APPLICATIONS OF THE BIRKHOFF ERGODIC THEOREM

WEIAN WANG

ABSTRACT. Ergodic theory studies the long-term averaging properties of measure-preserving dynamical systems. In this paper, we state and present a proof of the ergodic theorem due to George Birkhoff, who observed the asymptotic equivalence of the time-average and space-average of a point x in a finite measure space. Then, we examine a number of applications of this theorem in number-theoretic problems, including a study of normal numbers and of Lüroth series transformations.

CONTENTS

A **discrete dynamical system** consists of a space X and a transformation T which maps the space onto itself i.e. $T\colon X \to X$. We assume T is a measurable function. When studying discrete systems, we consider how a point in the space moves over discrete time intervals. We can think of the X as the space of all possible states of some system, where T specifies how the state changes changes over a specific time interval. For some $y \in X$, we define the **orbit of** y **under** T to be the sequence $y, T(y), T^2(y), \cdots, T^n(y), \cdots$. If T is one-to-one and onto, then we say T is an **invertible transformation**.

Ergodic theory studies a particular subset of these dynamical systems–those which are measure-preserving. In the paper, we define and present a number of characteristics pertaining to these measure-preserving dynamical systems, such as randomness and recurrence. We assume a background in basic notions of measure theory and Lebesgue integration. The relevant background information can be found in most real analysis textbooks, such as *Real and Complex Analysis* by Walter Rudin. The reference we derive the conventions and notations in the subsequent section from is *An Invitation to Ergodic Theory* by C.E. Silva.

Date: DEADLINES: Draft AUGUST 17 and Final version AUGUST 28, 2015.

1. Recurrence and Ergodicity in Dynamical Systems

Suppose we have a nonempty set X and a σ-algebra $\mathcal{B}$ in X. A **measure space** is defined to be a triple $(X, \mathcal{B}, \mu)$ where μ is a measure on $\mathcal{B}$. We say a set A is a **measurable set** if $A \subset X$ and $A \in \mathcal{B}$. A **probability space** is a measure space $(X, \mathcal{B}, \mu)$ such that $\mu(X) = 1$. We say a measure space is a **finite measure space** if $\mu(X) < \infty$ and is **σ-finite** if there exists a sequence of measurable sets A_n of finite measure such that $X = \cup_{n=1}^{\infty} A_n$.

Definition 1.1. Let $(X, \mathcal{B}, \mu)$ be a probability space. We say the transformation $T: X \to X$ is **measure-preserving** (with respect to μ) and that μ is **T-invariant** if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.

If T is measure-preserving, then we refer to the dynamical system $(X, \mathcal{B}, \mu, T)$ as a measure-preserving dynamical system.

We take a measurable set A of positive measure in $\mathcal{B}$ and consider the orbit of points in A. Specifically, we ask whether the points in A will return to the set A and if so, how often will they return. We say a measure-preserving transformation T defined on a measure space $(X, \mathcal{B}, \mu)$ is **recurrent** if for every measurable set A of positive measure, there is a null set $N \subset A$ such that for all $x \in A \setminus N$, there exists an integer $n = n(x) > 0$ with $T^n(x) \in A$. In order words, if, for every measurable set A of positive measure, every point in that set, except points in a set of measure zero, eventually returns to A under T, then T is recurrent, and the system is a recurrent dynamical system.

The following theorem is a theorem due to Poincaré who proved a property relating to finite measure spaces. The statement and full proof of this theorem can be found in [5].

Theorem 1.2 (Poincaré Recurrence Theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a finite measure space. If $T: X \to X$ is a measure-preserving transformation, then T is recurrent.*

We define a point which is not recurrent to be a **wandering point**. Formally, a wandering point is a point in A such that $\forall n \in \mathbb{N}$, $T^n(x) \notin A$. By Poincaré recurrence, for a finite measure space with a measure-preserving transformation T, the wandering set (i.e. the set of all wandering points in this system) is a set of measure zero.

Many of the properties we study that hold for recurrent and ergodic systems hold outside of a set of measure zero; we characterize these properties as holding *almost everywhere*. To generalize this notion, we define an invariant set. We say a set A is **positively invariant** if $A \subset T^{-1}(A)$ and **strictly invariant** or simply **T-invariant** if $A = T^{-1}(A)$. Studying ergodic properties on T-invariant sets allows us to disregard sets of measure zero.

Recurrent transformations on a finite measure space are said to be **ergodic** if they satisfy the following property.

Definition 1.3. A measure-preserving transformation T is **ergodic** if whenever A is a strictly T-invariant measurable set, then either $\mu(A) = 0$ or $\mu(A^c) = 0$.

The following lemma relates our notions of recurrence and ergodicity, and introduces some new properties of such systems.

Lemma 1.4. *Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space and let T be a measure-preserving transformation. Then the following are equivalent:*

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

1.1. **Approximation with Sufficient Semi-rings.** While we assume a background in fundamental notions of measure theory, we develop here some techniques of approximation with semi-ring structures.

We define a **semi-ring** $\mathcal{R}$ to be a collection of subsets of a nonempty space X such that

(1) $\mathcal{R}$ is nonempty.
(2) if $A, B, \in \mathcal{R}$, then $A \cap B \in \mathcal{R}$, and
(3) if $A, B \in \mathcal{R}$, then

$$A \setminus B = \bigsqcup_{j=1}^{n} E_j$$

where $E_j \in \mathcal{R}$ are disjoint.

We say a semi-ring $\mathcal{R}$ is a **sufficient semi-ring** if it satisfies that for every measurable set A in the σ-algebra,

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

There are several useful properties for studying these structures. In particular, we have that if a measurable set can be written as a countable union of elements of a semi-ring $\mathcal{C}$, then it can be written as a countable union of disjoint elements of the semi-ring.[1] We also have the property that every finite measurable set can be approximated, for every $\epsilon > 0$, by a finite union of disjoint elements of a sufficient semi-ring where the symmetric difference between the finite measurable set and the sufficient semi-ring is less than ϵ. In this examination of ergodic dynamical systems, particularly relevant sufficient semi-rings are the intervals and the dyadic intervals, or the set of all intervals of the form $[\frac{k}{2^n}, \frac{k+1}{2^n})$ where $n > 0$ and $k = 0, 1, \cdots, 2^n - 1$.

Lemma 1.5. *Let $(X, \mathcal{B}, \mu)$ be a measure space with a sufficient semi-ring $\mathcal{C}$. Let A be a measurable set, $\mu(A) < \infty$, and let $\epsilon > 0$. Then there exists a set H^* that is a finite union of disjoint elements of $\mathcal{C}$ such that $\mu(A \Delta H^{*2}) < \epsilon$.*[3]

This lemma implies that every element of a sufficient semi-ring is a measurable set. When proving properties of measure-preserving systems, it is sufficient to prove them only for the measurable sets which are elements of a sufficient semi-ring in order to verify the properties for all measurable sets in a σ-algebra, i.e.

[1] This statement is Proposition 2.7.1 in [5].
[2] We use Δ to denote symmetric difference.
[3] The lemma is a statement of the first of Littlewood's Three Principles of Real Analysis. Its statement and full proof can be found in [5].

Theorem 1.6. *Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space with a sufficient semi-ring $\mathcal{C}$. If for all $I \in \mathcal{C}$,*

1. *$T^{-1}(I)$ is a measurable set, and*
2. *$\mu(T^{-1}(I)) = \mu(I)$*

then T is a measure-preserving transformation.

2. The Birkhoff Ergodic Theorem

The Ergodic Theorem, due to Birkhoff in 1931, relates the time-average of a transformation and the measure of the space. By time-average, we refer to the limit of the average number of times the elements of the sequence $x, T(x), T^2(x), \cdots$ are in A, or the average number of times x visits A in its orbit. This theorem states that for an ergodic system, for any measurable set A and almost every x in the full space X, the limit of the average recurrence frequency of x in A is asymptotically equal to the measure of A.

The following proof of the theorem follows [2] and [4]. In particular, the combinatorial trick used to prove the Maximal Ergodic Theorem follows the one presented in [2] due to Riesz.

Definition 2.1. Suppose we have a finite sequence of real numbers $a_1, a_2, \ldots, a_n$. We say the term a_k is an **m-leader** if there exists a positive integer p where $1 \leq p \leq m$ such that $a_k + \cdots + a_{k+p-1} \geq 0$.

Lemma 2.2. *The sum of all m-leaders is nonnegative.*

Proof. Let a_k be the first m-leader of the finite sequence of reals $a_1, \ldots, a_n$ and let p be the smallest integer $p \leq m$ for which $a_k + \cdots + a_{k+p-1} \geq 0$.

It follows that every a_h such that $k \leq h \leq k+p-1$ must be an m-leader as well. If not, then $a_h + \cdots + a_{k+p-1} < 0 \Rightarrow a_k + \cdots + a_{h-1} > 0$, which contradicts that p is the smallest integer $p \leq m$ for which $a_k + a_{k+1} + \cdots + a_{k+p-1} \geq 0$.

Since each a_h where $k \leq h \leq k+p-1$ is an m-leader, the sum of these terms is the sum $a_k + \cdots + a_{k+p-1}$, which by assumption, is nonnegative. We repeat this process inductively for the rest of the terms of the sequence $a_{k+p}, \ldots, a_n$ and the result follows. □

We will denote

$$f_n(x) = \sum_{k=0}^{n-1} f(T^k(x))$$

Lemma 2.3 (Maximal Ergodic Theorem)**.** *Suppose we have a probability space $(X, \mathcal{B}, \mu)$ and a measure-preserving transformation $T \colon X \to X$. Let $f \colon X \to \mathbb{R}$ be an integrable function and define*

$$G(f) = \{x \in X \colon f_n(x) \geq 0 \textit{ for some } n > 0\}.$$

Then,

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

Proof. Let m be a positive integer. We define G_m as follows:

$$G_m = \{x \in X \colon f_k(x) \geq 0 \text{ for some } k, 1 \leq k \leq m\}.$$

Let n be an arbitrary positive integer. Consider for each x, the m-leaders of the sequence $f(x), f(T(x)), \ldots, f(T^{n+m-1}(x))$. We define $s_m(x)$ to be the sum of these m-leaders.

We define B_k to be the set of $x \in X$ for which $f(T^k(x))$ is an m-leader of the sequence $f(x), f(T(x)), \ldots, f(T^{n+m-1}(x))$. From our definitions, it is clear s_m is measurable and integrable.

By Lemma 2.2, we see that $s_m \geq 0$, and so,

$$0 \leq \int_{B_k} s_m d\mu = \sum_{k=0}^{n+m-1} \int_{B_k} f \circ T^k d\mu$$

We notice that if $k = 1, 2, \cdots, n-1$, $x \in B_k \iff T(x) \in B_{k-1}$, and equivalently, $B_k = T^{-1}(B_{k-1}) \iff B_k = T^{-k}(B_0)$. By a change of variables,

$$\int_{B_k} f \circ T^k d\mu = \int_{T^-k(B_0)} f \circ T^k d\mu = \int_{B_0} f d\mu.$$

As $G_m = B_0$ and T is measure-preserving, it follows

$$\begin{aligned} 0 \leq \sum_{k=0}^{n+m-1} \int_{B_k} f \circ T^k d\mu &= \sum_{k=0}^{n-1} \int_{B_0} f d\mu + \sum_{k=n}^{n+m-1} \int_{B_k} f \circ T^k d\mu \\ &\leq n \int_{G_m} f d\mu + m \int |f| d\mu \end{aligned}$$

If we divide through by n and let $n \to \infty$, we are left with

$$\int_{G_m} f d\mu \geq 0.$$

Consider $f\chi_{G_n}$. We find that as $G_m \subset G_{m+1}$, this is an increasing sequence. As $G(f) = \bigcup_{m \geq 1} G_m$, we observe $\lim_{n\to\infty} f\chi_{G_n} = f\chi_{G(f)}$. Since $|f\chi_{G_n}| \leq |f|$, by the Dominated Convergence Theorem, we have

$$0 \leq \lim_{n \to \infty} \int f\chi_{G_n} d\mu = \int_{G(f)} f d\mu$$

□

Theorem 2.4 (Birkhoff Ergodic Theorem)**.** *Suppose we have a probability space $(X, \mathcal{B}, \mu)$ and a measure-preserving transformation $T\colon X \to X$. If $f\colon X \to \mathbb{R}$ is an integrable function, then*

(1) $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$ *exists for almost all $x \in X$. Denote this limit as $\tilde{f}(x)$.*
(2) $\tilde{f}(Tx) = \tilde{f}(x)$ *a.e.*
(3) *For any measurable set A that is T-invariant,*

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

If T is ergodic, then

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

Proof. (1) We denote

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

and

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

For $\alpha, \beta \in \mathbb{R}$, we denote

$$E_{\alpha,\beta} = \{x \in X \colon f_*(x) < \alpha < \beta < f^*(x)\}.$$

To prove the existence of the limit a.e., we want to show that for almost every $x \in X$, $f_*(x) = f^*(x)$. To do so, we will show that $E_{\alpha,\beta}$, i.e. the set of points where $f_*(x) = f^*(x)$ differ, is a set of measure zero.

First, we want to show that our set $E_{\alpha,\beta}$ is T-invariant. We claim f^* and f_* are T-invariant. We take $\liminf$ as n approaches infinity of the following expression

$$\frac{1}{n} f_n(T(x)) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(T(x))) = \frac{n+1}{n} f_{n+1}(x) - \frac{1}{n} f(x)$$

and find

$$f_*(T(x)) = \liminf_{n\to\infty} \frac{n+1}{n} f_{n+1}(x) - \liminf_{n\to\infty} \frac{1}{n} f(x) = f_*(x)$$

which proves $f_* \circ T = f_*$ i.e. f_* is T-invariant. A similar argument shows $f^* \circ T = f^*$ i.e. f^* is T-invariant. Consequently, $E_{\alpha,\beta}$ is T-invariant. We define $G(f-\beta) = \{x \in X \colon (f-\beta)_n \geq 0 \text{ for some } n > 0\}$. Next, we consider the set of all x such that $f^*(x) > \beta$. There exists an $N \in \mathbb{N}$ such that $\frac{1}{N}\sum_{i=0}^{N-1} f(T^i(x)) > \beta \Rightarrow \sum_{i=0}^{N-1} f(T^i(x)) - N\beta \geq 0$ exactly if $\sum_{i=0}^{N-1}(f-\beta)_i(x) \geq 0$, so $x \in G(f-\beta)$.

In particular, we find $E_{\alpha,\beta} \subset G(f-\beta)$. We apply the Maximal Ergodic Theorem to T restricted to $E_{\alpha,\beta}$ and to $f-\beta$, which gives us

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

Using that $f_*(x) < \alpha \implies -f^*(x) > -\alpha$, we similarly find $E_{\alpha,\beta} \subset G(\alpha - f)$. By an application of the Maximal Ergodic Theorem to T restricted to $E_{\alpha,\beta}$ and to $\alpha - f$, it follows

$$\int_{E_{\alpha,\beta}} -f d\mu \geq -\alpha\mu(E_{\alpha,\beta}) \Rightarrow \int_{E_{\alpha,\beta}} f \leq \alpha\mu(E_{\alpha,\beta}).$$

Then, as $\alpha < \beta$ by assumption, and

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

it follows $\mu(E_{\alpha,\beta}) = 0$. Hence, as this holds for all rational α, β, $f^* = f_*$ a.e.

(2) Next, we want to show $\tilde{f}(T(x)) = \tilde{f}(x)$ a.e.

The proof that $\tilde{f}$ is T-invariant comes directly from the definition of the limit. We find

$$\begin{aligned}\tilde{f}(T(x)) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) &= \lim_{n\to\infty} \frac{1}{n} \left(\sum_{k=0}^{n-2} f(T^k(T(x)) + f(x) \right) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-2} f(T^k(T(x)) + \lim_{n\to\infty} \frac{1}{n} f(x) \\ &= \tilde{f}(x)\end{aligned}$$

(3) Finally, we will show that $\int_A f d\mu = \int_A \tilde{f} d\mu$ for any measurable set A.

We begin by defining $A_{n,k} = \{x \in A\colon \frac{k}{2^n} \le \tilde{f}(x) < \frac{k+1}{2^n}\}$ where $n = 0, 1, \ldots,$ and $k = 0, \pm 1, \pm 2, \ldots$. By (2), $A_{n,k}$ is a T-invariant set for each n and k. We observe that for each n, $X = \bigcup_k A_{n,k}$.

Fix $\epsilon > 0$. In (1), we found $\tilde{f}(x) = f^*(x)$, as the limit exists, and so, given $\epsilon > 0$, it is true that $\tilde{f}(x) \ge \frac{k}{2^n} \Rightarrow f^*(x) - \frac{k}{2^n} + \epsilon > 0$. We apply the Maximal Ergodic Theorem to T restricted to $A_{n,k}$ and to $f(x) - \frac{k}{2^n} + \epsilon$ as we did in (1) and get

$$\int_{A_{n,k}} f d\mu \ge \left(\frac{k}{2^n} - \epsilon \right) \mu(A_{n,k})$$

We provide a similar argument to handle the right-hand-side inequality i.e. $\tilde{f}(x) < \frac{k+1}{2^n}$. By the existence of the limit, $\tilde{f}(x) = f_*(x)$, from which it follows $\tilde{f}(x) < \frac{k+1}{2^n} \Rightarrow f^*(x) < \frac{k+1}{2^n} \Rightarrow -f_*(x) > -(\frac{k+1}{2^n})$. This gives us $-f^*(x) + (\frac{k+1}{2^n}) > 0$ for all $x \in A_{n,k}$. As in (1), we find $A_{n,k} \subset G(\frac{k+1}{2^n} - f)$, and we apply the Maximal Ergodic Theorem to $\frac{k+1}{2^n} - f$ and $A_{n,k}$, which gives us

$$\int_{A_{n,k}} -f d\mu \ge -\left(\frac{k+1}{2^n} \right) \mu(A_{n,k})$$

Then, we let $\epsilon \to 0$, and it follows

$$\frac{k}{2^n} \mu(A_{n,k}) \le \int_{A_{n,k}} f d\mu \le \frac{k+1}{2^n} \mu(A_{n,k}).$$

Our definition of $A_{n,k}$ gives us the same inequality expression for $\tilde{f}$ i.e. $\frac{k}{2^n} \mu(A_{n,k}) \le \int_{A_{n,k}} \tilde{f} d\mu \le \frac{k+1}{2^n} \mu(A_{n,k})$. Then,

$$\int_{A_{n,k}} |f - \tilde{f}| d\mu \le \frac{1}{2^n} \mu(A_{n,k})$$

We sum over k and find

$$\int_A |f - \tilde{f}| d\mu \le \frac{1}{2^n} \mu(A).$$

We let n go to infinity, which gives us

$$\int_A |f - \tilde{f}| d\mu = 0 \Rightarrow \int_A f d\mu = \int_A \tilde{f} d\mu.$$

□

3. Lüroth Series Transformations

We apply our understanding of ergodic theory to study properties of one particular class of transformations known as *Lüroth series transformations.* A Lüroth series transformation is a transformation on $[0,1)$ that arises as follows: there is a partition of $[0,1)$ into intervals $\{J_n : n \in A\}$, where A is $\mathbb{N}$ or a finite subset of $\mathbb{N}$ such that on each J_n, T is an increasing linear function whose range is an interval with endpoints 0 and 1.

The classical example of one such transformation is the following map $T\colon [0,1) \to [0,1)$ defined by

$$T(x) = \begin{cases} n(n+1)x - n, & \text{x} \in [\frac{1}{n+1}, \frac{1}{n}) \\ 0, & \text{x} = 0. \end{cases} \tag{3.1}$$

In this section, we will show that each $x \in [0,1)$ admits a unique, finite or infinite, Lüroth series expansion for this particular T and illustrate some properties of the dynamics of such systems with this particular transformation T, including a property regarding the recurrence of each $k \in \mathbb{N}$ which results from the Birkhoff Ergodic Theorem.

A point $x \in [0,1)$ is said to have a finite Lüroth transformation if there is some k for which $T^{k-1}(x) = 0$. The set of all points in $[0,1)$ with finite expansion is a subset of the rational numbers, and thus has Lebesgue measure zero.

Remark 3.2. In the following sections, we work exclusively with the set of all points $x \in [0,1)$ such that x has an infinite Lüroth expansion i.e. for all k, $T^{k-1}(x) \neq 0$. We define our space X to be the set of these points; we observe $\mu(X) = 1$.

We suppose $x \neq 0$ and for all $k \geq 1$, $T^{k-1}(x) \neq 0$. We define $a_n = a_n(x)$ by

$$a_k(x) = a_1(T^{k-1}(x))$$

where $a_1(x) = n+1$ if $x \in [\frac{1}{n+1}, \frac{1}{n}), n \geq 1$. For convenience, we will write a_1 in place of $a_1(x)$. We redefine our transformation T with these conventions:

$$T(x) = \begin{cases} a_1(a_1-1)x - (a_1 - 1), & \text{x} \neq 0 \\ 0, & \text{x} = 0. \end{cases}$$

It follows that

$$x = \frac{1}{a_1} + \frac{T(x)}{a_1(a_1-1)}$$

and that

$$T(x) = \frac{1}{a_1(T(x))} + \frac{T(T(x))}{a_1(T(x))(a_1(T(x))-1)}.$$

Given $a_2(x) = a_1(T(x))$, we observe

$$\begin{aligned} x &= \frac{1}{a_1} + \frac{1}{a_1(a_1-1)}\left(\frac{1}{a_2} + \frac{T^2(x)}{a_2(a_2-1)}\right) \\ &= \frac{1}{a_1} + \frac{1}{a_1(a_1-1)a_2} + \frac{T^2(x)}{a_1(a_1-1)a_2(a_2-1)}. \end{aligned}$$

For all $k \geq 1$, we proceed inductively and have an infinite series expansion

$$x = \frac{1}{a_1} + \frac{1}{a_1(a_1-1)a_2} + \cdots + \frac{1}{a_1(a_1-1)\cdots a_{n-1}(a_{n-1}-1)a_n} + \cdots$$

where $a_k \geq 2$ for each $k \geq 1$. We show that the series does indeed converge to x. If $S_k(x)$ denotes the sum of the first k terms of the series, then

$$x = S_k(x) + \frac{T^{k-1}(x)}{a_1(a_1-1)\cdots a_{k-1}(a_{k-1}-1)a_k}.$$

Our transformation T is bounded above by 1. We observe also for each k, as $a_k \geq 2$,

$$\frac{1}{a_k(a_k-1)} \leq \frac{1}{2}$$

Therefore,

$$|x - S_k(x)| = \left| \frac{T^{k-1}(x)}{a_1(a_1-1)\cdots a_{k-1}(a_{k-1}-1)a_k} \right| \leq \frac{1}{2^k}$$

Taking the limit as k approaches infinity verifies the convergence.

The following proof is due to [6].

Proposition 3.3. *The Lüroth expansion for T is unique.*

Proof. For convenience, we will denote the Lüroth expansion as an infinite string of digits $d_1d_2d_3\cdots$. Suppose we have two different Lüroth expansions under T for $x \in [0,1)$ i.e. we have two expansions $a_1a_2a_3\cdots$ and $b_1b_2b_3\cdots$ of x such that there exists at least one $N \in \mathbb{N}$ where $a_N \neq b_N$. Let $N \in \mathbb{N}$ be the first digit in the sequence where the two expansions differ.

WLOG, suppose $a_N < b_N$. We denote

$$S_{N-1} = \frac{1}{a_1} + \frac{1}{a_1(a_1-1)a_2} + \cdots + \frac{1}{a_1(a_1-1)\cdots a_{N-1}(a_{N-1}-1)}$$

and define the difference δ between the two expansions:

$$\begin{aligned}
\delta &= S_{N-1}\left(\left(\frac{1}{a_N} - \frac{1}{b_N}\right) + \left(\frac{1}{a_N(a_N-1)a_{N+1}} - \frac{1}{b_N(b_N-1)b_{N+1}}\right) + \cdots\right) \\
&= S_{N-1}\left(\left(\frac{1}{a_N} - \frac{1}{b_N}\right) + \sum_{k=1}^{\infty} \frac{1}{a_N(a_N-1)\cdots a_{N+k}} - \sum_{k=1}^{\infty} \frac{1}{b_N(b_N-1)\cdots b_{N+k}}\right) \\
&> S_{N-1}\left(\left(\frac{1}{a_N} - \frac{1}{b_N}\right) - \sum_{k=1}^{\infty} \frac{1}{b_N(b_N-1)\cdots b_{N+k}}\right)
\end{aligned}$$

For each $k \geq 1$, as $a_k \geq 2$, we observe

$$\frac{1}{a_m(a_m-1)\cdots a_{m+k}} \leq \frac{1}{2^k}$$

It follows

$$\delta > S_{N-1}\left(\frac{b_N - a_N}{a_N b_N} - \frac{1}{b_N(b_N-1)} \sum_{k=1}^{\infty} \frac{1}{2^k}\right) \geq S_{N-1}\left(\frac{1}{a_N b_N} - \frac{1}{b_N(b_N-1)}\right) \geq 0$$

We find that the difference between the two expansions is positive, which contradicts that both expansions converge to x. □

We can think of this series expansion as an approximation of x by intervals $[\frac{1}{n+1}, \frac{1}{n})$. We observe $(0,1) = \bigcup_{n\geq 1}[\frac{1}{n+1}, \frac{1}{n})$. Hence, for all nonzero $x \in [0,1)$, x will fall in one such interval for some $n \geq 1$; we can see $x \geq \frac{1}{n+1}$. What the Lüroth transformation T does is it determines how much greater x is than $\frac{1}{n+1}$ and returns some value $T(x) \in [0,1)$ that indicates the proportion of the interval $[\frac{1}{n+1}, \frac{1}{n})$ where that difference $T(x)$ lies. Iterating this T generates the Lüroth expansion for T.

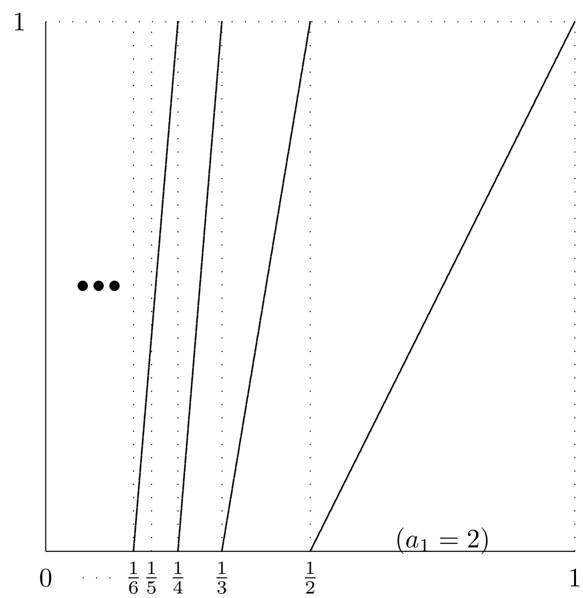

FIGURE 1. The Lüroth series transformation T from [1]

If we were to represent the expansion under T of a point $x \in [0.1)$, we could view the map T as a symbolic "shift" map. For instance, if $x \in [0,1)$ had the expansion $a_1a_2a_3\cdots$, then $T(a_1a_2a_3\cdots) = a_2a_3\cdots$.

3.1. **Ergodic Properties of Lüroth transformation T.** Consider the dynamical system $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a probability space with Lebesgue measure μ, where X is the set of points in $[0,1)$ with infinite expansion, and T is the defined Lüroth transformation. We will compute the average frequency of the appearance of a single positive integer $k \geq 2$ in the expansion of each irrational $x \in [0,1)$.

Proposition 3.4. *T is measure-preserving with respect to Lebesgue measure μ.*

Proof. Suppose $(a,b) \subset [0,1)$. We consider $T^{-1}(a,b) = \{x \in X : T(x) \in (a,b)\}$. We observe for $x \in T^{-1}(a,b)$, $a < T(x) < b$, thus for $n = n(x)$,

$$\frac{1}{n+1} + \frac{a}{n(n+1)} < \frac{1}{n+1} + \frac{T(x)}{n(n+1)} < \frac{1}{n+1} + \frac{b}{n(n+1)}.$$

Given $x = \frac{1}{n+1} + \frac{T(x)}{n(n+1)}$, we see

$$x \in \left(\frac{1}{n+1} + \frac{a}{n(n+1)}, \frac{1}{n+1} + \frac{b}{n(n+1)}\right).$$

Hence,

$$T^{-1}(a,b) = \bigcup_{n \geq 1} \left(\frac{1}{n+1} + \frac{a}{n(n+1)}, \frac{1}{n+1} + \frac{b}{n(n+1)}\right)$$

.

Each of these intervals in the union is disjoint. By the σ-additivity of μ, we find

$$\begin{aligned}\mu(T^{-1}(a,b)) &= \mu\left(\bigcup_{n\geq 1}\left(\frac{1}{n+1}+\frac{a}{n(n+1)},\frac{1}{n+1}+\frac{b}{n(n+1)}\right)\right)\\ &= \sum_{n=1}^{\infty}\mu\left(\frac{1}{n+1}+\frac{a}{n(n+1)},\frac{1}{n+1}+\frac{b}{n(n+1)}\right)\\ &= \sum_{n=1}^{\infty}\frac{b-a}{n(n+1)} = (b-a) = \mu(a,b)\end{aligned}$$

□

We want next to prove that T is ergodic. To do so, we need a lemma, as well as the notion of a *cylinder set* and a few of its properties.

Definition 3.5. A *cylinder set of rank n*, also known as a fundamental interval of rank, or order, n, $\Delta(i_1,\cdots,i_n)$ is the set of all $x\in X$ such that $a_1(x)=i_1$, $a_2(x)=i_2,\cdots,a_n(x)=i_n$.

Recall that X is the set of x with infinite Lüroth expansions under T. Cylinder sets of rank n in the context of the Lüroth transformation T represent the n-th interval approximation of some $x\in X$. Explicitly, for $x\in X$, if we have $A=\frac{1}{i_1}+\frac{1}{i_1(i_1-1)i_2}+\ldots+\frac{1}{i_1(i_1-1)\cdots i_{n-1}(i_{n-1}-1)}$, the cylinder set of rank n $\Delta(i_1,i_2,\cdots,i_n)$ is the interval

$$\left(A, A+\frac{1}{i_1(i_1-1)\cdots i_n(i_n-1)}\right)\cap X.$$

Next, we have a lemma which illustrates a property of the n-th iterate of T applied to a cylinder set of rank n.

Proposition 3.6. $T^n(\Delta(i_1,\cdots,i_n))=[0,1)$.

Proof. The proof of this follows from the fact that T applied to a cylinder set of rank 1 returns $[0,1)$. We assume the proposition holds for n i.e. $T^n(\Delta(i_1,\cdots,i_n))=[0,1)$. Then, we consider $T^{n+1}(\Delta(i_1,\cdots,i_n,i_{n+1})=T^1(T^n(\Delta(i_1,\cdots,i_{n+1}))$, where $T^n(\Delta(i_1,\cdots,i_{n+1})$ is a cylinder set of rank 1, specifically $\Delta(i_{n+1})$. Consequently, $T^{n+1}(\Delta(i_1,\cdots,i_n,i_{n+1})=[0,1)$, and the proposition is proven by induction. □

Next, we introduce a lemma. This lemma is a modified version of a lemma due to Knopp, which can be found in [1]. We notice that the first assumption in the lemma holds for all sufficient semi-rings.

Lemma 3.7. *Suppose $(X,\mathcal{B},\mu)$ is a probability space. Let $B\in\mathcal{B}$ and $\mu(B)>0$. If we have a collection $\mathcal{C}$ of subintervals of $[0,1)$ such that*

(a) *given $\epsilon>0$, for every $A\in\mathcal{B}$, there exists a countable union of disjoint elements of $\mathcal{C}$, denoted C^* such that $\mu(A\Delta C^*)<\epsilon$, and*
(b) *for every $C\in\mathcal{C}$, $\mu(C\cap B)\geq\gamma\mu(C)$, where $\gamma>0$ and is independent of C,*

then $\mu(B)=1$.

Proof. Let E_ϵ be the countable union of sets of $\mathcal{C}$ guaranteed by property (a) i.e. $\mu(B^c\Delta E_\epsilon)<\epsilon$; let $\{S_i\}_{i\in\mathbb{N}}$ be the collection of sets of $\mathcal{C}$ such that $E_\epsilon=\sqcup_{i\in\mathbb{N}}S_i$.

We observe $B \cap E_\epsilon \subset B^c \Delta E_\epsilon \implies \mu(B \cap E_\epsilon) \leq \mu(B^c \Delta E_\epsilon)$. By the σ-additivity of μ and property (b), we find

$$\mu(B \cap E_\epsilon) = \sum_{i=1}^{\infty} \mu(B \cap S_i) \geq \sum_{i=1}^{\infty} \gamma\mu(S_i) = \gamma \sum_{i=1}^{\infty} \mu(S_i) = \gamma\mu(E_\epsilon)$$

It is clear $\gamma\mu(E_\epsilon) \geq \gamma\mu(E_\epsilon \cap B)$ and $\gamma\mu(E_\epsilon) \geq \gamma\mu(E_\epsilon \cap B^c)$. Given $B^c \setminus (B^c \cap E_\epsilon) \subset B^c \Delta E_\epsilon$, it follows

$$\gamma\mu(B^c) = \gamma\left(\mu(B^c \cap E_\epsilon) + \mu(B^c \setminus (B^c \cap E_\epsilon))\right) < \gamma \cdot \epsilon + \epsilon$$

As ϵ is arbitrary and $\gamma > 0$, we have $\mu(B^c) = 0 \Rightarrow \mu(B) = 1$. □

Lemma 3.8. *For every open subinterval (a, b) of $[0, 1)$, $(a, b) \cap X$ is an at most countable union of disjoint cylinder sets.*

Proof. For points $x \in [0, 1)$ with finite expansion of length n, we denote for all $k > n$, $a_k(x) = \infty$. Take $x \in (a, b) \cap X$. Our transformation T gives us that $(a_k)_{k=1}^{\infty}(a) \succ (a_k)_{k=1}^{\infty}(x) \succ (a_k)_{k=1}^{\infty}(b)$.[4] There exists $N \in \mathbb{N}$ such that $\forall n < N$, $a_n(a) = a_n(x)$ and $a_N(a) > a_N(x)$; similarly, there exists $M \in \mathbb{N}$ such that $\forall m < N$, $a_m(x) = a_m(b)$ and $a_M(x) > a_M(b)$. We find that $x \in (a, b)$ exactly if

$$x \in (\bigcup_{N \in \mathbb{N}} \bigcup_{i < a_N(a)} \Delta(a_1(a), \cdots, a_{N-1}(a), i)) \cap (\bigcup_{M \in \mathbb{N}} \bigcup_{j > a_M(b)} \Delta(a_1(b), \cdots, a_{M-1}(b), j))$$

and thus,

$$x \in (\bigcup_{M,N} \bigcup_{i < a_N(a)} \bigcup_{j > a_M(a)} \Delta(a_1(a), \cdots, a_{N-1}(a), i) \cap \Delta(a_1(b), \cdots, a_{M-1}(b), j))$$

As the intersection of two cylinder sets is either another cylinder set or empty, this implies $(a, b) \cap X$ is at most a countable union of disjoint cylinder sets. □

We prove the ergodicity of T.

Theorem 3.9. *T is ergodic.*

Proof. Let B be a T-invariant measurable set and $\mu(B) > 0$. Let $\mathcal{C}$ be the collection of all cylinder sets in $[0, 1)$. Fix $\epsilon > 0$. For any $A \in \mathcal{B}$, there exists a finite sequence of disjoint intervals $I_1, \cdots, I_n$ such that $\mu(A \Delta \sqcup_{i=1}^{n} I_i) < \epsilon$ by Lemma 1.5 as the set of all intervals is a sufficient semi-ring. By Lemma 3.8, for all i such that $1 \leq i \leq n$, I_i is an at most countable union of disjoint cylinder sets i.e. $I_i = \sqcup_{j \in \mathbb{N}} C_{i,j}$ where $C_{i,j} \in C$ for all $j \in \mathbb{N}$. Then, we have

$$\mu(A \Delta \bigsqcup_{i=1}^{n} \bigsqcup_{j \in \mathbb{N}} C_{i,j}) = \mu(A \Delta \bigsqcup_{i=1}^{n} I_i) < \epsilon.$$

Property (a) of Lemma 3.7 is satisfied.

To prove property (b), we observe that T^n is linear on a given cylinder set $A \in \mathcal{C}$ of rank n, and thus of constant slope. So, we find

$$\frac{\mu(T^{-n}(B) \cap A)}{\mu(A)} = \frac{\mu(B \cap T^n(A)}{\mu(T^n(A))} = \mu(B).$$

[4] Here, the symbol $\succ$ indicates lexicographical order.

As B is T-invariant, it follows

$$\frac{\mu(B\cap A)}{\mu(A)} = \frac{\mu(T^{-n}(B)\cap A)}{\mu(A)}.$$

This implies $\mu(A\cap B) = \mu(A)\mu(B)$. Put $\gamma = \mu(B) > 0$, which does not change depending on $A \in \mathcal{C}$. This satisfies property (b) of Lemma 3.7. Then, we apply the lemma and find $\mu(B) = 1$, which proves T is ergodic. □

We can now apply the Birkhoff Ergodic Theorem to arrive at a result about the recurrence of a given integer $k \geq 2$ in the series expansion generated by T. We define the following function

$$f(x) = \begin{cases} 1, & \mathrm{a}_1(x) = k \\ 0, & otherwise \end{cases}$$

Our function f is a characteristic function which is non-zero when $x \in [\frac{1}{k-1}, \frac{1}{k})$. Given T ergodic, we apply the theorem and obtain

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^k(x)) &= \int f d\mu \\ &= \int_{[\frac{1}{k-1},\frac{1}{k})} \mathbb{1} d\mu \\ &= \mu\left[\frac{1}{k-1}, \frac{1}{k}\right) = \frac{1}{k(k-1)} \end{aligned}$$

The average recurrence of a given integer $k \geq 2$ for every $x \in X$ (almost every $x \in [0,1)$) is $\frac{1}{k(k-1)}$. In other words, every $x \in [0,1)$ with an infinite Lüroth expansion under T has asymptotically the same proportion of k in its Lüroth series expansion under T for all $k \geq 2$ i.e. $\frac{1}{2}$ of the digits in the expansion will be 2, $\frac{1}{3(2)} = \frac{1}{6}$ as 3, etc.

This transformation T can be further generalized. Instead of considering fixed partitions $[\frac{1}{n+1}, \frac{1}{n})$, we consider a digit set A, which is an at most countable subset of $\mathbb{N}$ and the partition of $[0,1)$ into intervals $\{L_n = (l_n, r_n) \colon n \in A\}$ such that on each L_n, T is an increasing linear function whose range is an interval with endpoints 0 and 1. We define the corresponding transformation that arises to be a *generalized Lüroth series* transformation. The same ergodic properties of T as defined in (3.1) apply; proofs of such properties and a closer study of these generalized Lüroth series transformations can be found in [1].

4. The Normal Number Theorem

Another application of Birkhoff Ergodic Theorem is the study of normal numbers. We say a number is a *simply normal to base b* if, for every digit $k \in \{0, 1, \cdots, b-1\}$, the average frequency of occurrence of k in the base b expansion of x is $\frac{1}{b}$. We say that a number is *normal to base b* if, for every sequence of m digits, $m \in \mathbb{N}$, the average frequency of occurrence of that sequence in the b-expansion of x is $\frac{1}{b^m}$

Every number which is normal to base b is simply normal to base b by its definition. This result and the definition of a normal number is due to Borel. Borel further proved in 1909 that except for a subset of measure zero, every $x \in [0,1)$ is normal. We will prove this result in this section using ergodic theory.

We formalize our definition of the base b series expansion of $x \in [0,1)$ by defining a transformation

$$T(x) = bx \mod 1 = \begin{cases} bx & \text{x} \in [0, \frac{1}{b}) \\ bx - 1 & \text{x} \in [\frac{1}{b}, \frac{2}{b}) \\ \vdots & \vdots \\ bx - (b-1) & \text{x} \in [\frac{b-1}{b}, 1) \end{cases}$$

Our base b expansion results from iterating this map T. We define $a_1(x) = \lfloor bx \rfloor$ and $a_k(x) = \lfloor bT^{k-1}(x) \rfloor = a_1(T^{k-1}(x))$. From our definition, we have $x = \frac{a_1}{b} + \frac{T(x)}{b}$, and we find

$$\begin{aligned} x = \frac{a_1}{b} + \frac{T(x)}{b} &= \frac{a_1}{b} + \frac{a_2}{b^2} + \frac{T^2}{b^2} \\ &= \frac{a_1}{b} + \frac{a_2}{b^2} + \cdots + \frac{a_k}{b^k} + \cdots \\ &= \sum_{j=1}^{\infty} \frac{a_j(x)}{b^j} \end{aligned}$$

The proof that this series converges to x is straightforward and similar to the case regarding the Lüroth transformation. In fact, the transformation presented in the previous section can be seen as a "generalization" of T, the previously defined map of the base b expansion. The example transformation presented in Section 3 simply does not require that the interval partitions be of equal length.

We revisit the definitions of (simply) normal numbers to b and formalize them. Suppose we have $x \in [0,1)$. For these definitions, we let $N(k,n)$ denote the number of occurrences of k in n digits of the base b series expansion.

Definition 4.1. A number $x \in [0,1)$ is *simply normal to base b* if for every $k \in \{0, 1, \cdots, b-1\}$,

$$\lim_{n\to\infty} \frac{N(k,n)}{n} = \frac{1}{b}$$

A normal number generalizes the previous definition for a finite sequence of digits, as opposed to a singular digit.

Definition 4.2. A number $x \in [0,1)$ is *normal to base b* if for every m-length sequence of digits $k_1 k_2 \cdots k_m$, where $k_1, \cdots, k_m \in \{0, 1, \cdots, b-1\}$,

$$\lim_{n\to\infty} \frac{1}{n} \#\{r \colon 1 \leq r \leq n \text{ and } a_r(x) = k_1, \cdots, a_{r+m-1} = k_m\} = \frac{1}{b^m}.$$

We verify T is indeed measure-preserving and ergodic before we apply the Ergodic theorem.

Proposition 4.3. *T is measure-preserving with respect to Lebesgue measure μ.*

Proof. To show T is Lebesgue measure-preserving, it suffices to show T preserves measure for any open $(a,b) \subset [0,1)$. A straightforward manipulation of our definition T gives us

$$T^{-1}(a,b) = \cup_{i=0}^{n-1} (\frac{i}{n} + \frac{a}{n}, \frac{i}{n} + \frac{b}{n})$$

, which then implies

$$\mu(T^{-1}(a,b)) = (b-a)\sum_{i=0}^{n-1}\frac{1}{n} = b-a = \mu(a,b).$$

□

Proposition 4.4. *T is ergodic with respect to Lebesgue measure μ.*

Proof. We consider T when $b = 2$. This is the transformation that generates the base 2 series expansion, which is also referred to as the "doubling map."

We will use our Lemma 3.11 to prove this result, in a similar fashion as we did in the previous section for the Lüroth transformation T. Property (a) of Lemma 3.7 follows from Lemma 1.5.

To prove property (b), we let $\mathcal{C}$ be the collection of dyadic intervals, where the dyadic interval is defined as $D_{n,k} = \{\frac{k}{2^n}, \frac{k+1}{2^n}\}$, for $n > 0$, $k = 0, 1, \cdots, 2^n - 1$. We observe several properties about these intervals.

We find $T^n(D_{n,k}) = [0,1)$, and thus, $T^{-n}(D_{n,k})$ consists of 2^n disjoint dyadic intervals, each of length 2^{-2n}. Next, we claim for any measurable set A,

$$\mu(T^{-n}(A) \cap D_{n,k}) = \mu(A)\mu(D_{n,k}).$$

We proceed by induction on n. Consider

$$\mu(T^{-1}(A) \cap D_{1,k}) = \mu(T^{-1}(A) \cap [\frac{k}{2}, \frac{k+1}{2}))$$

where $k = 0, 1$. Since A is Lebesgue measurable, it can be approximated up to a set of measure zero by the union of a countable disjoint collection of open intervals $I_1, \cdots, I_n, \cdots$. We denote $I_i = (a_i, b_i)$. Then,

$$T^{-1}(I_i) = T^{-1}(a_i, b_i) = (\frac{a_i}{2}, \frac{b_i}{2}) \cup (\frac{a_i}{2} + \frac{1}{2}, \frac{b_i}{2} + \frac{1}{2}).$$

Depending on k, either $(\frac{a_i}{2}, \frac{b_i}{2}) \subset D_{1,k}$ and $(\frac{a_i}{2} + \frac{1}{2}, \frac{b_i}{2} + \frac{1}{2}) \cap D_{1,k} = \emptyset$ or $(\frac{a_i}{2} + \frac{1}{2}, \frac{b_i}{2} + \frac{1}{2}) \subset D_{1,k}$ and $(\frac{a_i}{2}, \frac{b_i}{2}) \cap D_{1,k} = \emptyset$. So, it follows

$$\mu(T^{-1}(I_i) \cap D_{1,k}) = \frac{1}{2}\mu(I_i) = \mu(D_{1,k})\mu(I_i)$$

which implies

$$\begin{aligned}\mu(T^{-1}(A) \cap D_{1,k}) = \mu(T^{-1}(\bigcup_{i=1}^{n} I_i) \cap D_{1,k}) &= \mu(\bigcup_{i=1}^{\infty}(T^{-1}(I_i) \cap D_{1,k}))\\ &= \frac{1}{2}\mu(\bigcup_{i=1}^{\infty} I_i)\\ &= \frac{1}{2}\mu(A) = \mu(D_{1,k})\mu(A).\end{aligned}$$

We suppose the claim holds for some $n \in \mathbb{N}$. We consider

$$\mu(T^{-(n+1)}(A) \bigcap D_{n+1,k}) = \mu(T^{-n}(T^{-1}(A)) \cap D_{n+1,k}).$$

We observe

$$D_{n,k} = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) = \left[\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}\right) \cup \left[\frac{k+1}{2^{n+1}}, \frac{k+2}{2^{n+1}}\right) = D_{n+1,k} \cup D_{n+1,k+1}$$

where $D_{n+1,k}$ and $D_{n+1,k+1}$ are clearly disjoint and of the same length. It follows

$$\begin{aligned}\mu(T^{-n}(T^{-1}(A)) \cap D_{n,k}) &= \mu(T^{-n}(T^{-1}(A)) \cap (D_{n+1,k} \cup D_{n+1,k+1})) \\ &= \mu(T^{-n}(T^{-1}(A)) \cap D_{n+1,k}) + \mu(T^{-n}(T^{-1}(A)) \cap D_{n+1,k+1}).\end{aligned}$$

As the claim holds for arbitrary k, this implies

$$\begin{aligned}\mu(T^{-n}(T^{-1}(A)) \cap D_{n+1,k}) &= \frac{1}{2}\mu(T^{-n}(T^{-1}(A)) \cap D_{n,k}) \\ &= \frac{1}{2}\mu(D_{n,k})\mu(T^{-1}(A)) \\ &= \mu(D_{n+1,k})\mu(A).\end{aligned}$$

By induction on n, the claim is proven.

Suppose A is a T-invariant set i.e. $T^{-1}(A) = A$ such that $\mu(A) > 0$. Then

$$\mu(A \cap D_{n,k}) = \mu(A)\mu(D_{n,k}).$$

We let $\gamma = \mu(A)$ as in Lemma 3.7, and property (b) holds. Then, by Lemma 3.7, $\mu(A) = 1$, and T is ergodic when $n = 2$.

The case for the general base b expansion follows similarly. In place of the dyadic interval, we define an interval $A_{b,n,k} = [\frac{k}{b^n}, \frac{k+1}{b^n})$ where $n > 0$, $k = 0, 1, \cdots, b^n - 1$. The collection $\mathcal{C}$ of all such intervals satisfies property (a), and we find

$$\mu(A \cap A_{b,n,k}) = \mu(A)\mu(A_{b,n,k})$$

by induction. An application of Lemma 3.7 again proves that for a general $b \geq 2$, the base b transformation map is ergodic with respect to Lebesgue measure μ. □

We apply the Birkhoff Ergodic Theorem.

Theorem 4.5. *Almost every real $x \in [0, 1)$ is normal.*

Proof. Given $a_k(x) = a_1(T^{k-1}(x))$ for $k \geq 2$, we have $a_j(T^k(x)) = a_{j+k}(x)$. For any $n \in \mathbb{N}$,

$$\begin{aligned}a_j(x) = k_1 &\iff a_1(T^{j-1}(x)) = k_1, \\ a_{j+1}(x) = k_2 &\iff a_2(T^{j-1}(x)) = k_2, \\ &\vdots \\ a_{j+n-1}(x) = k_n &\iff a_n(T^{j-1}(x)) = k_n\end{aligned}$$

By our definition, $a_1(T^{j-1}(x)) = k_1 \iff T^{j-1}(x) \in [\frac{k_1}{b}, \frac{k_1+1}{b})$ and $a_n(T^{j-1}(x)) = k_n \iff T^{j-1}(x) \in [\frac{k_n}{b^n}, \frac{k_n+1}{b^n})$. Equivalently, $T^{j-1}(x) \in [\sum_{i=1}^n \frac{k_i}{b^i}, \sum_{i=1}^n \frac{k_i}{b^i} + \frac{1}{b^n})$. We define a characteristic function

$$f(x) = \begin{cases} 1 & \mathrm{T}^{j-1}(x) \in \left[\sum_{i=1}^n \frac{k_i}{b^i}, \sum_{i=1}^n \frac{k_i}{b^i} + \frac{1}{b^n}\right) \\ 0 & otherwise \end{cases}$$

We apply the Birkhoff Ergodic Theorem to f and arrive at the following:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^k(x)) = \int f d\mu = \mu\left(\left[\sum_{i=1}^n \frac{k_i}{b^i}, \sum_{i=1}^n \frac{k_i}{b^i} + \frac{1}{b^n}\right)\right) = \frac{1}{b^n}$$

That the average occurrence of every n-length finite sequence of digits $k_1 k_2 \cdots k_n$ is $\frac{1}{b^n}$ proves that almost every $x \in [0, 1)$ is normal. □

Acknowledgements

It is my pleasure to thank my mentor Jonathan Sheperd for providing very thorough feedback for my many drafts and answering my many questions. I also would like to thank Kathryn Lindsey for introducing me to ergodic theory and for working with me through learning the material. This paper would not have been possible without their guidance.

List of Figures

References

[1] Karma Dajani and Cor Kraaikamp. Ergodic Theory of Numbers. Mathematical Association of America. 2002.
[2] Paul R. Halmos. Lectures on Ergodic Theory. Chelsea Publishing Company. 1956.
[3] Rodney Nillsen Randomness and Recurrence in Dynamical Systems. Mathematical Association of America. 2010.
[4] Karl Petersen. Ergodic theory. Cambridge University Press. 1983.
[5] C.E. Silva. Invitation to Ergodic Theory. American Mathematical Society. 2008.
[6] Yulia Zhykharyeva and Mykola Pratsyioviti. *Expansions of numbers in positive Lüroth series and their applications to metric, probabilistic and fractal theories of numbers.* Algebra and Discrete Mathematics. **40** (2012), 145-160.