A FORAY INTO ERGODIC THEORY

AHMAD ALKADRI

ABSTRACT. Following the introduction to Lebesgue integration and L^p spaces from Analysis III (Math 545 at the University of Calgary), an introduction to Ergodic Theory is given. The central notion of a measure-preserving map is discussed at length, and a systematic pattern between various ideas in Ergodicity and Mixing are laid out. The end-goal of this paper is a proof of the Ergodic Theorem, which makes mathematically precise the idea that a time-average should be equal to its ensemble average. A complete proof of the Birkhoff (or pointwise) Ergodic Theorem is given, and physical applications of Ergodic Theory are briefly introduced.

CONTENTS

Date: December 22, 2018.

1. Introduction

This paper assumes a working knowledge of the basics of measure theory, including the notion of a σ-algebra, measures, and Lebesgue integration. Section 2 contains a number of definitions and theorems that should be good indicators of whether or not the reader will be able to successfully interpret the proofs laid out in the sections to follow. Basic convergence-related theorems like Fatou's Lemma are also assumed.

Section 3 introduces the fundamental idea of a measure-preserving map. Essentially, maps that do not alter the measure of a set of a σ-algebra are central are discussed, and techniques for checking wheter a transformation is a map of this kind is given. Section 4 then takes this idea, and expands it to the L^p function spaces. The result here is essentially that every measure-preserving map induces an isometry of L^p spaces. The Theorem proved at the very end of this section is crucial to the final proof of the Ergodic Theorem.

Section 5 lays out the second key definition of ergodicity, and discusses many equivalent conditions for ergodicity. The subtleties in proving these equivalent conditions are helpful to get a grasp of the reasons for chosing the sets that we do during the proof of the Ergodic Theorem. Section 6 introduces a concept that is stronger than the idea of ergodicity, but is entirely optional when it comes to the proof of the Ergodic Theorem. It is included here as both as an introduction to this important topic in case the reader wants to begin a study of it, and also to illustrate how the concepts of Section 5 can be used to prove stronger theorems. Finally, in Section 7 we prove a number of results that lead up to the (Birkhoff or pointwise) Ergodic Theorem, including the important Maximal Ergodic Theorem.

2. Preliminary Definitions and Theorems

This section contains a number of definitions already seen in class, but included here once more as a refresher. It also enables this document to be read independently for the first-time reader.

We begin with the notion of an algebra.

Definition 2.1. Let Ω be a set, and $\mathcal{J}$ a family of subsets of Ω. Then $\mathcal{J}$ is a *semi-algebra of subsets of* Ω (or *semi-algebra* for short) if:

(1) $\emptyset \in \mathcal{J}$
(2) $A, B \in \mathcal{J} \Rightarrow A \cap B \in \mathcal{J}$
(3) $A \in \mathcal{J} \Rightarrow \Omega \setminus A = \bigcup_{i=1}^{n} E_i$ where $E_i \in \mathcal{J}$ and $E_1, E_2, \ldots, E_n$ are pairwise disjoint subsets of Ω.

Definition 2.2. Let Ω be a set and $\mathcal{F}$ a collection of subsets of Ω. Then Ω is an *algebra of subsets of* Ω (or just *algebra* for short) if:

(1) $\emptyset \in \mathcal{F}$
(2) $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$
(3) If $A_1, A_2, \ldots, A_n \in \mathcal{F}$ then $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$ (i.e., $\mathcal{F}$ is closed under finite unions).

Definition 2.3. Let Ω be a set and $\mathcal{F}$ a collection of subsets of Ω. Then Ω is a *σ-algebra of subsets of* Ω (or just *σ-algebra* for short) if:

(1) $\emptyset \in \mathcal{F}$
(2) $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$
(3) If $A_1, A_2, \dots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (i.e., $\mathcal{F}$ is closed under countable unions).

Theorem 2.1. *Let Ω be a set, and $\mathcal{F}$ an algebra of subsets of Ω. The $\mathcal{F}$ is a semi-algebra*

Proof. $\emptyset \in \mathcal{F}$ is immediate. Let $A, B \in \mathcal{F}$. Then $A \cap B = (A^c \cup B^c)^c \in \mathcal{F}$ since $\mathcal{F}$ is closed under complementation and finite unions. If $A \in \mathcal{F}$, then $A^c = A^c \cup \emptyset$ is a finite disjoint union of pairwise disjoint subsetes of Ω. Hence, $\mathcal{F}$ is a semi-algebra. $\square$

Although many of the theorems in the next chapters can be phrased more generally in terms of semi-algebras, we will use algebras in order to be more consistent with the presentation of the material in class. Following the definition of a σ-algebra we can define a measure.

Definition 2.4. Let $\mathcal{F}$ be a σ-algebra. A *measure* on $\mathcal{F}$ is a non-negative, extended real-valued function that is countably additive on a disjoint collection of sets in $\mathcal{F}$. That is, $\mu : \mathcal{F} \to \overline{\mathbb{R}}$ is a *measure* if

(1) $\mu(A) \geq 0 \ \forall A \in \mathcal{F}$
(2) If $A_1, A_2, \ldots$ are disjoint sets in $\mathcal{F}$, then

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

For a construction of the Lebesgue measure and the Borel sets, see [1]. The Borel sets will be denoted here by $\mathcal{B}$.

Definition 2.5. Let μ be a measure on a σ-algebra $\mathcal{F}$ of subsets of a set Ω. Then μ is *finite* if $\mu(A) < \infty \ \forall A \in \mathcal{F}$. Observe that for measures, it is sufficient to have $\mu(\Omega) < \infty$ for μ to be finite. μ is *σ-finite* if $\mathcal{F}$ can be written as a countable union of sets of finite measure under μ.

Definition 2.6. If Ω is a set and $\mathcal{F}$ is a σ-algebra of subsets of Ω, then $(\Omega, \mathcal{F})$ is a *measurable space*. If μ is a measure on $\mathcal{F}$, then $(\Omega, \mathcal{F}, \mu)$ is a *measure space*.

We quickly recap on the definition of a couple other measure theory concepts, including the integral and an L^p space.

Definition 2.7. Let Ω_1 and Ω_2 be sets with corresponding σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$. Then $h : \Omega_1 \to \Omega_2$ is *measurable* if $h^{-1}(B) \in \mathcal{F}_1 \ \forall B \in \mathcal{F}_2$. If $\Omega_2 = \mathbb{R}$ (or $\overline{\mathbb{R}}$) and $\mathcal{F} = \mathcal{B}$, then h is said to be *Borel measurable*. A complex-valued function between measure spaces h is said to *Borel measurable* if its real and imaginary parts are Borel measurable.

As a note to the above,

Definition 2.8. If Ω is a set and $A \subseteq \Omega$ then χ_A, called the *indicator function on* A, is the function $\chi : \Omega \to \mathbb{R}$ satisfying $\chi_A(\omega) = 1$ if $\omega \in A$, and $\chi_A(\omega) = 0$ otherwise.

Definition 2.9. Let $(\Omega, \mathcal{F})$ be a measurable space. Then h is *simple* if it can be written as a finite linear combination of indicator functions on disjoint sets in $\mathcal{F}$. That is, if $h = \sum_{i=1}^{r} x_i \chi_{A_i}$ for disjoint sets $A_1, A_2, \ldots, A_r$ in $\mathcal{F}$ and $x_1, x_2, \ldots, x_r \in \mathbb{R}$ then h is simple.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space where μ is the Lebesgue measure.

Definition 2.10. If $h = \sum_{i=1}^{r} x_i \chi_{A_i}$ (i.e., h is simple) then the *Lebesgue integral* of h on Ω is denoted and defined by

$$\int_{\Omega} h d\mu = \sum_{i=1}^{r} x_i \mu(A_i)$$

Definition 2.11. Let h be a non-negative, Borel measurable function. Then the *Lebesgue integral* of h is denoted and defined by

$$\int_{\Omega} h d\mu = \sup\left\{\int_{\Omega} s d\mu \mid s \text{ is simple and } 0 \leq s \leq h\right\}$$

Definition 2.12. Let h be a Borel measurable function. Then the *Lebesgue integral* of h is denoted and defined by

$$\int_{\Omega} h d\mu = \int_{\Omega} h^+ d\mu - \int_{\Omega} h^- d\mu$$

(so long as this is not of the form $\infty - \infty$).

Definition 2.13. Let $p > 0$. If f is Borel measurable, define the *p-norm*

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

Definition 2.14. The space $L^p = L^p(\Omega, \mathcal{F}, \mu)$ denotes the collection of all Borel measurable functions with finite p-norm.

Definition 2.15. The space $L^0 = L^0(\Omega, \mathcal{F}, \mu)$ denotes the space of equivalence classes of Borel measurable functions that are equal to each other a.e..

When it comes to Ergodic theory, being able to approximate with the symmetric difference will be useful. For this, we recall the following theorem.

Theorem 2.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathcal{F}_0$ an algebra of subsets of Ω such that $\mathcal{F} = \sigma(\mathcal{F}_0)$. Suppose that μ is σ-finite when restricted to $\mathcal{F}_0$. Then $\forall \epsilon > 0$ and $A \in \mathcal{F}, \exists B \in \mathcal{F}_0 \ni$*

$$\mu(A \Delta B) < \epsilon$$

Proof. Omitted. See [1] for a full development. □

From functional analysis, we will need to use the norm of a linear operator, which may be expressed in a variety of equivalent ways. As a refresher, this so-called "sup norm" is given in the next definition

Definition 2.16. If A is a linear operator on a normed linear space L with norm $|| \; ||$, then the *norm* of A is denoted and defined by

$$||A|| = \sup\{||Ax|| \mid x \in L, ||x|| \leq 1\}$$

3. Measure Preserving Maps

In class, we have gone through an introduction to measure spaces $(\Omega, \mathcal{F}, \mu)$. For Ergodic Theory, it will be more convenient to work in *probability spaces*.

Definition 3.1. Let Ω be a set, $\mathcal{F}$ a σ-algebra of subsets of Ω, and P a probability measure on $\mathcal{F}$. That is, P is a measure on $\mathcal{F}$ that satisfies $P(\Omega) = 1$. Then $(\Omega, \mathcal{F}, P)$ is a *probability space*.

We begin by defining a central object of interest in Ergodic Theory.

Definition 3.2. Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be measure spaces, and $T : (\Omega_1, \mathcal{F}_1, \mu_1) \to (\Omega_2, \mathcal{F}_2, \mu_2)$ a measurable map. Then T is *measure-preserving* if $\mu_1(T^{-1}(B_2)) = \mu_2(B_2) \; \forall B_2 \in \mathcal{F}_2$. T is an *invertible measure-preserving transformation* if it is bijective and measure-preserving with T^{-1} measure-preserving.

A function between measure spaces T that is measure-preserving will also be referred to as a measure-preserving map or a measure-preserving transformation. In Ergodic Theory, the measure-preserving maps that are of interest are the ones defined from a measure space $(\Omega_1, \mathcal{F}_1, \mu_1)$ onto itself. In particular, we will be examining how the repeated application of T, denoted by $T^n := \underbrace{T \circ T \circ \cdots \circ T}_{n\text{-times}}$, behaves on a probability space $(X, \mathcal{F}, P)$.
The following theorem will be useful for checking whether a map is measure-preserving.

Theorem 3.1. *Suppose that $(X_1, \mathcal{F}_1, P_1)$ and $(X_2, \mathcal{F}_2, P_2)$ are probability spaces and $T : (X_1, \mathcal{F}_1, P_1) \to (X_2, \mathcal{F}_2, P_2)$ a map. Let $\mathcal{J}_2$ be an algebra of subsets of X_2. Suppose that $\sigma(\mathcal{J}_2) = X_2$. If $\forall A_2 \in \mathcal{J}_2$ we have $T^{-1}(A_2) \in \mathcal{F}_1$ and $P_1(T^{-1}(A_2)) = P_2(A_2)$ then T is measure-preserving.*

Proof. We use an argument that's reminiscent of the *good sets principle* to show the result. Let

$$\mathcal{C}_2 = \{A \in X_2 \mid T^{-1}(A) \in \mathcal{F}_1 \text{ and } P_1(T^{-1}(A)) = P_2(A)\}$$

Observe that $\mathcal{J}_2 \subseteq \mathcal{F}_2 \subseteq X_2$. We want to show that $\mathcal{C}_2$ is a σ-algebra. Observe that $T^{-1}(X_2) = X_1 \Rightarrow T^{-1}(X_2) \in \mathcal{F}_1$ and $P_1(T^{-1}(X_2)) = P_1(X_1) = 1 = P_2(X_2)$ so that $X_2 \in \mathcal{C}_2$. If $A_2 \in \mathcal{C}_2$, then $T^{-1}(A_2^c) = T^{-1}(A_2)^c \in \mathcal{F}_1$ since $T^{-1}(A_2) \in \mathcal{F}_1$. Now if $A_1, A_2, \ldots$ are sets in $\mathcal{C}_2$ then $T^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty} T^{-1}(A_i) \in \mathcal{F}_1$ since each $T^{-1}A_i \in \mathcal{F}_1$ and $\mathcal{F}_1$ is closed under countable unions.

Since $\mathcal{C}_2$ is a σ-algebra it follows that $\sigma(\mathcal{J}_2) = X_2 = \mathcal{C}_2$ so that T is measure-preserving by definition. □

4. Relation to L^p Spaces

With the knowledge that L^p spaces are also Banach spaces, we want to connect measure-preserving transformations on a measure space $(\Omega, \mathcal{F}, \mu)$ onto itself to isometries of the associated function space $L^p(\Omega, \mathcal{F}, \mu)$. The next definition considers the natural choice for the operator induced by a measure-preserving map between probability spaces.

Definition 4.1. Let $(X_1, \mathcal{F}_1, P_1)$ and $(X_2, \mathcal{F}_2, P_2)$ be probability spaces. Suppose that $T : X_1 \to X_2$ is a measure-preserving map. Then the *operature induced by* T is given by $U_T : L^0(X_2, \mathcal{F}_2, P_2) \to (X_1, \mathcal{F}_1, P_1)$ and

$$(U_T f)(x) = f(T(x)) \; \forall f \in L^0(X_1, \mathcal{F}_1, P_1), x \in X$$

Notice that if $f, g \in L^0(X_2, \mathcal{F}_2, P_2)$ and $a, b \in \mathbb{R}$ then $U_T(af + bg) = (af + bg)(T(x)) = af(T(x)) + bg(T(x))$ is immediate (that is, U_T is a linear operator). Furthermore, it is also clear that an element of $L^0(X_2, \mathcal{F}_2, P_2)$ is always mapped to $(X_1, \mathcal{F}_1, P_1)$ so that $U_T(L^0(X_2, \mathcal{F}_2, P_2)) \subseteq (X_1, \mathcal{F}_1, P_1)$. When it comes to L^p spaces, it is not so obvious that the function will still have finite p-norm in the target space. To prove this, we start with the more basic result for L^0.

Theorem 4.1. *Let $(X_1, \mathcal{F}_1, P_1)$ and $(X_2, \mathcal{F}_2, P_2)$ be probability spaces and $T : X_1 \to X_2$ measure-preserving. If $F \in L^0(X_2, \mathcal{F}_2, P_2)$ then*

$$\int_{X_1} U_T F dP_1 = \int_{X_2} F dP_2$$

Proof. We show the result for simple functions, from which the result will follow for non-negative measurable functions, and then directly for measurable functions as well.

Suppose that $s \in L^0(X_2, \mathcal{F}_2, P_2)$ is simple. Observe that for any $B \in \mathcal{F}_2$ and $x \in X_1$ we have $U_T \chi_B = \chi_B(Tx) = \chi_{T^{-1}(B)}(x)$. Since U_T is linear on elements of $L^0(X_2, \mathcal{F}_2, P_2)$ we have

$$\int_{X_1} U_T s dP_1 = \int_{X_1} U_T \left(\sum_{i=1}^{r} a_i \chi_{B_i} \right) dP_1$$

Where $a_i \in \mathbb{R}$ and $B_i \in X_2$ for $1 \leq i \leq r \in \mathbb{N}$

$$\begin{aligned}
&= \int_{X_1} \sum_{i=1}^{r} a_i U_T \chi_{B_i} dP_1 \text{ (by linearity)} \\
&= \int_{X_1} \sum_{i=1}^{r} a_i \chi_{T^{-1} B_i} dP_1 \\
&= \sum_{i=1}^{r} a_i P_1(T^{-1} B_i) \\
&= \sum_{i=1}^{r} a_i P_2(B_i) \text{ (since } T \text{ is measure-preserving)} \\
&= \int_{X_2} a_i \chi_{B_i} dP_2 \\
&= \int_{X_2} s dP_2
\end{aligned}$$

Now suppose that $f \in L^0(X_2, \mathcal{F}_2, P_2)$ is non-negative. Then for any sequence of simple functions $s_n \nearrow f$ we have,

$$\begin{aligned}
\int_{X_1} U_T f dP_1 &= \lim_{n \to \infty} \int_{X_1} U_T s_n dP_1 \\
&= \lim_{n \to \infty} \int_{X_2} s_n dP_2 \\
&= \int_{X_2} f dP_2
\end{aligned}$$

and so the result follows for all functions in $L^0(X_2, \mathcal{F}_2, P_2)$. □

Theorem 4.2. *Let $p \geq 1$. Consider the induced operator U_T on L^p defined in the same way as for L^0 (i.e., $U_T f = f \circ T$ where T is a measure-preserving map). Then $U_T L^p(X_2, \mathcal{F}_2, P_2) \subseteq L^p(X_1, \mathcal{F}_1, P_1)$ and $||U_T f||_p = ||f||_p \forall f \in L^p(X_2, \mathcal{F}, P_2)$.*

Proof. Let $f \in (X_2, \mathcal{F}, P_2)$. Then $U_T(f(x)) = f(T(x))\ \forall x \in X_1$ and it is clear that $U_T \in (X_1, \mathcal{F}_1, P_1)$ so that $L^p(X_2, \mathcal{F}_2, P_2) \subseteq L^p(X_1, \mathcal{F}_1, P_1)$. Now, observe that

$$\begin{aligned} ||U_T f||_p^p &= \int_{X_1} |U_T f|^p \, dP_1 \\ &= \int_{X_1} |f \circ T|^p \, dP_1 \end{aligned}$$

Let $F = |f|^p$. Then,

$$\begin{aligned} &= \int_{X_1} F \circ T dP_1 \\ &= \int_{X_1} U_T F dP_1 \\ &= \int_{X_2} F dP_2 \text{ (by Theorem 4.1)} \\ &= \int_{X_2} |f|^p dP_2 \\ \Rightarrow ||U_T f||_p &= ||f||_p \end{aligned}$$

□

Theorem 4.2 above is significant in that it shows that the linear operator U_T induced by a measure-preserving map $T : (X_1, \mathcal{F}_1, P_1) \to (X_2, \mathcal{F}_2, P_2)$ is actually an isometry between $L^p(X_2, \mathcal{F}_2, P_2) \to L^p(X_1, \mathcal{F}_1, P_1)$. This natural connection between L^p spaces and the notion of an isometry is good motivation to proceed further in our study of measure-preserving maps and the structure they imply about function spaces.

5. Ergodicity

Once again, although we can talk about measure-preserving transformations between two general measure spaces, we will examine measure-preserving maps $T : (\Omega, \mathcal{F}, \mu) \to (\Omega, \mathcal{F}, \mu)$ which we will simply call a measure-preserving map on $(\Omega, \mathcal{F}, \mu)$.

Definition 5.1. Let T be a measure-preserving map on $(\Omega, \mathcal{F}, \mu)$. Then $A \in \mathcal{F}$ is *invariant* under T if $A = T^{-1}A$. $A \in \mathcal{F}$ is called *almost invariant* under T if A and $T^{-1}A$ differ by a set of measure 0 (i.e., $\mu(A \Delta T^{-1}A) = 0$).
(Recall that for any two sets A, B we have $A \Delta B := (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$)

If we let $\mathscr{I}$ denote the set of all invariant sets under a measure-preserving map T on $(\Omega, \mathcal{F}, \mu)$, then observe that $\mathscr{I}$ is non-empty since $T^{-1}\Omega = \Omega \in \mathscr{I}$. Moreover, the fact that inverse images preserve complements and countable unions means that $\mathscr{I}$ will also be closed under complementation and countable union. In other words, $\mathscr{I}$ is a σ-algebra of subsets of Ω. The set of all almost invariant sets also form a σ-algebra of subsets of Ω by a similar argument.

Definition 5.2. Let T be a measure-preserving map on $(\Omega, \mathcal{F}, \mu)$. Then T is *Ergodic* if every $A \in \mathcal{F}$ that is invariant under T satisfies $\mu(A) = 0$ or $\mu(A^c) = 0$. Notice that if T is a measure-preserving map on a probability space $(X, \mathcal{F}, P)$, then T is ergodic if and only if every invariant $A \in \mathcal{F}$ under T satisfies $P(A) = 0$ or 1.

The notion of Ergodicity can be expressed in a variety of different ways for probability spaces. Before we showcase this, we need to connect the idea of an invariant set and an almost invariant set.

Theorem 5.1. *Let T be a measure-preserving map on a measure space $(\Omega, \mathcal{F}, \mu)$. Suppose that A is almost invariant under T. Then $\exists B \subseteq \mathcal{F} \ni B = T^{-1}B$ and $\mu(A \Delta B) = 0$.*

Proof. Suppose that $A \in \mathcal{F}$ is an almost invariant set under T. Consider

$$B = \limsup_{n \in \mathbb{N}} T^{-n}A := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-n}A$$

Then $B \in \mathcal{F}$ since every pre-image of A is, and

$$\begin{aligned} T^{-1}B &= T^{-1}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} T^{-n}A\right) \\ &= \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} T^{-(n+1)}A \\ &= \limsup_{n\in\mathbb{N}} T^{-(n+1)}A \\ &= B \end{aligned}$$

So B is invariant under T.
In order to make progress in evaluating the measure of $A\Delta B$, consider its disjoint parts $A\Delta B = (A \setminus B) \cup (B \setminus A)$. Suppose $\omega \in A \setminus B$. Then $\omega \notin \limsup_{n\in\mathbb{N}} T^{-n}A \Rightarrow \omega \notin T^{-n}A$ for finitely many $n \in \mathbb{N}$. Hence $\exists k \in \mathbb{N} \ni \omega \in T^{-k}A \setminus T^{-(k+1)}A$. Now suppose $\omega \in B \setminus A$. Then $T^n\omega A$ for infinitely many n, but $\omega \notin A$. Let $k \in \mathbb{N}$ be the smallest natural number that gives $T^{k+1}\omega \in A$. Then $\omega \in T^{-(k+1)}A \setminus T^{-k}A$. In either case, we have shown the existence of a $k \in \mathbb{N}$ such that

$$\begin{aligned} A\Delta B &\subseteq T^{-k}A \setminus T^{-(k+1)}A \text{ or } T^{-(k+1)}A \setminus T^{-k}A \\ \Rightarrow A\Delta B &\subseteq \bigcup_{k=0}^{\infty} (T^{-k}A \setminus T^{-(k+1)}A) \cup (T^{-(k+1)}A \setminus T^{-k}A) \\ &= \bigcup_{k=0}^{\infty} T^{-k}A\Delta T^{-(k+1)}A \end{aligned}$$

Finally, since T is measure-preserving we have $\mu(T^{-k}A\Delta T^{-(k+1)}A) = \mu(A\Delta T^{-1}A) = 0$ by assumption. Hence, $\mu(A\Delta B) = 0$ as well. □

As is common throughout measure theory, sets that differ by sets of measure zero usually give the same result. The same is true of Ergodicity.

Theorem 5.2. *A measure-preserving map T on a measure space $(\Omega, \mathcal{F}, \mu)$ is ergodic if and only if each almost invariant set A satisfies $\mu(A) = 0$ or $\mu(A^c) = 0$.*

Proof. ($\Leftarrow$). Since every invariant set is almost invariant, ergodicity is clear.

($\Rightarrow$). Suppose that A is an almost invariant set with $\mu(A) = 0$ or $\mu(A^c) = 0$. By Theorem 5.1 $\exists B \in \mathcal{F}$ such that B is invariant and $\mu(A\Delta B) = 0$. Now observe,

$$\begin{aligned} P(A) &= P((A \cap (A\Delta B)) \cup (A \cap B)) \\ &= P(A \cap (A\Delta B)) + P(A \cap B) \end{aligned}$$

But $0 \le P(A \cap (A\Delta B) \le P(A\Delta B) = 0 \Rightarrow P(A \cap (A\Delta B) = 0$, so

$$\Rightarrow P(A) = P(A \cap B)$$

By symmetry, we get that $P(B) = P(A \cap B)$ so that $P(A) = P(B)$. Then by the ergodicity of T, we conclude that $P(A) = P(B) = 0$ or $P(A^c) = P(B^c) = 0$. □

Theorems 5.1 and 5.2 shows that whenever we are dealing with an almost invariant set, we can simply use an invariant set that differs from the almost invariant set by a set of measure zero. This is useful in the following theorem.

Theorem 5.3. *Let T be a measure-preserving map on a probability space $(X, \mathcal{F}, P)$. Then the following are equivalent:*

(1) *T is ergodic.*
(2) *The only $A \in \mathcal{F}$ with $P(A\Delta T^{-1}A) = 0$ are those with $P(A) = 0$ or $P(A) = 1$.*
(3) *$\forall A \in \mathcal{F}$ with $P(A) > 0$, we have $P(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$.*
(4) *$\forall A, B \in \mathcal{F}$ with $P(A) > 0, P(B) > 0, \exists n > 0 \ni P(B \cap T^{-n}A) > 0$.*

Proof. (1) ⇒ (2). Suppose that T is ergodic and $P(A\Delta T^{-1}A) = 0$ for some $A \in \mathcal{F}$ (that is, A is almost invariant under T). By Theorem 5.1, $\exists B \in \mathcal{F}$ such that B is invariant under T and $P(A\Delta B) = 0$. As seen in Theorem 5.2, this implies that $P(A) = P(B)$. But by ergodicity, $P(B) = 0$ or 1 so that $P(A) = 0$ or 1 as well.

(2) ⇒ (3). Let $A \in \mathcal{F}$ with $P(A) > 0$. For convenience, define

$$A_1 = \bigcup_{n=1}^{\infty} T^{-n}A$$

So that $T^{-1}A_1 = T^{-1}\left(\bigcup_{n=1}^{\infty} T^{-n}A\right) = \bigcup_{n=1}^{\infty} T^{-(n+1)}A = \bigcup_{n=2}^{\infty} T^{-n}A \subseteq \bigcup_{n=1}^{\infty} T^{-n}A = A_1$. Since T is measure preserving, we also have $P(T^{-1}A_1) = P(A_1) \Rightarrow P(T^{-1}A_1\Delta A_1) = 0$. Hence by (2), $P(A_1) = 0$ or 1. Since $T^{-1}A \subseteq A_1 \Rightarrow P(T^{-1}A) \leq P(A_1) \Rightarrow 0 < P(A) \leq P(A_1)$ (since T is measure-preserving) we have that $P(A_1) = 1$, as desired.

(3) ⇒ (4). Suppose that $A, B \in \mathcal{F}$ satisfy $P(A) > 0$ and $P(B) > 0$. BY (3), we have $P(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$. But then $P(\bigcup_{n=1}^{\infty} T^{-n}A) + P(B) = P(B \cup \bigcup_{n=1}^{\infty} T^{-n}A) + P(B \cap \bigcup_{n=1}^{\infty} T^{-n}A) \Rightarrow P(B) = P(B \cap \bigcup_{n=1}^{\infty} T^{-n}A)$ since P is a probability measure. Hence $P(B \cap \bigcup_{n=1}^{\infty} T^{-n}A) > 0$ so that $P(B \cap T^{-n}A) > 0$ for some $n \in \mathbb{N}$.

(4) ⇒ (1). We prove the contrapositive of this implication. Suppose that $A \in \mathcal{F}$ with A invariant under T and $0 < P(A) < 1$. Then $A \subset X$ so that $P(X \setminus A) > 0$. Observe,

$$\begin{aligned} 0 &= P(\emptyset) \\ &= P(A \cap (X \setminus A)) \\ &= P(T^{-n}A \cap (X \setminus A)) \end{aligned}$$

Which holds $\forall n \geq 1$. Hence we have the negation of (4), completing the proof. □

There are even more equivalent conditions for ergodicity if we consider functions that map a measurable space to the measure space $(\mathbb{R}, \mathcal{B})$ (where $\mathcal{B}$ stands for the Borel sets).

Definition 5.3. Let $(\Omega, \mathcal{F})$ be a measurable space and $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$. Let T be a measure-preserving map on $(\Omega, \mathcal{F})$. Then g is *invariant* (or *almost invariant*) if $g(T\omega) = g(\omega)$ for all (or almost all) $\omega \in \Omega$. In the language of Section 4, g is invariant if $U_T g = g$ (i.e., it is invariant under the induced operator of T).

Once again, we desire a theorem to unite the concept of a function being invariant and almost invariant. When it comes to ergodicity, it turns out that it does not matter which kind of invariance a function has.

Theorem 5.4. *Let T be a measure-preserving map on $(\Omega, \mathcal{F}, \mu)$. Then the following are equivalent:*

1. *T is ergodic.*
2. *Every almost invariant function $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is constant a.e..*
3. *Every invariant function $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is constant a.e..*

Proof. (1) ⇒ (2). Suppose that T is ergodic and $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is almost invariant. For $\lambda \in \mathbb{R}$ consider

$$A_\lambda = \{\omega \in \Omega \mid g(\omega) \leq \lambda\}$$

Then $TA_\lambda = \{T(\omega) \mid \omega \in \Omega \text{ and } g(\omega) \leq \lambda\}$ is almost equal to $\{T(\omega) \mid \omega \in \Omega \text{ and } g(T\omega) \leq \lambda\}$ since $g(\omega) = g(T\omega)$ a.e.. Hence, A_λ is almost invariant. By Theorem 5.2 and the fact that T is ergodic, we have that $\mu(A) = 0$ or $\mu(A^c) = 1$. Now, let

$$c = \sup\{\lambda \mid \mu(A_\lambda = 0\}$$

If μ is trivially always zero then $c = \infty$. Disregarding this case, we have that $A_\lambda \nearrow \Omega$ as $\lambda \to \infty$ and $A_\lambda^c \nearrow \Omega$ as $\lambda \to -\infty$. Hence, if Ω has finite measure (like in the case where we have a probability space), then c must take on a finite value. That is, $c \in \mathbb{R}$. Putting this all together, we see that

$$\begin{aligned}\mu\{\omega \in \Omega \mid g(\omega) < c\} &= \mu\left(\bigcup_{n=1}^{\infty}\{\omega \in \Omega \mid g(\omega) \leq c - \frac{1}{n}\right) \\ &= \mu\left(\bigcup_{n=1}^{\infty} A_{c-\frac{1}{n}}\right) \\ &= 0\end{aligned}$$

Through a similar argument we see that $\mu\{\omega \in \Omega \mid g(\omega) > c\} = 0$ as well so that $g = c$ a.e..

(2) $\Rightarrow$ (3). Since every invariant function also satisfies the definition of an almost invariant function, this implication follows immediately.

(3) $\Rightarrow$ (1). Suppose that every invariant function from $(\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is constant a.e.. Let A be an invariant set. Recall from Theorem 4.1 that $U_T\chi_{A_i} = U_{T^{-1}A}$. Since A is invariant, however, we have that $T^{-1}A = A$ so that χ_A is invariant. Assuming (3) gives us that χ_A is constant a.e., so $\chi_A \equiv 0$ a.e. or 1 a.e.. If $\chi_A = 0$ a.e., then $\int_\Omega \chi_A d\mu = \mu(A) = 0$. Similarly, if $\chi_A = 1$ a.e. then $\chi_{A^c} = 0$ a.e. so that $\int_\Omega \chi_{A^c} d\mu = \mu(A^c) = 0$. ☐

As a note to Theorem 5.4, we can actually have $g \in L^p(\Omega, \mathcal{F}, \mu)$ and if $U_T g = g$, then g is ergodic (the proof is very similar to (3) $\Rightarrow$ (1) above).

6. Mixing

The concept of mixing is of interest because it relates closely to the many applications of Ergodic Theory. In fact, the idea of Ergodicity is simply inadequate when it comes to many applications in Thermodynamics (see Section 8). Throughout our discussion on Mixing, we will deal with a Probability space $(X, \mathcal{F}, P)$ and a measure-preserving map T on this space.

Definition 6.1. Let T be a measure-preserving map on a probability space $(X, \mathcal{F}, P)$. Then T is *mixing* if $\forall A, B \in \mathcal{F}$ we have

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

It turns out that every map that is mixing is also ergodic.

Theorem 6.1. *Let T be mixing on$(X, \mathcal{F}, P)$. Then T is ergodic.*

Proof. Suppose that B is an invariant set. Since T is measure-preserving we have that $B = T^{-n}B$. Now let $A \in \mathcal{F}$. Then $P(A \cap B) = (A \cap T^{-n}B)$ so if we take limits as $n \to \infty$ then $P(A \cap B) = P(A)P(B)$ by mixing. Consider the case where $A = B$. Then

$$P(B \cap B) = P(B)P(B) \Rightarrow P(B) = P(B)^2 \Rightarrow P(B)(P(B) - 1) = 0$$

So that $P(B) = 0$ or 1. It follows that T is ergodic. ☐

As we saw with measure-preserving maps (and all throughout measure theory), if a property holds for an algebra that generates a σ-algebra, then it holds for the entire σ-algebra. The case with mixing is no different. Unlike previous cases, however, the argument for why this is true involves much more than just the good sets principle.

Theorem 6.2. *Let T be a measure-preserving map on $(X, \mathcal{F}, P)$ and $\mathcal{F}_0$ an algebra of subsetes of X satisfying $\mathcal{F} = \sigma(\mathcal{F}_0)$. If T is mixing when it is restricted to $\mathcal{F}_0$, then it is mixing on all of $\mathcal{F}$.*

Proof. Suppose that $A, B \in \mathcal{F}$. Then by Theorem 2.2 we can find sequences of sets $\{A_k\}_{k\in\mathbb{N}}, \{B_k\}_{k\in\mathbb{N}} \subseteq \mathcal{F}_0$ such that $P(A\Delta A_k), P(B\Delta B_k) \to 0$ as $k \to \infty$. We are interested in evaluating the limit of $P(A \cap T^{-n}B)$

as $n \to \infty$. To do this, we consider the (symmetric) difference between the sets $A \cap T^{-n}B$ and $A_k \cap T^{-n}B_k$ and combine it with the fact that A_k (or B_k) approximates A (or B). Observe that

$$\begin{aligned}(A \cap T^{-n}B)\Delta(A_k \cap T^{-n}B_k) &= ((A \cap T^{-n}B)\Delta A_k) \cap ((A \cap T^{-n}B)\Delta T^{-n}B_k)\\ &\subseteq ((A \cap T^{-n}B)\Delta A_k) \cup ((A \cap T^{-n}B)\Delta T^{-n}B_k)\\ &\subseteq (A\Delta A_k) \cup (T^{-n}B\Delta T^{-n}B_k)\\ &= (A\Delta A_k) \cup T^{-n}(B\Delta B_k)\\ \Rightarrow P\left((A \cap T^{-n}B)\Delta(A_k \cap T^{-n}B_k)\right) &\leq P(A\Delta A_k) + P\left(T^{-n}(B\Delta B_k)\right)\end{aligned}$$

And since T is measure-preserving on all of $\mathcal{F}$,

$$\Rightarrow P\left((A \cap T^{-n}B)\Delta(A_k \cap T^{-n}B_k)\right) \leq P(A\Delta A_k) + P(B\Delta B_k)$$

From which we can now observe that $P(A\Delta A_k) + P(B\Delta B_k) \to 0$ as $k \to \infty$ (independently of choice of n). Thus, $P((A \cap T^{-n}B)\Delta(A_k \cap T^{-n}B_k)) \to 0$ so that $P(A_k \cap T^{-n}B_k) \to P(A \cap T^{-n}B)$ as $k \to \infty$ by what we have established for symmetric differences in Theorem 5.2. Now, since T is mixing on $\mathcal{F}_0$, we also have $P(A_k \cap T^{-n}B_k) \to P(A_k)P(B_k)$ as $n \to \infty$. Hence,

$$\begin{aligned}\lim_{n\to\infty} P(A \cap T^{-n}B) &= \lim_{n\to\infty}\lim_{k\to\infty} P(A_k \cap T^{-n}B_k)\\ &= \lim_{k\to\infty}\lim_{n\to\infty} P(A_k \cap T^{-n}B_k)\\ &= \lim_{k\to\infty} P(A_k)P(B_k)\\ &= P(A)P(B)\end{aligned}$$

as desired.

□

7. The Birkhoff Ergodic Theorem

This section will see us proving the big result of this paper, which we will call "The Birkhoff Ergodic Theorem" (Theorem 7.4) since it was first put forth and proved by George David Birkhoff in 1931 [1]. The road to this Theorem, however, is a bit technical so we will step through a variety of theorems to get there. It is recommended that the reader go through the set-up put forth in the preceding sections in order to understand the proofs of this section. Otherwise, the statement can more or less be interpreted without a comprehensive reading of every preceding definition (which perhaps showcases the power of The Ergodic Theorem).

For a measure-preserving transformation T, recall the induced (linear) operator U_T from 4. We will be interested in the behaviour of general linear operators, which we will denote by U. We begin with the following definition to clear up some notation.

Definition 7.1. Let An operator $U : L^1(\Omega, \mathcal{F}, \mu) \to L^1(\Omega, \mathcal{F}, \mu)$ is *positive* if $\forall f \in L^1(\Omega, \mathcal{F}, \mu)$, $f \geq 0 \Rightarrow Uf \geq 0$.

In the following sections L^1 will represent the Borel-measurable functions that are **real-valued**.

Theorem 7.1. *(Maximal Ergodic Theorem) Let $U : L^1(\Omega, \mathcal{F}, \mu) \to L^1(\Omega, \mathcal{F}, \mu)$ be a positive linear operator with $||U|| \leq 1$ (where $||\ ||$ is the usual $\sup$ norm given by Definition 2.16). Let $0 < N \in \mathbb{N}$ and $f \in L^1(\Omega, \mathcal{F}, \mu)$. Set $f_0 = 0$, $f_n = f + Uf + U^2f + \cdots + U^{n-1}f$ for $n \geq 1$, and $F_N = \max\{f_n \mid 0 \leq n \leq N\}$. Let $A = \{\omega \in \Omega \mid F_N(x) > 0\}$. Then $\int_A f d\mu \geq 0$.*

Proof. First observe that $F_N \geq 0$ by definition. Moreover, since $f \in L^1(\Omega, \mathcal{F}, \mu)$ and $||U|| \leq 1$, repeated applications of U will remain bounded. Since F_N is a finite linear combination of terms of the form $U^k f$, we have that $F_N \in L^1(\Omega, \mathcal{F}, \mu)$ as well.

It is also clear that for $0 \leq n \leq N$ we have $F_n \geq f_n$ so that $UF_N \geq Uf_n$ since U is positive. Then since

$$Uf_n = Uf + U^2f + U^3f + \cdots + U^nf = Uf_{n+1} - f$$

we have $UF_N + f \geq f_{n+1}$. Thus, $\forall \omega \in \Omega$ we have

$$\begin{aligned} UF_N(\omega) + f(\omega) &\geq \max_{0 \leq n \leq N-1} f_{n+1}(\omega) \\ &= \max_{1 \leq n \leq N} f_n(\omega) \\ &= \max_{0 \leq n \leq N} f_n(\omega) \text{ whenever } F_N(x) > 0 \ (\textbf{a key subtlety}) \\ &= F_N(x) \end{aligned}$$

Which shows that $f \geq F_N - UF_N$ on A. Hence,

$$\begin{aligned} \int_A f d\mu &\geq \int_A (F_N - UF_N) d\mu \\ &= \int_A F_N d\mu - \int_A UF_N d\mu \end{aligned}$$

And since $F_N(\omega) = 0 \ \forall \omega \in A^c$, we have $\int_A F_N d\mu = \int_\Omega F_N d\mu$. Moreover, $F_N \geq 0 \Rightarrow UF_N \geq 0$ since U is positive, which implies that $\int_A UF_N d\mu \leq \int_\Omega UF_N d\mu$. Putting this all together,

$$\Rightarrow \int_A f d\mu \geq \int_\Omega F_N d\mu - \int_\Omega UF_N d\mu$$

Finally, since $||U|| \leq 1$, we have that $\int_\Omega F_N d\mu \geq \int_\Omega UF_N d\mu$ so that

$$\int_A f d\mu \geq 0$$

as desired. □

The usefulness of the Maximal Ergodic Theorem (Theorem 7.1) comes in setting up the next theorem. In the Maximal Ergodic Theorem, observe that we can write the set A as

$$A = \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \sum_{k=0}^{n-1} (U^k f)(\omega) > 0 \right\}$$

It will be convenient to consider sets of the similar form

$$B_\lambda = \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} (U^k f)(\omega) > \lambda \right\}$$

where $\lambda \in \mathbb{R}$.

Theorem 7.2. *Let T be a measure-preserving map. Consider $h \in L^1(\Omega, \mathcal{F}, \mu)$ and A and B_λ ($\lambda \in \mathbb{R}$) as defined previously:*

$$\begin{aligned} A &= \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \sum_{k=0}^{n-1} (U_T^k h)(\omega) > 0 \right\} \\ B_\lambda &= \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} (U_T^k h)(\omega) > \lambda \right\} \end{aligned}$$

If A is invariant and $\mu(A) < \infty$ then

$$\int_{B_\lambda \cap A} h d\mu \geq \lambda \mu(B_\lambda \cap A)$$

Proof. Fix $\lambda \in \mathbb{R}$. First consider the case where $A = \Omega$ (for which A invariant is already true) and $\mu(\Omega) < \infty$. Define $f = h - \lambda$. Then,

$$\begin{aligned} B_\lambda &= \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} (U_T^k h)(\omega) > \lambda \right\} \\ &= \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \frac{1}{n} \sum_{k=0}^{n-1} (U_T^k f)(\omega) + \frac{1}{n} \sum_{k=0}^{n-1} \lambda > \lambda \right\} \\ &= \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \sum_{k=0}^{n-1} (U_T^k f)(\omega) > 0 \right\} \\ &= \left\{ \omega \in \Omega \mid \sup_{n \geq 1} f_n(\omega) > 0 \right\} \\ &= \bigcup_{N=0}^{\infty} \{\omega \in \Omega \mid F_N(\omega) > 0\} \end{aligned}$$

Hence by the Maximal Ergodic Theorem (Theorem 7.1), we have that

$$\begin{aligned} & \int_{B_\lambda} f d\mu > 0 \\ \Rightarrow & \int_{B_\lambda} (h - \lambda) d\mu > 0 \\ \Rightarrow & \int_{B_\lambda} h d\mu > \lambda \mu(B_\lambda) \end{aligned}$$

Now for the general case for $A \in \mathcal{F}$ with A invariant and $\mu(A) < \infty$, we restrict the measure-preserving map T to A and view A as the entire space. Since we satisfy the assumptions of finite measure and invariance under T, it follows that

$$\int_{B_\lambda \cap A} h d\mu > \lambda \mu(B_\lambda \cap A)$$

□

We have just one more Theorem to set up before we move on to the promised proof of the Ergodic Theorem. The next Theorem actually does most of the heavy-lifting in terms of characterizing the points that non-trivialize the ergodic theorem.

Theorem 7.3. *Let $f \in L^1(\Omega, \mathcal{F}, \mu)$. Let T be a measure-preserving map and define $f_n = \sum_{k=0}^{n-1} U_T^k f$ just like before for $n \geq 1$ (and $f_0 = 0$). Let $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$, and define also*

$$E_{\alpha,\beta} = E_{\alpha,\beta}(f) = \left\{ \omega \in \Omega \mid \liminf_{n \to \infty} \frac{f_n(\omega)}{n} < \alpha < \beta < \limsup_{n \to \infty} \frac{f_n(\omega)}{n} \right\}$$

With the same notation as before, we also have

$$B_\beta = \left\{ \omega \in \Omega \mid \sup_{n \geq 1} \frac{1}{n} f_n(\omega) > \beta \right\}$$

Then,

(1) *$E_{\alpha,\beta}$ is almost invariant*
(2) $\mu(E_{\alpha,\beta}) < \infty$
(3) *Even stronger, $\mu(E_{\alpha,\beta}) = 0$*

Proof. (1). Observe,

$$\begin{aligned}
U_T\left(\frac{1}{n}f_n(\omega)\right) &= \frac{1}{n}U_T f_n(\omega) \\
&= \frac{1}{n}U_T\left(\sum_{k=0}^{n-1}U_T^k f(\omega)\right) \\
&= \frac{1}{n}\sum_{k=0}^{n-1}U_T^{k+1}f(\omega) \\
&= \frac{1}{n}\sum_{k=1}^{n}U_T^k f(\omega) \\
&= \frac{1}{n}\sum_{k=0}^{n}\left(U_T^k f(\omega)\right) - \frac{1}{n}f(\omega) \\
&= \frac{n+1}{n}\left(\frac{1}{n+1}f_{n+1}(\omega)\right) - \frac{1}{n}f(\omega)
\end{aligned}$$

Since $f \in L^1(\Omega, \mathcal{F}, \mu)$, $f(\omega)$ must be finite a.e.. Thus,

$$\begin{aligned}
\liminf_{n\to\infty} U_T\left(\frac{1}{n}f_n(\omega)\right) &= \liminf_{n\to\infty}\left(\frac{n+1}{n}\left(\frac{1}{n+1}f_{n+1}(\omega)\right) - \frac{1}{n}f(\omega)\right) \\
&= \liminf_{n\to\infty}\left(\frac{1}{n+1}f_{n+1}(\omega)\right) \\
&= \liminf_{n\to\infty}\left(\frac{1}{n}f_n(\omega)\right)
\end{aligned}$$

holds a.e.. An identical argument can be made for the $\limsup$. It follows that $E_{\alpha,\beta}$ is almost invariant.

(2). Suppose that $E \in \mathcal{F}$ satisfies $E \subseteq E_{\alpha,\beta}$ and $\mu(E) < \infty$. Similar to the definition of B_β (and what was done in the proof of the Maximal Ergodic Theorem), define

$$F_\beta = \left\{\omega \in \Omega \mid \sup_{n\geq 1}(f - \beta\chi_E)_n(\omega) > 0\right\}$$

Observe that by definition, $E_{\alpha,\beta} \subseteq B_\beta$. Hence if $\omega \in E_{\alpha,\beta}$ then $\exists n \in \mathbb{N}$ such that

$$\frac{1}{n}\sum_{k=0}^{n-1}U_T^k f(\omega) > \beta$$

But then since $0 \leq \chi_E \leq 1$ we clearly have

$$\beta \geq \frac{1}{n}\sum_{k=0}^{n-1}\beta U_T^k \chi_E(\omega)$$

so that

$$\begin{aligned}
&\frac{1}{n}\sum_{k=0}^{n-1}U_T^k\left(f(\omega) - \beta\chi_E\right) > 0 \\
&\Rightarrow \sup_{n\geq 1}(f - \beta\chi_E)_n(\omega) > 0
\end{aligned}$$

so $\omega \in F_\beta$ as well. Thus,

$$E \subseteq E_{\alpha,\beta} \subseteq B_\beta \subseteq F_\beta$$

In particular, we have $E \cap F_\beta = E$. We can now apply the Maximal Ergodic Theorem (Theorem 7.1) to get

$$\begin{aligned} \int_{F_\beta} (f - \beta\chi_E) d\mu \geq 0 \\ \Rightarrow \int_{F_\beta} f d\mu &\geq \beta\mu(F_\beta \cap E) \\ &= \beta\mu(E) \end{aligned}$$

And since $\int_{F_\beta} |f| d\mu \geq \int_{F_\beta} f d\mu$, we have

$$\mu(E) \leq \frac{1}{\beta}||f||_1 < \infty$$

since $f \in L^1(\Omega, \mathcal{F}, \mu)$. Note in the above that $\beta = 0 \iff f \equiv 0$ so we can disregard the trivial case where $f = 0$. Now, for each $n \in \mathbb{N}$ recall that Theorem 4.2 gives

$$||U_T^n f||_1 = ||f||_1 < \infty$$

so that $\{\omega \in \Omega \mid |U_T^n f(\omega)| > 0\}$ may be written as a countable union of sets of finite measure. Since $E_{\alpha,\beta} \subseteq \{\omega \in \Omega \mid |U_T^n f(\omega)| > 0\}$ it follows that $E_{\alpha,\beta}$ may also be written as a countable union of sets of finite measure. We now have,

$$\mu(E_{\alpha,\beta}) = \sup \{\mu(E) \mid E \in \mathcal{F}, E \subseteq E_{\alpha,\beta}, \mu(E) < \infty\}$$

We just showed, however, that every $\mu(E) \in \{\mu(E) \mid E \in \mathcal{F}, E \subseteq E_{\alpha,\beta}, \mu(E) < \infty\}$ satisfies $\mu(E) \leq \frac{1}{\beta}||f||_1$. Hence,

$$\mu(E_{\alpha,\beta}) \leq \frac{1}{\beta}||f||_1 < \infty$$

Which finishes (2).

(3). Now, restrict T to $E_{\alpha,\beta}$, $\mathcal{F}$ to elements contained in $E_{\alpha,\beta}$, and μ to $E_{\alpha,\beta}$. That is, view $E_{\alpha,\beta}$ as the universal space. By (1) and (2) we have that $E_{\alpha,\beta}$ is invariant and has finite measure. Thus we can apply Theorem 7.2 to obtain

$$\int_{E_{\alpha,\beta}} f d\mu \geq \beta\mu(E_{\alpha,\beta})$$

just as in (2). By replacing f, α, β with $-f, -\beta, -\alpha$ respectively and recalling that $\limsup(-f) = -\liminf f$ (the details are identical to the presentation in (2)) we obtain

$$\int_{E_{\alpha,\beta}} f d\mu \leq \alpha\mu(E_{\alpha,\beta})$$

so that

$$\beta\mu(E_{\alpha,\beta}) \leq \alpha\mu(E_{\alpha,\beta})$$

Since we also have $\mu(E_{\alpha,\beta}) < \infty$, we conclude

$$\beta \leq \alpha \text{ or } \mu(E_{\alpha,\beta}) = 0$$

But since $\alpha < \beta$, we must have $\mu(E_{\alpha,\beta}) = 0$, as desired. □

We are finally ready to prove the Birkhoff Ergodic Theorem (sometimes also called the Pointwise Ergodic Theorem).

Theorem 7.4. *(Birkhoff Ergodic Theorem) Let $T : (\Omega, \mathcal{F}, \mu) \to (\Omega, \mathcal{F}, \mu)$ be a measure-preserving map and $f \in L^1(\Omega, \mathcal{F}, \mu)$. Then*

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

converges a.e. to a function $f^* \in L^1(\Omega, \mathcal{F}, \mu)$. Moreover, $f^* \circ T = f^*$ a.e., and $\int_\Omega f^* d\mu = \int_\Omega f d\mu$ if $\mu(\Omega) < \infty$.

Proof. The key to this proof is realizing that the set $E_{\alpha,\beta}$ defined in Theorem 7.3 gives the points where the above sum fails to converge. That is, if

$$D = \left\{ \omega \in \Omega \mid \frac{1}{n} f_n(\omega) \text{ fails to converge} \right\}$$

Then we can use the density and countability of the rationals to write

$$D = \bigcup_{\alpha,\beta \in \mathbb{Q}} E_{\alpha,\beta}(f)$$

Then by Theorem 7.3, D is a (countable) union of sets of measure zero so that $\mu(D) = 0$. That is, $\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(\omega))$ converges a.e. to a function f^*. We define $f^*(d) = 0 \; \forall d \in D$.

We now show that $f^* \in L^1(\Omega, \mathcal{F}, \mu)$. First observe that since $|\frac{1}{n} f_n(\omega)| \geq 0$ we can use Fatou's Lemma to obtain

$$\int_\Omega \liminf_{n\to\infty} \left| \frac{1}{n} f_n(\omega) \right| d\mu \leq \liminf_{n\to\infty} \int_\Omega \left| \frac{1}{n} f_n(\omega) \right| d\mu < \infty$$

Since $f_n \in L^1(\Omega, \mathcal{F}, \mu) \; \forall n \in \mathbb{N}$. But we also have

$$||f^*||_1 = \int_\Omega |f^*| d\mu = \int_\Omega \liminf_{n\to\infty} \left| \frac{1}{n} f_n(\omega) \right| d\mu$$

so that

$$||f^*||_1 < \infty$$

and thus $f^* \in L^1(\Omega, \mathcal{F}, \mu)$.

$f^* \circ T = f^*$ is clear from the same argument made in Theorem 7.3 part (2).

Finally, suppose that $\mu(\Omega) < \infty$. Consider the set

$$D_k^n = \left\{ \omega \in \Omega \mid \frac{k}{n} \leq \limsup_{i\to\infty} \frac{1}{i} f_i(\omega) \leq \frac{k+1}{n} \right\}$$

for $k \in \mathbb{Z}$ and $n \geq 1$. With the original definition of B_λ we take $\lambda = \frac{k}{n}$ so that

$$B_{\frac{k}{n}} = \left\{ \omega \in \Omega \mid \sup_{i \geq 1} \frac{1}{i} f_i(\omega) > \frac{k}{n} \right\}$$

Then $\forall \epsilon > 0$ we have $D_k^n \subseteq B_{\frac{k}{n}-\epsilon}$ so that $D_k^n \cap B_{\frac{k}{n}-\epsilon} = D_k^n$. Applying Theorem 7.2 once more, we get

$$\int_{D_k^n} f d\mu \geq \left(\frac{k}{n} - \epsilon \right) \mu(D_k^n)$$

And since ϵ was arbitrary,

$$\int_{D_k^n} f d\mu \geq \left(\frac{k}{n} \right) \mu(D_k^n)$$

Now by the definition of D_k^n we have

$$\begin{aligned}
&\limsup_{i\to\infty} \frac{1}{i} f_i(\omega) \leq \frac{k+1}{n} \\
\Rightarrow &\int_{D_k^n} \limsup_{i\to\infty} \frac{1}{i} f_i(\omega) d\mu \leq \left(\frac{k+1}{n} \right) \int_{D_k^n} d\mu \\
&= \left(\frac{k+1}{n} \right) \mu(D_k^n) \\
&= \frac{k}{n} \mu(D_k^n) + \frac{1}{n} \mu(D_k^n) \\
&\leq \int_{D_k^n} f d\mu + \frac{1}{n} \mu(D_k^n)
\end{aligned}$$

Where the last step comes from the inequality above. So far, we have

$$\int_{D_k^n} f^*(\omega)d\mu \leq \int_{D_k^n} f d\mu + \frac{1}{n}\mu(D_k^n)$$

Summing the above inequality over $k \in \mathbb{Z}$ (possible since $\mathbb{Z}$ is countable and each D_k^n is disjoint for different $k \in \mathbb{Z}$ except for at the endpoints, which form a set of measure zero) we obtain

$$\int_\Omega f^*(\omega)d\mu \leq \int_\Omega f d\mu + \frac{1}{n}\mu(\Omega)$$

Now letting $n \to \infty$ we get

$$\int_\Omega f^*(\omega)d\mu \leq \int_\Omega f d\mu$$

since $\mu(\Omega) < \infty$. To obtain the reverse inequality, simply apply the above argument to $-f$ to obtain

$$\int_\Omega \limsup_{i\to\infty} \frac{-1}{i} f_i(\omega)d\mu \leq \int_\Omega -f d\mu$$

$$\Rightarrow -\int_\Omega \liminf_{i\to\infty} \frac{1}{i} f_i(\omega)d\mu \leq -\int_\Omega f d\mu$$

$$\Rightarrow \int_\Omega f^*(\omega)d\mu \geq \int_\Omega f d\mu$$

It now follows that

$$\int_\Omega f^*(\omega)d\mu = \int_\Omega f d\mu$$

which completes the proof of the Ergodic Theorem. □

8. Applications of Ergodic Theory in Thermodynamics

Ergodic Theory has a place in thermodynamics (and in particular, statistical thermodynamics) when one is examining the properties of an example and wanting to compare the ensemble average to the time average of a particular particle. For example, if one considers a box full of gas particles and one wants to measure the temperature of the gas, most individuals will want just a single number that is representative of the temperature of every gas particle in the box. Indeed, if one inserts a thermometer into the box, then one will get but a single number for the temperature. Physically, however, every particle in the box can have different kinetic energy, and hence varying temperatures. It is only the time average of particles hitting the thermometer at varying kinetic energies that an individual is reading. Moreover, we seem to be comfortable in assuming that this time-average reading is exactly the same as the ensemble average (i.e., the average kinetic energy of every particle in the box) without ever referring to the concept of Ergodicity! The same can be said about the measurement of pressure. When it comes to other thermodynamic variables like enthalpy and entropy, the situation very quickly becomes hard to handle without the use of any mathematical tools. Once non-ideal fluids begin to get introduced, we are forced to look for something better than "intuition". What we developed in the preceding sections is exactly this, and we will illustrate the usefulness of some of these concepts in this short excursion into statistical mechanics.

The first axiom that is assumed when dealing with thermodynamic systems is that they're equivalent to measure spaces. The second part of the definition/axiom is that they have an associated function that lets us compute density (intuitively speaking, since we can have density at a point we require some way of talking about density when we consider a finite number of particles).

Definition 8.1. A *thermodynamic system* is a measure space $(\Omega, \mathcal{F}, \mu)$ where Ω is called the *phase space*. Every thermodynamic system has a characteristic *density* which we will denote by $f : \Omega \to \Omega$.

Observe that the density of a thermodynamic system induces a measure given by $\mu_f(A) = \int_A f(\omega)d\mu(\omega)$ for $A \in \Omega$. The fact that this is indeed a measure comes from all the hard work we put in to measure theory early on in the semester.

Next we will be interested in describing how a thermodynamic system evolves with time. For this, we simply apply a function each time we wish to know how the phase space changes.

Definition 8.2. A *dynamical law* is a map $S_t : (\Omega, \mathcal{F}, \mu) \to (\Omega, \mathcal{F}, \mu)$. t may either be continuous (real-valued) or discrete (integer valued). A *dynamical system* is a sequence $\{S_t\}_{t\in\mathbb{R}}$ or $\{S_t\}_{t\in\mathbb{Z}}$ (depending on whether we have a continuous or discrete dynamical law).

Now notice that if we have an ensemble of N particles, and we have all intensive variables specified according to the Gibbs Phase Rule (see [4] for a lengthy treatment) then we have a thermodynamic system. Hence, we should have some way of computing each thermodynamic variable using this system. In the case of entropy, we have

Definition 8.3. The *enropy*, $S(f)$ of a thermodynamic system $(\Omega, \mathcal{F}, \mu)$ is given by

$$S(f) = -\int_{\Omega} f(\omega) \log f(\omega) d\mu(\omega)$$

Where "log" is the usual natural logarithm.

The idea of a Canonical ensemble may then be defined, and then the exploration into thermodynamics may be taken further. In order to avoid bringing in too much physics in this paper, we stop at this point. In order to proceed further, however, the reader is advised to first read about the applications of Ergodic Theory to Markov Chains, and then pursue the study of Thermodynamics through one of the references listed at the end.

References

[1] R. B. Ash and C. A. Doléans-Dade, *Probability and measure theory.* Academic Press, 2000.
[2] P. Walters, *An introduction to ergodic theory.* Springer Science & Business Media, 2000, vol. 79.
[3] M. C. Mackey. The second law of thermodynamics: Comments from ergodic theory. Accessed: December 18, 2018. [Online]. Available: https://www.mcgill.ca/mathematical-physiology-lab/files/mathematical-physiology-lab/karpacz.pdf
[4] M. D. Koretsky, *Engineering and chemical thermodynamics.* Wiley New York, 2004, vol. 2.
[5] J. S. Rosenthal, *A first look at rigorous probability theory.* World Scientific Publishing Company, 2006.
[6] R. Becker, "Thermodynamics of chaotic systems. c beck and f schlogl," *APPLIED MECHANICS REVIEWS*, vol. 48, pp. B166–B166, 1995.
[7] M. C. Mackey. (2001, Sep.) Microscopic dynamics and the second law of thermodynamics. Accessed: December 18, 2018. [Online]. Available: https://www.mcgill.ca/mathematical-physiology-lab/files/mathematical-physiology-lab/2001_mcm__microscopic_dynamics_2nd_law_thermodynamics_newfinal.pdf
[8] A. Ray. (2015, Mar.) Chapter 05: Introduction to ergodic theory. Accessed: December 18, 2018. [Online]. Available: https://www.mne.psu.edu/ray/ME_Math_Phy_597_ChaoticDynamics/LaTexNotes/chapter05_ErgodicSystems/chapter05_Ergodicity.pdf

Email address: ahmad.alkadriarafa@ucalgary.ca