THE BIRKHOFF ERGODIC THEOREM

EZRA FURTADO-TIWARI

Abstract.
In this paper we build up Ergodic theory, using tools from measure theory and analysis. We then go on to introduce the Birkhoff Ergodic Theorem and its applications.

1. Introduction

Ergodic theory allows us to study the properties of transformations on dynamic systems.It has applications in mathematical physics, number theory, analysis, probability, statistics, and multiple other fields.

In this paper, we will look into the Birkhoff Ergodic Theorem, which states that a transformation on a set X is ergodic if and only if the limit of the average number of times that an iteration of a measurable transformation T lands in a set A is equivalent to the size of the set for each set A and point $x \in X$ which is not in a null set of X. In simpler words, it means on that on average, an ergodic process acting on a set X should visit a subset $X \subseteq A$ with size k approximately k times.

We will make this notion rigorous in the rest of the paper, using techniques from measure theory and analysis to classify our sets.

2. The Lebesgue Outer Measure

In order to understand how a process can act on a set, we first need to rigorously define what the size of a set is. First, we consider the sets $A, B \in \mathbb{R}$ such that $A = (0, 1)$ and $B = [0, 1]$. At first glance, the sizes of these sets appear to be different, since A is open while B is closed. However, the difference in the sizes of these sets is infinitesmally small, because they differ only by two points (which are of infinitesmally small length).

At this point we may choose to consider the length of the open set for both the open and closed sets. More generally, we may try to approximate the length of the closed set using the length of the union of multiple open sets. To do this, we introduce the Lebesgue outer measure.

Definition 2.1. We define the *Lebesgue outer measure* of a set $A \in \mathbb{R}$ to be

$$\lambda(A) = \inf \left\{ \sum_{j=1}^{\infty} |I_j| : A \subset \bigcup_{j=1}^{\infty} I_j \right\}$$

where the I_j are a sequence of bounded intervals.

There are a few properties of the Lebesgue outer measure that help us to understand it better.

Proposition 2.2. *The I_j can be assumed to be open intervals.*

Date: July 10, 2022.

Proof. We provide a bounding argument for this. Let $\alpha(A)$ be the outer measure obtained using open sets. Then we have $\alpha(A) \geq \lambda(A)$. Next, we construct a sequence $\{K_j\}$ of open intervals such that $I_j \subseteq K_j$ and

$$|K_j| < |I_j| + \frac{\epsilon}{2^j}$$

for $\epsilon > 0$. Then

$$\sum_{j=1}^{\infty} |K_j| < \sum_{j=1}^{\infty} |I_j| + \sum_{j=1}^{\infty} \frac{\epsilon}{2^j},$$

and since

$$\sum_{j=1}^{\infty} |I_j| + \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \sum_{j=1}^{\infty} |I_j| + \epsilon,$$

we find that

$$\alpha(A) = \inf\left\{\sum_{j=1}^{\infty} |K_j|\right\} \leq \inf\left\{\sum_{j=1}^{\infty} |I_j| + \epsilon\right\} = \lambda(A).$$

Combining these inequalities gives $\lambda(A) = \alpha(A)$, so it is sufficient to assume that the I_j are open intervals. ■

Using this fact, we can rephrase the Lebesgue outer measure of a set A as the combined lengths of open sets approximating A.

In addition, we can show that the sets I_j can be made infinitesmally small.

Proposition 2.3. *Let $\epsilon > 0$ be a constant. Then $|I_j| < \delta$ for all $j \geq 1$.*

Proof. This follows from the fact that we can split the I_j into smaller subintervals and preserve the measure. ■

Our last property concerns the union of multiple sets.

Proposition 2.4 (Countable Subadditivity)**.** *For any sequence $\{A_j\}$ of sets, we have*

$$\lambda\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \lambda\left(A_j\right).$$

Proof. Using a similar technique to the one that we used in Proposition 2.2, we let $\{I_{j,k}\}$ be a sequence of intervals such that

$$A_j \subset \bigcup_{k=1}^{\infty} I_{j,k}.$$

Then

$$\sum_{k=0}^{\infty} |I_{j,k}| < \lambda\left(A_j\right) + \frac{\epsilon}{2^j}.$$

Then

$$A = \bigcup_{j=1}^{\infty} A_j \subset \bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} I_{j,k},$$

so summing over j and k gives

$$\lambda(A) \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |I_{j,k}| < \epsilon + \sum_{j=0}^{\infty} \lambda\left(A_j\right).$$

Figure 1. The Cantor set.

Since our proof holds for all $\epsilon > 0$, our inequality also holds. ■

With these definitions and properties, we can look further into a few interesting sets and their outer measures.

3. Lebesgue Measurable Sets

We might wonder what happens when a set isn't empty, but still has outer measure 0. We call these sets null sets.

Definition 3.1. A *null set* N is a set of intervals $\{I_j\}$ such that for any $\epsilon > 0$, we have

$$N \subset \bigcup_{j=1}^{\infty} I_j$$

and

$$\sum_{j=1}^{\infty} |I_j| < \epsilon.$$

Let's look at an example of a null set, called the Cantor set.

Definition 3.2. Let $F = [0, 1]$ and

$$\{G_j\} = \bigcup_{k=1}^{3^{n-1}-1} \left(\frac{3k+1}{3^n}, \frac{3k+2}{3^n} \right).$$

We define the *Cantor set* to be

$$K = F \setminus \left(\bigcup_{j=1}^{\infty} G_j \right).$$

While the Cantor set is a null set, it also contains uncountably many elements. This gives us an idea of how strong results dependent on the Lebesgue outer measure can be; while we might be able to guarantee a result for most sets, it is possible that there will be an uncountable number of elements which do not satisfy it. A visualization of the unit interval excluding the unions of the first few G_i is shown in Figure 1.

Definition 3.3. A set A is *Lebesgue measurable* if there exists an open set $G = G_\epsilon$ such that $A \subset G$ and

$$\lambda(G \setminus A) < \epsilon$$

for all $\epsilon > 0$.

In other words, a set is Lebesgue measurable if it is well approximated by open sets.

Proposition 3.4. *Any open or null set is measurable.*

Proof. First we consider open sets. If A is open, then we can let $G = A$, so clearly A is Lebesgue measurable. Now, assume that A is a null set. Then we have

$$A \subset \bigcup_{j=1}^{\infty} I_j,$$

so we let $G = \bigcup_{j=1}^{\infty} I_j$. Then

$$\lambda(G \setminus A) \leq \lambda(G) \leq \sum_{j=1}^{\infty} |I_j| < \epsilon,$$

so A is measurable. ■

Proposition 3.5. *If $\{A_j\}$ is a sequence of measurable sets, then*

$$A = \bigcup_{j=1}^{\infty} A_j$$

is also measurable.

Proof. Let $\epsilon > 0$. Then for each n there exists some set G_n such that

$$\lambda(G_n \setminus A_n) < \frac{\epsilon}{2^n}.$$

Letting $G = \bigcup_{j=1}^{\infty} G_n$, we find that

$$\lambda(G \setminus A) \leq \lambda\left(\bigcup_{n=1}^{\infty} (G_n \setminus A_n)\right) \leq \sum_{n=1}^{\infty} \lambda(G_n \setminus A_n) < \epsilon.$$

Thus we find that A is a measurable set. ■

Proposition 3.6. *If A is a bounded closed set and B is an open set such that $A \subset B$,*

$$\lambda(B \setminus A) = \lambda(B) - \lambda(A).$$

Proof. We first note that by countable subadditivity, we have

$$\lambda(B \setminus A) \geq \lambda(B) - \lambda(A).$$

Then since $B \setminus A$ is open, there exist open intervals $\{I_j\}$ such that

$$B \setminus A = \bigcup_{j=1}^{\infty} I_j.$$

Then for $N \geq 1$, we have

$$B \subset \left(\bigcup_{j=1}^{N} I_j\right) \sqcup A.$$

Thus we find that

$$\lambda(B) \geq \lambda\left(\bigcup_{j=1}^{N} I_j\right) + \lambda(A)$$
$$= \sum_{j=1}^{N} \lambda\left(I_j\right) + \lambda(A).$$

Taking the limit as $N \to \infty$ gives

$$\lambda(B) \geq \lambda(B \setminus A) + \lambda(A),$$

so we are done. ■

Proposition 3.7. *All closed sets are measurable.*

Proof. Let B be some closed set, and assume that it is bounded. Then for some $\epsilon > 0$, there exist sets I_j such that

$$B \subset \bigcup_{j=1}^{\infty} I_j$$

and

$$\lambda\left(\bigcup_{j=1}^{\infty} I_j\right) < \lambda(B) + \epsilon.$$

Then let

$$I = \bigcup_{j=1}^{\infty} I_j,$$

so that I is an open set. Then

$$\lambda(I \setminus B) = \lambda(I) - \lambda(B) < \epsilon,$$

so B is a measurable set.

In the case of unbounded sets B, we let $B_n = B \cap [-n, n]$. Then

$$B = \bigcup_{n=-\infty}^{\infty} B_n$$

is a countable union of closed sets, so it is also measurable. ■

4. σ-Algebras and Measure Spaces

Definition 4.1. Let X be a nonempty set. A *σ-algebra* on X is a collection $\mathcal{S}$ of subsets of X satisfying the following properties:

(1) $\mathcal{S}$ is nonempty.
(2) If $A \in \mathcal{S}$, then $X \setminus A \in X$.
(3) If $A_n \in \mathcal{S}$ for $n \geq 1$, then

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{S}.$$

In other words, a σ-algebra is a nonempty collection of subsets which is closed under countable unions and complements.

Definition 4.2. A *measure* on a σ-algebra $\mathcal{S}$ is a function $\mu : \mathcal{S} \to [0, \infty)$ such that $\mu(\emptyset) = 0$ and

$$\mu\left(\bigsqcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu\left(A_j\right).$$

Measures generalize the notion of length. For instance, if μ is a measure on a set X, then $\mu' = 2 \cdot \mu$ is also a measure. However, the general properties of length must be preserved, meaning that an empty set must have length 0, and the sum of the lengths of some collection of disjoint sets must be equal to the length of the union.

Definition 4.3. A *measure space* is a triple $(X, \mathcal{S}, \mu)$ where X is a nonempty set, $\mathcal{S}$ is a σ-algebra on X, and μ is a measure on $\mathcal{S}$. A *probability space* satisfies $\mu\left(X\right) = 1$.

Measure spaces allow us to rigorously define the spaces which we can use transformations on.

Definition 4.4. Let X be a finite interval. We let $\mathfrak{L}(X)$ be the set of all Lebesgue-measurable sets in X.

This is a useful construction, since it allows us to use the Lebesgue outer measure as a measure on any finite interval. As we will observe shortly, this is a σ-algebra on X, so we can treat $(X, \mathfrak{L}(X), \lambda)$ as a measure space.

Proposition 4.5. *Let X be a finite interval. Then $\mathfrak{L}(X)$ is a σ-algebra on X.*

Proof. Clearly $\mathfrak{L}(X)$ is nonempty. Thus it suffices to prove that $\mathfrak{L}(X)$ is closed under complements and countable unions. Clearly the countable union of Lebesgue measurable sets is also Lebesgue measurable, so we need to prove that if A is a Lebesgue measurable set, then $(X \setminus A)$ is also Lebesgue measurable.

To prove this, we notice that $(X \setminus A)$ is a union of multiple intervals, which we proved earlier are Lebesgue measurable. Since the countable union of Lebesgue measurable sets is also Lebesgue measurable, $\mathfrak{L}(X)$ is closed under complements and is thus a σ-algebra. ∎

Definition 4.6. A *canonical nonatomic Lebesgue measure space* is the measure space $(X, \mathfrak{L}(X), \lambda)$ where λ is the Lebesgue outer measure.

Proposition 4.7. *Let $(X, \mathcal{S}, \mu)$ be a measure space. Then if $\{A_n\}$ is a sequence of measurable sets satisfying*

$$A_n \subset A_{n+1}$$

for $n \geq 1$, then

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

Proof. First, we note that

$$\bigcup_{n=1}^{\infty} A_n = \bigsqcup_{n=1}^{\infty} \left(A_{n+1} \setminus A_n\right).$$

Then we have

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigsqcup_{n=1}^{\infty} \left(A_{n+1} \setminus A_n\right)\right) \\
&= \sum_{n=1}^{\infty} \mu\left(A_{n+1} \setminus A_n\right) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \mu\left(A_{n+1} \setminus A_n\right) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \mu\left(A_{n+1}\right) - \mu\left(A_n\right). \\
&= \lim_{N\to\infty} \mu\left(A_N\right)
\end{aligned}$$

as desired.

■

Proposition 4.8. *Let $(X, \mathcal{S}, \mu)$ be a measure space. If $\{B_n\}$ is a sequence of measurable sets satisfying*

$$B_n \supset B_{n+1}$$

for $n \geq 1$ and $\mu(B_1) < \infty$, then

$$\mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu\left(B_n\right).$$

Proof. First, we let

$$B = \bigcup_{n=1}^{\infty} B_n.$$

Then we find that

$$B_1 = B \sqcup \bigcup_{n=1}^{\infty} \left(B_{n+1} \setminus B_n\right),$$

and thus

$$\begin{aligned}\mu(B_1) &= \mu(B) + \mu\left(\bigcup_{n=1}^{\infty}(B_{n+1} \setminus B_n)\right) \\ &= \mu(B) + \sum_{n=1}^{\infty} \mu\left(B_{n+1} \setminus B_n\right) \\ &= \mu(B) + \sum_{n=1}^{\infty} \mu\left(B_{n+1}\right) - \mu\left(B_n\right) \\ &= \mu(B) + \sum_{n=1}^{\infty} \mu\left(B_{n+1}\right) - \mu\left(B_n\right) \\ &= \mu(B) + \lim_{n\to\infty} \mu(B_1) - \mu\left(B_n\right) \\ &= \mu(B) + \mu(B_1) - \lim_{n\to\infty} \mu\left(B_n\right).\end{aligned}$$

Then we have

$$\mu(B_1) + \lim_{n\to\infty} \mu\left(B_n\right) = \mu(B_1) + \mu(B),$$

and it follows that

$$\lim_{n\to\infty} \mu\left(B_n\right) = \mu(B)$$

as desired. ■

5. Outer Measure in $\mathbb{R}^d$

Definition 5.1. Let X be a nonempty set. An *outer measure* ν is a function from subsets of X to $[0, \infty)$ such that $\nu(\emptyset) = 0$, if $A \subset B$ then $\nu(A) \leq \nu(B)$, and for sequences $\{A_j\}$ of measurable sets,

$$\nu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \nu\left(A_j\right).$$

An outer measure is slightly less strict than a measure, while still adhering to the same general properties. However, instead of countable additivity, the outer measure only needs to satisfy countable subadditivity.

Definition 5.2. The *Lebesgue outer measure* in $\mathbb{R}^d$ is defined as

$$\inf\left\{\sum_{j=1}^{\infty} |I_j|_d : A \subset \bigcup_{j=1}^{\infty} I_j\right\},$$

where I_j are d-rectangles.

This is a natural generalization; for instance, in $\mathbb{R}^3$, we can consider the intersection of the xy and yz planes, and consider rectangular prisms with sides parallel to these two planes. This is equivalent to considering some combination of the analog of the outer measure in $\mathbb{R}^2$ for both planes.

6. Measure-Preserving Transformations

In this section we introduce a special kind of transformation, called a measurable transformation. However, we first need to introduce some prerequisite definitions.

Definition 6.1. An *invertible transformation* is one that is bijective.

In other words, the inverse of the transformation should also be a transformation. Precisely, this means that the transformation is both injective and surjective.

Definition 6.2. Let T be an invertible transformation on A. We let the *pre-image* T^{-1} satisfy

$$T^{-1}(A) = \{x : T(x) \in A\}.$$

We denote the n-th iteration of the pre-image by $T^{-n}(A)$.

In other words, the pre-image is just the inverse of the transformation T.

Definition 6.3. Let $(X, \mathcal{S}, \mu)$ be a measure space. A transformation $T : X \to X$ is measurable if $T^{-1}(A) \in \mathcal{S}$ for all $A \in \mathcal{S}(X)$.

In other words, the σ-algebra $\mathcal{S}$ is closed under all measurable transformations.

Definition 6.4. An *invertible measurable transformation* is a transformation T such that T and T^{-1} are both measurable.

This means that both T and T^{-1} would be transformations on the same σ-algebra, so we can apply both of these transformations to the same measure space.

Definition 6.5. A transformation is *measure-preserving* if

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

for all $A \in X$. In this case we call μ an *invariant measure* for T.

In other words, a transformation is measure-preserving if it takes a unit interval to another unit interval.

7. Recurrence

Definition 7.1. Let $(X, \mathcal{S}, \mu)$ be a measure space. A transformation T on X is said to be *recurrent* if for every measurable set A of positive measure, there exists a null set $N \subset A$ such that for every $x \in A \setminus N$, there is a positive integer $n = n(x) > 0$ such that

$$T^n(x) \in A.$$

In other words, the transformation is recurrent if at some point it returns points $x \in A$ to A.

Lemma 7.2. *Let $(X, \mathcal{S}, \mu)$ be a measure space, and let $T : X \to X$ be a recurrent measure preserving transformation. Then for every set A of positive measure, there exists a null set N such that for all $x \in A \setminus N$ there is an increasing sequence $n_i > 0$ such that*

$$T^{n_i}(X) \in A \setminus N$$

for $i \geq 1$.

Proof. There exists a null set N_1 such that for $x \in A \setminus N_1$, there exists $n_1 = n_1(x) > 0$ such that $T^{n_1}(x) \in A$. Then let

$$N = \bigcup_{j=1}^{\infty} T^{-k}(x),$$

and since $\mu\left(T^{-k}(x)\right) = 0$ for all x, we find that $\mu(N) = 0$ as well. Then for $x \in A \setminus N \subset A \setminus N_1$ there exists $n_2(x) > 0$ such that $T^{n_2(x)}(x) \in A \setminus N$. Then if we let $n_3 = n_1 + n_2$, then $T^{n_3(x)} = T^{n_1(x)} \in A \setminus N$, and we can repeat this process to construct the n_i. ■

Definition 7.3. A measure preserving transformation is said to be *conservative* if for any set A of positive measure there exists some $n > 0$ satisfying

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

Lemma 7.4. *Let $(X, \mathcal{S}, \mu)$ be a measure space. A measure-preserving transformation T is recurrent if and only if it is conservative.*

Proof. First, we note that T is recurrent if and only if

$$\mu\left(A \setminus \bigcup_{n=1}^{\infty} T^{-n}(A)\right) = 0$$

for all sets A such that $\mu(A) > 0$. Now assume that T is recurrent. Then for sets A such that $\mu(A) > 0$, we have

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

Then for some positive integer n we must have $\mu(A \cap T^{-n}(A)) > 0$.

For the converse, we let A be a measurable set such that $\mu(A) > 0$ and

$$B = A \setminus \bigcup_{n=1}^{\infty} T^{-n}(A).$$

Then if B is a set of positive measure, there is some integer such that

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

which is a contradiction since this means that there exists $x \in B$ such that $T^{-n}(x) \in B$. Thus $\mu(B) = 0$ and thus T is recurrent. ■

The Poincaré Recurrence Theorem was perhaps the first theorem discovered in Ergodic Theory, which was proven in 1899.

Theorem 7.5 (Poincaré Recurrence)**.** *Let $(X, \mathcal{S}, \mu)$ be a finite measure space. If $T : X \to X$ is a finite measure-preserving transformation, then T is recurrent.* [Sil07]

Proof. It is sufficient to show that for any measurable set A of positive measure, there exists some $n > 0$ such that $\mu(A \cap T^{-n}(A)) > 0$. Then assume that there is no set A satisfying this.

Then for integers x, y such that $x \neq y$, we can let $x = n + y$ for $n > 0$, and thus we have

$$\begin{aligned}\mu\left(T^{-x}(A) \cup T^{-y}(A)\right) &= \mu\left(T^{-n-y}(A) \cup T^{-y}(A)\right)\\ &= \mu\left(T^{-y}\left(T^{-n}(A) \cup (A)\right)\right)\\ &= \mu\left(T^{-n}(A) \cup (A)\right)\\ &= 0.\end{aligned}$$

Then for any integers p, q, we have that $T^{-p}(A) \cap T^{-q}(A) = \emptyset$, so

$$\begin{aligned}\mu\left(\bigcup_{n=1}^{\infty} T^{-n}(A)\right) &= \sum_{n=1}^{\infty} \mu\left(T^{-n}(A)\right)\\ &= \sum_{n=1}^{\infty} \mu(A)\\ &= \infty.\end{aligned}$$

This is a contradiction to the assumption that X is finite, so we have

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

for some positive integer n. ■

Note that the argument we used is similar to the pigeonhole principle, but it applies for infinite sets as well.

Definition 7.6. We consider two functions f and g to be *almost equal* if

$$\mu\{x : f(x) \neq g(x)\} = 0.$$

We denote two almost equal functions by

$$f = g \text{ a.e.}$$

Most of the results that we will prove in this paper hold a.e. This means that while we can say the result holds for a general set, there is nothing that we can show about individual elements in these sets.

Definition 7.7. Let $(X, \mathcal{S}, \mu)$ be a σ-finite measure space, and let $T : X \to X$ be a recurrent measure-preserving transformation. Then for every measurable set A there exists some null set $N \subset A$ such that $x \in A \setminus N$, there is an integer $n = n(x) > 0$ such that

$$T^n(x) \in A.$$

We call

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

the *first return time* to A.

In other words, this is the first time that the transformation takes a state back to itself.

Definition 7.8. We let the *induced transformation* T_A satisfy

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

Lemma 7.9. *Let $(X, \mathcal{S}, \mu)$ be a measure space and let $T : X \to X$ be an ergodic transformation. Then if A is a measurable set of positive measure, the transformation T_A is ergodic on A.*

Proof. Since A has positive measure, there exist sets $B, C \in A$ which also have positive measure. Since T is ergodic, there must be $n > 0$ satisfying $\mu(T^n(B) \cap C) > 0$. Then there exists some $x \in B$ such that $T^n(x) \in C$. Then we let n_i satisfy

$$n_1 = n_A(x), n_2 = n_A(T^{n_1}(x)), \ldots,$$

such that k is the first integer such that $T^{n_k}(x) \in C$. Then $T_A^{n_k}(x) = T^n(x) \in C$, so $T^{n_k}(x) \cap B \neq \emptyset$, and thus T is ergodic. ∎

8. Ergodicity

Definition 8.1. Let $(X, \mathcal{S}, \mu)$ be a measure space, and let $T : X \to X$ be a transformation. A subset $A \subset X$ is *positively invariant* if

$$T(x) \in A$$

for $x \in A$.

Definition 8.2. Let $(X, \mathcal{S}, \mu)$ be a measure space, and let $T : X \to X$ be a transformation. A subset $A \subset X$ is *strictly invariant* if

$$T^{-1}(A) = A.$$

Definition 8.3. A set A is said to be *strictly invariant* mod μ if

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

Definition 8.4. Let $(X, \mathcal{S}, \mu)$ be a measure space, and let T be a measure-preserving transformation. T is said to be *ergodic* if when A is a strictly invariant measurable set, either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

Definition 8.5. Let a rational number $x = \frac{p}{q}$ be a *dyadic rational* if $q = 2^n$ for some n.

To understand transformations better, we can look at an example of one, called the Baker's Transformation.

Definition 8.6. The *Baker's transformation* is the transformation

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

Consider a rectangle, split into two regions A and B of equal area. Then the transformation is equivalent to folding the retangle in half after performing a 180 degree twist, and then flattening out the dough. A visualization of two iterations of this process is shown in Figure 2.

We might notice that half of the region which was originally marked region B (the left side of the rectangle) is always occupied by the black region, regardless of the number of transformations we apply to the rectangle. To formally understand this, we first need the notion of a dyadic rectangle.

Definition 8.7. We define a *dyadic rectangle* to be a rectangle whose sides are bounded by dyadic rationals.

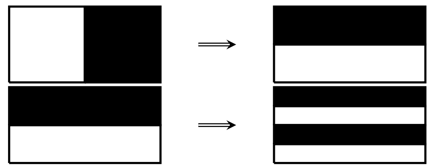

Figure 2. The Baker's Transformation.

In the case of the Baker's transformation, we noticed that the region originally occupied by B is always half occupied by B and half occupied by A. In fact, if we let the rectangle have area 1, half of the area originally occupied by B is equivalent to the products of the original areas of A and B. Transformations like these are called mixing transformations.

Definition 8.8. A transformation T is *mixing* if for any dyadic rectangles A, B, we have

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

A transformation might not be mixing, but there might be some sequence of n_i which seem to satsify a similar property. Thus we introduce the concept of weak mixing.

Definition 8.9. A transformation T is *weakly mixing* if there exists some increasing sequence n_i such that

$$\lim_{i\to\infty} \lambda(T^{n_i}(A) \cap B) = \lambda(A)\lambda(B).$$

9. The Lebesgue Integral

In this section, we address functions which can quantify different properties of our sets. More generally, we consider functions $f : X \to \mathbb{R}$ on measure spaces $(X, \mathcal{S}, \mu)$.

Definition 9.1. Let $(X, \mathcal{S}, \mu)$ be a measure space. A *measurable function* on the space is a function $f : X \to \mathbb{R}$ such that for all $z \in \mathbb{R}$ we have $\{x \in X : f(x) \leq z\} \in \mathcal{S}$.

Not every function is a measurable function; as we will see shortly, the characteristic function of a set is not measurable.

Definition 9.2. Let A be a measurable set. The *characteristic function* $\mathbb{1}_A(x)$ is equal to 1 for $x \in A$, and 0 otherwise.

In other words, the characteristic function of a set tells us whether a value is in a set of not.

With this notion, we can define the Lebesgue integral.

Definition 9.3. We let the *Lebesgue integral* of a characteristic function be

$$\int \mathbb{1}_A \, d\mu = \mu(A).$$

This should naturally be true, since the characteristic function takes on the value 1 if and only if $x \in A$, and takes on 0 otherwise.

To develop the Lebesgue integral for more complicated functions, we have to first simplify functions that are Lebesgue integrable.

Definition 9.4. A function g is a *simple function* if

$$g(x) = \sum_{i=1}^{k} a_i \mathbb{1}_{A_i}$$

for measurable sets $A_1, A_2, A_3, \ldots, A_k$ and real $a_1, a_2, \ldots a_k$.

In other words, if a function is in some set Y of intervals, it will take on the sum of the values indicated by each element $y \in Y$.

Definition 9.5. Assume that a simple function g takes on the values $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_k$. Then we can split the range of the function into disjoint sets E_j satisfying

$$E_j = \{x : g(x) = \alpha_j\}.$$

We let the sum

$$g(x) = \sum_{i=1}^{k} \alpha_j \mathbb{1}_{E_j}$$

be the *canonical representation* of g.

In other words, we can split the range of any simple function into finite disjoint sets, and say that the function $g(x)$ takes on some value α_j when $x \in E_j$.

Definition 9.6. We let the *Lebesgue integral* of a simple function g be

$$\int g \, d\mu = \sum_{j=1}^{n} \alpha_j \mu(E_j),$$

where α_j and E_j are as in the canonical representation of g.

It is important to note here that when $\alpha_j = 0$ and $\mu(E_j) = \infty$, we consider the product $\alpha_j \mu(E_j)$ to be equal to 0.

Notice that this integral seems to agree with the Riemann integral, and in fact, it always does. Notice that in this case, however, we split the range into chunks rather than the domain.

An example of the split range for calculation of the Lebesgue integral of a simple function is shown in 3.

Proposition 9.7. *Let g_1 and g_2 be nonnegative simple functions. Then*

$$\int \alpha g_1 + \beta g_2 \, d\mu = \alpha \int g_1 \, d\mu + \beta \int g_2 \, d\mu.$$

Proof. First, we represent both g_1 and g_2 in canonical form so that

$$g_1(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{E_i}, g_2(x) = \sum_{j=1}^{m} \mathbb{1}_{F_j},$$

where

$$\bigcup_{i=1}^{n} E_i = \bigcup_{j=1}^{m} F_j = X.$$

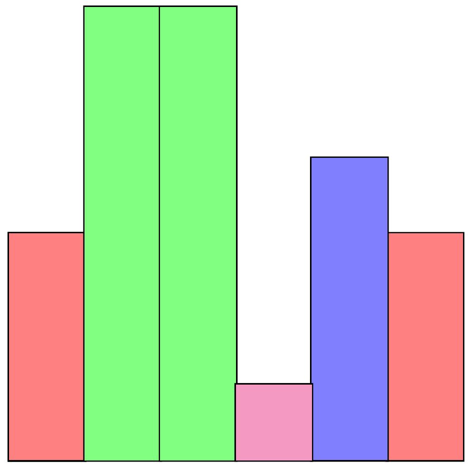

Figure 3. To calculate the Lebesgue integral, the range is partitioned rather than the domain.

Then we note that the sets $E_i \cap F_j = G_k$ are also disjoint and their union is X, so we can write our functions as

$$g_1(x) = \sum_{k=1}^{n \cdot m} y_k \mathbb{1}_{G_k}, g_2(x) = \sum_{k=1}^{n \cdot m} z_k \mathbb{1}_{G_k}.$$

Then we have

$$\begin{aligned}
\int \alpha g_1 + \beta g_2 \, d\mu &= \int \alpha \sum_{k=1}^{n \cdot m} y_k \mathbb{1}_{G_k} + \beta \sum_{k=1}^{n \cdot m} z_k \mathbb{1}_{G_k} \, d\mu \\
&= \int \sum_{k=1}^{n \cdot m} (\alpha y_k + \beta z_k) \mathbb{1}_{G_k} \, d\mu \\
&= \sum_{k=1}^{n \cdot m} (\alpha y_k + \beta z_k) \mu(G_k) \\
&= \sum_{k=1}^{n \cdot m} \alpha y_k \mu(G_k) + \sum_{k=1}^{n \cdot m} \beta z_k \mu(G_k) \\
&= \alpha \sum_{k=1}^{n \cdot m} y_k \mu(G_k) + \beta \sum_{k=1}^{n \cdot m} z_k \mu(G_k) \\
&= \alpha \int s_1 \, d\mu + \beta \int s_2 \, d\mu.
\end{aligned}$$

This completes the proof. ■

Proposition 9.8. *Let g_1 and g_2 be nonnegative simple functions such that $g_1 \leq g_2$. Then*

$$\int g_1 d\mu \leq \int g_2 \, d\mu$$

Proof. Let the G_k be defined as in Proposition 9.7. Then we have

$$s_1(x) = \sum_{k=1}^{n\cdot m} y_k \mathbb{1}_{G_k}$$

and

$$s_2(x)\,d\mu = \sum_{k=1}^{n\cdot m} z_k \mathbb{1}_{G_k}.$$

Now, consider some G_k. for $x \in G_k$, we have

$$s_1(x) = y_k, s_2(x) = z_k.$$

However, we note that $y_k \leq z_k$. Then the respective Lebesgue integrals are

$$\int s_1\,d\mu = \sum_{k=1}^{n\cdot m} y_k \mu(G_k)$$

and

$$\int s_2\,d\mu = \sum_{k=1}^{n\cdot m} z_k \mu(G_k).$$

Since

$$y_k \mu(G_k) \leq z_k \mu(G_k),$$

we have

$$\int s_1\,d\mu \leq \int s_2\,d\mu$$

as desired. ■

Definition 9.9. Let f be a nonnegative function. Then

$$\int f\,d\mu = \sup\left\{\int g\,d\mu : g \text{ is simple and } 0 \leq g \leq f\right\}.$$

Essentially, we approximate f from below using simple functions, and then take the upper bound of the set of approximations, in a similar way to the way that we defined the Lebesgue outer measure.

We note here that the Lebesgue integral always agrees with the Reimann integral, for any nonnegative measurable function. We might notice that as we approach the Lebesgue integral from below the function, the rectangles which we use to approximate it should naturally become infinitely small, as in the Reimann integral. It is also important to note that the Lebesgue integral is defined for any nonnegative function (even those which are not bounded from above). This allows us to avoid using improper integrals to calculate integrals for functions such as $f(x) = \frac{1}{x}$.

Theorem 9.10 (Monotone Convergence)**.** *Let $f_1 \leq f_2 \leq f_3 \leq \ldots \leq f_n \leq \ldots$ be a sequence of nonnegative functions. If*

$$f(x) = \lim_{n\to\infty} f_n \ a.e,$$

Then

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

Proof. Notice that since $f_n \leq f$, we have

$$\begin{aligned}\lim_{n\to\infty} \sup \int f_n \, d\mu &\leq \int f \, d\mu \\ &\leq \int \lim_{n\to\infty} f \, d\mu = \int f \lim_{n\to\infty} \inf \, d\mu \\ &\leq \lim_{n\to\infty} \inf \int f \, d\mu.\end{aligned}$$

Then by the Squeeze Theorem we have

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu$$

as desired. ∎

Definition 9.11. Let f be an arbitrary function. We define the functions

$$f^+(x) = \begin{cases} f(x) & f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f^-(x) = \begin{cases} -f(x) & f(x) \leq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We let

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

Naturally, this makes sense. We take the regions in which f is negative, multiply the function in these regions by -1, turning them into nonnegative regions. We then take the integral of these regions as we did for any other nonnegative function, and then multiply this integral by -1 again, almost as if factoring out the -1 from the original integral.

10. The Birkhoff Ergodic Theorem

Theorem 10.1 (Birkhoff Ergodic Theorem)**.** *Let $(X, \mathcal{S}, \mu)$ be a measure space and let T be an ergodic transformation. Then if $f : X \to \mathbb{R}$ is a Lebesgue integrable function,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f\left(T^i(x)\right) = \int f \, d\mu. \quad \text{[Sil07]}$$

There are two parts of the proof that are necessary. The first part is to prove that the limit exists, and the second part of the theorem is to prove that the limit is in fact equal to the Lebesgue integral.

Definition 10.2. We first introduce some preliminary definitions. We define

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

and also let

$$f_*(x) = \lim_{n\to\infty} \inf \frac{1}{n} f_n(x)$$

and

$$f^*(x) = \lim_{n\to\infty} \sup \frac{1}{n} f_n(x).$$

We are now ready to begin our proof of the ergodic theorem.

Proof of the Ergodic Theorem. Let

$$A = \left\{ x : f_*(x) < \int f \, d\mu \right\}.$$

We will show that $\mu(A) = 0$. Assume that $\mu(A) > 0$. Then

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

Then we let

$$B_r = \left\{ f_*(x) < r < \int f \, d\mu \right\}.$$

Then there exists some rational number r such that $\mu(B_r) > 0$. Since T is ergodic, we have $\mu(B_r) = 1$. Then let

$$E_p^r = \{x : \frac{1}{n} \sum_{i=0}^{n-1} f\left(T^i(x)\right) \geq r, \, 1 \leq n \leq p.$$

Since $\mu(B_r) = 1$, we have

$$\mu\left(\bigcup_{p=1}^{\infty} E_p^r\right) = 0.$$

Thus $\lim_{n\to\infty} \mu\left(E_n^r\right) = 0$ and therefore

$$\int f \, d\mu \leq r,$$

which is a contradiction. Thus we have $\mu(A) = 0$. Then we have

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

Applying this expression to $-f$ gives

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

so that

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

Then we have

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

as desired. ∎

11. Normal Numbers

Definition 11.1. Let x be a real number such that $x \in [0, 1)$. We let

$$M(x, b) = x \pmod 1_b = b \cdot x - \cdot \lfloor b \cdot x \rfloor .$$

Definition 11.2. We let a real number $x \in [0, 1)$ be *normal* in base b if for every $y \in \{0, 1, \ldots, b-1\}$, we have

$$\frac{1}{n} \lim_{n\to\infty} \sum_{i=0}^{n-1} M^i(x, b) \mathbb{1}_{[\frac{y}{b}, \frac{y+1}{b})} = \frac{1}{b}.$$

Theorem 11.3. *Almost every number $x \in [0, 1)$ is normal in every base b.* [KD02]

Proof. Once we show that the transformation $M(x, b)$ is measure preserving, this result follows as a straightforward application of the ergodic theorem. Here we note that a transformation is ergodic if when A is a strictly invariant measurable set, then $\mu(A^c) = 0$. This is true in the case of $M(x, b)$, so it is measure preserving. Thus the result holds. ■

While we cannot prove anything about an individual number like π or e, we can see that the result should hold for most numbers by looking at an example.

Consider the number

$$X = 0.012345678910111213\ldots.$$

We know that this is normal in base 10, and for the sake of approximation, we consider truncating this number after 99. We can consider the number in a few other bases.

First, we convert X to binary, and then check the number of 1s in the representation of the number. Truncating this at different lengths gives the results in the table shown in Figure 1.

100	46
150	67
200	94.

Table 1. Truncating the binary representation at 100, 150, and 200 digits gives the following results for the number of 1s.

Applying a similar strategy for base 3, we check the number of 1s and 2s in the representation of the number. Truncating again at different lengths gives the results in the table shown in Figure 2.

100	23	36
150	43	48
200	60	65.

Table 2. Truncating the ternary representation at 100, 150, and 200 digits gives the following results for the number of 1s and 2s.

12. Continued Fractions

In order to work with continued fractions, we first need to find a transformation representing continued fractions.

Definition 12.1. We define the *Gauß map* to be the transformation

$$T(x) = \begin{cases} \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Proposition 12.2. *The Gauß map is measure-preserving with respect to the measure*

$$\nu(x) = \frac{1}{\log 2} \int \frac{1}{1+x} \, d\lambda(x).$$

Proof. It is sufficient to prove that the map is measurable with respect to all intervals (a, b). First, we can rewrite these intervals as $[0, b) \setminus [0, a)$. Then we need to show that for $b \in (0, 1)$, we have

$$\begin{aligned} \nu\left(T^{-1}([0, b))\right) &= \frac{1}{\log 2} \int_{[0,b)} \frac{1}{x+1} \, d\lambda(x) \\ &= \frac{1}{\log 2} \log(x+1)\Big|_0^b \\ &= \frac{1}{\log 2} \log(b+1). \end{aligned}$$

Note that we have

$$T^{-1}([0, b)) = \bigsqcup_{n=1}^{\infty} \left[\frac{1}{n+b}, \frac{1}{n}\right) \mod \lambda.$$

Then we have

$$\begin{aligned} \nu\left(T^{-1}([0, b))\right) &= \frac{1}{\log 2} \int_{\bigsqcup_{n=1}^{\infty} \left[\frac{1}{n+b}, \frac{1}{n}\right)} \frac{1}{x+1} \, d\lambda \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{\left[\frac{1}{n+b}, \frac{1}{n}\right)} \frac{1}{x+1} \, d\lambda \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \log\left(\frac{1+\frac{1}{n}}{1+\frac{1}{n+b}}\right) \\ &= \frac{1}{\log 2} \lim_{N\to\infty} \sum_{n=1}^{N} \log\left(\frac{n+1}{n}\right) - \log\left(\frac{n+b}{n+b+1}\right) \\ &= \frac{1}{\log 2} \lim_{N\to\infty} \prod_{n=1}^{N} \log\left(\frac{n+1}{n}\right) \cdot \log\left(\frac{n+b}{n+b+1}\right) \\ &= \frac{1}{\log 2} (b+1) \end{aligned}$$

as desired. ∎

Proposition 12.3. *Let $x \in [0,1)$ be a real number and let*

$$[a_0, a_1, a_2, \ldots, a_n, \ldots]$$

be the coefficients of its continued fraction representation. Then

$$a_{n+1} = \left\lfloor \frac{1}{T^n(x)} \right\rfloor,$$

where $T(x)$ is the Gauß map.

This intuitively makes sense, since when computing the continued fraction of a number, we repeatedly remove the whole part of the number and perform the transformation on the fractional part that we have left.

Theorem 12.4. *Let x be a real number such that $x \in [0,1)$, and let $[a_0, a_1, a_2, \ldots, a_n, \ldots]$ be its continued fraction representation. Then*

$$\mathbb{P}(a_i = k) = \frac{1}{\log 2} \log\left(\frac{(k+1)^2}{k(k+2)}\right). \text{ [Hin15]}$$

Proof. Applying the ergodic theorem with $f = \mathbb{1}_{\left(\frac{1}{k+1}, \frac{1}{k}\right]}$, we have

$$\begin{aligned} \mathbb{P}(a_i = k) &= \frac{1}{\log 2} \int_{\left(\frac{1}{k+1}, \frac{1}{k}\right]} \frac{1}{1+x} \, d\lambda \\ &= \frac{1}{\log 2} \left(\frac{(k+1)^2}{k(k+2)}\right). \end{aligned}$$

■

Thus we have ≈ 41.5037% 1s, ≈ 16.9925% 2s, ≈ 9.31094% 3s, ≈ 5.88937% 4s, ≈ 4.0642% 5s, and so on.

13. Conclusion

Ergodic theory has many applications, particularly in probability on infinite spaces. Using ergodic theory, we can investigate many more related problems, such as "What is the probability that an arbitrary power of 2 starts with a 7?". Similar problems to this one are addressed further in [Sil16], and the reader is encouraged to consult this resource to explore this topic. Ergodic theory also has applications in many other fields, and open problems include applications in Ramsey theory, among others.

References

[Hin15] Robert Hines. Ergodic theory of simple continued fractions. 2015.

[KD02] Cor Kraaikamp Karma Dajani. *Ergodic Theory of Numbers*. Mathematical Association of America, 2002.

[Sil07] Cesar E. Silva. *Invitation to Ergodic Theory*. American Mathematical Society, 2007.

[Sil16] Cesar E. Silva. What is an ergodic transformation? *Notices of the American Mathematical Society*, 2016.