T^{-i}\mathcal{A}) + H(\mathcal{A}| \bigvee_{i=1}^{p} T^{-i}\mathcal{A})$$

using the induction hypothesis,

$$= H(\mathcal{A}) + \sum_{j=1}^{p} H(\mathcal{A}| \bigvee_{i=1}^{j} T^{-i}\mathcal{A})$$

So by induction,

$$H(\bigvee_{i=o}^{n-1} T^{-i}\mathcal{A}) = H(\mathcal{A}) + \sum_{j=1}^{n-1} H(\mathcal{A}| \bigvee_{i=1}^{j} T^{-i}\mathcal{A})$$

Now take $n \to \infty$ after multiplying $\frac{1}{n}$ to both sides, would give,

$$h(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} H(\mathcal{A}| \bigvee_{i=1}^{j} T^{-i}\mathcal{A})$$

As for a real number, Cesaro limit of convergent sequence is equivalent to the usual limit, the proof is completed. □

Corollary 5.4.9 (i): For a measure-preserving transformation T over a probability space $(X, \mathcal{F}, \mu)$and a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$,

$$h(T, \mathcal{A}) = 0 \qquad \iff \qquad \mathcal{A} \dot{\subseteq} \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}$$

Proof. Suppose $h(T, \mathcal{A}) = 0$. Then from Theorem 5.4.9,

$$\begin{aligned} 0 &= h(T, \mathcal{A}) \\ &= \lim_{n\to\infty} H(\mathcal{A}|\bigvee_{i=1}^{n} T^{-i}\mathcal{A}) \\ &= H(\mathcal{A}|\bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}) \end{aligned}$$

From Theorem 5.3.8,

$$H(\mathcal{A}|\mathcal{C}) = 0 \implies \mathcal{A}\dot{\subseteq}\mathcal{C}$$

For the converse, with $\mathcal{A}\dot{\subseteq} \bigvee_{i=0}^{\infty} T^{-i}\mathcal{A}$, using Theorem 5.4.9,

$$h(T, \mathcal{A}) = H(\mathcal{A}|\bigvee_{i=0}^{\infty} T^{-i}\mathcal{A})$$

By Theorem 5.3.8,

$$= 0$$

□

Corollary 5.4.9 (ii): For a measure-preserving transformation T over a probability space $(X, \mathcal{F}, \mu)$, $h(T) = 0$ if and only if for every finite sub-algebra $\mathcal{A}$ of $\mathcal{F}$, $\mathcal{A}\dot{\subseteq} \bigvee_{i=0}^{\infty} T^{-i}\mathcal{A}$.

Corollary 5.4.9 (iii): For a measure-preserving transformation T over a probability space $(X, \mathcal{F}, \mu)$,$h(T) = 0$ then $T^{-1}\mathcal{F}\doteq\mathcal{F}$.

Proof. Let $F \in \mathcal{F}$. Define a finite algebra as

$$\mathcal{A} := \{\phi, F, X \setminus F, X\}$$

Then by Corollary 5.4.9(i),

$$h(T) = 0 \implies \mathcal{A}\dot{\subseteq} \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}$$

But,

$$\bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}\dot{\subseteq}T^{-1}\mathcal{F}$$

So,

$$\mathcal{A}\dot{\subseteq}T^{-1}\mathcal{F}$$

As F was chosen arbitrarily to construct $\mathcal{A}$,

$$\begin{aligned} &\mathcal{F}\dot{\subseteq}T^{-1}\mathcal{F} \\ \implies &\mathcal{F}\doteq T^{-1}\mathcal{F} \end{aligned}$$

□

Corollary 5.4.9 (iv): For a measure-preserving transformation T over a probability space $(X, \mathcal{F}, \mu)$, let $h(T) = 0$. Then for a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$ with $T^{-1}\mathcal{A}\dot{\subseteq}\mathcal{A}$, implies $T^{-1}\mathcal{A}\doteq\mathcal{A}$.

Proof. Since it is given that $T^{-1}\mathcal{A}\dot{\subseteq}\mathcal{A}$, we can restrict the measure-preserving transformation T to $\mathcal{A}$, as $T|_{(X,\mathcal{A},\mu)}$. It can be seen that over this restriction the entropy is also zero for T. Then applying Corollary 5.4.9(iii),

$$T^{-1}\mathcal{A}\doteq\mathcal{A}$$

□

5.5 Methods for Calculating the Entropy of a Transformation

The way we defined entropy requires to calculate the entropy associated with every finite sub-σ-algebra. In the practical sense this is impossible to calculate. So there has to be some way to ease the process of calculating the entropy of T. Ideally we wish to calculate to the entropy of T with just one finite sub-σ-algebra $\mathcal{A}$. Hence this section will explore possibility of the existence of such a finite sub-σ-algebra $\mathcal{A}$.

Lemma 5.5.1: Suppose $r \geq 1$ be an integer. For any arbitrary $\epsilon > 0$, there exists $\delta > 0$ such that for any two partitions, $\xi = \{A_1, A_2, ..., A_r\}$ and $\eta = \{B_1, B_2, ..., B_r\}$ of $(X, \mathcal{F}, \mu)$with $\sum_{i=0}^{r} \mu(A_i \Delta B_i) < \delta$, we have

$$H(\xi|\eta) + H(\eta|\xi) < \epsilon$$

Proof. Let $\epsilon > 0$ be some arbitrary number. Choose δ such that,

$$0 < \delta < \frac{1}{4} \qquad \text{and} \qquad -r(r-1)\ .\ \delta \log \delta - (1-\delta)\log(1-\delta) < \frac{\epsilon}{2}$$

Construct a partition Ω such that it consists of the sets $A_i \cap B_j$ (where $i \neq j$)and $\cup_{i=1}^{r}(A_i \cap C_i)$. Then clearly we can see,

$$\xi \vee \eta = \xi \vee \Omega \ = \eta \vee \Omega$$

Since for $i \neq j,\ A_i \cap C_j \subset \cup_{n=1}^{r}(A_n \Delta B_n)$ and $\sum_{i=1}^{r} \mu(A_i \Delta B_i) < \delta$ from the hypothesis,

$$\mu(A_i \cap C_j) < \delta$$

With $\sum_{i=1}^{r} \mu(A_i \Delta B_i) < \delta$,

$$\mu(\cup_{i=1}^{r}(A_i \Delta B_i)) < \delta$$

or,

$$1 - \mu(\cup_{i=1}^{r}(A_i \cap B_i)) < \delta$$

or,

$$\mu(\cup_{i=1}^{r}(A_i \cap B_i)) > 1 - \delta$$

From the way Ω was constructed,

$$H(\Omega) = -\sum_{i,j,i\neq j} \mu(A_i \cap B_j)\log \mu(A_i \cap B_j) - \mu(\cup_{i=1}^{r}(A_i \cap B_i))\log \mu(\cup_{i=1}^{r}(A_i \cap B_i))$$

But with $\mu(A_i \cap C_j) < \delta$ for $i \neq j$,

$$-\sum_{i,j,i\neq j} \mu(A_i \cap B_j)\log \mu(A_i \cap B_j) < -r(r-1)\ .\ \delta \log \delta \tag{5.7}$$

Eqn.5.7 holds because $\delta < \frac{1}{4}$ and the function $-x \log x$ is increasing for $x \in [0, \frac{1}{e}]$. And with $\mu(\cup_{i=1}^{r}(A_i \cap B_i)) > 1 - \delta$,

$$-\mu(\cup_{i=1}^{r}(A_i \cap B_i))\log \mu(\cup_{i=1}^{r}(A_i \cap B_i)) < -(1-\delta)\log(1-\delta) \tag{5.8}$$

Eqn.5.8 holds because $1-\delta > \frac{3}{4}$ and the function $-x\log x$ is increasing for $x \in [\frac{1}{e}, \infty]$.
Therefore from equations 5.7 and 5.8,

$$H(\Omega) < -r(r-1)\ .\ \delta\log\delta - (1-\delta)\log(1-\delta) < \frac{\epsilon}{2}$$

Hence,

$$\begin{aligned} H(\xi) + H(\eta|\xi) &= H(\xi \vee \eta) \\ &= H(\xi \vee \Omega) \\ &\leq H(\xi) + H(\Omega) \\ &\leq H(\xi) + \frac{\epsilon}{2} \\ \implies H(\eta|\xi) &< \frac{\epsilon}{2} \end{aligned} \tag{5.9}$$

Similarly,

$$\begin{aligned} H(\eta) + H(\xi|\eta) &= H(\eta \vee \xi) \\ &= H(\eta \vee \Omega) \\ &\leq H(\eta) + H(\Omega) \\ &\leq H(\eta) + \frac{\epsilon}{2} \\ \implies H(\xi|\eta) &< \frac{\epsilon}{2} \end{aligned} \tag{5.10}$$

So from eqn. 5.9 and 5.10,

$$\implies H(\xi|\eta) + H(\eta|\xi) < \epsilon$$

$\square$

Theorem 5.5.2: On a probability space $(X, \mathcal{F}, \mu)$, let $\mathcal{F}_0$ be an algebra such that the σ-algebra generated by $\mathcal{F}_0$, $\mathscr{B}(\mathcal{F}_0) \doteq \mathcal{F}$. Let $\mathcal{C}$ be a finite sub-algebra of $\mathcal{F}$. Then for any arbitrary $\epsilon > 0$, there exists a finite sub-algebra $\mathcal{D} \dot{\subseteq} \mathcal{F}_0$ such that $H(\mathcal{C}|\mathcal{D}) + H(\mathcal{D}|\mathcal{C}) < \epsilon$.

Proof. Let $\mathcal{C} = \{C_1, C_2, ..., C_r\}$. Fix a number ω and choose λ such that ,

$$\lambda\ (r-1)[1 + r(r-1)] < \omega$$

Choose $F_i \in \mathcal{F}_0$ for each i, such that $\mu(C_i \Delta F_i) < \lambda$. For $i \neq j$,

$$F_i \cap F_j \subset (C_i \Delta F_i) \cup (C_j \Delta F_j)$$

so,

$$\mu(F_i \cap F_j) < 2\ \lambda$$

Then if we set $R = \bigcup_{i \neq j}(F_i \cap F_j)$, then $R < r\ (r-1)\ \lambda$. So assign for $1 \leq i < r$, $D_i = F_i \setminus R$ and $D_r = X \setminus \cup_{i=1}^{r-1} D_i$. Clearly $\mathcal{D} = \{D_1, D_2, ..., D_r\}$ forms a partition over X. Note that for each i, $D_i \in \mathcal{F}_0$.

For $i < r$,

$$\begin{aligned} C_i \Delta D_i &\subset R \cup (C_i \Delta F_i) \\ \implies \mu(C_i \Delta D_i) &< \lambda\ [1 + r(r-1)] < \omega \end{aligned}$$

For $i = r$,

$$\begin{aligned} C_r \Delta D_r &\subset \bigcup_{i=1}^{r-1} (C_i \Delta D_i) \\ \implies \mu(C_r \Delta D_r) &< \lambda\ (r-1)[1 + r(r-1)] < \omega \end{aligned}$$

Hence for each ω we can find a partition $\mathcal{D}$ such that for every $0 < i \leq r$, $\mu(C_i \Delta D_i) < \omega$. Then the proof is complete by invoking Lemma 5.5.1.

$\square$

Corollary 5.5.2: Suppose there is an increasing sequence $\{\mathcal{A}_n\}$ of finite sub-algebras of $\mathcal{F}$. Let $\mathcal{C}$ be a finite sub-algebra of $\mathcal{F}$ with $\mathcal{C}\dot{\subseteq}\bigvee_n \mathcal{A}$. Then $H(\mathcal{C}\setminus\mathcal{A}_n)\to 0$ as $n\to\infty$.

Proof. Let $\mathcal{F}_0=\cup_{i=1}^{\infty}\mathcal{A}_i$. Clearly $\mathcal{F}_0$ is an algebra. Since $\mathcal{C}\dot{\subseteq}\bigvee_n \mathcal{A}$, $\mathcal{C}\dot{\subseteq}\mathscr{B}(\mathcal{F}_0)$. From Theorem 5.5.2, for any $\epsilon>0$ there is a finite sub-algebra $\mathcal{D}_\epsilon$ of $\mathcal{F}_0$ such that $H(\mathcal{C}\setminus\mathcal{D}_\epsilon)<\epsilon$. But $\mathcal{D}_\epsilon$ being finite there is some α such that $\mathcal{D}_\epsilon\dot{\subseteq}\mathcal{A}_\alpha$. Then if we choose any $i>\alpha$, $\mathcal{D}_\epsilon\dot{\subseteq}\mathcal{A}_i$. Hence using the fact that $\{\mathcal{A}_n\}$ is an increasing sequence,

$$\begin{aligned}H(\mathcal{C}\setminus\mathcal{A}_i)&\le H(\mathcal{C}\setminus\mathcal{A}_\alpha)\\&\le H(\mathcal{C}\setminus\mathcal{D}_\epsilon)\\&\le\epsilon\end{aligned}$$

Then as $n\to\infty$,

$$H(\mathcal{C}\setminus\mathcal{A}_n)<\epsilon$$

□

The following theorem gives to the solution of finding a finite sub-algebra such that the entropy of the transformation is same as the entropy of the same associated with this sub-algebra.

Theorem 5.5.3: (Kolmogorov-Sinai Theorem) For a measure-preserving transformation T over a probability space $(X,\mathcal{F},\mu)$ and a finite sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$, suppose $\bigvee_{n=-\infty}^{\infty}T^n\mathcal{A}\doteq\mathcal{F}$. Then,

$$h(T)=h(T,\mathcal{A})$$

Proof. Let $\mathcal{C}$ be some finite sub-algebra of $\mathcal{F}$. If the entropy associated with $\mathcal{A}$ is a supremum over the values of entropy with respect to a sub-algebra, then the proof is done.
Using the property $h(T,\mathcal{A})\le h(T,\mathcal{C})+H(\mathcal{A}|\mathcal{C})$ and $n\ge 1$,

$$\begin{aligned}h(T,\mathcal{C})&\le h(T,\bigvee_{i=-n}^{n}T^i\mathcal{A})+H(\mathcal{C}|\bigvee_{i=-n}^{n}T^i\mathcal{A})\\&=h(T,\mathcal{A})+H(\mathcal{C}|\bigvee_{i=-n}^{n}T^i\mathcal{A})\end{aligned}$$

Set $\mathcal{A}_n:=\bigvee_{i=-n}^{n}T^i\mathcal{A}$. Note that $\bigvee_{i=-n}^{n}T^i\mathcal{A}\doteq\mathcal{F}$. So clearly, $\mathcal{C}\dot{\subseteq}\bigvee_{i=-n}^{n}T^i\mathcal{A}$.Thus from Corollary 5.5.2, $H(\mathcal{C}\setminus\mathcal{A}_n)\to 0$ as $n\to\infty$. So for any finite sub-algebra $\mathcal{C}$, $h(T,\mathcal{C})\le h(T,\mathcal{A})$.

$$\therefore h(T)=h(T,\mathcal{A})$$

□

The same result holds when a measure-preserving transformation T is not invertible.

Corollary 5.5.3: For an invertible measure-preserving transformation T over a probability space $(X,\mathcal{F},\mu)$ and $\bigvee_{n=0}^{\infty}T^n\mathcal{A}\doteq\mathcal{F}$ for some finite sub-algebra $\mathcal{A}$, then $h(T)=0$.

Proof. From Theorem 5.5.3,

$$\begin{aligned}h(T)&=h(T,\mathcal{A})\\&=\lim_{n\to\infty}H(\mathcal{A}|\bigvee_{i=0}^{n}T^{-i}\mathcal{A})\end{aligned}$$

But $\bigvee_{i=0}^{n}T^{-i}\mathcal{A}\doteq T^{-1}\mathcal{F}\doteq\mathcal{F}$. Set $\mathcal{A}_n:=\bigvee_{i=1}^{n}T^{-i}\mathcal{A}$, then $\mathcal{A}_1\dot{\subseteq}\mathcal{A}_2\dot{\subseteq}$..... and $\bigvee_{n=1}^{\infty}\mathcal{A}_n\doteq\mathcal{F}$. Then by Corollary 5.5.2, $H(\mathcal{A}|\mathcal{A}_n)\to 0$. Thus,

$$h(T)=0$$

□

Definition 5.5.4: A countable partition ξ of X is called a **generator** for an invertible measure-preserving transformation T if

$$\bigvee_{n=-\infty}^{\infty} T^n\mathcal{A}(\xi)\dot{=}\mathcal{F}$$

It can be shown that over a Lebesgue space, an ergodic invertible measure-preserving transformation will have a generator.

Theorem 5.5.5: Consider a probability space $(X, \mathcal{F}, \mu)$. Let $\mathcal{F}_0$ be a sub-algebra of $\mathcal{F}$ with $\mathscr{B}(\mathcal{F}_0)\dot{=}\mathcal{F}$. Then for each measure-preserving transformation,

$$h(T) = \sup h(T, \mathcal{A})$$

where supremum is taken over all the finite sub-algebras of $\mathcal{A} \in \mathcal{F}_0$.

Proof. Suppose $\mathcal{C}$ be some finite sub-algebra of $\mathcal{F}$. Then for any $\epsilon > 0$ there exists a finite sub-algebra $\mathcal{D}_\epsilon \in \mathcal{F}_0$ such that $H(\mathcal{C}|\mathcal{D}) + H(\mathcal{D}|\mathcal{C}) < \epsilon$.

Thus,

$$\begin{aligned}
h(T, \mathcal{C}) &\leq h(T, \mathcal{D}_\epsilon) + H(\mathcal{C}|\mathcal{D}_\epsilon) \\
&\leq h(T, \mathcal{D}_\epsilon) + \epsilon \\
\therefore h(T, \mathcal{C}) &\leq \sup\{h(T, \mathcal{D}) : \mathcal{D}\dot{\subseteq}\mathcal{F}_0, \mathcal{D} \text{ is finite}\}
\end{aligned}$$

The other way of the inequality can also be shown as,

$$\begin{aligned}
h(T, \mathcal{D}_\epsilon) &\leq h(T, \mathcal{C}) + H(\mathcal{D}_\epsilon|\mathcal{C}) \\
&\leq h(T, \mathcal{C}) + \epsilon
\end{aligned}$$

□

Theorem 5.5.6: Consider two probability spaces, $(X_1, \mathcal{F}_1, \mu_1)$ and $(X_2, \mathcal{F}_2, \mu_2)$ let $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ be their respective measure-preserving transformations. Then,

$$h(T_1 \times T_2) = h(T_1) + h(T_2)$$

Proof. Let $\mathcal{A}_1\dot{\subseteq}\mathcal{F}_1$ and $\mathcal{A}_2\dot{\subseteq}\mathcal{F}_2$ be finite. Then $\mathcal{A}_1 \times \mathcal{A}_2$ is also finite. The partition generated by $\mathcal{A}_1 \times \mathcal{A}_2$ is given as,

$$\xi(\mathcal{A}_1 \times \mathcal{A}_2) = \{A_1 \times A_2 : A_1 \in \xi(\mathcal{A}_1) \text{ and } A_2 \in \xi(A_2)\}$$

Hence the entropy of $T_1 \times T_2$ is given by,

$$h(T_1 \times T_2) = \sup\{h(T_1 \times T_2, \mathcal{A}_1 \times \mathcal{A}_2)\}$$

And so,

$$h(\bigvee_{i=0}^{n-1} (T_1 \times T_2)^{-i}(\mathcal{A}_1 \times \mathcal{A}_2)) = H(\bigvee_{i=0}^{n-1} T_1^{-i}\mathcal{A}_1 \ \times \ \bigvee_{i=0}^{n-1} T_2^{-i}\mathcal{A}_2)$$

With $\{C_k\} \in \xi(\bigvee_{i=0}^{n-1} T_1^{-i}\mathcal{A}_1)$ and $\{D_j\} \in \xi(\bigvee_{i=0}^{n-1} T_2^{-i}\mathcal{A}_2)$,

$$\begin{aligned}
&= -\sum_{k,j} (\mu_1 \times \mu_2)(C_k \times D_j) \log[(\mu_1 \times \mu_2)(C_k \times D_j)] \\
&= -\sum_k \sum_j \mu_1(C_k)\ \mu_2(D_j) \log[\mu_1(C_k)\ \mu_2(D_j)] \\
&= -\sum_k \mu_1(C_k) \log \mu_1(C_k) - \sum_j \mu_2(D_j) \log \mu_2(D_j)
\end{aligned}$$

Thus,

$$\begin{aligned}
H(T_1 \times T_2, \mathcal{A}_1 \times \mathcal{A}_2) &= H(T_1, \mathcal{A}_1) + H(T_2, \mathcal{A}_2) \\
\implies h(T_1 \times T_2) &= h(T_1) + h(T_2)
\end{aligned}$$

□

5.6 Examples

1. For the identity map I over the space $(X, \mathcal{F}, \mu)$, for every sub-algebra $\mathcal{A}$,

$$h(I, \mathcal{A}) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{A}) = 0$$

Hence the entropy of an identity map is zero. Also for a transformation T with a period $p \neq 0$, i.e., $T^p = I$, the entropy is zero. This is because from the Theorem 5.4.8,

$$0 = h(T^p) = |p| h(T)$$

2. Theorem 5.6.1: Any rotation $T(z) = az$ on a unit circle, has zero entropy.

Proof. To show this we will consider two cases. First suppose that the set $\{a^n : n \in \mathbb{Z}\}$ is not dense. This means that a is root of unity. So there is some $p \neq 0$ such that $a^p = 1$. This is same as writing,

$$\begin{aligned}
T^p(z) &= z \\
\implies h(T) &= 0
\end{aligned}$$

The other case is where the set $\{a^n : n \in \mathbb{Z}\}$ is dense in the unit circle.

Construct a partition on the circle as $\xi = \{A_1, A_2\}$, where A_1 is the upper half circle and A_2 is lower half circle, as given in the figure. Then the transformations $T^{-n}(\xi)$ for some $n \geq 0$, take the form of semi circles having values from $-a^n$ to a^n.

Note that $\bigvee_{n=0}^{\infty} T^{-n} \mathcal{A}(\xi)$contains all arcs possible on the unit circle. This means,

$$\mathcal{F} \doteq \bigvee_{n=0}^{\infty} T^{-n} \mathcal{A}(\xi)$$

But $\mathcal{A}(\xi)$ is a finite sub-algebra and by Corollary 5.5.3,

$$h(T) = 0$$

□

3. Consider an n-torus, $X = K^n$ with the transformation,

$$T(z_1, z_2, ..., z_n) = (a_1 z_1, a_2 z_2, ..., a_n z_n)$$

Here, $T = T_1 \times T_2 \times ... \times T_n$ and for all i, $T_i(z) = a_i z$. But from the above example, $h(T_i) = 0$ for every i. Therefore, from Theorem5.5.6,

$$h(T) = \sum_{i=0}^{n} h(T_i) = 0$$

4. Entropy of a doubling map is non-zero. Let T be such a map, i.e.,

$$\begin{aligned}
T &: [0, 1) \to [0, 1) \\
T(x) &= 2x \mod 1
\end{aligned}$$

Let $\alpha = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$. Then,

$$\alpha \vee T^{-1}(\alpha) = \{[0, \frac{1}{4}), [\frac{1}{4}, \frac{1}{2}), [\frac{1}{2}, \frac{3}{4}), [\frac{3}{4}, 1)\}$$

More generally, $\alpha \vee T^{-1}(\alpha) \vee T^{-1}(\alpha) \vee ... \vee T^{-(n-1)}(\alpha)$ forms a partition of $[0, 1)$ into 2^n intervals of equal length $\frac{1}{2^n}$. Thus,

$$
\begin{aligned}
H(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)) &= -\sum_{i=0}^{2^{n-1}} \mu(A_i) \log \mu(A_i) \\
&= -\sum_{i=0}^{2^{n-1}} \frac{1}{2^n} \log \frac{1}{2^n} \\
&= \frac{2^n}{2^n}\ n \log 2
\end{aligned}
$$

Hence,

$$
\begin{aligned}
h(T, \mathcal{A}) &= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}(\alpha)) \\
&= \lim_{n\to\infty} \frac{1}{n}\ .\ n \log 2 \\
&= \log 2
\end{aligned}
$$

But $h(T) \geq h(T, \mathcal{A})$ for all $\mathcal{A}$. So,

$$h(T) > 0$$

5.7 Topological Dynamics

Before heading into explaining how topological entropy is defined, certain topics needed to be discussed to fully grasp the notions behind Adler's Topological Entropy.

For the ease of notations, from here on wards the space X *will denote a compact metric space unless stated otherwise.*

Definition 5.7.1: A homeomorphism $T : X \to X$ is said to be **minimal** if for all $x \in X$, the set $\{T^n(x) : n \in \mathbb{Z}\}$ is dense in X.

Recall that the orbit of $x \in X$ under T was defined as $O_T(x) = \{T^n(x) : n \in \mathbb{Z}\}$. Thus if X is minimal, then all the orbits of T must be dense.

Theorem 5.7.2: Consider homeomorphism $T : X \to X$. Then the following are equivalent,

1. T is minimal.
2. $\emptyset$ and X are the only closed subsets E of X with $TE = E$.
3. If U is a non-empty open subset of X then $\bigcup_{-\infty}^{\infty} T^n(U) = X$.

Proof. For (1) $\implies$ (2), suppose that E is closed but $E \neq \emptyset$ and $TE = E$. Clearly if $x \in E$ then the orbit, $O_T(x) \subset E$. But $O_T(x)$ is dense in X. That is, $X = \overline{O_T(x)} \subset E$. Hence $E = X$.

For (2) $\implies$ (3), Let U be a non-empty open set. Then clearly $E = X \setminus \cup_{n=-\infty}^{\infty} T^n(U)$ is closed and $TE = E$. This implies either E is X or $\emptyset$. But $E \neq X$ so $E = \emptyset$. Thus $\cup_{n=-\infty}^{\infty} T^n(U) = X$.
For (3) $\implies$ (1), Suppose $x \in X$ and U be a non-empty open subset of X. Then $x \in T^n(U)$ for some $n \in \mathbb{Z}$. So, $T^{-n}(x) \in U$. Since this holds for any open subset U, the orbit, $O_T(x)$ is dense in X and thus T is minimal.

□

Note that for a subet E of X which is closed and T-invariant, $T|_E$ is a homeomorphism of the compact metric space E.

Definition 5.7.3: For a homeomorphism $T : X \to X$, a T-invariant closed set E of X is said to be a minimal set with respect T if $T|_E$ is minimal.

It can be shown with the help of Zorn's lemma that any homeomorphism $T : X \to X$ has a minimal set.

Theorem 5.7.4: Let G be a compact metric group and $T(x) = ax$. Then T is minimal if and only if $\{a^n : n \in \mathbb{Z}\}$ is dense in G.

Proof. Let the identity element of G be denoted as e. Then the orbit of e, $O_T(e) = \{a^n \,.\, e = a^n : n \in \mathbb{Z}\}$. And the orbit of e is dense because T is minimal.

For the converse, suppose the set $\{a^n : n \in \mathbb{Z}\}$ is dense in G. Let $x, y \in G$.Then there is sequence $\{n_i\}$ such that,

$$a^{n_i} \to y\ x^{-1}$$

or,

$$a^{n_i}\ x \to y$$

or,

$$T^{n_i}(x) \to y$$

Therefore for all $x \in G$,

$$O_T(x) \text{ is dense in } G$$

□

The following is a particular transformation which will be used as example for wandering sets.

5.7.1 The North-South Map

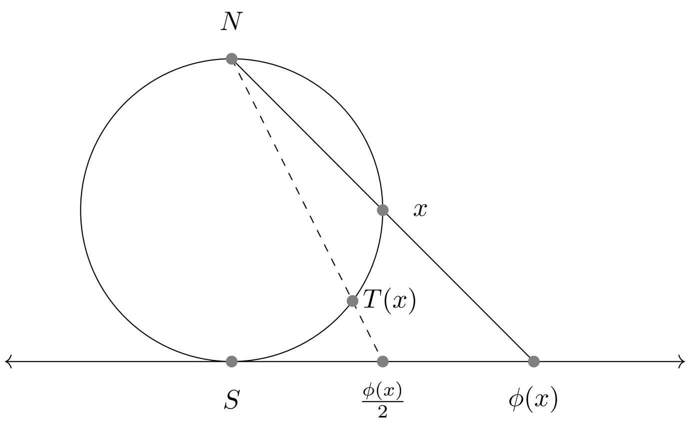


Figure 5.2: The North-South Map

Consider a unit circle in $\mathbb{R}^2$, centered at $(0,1)$. Let the north pole , $N = (0,2)$ and the south pole, $S = (0,0)$. Let ϕ be a map defined as, $\phi : X \setminus \{N\} \to \mathbb{R} \times \{0\}$, which takes maps a point $x \in X \setminus \{N\}$on the unit circle to the point on the x-axis intersected by the line that joins the north pole and x. Then we define the north-south map as,

$$T : X \to X$$
$$T(x) = \begin{cases} \phi^{-1}(\frac{1}{2}\phi(x)), \text{ if } x \in X \setminus \{N\} \\ N, \text{ if } x = N \end{cases}$$

It can be observed that $T(N) = N$ and $T(S) = S$. And when $x \neq N, S$, then $T^n(x) \to S$ as $n \to \infty$. Also see that this transformation is not minimal as the orbit of the north pole is not dense.

5.7.2 The Non-Wandering Sets

Definition 5.7.5: For a continuous transformation T on a compact metric space X and $x \in X$, the ω-limit set of x is the collection of all the limit points of $\{T^n(x) : n \geq 0\}$, i.e.,

$$\omega(x) = \{y \in X : \exists n_i \to \infty \text{ with } T^{n_i}(x) \to y\}$$

Theorem 5.7.6: For a continuous transformation T on a compact metric space X and $x \in X$,

1. $\omega(x) \neq \emptyset$
2. $\omega(x)$ is a closed subset of x.
3. $T(\omega(x)) = \omega(x)$

Proof. For (1) the proof is obvious as T is continuous transformation.
For (2), let $y_k \in \omega(x)$ with $y_k \to y \in X$. For each$i \geq 1$, choose a k_i such that.

$$d(y_{k_j}, y) < \frac{1}{i}$$

And choose n_i such that.

$$d(T^{n_i}(x), y_{k_i}) < \frac{1}{i}$$

Then,

$$d(T^{n_i}(x), y) < \frac{1}{i}$$
$$\implies y \in \omega(x)$$

For (3), it is clear that $T(\omega(x)) \subset \omega(x)$. If $y \in \omega(x)$ then there exist some sequence $\{n_i\}$ such that $T^{n_i}(x) \to y$. Also there must be a subsequence $\{n_{i_j}\}$ such that $T^{n_{i_j}-1}(x) \to z$ for some $z \in X$. .Then, $T^{n_{i_j}}(x) \to T(z)$ so that $T(z) = y$. Clearly $z \in \omega(x)$. Then

$$y = T(z) \in T(\omega(x))$$
$$\implies \omega(x) \subset T(\omega(x))$$
$$\therefore \omega(x) = T(\omega(x))$$

□

Definition 5.7.7: Suppose T is a continuous transformation on X. An $x \in X$ is called wandering for T if there is an open set U such that for all the sets $T^{-n}U$ are mutually disjoint. Hence the non-wandering sets for T , denoted by, $\Omega(T)$, are

$$\Omega(T) = \{x \in X : \text{ for every neighbourhood } U \text{ of } x, \ \exists n \geq 1 \text{ with } T^{-n}U \cap U \neq \emptyset\}$$

Observe that for a homeomorphism T,

$$T^{-n}(U) \cap U = T^{-n}(U \cap T^n U)$$

So,

$$\Omega(T^{-1}) = \Omega(T)$$

Theorem 5.7.8: For a continuous transformation T on X,

1. $\Omega(T)$ is closed.

2. $\cup_{x\in X}\omega(x) \subset \Omega(T)$; in particular, $\Omega(T) \neq \emptyset$.

3. $\Omega(T)$ contains all periodic points of X for T.

4. $T(\Omega(T)) \subset \Omega(T)$ and the equality holds when T is a homeomorphism.

Proof. For (1), $X \setminus \Omega(T)$ is open from the definition of $\Omega(T)$. Therefore $\Omega(T)$ is closed.

For(2), Suppose $x \in X$ and $y \in \omega(x)$, Suppose V is some neighbourhood of y. If we can show that for some $n \geq 1$, $T^n V \cap V \neq \emptyset$ then we are done. For this, consider a subsequence $\{n_i\}$ such that $T^{n_i}(x) \to y$. Choose n_{i_a} and n_{i_b} from this subsequence such that $n_{i_a} < n_{i_b}$ with $T^{n_{i_a}}(x) \in V$ and $T^{n_{i_b}}(x) \in V$. Set $n = n_{i_a} - n_{i_b}$ and $z = T^{n_{i_a}}(x)$. So we can see that,

$$z = T^{n_{i_a}}(x) \in V$$

and,

$$T^n(z) = T^{n+n_{i_a}}(x) = T^{n_{i_b}}(x) \in V$$

For(3), if x is a periodic point and thus for some $n \geq 1$, $T^n(x) = x$. So for a neighbourhood V of x,

$$x \in T^{-n}V \cap V$$

$$\implies x \in \Omega(T)$$

For (4), let $x \in \Omega(T)$. Suppose V is a neighbourhood of $T(x)$. Then obviously, $T^{-1}V$ is a neighbourhood of x. Then there is some $n \geq 1$ such that,

$$\begin{aligned} T^{-(n+1)}V \cap T^{-1}V &\neq \emptyset \\ \therefore T^{-n}V \cap V &\neq \emptyset \\ &\implies T(x) \in \Omega(T) \end{aligned}$$

And if T is homeomorphic, $\Omega(T) = \Omega(T^{-1})$. So

$$\begin{aligned} T^{-1}\Omega(T) &\subset \Omega(T) \\ \therefore T(\Omega(T)) &= \Omega(T) \end{aligned}$$

□

The points from the north-south map apart from the poles are wandering. To show this, let y be point lying between $T^{-1}(x)$ and x . Then clearly $T(y)$ will lie between x and $T(x)$.

Suppose U is an open arc between y and $T(y)$. Then U is a neighbourhood of x. So for any $n \in \mathbb{Z}$, T^nU is an open arc between T^nU and $T^{n+1}U$. Hence, the collection $\{T^n(U)\}_{-\infty}^{\infty}$ are pair-wise disjoint. Therefore x is wandering and $\Omega(T) = \{N, S\}$.

5.8 Invariant Measures for Continuous Transformations

Definition 5.8.1: Let $M(X)$ denote the collection of all probability measures defined on the measurable space $(X, \mathscr{B}(X))$, where $\mathscr{B}(X)$ is the Borel σ-algebra generated on X.. The elements of $M(X)$ are called the Borel Probability Measures.

For a continuous transformations T over the compact metrisable space X, the collection $\{A \in \mathscr{B}(X) : T^{-1}(A) \in \mathscr{B}(X)\}$ forms a σ-algebra and contains all opens sets. So,

$$T^{-1}\mathscr{B}(X) \subset \mathscr{B}(X)$$

in other words, T is measurable. This helps us define the following map,

$$\begin{aligned}\tilde{T} &: M(X) \to M(X)\\ (\tilde{T}\mu)(A) &= \mu(T^{-1}(A)\end{aligned}$$

Lemma 5.8.2: For a continuous transformations T over the compact metrisable space X, and let $\tilde{T} : M(X) \to M(X)$ as defined above. Then for every $f \in C(X)$,

$$\int f \ d(\tilde{T}\mu) = \int f \circ T \ d\mu$$

From definition of $\tilde{T}$, integrating the characteristic function χ_A under the measure $\tilde{T}\mu$ is same as integrating $\chi_A \circ T$ under μ, i.e.,

$$\int \chi_A \ d(\tilde{T}\mu) = \int \chi_A \circ T \ d\mu$$

This would hold for a simple function as it can be seen as the limit of an increasing sequence of characteristic functions. Hence for any non-negative measurable functions f would satisfy this relation. Since any functions $f \in C(X)$ can be split into $f = f^+ + f^-$, the same relation holds for any f.

Theorem 5.8.3: The map $\tilde{T}$ is continuous and affine.

Proof. Suppose there is sequence $\{\mu_n : \mu_n \in M(X)\}$ such that $\mu_n \to \mu \in M(X)$. From the previous lemma 5.8.2, for $f \in C(X)$, $\int f \ d(\tilde{T}\mu) = \int f \circ T \ d\mu$. So ,

$$\int f \ d(\tilde{T}\mu_n) = \int f \circ T \ d\mu_n \to \int f \circ T \ d\mu = \int f \ d(\tilde{T}\mu)$$

Then,

$$\tilde{T}\mu_n \to \tilde{T}\mu$$

Thus shows that $\tilde{T}$ is continuous.

Let $\mu_1, \mu_2 \in M(X)$ and $p \in [0, 1]$. So, for all $A \in \mathscr{B}(X)$,

$$\begin{aligned}\tilde{T}(p\mu_1 + (1-p)\mu_2)(B) &= p \ \mu_1(T^{-1}A) + (1-p) \ \mu_2(T^{-1}A)\\ &= (p \ \tilde{T}\mu_1 + (1-p) \ \tilde{T}\mu_2)(A)\end{aligned}$$

Thus $\tilde{T}$ is affine. □

Definition 5.8.4: Define $M(X, T) := \{\mu \in M(X) : \tilde{T}\mu = \mu\}$. This set consists of all the measures that make T measure-preserving for the the measure space, $(X, \mathscr{B}(X), \mu)$.

Theorem 5.8.5: For a continuous transformations T over the compact metrisable space X, consider a sequence $\{\nu_n\}_{n=1}^{\infty}$ in $M(X)$. Define a new set sequence from $\{\nu_n\}_{n=1}^{\infty}$ as $\{\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} \tilde{T}^i \nu_n\}$. Then $M(X, T)$ contains all the limit points of $\{\mu_n\}_{n=1}^{\infty}$.

Proof. Note that the existence of the limit point is ensured with the compactness of $M(X)$and $M(X)$ has weak-* topology. Let μ be a limit point in $M(X)$. Then there is some sequence $\{n_i\}$ such that

$\mu_{n_i} \to \mu$. For $f \in C(X)$,

$$
\begin{aligned}
|\int f \circ T \, d\mu - \int f \, d\mu| &= \lim_{j\to\infty} |\int f \circ T \, d\mu_{n_i} - \int f \, d\mu_{n_i}| \\
&= \lim_{j\to\infty} |\int (\frac{1}{n_i}\sum_{j=0}^{n_i-1} f \circ T^{j+1}) \, d\nu_{n_i} - (\frac{1}{n_i}\sum_{j=0}^{n_i-1} f \circ T^{j}) \, d\nu_{n_i}| \\
&= \lim_{i\to\infty} |\frac{1}{n_i}\int \sum_{j=0}^{n_i-1} (f \circ T^{j+1}) - f \circ T^{j})) \, d\nu_{n_i}| \\
&= \lim_{i\to\infty} |\frac{1}{n_i}\int (f \circ T^{n_i} - f) \, d\nu_{n_i}| \\
&\leq \lim_{i\to\infty} \frac{2||f||}{n_i} \\
&= 0
\end{aligned}
$$

Thus $\mu \in M(X, T)$. ☐

As a result of this we get the following result.

Corollary 5.8.5: (Krylov and Bogolioubov Theorem) For a continuous transformations T over the compact metric space X, $M(X, T)$ is non-empty.

Lemma 5.8.6: For a continuous transformations T over the compact metrisable space X, μ is an extreme point of $M(X, T)$ if and only if T is an ergodic measure-preserving transformation of $(X, \mathscr{B}(X), \mu)$.

Proof. Suppose $\mu \in M(X, T)$ is not ergodic. Then for an T-invariant Borel measure set E will satisfy, $0 < \mu(E) < 1$. Let for $A \in \mathscr{B}(X)$,

$$
\mu_1(A) = \frac{\mu(A \cap E)}{\mu(E)} \qquad \text{and} \qquad \mu_2(A) = \frac{\mu(A \cap (X \setminus E))}{\mu(X \setminus E)}
$$

It can be seen that μ_1 and $\mu_2 \in M(X, T)$. Clearly $\mu_1 \neq \mu_2$ and $\mu(A) = \mu(E)\ \mu_1(A) + (1 - \mu(E))\ \mu_2(A)$. This says that μ is not an extreme point of $M(X, T)$.

For the converse, let $\mu \in M(X, T)$ be ergodic. Let $\mu_1, \mu_2 \in M(X, T)$ such that

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

where $p \in [0, 1]$. If $\mu_1 = \mu_2$ then clearly μ is an extreme point. Also μ is absolutely continuous with respect to μ_1. Therefore the Radon Nikodym derivative $\frac{d\mu_1}{d\mu}$ exists. So for any $A \in \mathscr{B}(X)$,

$$
\mu_1(A) = \int_A \frac{d\mu_1(x)}{d\mu} d\mu(x)
$$

The derivative is non-negative. So if we choose the set A as,

$$
A = \{x : \frac{d\mu_1}{d\mu} < 1\}
$$

then,

$$
\begin{aligned}
\int_{A\cap T^{-1}A} \frac{d\mu_1}{d\mu}\, d\mu + \int_{A\setminus T^{-1}A} \frac{d\mu_1}{d\mu}\, d\mu &= \mu_1(A) = \mu_1(T^{-1}A) \\
&= \int_{A\cap T^{-1}A} \frac{d\mu_1}{d\mu}\, d\mu + \int_{T^{-1}A\setminus A} \frac{d\mu_1}{d\mu}\, d\mu \\
\implies \int_{A\setminus T^{-1}A} \frac{d\mu_1}{d\mu}\, d\mu &= \int_{T^{-1}A\setminus A} \frac{d\mu_1}{d\mu}\, d\mu
\end{aligned}
$$

With,

$$
\frac{d\mu_1}{d\mu} < 1 \text{ on } A \setminus T^{-1}A \qquad \text{and} \qquad \frac{d\mu_1}{d\mu} > 1 \text{ on } T^{-1}A \setminus A
$$

and,

$$\begin{aligned}\mu(A \setminus T^{-1}A) &= \mu(A) - \mu(A \cap T^{-1}A) \\ &= \mu(T^{-1}A) - \mu(T^{-1}A \cap A) \\ &= \mu(T^{-1}A \setminus A)\end{aligned}$$

would give,

$$\mu(A \setminus T^{-1}A) = \mu(T^{-1}A \setminus A) = 0$$

Hence $\mu(A\Delta T^{-1}A) = 0$. This shows that A is T-invariant so $\mu(A) = 1$ or $\mu(A) = 0$. But if $\mu(A) = 1$, then $\mu(X) = \int_A \frac{d\mu_1}{d\mu}\, d\mu < \mu(A) = 1$. This is a contradiction to the definition of μ_1 and the fact that $\mu_1(X) = 1$. Thus the only possible case is $\mu(A) = 0$.

Similarly if we choose the set B as,

$$B = \{x : \frac{d\mu_1}{d\mu} > 1\}$$

then $\mu(B) = 0$.

Therefore,

$$\frac{d\mu_1}{d\mu} = 1$$

hence,

$$\mu_1 = \mu$$

i.e., $\mu_1 = \mu_2$ or μ is an extreme point of $M(X,T)$.

□

Definition 5.8.7: A continuous transformation T over the compact metrisable space X is called **uniquely ergodic** if $M(X,T)$ consists of only a single point. In other words There is only one T-invariant Borel probability measure.

Clearly if there is only a single point $M(X,T)$ then that point is an extreme point making T ergodic from theorem 5.8.6.

Theorem 5.8.8: For a homeomorphic transformation T over the compact metrisable space X, let T be uniquely ergodic which makes $M(X,T) = \{\mu\}$. Then T is minimal if and only if for all non-empty open sets U, $\mu(U) > 0$.

Proof. Let T be minimal. So for an open set U,

$$\bigcup_{n=1}^{\infty} T^n U = X$$

so if U has $\mu(U) = 0$ then $\mu(X) = 0$ which is a contradiction.

For the converse, suppose that T is not minimal. Then there is some closed set $K \subset X$ and an integer $n > 0$ such that $T^n K = K$. Note that $\emptyset \neq K \neq X$. Looking at the restriction of T over the closed set K, using the Corollary 5.8.5, $M(K,T)$ is non-empty. Let μ_K be a non-invariant Borel probability measure on K.
Define a new measure $\tilde{\mu}$ on X as,

$$\tilde{\mu}(A) = \mu_K(K \cap A)$$

for all $A \in \mathscr{B}(X)$. Then clearly $\tilde{\mu} \in M(X,T)$ and $\tilde{\mu} \neq \mu$ as $\mu(X \setminus K) > 0$ while $\tilde{\mu}(X \setminus K) = 0$. This says that there are more than one T-invariant Borel probability measure which implies T is not uniquely ergodic. □

Theorem 5.8.9: Let $T(x) = a\ x$ be a rotation on a compact metrisable group G. Then T is uniquely ergodic if and only if T is minimal. The Haar measure is the only invariant measure here.

Proof. The first part of the theorem is proved by the Theorem 5.8.8. Only thing that remains is to show that the Haar measure is the only invariant measure. Note that the Haar measure assigns an invariant volume to the subsets of a locally compact topological groups.

With T being minimal, the set $\{a^n\}_{-\infty}^{\infty}$ is dense in G. For $\mu \in M(G,T)$, every $f \in C(X)$ and every $n \in \mathbb{Z}$,

$$\int f(a^n\ x)\ d\mu(x) = \int f(x)\ d\mu(x)$$

And if $b \in G$, there exists a sequence such that $a^{n_i} \to b$. Applying the dominated convergence theorem here, for every $f \in C(X)$,

$$\begin{aligned}\int f(bx)\ d\mu(x) &= \lim_{i\to\infty} \int f(a^{n_j}x)\ d\mu \\ &= \int f(x)\ d\mu(x)\end{aligned}$$

Hence we can see an invariant volume under every rotation of G. This unique measure μ is the Haar measure. $\square$

5.9 Topological Entropy

Like how partitions helped us in defining entropy of a transformation with respect to a sub-algebra, here we look into how open covers help in defining entropy. Recall that X is a compact metric space.

Definition 5.9.1: For two open covers α and β, over X, their join $\alpha \vee \beta$ is the open made by the sets of the form $A \cap B$ where $A \in \alpha$ and $B \in \beta$.

Definition5.9.2: For two open covers α and β, over X, β is said to be a refinement of α if every member of $\beta$is a subset of α.This can be notated as $\alpha < \beta$.

Definition 5.9.3: For an open cover α over X, if $N(\alpha)$ denote the number of sets in a finite sub-cover of α which has the smallest cardinality. The **entropy of** α is defined as,

$$H(\alpha) = \log N(\alpha)$$

As we can see the open covers are analogous to the partitions discussed earlier in this chapter. Similarly $H(\alpha)$ have all the properties $H(\mathcal{A})$ had.

Definition 5.9.4: For an open cover α over X, if $T : X \to X$ is a continuous map then the **entropy of** T **with respect to the open cover** α is given as,

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

Definition 5.9.5: For a continuous map, $T : X \to X$, the **topological entropy** is,

$$h(T) = \sup_{\alpha} h(T,\alpha)$$

Here the supremum is taken over from all the open covers α of X.

Theorem 5.9.6: For a homeomorphism, $T : X \to X$,

$$h(T) = h(T^{-1})$$

Proof.

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

Using Theorem 5.3.2 (10),

$$\begin{aligned}
&= \lim_{n\to\infty} \frac{1}{n} H(T^{n-1} \bigvee_{i=0}^{n-1} T^{-i}(\alpha)) \\
&= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{i}(\alpha)) \\
&= h(T^{-1},\alpha)
\end{aligned}$$

□

Again here, the definition for the entropy is not easy to calculate. So like before we need to find sub-covers for T whose entropy equals the entropy of T.

Definition 5.9.7: For a compact metric space X, let T be a homeomorphic map. A finite cover α of X is said to be a **generator** for T if for every $\{A_n\}_{-\infty}^{\infty}$ of the members of α the set $\bigcap_{n=-\infty}^{\infty} T^{-n}A_n$ contains at most one point of X.

Definition 5.9.8: A homeomorphism T is said to be **expansive** if there is some $\delta > 0$ with the property that if $x \neq y$ then there exists $n \in \mathbb{Z}$ such that,

$$d(T^n(x), T^n(y)) > \delta$$

Here δ is called the expansive constant for T.

It can be seen that the rotation map is not expansive. This is because if we choose some $x, y \in X$ such that $d(x,y) < \delta$, then for all $n \in \mathbb{Z}$,

$$d(T^n(x), T^n(y)) < \delta$$

Theorem 5.9.9: For an expansive homeomorphism T over a compact metric space X, if α is a generator of T then $h(T) = h(T,\alpha)$.

Proof. Let δ be the expansive constant. Let β be an open cover such that δ is the Lebesgue number of β. Choose $N > 0$ such that the elements of $\bigvee_{n=-N}^{N} T^{-n}\alpha$. Then $\bigvee_{n=-N}^{N} T^{-n}\alpha$ is a refinement of β. So,

$$\begin{aligned}
h(T,\beta) &\leq h(T, \bigvee_{n=-N}^{N} T^{-n}\alpha) \\
&= \lim_{k\to\infty} \frac{1}{k} H(\bigvee_{i=0}^{k} T^{-i}(\bigvee_{n=-N}^{N} T^{-n}\alpha)) \\
&= \lim_{k\to\infty} \frac{1}{k} H(\bigvee_{i=-N}^{N+k-1} T^{-n}\alpha) \\
&= \lim_{k\to\infty} \frac{1}{k} H(\bigvee_{i=0}^{2N+k-1} T^{-n}\alpha) \\
&= \lim_{k\to\infty} \frac{2N+k-1}{k} \frac{1}{2N+k-1} H(\bigvee_{i=0}^{2N+k-1} T^{-n}\alpha) \\
&= h(T,\alpha)
\end{aligned}$$

Since this holds for all open covers β,

$$h(T) = h(T,\alpha)$$

□

5.10 Relating Topological Entropy and Measure-Theoretic Entropy

The relationship between topological entropy and measure-theoretic entropy is also called the variation principle. But before showing the relation, let us define an entropy map.

Definition 5.10.1: The **entropy map** of a continuous transformation $T : X \to X$ is the map that takes $\mu \to h_\mu(T)$which is defined on $M(X, T)$ and has values in $[0, \infty)$.

Theorem 5.10.2: Let $T : X \to X$ be a continuous map for a compact metric space X. Then

$$h(T) = \sup\{h_\mu(T) : \mu \in M(X, T)\}$$

The proof can be referred from Peter Walters' Introduction to Ergodic Theory[1], where the author uses the proof done by M Misiurewicz.

Definition 5.10.3: Let $T : X \to X$ be a continuous transformation on a compact metric space X. A member μ of $M(X, T)$ is called a measure of maximal entropy for T if

$$h_\mu(T) = h(T)$$

Further Reading

The reader can refer to further topics from Walter's Introduction to Ergodic Theory [1]. But I will give a brief overview of subjects that proceed from where I have stopped.

Bowen's definition for topological entropy using separating and spanning sets helped in defining entropy for non-compact spaces.

With Bowen's method one can explore about topological pressure which is a generalization of topological entropy. Similarly there is a measure-theoretic pressure for Borel probability measures. A variation principle similar to that of entropy can be shown to relate the topological and measure-theoretic entropies.

Several ergodic theorems were spawned from Brikhoff's Ergodic theorem and Von Neumann's theorem. Once such is the subadditive ergodic theorem which is a generalization of Birkhoff's theorem. The Oseledet's multiplicative theorem helps in calculating the Lyapunov constants associated with non-linear dynamical systems.

The quasi-invariant measures are measures which preserves sets of measure zero while having a transformation that is not measure-preserving. On topological groups these measures can be seen as a generalization of Haar measures. They are useful in the study of differential dynamical systems and geometry of manifolds.

One could also look into the applications of ergodic theory in Riemannian Geometry such as the ergodicity of the geodesic flow on compact Riemann surfaces.

References

[1] Peter Walters. *An introduction to ergodic theory*, volume 79. Springer Science & Business Media, 2000.

[2] César Ernesto Silva. *Invitation to ergodic theory*, volume 42. American Mathematical Soc., 2008.

[3] Walter Rudin. *Real and complex analysis*. Tata McGraw-hill education, 2006.

[4] G de Barra. Measure theory and integration. *John Willy and sons*, 1981.

[5] Robert A Meyers. *Mathematics of complexity and dynamical systems*. Springer Science & Business Media, 2011.

[6] Claude E Shannon and Warren Weaver. The mathematical theory of information. *Urbana: University of Illinois Press*, 97, 1949.

[7] A Ya Khinchin. *Mathematical foundations of information theory*. Courier Corporation, 2013.

[8] Amie Bowles, Lukasz Fidkowski, Amy E Marinello, Cesar E Silva, et al. Double ergodicity of nonsingular transformations and infinite measure-preserving staircase transformations. *Illinois Journal of Mathematics*, 45(3):999–1019, 2001.

[9] Daniel J Rudolph. *Fundamentals of measurable dynamics: Ergodic theory on Lebesgue spaces*. Clarendon Press Oxford, UK, 1990.

[10] P Billingsley. Probability and measure. 3rd wiley. *New York*, 1995.