On Cutoff for the Asymmetric Riffle Shuffle

Sophia Zhou

December 22, 2024

1 Introduction

1.1 What is a p-shuffle?

The riffle shuffle is one of the most common methods for shuffling a deck of cards. Gilbert, Shannon, and Reeds parameterized the riffle shuffle into something called a p-shuffle, which operates as follows: Given a deck of N cards, we take the top $Bin(N,p)$ cards off, and interleave the two piles such that if the left pile has A cards and the right has B cards, the next card will be dropped from the left pile with probability $\frac{A}{A+B}$ and from the right pile with probability $\frac{B}{A+B}$. We can see that conditioned on pile size, this gives a uniformly random interleaving.

Bayer and Diaconis [1] have shown that for a deck of size N undergoing $\frac{1}{2}$-shuffles, that $(\frac{3}{2\log 2} \pm o(1))\log N$, log being the natural log, shuffles are necessary and sufficient to randomize the deck. More precisely, for any $\epsilon > 0$, the total variation distance tends to 1 after $\lfloor(\frac{3}{2\log 2} - \epsilon)\log N\rfloor$ and tends to 0 after $\lfloor(\frac{3}{2\log 2} + \epsilon)\log N\rfloor$ shuffles as $N \to \infty$. Though we have these results for $p = \frac{1}{2}$, the total variation distance cutoff has not been well-understood for general p-shuffles.

1.2 Description of main result

Sellke in "Cutoff for the Asymmetric Riffle Shuffle" [4] shows that all p-shuffles exhibit cutoff, and gives an upper and lower bound for the cutoff time of general p-shuffles.

Precisely, for a deck of size N, Sellke proves cutoff bounds for $\mathbf{p}$-shuffles: let $\mathbf{p} = (p_1, ..., p_k), p_i > 0 \forall i, \sum_i p_i = 1$. Then, generate a multinomial vector $(n_1, ..., n_k)$ such that each $n_i \sim Bin(N, p_i)$ marginally and $\sum_i n_i = N$ a.s.. Then, split the cards by taking the first n_1 cards to form the first pile, the next n_2 cards for the second, and so on. Then, interleave the cards such that the next card has $\frac{A_i}{\sum_j A_j}$ chance from falling from the i-th pile, where the current size of pile i is A_i. When $\mathbf{p} = (\frac{1}{k}, ..., \frac{1}{k})$, the cutoff time is $\frac{3\log N}{2\log k} + O(1)$ by the same rising sequence analysis as in the $p = \frac{1}{2}$ case analyzed by Bayer and Diaconis [1].

2 Preliminaries

2.1 General background

Markov chain: A markov chain is a sequence of random variables (X_n) on a probability space $(\Omega, \mathcal{X}, P)$ such that $P(X_{n+1} = a_{n+1} | X_n = a_n, X_{n-1} = a_{n-1}, ...) = p(a_n, a_{n+1})$, where $p(a_n, a_{n+1})$ is a constant transition probability (between values/states a_n and a_{n+1}). Markov chains are defined by their transition matrix.

Symmetric group: The symmetric group S_n is the group of permutations from $[n]$ to itself. For example, the identity element in S_n is the map $x \mapsto x, x \in [n]$.

Total variation distance: Given a measure space $(\Omega, \mathcal{X})$, and two probability distributions p, Q on $(\Omega, \mathcal{X})$, the total variation distance $||p - Q||_{TV} := \frac{1}{2} \sum_{x \in \Omega} |p(x) - Q(x)|$. For a Markov chain transition matrix denoted by P, the distance to stationarity is $d(t) := \max_{x \in \Omega} ||P^t(x, \cdot) - Q(\cdot)||_{TV}$.

Total variation distance mixing time and cutoff: Given a Markov chain X_n, the total variation distance mixing time with respect to some $\epsilon > 0$ and probability distribution π is $t_{mix}(\epsilon) := \min\{t : d(t) < \epsilon\}$. A Markov chain X_n is said to exhibit cutoff if $\exists$ a sequence t_n such that $d((1-\epsilon)t_n) \to 1$ and $d((1+\epsilon)t_n) \to 0$ as $n \to \infty$ for all $\epsilon > 0$.

Big O and small o notation: For a sequence of random variables (X_n) on a probability space $(\Omega, \mathcal{X}, P)$, and a sequence of real numbers (a_n), $X_n = o(a_n) \iff \lim_{n \to \infty} P(|\frac{X_n}{a_n}| > \epsilon) = 0, \forall \epsilon > 0$. Also, $X_n = O(a_n) \iff \exists N, M \in \mathbb{R}$ s.t. $\lim_{n \to \infty} P(|\frac{X_n}{a_n}| > M) < \epsilon, \forall n \geq N, \forall \epsilon > 0$. So, $X_n = o(a_n) \implies X_n = O(a_n)$, but $X_n = O(a_n)$ does not necessarily imply $X_n = o(a_n)$, since the former is continuity while the latter is uniform continuity.

Notice also that if we write $X_n = o(1)$, then $\lim_{n \to \infty} P(|X_n| > \epsilon) = 0 \iff X_n$ converges to 0 in probability.

2.2 Strings, shuffle graphs, and notation

Let $P_{\mathbf{p}}$ be the probability measure on S_n after applying a single $\mathbf{p}$-shuffle to the identity.

Proposition 2.1. *Let $\boldsymbol{p} = (p_0, ..., p_{k-1})$, $\boldsymbol{q} = (q_0, ..., q_{l-1})$ be discrete probability vectors. Then, performing a $\boldsymbol{q}$-shuffle followed by a $\boldsymbol{p}$-shuffle is equivalent to performing a $\boldsymbol{p} * \boldsymbol{q}$-shuffle, where $\boldsymbol{p} * \boldsymbol{q} = (p_0 q_0, p_0 q_1, ..., p_0 q_{l-1}, p_1 q_0, ..., p_{k-1} q_{l-1})$.*

Proof. This proposition was proved in “Analysis of Top To Random Shuffles” by Diaconis, Fill, and Pitman [2]. □

We are interested in composing a $\mathbf{p}$-shuffle K times with itself, denoted as a $\mathbf{p}^{*K}$-shuffle. Using Proposition 2.1, we can derive an explicit formula for the probability distribution $P_{\mathbf{p}^{*K}}$, the distribution over S_n after K $\mathbf{p}$-shuffles.

To do this, we first generate a set of length K strings. We generate N length K strings with digits in $[k]_0 = \{0, ..., k-1\}$, where each letter in the strings is $\mathbf{p}$-random and i.i.d.. Then, we sort these strings, call them s_i, such that $s_1 \leq_{lex} ... \leq_{lex} s_N$. Let $S = (s_1, ..., s_N)$.

Next, we define a shuffle graph $G = G(S)$ on the vertex set $[N]$ such that $(i, i+1) \in E(G) \iff s_i = s_{i+1}$. So, G will look like a set of disjoint paths.

Finally, we choose a uniform random permutation $\pi \in S_n$ and define its G modification π^G as follows: within each path $\{s_i, ..., s_j\}$, we sort the values $\pi(i), ..., \pi(j)$ in increasing order and reassign them to $\pi^G(i), ..., \pi^G(j)$ such that $\pi^G(i) < ... < \pi^G(j)$. An example of S, G, π, and π^G are shown below.

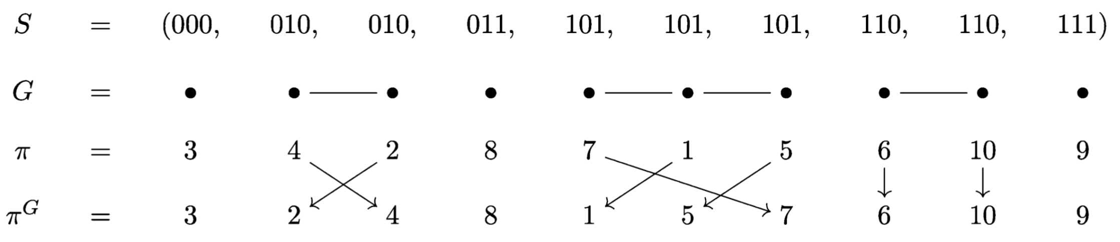
S = (000, 010, 010, 011, 101, 101, 101, 110, 110, 111)
G =
π = 3 4 2 8 7 1 5 6 10 9
π^G = 3 2 4 8 1 5 7 6 10 9

Figure 1: An example of S, G, π, π^G for $N = 10$, $[k]_0^K = [2]_0^3$.

Here is some intuition for why we are generating these strings and graphs. We can view each string s_i as denoting the sequence of piles that the i-th card visits. Observe that we sort within the G components for the fact that if two cards have the same exact orbit through the piles of the deck, then their relative order must be preserved. This motivates the following proposition.

Proposition 2.2. *For $S_,G, \pi, \pi^G$ as defined above for some constants N, K, the distribution of all $(\pi^G)^{-1}$ is exactly $P_{\mathbf{p}}^{*K}$. As a result, the total variation distance of $(\pi^G)^{-1}$, or interchangeably, π^G, from uniform equals $d_N(K)$, where $d_N(K)$ denotes the total variation distance from uniform after $\mathbf{p}$-shuffling K times.*

In other words, after more and more shuffles, the increasing subintervals in π^G should decrease in size, so that π^G looks roughly uniform.

Proof. This proposition was proved as lemma 3 by Lalley in "On the Rate for Mixing for p-shuffles" [3]. □

There is another natural parametrization for generating such S, G, and π^G. This parameterization is used to prove lemma 4.3 at the end of section 4.

Let $\mu_{\mathbf{p},M}$ be the probability measure on $[k]_0^M$ where each digit is independently $\mathbf{p}$-random. Let $\mathcal{S} \subset ([k]_0^K)^N$ be the set of all lexicographically nondecreasing sequences of N strings of length K each. Let $\mathcal{G}$ be the set of shuffle graphs with N vertices, that is, subgraphs of the path graph on N vertices.

Let $\mathbb{E}^\sigma, \mathbb{E}^\pi, \mathbb{P}^\sigma, \mathbb{P}^\pi$ denote expectations or probabilities taken over uniformly random permutations σ or π in S_N. Similarly, $\mathbb{E}^S$ is the expectation over $S \sim \mu_{\mathbf{p},K}$.

Let $E(G, G') = E(G) \cap E(G')$, denoting the edge intersection of graphs $G, G' \in \mathcal{G}$.

Square brackets denote substrings or indexes into strings. For example, if $x \in [k]_0^M$, then $x[i]$ denotes the i-th digit of x. We also have that $[00]$ is the string containing two consecutive 0's, for example, and $[(k-1)(k-1)]$ is the string containing two consecutive $(k-1)$'s.

The new parameterization is as follows. For each string $x \in [k]_0^M$, $M \geq 1$, we define,

$$t_x := \mathbb{P}^{y \sim \mu_{\mathbf{p}}}[y <_{lex} x],$$

$$\lambda_x := \mathbb{P}^{y \sim \mu_{\mathbf{p}}}[y = x] = \Pi_i p_x[i],$$
$$J_x := [t_x, t_x + \lambda_x).$$

We can see that the intervals (J_x) partition $[0,1)$ (also see figure below). Note that $t_x + \lambda_d = \mathbb{P}^{y \sim \mu_{\mathbf{p}}}[y \leq_{lex} x]$.

Thus, to sample a **p**-random string $x \in [k]_0^M$. we can equivalently sample a uniform random variable $a \in [0,1]$ and take the unique x where $a \in J_x$. So, to sample $(s_1, ..., s_N) \in \mathcal{S}$, we can simply sample uniform i.i.d. $a_1, ..., a_N \in [0,1]$, sort them in increasing order, and choose the strings s_i such that $a_i \in J_{s_i}$.

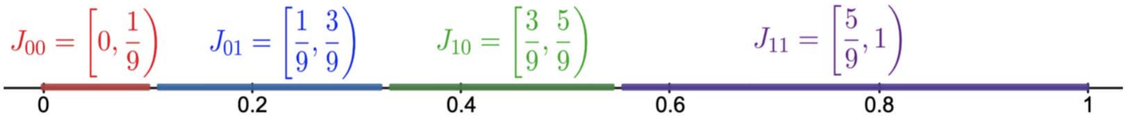
$J_{00} = \left[0, \frac{1}{9}\right)$
$J_{01} = \left[\frac{1}{9}, \frac{3}{9}\right)$
$J_{10} = \left[\frac{3}{9}, \frac{5}{9}\right)$
$J_{11} = \left[\frac{5}{9}, 1\right)$
0
0.2
0.4
0.6
0.8
1

Figure 2: The partition $U_i J_{s_i}$ with $k = 2, M = 2, (p_0, p_1) = (\frac{1}{3}, \frac{2}{3})$.

2.3 Constants for statement of main result

We define some important constants that will be referenced in the proofs.

$$p_{min} = \min_i (p_i)$$

$$p_{max} = \max_i (p_i)$$

$$\phi_{\mathbf{p}}(t) = \sum_{i=1}^{k} p_i^t$$

$$\psi_{\mathbf{p}}(t) = -\log \phi_{\mathbf{p}}(t)$$

$$\theta_{\mathbf{p}} \in \mathbb{R} \text{ s.t. } \psi_{\mathbf{p}}(\theta_{\mathbf{p}}) = 2\psi_{\mathbf{p}}(2)$$

By raising e to the LHS and RHS of the above equation, it follows that $\phi_{\mathbf{p}}(\theta_{\mathbf{p}}) = \phi_{\mathbf{p}}(2)^2$.

$$C_{\mathbf{p}} = \frac{3 + \theta_{\mathbf{p}}}{4\psi_{\mathbf{p}}(2)} = \frac{3 + \theta_{\mathbf{p}}}{2\psi_{\mathbf{p}}(\theta_{\mathbf{p}})}$$

$$\widetilde{C_{\mathbf{p}}} = \frac{1}{\log(\frac{1}{p_{max}})}$$

$$\overline{C_{\mathbf{p}}} = max(\widetilde{C_{\mathbf{p}}}, C_{\mathbf{p}})$$

$\overline{C_{\mathbf{p}}}$ is a constant that we weight with $\log N$ to get the mixing time cutoff. $\widetilde{C_{\mathbf{p}}}, C_{\mathbf{p}}$ give two separate bounds of mixing time because of two main obstructions: when $K \leq (\widetilde{C_{\mathbf{p}}} - \epsilon) \log N$, some strings will occur many times, and in general there is a fractal set of locations which tend to contain many G edges. Sellke uses an independent point process heuristic to analyze this fractal set, but I will not be going over this proof in this report.

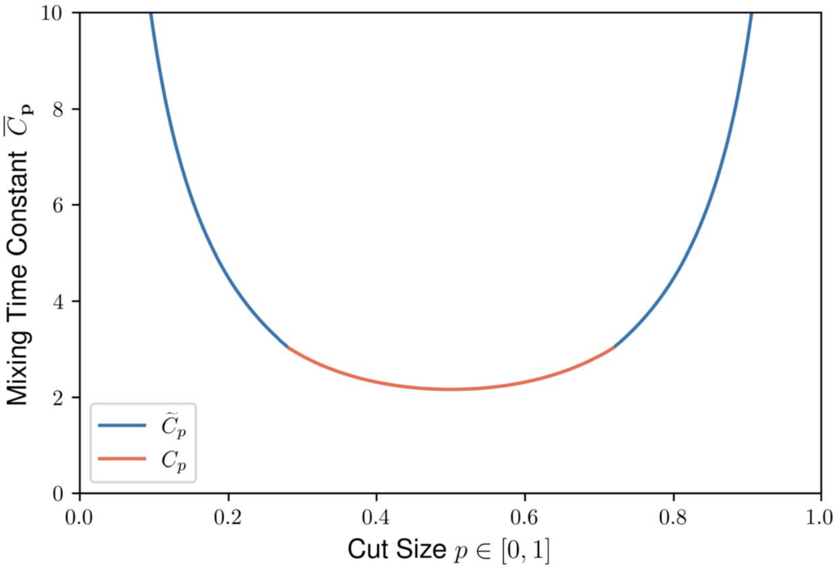

Figure 3: The values for $\overline{C_{\mathbf{p}}}$ for $\mathbf{p} = (p, 1-p)$. The transitions between $C_{\mathbf{p}}$ and $\widetilde{C_{\mathbf{p}}}$ occur at approximately 0.28 and 0.72.

3 Statement of main result

Theorem 3.1. *For a deck of size N, a $\boldsymbol{p}$-shuffle undergoes total variation cutoff after $\overline{C_{\boldsymbol{p}}}\log(N)$ shuffles. Precisely,*

$$\lim_{N\to\infty} d_N(\lfloor(1-\epsilon)\overline{C_{\boldsymbol{p}}}\log(N)\rfloor) = 1 \tag{1}$$

$$\lim_{N\to\infty} d_N(\lfloor(1+\epsilon)\overline{C_{\boldsymbol{p}}}\log(N)\rfloor) = 0, \tag{2}$$

for all $\epsilon > 0$.

Most of the paper is devoted to proving Theorem 3.1.2, the upper bound. In the rest of the report, I prove most of the results used in the proof of the upper bound, and finally the proof of the upper bound itself.

4 Lemmas for proof of upper bound

4.1 Proof strategy

First, we can write an explicit formula for the total variation distance, and compute some important associated values. In particular, an important value is $f_{G,G'}$, which the core of the proofs works on bounding.

Next, we analyze exceptional sequences in the set of lexicographically sorted sequences of strings and their corresponding shuffle graphs, and eventually come up with a bound for $f_{G,G'}$. These results complete the proof of the mixing time upper bound!

4.2 Useful derivations

We begin by examining the Radon-Nikodym derivative between the distributions of π^G and π. This will help us define a function $f_{G,G'}$, which will help us find an explicit formula for $d_N(K)$. For each $G \in \mathcal{G}$, let $C(G) = \{G_1, ..., G_j\}$ be the connected components of G, and let v_i be the number of vertices in G_i. Then, the map $\alpha : S_N \to S_n$ given by $\pi \mapsto \pi^G$ is $\Pi_{i=1}^{j} v_i!$ to 1, since all the permutations π containing the same vertices within the components in any order will map to the same π^G. Moreover, notice that $Im(\alpha)$ consists of $\sigma \in S_n$ where $\sigma^G = \sigma$. Then,

$$\begin{aligned}
\mathbb{P}^{\pi}[\pi^G = \sigma] &= \frac{\Pi_{i=1}^{j} v_i!}{N!}, \sigma \in Im(\alpha) \\
&= 1_{\sigma = \sigma^G} \cdot \frac{\Pi_{i=1}^{j} v_i!}{N!}, \sigma \in S_N.
\end{aligned}$$

Thus, for a fixed $G \in \mathcal{G}$, the R-N derivative $f_{G,\sigma}$ of π^G w.r.t. π is given as

$$\begin{aligned}
f_{G,\sigma} &:= \frac{\mathbb{P}^{\pi}[\pi^G = \sigma]}{\mathbb{P}^{\pi}[\pi = \sigma]} \\
&= N! \cdot \mathbb{P}^{\pi}[\pi^G = \sigma] \\
&= 1_{\sigma = \sigma^G} \cdot \Pi_{i=1}^{j} v_i! \\
&= \frac{1_{\sigma = \sigma^G}}{\mathbb{P}^{\pi}[\pi^G = \pi]}.
\end{aligned}$$

Thus, notice that for a fixed $G \in \mathcal{G}$, $\mathbb{E}^{\sigma}[f_{G,\sigma}] = \mathbb{E}^{\sigma}[\frac{1_{\sigma = \sigma^G}}{\mathbb{P}^{\pi}[\pi^G = \pi]}] = \frac{\mathbb{P}^{\sigma}[\sigma^G = \sigma]}{\mathbb{P}^{\pi}[\pi^G = \pi]} = 1$.

On the other hand, for fixed $\sigma \in S_N$ and $\mu_{\mathbf{p},K}$-random $G = G(S)$, we can apply the law of total expectation to get,

$$\begin{aligned}
\mathbb{P}^{\pi,S}[\pi^{G(S)} = \sigma] &= \mathbb{E}^S[f_{G(S),\sigma} \cdot \mathbb{P}^{\pi}[\pi = \sigma]] \\
&= \frac{\mathbb{E}^S[f_{G(S),\sigma}]}{N!}.
\end{aligned}$$

Finally, we can formulate the total variation distance to uniform after K shuffles with the aforementioned constants and formulas. Specifically, if we let the probability measure over π^G be P and the uniform probability measure over π be $Q = \frac{1}{N!}$, then the total variation distance to uniform is,

$$\begin{aligned}
d_N(K) &= \frac{1}{2} \cdot \sum_{\sigma \in S_N} |P(\sigma) - Q(\sigma)| \\
&= \frac{1}{2} \cdot \sum_{\sigma} |\frac{\mathbb{E}^S[f_{G(S),\sigma}]}{N!} - \frac{1}{N!}| \\
&= \frac{1}{2} \sum_{\sigma} \frac{1}{N!} |\mathbb{E}^S[f_{G(S),\sigma} - 1]| \\
&= \frac{1}{2} \cdot \mathbb{E}^{\sigma} |\mathbb{E}^S[f_{G(S),\sigma} - 1]|.
\end{aligned}$$

Now, we use a chi-squared upper bound for total variation distance after removing a certain exceptional subset of sequences from $\mathcal{S}$. We will be defining the conditions that the "typical" sequences in $\mathcal{S}$ must satisfy shortly. To make this upper bound, first partition $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_0$, where $\mathcal{S}_1$ contains "typical" sequences. Then,

$$\begin{aligned}
\mathbb{E}^{\sigma}|\mathbb{E}^{S}[f_{G(S),\sigma} - 1]| &\leq \mathbb{E}^{\sigma}|\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{S\in\mathcal{S}_1}]| + \mathbb{E}^{\sigma}|\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{S\in\mathcal{S}_0}]| \\
&\leq \mathbb{E}^{\sigma}|\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{S\in\mathcal{S}_1}]| + \mathbb{E}^{\sigma,S}[(f_{G(S),\sigma} + 1)1_{S\in\mathcal{S}_0}] \\
&\underset{\mathbb{E}^{\sigma}[f_{G,\sigma}]=1}{=} \mathbb{E}^{\sigma}|\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{S\in\mathcal{S}_1}]| + \mathbb{E}^{S}[(1+1)1_{S\in\mathcal{S}_0}] \\
&= \mathbb{E}^{\sigma}|\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{S\in\mathcal{S}_1}]| + 2\mu_{\mathbf{p}}(\mathcal{S}_0).
\end{aligned}$$

Now, take S' to be an independent copy of S, and define for any shuffle graphs $G, G' \in \mathcal{G}$, $f_{G,G'} := \mathbb{E}^{\sigma}[f_{G,\sigma}f_{G',\sigma}]$. Next, we use Cauchy-Schwarz to upper-bound the main term of the equation from above,

$$\begin{aligned}
(\mathbb{E}^{\sigma}|\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{s\in\mathcal{S}_1}]|)^2 &\leq \mathbb{E}^{\sigma}[(\mathbb{E}^{S}[(f_{G(S),\sigma} - 1)1_{s\in\mathcal{S}_1}])^2] \\
&\underset{\text{expand squared term}}{=} \mathbb{E}^{\sigma}\mathbb{E}^{S,S'}[(f_{G(S),\sigma} - 1)(f_{G(S'),\sigma} - 1)1_{S,S'\in\mathcal{S}_1}] \\
&= \mathbb{E}^{\sigma}\mathbb{E}^{S,S'}[(f_{G(S),\sigma}f_{G(S'),\sigma} - f_{G(S),\sigma} - f_{G(S'),\sigma} + 1)1_{S,S'\in\mathcal{S}_1}] \\
&= \mathbb{E}^{\sigma}\mathbb{E}^{S,S'}[(f_{G(S),\sigma}f_{G(S'),\sigma} - 1 - 1 + 1)1_{S,S'\in\mathcal{S}_1}] \\
&= \mathbb{E}^{\sigma}\mathbb{E}^{S,S'}[(f_{G(S),\sigma}f_{G(S'),\sigma} - 1)1_{S,S'\in\mathcal{S}_1}] \\
&= \mathbb{E}^{S,S'}[(f_{G,G'} - 1)1_{S,S'\in\mathcal{S}_1}].
\end{aligned}$$

Recall that this quantity must be non-negative since it is at least as big as the squared quantity we started with on the LHS. So, when $S, S' \in \mathcal{S}_\infty$, $f_{G,G'} \geq 1$. So, to establish mixing, or to show $d_N(K) \to 0$ as $N \to \infty$, it remains to show that $f_{G,G'}$ rarely exceeds 1 in in a L^1 sense.

So, we now have,

$$d_N(K) \leq \frac{1}{2}\sqrt{\mathbb{E}^{S,S'}[(f_{G,G'} - 1)1_{S,S'\in\mathcal{S}_1}]} + \mu_{\mathbf{p}}(\mathcal{S}_0).$$

For the remainder of the paper, we set $G = G(S), G' = G(S')$, since that will be the only context in which we refer to G, G'. For any shuffle graphs G, G', define a new shuffle graph U to be their edge-union with components $C(U)$.

4.3 Upper-bounding $f_{G,G'}$

We now move on to upper-bounding $f_{G,G'}$ using $|E(G,G')|$. To motivate why such a relationship should exist, observe that if no vertex $i \in [N]$ is incident to both a G-edge and a G'-edge, then $f_{G,\sigma} = \frac{\mathbb{P}^{\pi}[\pi^G=\sigma]}{\mathbb{P}^{\pi}[\pi=\sigma]}$, $f_{G',\sigma} = \frac{\mathbb{P}^{\pi}[\pi^{G'}=\sigma]}{\mathbb{P}^{\pi}[\pi=\sigma]}$ are exactly independent for $\sigma \in S_N$ uniformly random, e.g. the set of π such that $\pi^G = \sigma$ intersects with the set of π such that $\pi^{G'} = \sigma$ exactly in the set where $\pi^U = \sigma$. In the independent case, we would have the exact equality $f_{G,G'} = \mathbb{E}^{\sigma}[f_{G,\sigma}f_{G',\sigma}] = \mathbb{E}^{\sigma}[f_{G,\sigma}]\mathbb{E}^{\sigma}[f_{G',\sigma}] = 1$.

Lemma 4.1. *Suppose $C(U)$ has vertex sizes $(u_1, ..., u_c)$. Then,*

$$f_{G,G'} \leq \Pi_{1\leq i\leq c, E(U_i)\cap E(G,G')\neq\emptyset} u_i!.$$

We will go over the proof of this lemma after going through some results based on this lemma.

Here is the first condition that "typical" sequences, e.g. those in $\mathcal{S}_1$, must satisfy. The objective is to ensure that the u_i in lemma 4.1 are uniformly bounded by some constant $L = L(\mathbf{p}, \epsilon)$.

Definition 4.1. *For $L \geq 10$ a positive integer, a shuffle graph G is L-sparse if within any interval $\{i, ..., i+L-1\} \subset [N]$ of L consecutive vertices, at most, $\frac{L}{3}$ of the possible $L-1$ edges are in $E(G)$.*

Lemma 4.2. *Suppose G, G' are L-sparse shuffle graphs. Then,*

$$f_{G,G'} \leq (L!)^{|E(G,G')|}.$$

Proof. We claim that $\max_i(u_i) \leq L$, that is, each U-component (recall U is the edge union of G, G') contains at most L vertices. Indeed, by L-sparsity, U contains at most $\frac{L}{3} + \frac{L}{3} < L - 1$ edges within each subinterval of L vertices, so no such interval can be a connected component of U. Thus, lemma 4.1 implies that,

$$f_{G,G'} \leq \Pi_{E(U_i) \cap E(G,G') \neq \emptyset}(L!).$$

By definition, the number of components U_i satisfying $E(U_i) \cap E(G, G') \neq \emptyset$ can be at most $|E(G, G')|$, e.g. if each edge in $E(G, G')$ corresponds with a component in U, so we have the result as desired. □

Now, we define a second condition for the definition of $\mathcal{S}_1$, called regularity. This will help us create a cover of $E(G, G')$, and eventually bound the (truncated) exponential moments of $|E(G, G')|$, which is our remaining task. Intuititvely, regularity amounts to requiring that both prefixes [00] (two consecutive 0 digits) and $[(k-1)(k-1)]$ (two consecutive $k-1$ digits) appear with roughly the expected frequency among $S = (s_1, ..., s_N)$.

Definition 4.2. *Recall $\boldsymbol{p} = (p_0, ..., p_{k-1})$. The sequence $S = (s_1, ..., s_N)$ is regular if at most $(p_0^2 + \frac{p_0 p_{k-1}}{2})N$ strings s_i begin with [00], and at most $(p_{k-1}^2 + \frac{p_0 p_{k-1}}{2})N$ strings begin with $[(k-1)(k-1)]$.*

Lemma 4.3. *For any $\boldsymbol{p}$ and $\epsilon > 0$, there exist $L = L(\boldsymbol{p}, \epsilon) \in \mathbb{Z}^+$ and $\delta = \delta(\boldsymbol{p}, \epsilon) > 0$ such that the following holds. Consider a $\boldsymbol{p}$-random sequence $S = (s_1, ..., s_N)$ of strings of length $K \geq (\widetilde{C_{\boldsymbol{p}}} + \epsilon) \log N$. Then, with probability $1 - O(N^{-\delta})$, S is regular and $G(S)$ is L-sparse.*

We will prove this lemma later in this report, and will first go over the results that it implies.

We now precisely define $\mathcal{S}_1$ as the set of regular sequences S for which $G(S)$ is L-sparse for the $L = L(\mathbf{p}, \epsilon)$ that exists by lemma 4.3. Then, lemma 4.3 exactly states that,

$$\mu_{\mathbf{p}}(\mathcal{S}_0) = 1 - \mu_{\mathbf{p}}(\mathcal{S}_1) = O(N^{-\delta}),$$

for some small $\delta = \delta(\mathbf{p}, \epsilon)$.

Next, we show how to cover $E(G, G')$ when S, S' are regular.

Definition 4.3. *Let $E_{for}(G)$ consist of all edges $(i, i+1) \in E(G)$ for which the strings $s_i = s_{i+1}$ do not begin with $[(k-1)(k-1)]$. Let $E_{for}(G, G') := E_{for}(G) \cap E_{for}(G')$. Define $E_{back}(G, G')$ in the same way but replacing $[(k-1)(k-1)]$ with $[00]$.*

Lemma 4.4. *If $S, S' \in \mathcal{S}$ are regular, then,*

$$|E(G, G')| \leq |E_{for}(G, G')| + |E_{back}(G, G')|.$$

Proof. The lemma is not immediately clear because S and S' are not necessarily the same, e.g. we are afraid of filtering out too many edges when taking $E_{for}(G) \cap E_{for}(G')$ and $E_{back}(G) \cap E_{back}(G')$.

Regularity of S, S' implies that $E_{for}(G, G')$ contains all edges $(i, i+1) \in E(G, G')$ with

$$i \leq N - (p_{k-1}^2 + \frac{p_0 p_{k-1}}{2})N.$$

To clarify, $E_{for}(G, G')$ could contain more edges in $E(G, G')$ because of the "at most" condition of regularity. Similarly, $E_{back}(G, G')$ contains all shared edges $(i, i+1) \in E(G, G')$ with

$$i \geq (p_0^2 + \frac{p_0 p_{k-1}}{2})N.$$

Since $p_0^2 + p_0 p_{k-1} + p_{k-1}^2 < (p_0 + p_{k-1})^2 \leq 1$, we get,

$$(p_0^2 + \frac{p_0 p_{k-1}}{2} + p_{k-1}^2 + \frac{p_0 p_{k-1}}{2})N < N$$
$$\implies (p_0^2 + \frac{p_0 p_{k-1}}{2})N < (1 - p_{k-1}^2 - \frac{p_0 p_{k-1}}{2})N.$$

Therefore, there is overlap between $E_{for}(G, G')$ and $E_{back}(G, G')$, implying $E_{for}(G, G') \cup E_{back}(G, G') = E(G, G')$, giving us the desired result. □

Finally, to establish the mixing time upper bound in Theorem 3.1.2, it remains to prove the following lemma.

Lemma 4.5. *For any $\boldsymbol{p}$ and positive reals ϵ, t, there exists $\delta = \delta(\boldsymbol{p}, \epsilon, t)$ such that if $K \geq (\overline{C_{\boldsymbol{p}}} + \epsilon \log N$, then,*

$$\mathbb{E}[e^{t \cdot |E_{for}(G, G')|}] \leq 1 + O(N^{-\delta}).$$

We will not prove this lemma in this report, but we prove the mixing time upper bound using the above results.

Proof of Theorem 3.1.2. Recall that from above we have

$$d_N(K) \leq \frac{1}{2} \cdot \sqrt{\mathbb{E}^{S,S'}[(f_{G,G'} - 1)1_{S,S' \in \mathcal{S}_1}]} + \mu_{\mathbf{p}}(\mathcal{S}_0).$$

Since $\mu_{\mathbf{p}}(\mathcal{S}_0) = O(N^{-\delta})$, it remains to estimate $\mathbb{E}^{S,S'}[(f_{G,G'} - 1)1_{S,S' \in \mathcal{S}_1}]$.

Let $\delta > 0$ be sufficiently small, depending on $(\mathbf{p}, \epsilon, L, t)$. Using lemmas 4.2, 4.4, and 4.5 with $t = 2\log(L!)$, we get,

$$\begin{aligned}
\mathbb{E}^{S,S'}[(f_{G,G'} - 1)1_{S,S' \in \mathcal{S}_1}] &\underset{(4.2)}{\leq} \mathbb{E}^{S,S'}[((L!)^{|E(G,G')|} - 1)1_{S,S' \in \mathcal{S}_1}] \\
&\underset{(4.4)}{\leq} \mathbb{E}^{S,S'}[((L!)^{|E_{for}(G,G')|+|E_{back}(G,G')|} - 1)1_{S,S' \in \mathcal{S}_1}] \\
&\underset{(4.5)}{\leq} \frac{\mathbb{E}[e^{2|E_{for}(G,G')|}] + \mathbb{E}[e^{2|E_{back}(G,G')|}]}{2} - 1 \\
&= O(N^{-\delta}).
\end{aligned}$$

Finally, by the above lemma, we get that $d_N(K) \leq O(N^{-\delta})$ when $K \geq (\overline{C_\mathbf{p}} + \epsilon)\log N$. □

This is a complete proof for the upper bound, except that lemmas 4.1, 4.3, and 4.5 are yet to be proved. In this report, I will continue to prove lemmas 4.1 and 4.3.

4.4 Proof of lemmas 4.1 and 4.3

Proof of Lemma 4.1. Let $(v_1, ..., v_a)$ be the vertex sizes of the G-components, $(w_1, ..., w_b)$ be the vertex sizes of the G'-components, and $(u_1, ..., u_c)$ be the sizes of the U-components. We claim that,

$$f_{G,G'} = \frac{(\Pi_{i=1}^a v_i!) \cdot (\Pi_{j=1}^b w_j!)}{\Pi_{l=1}^c u_l!}.$$

Indeed, this follows by,

$$\begin{aligned}
f_{G,G'} &= \mathbb{E}^\sigma[f_{G,\sigma} g_{G',\sigma}] \\
&= \mathbb{E}^\sigma[1_{\sigma^G = \sigma} \cdot 1_{\sigma^{G'} = \sigma} \cdot (\Pi_{i=1}^a v_i!) \cdot (\Pi_{j=1}^b w_j!)] \\
&= \mathbb{E}^\sigma[1_{\sigma^G = \sigma^{G'} = \sigma}] \cdot (\Pi_{i=1}^a v_i!) \cdot (\Pi_{j=1}^b w_j!) \\
&= \mathbb{E}^\sigma[1_{\sigma^U = \sigma}] \cdot (\Pi_{i=1}^a v_i!) \cdot (\Pi_{j=1}^b w_j!) \\
&= \frac{(\Pi_{i=1}^a v_i!) \cdot (\Pi_{j=1}^b w_j!)}{\Pi_{l=1}^c u_l!}.
\end{aligned}$$

Indeed, $1_{\sigma^U = \sigma} = 1_{\sigma^G = \sigma^{G'} = \sigma}$, since the LHS implies the RHS and the RHS implies the LHS (U contains precisely all of the edges and thus connected components in G and G').

We define

$$f_{G,G',U_l} := \frac{(\Pi_{i:G_i \subset U_l} v_i!) \cdot (\Pi_{j:G_j \subset U_l} w_j!)}{u_l!}.$$

Then, we can decompose $f_{G,G'}$ as $f_{G,G'} = \Pi_l f_{G,G',U_l}$.

Finally, observe that for any general positive integers $m_1, ..., m_n, M$ where $\sum_i (m_i - 1) \leq M - 1$, we have that $\Pi_{i=1}^n m_i! \leq M!$, since both sides can be written as a product of at most $M - 1$ integers at least 2, and the $M - 1$ numbers appearing in the product for $M!$ are clearly larger. In particular, this holds for $M = u_l$, the total number of vertices in U_l, where $m_1, ..., m_n$ are vertex sizes of edge-disjoint subinterval graphs of U_l. We thus have,

$$\Pi_{i:G_i \subset U_l} v_i! \leq u_l!,$$

$$\Pi_{j:G_j \subset U_l} w_j! \leq u_l!,$$

and so it follows that $f_{G,G',U_l} \leq u_l!$. Moreover, if the component U_l does not contain any edges in both G and G',, then the $G_i \subset U_l$ and $G'_j \subset U_l$ are collectively disjoint, and we have,

$$(\Pi_{i:G_i \subset U_l} v_i!) \cdot (\Pi_{j:G_j \subset U_l} w_j!) \leq u_l!,$$

implying $f_{G,G',U_l} \leq 1$. Substituting these into $\Pi_l f_{G,G',U_l}$ gives the desired result. □

The next lemma is used to prove lemma 4.3.

Lemma 4.6. *For $K \geq (\widetilde{C_{\boldsymbol{p}}} + \epsilon) \log N$, there is $\delta(\boldsymbol{p}, \epsilon) > 0$ such that the following holds. Conditioned on any initial strings $s_1, ..., s_i$, none of which begin with $[(k-1)(k-1)]$, the conditional probability that $s_i = s_{i+1}$ is at most $O(N^{-\delta})$.*

Proof. Recall the definitions from preliminaries and the alternative parameterization for generating a sequence of strings S. Let $a_1, ..., a_N$ be our i.i.d. uniform random variables on $[0, 1]$, where we've sorted them such that $0 \leq a_1 \leq ... \leq a_N \leq 1$, and then we choose s_j such that $a_j \in J_{s_j}$ for each $1 \leq j \leq N$.

Consider conditioning on the value $a_i \in J_{s_i}$. Then, the numbers $a_j, j > i$ are conditionally i.i.d. in $[a_i, 1]$. Then, notice that the interval $[a_i, 1]$ has length lower-bounded by $1 - a_i \geq$ length of interval corresponding to strings starting with $[(k-1)(k-1)] = p_{k-1}^2 \geq p_{min}^2$.

It then follows that the conditional probability for any single $j > i$ to have $s_j = s_i$ is $\geq \frac{\lambda_{s_i}}{1-a_i} \geq \frac{\lambda_{s_i}}{p_{min}^2}$, and so we get that the conditional distribution for the number of $j > i$ with $s_j = s_i$ is stochastically dominated by a $Bin(N, \frac{\lambda_{s_i}}{p_{min}^2})$ random variable.

Finally, since $K \geq (\widetilde{C_{\mathbf{p}}} + \epsilon) \log N$ was assumed,

$$\begin{aligned} \lambda_{s_i} &\leq (p_{max})^K \\ &\leq p_{max}^{(\widetilde{C_{\mathbf{p}}}+\epsilon) \log N} \\ &= (p_{max}^{\frac{1}{-\log p_{max}}+\epsilon})^{\log N} \\ &= (\frac{1}{e} \cdot p_{max}^{\epsilon})^{\log N} \\ &= N^{-1} \cdot p_{max}^{\epsilon \log N} \\ &\leq N^{-1-\delta}, \end{aligned}$$

by choosing the appropriate δ. Finally, by using the binomial distribution from above, we get $\mathbb{P}[s_{i+1} = s_i | (s_1, ..., s_i)] \leq N \cdot \frac{\lambda_{s_i}}{p_{min}} \leq N \cdot N^{-1-\delta} \cdot p_{min}^{-2} = O(N^{-\delta})$, which is our desired result. □

Proof of Lemma 4.3. We focus on L-sparsity, since the regularity of S follows from Chernoff estimates. First, the above lemma implies that $\mathbb{P}[s_{i+1} = s_i | (s_1, ..., s_i)] \leq O(N^{-\delta}$ when $s_i <_{lex} [(k-1)(k-1)]$. This implies that the set of edges formed by strings $s_i <_{lex} [(k-1)(k-1)]$ is stochastically dominated by choosing each edge independently with probability $O(N^{-\delta})$. By symmetry, this also holds for edges formed by strings starting with $[(k-1)(k-1)]$ (we can do the same analysis as above for the λ_{s_i} where $s_i \geq_{lex} [(k-1)(k-1)]$). We call these ordinary edges and final edges, respectively.

Finally, a Chernoff bound implies that for $L \geq 1000\delta^{-1}$, each interval of L consecutive vertices contains, at most, $L/6$ ordinary edges and $L/6$ final edges with probability $1 - O(N^{-2})$, O taken w.r.t. L. Finally, we can see that over, at most, N such intervals, L-sparsity holds with probability at least $1 - O(N^{-1}) \geq 1 - O(N^{-\delta})$. □

I have not dived deep into Chernoff or Chi-squared bounds since I determined to regulate the scope of this report to what I have so far.

References

[1] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2), 1992.

[2] Persi Diaconis, James Allen Fill, and Jim Pitman. Analysis of top to random shuffles. *Combinatorics, Probability and Computing*, 1992.

[3] Steven P. Lalley. On the rate of mixing for p-shuffles. *The Annals of Applied Probability*, 10(4), 2000.

[4] Mark Sellke. Cutoff for the asymmetric riffle shuffle. *The Annals of Probability*, 50(6), 2022.