# Shuffling

Shahrzad Haddadan

March 7, 2013

**Abstract**

In this paper we will talk about two well-known shuffling methods, the "Top to Random" and the "Riffle Shuffle". We are interested in the number of shuffles that will make the deck of $n$ cards uniformly random.

## 1 Introduction

Shuffling cards has been interesting to mathematicians for more than fifty years. Other than its mathematical beauty, shuffling is also studied regarding some applications in other fields such as Biology ([2],[3]) and also Cryptography ([4]).

In this paper, we will talk about the "Top to random" shuffle and "Riffle shuffle". We will try to show their exact mixing time and also bound the deviation and separation distances.

## 2 Preliminaries

Let's say we have a deck of size $n$. Any arrangement of $n$ cards is essentially a permutation of the numbers $1 \dots n$.

During this talk, $S_n$ equals the set of all permutations of $1 \dots n$. Shuffling is nothing but a random walk on $S_n$ with uniform stationary distribution. The question is, given a random walk, how many steps will make the distribution close to uniform distribution. In order to measure being close we need the following definitions [1]:

**Definition 2.1.** Given two distributions $\mu$ and $\gamma$ on some set $S$, we define the standard deviation to be $||\mu - \gamma||_{TV} = \frac{1}{2} \sum_{x \in S} |\mu(x) - \gamma(x)|$.

[1] The definitions and lemmas of this section can be found in [1] and [5].

Given $P$ the transition matrix of a random walk on state space $S$, $x \in S$ the starting state, $t$ an integer representing the number of steps and $\pi$ the stationary distribution of the random walk, we denote the deviation distance by $d(t)$ which is defined by $d(t) := \max_{x \in S} ||P^t(x, .) - \pi||_{TV}$.

We say a chain is mixed at time $t$ if $d(t) \leq 1/4$.

We also define the *separation distance* $s_x(t)$ by, $s_x(t) := \max_{y \in S}[1 - \frac{P^t(x,y)}{\pi(y)}]$.

**Lemma 2.1.** *Let $P$ be the transition matrix of a Markov chain on state space $S$, $x$ the starting state and $t$ the number of steps. We have $||P^t(x, .) - \pi||_{TV} \leq s_x(t)$.*

Proof. $||P^t(x, .) - \pi||_{TV} = \sum_{\substack{y \in S \\ P^t(x,y) < \pi(y)}} [\pi(y) - P^t(x, y)] = \sum_{\substack{y \in S \\ P^t(x,y) < \pi(y)}} \pi(y)[1 - \frac{P^t(x, y)}{\pi(y)}] \leq \max_y [1 - \frac{P^t(x, y)}{\pi(y)}] = s_x(t)$.

We also need some definitions regarding exact mixing time.

**Definition 2.2.** Given a Markov chain $M$ on a state space $S$, let $W^*$ be the space of all finite walks of $M$. A *stopping rule* is map $\Gamma : W^* \to [0, 1]$ which shows the probability of continuing a walk $w = (w_1, \ldots, w_t) \in W^*$. We define *stopping time* to be the random variable $\tau$ with values $\{0, 1, \ldots\}$ that is equal to the random time that $\Gamma$ stops.[2]

The *exact mixing time* of a chain is the expected stopping time for an optimal stopping rule.

**Definition 2.3.** For the chain $M = (M_t)$, a *strong stationary time*, $\tau$ is a stopping time which satisfies, $Pr_x(\tau = t, M_\tau = y) = Pr_x(\tau = t)\pi(y)$. Where $\pi$ is the stationary distribution.

**Theorem 2.2.** (Lovász and Winkler [5])[3]

*An stopping rule is optimal if and only if it has a halting state. A halting state is a state that given that we already stopped we know it was never exited.*

**Lemma 2.3.** (Aldous and Diaconis [6])

*Let $x$ be the starting state of a random walk $(X_t)$ and $\tau$ a strong stationary time, then $s_x \leq Pr_x(\tau > t)$.*

[2]Check out [5] for more details on stopping rules and stopping time.

[3]We don't bring the proof here. However, it can be found in [5]

Proof. Let $y$ be the state for which the maximum of $1 - Pr_x(X_t = y)/\pi(y)$ is achieved. Then,

$$s_x(t) = 1 - \frac{Pr_x(X_t = y)}{\pi(y)} \leq 1 - \frac{Pr_x(X_t = y, \tau \leq t)}{\pi(y)} = 1 - \frac{Pr_x(\tau \leq t)\pi(y)}{\pi(y)} = Pr_x(\tau > t). \quad \square$$

# 3 Top to Random Shuffle

Consider the following method of shuffling for a deck of size $n$. At each step, take the first card and insert it uniformly in any of $n$ places that is left in the deck. Formally, we have a random walk on $S_n$ with the following transition matrix:

$$Pr(\sigma, \gamma) = \begin{cases} 1/n & \text{if } \gamma = (\sigma_2, \sigma_3, \ldots, \sigma_i, \sigma_1, \sigma_{i+1}, \ldots, \sigma_n), \quad 1 \leq i \leq n \\ 0 & \text{otherwise.} \end{cases}$$

## 3.1 Exact mixing

**Lemma 3.1.** *For a deck of $n$ cards, the exact mixing time for top to random shuffle is $n(H_{n-1}-1)+1$. Where $H_n = \sum_{i=1}^{n} 1/i$. When $n \to \infty$ the exact mixing time converges to $n \log(n-1) - n+1$.*

Proof. Following is a an optimal stopping rule for top to random shuffle. Let's say we start from $\sigma = (\sigma_1, \ldots, \sigma_n)$. Mark the card $\sigma_{n-1}$. Shuffle until the marked card gets to the top. Do one more shuffle. Stop. It is easy to see that when we stop we are in uniform distribution. Moreover, this rule is optimal since any state in which the deck has $\sigma_n$ on top is a halting state. Now, we need to calculate the expected time that takes the rule to stop.

Let $T_i$ be time that it takes until $i$ cards get under card $\sigma_{n-1}$. We know that $E[T_1] = 0$. Now consider the time $T_{i+1}-T_i$. This is the time that is needed for another card to get under the card $\sigma_{n-1}$ given that there are already $i$ cards below $\sigma_{n-1}$. Note that $T_{i+1}-T_i$ has geometric distribution with parameter $(i+1)/n$. Therefore, we have $E[T_{i+1}-T_i] = n/(i+1)$. Let $\tau$ be the stopping time for top to random shuffle. We have, $E[\tau] = E[T_{n-1}]+1 = E[T_1]+E[T_2-T_1]+\ldots+E[T_{n-1}-T_{n-2}]+1 = n\sum_{i=2}^{n-1}(1/i)+1 = n(H_{n-1}-1)+1$. $\square$

## 3.2 Conventional mixing time and bounds on separation distance and deviation distance

**Lemma 3.2.** *Given a deck of size $n$, the mixing time of the top to random shuffle is less than $n \log n$.*

Proof. The stopping time $\tau$ that we gave in proof of Lemma 3.1 is in fact a strong stationary time. Let $P$ be the transition matrix of the chain and $U$ the uniform distribution. We have $||P^t(\sigma, .)-U|| \leq s_\sigma(t) \leq Pr(\tau > t)$.

Claim. $Pr(\tau > n \log n+cn) \leq e^{-c}$. Consider the coupon collector problem[4]. Notice that for $T_i$s in proof of Lemma 3.1 we have $Pr(T_i-T_{i-1} = j) = \frac{i}{n}(1 - \frac{i}{n})^{j-1}$ which is the same probability of how long it takes for the coupon collector to collect the $n-i$+1st coupon after collecting the $n-i$th one. The stopping time $\tau$ in the above proof is in fact equal to the time it takes for the coupon collector to collect the last $n-1$st cards(2nd, 3rd ,... $n$th card) plus one extra step which is equal to the time needed to collect all $1, \ldots, n$ coupons. Now, let's try to upper bound $Pr(\tau > n \log n + cn)$. Let $A_i$ be the event that the collector does not collect the coupon number $i$ till time $n \log n + cn$. We have

$$Pr(\tau > n \log n+cn) \leq \sum_{i=1}^{n} Pr(A_i) = \sum_{i=1}^{n} (1 - \frac{1}{n})^{n \log n+cn} \leq n \exp(-\frac{n \log n + cn}{n}) = e^{-c}.$$

# 4 Riffle Shuffle

Riffle shuffle is a very common way of shuffling. In Riffle shuffle one divides the deck to two piles and successively drop cards from the bottom of each pile. In 1955, Gilbert and Shannon and independently Reeds in 1981 established a good mathematical modeling of the problem. In 1992 Diaconis and Bayer analysed the Riffle shuffle. Here, as we did for top to random case, we will discuss the exact mixing time and bounds on deviation and separation distances. Then, we will talk about the famous result of Bayer and Diaconis that has been famous in news as "7 shuffles is enough".

## 4.1 Modeling of the problem

**Definition 4.1.** The following four modelings of Riffle shuffle are equivalent:

[4]To know more about the coupon collector problem, read [1] Section 2.2 or [6].

I. Let $m$ be taken randomly from $Binomial(n, 1/2)$. Split the card to piles of size $m$ and $n-m$. Let a card drop from left pile with probability $a/(a+b)$ and from right pile with probability $b/(a+b)$, where $a$ is the number of cards left in left pile and $b$ is the number of cards left in right pile.

II. Let $m$ be taken randomly from $Binomial(n, 1/2)$. Split the card to piles of size $m$ and $n-m$. Choose one of the $\binom{n}{m}$ possible arrangements of these card uniformly at random.

III. Place $n$ points $x_1, \ldots, x_n$ uniformly and independently in unit interval. Assign the cards in their order to $x_1, \ldots, x_n$. Apply the mapping $x \to 2x$ to the points. Rearrange the cards according to the new ordering.

Here, we don't give a proof of these definitions being equivalent but it can be easily checked that all of them will yield the following distribution. [5]

$$Pr(\sigma, \gamma) = \begin{cases} (n+1)/2^n & \text{if } \gamma = \sigma \\ 1/2^n & \text{if } \gamma = \sigma \circ \lambda \text{ and } \lambda \text{ has exactly two rising sequences(Definition 4.2).} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

**Definition 4.2.** A rising sequence in a permutation is the maximal set of consecutive numbers that occur in the correct order. For example $(2, 3, 1, 4, 6, 5, 7)$ has three rising sequences $\{2, 3, 4, 5\}$, $\{1\}$, $\{6, 7\}$.

## 4.2 Exact mixing and bounds on separation and deviation distances

In order to analyse the Riffle shuffle it is easier to look at its reverse. According to a theorem by Winkler and Lovász ([8]) the exact mixing time of a chain and its reverse are equal. The reverse of Riffle shuffle is famous as unshuffle. A step of unshuffle is performed by:

**Definition 4.3.** (Unshuffle)

To each card in the deck assign a uniformley random bit (0 or 1). Pull the cards with label 0 to top of the deck preserving their relative order. The cards with label 1 will stay at the bottom preserving their relative order.

[5] For more details please check [1], Chapter 8, Section 8.3 ,[7] or [9]

Considering definition (II) it is not hard to see that unshuffle is the reverse of shuffle.

**Lemma 4.1.** *Let $\tau$ to be an optimal stopping time for Riffle shuffle and $\bar{\tau}$ to be an optimal stopping time for unshuffle. We have $E[\tau] = E[\bar{\tau}] \leq 2\log n$.*

Proof. The following is an optimal stopping rule for unshuffle. Unshuffle the cards and at each step, keep track of the bits that are assigned to each card. After $t$ steps any card will be associated with a length $t$ binary number. Stop when all $n$ numbers are different. It is easy to check that this stopping rule generates the uniform distribution and the inverse of starting permutation is the halting state. Now, we should calculate the expectation of stopping time. Notice that we stop after $t$ steps if we got $n$ different numbers when we are allowed to choose from $\{0, 1, \ldots, 2^t - 1\}$. Therefore, we have an instance of the Birthday problem. We use the results from Birthday problem to bound the expected stopping time.
$Pr(\tau \leq t) = Pr(\text{we have } n \text{ distinct numbers in range } \{1, \ldots 2^t\}) = (1 - \frac{1}{2^t})(1 - \frac{2}{2^t}) \ldots (1 - \frac{n}{2^t}) \simeq \Pi_{i=1}^{n}(e^{-i/2^t}) = e^{(-\frac{1}{2^t})\sum_{i=1}^{n} i} \simeq e^{(n^2/2^t)}$.

Therefore, we have $E(\tau) = \sum_{t=1}^{\infty} Pr(\tau \geq t) \simeq \sum_{t=1}^{\infty}(1 - e^{(n^2/2^t)}) = \sum_{t=1}^{\log n^2}(1 - e^{(n^2/2^t)}) + \sum_{t=\log n^2+1}^{\infty}(1 - e^{(n^2/2^t)}) \leq \log n^2 - \sum_{\log n+1}^{\infty}(n^2/2^t) \simeq 2\log(n)$.

Using the result from Lovász and Winkler ([8])[6], we know that the above exact mixing time for unshuffle is also the exact mixing time of shuffle. □

**Corollary.** For the Riffle shuffle the conventional mixing time is bounded by $2\log n$.

Proof. Let $X^t$ be distribution of the deck after $t$ Riffle shuffles. Since $\tau$ is strong stationary time, we have $||X^t - U|| \leq Pr(\tau > t)$. For $t = \log(n^2/c)$, we have $Pr(\tau > \log(n^2/c)) \simeq 1 - e^c \leq c$. □

## 4.3 Mixing time of Riffle shuffle. Is seven shuffles enough?

Now, we know the exact mixing time of the Riffle shuffle and it also gives us a bound on mixing time. However, our discussion will not be complete unless we discuss the famous paper of Bayer and Diaconis, "Trailing the Dovetail shuffle to its lair"([7]). This result has been in the news[7] and famous as "In shuffling cards, 7 is the wining number". As you will see in the

[6]The exact mixing time of a chain and its reversal are equal.

[7]Kolata, Gina (January 09, 1990). "In Shuffling Cards, 7 Is Winning Number". New York Times. Retrieved 2012-11-14.

following, the fact that seven shuffles is enough to make a card random has been questioned a lot. We will also discuss how number "seven" might be significant in shuffling 52 cards although it is NOT specifically discussed in the paper. Here is the most important theorem in [7].

**Theorem 4.2.** *If $n$ cards are shuffled $m$ times, then the chance that the deck is in arrangement $\pi$ is $\binom{2^m+n-r}{n}/2^{mn}$, where $r$ is the number of rising sequences in $\pi$.*

Sketch of Proof. First consider a generalization of Riffle shuffle to $a$-shuffle where the deck is cut to $a$ piles and then the piles will be interleaving into each other. Definition 4.1 will have the following formulation in general case.

**Definition 4.4.** The following four modelings of $a$-shuffle are equivalent:

I. Let $m_1, \ldots, m_a$ be taken randomly from multinomial $(n, 1/a, \ldots, 1/a)$. Let a card drop from pile $i$ with probability $x_i/(x_1 + \cdots + x_a))$ , where $x_i$ is the number of cards left in pile $i$.

II. Let $m_1, \ldots, m_a$ be taken randomly from multinomial $(n, 1/a, \ldots, 1/a)$. Split the card to piles of size $m_1, \ldots, m_a$. Choose one of the $\binom{n}{m_1,\ldots,m_a}$ possible arrangements of these card uniformly at random.

III. Place $n$ points $x_1, \ldots, x_n$ uniformly and independently in unit interval. Assign the cards in their order to $x_1, \ldots, x_n$. Apply the mapping $x \to ax$ to the points. Rearrange the cards according to the new ordering.

Considering part III of the definition we can see an $ab$-shuffle is equivalent to performing an $a$-shuffle first and then a $b$-shuffle. As a result, in order to find out the distribution after $k$, 2-shuffles, it will be enough to calculate the distribution of the deck after one $2^m$ shuffle. The proof of Theorem 4.2 will then be a result of the following lemma[8]:

**Lemma 4.3.** *If an $a$-shuffle is performed on a deck of $n$ cards, then the chance that the deck is in arrangement $\pi$ is $\binom{a+n-r}{n}/a^n$, where $r$ is the number of rising sequences in $\pi$.*

[8]For more details please read [7] and [9]. □

Proof of lemma: Any rearrangement consists of a cut and interleave. The probability of a cut and interleave is $\frac{1}{\binom{n}{m_1,\ldots,m_a}} \times \binom{n}{m_1,\ldots,m_a}/a^n = 1/a^n$. Therefore, it suffices to count the number of rearrangements that will generate $\pi$. Notice that when the cut is specified the interleaving will be forced and each of the rising sequences is a union of some piles. Therefore, $r-1$ cuts are forced. For the rest $a-r+1$, the number of possible cuts will be $\binom{a+n-r}{n}$ $\square$.

**Corollary.** Let $Q^t$ be the distribution of the deck after $t$ shuffles. As a result of Theorem 4.2 we have

$$||Q^t - U|| = \frac{1}{2}\sum_{r=1}^{n} A_{n,r}\left|\binom{2^t+n-r}{n}/2^{tn} - \frac{1}{n!}\right|$$

Where $A_{n,r}$ is the number of permutations of $1\ldots n$ with $r$ rising sequences which is known as Eulerian numbers.

Having the above formulation, for arbitrary $n$, Bayer and Diaconis approximate the deviation distance and they observe that it does not change much till it reaches the point $t = \frac{3}{2}\log n$ then at this point it drops in a considerable amount. They call this phenomenon, the *cut off* phenomenon which is a more powerful concept than mixing. You can read more about the cut off phenomenon in Chapter 18 of [1] and also [7].

**Note.** For a deck of 52 cards we can calculate the deviation distance after $t = 1,\ldots 10$ Riffle shuffles. The following table can be found in [7].

| t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| dv(t) | 1.0 | 1.0 | 1.0 | 1.0 | 0.92 | 0.61 | 0.33 | 0.16 | 0.8 | 0.04 |

We notice that before the 7th step the deviation distance is not changing much. However, at the 7th shuffle it almost halves and it continues getting half of its previous amount after each single shuffle. Therefore, time 7 can be considered as the "cut off" point. This might be the reason that the result has been famous as "seven shuffles is enough". However, we definitely know that the distribution after 7 shuffles is still far from being uniform. Also, note that there are 11 steps needed if you consider exact mixing. In this regard, Peter Boyle[9] has constructed a game on which one's chance of winning after seven shuffles is 0.8 although it is 0.5 with uniform deck.

[9]You can read about Doyle's game of New Age Solitaire in [9]

# 5 Conclusion and other works

There are some other papers of Diaconis et al looking at the Riffle shuffle when we have repeated cards([10], [11]). This is specially interesting because in most of the card games some of the cards are treated equally. They show that if the deck consists of two different type of cards, the mixing time will be $\log n + c$[10]. Specifically, for a deck of 52 cards 4 shuffles will be enough[11].

What came here was a very concise talk about two well known shuffles. There are still much more ways of shuffling that have been studied by mathematicians. The interested reader should also check out the random transposition(Chapter 8 of [1]) and Thrope shuffle([12]). Also, if you want to know more about Riffle shuffle, check Persi Diaconis's website[12].

# References


[1] D. Levin, Y. Peres, E. Wilmer (2008), Markov Chains and Mixing Times, *American Mathematical Society Press.*

[2] D. Kandel, Y. Matias, R. Unger and P. Winkler (1996), Shuffling Biological Sequences, *Disc. Appl. Math.* **71**.

[3] R. Durret (2003), Shuffling Chromosomes, *J. Theoret. Probab.* **16**, 725-750.

[4] B. Morris (2012), P. Rogaway and V. T. Hoang An Enciphering Scheme Based on a Card Shuffle. *CRYPTO.*

[5] L. Lovász and P. Winkler (1998), Mixing times, *Microsurveys in Discrete Probability,* D. Aldous and J. Propp, eds., *DIMACS Series in Discrete Math. and Theoretical Computer Science* **41**, Amer. Math. Soc., Providence RI, pp. 85-134

[6] D. Aldous, P. Diaconis (1986), Shuffling Cards and Stopping Times. *Math'l Monthly*, **93(5)**:333-348.


[10] [10]

[11] Shuffling the cards: Math does the trick". Science News. Friday, November 7, 2008. Retrieved 14 November 2008. "Diaconis and his colleagues are issuing an update. When dealing many gambling games, like blackjack, about four shuffles are enough.

[12] http://www-stat.stanford.edu/ cgates/PERSI/index.html

[7] D. Bayer, and P. Diaconis (1992), Trailing the Dovetail shuffle to its lair, *Ann. Appl. Probab* **2**, 294-313.

[8] L. Lovász and P. Winkler (1998), Reversal of Markov chains and the forget time, *Combinatorics, Probability and Computing* **7:1**, pp. 189-204.

[9] B. Mann (1995), How many times should you shuffle a deck of cards? *In Topics in Contemporary Probability and its Applications* (J.L. Snell, ed.), CRC Press, Boca Raton.

[10] P. Diaconis, K. Assaf, S. Soundararajan (2008), A Rule of Thumb for Riffle Shuffling. *Annals of Applied Probability*, **21**(3):843-875. arXiv:0908.3462v1.

[11] P. Diaconis, K. Assaf, S. Soundararajan (2009), Riffle shuffles of a deck with repeated cards. *DMTCS Proceedings, 21st International Conference on Formal Power Series and Algebraic Combinatorics* (FPSAC 2009),**0**(1):89-102

[12] B. Morris (2005), The mixing time of the Thorp shuffle. *SIAM journal on computing, STOC.*