Entropy in dynamical systems

Yariv Barsheshat
Department of Mathematics and Statistics
McGill University

March, 2015

A thesis submitted to McGill University in partial fulfillment
of the requirements of a degree of Master of Science.

Résumé

La théorie des systèmes dynamiques était recherchée par des scientifiques et des mathématiciens au cours des derniers siècles. Aussi, l'entropie de Shannon et Gibbs, qui était introduit en premier par Claude Shannon dans son célèbre papier de 1948 [1], a commencé une ére de la recherche des systèmes dynamiques, et, sans doute, était la source de la théorie de l'information. Dans ce thèse, on commence avec une introduction assez complète de la théorie des systèmes dynamiques dans la guise de la théorie de la mesure. On continue avec la définition de l'entropie de Shannon et Gibbs, et nous montrons quelques théorèmes fondamentaux qui vont élucider son connection au systèmes dynamiques. Nous terminons en discutant certaines applications aux domaines de codage et compression.

Abstract

Dynamical systems theory has been present at the forefront of research by scientists and mathematicians for the past few centuries. Furthermore, the Shannon-Gibbs entropy, first proposed by Claude Shannon in his celebrated paper from 1948 [1], helped usher in a new era of dynamical systems research and can arguably be hailed as the source of information theory. In this thesis, we begin with a fairly comprehensive introduction to modern, measure-theoretic dynamical systems theory. We then move on to define the Shannon-Gibbs entropy and prove some fundamental theorems which elucidate its connection to dynamical systems. We finish by discussing some applications to the fields of coding and compression.

Acknowledgments

First and foremost, I would like to sincerely thank my supervisor, Professor Vojkan Jakšić, for his time and patience in guiding me during my time as a student at McGill University. Apart from working with me in building my thesis project at every step, Prof. Jakšić has advised me on many matters, and has gone above and beyond his duties to ensure that my time spent in the Masters' program has been both fruitful and enjoyable.

I would also like to thank Professors Yan Pautrat and Claude-Alain Pillet for their input and suggestions for my thesis. Discussions with them were definitely essential for my understanding of the theory. As well, contributions and corrections by students Jane Panangaden and Sherry Chu were extremely helpful and worthy of gratitude.

Acknowledgment must also be given to the Natural Sciences and Engineering Research Council of Canada (NSERC), whose funding made my research possible.

Last, but certainly not least, I would like to thank my parents, Mia and Menahem, and my sisters, Tal and Noam, for their constant moral and emotional support. Without them, I would be lost.

Contents

1 Introduction

A dynamical system is simply a set of points, referred to as a collection of states, or a state space, with a rule describing how these states evolve in time.

The formal study of dynamical systems from a mathematical perspective has a long and colourful history, arguably beginning with Sir Isaac Newton's development of classical mechanics and calculus in his famous work "Mathematical principles of natural philosophy" published in 1687. Newton went further than any scientist who came before him in describing the motion and interactions of objects in our universe.

Further developments and abstractions were made in the following centuries. Notable works by Poincaré [2], Birkhoff [3] and others helped branch out the study into other areas of focus, such as chaos theory and ergodic theory.

In the late nineteenth century, famed physicist Rudolf Clausius introduced the concept of entropy in his famous book, *The mechanical theory of heat* [4]. Later on, while developing the theory of statistical mechanics, Austrian physicist Ludwig Boltzmann reinterpreted entropy as a measure of disorder of a system in some of his writings [5]. His formula for entropy (which is now famously engraved on his tombstone), is

$$S = K_B \ln W \tag{1.1}$$

where K_B is a physical constant referred to as "Boltzman's constant" and W represents the number of microstates associated to a given macrostate of a system.

Claude Shannon extended this notion of disorder when he defined his information-theoretic version of entropy [1]. Partially inspired by Alan Turing's wartime research into automation, Shannon developed his theory by considering messages transmitted with finite alphabets and attempting to discover a way to both decipher and compress this information as efficiently as possible. His work is considered the foundation of modern information theory.

In many approaches to the study of dynamical systems, one takes a manifold M for a state space. Time is modelled by a monoid action on the manifold (either $\mathbb{R}^+$ or $\mathbb{N} \cup \{0\}$). More specifically, we take time as a point t in a monoid $\mathcal{T}$ ($\mathcal{T} = \mathbb{R}^+$ or $\mathcal{T} = \mathbb{N} \cup \{0\}$) and use it to define a collection of smooth functions $\{\phi^t\}_{t \in \mathcal{T}}$, $\phi^t : M \to M$, which preserve the structure of the monoid. In other words, $\phi^s \circ \phi^t = \phi^{s+t}$.

The approach that we shall study extensively involves a measure-theoretic interpretation of a dynamical system, where our state space is taken to be a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and time evolution is modelled through a collection of measurable functions $T^t : \Omega \to \Omega$ where the monoid structure of $\mathcal{T}$ is once again preserved: $T^s \circ T^t = T^{s+t}$.

In this thesis, we will focus our efforts specifically on studying discrete-time dynamical systems.

The consistency condition derived from the monoid structure of $\mathcal{T} = \mathbb{N} \cup \{0\}$ implies the existence of an operator $T : \Omega \to \Omega$ (which is the 'one-step time-evolution operator') which generates the rest of the time evolution through simple iteration.

The outline of this thesis is as follows: We shall first present some preliminary results from measure theory and analysis. We will then go on to present our standard working definition of a dynamical system, and develop some of the basic theory and properties. We will then introduce the concept of entropy, provide some basic properties, and discuss the fundamental theorems relating entropy to dynamical systems. We will conclude by looking at specific applications of this theory to the study of coding and compression.

2 Preliminaries

Before we begin with our study of dynamical systems, we will introduce some standard results of measure theory and probability which will prove useful in our analysis.

The first result is the Radon-Nikodym theorem which provides a way of representing certain measures as integrals with respect to some reference measure. This notion shall be clarified in the proceeding subsection.

The next one is a widely used result from probability theory known as the Borel-Cantelli Lemma, which provides a neat method for computing the probabilities of sequences of events.

We shall then move on to the definition of the concept of conditional expectation, a widely used notion in probability theory which will provide a useful framework for understanding many of the results in our study of dynamical systems.

We shall conclude this section with a brief overview of measure-preserving transformations, which we shall use to model the dynamics of our system.

2.1 The Radon-Nikodym theorem

Proved for $\mathbb{R}^n$ by Johann Radon in 1913, and extended to measure spaces by Otto Nikodym in 1930 [6], the theorem we present in this section is central to the study of measure theory. The version we present has been extended to general measure spaces.

Before we state the theorem, we recall a couple of basic definitions relating to measures.

Definition 2.1. Let $(X, \mathcal{F}, \mu)$ be a measure space, and let λ be another measure (either positive or complex) on $(X, \mathcal{F})$. We say that λ is *absolutely continuous* with respect to μ, and write $\lambda \ll \mu$ if, for all $E \in \mathcal{F}$,

$$\mu(E) = 0 \Rightarrow \lambda(E) = 0 \tag{2.1}$$

Definition 2.2. Let $(X, \mathcal{F})$ be a measurable space, and let μ be a (positive or complex) measure on $(X, \mathcal{F})$. We say that μ is *concentrated* on a set $A \in \mathcal{F}$ if $\mu(E) = \mu(A \cap E)$ for all $E \in \mathcal{F}$, or in other words, if $\mu(E) = 0$ for all $E \in \mathcal{F}$ such that $A \cap E = \emptyset$.

If μ_1 and μ_2 are both measures on $(X, \mathcal{F})$ such that μ_1 is concentrated on $A \in \mathcal{F}$, μ_2 is concentrated on $B \in \mathcal{F}$, and $A \cap B = \emptyset$, then we say that μ_1 and μ_2 are *mutually singular*, and we write

$$\mu_1 \perp \mu_2 \tag{2.2}$$

We are now ready to state the celebrated Radon-Nikodym theorem.

Theorem 2.3. *Let $(X, \mathcal{F}, \mu)$ be a σ-finite measure space, and let λ be a complex measure on $(X, \mathcal{F})$.*

(a) There exists a unique pair of complex measures, λ_a and λ_s such that $\lambda = \lambda_a + \lambda_s$, and $\lambda_a \ll \mu$, while $\lambda_s \perp \mu$.

(b) There exists a unique function $f \in L^1(\mu)$ such that

$$\lambda_a(E) = \int_E f \, d\mu \tag{2.3}$$

for all $E \in \mathcal{F}$

The uniqueness of the function f in part (b) of Theorem 2.3 is an easy consequence of a basic fact about L^1 functions on measure spaces. This fact is fundamental in measure theory and will prove useful to us later on in this thesis. As such, we present it here as a lemma.

Lemma 2.4. *Let $(X, \mathcal{F}, \mu)$ be a measure space, and $f, g \in L^1(\mu)$. $f = g$ if and only if, for all $E \in \mathcal{F}$,*

$$\int_E f \, d\mu = \int_E g \, d\mu \tag{2.4}$$

It is important to note here that f and g are actually equivalence classes of functions which are equal 'almost everywhere'. When we say $f = g$, we really mean that f and g belong to the same equivalence class.

Proof of Lemma 2.4.
The "only if" part of the proof is obvious, so we need only prove the "if" direction. Also, we shall restrict our proof to real-valued functions, as the extension to complex-valued functions is also obvious.

Let f and g be in $L^1(\mu)$ such that for all $E \in \mathcal{F}$,

$$\int_E f \mathrm{d}\mu = \int_E g \mathrm{d}\mu \tag{2.5}$$

Consider sets of the form

$$B_n = \left\{ x \in X : f(x) > g(x) + \frac{1}{n} \right\} \tag{2.6}$$

for $n \in \mathbb{N}$. each B_n is measurable, since

$$B_n = (f - g)^{-1} \left[\left(\frac{1}{n}, \infty \right) \right] \tag{2.7}$$

Thus,

$$\begin{aligned} 0 &= \int_{B_n} (f - g) \mathrm{d}\mu \\ &\geq \int_{B_n} \frac{1}{n} \mathrm{d}\mu \end{aligned} \tag{2.8}$$

where the equality follows from our assumption, and the inequality is a consequence of the definition of B_n.

It follows that $\mu(B_n) = 0$. Defining $B = \cup_{n=1}^{\infty} B_n$, we see that $\mu(B) = 0$ and that B can also be written as $B = \{x \in X : f(x) > g(x)\}$.

An analogous argument shows that the set $A = \{x \in X : f(x) > g(x)\}$ has measure zero and thus

$$\mu\left(\{x \in X : f(x) \neq g(x)\}\right) = \mu(A) + \mu(B) = 0 \tag{2.9}$$

This completes the proof. □

2.2 The Borel-Cantelli lemma

Another important result that we shall use comes from probability theory. Named after mathematicians Émile Borel and Francesco Cantelli for discovering it at the turn of the twentieth century (see [7] and [8]), it is a result which helps us to compute probabilities associated with sequences of events.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (i.e. a measure space such that $\mathbb{P}(\Omega) = 1$), we may consider

a sequence of events $(E_n)_{n=1}^{\infty} \subset \mathcal{F}$. We define the following set:

$$\limsup_{n\to\infty} E_n \equiv \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k \tag{2.10}$$

We clearly have $\limsup_{n\to\infty} E_n \in \mathcal{F}$ by the axioms of a σ-algebra. $\limsup_{n\to\infty} E_n$ can be interpreted as the set of all outcomes such that infinitely many E_n are achieved. In other words,

$$\limsup_{n\to\infty} E_n = \{\omega \in \Omega : \omega \in E_n \text{ for infinitely many } n \in \mathbb{N}\} \tag{2.11}$$

We can now state the result concisely.

Lemma 2.5 (Borel-Cantelli)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(E_n)_{n=1}^{\infty}$ be a sequence of events in $\mathcal{F}$ such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty.$$

We must then have

$$\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 0$$

Proof. We begin by defining the sets B_n as follows:

$$B_n \equiv \bigcap_{k=n}^{\infty} E_k \tag{2.12}$$

Immediately from the definition, we see that

1. $B_n \in \mathcal{F}$ for all $n \in \mathbb{N}$
2. $B_{n+1} \subset B_n$ for all $n \in \mathbb{N}$.

The second fact (i.e. that B_n is a decreasing sequence) implies that $\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right)$. This is a basic result from probability theory and can be found in any introductory textbook (see [9], for example).

We can now attempt to estimate the probability of $\limsup E_n$:

$$\begin{aligned}
\mathbb{P}(\limsup_{n\to\infty} E_n) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} E_k\right) \\
&= \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) \\
&= \lim_{n\to\infty} \mathbb{P}(B_n) \\
&\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(E_k) \\
&= 0
\end{aligned} \tag{2.13}$$

where in the last step, we have implicitly used the fact that $\sum_{k=1}^{\infty} \mathbb{P}(E_k) < \infty$. □

2.3 Conditional expectation

Throughout the rest of this thesis, we shall be considering a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we shall define *expectation* as integration with respect to $\mathbb{P}$, namely, (whenever it makes sense),

$$\mathbb{E}[f] = \int_{\Omega} f \mathrm{d}\mathbb{P} \tag{2.14}$$

We suppose $f \in L^1(\mathbb{P})$ and consider a sub σ-algebra $\mathcal{A} \subset \mathcal{F}$. While f is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$, it need not be measurable from the reduced measurable space, $(\Omega, \mathcal{A})$. It would be useful if one could find an $\mathcal{A}$-measurable function $\tilde{f}$ which could act as an 'approximation' to f. We will proceed by showing the existence such a function.

We consider the reduced measure space $(\Omega, \mathcal{A}, \mathbb{P}_{\mathcal{A}})$, where $\mathbb{P}_{\mathcal{A}}$ is simply the restriction of $\mathbb{P}$ to $\mathcal{A}$, and define a complex measure $\lambda : \mathcal{A} \to \mathbb{C}$ given by

$$\lambda(A) \coloneqq \int_A f \mathrm{d}\mathbb{P} \tag{2.15}$$

one quickly verifies that λ is absolutely continuous with respect to $\mathbb{P}_{\mathcal{A}}$, and since $\mathbb{P}_{\mathcal{A}}$ is trivially σ-finite, Theorem 2.3 implies the existence of a unique (up to a set of measure zero) function $\tilde{f}$, an L^1 **$\mathcal{A}$-measurable** function which satisfies

$$\lambda(A) = \int_A \tilde{f} \mathrm{d}\mathbb{P}_A \tag{2.16}$$

for each $A \in \mathcal{A}$. Comparing this to the definition of λ reveals that, for all $A \in \mathcal{A}$

$$\int_A f \mathrm{d}\mathbb{P} = \int_A \tilde{f} \mathrm{d}\mathbb{P} \tag{2.17}$$

Thus $\tilde{f}$ is an approximation of f in the sense that their expectations agree over any $\mathcal{A}$-measurable set. $\tilde{f}$ is called *the conditional expectation of f with respect to $\mathcal{A}$*, or *the expectation of f conditioned on $\mathcal{A}$*, and is commonly denoted by $\mathbb{E}[f|\mathcal{A}]$.

We now list some basic properties of conditional expectation.

Proposition 2.6. *Let $f, g \in L^1(\mathbb{P})$ and $\mathcal{A} \subset \mathcal{F}$, the following properties hold $\mathbb{P}$-almost everywhere:*

(a) for $\alpha, \beta \in \mathbb{C}$, we have $\mathbb{E}[\alpha f + \beta g|\mathcal{A}] = \alpha\mathbb{E}[f|\mathcal{A}] + \beta\mathbb{E}[g|\mathcal{A}]$

(b) if f is $\mathcal{A}$-measurable, then $\mathbb{E}[f|\mathcal{A}] = f$

(c) if f is $\mathcal{A}$-measurable, then $\mathbb{E}[fg|\mathcal{A}] = f\mathbb{E}[g|\mathcal{A}]$.

(d) if $f \in L^p(\mathbb{P})$ for $p \in [1, \infty)$, then $\mathbb{E}[f|\mathcal{A}] \in L^p(\mathbb{P})$ and $\|\mathbb{E}[f|\mathcal{A}]\|_p \leq \|f\|_p$

(e) if $\mathcal{A} = \{\emptyset, \Omega\}$ is the trivial σ-algebra, then $\mathbb{E}[f|\mathcal{A}] = \mathbb{E}[f]$.

The proofs of these facts are elementary and can be found in Chapter 15 of [10]. We shall now make some illuminating remarks about some of these properties.

Firstly, for property (d), one sees that $\mathbb{E}[f|\mathcal{A}]$ is well-defined for $f \in L^p(\mathbb{P})$ since it is a well-known fact that in a probability space, $L^p(\mathbb{P}) \subset L^q(\mathbb{P})$ for all $q \leq p$.

Also, properties (a) and (d) together imply that the mapping $f \mapsto \mathbb{E}[f|\mathcal{A}]$ is a bounded linear operator from $L^p(\mathbb{P})$ to itself for any p.

Property (e) follows from the fact that if $\mathcal{A}$ is the trivial σ-algebra, then all $\mathcal{A}$-measurable functions must be constant functions. In fact, this property can be generalized to slightly more complicated σ-algebras.

2.4 Measure-preserving transformations

Here we shall introduce the concept of a measure-preserving transformation, and prove a basic result that we will be extremely important to us throughout our study of dynamical systems.

Definition 2.7. Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be two probability spaces, and let $T : \Omega_1 \to \Omega_2$ be a measurable transformation, namely $T^{-1}(E) \in \mathcal{F}_1 \quad \forall E \in \mathcal{F}_2$. T is called *measure-preserving* if, for all $E \in F_2$, $\mathbb{P}_1(T^{-1}(E)) = \mathbb{P}_2(E)$

For the purpose of studying a dynamical system, T will be understood as an operator which will take a state $\omega \in \Omega$ to the next state in time. As such, $T : \Omega \to \Omega$ will be a measure-preserving transformation.

We now present an important result relating integration to measure-preserving transformations.

Theorem 2.8. *Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be two probability spaces, and let $T : \Omega_1 \to \Omega_2$ be a measure-preserving transformation. Given $g : \Omega_2 \to \mathbb{C}$, a measurable function, then $f = g \circ T$ defines a measurable function from Ω_1 to $\mathbb{C}$ and for any $E \in \mathcal{F}_2$, we have*

$$\int_E g d\mathbb{P}_2 = \int_{T^{-1}(E)} f d\mathbb{P}_1 \tag{2.18}$$

Proof.
First off, the measurability of $f = g \circ T$ is easily verified, and we will omit it here. Secondly, we prove this result when g is taken to be a characteristic function, $g = \chi_B$ for some $B \in \mathcal{F}_2$. The general result follows from the linearity of the integral and simple applications of monotone convergence theorem (MCT).

If $f = g \circ T$, we see that

$$f(\omega) = \chi_B(T(\omega)) = \chi_{T^{-1}(B)}(\omega) \tag{2.19}$$

thus, we easily verify that

$$\begin{aligned}
\int_{T^{-1}(E)} f \mathrm{d}\mathbb{P}_1 &= \int_{T^{-1}(E)} \chi_{T^{-1}(B)} \mathrm{d}\mathbb{P}_1 \\
&= \mathbb{P}_1(T^{-1}(E) \cap T^{-1}(B)) \\
&= \mathbb{P}_1(T^{-1}(E \cap B)) \\
&= \mathbb{P}_2(E \cap B) \\
&= \int_E g \mathrm{d}\mathbb{P}_2
\end{aligned} \tag{2.20}$$

where the fourth equality above is due to the measure-preserving assumption on T. □

One important question one asks is "Why should we assume that T is measure preserving?". The origin of this assumption is Liouville's theorem from Hamiltonian mechanics, which says that the volume (i.e. Lebesgue measure) of a distribution on phase space is preserved over time. Within the context of probability theory, the assumption has another interpretation.

Since T is a measurable transformation, one can view it as the random variable representing the immediate future. It's distribution on $\mathcal{F}$ is therefore given by

$$\mathbb{P}_T(E) \coloneqq \mathbb{P}(T^{-1}(E)) \tag{2.21}$$

for all $E \in \mathcal{F}$. By the measure preserving property of T, however, we see that $\mathbb{P}_T = \mathbb{P}$, which is equivalent to saying that the distribution of states is time-invariant.

3 Abstract dynamical systems

3.1 Definitions and first examples

An *abstract dynamical system* is a grouping $(\Omega, \mathcal{F}, \mathbb{P}, T)$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $T : \Omega \to \Omega$ is a measure-preserving transformation. As mentioned in the introduction, T can be seen as the generator of a monoid, equipped with the operation of composition, so that $\{T^i\}_{i \in \mathbb{N} \cup \{0\}}$ (with T^0 being the identity) represents the dynamics of the system.

Given a state $\omega \in \Omega$, we define its *orbit* by the set of points $\{T^i(\omega)\}_{i \in \mathbb{N} \cup \{0\}}$, which represent the trajectory of ω as the system evolves in time.

We say that T is *invertible* if T^{-1} exists and is a measurable function. In this case, it is easy to show that T^{-1} is also measure-preserving and thus $(\Omega, \mathcal{F}, \mathbb{P}, T^{-1})$ defines another dynamical system. Equivalently, we can extend the monoid to a full group structure by taking $\mathbb{N} \cup \{0\}$ to $\mathbb{Z}$.

The first and perhaps easiest example of an abstract dynamical system is to consider a fair coin toss as our probability space, namely $\Omega = \{0, 1\}$ and $\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = \frac{1}{2}$, and to take a flip as our generator, namely

$$T(0) = 1, \quad T(1) = 0 \tag{3.1}$$

One easily verifies that T is measure-preserving, and in fact T is also invertible (namely it is its own inverse). However this dynamical system is not so interesting, seeing as how any orbit in this system is just an alternating sequence of 1's and 0's.

A second fundamental example of an abstract dynamical system is the rotation around a circle. We take as our state space the unit circle in the complex plane with the normalized arc length as our probability measure, combined with a rotation as our generator of dynamics. Formally,

$$\begin{aligned} \Omega &= \{z \in \mathbb{C} : \|z\| = 1\} \\ \mathcal{F} &= \mathcal{B}(\Omega) \\ \mathrm{d}\mathbb{P} &= \frac{\mathrm{d}\theta}{2\pi} \\ T(z) &= e^{i2\pi\alpha} z, \quad \alpha \in [0, 1) \end{aligned} \tag{3.2}$$

Since arc length is preserved through rotations, we can see that T is clearly measure preserving,

as is illustrated in Figure 3.1. Furthermore, T is easily seen to be invertible, with its inverse given by a rotation in the opposite direction. We shall revisit this example throughout this thesis.

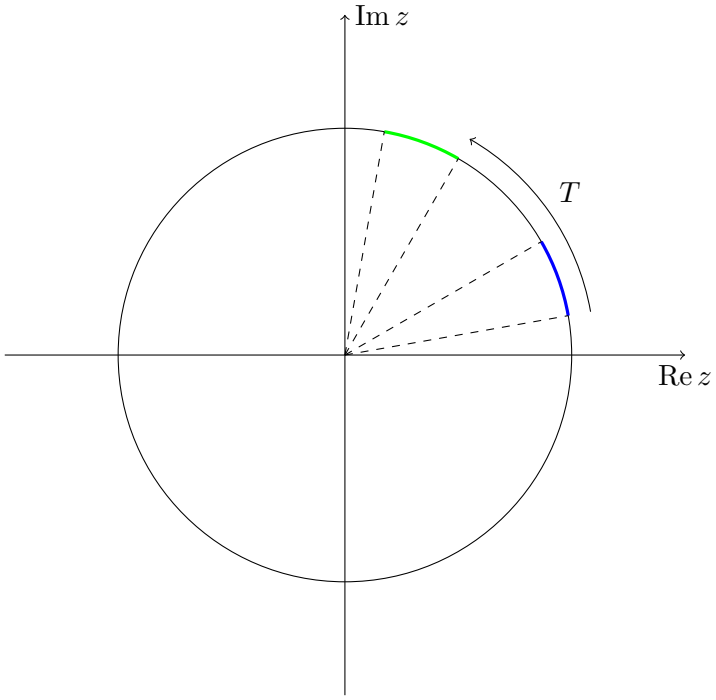

Figure 1: T preserves the arc length as the blue segment is sent to the green one

3.2 The Poincaré recurrence theorem

The first major result of dynamical systems theory that we present is originally due to Henri Poincaré in 1890 and is quite astonishing [2]. Intuitively, it states that a state will almost surely return arbitrarily close to its initial state infinitely many times. We give the formal statement below.

Theorem 3.1. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system, and let $E \in \mathcal{F}$ be such that $\mathbb{P}(E) > 0$. Then $\mathbb{P}$-almost all points in E return to E infinitely many times.*

In mathematical terms, there exists $F \in \mathcal{F}$, $F \subset E$, with $\mathbb{P}(F) = \mathbb{P}(E)$ such that for each $\omega \in F$ there exists a sequence $\{n_k\}_{k=1}^{\infty} \subset \mathbb{N}$ such that $T^{n_k}(\omega) \in E \quad \forall k \in \mathbb{N}$.

Proof.
We let $F_n = \cup_{k=n}^{\infty} T^{-k}(E)$, namely the set of all points in Ω which eventually land in E after at least n time steps. We proceed to define F_∞ as

$$F_\infty = \bigcap_{n=0}^{\infty} F_n = \{\omega \in \Omega : \exists \{n_k\}_{k=1}^{\infty} \subset \mathbb{N},\ T^{n_k}(\omega) \in E\ \forall k \in \mathbb{N}\} \tag{3.3}$$

Thus the desired set is $F = E \cap F_\infty$, and it remains to show that $\mathbb{P}(F) = \mathbb{P}(E)$. This is of course equivalent to showing that $\mathbb{P}(E \setminus F_\infty) = 0$ since E can be written as the following disjoint union

$$E = (E \cap F_\infty) \bigsqcup (E \setminus F_\infty) \tag{3.4}$$

We will thus proceed by showing that $E \setminus F_\infty$ has measure zero. To begin with, we notice that $T^{-j}(F_n) = F_{n+j}$, and thus $\mathbb{P}(F_i) = \mathbb{P}(F_j),\ \forall i, j \geq 0$.

We now wish to calculate $\mathbb{P}(E \setminus F_n)$ for each n. To do this, we first note that $E \setminus F_n \subset F_0 \setminus F_n$ (since $E \subset F_0$). Thus,

$$\mathbb{P}(E \setminus F_n) \leq \mathbb{P}(F_0 \setminus F_n) = \mathbb{P}(F_0) - \mathbb{P}(F_n) = 0 \tag{3.5}$$

Thus, we find that

$$\begin{aligned} \mathbb{P}(E \setminus F_\infty) &= \mathbb{P}(E \setminus \cap_{n=0}^{\infty} F_n) \\ &= \mathbb{P}(\cup_{n=0}^{\infty} E \setminus F_n) \\ &\leq \sum_{n=0}^{\infty} \mathbb{P}(E \setminus F_n) \\ &= 0 \end{aligned} \tag{3.6}$$

which completes the proof. □

3.3 Birkhoff's ergodic theorem

In 1931, George David Birkhoff proved another major result of dynamical systems theory [3]. It establishes the existence of the 'time average' of integrable functions on a dynamical system and provides a probabilistic interpretation.

Theorem 3.2. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system, and let $f \in L^1(\mathbb{P})$. Then*

$$f_T(\omega) \coloneqq \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(\omega) \tag{3.7}$$

exists for $\mathbb{P}$-almost all ω. If we define the following subset of $\mathcal{F}$,

$$\mathcal{A}_T = \{A \in \mathcal{F} : T^{-1}(A) = A\} \tag{3.8}$$

then $\mathcal{A}_T$ is a σ-algebra, and we have that

$$f_T = \mathbb{E}[f|\mathcal{A}_T] \tag{3.9}$$

$\mathbb{P}$-*almost everywhere. Moreover, if* $f_T \in L^p(\mathbb{P})$ *for* $p \in [1, \infty)$*, then*

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(\omega) - f_T\right|^p\right] = 0 \tag{3.10}$$

In this subsection, we will provide a similar proof to the theorem as the one found in Chapter 4 of [11]

Before we prove this theorem we state and prove a small lemma.

Lemma 3.3. *If* $f \in L^1(\mathbb{P})$ *then* $\mathbb{E}[f|\mathcal{A}_T]$ *is* T*-invariant* $\mathbb{P}$*-almost everywhere*

Proof of Lemma 3.3.
We know that both $\mathbb{E}[f|\mathcal{A}_T]$ and $\mathbb{E}[f|\mathcal{A}_T] \circ T$ are $\mathcal{A}_T$-measurable functions. By Lemma 2.4, it suffices then to show that integrals of the functions over sets from $\mathcal{A}_T$ agree with each other. To that end, we observe for an arbitrary set $E \in \mathcal{A}_T$

$$\begin{aligned}\int_E \mathbb{E}[f|\mathcal{A}_T]\mathrm{d}\mathbb{P} &= \int_{T^{-1}(E)} \mathbb{E}[f|\mathcal{A}_T] \circ T\mathrm{d}\mathbb{P} && \text{(by Theorem 2.8)} \\ &= \int_E \mathbb{E}[f|\mathcal{A}_T] \circ T\mathrm{d}\mathbb{P} && \text{(since } E = T^{-1}(E)\text{)} \end{aligned} \tag{3.11}$$

Since E was arbitrarily chosen from $\mathcal{A}_T$, the lemma is proven. □

Proof of Theorem 3.2.
By the linearity of the conditional expectation, it is sufficient to prove the result for real-valued f.

We take $g \in L^1(\mathbb{P})$ another real-valued function (for now, we will not precisely define g, but interpret it as some arbitrary, real-valued L^1 function). For $\omega \in \Omega$, we define $G_n(\omega)$ as

$$G_n(\omega) \coloneqq \max_{1\le k\le n} \sum_{i=0}^{k-1} g \circ T^k(\omega) \tag{3.12}$$

It easily follows from the definition that, for fixed $\omega \in \Omega$, $G_n(\omega)$ is an increasing sequence of real numbers, and thus its limit exists (either in $\mathbb{R}$ or it is ∞). Let A be the following set in $\mathcal{F}$:

$$A \coloneqq \left\{\omega \in \Omega : \lim_{n\to\infty} G_n(\omega) = \infty\right\} \tag{3.13}$$

The first observation we make is that $A \in \mathcal{A}_T$. To show this, we must prove that $\lim_{n\to\infty} G_n(\omega) = \infty$ if and only if $\lim_{n\to\infty} G_n(T\omega) = \infty$.

To do this, we manipulate the expression $G_{n+1}(\omega) - G_n(T\omega)$:

$$\begin{aligned} G_{n+1}(\omega) - G_n(T\omega) &= \max_{1\le k\le n+1} \sum_{i=0}^{k-1} g\circ T^k(\omega) - G_n(T\omega) \\ &= \max\left\{ g(\omega), g(\omega) + \max_{2\le k\le n+1} \sum_{i=0}^{k-1} g\circ T^k(\omega) \right\} - G_n(T\omega) \\ &= g(\omega) + \max\left\{ 0, \max_{2\le k\le n+1} \sum_{i=0}^{k-1} g\circ T^k(\omega) \right\} - G_n(T\omega) \\ &= g(\omega) - \min\left\{0, G_n(T\omega)\right\} \end{aligned} \tag{3.14}$$

To summarize, we obtain

$$G_{n+1}(\omega) - G_n(T\omega) = g(\omega) - \min\left\{0, G_n(T\omega)\right\} \tag{3.15}$$

Equation 3.15 provides us with a lot of insight into G_n. Firstly, one easily establishes that either both $G_n(\omega)$ and $G_n(T\omega)$ converge to a real number, or both diverge to infinity, and thus $T^{-1}(A) = A$

Secondly, since $G_n(T\omega)$ is an increasing sequence, one establishes that $G_{n+1}(\omega) - G_n(T\omega)$ must be a decreasing sequence of real numbers, whose absolute value is bounded by $2|g(\omega)|$.

If $\omega \in A$, one also sees from Equation 3.15 that $\lim_{n\to\infty} (G_{n+1}(\omega) - G_n(T\omega)) = g(\omega)$.

Furthermore, for all ω, $G_{n+1}(\omega) - G_n(\omega) \ge 0$ (since $G_n(\omega)$ is increasing), thus one has the following:

$$\begin{aligned} 0 &\le \int_A (G_{n+1} - G_n)\,\mathrm{d}\mathbb{P} \\ &= \int_A G_{n+1}\mathrm{d}\mathbb{P} - \int_{T^{-1}(A)} G_n\circ T\mathrm{d}P \\ &= \int_A (G_{n+1} - G_n\circ T)\,\mathrm{d}\mathbb{P} \end{aligned} \tag{3.16}$$

where the second line follows from Theorem 2.8, and the third line follows from the fact that $A \in \mathcal{A}_T$. Thus we have that

$$\int_A (G_{n+1} - G_n\circ T)\,\mathrm{d}\mathbb{P} \ge 0 \tag{3.17}$$

By taking n to infinity on both sides of the relation, and applying dominated convergence

theorem, we see that

$$\int_A g(\omega)\mathrm{d}\mathbb{P} \geq 0 \tag{3.18}$$

and, since $A \in \mathcal{A}_T$, the definition of conditional expectation implies that

$$\int_A \mathbb{E}[g|\mathcal{A}_T]\mathrm{d}\mathbb{P} \geq 0 \tag{3.19}$$

At this stage, we pick $\epsilon > 0$ and make the following choice of $g : \Omega \to \mathbb{R}$:

$$g = f - \mathbb{E}[f|\mathcal{A}_T] - \epsilon \tag{3.20}$$

This choice of g is integrable, since $f, \mathbb{E}[f|\mathcal{A}_T]$ and the constant function ϵ are all L^1 functions. Moreover, by linearity of the conditional expectation, we have that

$$\mathbb{E}[g|\mathcal{A}_T] = \mathbb{E}[f|\mathcal{A}_T] - \mathbb{E}[f|\mathcal{A}_T] - \epsilon = -\epsilon < 0 \tag{3.21}$$

Combining this fact with Equation 3.19 shows us that

$$\int_A \mathbb{E}[g|\mathcal{A}_T]\mathrm{d}\mathbb{P} = 0 \tag{3.22}$$

and since $\mathbb{E}[g|\mathcal{A}_T]$ is strictly negative (for our choice of g), this means that $\mathbb{P}(A) = 0$.

So, for our choice of g, $G_n(\omega)$ is a convergent sequence for $\mathbb{P}$-almost all ω. In particular, we have that

$$\lim_{n\to\infty} \frac{G_n(\omega)}{n} = 0 \tag{3.23}$$

for $\mathbb{P}$-almost all ω. Thus, we have

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k \leq \limsup_{n\to\infty} \frac{G_n}{n} = 0 \tag{3.24}$$

$\mathbb{P}$-almost everywhere. Substituting our specific choice of g into the above inequality gives

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \Big[f \circ T^k - \mathbb{E}[f|\mathcal{A}_T] \circ T^k - \epsilon\Big] \tag{3.25}$$

Using Lemma 3.3, we may rearrange the above to obtain

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \leq \mathbb{E}[f|\mathcal{A}_T] - \epsilon \tag{3.26}$$

And since $\epsilon > 0$ was chosen arbitrarily, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \leq \mathbb{E}[f|\mathcal{A}_T] \tag{3.27}$$

$\mathbb{P}$-almost everywhere.

Now, if we had instead chosen $g = (-f) - \mathbb{E}[(-f)|\mathcal{A}_T] - \epsilon$ for some $\epsilon > 0$, an equivalent derivation to the one above would lead to the following analogous inequality,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (-f) \circ T^k \leq \mathbb{E}[-f|\mathcal{A}_T] \tag{3.28}$$

However, this is easily rearranged (through basic properties of lim sup and lim inf, as well as the linearity of conditional expectation) to obtain

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \geq \mathbb{E}[f|\mathcal{A}_T] \tag{3.29}$$

Combining (3.27) and (3.29) establishes the almost-sure convergence of $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$.

To prove the last part of the theorem, we consider the operator $U_T : L^p(\mathbb{P}) \to L^p(\mathbb{P})$ defined by

$$U_T(f) \coloneqq f \circ T \tag{3.30}$$

This is known as the Koopman operator and shall be explored in greater detail in Section 5. Theorem 2.8 implies that

$$\|U_T(f)\|_p = \|f\|_p \tag{3.31}$$

where $\|\cdot\|_p$ is the L^p norm, namely

$$\|f\|_p \coloneqq \left(\int_\Omega |f|^p \, \mathrm{d}\mathbb{P} \right)^{\frac{1}{p}} \tag{3.32}$$

For $n \in \mathbb{N}$, we also define the n-th averaging operator $A_n : L^p(\mathbb{P}) \to L^p(\mathbb{P})$,

$$A_n(f) \coloneqq \frac{1}{n}\sum_{k=0}^{n-1} U_T^k(f) = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \tag{3.33}$$

From what was proved earlier, we have, for $\mathbb{P}$-almost all $\omega \in \Omega$,

$$\lim_{n\to\infty} A_n(f)(\omega) = \mathbb{E}[f|\mathcal{A}_T](\omega) \tag{3.34}$$

We also easily verify that $A_n(f)$ is bounded in norm:

$$\begin{aligned} \|A_n(f)\|_p &= \left\| \frac{1}{n}\sum_{k=0}^{n-1} U_T^k(f) \right\|_p \\ &\leq \frac{1}{n}\sum_{k=0}^{n-1} \left\| U_T^k(f) \right\|_p \\ &= \|f\|_p \end{aligned} \tag{3.35}$$

Thus, for all n, A_n is a contraction.

To show L^p convergence of $A_n(f)$, we first assume that $f \in L^p(\mathbb{P})$ is bounded, i.e. $|f(\omega)| \leq M$ for some $M \in [0, \infty)$. In that case, the almost-sure convergence established in (3.34) allows us to use the Lebesgue dominated convergence theorem to conclude that

$$\lim_{n\to\infty} \|A_n(f) - \mathbb{E}[f|\mathcal{A}_T]\|_p = 0 \tag{3.36}$$

In the general case, when $f \in L^p(\mathbb{P})$, we fix an arbitrary $\varepsilon > 0$. it is a well known fact that we can choose a bounded approximation to f, say $g \in L^p(\mathbb{P})$ with $|g(\omega)| \leq M \in [0, \infty)$ for all $\omega \in \Omega$, such that

$$\|f - g\|_p < \frac{\varepsilon}{4} \tag{3.37}$$

Equation (3.36) establishes the L^p convergence of $A_n(g)$ to $\mathbb{E}[g|\mathcal{A}_T]$. Thus, if we choose n and m in $\mathbb{N}$, we can verify whether or not the desired sequence has the Cauchy property:

$$\begin{aligned} \|A_n(f) - A_m(f)\|_p &= \|A_n(f) - A_n(g) + A_n(g) - A_m(g) + A_m(g) - A_m(f)\|_p \\ &\leq \|A_n(f-g)\|_p + \|A_m(f-g)\|_p + \|A_n(g) - A_m(g)\|_p \\ &\leq 2\|f-g\|_p + \|A_n(g) - A_m(g)\|_p \\ &< \frac{\varepsilon}{2} + \|A_n(g) - A_m(g)\|_p \end{aligned} \tag{3.38}$$

As $A_n(g)$ is an L^p-converging sequence, we may choose n and m large enough so that

$$\|A_n(g) - A_m(g)\|_p < \frac{\varepsilon}{2} \tag{3.39}$$

This allows us to conclude that $A_n(f)$ is a Cauchy sequence in the L^p sense, and by the completeness of L^p, we must have that $A_n(f)$ converges in the L^p sense. This completes the proof of Birkhoff's ergodic theorem. □

3.4 Ergodicity

A large focus in dynamical systems theory is on the notion of ergodicity. The concept, which was first explored in detail by Boltzmann in relation to statistical mechanics, is central in many well-studied examples, including those mentioned above.

Definition 3.4. An abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is said to be *ergodic* if

$$\mathbb{P}(A) \in \{0, 1\} \tag{3.40}$$

for all $A \in \mathcal{A}_T$, the σ-algebra of all T-invariant sets in $\mathcal{F}$.

Intuitively, ergodicity in an abstract dynamical system is a statement about its irreducibility, in the sense that the system cannot be decomposed into 'loops' of significant size.

The following theorem provides further interpretations of ergodicity

Theorem 3.5. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system. The following statements are equivalent:*

(a) $(\Omega, \mathcal{F}, \mathbb{P}, T)$ *is ergodic*

(b) $\mathbb{P}(E) \in \{0, 1\}$ *for all* $E \in \mathcal{F}$ *such that* $\mathbb{P}(T^{-1}(E)\Delta E) = 0$

(c) $\mathbb{P}(E) \in \{0, 1\}$ *for all* $E \in \mathcal{F}$ *such that* $E \subset T^{-1}(E)$

(d) $\mathbb{P}(E) \in \{0, 1\}$ *for all* $E \in \mathcal{F}$ *such that* $T^{-1}(E) \subset E$

(e) if $E \in \mathcal{F}$ *is such that* $\mathbb{P}(E) > 0$, *then* $\mathbb{P}(\cup_{i=0}^{\infty} T^{-i}(E)) = 1$

(f) if $E, F \in \mathcal{F}$ *such that* $\mathbb{P}(E) > 0$ *and* $\mathbb{P}(F) > 0$, *then* $\exists n \in \mathbb{Z}^+$ *such that* $\mathbb{P}(T^{-n}(E) \cap F) > 0$

Prior to proving this theorem, we state (without proof) a couple of lemmas relating to the symmetric difference, Δ, which will be useful.

Lemma 3.6. *For any sets A, B, and C, we have*

$$A\Delta B \subset (A\Delta C) \cup (C\Delta B) \tag{3.41}$$

Lemma 3.7. *Let $(X, \mathcal{F}, \mu)$ be a measure space, and let $A, B \in \mathcal{F}$ be such that $\mu(A \Delta B) = 0$. Then*

$$\mu(A) = \mu(B) = \mu(A \cap B) \tag{3.42}$$

Proof of Theorem 3.5.
((a) $\Rightarrow$ (b)): If $E \in \mathcal{F}$ such that $\mathbb{P}(T^{-1}(E) \Delta E) = 0$, then by the subadditivity of $\mathbb{P}$, and by iterating Lemma 3.6, we see that

$$\begin{aligned}
\mathbb{P}(T^{-n}(E) \Delta E) &\leq \mathbb{P}\left(\bigcup_{i=1}^{n} T^{-i}(E) \Delta T^{-(i-1)}(E) \right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}(T^{-i}(E) \Delta T^{-(i-1)}(E)) \\
&= \sum_{i=1}^{n} \mathbb{P}\left(T^{-(i-1)} \left(T^{-1}(E) \Delta E \right) \right) = 0
\end{aligned} \tag{3.43}$$

where in the last line, we used the measure-preserving property of our ADS. Thus for any $n \in \mathbb{Z}^+$, we have

$$\mathbb{P}(T^{-n}(E) \Delta E) = 0 \tag{3.44}$$

Intuitively, this tells us that the set of points which are mapped to E after any fixed number of time-steps is 'almost' (in the measure-theoretic sense) the same set as E itself. Also, Equation 3.44 immediately implies that

$$\mathbb{P}(T^{-n}(E) \setminus E) = 0, \quad \mathbb{P}(E \setminus T^{-n}(E)) = 0 \quad \forall n \in \mathbb{Z}^+ \tag{3.45}$$

We now consider the set of all points in Ω which are mapped to E infinitely often, namely

$$E_\infty = \limsup_{n \to \infty} T^{-n}(E) = \bigcap_{n=0}^{\infty} \left(\bigcup_{k=n}^{\infty} T^{-k}(E) \right) \tag{3.46}$$

It is not hard to show that $T^{-1}(E_\infty) = E_\infty$, which, by the assumption of ergodicity, implies that $\mathbb{P}(E_\infty) \in \{0, 1\}$. Thus, if we can show that $\mathbb{P}(E \Delta E_\infty) = 0$, we can apply Lemma 3.7 to prove $P(E) \in \{0, 1\}$.

We first examine $\mathbb{P}(E_\infty \setminus E)$ to find that

$$\begin{aligned}
\mathbb{P}(E_\infty \setminus E) &= \mathbb{P}\left(\bigcap_{n=0}^{\infty}\left(\bigcup_{k=n}^{\infty} T^{-k}(E)\right) \setminus E\right) \\
&= \mathbb{P}\left(\bigcap_{n=0}^{\infty}\left(\bigcup_{k=n}^{\infty}\left(T^{-k}(E) \setminus E\right)\right)\right) \\
&\leq \mathbb{P}\left(\bigcup_{k=n}^{\infty}\left(T^{-k}(E) \setminus E\right)\right) \\
&\leq \sum_{k=n}^{\infty} \mathbb{P}(T^{-k}(E) \setminus E) = 0
\end{aligned} \tag{3.47}$$

where we have used the subadditivity of $\mathbb{P}$ as well as basic properties of the symmetric difference.

In a similar fashion, we look at $\mathbb{P}(E \setminus E_\infty)$ to find

$$\begin{aligned}
\mathbb{P}(E \setminus E_\infty) &= \mathbb{P}\left(E \setminus \bigcap_{n=0}^{\infty}\left(\bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right) \\
&= \mathbb{P}\left(\bigcup_{n=0}^{\infty}\left(\bigcap_{k=n}^{\infty}\left(E \setminus T^{-k}(E)\right)\right)\right) \\
&\leq \sum_{n=0}^{\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty}\left(E \setminus T^{-k}(E)\right)\right) \\
&\leq \sum_{n=0}^{\infty} \mathbb{P}\left(E \setminus T^{-n}(E)\right) = 0
\end{aligned} \tag{3.48}$$

Thus, we have

$$\mathbb{P}(E_\infty \Delta E) = \mathbb{P}(E_\infty \setminus E) + \mathbb{P}(E \setminus E_\infty) = 0 \tag{3.49}$$

and thus, by Lemma 3.7, we have $\mathbb{P}(E) = \mathbb{P}(E_\infty) \in \{0, 1\}$.

((b) $\Rightarrow$ (c)): If $E \in \mathcal{F}$ such that $E \subset T^{-1}(E)$, then trivially, we see that $\mathbb{P}(E \setminus T^{-1}(E)) = 0$.

Also, $\mathbb{P}(T^{-1}(E) \setminus E) = \mathbb{P}(T^{-1}(E)) - \mathbb{P}(E) = 0$. This, and the above fact, imply that $\mathbb{P}(E \Delta T^{-1}(E)) = 0$ and by (b), we have that $P(E) \in \{0, 1\}$.

((c) $\Rightarrow$ (d)): If $E \in \mathcal{F}$ is such that $T^{-1}(E) \subset E$, then E^c satisfies the hypothesis of (c), so that $\mathbb{P}(E^c) \in \{0, 1\}$ which implies that $\mathbb{P}(E) \in \{0, 1\}$.

((d) $\Rightarrow$ (e)): If $E \in \mathcal{F}$ such that $\mathbb{P}(E) > 0$, then one easily verifies that

$T^{-1}(\cup_{n=0}^{\infty} T^{-n}(E)) = \cup_{n=1}^{\infty} T^{-n}(E) \subset \cup_{n=0}^{\infty} T^{-n}(E)$, and by (d), we have that

$$\mathbb{P}(\cup_{n=0}^{\infty} T^{-n}(E)) \in \{0, 1\} \tag{3.50}$$

but since $\mathbb{P}(E) > 0$, and $E \subset \cup_{n=0}^{\infty} T^{-n}(E)$, we see that we must have

$$\mathbb{P}(\cup_{n=0}^{\infty} T^{-n}(E)) = 1 \tag{3.51}$$

((e) $\Rightarrow$ (f)): If $E, F \in \mathcal{F}$ are such that $\mathbb{P}(E), \mathbb{P}(F) > 0$, then by (e), we have that $\mathbb{P}(\cup_{n=0}^{\infty} T^{-n}(E)) = 1$, thus

$$\begin{aligned} 0 < \mathbb{P}(F) &= \mathbb{P}\left(\left(\bigcup_{n=0}^{\infty} T^{-n}(E)\right) \cap F\right) \\ &\leq \sum_{n=0}^{\infty} \mathbb{P}(T^{-n}(E) \cap F) \end{aligned} \tag{3.52}$$

Since the above sum is strictly positive, at least one term in the sequence must be strictly positive, thereby demonstrating (f).

((f) $\Rightarrow$ (a)): Suppose that (a) is not true. Then there exists a set $E \in \mathcal{F}$ such that $T^{-1}(E) = E$ and $0 < \mathbb{P}(E) < 1$. Then we must also have that $0 < \mathbb{P}(E^c) < 1$.

By applying (f) to E, and E^c, we must have, for some $n \in \mathbb{Z}^+$, that

$$\mathbb{P}(T^{-n}(E) \cap E^c) > 0 \tag{3.53}$$

However, $T^{-n}(E) = E$ by assumption, and so

$$\mathbb{P}(T^{-n}(E) \cap E^c) = \mathbb{P}(E \cap E^c) = \mathbb{P}(\emptyset) = 0 \tag{3.54}$$

which is a contradiction. □

Theorem 3.5 provides with further characterizations of ergodicity which give us a better understanding of its nature. For example, Theorem 3.5(e) tells us that almost all points in Ω map to any set E of arbitrary positive measure in an ergodic abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$.

Another way to characterize an ergodic system is through its collection of T-invariant functions. The following theorem clarifies the connection.

Theorem 3.8. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system. The following are equivalent:*

(a) $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is an ergodic system

(b) for all measurable $f : \Omega \to \mathbb{C}$, $f = f \circ T$ $\mathbb{P}$-almost everywhere implies f is constant $\mathbb{P}$-almost everywhere.

(c) for all $f \in L^2(\mathbb{P})$, $f = f \circ T$ $\mathbb{P}$-almost everywhere implies f is constant $\mathbb{P}$-almost everywhere.

Proof.
((a) $\Rightarrow$ (b)): Firstly, we note that since any measurable complex function can be split up into its real and imaginary parts, each of which are real, measurable functions, it suffices to prove the implication for real-valued functions only.

We consider some measurable function $f : \Omega \to \mathbb{R}$ such that $f(\omega) = f(T\omega)$ for $\mathbb{P}$-almost all $\omega \in \Omega$.

For $m \in \mathbb{Z}$, and $n \in \mathbb{N}$, we define the following set

$$A_{m,n} = f^{-1}\left(\left[\frac{m}{n}, \frac{m+1}{n}\right)\right) = \left\{\omega \in \Omega : \frac{m}{n} \leq f(\omega) < \frac{m+1}{n}\right\} \tag{3.55}$$

The sets $A_{m,n}$ satisfy the following properties:

- for fixed $n \in \mathbb{N}$, we have $\bigsqcup_{m\in\mathbb{Z}} A_{m,n} = \Omega$, i.e. Ω is the disjoint union of the sets $A_{m,n}$ for all $m \in \mathbb{Z}$.
- by the T-invariance of f, we have that $\mathbb{P}(T^{-1}(A_{m,n})\Delta A_{m,n}) = 0$

These facts allow us to conclude that

$$\sum_{m\in\mathbb{Z}} \mathbb{P}(A_{m,n}) = 1 \tag{3.56}$$

for fixed $n \in \mathbb{Z}$, and that

$$\mathbb{P}(A_{m,n}) \in \{0, 1\} \quad \forall m \in \mathbb{Z}, n \in \mathbb{N} \tag{3.57}$$

From equations 3.56 and 3.57, we can see that for each $n \in \mathbb{N}$, we must have $m_n \in \mathbb{Z}$ such that

$$\mathbb{P}(A_{m_n,n}) = 1 \tag{3.58}$$

and for all $m \in \mathbb{Z}$ such that $m \neq m_n$, we have

$$\mathbb{P}(A_{m,n}) = 0 \tag{3.59}$$

Also, one easily verifies that $A_{m_n,n}$ is a decreasing sequence of sets, namely

$$A_{m_{n+1},n+1} \subset A_{m_n,n} \quad \forall n \in \mathbb{N} \tag{3.60}$$

Thus

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_{m_n,n}\right) = \lim_{n\to\infty} \mathbb{P}(A_{m_n,n}) = 1 \tag{3.61}$$

Thus, in the probabilistic sense, almost all points of Ω lie in $A \coloneqq \bigcap_{n\in\mathbb{N}} A_{m_n,n}$.

We now claim that f is constant on A. Suppose for the sake of contradiction that this is not true. In particular, suppose there exist ω_1 and ω_2 in A such that $f(\omega_1) \neq f(\omega_2)$.

Choose $n \in \mathbb{N}$ large enough so that

$$\frac{1}{n} < |f(\omega_1) - f(\omega_2)| \tag{3.62}$$

Since $\omega_1, \omega_2 \in A$, then $\omega_1, \omega_2 \in A_{m_n,n}$. Thus, by our definition of the set,

$$\frac{m_n}{n} \leq f(\omega_1) < \frac{m_n+1}{n}, \quad \frac{m_n}{n} \leq f(\omega_2) < \frac{m_n+1}{n} \tag{3.63}$$

Therefore,

$$|f(\omega_1) - f(\omega_2)| < \left|\frac{m_n}{n} - \frac{m_n+1}{n}\right| = \frac{1}{n} < |f(\omega_1) - f(\omega_2)| \tag{3.64}$$

which is a contradiction. We may then conclude that f is constant on A which shows that (a) implies (b).

((b) $\Rightarrow$ (c)): This is obvious.

((c) $\Rightarrow$ (a)): We take a T-invariant set $E \in \mathcal{F}$ and would like to show that we necessarily have $\mathbb{P}(E) \in \{0, 1\}$.

To do this, we consider the indicator function on E,

$$\chi_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E \end{cases} \tag{3.65}$$

Obviously, $\chi_E \in L^2(\mathbb{P})$ and furthermore, we see that

$$\begin{aligned}\chi_E(T\omega) &= \begin{cases} 1 & T\omega \in E \\ 0 & T\omega \notin E \end{cases} \\ &= \begin{cases} 1 & \omega \in T^{-1}(E) \\ 0 & \omega \notin T^{-1}(E) \end{cases} \\ &= \chi_{T^{-1}(E)}(\omega) = \chi_E(\omega) \end{aligned} \tag{3.66}$$

Thus, χ_E satisfies the assumptions of (c), and therefore χ_E is constant $\mathbb{P}$-almost everywhere. Thus, for $\mathbb{P}$-almost all ω, $\chi_E(\omega) = 1$, or alternatively, $\chi_E(\omega) = 0$ for $\mathbb{P}$-almost all ω. In either case,

$$\mathbb{P}(E) = \int_\Omega \chi_E \mathrm{d}\mathbb{P} \in \{0, 1\} \tag{3.67}$$

□

One important fact that is worth noting is that for any ADS $(\Omega, \mathcal{F}, \mathbb{P}, T)$ and any function $f \in L^1(\mathbb{P})$, $\mathbb{E}[f|\mathcal{A}_T]$ is a T-invariant function (as we saw in the proof of Birkhoff's ergodic theorem, Theorem 3.2). We therefore have an important corollary.

Corollary 3.9 (corollary to Theorems 3.2 and 3.8)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an ergodic abstract dynamical system, and $f \in L^1(\mathbb{P})$, then*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) \tag{3.68}$$

converges to a constant for $\mathbb{P}$-almost all ω, namely

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = \mathbb{E}[f] \coloneqq \int_\Omega f \, d\mathbb{P} \tag{3.69}$$

$\mathbb{P}$-almost surely.

This corollary tells us that for an ergodic system, the 'time-average' of a function is, regardless of initial state, equal to its 'space-average', namely its expectation over all possible states. This is an important characterization of ergodicity.

We conclude this section with a final interpretation of ergodicity, known as the mean sojourn time theorem, which is one of the initial formulations of the theory by Boltzmann in his development of statistical mechanics.

Definition 3.10. Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system, and consider $C \in \mathcal{F}$ and $\omega \in \Omega$. An *ocurrence time* for ω in C is a number $k \in \mathbb{Z}^+$ such that $T^k(\omega) \in C$.

We note that an occurence time may not exist for a given state ω and event $C \in \mathcal{F}$. We shall soon see however that ergodicity will guarantee their existence (at least 'almost-surely') for certain sets.

For fixed $\omega \in \Omega$,$C \in \mathcal{F}$ and $n \in \mathbb{N}$, we define the following quantity

$$r_n(\omega, C) \coloneqq \left|\left\{k \in \mathbb{Z}^+ : k \text{ is an occurence time for } \omega \text{ in } C, k < n\right\}\right| \tag{3.70}$$

which is the number of times that ω is evolved to a state in C for times strictly less than n. The following theorem relates the time averages of this quantity to ergodicity.

Theorem 3.11 (Mean sojourn time)**.** *An abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is ergodic if and only if for each $C \in \mathcal{F}$,*

$$\lim_{n\to\infty} \frac{r_n(\omega, C)}{n} = \mathbb{P}(C) \tag{3.71}$$

for $\mathbb{P}$-almost all $\omega \in \Omega$. This limit is referred to as the mean sojourn time of ω in C, as it can be interpreted as the average time that ω will spend in the set C, as time evolves indefinitely.

Proof.
($\Rightarrow$):
This direction is a fairly straightforward application of Birkhoff's ergodic theorem, since if our system is ergodic, and $C \in \mathcal{F}$, then one can easily verify that

$$r_n(\omega, C) = \sum_{k=0}^{n-1} \chi_C(T^k\omega) \tag{3.72}$$

where χ_C is the usual indicator function of C. One thus concludes the following

$$\lim_{n\to\infty} \frac{r_n(\omega, C)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_C(T^k\omega) = \int_\Omega \chi_C \mathrm{d}\mathbb{P} = \mathbb{P}(C) \tag{3.73}$$

where we have used Corollary 3.9.

($\Leftarrow$):
Suppose Equation 3.71 holds for each set $C \in \mathcal{F}$, and almost all $\omega \in \Omega$. If $E \in \mathcal{A}_T$ ($T^{-1}(E) = E$), we would like to show that $\mathbb{P}(E) \in \{0, 1\}$.

To do so, we consider χ_E. One notes that the T-invariance of E leads to the T-invariance of

χ_E and thus,

$$\lim_{n\to\infty} \frac{r_n(\omega, E)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T^k\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega \\ 0 & \text{if } \omega \notin \Omega \end{cases} \tag{3.74}$$

However, by Equation 3.71, we know that

$$\lim_{n\to\infty} \frac{r_n(\omega, E)}{n} = \mathbb{P}(E) \tag{3.75}$$

for $\mathbb{P}$-almost all $\omega \in \Omega$. Since we have shown that the mean sojourn time can only possibly take on the values of 0 or 1, we have finished the proof. □

4 The convex structure of invariant measures

In this section, we will broaden our view of dynamical systems theory. Instead of fixing an abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$, we will instead look at a fixed *measurable space* $(\Omega, \mathcal{F})$ (a sample space Ω with an associated σ-algebra) and then consider the collection $\mathcal{P}_{(\Omega,\mathcal{F})}$ of all probability measures on $(\Omega, \mathcal{F})$. If we further fix a measurable transformation $T : \Omega \to \Omega$, we can consider the collection of all T-invariant measures $\mathcal{P}^T_{(\Omega,\mathcal{F})}$, as well as $\mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})}$, the collection of all ergodic measures with respect to T.

An elementary observation is that the collections $\mathcal{P}_{(\Omega,\mathcal{F})}$ and $\mathcal{P}^T_{(\Omega,F)}$ are *convex*. To see this, we note that if $t \in (0, 1)$ and $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}_{(\Omega,\mathcal{F})}$, then the measure $\mathbb{P}$ defined by

$$\mathbb{P}(A) = t\mathbb{P}_1(A) + (1 - t)\mathbb{P}_2(A) \tag{4.1}$$

for each $A \in \mathcal{F}$ is also a probability measure, i.e. $\mathbb{P} \in \mathcal{P}_{(\Omega,\mathcal{F})}$. If we had further assumed that $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$, then we would find that $\mathbb{P} \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$ as well.

4.1 The structure of $\mathcal{P}^{T(\mathbf{erg})}_{(\Omega,\mathcal{F})}$

As we shall see in the remainder of this thesis, it will be useful to have a characterization of the structure of $\mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})}$ in addition to the convexity property we discussed for $\mathcal{P}_{(\Omega,\mathcal{F})}$ and $\mathcal{P}^T_{(\Omega,\mathcal{F})}$. To that end, we provide a definition for an *extremal measure* $\mathbb{P} \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$.

Definition 4.1. Given a measurable space $(\Omega, \mathcal{F})$, and a measurable transformation $T : \Omega \to \Omega$, a measure $\mathbb{P} \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$ is called *extremal* if it cannot be expressed as the non-trivial convex combination of two other T-invariant measures. In other words, if there exist $t \in (0, 1)$ and $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$ such

that

$$\mathbb{P} = t\mathbb{P}_1 + (1-t)\mathbb{P}_2 \tag{4.2}$$

then, necessarily, $\mathbb{P}_1 = \mathbb{P}_2 = \mathbb{P}$. We denote the collection of all extremal measures by $\mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$.

We now state a theorem which shows the equivalence between ergodic measures and extremal measures.

Theorem 4.2. *For a given measurable space $(\Omega, \mathcal{F})$ and a measurable transformation $T : \Omega \to \Omega$, we have that*

$$\mathcal{P}^{T(erg)}_{(\Omega,\mathcal{F})} = \mathcal{P}^{T(ext)}_{(\Omega,\mathcal{F})} \tag{4.3}$$

Proof.
($\mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})} \subset \mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})}$):
To prove the first part, we consider $\mathbb{P} \in \mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$ and would like to show that $\mathbb{P} \in \mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})}$.

Suppose for the sake of contradiction that there exists $E \in \mathcal{F}$ such that $T^{-1}(E) = E$ and $\mathbb{P}(E) \notin \{0, 1\}$. We may thus define the following conditional probabilities

$$\mathbb{P}_1(A) = \mathbb{P}(A|E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}, \quad \mathbb{P}_2(A) = \mathbb{P}(A|E^c) = \frac{\mathbb{P}(A \cap E^c)}{\mathbb{P}(E^c)} \tag{4.4}$$

for arbitrary $A \in \mathcal{F}$.

$\mathbb{P}_1$ and $\mathbb{P}_2$ are well-defined probability measures, since both $\mathbb{P}(E) > 0$ and $\mathbb{P}(E^c) > 0$. Furthermore, one can show that both $\mathbb{P}_1$ and $\mathbb{P}_2$ are T-invariant. We verify this for $\mathbb{P}_1$. We take an arbitrary set $A \in \mathcal{F}$ and compute

$$\begin{aligned}
\mathbb{P}_1(T^{-1}(A)) &= \frac{\mathbb{P}(T^{-1}(A) \cap E)}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(T^{-1}(A \cap E))}{\mathbb{P}(E)} \\
&= \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \mathbb{P}_1(A)
\end{aligned} \tag{4.5}$$

where in the second equality we used the T-invariance of the set E and in the third equality we used the T-invariance of the measure $\mathbb{P}$. $\mathbb{P}_2$ is verified to be T-invariant in an analogous manner. Furthermore, it is evident that $\mathbb{P}_1$ and $\mathbb{P}_2$ are each distinct from $\mathbb{P}$:

$$\begin{aligned}
\mathbb{P}_1(E^c) &= \frac{\mathbb{P}(E^c \cap E)}{\mathbb{P}(E)} = 0 \neq \mathbb{P}(E^c) \\
\mathbb{P}_2(E) &= \frac{\mathbb{P}(E \cap E^c)}{\mathbb{P}(E^c)} = 0 \neq \mathbb{P}(E)
\end{aligned} \tag{4.6}$$

Thus, if we take $t = \mathbb{P}(E) \in (0,1)$, then one easily verifies that

$$\mathbb{P} = t\mathbb{P}_1 + (1-t)\mathbb{P}_2 \tag{4.7}$$

which is a contradiction to the fact that $\mathbb{P} \in \mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$.

$(\mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})} \subset \mathcal{P}^{T(\text{ext})}_{(\Omega,f)})$:
Once again, we proceed with a proof by contradiction.

Suppose $\mathbb{P} \in \mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})}$ but $\mathbb{P} \notin \mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$. Then there exist $t \in (0,1)$ and $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ such that

$$\mathbb{P} = t\mathbb{P}_1 + (1-t)\mathbb{P}_2 \tag{4.8}$$

and both $\mathbb{P}_1$ and $\mathbb{P}_2$ are distinct from $\mathbb{P}$. One easily verifies that for the above relation to hold, we must have $\mathbb{P}_1 \ll \mathbb{P}$ and $\mathbb{P}_2 \ll \mathbb{P}$. Thus, by Theorem 2.3, we must have $h_1, h_2 \in L^1(\mathbb{P})$ such that

$$\begin{aligned} \mathbb{P}_1(E) &= \int_E h_1 \mathrm{d}P \\ \mathbb{P}_2(E) &= \int_E h_2 \mathrm{d}P \end{aligned} \tag{4.9}$$

Now it is left as an exercise to check that the T-invariance of the measures $\mathbb{P}_1$ and $\mathbb{P}_2$ imply the ($\mathbb{P}$-almost everywhere) T-invariance of h_1 and h_2. Thus, by Theorem 3.8, we must have that h_1 and h_2 are constant functions (except possibly on sets of $\mathbb{P}$-measure zero).

Since $\mathbb{P}_1$ and $\mathbb{P}_2$ are unit normalized (as they are probability measures), this necessarily implies that $h_1 = h_2 = 1$, and thus we find that $\mathbb{P}_1 = \mathbb{P}_2 = \mathbb{P}$, i.e. $\mathbb{P} \in \mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$. □

This theorem ties in nicely with our previous understanding of ergodicity; In the last section, we saw that ergodic systems were 'irreducible' in the sense that we cannot decompose them into disjoint 'loops' (T-invariant subsets) with positive measure. In this interpretation, we see that ergodic measures are those that are 'irreducible' in the sense that they cannot be written as convex combinations of other, distinct T-invariant measures.

4.2 Existence of invariant measures, the Bogoliubov-Krylov theorem

In the preceding parts of this section, we have almost implicitly been working under the assumption that, given a measurable space $(\Omega, \mathcal{F})$ and a measurable transformation $T : \Omega \to \Omega$, $\mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ is non-empty. In fact, one can easily show that this is not always the case.

We take, as an elementary example, the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the measurable transformation $T(x) = x + 1$ for $x \in \mathbb{R}$. If $\mathbb{P}$ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which is T-invariant, then we see

that for any interval of the form $[m, m+1)$ for some $m \in \mathbb{Z}$, we must have

$$\mathbb{P}([m, m+1)) = \mathbb{P}\left(T^{-1}\left([m, m+1)\right)\right) = \mathbb{P}([m-1, m)) \tag{4.10}$$

which shows that the $\mathbb{P}([m, m+1))$ is constant for all $m \in \mathbb{Z}$. Now since

$$\mathbb{R} = \bigsqcup_{m \in \mathbb{Z}} [m, m+1) \tag{4.11}$$

we have two possibilities. The first is that $\mathbb{P}([m, m+1)) = 0$, which would imply that $\mathbb{P}(\mathbb{R}) = 0$, i.e. $\mathbb{P}$ is the trivial measure, and not a probability measure. The second possibility is that $\mathbb{P}([m, m+1)) > 0$, which implies that

$$\mathbb{P}(\mathbb{R}) = \sum_{m \in \mathbb{Z}} \mathbb{P}([m, m+1)) = \infty \tag{4.12}$$

which also shows that $\mathbb{P}$ is not a probability measure, and thus, for our choice of T, $\mathcal{P}^T_{(\mathbb{R}, \mathcal{B}(\mathbb{R}))} = \emptyset$.

The following celebrated theorem, due to mathematicians N. Bogoliubov and N. Krylov [12], specifies conditions for which invariant measures are guaranteed to exist.

Theorem 4.3. *Let Ω be a compact metric space, and let $\mathcal{B}(\Omega)$ be the associated Borel σ-algebra. Then for any measurable transformation $T : \Omega \to \Omega$, we have that $\mathcal{P}^T_{(\Omega, \mathcal{F})} \neq \emptyset$.*

As a prerequisite to proving this theorem, we present a lemma which is an important consequence of the celebrated Stone-Weierstrass theorem.

Lemma 4.4. *If Ω is a compact metric space, then $C(\Omega)$, the space of all continuous, real-valued functions, equipped with the supremum norm, $\|f\|_\infty = \sup_{\omega \in \Omega} |f(\omega)|$, is a separable space.*

We shall not prove this lemma here, but it can be found in [13].

Proof of Theorem 4.3.
By the above lemma, we know that there exists a countable dense subset of $C(\Omega)$, which we shall denote by $\{f_n\}_{n=1}^\infty$. We now fix $\omega_0 \in \Omega$ and consider the sequence

$$\frac{1}{n} \sum_{k=0}^{n-1} f_1(T^k \omega_0) \tag{4.13}$$

for each $n \in \mathbb{N}$. By the compactness of Ω and continuity of f_1, it must be a bounded sequence of real numbers and thus there exists a subsequence which converges in $\mathbb{R}$. We may do the same for each function f_j and thus, using a standard diagonal argument, we may obtain a strictly increasing

sequence of natural numbers $(n_k)_{k=1}^{\infty}$ such that

$$\frac{1}{n_k}\sum_{i=0}^{n_k-1} f_j(T^i\omega_0) \tag{4.14}$$

converges as $k \to \infty$, for each $j \in \mathbb{N}$.

We now claim that for the above sequence converges for any $f \in C(\Omega)$. To see this, we fix $\epsilon > 0$ and choose $m \in \mathbb{N}$ such that $\|f - f_m\| < \epsilon/3$. Also since

$$\frac{1}{n_k}\sum_{i=0}^{n_k-1} f_m(T^i\omega_0) \tag{4.15}$$

converges as $k \to \infty$, we may choose k_1, k_2 large enough such that

$$\left|\frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} f_m(T^i\omega_0) - \frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} f_m(T^i\omega_0)\right| < \frac{\epsilon}{3} \tag{4.16}$$

thus by comparing the k_1 and k_2 terms in our sequence for f, we find

$$\begin{aligned}
&\left|\frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} f(T^i\omega_0) - \frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} f(T^i\omega_0)\right| \\
=&\left|\frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} f(T^i\omega_0) - \left(\frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} f_m(T^i\omega_0) - \frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} f_m(T^i\omega_0)\right)\right. \\
&\left. + \left(\frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} f_m(T^i\omega_0) - \frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} f_m(T^i\omega_0)\right) - \frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} f(T^i\omega_0)\right| \\
\leq&\frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} \|f - f_m\| + \left|\frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} f_m(T^i\omega_0) - \frac{1}{n_{k_1}}\sum_{i=0}^{n_{k_1}-1} f_m(T^i\omega_0)\right| + \frac{1}{n_{k_2}}\sum_{i=0}^{n_{k_2}-1} \|f - f_m\| \\
<&\frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\
=&\epsilon
\end{aligned} \tag{4.17}$$

This shows that the sequence is Cauchy, which proves the existence of the limit.

We proceed to define $J : C(\Omega) \to \mathbb{R}$ in the following manner:

$$J(f) = \lim_{k\to\infty} \frac{1}{n_k}\sum_{i=0}^{n_k-1} f(T^i\omega_0) \tag{4.18}$$

One easily verifies that

- J is a positive linear functional.
- $|J(f)| \leq \|f\|$
- $J(1) = 1$

Thus, by the Riesz representation theorem, there exists a measure $\mathbb{P}$ on $(\Omega, \mathcal{B}(\Omega))$ such that

$$J(f) = \int_\Omega f \mathrm{d}\mathbb{P} \tag{4.19}$$

from which we see that $\mathbb{P}(\Omega) = \int_\Omega 1 \mathrm{d}\mathbb{P} = J(1) = 1$ so that $\mathbb{P}$ is a probability measure. Furthermore, for any $f \in \mathbb{C}(\Omega)$, we verify that

$$\begin{aligned} J(f \circ T) &= \lim_{k\to\infty} \frac{1}{n_k} \sum_{i=1}^{n_k} f(T^i \omega_0) \\ &= \lim_{k\to\infty} \frac{1}{n_k} \sum_{i=0}^{n_k - 1} f(T^i \omega_0) = J(f) \end{aligned} \tag{4.20}$$

where we were able to shift the sum due to the fact that n_k grows to infinity and the terms at either end of the sequence are uniformly bounded. This allows us to conclude that

$$\int_\Omega f \circ T \mathrm{d}\mathbb{P} = \int_\Omega f \mathrm{d}\mathbb{P} \tag{4.21}$$

for all $f \in \mathbb{C}(\Omega)$. By the density of $\mathbb{C}(\Omega) \in L^1(\mathbb{P})$, the same must hold for all L^1 functions as well. In particular, for any Borel set $E \in \mathcal{B}(\Omega)$, we must have

$$\begin{array}{ccc} \int_\Omega \chi_E \mathrm{d}\mathbb{P} & = \int_\Omega \chi_E \circ T \mathrm{d}\mathbb{P} = & \int_\Omega \chi_{T^{-1}(E)} \mathrm{d}\mathbb{P} \\ \| & & \| \\ \mathbb{P}(E) & & \mathbb{P}(T^{-1}(E)) \end{array} \tag{4.22}$$

This establishes the T-invariance of $\mathbb{P}$. Thus $\mathcal{P}^T_{(\Omega,\mathcal{B}(\Omega))}$ is non-empty. □

4.3 The weak topology and existence of ergodic measures

We would now like to continue our study of dynamical systems by showing that under the same assumptions as the Bogoliubov-Krylov theorem, there exists an ergodic measure with respect to any measurable transformation T.

Before we can show this, we will need some preliminary results from topology.

Definition 4.5. Let $(\Omega, \mathcal{F})$ be a measurable space. The weak topology τ on $\mathcal{P}_{(\Omega,\mathcal{F})}$ is the minimal

topology such that the maps

$$\mu \mapsto \int_\Omega f \mathrm{d}\mu \tag{4.23}$$

are continuous for all $f \in C(\Omega)$.

With this definition in hand, one sees that if a sequence $\{\mu_n\}_{n=1}^\infty \subset \mathcal{P}_{(\Omega,\mathcal{F})}$ converges (in the weak topology) to $\mu \in \mathcal{P}_{(\Omega,\mathcal{F})}$, this is equivalent to saying that

$$\lim_{n\to\infty} \int_\Omega f \mathrm{d}\mu_n = \int_\Omega f \mathrm{d}\mu \tag{4.24}$$

for each $f \in C(\Omega)$.

With this definition, we are now prepared to present two important results, which shall be useful in proving the existence of an ergodic measure.

Theorem 4.6. *Let Ω be a compact metric space, and let $\mathcal{F} \coloneqq \mathcal{B}(\Omega)$ be its associated Borel σ-algebra. Then $\mathcal{P}_{(\Omega,\mathcal{F})}$ is compact in the weak topology.*

The proof of this theorem can be found in [14] and [15].

Lemma 4.7. *With the same assumptions as Theorem 4.6, we have that $\mathcal{P}^T_{(\Omega,\mathcal{F})}$ is also compact.*

Proof.
Since $\mathcal{P}^T_{(\Omega,\mathcal{F})} \subset \mathcal{P}_{(\Omega,\mathcal{F})}$, it suffices to show that $\mathcal{P}^T_{(\Omega,\mathcal{F})}$ is closed.

To do this, we consider a sequence $(\mu_n)_{n=1}^\infty$ in $\mathcal{P}^T_{(\Omega,\mathcal{F})}$ which converges (in the weak sense) to μ, and we would like to show that $\mu \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$ as well.

By our definition of the weak topology, we must have that

$$\lim_{n\to\infty} \int_\Omega f \mathrm{d}\mu_n = \int_\Omega f \mathrm{d}\mu \tag{4.25}$$

for all $f \in C(\Omega)$. Also, since $\mu_n \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$ for each $n \in \mathbb{N}$, we must have

$$\int_\Omega f \circ T \mathrm{d}\mu_n = \int_\Omega f \mathrm{d}\mu_n \tag{4.26}$$

for all $f \in C(\Omega)$. Thus

$$\int_\Omega f \circ T \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega f \circ T \mathrm{d}\mu_n = \lim_{n\to\infty} \int_\Omega f \mathrm{d}\mu_n = \int_\Omega f \mathrm{d}\mu \tag{4.27}$$

for each $f \in \mathcal{F}$. By the density of $C(\Omega)$ in $L^1(\mu)$, we must have T-invariance of μ. □

We now establish the main result

Theorem 4.8. *Let Ω be a compact metric space and $\mathcal{F}$ its associated Borel σ-algebra, then $\mathcal{P}^{T(erg)}_{(\Omega,\mathcal{F})} \neq \emptyset$.*

Proof.
We consider a countable dense subset of $C(\Omega)$, namely $\{f_n\}_{n=1}^{\infty}$, and define the following sequence of subsets of $\mathcal{P}^{T}_{(\Omega,\mathcal{F})}$:

$$\begin{aligned}\mathcal{P}_0 &= \mathcal{P}^{T}_{(\Omega,\mathcal{F})} \\ \mathcal{P}_n &= \left\{ \mu \in \mathcal{P}_{n-1} : \int_\Omega f_n \mathrm{d}\mu = \sup_{\nu \in \mathcal{P}_{n-1}} \int_\Omega f_n \mathrm{d}\nu \right\}\end{aligned} \tag{4.28}$$

We now list some important facts about this sequence. For each $n \in \mathbb{Z}^+$, we have:

(a) $\mathcal{P}_n \supset \mathcal{P}_{n+1}$

(b) $\mathcal{P}_n$ is non-empty

(c) $\mathcal{P}_n$ is closed, and therefore compact.

Fact (a) is obvious, and we prove the other two facts by induction.

To begin with, we see that in the base case ($n = 0$), $\mathcal{P}_0 = \mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ is both closed by Lemma 4.7 and non-empty by Theorem 4.3.

For the induction step, we assume $\mathcal{P}_n$ is both non-empty and closed. To help verify that this implies that $\mathcal{P}_{n+1}$ is non-empty and closed as well, we introduce the following collection of mappings

$$\begin{aligned}U_{f_n} &: \mathcal{P}_{(\Omega,\mathcal{F})} \to \mathbb{R} \\ U_{f_n}(\mu) &= \int_\Omega f_n \mathrm{d}\mu\end{aligned} \tag{4.29}$$

which are easily seen to be continuous with respect to the weak topology. We may now write $\mathcal{P}_{n+1}$ as

$$\mathcal{P}_{n+1} = U_{f_n}^{-1}\left(\left\{ \sup_{\nu \in \mathcal{P}_{n-1}} \int_\Omega f_n \mathrm{d}\nu \right\} \right) \cap \mathcal{P}_n \tag{4.30}$$

From this we see that $\mathcal{P}_{n+1}$ is the intersection of two closed sets and is thus closed as well. Furthermore, U_{f_n} must attain its maximum on $\mathcal{P}_n$ by our assumption of closedness and non-emptiness, which demonstrates that $\mathcal{P}_{n+1}$ is non-empty.

We thus have a decreasing sequence of compact, non-empty sets, which implies that

$$\mathcal{P} = \bigcap_{n=0}^{\infty} \mathcal{P}_n \tag{4.31}$$

is non-empty as well. If we can show that $\mathcal{P} \subset \mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})}$, then we will have completed the proof. To do so, we will show that any element of $\mathcal{P}$ is an extremal measure (recall that in Theorem 4.2, we showed that $\mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})} = \mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$).

Take $\mathbb{P} \in \mathcal{P}$, and suppose that there exist $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ and $t \in (0,1)$ such that

$$\mathbb{P} = t\mathbb{P}_1 + (1-t)\mathbb{P}_2 \tag{4.32}$$

Then, we must have

$$\int_\Omega f_1 \mathrm{d}\mathbb{P} = t\int_\Omega f_1 \mathrm{d}\mathbb{P}_1 + (1-t)\int_\Omega f_1 \mathrm{d}\mathbb{P}_2 \tag{4.33}$$

However, since $\mathbb{P} \in \mathcal{P}_1$, we must also have that

$$\int_\Omega f_1 \mathrm{d}\mathbb{P} = \sup_{\nu \in \mathcal{P}^{T}_{(\Omega,\mathcal{F})}} \int_\Omega f_1 \mathrm{d}\nu \tag{4.34}$$

The only possibility then is that

$$\int_\Omega f_1 \mathrm{d}\mathbb{P}_1 = \int_\Omega f_1 \mathrm{d}\mathbb{P}_2 = \int_\Omega f_1 \mathrm{d}\mathbb{P} \tag{4.35}$$

which shows that $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}_1$. We repeat this argument inductively to find that

- $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}$
- $\int_\Omega f_n \mathrm{d}\mathbb{P}_1 = \int_\Omega f_n \mathrm{d}\mathbb{P}_2 = \int_\Omega f_n \mathrm{d}\mathbb{P} \quad \forall n \in \mathbb{N}$

The density of $\{f_n\}_{n=1}^{\infty}$ in $C(\Omega)$, and subsequently the density of $C(\Omega)$ in $L^1(\mathbb{P})$, show that $\mathbb{P}_1 = \mathbb{P}_2 = \mathbb{P}$, showing us that $\mathbb{P} \in \mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})}$. □

4.4 Unique ergodicity

In this section we examine the properties of a system with exactly one invariant (and therefore ergodic measure).

Definition 4.9. Let $(\Omega, \mathcal{F})$ be a measurable space and let $T : \Omega \to \Omega$ be a measurable transformation. T is said to be *uniquely ergodic* if $\mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ consists of exactly one element. In that

case,

$$\mathcal{P}^{T(\text{ext})}_{(\Omega,\mathcal{F})} = \mathcal{P}^{T(\text{erg})}_{(\Omega,\mathcal{F})} = \mathcal{P}^{T}_{(\Omega,\mathcal{F})} = \{\mathbb{P}\} \tag{4.36}$$

where $\mathbb{P}$ is the sole T-invariant and sole ergodic probability measure on $(\Omega, \mathcal{F})$.

A system that is uniquely ergodic is of special interest, since a natural selection of a measure is apparent. In this subsection, we will demonstrate sufficient conditions for unique ergodicity in the case of a compact metric space.

Theorem 4.10. *Let Ω be a compact metric space, $\mathcal{F}$ its associated Borel σ-algebra, and $T : \Omega \to \Omega$ a measurable transformation. If there exists a collection of functions $\Phi \subset C(\Omega)$ which is dense in $C(\Omega)$ such that for all $f \in \Phi$, and $\omega \in \Omega$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = c_f \tag{4.37}$$

where $c_f \in \mathbb{C}$ is a constant (depending on f), then T is uniquely ergodic.

Proof.
We need to prove that $\mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ consists of exactly one element. We first note that for $\mu \in \mathcal{P}^{T}_{(\Omega,\mathcal{F})}$, and for $f \in \Phi$, we have by Birkhoff's ergodic theorem that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \mathbb{E}[f|\mathcal{A}_T](\omega) \tag{4.38}$$

for μ-almost all ω. However, by assumption,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = c_f \tag{4.39}$$

for all $\omega \in \Omega$. Thus, we must have that

$$\int_\Omega \mathbb{E}[f|\mathcal{A}_T]\mathrm{d}\mu = \int_\Omega c_f \mathrm{d}\mu = c_f \tag{4.40}$$

which implies that

$$\int_\Omega f \mathrm{d}\mu = c_f \tag{4.41}$$

for any $\mu \in \mathcal{P}^{T}_{(\Omega,\mathcal{F})}$ and any $f \in \Phi$.

Suppose $\mu_1, \mu_2 \in \mathcal{P}^T_{(\Omega,\mathcal{F})}$, then by the above fact,

$$\int_\Omega f \mathrm{d}\mu_1 = c_f = \int_\Omega f \mathrm{d}\mu_2 \tag{4.42}$$

for any $f \in \Phi$. Since Φ is dense in $C(\Omega)$, then

$$\int_\Omega f \mathrm{d}\mu_1 = \int_\Omega f \mathrm{d}\mu_2 \tag{4.43}$$

for all $f \in C(\Omega)$. Then, by the density of $C(\Omega)$ in L^1, we must have $\mu_1 = \mu_2$. Thus $\mathcal{P}^T_{(\Omega,\mathcal{F})}$ consists of a single element. □

4.5 Rotation on unit circle, Kronecker-Weyl theorem

We return now to the example mentioned in Section 3.1, the rotation on the unit circle. We recall that

$$\begin{aligned} \Omega &= \{z \in \mathbb{C} : |z| = 1\} \\ \mathcal{F} &= \mathcal{B}(\Omega) \\ \mathrm{d}\mathbb{P} &= \frac{\mathrm{d}\theta}{2\pi} \\ T_\alpha(z) &= e^{i(2\pi\alpha)}z, \quad \alpha \in [0,1) \end{aligned} \tag{4.44}$$

defines an abstract dynamical system. Also, since Ω is a bounded, closed subset of $\mathbb{C}$, it must be compact.

The obvious question to ask at this point is whether or not $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is ergodic. This is resolved by the following theorem, originally proven by Hermann Weyl in [16].

Theorem 4.11 (Kronecker-Weyl)**.** *T_α is uniquely ergodic for $(\Omega, \mathcal{F})$ if and only if α is irrational.*

Proof.
We only prove one direction of the implication. The other is left as an exercise that will be more precisely formulated at the end of the proof.

We assume α is irrational and would like to prove unique ergodicity of T_α. To do so, we will attempt to apply Theorem 4.10 to our system.

Let $\Phi = \{f_m \in C(\Omega) : m \in \mathbb{Z}\}$, where

$$f_m(z) = z^m \tag{4.45}$$

By Stone-Weierstrass theorem, Φ is dense in $C(\Omega)$. Fixing some $m \in \mathbb{Z}$, we examine the magnitude

of the n-th term in the ergodic average of f_m:

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=0}^{n-1} f_m(T^k z)\right| &= \left|\frac{1}{n}\sum_{k=0}^{n-1}\left(e^{i(2\pi\alpha)k}z\right)^m\right| \\ &= \left|\frac{z^m}{n}\sum_{k=0}^{n-1}\left(e^{i(2\pi m\alpha}\right)^k\right| \\ &= \left|\frac{z^m}{n}\frac{1-\left(e^{i(2\pi\alpha m)}\right)^n}{1-e^{i(2\pi\alpha m)}}\right| \\ &\leq \frac{|z|^m}{n}\frac{2}{\left|1-e^{i(2\pi\alpha m)}\right|} \end{aligned} \tag{4.46}$$

Since $\alpha \notin \mathbb{Q}$, we must have that $\left|1-e^{i(2\pi\alpha m)}\right| \neq 0$ for any $m \in \mathbb{Z}$. Thus, the above calculation shows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f_m(T^k z) = 0 \tag{4.47}$$

for each $m \in \mathbb{Z}$ and $z \in \Omega$, and thus, by Theorem 4.10, T_α is uniquely ergodic. □

Since $\mathrm{d}\mathbb{P} = \frac{\mathrm{d}\theta}{2\pi}$ is an invariant measure, it is the thus the only invariant measure when α is irrational and thus $(\Omega, \mathcal{F}, \mathbb{P}, T_\alpha)$ is an ergodic ADS.

5 Koopmanism and spectral theory

In this section, we explore a deep connection shared between dynamical systems theory and spectral theory on Hilbert spaces. Specifically, we will see how the dynamics of a given system can be mapped onto the Hilbert space of L^2 functions on our system.

5.1 The Koopman operator

Given an abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$, we consider the Hilbert space of square-integrable functions on $(\Omega, \mathcal{F}, \mathbb{P})$, namely $L^2(\mathbb{P})$ equipped with the inner product

$$\langle f, g\rangle = \int_\Omega \overline{f} g \mathrm{d}\mathbb{P} \tag{5.1}$$

On this space, we define the operator $U_T : L^2(\mathbb{P}) \to L^2(\mathbb{P})$ in the following way

$$U_T(f) = f \circ T \tag{5.2}$$

which we call the Koopman operator. In simple terms, the Koopman operator sends a function forward one step in time, by composing it with the generator of the dynamics.

With the knowledge that we have already obtained on dynamical systems, we can make some important observations about the Koopman operator.

Lemma 5.1. *The following statements hold for an abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ and its associated Koopman operator, U_T.*

(a) for any $f, g \in L^2(\mathbb{P})$, we have $\langle U_T(f), U_T(g)\rangle = \langle f, g\rangle$ (and in particular, if T is invertible, then U_T is unitary).

(b) 1 is an eigenvalue of U_T. Furthermore, 1 is a simple eigenvalue (i.e. its eigenspace is one-dimensional) if and only if T is ergodic.

(c) if λ is an eigenvalue of U_T, then $|\lambda| = 1$.

(d) If $f, g \in L^2(\mathbb{P})$ are eigenfunctions of U_T with different eigenvalues, then f and g are orthogonal, i.e. $\langle f, g\rangle = 0$.

Proof.

(a) To see this, we make a simple application of Theorem 2.8:

$$\begin{aligned}\langle U_T(f), U_T(g)\rangle &= \int_\Omega \overline{(f \circ T)} g \circ T \mathrm{d}\mathbb{P} \\ = \int_\Omega \left(\overline{f} g\right) \circ T \mathrm{d}\mathbb{P} &= \int_\Omega \overline{f} g \mathrm{d}\mathbb{P} = \langle f, g\rangle\end{aligned} \tag{5.3}$$

If T is invertible, then one readily verifies that the inverse Koopman operator is

$$U_T^{-1} = U_{T^{-1}} \tag{5.4}$$

which shows us that U_T is unitary.

(b) To see that 1 is always an eigenvalue of U_T, consider the constant function

$$\mathbb{1}(\omega) = 1 \quad \forall \omega \in \Omega \tag{5.5}$$

We easily see that $U_T(\mathbb{1}) = \mathbb{1}$, and thus 1 is an eigenvalue and its eigenspace contains the space of all constant functions. By Theorem 3.8, we know that ergodicity of our system is equivalent

to the fact that all T-invariant functions are constant, which is equivalent to the fact that the eigenspace of the eigenvalue 1 consists only of constant functions (i.e. it is one-dimensional).

(c) If $\lambda \in \mathbb{C}$ is an eigenvalue for U_T, then for some non-zero $f \in L^2(\mathbb{P})$, we have

$$U_T(f) = \lambda f \tag{5.6}$$

Thus, we have

$$\langle U_T(f), U_T(f) \rangle = \langle \lambda f, \lambda f \rangle = |\lambda|^2 \langle f, f \rangle = |\lambda|^2 \|f\|^2 \tag{5.7}$$

We also have, by property (a), that

$$\langle U_T(f), U_T(f) \rangle = \langle f, f \rangle = \|f\|^2 \tag{5.8}$$

and since f is non-zero, this necessarily implies that $|\lambda| = 1$.

(d) Suppose we have that

$$\begin{aligned} U_T(f) &= \lambda_1 f \\ U_T(g) &= \lambda_2 g \end{aligned} \tag{5.9}$$

where, $f, g \in L^2(\mathbb{P})$ are non-zero, $\lambda_1 \neq \lambda_2$, and by (c), we know that $|\lambda_1| = |\lambda_2| = 1$. An application of (a) shows us that

$$\langle U_T(f), U_T(g) \rangle = \langle f, g \rangle \tag{5.10}$$

but we also have that

$$\langle U_T(f), U_T(g) \rangle = \langle \lambda_1 f, \lambda_2 g \rangle = \lambda_1^{-1} \lambda_2 \langle f, g \rangle \tag{5.11}$$

We are thus able to conclude that

$$\langle f, g \rangle = \lambda_1^{-1} \lambda_2 \langle f, g \rangle \tag{5.12}$$

However, by assumption, $\lambda_1 \neq \lambda_2$ which implies that $\lambda_1^{-1} \lambda_2 \neq 1$. Thus, we are left to conclude that $\langle f, g \rangle = 0$.

□

5.2 Koopmanism and ergodicity

In Lemma 5.1, we saw that ergodicity was equivalent to the simplicity of the eigenvalue 1. In this next theorem, we will show that ergodicity implies some really elegant spectral properties of the Koopman operator.

Theorem 5.2. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an ergodic abstract dynamical system, and let U_T be its associated Koopman operator. Then its eigenvalues are all simple and form a subgroup of $\{z \in \mathbb{C} : |z| = 1\}$ (with respect to complex multiplication)*

Proof.
To begin our proof, we will show that any eigenvalue λ of the Koopman operator must be simple.

Let $\lambda \in \mathbb{C}$ be an eigenvalue, and E_λ its associated eigenspace. We would like to show that $\dim E_\lambda = 1$. To do so, we consider non-zero $f, g \in E_\lambda$ and we shall show that f and g are simply scalar multiples of each other.

Since $f, g \in E_\lambda$, we have that

$$\begin{aligned} U_T(f) &= \lambda f \\ U_T(g) &= \lambda g \end{aligned} \tag{5.13}$$

Moreover, if we consider their absolute values,

$$\begin{aligned} U_T(|f|) &= |f \circ T| = |\lambda|\,|f| = |f| \\ U_T(|g|) &= |g \circ T| = |\lambda|\,|g| = |g| \end{aligned} \tag{5.14}$$

Thus $|f|, |g|$ are eigenfunctions of U_T associated to the eigenvalue 1, and from Lemma 5.1 (b), we know that the ergodicity of our system implies that $|f|$ and $|g|$ must be constant functions. Thus, since f and g are non-zero, we must have $|f|, |g| > 0$ on all of Ω, and thus

$$\frac{f}{g}(\omega) = \frac{f(\omega)}{g(\omega)} \tag{5.15}$$

must be an element of $L^2(\mathbb{P})$. We then verify that

$$U_T\left(\frac{f}{g}\right) = \frac{f \circ T}{g \circ T} = \frac{\lambda}{\lambda}\frac{f}{g} = \frac{f}{g} \tag{5.16}$$

which shows that $\frac{f}{g} \in E_1$, and thus is a constant function. This is equivalent to saying that f and g are scalar multiples of each other. Thus $\dim E_\lambda = 1$.

To see that the set of eigenvalues of U_T forms a group, we need only check that it is closed under multiplication, and that for each eigenvalue, its multiplicative inverse is also an eigenvalue.

This is easily done by verifying that if $f \in E_{\lambda_1}$ and $g \in E_{\lambda_2}$ are both non-zero functions for some $\lambda_1, \lambda_2 \in \mathbb{C}$, then $fg \in L^2(\mathbb{P})$ defines an eigenfunction with respect to $\lambda_1\lambda_2$, and that $\frac{1}{f} \in L^2(\mathbb{P})$ is an eigenfunction for the eigenvalue λ_1^{-1} □

5.3 Koopman spectrum for circle rotation

In this section, we return to our classic example of the rotation on the unit circle look at the Koopman spectrum for the dynamical system.

We recall that our system is given by

$$\begin{aligned} \Omega &= \{z \in \mathbb{C} : |z| = 1\} \\ \mathcal{F} &= \mathcal{B}(\Omega) \\ \mathrm{d}\mathbb{P} &= \frac{\mathrm{d}\theta}{2\pi} \\ T_\alpha(z) &= e^{i(2\pi\alpha)}z, \quad \alpha \in [0,1) \end{aligned} \tag{5.17}$$

and that this system is *uniquely ergodic* for irrational α.

In the proof of unique ergodicity, we introduced the collection of functions $\mathscr{F} = \{f_m : m \in \mathbb{Z}\}$ on Ω where

$$f_m(z) = z^m \tag{5.18}$$

We recall that $\mathscr{F}$ has a (complex)-linear span which forms a dense set in $C(\Omega)$ and thus in $L^2(\Omega)$. We now look at the action of the Koopman operator on a function $f_m \in \mathscr{F}$

$$\begin{aligned} U_{T_\alpha}(f_m)(z) &= f_m \circ T(z) \\ &= \left(e^{i(2\pi\alpha)}z\right)^m \\ &= e^{i(2\pi m\alpha)}z^m \\ &= e^{i(2\pi m\alpha)}f_m(z) \end{aligned} \tag{5.19}$$

Thus, we see that $\{e^{2\pi m\alpha} : m \in \mathbb{Z}\}$ is a collection of eigenvalues of U_{T_α}. In the following theorem, we show that this collection is in fact the complete set of eigenvalues.

Theorem 5.3. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T_\alpha)$ be the dynamical system as given above. Then*

$$\sigma_p(U_T) = \{e^{2\pi m\alpha} : m \in \mathbb{Z}\} \tag{5.20}$$

where $\sigma_p(U_T)$ denotes the point spectrum (eigenvalues) of U_T.

Proof.
We already know that $\{e^{i(2\pi m\alpha)} : m \in \mathbb{Z}\} \subset \sigma_p(U_T)$ since $\mathscr{F}$ is a collection of eigenfunctions of U_T. Thus, we must show that if $\lambda \in \mathbb{C}$, $\lambda \notin \{e^{2\pi m\alpha} : m \in \mathbb{Z}\}$, then λ must not be an eigenvalue.

Suppose $U_T(f) = \lambda f$ for some $\lambda \notin \{e^{i(2\pi m\alpha)} : m \in \mathbb{Z}\}$ and some $f \in L^2(\mathbb{P})$. We would like to show that $f = 0$. By Lemma 5.1 (d), we have for any $f_m \in \mathscr{F}$, that

$$\langle f, f_m \rangle = 0 \tag{5.21}$$

Which allows us to conclude that for any g in the linear span of $\mathscr{F}$, we also have

$$\langle f, g \rangle = 0 \tag{5.22}$$

A simple density argument allows us to then conclude that for any g in $L^2(\mathbb{P})$, we have that

$$\langle f, g \rangle = \langle g, f \rangle = 0 \tag{5.23}$$

Thus, in particular, (taking g to be an indicator function), we see that for any $E \in \mathcal{F}$, we have that

$$\int_E f \mathrm{d}\mathbb{P} = 0 \tag{5.24}$$

Which implies (by Lemma 2.4) that $f = 0$. □

6 Entropy in dynamical systems

Entropy is a concept that is fundamental in dynamical systems theory and information theory. It provides a numerical measure of the 'disorder' or 'uncertainty' in a system.

The notion of entropy was first introduced by Rudolf Clausius in a series of papers he published in the 1850's and 1860's [4]. He used it to describe the heat that is dissipated as energy is transferred from one system to another. It is precisely this phenomenon that led Clausius to choose the name entropy; it combines the words 'energy' and '*tropos*', the Greek word for a change or transformation.

6.1 Independence and refinement of partitions

It is not possible to properly define the Shannon-Gibbs entropy without the notion of a partition. A partition of a probability space is intuitively just a splitting of the space into measurable subsets. We present a formal definition below.

Definition 6.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A collection $\mathcal{C} = \{C_i\}_{i \in I} \subset \mathcal{F}$ is called a *partition* if the following are true:

(a) I is countable.

(b) $C_i \cap C_j = \emptyset$ for all $i \neq j \in I$

(c) $\bigcup_{i \in I} C_i = \Omega$

It turns out that the study of entropy can, without loss of generality, be restricted to more 'well-behaved' partitions, and that the analysis becomes greatly simplified when one does so. To that end, we provide the definition of a normal partition below.

Definition 6.2. A partition $\mathcal{C} = \{C_i\}_I$ of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *normal* if I is a finite index set, and $\mathbb{P}(C_i) > 0$ for each $i \in I$. The collection of all normal partitions of the space is denoted by $\text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$.

As an abuse of notation, the trivial partition containing only the whole space itself will be denoted by Ω.

A useful way of thinking about partitions is to imagine them as instruments which can 'observe' the system, and depending on what state it is in, tell you which element of the partition it belongs to.

With this interpretation in mind, we introduce two concepts which will be central to understanding the notion of entropy.

Definition 6.3. Given two partitions $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$, we say that $\mathcal{C}$ *is refined by* $\mathcal{D}$, and write $\mathcal{C} \prec \mathcal{D}$, if each element of $\mathcal{C}$ can be written as a union of elements from $\mathcal{D}$.

Equivalently, we say that $\mathcal{D}$ *is a refinement of* $\mathcal{C}$, and write $\mathcal{D} \succ \mathcal{C}$.

One easily verifies that the symbols $\prec$ and $\succ$ define a partial order on $\text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$ with Ω as a minimal element, i.e.

$$\Omega \prec \mathcal{C} \quad \forall \mathcal{C} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P}) \tag{6.1}$$

Another useful relation between two partitions is that of independence.

Definition 6.4. Two partitions $\mathcal{C}$ and $\mathcal{D}$ in $\text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$ are said to be *independent* of each other if

$$\mathbb{P}(C_i \cap D_j) = \mathbb{P}(C_i)\mathbb{P}(D_j) \quad \forall C_i \in \mathcal{C}, D_j \in \mathcal{D} \tag{6.2}$$

In this case, we write $\mathcal{C} \perp \mathcal{D}$.

The final definition we present in this section is that of a *common refinement*. It specifies a canonical way of taking two partitions and constructing a third, which refines both of them.

Definition 6.5. Given two partitions, $\mathcal{C} = \{C_i\}_{i\in I}$ and $\mathcal{D} = \{D_j\}_{j\in J}$ in $\text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$, the *common refinement* of $\mathcal{C}$ and $\mathcal{D}$, denoted by $\mathcal{C} \vee \mathcal{D}$ is defined as

$$\mathcal{C} \vee \mathcal{D} \coloneqq \{C_i \cap D_j : i \in I, j \in J\} \tag{6.3}$$

We now present some basic properties of partitions.

Proposition 6.6. *Given arbitrary partitions $\mathcal{C}, \mathcal{D}, \mathcal{E} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$, the following hold:*

(a) $\mathcal{C} \prec \mathcal{C} \vee \mathcal{D}, \quad \mathcal{D} \prec \mathcal{C} \vee \mathcal{D}$.

(b) Associativity: $(\mathcal{C} \vee \mathcal{D}) \vee \mathcal{E} = \mathcal{C} \vee (\mathcal{D} \vee \mathcal{E})$

(c) $\Omega \vee \mathcal{C} = \mathcal{C}$

(d) $\Omega \perp \mathcal{C} \quad (\forall \mathcal{C} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P}))$

(e) $(\mathcal{C} \perp \mathcal{D}, \quad \mathcal{C} \prec \mathcal{D}) \Rightarrow \mathcal{C} = \Omega$

Proof.
Properties (a),(b),(c), and (d) are trivial to verify, and thus we will only prove property (e).

Suppose that we have two partitions, $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$ satisfying

$$\mathcal{C} \perp \mathcal{D}, \quad \mathcal{C} \tag{6.4}$$

Let C be an element of $\mathcal{C}$. Since $\mathcal{D}$ refines $\mathcal{C}$, we must have that

$$C = \bigcup_{i=1}^{n} D_i \tag{6.5}$$

for some collection $\{D_i\}_{i=1}^n \subset \mathcal{D}$. Furthermore, by independence, we have

$$\mathbb{P}(D_i) = \mathbb{P}(C \cap D_i) = \mathbb{P}(C)\mathbb{P}(D_i) \tag{6.6}$$

We must therefore have that $\mathbb{P}(C) = 1$ and since $\mathcal{C}$ is a normal partition, we must have $C = \Omega$. □

Proposition 6.6 (b) implies that we may write the common refinement of multiple partitions in a relatively simply way: Given a finite index set I, and a collection $\{\mathcal{C}_i\}_{i\in I}$ of normal partitions, we use the notation

$$\bigvee_{i\in I} \mathcal{C}_i \tag{6.7}$$

to denote the common refinement of all the partitions in the collection.

6.2 Shannon-Gibbs Entropy

We now introduce the Shannon-Gibbs entropy in the formalism that we have developed thus far.

Definition 6.7. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a partition $\mathcal{C} \in \text{Part}(\Omega, \mathcal{F}, \mathbb{P})$, the entropy of $\mathcal{C}$, denoted by $h(\mathcal{C})$, is defined as

$$h(\mathcal{C}) \coloneqq -\sum_{C_i \in \mathcal{C}} \mathbb{P}(C_i) \ln(\mathbb{P}(C_i)) \tag{6.8}$$

One should think of entropy as a measure of uncertainty, or alternatively, the expected 'information' gained from observing the state of the system. These interpretations will become more clear as we reveal the properties of entropy.

To elucidate the interpretation of entropy as the expected information, we define the information function on Ω for a partition.

Definition 6.8. Given a partition $\mathcal{C} \in \text{Part}(\Omega, \mathcal{F}, \mathbb{P})$, the information function with respect to $\mathcal{C}$ is given by

$$I_{\mathcal{C}}(\omega) \coloneqq -\ln(\mathbb{P}(C_i)) \quad \text{if } \omega \in C_i \tag{6.9}$$

The information thus measures the following: if $\omega \in \Omega$ belongs to a 'small' part of $\mathcal{C}$ (in the probabilistic sense), the information function returns a high value. It only returns 0 if $\mathcal{C} = \Omega$. Also, with this definition, one easily sees that

$$h(\mathcal{C}) = \int_{\Omega} I_{\mathcal{C}} \mathrm{d}\mathbb{P} = \mathbb{E}[I_{\mathcal{C}}] \tag{6.10}$$

This validates our interpretation of entropy as the average information contained in a state.

6.3 Conditional entropy

Another important concept in our study is that of conditional information and conditional entropy. Above, we interpreted entropy of a partition as the average information obtained from determining in which component of the partition a given state lies. The conditional entropy is a measure which relates the information of two given partitions. Specifically, it measures how much *more* information is obtained from determining the relevant component of one partition, given that we already have obtained information from another partition.

Definition 6.9. Given two partitions, $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$, the conditional information function of $\mathcal{C}$, given $\mathcal{D}$, is denoted by $I_{\mathcal{C}|\mathcal{D}}$ and is defined on Ω by

$$I_{\mathcal{C}|\mathcal{D}}(\omega) \coloneqq -\ln\left(\mathbb{P}(C_i|D_j)\right) = -\ln\left(\frac{\mathbb{P}(C_i \cap D_j)}{\mathbb{P}(D_j)}\right) \quad \text{for } \omega \in C_i \cap D_j \tag{6.11}$$

With Definition 6.9, we have a convenient way of defining conditional entropy.

Definition 6.10. The *conditional* entropy of $\mathcal{C}$ given $\mathcal{D}$, or the entropy of $\mathcal{C}$ conditioned on $\mathcal{D}$ (where $\mathcal{C}$ and $\mathcal{D}$ are two normal partitions of $(\Omega, \mathcal{F}, \mathbb{P})$) is simply the expectation of the conditional information:

$$h(\mathcal{C}|\mathcal{D}) \coloneqq \mathbb{E}[I_{\mathcal{C}|\mathcal{D}}] = -\sum_{C_i \in \mathcal{C}, D_j \in \mathcal{D}} \mathbb{P}(C_i \cap D_j) \ln\left(\mathbb{P}(C_i|D_j)\right) \tag{6.12}$$

We remark that the specific case of conditioning on the trivial partion reduces these definitions to the more basic notions of entropy and information:

$$h\left(\mathcal{C}|\Omega\right) = h\left(\mathcal{C}\right), \quad I_{\mathcal{C}|\Omega} = I_{\mathcal{C}} \tag{6.13}$$

We now present a number of fundamental properties of information and entropy

Proposition 6.11. *Consider $\mathcal{C}, \mathcal{D}, \mathcal{E} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$ and $\omega \in \Omega$. The following hold:*

(a) $h(\mathcal{C}) \geq 0$ *and* $h(\mathcal{C}) = 0 \Leftrightarrow \mathcal{C} = \Omega$

(b) If $\mathcal{C}$ *has* n *elements, then* $h(\mathcal{C}) \leq \ln(n)$ *and* $h(\mathcal{C}) = \ln(n) \Leftrightarrow \mathbb{P}(C_i) = \frac{1}{n} \quad \forall C_i \in \mathcal{C}$

(c) $\mathcal{C} \prec \mathcal{D} \Rightarrow h(\mathcal{C}) \leq h(\mathcal{D})$

(d) $h(\mathcal{C}|\mathcal{D}) \leq h(\mathcal{C})$ *and* $h(\mathcal{C}|\mathcal{D}) = h(\mathcal{C}) \Leftrightarrow \mathcal{D} \perp \mathcal{C}$

(e) $I_{\mathcal{C}\vee\mathcal{D}} = I_{\mathcal{D}} + I_{\mathcal{C}|\mathcal{D}}, \quad (h(\mathcal{C} \vee \mathcal{D}) = h(\mathcal{D}) + h(\mathcal{C}|\mathcal{D}))$

(f) $I_{\mathcal{C}\vee\mathcal{D}|\mathcal{E}} = I_{\mathcal{D}|\mathcal{E}} + I_{\mathcal{C}|\mathcal{D}\vee\mathcal{E}}, \quad (h(\mathcal{C} \vee \mathcal{D}|\mathcal{E}) = h(\mathcal{D}|\mathcal{E}) + h(\mathcal{C}|\mathcal{D} \vee \mathcal{E}))$

Prior to proving Proposition 6.11, we will present the celebrated Jensen's inequality for convex functions, named after the mathematician Johan Jensen who proved it in 1906 [17]. We shall see that it is fundamental to understanding and interpreting entropy.

Lemma 6.12. *Let $\varphi : (a, b) \to \mathbb{R}$ be a convex function and $x_1, \ldots, x_n$ a finite sequence in (a, b). Furthermore, let $a_1, \ldots, a_n$ be a sequence of strictly positive numbers whose sum is 1. Then,*

$$\varphi\left(\sum_{i=1}^{n} a_i x_i\right) \leq \sum_{i=1}^{n} a_i \varphi(x_i) \tag{6.14}$$

Furthermore, if φ is strictly convex, then equality holds only when $x_1 = x_2 = \cdots = x_n$.

Jensen's inequality is proven in much greater generality in [18]. We have presented a much simpler version as it will suffice in our applications of it throughout this thesis.

Proof of Proposition 6.11.

(a) The positivity of entropy is trivial, and moreover holds for the information function: $I_{\mathcal{C}}(\omega) \geq 0 \quad \forall \omega \in \Omega$.

If $h(\mathcal{C}) = 0$ then

$$\sum_{C_i \in \mathcal{C}} \mathbb{P}(C_i) \ln(\mathbb{P}(C_i)) = 0 \tag{6.15}$$

The only possibility is that $P(C_i) \in \{0, 1\}$ for each element $C_i \in \mathcal{C}$, and since $\mathcal{C}$ is normal, we must have $\mathcal{C} = \Omega$.

(b) Suppose $\mathcal{C} = \{C_i\}_{i=1}^n$. We then see that

$$\begin{aligned} -h(\mathcal{C}) &= \sum_{i=1}^n \mathbb{P}(C_i) \ln(\mathbb{P}(C_i)) = \sum_{i=1}^n \mathbb{P}(C_i) \left(- \ln \left(\frac{1}{\mathbb{P}(C_i)} \right) \right) \\ &\geq - \ln \left(\sum_{i=1}^n \frac{\mathbb{P}(C_i)}{\mathbb{P}(C_i)} \right) = - \ln(n) \end{aligned} \tag{6.16}$$

where the inequality was obtained by applying Lemma 6.12 to the strictly convex function $-\ln$. Negating the above proves the first part of (b) and once again by Lemma 6.12 we see that equality can only be achieved when $\mathbb{P}(C_1) = \mathbb{P}(C_2) = \cdots = \mathbb{P}(C_n)$, i.e. when

$$\mathbb{P}(C_i) = \frac{1}{n} \quad \text{for } i = 1, \ldots, n \tag{6.17}$$

(c) If $\mathcal{C} \prec \mathcal{D}$, then for each $C_i \in \mathcal{C}$, we have a finite sequence $D_{(i,1)}, \ldots, D_{(i,k_i)}$ of elements of $\mathcal{D}$ such that

$$\bigcup_{j=1}^{k_i} D_{(i,j)} = C_i \tag{6.18}$$

Monotonicity of the probability measure implies that $\mathbb{P}(D_{i,j}) \leq \mathbb{P}(C_i)$ for each j, and therefore $-\ln(P(D_{i,j})) \geq -\ln(\mathbb{P}(C_i))$. Thus, for $\omega \in C_i$,

$$I_{\mathcal{C}}(\omega) = -\ln(\mathbb{P}(C_i)) \leq -\ln(P(D_{i,j})) = I_{\mathcal{D}}(\omega) \tag{6.19}$$

and since each ω in Ω must belong to some $C_i \in \mathcal{C}$, this proves (c)

(d) The proof of this fact is a fairly simple application of Lemma 6.12:

$$\begin{aligned} h(\mathcal{C}|\mathcal{D}) &= -\sum_{i,j} \mathbb{P}(C_i \cap D_j) \ln(\mathbb{P}(C_i|D_j)) \\ &= \sum_i \mathbb{P}(C_i) \sum_j \mathbb{P}(D_j|C_i) \ln\left(\frac{1}{\mathbb{P}(C_i|D_j)}\right) \\ &\leq \sum_i \mathbb{P}(C_i) \ln\left(\sum_j \frac{\mathbb{P}(D_j)}{\mathbb{P}(C_i)}\right) \\ &= -\sum_i \mathbb{P}(C_i) \ln(\mathbb{P}(C_i)) = h(\mathcal{C}) \end{aligned} \tag{6.20}$$

equality is achieved if and only if $\mathbb{P}(C_i|D_j) = \mathbb{P}(C_i|D_k)$ for any i, j, k, which is equivalent to saying that $\mathbb{P}(C_i|D_j) = \mathbb{P}(C_i)$. This is an equivalent condition for independence of $\mathcal{C}$ and $\mathcal{D}$

(e) This part follows from (f) (which shall be proven below) by taking $\mathcal{E} = \Omega$.

(f) For $\omega \in C_i \cap D_j \cap E_k$, where $C_i \in \mathcal{C}$, $D_j \in \mathcal{D}$, and $E_k \in \mathcal{E}$, we have

$$\begin{aligned} I_{\mathcal{D}|\mathcal{E}}(\omega) + I_{\mathcal{C}|\mathcal{D}\vee\mathcal{E}}(\omega) &= -\ln\left(\frac{\mathbb{P}(D_j \cap E_k)}{\mathbb{P}(E_k)}\right) - \ln\left(\frac{\mathbb{P}(C_i \cap D_j \cap E_k)}{\mathbb{P}(D_j \cap E_k)}\right) \\ &= -\ln\left(\frac{\mathbb{P}(D_j \cap E_k)}{\mathbb{P}(E_k)} \cdot \frac{\mathbb{P}(C_i \cap D_j \cap E_k)}{\mathbb{P}(D_j \cap E_k)}\right) \\ &= -\ln\left(\frac{\mathbb{P}(C_i \cap D_j \cap E_k)}{\mathbb{P}(E_k)}\right) = I_{\mathcal{C}\vee\mathcal{D}|\mathcal{E}}(\omega) \end{aligned} \tag{6.21}$$

Since C_i, D_j and E_k were arbitrary elements of their respective partitions, the equality above must hold everywhere on Ω. Thus (e), and the entire lemma have been proved.

□

Proposition 6.11 provides us with invaluable intuition on how entropy measures the information contained in a partition. For instance, part (c) of the lemma tells us that a more refined partition allows us to distinguish a given state more precisely and thus one obtains necessarily more information then in a 'coarser' partition. Another example can be found in part (e), which tells us that the information gained by measuring against two partitions simultaneously is equivalent to the information gained from measuring one partition plus the information gained from measuring the second partition conditioned on the first.

The lemma can also be used to derive further interesting properties of information and entropy. For instance, parts (d) and (e) together imply the following important property.

Corollary 6.13. *Entropy is subadditive: Given two normal partitions $\mathcal{C}$ and $\mathcal{D}$, we have*

$$h(\mathcal{C} \vee \mathcal{D}) \leq h(\mathcal{C}) + h(\mathcal{D}) \tag{6.22}$$

with equality if and only if $\mathcal{C}$ and $\mathcal{D}$ are independent of each other.

6.4 An axiomatization of entropy

As early as Shannon's seminal paper from 1948 [1], efforts were made to mathematically axiomatize the concept of entropy (see for instance [19] and [20]). Khinchin's book, *Mathematical foundations of information theory*, was among the first to provide both concise and physically meaningful axioms to uniquely specify an entropy functional.

We first note that for a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and partition $\mathcal{C} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$, the only necessary information for defining entropy are the individual probabilities of each of the elements of $\mathcal{C}$. Thus, we can simply reinterpret entropy as a function on probability vectors from $\mathbb{R}^n$.

Formally speaking, if we define

$$\mathcal{P}_n = \left\{ p = (p_1, \ldots, p_n) \in \mathbb{R}^n : p_i \geq 0 \forall i\,,\ \sum_{i=1}^{n} p_i = 1 \right\} \tag{6.23}$$

and

$$\mathcal{P} = \bigcup_{i=1}^{\infty} \mathcal{P}_n \tag{6.24}$$

then we may re-envision the Shannon-Gibbs entropy defined in Section 6.2 as a function from $\mathcal{P}$ to $\mathbb{R}^+$. Specifically, we see that

$$\begin{aligned} h &: \mathcal{P} \to \mathbb{R}^+ \\ h(p_1, \ldots, p_n) &= -\sum_{i=1}^{n} p_i \ln p_i \end{aligned} \tag{6.25}$$

(with the convention $0 \ln 0 = 0$) is an extension of Definition 6.7, with the identification of $\text{Part}(\Omega, \mathcal{F}, \mathbb{P})$ to a subset of $\mathcal{P}$. Stating and proving uniqueness theorems about entropy in this formalism is in some sense the most universal and 'natural'.

Khinchin's five axioms are straightforward properties that we have already seen are satisfied by the Shannon-Gibbs entropy: If $H : \mathcal{P} \to \mathbb{R}$ is a function such that

(i) H is symmetric in its arguments.

(ii) for fixed n, $H : \mathcal{P}_n \to \mathbb{R}$ is a continuous function

(iii) H is maximized on $\mathcal{P}_n$ by $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$

(iv) $H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n)$

(v) for fixed $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$, we have that $H(\mathcal{C} \vee \mathcal{D}) = H(\mathcal{C}) + H(\mathcal{C}|\mathcal{D})$ (see Proposition 6.11 (e))

then $H = c \cdot h$ for some $c \geq 0$.

The uniqueness theorem that we shall prove in this section is particularly impressive in that it can be succinctly described with 2 axioms. We shall follow the proof as completed in a set of project notes by student Sherry Chu [21].

To be able to properly describe the axioms, we introduce some time-saving notation.

Definition 6.14. Let Ω_i is a finite probability space for each $i \in \{1, \ldots, m\}$, with size $|\Omega_i| = n_i$ and associated probability vector $(p_{(\Omega_i,1)}, \ldots, p_{(\Omega_i,n_i)})$ for each Ω_i. Furthermore, let $p = (p_1, \ldots, p_m) \in \mathcal{P}_m$. Then the p-mixing of the collection $\Omega_{i\,i=1}^{m}$, denoted by $\bigoplus_{i=1}^{m} p_i\Omega_i$, is a probability space with $\sum_{i=1}^{m} n_i$ elements, and probabilities given by

$$p_{ij} = p_i \cdot p_{(\Omega_i,j)} \tag{6.26}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n_i$.

If we take a p-mixing as in the above definition, we see that we can rewrite its Shannon-Gibbs entropy in terms of the relevant entropies of its components.

$$\begin{aligned}
h\left(\bigoplus_{i=1}^{m} p_i\Omega_i\right) &= -\sum_{i=1}^{m}\sum_{j=1}^{n_i} p_{ij} \ln p_{ij} \\
&= -\sum_{i=1}^{m}\sum_{j=1}^{n_i} p_{ij} \left(\ln p_i + \ln p_{(\Omega_i,j)}\right) \\
&= -\sum_{i=1}^{m} p_i \ln p_i \sum_{j=1}^{n_i} p_{(\Omega_i,j)} - \sum_{i=1}^{m} p_i \sum_{j=1}^{n_i} p_{(\Omega_i,j)} \ln p_{(\Omega_i,j)} \\
&= -\sum_{i=1}^{m} p_i \ln p_i + \sum_{i=1}^{m} p_i \cdot h(\Omega_i) \\
&= \sum_{i=1}^{m} p_i \cdot h(\Omega_i) + h(p_1, \ldots, p_n)
\end{aligned} \tag{6.27}$$

As we see above, the entropy of the p-mixing of a collection $\{\Omega_i\}_{i=1}^{m}$ is just the weighted average of the entropies of the components Ω_i, plus the entropy gained from the mixing itself, namely $h(p_1, \ldots, p_n)$. We shall aptly call this the *mixing property* of entropy.

We are now prepared to state and prove the main theorem.

Theorem 6.15. *Let $H : \mathcal{P} \to \mathbb{R}$ be a function such that*

(a) if we restrict the domain to $\mathbb{R}^2$, then $H : \mathcal{P}_2 \to R$ is a continuous function.

(b) the mixing property holds: for a collection $\Omega_i{}_{i=1}^m$ of probability spaces, and $p = (p_1, \ldots, p_m) \in \mathcal{P}_n$, we have

$$H\left(\bigoplus_{i=1}^{m} p_i \Omega_i\right) = \sum_{i=1}^{m} p_i H(\Omega_i) + H(p_1, \ldots, p_m)$$

(c) $H\left(\frac{1}{2}, \frac{1}{2}\right) > 0$

then $H = c \cdot h$ for some $c > 0$.

We remark that in the standard literature, a continuity assumption is generally made for the whole function $H : \mathcal{P} \to \mathbb{R}$ as opposed to simply the restricted function $H : \mathcal{P}_2 \to \mathbb{R}$. We shall see in the following proof that continuity is only required in two variables, and the rest follows from and induction proof which has no argument based on continuity.

Proof. For $n \in \mathbb{N}$, we define

$$f(n) \equiv H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \tag{6.28}$$

As a direct result of the mixing property, we have that

$$f(nm) = f(n) + f(m) \tag{6.29}$$

for $n, m \in \mathbb{N}$. Applying this inductively, we see that for $n, k \in \mathbb{N}$, we have $f(n^k) = kf(n)$.

We would now like to show that $f(n) = c \ln n$ for some $c > 0$. We shall do so in a few steps. We will first begin by showing that the difference between successive points of $f(n)$ tends to 0, and then use this fact among others to show that $f(n)$ must be of the desired form.

We begin with a small lemma:

Lemma 6.16. *If H satisfies the mixing property, then $H(1) = H(0, 1) = 0$.*

Proof of Lemma 6.16. We first show $H(1) = 0$. Take any vector $\rho = (\rho_1, \rho_2) \in \mathcal{P}_2$, then we apply the mixing property to find that

$$\begin{aligned} H(\rho_1, \rho_2) &= \rho_1 H(1) + \rho_2 H(1) + H(\rho_1, \rho_2) \\ &= H(1) + H(\rho_1, \rho_2) \end{aligned}$$

which, upon rearranging, implies that $H(1) = 0$. Next, we consider the vector $p = (0, p_1, p_2) \in \mathcal{P}_3$ and apply the mixing property in two different ways:

$$\begin{aligned} H(0, p_1, p_2) &= p_1 H(0, 1) + p_2 H(1) + H(p_1, p_2) \\ &= p_2 H(0, 1) + p_1 H(1) + H(p_1, p_2) \end{aligned}$$

Using the already proven fact that $H(1) = 0$, we find that

$$(p_2 - p_1) H(1, 0) = 0$$

Since we are free to choose p_1 and p_2 in such a way that $p_1 \neq p_2$, we must have $H(1, 0) = 0$. $\square$

We now proceed with showing that $\lim_{n\to\infty} f(n) - f(n-1) = 0$.

We define $d_n \equiv f(n) - f(n-1)$ and $\delta_n \equiv H\left(\frac{1}{n}, 1 - \frac{1}{n}\right)$. by the continuity of H on $\mathcal{P}_2$, we have $\delta_n \to H(1, 0) = 0$ as $n \to \infty$. If we can then show that $|d_n - \delta_n| \to 0$, then this will imply that $d_n \to 0$, and we will have completed the first half of this proof.

Before we estimate $|d_n - \delta_n|$, it is useful to remark that $f(1) = 0$, implying that $d_2 = f(2) - f(1) = f(2)$, and inductively, one easily finds that

$$f(n) = \sum_{i=2}^{n} d_i \tag{6.30}$$

Furthermore, by carefully applying the mixing property with $\left(\frac{1}{n}, 1 - \frac{1}{n}\right)$, we find

$$\begin{aligned} d_n &= f(n) - f(n-1) \\ &= H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) - H\left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}\right) \\ &= \frac{1}{n} H(1) + \left(1 - \frac{1}{n}\right) H\left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}\right) \\ &\quad + H\left(\frac{1}{n}, 1 - \frac{1}{n}\right) - H\left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}\right) \qquad \text{(by mixing property)} \\ &= \delta_n - \frac{\sum_{i=2}^{n-1} d_i}{n} \end{aligned} \tag{6.31}$$

Using equation (6.31) inductively, one can prove for any $N \in \mathbb{N}$ ($N \geq 2$), that

$$\sum_{n=2}^{N} d_n = \frac{1}{N} \sum_{n=2}^{N} n \delta_n \tag{6.32}$$

This leads to the following expression for d_N:

$$\begin{aligned} d_N &= \sum_{n=2}^{N} d_n - \sum_{n=2}^{N-1} d_n \\ &= \delta_N + \frac{1}{N}\sum_{n=2}^{N-1} n\delta_n - \sum_{n=2}^{N-1} d_n \\ &= \delta_N + \frac{1}{N}\sum_{n=2}^{N-1} n\delta_n - \frac{1}{N-1}\sum_{n=2}^{N-1} n\delta_n \\ &= \delta_N - \frac{1}{N(N-1)}\sum_{n=2}^{N-1} n\delta_n \end{aligned} \tag{6.33}$$

Thus, we may finally estimate the distance between d_N and δ_N:

$$\begin{aligned} |d_N - \delta_N| &= \left| \frac{1}{N(N-1)}\sum_{n=2}^{N-1} n\delta_n \right| \\ &\leq \frac{1}{N(N-1)}\left(\sum_{n=2}^{\sqrt{N}-1} n\,|\delta_n| + \sum_{n=\sqrt{N}}^{N-1} n\,|\delta_n| \right) \\ &\leq \frac{1}{N(N-1)}\left(N \sup_n |\delta_n| + (N-\sqrt{N})N \sup_{n\geq N} |\delta_n| \right) \end{aligned} \tag{6.34}$$

In particular, since $\delta_n \to 0$, we must have that $\sup_n |\delta_n| < \infty$ and $\sup_{n\geq N} |\delta_n| \to 0$ as $N \to \infty$. Thus, the expression on the right hand side of equation (6.34) goes to 0 as N goes to ∞. We thus have the desired result: $\lim_{n\to\infty} f(n) - f(n-1) = 0$.

We would now like to show that $f(n) = c\ln n$ for some $c > 0$. Clearly, if such a c exists, then we must have

$$\begin{aligned} c &= \frac{f(n)}{\ln n} \quad \forall n \geq 2 \\ \implies c &= \frac{f(2)}{\ln 2} \end{aligned} \tag{6.35}$$

Thus, we define $c \equiv f(2)/\ln 2$. By property (c) of H, we must have $c > 0$. Furthermore, since $f(n^k) = kf(n)$ for any $k, n \in \mathbb{N}$, we can see that

$$\lim_{n\to\infty} \frac{f(n)}{\ln n} = c \iff f(n) = c\ln n \quad \forall n \in \mathbb{N} \tag{6.36}$$

$f(n) = c\ln n \implies \lim_{n\to\infty} f(n)/\ln n = c$ is obvious. To prove the other direction, we suppose for contradiction that $\exists n_0 \in \mathbb{N}$ such that $f(n_0) \neq c\ln n_0$. in this case, if we consider taking the limit

along the subsequence n_0^k for $k = 1, 2, \ldots$, then we see that

$$\begin{aligned}\lim_{n\to\infty} \frac{f(n)}{\ln n} &= \lim_{k\to\infty} \frac{f(n_0^k)}{\ln n_0^k} \\ &= \frac{f(n_0)}{\ln n_0} \neq c\end{aligned}$$

which leads to a contradiction. Thus, we need only show that $\lim_{n\to\infty} f(n)/\ln n = c$. Furthermore, if we define the function

$$g(n) \equiv \begin{cases} 0 & \text{if } n = 1 \\ f(n) - c\ln n & \text{if } n \geq 2 \end{cases} \tag{6.37}$$

Then our goal becomes to show that

$$\lim_{n\to\infty} \frac{g(n)}{\ln n} = 0 \tag{6.38}$$

To do this, we first note that $g(2) = 0$ and we define $\epsilon_k = g(k+1) - g(k)$. We can see from the definition of $g(n)$ and the fact that $f(n) - f(n-1)$ tends to zero that $\epsilon_k \to 0$ as $k \to \infty$ For given $n \geq 2$, we may write $n = 2n_1 + r_1$ where $r_1 \in \{1, 2\}$. Specifically, for fixed n, we must have

$$\begin{aligned} n_1 &= \left\lceil \frac{n}{2} \right\rceil - 1 \\ r_1 &= n - 2n_1 \end{aligned} \tag{6.39}$$

We use this to re-express $g(n)$ in a helpful way:

$$g(n) = \sum_{k=2n_1}^{n-1} \epsilon_k + g(2n_1) \tag{6.40}$$

We note that $2n_1 \leq n - 1$ so that the expression above makes sense.

We can also re-express $g(2n_1)$ in a more useful manner:

$$\begin{aligned} g(2n_1) &= f(2n_1) - \frac{f(2)}{\ln 2}\ln(2n_1) \\ &= f(2) + f(n_1) - \frac{f(2)}{\ln 2}(\ln 2 + \ln n_1) \\ &= f(n_1) - c\ln n_1 \\ &= g(n_1) \end{aligned} \tag{6.41}$$

Thus, equation (6.40) becomes

$$g(n) = \sum_{k=2n_1}^{n-1} \epsilon_k + g(n_1) \tag{6.42}$$

We repeat this process, expressing $n_1 = 2n_2 + r_2$ with $r_2 \in \{1, 2\}$. More specifically, analogous to equation (6.42), we have

$$g(n_1) = \sum_{k=2n_2}^{n_1-1} \epsilon_k + g(n_2) \tag{6.43}$$

Iterating this process inductively, one sees that after k_0 steps, we eventually have $n_{k_0} = 1$ with the following recursive relations

$$\begin{aligned} n_k &= 2n_{k+1} + r_{k+1} \\ g(n_k) &= \sum_{i=2n_{k+1}}^{n_k-1} \epsilon_i + g(n_{k+1}) \end{aligned} \tag{6.44}$$

Using these relations, and taking $n_0 \equiv n$, one can neatly expand equation (6.42) as follows:

$$g(n) = \sum_{k=1}^{k_0} \sum_{i=2n_k}^{n_{k-1}-1} \epsilon_i \tag{6.45}$$

where we have implicitly used the fact that $g(2) = g(1) = 0$. Furthermore, one deduces from the first relation in equation (6.44) that the number of iterations, k_0, cannot exceed $\log_2 n = \frac{\ln n}{\ln 2}$.

We now proceed to estimate the size of $g(n)/\ln(n)$ using the results we have shown thus far. We first fix some $\varepsilon > 0$. Since $\lim_{i\to\infty} \epsilon_i = 0$, we may choose $N_1 \in \mathbb{N}$ large enough so that $|\epsilon_n| < \left(\frac{\ln 2}{2}\right) \frac{\varepsilon}{4}$.

$$\begin{aligned} \left|\frac{g(n)}{\ln n}\right| &= \frac{1}{\ln n} \left| \sum_{k=1}^{k_0} \sum_{i=2n_k}^{n_{k-1}-1} \epsilon_i \right| \\ &\leq \frac{1}{\ln n} \sum_{k=1}^{k_0} \left| \sum_{i=2n_k}^{n_{k-1}-1} \epsilon_i \right| \\ &< \frac{1}{\ln n} \left| \sum_{k=1}^{N_1} \epsilon_k \right| + \left(\frac{(k_0+1)2}{\ln n} \right) \left(\frac{\ln 2}{2} \right) \frac{\varepsilon}{4} \\ &\leq \frac{1}{\ln n} \left| \sum_{k=1}^{N_1} \epsilon_k \right| + \left(\frac{\ln n + \ln 2}{\ln n} \right) \frac{\varepsilon}{4} \end{aligned} \tag{6.46}$$

where in the third line, we have used the fact there are at most $2(k_0 + 1)$ total terms in the summation expression for $g(n)$ from equation (6.45), and in the fourth line, have used the fact that

$k_0 \leq \frac{\ln n}{\ln 2}$.

Since $\left|\sum_{k=1}^{N_1} \epsilon_k\right|$ is a finite constant, we may choose $N_2 \in \mathbb{N}$ large enough so that

$$\frac{1}{\ln n}\left|\sum_{k=1}^{N_1} \epsilon_k\right| < \frac{\varepsilon}{2} \quad \forall n \geq N_2 \tag{6.47}$$

Furthermore, we may choose $N_3 \in \mathbb{N}$ such that

$$\frac{\ln n + \ln 2}{\ln n} < 2 \quad \forall n \geq N_3 \tag{6.48}$$

Thus, if we choose $n \geq \max\{N_1, N_2, N_3\}$, equations (6.46), (6.47), and (6.48) imply that

$$\left|\frac{g(n)}{\ln n}\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \tag{6.49}$$

This finally demonstrates that $\lim_{n\to\infty} g(n)/\ln n = 0$, and thus $f(n) = c \ln n$, where $c = H(1/2, 1/2)/\ln 2$.

We now move on to the final part of this proof, showing that $H = c \cdot h$. We begin by first showing that this is true when we restrict the functions to $\mathcal{P}_2$. In other words, we would like to show that

$$H(p, 1-p) = c \cdot h(p, 1-p) = -c\left(p \ln p + (1-p)\ln(1-p)\right) \tag{6.50}$$

By property (a), it suffices for us to prove the above statement for $p \in \mathbb{Q} \cap (0, 1)$.

We thus consider $p = \frac{n}{m}$ for some $n, m \in \mathbb{N}$, $m > n$. Using the mixing property, we shall re-express $f(m)$ in terms of $f(n)$.

$$\begin{aligned} f(m) &= H\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) \\ &= p \cdot H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) + (1-p)H\left(\frac{1}{m-n}, \ldots, \frac{1}{m-n}\right) + H(p, 1-p) \\ &= pf(n) + (1-p)f(m-n) + H(p, 1-p) \end{aligned} \tag{6.51}$$

We may now rearrange equation (6.51) to obtain an expression for $H(p, 1-p)$:

$$\begin{aligned} H(p,1-p) &= f(m) - pf(n) - (1-p)f(m-n) \\ &= c\left(\ln m - p\ln n + (1-p)\ln(m-n)\right) \\ &= c\left((p+(1-p))\ln m - p\ln n + (1-p)\ln(m-n)\right) \\ &= c\left(p\left(\ln\frac{m}{n}\right) + (1-p)\left(\ln\frac{m}{m-n}\right)\right) \\ &= -c\left(p\ln p + (1-p)\ln(1-p)\right) \\ &= c\cdot h(p,1-p) \end{aligned} \tag{6.52}$$

Where we have used the fact that $p = n/m$ and $1-p = (m-n)/m$.

It is now left for us to prove the correct form for a general probability vector from $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$. We do this by induction on n.

We consider an arbitrary probability vector $(p_1, \ldots, p_n) \in \mathcal{P}_n$ (where $n > 2$), and we assume that for any $p \in \bigcup_{k=1}^{n-1} \mathcal{P}_k$, we have $H(p) = c \cdot h(p)$. We shall now apply the mixing property with the vector $(p_n, 1-p_n) \in \mathcal{P}_2$, along with the induction hypothesis to obtain the desired expression for H:

$$\begin{aligned} H(p_1,\ldots,p_n) &= p_n H(1) + (1-p_n)H\left(\frac{p_1}{1-p_n},\ldots,\frac{p_{n-1}}{1-p_n}\right) + H(p_n, 1-p_n) \\ &= -c\left[(1-p_n)\sum_{i=1}^{n-1}\frac{p_i}{1-p_n}\ln\left(\frac{p_i}{1-p_n}\right) + p_n\ln p_n + (1-p_n)\ln(1-p_n)\right] \\ &= -c\left[\sum_{i=1}^{n-1} p_i\ln p_i - (1-p_n)\ln(1-p_n) + p_n\ln p_n + (1-p_n)\ln(1-p_n)\right] \\ &= -c\sum_{i=1}^{n} p_i\ln p_i \\ &= c\cdot h(p_1,\ldots,p_n) \end{aligned} \tag{6.53}$$

where we have used the fact that $\sum_{i=1}^{n-1} p_i = 1 - p_n$. This completes the proof. □

6.5 Measurability and information

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, it is obvious from Definitions 6.8 and 6.9 that the information function and conditional information function for given partitions are measurable with respect to $\mathcal{F}$. However, we shall see that the structure of a normal partition leads to more specific measurability conditions for information.

Definition 6.17. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a normal partition $\mathcal{C}$, the σ-algebra generated by $\mathcal{C}$ shall be denoted by $\sigma(\mathcal{C})$

Since a partition $\mathcal{C}$ is a finite collection of disjoint, measurable sets whose union is Ω, we see that $\sigma(\mathcal{C})$ is simply the collection of unions of elements in the partition (with the empty set included) and is in bijective correspondence with the power set of $\mathcal{C}$. Furthermore, we clearly have $\mathcal{C} \prec \mathcal{D} \Rightarrow \sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$ for partitions $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$.

The following lemma tells us how the measurability of the information and conditional information functions are related to the partitions on which they are defined.

Lemma 6.18. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$ then $I_{\mathcal{C}}$ is measurable with respect to $\sigma(\mathcal{C})$ and $I_{\mathcal{C}|\mathcal{D}}$ is measurable with respect to $\sigma(\mathcal{C} \vee \mathcal{D})$.*

Let us now consider three normal partitions on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{C}$,$\mathcal{D}$, and $\mathcal{E}$, and suppose that $\mathcal{C} \prec \mathcal{D}$. From our established interpretation of entropy and information, we should expect that if we are already given the information from $\mathcal{D}$, then the expected additional information gained by measuring $\mathcal{E}$ would be less than the information gained had we only been given the information from $\mathcal{C}$. This intuition is formalized in the lemma below.

Lemma 6.19. *Let $\mathcal{C}$,$\mathcal{D}$, and $\mathcal{E}$ be normal partitions of $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathcal{C} \prec \mathcal{D}$. Then*

$$\mathbb{E}[I_{\mathcal{E}|\mathcal{D}} | \sigma(\mathcal{C} \vee \mathcal{E})] \leq I_{\mathcal{E}|\mathcal{C}} \tag{6.54}$$

and in particular,

$$h(\mathcal{E}|\mathcal{D}) \leq h(\mathcal{E}|\mathcal{C}) \tag{6.55}$$

Proof.
We first remark that since any element of $\sigma(\mathcal{C} \vee \mathcal{E})$ is a union of elements in $\mathcal{C}$ (which are disjoint), it suffices to show that for an arbitrary element $C_i \cap E_k \in \mathcal{C} \vee \mathcal{D}$, we have

$$\int_{C_i \cap E_k} I_{\mathcal{E}|\mathcal{D}} \mathrm{d}\mathbb{P} \leq \int_{C_i \cap E_k} I_{\mathcal{E}|\mathcal{C}} \mathrm{d}\mathbb{P} \tag{6.56}$$

Since $\mathcal{C} \prec \mathcal{D}$, we may write $C_i = \bigcup_{j=1}^{m} D_{(i,j)}$ with $\{D_{(i,j)}\}_{j=1}^{m} \subset \mathcal{D}$ so that

$$\sum_{j=1}^{m} \mathbb{P}(D_{(i,j)}) = \mathbb{P}(C_i) \tag{6.57}$$

We may therefore expand the left hand side of equation (6.56) in the following manner:

$$\begin{aligned}\int_{C_i\cap E_k} I_{\mathcal{E}|\mathcal{D}}\mathrm{d}\mathbb{P} &= \sum_{j=1}^{m}\int_{D_{(i,j)}\cap E_k} I_{\mathcal{E}|\mathcal{D}}\mathrm{d}\mathbb{P} = -\sum_{j=1}^{m}\mathbb{P}(D_{(i,j)}\cap E_k)\ln\left(\frac{\mathbb{P}(D_{(i,j)}\cap E_k)}{\mathbb{P}(D_{(i,j)})}\right)\\ &= -\sum_{j=1}^{m}\mathbb{P}(D_{(i,j)})\mathbb{P}(E_k|D_{(i,j)})\ln(\mathbb{P}(E_k|D_{(i,j)}))\\ &= -\mathbb{P}(C_i)\sum_{j=1}^{m}\left(\frac{\mathbb{P}(D_{(i,j)})}{\mathbb{P}(C_i)}\right)\mathbb{P}(E_k|D_{(i,j)})\ln(\mathbb{P}(E_k|D_{(i,j)}))\end{aligned} \tag{6.58}$$

Now, by taking $\varphi(x) = x\ln(x)$, with $a_j = \frac{\mathbb{P}(D_{(i,j)})}{\mathbb{P}(C_i)}$ and $x_j = \mathbb{P}(E_k|D_{(i,j)})$, we may apply Lemma 6.12 (since φ is strictly convex on $(0,\infty)$) to find that $-\sum_{j=1}^{m} a_j\varphi(x_j) \leq -\varphi(\sum_{j=1}^{m} a_j\cdot x_j)$, which expands as follows,

$$\begin{aligned}&-\mathbb{P}(C_i)\sum_{j=1}^{m}\left(\frac{\mathbb{P}(D_{(i,j)})}{\mathbb{P}(C_i)}\right)\mathbb{P}(E_k|D_{(i,j)})\ln(\mathbb{P}(E_k|D_{(i,j)}))\\ &\leq -\mathbb{P}(C_i)\left(\sum_{j=1}^{m}\frac{\mathbb{P}(D_{(i,j)})}{\mathbb{P}(C_i)}\mathbb{P}(E_k|D_{(i,j)})\right)\ln\left(\sum_{j=1}^{m}\frac{\mathbb{P}(D_{(i,j)})}{\mathbb{P}(C_i)}\mathbb{P}(E_k|D_{(i,j)})\right)\\ &= -\mathbb{P}(C_i)\mathbb{P}(E_k|C_i)\ln(\mathbb{P}(E_k|C_i))\\ &= -\mathbb{P}(C_i\cap E_k)\ln(\mathbb{P}(E_k|C_i)) = \int_{C_i\cap E_k} I_{\mathcal{E}|\mathcal{C}}\mathrm{d}\mathbb{P}\end{aligned} \tag{6.59}$$

Summarizing the work above,

$$\int_{C_i\cap E_k} I_{\mathcal{E}|\mathcal{D}}\mathrm{d}\mathbb{P} \leq \int_{C_i\cap E_k} I_{\mathcal{E}|\mathcal{C}}\mathrm{d}\mathbb{P} \tag{6.60}$$

which was precisely what we needed to show. □

6.6 Kolmogorov-Sinai entropy

Our discussion about entropy thus far has not included any mention of dynamics. We shall now see how one may suitably define a quantity which we may call the dynamical entropy of a system.

Appearing first in papers by Russian mathematician Andrei Kolmogorov ([22] and [23]), and expanded upon by his student Yakov Sinai [24], this concept of a dynamical entropy has proven useful in classifying types of dynamical systems.

Given an ADS $(\Omega,\mathcal{F},\mathbb{P},T)$, a partition $\mathcal{C}\in\mathrm{Part}\,(\Omega,\mathcal{F},\mathbb{P})$, and a number $n\in\mathbb{N}$ one easily verifies that the collection $T^{-n}\mathcal{C}$, defined by

$$T^{-n}\mathcal{C} \coloneqq \{T^{-n}(C) : C\in\mathcal{C}\} \tag{6.61}$$

defines a normal partition, as the operation $C \mapsto T^{-n}(C)$ preserves intersections and unions. We shall call $T^{-n}\mathcal{C}$ the *n-th pull-back of* $\mathcal{C}$ If we further assume that T is invertible, then $T^n\mathcal{C}$ is a partition for any $n \in \mathbb{Z}$ ($T^0\mathcal{C} = \mathcal{C}$). For n positive, we shall call $T^n(\mathcal{C})$ the *n-nth push-foward of* $\mathcal{C}$. In what follows, we shall assume for simplicity that T is invertible.

A simple observation one can make about the 'pull-back' and 'push-forward' actions is that they preserve entropy: Given partitions $\mathcal{C}, \mathcal{D} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$,

$$h(\mathcal{C}|\mathcal{D}) = h(T^n\mathcal{C}|T^n\mathcal{D}) \quad \forall n \in \mathbb{Z} \tag{6.62}$$

A similar relation holds for the information and conditional information functions,

$$I_{\mathcal{C}|\mathcal{D}}(\omega) = I_{T^n\mathcal{C}|T^n\mathcal{D}}(T^n\omega) \quad \forall n \in \mathbb{Z}, \omega \in \Omega \tag{6.63}$$

Definition 6.20. Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system, and $\mathcal{C} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$. The dynamical entropy of $\mathcal{C}$, denoted by $h_T(\mathcal{C})$, is given by

$$h_T(\mathcal{C}) \coloneqq \lim_{n\to\infty} \frac{1}{n} h\left(\bigvee_{i=-(n-1)}^{0} T^i\mathcal{C}\right) \tag{6.64}$$

The Kolmogorov-Sinai entropy, denoted by h_T is then defined as the supremum of the dynamical entropy over all normal partitions,

$$h_T \equiv \sup_{\mathcal{C}\in\text{Part}\,(\Omega,\mathcal{F},\mathbb{P})} h_T(\mathcal{C}) \tag{6.65}$$

It is not clear from the definition that the dynamical entropy should even exist for a given partition. However, its existence is a direct consequence of the subadditivity property from Corollary 6.13 and Fekete's lemma, a simple, yet famous and important result about subadditive sequences of real numbers, proven by Michael Fekete in 1923 [25]. We state the lemma without proof here.

Lemma 6.21. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a subadditive sequence of real numbers, i.e.*

$$a_{n+m} \leq a_n + a_m \quad \forall n, m \in \mathbb{N} \tag{6.66}$$

then we have that $\lim_{n\to\infty} \frac{a_n}{n}$ *exists (could be* $-\infty$*), and*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\in\mathbb{N}} \frac{a_n}{n} \tag{6.67}$$

Of course, since entropy is a positive quantity for any partition, the limit defined in equation

(6.64) not only exists but is in fact finite, and we have that

$$h_T(\mathcal{C}) \leq h(\mathcal{C}) \tag{6.68}$$

We shall now introduce some shorthand notation to help in our discussion of dynamical entropy. Assuming the ADS $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is invertible, and given a partition $\mathcal{C} \in \text{Part}(\Omega, \mathcal{F}, \mathbb{P})$, and $m, n \in \mathbb{Z}$ with $m \leq n$, we denote by $\mathcal{C}_m^n$ the following partition,

$$\mathcal{C}_m^n = \bigvee_{i=m}^{n} T^i \mathcal{C} \tag{6.69}$$

We will also define, for $n \in \mathbb{N}$, $\mathcal{C}^n \equiv \mathcal{C}_{-(n-1)}^0$. With this, we observe that equation (6.64) may be rewritten as

$$h_T(\mathcal{C}) = \lim_{n\to\infty} \frac{1}{n} h(\mathcal{C}^n) \tag{6.70}$$

Another lemma about sequences of real numbers will prove useful in providing more interpretation for the dynamical entropy of a system:

Lemma 6.22. *If $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that $\lim_{n\to\infty} a_n$ exists, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i \tag{6.71}$$

exists as well and

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i \tag{6.72}$$

To help understand what it is exactly that the dynamical entropy is measuring about a system, we consider the sequence of real numbers

$$h_n \equiv h(\mathcal{C}^n), \quad h_0 = 0 \tag{6.73}$$

for some $\mathcal{C} \in \text{Part}(\Omega, \mathcal{F}, \mathbb{P})$. Looking at the difference of successive terms,

$$h_n - h_{n-1} = h(\mathcal{C}^n) - h\left(\mathcal{C}^{n-1}\right) = h\left(\mathcal{C}|\mathcal{C}_{-(n-1)}^1\right) \tag{6.74}$$

where in the last equality, we applied Proposition 6.11 (e). One verifies that $h_n - h_{n-1}$ is a decreasing sequence of positive numbers (as a consequence of both Proposition 6.11 and Lemma 6.19) and

thus has a finite limit. Thus by Lemma 6.22, we have that

$$\lim_{n\to\infty} h\left(\mathcal{C}|\mathcal{C}_{-(n-1)}^{1}\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(h_n - h_{n-1}) = \lim_{n\to\infty}\frac{1}{n}h_n = h_T(\mathcal{C}) \tag{6.75}$$

In this way, a physical interpretation becomes clear. The dynamical entropy of a partition is simply the expected additional information gained from observing the system with this partition, given (conditioned on) its entire history.

Of course, when T is invertible, one makes the relatively simple observation, that $h_T(\mathcal{C}) = h_{T^{-1}}(\mathcal{C})$, and thus

$$h_T(\mathcal{C}) = \lim_{n\to\infty} h\left(\mathcal{C}|\mathcal{C}_1^{n-1}\right) \tag{6.76}$$

as well.

6.7 The Shannon-McMillan-Breiman theorem

The Shannon-McMillan-Breiman theorem is considered by many to be the cornerstone of information theory. It relates the notion of the ergodicity of a system to its dynamical entropy and in many ways, is an analog of Birkhoff's ergodic theorem.

Claude Shannon first proved the theorem in the specific case of Markhov processes in his classic paper from 1948, which is considered to be the starting point of information theory [1]. This theorem was expanded upon to include general ergodic dynamical systems by McMillan (mcmill ref) and later proved for almost-everywhere convergence by Leo Breiman in 1957 [26].

In this subsection we state the theorem, as well as an important corollary known as the asymptotic equipartition property, and in the following two subsections, we shall provide two proofs of the theorem, whose approaches are different enough so that presenting both of them is justified from a pedagogical standpoint.

Theorem 6.23 (Shannon, McMillan, Breiman)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an ergodic, invertible ADS. Let $\mathcal{C} \in \text{Part}(\Omega, \mathcal{F}, \mathbb{P})$. Then, for $\mathbb{P}$-almost all ω, we have*

$$\lim_{n\to\infty}\frac{1}{n}I_{\mathcal{C}^n}(\omega) = h_T(\mathcal{C}) \tag{6.77}$$

Moreover, convergence is in $L^1(\mathbb{P})$ also.

6.8 Proof with martingales

In our first proof of the theorem, we shall rely on an application of one of Doob's martingale convergence theorems. Before we proceed with the proof of Theorem 6.23, we recall the definitions

of filtrations, stochastic processes, and martingales.

Definition 6.24. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sequence $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ of sub-σ-algebras of $\mathcal{F}$ is called a filtration if $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for each $n \in \mathbb{N}$. In other words, a *filtration* is an 'increasing' sequence of σ-algebras, compatible with the space $(\Omega, \mathcal{F}, \mathbb{P})$.

Filtrations are most commonly used to represent the way information is acumulated as one makes measurements over time. However, on its own, a filtration is insufficient to model a sequence of measurements. To that end, we introduce the notion of a stochastic process.

Definition 6.25. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a *real-valued (or complex-valued) stochastic process* is a sequence $(f_n)_{n=1}^{\infty}$ of real-valued (or complex-valued) random variables on Ω. That is to say that $f_n : \Omega \to \mathbb{R}$ (or $f_n : \Omega \to \mathbb{C}$) is $\mathcal{F}$-measurable for each $n \in \mathbb{N}$. One says that a stochastic process $(f_n)_{n=1}^{\infty}$ is *adapted* to the filtration $(\mathcal{F}_n)_{n=1}^{\infty}$ if each f_n is $\mathcal{F}_n$-measurable.

A stochastic process allows one to represent a dynamical system with a probabilistic time evolution, as opposed to the theory we already developed where the time evolution is generated by a fixed, deterministic evolution operator, T. The sequence of random variables (measurable functions) represents measurements of some system and are indexed by a discrete time.

Moreover, if a stochastic process $(f_n)_{n=1}^{\infty}$ is adapted to a filtration $(\mathcal{F}_n)_{n=1}^{\infty}$, then the filtration helps capture how information is gained as measurements are sequentially carried out. $\mathcal{F}_n$ contains the events that can be tested once the measurements up to time n have been made.

It is important to note that for any stochastic process $(f_n)_{n=1}^{\infty}$, a natural filtration (to which $(f_n)_{n=1}^{\infty}$ is adapted) is given by

$$\mathcal{F}_n = \sigma(f_1, f_2, \ldots, f_n) \tag{6.78}$$

where $\sigma(f_1, f_2, \ldots, f_n)$ is the minimal σ-algebra such that $f_1, \ldots, f_n$ are measurable functions.

Stochastic processes are used as models in a wide variety of settings, and without further imposed structure, very little can be said about the asymptotic behaviour of an arbitrary stochastic process. However, by requiring it to have a certain expected monotonic behaviour, a lot of knowledge can be gleaned.

Definition 6.26. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A real-valued stochastic process $(f_n)_{n\in\mathbb{N}}$ is called a *supermartingale* with respect to a filtration $(\mathcal{F}_n)_{n=1}^{\infty}$ if, for all $n \in \mathbb{N}$, we have that $f_n \in L^1(\mathbb{P})$, and

(i) $(f_n)_{n=1}^{\infty}$ is adapted to $(\mathcal{F}_n)_{n=1}^{\infty}$

(ii) $\mathbb{E}[f_{n+1}|\mathcal{F}_n] \leq f_n \quad \forall n \in \mathbb{N}$

If equality holds in (ii), then we call $(f_n)_{n=1}^{\infty}$ a *martingale*. A stochastic process $(g_n)_{n\in\mathbb{N}}$ is called a *submartingale* if the stochastic process $(-g_n)_{n\in\mathbb{N}}$ is a supermartingale.

Martingales (as well as super- and sub-martingales) have many nice properties, specifically relating to their convergence. We now present a theorem, named after the American mathematician Joseph Doob, which deals with two types of convergence for martingales.

Theorem 6.27 (Doob's martingale convergence theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(f_n)_{n=1}^{\infty}$ be a supermartingale with respect to filtration $(\mathcal{F}_n)_{n=1}^{\infty}$. suppose further that $(f_n)_{n=1}^{\infty}$ is uniformly integrable, that is to say*

$$\lim_{K\to\infty}\left(\sup_{n\in\mathbb{N}}\int_{\{|f_n|>K\}}|f_n|d\mathbb{P}\right)=0 \tag{6.79}$$

Then there exists a function $f \in L^1(\mathbb{P})$ such that $f_n \to f$ $\mathbb{P}$-almost everywhere and in $L^1(\mathbb{P})$, that is, for almost all ω, we have

$$\lim_{n\to\infty} f_n(\omega) = f(\omega) \tag{6.80}$$

and furthermore,

$$\lim_{n\to\infty}\mathbb{E}[|f_n - f|] = 0 \tag{6.81}$$

A complete proof of this theorem can be found in [27].

With the concept of martingales and with Theorem 6.27 at our disposal, we are now in a position to prove Theorem 6.23.

Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an invertible, ergodic abstract dynamical system, and $\mathcal{C} \in \text{Part}\,(\Omega, \mathcal{F}, \mathbb{P})$. We would like to show that

$$\frac{1}{n} I_{\mathcal{C}^n} \to h_T(\mathcal{C}) \tag{6.82}$$

To do so, we inductively apply Proposition 6.11, parts (e) and (f) to $I_{\mathcal{C}^n}$ to find

$$\begin{aligned} I_{\mathcal{C}^n} &= I_{\mathcal{C}^0_{-(n-1)}} = I_{\mathcal{C}} + I_{\mathcal{C}^{-1}_{-(n-1)}|\mathcal{C}} \\ &= I_{\mathcal{C}} + I_{\mathcal{C}^0_{-(n-2)}|T\mathcal{C}} \circ T \\ &= \cdots \\ &= I_{\mathcal{C}} + \sum_{k=1}^{n-1} I_{\mathcal{C}|\mathcal{C}_1^k} \circ T^k \end{aligned} \tag{6.83}$$

Thus, by defining

$$\begin{aligned} f_0 &= I_{\mathcal{C}} \\ f_k &= I_{\mathcal{C}|\mathcal{C}_1^k} \end{aligned} \tag{6.84}$$

we see that

$$\frac{1}{n} I_{\mathcal{C}^n} = \frac{1}{n} \sum_{k=0}^{n-1} f_k \circ T^k \tag{6.85}$$

It is obvious that $(f_k)_{k=0}^{\infty}$ is a sequence of positive functions. We would now like to show that $(f_k)_{k=0}^{\infty}$ is a (uniformly integrable) supermartingale with respect to the filtration

$$\mathcal{F}_k = \sigma(\mathcal{C}_0^k) \tag{6.86}$$

The fact that its a supermartingale follows simply from Lemma 6.19. To show uniform integrability, we present the following lemma.

Lemma 6.28.

$$\sup_{k \geq 0} f_k \in L^1(\mathbb{P}) \tag{6.87}$$

Proof of Lemma 6.28.
Showing the integrability of $\sup_{k\geq 0} f_k$ is equivalent to showing that the function $g : [0, \infty) \to [0, \infty)$ defined by

$$g(x) = \mathbb{P}\left(\left\{ \sup_{k \geq 0} f_k > x \right\} \right) \tag{6.88}$$

is integrable with respect to the Lebesgue measure on $[0, \infty)$. To see why this is true, we rewrite the integral of g and apply Fubini's theorem:

$$\begin{aligned} \int_0^{\infty} g(x) \mathrm{d}x &= \int_0^{\infty} \mathbb{P}\left(\left\{ \sup_{k \geq 0} f_k > x \right\} \right) \mathrm{d}x \\ &= \int_0^{\infty} \left(\int_{\Omega} \chi_{\left\{ \sup_{k\geq 0} f_k > x \right\}} \mathrm{d}\mathbb{P} \right) \mathrm{d}x \\ &= \int_{\Omega} \left(\int_0^{\infty} \chi_{\left\{ \sup_{k\geq 0} f_k > x \right\}} \mathrm{d}x \right) \mathrm{d}\mathbb{P} \\ &= \int_{\Omega} \sup_{k \geq 0} f_k \mathrm{d}\mathbb{P} \end{aligned} \tag{6.89}$$

Thus we have equivalence of integrability. We now proceed to show that g is integrable.

We begin by noting that $\mathcal{C}$ is a finite partition, so we denote its size by N, and thus have $\mathcal{C} = \{C_i\}_{i=1}^N$. Now, we fix $x \in [0, \infty)$ and we define

$$B(x) \coloneqq \left\{ \omega \in \Omega : \sup_{k \geq 0} f_k > x \right\} \tag{6.90}$$

so that $g(x) = \mathbb{P}(B(x))$. We now look at the elements of the partition $\mathcal{C}$ and examine how they intersect with $B(x)$.

Suppose $C_i \in \mathcal{C}$ such that $\mathbb{P}(C_i) < e^{-x}$, then, for $\omega \in C_i$, we have

$$f_0(\omega) = -\ln(\mathbb{P}(C_i)) > x \tag{6.91}$$

thus implying that $C_i \subset B(x)$. Let $I = \{i : \mathbb{P}(C_i) < e^{-x}\}$ be the collection of all indices for which the corresponding elements have the appropriate bound on their probabilities. Then it follows that

$$\bigcup_{i \in I} C_i \subset B(x) \tag{6.92}$$

If we consider the set of remaining indices, $J = \{1, \ldots, N\} \setminus I$, then for $j \in J$, we clearly have $\mathbb{P}(C_j) \geq e^{-x}$. We consider, for fixed $j \in J$, the intersection

$$B(x) \cap C_j \tag{6.93}$$

Without loss of generality, we may assume that $B(x) \cap C_j \neq \emptyset$. For $\omega \in B(x) \cap C_j$, we define

$$k(\omega) \equiv \min\{k \in \mathbb{N} : f_k(\omega) > x\} \tag{6.94}$$

and define $D(\omega)$ as the unique element of $\mathcal{C}_1^{k(\omega)}$ such that $\omega \in D(\omega)$. Now we present some basic facts about $D(\omega)$ for $\omega \in B(x) \cap C_j$.

(1) $C_j \cap D(\omega) \subset B(x) \cap C_j$

(2) $\mathbb{P}(C_j \cap D(\omega)) \leq e^{-x}\mathbb{P}(D(\omega))$

(3) for $\omega_1, \omega_2 \in B(x) \cap C_j$, either $D(\omega_1) = D(\omega_2)$, or $D(\omega_1) \cap D(\omega_2) = \emptyset$

Fact (1) is obvious. to prove fact (2), we simply note that by our definitions of $k(\omega)$ and $D(\omega)$, we have

$$f_{k(\omega)}(\omega) > x \tag{6.95}$$

which is equivalent to

$$\ln\left(\frac{\mathbb{P}(C_j \cap D(\omega))}{\mathbb{P}(D(\omega))}\right) > x \tag{6.96}$$

We thus algebraically manipulate the above to obtain

$$\mathbb{P}(C_j \cap D(\omega)) < e^{-x}\mathbb{P}(D(\omega)) \tag{6.97}$$

To see why fact (3) is true, we assume that $D(\omega_1) \neq D(\omega_2)$ and we first consider the case when $k(\omega_1) = k(\omega_2) \equiv k$ for some $k \in \mathbb{N}$. In this case, we see that $D(\omega_1), D(\omega_2) \in \mathcal{C}_1^k$, and since they are distinct elements of the same partition, they must be disjoint.

The other case we must consider is when $k(\omega_1) \neq k(\omega_2)$. Without loss of generality assume $k(\omega_1) < k(\omega_2)$. Then we have that $\mathcal{C}_1^{k(\omega_1)} \prec \mathcal{C}_1^{k(\omega_2)}$, and in particular, either $D(\omega_2) \subset D(\omega_1)$ or $D(\omega_2) \cap D(\omega_1) = \emptyset$. the first possibility contradicts the minimality of $k(\omega_2)$, and thus $D(\omega_2) \cap D(\omega_1) = \emptyset$.

These three facts combined allow us to conclude that for each $j \in J$, there exists a countable index set K_j and a countable collection of sets

$$\{D_k^j\}_{k \in K_j} \subset \bigcup_{n \in \mathbb{N}} \mathcal{C}_1^n \tag{6.98}$$

such that $D_k^j \cap D_l^j = \emptyset \quad \forall k \neq l$, $\mathbb{P}(C_j \cap D_k^j) < e^{-x}\mathbb{P}(D_k^j)$, and

$$\bigsqcup_{k \in K_j} (C_j \cap D_k^j) = B(x) \cap C_j \tag{6.99}$$

Thus, given our knowledge of the index sets I and J, we can bound the probability of $B(x)$:

$$\begin{aligned}
g(x) &= \mathbb{P}(B(x)) \\
&= \sum_{i\in I} \mathbb{P}(B(x)\cap C_i) + \sum_{j\in J} \mathbb{P}(B(x)\cap C_j) \\
&= \sum_{i\in I} \mathbb{P}(B(x)\cap C_i) + \sum_{j\in J}\sum_{k\in K_j} \mathbb{P}(C_j\cap D_k^j) \\
&< \sum_{i\in I} e^{-x} + \sum_{j\in J}\sum_{k\in K_j} e^{-x}\mathbb{P}(D_k^j) \\
&= \sum_{i\in I} e^{-x} + \sum_{j\in J} e^{-x}\mathbb{P}(\bigcup_{k\in K_j} D_k^j) \\
&\leq \sum_{i=1}^{N} e^{-x} \\
&= Ne^{-x}
\end{aligned} \tag{6.100}$$

This completes the proof that $\sup_{n\geq 0} f_n \in L^1(\mathbb{P})$. □

From Lemma 6.28, it follows that the sequence f_n satisfies the uniform integrability condition from Theorem 6.27, since

$$\begin{aligned}
\lim_{K\to\infty}\sup_{n\geq 0}\int_{\{f_n>K\}} f_n \mathrm{d}\mathbb{P} &\leq \lim_{K\to\infty}\sup_{n\geq 0}\int_{\{\sup_{n\geq 0} f_n>K\}} f_n \mathrm{d}\mathbb{P} \\
&\leq \lim_{K\to\infty}\int_{\{\sup_{n\geq 0} f_n>K\}} \sup_{n\geq 0} f_n \mathrm{d}\mathbb{P}
\end{aligned} \tag{6.101}$$

where in the first step, we used the monotonicity of the Lebesgue integral on positive functions and in the second step, we used the fact that

$$\int_{\{\sup_{n\geq 0} f_n>K\}} \sup_{n\geq 0} f_n \mathrm{d}\mathbb{P} \geq \int_{\{\sup_{n\geq 0} f_n>K\}} f_n \mathrm{d}\mathbb{P} \tag{6.102}$$

for all $n \in \mathbb{N}$.

Since $\sup_{n\geq 0} f_n \in L^1(\mathbb{P})$, we have that $\lim_{K\to\infty}\int_{\{\sup_{n\geq 0} f_n>K\}} \sup_{n\geq 0} f_n \mathrm{d}\mathbb{P} = 0$, and thus, from (6.101) we see that f_n is uniformly integrable.

So, by Theorem 6.27, there exists $f \in L^1$ such that $f_n \to f$ $\mathbb{P}$-almost surely, and $\mathbb{E}[|f_n - f|] \to 0$

as $n \to \infty$. In particular, we have

$$\begin{aligned}\mathbb{E}[f] &= \lim_{n\to\infty} \mathbb{E}[f_n] \\ &= \lim_{n\to\infty} h\left(\mathcal{C}|\mathcal{C}_1^n\right) \\ &= h_{T^{-1}}\left(\mathcal{C}\right) \\ &= h_T\left(\mathcal{C}\right)\end{aligned} \tag{6.103}$$

We will now make use of f and its properties to rewrite and analyze the sequence of interest, $\frac{1}{n} I_{\mathcal{C}^n}$. Starting from equation (6.85), we have

$$\begin{aligned}\frac{1}{n} I_{\mathcal{C}^n} &= \sum_{k=0}^{n-1} f_k \circ T^k \\ &= \sum_{k=0}^{n-1} (f_k - f) \circ T^k + \sum_{k=0}^{n-1} f \circ T^k\end{aligned} \tag{6.104}$$

By Theorem 3.2, and the fact that $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is an ergodic system, we must have that

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} f \circ T^k = \mathbb{E}[f] = h_T\left(\mathcal{C}\right) \tag{6.105}$$

where this limit can be interpreted both $\mathbb{P}$-almost surely and in $L^1(\mathbb{P})$. It therefore remains to show that

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} (f_k - f) \circ T^k = 0 \tag{6.106}$$

both in $L^1(\mathbb{P})$, and $\mathbb{P}$-almost surely. To do so, we define a sequence of positive functions

$$G_N \equiv \sup_{k\geq N} |f_k - f| \tag{6.107}$$

for each $N \geq 0$. By the construction of f, we see that

$$\lim_{N\to\infty} G_N = 0 \tag{6.108}$$

$\mathbb{P}$-almost surely. Furthermore, one easily verifies that

$$G_N \leq |f| + \left| \sup_{k\geq 0} f_k \right| \tag{6.109}$$

Since both f and $\sup_{k\geq 0} f_k$ are $L^1(\mathbb{P})$, (6.109) and the Lebesgue dominated convergence theorem

imply that

$$\lim_{N\to\infty} \mathbb{E}[G_N] = 0 \tag{6.110}$$

or, stated in an alternative way, for any $\varepsilon > 0$, we may choose an m large enough so that

$$\mathbb{E}[G_m] < \varepsilon \tag{6.111}$$

To conclude the proof, we fix an arbitrary $\varepsilon > 0$ and choose m large enough so that (6.111) is satisfied. Now, looking at the magnitude of the n-th term in the sequence, where n is some number larger than m, we find

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=0}^{n-1}(f_k - f)\circ T^k\right| &\le \frac{1}{n}\sum_{k=0}^{n-1}\left|(f_k - f)\circ T^k\right| \\ &\le \frac{1}{n}\sum_{k=0}^{m-1}\left|(f_k - f)\circ T^k\right| + \frac{1}{n}\sum_{k=m}^{n-1} G_m \circ T^k \\ &= \frac{1}{n}\sum_{k=0}^{m-1}\left|(f_k - f)\circ T^k\right| + \frac{1}{n}\sum_{k=0}^{n-1} G_m \circ T^k - \frac{1}{n}\sum_{k=0}^{m-1} G_m \circ T^k \end{aligned} \tag{6.112}$$

Now by taking the 'lim sup' of the above (as n tends to infinity), we can easily see that the first and third terms in the last line tend to zero, and the middle term is simply an ergodic sum which tends to $\mathbb{E}[G_m]$. Thus,

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{k=0}^{n-1}(f_k - f)\circ T^k\right| \le \mathbb{E}[G_m] < \varepsilon \tag{6.113}$$

This establishes the desired convergence (as $\varepsilon > 0$ was arbitrarily chosen) and thus the theorem is proved. □

7 Subadditivity and entropy: Kingman's subadditive theorem and extensions

To broaden our perspective on Shannon entropy, we shall study a special class of random variable sequences, called *subadditive*.

Definition 7.1. Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an ADS. A sequence of real-valued random variables $(X_n)_{n=1}^{\infty}$ is called *subadditive* if, for all $n, m \in \mathbb{N}$, we have

$$X_{n+m} \le X_n + X_m \circ T^n \tag{7.1}$$

$\mathbb{P}$-almost everywhere.

In 1968, Sir John Kingman published two important papers relating to the convergence of subadditive sequences, and his work culminated in an important theorem, now referred to as Kingman's subadditive ergodic theorem [28]. Its implications are far-reaching and it provides an important interpretation for some of the convergence theorems we have seen thus far.

Kingman's theorem was extended in the works of Derriennic [29], and alternative and elegant proofs were presented in a paper by Avila and Bochi [30].

As a classical example of a subadditive sequence, we consider an L^1 function f and consider the sequence of its ergodic averages,

$$S_n = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^n \tag{7.2}$$

One easily verifies that S_n satisfies (7.1) and is thus a subadditive sequence. In fact equality holds, and such a sequence is called additive.

As we shall see, Kingman's original theorem suffices to imply Birkhoff's ergodic theorem, but does not however imply the Shannon-McMillan-Breiman theorem. This is because the sequence

$$I_n = \frac{1}{n} I_{\mathcal{C}^n} \tag{7.3}$$

for some partition $\mathcal{C}$ is not a subadditive sequence.

Nevertheless, subadditivity plays a key role in understanding why the above sequence converges. One of Deriennic's extensions, aptly named the almost subadditive ergodic theorem, suffices to prove the convergence of I_n.

In this section, we shall first present (without proof) Kingman's original theorem, and then present Deriennic's generalization, and provide a proof based on that of Avila and Bochi.

7.1 Kingman's subadditive theorem

Theorem 7.2. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an ADS, and let (X_n) be a sequence of real-valued measurable functions such that X_1^+ (the positive part of X_1) is integrable. If X_n is a subadditive sequence, then*

$$X = \lim_{n\to\infty} \frac{1}{n} X_n \tag{7.4}$$

exists almost everywhere, in $[-\infty, \infty)$. Moreover, X^+ is integrable and

$$\int_\Omega X \, d\mathbb{P} = \lim_{n\to\infty} \frac{1}{n} \int_\Omega X_n \, d\mathbb{P} = \inf_{n\in\mathbb{N}} \frac{1}{n} \int_\Omega X_n \, d\mathbb{P} \tag{7.5}$$

One can see that Kingman's theorem at least partially implies Birkhoff's ergodic theorem: Take an ADS $(\Omega, \mathcal{F}, \mathbb{P}, T)$ and consider an $L^1(\mathbb{P})$ function, say f. We then define the stochastic process $(X_n)_{n=1}^{\infty}$ in the following manner

$$X_n \equiv \sum_{k=0}^{n-1} f \circ T^k \tag{7.6}$$

It is then a trivial exercise to check that X_n is a subadditive sequence. In particular, it is *additive*, namely

$$\begin{aligned} X_{n+m} &= \sum_{k=0}^{n+m-1} f \circ T^k \\ &= \sum_{k=0}^{n-1} f \circ T^k + \sum_{k=n}^{n+m-1} f \circ T^k \\ &= X_n + X_m \circ T^n \end{aligned} \tag{7.7}$$

Since $(X_1)^+ = f^+$ is integrable, Theorem 7.2 shows us that

$$X = \lim_{n\to\infty} \frac{1}{n} X_n \tag{7.8}$$

exists $\mathbb{P}$-almost everywhere and that

$$\begin{aligned} \int_\Omega X \mathrm{d}\mathbb{P} &= \lim_{n\to\infty} \frac{1}{n} \int_\Omega X_n \mathrm{d}\mathbb{P} \\ &= \lim_{n\to\infty} \frac{1}{n} \int_\Omega \sum_{k=0}^{n-1} f \circ T^k \mathrm{d}\mathbb{P} \\ &= \int_\Omega f \mathrm{d}\mathbb{P} \end{aligned} \tag{7.9}$$

where in the last equality, we have used the measure-preserving nature of T.

As we shall see later on, this theorem is not sufficient however to prove the Shannon-McMillan-Breiman theorem. To circumvent this, we now proceed to prove a theorem which extends Kingman's original theorem into a result which is powerful enough to establish Shannon's theorem with the right convergence properties.

7.2 The almost-subadditive ergodic theorem

Over the years there have been numerous extensions made to Kingman's original theorem about subadditivity (see for example [31] and [32], among others). In this section we present a variant of the theorem in which we relax the requirement that the relevant sequence be subadditive. Instead,

we allow for an appropriate error term, which satisfies certain integrability and boundedness conditions. We shall see that the entropy as defined in Section 6 will satisfy our conditions and the Shannon-McMillan-Breiman theorem will be a corollary.

Theorem 7.3. *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an abstract dynamical system, and let $(X_n)_{n=1}^{\infty}$ be a real-valued stochastic process such that $\mathbb{E}[X_1^+] < \infty$.Also, let $Y_n : \Omega \to \mathbb{R}^+$ be a sequence of non-negative random variables such that $\sup_{n\in\mathbb{N}} \mathbb{E}[Y_n] < \infty$ and*

$$\lim_{n\to\infty} \frac{Y_n}{n} = 0 \tag{7.10}$$

$\mathbb{P}$-almost surely.

If (X_n) is almost-subadditive with respect to (Y_n), that is, for all $n, m \in \mathbb{N}$, we have

$$X_{n+m} \leq X_n + X_m \circ T^n + Y_m \circ T^n, \tag{7.11}$$

then the following results hold

(1) There exists $X : \Omega \to \mathbb{R}$ measurable, such that $\mathbb{E}[X^+] < \infty$ and

$$\lim_{n\to\infty} \frac{X_n(\omega)}{n} = X(\omega) \tag{7.12}$$

for $\mathbb{P}$-almost all $\omega \in \Omega$. Furthermore, X is T-invariant: $X = X \circ T$.

(2)

$$\mathbb{E}[X] = \lim_{n\to\infty} \frac{\mathbb{E}[X_n]}{n} = \inf_{n\in\mathbb{N}} \frac{\mathbb{E}[X_n]}{n} \tag{7.13}$$

(3) If $X_1 \in L^1(\mathbb{P})$, i.e. $\mathbb{E}[|X_1|] < \infty$, then $X \in L^1(\mathbb{P})$ and

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\frac{X_n}{n} - X\right|\right] = 0. \tag{7.14}$$

That is to say $\frac{X_n}{n} \to X$ in L^1.

Proof.
We start the proof with observation that the subadditivity assumption from equation (7.11) implies

$$X_n \leq \sum_{j=0}^{n-1} (X_1 \circ T^j + Y_1 \circ T^j),$$

and so

$$\frac{1}{n}\mathbb{E}(X_n^+) \leq \mathbb{E}(X_1^+) + \mathbb{E}(Y_1) < \infty. \tag{7.15}$$

Set

$$A_n = \mathbb{E}(X_n) \in [-\infty, \infty[.$$

The subadditivity implies

$$A_{n+m} \leq A_n + A_m + C$$

where $C = \sup_n \mathbb{E}(Y_n)$. We denote

$$L = \lim_{n\to\infty} \frac{1}{n} A_n = \inf_n \frac{1}{n} A_n \in [-\infty, \infty[.$$

Let

$$\underline{X}(\omega) = \liminf_{n\to\infty} \frac{1}{n} X_n(\omega), \qquad \bar{X}(\omega) = \limsup_{n\to\infty} \frac{1}{n} X_n.$$

The relation (7.15) and Fatou's Lemma imply

$$\mathbb{E}(\underline{X}^+) \leq \mathbb{E}(X_1^+) + \mathbb{E}(Y_1) < \infty.$$

The relation

$$X_n \leq X_1 + X_{n-1} \circ T + Y_{n-1} \circ T$$

implies

$$\underline{X} \leq \underline{X} \circ T, \qquad \bar{X} \leq \bar{X} \circ T,$$

One can then easily show that this implies

$$\underline{X} \circ T^m = \underline{X}, \qquad \bar{X} \circ T^m = \bar{X}. \tag{7.16}$$

We shall require the following result as it plays a key role in the rest of the proof:

Proposition 7.4.

$$\int_\Omega \underline{X} d\mathbb{P} = L. \tag{7.17}$$

(Proof of Proposition 7.4): We shall first establish (7.17) under the assumption that for some $C > 0$ and all n,

$$\frac{1}{n} X_n \geq -C. \tag{7.18}$$

Fatou's Lemma then implies

$$\int_\Omega \underline{X} \mathrm{d}\mathbb{P} \leq L. \tag{7.19}$$

To prove the opposite inequality, fix $\epsilon > 0$. For $k \geq 1$, set

$$E_k = \left\{ \omega \mid \frac{1}{j}(X_j(\omega) + Y_j(\omega)) \leq \underline{X}(\omega) + \epsilon \text{ for some } 1 \leq j \leq k \right\}.$$

This sequence of sets is increasing and $\cup_k E_k = \Omega$. Some of these sets might be empty and let k_0 be the first integer such that $P(E_{k_0}) > 0$. For $k \geq k_0$ we set

$$H_k = (\underline{X} + \epsilon)\chi_{E_k} + (X_1 + Y_1)\chi_{E_k^c},$$

and

$$R_k = \max(H_k, X_1 + Y_1).$$

One easily verifies that

$$\underline{X} + \epsilon \leq H_k. \tag{7.20}$$

We will prove that the following inequality holds for $n \geq k \geq k_0$:

$$X_n \leq \sum_{j=0}^{n-k-1} H_k \circ T^j + \sum_{j=n-k}^{n-1} R_k \circ T^j. \tag{7.21}$$

To prove this inequality, for given ω we defined inductively a sequence of integers

$$0 = m_0 \leq n_1 < m_1 \leq n_2 \leq \cdots$$

as follows. Suppose that m_{j-1} is defined. Let n_j be the least integer bigger or equal then m_{j-1} such that $T^{n_j}(\omega) \in E_k$ (recall Poincare Recurrence Theorem). Let $1 \leq l \leq k$ be such that

$$\frac{1}{l}(X_l(T^{n_j}\omega) + Y_l(T^{n_j}\omega)) \leq \bar{X}(T^{n_j}\omega) + \epsilon.$$

We then set

$$m_j = n_j + l.$$

Let now ℓ be the largest integer such that $m_\ell \leq n$. Iterating the inequality

$$X_n(\omega) \leq X_{n-1}(\omega) + X_1(T\omega) + Y_1(T\omega)$$

we derive

$$X_n(\omega) \leq X_{m_\ell}(\omega) + \sum_{j \in [m_\ell, n[} (X_1(T^j\omega) + Y_1(T^j\omega)).$$

Now,

$$X_{m_\ell}(\omega) \leq X_{n_\ell}(\omega) + X_{m_\ell - n_\ell}(T^{n_\ell}\omega) + Y_{m_\ell - n_\ell}(T^{n_l}\omega),$$

and another iteration gives

$$X_{n_\ell}(\omega) \le X_{m_{\ell-1}}(\omega) + \sum_{j\in[m_{\ell-1},n_l[} (X_1(T^j\omega) + Y_1(T^j\omega)).$$

Continuing this way, we derive

$$X_n(\omega) \le \sum_{j=1}^{\ell} X_{m_j-n_j}(T^{n_j}\omega) + \sum_{j\in S}(X_1(T^j\omega) + Y_1(T^j\omega)),$$

where

$$S = \bigcup_{j=0}^{\ell-1}[m_j, n_{j+1}[\cup [m_\ell, n[.$$

Now, the definition of m_j and relations (7.16), (7.20), imply

$$X_{m_j-n_j}(T^{n_j}\omega) \le (m_j - n_j)(\underline{X}(T^{n_j}\omega) + \epsilon) = \sum_{i\in[m_j,n_j[} (\underline{X}(T^i\omega) + \epsilon) \le \sum_{i\in[n_j,m_j[} H_k(T^\omega),$$

and so

$$\sum_{j=1}^{\ell} X_{m_j-n_j}(T^{n_j}\omega) \le \sum_{j\in D} H_k(T^j\omega)$$

where

$$D = \sum_{j=1}^{\ell}[n_j, m_j[.$$

Hence,

$$X_n(\omega) \le \sum_{j\in D} H_k(T^j\omega) + \sum_{j\in S}(X_1(T^j\omega) + Y_1(T^j\omega)). \tag{7.22}$$

Note that $X_1(T^j\omega) + Y_1(T^j\omega) = H_k(T^j\omega)$ for $j \in \bigcup_{i=0}^{\ell-1}[m_i, n_{i+1}[$. If $n \le n_{\ell+1}$, then $X_1(T^j\omega) + Y_1(T^j\omega) = H_k(T^j\omega)$ also for $j \in [m_\ell, n[$, and (7.21) follows from (7.22). If $n > n_{\ell+1}$, then (7.22) can be written as

$$X_n(\omega) \le \sum_{j=0}^{n_{\ell+1}-1} H_k(T^j\omega) + \sum_{j=n_{\ell+1}}^{n-1} (X_1(T^j\omega) + Y_1(T^j\omega).$$

By construction, $m_{\ell+1} - n_{\ell+1} \le k$, and by the choice of ℓ, $m_{\ell+1} > n$. We can now write (7.22) as

$$X_n(\omega) \le \sum_{j=0}^{n_{\ell+1}-1} H_k(T^j\omega) + \sum_{j=n_{\ell+1}}^{n} (X_1(T^j\omega) + Y_1(T^j\omega))$$

and (7.21) follows. Integrating (7.21) we derive

$$\frac{1}{n}\int_\Omega X_n \mathrm{d}\mathbb{P} \le \frac{n-k-1}{n}\int_\Omega H_k \mathrm{d}\mathbb{P} + \frac{k}{n}\int_\Omega R_k \mathrm{d}\mathbb{P}.$$

Since R_k is integrable,

$$L = \lim_{n\to\infty}\frac{1}{n}\int_\Omega X_n \mathrm{d}\mathbb{P} \le H_k \mathrm{d}\mathbb{P} = \int_{E_k}(\underline{X}+\epsilon)\mathrm{d}\mathbb{P} + \int_{E_k^c}(X_1+Y_1)\mathrm{d}\mathbb{P}.$$

Since $\underline{X}$, X_1+Y_1, are integrable and E_k is increasing sequence of sets satisfying $\cup_k E_k = \Omega$,

$$\lim_{k\to\infty}\int_{E_k}(\underline{X}+\epsilon)\mathrm{d}\mathbb{P} = \int_\Omega(\underline{X}+\epsilon)\mathrm{d}\mathbb{P}, \qquad \lim_{k\to\infty}\int_{E_k^c}(X_1+Y_1)\mathrm{d}\mathbb{P} = 0.$$

Hence,

$$L \le \int_\Omega(\underline{X}+\epsilon)\mathrm{d}\mathbb{P}.$$

Taking $\epsilon \downarrow 0$, we derive

$$L \le \int_\Omega \underline{X}\mathrm{d}\mathbb{P}.$$

This estimate and (7.19) yield (7.17) under the assumption (7.18).

We now use a limiting argument to show that (7.17) holds without the additional assumption (7.18). For $C>0$, let

$$X_n^C = \max(X_n, -nC).$$

This sequence of random variables satisfies all assumptions of Theorem 7.3 (with the same Y_n) and the bound (7.18) holds. Moreover, if

$$\underline{X} = \liminf_{n\to\infty}\frac{1}{n}X_n, \qquad \underline{X}^C = \liminf_{n\to\infty}\frac{1}{n}X_n^C,$$

then $\underline{X}^C = \max(\underline{X}, -C)$. Since $\mathbb{E}(\underline{X}^+) < \infty$, an application of Monotone Convergence Theorem gives

$$\int_\Omega \underline{X}\mathrm{d}\mathbb{P} = \inf_{C>0}\int_\Omega(\underline{X},-C)\mathrm{d}\mathbb{P}.$$

On the other hand,

$$\inf_{C>0}\int_\Omega(\underline{X},-C)\mathrm{d}\mathbb{P} = \inf_{C>0}\int_\Omega \underline{X}_n^C\mathrm{d}\mathbb{P} = \inf_{C>0}\inf_n\frac{1}{n}\int_\Omega X_n^C\mathrm{d}\mathbb{P} = \inf_n\inf_{C>0}\frac{1}{n}\int_\Omega X_n^C\mathrm{d}\mathbb{P}.$$

Monotone Convergence Theorem again gives that

$$\inf_{C>0}\int_\Omega X_n^C\mathrm{d}\mathbb{P} = \int_\Omega X_n\mathrm{d}\mathbb{P}.$$

Hence,

$$L = \inf_n \frac{1}{n} \int_\Omega X_n d\mathbb{P} = \int_\Omega \underline{X} d\mathbb{P}.$$

□

We now continue with the rest of the proof. We must show:

Lemma 7.5. *For any positive integer k,*

$$\limsup_{n\to\infty} \frac{1}{n} X_{kn} = k \limsup_{n\to\infty} \frac{1}{n} X_n.$$

The inequality

$$\limsup_{n\to\infty} \frac{1}{n} X_{kn} \le k \limsup_{n\to\infty} \frac{1}{n} X_n.$$

is obvious. To prove the opposite inequality, write

$$n = km_n + r_n, \qquad 1 \le r_n \le k.$$

Then

$$X_n \le X_{km_n} + X_{r_n} \circ T^{km_n} + Y_{r_n} \circ T^{km_n} \le X_{km_n} + H \circ T^{km_n},$$

where

$$H = \sum_{j=1}^{k} (X_j^+ + Y_j^+).$$

Since H is integrable, Birkhoff's Theorem implies

$$\lim_{n\to\infty} \frac{1}{km_n} H \circ T^{km_n} = 0,$$

Hence,

$$\limsup_{n\to\infty} \frac{1}{n} X_n = \limsup_{n\to\infty} \frac{1}{km_n} X_n \le \frac{1}{k} \limsup_{n\to\infty} \frac{1}{km_n} X_{km_n} \le \frac{1}{k} \limsup_{n\to\infty} \frac{1}{n} X_{kn}.$$

We are now ready to complete the proof of the Theorem 7.3. Suppose that for some $C > 0$ and all n,

$$\frac{1}{n} X_n \ge -C. \tag{7.23}$$

Fix k and set

$$\mathcal{S}_n = -\sum_{j=0}^{n-1} (X_k + Y_k) \circ T^{jk}.$$

$\mathcal{S}_n$ is additive with respect to T^k,

$$\mathcal{S}_{n+m} = \mathcal{S}_n + \mathcal{S}_m \circ T^n,$$

and $\mathcal{S}_1 = -X_k \leq Ck$. Hence, Proposition 7.4 applies to $\mathcal{S}_n$ and

$$\int_\Omega \underline{\mathcal{S}} \mathrm{d}\mathbb{P} \geq \lim_{n\to\infty} \frac{1}{n} \int_\Omega \mathcal{S}_n \mathrm{d}\mathbb{P} = -\int_\Omega (X_k + Y_k) \mathrm{d}\mathbb{P}.$$

On the other hand, subadditivity and the previous Lemma yield

$$-\underline{S} = \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (X_k + Y_k) \circ T^{jk} \geq \limsup_{n\to\infty} \frac{1}{n} X_{kn} = k \limsup_{n\to\infty} \frac{1}{n} X_n = k\bar{X}.$$

Hence,

$$\frac{1}{k} \int_\Omega (X_k + Y_k) \mathrm{d}\mathbb{P} \geq \int_\Omega \bar{X} \mathrm{d}\mathbb{P}.$$

Taking $k \to \infty$, we get

$$L \geq \int_\Omega \bar{X} \mathrm{d}\mathbb{P}.$$

This estimate combined with Proposition 7.4 yields that $\underline{X} = \bar{X}$. This proves Parts (1) and (2) of Theorem 7.3 under the assumption (7.23). The remove this assumption, one argues as in the proof of the Proposition 7.4. If X_n^C, $\underline{X}^C$ and $\bar{X}^C$, are is in the proof of Proposition (7.4), then the above argument gives

$$\max(\underline{X}, -C) = \underline{X}^C = \bar{X}^C = \max(\bar{X}, C).$$

Taking $C \to \infty$ we derive $\underline{X} = \bar{X}$, and this yields Parts (1) and (2) of Theorem 7.3. Finally, to prove Part (3), note that the sequence $|X_n|$ also satisfies the assumptions of Theorem 7.3 and so

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}(|X_n|) = \inf_n \frac{1}{n} \mathbb{E}(|X_n|) = \mathbb{E}(|X|).$$

Part (3) of the theorem then follows from a fairly simple application of dominated convergence theorem. □

7.3 Corollary: Shannon-McMillan-Breiman theorem from almost-subadditivity

In this section, we show how Shannon-McMillan Breiman can be viewed as a direct consequence of Theorem 7.3.

To see this, we start with an ergodic abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ and a partition $\mathcal{C} \in \operatorname{Part} \Omega$. Defining

$$I_n(\omega) = I_{\mathcal{C}^n}(\omega). \tag{7.24}$$

We note that each of the I_n's is a strictly positive function.

If we can show that the sequence I_n satisfies the conditions of Theorem 7.3, then we shall be done, since this would imply that the sequence $\frac{I_n}{n}$ converges (both $\mathbb{P}$-almost surely and in L^1) to a

T-invariant function S. Since all T-invariant functions are constant in an ergodic ADS, we would then have

$$S = \lim_{n\to\infty} \frac{\mathbb{E}[I_n]}{n} \tag{7.25}$$

and in particular,

which is equivalent to Theorem 6.23.

Thus, we choose $n, m \in \mathbb{N}$ arbitrarily, and write

$$I_{n+m} = I_n + I_m \circ T^n + \underbrace{(I_{n+m} - I_n - I_m \circ T^n)}_{\text{(third term)}} \tag{7.26}$$

Now, by taking a closer look at the third term in the above expression,

$$\begin{aligned} I_{n+m} - I_n - I_m \circ T^n &= I_{\mathcal{C}^{n+m}} - I_{\mathcal{C}^n} - I_{\mathcal{C}^m} \circ T^n \\ &= I_{\mathcal{C}^{-n}_{-(n+m)-1}|\mathcal{C}^0_{n-1}} - I_{\mathcal{C}^m} \circ T^n \\ &= \left(I_{\mathcal{C}^0_{m-1}|\mathcal{C}^n_1} - I_{\mathcal{C}^0_{m-1}}\right) \circ T^n \\ &\leq \left(I_{\mathcal{C}^0_{m-1}|\mathcal{C}^n_1} - I_{\mathcal{C}^0_{m-1}}\right)^+ \circ T^n \end{aligned} \tag{7.27}$$

We now define

$$Y_m \equiv \left(\sup_{n\in\mathbb{N}} \left(I_{\mathcal{C}^m|\mathcal{C}^n_1}\right) - I_{\mathcal{C}^m}\right)^+ \tag{7.28}$$

First off, we see by definition that each Y_n is a measurable function with range in R^+. Moreover, by Lemma 6.28, we must have that $Y_m \in \mathcal{L}^1(\mathbb{P})$ for each $m \in \mathbb{N}$. Also, by equations (7.26) and (7.27), we have that

$$I_{n+m} \leq I_n + I_m \circ T^n + Y_m \circ T^n \tag{7.29}$$

Thus, if the sequence of random variables I_n is to satisfy the conditions of Theorem 7.3, it remains to show that $\sup_n \mathbb{E}[Y_n] < \infty$.

We shall first need the following lemma.

Lemma 7.6. *For any $A \in \mathcal{C}^m$, and $x \geq 0$, we have*

$$\mathbb{P}\left(A \cap \left\{\omega \in \Omega : \sup_{n\in\mathbb{N}} I_{\mathcal{C}^m|\mathcal{C}^n_1}(\omega) > x\right\}\right) \leq e^{-x} \tag{7.30}$$

Proof of Lemma 7.6.

Let $C(x) \equiv \left\{\omega \in \Omega : \sup_{n\in\mathbb{N}} I_{\mathcal{C}^m|\mathcal{C}_1^n}(\omega) > x\right\}$, and let $\omega \in A \cap C(x)$. For such an ω, we define

$$n(\omega) \equiv \min\left\{n \in \mathbb{N} : I_{\mathcal{C}^m|\mathcal{C}_1^n}(\omega) \geq x\right\}. \tag{7.31}$$

This quantity exists (is finite) as $\omega \in C(x)$.

We further define $D(\omega)$ to be the unique element of $\mathcal{C}_1^{n(\omega)}$ which contains ω. It is fairly easy to see from the definition of $D(\omega)$ that for $\omega_1 \neq \omega_2$, both elements of $A \cap C(x)$, that either $D(\omega_1) = D(\omega_2)$ or $D(\omega_1) \cap D(\omega_2) = \emptyset$. Also, for $\omega \in A \cap C(x)$. We must have

$$\begin{aligned} I_{\mathcal{C}^m|\mathcal{C}_1^{n(\omega)}}(\omega) = -\ln\left(\frac{\mathbb{P}(A \cap D(\omega))}{\mathbb{P}(D(\omega))}\right) &\geq x \\ \implies \mathbb{P}(A \cap D(\omega)) &\leq \mathbb{P}(D(\omega))e^{-x} \end{aligned} \tag{7.32}$$

The above facts can be neatly summarized in the following manner. We may write the set $A \cap C(x)$ as

$$A \cap C(x) = \bigcup_{i\in I} A \cap D_i \tag{7.33}$$

where I is some countable index set, $D_i \cap D_j = \emptyset$ for any $i \neq j$, and

$$\mathbb{P}(A \cap D_i) \leq \mathbb{P}(D_i)e^{-x} \tag{7.34}$$

Thus,

$$\begin{aligned} \mathbb{P}(A \cap C(x)) &= \mathbb{P}\left(\bigcup_{i\in I} A \cap D_i\right) \\ &= \sum_{i\in I} \mathbb{P}(A \cap D_i) \\ &\leq \sum_{i\in I} e^{-x}\mathbb{P}(D_i) \\ &= e^{-x}\sum_{i\in I} \mathbb{P}(D_i) \\ &= e^{-x}\mathbb{P}\left(\bigcup_{i\in I} D_i\right) \\ &\leq e^{-x} \end{aligned} \tag{7.35}$$

This completes the proof. □

The rest of the proof is fairly straightforward. We shall use Lemma 7.6 to place a bound $\mathbb{P}(Y_m > x)$ for $x \geq 0$.

If we choose $A \in \mathcal{C}^m$, then, using the previous lemma, we have

$$\begin{aligned}\mathbb{P}\left(A \cap \{Y_m > x\}\right) &= \mathbb{P}\left(A \cap \left\{\sup_{n\in\mathbb{N}} I_{\mathcal{C}^m|\mathcal{C}_1^n} - I_{\mathcal{C}^m} > x\right\}\right) \\ &= \mathbb{P}\left(A \cap \left\{\sup_{n\in\mathbb{N}} I_{\mathcal{C}^m|\mathcal{C}_1^n} > x - \ln(\mathbb{P}(A))\right\}\right) \\ &\leq e^{-(x-\ln(\mathbb{P}(A)))} \\ &= \mathbb{P}(A)e^{-x} \end{aligned} \tag{7.36}$$

Summing over all elements of $\mathcal{C}^m$ gives

$$\mathbb{P}(Y_m > x) \leq e^{-x}. \tag{7.37}$$

This gives us the desired integrability, as we see that

$$\mathbb{E}[Y_m] \leq \int_0^\infty e^{-x}\mathrm{d}x = 1 \tag{7.38}$$

for any $m \in \mathbb{N}$, and thus $\sup_m \mathbb{E}[Y_m] < \infty$.

Furthermore, if $\epsilon > 0$ is an arbitrary positive number, we have

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{Y_n}{n} > \epsilon\right) \leq \sum_{n=1}^{\infty} e^{-n\epsilon} < \infty \tag{7.39}$$

Thus, by the Borel-Cantelli lemma (Lemma 2.5), we have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{Y_n}{n} > \epsilon\right) = 0 \tag{7.40}$$

Also, since $\epsilon > 0$ was arbitrarily chosen, we must have

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{Y_n}{n} = 0\right) = 1 \tag{7.41}$$

Or, in other words, $\limsup_{n\to\infty} \frac{Y_n}{n} = 0$ $\mathbb{P}$-almost surely.

Thus, I_n satisfies the conditions of Theorem 7.3 and we have that equation (7.25) is satisfied.

8 Coding and entropy

One central area of study in coding theory is that of data compression. Given a source of data, in the form of character strings with varying lengths, a code will transcribe this data into another alphabet, and one generally tries to do so in such a way as to minimize the average length of the

obtained strings.

Claude Shannon, not surprisingly, was the main pioneer in the development of coding theory with the source coding theorem (i.e. a more specific version of Theorem 6.23, see [1]). Other advances were made by mathematicians and computer scientists over the second half of the twentieth century. Noted works include Huffman's celebrated paper on prefix-free coding [33], as well as an extension by Gray, Ornstein and Dobrushin [34].

In this section, we shall examine the relationship between entropy and compression, looking first at the notions of faithful and prefix-free coding, and then turning our attention to compression rate bounds and how they relate to our already-developed concept of entropy.

We remark that in this section, in contrast to the preceding sections, we shall use a base of 2 for the logarithm in the definition of entropy, as it has more physical significance to binary coding, the main study of compression theory. Nevertheless, the theory developed here has a direct analogue for n-ary coding.

8.1 Faithful codes and prefixes

Let A be a finite set. A shall be called an alphabet. An element $a \in A$ is called a character, or a letter. A^n will refer to the set of all length-n sequences $w = (w_i)_{i=1}^n$ with $w_i \in A$ for each i. Elements of A^n are called words of length n. $A^* = \cup_{n=1}^{\infty} A^n$ is the set of of all A-words of finite length.

If A and B are two alphabets. then a function $C : A \to B^*$ is called a 1-code, or simply a code, from A to B. A is called the source alphabet, while the range, $C(A)$ is called the codebook. An element of the codebook $C(a)$ for some $a \in A$ is called a codeword. In general, one takes A and B to be different sizes, and in many applications, B is taken to be the set $B = \{0, 1\}$. We shall do the same in the rest of this thesis, without losing any significant intuition about the subject.

A desirable feature of a code is called faithfulness. A code $C : A \to B^*$ is called faithful if it is injective. A faithful code ensures that no two letters of the source alphabet A are mapped to the same codeword.

To build upon the notion of coding, one can look at transcribing A-words of length greater than 1. One way of doing so is to use the binary operation of concatenation. Given two elements of A^*, say $u = (u_i)_{i=1}^n$ and $v = (v_i)_{i=1}^m$, the concatenation of u and v, denoted by uv, is an element of A^* of length $n + m$ with the following as characters:

$$(uv)_i = \begin{cases} u_i & 1 \leq i \leq n \\ v_{i-n} & n+1 \leq i \leq n+m \end{cases}$$

A canonical way of obtaining n-codes from 1-codes is through concatenation. We first look at

the case of 2-codes. If C is a 1-code, then one defines the 2-coding $C_2 : A^2 \to B^*$ for any sequence $ab \in A^2$ (a and b are arbitrary elements of A) as

$$C_2(ab) = C(a)C(b)$$

In other words, the concatenation of the letters a and b is encoded by the concatenation of the codewords $C(a)$ and $C(b)$. This method can be recursively applied to obtain n-codes.

While this method of building codes of words of A is intuitive, it unfortunately does not preserve faithfulness. As an example, we consider the alphabet $A = \{a, b, c\}$ and faithful 1-code $C : A \to B^*$ with $C(A) = \{01, 010, 10\}$, $C(a) = 01$, $C(b) = 010$, and $C(c) = 10$. If we look at the 2-code C_2 as defined above, we get that $C_2(ab) = 01010 = C_2(bc)$ and the ability to decode is lost. This problem can be resolved by requiring that an extra property hold on the 1-code C.

A non-empty word $u \in A^*$ is called a prefix of a word $w \in A^*$ if we have that $w = uv$ for some $v \in A^*$. u is called a proper prefix if v is also non-empty. A subset $W \subset A^*$ is called prefix-free if no member of W is a proper prefix of any other element in W.

A faithful coding is called a prefix coding, or, more aptly, a prefix-free coding, if $a = \tilde{a}$ whenever $C(a)$ is a prefix of $C(\tilde{a})$. This is equivalent to saying that the codebook $C(A)$ is prefix-free. This leads us to the first important result.

Lemma 8.1. *Let $C : A \to B^*$ be a faithful, prefix-free code. Then, for each $n \in \mathbb{N}$, the n-code $C_n : A^n \to B^*$ obtained by concatenating codewords of C is also faithful.*

Proof.
We shall look at the case $n = 2$. The rest follows from inductively applying the argument. Suppose, for the sake of contradiction, that the 2-coding C_2 is not faithful. Then there exists $ab, cd \in A^2$, with $ab \neq cd$, such that

$$C_2(ab) = C(a)C(b) = C(c)C(d) = C_2(cd)$$

Now, if $a = c$, then we'd have $C(a) = C(c) \Rightarrow C(b) = C(d)$ and by the faithfulness of C, we'd have that $a = c$ and $b = d$ and then we'd be done. Thus, we may assume that $a \neq c$. By the faithfulness of C, we must have that $C(a) \neq C(c)$. The only way this can be true, however, and still have that $C(a)C(b) = C(c)C(d)$, is if $C(a)$ is a prefix of $C(c)$ or vice-versa. This contradicts the prefix-free property of C. □

A more enlightening way of seeing why prefix-free concatenation code must preserve faithfulness is simply to make sense of the algorithm necessary to decode a codeword from an n-coding C_n obtained from a prefix-free code C.

As an example, suppose that $A = \{a, b, c\}$ and $C : A \to B^*$ is a faithful, prefix-free code, letting $C(a) = 010$, $C(b) = 1001$, and $C(c) = 1010$.

We now look at the word in B^* given by

$$w = 10100101001$$

To proceed with decoding, we start at the left of the sequence w, and start to test the prefixes of w (i.e. 1, 10, 101, etc...) against our codebook $C(A)$ until a valid codeword appears. In this case, the first is $1010 = C(c)$. Since $C(A)$ is prefix-free, no other letter could have produced the beginning of the sequence w, so we know that the first letter of our code is c. We then proceed to look at the sequence w without the first four letters, i.e. $w' = 0101001$. Repeating the process, we find that the next codeword is $010 = C(a)$, and then finally the last letter is $1001 = C(b)$, leading to the decoding of the word $cab \in A^3$.

The algorithm for decoding a word $w \in B^*$ (assuming it is a valid codeword, for some n-code concatenation of C) can be summarized as follows:

1. Start testing the prefixes of w in order until one of them appears in the codebook, $C(A)$, call it c_1.
2. c_1 corresponds to the first letter of the decoded word. There can be no ambiguity since the prefix-free property guarantees us that this prefix of w cannot be the prefix of another codeword in $C(A)$
3. Repeat the process on w, starting after the decoded prefix to obtain codes c_2, c_3... Proceed until the word is completely decoded.

8.2 Binary-tree representations and Kraft's inequality

A prefix-free set $W \subset B^*$ has the property that it can be effectively represented by a directed binary tree, which we shall label $T(W)$ with vertex set $V(W)$. It has the following rules:

1. Each vertex $v \in V(W)$ represents a prefix for codewords in W, except for the parent vertex v_0, which is just the empty word.
2. If a vertex v may be represented as $v = ub$ with u being another vertex, and b not being empty, then there is a unique directed path from u to v.

With these conditions, W happens to be the leaves of the tree, or in other words, the subset of $V(W)$ of vertices which do not have edges directed away from them. As an example, we look at the prefix-free set $W = \{00, 100, 101, 0100, 0101\}$. Included below is a figure of the binary-tree representation of W.

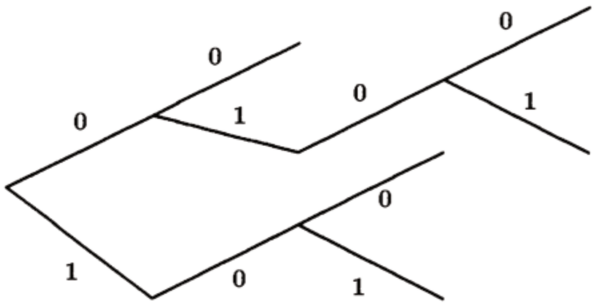

Figure 2: Binary-tree representation of set $W = \{00, 100, 101, 0100, 0101\}$. The edges represent the addition of a binary digit necessary to obtain the next vertex. Figure obtained from page 72 of [35]

One can see that the number of leaves at a given depth of the tree is bounded not only by the number of choices of 0's and 1's but also by the number of leaves at greater depths. This realization and a fairly simple combinatorial argument gives a result known as Kraft's inequality.

Lemma 8.2 (Kraft's inequality)**.** *Let $W \subset B^*$ be a finite prefix-free set. Then,*

$$\sum_{w \in W} 2^{-\ell(w)} \leq 1 \tag{8.1}$$

where $\ell : B^* \to \mathbb{N}$ represents the length of a sequence in B^*.

Proof.
Let m be the Lebesgue measure on the unit interval $[0, 1]$, and for $w \in W$ let I_w represent the closed sub-interval of length $2^{-\ell(w)}$ where the left endpoint has the dyadic expansion of w. Alternatively stated,

$$\inf I_w = \sum_{i=1}^{\ell(w)} w_i \left(\frac{1}{2^i}\right).$$

For example, if $w = 101$ then the left endpoint of I_w would be $\frac{1}{2} + \frac{1}{2^3} = \frac{5}{8}$ so that $I_w = \left[\frac{5}{8}, \frac{3}{4}\right]$.

An important fact is that each of these intervals is disjoint, i.e. $I_u \cap I_v = \emptyset$ for $u, v \in W$, and $u \neq v$. To see why, we note that if this were not true, then there would be $u, v \in W$ with $u \neq v$ such that $\inf I_v \in I_u$. since the interval I_u has length $2^{-\ell(u)}$, this would imply that the dyadic expansion of $\inf I_v$ would have the same first terms as the dyadic expansion of $\inf I_u$, but that would imply that u is a proper prefix of v, a contradiction.

Thus, we have a collection $\{I_w : w \in W\}$ of disjoint intervals, all contained in $[0, 1]$. Thus,

$$m\left(\bigcup_{w \in W} I_w\right) = \sum_{w \in W} m(I_w) = \sum_{w \in W} 2^{-\ell(w)} \leq m([0, 1]) = 1$$

□

Kraft's inequality shall play a vital role in establishing a bound on code-length, as will its important converse:

Lemma 8.3. *Let $1 \leq l_1 \leq l_2 \leq \ldots \leq l_n$ be a non-decreasing sequence of integers satisfying the Kraft inequality. That means*

$$\sum_{i=1}^{n} 2^{-l_i} \leq 1$$

Then, there exists a prefix-free subset of B^*, $W = \{w_i \in B^* : 1 \leq i \leq n\}$, such that $\ell(w_i) = l_i$.

This lemma will not be proved directly here, but we note that it is quite obvious given Kraft's inequality, and that the above proof of Lemma 8.2 can be easily reordered to prove Lemma 8.3. An alternative proof can be found in [35]

8.3 Entropy and average code length

Up to now, we have not discussed any properties one can assign to the source alphabet A in a given code. In general, if one considers realistic cases, we may find that certain characters in an alphabet are used more often then others. Intuitively, by assigning shorter codewords to the more frequent characters in A, one may hope to obtain shorter codes on average. In this chapter, we shall properly formalize this concept.

Let us suppose that we have a probability distribution on A, some finite alphabet. By this, we mean that there is a function $\mu : A \to [0, 1]$ with $\sum_{a \in A} \mu(a) = 1$. For each $a \in A$, $\mu(a)$ can be seen as the probability of obtaining a. Of course, without a notion of valid words in A^*, this doesn't really have any meaning, but for now, this will suffice as a good approximation of more complex and meaningful scenarios.

We may define what is known as the entropy of this distribution, $H(\mu)$, by

$$H(\mu) = -\sum_{a \in A} \mu(a) \log \mu(a) \tag{8.2}$$

where the base of the logarithm is assumed to be 2, as stated in the introduction of this section.

For a given code $C : A \to B^*$, one may define the code-length function, $\mathcal{L}_C : A \to \mathbb{N}$, given by

$$\mathcal{L}_C(a) = \ell(C(a))$$

In other words, the code-length function assigns to each $a \in A$ the length of its associated

codeword in $C(A)$. If C is prefix-free, then by Lemma 8.2, we have

$$\sum_{a\in A} 2^{-\mathcal{L}_C(a)} \leq 1$$

If our source alphabet A has a probability distribution μ associated with it, one may further define the average, or expected code-length:

$$\mathbb{E}_\mu\left[\mathcal{L}_C\right] = \sum_{a\in A} \mathcal{L}_C(a)\mu(a) \tag{8.3}$$

As stated previously, one main goal of coding theory is to try and find effective ways of minimizing this quantity. The following theorem provides an important bound on the expected code-length of prefix codes, and also shows a way to obtain codes with expected lengths very close to this bound.

Theorem 8.4. *Let A be a finite alphabet equipped with probability distribution μ. Then,*

(i) $H(\mu) \leq \mathbb{E}\left[\mathcal{L}_C\right]$ for all prefix-free codes $C : A \to B^*$.

(ii) There exists a prefix-free code $C : A \to B^*$ such that $\mathbb{E}\left[\mathcal{L}_C\right] < H(\mu) + 1$

Before we prove this theorem, we present a necessary lemma.

Lemma 8.5. *let $p = (p_i)_{i=1}^n$ be a probability vector and let $q = (q_i)_{i=1}^n$ be a sub-probability vector. This means $p_i, q_i \geq 0$ for all i and that $\sum_{i=1}^n q_i \leq \sum_{i=1}^n p_i = 1$. Then we have*

$$\sum_{i=1}^n p_i \log \frac{p_i}{q_i} \geq 0 \tag{8.4}$$

with equality if and only if $p_i = q_i$ for each i.

Proof of Lemma 8.5.
the function $f(x) = -\ln x$ is strictly convex, such that $-\ln x \geq 1 - x$, with equality if and only if $x = 1$. With this in mind we consider the following sum:

$$\begin{aligned}\sum_{i=1}^n p_i \ln \frac{p_i}{q_i} &= -\sum_{i=1}^n p_i \ln \frac{q_i}{p_i} \geq \sum_{i=1}^n p_i \left(1 - \frac{q_i}{p_i}\right) \\ &= \sum_{i=1}^n p_i - \sum_{i=1}^n q_i \geq 0\end{aligned}$$

Since $\log_b x = \frac{\ln x}{\ln b}$, the same inequality is true by replacing ln with log with any base. □

With this lemma, we are now capable of proving our main theorem.

Proof of Theorem 8.4.
We first let $C : A \to B^*$ be some prefix-free code and examine its expected code-length.

$$\begin{aligned}\mathbb{E}_\mu[\mathcal{L}_C] &= \sum_{a\in A} \mathcal{L}_C(a)\mu(a) = \sum_{a\in A} \log 2^{\mathcal{L}_C(a)}\mu(a) \\ &= \sum_{a\in A} \mu(a)\log\frac{\mu(a)}{2^{-\mathcal{L}_C(a)}} - \sum_{a\in A}\mu(a)\log\mu(a) \\ &= \sum_{a\in A} \mu(a)\log\frac{\mu(a)}{2^{-\mathcal{L}_C(a)}} + H(\mu)\end{aligned}$$

By Kraft's inequality, $\sum_{a\in A} 2^{-\mathcal{L}_C(a)} \leq 1$ and thus, by Lemma 8.5, the first term in the last line above is positive, therefore proving the first part of the theorem.

To show the second part of the theorem, we define the following function $\mathcal{L} : A \to \mathbb{N}$ so that $\mathcal{L}(a) = \lceil -\log\mu(a)\rceil$ where $\lceil x\rceil$ is the smallest integer which is greater than or equal to x. With this definition we can see that

$$\sum_{a\in A} 2^{-\mathcal{L}(a)} \leq \sum_{a\in A} 2^{\log\mu(a)} = \sum \mu(a) = 1$$

By Lemma 8.3, we may construct a prefix-free code $C : A \to B^*$ such that $\mathcal{L}_C(a) = \mathcal{L}(a)$ for each $a \in A$. Noting that $\mathcal{L}(a) < 1 - \log\mu(a)$, we proceed to examine the expected code-length of C.

$$\begin{aligned}\mathbb{E}_\mu[\mathcal{L}_C] &= \sum_{a\in A}\mathcal{L}(a)\mu(a) \\ &< \sum_{a\in A}(1-\log\mu(a))\mu(a) = \sum_{a\in A}\mu(a) - \sum_{a\in A}\mu(a)\log\mu(a) \\ &= 1 + H(\mu)\end{aligned}$$

This proves the second part of the theorem. □

It is important to note the details of how the prefix-free code was chosen so as to obtain the right bound. for characters $a \in A$ with larger frequency $\mu(a)$, the lengths of the codewords $\mathcal{L}(a)$ were chosen to be shorter. If each of these frequencies happened to be exact powers of $\frac{1}{2}$, then it would be possible to choose a code satisfying $\mathbb{E}_\mu[\mathcal{L}_C] = H(\mu)$.

8.4 n-codes and asymptotic compression rates

We now begin to consider codes n-length words formed with our alphabet A. Of course, since A^n is just a finite set (A is assumed to be finite), we see that n-codes $C_n : A^n \to B^*$ are conceptually no different from simple codes from A to B^*.

More specifically, if μ_n is a probability distribution on A^n, then, as before, one defines the entropy of the distribution, $H(\mu_n)$, as follows

$$H(\mu_n) = -\sum_{x \in A^n} \mu_n(x) \log(\mu_n(x))$$

also, for any code $C_n : A^n \to B^*$, the code-length function, $\mathcal{L}_{C_n}$, is defined as before.

We now consider a probability measure μ on A^∞, the set of all sequences $x = (x_i)_{i=1}^\infty$ with each x_i in A.

To clarify what we mean by this, we shall introduce some notation. If $x = (x_i)_{i=1}^\infty \in A^\infty$, then x_n^m is the word of length $(m-n)+1$ representing the n-th to m-th elements of the sequence x. Also, for $x \in A^\infty$, the 'n to m cylinder set of x', denoted by $[x_n^m]$, is defined by

$$[x_n^m] = \{y \in A^\infty : y_n^m = x_n^m\}$$

With these definitions, we let $\mathcal{F}$ be the σ-algebra generated by all finite intersections of cylinder sets. Thus, $(A^\infty, \mathcal{F})$ is a measurable space and the notion of a probability measure μ on $(A^\infty, \mathcal{F})$ is well-defined.

The n-th marginal of μ, denoted by μ_n, is a probability distribution on A^n, defined in the following manner: If $a = a_1^n$ is an element of A^n, then we define its probability by

$$\mu_n(a) = \mu([a_1^n])$$

one can easily verify that the additivity and normalization properties of a probability distribution are satisfied.

It is reasonable to ask how one may define the notion of entropy of a measure μ on A^∞. To do so, we first define the per-symbol entropy of its marginal, μ_n,

$$h_n(\mu_n) = \frac{1}{n} H(\mu_n) \tag{8.5}$$

The entropy of the measure, $h(\mu)$ is then defined with an appropriate limiting procedure.

$$h(\mu) = \limsup_{n \to \infty} h_n(\mu_n) \tag{8.6}$$

It turns out if μ is a stationary process, i.e. measure-invariant with respect to the left-shift operator, then equation (8.6) is simply the Kolmogorov-Sinai entropy (see Section 6.6) of the associated dynamical system.

Now, suppose we have a measure μ on A^∞, and an n-code $C_n : A^n \to B^*$, then another quantity

of interest is that of the expected per-symbol code-length,

$$\frac{1}{n}\mathbb{E}_{\mu_n}[\mathcal{L}_{C_n}] = \frac{1}{n}\sum_{a\in A^n}\mathcal{L}_{C_n}(a)\mu_n(a) \tag{8.7}$$

Equations (8.5) and (8.7), along with Theorem 8.4 imply that for a given probability distribution μ_n on A^n, all prefix-free n-codes C_n obey

$$h_n(\mu_n) \leq \frac{1}{n}\mathbb{E}_{\mu_n}[\mathcal{L}_{C_n}] \tag{8.8}$$

They also imply the existence of a prefix-free n-code $\tilde{C_n}$ such that

$$\frac{1}{n}\mathbb{E}_{\mu_n}[\mathcal{L}_{C_n}] \leq h_n(\mu_n) + \frac{1}{n} \tag{8.9}$$

We may now look at sequences of prefix codes and ask how they might behave as the length of source words grows. Let $\{C_n\}_{n=1}^{\infty}$ be a sequence of codes with C_n being a prefix code from A^n to B^*. Given a probability measure μ on A^∞, we may define the (asymptotic) compression rate of the sequence, $R_\mu(\{C_n\})$ by

$$R_\mu(\{C_n\}) = \limsup_{n\to\infty}\frac{1}{n}E_{\mu_n}[\mathcal{L}_{C_n}] \tag{8.10}$$

which is an approximate measure of the ability of our code to 'shorten' words from our source alphabet as their length gets arbitrarily long. With this definition, we easily obtain the following theorem from equations (8.8) and (8.9).

Theorem 8.6. *For a given probability measure μ on $(A^\infty, \mathcal{F})$ the following are true.*

(a) There exist prefix sequences $\{C_n\}$ such that $R_\mu(\{C_n\}) = h(\mu)$

(b) There is no prefix sequence $\{\tilde{C_n}\}$ such that $R_\mu(\{\tilde{C_n}\}) < h(\mu)$

This result helps illuminate the direct connection between the notion of code-length and entropy. In Section 8.3 we showed how the entropy of a given distribution on A acted as a lower bound for the average code-length of prefix codes and showed how one can construct a prefix code whose average code-length is 'nearly optimal', within one of the entropy. By focusing our attention on the asymptotic results, Theorem 8.6 gives us a tight bound, solidifying the notion that entropy is indeed a measure of compressibility of information.

References

[1] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal, The*, 27(3):379–423, 1948.

[2] Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica*, 13(1):A3–A270, 1890.

[3] George D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.

[4] R. Clausius and W.R. Browne. *The Mechanical Theory of Heat*. Macmillan and Company, 1879.

[5] Ludwig Boltzmann. *Vorlesungen über Gastheorie*, volume 1. JA Barth, 1896.

[6] Otto M. Nikodym. Sur une généralisation des intégrales de M.J. Radon. *Fundamenta Mathematicae*, 15(1):131–179, 1930.

[7] M. Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27(1):247–271, 1909.

[8] Francesco Paolo Cantelli. Sulla probabilita come limite della frequenza. *Atti Accad. Naz. Lincei*, 26(1):39–45, 1917.

[9] V.K. Rohatgi and A.K.M.E. Saleh. *An Introduction to Probability and Statistics*. Wiley Series in Probability and Statistics. Wiley, 2011.

[10] C.W. Burrill. *Measure, integration, and probability*. McGraw-Hill, 1972.

[11] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1997.

[12] Nicolas Krylov and Nicolas Bogoliubov. La théorie generale de la mesure dans son application a l'étude des systèmes dynamiques de la mécanique non linéaire. *Annals of Mathematics*, 38(1):pp. 65–113, 1937.

[13] Massimiliano Gubinelli. Topological preliminaries.

[14] W. Rudin. *Functional analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991.

[15] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley, 1999.

[16] Hermann Weyl. ber die gleichverteilung von zahlen mod. eins. *Mathematische Annalen*, 77(3):313–352, 1916.

[17] J.L.W.V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.

[18] W. Rudin. *Real and complex analysis*. Mathematics series. McGraw-Hill, 1987.

[19] Alfred Renyi. On measures of entropy and information. In *Fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 547–561, 1961.

[20] Dmitrii Konstantinovich Faddeev. On the concept of entropy of a finite probabilistic scheme. *Uspekhi Matematicheskikh Nauk*, 11(1):227–231, 1956.

[21] Sherry Chu. Lectures on entropy: Project solutions. 2013.

[22] Andrei Nikolaevich Kolmogorov. A new metric invariant of transient dynamical systems and automorphisms in lebesgue spaces. In *Dokl. Akad. Nauk. SSSR*, volume 119, pages 861–864, 1958.

[23] Andrei Nikolaevitch Kolmogorov. Entropy per unit time as a metric invariant of automorphisms. In *Dokl. Akad. Nauk SSSR*, volume 124, pages 754–755, 1959.

[24] Yakov Sinai. On the notion of entropy of a dynamical system. In *Dokl Akad Nauk SSSR*, volume 124, pages 768–771, 1959.

[25] M. Fekete. ber die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Mathematische Zeitschrift*, 17(1):228–249, 1923.

[26] Leo Breiman. The individual ergodic theorem of information theory. *The Annals of Mathematical Statistics*, 28(3):809–811, 09 1957.

[27] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2013.

[28] John F.C. Kingman. The ergodic theory of subadditive stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 499–510, 1968.

[29] Yves Derriennic. Un théorème ergodique presque sous-additif. *The Annals of Probability*, 11(3):669–677, 08 1983.

[30] Artur Avila and Jairo Bochi. On the subadditive ergodic theorem. *preprint*, 2009.

[31] Thomas M. Liggett. An improved subadditive ergodic theorem. *Ann. Probab.*, 13(4):1279–1285, 11 1985.

[32] Klaus Schurger. Almost subadditive extensions of kingman's ergodic theorem. *Ann. Probab.*, 19(4):1575–1586, 10 1991.

[33] D.A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, Sept 1952.

[34] R. M. Gray, D. S. Ornstein, and R. L. Dobrushin. Block synchronization, sliding-block coding, invulnerable sources and zero error codes for discrete noisy channels. *Ann. Probab.*, 8(4):639–674, 08 1980.

[35] P.C. Shields. *The Ergodic Theory of Discrete Sample Paths*. Graduate studies in mathematics. American Mathematical Society, 1996.