ABSTRACT

MARKOV CHAINS, RANDOM WALKS, AND CARD SHUFFLING

by

Nolan Outlaw

May 2015

Chair: Dr. Gail Ratcliff, PhD

Major Department: Mathematics

A common question in the study of random processes pertains to card shuffling. Whether or not a deck of cards is random can have huge implications on any game being played with those particular cards. This thesis explores the question of randomness by using techniques established through analysis of Markov chains, random walks, computer simulations, and some basic shuffling models. Ultimately, the aim is to explore the cutoff phenomenon, which asserts that at some point during the shuffling process there is a sharp decline in the shuffled deck's distance from random.

MARKOV CHAINS, RANDOM WALKS, AND CARD SHUFFLING

A Thesis

Presented to

The Faculty of the Department of Mathematics

East Carolina University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in Mathematics

by

Nolan Outlaw

May 2015

MARKOV CHAINS, RANDOM WALKS, AND CARD SHUFFLING

by

Nolan Outlaw

APPROVED BY:

DIRECTOR OF THESIS:

Dr. Gail Ratcliff, PhD

COMMITTEE MEMBER:

Dr. Jungmin Choi, PhD

COMMITTEE MEMBER:

Dr. Chris Jantzen, PhD

COMMITTEE MEMBER:

Dr. Heather Ries, PhD

CHAIR OF THE DEPARTMENT
OF MATHEMATICS:

Dr. Johannes Hattingh, PhD

DEAN OF THE
GRADUATE SCHOOL:

Dr. Paul Gemperline, PhD

ACKNOWLEDGEMENTS

I would like to thank Dr. Gail Ratcliff for her work as my advisor. Dr. Ratcliff has served an essential role in my education, both as a thesis advisor and as a professor, and her constant guidance, insight, and encouragement is greatly appreciated. I am grateful to Dr. Jungmin Choi, Dr. Chris Jantzen, and Dr. Heather Ries for agreeing to serve on this committee. I thank the entire staff and faculty of the Mathematics Department for the education I received throughout my time at East Carolina University. Finally, I would like to thank my family and friends for their support during my time working on this thesis.

TABLE OF CONTENTS

CHAPTER 1: Introduction

Why does it matter if a deck of cards is random? When playing a multi-person card game, someone could gain a significant advantage over the other players if that person noticed a lack of randomness in shuffled cards. Suppose that a person is playing a game in a casino - blackjack, for example. Again, if this player were to notice that they were playing with a non-random deck of cards, they could gain a meaningful advantage in the game. This time, however, there could be financial implications. A non-random deck could influence the player's betting, which may ultimately lead to larger wins for the player. Ensuring that a deck of cards is random promotes fairness in the game being played (or in the case of the casino, it ensures that the odds are in the casino's favor). Determining the appropriate number of shuffles needed to make a deck random can be done mathematically. Some of the methods used to determine the necessary number of shuffles to make a deck random are presented here.

The measurement that we use to determine if a deck of cards is random is total variation. For two probability measures μ and ν, total variation is defined by the equation

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{i=1}^{k} |\mu_i - \nu_i| .$$

What we observe is that there is a point at which a sharp decline occurs in the total variation between the random deck and the shuffled deck. The total variation remains fairly constant for several shuffles and then suddenly drops. This is called the cutoff phenomenon, and one of our goals is to replicate this occurrence through computer simulations. Figure 1.1 shows the appearance of the cutoff phenomenon at k_0.

The most common type of shuffle is called the riffle shuffle, and this is the shuffle that we ultimately seek to evaluate. For this shuffling process, a deck of n cards is

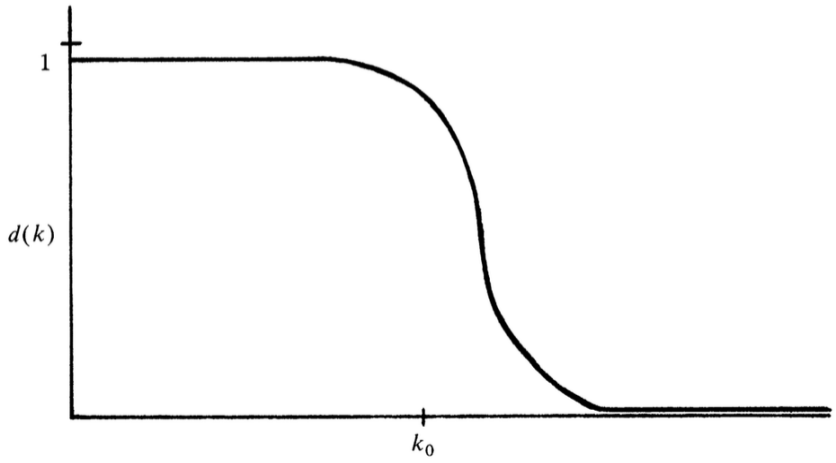


Figure 1.1: Cutoff phenomenon [1].

separated into 2 packets of size c and $n - c$. The packets are then riffled together and combined into one final packet. The mathematical formulation of this process is attributed to Edgar Gilbert, Claude Shannon, and J. Reeds. Gilbert and Shannon proposed the model in 1955, and Reeds proposed it separately in an unpublished manuscript in 1981. The model is referred to as the Gilbert-Shannon-Reeds (GSR) model and is explored further in Chapter 5.

In order to examine card shuffling, we use techniques established through the analysis of Markov chains and random walks. In Chapter 2, we explore some fundamental ideas and properities of Markov chains that create a foundation for our analysis of card shuffling. Examples of different Markov chains are discussed throughout the chapter as new ideas are introduced.

In Chapter 3 we discuss some generalities of random walks and then, more specifically, random walks on finite groups. We develop criteria for ergodicity and limiting distributions.

In Chapter 4 we discuss the total variation distance between two probability distribution and show that the probability distribution for an ergodic random walk on

a group converges to the uniform distribution. Stopping times are used to estimate total variation.

In Chapter 5, we discuss various mathematical models for card shuffling, stopping times, and the cutoff phenomenon. Methods established in earlier chapters are used to analyze the card shuffling models.

Appendix A includes Maple programs that were used to deepen our understanding of random walks, along with some additional programs that pertain to our observation of the cutoff phenomenon.

CHAPTER 2: Markov Chains

2.1 Markov Chains

The general definitions and theorems presented in this section can be found in [7] and in [9]. We begin our discussion of Markov chains with the following example.

Example 2.1. Consider a square with corners labeled s_1, s_2, s_3, s_4 in a counter-clockwise manner beginning in the upper right corner. Suppose that there is particle located at the corner labeled s_1 at time $t = 0$. For each time $t = 1, 2, \ldots$, the particle moves from its current location to one of the neighboring corners with equal probability.

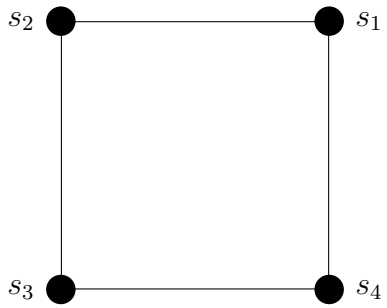


Figure 2.1: A particle moving around a square.

Denote by $S = \{s_1, s_2, s_3, s_4\}$ the state space of this process and denote the location of the particle at time k by X_k. Using this notation, since the particle starts at s_1 for $k = 0$, we have

$$\mathbf{P}(X_0 = s_1) = 1.$$

At $k = 1$ the particle will move to either s_2 or s_4, each with probability $1/2$. Therefore,

$$\mathbf{P}(X_1 = s_2) = 1/2 \quad \text{and} \quad \mathbf{P}(X_1 = s_4) = 1/2.$$

Under the conditions outlined, observe that the next position of the particle depends only on its current position and not on any of its past positions. This memoryless property is referred to as the *Markov property*. Any stochastic process that satisfies the Markov property is considered a *Markov process*. We thus see that the particle moving around the square in Figure 2.1 represents a Markov process. To further illustrate this notion, if the particle is at s_2 at time n, then

$$\mathbf{P}(X_{n+1} = s_1) = 1/2 \quad \text{and} \quad \mathbf{P}(X_{n+1} = s_3) = 1/2,$$

independent of past positions.

We can create a transition matrix, P, modeling this process by letting $P_{i,j}$, the i, j^{th} entry of P, represent the probability of moving from state i to state j. Thus, the transition matrix for the movement of the particle is given by

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

Definition 2.2. Let P be a $k \times k$ matrix with entries $P_{i,j}$ for $i, j = 1, \ldots, k$. A random process $(X_0, X_1, \ldots)$ with finite state space $S = \{s_1, \ldots, s_k\}$ is a *Markov chain* with transition matrix P if for all n, all $i, j \in \{1, \ldots, k\}$, and all $i_0, \ldots, i_{n-1} \in \{1, \ldots, k\}$, we have

$$\begin{aligned} &\mathbf{P}(X_{n+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, \ldots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) \\ &\qquad = \mathbf{P}(X_{n+1} = s_j | X_n = s_i) \\ &\qquad = P_{i,j}. \end{aligned}$$

Each $P_{i,j}$ is the conditional probability of going from state i to state j. Moreover, we

have that $P_{i,j} \geq 0$ for all $i, j \in \{1, \ldots, k\}$ and $\sum_{j=1}^{k} P_{i,j} = 1$ for all $i \in \{1, \ldots, k\}$.

From this definition, we see that the particle moving about the square in Example 2.1 can be represented by a Markov chain. A traditional method for visualizing Markov chains is a phase diagram. This diagram represents the possible states in the process, as well as the probability of transitioning from each state to the next possible states. Figure 2.2 shows the transition diagram for the particle moving about the square in Example 2.1.

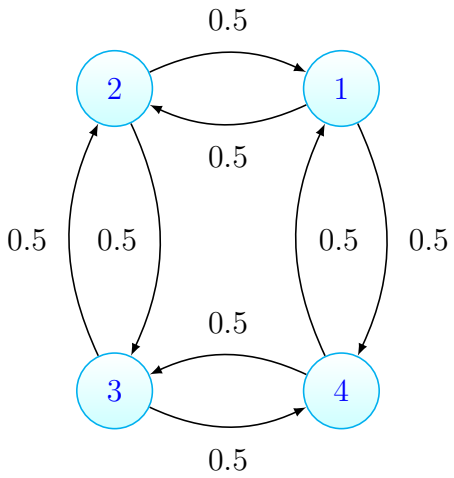


Figure 2.2: Transition diagram.

There are several techniques which can be used to further examine the behavior of Markov chains. Rather than specifying an initial state, we can begin with a probability distribution $\mu^{(0)}$ for the initial state. Then $\mu^{(n)} = \mu^{(0)} P^n$ is the probability distribution for the process after n steps, with $\mu_i^{(n)}$ denoting the probability of being at state s_i.

One of the most desired results is to determine the probability distribution of the particle after some period of time, and ultimately, the focus is whether or not the probability distribution has a limiting distribution. The exploration of this idea

involves the use of several matrix techniques.

Theorem 2.3. *Suppose that* $(X_0, X_1, \dots)$ *is a Markov chain with finite state space* $S = \{s_1, \dots, s_k\}$*, initial distribution* $\mu^{(0)}$*, and transition matrix* P*. For any* n*, the vector* $\mu^{(n)}$ *satisfies*

$$\mu^{(n)} = \mu^{(0)} P^n.$$

Proof. The proof of the theorem proceeds by induction. First, assume that $n = 0$. Then

$$\mu^{(0)} = \mu^{(0)} P^0 = \mu^{(0)} I,$$

where I is the identity matrix. Now, suppose for some $n \geq 0$ that $\mu^{(n)} = \mu^{(0)} P^n$. Then

$$\begin{aligned}
\mu_j^{(n+1)} &= \mathbf{P}(X_{n+1} = s_j) \\
&= \sum_{i=1}^{k} \mathbf{P}(X_n = s_i \text{ and } X_{n+1} = s_j) \\
&= \sum_{i=1}^{k} \mathbf{P}(X_n = s_i) \cdot \mathbf{P}(X_{n+1} = s_j | X_n = s_i) \\
&= \sum_{i=1}^{k} \mu_i^{(n)} P_{i,j} \\
&= (\mu^{(n)} P)_j,
\end{aligned}$$

which is the j^{th} entry of the vector of $\mu^{(n)} P$. By the induction hypothesis, it follows that

$$\mu^{(n+1)} = \mu^{(n)} P = \mu^{(0)} P^n P = \mu^{(0)} P^{n+1}.$$

□

2.2 Linear Algebra

In preparation for Markov chain analysis, it is important to establish some basic definitions and results concerning matrices. The general linear algebra definitions and theorems in this section can be found in [6]. Some of the most important results relating to the analysis of Markov chains come from the eigenvalues and eigenvectors of the transition matrix of the Markov process.

Definition 2.4. Given a $k \times k$ matrix A, a nonzero vector v is called an *eigenvector* of A if there exists a scalar λ such that $Av = \lambda v$. The scalar λ is called the *eigenvalue* corresponding to the eigenvector v.

Theorem 2.5. *The scalar λ is an eigenvalue of a $k \times k$ matrix A if and only if $\det(A - \lambda I) = 0$, where I is the identity matrix.*

Proof. Let A be a $k \times k$ matrix. A scalar λ is an eigenvalue of A if and only if there exists a vector $v \neq 0$ such that $Av = \lambda v$. This is true, however, if and only if $\ker(A - \lambda I)$ is non-trivial, and hence, $A - \lambda I$ is not invertible. Therefore, λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$. □

The polynomial $f(t) = \det(A - tI)$ is called the characteristic polynomial, and from the above theorem, we see that the zeros of the characteristic polynomial correspond to the eigenvalues of the matrix A. Note that the eigenvalues are the same for both left and right eigenvectors.

Proposition 2.6. *Let P be the transition matrix of an finite state Markov chain. For each eigenvalue λ we have $|\lambda| \leq 1$.*

Proof. Suppose that λ is an eigenvalue for the matrix P with associated right eigenvector v. Then

$$|\lambda v_i| = |(Pv)_i| = \left| \sum_j P_{i,j} v_j \right|.$$

Summing over i yields

$$\begin{aligned}\sum_i |\lambda v_i| &= \sum_i \left| \sum_j P_{i,j} v_j \right| \\ &\leq \sum_i \sum_j P_{i,j} |v_j| \\ &= \sum_j |v_j|\end{aligned}$$

Thus $\sum_i |\lambda v_i| \leq \sum_j |v_j|$, and the result follows. □

A second definition that is beneficial is the definition of a diagonal matrix. Diagonalization of a matrix readily simplifies the calculation of powers of that matrix, which makes diagonal matrices efficient tools for calculating the long-term distribution of a Markov chain.

Definition 2.7. A matrix A is said to be a *diagonal* matrix if $A_{i,j} = 0$ for all $i \neq j$, meaning that all entries off the main diagonal must be 0.

Definition 2.8. A matrix A is *diagonalizable* if and only if there exists some matrix U such that $A = UDU^{-1}$, where D is a diagonal matrix.

Lemma 2.9. *If a matrix A is diagonalizable, then $A^n = UD^nU^{-1}$.*

Lemma 2.10. *A matrix A is diagonalizable if and only if there exists a basis consisting of eigenvectors of A.*

2.3 Limiting Distributions and Stationary Distributions

A limiting distribution can be thought of as the "final state" of a Markov process. That is, for an initial distribution, $\mu^{(0)}$, of a Markov process with a limiting distribution σ, as $n \to \infty$, then $\mu^{(n)} \to \sigma$. Not every Markov chain has a limiting distribution,

but if one exists, then the chain *must* converge to that distribution, regardless of the initial distribution. Furthermore, if a limiting distribution exists, it will be what is known as a *stationary distribution.*

Definition 2.11. Let $(X_0, X_1, \dots)$ be a Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. A row vector $\pi = (\pi_1, \dots, \pi_k)$ is said to be a *stationary distribution* for the Markov chain if it satisfies each of the following:

(i) $\pi_i \geq 0$ for $i = 1, \dots, k$

(ii) $\sum_{i=1}^{k} \pi_i = 1$

(iii) $\pi P = \pi$.

Items (i) and (ii) mean that the probabilities must be non-negative and sum to 1, the requirements of any probability distribution. Item (iii) states that if π is a stationary distribution, then once stationary has been attained, subsequent steps of the Markov process do not alter the probability distribution. Thus the process maintains the same probability distribution. From (iii), one can see that π is a left eigenvector of the transition matrix P with eigenvalue 1.

Lemma 2.12. *If σ is a limiting distribution, then σ is a stationary distribution.*

Proof.

$$\sigma P = \lim_{n\to\infty} (\sigma P^n)P = \lim_{n\to\infty} \sigma P^{n+1} = \sigma.$$

□

Remark 2.13. It is not necessarily the case that a stationary distribution will be a limiting distribution. A distribution π may satisfy Definition 2.11 but may fail to satisfy $\mu^{(n)} \to \pi$ as $n \to \infty$.

Example 2.14. Consider a random particle as described in Example 2.1. This time, though, suppose that there is a fifth state, s_5, available to the particle that is located in the center of the square. From any given location, the probabilities for the next state of the particle are taken to be uniform. Figure 2.3 depicts this process.

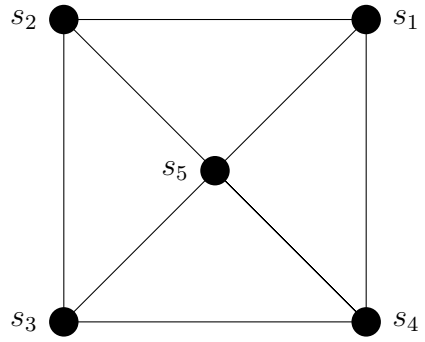


Figure 2.3: A particle moving around a square with a state in the center.

The transition matrix corresponding to this process is

$$P = \begin{pmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{pmatrix}.$$

This Markov chain exhibits a unique stationary distribution. In order to calculate this distribution, we first find the eigenvectors of P. The eigenvectors and their corresponding eigenvalues are summarized below. Note the fact of our earlier claim that $|\lambda_n| \leq 1$.

$$\lambda_1 = 0, \;\; v_1 = [\; 0 \quad -1 \quad 0 \quad 1 \quad 0 \;]$$

$$\lambda_2 = 0, \;\; v_2 = [\; -1 \quad 0 \quad 1 \quad 0 \quad 0 \;]$$

$$\lambda_3 = -2/3, \;\; v_3 = [\; -1 \quad 1 \quad -1 \quad 1 \quad 0 \;]$$

$$\lambda_4 = -1/3, \;\; v_4 = [\; -1/3 \quad -1/3 \quad -1/3 \quad -1/3 \quad 1 \;]$$

$$\lambda_5 = 1, \;\; v_5 = [\; 1 \quad 1 \quad 1 \quad 1 \quad 1 \;]$$

We now create a diagonal matrix D of eigenvalues and a matrix U of the associated eigenvectors. By Definition 2.7, we use D and U to calculate the powers of P, and from the results of Theorem 2.3, we observe the long-term distribution of the particle having transition matrix P.

For D, U, and U^{-1}, we have

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2/3 & 0 & 0 \\ 0 & 0 & 0 & -1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & -1 & -1 & -1/3 & 1 \\ -1 & 0 & 1 & -1/3 & 1 \\ 0 & 1 & -1 & -1/3 & 1 \\ 1 & 0 & 1 & -1/3 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

and

$$U^{-1} = \begin{pmatrix} 0 & -1/2 & 0 & 1/2 & 0 \\ -1/2 & 0 & 1/2 & 0 & 0 \\ -1/4 & 1/4 & -1/4 & 1/4 & 0 \\ -3/16 & -3/16 & -3/16 & -3/16 & 3/4 \\ 3/16 & 3/16 & 3/16 & 3/16 & 1/4 \end{pmatrix}.$$

Since $P^n = UD^nU^{-1}$, by looking at the limit as $n \to \infty$, observe below that every row of P^n converges to the same probabilities, and by definition, those probabilities represent the limiting distribution.

$$\begin{aligned} \lim_{n\to\infty} P^n &= \lim_{n\to\infty} UD^nU^{-1} \\ &= \lim_{n\to\infty} U \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (-2/3)^n & 0 & 0 \\ 0 & 0 & 0 & (-1/3)^n & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} U^{-1} \\ &= U \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} U^{-1} \end{aligned}$$

$$= \begin{pmatrix} 3/16 & 3/16 & 3/16 & 3/16 & 1/4 \\ 3/16 & 3/16 & 3/16 & 3/16 & 1/4 \\ 3/16 & 3/16 & 3/16 & 3/16 & 1/4 \\ 3/16 & 3/16 & 3/16 & 3/16 & 1/4 \\ 3/16 & 3/16 & 3/16 & 3/16 & 1/4 \end{pmatrix}$$

Recall that if there is a limiting distribution for a Markov chain, then that distribution will also be the stationary distribution, and thus π is given by

$$\pi = [\ 3/16 \quad 3/16 \quad 3/16 \quad 3/16 \quad 1/4\ ].$$

We point out that here π satisfies all conditions from Definition 2.11. Requirements (i) and (ii) are readily observable, and after a quick calculaiton, note that

$$\pi P = [\ 3/16 \quad 3/16 \quad 3/16 \quad 3/16 \quad 1/4\ ]\, P = \pi,$$

as necessary. Hence π is a stationary distribution for this Markov chain.

This is an example of how to determine the stationary distribution of a Markov process. In practice, however, not every Markov process converges to a stationary distribution, and for some processes there may exist multiple distributions satisfying Definition 2.11. The remainder of this section explains this idea further and provides some useful definitions and theorems. Before moving on, though, we now calculate the stationary distribution for the particle moving around the square in Example 2.1.

Example 2.15. Recall that the transition matrix for Example 2.1 is given by

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

Calculating the eigenvectors of P gives

$$\lambda_1 = 0,\ v_1 = [\ 0 \quad -1 \quad 0 \quad 1\ ]$$

$$\lambda_2 = 0,\ v_2 = [\ -1 \quad 0 \quad 1 \quad 0\ ]$$

$$\lambda_3 = -1,\ v_3 = [\ -1 \quad 1 \quad -1 \quad 1\ ]$$

$$\lambda_4 = 1,\ v_4 = [\ 1 \quad 1 \quad 1 \quad 1\ ]$$

Now we create a diagonal matrix D of eigenvalues and a matrix U of associated eigenvectors and use Theorem 2.3 to observe the long-term distribution of the particle around the square. Note that there are two eigenvalues, namely λ_3 and λ_4, that have modulus 1. This will present some interesting differences between Example 2.14 which had only a single eigenvalue of modulus 1.

For D, U, and U^{-1}, we have

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & -1 & 1 & -1 \\ -1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & 1 \end{pmatrix},$$

and

$$U^{-1} = \begin{pmatrix} 0 & -1/2 & 0 & 1/2 \\ -1/2 & 0 & 1/2 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ -1/4 & 1/4 & -1/4 & 1/4 \end{pmatrix}.$$

Since $P^n = UD^nU^{-1}$, taking the limit as $n \to \infty$ gives the long-term behavior of the particle.

$$\begin{aligned} \lim_{n\to\infty} P^n &= \lim_{n\to\infty} UD^nU^{-1} \\ &= \lim_{n\to\infty} U \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & (1)^n & 0 \\ 0 & 0 & 0 & (-1)^n \end{pmatrix} U^{-1} \end{aligned}$$

In this case, contrary to Example 2.14, note the distinction between P^n for odd and even values of n. For even n,

$$UD^nU^{-1} = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}$$

and for odd n, we have

$$UD^nU^{-1} = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}.$$

In this example, there is no limiting distribution since in the limit, we oscillate between two states, $v_4 \pm v_3$. One sees that the Markov process with probability matrix P^2 has two stationary states, v_3 and v_4.

In Example 2.14, however, $\lambda_5 = 1$, with $|\lambda_i| < 1$ for $i \neq 5$, resulting in $D^n_{i,i} \to 0$ as $n \to \infty$. Note that the stationary distribution is approached exponentially as we take powers of the transition matrix P. What dictates how quickly stationarity is reached is the modulus of the second largest eigenvalue. The coefficients of the characteristic polynomial are determined by the eigenvalues of the transition matrix. As such, smaller eigenvalues result in lower-degree terms tending to 0 more quickly than would occur with larger eigenvalues.

For our purposes we focus our attention on those Markov processes which do possess a unique stationary distribution. The types of Markov chains that possess a unique stationary distribution must now be established, and that is done via the following definitions.

Definition 2.16. Suppose that $(X_0, X_1, \dots)$ is a Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. For any two states $s_i, s_j \in S$, it is said

that s_i *communicates* with s_j if there exists $n > 0$ such that

$$\mathbf{P}(X_n = s_j | X_0 = s_i) > 0.$$

Communication between states is denoted by $s_i \to s_j$. Moreover, the situation when $s_i \to s_j$ and $s_j \to s_i$ is represented by $s_i \leftrightarrow s_j$, and the states s_i, s_j are said to *intercommunicate*.

Definition 2.17. Let $(X_0, X_1, \dots)$ be a Markov chain with transition matrix P and state space $S = \{s_1, \dots, s_k\}$. If $s_i \leftrightarrow s_j$ for all $s_i, s_j \in S$, then the Markov chain is said to be *irreducible*.

In other words, Definition 2.17 says that there is positive probability of reaching state s_j from any current state s_i for all $s_i, s_j \in S$ of an irreducible Markov chain. To formalize this notion, given an irreducible Markov chain, for any $s_i, s_j \in S$, there exists $m, n > 0$ such that

$$\mathbf{P}(X_m = s_j | X_0 = s_i) > 0 \text{ and } \mathbf{P}(X_n = s_i | X_0 = s_j) > 0.$$

By looking at whether states communicate, the states of any finite state Markov chain can be divided into groups: those that communicate with a given state and those that do not. A *class* $\mathcal{C}$ of states is defined to be a non-empty subset of states such that each $s_i \in \mathcal{C}$ communicates with all other $s_j \in \mathcal{C}$, and s_i does not communicate with $s_j \notin \mathcal{C}$. Moreover, a state s_i is *recurrent* if $s_i \to s_j$ implies that $s_j \to s_i$. If a state s_i is not recurrent, then it is called *transient*. The following theorem provides an interesting result concerning classes of recurrent and transient states for a finite state Markov chain.

Theorem 2.18. *Given a class $\mathcal{C}$ of states in a finite state Markov chain, either all states in $\mathcal{C}$ are recurrent or all states are transient.*

Proof. Suppose that $\mathcal{C}$ is a class of states. Let $s_i \in \mathcal{C}$ be transient. Then there exists $s_j \in \mathcal{C}$ such that $s_i \to s_j$ but $s_j \not\to s_i$. Now let $s_m \in \mathcal{C}$ with s_m recurrent. Therefore, $s_i \leftrightarrow s_m$ and $s_j \leftrightarrow s_m$. Observe that $s_m \to s_i$ and $s_i \to s_j$, which implies that $s_m \to s_j$. However, since $s_j \to s_m$ and $s_m \to s_i$ then it would be the case that $s_j \to s_i$. This is contrary to the assumption that $s_j \not\to s_i$. Hence s_m is not recurrent and must be transient. Thus all states in $\mathcal{C}$ must be transient. A similar argument shows that if $s_i \in \mathcal{C}$ is recurrent then all other states in $\mathcal{C}$ are also be recurrent. □

Following from Theorem 2.18, one can see that a Markov chain may eventually return to state s_i having started from s_i. For a state $s_i \in S$, if there exists n such that $\mathbf{P}(X_n = s_i | X_0 = s_i) > 0$, then n is called a *return time* for state s_i. The *period* of a state s_i, denoted by $d(s_i)$, is defined to be

$$d(s_i) = gcd(n \geq 1 : (P^n)_{i,i} > 0),$$

which is the greatest common divisor of all return times for s_i.

Definition 2.19. If $d(s_i) = 1$, then state s_i is said to be *aperiodic*. Furthermore, if all of the states of a Markov chain are aperiodic, then that Markov chain is said to be aperiodic. Otherwise, the chain is periodic.

2.4 Ergodicity

Now that it has been established what it means to be aperiodic and irreducible, the idea of ergodicity can be approached. Ergodicity is an important concept when

examining Markov chains, and it has a key role in whether or not a Markov chain has a limiting distribution.

Definition 2.20. Suppose that $(X_0, X_1, \dots)$ is a Markov chain with state space $S = \{s_1, \dots, s_k\}$. This Markov chain is said to be *ergodic* if s_i is both aperiodic and irreducible for all i.

Two examples of Markov chains have been presented - one that had a unique stationary distribution and one that did not. The remainder of this section explores three important properties of an ergodic Markov chain:

- the existence of a stationary distribution
- the existence of a limiting distribution
- the uniqueness of the stationary distribution.

Once each of these has been proved, it can then be assumed that there exists a unique stationary distribution for any process that can be modeled by an ergodic Markov chain. By using this unique stationary distribution, the number of transitions needed in order to reach the stationary distribution can be estimated.

First, it must be established that there exists a stationary distribution for any ergodic Markov chain. In order to do this, we use the idea of the hitting time for a given state.

Definition 2.21. Let $(X_0, X_1, \dots)$ be a Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. Suppose that $X_0 = s_i$. The *hitting time* $T_{i,j}$ for the state s_j is the random variable defined by

$$T_{i,j} = \min\{n \geq 1 : X_n = s_j\}$$

Moreover, the *mean hitting time*, $\tau_{i,j}$, is the expected time to travel to state s_j starting at state s_i. This is denoted by

$$\tau_{i,j} = \mathbf{E}[T_{i,j}].$$

Furthermore, $\tau_{i,i}$ is called the *mean return time* for state s_i.

Theorem 2.22. *Suppose that $(X_0, X_1, \dots)$ is an ergodic Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. There exists $N < \infty$ such that*

$$(P^n)_{i,j} > 0$$

for all $i \in \{1, \dots, k\}$ and $n \geq N$.

Proof. Consider the following lemma from Number Theory.

Lemma 2.23. *Let $A \subseteq \mathbb{N}$ be a set of positive integers such that*

(i) $gcd(A) = d$

(ii) A is closed under addition.

Let d be the minimum of the pair-wise greatest common divisors of elements in A. Then there exists $N < \infty$ such that $nd \in A$ for all $n \geq N$.

Proof. Suppose that $d = \min\{gcd(a, b) : a, b \in A\}$. Then there exists $a, b \in A$ so that $gcd(a, b) = d$. Moreover, there exists $x, y \in \mathbb{Z}$ such that

$$ax + by = d.$$

Pick $x', y' \in \mathbb{N}$ such that $x + x' > 0$ and $y + y' > 0$. By (ii), note that $ax' + by'$ is in A and is also divisible by d, so that

$$ax' + by' = md$$

for some $m \in \mathbb{N}$. By adding these two equations,

$$a(x + x') + b(y + y') = (m + 1)d.$$

Now suppose that $n \geq m^2$ with $n = qm + r$ for $0 \leq r < m$ and $q \geq m$. Then

$$\begin{aligned} n &= qm + r \\ &= qm + r(m + 1) - rm \\ &= (q - r)m + r(m + 1). \end{aligned}$$

Hence $nd = (q - r)md + rd(m + 1)$ is in A since $q - r > 0$, $m + 1 > 0$, and $md, (m + 1)d \in A$. Therefore $nd \in A$ for all $n \geq m^2$.

□

By the definition of ergodic, the given Markov chain is aperiodic, which means that $gcd\{n \geq 1 : (P^n)_{i,i} > 0\} = 1$. Let $s_i \in S$ and $A_i = \{n \geq 1 : (P^n)_{i,i} > 0\}$ be the set of all possible return times for s_i. Further, A_i is closed under addition since if $a, a' \in A_i$, then $\mathbf{P}(X_a = s_i | X_0 = s_i) > 0$ and $\mathbf{P}(X_{a+a'} = s_i | X_a = s_i) = \mathbf{P}(X_{a'} = s_i | X_0 = s_i) > 0$. Thus,

$$\begin{aligned} \mathbf{P}(X_{a+a'} = s_i | X_0 = s_i) &\geq \mathbf{P}(X_a = s_i \text{ and } X_{a+a'} = s_i | X_0 = s_i) \\ &= \mathbf{P}(X_a = s_i | X_0 = s_i) \cdot \mathbf{P}(X_{a+a'} = s_i | X_a = s_i) \\ &> 0 \end{aligned}$$

Thus A_i also satisfies (ii) from the lemma, and hence, for each i there exists some $N_i < \infty$ such that $(P^n)_{i,i} > 0$ for all $n \geq N_i$. By letting $N = \max\{N_1, \ldots, N_k\}$, the theorem is proved. □

Lemma 2.24. *Suppose that $(X_0, X_1, \dots)$ is an ergodic Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. Given any two states $s_i, s_j \in S$, if the chain starts in s_i, then*

$$\boldsymbol{P}(T_{i,j} < \infty) = 1.$$

Moreover, the mean hitting time is finite. That is, $\tau_{i,j} < \infty$.

Proof. As the Markov chain is ergodic, there exists $N < \infty$ such that $(P^N)_{i,j} > 0$ for all $i, j \in \{1, \dots, k\}$. Let $\alpha = \min\{(P^N)_{i,j} : i, j \in \{1, \dots, k\}\}$, and so $\alpha > 0$. Suppose that $X_0 = s_i$. Since $(P^N)_{i,j} \geq \alpha$ for all i and j, then $\mathbf{P}(X_N = s_j) \geq \alpha$. Therefore

$$\mathbf{P}(T_{i,j} > N) \leq \mathbf{P}(X_N \neq s_i) \leq (1 - \alpha).$$

Further, note that $\mathbf{P}(X_{2N} = s_j | X_N = s_k) = \mathbf{P}(X_N = s_j | X_0 = s_k) = (P^N)_{k,j} \geq \alpha$. This means that given the state at time N, there is conditional probability at least α of being in state s_j at time $2N$. Thus,

$$\begin{aligned}
\mathbf{P}(T_{i,j} > 2N) &= \mathbf{P}(T_{i,j} > N) \cdot \mathbf{P}(T_{i,j} > 2N | T_{i,j} > N) \\
&\leq \mathbf{P}(T_{i,j} > N) \cdot \mathbf{P}(X_{2N} \neq s_j | T_{i,j} > N) \\
&\leq (1 - \alpha)^2.
\end{aligned}$$

This argument can be iterated in order to conclude that for any l, it is the case that

$$\mathbf{P}(T_{i,j} > lN) \leq (1 - \alpha)^l.$$

As $0 \leq (1 - \alpha) < 1$, observe that as $l \to \infty$ then $(1 - \alpha)^l \to 0$. Moreover, note that $\mathbf{P}(T_{i,j} = \infty) = 0$, which shows $\mathbf{P}(T_{i,j} < \infty) = 1$.

For the mean hitting time,

$$\begin{aligned}
\tau_{i,j} &= \mathbf{E}[T_{i,j}] \\
&= \sum_{m=0}^{\infty} m\mathbf{P}(T_{i,j} = m) \\
&= \sum_{m=0}^{\infty} \sum_{n=0}^{m-1} \mathbf{P}(T_{i,j} = m) \\
&= \sum_{n=0}^{\infty} \sum_{m=n+1}^{\infty} \mathbf{P}(T_{i,j} = m) \\
&= \sum_{n=0}^{\infty} \mathbf{P}(T_{i,j} > n)
\end{aligned}$$

Thus,

$$\begin{aligned}
\tau_{i,j} &= \sum_{n=0}^{\infty} \mathbf{P}(T_{i,j} > n) \\
&= \sum_{l=0}^{\infty} \sum_{n=lN}^{(l+1)N-1} \mathbf{P}(T_{i,j} > n) \\
&\leq \sum_{l=0}^{\infty} \sum_{n=lN}^{(l+1)N-1} \mathbf{P}(T_{i,j} > lN) \\
&= N \sum_{l=0}^{\infty} \mathbf{P}(T_{i,j} > lN) \\
&\leq N \sum_{l=0}^{\infty} (1-\alpha)^l \\
&= N \frac{1}{1-(1-\alpha)} \\
&= \frac{N}{\alpha} \\
&< \infty
\end{aligned}$$

□

The results of Theorem 2.22 and Lemma 2.24 provide the conclusion that after a finite number of steps all probabilities of P will be positive. Furthermore, regardless of the starting position, it can be expected that the ergodic Markov process will reach any state $s_i \in S$ after a finite number of steps. These two conclusions are key findings for ergodic Markov chains.

Theorem 2.25. *For any ergodic Markov chain, there exists at least one stationary distribution.*

Proof. We follow the proof outlined in [9]. Let $(X_0, X_1, \dots)$ be an ergodic Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. Suppose that $X_0 = s_1$. For each $s_i \in S$, define

$$\rho_i = \sum_{n=0}^{\infty} \mathbf{P}(X_n = s_i \text{ and } T_{1,1} > n).$$

Thus ρ_i is the expected number of visits to s_i before the first return time. The mean return time $\tau_{1,1}$ is finite, and $\rho_i < \tau_{1,1}$, which implies that ρ_i is also finite.

To obtain a stationary distribution, consider π given by

$$\pi = (\pi_1, \dots, \pi_k) = \left(\frac{\rho_1}{\tau_{1,1}}, \frac{\rho_2}{\tau_{1,1}}, \dots, \frac{\rho_k}{\tau_{1,1}} \right).$$

It needs to be verified that π satisfies the two requirements from Definition 2.11. First, suppose that $j \neq 1$. Then

$$\begin{aligned} \pi_j = \frac{\rho_j}{\tau_{1,1}} &= \frac{1}{\tau_{1,1}} \sum_{n=0}^{\infty} \mathbf{P}(X_n = s_j \text{ and } T_{1,1} > n) \\ &= \frac{1}{\tau_{1,1}} \sum_{n=1}^{\infty} \mathbf{P}(X_n = s_j \text{ and } T_{1,1} > n) \\ &= \frac{1}{\tau_{1,1}} \sum_{n=1}^{\infty} \mathbf{P}(X_n = s_j \text{ and } T_{1,1} > n-1) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\tau_{1,1}} \sum_{n=1}^{\infty} \sum_{i=1}^{k} \mathbf{P}(X_{n-1} = s_i \text{ and } X_n = s_j \text{ and } T_{1,1} > n-1) \\
&= \frac{1}{\tau_{1,1}} \sum_{n=1}^{\infty} \sum_{i=1}^{k} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} > n-1) \mathbf{P}(X_n = s_j | X_{n-1} = s_i) \\
&= \frac{1}{\tau_{1,1}} \sum_{n=1}^{\infty} \sum_{i=1}^{k} P_{i,j} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} > n-1) \\
&= \frac{1}{\tau_{1,1}} \sum_{i=1}^{k} P_{i,j} \sum_{n=1}^{\infty} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} > n-1) \\
&= \frac{1}{\tau_{1,1}} \sum_{i=1}^{k} P_{i,j} \sum_{m=0}^{\infty} \mathbf{P}(X_m = s_i \text{ and } T_{1,1} > m) \\
&= \frac{\sum_{i=1}^{k} \rho_i P_{i,j}}{\tau_{1,1}} \\
&= \sum_{i=1}^{k} \pi_i P_{i,j}
\end{aligned}
$$

For $j = 1$, note that $\rho_1 = 1$, since ρ_1 is the expected number of visits to state s_1 up to time $T_{1,1}$. Therefore, we have

$$
\begin{aligned}
\rho_1 = 1 &= P(T_{1,1} < \infty) \\
&= \sum_{n=1}^{\infty} \mathbf{P}(T_{1,1} = n) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} = n) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} > n-1) \mathbf{P}(X_n = s_1 | X_{n-1} = s_i) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k} P_{i,1} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} > n-1) \\
&= \sum_{i=1}^{k} P_{i,1} \sum_{n=1}^{\infty} \mathbf{P}(X_{n-1} = s_i \text{ and } T_{1,1} > n-1)
\end{aligned}
$$

$$= \sum_{i=1}^{k} P_{i,1} \sum_{m=0}^{\infty} \mathbf{P}(X_m = s_i \text{ and } T_{1,1} > m)$$
$$= \sum_{i=1}^{k} \rho_i P_{i,1}$$

Hence $\pi_1 = \frac{\rho_1}{\tau_{1,1}} = \sum_{i=1}^{k} \frac{\rho_i P_{i,1}}{\tau_{1,1}} = \sum_{i=1}^{k} \pi_i P_{i,1}$. Thus, $\pi P = \pi$.

Observe that $\pi_i \geq 0$ for $i = 1, \ldots, k$. To see that $\sum_{i=1}^{k} \pi_i = 1$,

$$\begin{aligned}
\tau_{1,1} = \mathbf{E}[T_{1,1}] &= \sum_{n=0}^{\infty} \mathbf{P}(T_{1,1} > n) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k} \mathbf{P}(X_n = s_i \text{ and } T_{1,1} > n) \\
&= \sum_{i=1}^{k} \sum_{n=1}^{\infty} \mathbf{P}(X_n = s_i \text{ and } T_{1,1} > n) \\
&= \sum_{i=1}^{k} \rho_i
\end{aligned}$$

Thus $\sum_{i=1}^{k} \pi_i = \frac{1}{\tau_{1,1}} \sum_{i=1}^{k} \rho_1 = 1$. Hence π is a stationary distribution for the given Markov chain. □

We may now conclude that there exists a stationary distribution for any ergodic Markov chain. As illustrated by Example 2.15, not all Markov chains have a limiting distribution; however, ergodic Markov chains do.

To this end, an important result for ergodic Markov chains is that given any column in a transition matrix P for the Markov chain, for P^n, the largest element of the column is non-increasing with n and the smallest element is non-decreasing with n. This is captured by the following lemma.

Lemma 2.26. *Suppose that $(X_0, X_1, \ldots)$ is an ergodic Markov chain with state space $S = \{s_1, \ldots, s_k\}$ and transition matrix P. Suppose that $X_0 = s_i$. For any $s_j \in S$ and*

$n \geq 1$, the following inequalities hold.

$$\min_i (P^{n+1})_{i,j} \geq \min_\ell (P^n)_{\ell,j} \quad \textit{and} \quad \max_i (P^{n+1})_{i,j} \leq \max_\ell (P^n)_{\ell,j}.$$

This lemma provides the conclusion that for any Markov chain, one of two situations occurs. One case is that the sequence of decreasing maximal elements in each column has the same limit as the sequence of increasing minimal elements in each column. In this instance, note that the elements in each column are converging to a constant value, and therefore, the P converges to a matrix with constant columns. On the other hand, it may be the case that the sequence of decreasing maximal elements converges to some limit that is strictly greater than the increasing sequence of minimal elements. In this case P does not converge to a matrix with constant columns.

The determining factor in whether or not the former or latter case applies to a given Markov chain is whether or not the chain is ergodic. By Theorem 2.22 there exists N making all entries of P^n positive for $n \geq N$. This fact, along with the previous lemma provides basis for the following theorem.

Theorem 2.27. *Suppose that $(X_0, X_1, \dots)$ is an ergodic Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P. Let $\alpha = \min\{P_{i,j} : i, j \in \{1, \dots, k\}\}$. Then for all $s_j \in S$ and $n \geq N$,*

$$\max_\ell (P^n)_{\ell,j} - \min_\ell (P^n)_{\ell,j} \leq (1 - 2\alpha)^n.$$

Moreover,

$$\lim_{n\to\infty} \max_\ell (P^n)_{\ell,j} = \lim_{n\to\infty} \min_\ell (P^n)_{\ell,j} > 0.$$

Proof. Let s_j and n be fixed. Denote by $\ell_{\min}$ the value of ℓ that minimizes $(P^n)_{\ell,j}$,

and define $\ell_{\max}$ similarly. Let $\alpha(j,n)$ be the minimum and $\beta(j,n)$ be the maximum of $\{(P^n)_{\ell,j} : \ell = 1,\ldots,k\}$. Then

$$\begin{aligned}
(P^{n+1})_{i,j} &= \sum_k P_{i,k}(P^n)_{k,j} \\
&= \sum_{k\neq\ell_{\min}} P_{i,k}(P^n)_{k,j} + P_{i,\ell_{\min}}\alpha(j,n) \\
&\leq \sum_{k\neq l_{\min}} P_{i,k}\beta(j,n) + P_{i,\ell_{\min}}\alpha(j,n) \\
&= (1-P_{i,\ell_{\min}})(\beta(j,n)) + P_{i,\ell_{\min}}\alpha(j,n) \\
&= \beta(j,n) - P_{i,\ell_{\min}}(\beta(j,n)-\alpha(j,n)) \\
&\leq \beta(j,n) - \alpha(\beta(j,n)-\alpha(j,n))
\end{aligned}$$

By a similar argument, substituting ℓ_{max} for ℓ_{min} yields the following inequality

$$(P^{n+1})_{i,j} \geq \alpha(j,n) + \alpha\left(\beta(j,n)-\alpha(j,n)\right).$$

These two inequalities establish upper and lower bounds for the maximum and minimum elements for each column in (P^{n+1}). By subtracting the lower bound from the upper bound,

$$\max_i(P^{n+1})_{i,j} - \min_i(P^{n+1})_{i,j} \leq \left(\max_\ell(P^n)_{\ell,j} - \min_\ell(P^n)_{\ell,j}\right)(1-2\alpha).$$

Note

$$\alpha \leq P_{\ell,j} \leq 1-\alpha.$$

Hence

$$\beta(j,1)-\alpha(j,1) \leq 1-2\alpha$$

and further,

$$\beta(j, n+1) - \alpha(j, n+1) \leq (\beta(n+1) - \alpha(n+1))(1 - 2\alpha).$$

Therefore $\beta(j, n) - \alpha(j, n) \leq (1 - 2\alpha)^n$. As $(1 - 2\alpha) < 1$, and $(P^n)_{i,j} > 0$ for all i and j,

$$\lim_{n\to\infty} \max_i (P^n)_{i,j} = \lim_{n\to\infty} \min_i (P^n)_{i,j} > 0.$$

□

Thus it is clear that for any ergodic Markov chain, the lower bound for the minimal elements and the upper bound for the maximal elements both converge exponentially in n. Now the task is to describe the limit of the entries in the powers of the transition matrix. Quite simply, the rows are converging to a (unique) stationary distribution of that Markov chain (see Theorem 2.29). Define π_j by

$$\pi_j = \lim_{n\to\infty} \max_i (P^n)_{i,j} = \lim_{n\to\infty} \min_i (P^n)_{i,j} > 0.$$

Therefore it is clear that $\min_i (P^n)_{i,j} \leq \pi_j \leq \max_i (P^n)_{i,j}$ for all n. Thus it is now observable that for all n

$$|(P^n)_{i,j} - \pi_j| \leq |\max(P^n)_{i,j} - \min(P^n)_{i,j}| \leq (1 - 2\alpha)^n.$$

This offers the conclusion that

$$\lim_{n\to\infty} (P^n)_{i,j} = \pi_j.$$

This says that, for an ergodic Markov chain the rows of P^n converge to the limiting

distribution. This limit is approached exponentially in n as the i, j entries of the matrix approach π_j for each i, j. The results are further summarized by Theorem 2.28

Theorem 2.28. *Suppose that $(X_0, X_1, \dots)$ is an ergodic Markov chain with state space $S = \{s_1, \dots, s_k\}$ and transition matrix P having $\pi > 0$ as the limiting distribution. Then π satisfies each of the following:*

- $\lim_{n\to\infty}(P^n)_{i,j} = \pi_j$ *for all* j
- $\lim_{n\to\infty} P^n = \boldsymbol{e}\pi$, *where* $\boldsymbol{e} = (1, 1, \dots, 1)^T$.

This provides a method for determining a specific stationary distribution for any ergodic Markov chain. There remains, however, the concern of uniqueness of this distribution.

Can there be more than one stationary distribution? By Theorem 2.25, it is the case that there exists at least one stationary distribution for an ergodic Markov chain. It is, in fact, the case that there is a unique stationary distribution for an ergodic Markov chain. Recall the random particle from Example 2.1 which led to the existence of a Markov process with two stationary distributions. This result is attributed to the fact that the states of that process are periodic as

$$d(s_i) = 2$$

for all $s_i \in S$ and, as such, is not ergodic. This illustrates the strength of ergodicity when examining finite state Markov chains. It is a desirable quality that provides significant uses when exploring long-term probabilities of the Markov chain. The following theorem provides uniqueness to this stationary distribution and concludes the current discussion of Markov chains.

Theorem 2.29. *Suppose that* $(X_0, X_1, \dots)$ *is an ergodic Markov chain with state space* $S = \{s_1, \dots, s_k\}$ *and transition matrix* P*. For this Markov chain, there exists a unique stationary distribution* $\pi > 0$ *satisfying Definition 2.11 and Theorem 2.28.*

Proof. Let π be the limiting distribution satisfying Definition 2.11 and Lemma 2.28. Suppose that μ is any stationary distribution. Then $\mu = \mu P^n$ for all $n \geq 1$. By letting $n \to \infty$

$$\mu = \mu \lim_{n\to\infty} P^n = \mu \mathbf{e}\pi = \pi.$$

Hence π is unique. □

CHAPTER 3: Random Walks

The general definitions and theorems throughout this chapter can be found in [3], [8], and [10].

3.1 Random Walks on $\mathbb{R}^m$

At its core, a random walk is the mathematical representation of a series of random steps. Each of these steps has a given probability of occurring, and for this reason many random walks can be modeled by Markov chains. In Example 2.1. the Markov process, in fact, represents a random walk of the particle around the four vertices of the square. Note that each movement is random, and there are given probabilities for each transition. For lower dimensions, it is possible to visualize random walks through some common comparisons. For example random walks on $\mathbb{R}^m$ for $m = 1, 2, 3$ can be thought of as a random walk on a line, in a city, and in a jungle gym, respectively. For higher dimensional random walks, it is not possible to have such visualizations.

In representing a random walk, let $\xi_1, \ldots, \xi_n \in \mathbb{R}^m$ denote independent and identically distributed discrete random variables with distribution function p, and denote by X_n the sum

$$X_n = \sum_{k=1}^{n} \xi_k.$$

The sequence $\{X_n\}_{n=1}^{\infty}$ is called a random walk, and moreover, the range of the ξ_k's determines the space of the walk. For example, if $\{\xi_k\} \subseteq \mathbb{R}^m$, then X_n is said to be a random walk on $\mathbb{R}^m$. Note that X_n represents the location of the walk after n steps, with the convention that $X_0 = 0$. This discussion of random walks begins by first considering results on $\mathbb{R}^1$. Once these results for $\mathbb{R}^1$ are established, we extrapolate to results in higher dimensions.

3.2 Random Walks on $\mathbb{R}^1$

In $\mathbb{R}^1$, the distribution function of the ξ_k's is modeled by the equation

$$p(\xi) = \begin{cases} 1/2, & \xi = \pm 1 \\ 0, & \text{otherwise.} \end{cases}$$

When studying random walks, a topic of concern is the equalization of the random walk. Equalization refers to the return of the walk to the origin. If the random walk returns to the origin at time n, say $X_n = 0$, then an equalization occurs at time n. It is noteworthy to point out that in the case of $\mathbb{R}^1$ all equalizations must occur at even times. To justify this, observe that to have an equalization in $\mathbb{R}^1$, there must be exactly m steps in the $+1$ direction and exactly m steps in the -1 direction. Thus, $2m$ steps are needed for an equalization.

We are interested in the probability of having a return time, which is synonomous with equalization, at a given time $2m$. In order to calculate the probability that $2m$ is a return time, consider the number of paths having length $2m$ with $X_0 = X_{2m} = 0$ and no further conditions on time n with $0 < n < 2m$. (Later we introduce the requirement that $X_n \neq 0$ for all $0 < n < 2m$.) The number of paths achieving an equalization at time $2m$ is given by $\binom{2m}{m}$, with probability of 2^{-2m} on each of these paths.

Definition 3.1. Given a random walk, we denote the probability that an equalization occurs at time $2m$ by u_{2m}.

Definition 3.2. Given a random walk, we denote the probability that a first equalization occurs at time $2m$ by f_{2m}. That is, $X_{2m} = 0$ and $X_{2k} \neq 0$ for all $k < m$.

Using this notation, the relationship between u_{2k} and f_{2k} is expressed by the

following lemma.

Lemma 3.3. *For $n \geq 1$, the following relationship exists between $\{u_{2k}\}$ and $\{f_{2k}\}$*

$$u_{2n} = f_0 u_{2n} + f_2 u_{2n-2} + \cdots + f_{2n} u_0.$$

Proof. There are $u_{2n}2^{2n}$ number of ways to return to the origin in $2n$ steps. For any such walk, there will be a first return at time $2k$ for some $k \leq n$. There are $f_{2k}2^{2k}$ paths having a first return time at $2k$, and there are $u_{2n-2k}2^{2(n-k)}$ ways to complete the remainder of the walk. Hence, there are $f_{2k}2^{2k}u_{2n-2k}2^{2(n-k)} = f_{2k}u_{2n-2k}2^{2n}$ paths that satisfy a return time at $2n$, with a first return time at $2k$. We take the sum over $k \leq n$ to obtain the result. □

Corollary 3.4. *Let $U(x), F(x)$ be generating functions for u_{2m} and f_{2m} so that*

$$U(x) = \sum_{n=0}^{\infty} u_{2n} x^n$$

and

$$F(x) = \sum_{n=0}^{\infty} f_{2n} x^n.$$

Then by Lemma 3.3

$$F(x) = \frac{U(x) - 1}{U(x)} = 1 - \frac{1}{U(x)}.$$

Given this relationship between return time and first return time, we determine the probability of a first return time at $2n$ in the next theorem.

Theorem 3.5. *For $n \geq 1$, the probability of a first return time at $2n$ is given by*

$$f_{2n} = \frac{\binom{2n}{n}}{(2n-1)4^n}.$$

Proof. Let U and F be generating functions as in Corollary 3.4. Recall that the probability of a return to the origin at time $2n$ is given by $u_{2n} = \binom{2n}{n}2^{-2n}$. Substituting, yields

$$U(x) = \sum_{n=0}^{\infty} \binom{2n}{n} \frac{x^n}{4^n} = \frac{1}{\sqrt{1-x}}.$$

Note that

$$F(x) = 1 - \frac{1}{U(x)} = 1 - \sqrt{1-x}$$

and so

$$F'(x) = \frac{1}{2\sqrt{1-x}} = \frac{1}{2}U(x).$$

To obtain the coefficients for f_{2n}, integrate the function $U(x)$.

$$\begin{aligned} \int U(x)dx &= \sum_{n=0}^{\infty} \binom{2n}{n} \frac{x^{n+1}}{2^{2n}(n+1)} \\ &= \sum_{n=1}^{\infty} \binom{2(n-1)}{n-1} \frac{x^n}{n2^{n-1}} + C. \end{aligned}$$

Hence

$$\begin{aligned} f_{2n} &= \binom{2n-2}{n-1} \frac{1}{n2^{2n-1}} \\ &= \left(\frac{(2n-2)!}{(n-1)!(n-1)!}\right)\left(\frac{2n}{2n}\right)\left(\frac{2n-1}{2n-1}\right)\left(\frac{n}{n}\right)\left(\frac{1}{n2^{2n-1}}\right) \\ &= \left(\frac{(2n)!}{n!n!}\right)\left(\frac{1}{2^{2n-1}(2)(2n-1)}\right) \\ &= \binom{2n}{n}\left(\frac{1}{(2n-1)4^n}\right) \end{aligned}$$

□

This establishes the result for the probability that a first return occurs at time $2n$. However, interest in the probability that a first return to the origin occurs at

time $2n$, in some sense, offers the question "has *any* return to the origin occured by time $2n$?" It is an interesting notion, and it does not have a trivial answer. As this idea is explored, denote the probability w_{2n} to be the probability that the walk has experienced a first return to the origin at a time no later than $2n$. In order to calculate w_{2n} one must consider the set of all random walks of length $2n$.

Definition 3.6. Given a random walk in $\mathbb{R}^m$, we denote the probability that a first return has occurred no later than time $2n$ by w_{2n}. Moreover, define

$$w_* = \lim_{n\to\infty} \sum_{i=1}^{n} f_{2i}$$

to be the probability of an *eventual return*. The relationship between the probability of eventual return to the probability of a first return is given by

$$w_{2n} = \sum_{i=1}^{n} f_{2i}.$$

It happens that examining the notion of eventual return shows different results for different dimensions. One thing is clear, however, and that is the fact that $w_* \leq 1$. Moreover, the probability of eventual return is inversely related to the dimension where the walk takes places. Intuitively, one can attribute this fact to the increase in the number of possible movements of a random particle traveling in $\mathbb{R}^1$ versus in $\mathbb{R}^2$. In $\mathbb{R}^m$, as there are 2^m possible transitions from one state to the next. Therefore, as m grows, it becomes considerably less likely that a random particle will ever return to the origin.

Theorem 3.7. *For a random walk on $\mathbb{R}^1$, the probability of eventual return is* 1.

Proof. Let f_{2n} denote the probability of a first return at $2n$. It suffices to show that the sum of all first return times is 1. To do this, we first use Stirling's approximation

for factorials. Stirling's approximation says

$$\sqrt{2\pi}n^{n+1/2}e^{-n} \leq n! \leq n^{n+1/2}e^{1-n}.$$

Using this gives

$$\begin{aligned}\binom{2n}{n} &\leq \frac{(2n)^{n+1/2}e^{1-2n}}{(\sqrt{2\pi}n^{n+1/2}e^{-n})^2} \\ &= \frac{2^{2n+1/2}n^{2n+1/2}e^{1-2n}}{2\pi n^{2n+1}e^{-2n}}\end{aligned}$$

This shows that

$$\frac{1}{2n-1}\binom{2n}{n}\frac{1}{4^n} < c\frac{1}{n^{3/2}}$$

for some constant c, and we know that $\sum \frac{1}{n^{3/2}}$ converges, and so $\sum f_{2n}$ converges. Now,

$$F(x) = \sum f_{2n}x^n = 1 - \sqrt{1-x},$$

and therefore,

$$\lim_{x\to 1^-} \sum f_{2n}x^n = \sum f_{2n} = F(1) = 1.$$

□

3.3 Random Walks on $\mathbb{R}^m$, $m \geq 2$

In [8], Grinstead and Snell show that the probability of eventual return in $\mathbb{R}^2$ is also equal to 1. However, the probability of eventual return on $\mathbb{R}^3$ is approximately 0.65. As previously noted, the probability of eventual return decreases as dimension increases.

3.4 Random Walks on Finite Groups

Definition 3.8. Suppose that G is a finite group with order $|G|$, and p is a probability measure on G with the support of p being Σ. That is, if $p(\xi) > 0$, then $\xi \in \Sigma$. The *left-invariant random walk* on G driven by p is the Markov chain with state space $S = G$ and transition kernel P. In this Markov chain, we have $X_n = \xi_0\xi_1 \cdots \xi_n$ where each $\xi_j \in \Sigma$, and

$$p(x) = \mathbf{P}(\xi = x)$$

for all $x \in G$. We transition from x to y by $x(x^{-1}y)$, and thus the probability of this transition is

$$P(x, y) = p(x^{-1}y).$$

Moreover, the *iterated kernel* $P^n(x, y)$ is given recursively by

$$P^n(x, y) = \sum_{z \in G} P^{n-1}(x, z)P(z, y)$$

with the convention that $P^1 = P$.

Definition 3.9. Given two probability measures p and q on a group G, define the *group convolution* $p * q$ by

$$(p * q)(x) = \sum_{y \in G} p(y)q(y^{-1}x).$$

Thus, we obtain

$$\begin{aligned} P^2(x, y) &= \sum_{z \in G} P(x, z)P(z, y) \\ &= \sum_{z \in G} p(x^{-1}z)p(z^{-1}y) \end{aligned}$$

$$= \sum_{z \in G} p(z)p(z^{-1}x^{-1}y), \text{ obtained by changing } z \text{ to } xz$$
$$= p * p(x^{-1}y)$$

Hence the iterated kernel for $P^n(x, y)$ is given by the *iterated group convolution* $p^{*n}(x^{-1}y)$, where

$$p^{*n} = \underbrace{p * \cdots * p}_{n \text{ times}}.$$

Let (ξ_i), $i = 0, 1, \ldots$ be a sequence of independent $G-$valued random variables. Let ξ_0 follow the probability measure ν and ξ_i follow the probability measure p for $i \geq 1$. Thus, the random walk on G, has probability distribution νP^n at time n, where $P^n(x, y) = p^{*n}(x^{-1}y)$.

The particle moving in Example 2.1 is an example of a random walk on the cyclic group C_4. We have already shown that the Markov process modeling the particle in Example 2.1 is not ergodic. In general, the random walk on C_n is ergodic for n odd and is not ergodic for n even. This is because the set of return times on C_n for n even are all even, making a random walk on C_n periodic for even n.

Example 3.10. Let $Q_n = \{0, 1\}^n = \{(a_1, \ldots, a_n) : a_k \in \{0, 1\}, k = 1, \ldots, n\}$, which is the $n-$dimensional hypercube. A random walk on the hypercube is a random walk on $\mathbb{Z}_2^n$. The progression of steps throughout the random walk is modeled by a transition kernel with transition probabilities independent of previous transitions and each transition occurring randomly.

For a random walk on the $n-$dimensional hypercube, there are 2^n possible states. For simplicity, we consider the 3-cube, Q_3, whose graph is given in Figure 3.1. For each transition, the walk will either remain in its current location or will move to one of the three adjacent vertices. Each of these four events have probability 1/4. In

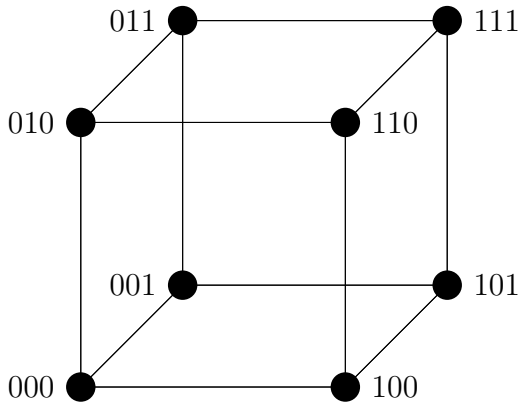


Figure 3.1: The hypercube, Q_3

general, a random walk on the $n-$dimensional hypercube will have probability $\frac{1}{n+1}$ of remaining at its current location and probability $\frac{1}{n+1}$ of transitioning to any given neighboring vertex. It is worth mentioning that we explored an example a random walk on the $2-$cube in Example 2.1. In this previous example there was not the option of remaining in place.

A random walk is a Markov process, and we have already pointed out that a Markov chain will exhibit a unique stationary distribution if the Markov chain is ergodic (i.e., irreducible and aperiodic). We now examine when a random walk on a finite group G is ergodic.

Theorem 3.11. *Let G be a finite group with p a probability measure on G, and let $\Sigma = \{x \in G : p(x) > 0\}$. Then*

(i) The random walk driven by p is irreducible if and only if Σ generates G.

(ii) Assuming that Σ generates G, then the random walk is aperiodic if and only if Σ is not contained in a coset of a proper normal subgroup of G.

For intuition purposes, consider the following two examples to illustrate Theorem 3.11.

Example 3.12. Consider $G = \mathbb{Z}_n$ for n even, and let p be the probability measure on G so that $p(x) = 1/2$ for $x = \pm 1$ and 0 otherwise. Let $J = \{n : p^{(n)}(e) > 0\}$ be the set of return times in $\mathbb{Z}_n$. The support of p is $\Sigma = \{1, -1\}$. As n is even, the return times must all be even - this was pointed out earlier - and so the random walk on G is periodic since $gcd\{n : n \in J\} = 2$. Thus the random walk is not ergodic. Let $H = \{n : n \in \mathbb{Z}_n \text{ and } n \text{ is even}\}$, which is a proper normal subgroup of G. According to the theorem, the random walk driven by p is aperiodic if and only if Σ is not contained in a coset of a proper normal subgroup of G. It is clear here that Σ violates this because $\Sigma = \{-1, 1\} \subset H + 1$. It is worth mentioning that if we allow q to be a probability measure on G such that $q(0) = q(1) = q(-1) = 1/3$, then the support of G is $\Sigma = \{-1, 0, 1\}$. Now Σ is not contained in a coset of a proper normal subgroup, making the random walk driven by q ergodic.

Example 3.13. Consider $G = S_n$, and let p be the uniform probability measure on the set Σ of all transpositions. Let ξ_j be one step in a random walk on G driven by p. For a series of steps to result in a return to the origin, we need $\xi_1 \xi_2 \cdots \xi_n = e$. Since each ξ_j represents a transposition, then each return time must contain an even number of transpositions. In this example, as in the last, there is not a unique stationary distribution. If we let $J = \{k : p^{(k)}(x^{-1}y) > 0\}$, then $gcd\{k : k \in J\} = 2$. To see how this example relates to the theorem, we let A_n be the set of even permutations, which is a proper normal subgroup of G. Then we see that $\Sigma \subset (12)A_n$, and so the walk driven by p is not ergodic. As in the first example, this walk on G can be made ergodic if we make $p(e) \neq 0$, where e is the identity element of G. By allowing $p(e) > 0$, we prevent Σ from being contained in a coset of a proper normal subgroup.

Note that the random walks in the above examples are not ergodic, but they become ergodic once we allow the walk to remain in its current position (by allowing

$p(e) > 0$). By doing this, the support of the function is no longer contained in a proper normal subgroup of G. We showed earlier that the Markov chain in Example 2.1, which is a random walk on the group $G = \mathbb{Z}_2^2$, is not ergodic.

We now prove Theorem 3.11.

Proof. Suppose that the random walk driven by p is irreducible. Then $e \to g$ for all $g \in G$. Therefore, Σ must generate G. Now suppose that Σ generates G. If we are at position g of the walk driven by p, we can surely extend the walk to some position h since Σ generates G. Hence the random walk driven by p is irreducible.

Now, assume that (i) of Theorem 3.11 is satisfied. Let H be a proper normal subgroup of G, and suppose that $\Sigma \subset gH$. Suppose that $\xi_1 \cdots \xi_n = e$ with each $\xi_j \in \Sigma$. Then $\xi_j = gh_j$ for some $h_j \in H$. Thus

$$e = \xi_1 \cdots \xi_n \in (gH)^n = g^n H.$$

Hence $e \in g^n H$, and thus, $g^n H = H$. This implies that n is divisible by $o(gH) > 1$, which implies that the walk driven by p is periodic.

Conversely, suppose that the walk on G driven by p is periodic. The walk will experience a return time when there is a sequence of transitions that result in the identity. Since the random walk on G is periodic, the greatest common divisor of these return times will not be 1. Define

$$m = gcd\{n : \xi_1 \cdots \xi_n = e,\ \xi_j \in \Sigma\} > 1.$$

For $k = 0, \ldots, m-1$, define

$$\Sigma^k = \{\xi_1 \cdots \xi_K : \xi_j \in \Sigma,\ K \equiv k \mod n\}.$$

Then for $\xi \in \Sigma$, we have $\xi^{-1} \in \Sigma^{m-1}$. To show that the Σ^k's are disjoint, suppose that $\xi_1 \dots \xi_p = \xi'_1 \dots \xi'_q$. Then $e = \xi_p^{-1} \dots \xi_1^{-1} \xi'_1 \dots \xi'_q$ is in $\Sigma^{p(m-1)+q}$. Since any product giving the identity must be in Σ^0, we have $p(m-1)+q \equiv 0 \mod m$, hence $p \equiv q \mod m$. We also see that $x \in \Sigma^p$ and $y \in \Sigma^q$ implies $xy \in \Sigma^{p+q \mod m}$.

Consider the set $H = \Sigma^0$. Clearly H is a subgroup of G. If $x \in \Sigma^k$ and $h \in \Sigma^0$, then $x^{-1}hx \in \Sigma^{n-k+0+k \mod m} = \Sigma^0$, so H is a proper normal subgroup of G. We also see that $\Sigma \subset \Sigma^1$, so Σ is contained in a coset of the proper normal subgroup H. $\square$

CHAPTER 4: Total Variation

4.1 Basics of Total Variation

Total variation is used to determine the distance between two probability distributions. In our continuing analysis of random walks, we use total variation to measure the difference between the probability distribution of a random walk on a finite group and the uniform distribution. We have the following definition for total variation.

Definition 4.1. Given two probability measures μ and ν on a finite group G, define total variation by

$$d_{TV}(\mu, \nu) = \max_{A \subseteq G} \left| \sum_{x \in A} (\mu(x) - \nu(x)) \right|.$$

Proposition 4.2. *Let $\|\mu - \nu\|_1 = \sum_x |\mu(x) - \nu(x)|$ denote the standard L^1 norm. The following is an equivalent definition for total variation*

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{x \in G} |\mu(x) - \nu(x)|.$$

Proof. To show this, let $A \subseteq G$ be the set of all x such that $\mu(x) - \nu(x) \geq 0$. Further, note that since μ and ν are probability measures, then

$$\sum_{x \in G} \mu(x) = \sum_{x \in G} \nu(x) = 1.$$

Thus

$$\sum_{x \in A} \mu(x) + \sum_{x \in A^c} \mu(x) = \sum_{x \in A} \nu(x) + \sum_{x \in A^c} \nu(x)$$

and

$$\sum_{x \in A} (\mu(x) - \nu(x)) = \sum_{x \in A^c} (\nu(x) - \mu(x)) = d_{TV}(\mu, \nu).$$

Now

$$\begin{aligned}\|\mu-\nu\|_1 &= \sum_{x\in A}|\mu(x)-\nu(x)| \\ &= \sum_{x\in A}\mu(x)-\nu(x)+\sum_{a\in A^c}\nu(x)-\mu(x) \\ &= 2d_{TV}(\mu,\nu)\end{aligned}$$

Hence,

$$d_{TV}(\mu,\nu)=\frac{1}{2}\|\mu-\nu\|_1.$$

□

Let G be a group, and suppose that μ represents the uniform distribution on G. Thus

$$\mu(x)=\frac{1}{|G|},$$

for $x\in G$, where $|G|$ is the order of G. By definition of stationary distribution, we note that $\mu(x)$ is a stationary distribution for any random walk on G by noting that $\mu(x)\geq 0$ for all $x\in G$, and

$$\sum_{x\in G}\frac{1}{|G|}p(x^{-1}y)=\frac{1}{|G|},$$

for any probability measure p. We summarize this:

Theorem 4.3. *Let p be a probability measure on a finite group G. The random walk on G that is driven by p exhibits the uniform distribution $\mu=\frac{1}{|G|}$ as a stationary distribution.*

We now present the following example to illustrate a calculation for total variation.

Example 4.4. Consider a random walk on $G = \mathbb{Z}_{10}$ with μ the uniform distribution on $\mathbb{Z}_{10}$ and p a probability distribution on $\mathbb{Z}_{10}$. Note that $\mu(x) = 1/10$ for $x \in G$. Define p by:

$$p(0) = p(1) = p(2) = p(3) = 1/4$$

and

$$p(4) = p(5) = \cdots = p(9) = 0.$$

Then

$$\|p - \mu\|_{TV} = \frac{1}{2}\left(4\left|\frac{1}{4} - \frac{1}{10}\right| + 6\left|0 - \frac{1}{10}\right|\right) = \frac{3}{5}.$$

Moreover, to illustrate this computation for a convolution, we compute the total variation of the convolution $p^{*2} = p * p$:

$$p^{*2}(6) = p^{*2}(0) = \frac{1}{16},\ p^{*2}(1) = p^{*2}(5) = \frac{2}{16},$$

$$p^{*2}(2) = p^{*2}(4) = \frac{3}{16},\ p^{*2}(3) = \frac{4}{16},$$

with all other probabilities being 0. Thus,

$$\begin{aligned}\|p^{*2} - \mu\|_{TV} =& \frac{1}{2}\left(2\left|\frac{1}{16} - \frac{1}{10}\right| + 2\left|\frac{2}{16} - \frac{1}{10}\right| + 2\left|\frac{3}{16} - \frac{1}{10}\right| + \left|\frac{4}{16} - \frac{1}{10}\right| + 3\left|0 - \frac{1}{10}\right|\right) \\ =& \frac{3}{8}.\end{aligned}$$

In the example, we compute the distance of the given distribution from the uniform distribution on $\mathbb{Z}_{10}$. As observed above, p^{*2} is closer to uniform than p since $d_{TV}(p^{*2}, \mu) < d_{TV}(p, \mu)$, which illustrates an important result for repeated convolutions.

Theorem 4.5. *Let p be a probability distribution on a finite group G with associated*

random walk. Let μ be the uniform distribution on G. Then

$$d_{TV}(p^{*k+1}, \mu) \leq d_{TV}(p^{*k}, \mu)$$

Proof.

$$\begin{aligned}
d_{TV}(p^{*k+1}, \mu) &= \frac{1}{2} \sum_{x \in G} \left| p^{*k+1}(x) - \frac{1}{|G|} \right| \\
&= \frac{1}{2} \sum_{x \in G} \left| \left(\sum_{y \in G} p^{*k}(y) p(y^{-1}x) \right) - \frac{1}{|G|} \right| \\
&= \frac{1}{2} \sum_{x \in G} \left| \sum_{y \in G} (p^{*k}(y) p(y^{-1}x)) - \sum_{y \in G} \frac{1}{|G|} p(y^{-1}x) \right| \\
&\leq \frac{1}{2} \sum_{x \in G} \sum_{y \in G} \left| p^{*k}(y) - \frac{1}{|G|} \right| p(y^{-1}x) \\
&= \frac{1}{2} \sum_{y \in G} \left| p^{*k}(y) - \frac{1}{|G|} \right| \\
&= d_{TV}(p^{*k}, U)
\end{aligned}$$

By an induction process, we can conclude that for any $j > k$ we have

$$d_{TV}(p^{*j}, \mu) \leq d_{TV}(p^{*k}, \mu).$$

□

For an ergodic Markov chain on a finite group, the uniform distribution is the unique stationary distribution, hence the limiting distribution. Thus, $P^n(x, y) \to \frac{1}{|G|}$ as $n \to \infty$, and therefore

$$p^{*k}(x) \to \mu(x) \text{ as } k \to \infty,$$

implying that

$$d_{TV}(p^{*k}, \mu) \to 0 \text{ as } k \to \infty.$$

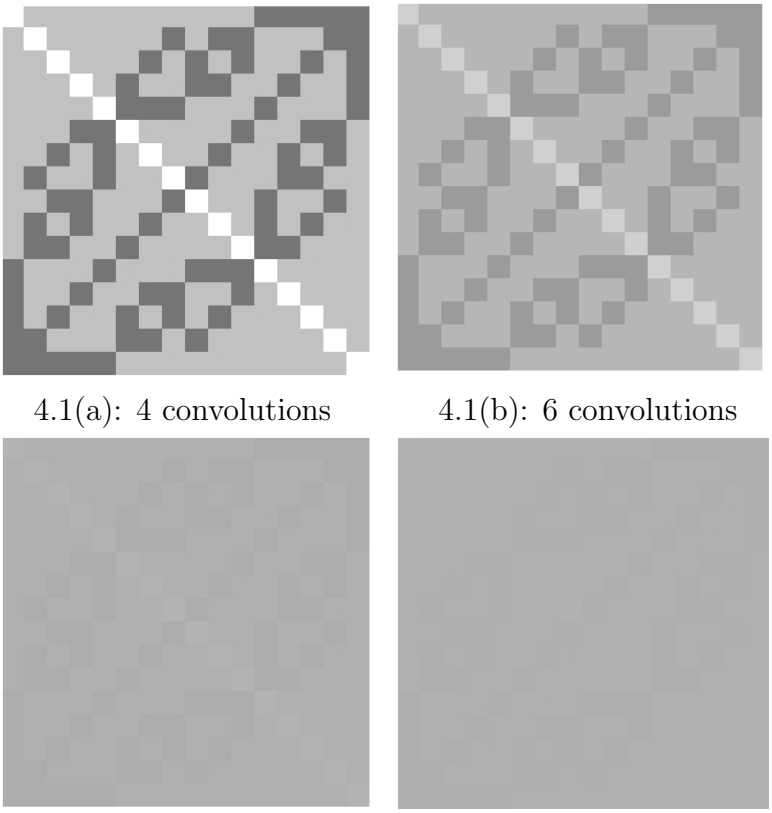
We have discussed the $n-$dimensional cube, which, in fact, makes a great example for observing the convergence to uniform. Figure 4.1 was created using powers of the transition matrix, which were converted to an image where shades of gray are based on the value of each matrix entry. In this scale, 0 is black and 1 is white. As the powers of the transition matrix increase, the rows of the matrix converge to the uniform distribution. This occurrence is observed as a uniform gray is approached. We can compare Figure 4.1 to Figure 4.2, which are image representation of the convolutions of the transition matrix for the $4-$cube and $7-$cube, respectively.

4.2 Stopping Rules and Stopping Times

To begin the discussion of stopping rules and stopping times, consider the following example. It will provide a useful reference when considering definitions that follow.

Example 4.6. Consider the $n-$dimensional hypercube with associated random walk as described in the previous chapter. A stopping rule is a rule that specifies when to end the walk. Suppose our stopping rule says to stop the walk on the hypercube once all coordinates have been changed at least once. The point at which the stopping rule is satisfied is called the stopping time for the walk. In our example, the time when the final unchanged coordinate is changed would be the stopping time.

The stopping rule presented above is a bit too relaxed for our interests, and we wish to impose stricter requirements on a stopping time for our purposes. We are not only interested in a stopping time, but we need one that provides equal likelihood for all possible positions of the walk.



4.1(a): 4 convolutions

4.1(b): 6 convolutions

4.1(c:) 10 convolutions

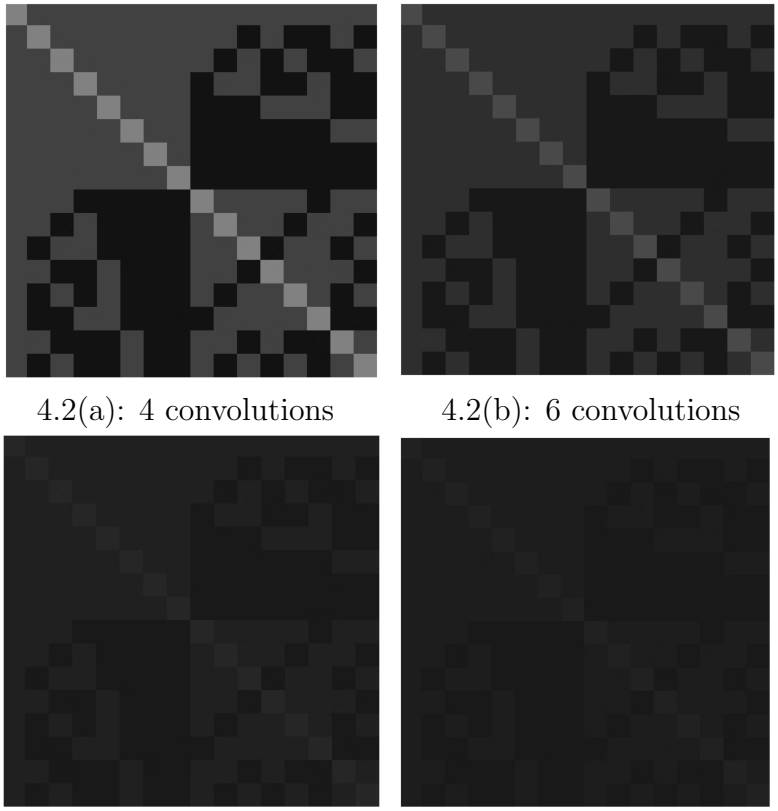
4.1(d): 12 convolutions

Figure 4.1: The 4−cube convergence to uniform.

Definition 4.7. Consider a random walk on a finite group G. We define a *stopping rule* to be a rule that determines when to stop the walk. If the walk is terminated at time T, then T is called a *stopping time*. Note that T is a random variable. For $k \in \mathbb{N}$ and $x \in G$, the time T is considered to be a *strong uniform time* if

$$\mathbf{P}(T = k \text{ and } X_k = x)$$

does not depend on x. In other words, the probability that k is a strong uniform time does not depend on the final location of the walk on G. Equivalently, T is a strong



4.2(a): 4 convolutions 4.2(b): 6 convolutions

4.2(c:) 10 convolutions 4.2(d): 12 convolutions

Figure 4.2: The 7−cube convergence to uniform.

uniform time if for each $k \in \mathbb{N}$

$$\mathbf{P}(X_k = x | T = k) = 1/|G|.$$

4.3 Upper Bound on Total Variation

We have established that for repeated convolutions of a probability distribution p on a finite group G, as k increases, p^{*k} gets closer to uniform. Just how quickly this convergence occurs depends on the transition matrix.

The following theorem establishes an upper bound on total variation that is related to stopping times.

Theorem 4.8. *Let p be a probability distribution on a group G with uniform distribution μ. Let T be a strong uniform time for a random walk on G driven by p. Then*

$$d_{TV}(p^{*k}, \mu) \leq \boldsymbol{P}(T > k), \text{ for all } k \geq 0.$$

Proof. Let $A \subseteq G$. Then

$$\begin{aligned}
p^{*k}(A) &= \sum_{x \in A} p^{*k}(x) \\
&= \mathbf{P}(X_k \in A) \\
&= \sum_{j \leq k} \mathbf{P}(X_k \in A \text{ and } T = j) + \mathbf{P}(X_k \in A \text{ and } T > k) \\
&= \sum_{j \leq k} \mu(A)\mathbf{P}(T = j) + \mathbf{P}(X_k \in A | T > k)\mathbf{P}(T > k) \\
&= \mu(A)P(T \leq k) + \mathbf{P}(X_k \in A | T > k)\mathbf{P}(T > k) \\
&= \mu(A)(1 - \mathbf{P}(T > k)) + \mathbf{P}(X_k \in A | T > k)\mathbf{P}(T > k) \\
&= \mu(A) + \mathbf{P}(T > k) \left(\mathbf{P}(X_k \in A | T > k) - \mu(A)\right)
\end{aligned}$$

We have

$$p^{*k}(A) = \mu(A) + \mathbf{P}(T > k) \left(\mathbf{P}(X_k \in A | T > k) - \mu(A)\right).$$

Subtracting $\mu(A)$ yields

$$p^{*k}(A) - \mu(A) = \mathbf{P}(T > k) \left(\mathbf{P}(X_k \in A | T > k) - \mu(A)\right).$$

Note that $\left|\sum_{x \in A}(p^{*k}(x) - \mu(x))\right| = \left|p^{*k}(A) - \mu(A)\right|$. Thus we conclude

$$d_{TV}(p^{*k}, \mu) \leq \mathbf{P}(T > k).$$

□

By using an upper bound like this one, we are able to estimate total variation for a random walk on $\mathbb{Z}_2^n$. Note that this upper bound is simply an inequality, and there is the chance of error with the estimate. Of course some estimates are better than others.

We analyze the following model for a random walk on the $5-$cube. For each step of the walk, we randomly choose a coordinate p of the vector, and then randomly reassign coordinate p to 0 or 1. Under this model, there is probability of $1/2$ that the random walk remains in its current location, and there is probability of $\frac{1}{10}$ that the walk moves to each of the neighboring vertices.

The first simulation uses a stopping time beginning with the zero vector $\mathbf{s}$ and a randomly generated vector $\mathbf{v}$. The two vectors proceed independently, and we stop the walk once they land in the same position. This defines a strong uniform stopping time. The stopping times for 200 trials were recorded, and we produced a cumulative frequency graph.

The second method we tested was a coupling method. In this process, we start with the zero vector $\mathbf{s}$ and a randomly generated vector $\mathbf{v}$. At each time, we pick a random coordinate p and change coordinate p of both $\mathbf{s}$ and $\mathbf{v}$ to the same value (either 0 or 1 chosen randomly). Once changed, coordinate p remains the same for both vectors.

As discussed in Theorem 4.8, the cumulative frequency $\mathbf{P}(T > k)$ is an upper bound for the total variation distance $\|p^{*k} - \mu\|$ between the probability distribution of the random walk after k steps and the uniform distribution.

In Figure 4.3, the top plot shows the first method using a stopping time. The middle plot shows method two, the coupling method. The bottom graph shows the

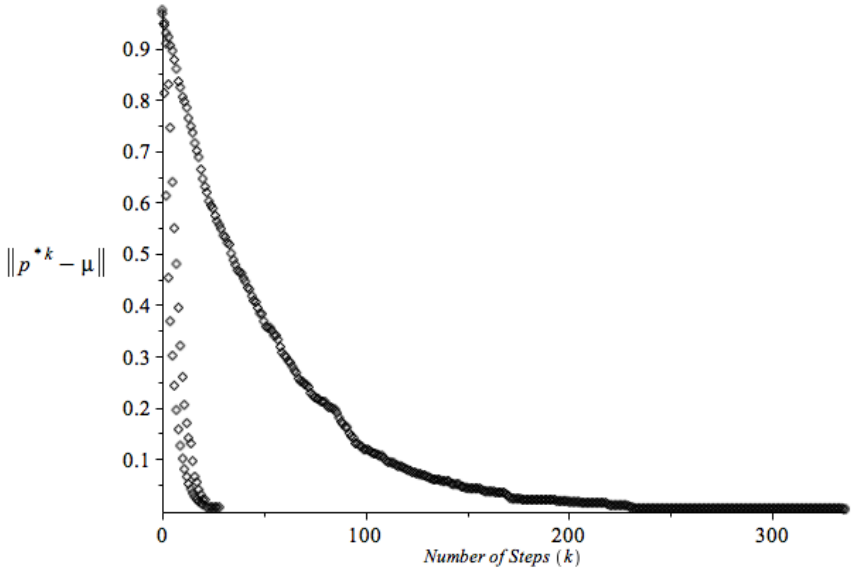


Figure 4.3: Comparison of upper bound estimates for the 5−cube.

actual total variation calculated through powers of the transition matrix. Clearly the coupling method is the better estimate for total variation. Figure 4.4 shows a graph containing the coupling method and total variation calculation only.

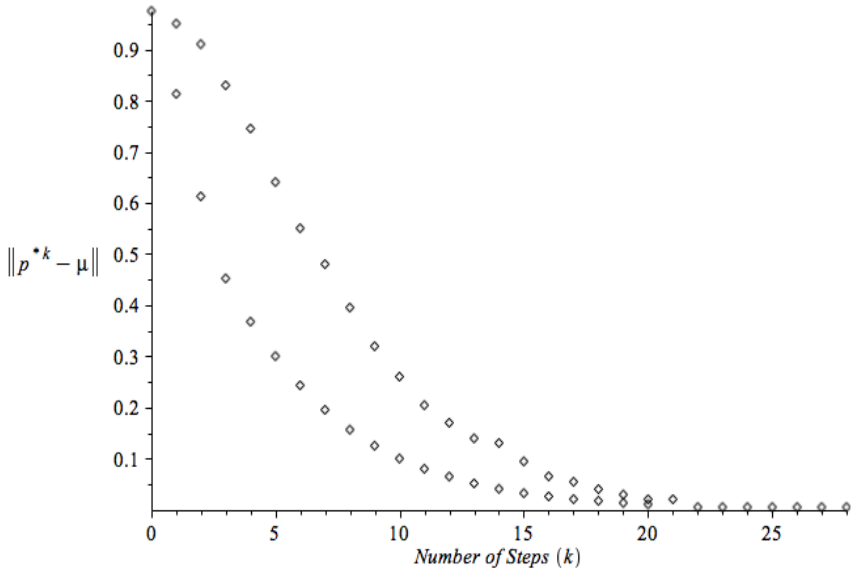


Figure 4.4: Coupling estimate compared to true total variation.

CHAPTER 5: Cutoff Phenomenon and Analysis of Shuffles

Recall Figure 1.1. The graph represents the total variation for a shuffling model. Notice that the graph remains constant for a time, and then we observe a sharp decline in $d_{TV}(p^{*k}, \mu)$. We identified this as a cutoff in Chapter 1, and in this chapter, we will elaborate on this notion. The discussion concludes with analysis of some examples of shuffling procedures.

5.1 Cutoff Phenomenon

We established in Chapter 2 that some Markov chains exhibit a stationary distribution. For shuffling a deck of cards, the Markov chains we are interested in are those for random walks on the group S_{52}. Since the group S_{52} is so large, simulations are almost impossible unless we use stopping times. These random walks are ergodic and exhibit a unique stationary distribution, which is the uniform distribution on S_{52}. Since the random walk shows a unique stationary distribution, as the number of shuffles increases, the probability distribution on the group approaches uniform. This convergence is slow for a while but then increases rapidly, showing a sharp decline in the total variation between between the probability distribution of the shuffled deck and the uniform distribution. This occurrence is called the *cutoff phenomenon*, and it shows that after an appropriate number of shuffles, the total variation between the shuffled deck and the uniform distribution is arbitrarily small, and hence, the deck is sufficiently shuffled.

5.2 Top-to-Random Shuffle

Consider a deck of n cards. The top-to-random shuffle is performed by taking the top card of the deck and inserting it at a random location within the remaining $n-1$ cards. With $n-1$ cards, there are n possible locations for the top card to be inserted, and thus the probability that the top card will be inserted at any given location (including back at the top) is $1/n$ for each position.

The number of times this shuffle must be performed before the deck is random is the topic of interest. To determine this, we need a strong uniform time. A strong uniform time for this shuffle is one step after the original bottom card has reached the top of the deck.

Proposition 5.1. *Define a stopping rule for the top to random shuffle to be one step after the original bottom card reaches the top of the deck. This stopping rule defines a stopping time T, which is a strong uniform time.*

Proof. Suppose that the top to random shuffle is performed on a deck of n cards. Let T be the stopping time associated with the stopping rule from the proposition. At some time, a card will be inserted below the original bottom card. When a second card is inserted below the original bottom card, it is equally likely to be placed in any of the two possible locations. Continuing in this manner, we find that when the k^{th} card is inserted below the original bottom card, the inserted card has equal probability of being inserted in any of the k locations below the original bottom card. Therefore any arrangement of cards below the original bottom card is equally likely. By the time that the original bottom card reaches the top of the deck, all $(n-1)!$ arrangements of the other cards are equally likely. The original bottom card is then taken from the top of the deck and inserted in a random position with equal probability on all positions, thus making all $n!$ arrangements of the cards equally likely. Hence T is a

strong uniform time. □

Proposition 5.2. *The top to random shuffle is ergodic with the uniform distribution as a stationary distribution.*

Proof. Consider the proof of Proposition 5.1. In this proof we showed that after a sufficient number of steps, all arrangements of the deck are equally likely. Thus, one makes the observation that the support Σ of the random walk must generate the group S_n. Furthermore, the probability that a shuffle corresponds to the identity is positive, thus $p(e) > 0$. For this reason, Σ is not contained in a coset of a proper normal subgroup of S_n (as illustrated in Example 3.12 and Example 3.13). Referencing Theorem 3.11, the random walk for this shuffle satisfies the requirements of the theorem, so the walk is ergodic on S_n. □

5.3 Riffle and Reverse Riffle

To determine the number of shuffles needed to make a deck of cards random, we need a model that accurately represents the shuffling process that humans use. The model that we use is the GSR model, which was outlined in Chapter 1.

Definition 5.3. Consider a deck of n cards. The *GSR* model is defined in the following way. The cards are cut into two packets, A and B, according to the binomial distribution. The probability that packet A contains c cards is given by $\binom{n}{c}2^{-n}$. So the deck is separated into two packets, A and B, with c and $n-c$ cards, respectively. Next, the packets are riffled together as cards are dropped one at a time from each packet with probabilities corresponding to the number of cards remaining in each packet. This means that the probability of dropping the first card from packet A is given by c/n, and for packet B the probability is $(n-c)/n$. The sizes of packets A and

B will be altered throughout the shuffling process, resulting in changing probabilities after each card is dropped.

Remark 5.4. Once the riffle shuffle is completed, the ordering within each packet is unchanged.

Remark 5.5. The top to random shuffle is a shuffle that could result from the GSR model. The top to random shuffle occurs as a riffle shuffle when packet A contains a single card. Then, for $c \leq n-1$, a total of c cards are dropped from packet B, followed by the single card from packet A, and then the remaining $n-c-1$ cards from packet B. For this reason, properties of the riffle shuffle also apply to the top to random shuffle.

We wish to determine how many steps are necessary in the random walk to obtain a random deck. Since $d_{TV}(p^{*k}, \mu) \to 0$, the concern now is to determine what k is sufficient to merit randomness for $G = S_{52}$. Thus, in order to determine the desired number of shuffles for a deck of cards, we use the upper bound

$$d_{TV}(p^{*k}, \mu) \leq \mathbf{P}(T > k),$$

which was given in Chapter 4. A strong uniform time is used to estimate the total variation between our shuffled deck and the uniform distribution. In order to do this, we use what is known as the reverse riffle shuffle since the reverse riffle shuffle offers a more straightforward model for simulation. We now define the reverse riffle shuffle and describe the strong uniform time used in our estimate.

Definition 5.6. Consider a deck of n cards. Randomly assign a value of 0 or 1 to each card. Next, place the cards labeled with a 0 on the top of the deck and the cards

labeled with a 1 on the bottom of the deck, in ascending order from top to bottom. This process defines the *reverse riffle* shuffle.

Remark 5.7. Once the reverse riffle shuffle is completed, the cards labeled 0 have the same relative ordering that they had before the shuffle. Likewise for the cards labeled 1.

Proposition 5.8. *The reverse riffle shuffle has the same probability distribution as the riffle shuffle.*

Proof. Consider a deck of n cards. The first step in a riffle shuffle is to divide the deck into two packets A and B of size c and $n-c$, respectively. As this is done, label the cards A or B, according to the packet. To perform the shuffle, the two packets are interleaved, as cards are dropped from each packet. The probability that a card is dropped from packet A is proportional to the number of cards remaining in each, and likewise for the probability that the next card dropped is from packet B.

Using labels A and B (rather and 0 and 1) for the reverse riffle, the probability that c cards are labeled A is given by

$$\mathbf{P}(|A| = c) = \frac{1}{2^n}\binom{n}{c}.$$

Note that this corresponds to the probability that packet A has c cards in the riffle shuffle. For the reverse riffle, we have c cards labeled A and $n-c$ cards labeled B. Working up from the bottom of the deck, there is probability c/n that the bottom card is labeled A and probability $(n-c)/n$ that it is labeled B. Continuing up the deck, the probability that the subsequent card is labeled A is proportional to the number of cards remaining in each packet. Similarly for cards labeled B.

Suppose that, in the riffle shuffle, packet A contains c cards. Then the probability

that the bottom card is labeled A in the reverse riffle is the same as the probability that the bottom card was dropped from packet A in the riffle shuffle. Working from the bottom up, the probability that a later card is labeled A in the reverse riffle shuffle is the same as the probability that the corresponding card came from packet A in the riffle shuffle. Hence the probability distributions are the same for each of these shuffles. □

As the reverse riffle shuffle has the same probability distribution as the riffle shuffle, results from the analysis of the reverse riffle shuffle also apply to the riffle shuffle. As mentioned, the reverse riffle shuffle is a less complicated model to analyze, and for this reason, it is the preferred model to use. To estimate the necessary number of shuffles, we use an upper bound, but in order to use this upper bound, we need a strong uniform time.

Working with a deck of n cards, suppose that the reverse riffle shuffle has been completed one time. Then each card has been assigned a value of either 0 or 1. The $0'$s have been placed on the top of the deck and the $1'$s have been placed on the bottom. For the next shuffle, new values of 0 or 1 are randomly assigned to each card, keeping track of the original number that was assigned in the first shuffle. Then the deck is re-ordered (based on the new numbers assigned) by placing cards labeled 0 on top and cards labeled 1 on the bottom. This process is shown below, completed for 2 shuffles.

Cards	**Random 0, 1**	**Re-order**	**Random 0,1**	**Re-order**
1	$1-0$	$1-0$	$1-0,0$	$1-0,0$
2	$2-1$	$4-0$	$4-0,1$	$6-0,0$
3	$3-1$	$6-0$	$6-0,0$	$7-0,0$
4	$4-0$	$7-0$	$7-0,0$	$2-1,0$
5	$5-1$	$2-1$	$2-1,0$	$5-1,0$
6	$6-0$	$3-1$	$3-1,1$	$9-1,0$
7	$7-0$	$5-1$	$5-1,0$	$4-0,1$
8	$8-1$	$8-1$	$8-1,1$	$3-1,1$
9	$9-1$	$9-1$	$9-1,0$	$8-1,1$
10	$10-1$	$10-1$	$10-1,1$	$10-1,1$

Reverse Riffle Shuffle.

Proposition 5.9. *Define a stopping rule for the reverse riffle shuffle as the first time when there are no repeated histories of card locations. The stopping time associated with this stopping rule is a strong uniform time [1].*

Proof. Let T be the stopping time defined above, and consider the history of card locations to be vectors. After each shuffle, these vectors are sorted in lexicographic order from right to left. This can be seen in the example above. When time T is reached, all vectors are distinct and are equally likely to be associated with a given card. Hence all $n!$ arrangements of the deck are equally likely, so T is a strong uniform time. □

Since T is a strong uniform time, it can be used as an upper bound estimate for total variation distance. Using Maple, we estimate the total variation distance of the distribution of a deck of n cards after k shuffles from the uniform distribution. Figure 5.2 presents this curve, which includes the cutoff phenomenon, for a deck of 52 cards. The program can be found in Appendix A.

In [1] Aldous and Diaconis explicitly calculate the upper bound on total variation

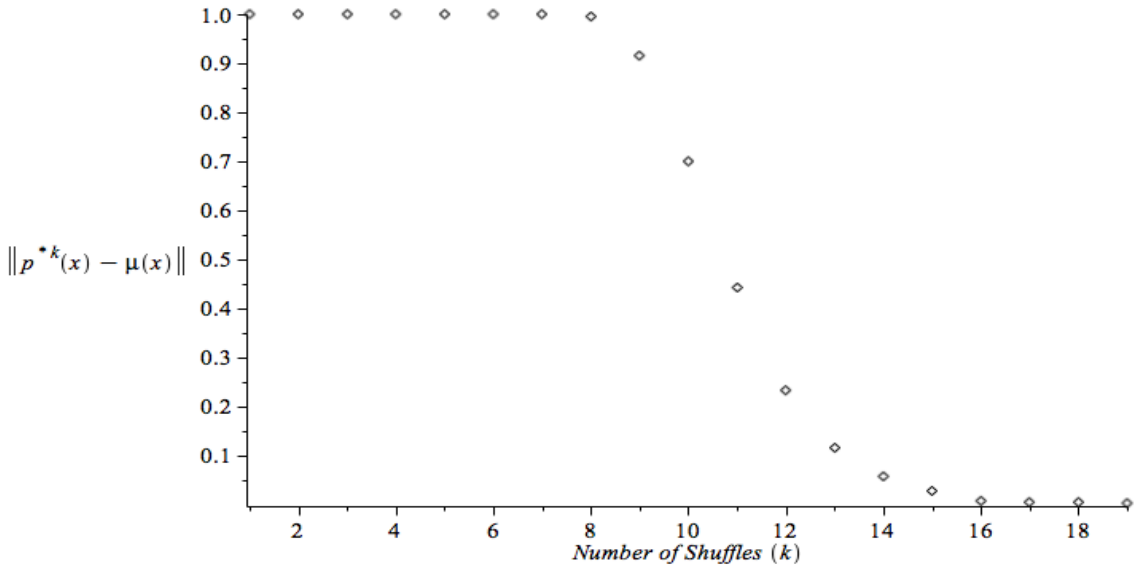


Figure 5.1: Total variation and cutoff.

distance for a deck of 52 cards, and those results are given below:

Number of Shuffles	**Upper Bound**
10	0.73
11	0.48
12	0.28
13	0.15
14	0.08

REFERENCES

[1] Aldous, D. and Diaconis, P. (1986). Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348.

[2] Bayer, D. and Diaconis, P. (1992). Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313.

[3] Ceccherini-Silberstein, T., Scarabotti, F., and Tolli, F. (2008). *Harmonic Analysis on Finite Groups*. Cambridge University Press.

[4] Diaconis, P. (1996). The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences of the United States of America*, 93:1659–1664.

[5] Diaconis, P. (2011). The mathematics of mixing things up. *Journal of Statistical Physics*, 144:445–458.

[6] Friedberg, S. H., Insel, A. J., and Spence, L. E. (2002). *Linear Algebra*. Pearson, 4th edition.

[7] Gallager, R. G. (2014). *Stochastic Processes: Theory for Application*. Cambridge University Press.

[8] Grinstead, C. M. and Snell, J. L. (1997). *Introduction to Probability*. American Mathematical Society.

[9] Haggstrom, O. (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press.

[10] Hildebrand, M. (2005). A survey of results on random walks on finite groups. *Probability Surveys*, 2:33–63.

[11] Saloff-Coste, L. (2004). Random walks on finite groups. In Kesten, H., editor, *Probability on Discrete Structures*, volume 110, chapter Random Walks on Finite Groups, pages 263–346. Springer Berlin Heidelberg.

APPENDIX A: Maple programs

A.1 Random Walk on the Hypercube

```
with(LinearAlgebra) : with(ListTools) : with(ImageTools) : with(plots) :
dim := 7 :
> Sets the dimension for the group G, the hypercube
elements :=proc(G)local H, J;
 H := map(G→[op(G), 0], G) :
 J := map(G→[op(G), 1], G) :
 sort([op(H), op(J)], length) :
 end proc:
> Generates the elements of G
Pr :=proc(dim)local A, n; global G;
 A := [[0], [1]];
 for n from 1 to dim − 1 do;
 A := elements(A);
 end do; end proc:
> Sorts the elements of G in ascending order
G := Pr(dim) :
nops(%) :
s := G[1] :
mu :=proc(x) local s :
 s := add(x[i], i = 1 ..dim) :
 if s = 0 then 1/(dim + 1) else if s = 1 then 1/(dim + 1) :
 else 0 :
 end if: end if: end proc:
> Creates probabilities for the walk on G
map(mu, G) :
p := (x, y)→mu(G[x] − G[y]mod 2) :
A := Matrix(2^dim, 2^dim, p) :
> mu2 := y→add(mu(y − G[i]mod 2)·mu(G[i]), i = 1 ..2^dim) :
> Probabilities for trantisions after 2 steps
> map(mu2, G) :
> mu1 := map(mu, G) :
> MatrixPower(A, 2) :
```

```
> cmu := nu→
  map(y→add(mu(y + G[i] mod 2)·nu[i], i = 1..2^dim), G):
> Defines the convolution
> cmu(mu1):
> Convolution of mu1 with itself
> cmu(%):
> cmu(cmu(mu1)):
> 2 convolutions of mu1 with itself
> tv := u→evalf(1/2·add(abs(u[i] − 1/2^dim), i = 1..2^dim)):
> Calculates the total variation
> N := 30:
> L := [A]:
  for n from 1 to N − 1 do:
  L := [op(L), A.L[−1]]:
  end do:
> Creates a list of matrices
R := map(x→Row(x, 1), L):
> Converts elements of the matrix into a vector
t := map(tv, R):
> Calculates total variation between elements of
  R and the uniform distribution
> pts := [seq([n, t[n]], n = 1..N)]:
> Plots the total variation
> pointplot(pts, view = 0..1):
>
> Am := m→
  Create(159, 159, (x, y)→10· evalf(m[floor(x/10 + 1), floor(y/10 + 1)])):
> The procedure Am(m) creates an image from the matrix m
> View([Am(L[4]), Am(L[6]), Am(L[10]), Am(L[12]), Am(L[20])])
> We see that the distribution approaches a uniform gray
```

A.2 Total Variation Estimates for the Hypercube

```
> with(LinearAlgebra) : with(ListTools) : with(plots) :
> dim := 5 :
> Sets the dimension for the cube
> r := rand(0..dim) :
> Random number generator
> mo :=proc(s) local p, q, l, a :
    p := rand(1..nops(s))( ) :
    l := s :
    q := rand(0..1)( ) :
   l[p] := q :
    l;
    end proc:
> The procedure defining the walk
> r2 := rand(0..1) :
> Random number generator
> rv := dim → [seq(r2( ), k = 1..dim)] :
> Generated a vector of 0's and 1's
> trial :=proc(dim) local s1, s2, count, n :
    s1 := [seq(0, k = 1..dim)];
    s2 := rv(dim);
    count := 0 :
    for n from 1 to dim^4 while not(s1 = s2) do:
    count := count + 1;
    s1 := mo(s1);
    s2 := mo(s2);
    end do;
    count;
    end proc:
> Procedure for method number 1
> N := 200 :
> Number of trials
> L := [seq(trial(dim), n = 0..N)] :
> List of stopping times for method 1
```

```
> M := max(L) :
> Max of stopping times for method 1
> fr := n→nops( [SearchAll(n, L)] ) :
> Frequencies of stopping times for method 1
> frL := [seq(fr(n), n = 1..M)] :
> List of frequencies for method 1
> cumf := n→add( fr(k)/N , k = n + 1..M ) :
> Cumulative frequencies of stopping times
> trialpts := [seq( [n, cumf(n)], n = 0..M − 1 )] :
> Creates points for method 1
> pointplot(trialpts) :
> Plot of stopping times for method 1
> ------------------------------------------------
> alt := proc( ) local s1, s2, count, n, p, q :
    s1 := [seq(0, k = 1..dim)];
    s2 := [seq(r2( ), k = 1..dim)];
    count := 0 :
    for n from 1 to dim^4 while not(s1 = s2) do;
   count := count + 1;
    p := rand(1..dim)( ) :
    q := r2( ) :
    s1[p] := q;
    s2[p] := q;
    end do;
    count;
    end proc:
> Procedure for coupling method
> L := [seq(alt( ), n = 1..N)] :
> List of stopping times for method 2
> M := max(L) :
> Max of stopping times for method 2
> [seq(fr(n), n = 0..M)] :
> Frequencies of each stopping time
> altpts := [seq( [n, cumf(n)], n = 0..M − 1 )] :
> Creates points for method 2
```

```
> pointplot(altpts) :
> Plot of stopping times for method 2
> ----------------------------------------------------
elements :=proc(G)local H, J;
 H := map(G→[op(G), 0], G) :
 J := map(G→[op(G), 1], G) :
 sort([op(H), op(J)], length) :
 end proc:
> Process for generating the vectors of
  0's and 1's
> Pr :=proc(dim) local A, n global G;
   A := [[0], [1]];
   for n from 1 to dim − 1 do;
   A := elements(A);
   end do; end proc:
> Generates vectors based on a given dimension
> G := Pr(dim) :
> Generates elements for G
> G := sort(G, length) :
> Sorts the elements of G
> mu :=proc(x) local s :
   s := add(x[i], i = 1 ..dim) :
   if s = 0 then 1/2 :
   else if  s = 1 then 1/(2·dim) :
   else 0 :
   end if: end if: end proc:
> Defines transition probabilities
> map(mu, G) :
> Creates a vector of the probabilities
> p := (x, y)→mu(G[x] − G[y] mod 2) :
> Defines the probability of going from x to y
> P := Matrix(2^dim, 2^dim, p) :
> Creates a matrix based on the dimension and
  entries of p from above
```

```
> tv := u -> 1/2 * add( abs( u[i] - 1/2^dim ), i = 1..2^dim ) :
> Defines total variation between u and uniform
> L := [ seq( MatrixPower(P, n), n = 1..20 ) ] :
> List of powers of P
> Ra := map( x -> Row(x, 1), L ) :
> Selects the first row of each power of P
  and makes a vector
> T := map( tv, Ra ) :
> Calculates total variation between each vector
  and uniform
> tp := [ seq( [n, T[n]], n = 1..20 ) ] :
> Creates points for total variation
> pointplot( tp ) :
> Plot of actual total variation
> pointplot( [ op(tp), op(cf), op(altpts) ] ) :
> Plot of total variation, method 1, coupling method
> pointplot( [ op(tp), op(altpts) ] ) :
> Plot of total variation and coupling method
```

A.3 Upper Bound Estimate for Reverse Riffle

```
with(LinearAlgebra) : with(ListTools) :
with(plots) :
N := 52 :
Sets the number of cards to shuffle
r2 := rand(0..1) :
> Random number generator
stime := proc(N) local d, k, L, p0, p1, M :
d := [seq(n, n = 1..N)] :
L := [seq(r2( ), n = 1..N)];
p0 := [SearchAll(0, L)];
p1 := [SearchAll(1, L)];
d := [op(map(x → [op(d[x]), 0], p0)), op(map(x → [op(d[x]), 1], p1))];
M := map(x → subsop(1 = NULL, x), d);
for k from 1 to 100 while not (evalb(nops(M) = nops(convert(M, set)))) do:
L := [seq(r2( ), n = 1..N)];
p0 := [SearchAll(0, L)];
p1 := [SearchAll(1, L)];
d := [op(map(x → [op(d[x]), 0], p0)), op(map(x → [op(d[x]), 1], p1))];
M := map(x → subsop(1 = NULL, x), d);
end do:
nops(M[1]);
end proc:
> The shuffling procedure
> NT := 400 :
> The number of times to perform a new shuffle
> T := [seq(stime(N), n = 1..NT)] :
> The sequence of stopping times
> M := max(T) :
> The largest stopping time
> fr := [seq(nops([SearchAll(n, T)]), n = 1..M)] :
> Frequencies of each stopping time
> cfr := [seq(add(fr[k], k = j + 1..M), j = 1..M − 1)] :
> Cumulative frequencies of each stopping time
> pts := [seq([n, cfr[n]/NT], n = 1..M − 1)] :
> Points for the cumulative frequencies graph
> pointplot(pts) :
> Graph of cumulative frequencies
```