ANALYSIS OF THE RIFFLE

Jack Hsieh

November 22, 2020

1 Abstract.

This paper explores the mixing times related to the riffle shuffle as explored by Bayer and Diaconis. We define the riffle shuffle with the commonly-accepted Gilbert–Shannon–Reeds model and follow the analysis of the mode, the major conclusion of which is the following theorem.

Theorem 1. If n cards are riffled m times, then the probability that the deck contains r rising sequences is

$$\frac{\binom{2^m+n-r}{n}}{2^{mn}}.$$

We then analyze the total variation distance between the probability distribution effected by consecutive riffles and the uniform distribution, intending to arrive at Bayer and Diaconis's principal result.

Theorem 2. For $m = \frac{3}{2}\log_2 n + \theta$ riffles, then for large n,

$$\left\|Q^{(m)} - U\right\| = 1 - 2\phi\left(-\frac{2^{-\theta}}{\sqrt{3}}\right) + O\left(n^{-\frac{1}{4}}\right)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$.

Though we do not detail the complete proof, we outline the necessary simplifications that reduce the proof of the above theorem to the collection of known asymptotic bounding. We then interpret the theorem in more familiar terms.

We then demonstrate by calculation Bayer and Diaconis's famed result that approximately seven riffles provides an adequate mixing of a standard deck of 52 cards with respect to total variation.

2 Definitions.

2.1 The riffle

The first order of business is to define the model with which we work. We represent decks of n cards as permutations of n elements, or the symmetric group

S_n. We start with the identity permutation $\sigma_0 = (1, 2, ..., n)$, an ordered deck, and repeatedly apply a non-deterministic shuffle, thereby generating σ_{i+1} from σ_i. The particular shuffle we use is the Gilbert-Shannon-Reeds riffle, defined to simulate the real-world riffle or dovetail shuffle:

Definition 1 (The Gilbert-Shannon-Reeds riffle)**.** Consider a deck of n cards. To perform the riffle,

1. Split the deck into a bottom packet P_{bottom} and top packet P_{top}, selecting the size of the P_{top} according to the binomial distribution. That is, $\mathbb{P}(P_{\text{top}} \text{ contains exactly } k \text{ cards}) = \binom{n}{k} 2^{-n}$.

2. Where B is the number of cards remaining in P_{bottom} and T is the number of cards remaining in P_{top}, with probability $\frac{B}{B+T}$, remove the top card in P_{bottom} and add it to the top of the new deck. Otherwise, remove the top card in P_{top} and add it to the top of the new deck. Repeat until the new deck contains all n cards.

2.2 The a-shuffle

Critical to the analysis of the Gilbert-Shannon-Reeds model is the generalization of the riffle to the generalized a-shuffle. While Bayer and Diaconis provide four equivalent descriptions of the generalized a-shuffle, here we only present the most relevant and useful of the four, describing the a-shuffle by its inverse:

Definition 2 (Inverse a-shuffle)**.** Consider a deck of n cards. To perform the inverse a-shuffle,

1. Take the top card of the original deck and place it into a uniformly-randomly chosen packet P_i, where $1 \leq i \leq a$. Repeat until the original deck is empty.

2. When all the packets are filled, stack each packet in order so that P_1 is the bottom and P_a is on top.

It can easily be seen that the inverse 2-shuffle yields the inverse of the Gilbert-Shannon-Reeds riffle, as expected. Note the implication that all sequential divisions into packets and interleavings are equally likely. The most important revelation is that performing an a-shuffle followed by a b-shuffle is an equivalent to performing a single ab-shuffle.

2.3 Total variation distance

The second order of business is to define our measure of success. In this case, we consider metrics we use to declare when a deck has been properly shuffled. Our first instinct is the total variation distance:

Definition 3 (Total variation distance)**.** The total variation distance between probability distributions V_1 and V_2 on state space Ω is defined as

$$\|V_1 - V_2\| = \max_{A \subseteq \Omega}(V_1(A) - V_2(A)),$$

or, equivalently,

$$\|V_1 - V_2\| = \sum_{\substack{a \in \Omega \\ V_1(a) \geq V_2(a)}} V_1(a) - V_2(a).$$

The equivalence of the two definitions is trivial to show. We let the probability distribution generated by m riffles be $Q^{(m)}$ such that $Q^{(m)}(\pi) = \mathbb{P}(\sigma_m = \pi)$. As our standard of randomness, we compare $Q^{(m)}$ to the uniform distribution U where $U(\sigma) = \frac{1}{n!}$ (i.e. the probability distribution under which all permutations are equally likely). As an illustrative principle, we should expect that as m increases, $\left\|Q^{(m)} - U\right\|$ approaches 0.

3 Results.

3.1 Gilbert-Shannon-Reeds model analysis.

An essential tool for characterizing card permutations is the notion of a rising sequence.

Definition 4 (Rising sequence)**.** A rising sequence r in a permutation of cards is a maximal consecutively-increasing sequence.

For example, a deck $(4, 1, 2, 5, 6, 3)$ contains two rising sequences: $4, 5, 6$ and $1, 2, 3$. The key observation is that a rising sequence tells us where the start and end of some packets must be. This logic yields the following:

Theorem 3 (Probability of a permutation in relation to the number of rising sequences)**.** Suppose permutation π has r rising sequences. Then the probability that an a-shuffle will result in π is

$$\frac{\binom{a+n-r}{n}}{a^r}.$$

Proof. Note that a set of packets and a particular interleaving of those packets identifies a (not-necessarily-unique) permutation. Note that for any packet division, there is either one or no interleavings that yield the permutation π; thus it suffices to simply count the number of packet divisions that could yield the permutation π and divide by the total number of packet divisions and interleavings. Also note that a rising sequence can only arise from the union of consecutive packets. Clearly, there are $a - 1$ packet dividers in total. The location of $r - 1$ dividers are fixed by the specific rising sequences of π. However, the remaining $a - r$ dividers can be placed in n positions, lending to a total of $\binom{n+a-r}{n}$ possible divisions into packets. It is easy to see through the inverse shuffle that the total number of packet divisions and interleavings is a^n. Thus, the probability of the a-shuffle result in π is $\frac{\binom{a+n-r}{n}}{a^r}$, as desired. □

Theorem 1 easily follows by combining the above theorem to the equivalence of an a shuffle followed by a b shuffle and an ab shuffle. For convenience, we let $Q_r^{(m)} = \mathbb{P}(\sigma_m \text{ has } r \text{ rising sequences}) = \frac{\binom{2^n+n-r}{n}}{2^{mn}}$.

3.2 Asymptotic bounds on the mixing time.

Proposition 1. Let $r = \frac{n}{2} + h$, where $1 - \frac{n}{2} \leq h \leq \frac{n}{2}$. Let $m = \log_2(n^{3/2}c)$, where $c > 0$ is fixed and finite. Then

$$Q_r^{(m)} = \frac{1}{n!} \exp\left(\frac{1}{c\sqrt{n}} \left(-h + \frac{1}{2} + O_c\left(\frac{h}{n}\right)\right) - \frac{1}{24c^2} - \frac{1}{2}\left(\frac{h}{cn}\right)^2 + O_c\left(\frac{1}{n}\right)\right)$$

Proof.

$$\begin{aligned} Q_r^{(m)} &= \frac{1}{n!} \prod_{i=1}^{n} \frac{2^n + i - r}{2^m} \\ &= \frac{1}{n!} \exp\left(\sum_{i=0}^{n-1} \log\left(1 + x_i\right)\right) \end{aligned}$$

where

$$x_i = \frac{\frac{n}{2} - h - i}{cn^{\frac{3}{2}}}.$$

The power series bounds log:

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \leq \log(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}.$$

Sum formulas over the sum of the appropriately weighted corresponding powers of x_i yields the proposition. □

The following relationship is helpful for calculating the total variation distance:

Proposition 2. Let h^* be the integer such that $Q_{\frac{n}{2}+h}^{(m)} \geq \frac{1}{n!}$ if and only if $h \leq h^*$. Then, for any fixed c, as $n \to \infty$

$$h^* = \frac{-\sqrt{n}}{24c} + \frac{1}{12c^3} + B + O_c\left(\frac{1}{\sqrt{n}}\right)$$

where $|B| \leq 1$.

Proof. Note that $Q_{\frac{n}{2}+h}^{(m)} \leq \frac{1}{n!}$ if and only if the power in the above expression is non-negative. Setting the exponent to be 0 and solving for h yields the above expression for h^*, corrected by B. □

Theorem 4. Let $r = \frac{n}{2} + h$, where $1 - \frac{n}{2} \leq h \leq \frac{n}{2}$. Let $m = \log_2(n^{3/2}c)$, where $c > 0$ is fixed and finite. Then

$$\left\| Q^{(m)} - U \right\| = 1 - 2\phi\left(-\frac{1}{4c\sqrt{3}}\right) + O_c\left(n^{-\frac{1}{4}}\right)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$.

Proof. We will not attempt here to prove this theorem in its entirety. However, we can make the following simplifications with ease. Let $R_{n,h}$ be the number of permutations with $r = \frac{n}{2} + h$ rising sequences. Then

$$\left\|Q^{(m)} - U\right\| = \sum_{-\frac{n}{2} < h \leq h^*} R_{n,h} \left(Q_r^{(m)} - \frac{1}{n!}\right).$$

Note the usage of h^* to only select the range where the summand is non-negative.

Define a descent as a position i in a permutation π where $\pi(i) > \pi(i{+}1)$. The position of the value i in π is given by $\pi^{-1}(i)$, so a descent in $\pi^{-}1$ corresponds with the start of a new rising sequence in π. Thus a permutation π has r rising sequences if and only if π^{-1} has $r - 1$ descents. The number of permutations of n elements with exactly $r - 1$ descents is counted by the well-known Eulerian numbers $A_{n,r-1}$, yielding

$$\left\|Q^{(m)} - U\right\| = \sum_{-\frac{n}{2} < h \leq h^*} A_{n,r-1} \left(Q_r^{(m)} - \frac{1}{n!}\right).$$

Certain asymptotic bounds as well as explicit formulas for the Eulerian numbers are known, so we have tremendously simplified the problem of calculating the total variation distance. The remainder of the proof combines the earlier lemmas as well as the aforementioned bounds on the Eulerian numbers. For the specifics, the curious can consult Bayer and Diaconis's paper. □

Theorem 2 easily follows by defining $\theta = \log_2 c$, thus proving a main object of interest in this paper while also providing a more precise behavioral description. Note that our result stands at time $m = \frac{3}{2}\log_2 n + \theta)$, where θ represents the number of riffles after $\frac{3}{2}\log_2 n$ riffles. It can be shown that

$$\left\|Q^{(m)} - U\right\| \approx 1 - 2\phi\left(-\frac{-1}{4c\sqrt{3}}\right) \sim \frac{1}{2c\sqrt{6\pi}}$$

as $c \to \infty$. In other words, with more steps *beyond* $\frac{3}{2}\log_2 n$ steps, the shuffled probability distribution approaches the uniform distribution exponentially fast. Furthermore,

$$\left\|Q^{(m)} - U\right\| \approx 1 - 2\phi\left(-\frac{-1}{4c\sqrt{3}}\right) \; 1 - \frac{4c\sqrt{3}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(-\frac{1}{4c\sqrt{3}}\right)^2\right)$$

as $c \to 0$. In other words, with more steps *before* $\frac{3}{2}\log_2 n$ steps, the shuffled probability distributions diverges from the uniform distribution exponentially fast. One interpretation is that the transition experiences rapid cutoff.

3.3 Calculated example for standard 52-card deck.

The above theorem gives asymptotic bounds. However, our simplifications made in our introduction to the proof of **Theorem 4** makes calculating the total variation distance simple for (small) values of n.

n \ m	1	2	3	4	5	6	7	8	9	10
25	1.000	1.000	0.999	0.775	0.437	0.231	0.114	0.056	0.028	0.014
32	1.000	1.000	1.000	0.929	0.597	0.322	0.164	0.084	0.042	0.021
52	1.000	1.000	1.000	1.000	0.924	0.614	0.334	0.167	0.085	0.043
78	1.000	1.000	1.000	1.000	1.000	0.893	0.571	0.307	0.153	0.078
104	1.000	1.000	1.000	1.000	1.000	0.988	0.772	0.454	0.237	0.119
208	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.914	0.603	0.329
312	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.883	0.565

Table 1. Total variation distance calculated for for varying values of n and m.

For the standard 52-card deck, the steep drop at $m = 7$ leads to the famous conclusion that seven riffles sufficiently shuffles the deck.

4 References.

BAYER, D. and DIACONIS, P. (1992) Trailing the Dovetail Shuffle to Its Lair. The Annals of Applied Probability, 2 294-313.