

# ERGODICITY AND ENTROPY IN SEQUENCE SPACES

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Applied Mathematics.

By

Christopher Miglino

Director: Dr. Julia Barnes  
Professor of Mathematics  
Department of Mathematics and Computer Science

Committee Members:  
Dr. Risto Atanasov  
Assistant Professor of Mathematics  
Department of Mathematics and Computer Science

Dr. Geoff Goehle  
Assistant Professor of Mathematics  
Department of Mathematics and Computer Science

April 1, 2013

## ACKNOWLEDGMENTS

I would like to thank my thesis adviser, Dr. Julia Barnes for her tireless commitment to helping me achieve my goal. Her advice and guidance were invaluable. I would also like to thank my committee members, Dr. Geoff Goehle and Dr. Risto Atanasov for their time and attention. Finally, I would like to thank my wife, Mrs. Sabrina Miglino for her production of the game board images found in Appendix A.

## TABLE OF CONTENTS

ABSTRACT . . . . .	iv
1 INTRODUCTION . . . . .	5
2 MEASURE THEORY . . . . .	7
2.1 Measurable Spaces . . . . .	7
2.2 Sequence Spaces and Cylinder Sets . . . . .	11
2.3 Shift Transformations . . . . .	14
3 ERGODICITY . . . . .	16
3.1 Ergodic Theory . . . . .	16
3.2 Bernoulli Shift . . . . .	21
3.3 Markov Shift . . . . .	23
4 ENTROPY . . . . .	29
4.1 Partitions . . . . .	29
4.2 Calculating Entropy . . . . .	30
4.3 Entropy of Bernoulli and Markov Shifts . . . . .	34
5 NEW GAMES . . . . .	39
5.1 Two Towers . . . . .	39
5.2 Sink Hole . . . . .	40
5.3 Medieval Game of Life . . . . .	43
6 CONCLUSIONS . . . . .	47
A GAME BOARDS . . . . .	48
A.1 Two Towers . . . . .	48
A.2 Sink Hole . . . . .	49
A.3 Sink Hole 2 . . . . .	49
A.4 Medieval Game of Life . . . . .	50
A.4.1 2 Loops . . . . .	50
A.4.2 3 & 4 Loops . . . . .	51
B TRANSITION MATRICES . . . . .	52
B.1 Monopoly Die-Roll-Only Matrix . . . . .	52
B.2 Monopoly Transition Matrix . . . . .	53
B.3 Two Towers Transition Matrix . . . . .	54
B.4 Sink Hole Transition Matrix . . . . .	55
B.5 Sink Hole 2 Transition Matrix . . . . .	56
B.6 Medieval Game of Life Transition Matrix (2 loops) . . . . .	57
B.7 Medieval Game of Life Transition Matrix (3 loops) . . . . .	58
B.8 Medieval Game of Life Transition Matrix (4 loops) . . . . .	59
C MATLAB CODE . . . . .	60
C.1 MATLAB Code for calculating Entropy of a Markov Shift . . . . .	60
C.2 MATLAB Code for finding a steady state vector in a Markov Chain. . . . .	61

## ABSTRACT

## ERGODICITY AND ENTROPY IN SEQUENCE SPACES

Christopher Miglino

Western Carolina University (April 2013)

Director : Dr. Julia Barnes

The infinite permutations of possible moves in a game, or positions on a game board, form a one-sided sequence space. We are working with a probability measure on the space of measurable subsets of the sequence space. We are studying a shift transformation on this space, which is measure preserving. We explore conditions under which the shift transformation is ergodic and calculate the entropy of the shift that is associated with the steady state of the game where applicable. These concepts are exemplified by the games Rock, Paper, Scissors and Monopoly. We then create new games and study how the properties of ergodicity and entropy change with respect to different aspects of the games.

## 1 INTRODUCTION

Ergodic Theory was first introduced by Ludwig Boltzmann via the field of statistical mechanics. It is from the theorems of George David Birkhoff and John von Neumann, in the 1930's, that the mathematics of ergodic theory has arisen [12]. In the eighty years following the establishment of these theorems, applications of ergodic theory to a variety of dynamical systems have abounded [2].

In this thesis we analyze various games using ergodic theory on the sequence space of all possible moves that one player can make. There are two main types of games that we will be studying. One type is the old children's game Rock, Paper, Scissors, and a newer version of the game with two more choices, Rock, Paper, Scissors, Lizard, Spock. The second are standard board games where a token is moved around a path by rolling dice or spinning a dial. A classic example of a board game of this nature is Monopoly which has already been analyzed in the literature [9, 10].

Our objective is to examine the conditions under which a shift transformation on the sequence space generated by a game is not ergodic and how changes to the game will effect the entropy of the shift. New games that we have developed in order to explore the effect of various configurations of a game board are introduced in Section 5. The new games are Two Towers, Sink Hole, Sink Hole 2, and three versions of the Medieval Game of Life. Each game is described with images of the game boards shown in Appendix A. We have analyzed the games that we have created and discuss their ergodicity and entropy. To reach our goal, we survey some necessary background.

In Section 2 of this paper we introduce measurable spaces and the  $\sigma$ -algebra. We show how cylinder sets generate a  $\sigma$ -algebra and how a natural probability on these sets can be extended to a probability measure for the entire space. Many of our measures are Markov; an explanation of Markov Chains is given. We then define a

shift transformation on the sequence space which move one space along an infinite sequence. The two shift transformations under study are the Bernoulli and Markov Shifts [6]. The mode of action in both of these transformations is the same.

Section 3 introduces the concept of ergodicity with several theorems stated and proven. Many are known results. However, the proofs have been modified to sequence spaces and, in many instances, significantly expanded with the intent of increasing the clarity of the arguments. Building on these theorems, we state and prove theorems regarding the ergodicity of Bernoulli and Markov Shifts. Entropy is introduced in Section 4 beginning with the entropy of a partition and culminating with the entropy of a shift transformation. In particular, the  $(\mathbf{p}, P)$  Markov shift is discussed as it relates to the steady state of a Markov Chain. The game of Monopoly is used as an example of a Markov system, as described by Ian Stewart [9, 10].

## 2 MEASURE THEORY

Any study of ergodic theory is done on a *measurable space*,  $(X, \mathcal{B}, m)$ , which consists of a set,  $X$ , a collection of subsets,  $\mathcal{B}$  satisfying particular properties, and a measure,  $m$ . A measure is a function from the space to the positive real numbers with properties we often think of when we measure the length of a real object with a ruler. For example, the measure of the empty set is zero, and if we break a set up into disjoint pieces, the measure of the set is the sum of the measure of its parts. The collection of subsets,  $\mathcal{B}$ , is precisely the collection of sets which can be measured.

Measure Theory is an abstraction of a concept with which most people are already familiar: how to measure something. We often think of measures in terms of length, area, and volume. While these concepts are certainly present in measure theory, there are also many other examples of measures. In order to define a measure, we must begin with a *measurable space*.

### 2.1 Measurable Spaces

We call the pair  $(\Omega, \mathcal{B})$  a measurable space, where  $\Omega$  is a set and  $\mathcal{B}$  is a collection of subsets called a  *$\sigma$ -algebra*. A  $\sigma$ -algebra is a collection of subsets,  $\mathcal{B}$ , that satisfy the following:

1.  $\Omega \in \mathcal{B}$ ;
2. If  $B \in \mathcal{B}$ , then  $B^C \in \mathcal{B}$  where  $B^C$  is the complement of  $B$ ;
3. If  $B_n \in \mathcal{B}$  for  $n \geq 1$ , then  $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$ .

It is not always possible to explicitly define a  $\sigma$ -algebra. Frequently, we start with a collection of subsets and generate a  $\sigma$ -algebra by beginning with the sets in which

we are interested and including sets necessary to meet the above conditions. Some collections of subsets that may be considered en route to a  $\sigma$ -algebra are:

A collection of subsets,  $\mathcal{L}$ , is a *semi-algebra* if

1.  $\emptyset \in \mathcal{L}$ ;
2. If  $A, B \in \mathcal{L}$ , then  $A \cap B \in \mathcal{L}$ ;
3. If  $A \in \mathcal{L}$ , then  $A^C = \bigcup_{i=1}^n E_i$  where each  $E_i \in \mathcal{L}$  and  $E_1, \dots, E_n$  are pairwise disjoint.

Note that a collection of sets  $\{E_1, \dots, E_n\}$  is *pairwise disjoint* if for  $i \neq j$ ,  $E_i \cap E_j = \emptyset$ .

A collection of subsets,  $\mathcal{A}$ , is an *algebra* if

1.  $\emptyset \in \mathcal{A}$ ;
2. If  $A, B \in \mathcal{A}$ , then  $A \cap B \in \mathcal{A}$ ;
3. If  $A \in \mathcal{A}$ , then  $A^C \in \mathcal{A}$ .

Every algebra is a semi-algebra, and every  $\sigma$ -algebra is an algebra [11]. In order to generate a  $\sigma$ -algebra from an algebra, we need the following lemma.

**Lemma 2.1.** *If  $\mathcal{A}$  is an algebra on  $\Omega$ , the following statements are equivalent.*

- i. If  $A, B \in \mathcal{A}$ , then  $A \cap B \in \mathcal{A}$ ;*
- ii. If  $A_1, A_2, \dots, A_n \in \mathcal{A}$ , then  $\bigcup_{i=1}^n A_i \in \mathcal{A}$ .*

*Proof.* Let  $\mathcal{A}$  be an algebra on  $\Omega$ .

*i.  $\Rightarrow$  ii.* Assume for any subsets  $A, B \in \mathcal{A}$ ,  $A \cap B \in \mathcal{A}$ . Let  $A_1, A_2, \dots, A_n$  be



an arbitrary collection of elements of  $\mathcal{A}$ . Since  $\mathcal{A}$  is an algebra, for each  $A_i \in \mathcal{A}$ ,  $A_i^C \in \mathcal{A}$ . We will show by induction that  $\bigcap_{i=1}^n A_i^C \in \mathcal{A}$  and use this fact to show closure under finite unions.

As a base case we have, by hypothesis,  $A_1^C \cap A_2^C \in \mathcal{A}$ . Then assume that for  $k \in \mathbb{N}$ ,  $\bigcap_{i=1}^k A_i^C \in \mathcal{A}$ . Then,

$$\bigcap_{i=1}^{k+1} A_i^C = \left( \bigcap_{i=1}^k A_i^C \right) \cap A_{k+1}^C.$$

Since  $\bigcap_{i=1}^k A_i^C \in \mathcal{A}$ , and  $\mathcal{A}$  is an algebra,  $\bigcap_{i=1}^{k+1} A_i^C \in \mathcal{A}$ . Hence  $\bigcap_{i=1}^n A_i^C \in \mathcal{A}$ . Since  $\mathcal{A}$  is closed under complementation, by DeMorgan's Law

$$\left( \bigcap_{i=1}^n A_i^C \right)^C = \bigcup_{i=1}^n (A_i^C)^C = \bigcup_{i=1}^n A_i \in \mathcal{A}.$$

Hence  $\mathcal{A}$  is closed under finite unions.

*ii.  $\Rightarrow$  i.* Assume that if  $\{A_1, \dots, A_n\} \in \mathcal{A}$ , then  $\bigcup_{i=1}^n A_i \in \mathcal{A}$ . Let  $A, B \in \mathcal{A}$ . Since  $\mathcal{A}$  is an algebra,  $A^C, B^C \in \mathcal{A}$ . By assumption,  $A^C \cup B^C \in \mathcal{A}$ . Then  $(A^C \cup B^C)^C \in \mathcal{A}$ . By DeMorgan's Law,

$$(A^C \cup B^C)^C = (A^C)^C \cap (B^C)^C = A \cap B.$$

Indeed,  $A \cap B \in \mathcal{A}$ .

□

Since the inclusion of set compliments and intersections in an algebra guarantees the inclusion of finite unions, by allowing countable unions, we can define a  $\sigma$ -algebra on  $\Omega$ . This distinction is necessary to the construction of a measurable space.

A measure is a function from some subset of a measurable space,  $(\Omega, \mathcal{B})$ , to the non-negative real numbers that satisfies conditions 2 and 3 below. A *probability*

measure on  $(\Omega, \mathcal{B})$  is a function  $m : \mathcal{B} \rightarrow [0, 1]$ , where

1.  $m(\Omega) = 1$ ;
2.  $m(\emptyset) = 0$ ;
3.  $m(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} m(B_i)$  where  $\{B_i\}_{i=1}^{\infty}$  is a sequence of pairwise disjoint members of  $\mathcal{B}$ .

A *probability space* is a triple  $(\Omega, \mathcal{B}, m)$ , where  $(\Omega, \mathcal{B})$  is a measurable space and  $m$  is a probability measure on  $(\Omega, \mathcal{B})$ . All measures referred to in this paper are probability measures. The following theorems from Resnick [8] will be used.

**Theorem 2.2.** *Let  $\mathcal{L}$  be a semi-algebra of subsets of  $\Omega$ . The algebra,  $\mathcal{A}(\mathcal{L})$ , generated by  $\mathcal{L}$  consists precisely of those subsets of  $\Omega$  that can be written in the form  $E = \bigcup_{i=1}^n A_i$  where each  $A_i \in \mathcal{L}$  and  $A_1, A_2, \dots, A_n$  are disjoint subsets of  $\Omega$ .*

If  $\mathcal{L}$  is a semi-algebra of subsets of  $\Omega$ , a function  $\mu : \mathcal{L} \rightarrow \mathbb{R}^+$  is *finitely additive* if  $\mu(\emptyset) = 0$  and  $\mu(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n \mu(E_i)$  where  $E_1, \dots, E_n \in \mathcal{L}$  are pairwise disjoint subsets of  $\Omega$ . The measure  $\mu$  is *countably additive* if  $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$  [11].

**Theorem 2.3.** *If  $\mathcal{L}$  is a semi-algebra of subsets of  $\Omega$  and  $\mu : \mathcal{L} \rightarrow \mathbb{R}^+$  is finitely additive, then there is a unique finitely additive function  $m : \mathcal{A}(\mathcal{L}) \rightarrow \mathbb{R}^+$  which is an extension of  $\mu$ . That is,  $m$  restricted to  $\mathcal{L}$  is equal to  $\mu$ . If  $\mu$  is countably additive, then so is  $m$ .*

In our study, we will focus on specific probability spaces and accompanying probability measures as defined in Section 2.2.

## 2.2 Sequence Spaces and Cylinder Sets

The space in which we are working is composed of infinite sequences. A point in the space is a *one-sided sequence*,  $\{a_i\}_{i=0}^{\infty}$ , formed from a finite set of distinct elements,  $\{a_0, a_1, \dots, a_n\}$ , called an alphabet. A *sequence space*,  $X$ , is the set consisting of all possible sequences of the elements of an alphabet.

The sequences with which we are working are the set of all possible combinations of the various states of certain games. One such game is the classic “Rock, Paper, Scissors” (RPS). Should the reader be unfamiliar with RPS, a comprehensive explanation is available at [1]. With each turn, either Rock ( $R$ ), Paper ( $P$ ), or Scissors ( $S$ ) occurs for each player with a probability of  $\frac{1}{3}$ . Imagine playing an infinite game of RPS and recording all of the moves that one player makes. For example, one sequence that could be formed is  $x = \{R, R, R, R, \dots\}$ ; another is  $y = \{R, P, R, P, \dots\}$ . Observe that both of these sequences are elements of the set  $A_R = \{\{a_i\}_{i=0}^{\infty} : a_0 = R\}$ , the set of all sequences that begin with  $R$ ; i.e.,  $x, y \in A_R$ . Since the probability of observing  $R$  on the first move is  $\frac{1}{3}$ , we can intuit that one-third of all sequences begin with  $R$ . If we define  $A_{RR} = \{\{a_i\}_{i=0}^{\infty} : a_0 = a_1 = R\}$ , we have the set of all sequences that begin with  $R, R$ . Notice that  $x \in A_{RR}$ , but  $y \notin A_{RR}$ .

With each turn,  $R$ ,  $P$ , and  $S$  occur independently. Recall one of the rules of elementary probability: the probability of independent events occurring successively is the product of their respective probabilities. Since  $P[a_0 = R] = \frac{1}{3}$  and  $P[a_1 = R] = \frac{1}{3}$ , we say that the probability of a random sequence belonging to  $A_{RR}$  is  $\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$ . Another tenant of basic probability is that the probability of observing either of two (or more) separate, independent events is the sum of their individual probabilities. For instance, the probability of observing either  $R$  or  $P$  on the second move is  $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$ . However, if all positions of a sequence are fixed, we would calculate the probability of

its occurrence as  $\lim_{n \rightarrow \infty} \left(\frac{1}{3}\right)^n = 0$ . In other words, the probability of any sequence occurring in which all positions are fixed is 0.

In order to talk about the measure of sets, we must first establish what sets in the sequence space,  $X$ , are measurable and how to assign a measure to them. The sets formed by fixing one or more consecutive positions of a sequence and considering all sequences that satisfy that condition are called *cylinder sets*. These sets will be used to generate a  $\sigma$ -algebra on  $X$ . To generate the  $\sigma$ -algebra, we begin by letting  $k$  be the number of possible states of a game. In the game of RPS, the states of the game refer to playing  $R$ ,  $P$ , or  $S$  at any given turn. In this instance,  $k = 3$ . We may consider  $Y = \{0, 1, \dots, k - 1\}$  to be an alphabet where  $i$  corresponds to the  $i^{\text{th}}$  state of a game. Let  $2^Y$  be the power set of  $Y$ , and let  $\mu$  be a measure defined by a probability vector,  $(p_0, p_1, \dots, p_{k-1})$ , where  $\mu(\{i\}) = p_i$ . A cylinder set  $A$  is the collection of all sequences where  $s$  consecutive positions are fixed elements of  $Y$ . The set  $A$  can be written as  $\{\{x_i\}_{i=0}^{\infty} : x_i = a_j \in Y \text{ for } r \leq j \leq r + s\}$ . We can denote  $A$  by  ${}_r[a_r, a_{r+1}, \dots, a_{r+s}]_{r+s}$  and call it a *block* beginning with the  $r^{\text{th}}$  position and ending with the  $(r + s)^{\text{th}}$  position.

Let  $\mathcal{L}$  be the collection of finite intersections of blocks of fixed positions. A member,  $A$ , of  $\mathcal{L}$  has the form  $A = \bigcap_{i=1}^n {}_{r_i}[a_{r_i}, \dots, a_{r_i+s_i}]_{r_i+s_i}$ . If we define an impossible set of fixed positions, such as those with  $a_1$  and  $a_2$  in the initial position with  $a_1 \neq a_2$ , then  $\emptyset \in \mathcal{L}$ . For any two members  $A$  and  $B$  of  $\mathcal{L}$ ,  $A \cap B$  is an intersection of a finite number of blocks of fixed positions. Hence,  $A \cap B \in \mathcal{L}$ . The complement of a block  ${}_r[a_r, \dots, a_{r+s}]_{r+s}$  is the finite union of  ${}_r[b_r, \dots, b_{r+s}]_{r+s}$  where  $b_i \neq a_i$  for at least one  $i$ . It follows that the complement of a member  $A = \bigcap_{i=1}^n {}_{r_i}[a_{r_i}, \dots, a_{r_i+s_i}]_{r_i+s_i}$  of  $\mathcal{L}$  is the union complements of each blocks in  $A$ . Therefore,  $\mathcal{L}$  satisfies the requirements of being a semi-algebra.

We consider the space  $X$  of all sequences  $x \in X$  where  $x = \{x_i\}_{i=0}^{\infty}$  and each  $x_i \in Y$ .

Since  $X$  is the collection of all sequences, we can denote  $X$  as

$$X = \prod_{i=0}^{\infty} Y = \{\{x_i\}_{i=0}^{\infty} : x_i \in Y\}.$$

More generally, we can also look at sets of the form

$$\prod_{i=0}^n A_j \times \prod_{j=n+1}^{\infty} Y \text{ where } A_j \in 2^Y.$$

That is, the elements in the  $j^{\text{th}}$  position are members of a set in the power set of  $Y$ , rather than single elements of  $Y$ . This is an example of something called a *product  $\sigma$ -algebra*,  $\mathcal{B}$ , on  $X$ . Then  $(X, \mathcal{B})$  is a measurable space denoted by

$$(X, \mathcal{B}) = \prod_{i=0}^{\infty} (Y, 2^Y).$$

The natural measure defined on the blocks is  $\mu(r[a_r, \dots, a_{r+s}]_{r+s}) = \prod_{i=r}^{r+s} p_i$ . The measure  $\mu$  on  $\mathcal{L}$  is the product of the measures of the blocks, i.e.

$$\mu \left( \bigcap_{i=1}^n r_i[a_{r_i}, \dots, a_{r_i+s_i}]_{r_i+s_i} \right) = \prod_{i=1}^n \mu(r_i[a_{r_i}, \dots, a_{r_i+s_i}]_{r_i+s_i}).$$

By Theorem 2.3,  $\mu$  can be extended to a measure  $m$  on  $\mathcal{B}$ . The measure  $m$  is called the  $(p_0, \dots, p_{k-1})$ -product measure [8].

As stated previously, it is the independence of the individual events that allow for the multiplication of their respective probability measures. We now cite a theorem from Walters [11] that establishes the existence of a unique probability measure on our sequence spaces.

**Theorem 2.4.** Fix  $k \geq 1$  and let  $Y = \{0, 1, \dots, k-1\}$  and  $(X, \mathcal{B}) = \prod_{i=0}^{\infty} (Y, 2^Y)$ . For each  $n \in \mathbb{N}$  and  $a_0, \dots, a_n \in Y$  suppose a non-negative real number  $p_n(a_0, \dots, a_n)$  is given so that

1.  $\sum_{a_0 \in Y} p_0(a_0) = 1$ , and
2.  $p_n(a_0, \dots, a_n) = \sum_{a_{n+1} \in Y} p_{n+1}(a_0, \dots, a_n, a_{n+1})$ .

Then there is a unique probability measure  $m$  on  $(X, \mathcal{B})$  with  $m([_h[a_h, \dots, a_l]_l]) = p_{l-h}(a_h, \dots, a_l)$  for all  $h \leq l$  and all  $a_i \in Y$ ,  $h \leq i \leq l$ .

The probability space  $(X, \mathcal{B}, m)$  is called the *direct product* of  $(Y, 2^Y, \mu)$ . It is precisely because each element in our sequences is an independent event that we can form the space using these products.

We now have that the set of all sequences formed from some finite set of possible states generates a probability space. Returning to the example at the beginning of this section, we may now revisit the sets,  $A_R$  (all sequences of RPS that begin with  $R$ ) and  $A_{RR}$  (all those that begin with  $R, R$ ). We can now measure these sets and confirm our earlier intuition. As expected,  $m(A_R) = \frac{1}{3}$  and  $m(A_{RR}) = \frac{1}{9}$ .

### 2.3 Shift Transformations

A shift transformation is a type of iterative process on a sequence or set of sequences. Let  $T : X \rightarrow X$  be a map such that for  $\{x_i\}_{i=0}^{\infty} \in X$ ,  $T(\{x_i\}_{i=0}^{\infty}) = \{x_{i+1}\}_{i=0}^{\infty}$ . The transformation,  $T$ , acts by “shifting” the sequence by one position. For instance, consider  $A_{RR} = \{\{a_i\}_{i=0}^{\infty} : a_0 = a_1 = R\}$ . If we apply the shift transformation  $T$  to the set  $A_{RR}$  the leading term “drops off” and we are left with the set  $A_R = \{\{a_i\}_{i=0}^{\infty} : a_0 = R\}$ . In other words, the second term has now shifted to become the initial term.

Consider the sets  $A_{PR} = \{\{a_i\}_{i=0}^{\infty} : a_0 = P, a_1 = R\}$ ,  $A_{SR} = \{\{a_i\}_{i=0}^{\infty} : a_0 = S, a_1 = R\}$ , and  $A_{RR} = \{\{a_i\}_{i=0}^{\infty} : a_0 = a_1 = R\}$ . By applying  $T$ , we shift the sequences in these sets; we have  $T(A_{PR}) = A_R$ ,  $T(A_{SR}) = A_R$ , and  $T(A_{RR}) = A_R$ . Under  $T$ ,  $A_{PR}$ ,  $A_{SR}$ , and  $A_{RR}$  all map into the set of sequences that start with

$R$ . Thus,  $A_{PR} \cup A_{SR} \cup A_{RR} \subseteq T^{-1}(A_R)$ . Since the first position  $x_0$  must be  $R$ ,  $P$ , or  $S$ ,  $T^{-1}(A_R) \subseteq A_{PR} \cup A_{SR} \cup A_{RR}$ . Then  $T^{-1}(A_R) = A_{PR} \cup A_{SR} \cup A_{RR}$ . Since  $X$  is a probability space and  $A_{PR} \cap A_{SR} \cap A_{RR} = \emptyset$ ,  $m(A_{PR} \cup A_{SR} \cup A_{RR}) = m(A_{PR}) + m(A_{SR}) + m(A_{RR})$ .

From Section 2.2 we know that  $m(A_{RR}) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$ . Similarly,  $m(A_{PR}) = m(A_{SR}) = \frac{1}{9}$ . Since  $m(A_{PR}) + m(A_{SR}) + m(A_{RR}) = \frac{1}{3}$ , we have that  $m(T^{-1}(A_R)) = \frac{1}{3}$ . Since  $m(A_R) = \frac{1}{3}$ , we see that the measure of the set,  $A_R$ , is preserved under the transformation  $T^{-1}$ .

**Definition.** A shift transformation  $T$  on a probability space  $(X, \mathcal{B}, m)$  is called *measure-preserving* if for  $B \in \mathcal{B}$ ,  $m(T^{-1}(B)) = m(B)$  [11].

All transformations in this study are of of this form and are measure-preserving. It is worth noting that  $m(T(A_{RR})) = m(A_R) = \frac{1}{3} \neq \frac{1}{9} = m(A_{RR})$ . Therefore, even though  $T$  is a measure-preserving transformation, measure is not preserved in the forward direction.

## 3 ERGODICITY

A model that describes the changes in a system as time progresses is called a *dynamical system*. Ergodic theory concerns the long-term behavior of dynamical systems. In our study, we are concerned with the ergodicity of the shift transformations described in Section 2.3. This leads us to the following definition.

**Definition.** A measure preserving transformation  $T$  of a probability space  $(\Omega, \mathcal{B}, m)$  is called *ergodic* if the only members  $B$  of  $\mathcal{B}$  with  $T^{-1}(B) = B$  satisfy  $m(B) = 0$  or  $m(B) = 1$  [11].

A set theoretic operation used in this study that deserves explanation is the *symmetric difference*. The symmetric difference of sets  $A$  and  $B$ , denoted  $A \triangle B$ , is exactly the portion of  $A$  that is not in  $B$  and of  $B$  that is not in  $A$ . Symbolically, we could write  $A \triangle B = (A \setminus B) \cup (B \setminus A)$ . If the measures of sets are equal up to sets of measure zero, we say that the measures are equal *almost everywhere* and use the notation, *a.e.*; that is,  $m(A) = m(B)$  if and only if  $m(A \triangle B) = 0$ . Also, function  $\chi : X \rightarrow \{0, 1\}$  is called the characteristic function. For  $B \in \mathcal{B}$ ,  $\chi_B$  is defined as follows.

$$\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

## 3.1 Ergodic Theory

Two main theorems that are central to ergodic theory are the Pointwise Ergodic Theorem by George Birkhoff, and another by Jon Von Neumann. We are concerned with the first, which is stated here without proof.

**Theorem 3.1** ([11];1.14). *Let  $(\Omega, \mathcal{B}, m)$  be a probability space and  $T : \Omega \rightarrow \Omega$  be a*



measure-preserving transformation. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \int_{\Omega} f dm \text{ a.e.}$$

The Pointwise Ergodic Theorem gives us that the long term average of a sequence of functions converges to some function, and that function is integrable. In this paper, the only function we will be integrating is the characteristic function. Integrating the characteristic function of a set with respect to its measure gives the measure of the set,  $\int_X \chi_B(x) dm(x) = m(B)$  [7]. In some of the following proofs we use the fact that the measure of any set in a probability space is bounded above by 1 and the convergence given by the Pointwise Ergodic Theorem to apply the following theorem.

**Theorem 3.2** ([11]; Dominated Convergence). *Let  $\{f_n\}$  be a sequence of real valued functions on  $(X, \mathcal{B}, m)$ . If  $\exists M \in \mathbb{R}^+$  such that  $\forall n \geq 1, |f_n| \leq M$  and  $\lim_{n \rightarrow \infty} f_n = f$  a.e. then  $f$  is integrable and  $\lim \int f_n dm = \int f dm$ .*

The following corollary to the Pointwise Ergodic Theorem is used frequently.

**Corollary 3.3** ([11];1.14.2). *Let  $(X, \mathcal{B}, m)$  be the probability space of sequences from Section 2.2 and let  $T : X \rightarrow X$  be the shift transformation from Section 2.3. Then  $T$  is ergodic if and only if  $\forall A, B \in \mathcal{B}$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(T^{-i}(A) \cap B) = m(A)m(B).$$

*Proof.* (of Corollary 3.3)

Let  $T$  be ergodic. Let  $f = \chi_A$ , the characteristic function on the set  $A$ . By Theorem 3.1, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) = \int_{x \in X} \chi_A(x) dm(x) = m(A).$$

Note that  $\int_{x \in X} \chi_A(x) dm(x) = m(A)$  because the integral of the characteristic function of a set gives the measure of the set. Multiplying through by  $\chi_B(x)$ , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) \chi_B(x) = m(A) \chi_B(x) \text{ a.e.}$$

Notice that  $|\chi_l(x)| \leq 1$  for any  $l$ , the average  $\frac{1}{n} \sum_{i=0}^{n-1} \chi_l(x) \leq 1$  for any  $n$ . Furthermore by the Pointwise Ergodic Theorem  $\frac{1}{n} \sum_{i=0}^{n-1} \chi_l(x)$  is equal to the measure of the set  $l$  almost everywhere. By the Dominated Convergence Theorem, the integral of the limit of these averages as  $n$  tends toward infinity is equal to the limit of the average of the integral of the characteristic function. Then

$$\begin{aligned} \int \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) \chi_B(x) dm(x) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \int \chi_A(T^i(x)) \chi_B(x) dm(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(T^{-i}(A) \cap B) = m(A)m(B). \end{aligned}$$

Conversely, assume the convergence property holds. Let  $T^{-1}(E) = E$  for  $E \in \mathcal{B}$ . Then  $T^{-1}(E) \cap E = E$ . Let  $A = B = E$  in the above convergence equations. Then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(T^{-i}(E) \cap E) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(E) = m(E)^2.$$

Moreover, notice that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(E) = \frac{n \cdot m(E)}{n} = m(E).$$

Then we have that  $m(E) = m(E)^2$ . Hence,  $m(E) = 0$  or  $m(E) = 1$ . By definition,  $T$  is ergodic.  $\square$

Commutativity of symmetric difference is used in the following proofs. Since the

symmetric difference between sets is essentially a union of their set differences, commutativity clearly holds.

These equivalent statements are used in proving the ergodicity of Bernoulli and Markov shifts later in this paper.

**Theorem 3.4** ([11];1.5). *If  $T : X \rightarrow X$  is a shift transformation on a probability space  $(X, \mathcal{B}, m)$ , then the following are equivalent:*

- i.  $T$  is ergodic.*
- ii. The only members  $B$  of  $\mathcal{B}$  with  $m(T^{-1}(B) \Delta B) = 0$  are those with  $m(B) = 0$  or  $m(B) = 1$ .*
- iii. For every  $A \in \mathcal{B}$  with  $m(A) > 0$  we have  $m(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$ .*
- iv. For every  $A, B \in \mathcal{B}$  with  $m(A) > 0$  and  $m(B) > 0$ , there exists an  $n > 0$  with  $m(T^{-n}(A) \cap B) > 0$ .*

*Proof.* Let  $(X, \mathcal{B}, m)$  be a probability space, and  $T : X \rightarrow X$  a shift transformation on  $(X, \mathcal{B}, m)$ .

*i.  $\Rightarrow$  ii.* Let  $T$  be ergodic. Let  $B \in \mathcal{B}$  such that  $m(B \Delta T^{-1}(B)) = 0$ . Since

$$T^{-n}(B) \Delta B \subset \bigcup_{i=0}^{n-1} T^{-(i+1)}(B) \Delta T^{-i}(B) = \bigcup_{i=0}^{n-1} T^{-i}(T^{-1}(B) \Delta B)$$

and  $T$  is a measure preserving transformation, then  $m(T^{-n}(B) \Delta B) \leq \sum_{i=0}^{n-1} m(T^{-i}(B) \Delta B) = nm(T^{-1}(B) \Delta B) = 0$ . Hence,  $\forall n \geq 0$ ,  $m(B \Delta T^{-n}(B)) = 0$ . It follows that

$$m(B \Delta \bigcup_{i=n}^{\infty} T^{-i}(B)) \leq \sum_{i=n}^{\infty} m(B \Delta T^{-1}(B)) = 0.$$

Let  $B_\infty = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-i}(B)$ . Note that as  $n$  increases,  $\bigcup_{i=n}^{\infty} T^{-i}(B)$  decreases. Also, since  $T$  is a measure preserving transformation,  $m(T^{-i}(B)) = m(B)$ ,  $\forall i \geq n$ .

By assumption, we have that  $m(B \Delta T^{-n}(B)) = 0$ . It follows that  $m(B \Delta B_\infty) = 0$ . Hence,  $m(B_\infty) = m(B)$ . Further note that

$$T^{-1}(B)_\infty = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-(i+1)}(B) = \bigcap_{n=0}^{\infty} \bigcup_{i=n+1}^{\infty} T^{-i}(B) = B_\infty.$$

Since  $T$  is ergodic,  $m(B_\infty) = 0$  or  $m(B_\infty) = 1$ . Then  $m(B) = 0$  or  $m(B) = 1$ .

*ii.  $\Rightarrow$  iii.* Let  $A \in \mathcal{B}$  such that  $m(A) > 0$ . Let  $A_1 = \bigcup_{n=1}^{\infty} T^{-n}(A)$ . Then  $T^{-1}(A)_1 \subset A_1$ . Since  $T$  is measure preserving,  $m(T^{-1}(A)_1) = m(A_1)$ . So,  $m(T^{-1}(A)_1 \Delta A_1) = 0$ . By ii., we have that  $m(A_1) = 0$  or  $m(A_1) = 1$ . Since  $m(A) = m(T^{-1}(A)) > 0$  and  $T^{-1}(A) \subset A_1$ ,  $m(A_1) \neq 0$ . Therefore  $m(A_1) = 1$ .

*iii.  $\Rightarrow$  iv.* Let  $A, B \in \mathcal{B}$  such that  $m(A) > 0$  and  $m(B) > 0$ . By iii., we have that  $m(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$ . So  $m(B)$  can be expressed as

$$0 < m(B) = m\left(B \cap \bigcup_{n=1}^{\infty} T^{-n}(A)\right) = m\left(\bigcup_{n=1}^{\infty} B \cap T^{-n}(A)\right).$$

Therefore,  $m(B \cap T^{-n}(A)) > 0$  for some  $n \geq 1$ .

*iv.  $\Rightarrow$  i.* Suppose that  $T^{-1}(B) = B$  and  $0 < m(B) < 1$ . Notice that  $m(B \cap B^C) = m(T^{-1}(B) \cap B^C) = 0$  for all  $n \geq 1$ . Since this is absurd,  $m(B) = 0$  or  $m(B) = 1$ .

□

### 3.2 Bernoulli Shift

A Bernoulli Shift is a shift transformation on a probability space, where  $\mathbf{p} = (p_1, \dots, p_k)$  is a probability vector; RPS is an example of a Bernoulli Shift. It is also called the  $(p_0, \dots, p_k)$ -shift. In the games which we are studying using the Bernoulli Shift, each event is equally likely to occur. In this case,  $p_i = \frac{1}{k}$  for  $1 \leq i \leq k$ . Each term of the sequences in our Bernoulli cylinder sets has equal measure. Flipping a coin forever would generate a Bernoulli sequence where each point would have  $p_i = \frac{1}{2}$ . A shift transformation on the space of all sequences consisting of “heads” or “tails” is an example of a  $(\frac{1}{2}, \frac{1}{2})$ -shift. The following theorem states that any Bernoulli Shift is ergodic.

**Theorem 3.5** ([11];1.12). *The  $(p_0, \dots, p_{k-1})$ -shift is ergodic.*

*Proof.* Let  $(X, \mathcal{B}, m)$  be a probability space and  $\mathcal{A}$  be an algebra of finite unions of cylinder sets. Let  $E \in \mathcal{B}$  and let  $T^{-1}(E) = E$ . Let  $\epsilon > 0$ . Let  $A \in \mathcal{A}$  such that  $m(E \Delta A) < \epsilon$ . Since  $(X, \mathcal{B}, m)$  is a probability space,

$$\begin{aligned} |m(E) - m(A)| &= |(m(E \cap A) + m(E \setminus A)) - (m(A \cap E) + m(A \setminus E))| \\ &= |m(E \cap A) + m(E \setminus A) - m(A \cap E) - m(A \setminus E)| \\ &= |m(E \setminus A) - m(A \cap E) + m(E \cap A) - m(A \setminus E)| \\ &= |m(E \setminus A) - m(A \setminus E)| \leq m(E \setminus A) + m(A \setminus E) = m(E \Delta A) < \epsilon. \end{aligned}$$

Choose  $n_0$  large enough that  $B = T^{-n_0}(A)$  depends on different coordinates from  $A$ . That is,  $B$  is disjoint from  $T^{-n_0}(A)$  and hence from  $E$ . If  $A$  is defined by a fixing a block  ${}_k[a_k, \dots, a_{k+l}]_l$ , then  $n_0 > l - k$ . Because  $m$  is a product measure and  $T$  is

measure preserving, we have  $m(B \cap A) = m(A)m(B) = m(A)^2$ . Then

$$m(E \Delta B) = m(T^{-n}(E) \Delta T^{-n}(A)) = m(E \Delta A) < \epsilon.$$

Notice that

$$\begin{aligned} E \Delta (A \cap B) &= (E \setminus (A \cap B)) \cup ((A \cap B) \setminus E) \\ &= ((E \setminus A) \cup (E \setminus B)) \cup ((A \setminus E) \cap (B \setminus E)) \\ &\subseteq (E \setminus A) \cup (E \setminus B) \cup (A \setminus E) \cup (B \setminus E) \\ &= (E \setminus A) \cup (A \setminus E) \cup (E \setminus B) \cup (B \setminus E) \\ &= (E \Delta A) \cup (E \Delta B). \end{aligned}$$

Then  $m(E \Delta (A \cap B)) < 2\epsilon$ . Hence  $|m(E) - m(A \cap B)| < 2\epsilon$ , and

$$\begin{aligned} |m(E) - m(E)^2| &\leq |m(E) - m(A \cap B)| + |m(A \cap B) - m(E)^2| \\ &\leq 2\epsilon + |m(A)^2 - m(E)^2| \\ &\leq 2\epsilon + m(A)|m(A) - m(E)| + m(E)|m(A) - m(E)| \\ &\leq 4\epsilon. \end{aligned}$$

Since  $\epsilon$  is arbitrary,  $m(E) = m(E)^2$ . Hence  $m(E) = 0$  or  $m(E) = 1$ . □

For the sake of brevity, we use the following definitions.

**Definition.** When the shift transformation associated with a particular game is found to be ergodic, we shall say “*the game is ergodic*”.

By Theorem 3.5, RPS is ergodic. Moreover, since a shift transformation on any sequence space of independent events is a Bernoulli shift, any alterations to RPS

will result in an ergodic game. For instance, consider the popular expansion of RPS that includes the terms “Lizard” and “Spock”, as played on the television program, “The Big Bang Theory”. Rock, Paper Scissors, Lizard, Spock is an example of a  $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ -shift and is clearly ergodic because it is a Bernoulli shift. Notice that the property of being ergodic does not depend on the measure of any  $p_i \in \mathbf{p}$ , only that  $\mathbf{p}$  is a probability vector. The effect of altering the values of the probabilities will be explored in Section 4.3.

### 3.3 Markov Shift

A sequence of random events, the state of which depends only upon the preceding event, is referred to as a Markov Chain. An example of a Markov Chain is the possible position of a player’s token on a board game. The space to which the token is moved depends on the space from which it is being moved. Due to its level of complexity, and the extensive studies that have been performed previously [9], the first board game used in this thesis to facilitate the discussion of Markov Chains is Monopoly. Because the position of a game token at a certain time depends on the previous position of the token, the probability of the token being at a particular location is conditional. We say that the probability is Markov.

Markov probability measures are calculated using a probability vector  $\mathbf{p}$  and a transition matrix  $P$ . The  $p_{ij}$  entry of  $P$  is the probability of moving from the  $i^{th}$  to the  $j^{th}$  state. For instance, consider a game token on the third position of a Monopoly board. Regardless of how the token arrived at that position, the probability that it will move to the fifth position is  $\frac{1}{36}$ . This is because the player would need to roll a two in order to move from the third to the fifth position. Because the probability of rolling a two on a pair of six sided dice is  $\frac{1}{36}$ ,  $p_{35} = \frac{1}{36}$ . A Markov Chain is formed by iterating  $\mathbf{p}^{(i-1)}P = \mathbf{p}^{(i)}$ . The result of left multiplication of  $P$  by  $\mathbf{p}$  is another

probability vector  $\mathbf{p}^{(1)}$ . The  $i^{\text{th}}$  entry of  $\mathbf{p}^{(1)}$  is the probability of being in the  $i^{\text{th}}$  state after the first iteration. After  $n$  iterations  $\mathbf{p}^{(n-1)}P = \mathbf{p}^{(n)}$  gives the probability of being in the  $i^{\text{th}}$  state as  $p_{i_n}$ .

According to Ian Stewart [10], the probabilities of being on any of the 40 positions on the Monopoly board approaches a steady state. This means that over time the probability of being on any position approaches a constant. If we assume that a player can move around the Monopoly board only as a result of rolling dice (and not by drawing a card or going to jail),  $P$  is the matrix in Appendix B.1. Note that for comparison, the transition matrix for the game including movements induced by the Chance and Community Chest cards is given in Appendix B.2. On the  $(\mathbf{p}, P)$  Markov Chain, where  $P$  is the transition matrix in Appendix B.1, all probabilities approach  $\frac{1}{40}$ . The steady state of the game when considering player movements resulting from the cards is found by calling the `steadyState` function in Appendix C.2 with the transition matrix in Appendix B.2. The function returns a similar probability vector, but with slightly more density around the positions to which a player may move as a result of the cards.

In the sequence space generated by all possible itineraries of the game when played forever, the set of all sequences that have the token being on Boardwalk at a fixed position in the sequence has measure equal to the long term probability of being on that spot. The set of all sequences that have Boardwalk followed by Atlantic Avenue is the intersection of the collection of sequences where the  $i^{\text{th}}$  position is Boardwalk and the  $i + 1^{\text{st}}$  position is Atlantic Avenue. Since we are using the measure defined on cylinder sets, the measure of such a set is the product of the probabilities of the token being in each position.

The existence of the limit of the long term average of a transition matrix is shown by the following lemma.



**Lemma 3.6** ([11];1.18). *Let  $P$  be a transition matrix and  $\mathbf{p}$  a strictly positive probability vector such that  $\mathbf{p}P = \mathbf{p}$ . Then  $Q = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} P^n$  exists.  $Q$  is a transition matrix and  $QP = PQ = Q$ . Moreover, any eigenvector of  $P$  corresponding to the eigenvalue 1 is also an eigenvector of  $Q$ .*

*Proof.*

Let  $m$  denote the  $(\mathbf{p}, P)$  Markov measure and  $T$  the  $(\mathbf{p}, P)$  Markov shift. Let  $\chi_i$  be the characteristic function of the cylinder set  ${}_0[i]_0 = \{(x_j)_{j=0}^{\infty} | x_0 = i\}$ . By Theorem 3.1 (Pointwise Ergodic Theorem),

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_j(T^k(x)) = \int_X \chi_j(x) dm(x) \text{ a.e.}$$

Multiplying by  $\chi_i(x)$  and integrating, we have

$$\int \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_j(T^k(x)) \chi_i(x) dm(x) = \int \chi_j(x) \chi_i(x) dm(x).$$

As stated previously, since each characteristic function is bounded above by 1, the average of the characteristic functions is also bounded above by 1. By the Pointwise Ergodic Theorem the limit of the average of the characteristic function is equal to the measure of the set on which the characteristic function is defined. Then we can apply the Dominated Convergence Theorem.

$$\begin{aligned} & \int \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_j(T^k(x)) \chi_i(x) dm(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \int \chi_j(T^k(x)) \chi_i(x) dm(x) \\ & = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} m({}_k[j]_k \cap {}_0[i]_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_i p_{ij}^k = \int \chi_i(x) \chi_j(x) dm(x) = p_i q_{ij}. \end{aligned}$$

Therefore,  $q_{ij} = \frac{1}{p_i} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^k = \frac{1}{p_i} \int \chi_i(x) \chi_j(x) dm(x)$  is the  $ij$ -entry of  $Q$ .

Multiplying  $Q$  on the right or left by  $P$  only changes the indexes of the summations;  $Q$  is unchanged and  $QP = PQ = Q$ . Let  $v$  be a left eigenvector of  $P$  corresponding to the eigenvalue 1. Then  $vP = 1v$ . Since  $Q = PQ$ ,  $vQ = vPQ = 1vQ = v$ . Hence  $v$  is a left eigenvector of  $Q$  corresponding to the eigenvalue 1.

The product of transition matrices is a transition matrix [3]. By the associativity of matrix multiplication we can easily see that for a transition matrix  $P$ ,  $P^n$  is also a transition matrix. Since  $P^n$  is a transition matrix, the rows of  $P^n$  sum to one for all  $n$ . Consider adding  $N$  transition matrices. The rows of the resultant matrix will sum to  $N$ . Multiplying this matrix by  $\frac{1}{N}$  results in a matrix whose rows sum to one. It follows that the rows of  $Q$  sum to one. Then the matrix  $Q$  is a transition matrix.  $\square$

We now show that a Markov shift is ergodic.

**Theorem 3.7** ([6];1.19). *Let  $T$  denote the  $(p, P)$  Markov Shift where can assume  $p_i > 0$  for each  $i$  and  $\mathbf{p} = (p_0, \dots, p_{k-1})$ . Let  $Q$  denote the matrix obtained in Lemma 3.6. The following statements are equivalent.*

- i.  $T$  is ergodic.*
- ii. All rows of the matrix  $Q$  are identical.*
- iii. 1 is a simple eigenvalue of  $P$ ; that is, 1 occurs with single multiplicity.*

*Proof.*

*i.  $\Rightarrow$  ii.* Let  $T$  be an ergodic measure preserving transformation. From the proof of Lemma 3.6 we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} m({}_k[j]_k \cap {}_0[i]_0) = m({}_0[i]_0)m({}_n[j]_n) = p_i q_{ij}.$$

Since  $m({}_0[i]_0) = p_i$  and  $m({}_n[j]_n) = p_j$ , we have that  $p_i q_{ij} = p_i p_j$ . Since  $p_i > 0$ ,  $q_{ij} = p_j$ . Hence each row of  $Q$  is identical.

*ii.*  $\Rightarrow$  *i.* Since the rows of  $Q$  are identical,  $q_{ij}$  depends only on  $i$ . Since  $QP = PQ = Q$ , it follows that  $q_{ij} = p_j$ . Let  $A_1, A_2$  be any two blocks of fixed letters in the  $\sigma$ -algebra.  $A_1$  and  $A_2$  have the form  $A_1 = {}_r[i_0, \dots, i_l]_{r+l}$ ,  $A_2 = {}_s[j_0, \dots, j_m]_{s+m}$ . For  $n > s+m-r$ ,

$$m(T^{-n}(A_1) \cap A_2) = p_{j_0} p_{j_0 j_1} \cdots p_{j_{m-1} j_m} p_{j_m i_0}^{(r+n-s-m)} p_{i_0 i_1} \cdots p_{i_{l-1} i_l}.$$

Since  $q_{ij} = p_j$ ,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} m(T^{-n}(A_1) \cap A_2) = p_{j_0} p_{j_0 j_1} \cdots p_{j_{m-1} j_m} p_{i_0} p_{i_0 i_1} \cdots p_{i_{l-1} i_l} = m(A_1) m(A_2).$$

By Corollary 3.3, we have that  $T$  is ergodic.

*ii.*  $\Rightarrow$  *iii.* Since the rows of  $Q$  are identical and  $QP = PQ = Q$ ,  $q_{ij} = p_j$ . Then, the only left eigenvectors of  $Q$  corresponding to the eigenvalue 1 are multiples of  $\mathbf{p}$ . This means that the eigenspace corresponding to the eigenvalue 1 has a dimension of 1. Since the dimension of the eigenspace is less than or equal to the multiplicity of its corresponding eigenvalue, 1 is a simple eigenvalue of  $Q$ . By Lemma 3.6, these are also the only left eigenvectors of  $P$  corresponding to the eigenvalue 1. It follows that 1 is a simple eigenvalue of  $P$ .

*iii.*  $\Rightarrow$  *ii.* Suppose 1 is a simple eigenvalue of  $P$ . Then the eigenspace corresponding to the eigenvalue 1 has a dimension of 1. If  $\mathbf{p}$  is a left eigenvector of  $P$  corresponding to the eigenvalue 1, then any other eigenvector is a multiple of  $\mathbf{p}$ . Since  $\mathbf{p}$  is a probability vector, the only valid multiple of  $\mathbf{p}$  is  $\mathbf{p}$  itself. Since  $QP = Q$ , each row of  $Q$  is a left eigenvector of  $P$ . Hence, each row of  $Q$  is identical.  $\square$

In the context of our study, a shift transformation,  $T$ , on the sequence space,  $X$ , of

all possible states of a board game is ergodic if a set,  $B$ , is equal to its pre-image,  $T^{-1}(B) = B$ , then  $B$  is either the entire game or some sets of measure zero. It is for this reason that, when played for a long enough time, a player will eventually land on every space of the board. In the long term, the probability that a player will be at a certain position on the board approaches a steady state [5]. The matrix,  $Q$ , described in Lemma 3.6 gives us these probabilities. Notice that  $Q$  is a transition matrix that represents the long term averages of the state transitions given by  $P$ . Since the long term probabilities approach a steady state and are Markov, the probability of moving to any position on the board becomes, over time, the same regardless of the previous position. This is why all rows of  $Q$  are identical.

We have shown in the proof of Theorem 3.7 that the ergodicity of  $T$  and identicalness the the rows of  $Q$  are necessary and sufficient conditions. We then showed that this property of the matrix  $Q$  is necessary and sufficient to guarantee the existence of 1 as a simple eigenvalue of  $P$ . While the ergodicity of  $T$  and the behavior of the long term average of  $P$  can only be shown algorithmically to within a set tolerance, it is a simple matter to find the eigenvalues of  $P$ . As we discuss the new games that we have created, we will look at the eigenvalues of their transition matrices to determine properties of each game.

Notice that ergodicity of a game does not depend on the values of the entries in  $P$ , but the behavior of the long term average of  $P$ . Though they have different transition matrices, the two representations of the game of Monopoly given in Appendix B, with and without the Chance and Community Chest cards, are both ergodic. What is important is that the game contains no backward invariant sub-loops. Monopoly clearly contains no invariant sub-loops whatsoever. In Section 5 we will introduce new games that do have sub-loops. Some of these are measure 0 or 1, while others have measures between 0 and 1. The former are ergodic games; the latter are not.

## 4 ENTROPY

Entropy describes the uncertainty that is present in a system. Our systems are sequence spaces with a shift transformation. We will begin by discussing partitions and then use them to define entropy, a measure of the uncertainty in a system.

## 4.1 Partitions

A *finite partition*  $\xi$  on a probability space on  $(X, \mathcal{B}, m)$  is a collection of disjoint subsets of  $\mathcal{B}$ ,  $\{A_1, \dots, A_k\}$  whose union is  $X$ . If  $\xi$  is a finite partition on  $(X, \mathcal{B}, m)$ , the collection of all elements of  $\mathcal{B}$  whose unions are elements of  $\xi$  is called a *finite sub- $\sigma$ -algebra* of  $\mathcal{B}$ , denoted  $\mathcal{A}(\xi)$ . Conversely, if a collection  $\mathcal{C} = \{C_i : i = 1, \dots, n\}$  is a finite sub- $\sigma$ -algebra of  $\mathcal{B}$ , then the non-empty sets of the form  $B_1 \cap \dots \cap B_n$  where  $B_i = C_i$  or  $X \setminus C_i$  form a partition,  $\eta$ , on  $(X, \mathcal{B}, m)$  denoted by  $\xi(\mathcal{C})$ . Then  $\mathcal{A}(\xi(\mathcal{C})) = \mathcal{C}$  and  $\xi(\mathcal{A}(\eta)) = \eta$ . Hence, there is a one-to-one correspondence between finite partitions and sub- $\sigma$ -algebras [11]. A  $\sigma$ -algebra,  $\mathcal{B}$ , may be viewed as the collection all possible events and combinations of events that can occur in a given probability space. A partition on  $\mathcal{B}$  is a set of events where one and only one event can happen at a time [4]. Such a collection is a sub- $\sigma$ -algebra of  $\mathcal{B}$ .

If  $\mathcal{C}$  and  $\mathcal{D}$  are sub- $\sigma$ -algebras of  $\mathcal{B}$  such that for every  $C \in \mathcal{C}$  there exists a  $D \in \mathcal{D}$ , and for every  $D \in \mathcal{D}$  there exists a  $C \in \mathcal{C}$  with  $m(C \triangle D) = 0$ , we write  $\mathcal{C} \doteq \mathcal{D}$ .

Let  $\xi = \{A_1, \dots, A_n\}$  and  $\eta = \{C_1, \dots, C_k\}$  be finite partitions of  $(X, \mathcal{B}, m)$ . The *join* of the partitions is

$$\xi \vee \eta = \{A_i \cap C_j : 1 \leq i \leq n, 1 \leq j \leq k\}.$$

Moreover, if  $\mathcal{A}$  and  $\mathcal{C}$  are finite sub- $\sigma$ -algebras of  $\mathcal{B}$ , then  $\mathcal{A} \vee \mathcal{C}$  is the smallest

sub- $\sigma$ -algebra of  $\mathcal{B}$  that contains both  $\mathcal{A}$  and  $\mathcal{C}$ . By the definition of a join,  $\mathcal{A} \vee \mathcal{C}$  consists of all unions of the form  $A \cap C$  where  $A \in \mathcal{A}$  and  $C \in \mathcal{C}$ .

## 4.2 Calculating Entropy

In order to define the term *entropy*, we will look at probability from an empirical perspective. Consider an experiment performed repeatedly. The probability of an outcome is a measure of the certainty of observing that outcome over repeated trials. Each possible outcome can then be assigned a probability of occurring, and the probabilities of all possible outcomes must sum to 1. Entropy is a measure of the amount of uncertainty present in performing the same experiment. For example, consider the interval  $[0, 1]$ . If we partition  $[0, 1]$  into two intervals of equal length and randomly choose a number in  $[0, 1]$ , it is equally likely to be in the lower half as the upper half. Our level of uncertainty is high. However, if we partition the interval such that  $\frac{1}{100}$ <sup>th</sup> of the length is on the lower end, a partition of  $[0, 1]$  would be  $\{[0, \frac{1}{100}], (\frac{1}{100}, 1]\}$ . We are far more certain that a randomly chosen number will be in the upper portion. Our level of uncertainty is much lower. Entropy is a quantity that represents our level of uncertainty.

A function that measures the amount of uncertainty in performing some experiment should satisfy certain requirements. For instance, if we are certain that one and only one outcome will occur, we have no uncertainty; the entropy should be zero. Moreover, if we also consider events that are impossible to occur, our level of uncertainty as to those that can occur should be unchanged.

The following theorem by A.I. Khinchin provides a function for calculating the entropy of a partition [4].

**Theorem 4.1** ([4];The Uniqueness Theorem). *Let  $\mathbf{p}$  be a strictly positive probability*

vector and  $H : \mathbf{p} \rightarrow \mathbb{R}$  be a continuous function with respect to all its arguments such that the following hold.

1.  $H(p_1, \dots, p_k) \geq 0$  and  $H(p_1, \dots, p_k) = 0$  if and only if some  $p_i = 1$ .
2.  $H(p_1, \dots, p_k, 0) = H(p_1, \dots, p_k)$ .
3. For each  $k \geq 1$ ,  $H$  has its largest value at  $(1/k, \dots, 1/k)$ .

Then there exists a positive number  $\lambda$  such that  $H(p_1, \dots, p_k) = -\lambda \sum_{i=1}^k p_i \log p_i$ .

Kinchin states on page 10 of [4], “This theorem shows that the expression for the entropy of a finite (partition) is the only one possible if we want to have certain general properties which seem necessary in view of the actual meaning of the concept of entropy.”

The measures of the sets in a finite sub- $\sigma$ -algebra form a probability vector. Theorem 4.1 gives us that the entropy of a finite sub- $\sigma$ -algebra  $\mathcal{A}$  of  $\mathcal{B}$  with  $\xi(\mathcal{A}) = \{A_1, \dots, A_k\}$  is

$$H(\mathcal{A}) = H(\xi(\mathcal{A})) = - \sum_{i=1}^k m(A_i) \log m(A_i).$$

The following theorem from Walters [11] gives us that  $H$  is strictly convex.

**Theorem 4.2** ([11];4.2). *The function  $\phi : [0, 1] \rightarrow \mathbb{R}$  defined by*

$$\phi(x) = \begin{cases} 0 & \text{if } x = 0 \\ x \log(x) & \text{if } x \neq 0 \end{cases}$$

*is strictly convex. That is, if  $\alpha \in [0, 1]$  and  $x, y \in [0, \infty)$ ,  $\phi(\alpha x + (1 - \alpha)y) \leq \alpha\phi(x) + (1 - \alpha)\phi(y)$ . We have equality only when  $x = y$  or  $\alpha = 0$ .*

*Proof.* To confirm the convexity of  $\phi$ , we use elementary calculus.

$$\begin{aligned}\phi'(x) &= 1 + \log(x), \text{ and} \\ \phi''(x) &= \frac{1}{|x|}.\end{aligned}$$

Fix  $\alpha \in [0, 1]$ . Let  $x, y \in [0, \infty)$ . By the Mean Value Theorem,

$\phi(y) - \phi(\alpha x + (1 - \alpha)y) = \phi'(z)\alpha(y - x)$  for some  $z$  between  $\alpha x + (1 - \alpha)y$  and  $y$ , and

$\phi(\alpha x + (1 - \alpha)y) - \phi(x) = \phi'(w)(1 - \alpha)(y - x)$  for some  $w$  between  $x$  and  $\alpha x + (1 - \alpha)y$ .

Since  $\phi''(x) > 0$ , we have  $\phi'(z) > \phi'(w)$  because  $z > w$ . Hence,

$$\begin{aligned}(1 - \alpha)(\phi(y) - \phi(\alpha x + (1 - \alpha)y)) &= \phi'(z)\alpha(1 - \alpha)(y - x) > \phi'(w)\alpha(1 - \alpha)(y - x) \\ &= \alpha(\phi(x + (1 - \alpha)y) - \phi(x)).\end{aligned}$$

Therefore  $\phi(\alpha x + (1 - \alpha)y) < \alpha\phi(x) + (1 - \alpha)\phi(y)$  if  $x, y > 0$ . By the piecewise definition of the function, it also holds if  $x, y \geq 0$  where  $x \neq y$ .

□

The corollary shows that maximum entropy occurs in an equi-probable system; that is where  $\mathbf{p} = (p_1, \dots, p_k)$  and  $p_i = \frac{1}{k}$  for  $1 \leq i \leq k$ .

**Corollary 4.3** ([11];4.2.1). *If  $\xi = \{A_1, \dots, A_k\}$ , then  $H(\xi) \leq \log(k)$ , and  $H(\xi) = \log(k)$  only when  $m(A_i) = \frac{1}{k}$  for all  $i$ .*

*Proof.* Let  $\mathbf{p} = (p_1, \dots, p_k)$  be a probability vector. Then there exists non-negative real numbers  $a_1, \dots, a_j$  where  $\sum_{i=1}^j \frac{a_i}{k} = 1$  such that  $p_1 = \frac{a_1}{k}, \dots, p_k = \frac{a_j}{k}$ . By Theorem 4.1, if  $a_j = 0$  then  $H(p_1, \dots, p_k) = H(p_1, \dots, p_{k-1})$ . Without loss of generality,



we can assume that  $a_i > 0$  for  $1 \leq i \leq j$ . Notice that

$$\begin{aligned} -\sum_{i=1}^j \frac{a_i}{k} \log \frac{a_i}{k} &= -\sum_{i=1}^j \frac{a_i}{k} (\log a_i - \log k) \\ &= \sum_{i=1}^j \left( \frac{a_i}{k} \log k - \frac{a_i}{k} \log a_i \right) \\ &= \log k \sum_{i=1}^j \frac{a_i}{k} - \sum_{i=1}^j \frac{a_i}{k} \log a_i. \end{aligned}$$

Since  $\sum_{i=1}^j \frac{a_i}{k} = 1$  and  $H$  is strictly convex, we have

$$-\sum_{i=1}^j \frac{a_i}{k} \log \frac{a_i}{k} = \log k - \sum_{i=1}^j \frac{a_i}{k} \log a_i \leq \log k.$$

Notice that equality occurs only when  $a_1 = \dots = a_j = 1$ , that is when  $j = k$  and  $p_i = \frac{1}{k}$ .  $\square$

**Definition.** Suppose  $T : X \rightarrow X$  is a measure-preserving transformation of the probability space  $(X, \mathcal{B}, m)$ . If  $\mathcal{A}$  is a finite sub- $\sigma$ -algebra of  $\mathcal{B}$  then

$$h(T, \xi(\mathcal{A})) = h(T, \mathcal{A}) = \lim_{n \rightarrow \infty} \frac{1}{n} H \left( \bigvee_{i=0}^{n-1} T^{-i} \mathcal{A} \right)$$

is called the *entropy* of  $T$  with respect to  $\mathcal{A}$ . Finally, the entropy of  $T$  is the supremum over all finite sub- $\sigma$ -algebras,  $\mathcal{A}$  of  $\mathcal{B}$ . We define  $h(T) = \sup h(T, \mathcal{A})$ . Since  $\xi(\mathcal{A}) = \mathcal{A}(\xi)$ , we have that  $h(T) = \sup h(T, \xi)$  where the supremum is taken over all finite partitions of  $(X, \mathcal{B}, m)$  [11].

Consider  $T^n$  to be the iteration of  $T$  over  $n$  increments of time. Then,  $h(T)$  represents the maximum average amount of information gained by performing  $T$  [6]. We now cite a theorem that allows us to equate the entropy of a measure preserving transformation  $T$  with the entropy of  $T$  with respect to  $\mathcal{A}$ . We state it here without proof; see [11]

for details.

**Theorem 4.4** ([11];4.18). *If  $T$  is a measure-preserving transformation (not necessarily invertible) of the probability space  $(X, \mathcal{B}, m)$  and if  $\mathcal{A}$  is a finite sub- $\sigma$ -algebra of  $\mathcal{B}$  with  $\bigvee_{i=0}^{\infty} T^{-i}\mathcal{A} \overset{\circ}{=} \mathcal{B}$ . Then  $h(T) = h(T, \mathcal{A})$ .*

### 4.3 Entropy of Bernoulli and Markov Shifts

We now show how to calculate the entropy of the measure-preserving transformations under study, the Bernoulli and Markov shifts. The Bernoulli shift is defined as the shift transformation,  $T$ , where  $\mathbf{p}$  is a probability vector. The following theorem defines a function to calculate the entropy for the shift.

**Theorem 4.5** ([11];4.26). *The  $(p_0, \dots, p_{k-1})$ -shift has entropy*

$$-\sum_{i=0}^{k-1} p_i \log p_i.$$

*Proof.* Let  $Y = \{0, 1, \dots, k-1\}$ ,  $X = \prod_0^{\infty} Y$ , and let  $T$  be the shift map. Let  $A_i = \{\{x_k\} : x_0 = i\}$ , for  $0 \leq i \leq k-1$ . Then  $\xi = \{A_0, \dots, A_{k-1}\}$  is a partition of  $X$ . Denote  $\mathcal{A}(\xi)$  by  $\mathcal{A}$ . Notice that for each  $n$ ,  $T^{-n}(\xi) = \{T^{-n}(A_0), \dots, T^{-n}(A_{k-1})\}$  is a partition with  $T^{-n}(A_i) = {}_n[i]_n$ . Then  $\bigvee_{i=r}^{r+s} T^{-n}(\xi)$  is the partition all blocks of the form  ${}_r[a_r, \dots, a_{r+s}]_{r+s}$ . Since these are the cylinder sets defined in Section 2 the  $\sigma$ -algebra generated by the join is the same as the  $\sigma$ -algebra generated by the cylinder sets. Hence,  $\mathcal{B} = \bigvee_{i=0}^{\infty} T^{-i}\mathcal{A}$ . By Theorem 4.4,

$$h(T) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{-(n-1)}\mathcal{A}).$$

A member of  $\xi(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{n-1}\mathcal{A})$  has the form

$$A_{i_0} \cap T^{-1}(A_{i_1}) \cap \dots \cap T^{-(n-1)}(A_{i_{n-1}}) = \{\{x_n\} : x_0 = i_0, x_1 = i_1, \dots, x_n = i_{n-1}\}$$

and has measure  $p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}$ . Notice that

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i} \mathcal{A}\right) &= -\sum (p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}) \log(p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}) \\ &= -\sum_{i_0, \dots, i_{n-1}=0}^{k-1} (p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}) [\log p_{i_0} + \cdots + \log p_{i_{n-1}}] \\ &= -n \sum_{i=0}^{k-1} p_i \log p_i. \end{aligned}$$

Then  $h(T) = h(T, \mathcal{A}) = -\sum_{i=0}^{k-1} p_i \log p_i$ . □

Calculating the entropy for a Bernoulli Shift is a simple matter when all probabilities are equal. The entropy is  $H(T) = \log k$ , where  $k$  is the number of states and each state has probability  $p_i = \frac{1}{k}$ . The entropy associated with Rock, Paper, Scissors, RPS, is equal to  $\log 3$ . Revisiting the example in Section 3.2 of an expanded RPS that includes the terms “Lizard” and “Spock”, we can see that this game has entropy equal to  $\log 5$ . The entropy is larger for the expanded version of the game because there are more moves that a player may make. Since these entropies are maximal, any changes we may impose on the probabilities will result in a decrease in entropy. We forgo further exploration of entropy in Bernoulli Shifts.

Since the shift transformation,  $T$ , is the same whether the measure,  $m$ , on the sequence space  $X$  is Markov or Bernoulli, the formula for finding the entropy of a Markov shift uses the same sequences space as the Bernoulli. The entropy of a Markov shift, is calculated in a similar way using the measure associated with Markov shifts. Note that the proof is similar to Theorem 4.5, though there are more computations because of the transition matrix.

**Theorem 4.6** ([11];4.27). *The  $(\mathbf{p}, P)$  Markov shift has entropy  $-\sum_{i,j} p_i p_{ij} \log p_{ij}$ .*

*Proof.* Let  $Y = \{0, 1, \dots, k-1\}$ , let  $X = \prod_0^\infty Y$ . Let  $T$  be the shift map. Let

$A_i = \{\{x_k\} : x_0 = i, 0 \leq i \leq k-1\}$ . Then  $\xi = \{A_0, \dots, A_{k-1}\}$  is a partition of  $X$ . Denote  $\mathcal{A}(\xi)$  by  $\mathcal{A}$ . Using the same notation as in the proof of Theorem 4.5,  $\mathcal{B} = \bigvee_{i=0}^{\infty} T^{-i}\mathcal{A}$ . We have that

$$\sum_{i=0}^{k-1} p_i p_{ij} = p_j.$$

A member of  $\xi(\mathcal{A} \vee T^{-1}\mathcal{A} \vee \dots \vee T^{n-1}\mathcal{A})$  has measure  $p_{i_0} p_{i_0 i_1} \dots p_{i_{n-2} i_{n-1}}$ . Since  $\sum_{i=0}^{k-1} p_i p_{ij} = p_j$  and  $\sum_{j=0}^{k-1} p_{ij} = 1$ ,

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right) &= - \sum_{i_0, \dots, i_{n-1}=0}^{k-1} (p_{i_0} p_{i_0 i_1} \dots p_{i_{n-2} i_{n-1}}) \log(p_{i_0} p_{i_0 i_1} \dots p_{i_{n-2} i_{n-1}}) \\ &= - \sum_{i_0, \dots, i_{n-1}=0}^{k-1} (p_{i_0} p_{i_0 i_1} \dots p_{i_{n-2} i_{n-1}}) \\ &\quad \cdot [\log p_{i_0} + \log p_{i_0 i_1} + \dots + \log p_{i_{n-2} i_{n-1}}] \\ &= \sum_{i=0}^{k-1} p_{i_0} \log p_{i_0} - (n-1) \sum_{i,j=0}^{k-1} p_i p_{ij} \log p_{ij}. \end{aligned}$$

By rearranging the terms in this sum, we obtain  $h(T) = - \sum_{i,j} p_i \cdot p_{ij} \log p_{ij}$ .  $\square$

When passed a probability vector  $\mathbf{p}$  and a transition matrix  $P$ , the `entropyMarkov` function in Appendix C.1 calculates the entropy of a shift transformation,  $T$ , as proscribed in Theorem 4.6.

We notice that in Theorem 3.7, it is necessary that the  $(\mathbf{p}, P)$  Markov system use a strictly positive probability vector,  $\mathbf{p}$ . By Theorem 4.4, the entropy of the shift transformation,  $T$ , is the limit supremum of the grand joins of partitions,  $\xi(\mathcal{A})$  being acted upon by  $T$ . Applying this result, we impose the condition that  $\mathbf{p}$  be the steady state vector of its Markov Chain. Since we are looking at the entropy of the long

term behavior of the system, the steady state probability vector is a logical choice. If a Markov Chain approaches a steady state, it will do so regardless of the starting conditions [5]. As all of our board games begin at the starting position, to find a steady state vector, we begin with a probability vector,  $\mathbf{p}$ , of length  $n$  with  $p_1 = 1$  and an  $n \times n$  transition matrix  $P$ . Using the `steadyState` function found in Appendix C.2, we generate a steady state probability vector,  $\mathbf{p}^*$ .

Notice that Theorem 4.6 imposes no restrictions on  $\mathbf{p}$ . For this reason, the entropy that is calculated is for one step, one “link” in the Markov Chain. We are concerned with the long term behavior of games, not just a single move. Therefore we can use the steady state vector,  $\mathbf{p}^*$ , and define the value returned by `entropyMarkov` to be the entropy of the game.

Consider the entropy for each of our representations of the game of Monopoly. The die-roll only version uses the transition matrix  $P$  given in Appendix B.1. The long term behavior of this game is that, over time, a player is equally likely to be in any position on the board. The `steadyState` function returns a vector of equal probabilities. With this  $\mathbf{p}^*$ , `entropyMarkov` returns a value of 2.270. This is the entropy of the game. The inclusion of Chance and Community Chest cards requires that  $P$  is the matrix given in Appendix B.2. Using the equal probability vector, `steadyState` returns a value of 2.307. While this vector may occur in the Markov Chain representation of the game, this is not the long term behavior of the game as a whole. Not only is this assertion supported by Ian Stewart in [10], but `steadyState` also confirms that the long term probabilities are affected by the inclusion of the cards; that is, the steady state vector of the  $(\mathbf{p}, P)$  Markov Chain when  $P$  is the transition matrix in Appendix B.2 is not the same as if  $P$  were the matrix in Appendix B.1. Using the steady state vector, `entropyMarkov` returns a value of 2.306, the true entropy for this rendition of the game. Since the steady state vectors were found to

within a tolerance of  $10^{-3}$ , the entropy associated with both version of Monopoly may be considered to be essentially equal. When studying new games, we will consider the entropy associated with the steady state vector.

**Definition.** The *entropy of a game* is the entropy of the shift transformation of the sequence space generated by the game that is calculated using the steady state vector of the Markov Chain for that game.

## 5 NEW GAMES

In Section 2.2 we established that the sequence space formed by playing a game forever and recording all of the moves that one player can make is a probability space,  $(X, \mathcal{B}, m)$ . We then defined a shift transformation,  $T : X \rightarrow X$ , on the sequence space; this was the focus of Section 2.3. Sections 3 and 4 defined ergodicity and entropy respectively, provided conditions under which  $T$  is ergodic, and showed how to find the entropy associated with  $T$ . We now turn our focus to new games that we have created with the purpose of exploring these topics. Where applicable, representations of each game are given in Appendix A and associated transition matrices are given in Appendix B. A special die called a *color die* is used in some of our games. Each face of the die is painted. Colors are evenly distributed across all of the faces; the number of colors is given in the name of the die. For instance, a two-color die consists of two colors. It has six sides, three sides per color. The die can have 6 or 8 sides to accommodate 2, 3, or 4 evenly distributed colors. A color die of any number is equally likely to show any of its colors.

### 5.1 Two Towers

The first new game that we created was Two Towers. The game board is shown in Appendix A.1. The object of the game is to conquer both towers. Players start in the lobby, the center. A two-color die determines whether the player moves into the right-hand loop or the left-hand loop. Two six-sided die are used to determine a player's movement around the board. Players continue around the same loop, consolidating their "tower of power" until landing on the starting position exactly. The two-color die is thrown to determine whether the player remains in the same loop, or moves into the opposite loop, at which point conquest of the other tower begins.

Let  $P$  be the Two Towers transition matrix shown in Appendix B.3. The state transition probabilities described above determined  $P$ . We find that 1 is a simple eigenvalue of  $P$ . By Theorem 3.7, a shift transformation  $T$  on the sequence space generated by Two Towers is ergodic. Hence, the game of Two Towers is ergodic.

Despite the fact that there are two separate loops in which a player may remain, it is not required that the player remain in one loop or the other. So, these loops are not backward invariant. A pre-image under  $T$  of a set of sequences in either of these loops is not equal to itself. This is because a player may move between loops. So, the only invariant sets are the whole game or those sequences where the player never moves between towers. The latter are sets of measure zero for the same reason that the probability of observing a sequence all of whose positions are fixed is zero.

Two Towers reaches a steady state. In the long term, players are equally likely to be in any position in either loop. Since the lobby is part of both towers, the long term probability that a player will be in the lobby is twice the probability of being at any other position. The `entropyMarkov` function returns a value of 2.330 for the entropy of Two Towers. Notice that Two Towers has slightly higher entropy than either rendition of Monopoly.

## 5.2 Sink Hole

The next game was designed with the specific intent of creating an invariant sub-loop. A simple design of the Sink Hole board is shown in Appendix A.2. Players begin at the starting position. The two-color die is thrown to determine whether the player will begin moving around the loop or immediately fall into the sink hole. If the player does not land in the sink hole, a six-sided die is thrown; the player moves accordingly. The player continues to move according to the roll of a six sided die until



landing again at the starting position. When the player next lands at the starting position, the two-color die is again thrown. The player will either continue around the board or fall into the “sink hole”. Once inside the sink hole, actual movement of the player’s game token ceases. However, the position of the token continues to be recorded forever. The positions in the sequence after the player has entered the sink hole will be constant.

Let  $P$  be the transition matrix for Sink Hole given in Appendix B.4. Since 1 is a simple eigenvalue of  $P$ , Sink Hole is an ergodic game. This was initially counter intuitive, since a forward invariant loop exists for a sequence landing in the sink hole. However, remaining in the sink hole is not backward invariant. If we consider the set of all sequences that enter the sink hole, we see immediately that it must have measure equal to that of the entire space. This is because all sequences except for a countable set of sequences land in the sink hole. The sequences that do not enter the sink hole are of zero measure. Then the backward invariant set of sequences that enter the sink hole has a measure of 1, while those that do not have measure 0.

The steady state vector for Sink Hole has most of the probability at the end of the vector. That is, the vector coordinate that corresponds to the sink hole. The `steadyState` function gives  $\mathbf{p}^*$  as a probability vector of length 11 where  $p_{11} = .998$ . This entry represents the probability that a player is in the sink hole. The remaining .002 is evenly distributed across the rest of the vector. As such, there should be very little uncertainty associated with this game. This assumption is confirmed when `entropyMarkov` returns an entropy value of 0.003. We are very certain that we are going to land in the sink hole! It is only because `steadyState` uses a set tolerance to find  $\mathbf{p}^*$  that there is any long term probability outside of the sink hole. By considering the limit to infinity, we expect that steady state of sink hole to be a vector of length 11 where the only non-zero entry is  $p_{11} = 1$ . In this case, by Theorem 4.1, the entropy

of Sink Hole is 0.

Since the existence of the sink hole resulted in the game being ergodic, we modified the game to include two sink holes to see if the new game would continue to be ergodic. The second version is called Sink Hole 2. The game board is shown in Appendix A.3. The game is played exactly the same way as the original version, but now there are two spaces where a player must consult the two-color die. On opposite sides of the game board, the player has two pitfalls to avoid. If the player lands on either space, the two-color die determines whether the player remains in the sink hole or not. There are now two invariant sets. These are the set of sequences that land in the first sink hole and the set that land in the second. Each of these has measure between 0 and 1.

Let  $P$  be the transition matrix in Appendix B.5. We notice that  $P$  has an eigenvalue of 1, but with double multiplicity; it is not a simple eigenvalue. By Theorem 3.7, Sink Hole 2 is not ergodic. Sink Hole 2 achieves a steady state vector. The `steadyState` function gives that the long term probability of being in the original sink hole is .722, while there is a long term probability of .267 of being in the new sink hole. The remaining probability density, .011, is distributed evenly across the rest of the vector. As was the case with the steady state vector for the original Sink Hole, the presence of any probability density outside of the sink hole positions is due to the set tolerance in the `steadyState` function. We notice that three quarters of the probability density is in one sink hole and one quarter in the other. This is because the first sink hole is at the starting position. At the beginning of the game, a player will move immediately into the sink hole with a probability of one-half. Assuming that play is allowed to continue, the probability of falling into either of the two sink holes is even. If the starting position was not associated with a sink hole, the long term probability of being in either hole would still be greater for the first hole that a player encounters.

Using  $\mathbf{p}^*$  as returned by `steadyState`, the `entropyMarkov` function returns a value of .002 for the entropy of Sink Hole 2. Since it is very unlikely that a player will not be in one of the sink holes there is very little uncertainty as to where a player will be. However, the steady state vector when taken to infinity would result in non-zero entropy. This is because eventually every sequence except those of measure zero will be in one of two spaces. Since the tolerance used in estimating the steady state vector was set at .0001, the difference in entropies between the two versions of Sink Hole is attributable to the tolerance level as well as the presence of another position where a player may be.

Both versions of Sink Hole are, admittedly trivial. They were designed solely with the purpose of exploring the requirements for ergodicity of a game. This begs the question, does the quality of being ergodic or not effect whether a game is fun to play? We leave the question as an exercise to the reader.

### 5.3 Medieval Game of Life

Noticing that the existence of two sink holes resulted in Sink Hole 2 being non-ergodic, we designed a game that was both non-trivial and non-ergodic. Our *pièces de résistance* is the Medieval Game of Life. The game board is shown in Appendix A.4. The Medieval Game of Life (MGOL) is an epic saga of a child born of a prince and a peasant woman. The player's childhood years are spent traversing the beginning path of the game. There are six positions that represent the player's childhood. A six-sided die thrown to determine the player's movement; players must land on the sixth space exactly. When the player reaches the sixth space of the board a decision is made. At this point in the player's life, the king will either reject the child, or accept him or her as a member of the royal family. Rolling the two-color die represents the decision of the king. If the player is accepted into the noble family, movement continues in

the loop to the left. If the player is rejected, movement is to the right. Players in the left-hand loop, or the royal family, have the responsibility of maintaining the rule of the monarchy and oppressing the peasantry. Players that were rejected by the king (despite their royal blood) are tasked with overthrowing the monarchy. The rebellion against the crown is born!

Let  $P$  be the transition matrix given in Appendix B.6. We notice that  $P$  has an eigenvalue of 1 with multiplicity of 2. By Theorem 3.7, MGOL is not an ergodic game. This is consistent with our results in Section 5.2. The loops in MGOL and the sink holes in Sink Hole 2 are both parts of the game out of which a player cannot move. The crucial difference is that the loops in MGOL are non-trivial. In fact, they are not pitfalls to be avoided. Rather, they dictate the course of the game.

The MGOL Markov Chain will approach a steady state. The `steadyState` function returns a vector,  $\mathbf{p}^*$ , that shows evenly distributed probability across the two loops,  $p_7$  to  $p_{16}$  and  $p_{17}$  to  $p_{26}$ . This is because the loops are of equal size and the player may enter either with equal probability. Since a player must land exactly on the decision spot to enter a loop,  $p_1$  to  $p_6$  have minuscule but non-zero probabilities that a player will enter neither loop. Using  $\mathbf{p}^*$ , `entropyMarkov` returns a value of 1.79, the entropy of MGOL.

We added another loop to the game to see how this effects ergodicity and entropy. The initial premise is the same, but upon reaching the decision spot a three-color die is thrown. The third option is that the player become a member of the clergy. Clerics will ally themselves with whichever faction they choose. Ultimately, since both oppression and rebellion cause great collateral damage to the people, the clergy is tasked with thwarting the efforts of both the royals and the rebels.

Let  $P$  be the transition matrix given in Appendix B.7. This matrix has an eigenvalue

of 1 with multiplicity of 3. Though not ergodic, we found it interesting that the multiplicity of the eigenvalue of 1 was equal to the number of loops. The transition matrix given in Appendix B.8 represents the same game with four loops. As expected, the eigenvalue 1 appears with multiplicity 4. The following theorem gives the multiplicity of 1 as an eigenvalue of a transition matrix for a game with  $k$  invariant loops.

**Theorem 5.1.** *The transition matrix,  $P$ , for a board game with  $k$  invariant loops has an eigenvalue of 1 with multiplicity  $k$ .*

*Proof.* Let  $P$  be the  $n \times n$  transition matrix associated with a game has  $k$  invariant loops and let  $P_1, \dots, P_k$  be sub-matrices of  $P$  corresponding to the loops. Renumbering the spaces of the board if necessary,  $P$  can be expressed as

$$P = \begin{pmatrix} P_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & P_k \end{pmatrix}.$$

Let  $\alpha_1, \dots, \alpha_k$  be the row dimensions of  $P_1, \dots, P_k$  respectively. The characteristic polynomial of  $P$  is

$$\det(P - \lambda I_n) = \det \begin{pmatrix} P_1 - \lambda I_{\alpha_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & P_k - \lambda I_{\alpha_k} \end{pmatrix} = \det(P_1 - \lambda I_{\alpha_1}) \cdots \det(P_k - \lambda I_{\alpha_k}).$$

Since each  $P_i$  represents a loop into which a player will move, there exists positive integers  $m_1, \dots, m_k$  such that  $P_i^{m_i}$  contains only positive entries. Then for each  $\det(P_i - \lambda I_{\alpha_i})$ , 1 occurs as an eigenvalue with a multiplicity of 1. Hence, 1 occurs as an eigenvalue of  $P$  with a multiplicity of  $k$ .  $\square$

Recall that the entropy of MGOL with two loops was found to be 1.79. The three

and four loop versions of the game also approach steady states. Using the respective steady state vectors for each matrix, `entropyMarkov` returns a value of 1.79 for both games. Since the probability of a player not entering one of the loops approaches zero, the only entropy that exists in the game occurs as a result of the loops. Equal amounts of uncertainty exist because the loops are of the same size.

## 6 CONCLUSIONS

We have shown the properties of ergodicity and entropy as they relate to some common games, and we have created new games that explore these properties. We have also shown how to calculate the entropy of a game. We found that the presence of loops in a game is not sufficient to guarantee non-ergodicity. If, however, the sets that these loops represent are backward invariant and have measure between 0 and 1, non-ergodicity is realized. We have provided a theorem that describes the behavior of the eigenvalue 1 for the transition matrices of games that contain backward invariant loops.

Finally, we noticed that the inclusion of additional loops of the same size will not change the entropy. This is because, with certainty, a player will be in one of the loops. The only uncertainty is where in the respective loop a player might be. Further research in this area may include a study of the entropy associated with games that contain invariant loops of different sizes or sink holes located inside of loops.

## A GAME BOARDS

## A.1 Two Towers

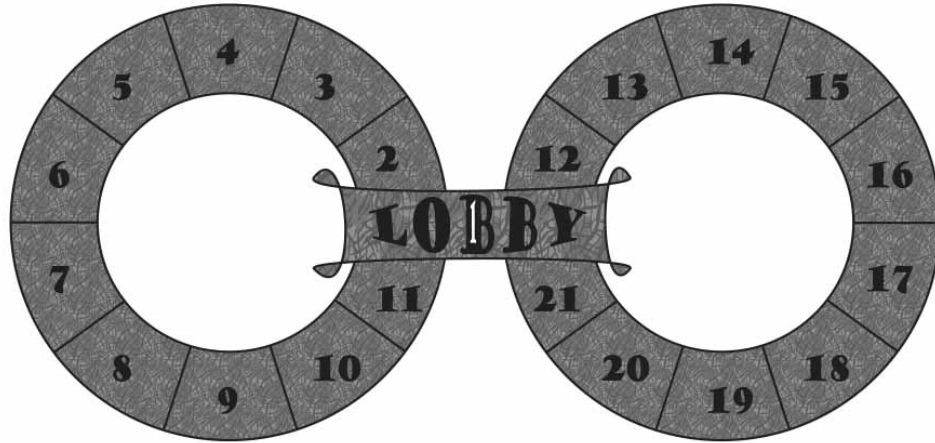


Figure A.1.1: Two Towers



## A.2 Sink Hole

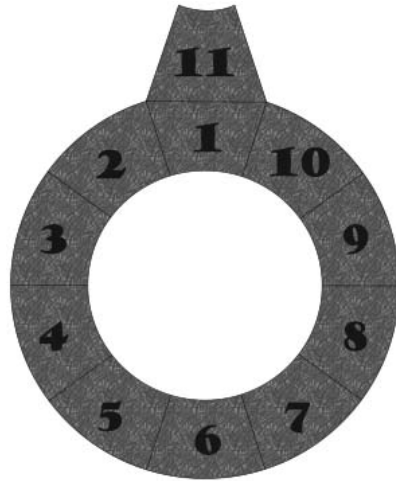


Figure A.2.1: Sink Hole

## A.3 Sink Hole 2

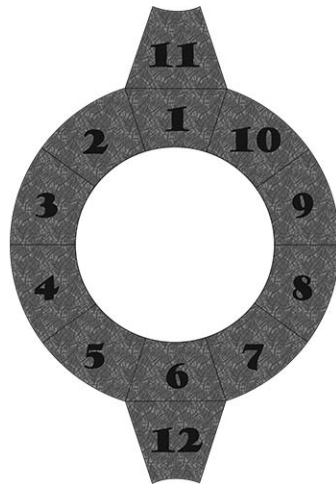


Figure A.3.1: Sink Hole 2

## A.4 Medieval Game of Life

## A.4.1 2 Loops

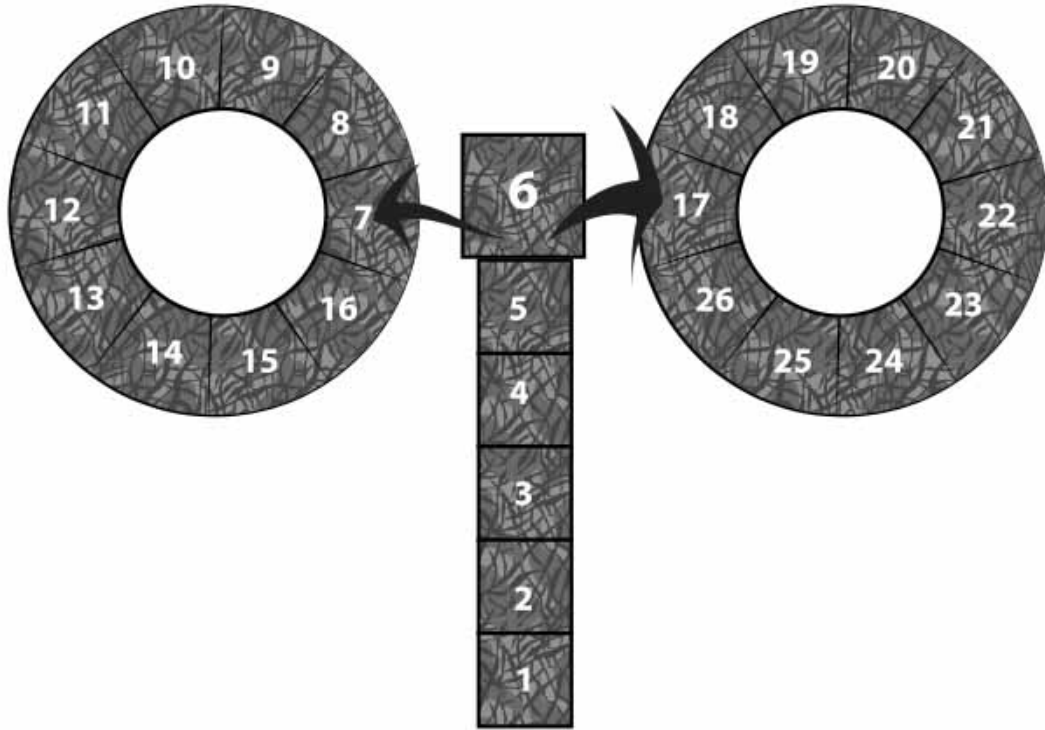


Figure A.4.1: The Medieval Game of Life

## A.4.2 3 &amp; 4 Loops

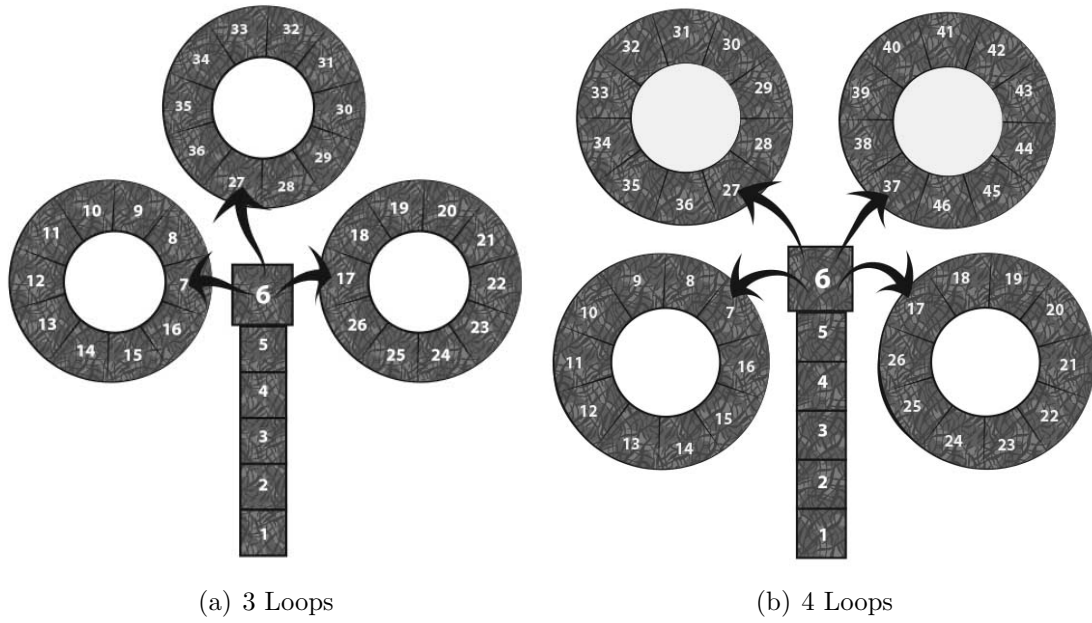


Figure A.4.2: Medieval Game of Life (3 &amp; 4 Loops)







## B.4 Sink Hole Transition Matrix

$$\begin{pmatrix}
 0 & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 & 0 & 0 & \frac{1}{2} \\
 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\
 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\
 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
 \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
 \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
 \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
 \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 \\
 \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 \\
 \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}$$

Sink Hole Transition Matrix

## B.5 Sink Hole 2 Transition Matrix

$$\begin{pmatrix} 0 & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ \frac{1}{12} & 0 & 0 & 0 & 0 & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & 0 & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Sink Hole 2 Transition Matrix









## C MATLAB CODE

## C.1 MATLAB Code for calculating Entropy of a Markov Shift

```

1 function [ent] = entropyMarkov(p,P)
2 %This function calculates the entropy of a Markov Shift.
3
4 [m,n] = size(P);
5 if m ~= n
6     fprintf('Size mismatch. Enter a square matrix. ')
7     return
8 end
9
10 ent = 0;
11 for i = 1:n
12     for j = 1:n
13         if p(i)==0 || P(i,j)==0
14             ent = ent + 0;
15         else
16             ent = ent + p(i)*P(i,j)*log(P(i,j));
17         end
18     end
19 end
20 ent = -ent;
21
22 end

```

## C.2 MATLAB Code for finding a steady state vector in a Markov Chain.

```
1 function [x] = steadyState(p,P,tol ,maxiter)
2 done = 0;
3 iter = 1;
4
5 while(~done && iter<maxiter)
6     pnew = p*P;
7     if norm(pnew-p)<tol
8         x = pnew;
9         break
10    else iter = iter +1;
11    end
12    p = pnew;
13 end
```

## REFERENCES

- [1] Erik Arneson. Board/card games. <http://http://boardgames.about.com/>. Last visited on 02/17/2013.
- [2] Julia Barnes and Lorelei Koss. The ergodic theory carnival. *Math Magazine*, 83:180–190, 2010.
- [3] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. *Linear Algebra*. Prentice Hall, 2002.
- [4] A.I. Khinchin. *Mathematical foundations of information theory*. Dover Publications, 1957.
- [5] J. R. Norris. *Markov Chains*. Cambridge University Press, 1997.
- [6] Karl Petersen. *Ergodic Theory*. Press Syndicate of University of Cambridge, 1983.
- [7] H. R. Pitt. *Measure, Integration, and Probability*. Dover Books, 2012.
- [8] Sidney I. Resnick. *A Probability Path*. Birkhauser Boston, 1999.
- [9] Ian Stewart. How fair is monopoly? *Scientific American*, 274:104, April 1996.
- [10] Ian Stewart. Monopoly revisited. *Scientific American*, 275:116, October 1996.
- [11] Peter Walters. *Introduction To Ergodic Theory*. Springer-Verlag, 1982.
- [12] Eric W. Weinstein. Ergodic theory From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/ErgodicTheory.html>. Last visited on 07/06/2012.