Mixing Time Estimates for the Riffle Shuffle

From factory order to (almost) complete randomness.

Shihan Kanungo

Euler Circle, Palo Alto, CA 94306

Abstract

In this paper we will talk about one of the most well-known shuffling methods, called the "riffle shuffle" or "dovetail shuffle". We are interested in the number of shuffles that will make the deck of n cards well-mixed, or "close" to uniformly random.

1 Motivation

One may ask, "Does it matter?" It seems to many people that if a deck of cards is shuffled 3 or 4 times, it will be quite mixed up for all practical purposes. However, there are patterns that will remain in the deck, unseen to the untrained eye. Magicians and card cheats have long taken advantage of such patterns. Suppose a deck of 52 cards in known order is shuffled 3 times and cut arbitrarily in between these shuffles. Then a card is taken out, noted and replaced in a different position. The noted card can be determined, quite unbelievably, with near certainty! Gardner (1977) describes card tricks based on the inefficiency of too few riffle shuffles. What is even more surprising is that a mathematical analysis of this seemingly innocuous question is linked to a wide range of mathematical questions, in areas such as Lie algebras, Hochschild homology, and random walks on graphs, and has been generalized in significant ways.

2 Introduction

It is natural to ask how many times a deck must be shuffled to mix it up? But first we must decide what we wish to achieve by shuffling. The answer is to get very close to total randomness, or, in other words, very close to the uniform density U on S_n, each permutation $\sigma \in S_n$ having probability

$$U(\sigma) = 1/|S_n| = 1/n!.$$

So when we ask how many times we need to shuffle, we are not asking how far to go in order to achieve total randomness, which would take infinite time, but rather to get close enough to randomness. So we must define what we mean by close, or far, i.e. we need a distance between densities.

Definition 2.1. Let P, Q be probability distributions on S_n. We define their *total variation distance* to be

$$\|P - Q\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\pi \in S_n} |P(\pi) - Q(\pi)|.$$

In this paper, we will use the total variation to measure the distance between two distributions. We will say that the distributions P and Q are "close" if $\|P-Q\|_{\mathrm{TV}}$ is small, say less than $\frac{1}{4}$.

3 What is a shuffle, really?

We define a shuffle, or method of shuffling, as a probability density on S_n, considering each permutation as a way of rearranging the deck. This means that each permutation is given a certain fixed probability of occurring, and that all such probabilities add up to one.

3.1. A few well known shuffles

Here are some examples of common shuffling methods.

Top-in at random shuffle. For each step of the shuffle, take the top card and insert it at a uniformly chosen random position.

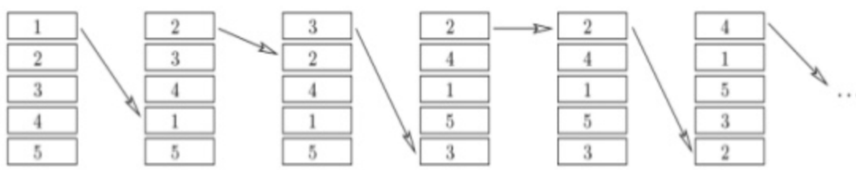


Figure 1. Repeated top-in at random shuffles of a 5-card deck

This means that the density on S_n is given by $1/n$ for each of the cyclic permutations

$$[23\cdots k\,1\,(k+1)\cdots n] \quad \text{for } 1 \le k \le n$$

and 0 for all other permutations. Table 1 lists this for a 3-card deck.

Permutation	[123]	[213]	[231]	[132]	[321]	[312]
Probability of occuring	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0	0

Table 1. Probability distribution of a 3-card deck after a single top-in at random shuffle.

We have seen in the notes (Week 7) that about $n\sum_{k=1}^{n-1}\frac{1}{k} \approx n\log n$ shuffles using the top-in method is sufficient to get the deck close to the uniform distribution.

The random transpositions shuffle. Here, the right and left hand independently each pick a uniformly chosen card. Then, unless both hands have chosen the same card, the two cards are interchanged. Figure 2 gives an example.

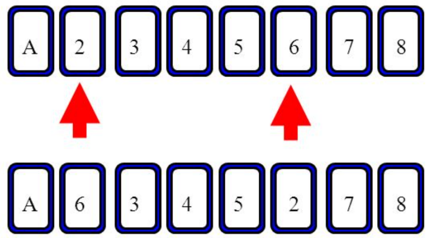


Figure 2. Random transposition shuffle of a 8-card deck

The overhand shuffle. This is our first model of a "real" shuffle, in the sense that people actually use it to mix a deck of cards.

An overhand shuffle transfers a few cards at a time from the dealer's right hand to his left. The dealer slides a few cards from the top of the deck in his right hand into his left hand. The process is repeated until all cards in his right hand are transferred. Consequently, cards that started near the top of the deck end up near the bottom, and vice versa. Mathematically,

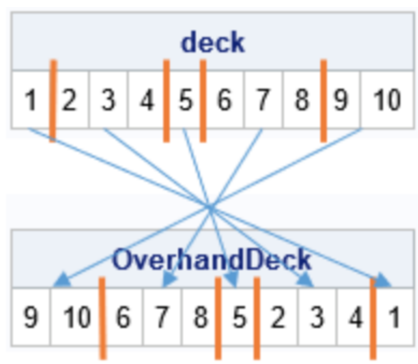


Figure 3. Overhand shuffle of a 10-card deck

you can model an overhand shuffle by randomly choosing k cut points that separate the deck into $k+1$ "packets" of contiguous cards. The size of a packet can vary, as can the number of packets. The overhand shuffle reverses the order of the packets while preserving the order of the cards within each packet. When the packet sizes are small, the cards are mixed better than when the packets are bigger.

Example. 1|234|5|678|90 results in 90|678|5|234|1.

3.2. Convolution of shuffles

What this definition of shuffle leads to, when the deck is repeatedly shuffled, is a random walk on the group of permutations S_n. Suppose you are given a method of shuffling Q, meaning each permutation π is given a certain probability $Q(\pi)$ of occuring. Start at the identity of S_n. Now take a step in the random walk, which means choose a permutation π_1 randomly, according to the probabilities specified by the density Q. Rearrange the deck as directed by π_1. Now repeat the procedure for a second step in the random walk, choosing another permutation π_2, again randomly according to the density Q (i.e. π_2 is a second, independent random variable with the same density as π_1). Rearrange the deck according to π_2. The effective rearrangement of the deck including both permutations is given by $\pi_1 \circ \pi_2$.

What is the probability of any particular permutation now, i.e what is the density for $\pi_1 \circ \pi_2$? Call this density $Q^{(2)}$. To compute it, note the probability of π_1 being chosen, and then π_2, is given by $Q(\pi_1) \cdot Q(\pi_2)$, since the choices are independent of each other. So for any particular permutation π, $Q(^{(2)}(\pi)$ is given by the sum

$$Q^{(2)}(\pi) = \sum_{\pi=\pi_1\circ\pi_2} Q(\pi_1)Q(\pi_2) = \sum_{\pi_1} Q(\pi_1)Q(\pi_1^{-1} \circ \pi)$$

This way of combining Q with itself is called a *convolution* and written $Q * Q = Q^{(2)}$. More generally, we may let each step be specified by a different density, say Q_1 and then Q_2. Then the resulting density is given by the convolution

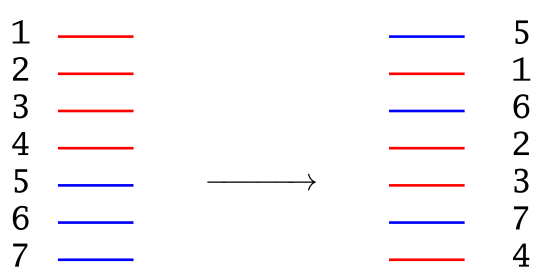
$$Q_1 * Q_2(\pi) = \sum_{\pi=\pi_1\circ\pi_2} Q_1(\pi_1)Q_2(\pi_2) = \sum_{\pi_1} Q_1(\pi_1)Q_2(\pi_1^{-1} \circ \pi)$$

In short, repeated shuffling corresponds to convoluting densities.

4 The Riffle Shuffle

The Riffle Shuffle is a popular way to shuffle a deck of cards. The shuffler divides the deck into two parts. The top part of the deck is placed in the left hand, the remaining stays in the right hand. Cards are then alternately interleaved from the left and right hands, but not necessarily perfectly, so several cards may be dropped from each hand at a time. This process is repeated a few times.

Figure 4 gives the result of a single riffle shuffle of a 7 card deck in the usual $i \to \pi(i)$ format



i	π(i)
1	2
2	4
3	5
4	7
5	1
6	3
7	6

Figure 4. A single riffle shuffle of a 7-card deck in the usual $i \to \pi(i)$ format

This shuffle is the result of cutting 4 cards off the top of a 7-card "deck" and riffling the packets together, first dropping card 4, then 7, followed by 3 and so on.

The "signature" of a riffle is provided by the rising sequences of the resulting permutation.

Definition 4.1. A *rising sequence* in a permutation is the maximal set of consecutive numbers that occur in the correct order. We denote by $r(\pi)$ the number of rising sequences in a permutation π.

Example. In the ordering $(2, 3, 5, 1, 4, 7, 6)$ there are 4 rising sequences

$$(1), (2, 3, 4), (5, 6), \text{ and } (7).$$

Since 2 is before 1, we see that the rising sequence starting at 1 is just (1). Then, starting at 2 we see that 3 occurs after 2 and 4 occurs after 3, but 5 occurs before 4, so the rising sequence starting at 2 ends at 4. Similarly we get $(5, 6)$ and (7) as the other rising sequences.

Example. The permutation in the figure has two rising sequences

$$\pi(1) < \pi(2) < \pi(3) < \pi(4) \quad \text{and} \quad \pi(5) < \pi(6) < \pi(7)$$

corresponding to the red pile and the blue pile because the cards in the red (resp. blue) pile retain their relative ordering in the shuffled deck. Equivalently, writing the resulting sequence as $(5, 1, 6, 2, 3, 7, 4)$ we get the two rising sequences $\{5, 6, 7\}$, and $\{1, 2, 3, 4\}$. This is immediately clear if we write the order of the deck this way,

$$516237,$$

indicating the pile each card came from.

The reason why rising sequences are essential to our analysis is because when we perform a shuffle, we can only at most double $r(\pi)$. The two rising sequences correspond to the two piles that started off the shuffle and ended up generating the final permutation π.

Note that rising sequences form a partition of the card labels $1, \ldots, n$. In general, a permutation π of n cards made by a single riffle shuffle will have exactly 2 rising sequences (unless it is the identity, which has 1). Conversely, any permutation of n cards with 1 or 2 rising sequences can be obtained by a physical riffle. Thus the mathematical definition of a *riffle shuffle* is "a permutation with 1 or 2 rising sequences".

Example. The density on S_n for a deck of size $n = 3$ after a single riffle shuffle is given by:

Permutation	[123]	[213]	[231]	[132]	[312]	[321]
Probability of occuring	$\frac{1}{2}$		$\frac{1}{8}$			0

Table 2. Probability distribution of a 3-card deck after a single riffle shuffle.

5 Modeling the Riffle

5.1. Gilbert-Shannon-Reeds model

The *Gilbert-Shannon-Reeds model* (GSR) is the first mathematically precise model of shuffling. It describes the most common strategy of card shuffling: a deck is cut into two heaps by the binomial distribution, meaning the probability of the cut occurring exactly after k cards is given by $2^{-n}\binom{n}{k}$. Then the cards are riffled together according to the rule: if the left packet has ℓ cards and the right has r cards, drop the next card from the left packet with probability $\ell/(\ell+r)$ (and from the right packet with probability $r/(\ell+r)$). Continue until all cards have been dropped.

5.2. Other Equivalent Models

The Gilbert-Shannon-Reeds model for shuffling has alternate descriptions, and a natural generalization to shuffles that begin with the deck being cut into a packets, with $a \geq 2$; the various packets are then riffled together.

Geometric Distribution. The geometric model begins by placing n points, uniformly and independently in the unit interval. The points are labeled in the order $x_1 < \cdots < x_n$. For positive integer a, the map $x \mapsto ax \pmod 1$ maps $[0,1]$ onto itself and preserves measure. This map rearranges the points x_i and so gives a measure on the symmetric group which will be called an a-shuffle.

A 2-shuffle is like an ordinary riffle shuffle: Points in $[0, \frac{1}{2}]$ and points in $[\frac{1}{2}, 1]$ are stretched out and interlaced.

Maximum Entropy Description. All possible ways of cutting a deck into a packets and then interleaving the packets are equally likely. Empty packets are allowed.

Inverse Description. All possible ways of pulling a shuffled deck back apart into a packets are equally likely. Empty packets are allowed. The following generates an inverse a-shuffle with the correct probability: A deck of n cards is held face down. Successive cards are turned face up and dealt into one of a piles uniformly and independently. After all cards have been distributed, the piles are assembled from left to right and the deck is turned face down.

Sequential Description. Choose integers $j_1, \ldots, j_a$ according to the multinomial distribution

$$\mathbb{P}(j_1, \ldots, j_a) = \binom{n}{j_1, \ldots, j_a} \frac{1}{a^n}$$

Given j_i cut off the top j_1 cards, the next j_2 cards and so on, producing a or fewer packets. Shuffle the first two packets using the GSR shuffle described in Section 4.1. Then shuffle this combined packet with packet 3, and so forth. This is equivalent to riffling all a packets together at once, where if there are A_i cards remaining in each heap, the chance that the next card will drop from heap i is $A_i/(A_1 + \cdots + A_a)$.

Lemma 5.1. *The four descriptions above generate the same permutation distribution.*

Proof. Left as an exercise. See Bayer and Diaconis [2] ■

Since one riffle shuffle gives rise to exactly one or two rising sequences, it should be clear that when we riffle shuffle t times, we obtain an ordering with at most 2^t rising sequences. In fact, it is not hard to see that each such ordering is the result of some sequence of t shuffles. The question becomes, then, in how many ways can an ordering with r rising sequences come about by applying t riffle shuffles to the identity ordering? In order to answer this question we turn to the ingenious idea of a-shuffles.

6 Probabilities of reaching a specific permutation

A generalization of the standard riffle shuffle to a-shuffles is introduced by Bayer-Diaconis.

Definition 6.1. For a deck of n cards, denote the a-shuffle in the following way: Take a stack of cards. Cut it into a packets by a multinomial distribution. Then, drop the cards from these packets with probability proportional to packet size: let b_i be the number of cards in the packet i at any moment; then the chance that the next card dropped will be from this packet is

$$\frac{b_i}{\sum_{j=1}^{a} b_j}$$

Considering the *geometric description*, we can see that an ab-shuffle is equivalent to performing an a-shuffle first and then a b-shuffle. Letting $Q_a(\sigma)$ denote this measure, we thus have

Theorem 6.2 (The multiplication theorem). *An a-shuffle followed by a b-shuffle is equivalent to a single ab-shuffle, in the sense that both processes give exactly the same resulting probability density on the set of permutations. In other words,*

$$Q_a * Q_b = Q_{ab}. \tag{1}$$

Thus it is enough to study a single a-shuffle. In order to find out the distribution after m consecutive 2-shuffles, it will be enough to calculate the distribution of the deck after a single 2^m shuffle.

There is another great advantage to considering a-shuffles. It turns out that when you perform a single a-shuffle, the probability of achieving a particular permutation π does not depend upon all the information contained in π, but only on the number of rising sequences that π has. In other words, we immediately know that the permutations [12534], [34512], [51234], and [23451] all have the same probability under any a-shuffle, since they all have exactly two rising sequences. Here is the exact result:

Theorem 6.3. *If an a-shuffle is performed on a deck of n cards, then the probability that it will result in a specific permutation π is*

$$\frac{1}{a^n}\binom{a+n-r(\pi)}{n},$$

where $r(\pi)$ is the number of rising sequences in π.

Proof. Using the *maximum entropy description*, this probability is determined by the number of ways of cutting an ordered deck into a packets, such that π is a possible interleaving. Because each packet stays in order as the cards are riffled together, each rising sequence in the shuffled deck is a union of packets. Thus, we want to count the number of ways of refining r rising sequences into a packets.

We emulate the classical "stars and bars" argument, counting arrangements of cuts on the ordered deck before shuffling: At least one cut must fall between each successive pair of rising sequences of π, but the remaining cuts can be located arbitrarily. Thus, the n

cards form dividers creating $n+1$ bins, into which the $a-r$ spare cuts are allocated. There are $\binom{a+n-r(\pi)}{n}$ ways of doing this. There are a^n possible a-shuffles in all, giving the stated probability. ■

Notice that a riffle shuffle is just a 2-shuffle, i.e. an a-shuffle with $a=2$. An application of Theorem 6.3 along with equation (1) then gives

Theorem 6.4. *If n cards are riffle shuffled t times, then the probability that it will result in a specific permutation π is*

$$\frac{1}{2^{tn}}\binom{2^t+n-r}{n},$$

where $r=r(\pi)$ is the number of rising sequences in π.

7 Putting It All Together

Let R^t be the distribution of the deck after t riffle shuffles. This is just t 2-shuffles, one after another. So by the multiplication theorem, this is equivalent to a single $2\cdot 2\cdots 2=2^t$ shuffle. Hence, in the R^t density, there is a $2^{-tn}\binom{2^t+n-r}{n}$ chance of a permutation with r rising sequences occurring, by our rising sequence formula. This now allows us to work on the variation distance $\|R^t-U\|$. For a permutation π with r rising sequences, we see that

$$|R^t(\pi)-U(\pi)| = \left|\frac{1}{2^{tn}}\binom{2^t+n-r}{n} - \frac{1}{n!}\right|$$

We must now add up all the terms like this, one for each permutation. We can group terms in our sum according to the number of rising sequences. The Eulerian numbers $A_{n,r}$ stand for the number of permutations of n cards that have exactly r rising sequences. Then the variation distance, in terms of the $A_{n,r}$ is given by

Theorem 7.1. *The variation distance $\|R^t-U\|_{\mathrm{TV}}$ between the t-th iterate of the riffle shuffle and the uniform probability distribution, U on S_n is given by*

$$\|R^t-U\|_{\mathrm{TV}} \;=\; \frac{1}{2}\sum_{r=1}^{n} A_{n,r}\left|\frac{1}{2^{tn}}\binom{2^t+n-r}{n} - \frac{1}{n!}\right| \tag{2}$$

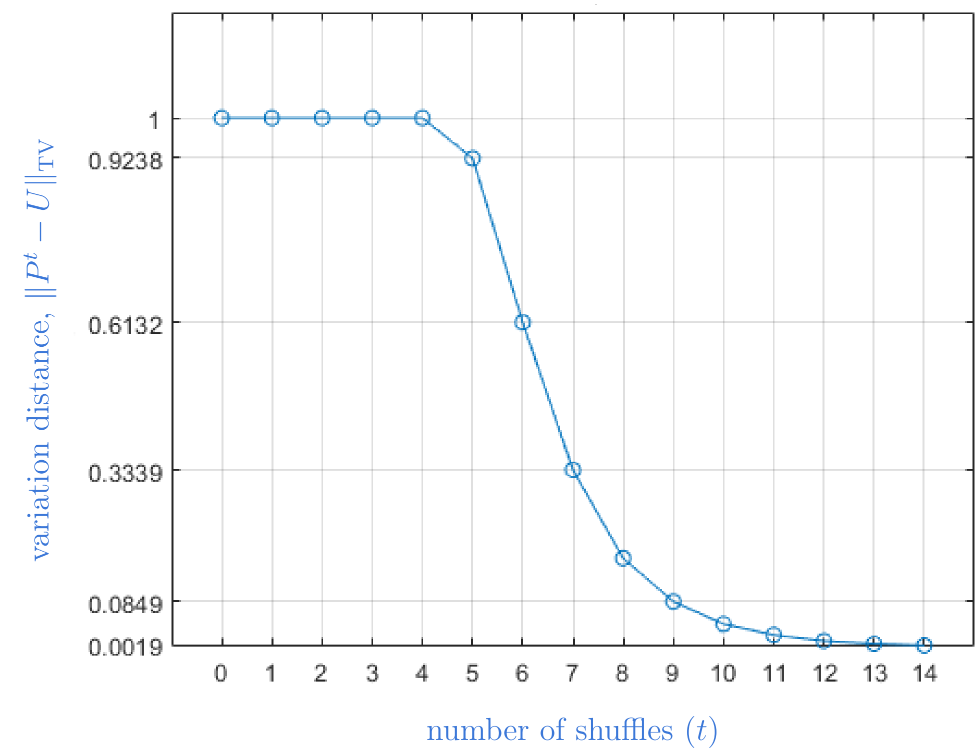
where the coefficients $A_{n,r}$ are the so-called Eulerian numbers; it is the number of permutations of $\{1,\ldots,n\}$ with r rising sequences. They can be obtained by the recursive formula $A_{n,1}=1$ and

$$A_{n,r} = r^n - \sum_{j=1}^{r-1}\binom{n+r-j}{n}A_{n,j}$$

Example. It is interesting to note that the Eulerian numbers are symmetric in the sense that $A_{n,r}=A_{n,n-r+1}$. So there are just as many permutations with r rising sequences as there are with $n-r+1$ rising sequences.

The expression (2) for total variation distance, although quite formidable, is well within the reach of computer evaluation and it is possible to graph $\|R^t - U\|_{\mathrm{TV}}$ versus t for any specific, moderately sized n. Even on the computer, however, this computation is tractable because we only have n terms, corresponding to each possible number of rising sequences. If we did not have the result on the invariance of the probability when the number of rising sequences is constant (Theorem 6.3), we would have $|S_n| = n!$ terms in the sum. For $n = 52$, this is approximately 10^{68}, which is much larger than any computer could handle.

Here is the graphical result of the calculations for $n = 52$. The horizontal axis is the number of riffle shuffles, and the vertical axis is the variation distance to uniform.



It is seen that $\|R^t - U\|_{\mathrm{TV}}$ is above 0.9 for $k \leq 5$, then decreases abruptly and is below 0.1 for $k = 10$, quickly approaching zero afterward. A good middle point seems to be $k = 7$, which justifies the claim that 7 shuffles are enough.

For arbitrary n, Bayer and Diaconis [2] do some asymptotic analysis which shows that when n, the number of cards, is large, approximately $t = \frac{3}{2}\log_2 n$ shuffles suffice. Around this point the variation distance drops by 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 [2] and also in Chapter 18 of [1].

8 Approach to Uniformity in the GSR Shuffling Model

Bayer and Diaconis [2] analyze the approach to uniformity for the GSR model. They derive approximations when n is large after m shuffles, with $m = \frac{3}{2}\log n + \theta$. For notational convenience, write $m = \frac{3}{2}\log(n^{3/2}c)$, so $c = 2^\theta$ satisfies $0 < c < \infty$. The arguments use the asymptotics of Eulerian numbers. Asymptotics and exact results are compared at the end of this section. They prove that this distribution converges to a uniform one. The next result is the main theorem of their paper.

Theorem 8.1. *Let Q^m be the Gilbert-Shannon-Reeds distribution on the symmetric group S_n, and U be the uniform distribution. Then for $m = \log_2(n^{3/2}c)$, with $0 < c < \infty$ fixed, as $n \to \infty$,*

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

with

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

This theorem also allows a corollary:

Corollary 8.2. *If n cards are shuffled m times with $m = \frac{3}{2}\log_2(n) + \theta$, then for large n,*

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

Therefore, if θ is large, the distance to uniformity approaches 0, while for θ small, it approaches 1. We can calculate the different variation distances for distinct numbers of cards. Then, we see that about $\frac{3}{2}\log_2 n$ shuffles are necessary for shuffling n cards.

TABLE 3
Total variation distance for m shuffles of* 25, 32, 52, 78, 104, 208 *or* 312 *distinct cards

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 3 gives exact computations of variation distance for a number of popular deck sizes. Each deck size shows the cutoff phenomenon. Variation distance decreases by a factor of 2 after each shuffle following the cutoff. For comparison, Table 4 gives $\frac{3}{2}\log n$ for these deck sizes.

TABLE 4
Shuffles needed to mix* 25, 32, 52, 78, 104, 208 *or* 312 *cards

n	25	32	52	78	104	208	312
$\frac{3}{2}\log_2 n$	6.97	7.50	8.55	9.43	10.05	11.55	12.43

9 So, which shuffle is the best?

We present below the number of repetitions required, on average, to sample a (almost) uniform permutation – say, within 20% of the uniform distribution.

Type of shuffle	repetitions	τ_{mix}
Random Transpositions	~ 100	$\tau_{\text{mix}} \sim \frac{1}{2} n \ln n$
Top-in at Random	~ 300	$\tau_{\text{mix}} \leq \frac{1}{2} n \ln n + \lceil (1 + \ln 2) n \rceil$
Overhand Shuffle	~ 2500	$\tau_{\text{mix}} \sim n^2 \ln n$
Riffle Shuffle	~ 8	$\tau_{\text{mix}} \sim \frac{3}{2} \ln n$

.

.

References

[1] D. A. Levin, Y. Peres and E. L. Wilmer. Markov Chains and Mixing Times

[2] D. Bayer and P. Diaconis. (1992) "Trailing the Dovetail Shuffle to its Lair" *Ann. Applied Prob.* **2** 294-314

[3] 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

[4] C. Moler. (2016), How Many Times Should You Shuffle the Cards? *Cleve's Corner: Cleve Moler on Mathematics and Computing*, The MathWorks, Inc., 15 Feb. 2016, `blogs.mathworks.com/cleve/2016/02/15/how-many-times-should-you-shuffle-the-cards-2/`

[5] R, Pemantle. 2020. Randomization Time For The Overhand Shuffle. *Journal of Theoretical Probahility*, Vol.2, **No.1**, 1989 `www.math.upenn.edu/~pemantle/papers/overhand2.pdf`

[6] S. Rubinstein-Salzedo, Markov Chains, *self-published*