AN ABSTRACT OF THE THESIS OF

Daniel Luke Rockwell for the degree of Doctor of Philosophy in Mathematics

presented on September 9, 2011.

Title: A Reinterpretation, and New Demonstrations of, the Borel Normal Number Theorem

Abstract approved: ______________________________

Robert M. Burton

The notion of a normal number and the Normal Number Theorem date back over 100 years. Émile Borel first stated his Normal Number Theorem in 1909. Despite their seemingly basic nature, normal numbers are still engaging many mathematicians to this day. In this paper, we provide a reinterpretation of the concept of a normal number. This leads to a new proof of Borel's classic Normal Number Theorem, and also a construction of a set that contains all absolutely normal numbers. We are also able to use the reinterpretation to apply the same definition for a normal number to any point in a symbolic dynamical system. We then provide a proof that the Fibonacci system has all of its points being normal, with respect to our new definition.

A Reinterpretation, and New Demonstrations of, the Borel Normal Number Theorem

by

Daniel Luke Rockwell

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 9, 2011
Commencement June 2012

Doctor of Philosophy thesis of Daniel Luke Rockwell presented on September 9, 2011

APPROVED:

Major Professor, representing Mathematics

Chair of the Department of Mathematics

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Daniel Luke Rockwell, Author

ACKNOWLEDGEMENTS

Academic

I would like to thank my advisor, Robert Burton, for providing the guidance I deeply needed in my mathematical research and in my life. I would also like to thank all of my classmates who energized my passion for mathematics and challenged me to be the best damn mathematician I can be.

Personal

I want to thank my beautiful wife, Jacqueline Rockwell, for her support, love, and compassion throughout my time at OSU. I also want to thank my daughter, Eleanor Rockwell, for giving me the inspiration to keep going in my darkest hours. I also want to thank my loving parents for all of the support and love they have given me. Also, my sister Kat Rockwell was invaluable in helping me edit my paper for grammatical errors.

I would also like to dedicate this work to my late brother, Eric Rockwell. I miss you every day.

TABLE OF CONTENTS

Page

1. INTRODUCTION 1

2. NORMAL NUMBERS 3

    2.1. Normal Numbers 3

3. ERGODIC THEORY 7

    3.1. Probability basics 7

    3.2. Ergodic Theory 14

4. SYMBOLIC DYNAMICAL SYSTEMS 17

    4.1. Symbolic Dynamical Systems 17

    4.2. A Proof of the Normal Number Theorem 21

    4.3. Entropy of a Partition 24

    4.4. Shift Spaces 28

    4.5. Entropy of Shift Spaces 32

5. INFINITE WORDS 34

    5.1. The Fibonacci System 34

    5.2. Sturmian Words 37

        5.2.1 Complexity 41

        5.2.2 Geometric Interpretation of Characteristic Words 42

    5.3. Proof That the Fibonacci System is Normal 43

        5.3.1 Connection to Certain Sturmian Words 60

TABLE OF CONTENTS (Continued)

	Page
6. CONCLUSION AND FUTURE WORK	62
BIBLIOGRAPHY	64
APPENDICES	66
A APPENDIX Some Results on Normal Numbers	67
A1 Finite Variations	67
A2 Equidistribution of Sequences	67
A3 Normality from Simple Normality	68
A4 Closed Under Multiplication by a Rational	68
B APPENDIX Fibonacci Numbers	69
C APPENDIX Fibonacci Blocks	70
INDEX	71

A REINTERPRETATION, AND NEW DEMONSTRATIONS OF, THE BOREL NORMAL NUMBER THEOREM

1. INTRODUCTION

When Émile Borel first postulated the Normal Number Theorem [1] did he imagine that over 100 years later mathematicians would still be tinkering with his novel concept? Regardless of what Borel imagined, mathematicians keep extending the idea of a normal number to other objects and studying them. Mathematicians have also worked hard to produce numbers that are normal to a base b. The author of this paper thinks that É. Borel did have a vision of this future, a future that exalts his Normal Number on high.

This paper will take the reader through a journey of sorts. The reader will come to learn exactly what a normal number is and why it is so interesting and special. We will explore the amazing properties of normal numbers and how they have helped mathematicians better understand the real number system. The main difficulty associated with normal numbers is determining whether or not a real number x is normal or not.

The Normal Number Theorem states that almost all points in the unit interval are normal to a base b. A real number is normal to the base b if the digits in the base b expansion occur in a uniform way. These statements will be made more precise in the sections to come. A normal number can be thought of as a number that has every pattern of digits occurring, and furthermore, the patterns occur with the proper frequency in the expansion.

We will see how symbolic dynamical systems are a natural way to represent the real numbers, as well as being able to represent many other spaces. This natural representation, using symbolic dynamical systems, of the real numbers gives way to a natural analogue of a normal number and the Normal Number Theorem. Given the more general nature of symbolic dynamical systems, our analogue of a normal number is extended. We will provide a unique proof of the Normal Number Theorem, constructing a set that contains all of the normal numbers.

Finally, we will discuss how entropy is related to normality and how certain systems with 0-entropy behave slightly differently under the normal number criteria. We will examine the Fibonacci dynamical system in detail, providing a proof that the Fibonacci system has every point being normal according to our new definition.

2. NORMAL NUMBERS

2.1. Normal Numbers

The idea of a normal number comes from a curious and persistent result that Borel proved in 1909[1]. What Borel's paper showed was that the set of normal numbers in the unit interval has Lebesgue measure 1. In other words, almost all real numbers between 0 and 1 are normal.

Theorem 2.1.0.1. ***(Normal Number Theorem)***

The set of normal numbers in the unit interval has Lebesgue measure 1.

What has proven to be continually intriguing about normal numbers is that, though they are easily shown to exist in large quantities, it has proven difficult to prove that a specific number has the normal number property. For instance, it remains unknown if π, e, or $\sqrt{2}$ are normal numbers. Some might say that this type of existence proof is the hallmark of 20^{th} century mathematics. That is, being able to show there is a large class of objects with a certain property but being unable to prove that a specific object has the purported property. Before we talk too much about normal numbers, we should first define what it means for a real number to be normal.

Let α be a real number, and let $b \geq 2$ be an integer. Then α has a unique b-adic expansion called the base b expansion and is of the form

$$\alpha = \lfloor \alpha \rfloor + \sum_{n=1}^{\infty} \frac{a_n}{b^n}.$$

Here the digits a_n are integers with the following constraints. For $n \geq 1$ we have $0 \leq a_n < b$ and $a_n < b-1$ for infinitely many n. Let a be a digit with respect to the base b and let N be a positive integer. Then define $\#_N(a, b, \alpha)$ to be the number of n with $1 \leq n \leq N$ such that $a_n = a$. We can extend this definition to blocks of digits $B_j = b_1 b_2 ... b_j$ in the following way: let $\#_N(B_j, b, \alpha)$ be the number of k, $1 \leq k \leq N - j + 1$, such that $a_{k+j-1} = b_k$ for $1 \leq k \leq j$. We also refer to blocks as words. If it is not clear we will say, B is a finite word, meaning a word with finite length. The function $\#_N(B_j, b, \alpha)$ can be viewed as counting the number of distinct occurrences (perhaps with overlap) of the block B_j in the first N digits of α's b-adic expansion.

Definition 2.1.0.1. *The real number α is called **simply normal** to the base b if*

$$\lim_{N \to \infty} \frac{\#_N(a, b, \alpha)}{N} = \frac{1}{b} \qquad \text{for } a \in \{0, 1, 2, ..., b-1\}.$$

Definition 2.1.0.2. *The real number α is called **normal to the base b** if*

$$\lim_{N \to \infty} \frac{\#_N(B_j, b, \alpha)}{N} = \frac{1}{b^j} \qquad \text{for all } j \geq 1 \text{ and all } B_j.$$

Definition 2.1.0.3. *The real number α is called **absolutely normal** if α is normal to every base $b \geq 2$.*

As we can see, the integer part of the real number α plays no role in whether or not α is normal. In fact, changing or removing any finite number of digits will maintain normality for the resulting sequence (See Appendix A). Therefore, we will limit ourselves to the set of real numbers in the unit interval, $[0, 1)$.

We will refer to the limit in definition 2.1.0.2 as the *frequency of the block* B_j in α denoted $Freq_{B_j}(\alpha)$ or more simply $Freq_B(\alpha)$. We will usually suppress the subscript on the block name and simply denote a block by B with the length inferred from context. When needed we will use $|B|$ to denote the length of the finite block B.

It is obvious that a number that is normal to the base b is also simply normal to the base b, since 1 is a possible length of a block. The converse is not true, however. Consider the one-sided infinite word $(0, 1, 0, 1, 0, 1, 0, ...)$ that alternates 0's and 1's. This number is clearly simply normal to base 2. However, it is not normal since the block $B = 11$ never occurs. In fact, there are uncountably many real numbers that are not normal. Consider a decimal expansion that has no 2's in it. The set of all of these decimals is uncountable and clearly contains only numbers that are not normal. A real number that is normal to any base is necessarily irrational, since rational numbers have finite or periodic expansions in any base. But it is not the case that all irrationals are normal or simply normal. For example, think of any irrational number in the base 3 Cantor set. It does not have a 1 in its base 3 expansion at all.

There are a few constructions of specific normal numbers that are known. The first widely known example is due to Champernowne [3] and is now referred to as Champernowne's number. It is simply the concatenation of the natural numbers in base 10, 0.123456789101112131415161718.... Champernowne [3] proved that this number is normal to the base 10. This type of construction can create a number normal to any integer base, $b \geq 2$. For instance, if we concatenate the natural numbers represented in base 2 we get 0.11011100101110... which is a real number that is normal to the base 2.

Champernowne conjectured, and it has since been proven in [6], that the concatenation of the primes, 0.23571113171923293137..., also known as the Copeland-Erdös constant, yields a normal number to the base 10. In a paper by Copeland and Erdös they classify the types of subsets of the natural numbers that when concatenated in increasing order produce a normal number.[6] These subsets are dense in the natural numbers. A set $D \subseteq \mathbb{N}$ is *dense* if, for every $\delta < 1$ and for all sufficiently large n, $|D \cap \{1, 2, ..., n\}| \geq n^{\delta}$. Their result states that taking the base b expansion of the elements of a dense set $D \subseteq \mathbb{N}$ and concatenating them in increasing order yields a number normal to the base b.

For a long time it was widely believed that there was not an example of an absolutely normal number. Even to this day some people still think that such an example does not exist. This is, however, untrue. In 1917 Sierpinski, see [2] and [7], proposed a construction of a number that would be absolutely normal. This construction does indeed provide a construction of an absolutely normal number.

Remark 2.1.0.1. *If we forget that the sequences $(a_n)_{n=1}^{\infty}$ can be viewed as b-adic expansions of real numbers, then normality can simply be viewed as a property of sequences of symbols from a finite set of size b. This is an extremely useful viewpoint to have for the discussion in section 4.*

3. ERGODIC THEORY

3.1. Probability basics

It was previously mentioned that the examination of normal numbers can be restricted to real numbers in the unit interval. The unit interval can be equipped with tools that make it a finite measure space. Moreover, the unit interval can easily be seen as a probability measure space. It is because of this deep connection that probability theory lends itself naturally to the ongoing study of normal numbers. In this section we will introduce the basics of probability theory in a way that will mesh with the theory we will use later in the paper. For a deeper discussion of probability theory please refer to [15] or [16].

As we introduce the basics of probability theory we will try to keep the terminology as general as possible to maintain a larger audience for this paper.

Definition 3.1.0.4. *Given a set X equipped with a σ-algebra, Σ, of subsets of X and a finite measure, μ, on X, (i.e. $\mu(X) < \infty$), we call the triple (X, Σ, μ) a* ***finite measure space****. Moreover, if we have that $\mu(X) = 1$ then we call the triple, (X, Σ, μ), a* ***probability measure space*** *or more simply a* ***probability space*** *and μ a* ***probability measure****.*

We can view any finite measure space as a probability measure space by simply defining a new measure $\nu = \frac{1}{\mu(X)}\mu$. The sets in the σ-algebra are called *measurable sets*. Given a measurable set E we think of $\mu(E)$ as the probability of E.

Definition 3.1.0.5. *Given two measurable sets E_1 and E_2, we say that E_1 and E_2 are* ***independent*** *if $\mu(E_1 \cap E_2) = \mu(E_1)\mu(E_2)$. For a finite collection of measurable* *sets $E_1, E_2, ..., E_k$ we say they are independent if*

$$\mu(E_{j_1} \cap E_{j_2} \cap ... \cap E_{j_m}) = \mu(E_{j_1})\mu(E_{j_2})...\mu(E_{j_m})$$

for all possible subsets $\{j_1, j_2, ..., j_m\} \subseteq \{1, 2, ..., k\}$, of distinct elements and $1 \leq m \leq k$.

Example 3.1.0.1. Suppose we had 3 sets, D, E, F. They would be independent if and only if all of the following four conditions are met:
(i) $\mu(D \cap E \cap F) = \mu(D)\mu(E)\mu(F)$, **(ii)** $\mu(D \cap E) = \mu(D)\mu(E)$
(iii) $\mu(E \cap F) = \mu(E)\mu(F)$, **(iv)** $\mu(D \cap F) = \mu(D)\mu(F)$.

In order to talk about the independence of an infinite collection of measurable sets we must consider all possible sub-collections. We refer the reader to [16] (pg. 68) for a more thorough treatise of this delicate concept.

Definition 3.1.0.6. *Given a finite measure space (X, Σ, μ), a function $f : X \to \mathbb{R}$ is **measurable** (with respect to μ) if for all measurable sets $A \subseteq \mathbb{R}$ then $f^{-1}(A) \in \Sigma$.*

Up to this point we have not discussed what the measurable sets of $\mathbb{R}$ are. In most cases it is of little importance to us which type of measurable sets of $\mathbb{R}$ are to be considered. We typically use one of two classical constructions for a measure on $\mathbb{R}$. The first is Borel measure and the Borel measurable sets. The second is Lebesgue measure and Lebesgue measurable sets. These are both useful because the measure of an interval is its length.

Definition 3.1.0.7. *A real valued **random variable**, f, is a measurable function from X into $\mathbb{R}$. Given a measurable set $A \subseteq \mathbb{R}$, we denote the set $\{x \in X \mid f(x) \in A\}$ as $\{f \in A\}$ for simplicity. For convenience we will use the following notation.*

$$\mu(f \in A) = \mu(\{x \in X \mid f(x) \in A\})$$

We refer to $\mu(f \in A)$ as the probability that f is in A.

There are a few specific random variables that are of special significance in probability theory, so much so that they have special names.

Definition 3.1.0.8. *The **indicator function** or **characteristic function** of a set A denoted $\mathbb{1}_A$ is defined by*

$$\mathbb{1}_A(x) = \begin{cases} 0 & \text{if, } x \notin A \\ \\ 1 & \text{if, } x \in A. \end{cases}$$

The indicator function is measurable, since a measurable set in $\mathbb{R}$ can either contain both 0 and 1 or contain just one of 0 or 1 or contain neither. These three different cases yield the measurable sets: X, A, A^c, and $\varnothing$. The indicator function is an example of a discrete random variable taking the values 0 or 1. The indicator function helps simplify notation for random variables that would otherwise be cumbersome to work with as we will see later in our proof of the Normal Number Theorem.

Definition 3.1.0.9. *A **discrete random variable**, $f : X \to \mathbb{R}$, takes only finitely many distinct values in $\mathbb{R}$.*

An example of a discrete random variable is a *Bernoulli random variable*. Given $0 \leq p \leq 1$ we define the Bernoulli random variable, f, such that $\mu(f \in \{1\}) = p$ and $\mu(f \in \{0\}) = 1 - p$.

Definition 3.1.0.10. *Let $\mathcal{M}$ be the set of measurable sets of $\mathbb{R}$. Then given two random variables f_1 and f_2. We say f_1 and f_2 are **independent random variables** if for all $E \in \mathcal{M}$ the sets $f_1^{-1}(E)$ and $f_2^{-1}(E)$ are independent.*

Similarly, we may extend this definition to finite collections of random variables and infinite collections of random variables. This is done in such a way as to be analogous to the definition of independent sets. We should also note that the σ-algebra being used plays a key role in whether a collection of random variables are independent or not.

Definition 3.1.0.11. *Given a random variable f we define the **distribution** of f by $\mu \circ f^{-1}$.*

Example 3.1.0.2. Let f be a discrete random variable taking values in a set $\{a_1, a_2, ..., a_n\}$. Then f has a *uniform distribution* if $\mu(f \in \{a_i\}) = \frac{1}{n}$ for all $1 \leq i \leq n$. Or we could say, the probability that $f = a_i$ is $\frac{1}{n}$ for all $1 \leq i \leq n$.

The uniform distribution is very appealing since it lends itself to many real world situations. For instance, the random variable that represents the outcome of a fair coin flip is a uniform distribution on the set $\{-1, 1\}$, where -1 stands for heads and 1 stands for tails. The uniform distribution can also be used for describing picking a card out of a well shuffled deck of playing cards. There is a $\frac{1}{52}$ chance of picking a card with a specific suit and number.

Lemma 3.1.0.1. *[16] The distribution of a random variable is a probability measure on $\mathbb{R}$.*

Again, in this we are not specifying which sets are measurable. Usually the context will make it clear which sets are measurable in $\mathbb{R}$. In fact, it will be the same sets that the random variable f is measurable with respect to. At the core of the proof for Lemma 3.1.0.1 is the fact that μ is a probability measure. All of the needed attributes are inherited from μ and are preserved by the measurable f.

A common abbreviation used in probability theory when dealing with a collection or sequence of *independent and identically distributed* random variables is i.i.d. This means that all of the random variables in the collection have identical distributions and the collection of random variables is independent.

Definition 3.1.0.12. *Given a random variable $f : X \to \mathbb{R}$, then the **expectation** of f is the integral of f with respect to the measure μ denoted*

$$\mathbb{E}(f) = \int_X f d\mu.$$

Proposition 3.1.0.1. *[19] Given independent random variables f and g,*

$$\mathbb{E}(fg) = \mathbb{E}(f)\mathbb{E}(g).$$

This proposition can be extended to finite and infinite collections of random variables. This extension is analogous to the definition of independence for collections of random variables.

Theorem 3.1.0.2. ***(Markov's Inequality)***

Given a probability measure space (X, Σ, μ), let f be a random variable on X then for all $\epsilon > 0$,

$$\mu(|f| > \epsilon) \leq \frac{\mathbb{E}(|f|)}{\epsilon}.$$

Proof. Let A be the set where $|f| > \epsilon$. Then $\mathbb{1}_A \leq \frac{|f|}{\epsilon}$. Taking expectations on both sides yields

$$\mathbb{E}(\mathbb{1}_A) \leq \frac{\mathbb{E}(|f|)}{\epsilon}.$$

Since $\mathbb{E}(\mathbb{1}_A) = \mu(A) = \mu(|f| > \epsilon)$ the proof is complete. □

Markov's inequality can be used to prove Chebyshev's inequality for even integer values of p. Since $|f|^p \geq \epsilon^p$ if and only if $|f| \geq \epsilon$ for p an even integer.

Theorem 3.1.0.3. *[19] (**Chebyshev's Inequality**) For all $p, \epsilon > 0$ and for all random variables $f \in L^p(\mu)$, i.e. $\mathbb{E}(|f|^p) < \infty$,*

$$\mu(|f| \geq \epsilon) \leq \frac{1}{\epsilon^p} \int_{\{|f| \geq \epsilon\}} |f|^p d\mu \leq \frac{(\mathbb{E}(|f|^p))}{\epsilon^p}.$$

Here $L^P(\mu)$ is the set of all random variables such that $\int |f|^p d\mu < \infty$. Instead of providing a proof of Chebyshev's inequality, we will provide a proof for a similar inequality that will be useful when we provide a proof of the Normal Number Theorem.

Lemma 3.1.0.2. *Given a probability measure space (X, Σ, μ) and a sequence of i.i.d. random variables $\{f_i\}_{i=0}^{\infty}$, let $S_N = f_0 + f_1 + ... + f_{N-1}$. Then for all $\epsilon > 0$,*

$$\mu\left(\left\{x \in X \;\middle|\; \left|\frac{S_N}{N} - \mathbb{E}(f_0)\right| > \epsilon\right\}\right) \leq \frac{\mathbb{E}(S_n^4)}{N^4\epsilon^4}.$$

Proof. Since $|S_N - N\mathbb{E}(f_0)|^4$ is a random variable we can apply Markov's inequality to obtain

$$\mu\left(\left\{x \in X \;\middle|\; |S_N - N\mathbb{E}(f_0)|^4 > N^4\epsilon^4\right\}\right) \leq \frac{\mathbb{E}(S_n^4)}{N^4\epsilon^4}.$$

This is equivalent to

$$\mu\left(\left\{x \in X \;\middle|\; \left|\frac{S_N}{N} - \mathbb{E}(f_0)\right|^4 > \epsilon^4\right\}\right) \leq \frac{\mathbb{E}(S_n^4)}{N^4\epsilon^4}.$$

Then since $\left|\frac{S_N}{N} - \mathbb{E}(f_0)\right|^4 > \epsilon^4$ if and only if $\left|\frac{S_N}{N} - \mathbb{E}(f_0)\right| > \epsilon$ we have

$$\mu\left(\left\{x \in X \;\middle|\; \left|\frac{S_N}{N} - \mathbb{E}(f_0)\right| > \epsilon\right\}\right) = \mu\left(\left\{x \in X \;\middle|\; \left|\frac{S_N}{N} - \mathbb{E}(f_0)\right|^4 > \epsilon^4\right\}\right).$$

Therefore

$$\mu\left(\left\{x \in X \;\middle|\; \left|\frac{S_N}{N} - \mathbb{E}(f_0)\right| > \epsilon\right\}\right) \leq \frac{\mathbb{E}(S_N^4)}{N^4\epsilon^4}.$$

and the proof is complete. $\square$

Lemma 3.1.0.3. *Given the hypotheses of lemma 3.1.0.2 and $f_i \in L^4(\mu)$*

$$\lim_{N\to\infty} \frac{\mathbb{E}(S_N^4)}{N^4\epsilon^4} = 0.$$

Proof. We can assume, without loss of generality, that $\mathbb{E}(f_i) = 0$. Otherwise we can replace the random variable f_i with the random variable $f_i - \mathbb{E}(f_i)$. Consider $\mathbb{E}(S_N^4) = \mathbb{E}((f_0 + f_1 + ... + f_{N-1})^4)$; the sum will have terms of the form $f_i^4,\ f_i f_j^3,\ f_i^2 f_j^2,\ f_i f_j f_k^2,$ and $f_i f_j f_k f_l$ with i, j, k, l all distinct. Since we have assumed the f_i's are independent and have expectation 0, most of the terms are zero. Terms of the form $f_i f_j^3,\ f_i f_j f_k^2,$ and $f_i f_j f_k f_l$ will have their expectation be zero. For a distinct pair i, j there are 6, which is "4 choose 2", terms of the form $f_i^2 f_j^2$. Thus we have

$$\mathbb{E}\left(S_N^4\right) = \mathbb{E}\left(\left(\sum_{i=0}^{N-1} f_i\right)^4\right) \tag{3.1}$$

$$= \mathbb{E}\left(\sum_{i=0}^{N-1} f_i^4 + 6\sum_{i\neq j} f_i^2 f_j^2\right) \tag{3.2}$$

$$= N\mathbb{E}\left(f_i^4\right) + 6\frac{N(N-1)}{2}\mathbb{E}\left(f_i^2 f_j^2\right). \tag{3.3}$$

Having $f_i \in L^4(\mu)$ implies that $f_i \in L^2(\mu)$. Also independence implies that $\mathbb{E}\left(f_i^2 f_j^2\right) = \left(\mathbb{E}\left(f_i^2\right)\right)^2$. Therefore we have $\mathbb{E}\left(S_N^4\right) = N\mathbb{E}\left(f_i^4\right) + 3N(N-1)\left(\mathbb{E}\left(f_i^2\right)\right)^2 < \infty$. The values $\mathbb{E}\left(f_i^4\right)$ and $\left(\mathbb{E}\left(f_i^2\right)\right)^2$ are constants. Let C be the maximum of the two values. Then

$$\lim_{N\to\infty} \frac{\mathbb{E}(S_N^4)}{N^4\epsilon^4} = \lim_{N\to\infty} \frac{NC + 3N(N-1)C}{N^4\epsilon^4} = 0.$$

$\square$

3.2. Ergodic Theory

The notion of probability has seeped so far into our language that it is not uncommon to hear someone say, "What are the chances that was going to happen?" or "What are the odds?". A naive answer to the first question is $\frac{1}{2}$; it either was going to happen or it wasn't. Describing why this answer is naive can be cumbersome. Sometimes the answer is correct. For example, if the question was referring to the result of a fair coin flip, then the answer of $\frac{1}{2}$ is completely correct. The probability of the coin landing with the heads side up is $\frac{1}{2}$, with the only other option being that the coin landed with the tails side up.

With great care, the naivety of the answer can be exposed and a deeper understanding of how randomness behaves can be found. There are quite a few different ways to define randomness. At the core of these definitions is the belief that randomness is a limitation of the ability to predict what is going to happen next or predict what object will be chosen out of a set.

Fair coin flips are inherently independent of the past. Knowing all the results of previous coin flips does not give a clue as to what the next coin flip will be. However, if it is unknown to an observer that a coin is biased to show heads 95% of the time and show tails the other 5% of the time, then the observer can infer, over the course of many coin flips, that a better guess for the result of the coin flip is heads. Since the observer will notice the disproportionate number of heads occurring, this would be seen by the observer as less random than a fair coin flip.

It is this idea that the repetition of a random experiment will give us a global understanding that is at the core of the ergodic theorem. In fact, it is often talked about as showing that the time average is the same as the space average. We will make all of this more precise with the following definitions. As we will see, the entropy of a system is a gauge as to how well we can predict the next outcome given the results of all previous outcomes.

Definition 3.2.0.13. *Given a finite measure space (X, Σ, μ) with $\mu(X) < \infty$ and a measurable transformation $T : X \to X$, T is said to be* ***measure preserving*** *if for all $E \in \Sigma$ we have that $\mu(T^{-1}(E)) = \mu(E)$.*

Notice that for a transformation to be measure preserving, it does not need to send the points of a set E to a set with the same measure as E. What needs to happen is that a set E needs to receive the proper amount of points under the transformation T. Also, we can conclude that $\mu(T^{-n}(E)) = \mu(E)$ for all n by repeatedly applying the definition.

Theorem 3.2.0.4. ***(Poincaré Recurrence)***

Given a finite measure space (X, Σ, μ) and a measure preserving transformation $T : X \to X$, for any $E \in \Sigma$ with $\mu(E) > 0$ then there exists some $x \in X$ such that $T^n(x) \in E$ for some $n > 0$.

Proof. Assume for a contradiction there is no point $x \in E$ such that $T^n(x) \in E$ for some n. Therefore $T^{-n}(E) \cap E = \varnothing$ for all $n > 0$. Also $T^{-n}(E) \cap T^{-m}(E) = \varnothing$ for all $n \neq m$. Therefore,

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

since $\mu(E) > 0$ and T is measure preserving. But we have that $\mu(X) < \infty$ and $\bigcup_{n=1}^{\infty} T^{-n}(E) \subseteq X$. Thus we have reached our contradiction. Therefore, the theorem holds.

$\square$

Corollary 3.2.0.1. *Assuming the hypotheses from theorem 3.2.0.4 then*

$$\mu(\{x \in E \mid T^n(x) \notin E \text{ for all } n > 0\}) = 0.$$

Note that this corollary is trivially true if we have a set E with $\mu(E) = 0$. Thus the hypotheses for the corollary can be relaxed to all measurable sets $E \in \Sigma$.

Proof. Let $N = \{x \in E \mid T^n(x) \notin E \text{ for all } n > 0\}$ with $\mu(E) > 0$. Clearly $N \subseteq E$ and no point in N will ever return to N under the transformation T, since the point never returns to E. Therefore, by the contrapositive of theorem 3.2.0.4 $\mu(N) = 0$. □

Definition 3.2.0.14. *Given a probability measure space (X, Σ, μ) and a measure preserving transformation $T : X \to X$, then T is said to be* ***ergodic*** *if for any $E \in \Sigma$ with $T^{-1}(E) = E$ either $\mu(E) = 0$ or $\mu(E) = 1$.*

Theorem 3.2.0.5. *[21]* ***The Ergodic Theorem***

Let T be an ergodic transformation on the measure space (X, Σ, μ) and let $f : X \to \mathbb{R}$ be a real-valued measurable function. Then for almost all $x \in X$ the following holds,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} f(T^{-j}(x)) = \frac{1}{\mu(X)} \int f d\mu.$$

If we take the function f to be the characteristic, or indicator, function of a subset $A \subset X$, then this theorem can be thought of as showing that the time average equals the space average for an ergodic transformation.

Therefore, if we knew the results of an infinite number of coin flips, we could accurately determine the probability of seeing heads or tails on the next coin flip. If we are in the setting of a probability space, then it also allows us to obtain very accurate estimates for the measure of a measurable set, A, by setting f equal to the indicator function of A.

4. SYMBOLIC DYNAMICAL SYSTEMS

4.1. Symbolic Dynamical Systems

In section 2.1. of this paper we introduced the idea of a normal number. In symbolic dynamical systems there is a finite alphabet set, a set of one-sided (or two-sided) infinite sequences of letters (called words). The alphabet set can be thought of as the digits with respect to some base b. The set of one-sided infinite sequences with entries from a finite set of size b could be thought of as the b-adic expansion.

One should notice that in the definition of a simply normal number, the limit is equal to the probability of any single digit being rolled if a b-sided fair die is rolled; in other words, using the uniform distribution on b elements. This holds true for blocks of digits, since the probabilities are multiplicative due to the independence of successive digits. But what if we did not use the uniform distribution to specify the probabilities of occurrence for single digits and blocks of digits? How can we define the "normality" of words that appear in a given symbolic dynamical system? It is simple, really. We will use the same definition, but instead of setting it equal to the probabilities given by the uniform distribution, we will use the probabilities given by the dynamical system via the measure of cylinder sets. But first we should define a few basic objects, including a symbolic dynamical system.

Definition 4.1.0.15. *Given a measure preserving probability measure space* (X, Σ, μ, T), *(i.e. T is measure preserving), and a finite partition $\mathcal{P} = \{P_0, P_1, ..., P_{b-1}\}$ of X, we will associate a unique symbol to each element of the partition. Usually we choose the symbols $\{0, 1, 2, ..., b-1\} = \mathcal{A}$. This set of symbols is the* ***alphabet*** *of the* *symbolic dynamical system. Given a point $x \in X$ we keep track of which partition element of $\mathcal{P}$ the T-orbit of x visits by writing down the associated symbol. The* ***$\mathcal{P}$-name*** *of x is an infinite list of letters from our alphabet, namely $\{a_k\}_{k=0}^{\infty}$ such that $T^k(x) \in P_{a_k}$.*

When working with a symbolic dynamical system, we want to keep the $\mathcal{P}$-names of distinct points distinct, at least up to a set of measure zero. The choice of partition, given a measure preserving transformation T, enables us to maintain the distinctness of distinct points in X.

Definition 4.1.0.16. *A partition, $\mathcal{P}$, is called a* ***generator*** *or* ***generating partition*** *if there exists a set $G \subseteq X$ with $\mu(G) = 1$ so that for all $x, y \in G$ with $x \neq y$ then x and y have distinct $\mathcal{P}$-names.*

We have the set X and the set of $\mathcal{P}$-names of points $x \in X$. When T acts on points in X it takes x to $T(x)$. The map that takes the $\mathcal{P}$-name of x to the $\mathcal{P}$-name for $T(x)$ is called the *shift map*, σ. If x has the $\mathcal{P}$-name $x_1x_2x_3...$ then $\sigma(x_1x_2x_3...) = x_2x_3...$ is the $\mathcal{P}$-name of $T(x)$.

The set of symbolic names along with the shift map is the symbolic system. It corresponds with the measure preserving probability measure space. As can be seen, it is cumbersome to talk about the point x and the $\mathcal{P}$-name of x. Therefore, we will think of $x = x_1x_2...$ as both the point in X and the associated $\mathcal{P}$-name of x. So we may also write and think $T(x) = \sigma(x)$ interchangeably.

Definition 4.1.0.17. *For any finite word $a_1a_2...a_k$ with $a_i \in \mathcal{A}$, we define a* ***length k cylinder set*** *to be all the $x \in X$ such that $x_i = a_i$ for all $1 \leq i \leq k$. Where $(x_i)_{i=1}^{\infty}$ is the $\mathcal{P}$-name of x, we will denote a length k cylinder set by $[a_1a_2...a_k]$.*

We will use the idea of a cylinder set when we define what it means for a symbolic dynamical system to be normal. An example of a cylinder set is the set of decimals that start with a 5 in base 10. These points are precisely the points in the interval $[5/10, 6/10)$. This is a cylinder set of length 1 using the appropriate partition for base 10 decimals.

Proposition 4.1.0.2. *Let $(X, \Sigma, \mu) = ([0, 1), \Sigma, \mu)$ with Σ the Lebesgue measurable sets and μ Lebesgue measure. Then the transformation $T : [0, 1) \to [0, 1)$ defined by $T(x) = bx - \lfloor bx \rfloor$ for some $2 \leq b \in \mathbb{N}$ is measure preserving.*

Proof. Given an arbitrary open interval $(c, d) \in [0, 1)$ we want to show that $\mu((c, d)) = \mu(T^{-1}(c, d))$. Since the Lebesgue measure of an open interval is the length of the interval, we have that $\mu((c, d)) = d - c$. Since T could be viewed as multiplication by $b \mod 1$, the only points that get mapped into (c, d) must have come from an interval of the form $(\frac{c}{b} + \frac{i}{b}, \frac{d}{b} + \frac{i}{b})$ with $i \in \{0, 1, 2, ..., b-1\}$. Therefore, $T^{-1}((c, d)) = \bigcup\limits_{i=0}^{b-1} (\frac{c}{b} + \frac{i}{b}, \frac{d}{b} + \frac{i}{b})$. This is a union of disjoint open intervals. Thus

$$\mu\left(\bigcup_{i=0}^{b-1}\left(\frac{c}{b} + \frac{i}{b}, \frac{d}{b} + \frac{i}{b}\right)\right) = \sum_{i=0}^{b-1} \mu\left(\left(\frac{c}{b} + \frac{i}{b}, \frac{d}{b} + \frac{i}{b}\right)\right) = \sum_{i=0}^{b-1}\left(\frac{d}{b} - \frac{c}{b}\right).$$

So that, finally, we have

$$\mu(T^{-1}((c, d))) = b\left(\frac{d}{b} - \frac{c}{b}\right) = d - c = \mu\left((c, d)\right).$$

□

The following example of a dynamical system is going to abstract the real numbers and create a measure preserving transformation that will create the base b expansion of a real number x as the $\mathcal{P}$-name of x in our dynamical system.

Example 4.1.0.3. Let $X = [0, 1)$. Then define

$$\mathcal{P} = \left\{ P_i = \left[\frac{i}{b}, \frac{i+1}{b}\right) \middle| \, i \in \{0, 1, 2, ..., b-1\} \right\}.$$

The partition elements $P_i = \left[\frac{i}{b}, \frac{i+1}{b}\right)$ make the symbolic dynamical system correspond to the base b expansion of $x \in [0, 1)$. Let Σ be the set of Lebesgue measurable sets and λ be Lebesgue measure. Let $T(x) = bx - \lfloor bx \rfloor$ which is the fractional part of bx. Therefore $T : [0, 1) \to [0, 1)$ and, as we have seen in proposition 4.1.0.2, T is measure preserving. So we have a measure preserving probability space (X, Σ, λ, T) and a partition $\mathcal{P}$. So we have a symbolic dynamical system.

Now let's examine what T does to a point $x \in [0, 1)$.

$$T(x) = bx - \lfloor bx \rfloor \Leftrightarrow bx = T(x) + \lfloor bx \rfloor \Leftrightarrow x = \frac{T(x) + \lfloor bx \rfloor}{b}$$

Let $\lfloor bx \rfloor = a_1 \leq b - 1$. So we have $x = (a_1 + T(x))/b$. We can iterate this and find that

$$T^2(x) = T(T(x)) = bT(x) - \lfloor (bT(x)) \rfloor \Leftrightarrow T(x) = \frac{a_2 + T^2(x)}{b}.$$

Where $a_2 = \lfloor bT(x) \rfloor \leq b - 1$. Therefore, we have

$$x = \frac{a_1 + T(x)}{b} = \frac{a_1 + \frac{(a_2 + T^2(x))}{b}}{b} = \frac{a_1}{b} + \frac{a_2}{b^2} + \frac{T^2(x)}{b^2}.$$

So in general we have

$$x = \left(\sum_{k=1}^{n} \frac{a_n}{b^n} \right) + \frac{T^n(x)}{b^n}.$$

Where $a_k = \lfloor bT^{k-1}(x) \rfloor \leq b - 1$. Since $\frac{1}{b^n} T^n(x) \leq \frac{1}{b^n} \to 0$ as $n \to \infty$ uniformly and exponentially fast, we have that $x = \sum_{n=1}^{\infty} \frac{a_n}{b^n}$ with $a_n \in \{0, 1, ..., b-1\}$.

Whenever an expansion of a real number x to a base b is defined, there is always the worry that x will have more than one representation. The author assumes the reader is familiar with this caveat and understands that the set of x that have 2 representations has measure zero. Recall that the number 1 can be written two different ways in base 10, i.e. $1 = .9999....$ Example 4.1.0.3 is the motivation for the following definitions.

Definition 4.1.0.18. *Let $\#_N([a], T, x)$ be the number of times $T^n(x) \in [a]$ for $1 \le n \le N$ and $a \in \mathcal{A}$. Similarly let $\#_N([a_1 a_2 ... a_k], T, x)$ be the number of times $T^n(x) \in [a_1 a_2 ... a_k]$ for $0 \le n \le N-1$ and $a_i \in \mathcal{A}$.*

This definition counts how many times the T-orbit of x visits a particular cylinder set.

Definition 4.1.0.19. *Given a symbolic dynamical system $(X, \Sigma, \mu, T, \mathcal{P})$, we say that the $\mathcal{P}$-name of $x \in X$ is* ***simply normal with respect to*** *$\mathcal{P}$ if*

$$\lim_{N\to\infty} \frac{\#_N([a], T, x)}{N} = \mu([a]) \quad \textit{for all } a \in \mathcal{A}.$$

Definition 4.1.0.20. *Given a symbolic dynamical system $(X, \Sigma, \mu, T, \mathcal{P})$,we say that the $\mathcal{P}$-name of $x \in X$ is* ***normal with respect to*** *$\mathcal{P}$ if*

$$\lim_{N\to\infty} \frac{\#_N([a_1 a_2 ... a_k], T, x)}{N} = \mu([a_1 a_2 ... a_k]) \quad \textit{for all } [a_1 a_2 ... a_k].$$

Using example 4.1.0.3 and definition 4.1.0.20 we can provide a basic proof of the Normal Number Theorem similar to something Borel could have come up with.

4.2. A Proof of the Normal Number Theorem

Recall the symbolic dynamical system from Example 4.1.0.3, $(X, \Sigma, \lambda, T, \mathcal{P})$. We showed that this system corresponds with the base b expansion of $x \in X = [0, 1)$.

We also have that $\lambda([a]) = b^{-1}$ for all $a \in \{0, 1, ..., b-1\}$. This is easily seen by the definition of the partition $\mathcal{P}$. One can easily show that $\lambda([a_1 a_2 ... a_k]) = b^{-k}$ for $a_i \in \{0, 1, ..., b-1\}$ and $1 \leq i \leq k$.

We can now state the Normal Number Theorem in the context of symbolic dynamical systems using example 4.1.0.3 as our system.

Theorem 4.2.0.6. *(**Normal Number Theorem**) Given the symbolic dynamical system from Example 4.1.0.3, the set of* $x \in [0, 1)$ *such that*

$$\lim_{N \to \infty} \frac{\#_N(C, T, x)}{N} = \lambda(C)$$

for all cylinder sets C has Lebesgue measure 1.

Proof. Let $C = [a_1 a_2 ... a_k]$ be a cylinder set for any $1 \leq k < \infty$ and any $a_i \in \{0, 1, 2, ..., b-1\}$, $1 \leq i \leq k$. Let $x \in [0, 1)$ have $\mathcal{P}$-name $x_1 x_2 ....$ with $x_i \in \{0, 1, ... b-1\}$. Recall that $\#_N(C, T, x)$ is the number of times $T^n(x)$ is in C for $0 \leq n \leq N-1$. Then for $\delta > 0$ define

$$D_N(C, T, \delta) = \left\{ x \in [0, 1) \,\middle|\, \left| \frac{\#_N(C, T, x)}{N} - \lambda(C) \right| \geq \delta \right\}.$$

For x to be normal it cannot belong to infinitely many $D_N(C, T, \delta)$. Note that A^c denotes the set complement of A. So the elements of the set

$$\bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} \left(D_N \left(C, T, \frac{1}{j} \right) \right)^c$$

are the x that have $|\#_N(C, T, x) - \lambda(C)| < \frac{1}{j}$ for all $N \geq N_0$. Therefore, if we take the intersection over all of the j we would have the set of $x \in [0, 1)$ for which $\lim_{N \to \infty} \frac{\#_N(C,T,x)}{N} = \lambda(C) = b^{-k}$. This set is

$$\bigcap_{j=1}^{\infty} \bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} \left(D_N \left(C, T, \frac{1}{j} \right) \right)^c.$$

The number of cylinder sets is countable so we can again take the intersection over all possible cylinder sets C of any length $1 \leq k < \infty$. Let the collection of all possible cylinder sets be denoted by $\mathcal{C}$. Then we have

$$\bigcap_{C \in \mathcal{C}} \bigcap_{j=1}^{\infty} \bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} \left(D_N \left(C, T, \frac{1}{j} \right) \right)^c$$

is the set of x that are normal with respect to base b. Notice that this set contains exactly all of the numbers normal to the base b with none missing. We can also take the intersection over all bases $b \geq 2$ and get the set of x that are absolutely normal

$$\bigcap_{b=2}^{\infty} \bigcap_{C \in \mathcal{C}} \bigcap_{j=1}^{\infty} \bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} \left(D_N \left(C, T, \frac{1}{j} \right) \right)^c . \tag{4.1}$$

To complete the proof we simply need to show that $\lambda \left(D_N \left(C, T, \frac{1}{j} \right) \right) \to 0$. We have that λ is a probability measure, so we can use results from probability theory. If we define the random variables $X_i : [0, 1) \to \{0, 1, ..., b-1\}$ where $X_i(x) = X_i(x_1 x_2 ...) = x_i$ with $x_1 x_2 ...$ being the $\mathcal{P}$-name of x. So the random variable X_i returns the i^{th} symbol in the $\mathcal{P}$-name of x. These random variables are independent and identically distributed with the uniform distribution on $\{0, 1, ..., b-1\}$. Therefore, $x \in C = [a_1 a_2 ... a_k]$ if and only if $X_1(x) = a_1$, $X_2(x) = a_2$, ..., $X_k(x) = a_k$. For notational convenience we will use $\mathbb{1}_C$ since it is equivalent to $X_1(x) = a_1$, $X_2(x) = a_2$, ..., $X_k(x) = a_k$.

Recall that $\lambda(C) = \int_0^1 \mathbb{1}_C \, d\lambda = \mathbb{E}(C)$. Noticing this will allow us to use probability when examining whether or not $\lambda \left(D_N \left(C, T, \frac{1}{j} \right) \right) \to 0$. Notice that $\#_N(C, T, x) = \sum_{n=0}^{N-1} \mathbb{1}_C(T^n(x))$ is the number of times that $T^n(x)$ is in C for $0 \leq n \leq N-1$. Let $S_N = \sum_{n=0}^{N-1} \mathbb{1}_C(T^n(x))$. Therefore, the set $D_N \left(C, T, \frac{1}{j} \right)$ is the set of $x \in [0, 1)$ such that $\left| \frac{S_N(x)}{N} - \lambda(C) \right| \geq \frac{1}{j}$. Therefore, by lemma 3.1.0.2, we have

$$\lambda \left(D_N \left(C, T, \frac{1}{j} \right) \right) \leq \frac{j^4 \, \mathbb{E}(S_N^4(x))}{N^4}.$$

Then lemma 3.1.0.3 gives us

$$\lim_{N\to\infty} \lambda \left(D_N \left(C, T, \frac{1}{j} \right) \right) = 0 \text{ for all } j > 0.$$

□

4.3. Entropy of a Partition

Entropy is a unitless quantity that tries to describe the amount of complexity or randomness of a certain system. The thermodynamic version of entropy is measuring the amount of energy in a system that cannot be used to do thermodynamic work. As time passes, the thermodynamic system tends toward an equilibrium state where the heat or energy is uniformly distributed. As we will see in our definitions, the equilibrium state is when the entropy is maximized. For us, it will be because we have minimized the certainty of seeing a particular outcome when the probabilities of outcomes are all equal (i.e., a uniform distribution of probabilities over the state space).

We will want to study the entropy of a symbolic dynamical system for a few reasons. The main reason is that entropy has been proven to be an isomorphism invariant for dynamical systems by Ornstein [13]. Another reason we want to study entropy is that it tells us if the system we are examining is deterministic or not. Determinism can be thought of as being able to predict the future given the past. A non-deterministic system has some amount of randomness associated with predicting the future knowing the past.

Definition 4.3.0.21. *Given a probability space,* (X, Σ, μ) *and a partition of* X, $\mathcal{P} = \{P_1, P_2, ..., P_n\}$, *with* $\mu(P_i) = p_i \geq 0$ *for* $1 \leq i \leq n$ *and* $\sum_{i=1}^{n} p_i = 1$. *We define* *the **entropy of the partition** $\mathcal{P}$ to be*

$$h(\mathcal{P}) = -\sum_{i=1}^{n} p_i \log(p_i).$$

We define $0\log(0) = 0$ if that case should arise.

We will always use the logarithm base 2. It is fairly arbitrary which base you pick for the logarithm. We chose base 2 so we can think of the entropy in terms of bits of information. If we think of $-\log(p_i)$ as the amount of "surprise" we experience when a point $x \in X$ is in P_i, then $h(\mathcal{P})$ is the expected amount of surprise the partition provides and can be viewed as the expectation of $-log(p_i)\mathbb{1}_{(P_i)}$.

Example 4.3.0.4. Let $\mu(P_i) = \frac{1}{n}$ for all $P_i \in \mathcal{P}$ with $|\mathcal{P}| = n$. Then

$$h(\mathcal{P}) = -\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) = \log(n).$$

For this example the logarithm is increasing in n, so as the size of the partition increases, the entropy increases. In addition, the maximum entropy for a partition of size n is $\log(n)$. The heuristic reason for this is that when n outcomes are equally likely then the expected amount of surprise is maximized.

Definition 4.3.0.22. *Given two finite partitions $\mathcal{P}$ and $\mathcal{Q}$ of a space X, we define the **span of $\mathcal{P}$ and $\mathcal{Q}$**, denoted $\mathcal{P} \vee \mathcal{Q}$, to be the set of all non-empty intersections of elements from $\mathcal{P}$ and $\mathcal{Q}$,*

$$\mathcal{P} \vee \mathcal{Q} = \{P \cap Q \mid P \in \mathcal{P} \ \&\ Q \in \mathcal{Q}\}.$$

Also given a family of finite partitions of X, $(\mathcal{P}_i)_{i=1}^{n}$, we define

$$\bigvee_{i=1}^{n} \mathcal{P}_i = \mathcal{P}_1 \vee \mathcal{P}_2 \vee ... \vee \mathcal{P}_n.$$

The entropy of a partition is only part of the story for a symbolic dynamical system. Recall that we want the symbolic system to keep points distinct. A generating partition was a partition that did exactly that, relative to what the transformation T is. For very similar reasons, if we want to talk about the entropy of a transformation T, we need to take into consideration the different partitions that are possible so that the entropy of T will not depend on the choice of partition.

Definition 4.3.0.23. *The* ***entropy of a finite partition*** $\mathcal{P}$ ***relative to*** T *is*

$$h(\mathcal{P}, T) = \limsup_{n\to\infty} \frac{1}{n} h\left(\bigvee_{k=0}^{n-1} T^{-k}(\mathcal{P})\right).$$

Now we can finally define the entropy of T.

Definition 4.3.0.24. *The* ***entropy of*** T *is*

$$h(T) = \sup_{\mathcal{P}} h(\mathcal{P}, T)$$

where the supremum is taken over all possible finite partitions of X*.*

Definition 4.3.0.25. *The* ***conditional entropy*** *of a partition* $\mathcal{P}$ *given a partition* $\mathcal{Q}$ *is*

$$h(\mathcal{P} \mid \mathcal{Q}) = h(\mathcal{P} \vee \mathcal{Q}) - h(\mathcal{Q}).$$

The conditional entropy can be thought of as the amount of expected surprise in seeing the next digit in a sequence if we have prior knowledge via the partition $\mathcal{Q}$. This heuristic is very helpful once we have the following proposition and theorem. It will help us better understand how the entropy relates to determinism.

Proposition 4.3.0.3. *Given a symbolic dynamical system* $(X, \Sigma, \mu, T, \mathcal{P})$

$$h(\mathcal{P}, T) = \limsup_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} h\left(\mathcal{P} \,\middle|\, \bigvee_{n=1}^{k} T^{-n}(\mathcal{P})\right). \tag{4.2}$$

Proof. From the definition of conditional entropy we have

$$h\left(\mathcal{P}\,\middle|\,\bigvee_{n=1}^{k} T^{-n}(\mathcal{P})\right) = h\left(\bigvee_{n=0}^{k} T^{-n}(\mathcal{P})\right) - h\left(\bigvee_{n=1}^{k} T^{-n}(\mathcal{P})\right).$$

Therefore the sum in equation 4.2 is a telescoping sum. Thus

$$\frac{1}{N}\sum_{k=1}^{N} h\left(\mathcal{P}\,\middle|\,\bigvee_{n=1}^{k} T^{-n}(\mathcal{P})\right) = \frac{1}{N} h\left(\bigvee_{n=0}^{N} T^{-n}(\mathcal{P})\right) - \frac{1}{N} h\left(\mathcal{P}\vee T^{-1}(\mathcal{P})\right),$$

since $h\left(\mathcal{P}\vee T^{-1}(\mathcal{P})\right)$ is some finite constant $\limsup_{N\to\infty} \frac{1}{N} h\left(\mathcal{P}\vee T^{-1}(\mathcal{P})\right) = 0$. Therefore,

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} h\left(\mathcal{P}\,\middle|\,\bigvee_{n=1}^{k} T^{-n}(\mathcal{P})\right) = \limsup_{N\to\infty} \frac{1}{N} h\left(\bigvee_{n=0}^{N} T^{-n}(\mathcal{P})\right) = h(\mathcal{P}, T).$$

$\square$

Theorem 4.3.0.7. *[21] (**Kolmogorov-Sinai**)*

Given a symbolic dynamical system $(X, \Sigma, \mu, T, \mathcal{P})$, if the partition $\mathcal{P}$ is a generator, then

$$h(T) = h(\mathcal{P}, T).$$

Thinking of $\bigvee_{k=1}^{\infty} T^{-k}(\mathcal{P})$ as the infinite past, then proposition 4.3.0.3, combined with the Kolmogorov-Sinai theorem, shows us that knowing the infinite past for a generating partition gives us the entropy. In other words, if the entropy is zero, knowing the past will tell us the future. This is determinism. If the entropy of a system is zero we will say that it is a *deterministic* system. In contrast, if a system has positive entropy, then it is not deterministic and must have some type of random structure.

4.4. Shift Spaces

Shift spaces are another way to describe symbolic dynamical systems. One could think of the shift space as the set of symbolic names with a shift operator. Just as before we have a finite set of states. Given some size for the alphabet, n, we will always use the finite set $\mathcal{A} = \{0, 1, ..., n-1\}$ for convenience. But using this set is not necessary; any finite set will work. An *infinite word* is a sequence $x = (x_k)_{k \in \mathbb{N}}$ where $x_k \in \mathcal{A}$. A *bi-infinite word* is a doubly infinite sequence $x = (x_k)_{k \in \mathbb{Z}}$ where $x_k \in \mathcal{A}$. We will talk about the bi-infinite shift spaces, but the ideas can easily be applied to the infinite shift spaces.

Definition 4.4.0.26. *Given a finite alphabet $\mathcal{A}$, the **full $\mathcal{A}$-shift** is the collection of all bi-infinite sequences of symbols from $\mathcal{A}$. If we are using the alphabet $\mathcal{A} = \{0, 1, ..., n-1\}$, we call this the **full n-shift**.*

We denote the full $\mathcal{A}$-shift by

$$\mathcal{A}^{\mathbb{Z}} = \{x = (x_k)_{k \in \mathbb{Z}} \mid x_k \in \mathcal{A} \text{ for all } k \in \mathbb{Z}\}.$$

If $\mathcal{A}$ has size $|\mathcal{A}| = n$ then there is a natural correspondence between the full n-shift and the full $\mathcal{A}$-shift. We will always think in terms of the full n-shift.

Blocks of consecutive symbols are central to the ideas we explore in this paper. A *block* or *word* over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$. We write the block without any spaces in between the symbols; for example, a block over the alphabet $\mathcal{A} = \{0, 1\}$ would look like 0101100111. This block would have a length of 10.

Much in the same way that the real number zero is important, we also need to include the *empty block* or *empty word*, which is the word containing no symbols, denoted by ϵ. The *length* of a block B is the number of symbols it contains and is denoted $|B|$. For example, if $B = b_1b_2...b_k$ is a non-empty block, then $|B| = k$ and we have that $|\epsilon| = 0$.

Definition 4.4.0.27. *The **shift map** σ on the full shift $\mathcal{A}^{\mathbb{Z}}$ maps a point x to a point y such that $y_i = x_{i+1}$.*

So the shift map shifts the symbols in the bi-infinite word to the left by one. This can be viewed as follows:

$$x = \dots x_{-2}x_{-1} \,.\, x_0x_1x_2 \dots$$

and

$$\sigma(x) = \dots x_{-1}x_0 \,.\, x_1x_2x_3 \dots$$

Notice the period that symbolizes an origin of sorts, without which we would be lost in a sea of symbols with no compass or sextant. In a very similar manner σ^{-1} shifts all of the symbols of a bi-infinite word to the right by one.

Definition 4.4.0.28. *A point $x \in \mathcal{A}^{\mathbb{Z}}$ is **periodic** if there exists some $n \in \mathbb{N}$ such that $\sigma^n(x) = x$. The least such n is called the least period of x. If $n = 1$ for some x we call x a **fixed point** of σ.*

The iteration of the shift map creates the dynamics of symbolic dynamical systems. Iterating the shift map k times will shift bi-infinite words to the left by k places. Similarly iterating the inverse shift map, σ^{-1}, k times will shift the bi-infinite word to the right by k places. The symbolic part of symbolic dynamics refers to the set of symbols used to form the sequences.

In some ways the full n-shift is boring because it contains no structure or constraints on possible words. For instance, if we think of the English language, our alphabet set contains the basic 26 letters, the 26 capital letters, punctuation marks, spaces, and perhaps some special characters. But we have a dictionary that tells us the finite words that we can make with these symbols i.e., "mathematics" is a valid "word" in the English language, but "Scramblepants" is not. So we see that being able to describe certain subsets of the full shift could prove useful for applying shift spaces to other more restrictive systems.

Instead of having a list of words that are allowed, it is easier to think of a list of words that are forbidden. If $x \in \mathcal{A}^{\mathbb{Z}}$ and B is a word over $\mathcal{A}$ with $|B| = k$, then the word B occurs in x if there exists an i such that $B = x_i x_{i+1} \ldots x_{i+k-1}$. We will call B a *factor* of x. Note that the empty word occurs as a factor in every x. Let $\mathcal{F}$ be a collection of finite words over $\mathcal{A}$ which we will think of as *forbidden words*. For any such $\mathcal{F}$, define $X_{\mathcal{F}}$ to be the subset of sequences from $\mathcal{A}^{\mathbb{Z}}$, such that the sequences in $X_{\mathcal{F}}$ do not contain any word in $\mathcal{F}$ as a factor.

Definition 4.4.0.29. *A **shift space** (or shift) is a subset X of a full $\mathcal{A}^{\mathbb{Z}}$ (or n-shift) such that $X = X_{\mathcal{F}}$ for some set $\mathcal{F}$ of forbidden words over $\mathcal{A}$.*

The collection of forbidden words $\mathcal{F}$ can be infinite or finite. If it is infinite it is at most countable, since you can list the elements in order of their lengths. When a shift space X is contained in a shift space Y we say that X is a *subshift* of Y. All shift spaces are subshifts of the full shift. So we may refer to shift spaces as subshifts. If the size of the forbidden word set is finite then we call that shift space a *subshift of finite type*. [11]

Example 4.4.0.5. Let X be the set of all binary sequences with no two 1's next to each other. We have $\mathcal{F} = \{11\}$ and $X = X_{\mathcal{F}}$. This shift is called the *golden mean shift*. This is a subshift of finite type.

A different but equivalent definition of a shift space involves the product topology. Sometimes it is convenient to look at the most basic structure underlying a space. Topology is good for doing such examinations. This makes this alternate definition useful for obtaining results that might have otherwise been obfuscated by higher order structure.

Definition 4.4.0.30. *A subset X of $\mathcal{A}^{\mathbb{Z}}$ is a **shift space** if it is closed with respect to the natural product topology of $\mathcal{A}^{\mathbb{Z}}$ and invariant under the shift map σ.*

Just as we defined the forbidden word set it is useful to define the set of words that can occur in a subshift. This is analogous to the dictionary for the English language.

Definition 4.4.0.31. *Given a subshift X of $\mathcal{A}^{\mathbb{Z}}$, let $\mathcal{L}_n(X)$ be the set of all n-words that occur as factors in points in X. Then the **language of** X is*

$$\mathcal{L}(X) = \bigcup_{n=0}^{\infty} \mathcal{L}_n(X).$$

The sizes of the sets $\mathcal{L}_n(X)$, $|\mathcal{L}_n(X)|$, can classify certain properties that X will have. For instance, given a binary system such that for all $x \in X$, x contains $n+1$ factors of length n and $\mathcal{L}_n(X) = n+1$. Then all $x \in X$ are balanced, meaning all factors of the same length have the property that the number of occurrences of a letter differ by at most 1. One way to study the properties of shift spaces is to look at the points and see what properties they possess.

Definition 4.4.0.32. *Let X be a subshift of $\mathcal{A}^{\mathbb{Z}}$. Then for $n \geq 1$ and $x \in X$ the **complexity function** $p_x(n)$ is the number of distinct words of length n in the sequence x.*

There is nothing in the definition of the complexity function requiring the factor to occur more than once. It only needs to occur once as a factor in x. Further definitions are needed to establish a factor that occurs infinitely many times and determine how long of a wait it is until the factor occurs as a factor of x again.

We clearly have $p_x(n) \leq |\mathcal{L}_n(X)|$ since $\mathcal{L}_n(X)$ is the number of length n words allowed in any $x \in X$. The complexity function can be defined for any sequence made up of symbols from a finite set. We will discuss complexity more in section 5.2.

4.5. Entropy of Shift Spaces

Entropy can be viewed as a measure of order or disorder. For shifts it is a measure of the variety of blocks that are possible. Given a shift space X, the number, $|\mathcal{L}_n(X)|$, of n-words that can appear as factors of points in X is an indicator of how complex the space X is. Instead of looking at the numbers $|\mathcal{L}_n(X)|$ themselves we can look at their growth rate to summarize their behavior.

Definition 4.5.0.33. *Given a shift space X, the **entropy of** X is defined by*

$$h(X) = \lim_{k\to\infty} \frac{1}{k} \log(|\mathcal{L}_k(X)|).$$

The logarithm we will always use is base 2. If we let X be the full n-shift we see that $|\mathcal{L}_k(X)| = n^k$. Thus $h(X) = \log(n)$. Then for all subshifts Y of X we have that $|\mathcal{L}_k(X)| \geq |\mathcal{L}_k(Y)|$ for all k, therefore $\log(n)$ is an upper bound for the entropy of a shift on a set of n symbols. The entropy of a shift is always non-negative since for non-trivial shifts $|\mathcal{L}_k(X)| > 1$ for all k.

If the growth of $|\mathcal{L}_k(X)|$ is sub-exponential then the entropy will be zero. In order to have positive entropy $|\mathcal{L}_k(X)|$ needs to grow exponentially. However, the rate of growth could be very small, which would indicate a small value for the entropy.

There is a connection between the entropy of a shift space and the entropy of the shift map on the shift space. We refer the reader to [11] (chapter 6) for a discussion of this.

5. INFINITE WORDS

5.1. The Fibonacci System

The Fibonacci system can be built in a variety of ways. One such construction is generated by the rotation $T : [0,1) \to [0,1)$ defined by $T(x) = x + \phi \pmod 1$ with $\phi = \frac{-1+\sqrt{5}}{2}$ and the partition of $[0,1)$ that consists of $P_1 = [0, 1-\phi)$ and $P_0 = [1-\phi, 1)$. If we label the larger interval 0 and the smaller interval 1 we can use T to generate binary sequences by looking at the orbit of any $x \in [0,1)$. This process actually generates an uncountable number of distinct binary sequences. This is due to the fact that T is non-periodic since ϕ is irrational. We call the $\mathcal{P}$-name of 0 the *infinite Fibonacci word*. This word is a fixed point of the following substitution system. For a more detailed look at substitution systems including precise definitions, see [4] or [14].

Another construction of the infinite Fibonacci word is given by the following substitution system. Using the substitution $\varphi = \begin{cases} 0 \to 01 \\ 1 \to 0 \end{cases}$, then we call $F = \varphi^{\omega}(0) = 0100101001001010010100...$ the Fibonacci word. It is the fixed point of the substitution φ.

If we define $F_1 = 1$ then the following sequence is built with φ.

$$\varphi^0(1) = F_1 = 1$$
$$\varphi^1(1) = F_2 = 0$$
$$\varphi^2(1) = F_3 = 01$$
$$\varphi^3(1) = F_4 = 010$$
$$\varphi^4(1) = F_5 = 01001$$
$$\varphi^5(1) = F_6 = 01001010$$
$$\varphi^6(1) = F_7 = 0100101001001$$
$$\varphi^7(1) = F_8 = 010010100100101001010$$

Theorem 5.1.0.8. *If we define $B_1 = 1$, $B_2 = 0$, and $B_n = B_{n-1}B_{n-2}$ for $n \geq 3$ then $F_{n+1} = \varphi^n(1) = B_{n+1}$ and $F_{n+2} = \varphi^n(0) = B_{n+2}$ for $n \geq 0$. Therefore, $F = \lim_{n\to\infty} B_n$ and for $n \geq 1$ we have that the length of B_n, $|B_n| = f_n$, the n^{th} Fibonacci number.*

We are showing that the infinite Fibonacci word can be constructed by continually concatenating successive finite Fibonacci blocks.

Proof. We will prove the theorem by induction on n. It is true that $B_1 = F_1$ and $B_2 = F_2$. Thus the theorem holds for $n = 0, 1$. Assume that $\varphi^k(1) = B_{k+1}$ and $\varphi^k(0) = B_{k+2}$ for all $k < n$. Consider $\varphi^n(1) = \varphi^{n-1}(\varphi(1)) = \varphi^{n-1}(0) = B_{n+1}$. Also consider $\varphi^n(0) = \varphi^{n-1}(\varphi(0)) = \varphi^{n-1}(01) = \varphi^{n-1}(0)\varphi^{n-1}(1) = B_{n+1}B_n = B_{n+1}$. Hence $F = \lim_{n\to\infty} F_n = \lim_{n\to\infty} B_n$. As for the claim about the lengths of the B_n being the Fibonacci numbers, we can see that by observing that $|B_1| = |B_2| = 1$ and $|B_n| = |B_{n-1}| + |B_{n-2}|$, which is the recurrence formula that generates the Fibonacci numbers. $\square$

It should be noted that the theorem above proved that $F_n = B_n$. This will be useful throughout the rest of the paper and F_n will be used in place of B_n for the rest of the paper. The ability to think of the Fibonacci system in terms of the substitution and symbolic dynamical system is very useful. It allows alternate perspectives to give different intuitions and reasons as to why certain properties of the system are the way they are.

Theorem 5.1.0.9. *For any $w \in \{0,1\}^{\mathbb{N}}$ let the map $s(w)$ interchange the last two letters of w, given that $\infty > |w| \geq 2$. Then for all $n \geq 1$ $F_n F_{n+1} = s(F_{n+1}F_n)$.*

It is worth noting that $s^2(w) = w$ for all w. Therefore $s(F_n F_{n+1}) = F_{n+1}F_n$ is an alternate statement of the result of the theorem. Also $s(wu) = ws(u)$ if $|u| \geq 2$. Therefore $s(F_{n-1}F_n) = F_{n-1}s(F_n)$ for $n \geq 3$.

Proof. We will use a proof by induction on n. For $n = 1$ we have that $F_1F_2 = 10 = s(01) = s(F_2F_1)$. for $n = 2$ we have that $F_2F_3 = 001 = s(010) = s(F_3F_2)$. Assume the theorem is true for all $k < n$ with $n \geq 3$. Then

$$F_nF_{n+1} = F_nF_nF_{n-1} = F_ns(F_{n-1}F_n) = F_nF_{n-1}s(F_n) = F_{n+1}s(F_n) = s(F_{n+1}F_n).$$

□

Lemma 5.1.0.4. *The last three digits of F_n are 001 if n is odd and 010 if n is even.*

This lemma is proven easily by induction and Theorem 5.1.0.8.

Lemma 5.1.0.5. *If the last two letters of F_n are removed the resulting finite word is a palindrome, i.e. it is the same if the order of the digits is reversed.*

Proof. We will prove this lemma by induction on n. Let $n \geq 3$. Then for $n = 3$ we have that $F_n = 010$. Clearly if the last two letters are removed, the resulting word is a palindrome of length 1. Now assume that for all $k \leq n$ the lemma holds true. Consider $F_{n+1} = F_n F_{n-1} = wab$ for a finite word w and letters $a, b \in \{0, 1\}$. We know from theorem 5.1.0.9 that $F_{n-1}F_n = wba$. Let $F_n = uba$ and $F_{n-1} = vab$. The inductive hypothesis tells us that u and v are palindromes. Let $r(\star)$ be the map that reverses the order of the digits of a finite word. For example, $r(0100) = 0010$. Then we have that $r(u) = u$ and $r(v) = v$. We want to show that $r(w) = w$. We have that $wab = F_n F_{n-1} = ubavab$, so $w = ubav$. But we also know that $wba = F_{n-1}F_n = vabuba$, so $w = vabu$. Therefore, $r(w) = r(ubav) = r(v)abr(u) = vabu = w$. Thus the lemma is proven by induction. □

In a paper by Mignosi[23] they show that the forbidden words for the Fibonacci system can be easily constructed from the palindromes that lemma 5.1.0.5 discusses. This is done by adding a 1 or a 0 at the beginning, and end, of the palindromes. Given a Fibonacci block F_k, we remove the last two digits to obtain a palindrome, w_k. If k is even, then $0w_k0$ is a forbidden word. If k is odd, then $1w_k1$ is a forbidden word. This list of forbidden words is another way to characterize the Fibonacci system. Knowing this list explicitly helps when working with the Fibonacci system. Here are the first few forbidden words:

$$\{11, 000, 10101, 00100100, ...\}.$$

5.2. Sturmian Words

We will start by first defining what a *characteristic* word is and then define what a *Sturmian* word is. This order has been chosen to start with the simplest idea and then abstract to a more general class of infinite words.

Definition 5.2.0.34. *A* ***characteristic word****, w, has digits defined in the following way. Let $0 < \alpha < 1$ be a real number. For $n \geq 1$, define a function $f_\alpha(n) := \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor$. Then $w = w_\alpha = f_\alpha(1)f_\alpha(2)f_\alpha(3)\ldots$. We sometimes call w the characteristic word with slope α.*

The characteristic word is an infinite word over the alphabet $\{0, 1\}$ since $0 < \alpha < 1$. We could alternately define

$$f_\alpha(n) = \begin{cases} 1 & \text{if } \{n\alpha\} \in [1-\alpha, 1), \\ 0 & \text{otherwise}, \end{cases}$$

where $\{n\alpha\}$ denotes the fractional part of α or $\{n\alpha\} = n\alpha \pmod 1$. To see that these two definitions are equivalent, note that

$$\lfloor (n+1)\alpha \rfloor = \lfloor n\alpha + \alpha \rfloor = \lfloor \lfloor n\alpha \rfloor + \{n\alpha\} + \alpha \rfloor = \lfloor n\alpha \rfloor + \lfloor \{n\alpha\} + \alpha \rfloor.$$

Thus $f_\alpha(n) = \lfloor \{n\alpha\} + \alpha \rfloor$. Therefore, $f_\alpha(n) = 1$ if and only if $\{n\alpha\} \in [1-\alpha, 1)$. This shows that the two definitions are equivalent. Using the two definitions, we see that the number of 1's in the first n digits of a characteristic word, w_α, is $\sum_{k=1}^{n} f_\alpha(k) = \lfloor (n+1)\alpha \rfloor$. Notice that this is a telescoping sum. It is then easy to see what the frequency of the digit 1 has in w_α. It is

$$\lim_{n\to\infty} \frac{\lfloor (n+1)\alpha \rfloor}{n} = \lim_{n\to\infty} \frac{(n+1)\alpha + \{(n+1)\alpha\}}{n} = \alpha.$$

Since characteristic words are composed of only two distinct digits, we know that the frequency of 0's is $1 - \alpha$. We now have a way to compute the frequency of single digits for all characteristic words.

Theorem 5.2.0.10. *The frequency of 1's in a characteristic word with slope $0 < \alpha < 1$ is α.*

Proof. The discussion above led us to this result. □

Corollary 5.2.0.2. *The frequency of* 0*'s in a characteristic word with slope* $0 < \alpha < 1$ *is* $1 - \alpha$.

Proof. This corollary is an immediate result of theorem 5.2.0.10 since Sturmian words are entirely composed of zeroes and ones. □

The generalization of the characteristic words are called *Sturmian* words.

Definition 5.2.0.35. *An infinite word,* $w_{\alpha,\rho}$*, is said to be* ***Sturmian*** *if it is of the form* $w = g_{\alpha,\rho}(1)g_{\alpha,\rho}(2)g_{\alpha,\rho}(3)...$ *with* $0 < \alpha < 1$ *and*

$$g_{\alpha,\rho}(n) := \lfloor (n+1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor, \ \ \textit{for } n \geq 1.$$

We call $w_{\alpha,\rho}$ *the Sturmian word with slope* α.

It is worth noting that Sturmian words are infinite words over the alphabet $\{0, 1\}$. We see that the essential difference from a characteristic word is that it has a "y-intercept" that is not zero. We will see later that a word, w, is *Sturmian* if and only if the complexity function of w is $n+1$. We can again easily see the number of 1's in the first n digits of a Sturmian word, $w_{\alpha,\rho}$, is $\displaystyle\sum_{k=1}^{n} g_{\alpha,\rho}(n) = \lfloor (n+1)\alpha + \rho \rfloor - \lfloor \alpha + \rho \rfloor$. Therefore, we can see that the frequency of the digit 1 in $w_{\alpha,\rho}$ is

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} g_{\alpha,\rho}(n)}{n} = \lim_{n\to\infty} \frac{\lfloor (n+1)\alpha + \rho \rfloor - \lfloor \alpha + \rho \rfloor}{n} \tag{5.1}$$

$$= \lim_{n\to\infty} \frac{(n+1)\alpha + \rho - \{(n+1)\alpha + \rho\} - (\alpha + \rho - \{\alpha + \rho\})}{n} \tag{5.2}$$

$$= \lim_{n\to\infty} \frac{n\alpha - \{(n+1)\alpha + \rho\} - \{\alpha + \rho\}}{n} \tag{5.3}$$

$$= \alpha. \tag{5.4}$$

In the next few lemmas we will prove a result that relates the characteristic words w_α and $w_{1-\alpha}$. What we find is that these two characteristic words are essentially the same up to a coding.

Definition 5.2.0.36. *Let C be defined by $C(0) = 1$ and $C(1) = 0$. We call C the **conjugate operation** on infinite words over the alphabet consisting of two letters, $\{0, 1\}$.*

Lemma 5.2.0.6. *Without loss of generality let $x \geq 0$.*

$$\lfloor x \rfloor + \lfloor -x \rfloor = \begin{cases} 0 & \text{if } x \in \mathbb{Z}, \\ -1 & \text{if } x \notin \mathbb{Z}. \end{cases}$$

Proof.

Case 1) Let $x \in \mathbb{Z}$. Then $\lfloor x \rfloor = x$ and $\lfloor -x \rfloor = -x$. So $\lfloor x \rfloor + \lfloor -x \rfloor = 0$.

Case 2) Let $x \notin \mathbb{Z}$. Thus $\{x\} \neq 0$. Then $\lfloor x \rfloor = x - \{x\}$ and $\lfloor -x \rfloor = -(x+1) - \{-x\} = -(x+1) - \{x\}$. Therefore $\lfloor x \rfloor + \lfloor -x \rfloor = -1$.

Hence the Lemma holds. $\square$

Lemma 5.2.0.7. *Given $0 < \alpha < 1$such that α is an irrational number, then $w_\alpha = C(w_{1-\alpha})$.*

Proof. By definition we have that

$$f_\alpha(n) = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor$$

and

$$f_{1-\alpha}(n) = \lfloor (n+1)(1-\alpha) \rfloor - \lfloor n(1-\alpha) \rfloor = \lfloor -(n+1)\alpha \rfloor - \lfloor -n\alpha \rfloor + 1.$$

Therefore,

$$f_\alpha(n) + f_{1-\alpha}(n) = \lfloor (n+1)\alpha \rfloor + \lfloor -(n+1)\alpha \rfloor - \lfloor -n\alpha \rfloor - \lfloor n\alpha \rfloor + 1.$$

With the result from the preceding lemma we have that

$$f_\alpha(n) + f_{1-\alpha}(n) = 1, \quad \text{for all } n.$$

Since these infinite words are composed of zeroes and ones exclusively, we have that the digits of $f_\alpha(n)$ are exactly the opposite of the digits of $f_{1-\alpha}(n)$, with 0 being the opposite of 1 and vice versa. □

5.2.1 Complexity

We have already given a definition of the complexity of a point in a shift space, $x \in \mathcal{A}^{\mathbb{Z}}$. For clarity we will define analogously the complexity for a point $x \in \mathcal{A}^{\mathbb{N}}$.

Definition 5.2.1.1. *Let $\mathcal{A}$ be a finite alphabet of size b. Let $x \in \mathcal{A}^{\mathbb{N}}$. Then we define $p_n(x)$ to be the number of distinct factors of length n in x.*

It is clear that $1 \leq p_n(x) \leq b^n$. For an example of an x such that $p_n(x) = 1$, consider x to be an infinite sequence of one symbol. An example of an x that has $p_n(x) = b^n$ is Champernowne's number in base b. Constructing sequences that have a certain complexity is not an easy task. For a general sequence x it may be impossible with current techniques to determine the complexity function of x. Sequences that are of great interest are those that have complexity function growing less than any exponential but greater than any polynomial. A paper by Cassaigne [22] constructs infinite words of what they call intermediate complexity, that is a complexity function that grows faster than any polynomial, but slower than any exponential.

Theorem 5.2.1.1. *[4] Given a Sturmian word w, then $p_n(w) = n + 1$.*

This theorem tells us that the Fibonacci word has complexity $n + 1$ for all n. If we recall the definition of Sturmian words it is clear that all the points in the

Fibonacci system are Sturmian, since $x+\phi$ can be seen as the Sturmian word $w_{\phi,x}$. Furthermore, in [4] it is shown that the factors of a Sturmian word $w_{\alpha,\rho}$ are exactly the same as the factors of the characteristic word w_α. Therefore, all of the points in the Fibonacci system have complexity $n+1$ and they all have the exact same factors. Therefore, the language of the Fibonacci system has $n+1$ words of length n for all $n \geq 0$. Thus the entropy of of the Fibonacci system in terms of definition 4.5.0.33 is

$$\lim_{n\to\infty} \frac{1}{n} \log(n+1) = 0.$$

Theorem 5.2.1.2. *[4] Almost all sequences w over a finite alphabet $\mathcal{A}$, with $|\mathcal{A}| = b$, satisfy $p_n(w) = b^n$ for all $n \geq 0$.*

The theorem above could be viewed as a weaker version of the Normal Number Theorem. If we use $\mathcal{A} = \{0, 1, ..., b-1\}$ and think of the elements of $\mathcal{A}^{\mathbb{N}}$ as the base b expansion of a real number in the unit interval then theorem 5.2.1.2 asserts that every finite factor occurs at least once in almost all real numbers in the unit interval. In order for this correspondence to be valid we need to consider what measures are being used in both cases. Theorem 5.2.1.2 uses a topological measure that is defined on cylinder sets, while the Normal Number Theorem uses Lebesgue measure. The topological measure gives the same measure to the cylinder sets as Lebesgue measure gives to the corresponding cylinder sets in the unit interval. It is in this way that the measures are comparable.

5.2.2 Geometric Interpretation of Characteristic Words

Consider the positive upper right quadrant of $\mathbb{R}^2$ with the lines $x = n$ and $y = m$ for all $n, m \in \mathbb{N} = \{1, 2, 3, 4, ...\}$. Let $\alpha > 0$ be an irrational number. Let L_α be the line $y = \alpha x$. The line L_α goes through the origin and considering only $x > 0$

we keep track of the order in which L_α intersects the various vertical and horizontal lines. We will keep track of this in the following way.

$$w_i = \begin{cases} 0 & \text{if } L_\alpha \text{ intersects a vertical line,} \\ 1 & \text{if } L_\alpha \text{ intersects a horizontal line.} \end{cases}$$

The word $w = w_1w_2w_3...$ we call the cutting sequence of α. This is a nice geometric representation of Sturmian words as lines with an irrational slope.

5.3. Proof That the Fibonacci System is Normal

As we have previously seen for a given α, the choice of ρ in a Sturmian word does not change the complexity of the word.[4] Furthermore, all Sturmian words with slope α have exactly the same factors regardless of the ρ used.[4] Thus we may consider only the characteristic word for Sturmian words, which in turn tells us that we can examine the orbit of 0 under the map $T(x) = x + \phi \pmod 1$, since the orbits of other points coincide with different choices of ρ. The orbit of 0 under T is the Fibonacci word.

Knowing that the Fibonacci word is Sturmian immediately tells us the frequency of 0 and of 1. The following theorem and corollary state these results.

Theorem 5.3.0.1. *The characteristic word with slope* $\phi = \dfrac{\sqrt{5}-1}{2} \approx .618...$, *(i.e. the conjugate of the Fibonacci word), has a frequency of* 1*'s equal to* $\phi = \dfrac{\sqrt{5}-1}{2}$. *Also the frequency of* 0*'s in the Fibonacci word is* $1 - \phi = \phi^2 = \dfrac{3-\sqrt{5}}{2}$.

Proof. Theorem 5.2.0.10 and its corollary have already proven this.

□

Corollary 5.3.0.1. *The Fibonacci word has the frequency of* 1*'s equal to* ϕ^2 *and frequency of* 0*'s equal to* ϕ*.*

Proof. This is a result from the fact that $w_\alpha = C(w_{1-\alpha})$ for C the conjugate operation.

□

We just showed what the frequency of single digits is in the Fibonacci word. We can extend this to factors of any finite length using the structure of the infinite word. Recall that a typical normal number used the idea of a symbolic dynamical system in its definition. Instead of using the symbolic dynamical system to construct the Fibonacci word we will use the substitution system. This will enable us to use the self similarity that is present using the substitution system.

We should reiterate the various ways that the infinite Fibonacci word can be constructed. The first is via the $\mathcal{P}$-name or symbolic name for the orbit of 0 under T. Second, is via the Sturmian word with slope $\phi^2 = \dfrac{3-\sqrt{5}}{2}$. Third, is via the substitution system defined by φ as seen at the beginning of section 5.1. We can show that certain blocks will occur in the same pattern and order that the 0's and 1's do. This is the self similarity we mentioned earlier. For instance, if we look at the Fibonacci factors of the infinite Fibonacci word:

$$\varphi^0(1) = F_1 = 1$$

$$\varphi^1(1) = F_2 = 0$$

$$\varphi^2(1) = F_3 = 01$$

$$\varphi^3(1) = F_4 = 010$$

$$\varphi^4(1) = F_5 = 01001$$

$$\varphi^5(1) = F_6 = 01001010$$

We can see that $\varphi^k(1) = F_{k+1}$ and $\varphi^k(0) = F_{k+2}$. Thus we can use successive Fibonacci factors to pave the infinite Fibonacci word in the following way.

If we replace every 0 with a F_{k+2} and every 1 with a F_{k+1} for any $k > 0$, we will again have the infinite Fibonacci word, since we have only applied the substitution an additional k times and the Fibonacci word is the fixed point of the substitution. So we can analyze the frequency with which these blocks will appear in the Fibonacci word. We can use this fact to create a paving of the Fibonacci word using successive Fibonacci blocks, F_k and F_{k-1}. A *paving* is a way to perfectly cover an infinite word with some finite set of its factors. The ability to pave the Fibonacci word in this way is extremely useful in our arguments.

We will now give a brief example of how our argument will work before we start proving anything.

Definition 5.3.0.1. *Let $\#_n(w \mid [a_1a_2...a_k])$ denote the number of times that the block $[a_1a_2...a_k]$ occurs in the first n digits of w and let*

$$Freq_{[a_1a_2...a_k]}(w) = \lim_{n\to\infty} \frac{1}{n}\#_n(w \mid [a_1a_2...a_k]).$$

As always, we let $\phi = \dfrac{\sqrt{5}-1}{2}$ and let F be the infinite Fibonacci word. Then in this new notation we have $Freq_{[0]}(F) = \phi$ and $Freq_{[1]}(F) = 1 - \phi = \phi^2$.

We can directly calculate the number of times that the finite Fibonacci words will be seen in the first f_n digits, where f_n is the n^{th} Fibonacci number. Let's consider $F_4 = 010$. If we calculate the frequencies for the first few values of f_n, we get

$$\begin{aligned}
\#_{f_4}(F \mid [010]) &= 1 = f_2 \\
\#_{f_5}(F \mid [010]) &= 1 = f_3 - 1 \\
\#_{f_6}(F \mid [010]) &= 3 = f_4 \\
\#_{f_7}(F \mid [010]) &= 4 = f_5 - 1 \\
\#_{f_8}(F \mid [010]) &= 8 = f_6 \\
\#_{f_9}(F \mid [010]) &= 12 = f_7 - 1 \\
\#_{f_{10}}(F \mid [010]) &= 21 = f_8.
\end{aligned}$$

There is a boundary condition that changes the counting, since every other finite Fibonacci word ends with a 01 and all finite Fibonacci numbers start with a 0. Also $|F_n| = f_n$ and for n odd F_n ends with 01. Also if n is even then F_n ends with 10. Therefore, when n is odd $F_n = F_{n-1}F_{n-2}$ has 10.0 at the concatenation point, where the period is where the two factors were concatenated. The counting misses the next occurrence of the factor by one digit. Given that $F_n = F_{n-1}F_{n-2}$, we find that the recursion for calculating Fibonacci numbers drives the recursion we are finding. So we have shown that for 010 the number of occurrences in the first f_{2n} digits of F is

$\#_{f_{2n}}(F \mid [010]) = f_{2n-2}$. If we analyze the frequency of this block we get:

$$Freq_{[010]}(F) = \lim_{n\to\infty} \frac{1}{f_{2n}} \#_{f_{2n}}(F \mid [010]) = \lim_{n\to\infty} \frac{f_{2n-2}}{f_{2n}} = \lim_{n\to\infty} \frac{f_{2n-2}}{f_{2n-1}} \frac{f_{2n-1}}{f_{2n}} = \phi^2.$$

Recall that $\lim_{n\to\infty} \dfrac{f_{n-1}}{f_n} = \phi$.

We can perform similar observations with the other finite Fibonacci words. We find that $\#_{f_n}(F \mid F_k) = f_{n-k+2}$ with k and n having the same parity. Therefore, we can see that $Freq_{F_k}(F) = \phi^{k-2}$ since $n - (n-k+2) = k-2$. The following lemmas will lead to a proof of this.

Lemma 5.3.0.1. *Given finite Fibonacci words F_k and F_{k-1} the concatenation $F_{k-1}F_k$ contains two factors of F_k's, i.e., the overlap produces a factor of F_k. On the other hand, the concatenation F_kF_{k-1} contains only one factor of F_k.*

Proof. To prove the first statement we start with the basic equality,

$$F_k = F_{k-1}F_{k-2} = F_{k-2}F_{k-3}F_{k-2}.$$

Then we have that $F_{k-1}F_k = F_{k-2}F_{k-3}F_{k-2}F_{k-3}F_{k-2}$ which has two and only two $F_{k-2}F_{k-3}F_{k-2}$ factors, thus proving the first statement.

The second statement could be rephrased to say the overlap of F_kF_{k-1} does not produce an additional factor of F_k. To see this we examine

$$F_kF_{k-1} = F_{k-2}F_{k-3}F_{k-2}F_{k-2}F_{k-3}.$$

Which as we see does not have a factor of F_k in the overlap. If it did, we would see a second $F_{k-2}F_{k-3}F_{k-2}$ as a factor, and we do not. Therefore, the statements are proven. □

Lemma 5.3.0.2. *Given $k \geq 2$. The number of 0's in F_k is f_{k-1} and the number of 1's is f_{k-2}.*

Proof. We will prove this lemma using induction on k. For the case $k = 2$ we have $F_2 = 0$. The statement clearly holds. Now assume that for all $m \leq k$ that the number of 0's in F_k is f_{k-1} and the number of 1's is f_{k-2}. Then $F_{k+1} = F_k F_{k-1}$ and therefore the number of zeroes in F_{k+1} is $f_k = f_{k-1} + f_{k-2}$, since those are the number of zeroes in F_k and F_{k-1}, respectively. The same argument gives the result for the number of 1's in F_{k+1}. □

Lemma 5.3.0.3. *If k is even then the number of factors, 10, in F_k is f_{k-2}. If k is odd then the number of factors, 10, in F_k is $f_{k-2} - 1$. In other words, the number of factors, 10, in F_k is equal to $f_{k-2} - (k \mod 2)$.*

Proof. It is easily verified that for k even F_k ends in 10 and for j odd F_j ends in 01. Let $k \geq 2$ be an integer. Then for the case when $k = 2$ we have that $F_2 = 0$ which has $0 = f_0$, 10, factors. For the case $k = 3$ we have $F_3 = 01$ which has $0 = f_1 - 1$, 10, factors. Now assume for all even $m \leq k$ that the number of factors 10 in F_m is f_{m-2} and for all odd $m \leq k$ that the number of factors 10 in F_m is $f_{m-2} - 1$. Then for k even $F_{k+2} = F_{k+1} F_k$. Since k is even, F_{k+1} ends in 01. So the overlap will produce an additional factor of 10. Therefore, the number of 10 factors in F_{k+2} is $f_{k-1} + f_{k-2} - 1 + 1$, where the plus one is from the overlap.

Now for the case when k is odd. Then $k + 1$ is even. So the overlap in $F_k = F_{k+1} F_k$ does not produce an additional factor of 10 since F_{k+1} would end in 10 and all finite Fibonacci words start with 0. Therefore the number of factors 10 in F_{k+2} is $f_{k-1} + f_{k-2} - 1 = f_k - 1$. This completes the proof of the lemma. □

Lemma 5.3.0.4. *For $k \geq 4$, the number of factors, 00, in F_k is $f_{k-3} - 1$ when k is even and is f_{k-3} when k is odd. In other words, the number of factors, 00, in F_k is $f_{k-3} - (k + 1 \mod 2)$. If $k < 4$ the number of, 00, factors is 0.*

Proof. The case when $k < 4$ is easily verified via direct observation. For $k = 4$ we have that $F_4 = 010$ which has $0 = f_1 - 1$ factors of, 00. For $k = 5$ we have that $F_5 = 01001$ which has $1 = f_2$ factors of, 00.

Now assume for all $k \leq n$ that the number of factors, 00, in F_k is $f_{k-3} - (k + 1 \mod 2)$. First let us consider n to be even. Then $n + 1$ is odd. We know $F_{n+1} = F_n F_{n-1}$. When n is even F_n ends with 10 and all of the F_k's begin with 0. So we gain one 00 factor in the overlap of F_n and F_{n-1}. Thus there are $f_{n-3} - 1 + f_{n-4} + 1 = f_{n-2}$ factors of 00 in F_{n+1} when n is even.

In the case that n is odd we have that $n + 1$ is even. Again we know that $F_{n+1} = F_n F_{n-1}$. We know that n is odd so F_n ends in 01, meaning there is no factor of 00 gained in the overlap of F_n and F_{n-1} when n is odd. Therefore, there are $f_{n-3} + f_{n-4} - 1 = f_{n-2} - 1$ factors of 00 in F_{n+1} when n is odd. □

Theorem 5.3.0.2. *Given n and k such that $n \geq k$, then the structure of F_n being viewed with factors F_k and F_{k-1} as 0's and 1's is of the form F_{n-k+2} where each 0 in F_{n-k+2} is replaced with an F_k and each 1 replaced with a F_{k-1}. Also for $n \geq k$ and $n - k = 0 \ (mod\ 2)$ we have $\#_{f_n}(F \mid F_k) = f_{n-k+2}$ and for $n \geq k$ with $n - k = 1 \ (mod\ 2)$ we have $\#_{f_n}(F \mid F_k) = f_{n-k+2} - 1$. In other words, $\#_{f_n}(F \mid F_k) = f_{n-k+2} - (n - k \mod 2)$.*

Proof. We will prove this theorem by induction on n. Fix $k \geq 0$. Then for the base case $n = k$ we have $F_{n-k+2} = F_{k-k+2} = F_2 = 0$. So F_k has structure in terms of F_k and F_{k-1} of F_k. Now assume that for all $k \leq m \leq n$ we have that the statement holds true. Then $F_{n+1} = F_n F_{n-1}$ and both of F_n and F_{n-1} have structures of F_{n-k+2} and F_{n-k+1} when paved with factors of F_k and F_{k-1}. Therefore, the structure of F_{n+1} when paved by F_k and F_{k-1} is of the form $F_{n-k+2} F_{n-k+1} = F_{n-k+3} = F_{(n+1)-k+2}$.

Therefore, the structure of F_n being viewed with factors F_k and F_{k-1} as 0's and 1's is of the form F_{n-k+2} where each 0 in F_{n-k+2} is replaced with an F_k and each 1 replaced with a F_{k-1} is proven by induction.

This allows us to use lemma 5.3.0.3. Next, we will prove the second statement by induction. We have that for $n = k$ that $\#_{f_n}(F \mid F_k) = \#_{f_k}(F \mid F_k) = f_{k-k+2} = f_2 = 1$. Now we assume for all $m \leq n$ that the second result holds. Since the second statement needs $n - k = 0 \pmod 2$, we will look at the $n+2$ case and show that the statement holds. We know from the first statement that F_{n+2} has the structure of $F_{(n+2)-k+2} = F_{n-k+4}$ when paved by F_k and F_{k-1}. From lemmas 5.3.0.2 and 5.3.0.3 from above, we have that the number of 0's in F_{n-k+4} is f_{n-k+3} and the number of 10's is f_{n-k+2}. So we have f_{n-k+2} occurrences of $F_{k-1}F_k$ in F_{n-k+4}. Therefore using lemma 5.3.0.1 there are $f_{n-k+3} + f_{n-k+2} = f_{(n+2)-k+1} + f_{(n+2)-k} = f_{(n+2)-k+2}$ occurrences of F_k as a factor in $F_{(n+2)-k+2}$. The argument above can be used for the case when $n - k = 1 \pmod 2$ and the only difference is which part of lemma 5.3.0.3 is used. Therefore $\#_{f_n}(F \mid F_k) = f_{n-k+2} - 1$. This completes the proof of the second statement. □

Corollary 5.3.0.2. *The* $\lim_{n\to\infty} \frac{1}{f_n} \#_{f_n}(F \mid F_k) = \phi^{k-2}$ *for* $k \geq 4$.

Proof. We will consider two subsequences of the sequence $\{f_n\}_{n=1}^{\infty}$, the sequence of the Fibonacci numbers. These two subsequences depend on the k we are considering. The first subsequence consists of the f_n such that $n - k = 0 \pmod 1$. The second subsequence consists of the f_n such that $n - k = 1 \pmod 1$. We will show that both have the same limit.

Let $n - k = 0 \pmod 1$. Then

$$\lim_{n\to\infty} \frac{\#_{f_n}(F \mid F_k)}{f_n} = \lim_{n\to\infty} \frac{f_{n-k+2} - (n - k \mod 2)}{f_n} = \lim_{n\to\infty} \frac{f_{n-k+2}}{f_n} - \frac{(n - k \mod 2)}{f_n}$$

$$= \lim_{n\to\infty} \frac{f_{n-k+2}}{f_{n-k+1}} \frac{f_{n-k+1}}{f_{n-k}} \ldots \frac{f_{n-1}}{f_n} = \phi^{k-2}.$$

Since $n-(k-2) = n-k+2$ and $n-(n-(k-2)) = k-2$, that is how many fractions of the form $\frac{f_{m-1}}{f_m}$ will be in the limit. We know $\lim_{m\to\infty} \frac{f_{m-1}}{f_m} = \phi$. We also know that $\lim_{n\to\infty} \frac{(n-k \mod 2)}{f_n} = 0$.

Therefore $\lim_{n\to\infty} \frac{1}{f_n} \#_{f_n}(F \mid F_k) = \phi^{k-2}$. □

Theorem 5.3.0.3. $\mathrm{Freq}_{F_k}(F) = \lim_{n\to\infty} \frac{1}{n} \#_n(F \mid F_k) = \phi^{k-2}$ *for* $k \geq 4$.

Proof.

Given k let $n > 2k$. Let $M > f_{10n}$. Consider the first M digits of F denote this finite block MF. We can pave MF with the blocks F_n and F_{n-1} almost perfectly. There may be some leftover digits at the end. Let M_0 be such that the first M_0 blocks are paved perfectly by the blocks F_n and F_{n-1} and is maximal. Then $M - M_0 < f_n$. Therefore

$$0 \leq \lim_{M\to\infty} \frac{M - M_0}{M} \leq \lim_{M\to\infty} \frac{f_n}{M} = 0 \tag{5.5}$$

and

$$\lim_{M\to\infty} \frac{M_0}{M} = \lim_{M\to\infty} \frac{M}{M} - \frac{M - M_0}{M} = 1. \tag{5.6}$$

It is easy to see that

$$\lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M} \leq \lim_{M\to\infty} \frac{\#_M(F \mid F_k)}{M} \leq \lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M} + \frac{f_n}{M}. \tag{5.7}$$

The right inequality in equation 5.7 can be seen by observing that the number of F_k blocks in the digits from M_0 to M is less than f_n, since f_n is the upper bound for the number of digits that are outside of the paving. Multiplying by $\frac{M_0}{M_0}$ gives the following:

$$\lim_{M\to\infty} \frac{M_0}{M} \frac{\#_{M_0}(F \mid F_k)}{M_0} \leq \lim_{M\to\infty} \frac{\#_M(F \mid F_k)}{M} \leq \lim_{M\to\infty} \frac{M_0}{M} \frac{\#_{M_0}(F \mid F_k)}{M_0} + \frac{f_n}{M}.$$

Now using equations 5.5 and 5.6 we obtain the following:

$$\lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M_0} \leq \lim_{M\to\infty} \frac{\#_{M}(F \mid F_k)}{M} \leq \lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M_0}.$$

Therefore

$$\lim_{M\to\infty} \frac{\#_{M}(F \mid F_k)}{M} = \lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M_0}. \tag{5.8}$$

Let A_0 be the number of F_n blocks in the paving and let A_1 be the number of F_{n-1} blocks in the paving. Thus

$$A_0 f_n + A_1 f_{n-1} = M_0 \quad \text{and} \quad \frac{A_0 f_n + A_1 f_{n-1}}{M_0} = 1.$$

Let $\alpha = \dfrac{A_0 f_n}{M_0}$ then $1-\alpha = \dfrac{A_1 f_{n-1}}{M_0}$.

We need to count how many blocks F_k occur in the first M_0 digits of F. Recall that there are f_{n-k+2} blocks F_k in F_n and f_{n-k+1} blocks F_k in F_{n-1}. Thus the number of F_k blocks that lie inside the F_n and F_{n-1} blocks in our paving is $A_0 f_{n-k+2} + A_1 f_{n-k+1}$. Now we need to count the number of F_k blocks that occur in the overlaps of the paving. For a similar reason as to why lemma 5.3.0.1 is true, the only time a F_k block occurs is in an overlap of the form $F_{n-1}F_n$. Since the F_{n-1} block is analogous to the 1 in F we know that every F_{n-1} block is followed by a F_n block in our paving. Hence we have either A_1 or $A_1 - 1$ as the number of F_k blocks that occur in the overlaps of F_n and F_{n-1}. Let $\delta_M \in \{0, -1\}$. Then

$$\#_{M_0}(F \mid F_k) = A_0 f_{n-k+2} + A_1 f_{n-k+1} + A_1 + \delta_M = A_0 f_{n-k+2} + A_1(f_{n-k+1} + 1) + \delta_M. \tag{5.9}$$

Therefore

$$\frac{\#_{M_0}(F \mid F_k)}{M_0} = \frac{A_0 f_{n-k+2} + A_1(f_{n-k+1}+1) + \delta_M}{M_0} \tag{5.10}$$

$$= \alpha \frac{f_{n-k+2}}{f_n} + (1-\alpha)\frac{f_{n-k+1}+1}{f_{n-1}} + \frac{\delta_M}{M_0}. \tag{5.11}$$

When $\delta_M = 0$ we have that $\dfrac{\#_{M_0}(F \mid F_k)}{M_0}$ is a convex combination of $\dfrac{f_{n-k+2}}{f_n}$ and $\dfrac{f_{n-k+1}+1}{f_{n-1}}$. Therefore $\dfrac{\#_{M_0}(F \mid F_k)}{M_0}$ is in between the two values. When $\delta_M = -1$ we have an extra $\dfrac{-1}{M_0}$ that prevents equation 5.11 from being a perfect convex combination. However, we have that $\lim_{M\to\infty} \dfrac{-1}{M_0} = 0$. Thus as $M \to \infty$ equation 5.11 approaches a convex combination of the two values. Thus we again have that $\dfrac{\#_{M_0}(F \mid F_k)}{M_0}$ is in between the two values for sufficiently large M.

Therefore

$$\lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M_0}$$

is trapped between

$$\lim_{n\to\infty} \frac{f_{n-k+2}}{f_n} = \phi^{k-2}$$

and

$$\lim_{n\to\infty} \frac{f_{n-k+1}+1}{f_{n-1}} = \phi^{k-2}.$$

We can let n go to infinity as M goes to infinity. Thus

$$\lim_{M\to\infty} \frac{\#_{M_0}(F \mid F_k)}{M_0} = \lim_{M\to\infty} \frac{\#_{M}(F \mid F_k)}{M} = \phi^{k-2}.$$

$\square$

Now we need to consider factors of the infinite Fibonacci word, F, that are not equal to F_k for all k. We want to eventually prove

$$Freq_B(F) = \lim_{n\to\infty} \frac{1}{n} \#_n(F \mid B) = \phi^j$$

for every finite word that exists as a factor of F. The j depends on what the smallest k is such that F_k has B as a factor, and how B fits inside of F_k. First we will discuss the two different ways that B can be a factor in F_k that yield different values for $Freq_B(F)$. Then we will examine the two different cases using arguments very similar to theorem 5.3.0.3 to find the values of $Freq_B(F)$.

Lemma 5.3.0.5. *Let $V_k = \{B \mid B \text{ is a factor of} F_k \text{ and} B \text{ is not a factor of} F_j$ for all $j < k\}$, then for all $B \in V_k$, B can only occur once in F_k.*

Proof. We will prove this lemma by induction. For $k = 1, 2$, and 3 the lemma is trivially true. Assume, for all $m \leq k-1$ the lemma holds. Let $B \in V_k$, then B is a factor of $F_k = F_{k-1}F_{k-2}$ such that B has the form $B'B'' = B$ with B' a factor of F_{k-1} and B'' a factor of F_{k-2}. Therefore, $B' \in V_i$ for some $i \leq k-1$ and $B'' \in V_j$ for some $j \leq k-2$. The following arguments are examining if a factor of F_iF_j can exist in F_{k-1}.

If $i = k-1$, the lemma holds for any j, since there is only one occurrence of F_{k-1} as a factor of $F_{k-1}F_{k-2}$ by lemma 5.3.0.1. If $j = k-2$, the lemma holds for any i, since the only other occurrence of F_{k-2} in $F_{k-1}F_{k-2} = F_{k-2}F_{k-3}F_{k-2}$ is as a prefix.

Let $i < k-1$ and $j < k-2$, then given that B' is a suffix of F_{k-1}, we can see that i must be equal to $k-l$ with l being odd. This puts constraints on the values j. In particular, $j \neq k-l-1$ for all l, because B would be a factor of F_{k-l+1}, contradicting that $B \in V_k$. Also, $j \neq i$ for $i \leq k-3$, since F_{k-1} contains a factor of F_iF_i for $i \leq k-3$.

Claim: If $i < k-1$ then $j = k-2$.

Proof of claim: Let $i = k-3$ and $j = k-5$, this is the largest value j can take when $i = k-3$, that is not $k-2$. However, there is a factor of $F_{k-3}F_{k-5}$ in $F_{k-1} = F_{k-3}F_{k-5}F_{k-6}F_{k-3}$. This contradicts that $B \in V_k$ or that $B'' \in V_j$. If the we consider smaller values for i and j, we will be inside the case above. Hence $j = k-2$

if $i < k-1$.

Therefore, B can only occur once in F_k.

$\square$

Lemma 5.3.0.6. *Let $V_k = \{B \mid B \text{ is a factor of } F_k \text{ and } B \text{ is not a factor of } F_j$ for all $j < k\}$, then the number of times B occurs in F_n is either $f_{n-k+2} + \delta$ or $f_{n-k+3} + \gamma$ for $\delta \in \{0, -1\}$ and $\gamma \in \{0, -1, -2, -3\}$ depending on k and for $n > k$.*

Proof. Fix k. Let $B \in V_k$. We already know how many times B occurs in F_n if it does not occur in the overlaps of blocks of the form F_kF_{k-1} (thus B will also not occur in the overlap of blocks of the form F_kF_k), since the counting will work exactly the same as it did in the proof of theorem 5.3.0.2. Recall that the block F_kF_{k-1} contains only one factor of F_k. Therefore, we need to classify when B can and cannot occur in the overlap of a block F_kF_{k-1}. The following argument works for both block types, F_kF_k and F_kF_{k-1}. We will only examine the F_kF_{k-1} case.

We claim that B cannot occur in the overlap of a block F_kF_{k-1} if $F_k = XBa$ or $F_k = YB$ for X and Y finite factors of F_k and $a \in \{0, 1\}$. Before we show this we should state some obvious, but important, facts. Recall that $F_k = F_{k-1}F_{k-2}$. For B to exist in V_k it must overlap the blocks $F_{k-1}F_{k-2}$, since otherwise it would be contained in some F_j with $j < k$. Let F'_{k-2} be such that $F_{k-2} = F'_{k-2}ab$ for some $a, b \in \{0, 1\}$ then $s(F_{k-2}) = F'_{k-2}ba$ by theorem 5.1.0.9. We will use parentheses only to help illustrate groupings from one line to the next. They do not have any special significance. Then

$$F_kF_{k-1} = (F_{k-1}F_{k-2})(F_{k-2}F_{k-3}) \tag{5.12}$$

$$= F_{k-1}F_{k-2}(F_{k-3}F_{k-4})F_{k-3} \tag{5.13}$$

$$= F_{k-1}(F_{k-2}F_{k-3})(F_{k-4}F_{k-3}) \tag{5.14}$$

$$= F_{k-1}F_{k-1}s(F_{k-3}F_{k-4}) \tag{5.15}$$

$$= F_{k-1}F_{k-1}s(F_{k-2}) \tag{5.16}$$

$$= F_{k-1}F_{k-1}F'_{k-2}ba. \tag{5.17}$$

So we can see that B can occur in the overlap of F_kF_{k-1} if it exists such that $F_k = LBK$ for factors L, B, K of F_k with $|K| \geq 2$. Otherwise B cannot occur in the overlap of F_kF_{k-1}.

Thus for B such that $F_k = LBK$ for factors L, B, K of F_k with $|K| \geq 2$, we have that B occurs $f_{n-k+2} + \delta$ times as a factor of F_n for $n > k$. For the case when B exists such that $F_k = XBa$ or $F_k = YB$ for X and Y finite factors of F_k and $a \in \{0, 1\}$, we have that the number of times B occurs as a factor of F_n is equal to the $F_{n-k+2} + \delta$ plus the number of times F_kF_{k-1} occurs in F_n plus the number of times F_kF_k occurs in F_n. The number of times F_kF_{k-1} occurs in F_n is $f_{n-k+1} + \delta$. The number of times F_kF_k occurs in F_n is $f_{n-k} + \delta$. Therefore the number of times B occurs as a factor of F_n is $f_{n-k+2} + \delta + f_{n-k+1} + \delta + f_{n-k} + \delta = f_{n-k+3} + \gamma$ for $\gamma \in \{0, -1, -2, -3\}$. We are abusing notation a little bit since the three δ's may be different. But this will not affect our argument as they will disappear when we take limits to obtain frequencies. $\square$

We are now ready to prove that all factors of the infinite Fibonacci word F have the appropriate frequency of occurrence.

Theorem 5.3.0.4. *Let* $V_k = \{B \mid B \textit{ is a factor of } F_k \textit{ and } B \textit{ is not a factor of } F_j$

for all $j < k\}$, then for all $k \geq 4$ and for all $B \in V_k$ we have

$$\mathrm{Freq}_B(F) = \lim_{n\to\infty} \frac{1}{n} \#_n(F \mid B) = \phi^{k-2}$$

or

$$\mathrm{Freq}_B(F) = \lim_{n\to\infty} \frac{1}{n} \#_n(F \mid B) = \phi^{k-1}.$$

Proof. We will use nearly the exact same argument that was used in theorem 5.3.0.3. The crux of that argument was putting finite Fibonacci blocks inside of bigger Fibonacci blocks that could be used to pave the first M digits of F with M very large.

Given k, let $B \in V_k$ and $n > 2k$. Let $M > f_{10n}$. Using the same notation used in theorem 5.3.0.3 and for the same reasons we immediately get

$$\lim_{M\to\infty} \frac{\#_M(F \mid B)}{M} = \lim_{M\to\infty} \frac{\#_{M_0}(F \mid B)}{M_0}. \tag{5.18}$$

The case where B occurs f_{n-k+2} times as a factor of F_n uses the exact same argument as theorem 5.3.0.3. Giving the result $Freq_B(F) = \phi^{k-2}$.

The case where B occurs f_{n-k+3} times as a factor of F_n uses a similar argument. But it is worth going over since it is not exactly the same nor is it perfectly obvious. Recall that we have a perfect paving by the blocks F_n and F_{n-1} covering the first M_0 digits of F.

We need to redefine what the A_i's are in comparison to the previous theorem. Let A_0 be the number of $F_{n-1}F_nF_{n-1}$ blocks in the paving up to M_0. Let A_1 be the number of F_{n-1} blocks in the paving and let A_2 be the number of F_nF_n blocks in the paving. Thus

$$M_0 = A_0 f_n + A_1 f_{n-1} + A_2 2 f_n \text{ and } 1 = \frac{(A_0 + 2A_2)f_n + A_1 f_{n-1}}{M_0}.$$

Let $\alpha = \dfrac{(A_0 + 2A_2)f_n}{M_0}$ then $1 - \alpha = \dfrac{A_1 f_{n-1}}{M_0}$.

We will pick up extra occurrences of B in the overlaps of $F_n F_n$ and $F_{n-1}F_nF_{n-1}$, getting one extra in the first overlap and two extra in the second overlap. So in the paving there will be A_2 occurrences of B in the first overlap type and A_0 occurrences of B in the second overlap type. Therefore,

$$\begin{aligned}
\#_{M_0}(F \mid B) &= A_0 f_{n-k+3} + A_1 f_{n-k+2} + 2A_2 f_{n-k+3} + A_0 + A_2 + \delta_M && (5.19)\\
&= (A_0 + 2A_2)(f_{n-k+3} + 1) + A_1 f_{n-k+2} + \delta_M, && (5.20)
\end{aligned}$$

where $|\delta_M| < 10$ is a term that takes into account the M_0 boundary where we might have over-counted by a small amount, since we do not know exactly how our paving ends.

Therefore,

$$\begin{aligned}
\frac{\#_{M_0}(F \mid B)}{M_0} &= (A_0 + 2A_2)(f_{n-k+3} + 1) + A_1 f_{n-k+2} + \delta_M && (5.21)\\
&= \alpha \frac{f_{n-k+3} + 1}{f_n} + (1 - \alpha)\frac{f_{n-k+2}}{f_{n-1}} + \frac{\delta_M}{M_0}. && (5.22)
\end{aligned}$$

We can again use the same convex combination argument used earlier to show that $\dfrac{\#_{M_0}(F \mid B)}{M_0}$ is between the two values $\dfrac{f_{n-k+3} + 1}{f_n}$ and $\dfrac{f_{n-k+2}}{f_{n-1}}$ for large enough M. Also

$$\lim_{n\to\infty} \frac{f_{n-k+2}}{f_{n-1}} = \lim_{n\to\infty} \frac{f_{n-k+3} + 1}{f_n} = \phi^{k-1}.$$

We can let n go to infinity as M goes to infinity. Therefore in this case

$$\lim_{M\to\infty} \frac{\#_{M_0}(F \mid B)}{M_0} = \phi^{k-1}.$$

This completes the proof.

□

Now we have that all factors of the infinite Fibonacci word, F, have a frequency of occurrence equal to a power of $\phi = \frac{\sqrt{5}-1}{2}$. The finite Fibonacci blocks F_n occur with frequency ϕ^{n-2} for $n \geq 4$. We already have the frequencies for $F_1 = 1$ and $F_2 = 0$. The frequencies of $F_3 = 01$ and 10 are the same as for the digit 1, since two 1's cannot appear next to each other.[23] The only other factor that may have not been covered by the theorems and lemmas is 00. However, 00 exists for the first time in F_5. Thus we have determined the frequency of occurrence of all factors of F.

The infinite Fibonacci word has $n + 1$ factors of length n. This applies to the dynamical system having $n + 1$ distinct cylinder sets of size n. It is easily seen that the length 1 cylinder sets of F have measures that are powers of ϕ. Also, the measures match with the frequencies of the digits. Since we have the relation $\phi^n = \phi^{n+1} + \phi^{n+2}$, or equivalently $\phi^n - \phi^{n+1} = \phi^{n+2}$, we can show that all cylinder sets have measures equal to powers of ϕ.

Viewing the Fibonacci system from the symbolic dynamical system perspective, we can use the fact that irrational rotations are ergodic[21] to apply the ergodic theorem to the Fibonacci system. The ergodic theorem yields the following: for almost all $x \in [0, 1)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_C(T^j(x)) = \mu(C).$$

Theorem 5.3.0.5. *The Fibonacci system,* $([0, 1), \Sigma, \mu, T, \mathcal{P})$*, has every point being normal.*

Proof. Let $x \in [0, 1)$, then there exists a $k \geq 0$ such that x and $T^k(0)$ are as close as we want, since the orbit of 0 is dense in $[0, 1)$.[17] Furthermore, let $y = T^k(0)$ then $T^j(x)$ and $T^j(y)$ are as close as we want for all $j \geq 0$. This can be seen by observing that T is simply a translation on the unit circle. Therefore, the closeness we speak of is in terms of the distance between points on the circle i.e., $|x-y| = \min\{|x-y|, |x-y+1|\}$.

It would be sufficient to show that x and y differ at a relatively small number of places so that the number of differences would not affect the frequencies of factors. The closer x and y are to each other the less often their $\mathcal{P}$-names will differ. Let $\{x_i\}_{i=1}^{\infty}$ and $\{y_i\}_{i=1}^{\infty}$ be the $\mathcal{P}$-names of x and y, respectively. If $x_l \neq y_l$ then $\mathbb{1}_{[x,y]}(T^{-l}(0)) + \mathbb{1}_{[x,y]}(T^{-l}(\phi^2)) = 1$. If $x_m = y_m$ then $\mathbb{1}_{[x,y]}(T^{-l}(0)) + \mathbb{1}_{[x,y]}(T^{-l}(\phi^2)) = 0$. Let $D_n(x,y) = \sum_{j=0}^{n-1} \mathbb{1}_{[x,y]}(T^{-j}(0)) + \mathbb{1}_{[x,y]}(T^{-j}(\phi^2))$. The function, $D_n(x,y)$, counts the number of differences in the first n symbols of the $\mathcal{P}$-names of x and y. By making x and y very close i.e., $\mu([x,y]) = \epsilon$, we can make $D_n(x,y)$ small enough so that $\lim_{n\to\infty} \frac{1}{n} D_n(x,y) = 0$. Therefore the frequencies of factors in x are the same as the frequencies of factors in y, which has the same frequencies of factors as the Fibonacci word. Therefore, all of the points in the Fibonacci system have frequencies of factors equal to the frequencies of factors in the Fibonacci word. The ergodic theorem gave us that these frequencies are equal to the measure of the cylinder sets. Hence the Fibonacci system has every point normal.

□

5.3.1 Connection to Certain Sturmian Words

The argument used to show that the Fibonacci word is normal can be applied to a specific subset of Sturmian words. There is a connection between the regular continued fraction expansion of a real number α and the Sturmian word with slope α. This connection is stated in the theorem below.

Theorem 5.3.1.1. *[4] Let $0 < \alpha < 1$ be an irrational number and let its regular* *continued fraction expansion be* $[0; 1 + a_1, a_2, ...]$. *Define* $X_0 = 1$, $X_1 = 0$, *and* $X_n = X_{n-1}^{a_{n-1}} X_{n-2}$ *for* $n \geq 2$. *Then the Sturmian word with slope* α, w_α, *is equal to* $\lim_{n\to\infty} X_n$.

If we take $\alpha = [0; 2, \overline{1}]$ we will have exactly the Fibonacci word, since $\phi^2 = \frac{3-\sqrt{5}}{2} = [0; 2, \overline{1}]$. Recall the recurrence we had for the Fibonacci word was $F_n = F_{n-1}F_{n-2}$ for $n \geq 2$. This recurrence matches what the theorem tells us it should be. The fact that this recurrence is the same for all n was an important part of the arguments used in proving what the frequencies of factors in the Fibonacci word are.

We can then generalize this argument to work with irrational numbers that have regular continued fraction expansions of the form $\alpha = [0; k + 1, \overline{k}]$. Continued fractions of this form will have the same recursion $X_n = X_{n-1}^k X_{n-2}$ for all $n \geq 2$. Therefore, the techniques we used can be modified slightly to take into account the slight variation in the recursion. Therefore, we can show that all Sturmian words with slopes having continued fraction expansions of the form $[0; k + 1, \overline{k}]$ will be normal with respect to the symbolic dynamical system the represent.

6. CONCLUSION AND FUTURE WORK

In this paper, we have taken the traditional approach to normal numbers and reinterpreted it to fit with the mechanics of measure preserving transformations. This allowed us to change the way we think about how a normal number behaves. Instead of having to walk along the base b expansion of a number, looking for certain blocks to appear, we bring the expansion to us. More formally, we look for numbers that visit cylinder sets the appropriate amount of time.

This change in view of a normal number led us to a novel proof of the Normal Number Theorem. Previous proofs can only claim that the set of full measure they produce will contain only absolutely normal numbers. Meaning, it is likely that there are some absolutely normal numbers that aren't captured. On the other hand, our proof constructs a set that contains all of the absolutely normal numbers. While this changes nothing in reference to the result of the Normal Number Theorem, it may prove useful to those studying the set of absolutely normal numbers, giving them a precise construction of the set.

We have also organically extended the idea of a normal number to be a property of any point in a symbolic dynamical system. Our reinterpretation allowed this extension to be a very natural one, where the definition does not need to change for different partitions and measure preserving transformations.

Currently there are many unanswered questions surrounding normal numbers. The looming question about how to show that a specific real number is normal, for instance π, may never be solved. There is, however, a conjecture mentioned in [20] that all irrational algebraic numbers are absolutely normal. There is still no known counterexample, so the conjecture remains valid. But approaching the problem from a dynamical system viewpoint may prove to be fruitful for number theorists.

The fact that the Fibonacci system has every point being normal shows there is a difference between systems with positive entropy and systems with 0-entropy. In general, the ergodic theorem can be used to show that ergodic systems have almost all points being normal.

As questions about dynamical systems get answered, these answers may help with our understanding of normal numbers, giving us deeper insight to which specific numbers are normal, perhaps answering the 100-year-old questions, is π normal to any base b? absolutely normal? The answers to these questions will undoubtably help some mathematicians sleep better at night, including the author of this paper.

BIBLIOGRAPHY

1. E. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Rend. Circ. Mat. Palermo 27, (1909), 247 - 271.

2. M.W. Sierpinski, *Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'un tel nombre*, Bull. Soc. Math. France 45, (1917), 127 - 132.

3. Champernowne, D.G., *The Construction of Decimals Normal in the Scale of Ten*, J. London Math. Soc. 8, (1933), 254 - 260.

4. Jean-Paul Allouche & Jeffrey Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, (2003).

5. Kuipers, L. and Niederreiter, H., *Uniform Distribution of Sequences*, Dover, Moneola, New York, (2006).

6. Copeland, A. H.; Erdós, P., *Note on normal numbers*, Bulletin of the American Mathematical Society 52, (1946), 857 - 860.

7. Becher, Verónica; Figueira, Santiago, *An Example of a Computable Absolutely Normal Number*, Theoretical Computer Science 270, (2002), 947 - 958.

8. Wall, D.D., *Normal Numbers*, Ph.D. thesis, Berkeley, California: University of California, (1949).

9. Birkhoff, G. D., *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U.S.A. 17, 656 - 660, (1931).

10. Keane, Michael and Perterson, Karl, *Easy and nearly simultaneous proofs of the Ergodic Theorem and Maximal Ergodic Theorem*, IMS Lecture NotesMonograph Series Dynamics & Stochastics Vol. 48, (2006), 248 - 251.

11. Lind, Douglas and Marcus, Brian, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, (1995).

12. Munkres, James, *Topology*, Prentice Hall, (2000).

13. Orenstein, Donald S., *Ergodic Theory, Randomness, and Dynamical Systems*, Yale University Press, (1974).

14. Queffelec, Martine, *Substitution Dynamical Systems - Spectral Analysis*, Springer-Verlag, (1987).

15. Billingsley, Patrick, *Probability and Measure*, John Wiley and Sons, Inc, (1979).

16. Khoshnevisan, Davar, *Probability*, American Mathematical Society, (2007).

17. Shields, Paul, *The Ergodic Theory of Discrete Sample Paths*, American Mathematical Society, (1996).

18. Petersen, Karl, *Ergodic Theory*, Cambridge University Press, (1983).

19. Ross, Sheldon, *A First Course In Probability*, Pearson Prentice Hall, (2006).

20. Bailey, David, and Crandall, Richard E., *On the Random Character of Fundamental Constant Expansions*, (2001).

21. Billingsley, Patrick, *Ergodic Theory and Information*, John Wiley & Sons, Inc, (1965).

22. Cassaigne, Julien, *Constructing infinite words of intermediate complexity*, Springer-Verlag, Proceedings of the 6th international conference on Developments in language theory, (2003).

23. Mignosi, F., Restivo, M., Sciortino, M., *Words and forbidden factors*, Theoretical Computer Science Volume 273, Issues 1-2, 28 February 2002, Pages 99-117.

APPENDICES

A APPENDIX Some Results on Normal Numbers

A1 Finite Variations

Any set of normal numbers, whether they are absolutely normal, simply normal, or normal to some base b, are closed under finite variations. This means that adding, removing, or changing any finite number of digits in the b-adic expansion leaves the property it started with intact.

Therefore, the first N digits of the b-adic expansion of some α has no impact on whether or not α is normal to the base b. Properties that do not depend on any finite number of beginning digits we call tail properties. In other words, being normal to the base b, or absolutely normal, or simply normal, is a tail property.

A2 Equidistribution of Sequences

For a more thorough treatise we refer the reader to [5].

A sequence of real numbers, ω, is *uniformly distributed modulo 1* if for every pair of a, b of real numbers with $0 \leq a < b \leq 1$ we have

$$\lim_{N\to\infty} \frac{C([a,b);N;\omega)}{N} = b - a$$

where $C([a,b);N;\omega)$ counts the number of terms of the sequence ω up to the N^{th} element of the sequence that has its floor in $[a,b)$.

A classical result involving normal numbers and equidistribution of sequences modulo 1 is that the sequence of real numbers $(b^k x)_{k=0}^{\infty}$ is uniformly distributed modulo 1 if and only if x is normal to the base b. This can be stated in a beautiful way using Weyl's criterion

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i m b^k x} = 0 \quad \text{for all integers } m \geq 1.$$

Being able to work with uniformly distributed sequences to prove things about normal numbers has led to some great results, including the following two theorems in the next section of this appendix.

A3 Normality from Simple Normality

The results we will discuss in this section are fully proven in [5].

Theorem A3.1. *Let $k \geq 2$ be an integer. A real number α is normal to the base b if and only if α is normal to the base b^k.*

If α is normal to the base b then all blocks of length k have the proper frequency of b^{-k}. If we view these blocks as the digits of the b^k-adic expansion of α, it is clear that α is simply normal to the base b^k. We can make the same assertions about blocks of length b^{nk} for all $n \geq 0$, which gives us that α is normal to the base b^k.

Theorem A3.2. *The real number α is normal to the base b if and only if it is simple normal to all of the bases $b, b^2, b^3, b^4, ...$*

If α is normal then it is clearly simply normal to $b, b^2, ...$ by the theorem above, since that theorem would guarantee that α is actually normal to the bases $b, b^2, ...$ a stronger result than needed. Now consider that α is simply normal to all the bases $b, b^2, ...$, then if we view the digits in base b^k as being blocks of length k in the base b, we see that we have that α is normal to the base b.

The discussion after each of these proofs was meant to be very brief and give a general idea of how one can think about why these theorems are true.

A4 Closed Under Multiplication by a Rational

A result due to Wall [8] (1949) states that given a normal number α and for all non-zero rational numbers x, we have that αx is normal.

B APPENDIX Fibonacci Numbers

This appendix will provide definitions for the Fibonacci numbers that this paper uses, to avoid confusion with definitions that may differ.

Let $f_0 = 0$ and $f_1 = 1$. Then the rest of the sequence $\{f_n\}_{n=0}^{\infty}$ via the recurrence

$$f_n = f_n + f_{n-1}.$$

The sequence is called the Fibonacci sequence and the entries are Fibonacci numbers.

For the reader's benefit here is a list of the first few Fibonacci numbers.

f_0	0
f_1	1
f_2	1
f_3	2
f_4	3
f_5	5
f_6	8
f_7	13
f_8	21
f_9	34
f_{10}	55
f_{11}	89
f_{12}	144
f_{13}	233

C APPENDIX Fibonacci Blocks

This appendix is provided as a reference for what the blocks F_n are defined to be.

Let $F_0 = 1$ and $F_1 = 0$. Define $F_n = F_{n-1}F_{n-2}$. The F_n's are referred to as Fibonacci blocks. The limit as n goes to infinity is known as the infinite Fibonacci word or more simply the Fibonacci word.

Below is a list of the first eleven Fibonacci blocks.

F_1	1
F_2	0
F_3	01
F_4	010
F_5	01001
F_6	01001010
F_7	0100101001001
F_8	010010100100101001010
F_9	0100101001001010010100100101001001
F_{10}	0100101001001010010100100101001001010010100100101001010
F_{11}	0100101001001010010100100101001001010010100100101001010010.. ..0101001001010010100100101001001

INDEX

$\mathcal{P}$-name, 18

Alphabet, 17

Champernowne's Number, 5

Complexity Function, 31

Copeland-Erdös Constant, 5

Cylinder Set, 17, 18

Dense Set of Natural Numbers, 6

Distribution, 10

Entropy

- Conditional, 26
- of finite partition, 26

Ergodic Theorem, 16

Expectation, 11

Finite Measure Space, 7

Fixed Point, 29

Forbidden Words, 30

Generating Partition, 18

Golden Mean Shift, 30

Indicator Function, 9

Language, 31

Length, of a Word, 29

Measurable Function, 8

Measure Preserving Transformation, 15

Normal Number, 3, 4

- Absolutely, 4
- Simply, 4

Periodic Sequence, 29

Probability Measure Space, 7

Random Variable, 8

- Bernoulli, 9
- Discrete, 9
- Independent, 9

Shift Map, 18, 29

Shift Spaces, 28

Subshift, 30

Subshift of Finite Type, 30

Symbolic Dynamical System, 17

Word

- Bi-Infinite, 28

Empty, 28

Finite, 28

Infinite, 28