Comparing the Mixing Rates of Neat Riffle Shuffles

University of Michigan

Mateo Palomares
Mentor: Mark Conger

August 2025

Abstract

The Gilbert-Shannon-Reeds model for the riffle shuffle has been extensively analyzed since its introduction in 1955. Sharp bounds on the mixing time of the model and methods for calculating the exact variation distance to uniform have been proven. In this report, we explore a modified model of shuffling that accounts for differences in the skill of a shuffler by introducing a parameter that describes the neatness of a shuffle. We introduce two techniques to analyze the new model. These techniques simplify the problem significantly and bring us closer to showing that GSR is not the optimal level of neatness for mixing rate.

1 Introduction

The mathematics behind card shuffling has been studied for over a century. One of the earliest investigations was done by Poincaré [9], who showed that any reasonable model of card shuffling will eventually converge to the uniform distribution on all possible permutations of the deck.

Riffle shuffling is one of the most common real-life methods to shuffle cards. We are interested in how many shuffles it takes to sufficiently randomize the deck. The standard model used to study riffle shuffling is the Gilbert-Shannon-Reeds model (GSR). Bayer and Diaconis [1] found that about $\frac{3}{2}\log_2 n$ shuffles are needed to mix up a deck of n cards using this model. However, the question remains whether this accurately models real riffle shuffling done by humans. Previous REU student Cope [2] collected data on human shufflers. Another previous REU student Drouillard [4] analyzed that data and determined that the standard model does not account for the various skills of different shufflers. When experienced shufflers perform a riffle shuffle, the cards alternate neatly between each packet, whereas beginners are more likely to keep cards clumped together.

A motivating question we have is which level of neatness is better for mixing cards. We know that if the shuffle is very clumpy many cards will maintain their order and the deck will not be randomized. On the other hand, a perfectly alternating deck does not randomize the deck well either, as it becomes deterministic. In fact, if you perfectly shuffle a 52-card deck eight times in a row the deck returns to its unshuffled state. So the natural question that arises is where the sweet spot of neatness is that results in the fastest mixing rate. Intuition may suggest that GSR is this sweet spot, as it assigns equal probability to every shuffle.

To account for differences in neatness between different shufflers, we use a model suggested by Diaconis [3], and developed and analyzed independently by Drouillard [4] and Jonasson and Morris [7]. When we look at the mixing rate for small deck sizes, we find an interesting pattern. Contrary to our intuition, choosing a neatness level slightly greater than GSR results in faster mixing. There is then a region of neatness that has a comparable mixing rate, and a sharp fall-off in neatness once the shuffling becomes too deterministic. However, there have been no non-trivial results proven that compare different neatness levels for arbitrary deck sizes. We seek to prove that the patterns found in the prior analysis hold for larger deck sizes.

First, we will explain the math behind riffle shuffling and the Gilbert-Shannon-Reeds model. Then we will discuss the new model and what we currently know. Finally, we will develop new methods for analyzing this model and explore potential routes we could use to prove our desired results.

2 Background

We study card shuffling using Markov chains:

Definition 1 (Discrete Time Markov Chain)**.** For a set $\mathcal{X}$, a **discrete time Markov chain** with state space $\mathcal{X}$ and transition matrix P is a sequence of $\mathcal{X}$-valued random variables $(X_0, X_1, \dots)$ such that for all $x, y \in \mathcal{X}$, all $t \geq 1$, and all events $H_{t-1} = \bigcap_{s=0}^{t-1}\{X_s = x_s\}$ satisfying $\mathbb{P}(H_{t-1} \cap \{X_t = x\}) > 0$, we have

$$\mathbb{P}\{X_{t+1} = y \mid H_{t-1} \cap \{X_t = x\}\} = \mathbb{P}\{X_{t+1} = y \mid X_t = x\} = P(x, y)$$

This property, often called the Markov property, simply means that when at state $x \in \mathcal{X}$, the next state is chosen based on a fixed distribution $P(x, \cdot)$ that only depends on the current state. That is, the system is memoryless.

Card shuffling happens to be an example of a special case of Markov chains, called a random walk on a group:

Definition 2 (Random Walk on a Group)**.** Given a probability distribution μ on a group G, a **random walk on** G with increment distribution μ is a Markov chain with transition matrix defined as follows:

$$P(g, hg) = \mu(h)$$

for all $g, h \in G$.

To move one step on the chain, you multiply the current group element on the left by another group element, choosing which group element according to the increment distribution. Then, we can simply write $P^k(g) := P^k(e, g)$ for the probability that the current state is g after k steps of the walk if we start at the identity. Any method for shuffling cards can be modeled as a random walk on the symmetric group S_n.

3 Riffle Shuffle

There are many methods of shuffling that have been studied mathematically. Two of the most popular methods are overhand shuffling and riffle shuffling. The overhand shuffle has been shown by Pemantle [8] to take thousands of repetitions in order to achieve a sufficient level of randomness, whereas riffle shuffling seems to be much more practical. In this paper we will discuss the riffle shuffle. To perform a riffle shuffle, we split the deck into two packets and interleave the two, maintaining the relative order of cards within the two packets.

We can describe any riffle shuffle of an n-card deck as an n-bit binary sequence. First, let c be the number of 1s in the sequence. Cut c cards from the top of the deck. For each bit reading left to right, drop the card from the top packet if the bit is a 1 and from the bottom packet if the bit is a 0.

Example: For the shuffle sequence "101100", start by cutting into three 0s and three 1s.

Card	1	2	3	4	5	6
Packet	0	0	0	1	1	1

Then, shuffle the cards into the given binary sequence:

Card	4	1	5	6	2	3
Packet	1	0	1	1	0	0

To model the process as a random walk on a group, we can define which of these 2^n possible shuffles result in which permutations.

Definition 3 (Riffle Shuffle Map)**.** The **riffle shuffle map** $S : \{0, 1, \ldots, 2^n - 1\} \to S_n$ maps a riffle shuffle (encoded as its binary shuffle sequence) to the permutation that the shuffle leads to. So for a shuffle $s \in \{0, \ldots, 2^n - 1\}$ written $b_1 b_2 \ldots b_n$ in binary with $c = |\{i : b_i = 1\}|$. We define the resulting permutation $S(b_1, b_2, \ldots, b_n) = \sigma$ as follows;

$$\sigma(i) := \begin{cases} |\{j \leq i : b_j = 0\}| & b_i = 0 \\ (n - c) + |\{j \leq i : b_j = 1\}| & b_i = 1 \end{cases}$$

This is equivalent to simply performing the shuffle described by the binary sequence.

An important point is that this map is neither surjective nor injective. To illustrate this for surjectivity, it is helpful to introduce an invariant of riffle shuffling.

Definition 4 (Rising Sequence)**.** A **rising sequence** in a permutation is any maximal consecutive sequence of elements that maintain their order in a permutation. For example, in the permutation $(0, 1, 3, 4, 2, 5)$ there are two rising sequences: $0, 1, 2$ and $3, 4, 5$.

Because the top and bottom packets maintain their relative order, they each form a rising sequence. Therefore a permutation resulting from a riffle shuffle can have at most two rising sequences. This means for $n > 2$ the reversed permutation $(n, n - 1, \ldots, 0)$, having n rising sequences, cannot be reached after a single riffle shuffle.

In the case of injectivity, note that any permutation with only 0s followed by 1s will result in the identity permutation, since all cards from the bottom packet are dropped before cards from the top packet. In fact, this is only the case for the identity permutation. Any permutation with two rising sequences will be the result of a unique interleaving of the top and bottom packets.

We can define a model for riffle shuffling by specifying the probability of performing a given binary shuffle sequence.

Definition 5 (Riffle Shuffle Random Walk on S_n)**.** For a given probability distribution ρ on all possible shuffles $\{0, 1, \ldots, 2^n - 1\}$, the **riffle shuffle random walk** on S_n induced by ρ is defined as the random walk with increment distribution

$$\mu(\sigma) = \sum_{\substack{s \in \{0, \ldots 2^n - 1\} \\ S(s) = \sigma}} \rho(s)$$

This means the probability of a permutation is the sum of the probabilities of performing shuffles that lead to that permutation.

4 Gilbert-Shannon-Reeds Model

The current accepted model used to study riffle shuffling is the Gilbert-Shannon-Reeds model (GSR) [5]. In this model, a deck of n cards is cut into two packets according to the binomial distribution. So the probability of cutting the deck after the c^{th} card is $\frac{1}{2^n}\binom{n}{c}$.

Then, we choose to drop top or bottom packet with probability proportional to the size of those packets. That is, if there are A cards in the top packet and B cards in the bottom packet, we drop a card into a new pile from the top packet with probability $\frac{A}{A+B}$ and from the bottom packet with probability $\frac{B}{A+B}$. Finally, we update the number of cards in each packet and repeat the process until every card is in the new pile. This completes one shuffle.

To calculate the probability of a given binary shuffle $b_1b_2\dots b_n$ with GSR, we can take the probability of cutting after c cards and multiply by the probability of dropping each card. Since every card must eventually be dropped from each packet, we calculate:

$$\begin{aligned}\rho_{\text{GSR}}(b_1b_2\dots b_n) &= \underbrace{\frac{1}{2^n}\binom{n}{c}}_{\text{cut probability}} \cdot \underbrace{\prod_{k=0}^{n-1}\frac{|\{i \geq k : b_i = b_k\}|}{n-k}}_{\text{drop probabilities}} \\ &= \frac{1}{2^n}\binom{n}{c}\left(\frac{c!\cdot(n-c)!}{n!}\right) = \frac{1}{2^n}\end{aligned}$$

So every binary shuffle has equal probability.

Example: To make this more clear, consider the shuffle 11001. There are $n = 5$ total cards and we cut $c = 3$, so we get

$$\begin{aligned}&\frac{1}{2^5}\binom{5}{3}\cdot\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}\cdot\frac{1}{1} \\ &= \frac{1}{2^5}\binom{5}{3}\cdot\frac{3!\cdot(5-3)!}{5!} = \frac{1}{2^5}\end{aligned}$$

Definition 6 (Gilbert-Shannon-Reeds Model)**.** The **Gilbert-Shannon-Reeds model (GSR)** for riffle shuffling an n-card deck is the riffle shuffle random walk on S_n induced by the uniform distribution on $\{0,\dots,2^n-1\}$. This gives us increment distribution

$$\begin{aligned}\mu(\sigma) &= \sum_{\substack{s\in\{0,\dots,2^n-1\}\\ S(s)=\sigma}} \frac{1}{2^n} \\ &= \begin{cases} \frac{n+1}{2^n} & \sigma = e \\ \frac{1}{2^n} & \sigma \neq e,\ \exists s \text{ such that } S(s) = \sigma \\ 0 & \text{otherwise}\end{cases}\end{aligned}$$

Dave Bayer and Persi Diaconis [1] found tight bounds on the mixing rate of GSR. This is where the popular notion that "seven shuffles suffice" comes from.

5 Geometric Model

GSR is a convenient way to analyze riffle shuffling that seems to approximate the randomness of card shuffling well for some people. However, a problem that arises is that not every riffle shuffler is the same. Beginner shufflers tend to lump cards together more often, whereas experienced shufflers will alternate drops between packets more neatly.

We generalize GSR by accounting for these different types of riffle shuffles. To choose which binary shuffle is performed, we introduce a "neatness" parameter $\alpha \in [0, 1]$. A shuffle close to $\alpha = 0$ models a very clumpy shuffle, whereas a shuffle with α near 1 models a very neat shuffle. Specifically, choose the binary shuffle performed using the following process:

1. The first bit has equal probability of being a 0 or 1.
2. Choose the next bit by switching to the other packet with probability α and dropping from the same packet with probability $(1 - \alpha)$.
3. Repeat the process until all n bits are chosen.

This gives us a probability of performing a shuffle $s = b_1b_2\dots b_n$ with a given number of switches between the top and bottom packet.

Definition 7 (Geometric Model)**.** The **Geometric model** for riffle shuffling an n-card deck with neatness parameter $\alpha \in [0, 1]$ is the riffle shuffle random walk on S_n induced by the distribution

$$\rho_\alpha(s) := \frac{1}{2}\alpha^j(1-\alpha)^{n-j-1}$$

if s is any binary shuffle $b_1b_2 \ldots b_n$ with $j = |\{i : b_i \neq b_{i+1}\}|$. For $\alpha = \frac{1}{2}$, the Geometric model is equivalent to GSR.

Previous REU student Cope [2], collected data on various shufflers. Another REU student Drouillard [4] determined based on that data that GSR does not account for neatness levels of all shufflers and analyzed mixing for small decks. REU students Lee, Lewandoski, and Reifor [6] continued this analysis. Jonasson and Morris [7] showed that for an equivalent model they call the Markovian model, the mixing time for any $\alpha \in (0, 1)$ is $O(\log^4(n))$ using an entropy technique.

6 Mixing Rate

A natural question to ask about the mathematics behind mixing a deck is how we quantify how mixed a deck is. Although there are many ways to measure this, one of the most commonly used is the total variation distance.

Definition 8 (Total Variation Distance)**.** The **total variation distance** between two distributions μ and ν is

$$\begin{aligned}\|\mu - \nu\|_{\text{TV}} &:= \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)| \\ &= \frac{1}{2}\sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|\end{aligned}$$

So after k shuffles of a deck, the total variation distance to uniform is

$$\|P^k - U\|_{\text{TV}} = \frac{1}{2}\sum_{\sigma \in S_n} \left|P^k(\sigma) - \frac{1}{n!}\right|$$

Definition 9 (Mixing Time)**.** The **mixing time** for a deck of cards is the amount of shuffles it takes for the total variation distance to drop below a given threshold $\varepsilon > 0$:

$$\tau_{\text{mix}}(\varepsilon) := \min\{k : \|P^k - U\|_{\text{TV}} < \varepsilon\}$$

or

$$\tau_{\text{mix}} := \tau_{\text{mix}}(1/4)$$

Another way to measure the rate at which the total variation distance converges to uniform is the eigenvalues of the transition matrix. Poincaré [9] showed that the largest eigenvalue is always 1, corresponding with the uniform eigenvector $\left(\frac{1}{n!}, \ldots, \frac{1}{n!}\right)$. The rest of the eigenvalues all have smaller magnitude, so we have

$$1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_{n!}| \geq 0$$

We are interested in $|\lambda_2|$ because it affects the mixing rate.

Lemma 6.1. *If P is a diagonalizable transition matrix of a random walk on S_n with eigenvalues $|\lambda_2| = |\lambda_3| = \cdots = |\lambda_\ell| > |\lambda_{\ell+1}|$,*

$$\|P^k - U\|_{TV} = C|\lambda_2|^k + O(\lambda_{\ell+1}^k)$$

Proof. With starting distribution $v = (1, 0, 0, \ldots, 0)$, the distribution after k shuffles is vP^k. So writing this as a linear combination of the eigenbasis, we have

$$\begin{aligned}vP^k - U &= a_1\lambda_1^k v_1 + \cdots + a_{n!}\lambda_{n!}^k v_{n!} - U \\ &= a_2\lambda_2^k + \cdots + a_{n!}\lambda_{n!}^k v_{n!} \\ &= a_2\lambda_2^k v_2 + \cdots + a_\ell\lambda_\ell^k v_\ell + O\left(\lambda_{\ell+1}^k\right)\end{aligned}$$

So applying this to the total variation, we get

$$\begin{aligned}
\|P^k - U\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{\sigma\in S_n} \left|\left(vP^k - U\right)(\sigma)\right| \\
&= \frac{1}{2}\sum_{\sigma\in S_n} \left|\left(a_2\lambda_2^k v_2 + \cdots + a_\ell \lambda_\ell^k v_\ell + O\left(\lambda_{\ell+1}^k\right)\right)(\sigma)\right| \\
&\leq \frac{1}{2}\sum_{\sigma\in S_n} \left|a_2\lambda_2^k v_2(\sigma)\right| + \cdots + \left|a_\ell\lambda_\ell^k v_\ell(\sigma)\right| + O\left(\lambda_{\ell+1}^k\right) \\
&= \frac{1}{2}\sum_{\sigma\in S_n} |a_2 v_2(\sigma)||\lambda_2|^k + \cdots + |a_\ell v_\ell(\sigma)||\lambda_\ell|^k + O\left(\lambda_{\ell+1}^k\right) \\
&= C|\lambda_2|^k + O\left(\lambda_{\ell+1}^k\right)
\end{aligned}$$

□

We are interested in the variation distance and studying it through the second largest magnitude eigenvalue. Specifically, we want to see how the mixing rate changes as we change the neatness parameter α of the Geometric model.

7 Eigenvalue Analysis

We are able to calculate the eigenvalues explicitly for small deck sizes $n < 9$, but for larger n the computation becomes too intensive. Figure 1 has $|\lambda_2|$ graphed for $\alpha \in [0,1]$ and $n = 3,4,5,6,7,8$ from left to right and top to bottom. As n grows larger, the eigenvalues seem to converge to a similar shape. This is mostly the shape that we expect; for low neatness levels the deck doesn't get sufficiently mixed because many cards stay together, and for very high neatness levels the shuffles become too deterministic and the deck isn't sufficiently randomized. This corresponds with the second eigenvalue being close to $|\lambda_1| = 1$, resulting in a slow decay to uniformity.

Some intuition may suggest that GSR, having an equal chance for any shuffle, would have the fastest mixing. However we find that at least for small n this does not seem to be true. For every n we analyzed, there is a region with α greater than GSR where $|\lambda_2|$ continues to decrease.

Conjecture 1 (Drouillard)**.** *For $n \geq 3$, the right-hand derivative of $|\lambda_2|$ with respect to α at $\alpha = 1/2$ is*

$$\left.\frac{\partial|\lambda_2|}{\partial\alpha}\right|_{\alpha=\frac{1}{2}^+} = -\frac{1}{n-1}$$

We can compute the eigenvalues symbolically for $n = 3,4,5$ and the result holds in these cases. Approximating by calculating the eigenvalues for α values close to $1/2$, the conjecture also seems to hold for n up to 10.

In Figure 2 we can see all eigenvalues graphed on the complex plane as α changes for $n = 5$. We can see some more patterns that arise.

Conjecture 2. *For all $n \geq 3$, there is some $\varepsilon > 0$ for which all eigenvalues of P_α are real, positive, and strictly decreasing for $\alpha \in [0, 0.5 + \varepsilon]$.*

Because of the computational limitations that we reach for larger n, it is necessary to use alternative methods in order to study these conjectures or compare mixing rates of different neatness levels further.

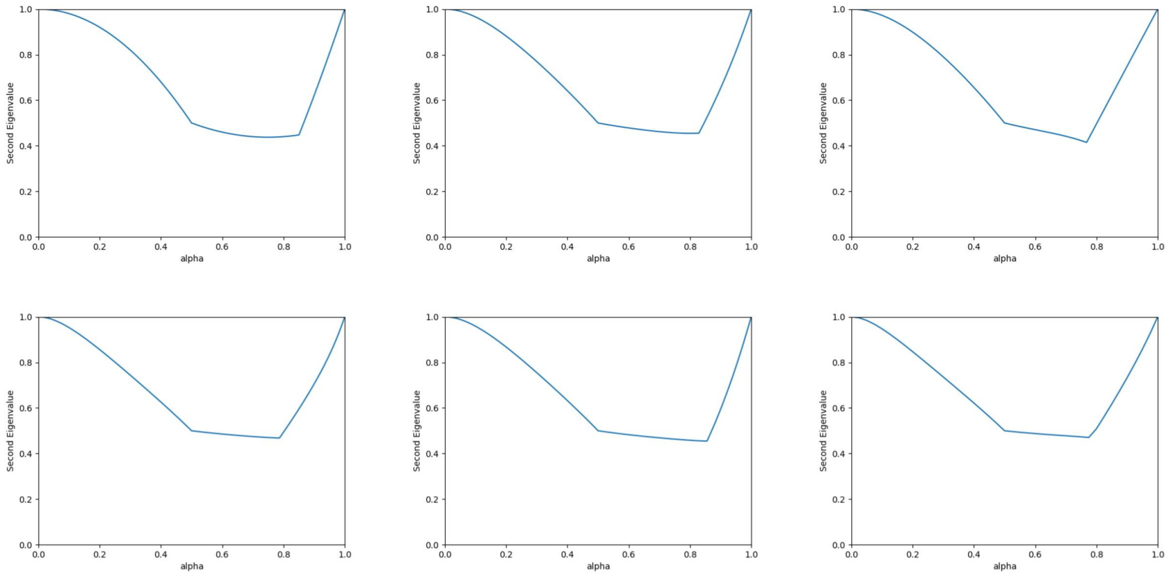

Figure 1: $|\lambda_2|$ for $n = 3, 4, 5, 6, 7, 8$

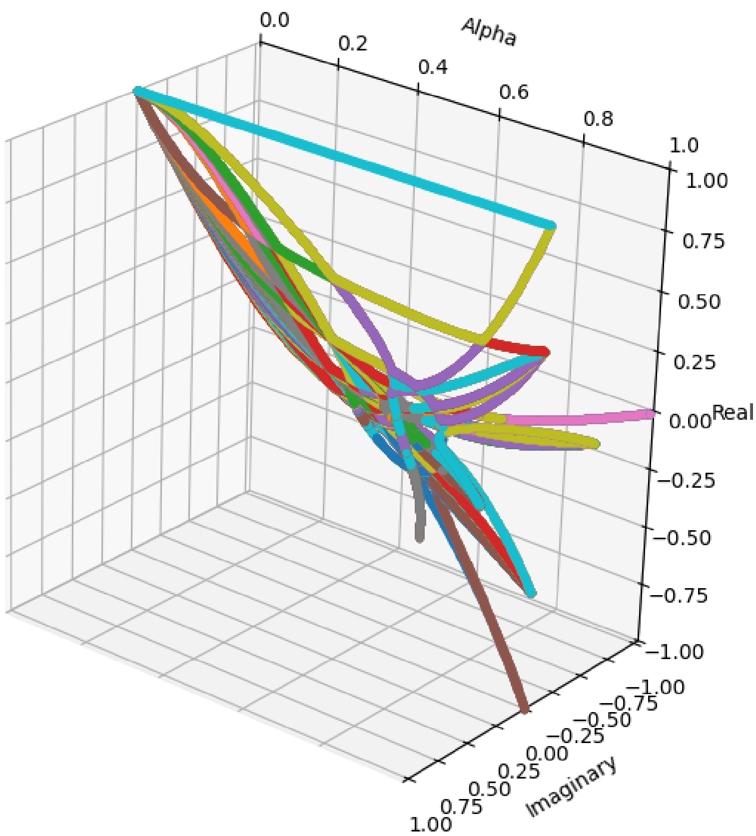

Figure 2: All eigenvalues for $n = 5$

8 New Methods

8.1 Trace Method

One way we can study the eigenvalues of the transition matrix is by using the trace. This way we can simply study the probability of being at the identity permutation after k shuffles.

Lemma 8.1.

$$|\lambda_2| = \lim_{k\to\infty} \frac{|\mathrm{tr}(P_\alpha^{k+1}) - 1|}{|\mathrm{tr}(P_\alpha^k) - 1|} = \lim_{k\to\infty} \frac{|n!P_\alpha^{k+1}(e) - 1|}{|n!P_\alpha^k(e) - 1|}$$

Proof. The diagonal entries of the transition matrix after k steps are simply $P_\alpha^k(\sigma,\sigma) = P_\alpha^k(e)$. This means the trace of the matrix is

$$\mathrm{tr}(P_\alpha^k) = n! \cdot P_\alpha^k(e).$$

Since the trace of a matrix is equal to the sum of its eigenvalues, we get

$$\mathrm{tr}(P) = n! \cdot P(e) = \sum_{i=1}^{n!} \lambda_i = 1 + \sum_{i=2}^{n!} \lambda_i.$$

It is a classical fact from linear algebra that if λ is an eigenvalue of a square matrix P, then λ^k is an eigenvalue of P^k. So we have

$$n! \cdot P^k(e) = 1 + \sum_{i=2}^{n!} \lambda_i^k$$

$$\implies \left|n! \cdot P^k - 1\right|^{1/k} = \left|\sum_{i=2}^{n!} \lambda_i^k\right|^{1/k}$$

Since each eigenvalue has magnitude less than 1, we know in the limit as k tends to infinity, the eigenvalue with the largest magnitude, λ_2 dominates the sum. Therefore,

$$|\lambda_2| = \lim_{k\to\infty} \left|\sum_{i=2}^{n!} \lambda_i^k\right|^{1/k} = \lim_{k\to\infty} \left|n! \cdot P^k(e) - 1\right|^{1/k}$$

So using some limit tests we get

$$|\lambda_2| = \lim_{k\to\infty} \frac{|n!P_\alpha^{k+1}(e) - 1|}{|n!P_\alpha^k(e) - 1|}$$

□

8.2 Total Variation Derivative

Another method we can use to show that there are neatness levels that have faster mixing than GSR is the total variation distance from uniform around $\alpha = 1/2$. First, we'll split up the variation distance into a couple parts.

$$\begin{aligned}\|P_\alpha^k - U\|_{\mathrm{TV}} &= \frac{1}{2} \sum_{\sigma\in S_n} \left|P_\alpha^k(\sigma) - \frac{1}{n!}\right| \\ &= \frac{1}{2} \sum_{\substack{\sigma\in S_n \\ P_\alpha^k(\sigma)\geq 1/n!}} \left(P_\alpha^k(\sigma) - \frac{1}{n!}\right) - \frac{1}{2} \sum_{\substack{\sigma\in S_n \\ P_\alpha^k(\sigma)< 1/n!}} \left(P_\alpha^k(\sigma) - \frac{1}{n!}\right)\end{aligned}$$

If the derivative of the total variation distance with respect to alpha after k shuffles is negative, this means that there is some $\varepsilon > 0$ such that

$$\|P_{0.5+\varepsilon}^k - U\|_{\mathrm{TV}} < \|P_{0.5}^k - U\|_{\mathrm{TV}}.$$

The derivative of the total variation distance is

$$\frac{\partial}{\partial\alpha}\|P_\alpha^k - U\|_{\mathrm{TV}} = \frac{1}{2}\sum_{\substack{\sigma\in S_n\\ P_\alpha^k(\sigma)\geq 1/n!}}\left(\frac{\partial}{\partial\alpha}P_\alpha^k(\sigma)\right) - \frac{1}{2}\sum_{\substack{\sigma\in S_n\\ P_\alpha^k(\sigma)<1/n!}}\left(\frac{\partial}{\partial\alpha}P_\alpha^k(\sigma)\right)$$

The probability of being at a given permutation is the sum of all "paths" of shuffles that lead to that permutation.

Definition 10 (Paths of Shuffles)**.** A path is a sequence $\Gamma = (s_1, s_2, \ldots, s_k)$ of shuffles with each $s_i \in \{0, 1, \ldots, 2^n - 1\}$. Let the length of the path be $|\Gamma| = k$. Then define the set of paths from $\sigma \in S_n$ to $\tau \in S_n$

$$\mathcal{P}_{\sigma,\tau}^k := \{\Gamma = (s_1, s_2, \ldots, s_k) : S(s_k)\cdots S(s_2)S(s_1)\sigma = \tau\}$$

The probability of performing a given path is simply the product of the probability of each shuffle. So if $\Gamma = (s_1, \ldots, s_k)$ is a path of shuffles with $j_1, j_2, \ldots, j_k$ switches, then the probability of performing the entire sequence of shuffles is

$$\begin{aligned}\rho_\alpha(s_1)\rho_\alpha(s_2)\cdots\rho_\alpha(s_k) &= \prod_{i=1}^k \frac{1}{2}\alpha^{j_i}(1-\alpha)^{n-1-j_i}\\ &= \frac{1}{2^k}\alpha^{\sum j_i}(1-\alpha)^{k(n-1)-\sum j_i}\end{aligned}$$

So if we define j to be the total switches between 0 and 1 in all shuffles of the path, we get the probability of a given path of shuffles $\frac{1}{2^k}\alpha^j(1-\alpha)^{k(n-1)-j}$. So writing the probability of a permutation as the sum of probabilities of all paths Γ with total switches j_Γ, we get

$$P_\alpha^k(\sigma) = \frac{1}{2^k}\sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k}\alpha^{j_\Gamma}(1-\alpha)^{k(n-1)-j_\Gamma}$$

So the derivative with respect to α is

$$\begin{aligned}\frac{\partial}{\partial\alpha}P_\alpha^k(\sigma) &= \frac{1}{2^k}\sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k}\left(j_\Gamma\alpha^{j_\Gamma-1}(1-\alpha)^{k(n-1)-j_\Gamma} - (k(n-1)-j_\Gamma)\alpha^{j_\Gamma}(1-\alpha)^{k(n-1)-j_\Gamma-1}\right)\\ &= \frac{1}{2^k}\sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k}\alpha^{j_\Gamma-1}(1-\alpha)^{k(n-1)-j_\Gamma-1}\left(j_\Gamma(1-\alpha) - (k(n-1)-j_\Gamma)\alpha\right)\end{aligned}$$

Since we are interested in the derivative at $\alpha = 1/2$, the expression simplifies greatly. Let $J_{e,\sigma}^k := \sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k} j_\Gamma$ be the sum of all switches in all paths.

$$\begin{aligned}\left.\frac{\partial}{\partial\alpha}P_\alpha^k(\sigma)\right|_{\alpha=\frac{1}{2}} &= \frac{1}{2^k}\sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k}\frac{1}{2^{k(n-1)-2}}\left(\frac{j_\Gamma}{2} - \frac{k(n-1)-j_\Gamma}{2}\right)\\ &= \frac{1}{2^k}\frac{1}{2^{k(n-1)-2}}\sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k}\left(j_\Gamma - \frac{k(n-1)}{2}\right)\\ &= \frac{1}{2^{kn-2}}\sum_{\Gamma\in\mathcal{P}_{e,\sigma}^k}\left(j_\Gamma - \frac{k(n-1)}{2}\right)\\ &= \frac{1}{2^{kn-2}}\left(J_{e,\sigma}^k - |\{\Gamma\in\mathcal{P}_{e,\sigma}^k\}|\frac{k(n-1)}{2}\right)\end{aligned}$$

$(n-1)$ is simply the total number of transitions between bits in a shuffle; the number of switches plus the number of stays. So $k(n-1)$ is the total number of transitions within the whole path of k shuffles. Let $T^k_{e,\sigma} := |\{\Gamma \in \mathcal{P}_{e,\sigma}\}| \cdot k(n-1)$ be the total number of transitions in all paths that lead to σ. If there are more total switches than stays in all paths to σ, then the derivative of the permutation probability will be positive, and if there are more stays than switches, it will be negative. Using this, we have the derivative for the total variation distance at $\alpha = 1/2$.

$$\frac{\partial}{\partial \alpha} \|P^k_\alpha - U\|_{\text{TV}}\bigg|_{\alpha=\frac{1}{2}} = \frac{1}{2} \sum_{\substack{\sigma \in S_n \\ P^k_{0.5}(\sigma) \geq 1/n!}} \left(\frac{\partial}{\partial \alpha} P^k_{0.5}(\sigma)\right) - \frac{1}{2} \sum_{\substack{\sigma \in S_n \\ P^k_{0.5}(\sigma) < 1/n!}} \left(\frac{\partial}{\partial \alpha} P^k_{0.5}(\sigma)\right)$$
$$= \frac{1}{2^{kn-1}} \left(\sum_{\substack{\sigma \in S_n \\ P^k_{0.5}(\sigma) \geq 1/n!}} \left(J^k_{e,\sigma} - \frac{T^k_{e,\sigma}}{2}\right) - \sum_{\substack{\sigma \in S_n \\ P^k_{0.5}(\sigma) \geq 1/n!}} \left(J^k_{e,\sigma} - \frac{T^k_{e,\sigma}}{2}\right) \right)$$

Lemma 8.2. *Suppose the deck size n is even. Let $R(\sigma)$ be the number of rising sequences in a permutation. For sufficiently large k,*

- *If $R(\sigma) > n/2$, then $P^k_{0.5}(\sigma) < 1/n!$.*
- *If $R(\sigma) \leq n/2$, then $P^k_{0.5}(\sigma) > 1/n!$.*

Bayer and Diaconis [1] showed that the probability of a given permutation σ with GSR is

$$P^k_{0.5}(\sigma) = \frac{\binom{2^k+n-R(\sigma)}{n}}{2^{kn}}$$

Since this quantity will strictly decrease as we increase $R(\sigma)$ we must prove the result holds for the edge cases $R(\sigma) = (n/2) + 1$ and $R(\sigma) = n/2$. It suffices to show that the ratio $\frac{P^k_{0.5}(\sigma)}{1/n!} = n! \cdot P^k_{0.5}(\sigma)$ is greater than or less than 1 depending on the case. Thus, we can reduce the problem to proving:

- For any k,
$$n! \cdot \binom{2^k + n - (\frac{n}{2} + 1)}{n} < 2^{kn}$$
- For sufficiently large k,
$$n! \cdot \binom{2^k + n - \frac{n}{2}}{n} > 2^{kn}$$

Proof. We will prove the first case, which holds for any k. Expanding, we get

$$n! \cdot \binom{2^k + n - R(\sigma)}{n} = n! \frac{(2^k + n - R(\sigma))!}{(2^k - R(\sigma))! \cdot n!}$$
$$= \left(2^k - \frac{n}{2} - 1 + 1\right) \cdot \left(2^k - \frac{n}{2} - 1 + 2\right) \cdots \left(2^k - \frac{n}{2} - 1 + n\right)$$
$$= \left(2^k - \frac{n}{2}\right) \cdot \left(2^k - \frac{n}{2} + 1\right) \cdots \left(2^k + \frac{n}{2} - 1\right)$$

We can group the factors by their distance from 2^k and apply difference of squares.

$$
\begin{aligned}
n! \cdot \binom{2^k + n - R(\sigma)}{n} &= 2^k \left(2^k - \frac{n}{2}\right) \prod_{d=1}^{\frac{n}{2}-1} \left(2^k + d\right)\left(2^k - d\right) \\
&= 2^k \underbrace{\left(2^k - \frac{n}{2}\right)}_{<2^k} \prod_{d=1}^{\frac{n}{2}-1} \underbrace{\left(2^{2k} - d^2\right)}_{<2^{2k}} \\
&< 2^k 2^k \left(2^{2k}\right)^{\frac{n}{2}-1} = \left(2^{2k}\right)^{n/2} = 2^{kn}
\end{aligned}
$$

□

Proof. For the case of $R(\sigma) = n/2$, we will choose a K so that the result holds for any $k \geq K$. Simplifying in the same way, we want to show that

$$
n! \cdot \binom{2^k + n - R(\sigma)}{n} = \left(2^k - \frac{n}{2} + 1\right) \cdot \left(2^k - \frac{n}{2} + 2\right) \cdots \left(2^k + \frac{n}{2}\right) > 2^{kn}
$$

Let $m = n/2 \in \mathbb{N}$. We will group the terms symmetrically.

$$
\begin{aligned}
n! \cdot \binom{2^k + n - R(\sigma)}{n} &= \prod_{i=1}^{n} \left(2^k - \frac{n}{2} + i\right) \\
&= \prod_{d=1}^{n/2} \left(2^k - d + 1\right)\left(2^k + d\right) \\
&= \prod_{d=1}^{n/2} 2^{2k} + d2^k - (d-1)2^k - d(d-1) \\
&= \prod_{d=1}^{n/2} 2^{2k} + \left(2^k - d(d-1)\right)
\end{aligned}
$$

Choose $K = \lceil \log_2(\frac{n}{2}(\frac{n}{2} - 1) + 1) \rceil$. Then if $k > K$, we have

$$
2^k \geq \frac{n}{2}\left(\frac{n}{2} - 1\right) + 1 > \frac{n}{2}\left(\frac{n}{2} - 1\right)
$$

So since $d \leq \frac{n}{2}$, for each factor we know $2^{2k} + (2^k - d(d-1)) > 2^{2k}$. Therefore,

$$
\begin{aligned}
n! \cdot \binom{2^k + n - R(\sigma)}{n} &= \prod_{d=1}^{n/2} 2^{2k} + \left(2^k - d(d-1)\right) \\
&> (2^{2k})^{n/2} = 2^{kn}
\end{aligned}
$$

□

Using this lemma, we can group permutations based on their rising sequence count. From Bayer and Diaconis [1], we know the total number of paths to σ is $|\{\Gamma \in \mathcal{P}^k_{e,\sigma}\}| = \binom{2^k - n + R(\sigma)}{n}$ and the total number of permutations with a given number of rising sequences r is $\left\langle {n \atop r-1} \right\rangle$, the eulerian number. Define the total switches in paths to permutations with r rising sequences as

$$
J^k_r := \sum_{\substack{\sigma \in S_n \\ R(\sigma) = r}} J^k_{e,\sigma}
$$

Also define the total number of transitions (switches or stays) in paths to permutations with r rising sequences as

$$
T^k_r := \sum_{\substack{\sigma \in S_n \\ R(\sigma) = r}} T^k_{e,\sigma} = \left\langle {n \atop r-1} \right\rangle \binom{2^k + n - r}{n} k(n-1)
$$

Then, for any even n and any $k \geq \lceil \log_2(\frac{n}{2}(\frac{n}{2}-1)+1) \rceil$ we have

$$\frac{\partial}{\partial \alpha} \|P_\alpha^k - U\|_{\mathrm{TV}} \Bigg|_{\alpha=\frac{1}{2}} = \frac{1}{2^{kn-1}} \left(\sum_{r=1}^{n/2} \left(J_r^k - \frac{T_r^k}{2} \right) - \sum_{r=\frac{n}{2}+1}^{n} \left(J_r^k - \frac{T_r^k}{2} \right) \right)$$

We want to show this derivative is negative for sufficient k. This simplifies the problem to simply showing that there are more stays than switches in all paths that lead to low rising sequence permutations, and that the opposite is true for paths that lead to high rising sequence permutation. In fact, since every possible sequence of binary shuffles is counted, there are equally many switches and stays in total. This means that we just need to show one of the two, and the other would follow.

9 Conclusion and Acknowledgments

In this report we investigated how the mixing rate of riffle shuffling changes depending on the neatness of the shuffle. We discussed the ways that our model has been explored in the past and we introduced new methods that can simplify the problem. The first new method shows that we can analyze the second eigenvalue simply through the probability of returning to the identity. The second new method simplifies our goal of showing that GSR is not optimal for mixing to counting the number of switches between packets in all shuffles that lead to certain permutations. We hope to explore these methods further in the future.

I would like to thank my mentor, Mark Conger, for guiding me throughout this research, and the University of Michigan REU organizers for providing me with this opportunity.

References

[1] Dave Bayer and Persi Diaconis, *Trailing the dovetail shuffle to its lair*, The Annals of Applied Probability (1992), 294–313.

[2] Alexander Cope, *Collecting data on how a riffle shuffle permutes a deck of cards*, University of Michigan, 2011.

[3] Persi Diaconis, *Mathematical developments from the analysis op riffle shuffling*, Groups, combinatorics and geometry: Durham 2001, 2003, pp. 73–97.

[4] Christopher Drouillard, *An alternative model for riffle shuffling*, University of Michigan, 2013.

[5] Edgar Gilbert, *Theory of shuffling*, Bell Labs, 1955.

[6] Lee Jieun, Janine Lewandoski, and Kevin Reifor, *Exploring the geometric model of riffle shuffling*, University of Michigan, 2014.

[7] Johan Jonasson and Benjamin Morris, *Rapid mixing of dealer shuffles and clumpy shuffles* (2015).

[8] Robin Pemantle, *Randomization time for the overhand shuffle*, Journal of Theoretical Probability **2** (1989), no. 1, 37–49.

[9] Henri Poincaré, *Calcul des probabilités*, Gauthier-Villars, 1912.