THE BIRKHOFF ERGODIC THEOREM WITH APPLICATIONS

DAVID YUNIS

Abstract. The Birkhoff Ergodic Theorem is a result in Ergodic Theory relating the spatial average of a function to its ”time” average under a certain kind of transformation. Though dynamics and Ergodic Theory seem at first removed from Number Theory, it turns out there are many basic applications that are nigh-immediate results of this theorem.

Contents

1. Introduction

What follows is a brief foray into Measure Theory and Ergodic Theory, which is like a study of the indivisible systems in measure theory. Ergodic systems are the measurable units that cannot be broken down further. Throughout this exploration I will give a proof of the Birkhoff Ergodic Theorem, and develop some seemingly unrelated and relatively surprising applications of it. Let’s start with its formal statement.

Theorem. (Birkhoff Ergodic Theorem): *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. For any $f \in \mathscr{L}^1_\mu$,*

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

converges almost everywhere to a T-invariant function $\bar{f} \in \mathscr{L}^1_\mu$, where

$$\int f d\mu = \int \bar{f} d\mu,$$

and if T is ergodic,

$$\int f d\mu = \bar{f}.$$

Of course there is a lot of terminology that remains to be defined, else I wouldn't have a paper, but the important thing to focus on is that there is a very nice way to relate what looks like the time average of any given function under a special transformation with its spatial average. Let's now begin at a hopefully more sensible beginning.

2. Measure Theory

2.1. **Basic Definitions.** We have a sort of intuitive sense of how large things are. For example, we feel in some sense that the interval $[0, 1]$ is bigger than $[0, \frac{1}{2}]$ or $\mathbb{Q} \cap [0, 1]$ (certainly when drawn one uses more ink than another). To make this intuition mathematically precise we introduce the definition of a *measure* on a space.

Definition 2.1. (Measure): A *measure* μ is a map $\mu : \mathscr{B} \to \mathbb{R} \cup \{\infty\}$, where $\mathscr{B}$ is a σ-algebra over a space X, such that for $B \in \mathscr{B}$,

- $\mu(\varnothing) = 0$.
- $\mu(B) \geq 0$.
- $\mu\left(\bigsqcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i)$, where $\{B_i\}_{i\in\mathbb{N}}$ are pairwise disjoint.

We also define what a σ-algebra is.

Definition 2.2. (σ-algebra): A set $\mathscr{B} \subseteq \mathbb{P}(X)$ is called a *σ-algebra* if

- $\varnothing \in \mathscr{B}$.
- $A, B \in \mathscr{B} \Rightarrow A \cap B \in \mathscr{B}$.
- $B \in \mathscr{B} \Rightarrow X \setminus B \in \mathscr{B}$.
- $B_1, B_2, ... \in \mathscr{B} \Rightarrow \bigcup_{i=1}^{\infty} B_i \in \mathscr{B}$.

The precise definition of σ-algebra isn't so important (I'll never require its specifics). It's a subset of the power set of X including the empty set and the whole space. We'd like to define a measure on all subsets of X, but in general this isn't possible, thus we restrict attention to smaller σ-algebras. A more specific σ-algebra is the Borel σ-algebra which contains all of the open sets of a given topological space. We will consider this specific example when looking at S^1.

Example 2.3. On $[0, 1]$ we can define the σ-algebra generated by all open subintervals $(a, b), a \leq b$ and the measure $\mu([a, b]) = b - a$. This is the natural measure we think of on intervals, also called the *Lebesgue* measure.

Example 2.4. We could also consider another measure on $[0,1]$, namely the δ_0 measure, where $\delta_0([a,b]) = 1$ if $0 \in [a,b]$ and $\delta_0([a,b]) = 0$ otherwise. One can see this fits the definition of measure.

Definition 2.5. (Probability Space): A triple $(X, \mathscr{B}, \mu)$ is a *finite measure space* if $\mathscr{B}$ is a σ-algebra and μ is a countably additive measure on $\mathscr{B}$ with $\mu(X) < \infty$. If $\mu(X) = 1$ then the triple is a *probability space*.

Note that any finite measure space with $\mu(X) < \infty$ can be scaled to a new measure ν where for $B \in \mathscr{B}$, $\nu(B) = \frac{\mu(B)}{\mu(X)}$. Since $\mu(X)$ is a constant, we see straight from the definition that ν is a valid measure, thus when talking about finite measure spaces it suffices to consider probability spaces. The σ-algebra of a probability space X makes rigorous the idea of a collection of possible "events" on X.

Definition 2.6. (Measure-Preserving): A function $f : X \to Y$ of probability spaces $(X, \mathscr{B}, \mu), (Y, \mathscr{C}, \nu)$ is *measurable* if for $C \in \mathscr{C}, f^{-1}C \in \mathscr{B}$. If we have this property for $T : X \to X$, then T is a *measurable transformation*. Also, we call T *measure-preserving* if for $B \in \mathscr{B}$,

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

For such a T we denote $(X, \mathscr{B}, \mu, T)$ a *measure-preserving system*.

Example 2.7. Consider the form of $S^1 = \mathbb{R}/\mathbb{Z}$, or alternatively, $[0,1]$ with $0 \sim 1$. The rotation on this S^1, $T_\alpha : x \mapsto x + \alpha \mod 1$ preserves the Lebesgue measure.

Example 2.8. Another map on the same characterization of S^1 that is also measure preserving is the doubling map $T : x \mapsto 2x \mod 1$. Note that for any open interval $(a,b) \in \mathscr{B}$, $T^{-1}(a,b) = (\frac{a}{2}, \frac{b}{2}) \cup (\frac{a+1}{2}, \frac{b+1}{2})$ so though the measure of any single open set is doubled, the image of T^{-1} is two intervals of half the length, thus it preserves Lebesgue measure.

The next example will require a bit of set up.

Example 2.9. Consider the set $\{0, 1, ..., n\}$, where the vector $\{p_0, p_1, ...p_n\}$ is the probability of each of these respective events occurring, so $\sum_{i=0}^{n} p_i = 1$. This gives us a general description of an $(n+1)$-sided die. Now suppose we throw that die an infinite amount of times, resulting in the space $X = \{0, 1, ..., n\}^{\mathbb{Z}}$ where a single element is any infinite string of integers in $\{0, 1, ..., n\}$. If we give this set the product topology, we can consider the smallest σ-algebra $\mathscr{B}$ containing all the open sets. Given any $A \subseteq \mathbb{Z}, |A| < \infty$ and a map $a : A \to \{0, 1, ..., n\}$ we define a cylinder set as $A_a = \{x \in X \mid x_i = a(i) \text{ for } i \in A\}$. Now we define a measure μ on X by its definition on cylinder sets: $\mu(A_a) = \prod_{i \in A} p_a(i)$. Consider the transformation $\sigma : X \to X$ that just shifts every element left, so that $\sigma(x)_i = x_{i-1}$. This obviously preserves the measure of all cylinder sets, which generate $\mathscr{B}$ so it is measure-preserving. We call such a σ a *Bernoulli Shift*.

2.2. **Required Results.** Here's a definition of function spaces that we'll be working with time and again.

Definition 2.10. **(L^p Space):** An $\mathscr{L}^p$ *space* on X is the space of functions f on X such that

$$\|f\|_p = \left(\int\limits_X |f|^p d\mu \right)^{\frac{1}{p}} < \infty.$$

The corresponding L^p *space* is the quotient $\mathscr{L}^p / \sim$ where $f \sim g$ if $\|f - g\|_p = 0$. The operator $\|\|_p$ is a norm on the L^p space. In addition, the space $L^\infty(X)$ is the set of functions f on X equipped with the norm

$$\|f\|_\infty = \inf_{C \in \mathbb{R}} \{C \geq 0 \mid |f| < C \text{ almost everywhere}\}$$

mod equivalence as before.

The next result is something that I'll use quite often, usually without even stating it.

Theorem 2.11. *Let $(X, \mathscr{B}, \mu)$ be a finite measure space. For any $1 \leq p \leq q \leq \infty$*

$$L^q(X, \mathscr{B}, \mu) \subseteq L^p(X, \mathscr{B}, \mu).$$

Here are a few other results in analysis that we will need for the rest of this paper. I'll state most of them without proof, but note when I use them.

Theorem 2.12. **Dominated Convergence:** *Let $g : X \to \mathbb{R}$ be an integrable function, i.e. $\int |g| d\mu < \infty$. Let $(f_n)_{n \geq 1}$ be a sequence of measurable real-valued functions which are dominated by g, meaning $|f_n| \leq g$ for all $n \geq 1$, where $\lim\limits_{n \to \infty} f_n = f$ exists almost everywhere. Then f is integrable with*

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

Theorem 2.13. *Let $(X, \mathscr{B}, \mu)$ be a measure space and let $f : X \to \mathbb{R}$ be a measurable function. Then there exists an increasing sequence $(f_n)_{n \in \mathbb{N}}$ of simple functions $f_n = \sum\limits_{i=0}^{n} a_i \chi_{B_i}$ such that $\lim\limits_{n \to \infty} f_n = f$ pointwise for any $f \in L^1_\mu$.*

Now we give an alternate characterization of measure-preserving that we will use again and again in the following pages.

Proposition 2.14. *A measure μ on a probability space X is preserved by $T : X \to X$ if and only if*

$$\int f d\mu = \int f \circ T d\mu \tag{2.15}$$

for all $f \in L^\infty$. In addition, if μ is preserved by T then 2.15 *holds for all $f \in L^1_\mu$.*

Proof. If 2.15 holds then take to $f = \chi_B$ the characteristic function for $B \in \mathscr{B}$. Thus

$$\mu(T^{-1}B) = \int \chi_{T^{-1}B} d\mu = \int \chi_B \circ T d\mu = \int \chi_B d\mu = \mu(B).$$

Now, if T preserves μ then 2.15 holds for any χ_B, so it holds for simple functions $\sum\limits_{i=1}^{n} a_i \chi_{B_i}$. By 2.13 we can take an increasing sequence (f_n) of such simple functions such that $\lim_{n\to\infty} f_n = f$ pointwise for any $f \in L^1_\mu$. Now we see that $(f_n \circ T)$ converges to $f \circ T$. By dominated convergence,

$$\int f d\mu = \lim_{n\to\infty} \int f d\mu = \lim_{n\to\infty} \int f \circ T d\mu = \int f \circ T d\mu.$$

□

The following theorems are more specific in their uses, and it will be noted when they're needed.

Theorem 2.16. Fubini-Tonelli: *Let f be a non-negative, integrable function on the product of two σ-finite measure spaces $(X, \mathscr{B}, \mu)$ and $(Y, \mathscr{C}, \nu)$. Then for almost every $x \in X$ and $y \in Y$,*

$$\begin{aligned} \int_{X\times Y} f(x,y) d(\mu \times \nu)(x,y) &= \int_X \left(\int_Y f(x,y) d\nu(y) \right) d\mu(x) \\ &= \int_Y \left(\int_X f(x,y) d\mu(y) \right) d\nu(x). \end{aligned}$$

Theorem 2.17. Riesz-Fischer: *Let $(X, \mathscr{B}, \mu)$ be a Lebesgue space. For any $1 \leq p < \infty$, the space L^p_μ is a separable Banach space with respect to the $\|\cdot\|_p$ norm. In addition, L^2_μ is a separable Hilbert space.*

3. Ergodicity and The Birkhoff Ergodic Theorem

Stronger than measure preserving is the *Ergodic* map. This kind of map lets us delineate the indivisible elements of measurable dynamical systems. Ergodic systems cannot be broken into further ergodic systems, but normal measure preserving ones can be broken into their ergodic components.

3.1. **Ergodicity and Examples.**

Definition 3.1. (Ergodic): A measure-preserving transformation $T : X \to X$ of a probability space $(X, \mathscr{B}, \mu)$ is *ergodic* if for $B \in \mathscr{B}$,

$$T^{-1}B = B \Rightarrow \mu(B) \in \{0, 1\}.$$

Thus we see that the notion of ergodicity makes rigorous some kind of uniform mixing of a dynamical system. What follows are some familiar examples.

Example 3.2. Consider the rotation map on S^1 given by $T_\alpha : x \mapsto x + \alpha \mod 1$. This map is ergodic with respect to the Lebesgue measure when α is irrational and is not when α isn't. In the case of $\alpha = \frac{1}{2}$, the set $B = [0, \frac{1}{4}] \cup [\frac{1}{2}, \frac{3}{4}]$ has the property $T^{-1}B = B$, and we see $\mu(B) = \frac{1}{2}$. So, generalizing, we see that for $\alpha = \frac{p}{q}$ the union of any q evenly spaced, disjoint intervals strictly contained in S^1 will violate the ergodic definition.

Proof. Now, for $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and for $\epsilon > 0$ we know we can find integers x, y, z with $x \neq y$ such that $|x\alpha - y\alpha - z| < \epsilon$. This means that mod 1 the set of $x\alpha$ is dense in S^1. If we assume that $B \in \mathscr{B}$ is T-invariant then for $\epsilon > 0$ choose a continuous function f such that $\|f - \chi_B\|_1 \leq \epsilon$. Since T_α is measure-preserving and B is

T-invariant we see that $\|f \circ T_\alpha^n - f\|_1 \leq 2\epsilon$ for $n \in \mathbb{Z}$. Combining this with the continuity of f, we get $\|f \circ T_a - f\|_1 \leq 2\epsilon$ for all $a \in \mathbb{R}$. Thus,

$$\begin{aligned}\left\| f(x) - \int f(a)da \right\|_1 &= \int \left| \int f(x) - f(x+a)da \right| dx \\ &\leq \int \int |f(x) - f(x+a)| dadx \\ &\leq 2\epsilon\end{aligned}$$

from the previous step and an application of Fubini's Theorem. So we know that

$$\begin{aligned}\|\chi_B - \mu(B)\|_1 &\leq \|\chi_B - f\|_1 + \left\| f - \int f(a)da \right\|_1 + \left\| \int f(a)da - \mu(B) \right\|_1 \\ &\leq \epsilon + 2\epsilon + \epsilon = 4\epsilon\end{aligned}$$

for $\epsilon > 0$. This means that $\chi_B = \mu(B)$, thus is constant almost everywhere. So $\chi_B = 0$ or 1 almost everywhere, thus $\mu(B) \in \{0, 1\}$, and T_α is ergodic. □

The next two examples require some alternate characterizations of ergodicity which we list here.

Proposition 3.3. *For a measure-preserving system $(X, \mathscr{B}, \mu, T)$, the following are equivalent:*

- *T is ergodic.*
- *For any $B \in \mathscr{B}$, if $\mu(T^{-1}B \triangle B) = 0$, then $\mu(B) = 0$ or $\mu(B) = 1$.*
- *For $f : X \to \mathbb{R}$ measurable, if $f \circ T = f$ almost everywhere, then f is constant almost everywhere.*

We now use a method to prove ergodicity that does not generalize well, but is nonetheless interesting.

Example 3.4. Recall the doubling map on S^1 given by $T : x \mapsto 2x \mod 1$. This map is ergodic with respect to the Lebesgue measure.

Proof. This will be proved via Fourier Analysis. Given some $f \in L^2(S^1)$ with $f \circ T = f$ we have a Fourier expansion $f(x) = \sum_{i \in \mathbb{Z}} a_i e^{2\pi i x}$, where $\|f\|_2^2 = \sum_{i \in \mathbb{Z}} |a_i|^2 < \infty$. Also we have

$$f(x) = \sum_{i \in \mathbb{Z}} a_i e^{2\pi i x} = \sum_{i \in \mathbb{Z}} a_i e^{2\pi i 2x} = f \circ T(x).$$

Hence, $a_i = a_{2i}$ for all i, but this contradicts the condition $\|f\|_2^2 < \infty$ except when $i = 0$. Thus f is constant almost everywhere so by Proposition 3.3, T is ergodic. □

Note that Fourier Series existence is somewhat of a strong condition that isn't always available to us. Thus this method is less general than the method we used to prove irrational rotation is ergodic, which argued more purely from measure theory.

Example 3.5. Recall our Bernoulli Shift map from Example 2.9. It is ergodic on the measure-preserving system defined there.

Proof. Let $B \in \mathscr{B}$ be an invariant set under the shift map σ. Since $\mathscr{B}$ is generated by the cylinder sets we can find a finite union of cylinder sets C such that $\mu(B \triangle C) < \epsilon$ for a fixed $0 < \epsilon < 1$. Thus $\mu(B) < \mu(C) + \epsilon$. Consider a shift of m large enough so that

$$\mu(\sigma^{-m}C \setminus C) = \mu(\sigma^{-m}C \cap X \setminus C) = \mu(\sigma^{-m}C)\mu(X \setminus C) = \mu(C)\mu(X \setminus C)$$

where the last step results from C being a cylinder set. Since B is σ-invariant by assumption, we know that $\mu(B \triangle \sigma^{-1}B) = 0$. So

$$\mu(\sigma^{-m}C \triangle B) = \mu(\sigma^{-m}C \triangle \sigma^{-m}B) = \mu(C \triangle B) < \epsilon,$$

thus, by the triangle inequality, $\mu(\sigma^{-m}C \triangle C) < 2\epsilon$. In addition,

$$\mu(\sigma^{-m}C \triangle C) = \mu(C \setminus \sigma^{-m}C) + \mu(\sigma^{-m}C \setminus C) < 2\epsilon.$$

So we see finally that

$$\begin{aligned} \mu(B)\mu(X \setminus B) &< (\mu(C) + \epsilon)(\mu(X \setminus C) + \epsilon) \\ &= \mu(C)\mu(X \setminus C) + \epsilon\mu(C) + \epsilon\mu(X \setminus C) + \epsilon^2 \\ &< \mu(C)\mu(X \setminus C) + 3\epsilon \\ &< 5\epsilon, \end{aligned}$$

which implies that either $\mu(B) = 0$ or $\mu(B) = 1$, meaning σ is ergodic. □

Ergodic maps have some very special properties which will shortly appear. Before proving the Birkhoff Ergodic Theorem, two general results will be required. The first will be a convergence result regarding the ergodic averages of a function, to be defined. The second will give a result bounding the integral of a function on some exceptional set related to the ergodic averages.

3.2. **The Mean Ergodic Theorem.** This first result characterizes the average convergence of a function under an ergodic transformation.

Theorem 3.6. (Mean Ergodic Theorem): *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. Define $U_T f = f \circ T$. Let $P_T : L^2_\mu \to I$ be the projection operator onto the closed subspace*

$$I = \{f \in L^2_\mu \mid U_T f = f\} \subseteq L^2_\mu.$$

Then for any $f \in L^2_\mu$,

$$\frac{1}{n}\sum_{i=0}^{n-1} U_T^i f \xrightarrow[L^2_\mu]{} P_T f.$$

Proof. We will show that each f decomposes as $f = P_T f + r$ where r is some remainder function. We do this by characterizing the orthogonal complement $I^\perp$. We first show that this set is

$$A = \{U_T f - f \mid f \in L^2_\mu\}.$$

If $f \in I$, then

$$\langle f, U_T g - g\rangle = \langle f, U_T g\rangle - \langle f, g\rangle = \langle U_T f, U_T g\rangle - \langle f, g\rangle = 0,$$

where the last step comes from Proposition 2.14, so $f \in A^\perp$. Now if $f \in A^\perp$, then for all $g \in L^2_\mu$

$$\langle U_T g - g, f\rangle = 0 \Rightarrow \langle U_T g, f\rangle = \langle g, f\rangle,$$

showing $U_T^* f = f$. This means

$$\begin{aligned}\|U_T f - f\|_2^2 &= \langle U_T f - f, U_T f - f\rangle \\ &= 2\|f\|_2^2 - \langle f, U_T^* f\rangle - \langle U_T^* f, f\rangle \\ &= 0,\end{aligned}$$

so $f \in I$. Now we see that $r \in \bar{A}$ and we want to show

$$\frac{1}{n}\sum_{i=0}^{n-1} U_T^i r \xrightarrow[L^2_\mu]{} 0.$$

If $r \in A$, then $r = U_T g - g$, so we see

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} U_T^i (U_T g - g)\right\|_2 = \left\|\frac{1}{n}(U_T^n g - g)\right\|_2 = \frac{1}{n}\|U_T^n g - g\|_2 \to 0.$$

Now we only know $r \in \bar{A}$, so consider a sequence $(r_i = U_T g_i - g_i)$ such that $\lim\limits_{i\to\infty} r_i = r$ in L^2_μ. We thus know that

$$\left\|\frac{1}{n}\sum_{j=0}^{n-1} U_T^j r\right\|_2 \le \left\|\frac{1}{n}\sum_{j=0}^{n-1} U_T^j (r - r_i)\right\|_2 + \left\|\frac{1}{n}\sum_{j=0}^{n-1} U_T^j r_i\right\|_2.$$

Now we fix $\epsilon > 0$ and pick n and i large such that

$$\|r - r_i\|_2 < \epsilon \text{ and } \left\|\frac{1}{n}\sum_{j=0}^{n-1} U_T^j r_i\right\|_2 < \epsilon.$$

By the triangle inequality,

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} U_T^i r\right\|_2 < 2\epsilon,$$

and we get our desired result. □

We have a notion of ergodic averages, defined to be $A_n^f = \frac{1}{n}\sum\limits_{i=0}^{n-1} f \circ T^i$. This terminology makes sense as we're averaging the iterations of f under the ergodic transformation T.

Corollary 3.7. *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system. For any $f \in L^1_\mu$ the ergodic averages A_n^f converge in L^1_μ to a T-invariant function $\bar{f} \in L^1_\mu$.*

Proof. By the Mean Ergodic Theorem we know for $g \in L^\infty_\mu \subseteq L^2_\mu$, the ergodic averages A_n^g converge in L^2_μ to $\bar{g} \in L^2_\mu$. We know $\|A_N^g\|_\infty \le \|g\|_\infty$ so for $B \in \mathscr{B}$ we see $|\langle A_n^g, \chi_B\rangle| \le \|g\|_\infty \mu(B)$. Since $A_n^g \xrightarrow[L^2_\mu]{} \bar{g}$ we know $|\langle \bar{g}, \chi_B\rangle| \le \|g\|_\infty \mu(B)$ so $\bar{g} \in L^\infty_\mu$. In addition, since we're on a finite measure space $L^2_\mu \subseteq L^1_\mu$ (Theorem 2.11), so $A_n^g \xrightarrow[L^1_\mu]{} \bar{g}$ as well. Now we would like to show the corollary holds for all $f \in L^1_\mu$, not just the dense set of $L^\infty_\mu \subseteq L^1_\mu$. For some $f \in L^1_\mu$, fix $\epsilon > 0$ and pick a $g \in L^\infty_\mu$ such that $\|f - g\|_1 < \epsilon$. We know then that

$$\|A_n^f - A_n^g\|_1 < \epsilon$$

and we can pick n sufficiently large such that

$$\|\bar{g} - A_n^g\|_1 < \epsilon.$$

So for n, m sufficiently large, we have

$$\begin{aligned}\|A_n^f - A_m^f\|_1 &\leq \|A_n^f - A_n^g\|_1 + \|A_n^g - \bar{g}\|_1 + \|\bar{g} - A_m^g\|_1 + \|A_m^g - A_m^f\|_1 \\ &< 4\epsilon.\end{aligned}$$

Since we get a Cauchy Sequence of a separable Banach space, by the Riesz-Fischer Theorem, it converges to $\bar{f}$ within the space. We now want to show $\bar{f}$ is T-invariant, so note that

$$\|A_n^f \circ T - A_n^f\|_1 = \left\|\frac{1}{n}(f \circ T^{n+1} - f)\right\|_1 < \frac{1}{n}\|2f\|_1$$

which goes to 0 as n grows. This shows $\bar{f}$ is T-invariant. □

This is one major result we will require in our proof of the Birkhoff Ergodic Theorem. The second shortly follows.

3.3. **The Maximal Ergodic Theorem.** We develop a general inequality for operators that then we can apply in the specific case of measure-preserving systems to get the result we desire.

Proposition 3.8. (Maximal Inequality): *Let $U : L^1_\mu \to L^1_\mu$ be a linear operator such that $\|U\| \leq 1$ and $f \geq 0 \Rightarrow Uf \geq 0$. For $f \in L^1_\mu$ define the functions*

$$f_n = f + Uf + U^2 f + ... + U^{n-1} f$$

for $n \geq 1$, with $f_0 = 0$, and let

$$F_N = \max_{0 \leq n \leq N} \{f_n\}$$

Then for all $N \geq 1$

$$\int\limits_{\{x | F_N(x) > 0\}} f d\mu \geq 0$$

Proof. Because of the properties of U we know $UF_N + f \geq Uf_n + f = f_{n+1}$ so $UF_N + f \geq \max_{1 \leq n \leq N} \{f_n\}$. Because $f_0 = 0$, on the set $E = \{x \in X \mid F_N(x) > 0\}$ this implies that $UF_N + f \geq F_N$, hence $f \geq F_N - UF_N$. We also note that $F_N \geq 0 \Rightarrow UF_N \geq 0$ always. Thus,

$$\int\limits_E f \geq \int\limits_E F_N - \int\limits_E UF_N = \int\limits_X F_N - \int\limits_X UF_N \geq 0$$

since $\|U\| \leq 1$. □

The result that follows is really more of a corollary of this lemma, but it is the result we will use to prove the Birkhoff Ergodic Theorem.

Theorem 3.9. (Maximal Ergodic Theorem): *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system on a probability space and let $f \in L^1_\mu$. For $\alpha \in \mathbb{R}$, let*

$$E_\alpha = \left\{x \in X \,\middle|\, \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i(x) > \alpha\right\},$$

then

$$\alpha \mu(E_\alpha) \leq \int\limits_{E_\alpha} f d\mu.$$

Also if $T^{-1}A = A$*, then*

$$\alpha\mu(E_\alpha \cap A) \leq \int_{E_\alpha \cap A} f d\mu.$$

Proof. Define $f = g - \alpha$ and let $Uf = f \circ T$, so

$$E_\alpha = \bigcup_{N=0}^{\infty} \{x \mid F_N(x) > 0\}.$$

By the Maximal Inequality, it follows that

$$\int_{E_\alpha} f d\mu \geq 0 \Rightarrow \int_{E_\alpha} g d\mu \geq \alpha\mu(E_\alpha).$$

Note that the second statement of the Theorem is obtained by changing the underlying probability space to $(A, \mathscr{B}|_A, \frac{1}{\mu(A)}\mu|_A, T|_A)$. □

3.4. **The Birkhoff Ergodic Theorem.** Our proof of the Birkhoff Ergodic Theorem follows roughly two steps: first, we must establish a sort of mean convergence of our function to the desired result, and second, we must show that any deviation from the result we like will be bounded by a small number using the Maximal Inequality. Ultimately, we would like the exceptional set upon which our estimate disagrees to be measure zero.

Theorem 3.10. (Birkhoff Ergodic Theorem): *Let* $(X, \mathscr{B}, \mu, T)$ *be a measure-preserving system. For any* $f \in L^1_\mu$*,*

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

converges almost everywhere to a T*-invariant function* $\bar{f} \in L^1_\mu$*, where*

$$\int f d\mu = \int \bar{f} d\mu,$$

and if T *is ergodic,*

$$\int f d\mu = \bar{f}.$$

Proof. Choose $g \in L^\infty_\mu$ first and apply the Mean Ergodic Theorem to see that $A^g_n \xrightarrow[L^1_\mu]{} \bar{g}$, where $\bar{g}$ is T-invariant. Now given $\epsilon > 0$, choose n sufficiently large so that $\|\bar{g} - A^g_n\|_1 < \epsilon^2$. By applying the Maximal Ergodic Theorem to $h = \bar{g} - A^g_n$ we see

$$\epsilon\mu(\{x \in X \mid \sup_{m\geq 1} |A_m(\bar{g} - A^g_n)| > \epsilon\}) \leq \epsilon^2.$$

Since $\bar{g}$ is T-invariant, we know $A_m(\bar{g}) = \bar{g}$. Also,

$$\begin{aligned}
A_m(A_n^g) &= \frac{1}{mn}\sum_{j=0}^{m-1}\sum_{i=0}^{n-1} g\circ T^{i+j} \\
&= \frac{1}{mn}\sum_{j=0}^{m-1}\sum_{i=0}^{n-1} g\circ T^{j} + (g\circ T^{i+j} - g\circ T^{j}) \\
&= A_m^g + \frac{1}{mn}\sum_{j=0}^{m-1}\sum_{i=0}^{n-1}(g\circ T^{i+j} - g\circ T^{j}) \\
&\leq A_m^g + \frac{1}{mn}\frac{n(n-1)}{2}2\|g\|_\infty \\
&\leq A_m^g + \frac{n-1}{m}\|g\|_\infty.
\end{aligned}$$

Thus,

$$A_m(A_n^g) = A_m^g + O_n\left(\frac{\|g\|_\infty}{m}\right),$$

so

$$|A_m(A_n^g) - A_m^g| \leq O_n\left(\frac{\|g\|_\infty}{m}\right) \to 0$$

for $m \to \infty$ and n fixed. So now we see now that

$$\begin{aligned}
\mu(\{x \mid \limsup_{m\to\infty} |\bar{g} - A_m^g| > \epsilon\}) &= \mu(\{x \mid \limsup_{m\to\infty} |\bar{g} - A_m(A_n^g)| > \epsilon\}) \\
&\leq \mu(\{x \mid \limsup_{m\to\infty} |A_m(\bar{g} - A_n^g)| > \epsilon\}) \\
&< \epsilon,
\end{aligned}$$

meaning $A_m^g \to \bar{g}$ almost everywhere. We now want to generalize to all of L^1_μ. Since we have the dense set $L^\infty_\mu \subset L^1_\mu$, for $f \in L^1_\mu$ fix $\epsilon > 0$ and find a $g \in L^\infty_\mu$ such that $\|f - g\|_1 < \epsilon^2$. Since $\|A_m^f - A_m^g\|_1 \leq \|f - g\|_1 < \epsilon^2$ we know $\|\bar{f} - \bar{g}\|_1 < \epsilon^2$. Thus,

$$\begin{aligned}
&\mu(\{x \mid \limsup_{m\to\infty} |\bar{f} - A_m^f| > 2\epsilon\}) \\
&\qquad \leq \mu(\{x \mid |\bar{f} - \bar{g}| + \limsup_{m\to\infty} |\bar{g} - A_m^g| + \sup_{m\geq 1} |A_m^g - A_m^f| > 2\epsilon\}) \\
&\qquad \leq \mu(\{x \mid |\bar{f} - \bar{g}| > \epsilon\}) + \mu(\{x | |A_m^g - A_m^f| > \epsilon\}) \\
&\qquad \leq \frac{\|\bar{f} - \bar{g}\|_1}{\epsilon} + \frac{2\|A_m^g - A_m^f\|_1}{\epsilon} \leq 2\epsilon,
\end{aligned}$$

where the last step comes from the Maximal Ergodic Theorem again. This shows that $A_m^f \xrightarrow[L^1_\mu]{} \bar{f}$ almost everywhere. Since T is measure-preserving, we see that

$$\int f d\mu = \int A_n^f d\mu = \int \bar{f} d\mu.$$

In addition, since $\bar{f}$ is T-invariant, when T is ergodic, we know $\bar{f}$ must be almost everywhere constant. Thus as $\mu(X) = 1$, we obtain

$$\int f d\mu = \int \bar{f} d\mu = \bar{f}\mu(X) = \bar{f}$$

when T is ergodic. □

Roughly what the theorem says is that, given an ergodic transformation on a space, the discrete time average of a function under this transformation converges to the space average of that function under the measure. We also see that the way in which the proof was accomplished was by two general results. One was the convergence, on average, of such a time average. The other was a bound on the measure of the sets in which this convergence was not assured. So, as long as we have two such results, the proof of the Birkhoff Ergodic Theorem in more general settings can be accomplished. Also we only gain almost everywhere convergence, so I'd like to provide an example in which that fails.

Example 3.11. Recall the circle doubling map $T : x \mapsto 2x \mod 1$. We have the statement from the Birkhoff Ergodic Theorem that

$$\int f d\mu = \bar{f}$$

almost everywhere, but we see $\bar{f}(0) = f(0)$. In the simple case that $f(x) = x$ we will have disagreement with the statement of the Birkhoff Ergodic Theorem at the point $x = 0$.

So is there a way to extend the statement of the Birkhoff Ergodic Theorem to everywhere on the measure space? It turns out we can define a stronger assumption on transformations, namely unique ergodicity, such that this stronger version will hold.

3.5. **Unique Ergodicity.**

Definition 3.12. (Uniquely Ergodic): We call a transformation T *uniquely ergodic* if there is only one measure μ that is T-invariant.

In order to prove the Birkhoff Ergodic Theorem for a uniquely ergodic map, we will require a theorem about the set of invariant measures of T, $\mathscr{M}^T(X)$.

Theorem 3.13. *Given a compact metric space X, let $T : X \to X$ be continuous, and let (ν_i) be any sequence of measures in $\mathscr{M}(X)$, the set of all measures on X. Any weak*-limit point of the sequence (μ_i) defined by $\mu_i = \frac{1}{n}\sum_{j=0}^{n-1} T^j \nu_i$ is contained in $\mathscr{M}^T(X)$.*

In our case, because X is a compact metric space, $\mathscr{M}(X)$ is weak*-compact which will be key. We now proceed to the proof of the Birkhoff Ergodic Theorem.

Proposition 3.14. *T is uniquely ergodic if and only if there exists $\mu \in \mathscr{M}^T(X)$ such that for $f \in L^1_\mu$,*

$$\int f d\mu = \lim_{n\to\infty} A_n^f$$

everywhere.

Proof. Note that since T is ergodic, we have for $x \in X$

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

by applying the Theorem 3.13 combined with the fact that $|\mathscr{M}^T(X)| = 1$. By integrating both sides with any f, we get

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

so we are done. For the other direction let $\mu, \mu^* \in \mathscr{M}^T(X)$, where μ is the measure such that the hypothesis holds. By the dominated convergence theorem and T-invariance

$$\int f d\mu^* = \lim_{n\to\infty} \int A_n^f d\mu^* = \int \lim_{n\to\infty} A_n^f d\mu^* = \int\int f d\mu d\mu^* = \int C d\mu^* = C$$

where C is a constant. So we see that μ and μ^* are equivalent, thus $|\mathscr{M}^T(X)| = 1$. □

To show an example of unique ergodicity we will return to the Fourier Analysis method we used in Example 3.4 for a familiar map.

Example 3.15. Irrational Rotation on S^1 given by $T_\alpha : x \mapsto x + \alpha \mod 1$ is uniquely ergodic.

Proof. Since $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ we know $e^{2\pi i\alpha} = 1$ only when $i = 0$. Given some $g(x) = e^{2\pi kx}$ where $k \in \mathbb{Z}$ we know

$$A_n^g = \frac{1}{n}\sum_{i=0}^{n-1} e^{2\pi k(x+i\alpha)} = \frac{1}{n}\sum_{i=0}^{n-1} e^{2\pi kx}e^{2\pi ki\alpha} = \frac{1}{n}e^{2\pi kx}\frac{e^{2\pi kn\alpha}-1}{e^{2\pi k\alpha}-1}$$

when $k \neq 0$ and 1 otherwise. Thus $A_n^g \to 0$ or 1 and so by linearity we can form Fourier approximations to functions f on S^1 and by applying the previous theorem we have unique ergodicity of T_α. □

4. Gelfand's Problem

One can make use of the Ergodic Theory we've developed to talk about problems in number theory. Consider the first digit of k^n where $n \in \mathbb{N}$. Is it possible for us to talk about the frequency with which the first digit is one particular number or another? This particular question was attributed to Gelfand, and since it is possible, what remains is how to structurally phrase it as an application of our theory.

Proposition 4.1. *The frequency, $P(i)$, of any particular digit $i \in \{1, 2, ..., 9\}$ appearing as the first digit of the powers k^n for $n \in \mathbb{N}$ is*

$$P(i) = \log\left(\frac{i+1}{i}\right).$$

Proof. Note that in base 10, x and $10x$ have the same first digit, so we would like to identify these two numbers in all cases because no additional information is gained. One way we might do this is by defining the map

$$T : [0, 1) \to [0, 1), T : x \mapsto \log_{10} x \mod 1.$$

Note that we have $T : S^1 \to S^1$ as in our previous rotation examples. Another similarity we might see is that if $k \neq 10^m$ then $\alpha = \log_{10} k$ is an irrational number. Note also that

$$\log_{10}(k^n) = n \log_{10} k = n\alpha.$$

Recall that the irrational rotation map

$$T_\alpha : x \mapsto x + \alpha \mod 1$$

is uniquely ergodic. What this tells us is that these $n\alpha$ mod 1 are equidistributed over $[0, 1)$. Also note that for

$$i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\},$$

if the first digit of k^n is i, then

$$\log_{10}(k^n) \in [\log_{10} i, \log_{10}(i + 1)).$$

Since we have equidistribution from ergodicity then we know that the frequency of any particular first digit i occurring, $P(i)$, is exactly the Lebesgue measure of this subinterval, thus

$$P(i) = \log_{10}(i + 1) - \log_{10} i = \log\left(\frac{i + 1}{i}\right).$$

□

It should be noted that this is true as long as $k \neq 10^m$. Thus the frequency of any first digit is not determined by k. In this sense k is the seed of a weighted random number generator; it doesn't affect the distribution, but it does affect the order of the output sequence.

5. Continued Fractions

We now switch gears to the domain of Continued Fractions. We need to develop some tools that allow us to turn this specific domain into a familiar setting so that, by applying the Birkhoff Ergodic Theorem, we can gain information about the speed of convergence of the continued fraction approximations of a large class of irrational numbers. In order to do this we'll also need to define a suitable measure and ergodic transformation. Putting that all to the side, though, we'll start with what exactly Continued Fractions are.

5.1. **Definitions and Properties.**

Definition 5.1. (Continued Fraction): A *continued fraction* is an expression of the form

$$a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}}},$$

denoted alternatively $[a_0; a_1, a_2, a_3, a_4, ...]$ where $a_0 \in \mathbb{N} \cup \{0\}$ and $a_n \in \mathbb{N}$ for $n \geq 1$. This expansion can be finite or infinite.

We can define the the rational numbers $\frac{p_n}{q_n} = [a_0; a_1, ..., a_n]$ as partial expansions, we call them the convergents for reasons we'll soon discuss. We should also note the recursive relation

$$p_{n+1} = a_{n+1}p_n + p_{n-1}, q_{n+1} = a_{n+1}q_n + q_{n-1}.$$

Proposition 5.2. *Given a sequence (a_n) not necessarily finite, where $a_n \in \mathbb{N}$, the rational numbers $\frac{p_n}{q_n}$ converge to an irrational number x given by*

$$x = [a_0; a_1, a_2, ...] = \lim_{n\to\infty} \frac{p_n}{q_n} = a_0 + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{q_{n-1}q_n}.$$

In particular we set $p_0 = a_0$ *and* $q_0 = 1$.

Proof. Suppose that $x = \frac{a}{b} \in \mathbb{Q}$. We know from the limit expression above and recursive relation that

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n a_{n+1} q_n} \leq \frac{1}{q_n^2}.$$

This implies that

$$|q_n a - p_n b| = q_n b \left| \frac{a}{b} - \frac{p_n}{q_n} \right| \leq \frac{b}{q_n}.$$

Since $q_n \to \infty$ by the recursive relation we see that $\left|x - \frac{p_n}{q_n}\right| = 0$. However, since the convergents are in lowest terms by definition, $|q_n a - p_n b| \neq 0$, thus we obtain a contradiction. □

Now we derive uniqueness of these expansions.

Proposition 5.3. *Every expansion represents a unique irrational number.*

Proof. Let $x = [a_0; a_1, a_2, ...]$. Then $x = a_0 + \frac{1}{[a_1; a_2, ...]}$ so

$$x \in (a_0, a_0 + \frac{1}{a_1}) \subseteq (a_0, a_0 + 1).$$

This means that x determines a_0, and now $\frac{1}{x - a_0} = [a_1; a_2, ...]$ so we can apply the partition again and we get our uniqueness inductively. □

Our measure preserving map will be $T(x) = \left\{\frac{1}{x}\right\} = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$ which is the fractional part of $\frac{1}{x}$. We just need a measure for it to preserve. If we consider the string expression of a continued fraction we see that $T([a_2, a_3, ...]) = [a_3, ...]$ so it acts like a shift map, not unlike our Bernoulli Shift. This suggests a similar method of proof for its ergodicity. First however, we see that all irrational numbers have a continued fraction expansion.

Proposition 5.4. *For any* $x \in [0, 1] \setminus \mathbb{Q}$ *the sequence of digits* $a_n(x) = \lfloor \frac{1}{T^{n-1}(x)} \rfloor$ *gives the continued fraction expansion of* $x = [a_1(x), a_2(x), a_3(x), ...]$.

Proof. Let $y = [a_1, a_2, a_3, ...]$. By Proposition 5.2,

$$[a_1, ..., a_{2n}] = \frac{p_{2n}}{q_{2n}} < y < \frac{p_{2n+1}}{q_{2n+1}} = [a_1, ... a_{2n+1}]$$

for all n. If we can show this is true for x for all n then we can conclude that $x = y$. Recall how we defined $p_0 = a_0 = 0$ and $q_0 = 1$. We also have $\frac{p_1}{q_1} = \frac{1}{a_1}$. Thus our inequality holds for x for $n = 0$. Assume now it holds for x for all $n \leq N$. Apply T to see

$$[a_2, ..., a_{2N+1}] < T(x) = \frac{1}{x} - a_1 < [a_2, ..., a_{2N+2}].$$

Thus,

$$a_1 + [a_2, ..., a_{2N+1}] < \frac{1}{x} < a_1 + [a_2, ..., a_{2N+2}].$$

So by inverting we see,

$$[a_1, a_2, ..., a_{2N+2}] < x < [a_1, a_2, ..., a_{2N+1}]$$

which shows the convergents oscillate around x as desired. If we apply T once more we will see the result for all $n \leq N + 1$. □

5.2. **Gauss Measure.** We have come to the task of defining our measure. It is somewhat strange, but it does serve the purpose.

Proposition 5.5. *Given $B \subseteq [0,1]$ measurable in the Borel σ-algebra, the continued fraction map $T(x) = \left\{\frac{1}{x}\right\}$ on $(0,1)$ preserves the Gauss measure*
$$\mu(B) = \frac{1}{\log 2} \int_B \frac{1}{1+x} dx$$

Proof. We show this is true for $[0,b]$ for all $b > 0$. Note
$$T^{-1}[0,b] = \{x | 0 \leq Tx \leq b\} = \bigsqcup_{n=1}^{\infty} \left[\frac{1}{b+n}, \frac{1}{n}\right].$$

Thus,
$$\begin{aligned}
\mu(T^{-1}[0,b]) &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{\frac{1}{b+n}}^{\frac{1}{n}} \frac{1}{1+x} dx \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log\left(1 + \frac{1}{n}\right) - \log\left(1 + \frac{1}{b+n}\right)\right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \log\left(\frac{(n+1)(b+n)}{(n)(b+n+1)}\right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log\left(1 + \frac{b}{n}\right) - \log\left(1 + \frac{b}{n+1}\right)\right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{\frac{b}{n+1}}^{\frac{b}{n}} \frac{1}{1+x} dx \\
&= \frac{1}{\log 2} \int_0^b \frac{1}{1+x} dx \\
&= \mu([0,b]).
\end{aligned}$$

So by taking intersections and unions of such intervals we are done. □

We now move to the stronger property of ergodicity.

Proposition 5.6. *The continued fraction map $T(x) = \left\{\frac{1}{x}\right\}$ on $(0,1)$ is ergodic with respect to the Gauss measure μ.*

Proof. Recall that T acts like a shift map on the continued fraction expansion of any particular x. Also recall that we proved the Bernoulli Shift is ergodic. All of this is to suggest that we'd like to pursue a similar method of proof: we want to control the size of the cylinder sets and their intersections with the hopes that it will let us prove ergodicity. Given the n-tuple $a = (a_1, ..., a_n) \in \mathbb{N}^n$, define the cylinder set
$$I_a = \{[x_1, x_2, ...] | x_i = a_i, 1 \leq i \leq n\} \subseteq \mathbb{N}^{\mathbb{N}}$$
We will start first with intervals $B = [\alpha, \beta] \in \mathscr{B}$ and the rest of the measurable sets will follow by generation. We need to develop some more machinery for continued

fractions. Denote the tail of the continued fraction expansion of x starting at index n by x_n. Thus when $x = [a_0; a_1, ...]$, $x_n = T^n x = [a_n; a_{n+1}, ...]$. One can derive from the recursive relations that

$$\frac{p_{n+m}}{q_{n+m}} = \frac{p_n \frac{p_{m-1}(x_{n+1})}{q_{m-1}(x_{n+1})} + p_{n-1}}{q_n \frac{p_{m-1}(x_{n+1})}{q_{m-1}(x_{n+1})} + q_{n-1}}.$$

So when $m \to \infty$ we get

$$x = \frac{p_n x_{n+1} + p_{n-1}}{q_n x_{n+1} + q_{n-1}}.$$

Now we see

$$x \in I_a \cap T^{-n}[\alpha, \beta] \iff x = [a_1, ..., a_n], x_n \in [\alpha, \beta].$$

Since T^n restricted to I_a is continuous and monotone (increasing when odd, decreasing when even), by putting the previous results together we get

$$I_a \cap T^{-n}[\alpha, \beta] = \left[\frac{p_n + p_{n-1}\alpha}{q_n + q_{n-1}\alpha}, \frac{p_n + p_{n-1}\beta}{q_n + q_{n-1}\beta}\right] \text{ or } \left[\frac{p_n + p_{n-1}\beta}{q_n + q_{n-1}\beta}, \frac{p_n + p_{n-1}\alpha}{q_n + q_{n-1}\alpha}\right].$$

Thus the Lebesgue measure, μ_L, of it is

$$\begin{aligned}\left|\frac{p_n + p_{n-1}\beta}{q_n + q_{n-1}\beta} - \frac{p_n + p_{n-1}\alpha}{q_n + q_{n-1}\alpha}\right| &= \left|\frac{(p_n + p_{n-1}\beta)(q_n + q_{n-1}\alpha) - (q_n + q_{n-1}\beta)(p_n + p_{n-1}\alpha)}{(q_n + q_{n-1}\beta)(q_n + q_{n-1}\alpha)}\right| \\ &= \left|\frac{p_{n-1}q_n\beta + p_n q_{n-1}\alpha - p_n q_{n-1}\beta - p_{n-1}q_n\alpha}{(q_n + q_{n-1}\beta)(q_n + q_{n-1}\alpha)}\right| \\ &= (\beta - \alpha)\left|\frac{p_{n-1}q_n - p_n q_{n-1}}{(q_n + q_{n-1}\beta)(q_n + q_{n-1}\alpha)}\right| \\ &= (\beta - \alpha)\frac{1}{(q_n + q_{n-1}\beta)(q_n + q_{n-1}\alpha)}.\end{aligned}$$

From the previous discussion, the Lebesgue measure of I_a is

$$\left|\frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right| = \left|\frac{p_{n-1}q_n - p_n q_{n-1}}{q_n(q_n + q_{n-1})}\right| = \frac{1}{q_n(q_n + q_{n-1})},$$

so we have that

$$\mu_L(I_a \cap T^{-n}A) = \mu_L(A)\mu_L(I_a)\frac{q_n(q_n + q_{n-1})}{(q_n + q_{n-1}\beta)(q_n + q_{n-1}\alpha)},$$

meaning the measures are equivalent up to a constant. Additionally,

$$\frac{\mu_L(B)}{2\log 2} \le \mu(B) = \frac{1}{\log 2}\int_B \frac{1}{1+x}dx \le \frac{\mu_L(B)}{\log 2}.$$

So combining these two results we get that

$$C_1\mu(I_a)\mu(B) \le \mu(I_a \cap T^{-n}B) \le C_2\mu(I_a)\mu(B)$$

where $C_1 > 0$ and $C_2 > 0$ are constants. By applying our recursive relation, and because every $a_i \in \mathbb{N}$ we know

$$q_n \ge 2^{\frac{n-2}{2}}, p_n \ge 2^{\frac{n-2}{2}}.$$

Thus $\mu(I_a) \le \frac{1}{2^{n-2}}$. As $n \to \infty$ this goes to 0 so the cylinder sets I_a generate the Borel σ-algebra. This is just as with Bernoulli Shifts. In turn we know that for $B, B^* \in \mathscr{B}$

$$C_1\mu(B)\mu(B^*) \le \mu(B \cap T^{-n}B^*) \le C_2\mu(B)\mu(B^*).$$

Consider such a $B^* = T^{-1}B^*$. Then $X \setminus B^* \in \mathscr{B}$ as well, so we know that

$$C_1\mu(X \setminus B^*)\mu(B^*) \leq \mu((X \setminus B^*) \cap B^*) \leq C_2\mu(X \setminus B^*)\mu(B^*),$$

thus $\mu(X \setminus B^*)\mu(B^*) = 0$, meaning $\mu(B^*) = 0$ or $\mu(X \setminus B^*) = 0$ so T is ergodic. □

With this, we derive a result regarding the rate of convergence of the continued fraction expansion

5.3. **Application of the Birkhoff Ergodic Theorem.** Perhaps the strangest thing about the following result is that it relates the rate of convergence of almost every irrational number to a ratio of transcendental numbers. It comes seemingly out of nowhere, where the real black box here is our Gauss measure. Without further ado...

Corollary 5.7. *For almost every $x = [a_1, a_2, a_3, ...] \in (0, 1)$, the rate of approximation of the continued fractions is given by*

$$\lim_{n\to\infty} \frac{1}{n} \log \left| x - \frac{p_n(x)}{q_n(x)} \right| \to \frac{-\pi^2}{6 \log 2}$$

Proof. We first note

$$\begin{aligned} \frac{p_n(x)}{q_n(x)} &= \frac{1}{a_1 + [a_2, ..., a_n]} \\ &= \frac{1}{a_1 + \frac{p_{n-1}(Tx)}{q_{n-1}(Tx)}} \\ &= \frac{q_{n-1}(Tx)}{a_1 q_{n-1}(Tx) + p_{n-1}(Tx)}. \end{aligned}$$

So we have equality of numerator and denominator on both sides because these expressions are always in lowest terms. In particular $p_n(x) = q_{n-1}(Tx)$. Because of the definition of the expansion, we have $p_1(x) = 1$ always, thus

$$\frac{1}{q_n(x)} = \frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)} \cdots \frac{p_{n-1}(Tx)}{q_{n-1}(Tx)} \frac{p_n(x)}{q_n(x)}.$$

Thus,

$$\frac{1}{n} \log \left(\frac{1}{q_n(x)} \right) = \frac{-1}{n} \log q_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \log \left(\frac{p_{n-i}(T^i x)}{q_{n-i}(T^i x)} \right)$$

Now let's define our function $f = \log x \in L^1_\mu$. Rewriting,

$$\begin{aligned} \frac{-1}{n} \log q_n(x) &= \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i(x) - \frac{1}{n} \sum_{i=0}^{n-1} \left[f \circ T^i(x) - \log \left(\frac{p_{n-i}(T^i x)}{q_{n-i}(T^i x)} \right) \right] \\ &= A_n^f - \frac{1}{n} r_n(x). \end{aligned}$$

Here $r_n(x)$ is a remainder term of some sort. We want to show that this remainder approaches 0 as $n \to \infty$. Remember that since every $a_i \in \mathbb{N}$, from our recursive relation we know that

$$q_n \geq 2^{\frac{n-2}{2}}, p_n \geq 2^{\frac{n-2}{2}}.$$

In addition, from the expression of an irrational number as limit of partial expansions, we know that

$$\begin{aligned}
\left|\frac{x}{\frac{p_n(x)}{q_n(x)}} - 1\right| &= \frac{q_n(x)}{p_n(x)}\left|x - \frac{p_n(x)}{q_n(x)}\right| \\
&= \frac{q_n(x)}{p_n(x)}\left|(-1)^{n+2}\left(\frac{1}{q_n q_{n+1}} - \frac{1}{q_{n+1}q_{n+2}} + ...\right)\right| \\
&< \frac{q_n(x)}{p_n(x)}\frac{1}{q_n q_{n+1}} \\
&= \frac{1}{p_n(x)q_{n+1}(x)} \\
&\leq \frac{1}{2^{n-1}}.
\end{aligned}$$

Another specific fact to note is that $|\log x| \leq 2|x-1|$ when $|x-1| \leq \frac{1}{2}$, which is true in our expansion for $r_n(x)$ whenever $i \leq n-2$. Applying this, we see

$$\begin{aligned}
|r_n(x)| &\leq \sum_{i=0}^{n-1}\left|\log\left(\frac{T^i x}{\frac{p_{n-i}(T^i x)}{q_{n-i}(T^i x)}}\right)\right| \\
&\leq \left|\log\left(\frac{T^{n-1}x}{\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)}}\right)\right| + \sum_{i=0}^{n-2} 2\left|\frac{T^i x}{\frac{p_{n-i}(T^i x)}{q_{n-i}(T^i x)}} - 1\right| \\
&\leq \left|\log\left(\frac{T^{n-1}x}{\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)}}\right)\right| + \sum_{i=0}^{n-2}\frac{2}{2^{n-i-1}} \\
&\leq \left|\log\left(\frac{T^{n-1}x}{\frac{p_1(T^{n-1}x)}{q_1(T^{n-1}x)}}\right)\right| + 2 \\
&\leq \left|\log(T^{n-1}(x)a_1(T^{n-1}(x)))\right| + 2 \\
&\leq \left|\log\left(\frac{a_1(T^{n-1}(x))}{a_1(T^{n-1}(x))+1}\right)\right| + 2 \\
&\leq \log 2 + 2.
\end{aligned}$$

Which shows $\lim_{n\to\infty} r_n(x) = 0$. From the limit of the partial expansions we know

$$\begin{aligned}
\frac{1}{q_n q_{n+1}} \geq \left|x - \frac{p_n(x)}{q_n(x)}\right| &\geq \frac{1}{q_n q_{n+1}} - \frac{1}{q_{n+1}q_{n+2}} \\
&= \frac{q_{n+2} - q_n}{q_n q_{n+1} q_{n+2}} \\
&= \frac{a_{n+1}q_{n+1} + q_n - q_n}{q_n q_{n+1} q_{n+2}} \\
&= \frac{a_{n+1}}{q_n q_{n+2}} \geq \frac{1}{q_n q_{n+2}}.
\end{aligned}$$

So when we take logs we obtain

$$-\log q_n - \log q_{n+2} \leq \log\left|x - \frac{p_n(x)}{q_n(x)}\right| \leq -\log q_n - \log q_{n+1}.$$

Now, by the Birkhoff Ergodic Theorem

$$\begin{aligned}\lim_{n\to\infty}\frac{-1}{n}\log q_n(x) &= \lim_{n\to\infty} A_n^f - \frac{1}{n}r_n(x)\\ &= \frac{1}{\log 2}\int_0^1 \frac{\log x}{1+x}dx - 0\\ &= \frac{-\pi^2}{12\log 2}\end{aligned}$$

almost everywhere. So combining this with the last statement we see finally that

$$\lim_{n\to\infty}\frac{1}{n}\log\left|x - \frac{p_n(x)}{q_n(x)}\right| = \frac{-\pi^2}{6\log 2}.$$

□

This is really a result about the speed of convergence because by moving some terms around we see that the quantity $\left|x - \frac{p_n(x)}{q_n(x)}\right|$ grows like $e^{\frac{-\pi^2}{6\log 2}}$ as $n \to \infty$. I should also emphasize that this holds almost everywhere, not everywhere.

Acknowledgments. This paper hoped to be a small walk through basic measurable dynamics, developing a major theorem whose applications include some domains that are at first glance unrelated. Thanks to my mentor, Brian Chung, for always being patient and helpful. Thanks, too, to all the faculty who kept the program chock-full. Finally, thanks to Peter May for all of his unending effort in this REU program.

References

[1] Manfred Einsiedler, Thomas Ward. Ergodic Theory with a view towards Number Theory. Springer, 2011.

[2] Jonathan L. King. Three Problems in Search of a Measure. The American Mathematical Monthly, vol 101, 1994, pp. 609-628.

[3] Elias M. Stein, Rami Shakarchi. Princeton Lectures in Analysis III: Real Analysis: Measure Theory, Integration, and Hilbert Spaces. Princeton University Press, 2005.