Shuffling Cards

Utrecht University
Mathematics
Bachelor's Thesis

D.J.W. Telkamp

Supervised by
Dr. K. Dajani

June 13, 2017

Contents

1 Introduction

Most people have played a decent amount of card games in their lives. Many shuffling methods have been shown to those occassional card players at some point (or the reader nay have been depended on other people to shuffle the cards). Certain shuffles of a deck of cards might have always seemed better than others. Is this truly the case? And what makes a shuffle "better than" another shuffle? What is the "fastest" way to mix up a deck of cards while preparing a card game? These seem to be natural questions, which most people do not worry about. However, It might make a game of cards (very) unfair. To answer all these questions, I will dive into the world of card decks, shuffles and "randomness". For this, a profound analysis of the mathematics behind these mentoined aspects is needed. This will all be presented by brightening examples, sometimes with numerical help.

2 Prerequisites

In order to establish sufficient shuffling techniques, we need to look into the mathematical tools needed for shuffling cards. A normal deck of cards contains 52 cards: 13 cards in 4 different kinds. But for an analysis of shuffling, we can also work with a deck of n cards. From now on, label the cards in order 1 to n, according to the order in which you find them. With [123], I mean the order of the cards in which one finds them (from left to right or from top to bottom). If $n > 10$, numbers will be separated like this: [(1)(2)(3)(4)...].

Example 1. Suppose we have all the cards of hearts from a regular deck, that is (in order):
$A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K$. We will number the numbered cards as they are and further: $J \to 11$, $Q \to 12$ $K \to 13$ and $A \to 1$. So the original deck is: [(1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12)(13)]. Suppose we mix up the cards and get the new order: $5, 9, 2, A, J, 3, 7, Q, 4, 6, 8, K, 10$. This corresponds to the deck ordering: [(5)(9)(2)(1)(11)(3)(7)(12)(4)(6)(8)(13)(10)]. We can also show this as a permutation (from the original deck):

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 5 & 9 & 2 & 1 & 11 & 3 & 7 & 12 & 4 & 6 & 8 & 13 & 10 \end{pmatrix}$$

Here, the permutation indicates that the card on place 1 was on place 5 before the reordering of the cards.

In the example, I spoke of "mix up the cards". I will now make more precise what I mean by this, through the notion of a shuffle.

Definition 1. Suppose we have a deck of n cards. A *shuffle* is a probability density on S_n.

Without using an aspect of randomness, shuffling cards would of course not be so fair (read: interesting). shuffles will be represented by permutations. With the permutation e, I mean the identity. The permutation (1234) means sending 1 to 2, 2 to 3, 3 to 4 and 4 to 1. So:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} = (1234)$$

I give an example of a shuffle.

Example 2. Suppose we have a deck of 4 cards. We number the cards accordingly, in order: [1234]. Define $Q : S_4 \to [0, 1]$ as:

$$Q(\pi) \to \begin{cases} \frac{1}{2} & \text{if } \pi = (1234) \\ \frac{1}{2} & \text{if } \pi = e \\ 0 & \text{else} \end{cases}$$

So in this example, there are two options. The first option is getting the original deck back. The second option is getting the deck [2341]. Both options appear with probability $\frac{1}{2}$.

What do we want of a shuffling technique? Well preferably to be random or at least as random as possible. That means we want it to be as close as possible to the uniform density U on S_n. Where $U(\pi) = \frac{1}{n!}$ for all permutations π. To determine what we mean by closeness, we need to define a metric on the probability space (of shuffles). There are many options, but in this thesis one will be used.

Definition 2. Suppose we have two probability densities Q_1 and Q_2 on S_n. The *variation distance*, denoted $\|.\|$ is defined as follows:

$$\|Q_1 - Q_2\| = \frac{1}{2} \sum_{\pi \in S_n} \left| Q_1(\pi) - Q_2(\pi) \right|$$

Here the factor $\frac{1}{2}$ is to scale the variation distance between 0 and 1. An equivalent definition is $\|.\|_1$, which can be calculated accordingly:

$$\|Q_1 - Q_2\|_1 = \max_{S \subset S_n} \left| \sum_{\pi \in S} Q_1(\pi) - \sum_{\pi \in S} Q_2(\pi) \right|$$

We see that the variation distance between two probability densities varies from 0 to 1. To see this, define $A = \{\pi \in S_n \mid Q_1(\pi) \geq Q_2(\pi)\}$. Then:

$$2\|Q_1 - Q_2\| = \sum_{\pi \in A} Q_1(\pi) - Q_2(\pi) + \sum_{\pi \in \overline{A}} Q_2(\pi) - Q_1(\pi) \leq \sum_{\pi \in A} Q_1(\pi) + \sum_{\pi \in \overline{A}} Q_2(\pi) \leq 2$$

If $\|Q_1 - Q_2\| = 0$, then the probability densities are the same. If $\|Q_1 - Q_2\| = 1$, then they are nothing alike. Notice that not shuffling at all (permutation e with probability 1) has the following variation distance from U on S_n:

$$\frac{1}{2}\Big(\Big|1 - \frac{1}{n!}\Big| + (n! - 1)\Big|\frac{1}{n!} - 0\Big|\Big) = 1 - \frac{1}{n!}$$

We see that not doing anything quickly becomes "less random", if n is big. Furthermore, the two variation distances are in fact the same, as we will now see. The proof of this lemma is inspired by [11].

Lemma 1. *The variation distances $\|.\|$ and $\|.\|_1$ are indeed the same.*

Proof. Suppose that Q, R are probability densities on S_n. We need to verify that $\|Q - R\| = \|Q - R\|_1$. First, set $A = \{\pi \mid Q(\pi) \geq R(\pi)\}$. Take $S \subset S_n$, then define: $Q(S) = \sum_{\pi \in S} Q(\pi)$. So:

$$Q(S) - R(S) = Q(S \cap A) - R(S \cap A) + Q(S \cap \overline{A}) - R(S \cap \overline{A}) \leq Q(S \cap A) - R(S \cap A)$$

The inequality is achieved since the if $\pi \in S \cap \overline{A}$, then $Q(\pi) - R(\pi) \leq 0$, since $\pi \in \overline{A}$. Moreover, if $\pi_1 \in S \cap A$, then $\pi_1 \in A$. So $Q(\pi_1) - R(\pi_1) \geq 0$. Together with the previous inequality, this results in:

$$Q(S) - R(S) \leq Q(S \cap A) - R(S \cap A) \leq Q(A) - R(A)$$

In a similar way we also achieve $R(S) - Q(S) \leq R(\overline{A}) - Q(\overline{A})$:

$$R(S) - Q(S) = R(S \cap A) - Q(S \cap A) + R(S \cap \overline{A}) - Q(S \cap \overline{A}) \leq R(S \cap \overline{A}) - Q(S \cap \overline{A})$$

Hence:

$$R(S) - Q(S) \leq R(S \cap \overline{A}) - Q(S \cap \overline{A}) \leq R(\overline{A}) - Q(\overline{A})$$

Note now that $Q(A) - R(A) = R(\overline{A}) - Q(\overline{A})$, since:

$$0 = Q(S_n) - R(S_n) = Q(A) - R(A) - R(\overline{A}) + Q(\overline{A})$$

Because S was arbitrary, we can conclude:

$$\|Q - R\|_1 = \max_{S \subset S_n} \big|Q(S) - R(S)\big| \leq Q(A) - R(A)$$

And so:

$$\|Q - R\|_1 \leq \frac{1}{2}\Big(Q(A) - R(A) + R(\overline{A}) - Q(\overline{A})\Big) = \|Q - R\|$$

So we proved that $\|Q-R\|_1 \leq \|Q-R\|$. For the other way around, notice that (since $A \subset S_n$):

$$\|Q-R\| = \frac{1}{2}\Big(Q(A) - R(A) + R(\overline{A}) - Q(\overline{A})\Big) = Q(A) - R(A)$$

$$= \big|Q(A) - R(A)\big| \leq \max_{S \subset S_n} \big|Q(S) - R(S)\big| = \|Q-R\|_1$$

So $\|Q-R\| \leq \|Q-R\|_1$ and we conclude: $\|Q-R\| = \|Q-R\|_1$. □

2.1 Problems with the variation distance

The variation distance seems fine, but as Mann in [13] already points out: it has a slight problem. Suppose we have a deck of n cards, which is randomly shuffled. So, we apply the aforementioned probability function U. Suppose now that the top card accidentally falls off and is shown to you. You place the card back, but in the top half of the deck. By doing this, you get a new distribution V. The variation distance between U and V is $\frac{1}{2}$, since half of the permutations have probability $\frac{2}{n!}$ and the other half 0. This is because you put the card in the top half, so every permutation where the top card is in the bottom half of the deck, is not possible. For the other permutations, it is obvious that they have equal probability of appearing. It was a uniform shuffle and the card is uniformly put in the top half of the deck. So for every permutation π, we establish that $U(\pi)$ differs $\frac{1}{2(n!)}$ from $V(\pi)$. Hence $\|U-V\| = \frac{1}{2}$. If n is very large, knowing just one card does not really matter for randomness, but the spread between U and V is still $\frac{1}{2}$. On the other hand, simulations turn out that the variation distance function (which is also the maximal variation distance possible) can be too forgiving in certain games. Doyle created such game which is "as unfair as possible" for the Riffle shuffle, as explained in [13]. However, the variation distance appears to be the most useful method of establishing whether probability densities are close together. And, more importantly, it is a great tool to come to a conclusion whether or not a shuffle is "random enough".

2.2 Multiple successive shuffles

It is of course practically not possible to perform a uniform shuffle U to mix up the deck (or is it? See section 7). It is therefore that we explore practically useful shuffles and try to minimize the variation distance. For shuffles, we need to dive into multiple shuffles following each other. We will denote this by $Q^{(k)} = Q \circ Q \cdots \circ Q$ (k-times). This is easily extended from what we already know:

$$Q^{(k)}(\pi) = \sum_{\pi_1 \circ \cdots \circ \pi_k = \pi} Q(\pi_1) \circ \cdots \circ Q(\pi_k)$$

It is apparent that we sum over all combinations $(\pi_1, \ldots, \pi_k)$ such that: $\pi_1 \circ \cdots \circ \pi_k = \pi$. I give an example, using the example used earlier (example 2).

Example 3. Suppose we seek $Q^{(3)}$. For this, let us first determine $Q^{(2)}$:

$$Q^{(2)}(\pi) \to \begin{cases} \frac{1}{4} & \text{if } \pi = (13)(24) \\ \frac{1}{2} & \text{if } \pi = (1234) \\ \frac{1}{4} & \text{if } \pi = e \\ 0 & \text{else} \end{cases}$$

From this we can derive $Q^{(3)}$:

$$Q^{(3)}(\pi) \to \begin{cases} \frac{1}{8} & \text{if } \pi = (1432) \\ \frac{3}{8} & \text{if } \pi = (1234) \\ \frac{3}{8} & \text{if } \pi = (13)(24) \\ \frac{1}{8} & \text{if } \pi = e \\ 0 & \text{else} \end{cases}$$

Note that, for example, to get the permutation $x = (1234)$, we have the following possibilities: $x \circ e \circ e$, $e \circ x \circ e$ and $e \circ e \circ x$, all with probability $\frac{1}{8}$.

These calculations seem a bit gruesome and indeed they turn out to be. Luckily there are other options for some natural shuffles, as we will find out later (see for example section 4).

3 Motivation: let's perform a card trick!

Before continuing with an analysis of certain shuffling methods, let me first stop for a moment to appreciate what we are doing. Why is it so important to inspect shuffling methods used in practice? Well, some shuffles are easy to spot as "bad". Every shuffle without an element of uncertainty is obviously not good in games when cards of opponents should remain hidden. However, some shuffles which on the eye seem to be good, are not so good after all. Certain shuffles are nice for magicians to use in card tricks, for example because they appear to be more random than they are. I will dedicate this section to the analysis of the Gilbreath shuffle. This shuffle has a lot of possible permutations of the deck in his arsenal, but is not as random as it seems.

3.1 The setup

First of all, we need to define the Gilbreath shuffle.

Definition 3. Say we have a deck of n cards as usual: $[(1)(2)\dots n]$. Pick any number $j \in \{0, 1, \dots, n\}$. It does not really matter for the analysis in this section, but let's say that:

$$\mathbb{P}(j = k) = \frac{\binom{n}{k}}{2^n}$$

This seems as a fair choice (binomial distribution), but as said: it does not really matter for this section. Now distribute the first j cards one-by-one into a new pile. This reverses the order in this new pile. We now have two piles (in order): $[(j+1)(j+2)\dots(n)]$ and $[(j)(j-1)\dots(1)]$. After this, riffle the cards together. Now: form a new deck using the two piles, where the relative order of the cards inside the piles should not be changed. Choose uniformly between the ways of "riffling" the piles together. There are $\binom{n}{j}$ ways to do this. The now obtained deck has been "Gilbreath shuffled".

This shuffle has a lot of common with the Riffle shuffle (see section 4). It has some nice properties, but let me give an example first.

Example 4. Suppose $n = 4$. I will handle all options for the Gilbreath shuffle.

j	pile 1	pile 2	possible decks	permutation
0	[1234]	$\emptyset$	[1234]	e
1	[234]	[1]	[1234],[2134],[2314],[2341]	e,(12),(132),(1432)
2	[34]	[21]	[2134],[2314],[2341], [3421],[3241],[3214]	(12),(132),(1432), (1423),(143),(13)
3	[4]	[321]	[3214],[3241],[3421],[4321]	(13),(143),(1423),(14)(23)
4	$\emptyset$	[4321]	[4321]	(14)(23)

Notice that there is no option in which we can get the top 2 or bottom 2 cards such that the numbers have the same parity.

What is so interesting about this Gilbreath shuffle? Well, as we saw in the example above, there is no option to get the top or bottom pair of the new deck with two cards of the same parity. This results holds in general, as we now witness in the form of a theorem.

Theorem 1. *Say we have a deck of n cards: $[(1)\dots(n)]$. Let π be a permutation of the deck. The following properties are equivalent:*
1) π is obtained through a Gilbreath shuffle

2) For every $j \in \mathbb{N}$, the top j cards after applying the permutation are distinct modulo j
3) For every $j \in \mathbb{N}$ and $k \in \mathbb{N}$ with $jk \leq n$, the cards on the places: $\{(k-1)j + 1, (k-1)j + 2, \ldots, kj\}$ after the permutation π, are distinct modulo j.

This is also known as the Ultimate Gilbreath principle, but I left one equivalent property out, since it is irrelevant for the card trick (for the interested reader, see [7]). If we take $n = 4$, $j = 2$ and $k = 1$ or $k = 2$, then property (3) of theorem 1 shows that modulo 2, the first and last two card numbers after a Gilbreath shuffle should be distinct. Hence they should have different parity, which we indeed saw in example 4. For the proof of theorem 1, we need some notation which will be familiar. With $\pi(i)$, the card on place i after the permutation π is applied to the deck, is meant. So, in example 4, if $k = 4$, then: $\pi(1) = 4$, $\pi(2) = 3$, $\pi(3) = 2$ and $\pi(4) = 1$. This is because the new deck is in order: [4321].

To prove theorem 1, we need a lemma. To prove this lemma, an equivalent explanation of the Gilbreath shuffle is needed. As we saw in example 4, every possible deck/permutation appears twice. These doubles appear exactly at consecutive j_1 and j_2. More importantly, we can calculate the number of distinct possible decks:

$$\frac{1}{2}\sum_{k=0}^{n}\binom{n}{k} = \frac{2^n}{2} = 2^{n-1}$$

Using the argument of Diaconis in [7], we can make exactly 2^{n-1} distinct Gilbreath permutations (i.e. a permutation obtained by a Gilbreath shuffle). Of course, this implies an equivalent explanation of the Gilbreath shuffle, since there are only 2^{n-1} distinct Gilbreath permutations. For every $j \in \{1, \ldots n\}$, pick all $S = \{s_1, \ldots, s_{j-1}\} \subset \{2, \ldots, n\}$ with $|S| = j-1$. Here the s_i are labeled such that: $s_1 < s_2 < \cdots < s_{j-1}$. There are exactly $\binom{n-1}{j-1}$ such subsets for every j, since $\big|\{2, \ldots, n\}\big| = n - 1$. Then place j on position 1 and place $j - i$ on position s_i. The positions that are still not filled, should be filled in increasing order. That is, starting from $j + 1$ up until n.

First of all, this indeed creates distinct decks. Secondly, the number of decks are:

$$\sum_{j=1}^{n}\binom{n-1}{j-1} = \sum_{j=0}^{n-1}\binom{n-1}{j} = 2^{n-1}$$

Thirdly, every permutation created this way is indeed a Gilbreath permutation. This is because for certain j and $S = \{s_1, \ldots, s_{j-1}\}$, this is the same as splitting the deck in the piles: $[(j+1)(j+2)\ldots(n)]$ and $[(j)(j-1)\ldots(1)]$. And after this, riffling the packs exactly together as needed (which is obviously possible).

Concluding, because we can indeed form 2^{n-1} distinct Gilbreath permutations, we indeed have an equivalent way of defining the Gilbreath permutations on a deck. To illustrate this, let's go back to example 4 and find all distinct permutations.

Example 5. Take $n = 4$, as in example 4. We then find:

j	$S \subset$ {2,3,4} such that $\lvert S\rvert = j-1$	corresponding new decks
1	$\emptyset$	[1234]
2	{2},{3},{4}	[2134],[2314,[2341]
3	{2,3},{3,4},{2,4}	[3214],[3421],[3241]
4	{2,3,4}	[4321]

A little explanation as to how to get to this table: take for example $j = 3$ and $S = \{2, 3\}$. Firstly, $j = 3$ should be on position 1 in the new deck. Then $j - 1 = 2$ should be on position 2 and $j - 2 = 1$ should be on position 3. Now, we can fill in the remaining numbered card 4 on the first empty position, which is position 4. This gives indeed [3214]. If we look at all newly formed decks, we see that these are exactly the $2^{n-1} = 2^3 = 8$ distinct decks found in example 4.

After this example, it is time to prove an important lemma to prove theorem 1.

Lemma 2. *The first j cards of a deck after a Gilbreath shuffle consist of j consecutive numbered cards. They need not be consecutive in the deck after the shuffle, but for every j you should be able to rearrange the first j numbered cards $\pi(1), \ldots, \pi(j)$ such that they are consecutive. That is, there exist $a_j = \min_{1 \leq i \leq j} \pi(i)$, such that:*

$$\bigcup_{i=1}^{j} \{\pi(i)\} = \{a_j, a_j + 1, \ldots, a_j + (j-1)\}$$

Alternatively, if after a permutation of the deck for every $j \in \{1, \ldots, n\}$ there exists such a_j, then the permutation is a Gilbreath permutation.

Proof. Suppose we have a deck of n cards and label it in its original order, to keep things simple. Suppose that the first card of the new deck after a Gilbreath permutation is: $\pi(1) = j$. Because of the explanation of the equivalent way of defining the Gilbreath permutations on a deck, we know that after j, there should a consecutive increasing sequence. If this sequence terminates, we must

have $j-1$. After that, this process continues. However, we can indeed always rearrange, because of the equivalent explanation of defining all the Gilbreath permutations.
To prove the converse, suppose that after a permutation π, for every $j \in \{1, n\}$, there exists $a_j = \min_{1 \leq i \leq j} \pi(i)$, such that:

$$\bigcup_{i=1}^{j} \{\pi(i)\} = \{a_j, a_j + 1, \ldots, a_j + (j-1)\}$$

Denote the place where card i is after the permutation as π_i. Take the first card of the deck after applying the permutation: $\pi(1)$. Assume that $\pi(1) = j$. Create a set $S \subset \{2, \ldots, n\}$ as follows: $S = \{\pi_{j-1}, \ldots, \pi_1\}$. If S is ordered $(\pi_{j-1} < \pi_{j-2} < \cdots < \pi_1)$, then we are done. This is, because the rest of the cards $(j+1, \ldots, n)$ must be filled in increasing order, otherwise the assumption would be violated. Suppose it is no Gilbreath permutation. Then, there exists a smallest $m > 1$, such that $\pi(m) \neq \pi(m-1) + 1$ and $\pi(j) \neq a_{m-1} - 1$. This necessarily means that (m was the smallest such that it did not hold!):

$$\bigcup_{i=1}^{m-1} \{\pi(i)\} \cup \{\pi(m)\} = \{a_{m-1}, a_{m-1} + 1, \ldots, a_{m-1} + (m-1)\} \cup \{\pi(m)\}$$

However, this certainly does not fulfill our assumption, hence completes a contradiction. Hence, all that remains is to show that: $\pi_{j-1} < \pi_{j-2} < \cdots < \pi_1$.

For this: suppose not. So, suppose there exists a biggest $i, l \in \{1, \ldots, j-2\}$, such that: $\pi_i < \pi_l$ (of course: $l > i$).
Firstly, if there is no smallest $k > i$, such that $\pi_m < \pi_i$ for all $m \geq k$, then the first $\pi_i - 1$ cards are $[j(j+1) \ldots (\pi_i - 1 - j + 1)] = [j(j+1) \ldots (\pi_i - j)]$, since otherwise our assumption of consecutive $\pi(k)$'s would not hold. Obviously, if we add card $i \leq j-2$ to these first cards, we get a contradiction. Since $i \leq j-2 < j-1$ and we would need to extend these first cards with a card with a value greater or equal than $j-1$.
On the other hand, suppose there exists a smallest $k > i$, such that $\pi_m < \pi_i$, for all $m \geq k$. Since these are the smallest respectively biggest such k and i, we know that the first $\pi_i - 1$ cards form a set:

$$\bigcup_{w=1}^{\pi_i - 1} \{\pi(w)\} = \{k, k+1 \ldots, k + \pi_i - 2\}$$

We can see this because k is the smallest fulfilling k, so $\pi_{j-1} < \pi_{j-2} < \cdots < \pi_k$. However, since the first π_i cards must also have consecutive values and $i < k$, we find that: $k = i + 1$. But, this results in: $\pi_m < \pi_i$, for all $m \geq i + 1$. So, there is no $l > i$, such that: $\pi_i < \pi_l$. All together, we can conclude that $\pi_{j-1} < \pi_{j-2} < \cdots < \pi_1$ and our proof is complete. □

We can use this lemma to prove theorem 1 (proof based on [7], thoroughly extended).

Proof. First of all: (3) $\Rightarrow$ (2), since we can take $k = 1$. On the other hand, we can show that (2) $\Rightarrow$ (3) by induction. Suppose (2) is true. From now on, take j fixed. Take any $k \geq 2$ ($k = 1$ is just property (2)) and suppose $jk \leq n$. Assume that for $1 \leq i \leq k-1$, the cards $\pi((i-1)j+1), \ldots, \pi(ij)$ are distinct modulo j. We know that the first jk are distinct modulo jk, from property (2). This means that of every equivalence class modulo jk, there is exactly one card. However, there should also be k cards of every equivalence class modulo j. Since we assumed by induction hypothesis that for $1 \leq i \leq k-1$, the cards $\pi((i-1)j+1), \ldots, \pi(ij)$ are distinct modulo j, we have exactly one card in every equivalence class modulo j left. Hence we conclude that the cards: $\{\pi((k-1)j+1), \pi((k-1)j+2), \ldots, \pi(kj)\}$ are distinct modulo j. So by induction: (2) $\Rightarrow$ (3) and the properties (2) and (3) are equivalent.

Now suppose that (1) holds and fix j. Using lemma (2), we have an $1 \leq a_j \leq n$ such that the first j cards are: $\{a_j, \ldots, a_j + (j-1)\}$. These cards are obviously distinct modulo j. Hence (1) $\Rightarrow$ (2). On the other hand, suppose (2) holds. We need to prove that for every $j \in \{1, \ldots, n\}$, the cards on position 1 to j in the deck are consecutive, up to order. Proceed by induction. If $j = 1$, then obviously the results hold. Suppose it holds for $j = k - 1 \geq 1$. So for some $a = \min_{1 \leq i \leq k-1} \pi(i)$:

$$\bigcup_{i=1}^{k-1} \{\pi(i)\} = \{a, a+1, \ldots, a+(k-2)\}$$

We know that the first k cards are distinct modulo k. So $\pi(k) = a-1+m(j+1) = a-1+mk$, with $m \in \mathbb{Z}$. Also: m should of course be such that $1 \leq \pi(k) \leq n$. Now, we exclude the case $m \notin \{0, 1\}$. If $m \geq 2$, then $mk - 1 > k$, since $k \geq 2$ and $m \geq 2$. It follows that: $k < \pi(k) - a \leq n - a < n$, because $a \geq 1$. So the first $\pi(k) - a$ cards can not be different modulo $\pi(k) - a$, since:

$$\pi(k) - a \equiv 0 \pmod{\pi(k) - a}$$

Notice that this can only be done since $k < \pi(k) - a < n$. This gives a contradiction, since we assumed property (2).

Suppose now that $m \leq -1$. We know that $a + (k-2) - \pi(k) < n$, since $a + (k-2) \leq n$ and $\pi(k) \geq 1$. On the other hand: $a + (k-2) - \pi(k) = (k-1) - mk = (1-m)k - 1 > k$, because $(1-m) \geq 2$ and $k \geq 2$. So the first $a+(k-2)-\pi(k)$ can not be different modulo $a+(k-2)-\pi(k)$, in the same sense as the case $m \geq 2$. This can now only be done because $k < a+(k-2)-\pi(k) < n$.

Concluding: $m = 0$ or $m = 1$. This means that $\pi(k) = a-1$ or $\pi(k) = a+(k-1)$. This means that the induction step is complete and that (2) $\Rightarrow$ (1) using lemma (2).

□

Looking at the theorem, it is obvious that the shuffle does not completely randomize the original deck, since a lot of permutations are not possible. In fact, we know that only 2^{n-1} permutations of a deck of n cards are possible. If we take $n = 52$, then only a fraction of $\frac{2^{51}}{52!} \approx 2.8 \times 10^{-53}$ of the permutations of the deck is possible. Also, the variation distance between the uniform shuffle U and the Gilbreath shuffle G on a deck of n cards is easy to approximate:

$$\|G - U\| = \frac{1}{2} \sum_{\pi \in S_n} \big|G(\pi) - U(\pi)\big| \geq \frac{1}{2} \frac{n! - 2^{n-1}}{n!}$$

Here we used that on at least $n! - 2^{n-1}$ permutations, the variation distance between U and G is $\frac{1}{n!}$. This is because G gives a probability of 0 to this amount of permutations. Even for $n = 10$, this gives: $\|G - U\| \geq 0.4999$. This is quite a rough estimation of the variation distance, but even then we see that the variation distance is not going to be better than 0.5 for $n \geq 10$.

3.2 The card trick

A lot of card tricks have been performed, using this Gilbreath principle. Say you are a magician and you have a standard deck of cards. That is: the numbers 1 to 10 and the cards J, Q, K and A (a total of 52 cards). These cards appear exactly once in the following 4 kinds: Spades (S), Hearts (H), Clubs (C) and Diamonds (D). Sort your deck using a specific order of the type of cards, for example: SHCDSHCD.... Because of the distinction between cards within a specific type, the deck might appear to be random for the non-trained eye. So you can briefly show the deck as "random" to the audience. Next, you perform the Gilbreath shuffle. Notice: every Spades card is numbered $x \equiv 1 \pmod 4$, every Hearts card is numbered $x \equiv 2 \pmod 4$, every Clubs card is numbered $x \equiv 3 \pmod 4$ and every Diamonds card is numbered $x \equiv 0 \pmod 4$. Now ask the audience for a number j, such that you can perform the Gilbreath shuffle as explained. After this, every new deck satisfies property (3) from theorem 1. To see how this helps in a card trick: take $j = 4$. Then the cards of the first quartet on top of the deck are distinct modulo 4, the cards of the second quartet on top of the deck (cards on position $5, 6, 7$ and 8) are distinct modulo 4, et cetera. Hence you can start showing quartet after quartet of cards. Obviously every quartet contains exactly one card of each type. This works, since there are exactly 52 cards in a regular deck and hence exactly 13 quartets. The audience

will, if they have a feeling for probability, be quite amazed. Why would they be so amazed? Well, the probability of getting 4 different type of cards in the first quartet in a deck of $4m$ cards (m cards of every type) is:

$$\frac{4m}{4m}\frac{3m}{4m-1}\frac{2m}{4m-2}\frac{m}{4m-3} = \frac{24m^4}{(4m)(4m-1)(4m-2)(4m-3)}$$

We can see this, because the first card does not matter. The second card we take from the remaining $4m-1$ cards. Here we can not get the type of the first card of which there are $m-1$ left. Hence $(4m-1)-(m-1) = 3m$. The third card we take from the remaining $4m-2$ cards. Here we can not get the type of the first card and the second card of which there are $2(m-1)$ left. Hence $(4m-2)-2(m-1) = 2m$. Similarly for the last card. Notice that $m = 1$ gives a probability of 1. So the probability of getting only quartets in a deck of 52 cards is easy to calculate, since we can just take a product:

$$\prod_{m=1}^{13} \frac{24m^4}{(4m)(4m-1)(4m-2)(4m-3)} = \Big(\prod_{m=1}^{13} 24m^4\Big)\frac{1}{52!} < 0.0000000001$$

So any person in the audience with a little bit of feeling for probability would see the remarkability in this trick. The trick also works for the red and black cards in a deck of cards, which might be even more entertaining. What can we learn from this? Well, never trust a shuffle which on the eye seems "random enough". This is a proper motivation to examine the variation distance between shuffles and what we want: the uniform shuffle U, since it is not that easy on the eye to spot "bad" shuffles.

4 Riffle shuffle

A natural way to shuffle a deck of cards, which is mathematically also very nice, is the Riffle shuffle. This goes as follows: divide the deck into two packs of size k and $n-k$. So $k \in \{0, 1, \ldots, n\}$, packs may be empty. Choose k according to the binomial distribution, with probability: $p = \frac{\binom{n}{k}}{2^n}$. Then split the deck in two packs: $\{1, 2, \ldots, k\}$ and $\{k+1, k+2, \ldots, n\}$. Now form a new deck, "riffling" the two packs together. The relative order of the two packs must be maintained, but cards in different packs need not be in their relative order from the original deck. There are $\binom{n}{k}$ possible ways to do this and each must have the same probability, so choose uniformly between those interleavings. The probability to pick a certain cut, followed by a certain interleaving is (because of uniformity):

$$\frac{p}{\binom{n}{k}} = \frac{1}{2^n}$$

Let's look at all the possible Riffle shuffles of a deck of size 4 through an example (inspired by [13]).

Example 6. Suppose $n = 4$. We have:

k	packs	cut probability	possible interleavings
0	| 1234	$\frac{1}{16}$	[1234]
1	1| 234	$\frac{1}{4}$	[1234], [2134], [2314], [2341]
2	12| 34	$\frac{3}{8}$	[1234], [3412], [1324], [1342], [3142], [3124]
3	123| 4	$\frac{1}{4}$	[1234], [1243], [1423], [4123]
4	1234|	$\frac{1}{16}$	[1234]

This gives the following shuffling distribution $Q : S_4 \to [0, 1]$:

$$Q(\pi) \to \begin{cases} \frac{5}{16} & \text{if } \pi = e \\ \frac{1}{16} & \text{if } \pi = (12) \\ \frac{1}{16} & \text{if } \pi = (132) \\ \frac{1}{16} & \text{if } \pi = (1432) \\ \frac{1}{16} & \text{if } \pi = (13)(24) \\ \frac{1}{16} & \text{if } \pi = (23) \\ \frac{1}{16} & \text{if } \pi = (243) \\ \frac{1}{16} & \text{if } \pi = (1243) \\ \frac{1}{16} & \text{if } \pi = (123) \\ \frac{1}{16} & \text{if } \pi = (34) \\ \frac{1}{16} & \text{if } \pi = (234) \\ \frac{1}{16} & \text{if } \pi = (1234) \\ 0 & \text{else} \end{cases}$$

The variation distance between randomness and this shuffle is easily determined:

$$\|Q - U\| = \frac{1}{2}\Big(\Big|\frac{5}{16} - \frac{1}{24}\Big| + 11\Big|\frac{1}{16} - \frac{1}{24}\Big| + 12\Big|\frac{1}{24}\Big|\Big) = \frac{3}{8}$$

This is better than the difference between doing nothing and uniform shuffling U, which is $1 - \frac{1}{4!} = \frac{23}{24}$. But still, multiple shuffles are needed, especially because for example, the order [4231] is not even possible yet.

To determine $Q^{(k)}$ for large k is, even with such small n, not very nice. Because of this, we want to extend our definition a bit to ensure an easier calculation.

4.1 The a-shuffle

Instead of the usual Riffle shuffle R, it is also possible to divide the deck of cards in a packs (the following explanation of the a-shuffle is inspired by [13]). These packs have size $n_1, \ldots, n_a$, where $n_i \geq 0$. So an 3-shuffle can have one or two packs, because $n_i = 0$ is allowed. The probability of picking a specific distribution of packs is:

$$p = \frac{\binom{n}{n_1,\ldots,n_a}}{a^n}$$

Note: the nominator is a multinomial coefficient, hence the packs are chosen by the multinomial distribution. After splitting the deck in a packs of size larger or equal to zero, we put the deck back together. While doing this, the relative order of cards inside a pack must be maintained as with the Riffle shuffle. A possible deck is chosen using the uniform distribution. The probability of a specific interleaving being chosen is therefore:

$$\frac{1}{\binom{n}{n_1,\ldots,n_a}}$$

The reason for this is because the relative order must be maintained. So we can colour all of the cards in a pack the same colour. We do not need to make a distinction between these cards, because their order has already been established in the deck after the cut. Then we determine all possible ways to divide the n cards back in the deck, this is just $\binom{n}{n_1,\ldots,n_a}$. Concluding, just as with the Riffle shuffle R, the probability of a specific cut and a following interleaving is:

$$\frac{p}{\binom{n}{n_1,\ldots,n_a}} = \frac{1}{a^n}$$

Notice that if we choose $a = 2$, then we end up with the Riffle shuffle. The notation for the a-shuffle is: R_a. This means that: $R_2 = R$. Let me illustrate the a-shuffle by a small example.

Example 7. Suppose $n = 4$ and $a = 3$. We cut the deck with $n_1 = n_3 = 1$ and $n_2 = 2$. This occurs with probability: $\binom{4}{1,2,1} \cdot \frac{1}{3^4} = \frac{4}{27}$. We have the following packs: $\{1\}$, $\{2, 3\}$ and $\{4\}$. There should be $\binom{4}{1,2,1} = 12$ possible ways to put a deck together. Indeed we have the possible decks:

$[1234], [1243], [1423], [2134], [2143], [2314], [2341], [2413], [2431], [4123], [4213], [4231]$

This is one possible way to obtain certain shuffles.

Notice that in the example there is no way in getting 3 as first card in the deck, due to the cut. Also, the probability of getting the original deck back after an a-shuffle, is bigger than any other permutation as we will find using rising sequences.

What is exactly the use for an a-shuffle? Practically, most people tend to use only the 2-shuffle, which is the practical Riffle shuffle. Well as it turns out, sequential Riffle shuffles are linked to an a-shuffle for certain a.

Theorem 2. *Suppose we have $a, b \in \mathbb{N}$. Performing an a-shuffle followed by a b-shuffle is equivalent to performing an ab-shuffle.*

This turns out to be a very nice and useful property, which is not true for most shuffles. To prove this theorem, firstly some background is needed. We want to define the inverse of an a-shuffle.

Definition 4. Define an a-unshuffle of a deck of n cards as follows. Form a numbered decks of cards (not necessary non-empty). Do this by taking firstly the top card of the deck and placing it with probability $\frac{1}{a}$ on one of the (now all empty) bottom of the stacks. Repeat this until the deck of cards is empty. Now place pile i on top of pile $i+1$ for all i with $1 \leq i \leq a$.

As there is only one way to cut the deck in a specific piles and there is only one way to interleave the cards to get the original deck before the unshuffle, every a-unshuffle is the unique inverse of a certain a-shuffle. We now have the necessary tools to complete the proof of the theorem (the proof is largely devoted to [8]).

Proof. Suppose we apply an ab-unshuffle to a deck. So we form ab stacks in the explained way. Label these stacks (in order) not $1, 2, \ldots, ab$, but with elements (x, y) of $\mathbb{Z}^2$. Here $x \in \{0, \ldots, a-1\}$ and $y \in \{0, \ldots, b-1\}$. Label each card accordingly, with the label of the pile it is in. Now we want to find a b-unshuffle and an a-unshuffle, such that applying the unshuffles after one another is the same as applying the ab-unshuffle.
To find the needed b-unshuffle, sort the deck into b stacks. In pile i, there should be exactly all cards where $y = i$. Do exactly the same after this for the needed a-unshuffle: in pile j should be exactly all cards with $x = j$. Now, after the b-unshuffle followed by the a-unshuffle, the cards lie in the following order: $(0,0), (0,1), \ldots, (0,b-1), (1,0), \ldots, (1,b-1), \ldots, (a-1,0), \ldots, (a-1,b-1)$. Conclude now that this is exactly the same as the ab-unshuffle. This completes the result, because $b^{-1}a^{-1}$ has a one-to-one correspondence with $(ab)^{-1}$. Hence an a-shuffle followed by a b-shuffle is the same as performing a ab-shuffle. □

This result might not seem useful now to the reader, but in that case I suggest reading on.

4.1.1 Rising sequences and their use

A nice way to look at a specific deck is to find its rising sequences. This turns out to be a useful tool in determining certain probabilities of a-shuffles.

Definition 5. Suppose we have a deck in specific ordering. Say $[x_1 \dots x_n]$. A *rising sequence* is a subsequence $A = (x_{i_k})_k$ of $(x_i)_i$ such that $x_{i_{k+1}} = x_{i_k} + 1$ for all k. Additionally there should not exist a sequence B such that $A \subset B$ and $A \neq B$ and B is also rising. Hence, it should be maximal.

Example 8. Consider the following deck ordering: [589214673]. Starting at 5, we find 6 and after that 7. Hence $(5, 6, 7)$ is a rising sequence. We could have obtained the sequence also by starting at 6 and counting up (spotting 7). After that returning to 6 and counting down. Moreover we find $(8, 9)$, $(2, 3)$, (1) and (4).

It appears that rising sequences partition the elements (cards) in an elegant way. Indeed this is the subject of the following theorem.

Theorem 3. *Given an ordering of a deck:* $[x_1 \dots x_n]$. *Every numbered card* x_i *is in exactly one rising sequence.*

Proof. Suppose we have such ordering of a deck: $[x_1 \dots x_n]$. First of all, we can construct a rising sequence with a given x_i. This is straightforward by first going to the right in the deck and counting up (looking for x_j such that $x_j = x_i + 1$). Continue, repeating this process until we come at x_n. After that go back to x_i and start counting down (looking for x_k such that $x_k = x_i - 1$). Also continue and repeat until we stop at x_1. This certainly gives a rising sequence (obviously maximal).
Suppose $A = (a_1, \dots a_{k-1}, x_i, a_{k+1}, \dots, a_{n_1})$ and $B = (b_1, \dots b_{m-1}, x_i, b_{m+1}, \dots, b_{n_2})$ are rising sequences. Now notice that $a_{k+1} = b_{m+1}$, since they are rising sequences. Continuing this way gives: $B = (b_1, \dots b_{m-1}, x_i, a_{k+1}, \dots, a_{n_1})$. The reason that $n_1 = n_2$ is that if $n_1 > n_2$ (or vice versa), then x_i is in the rising sequence B, but this can never be maximal. So it can not be a rising sequence. We can also do this the other way around, hence $B = (a_1, \dots a_{k-1}, x_i, a_{k+1}, \dots, a_{n_1}) = A$. □

The theorem establishes that a deck ordering can be partitioned into rising sequences. We also find that the number of rising sequences is limited to n, where n is the order of the deck. We can achieve this by this ordering: $[(n)(n-1)\dots(2)(1)]$. But what do rising sequences contribute to the a-shuffle R_a? Well, as it turns out, it makes calculations a lot easier. See the following theorem and its proof, which is a hybrid form of [13] and [5].

Theorem 4. *Suppose we are given a deck* $[(1)(2)\dots(n-1)(n)]$ *and we want to perform the a-shuffle* R_a*. The probability of a certain permutation* π *on this deck is:*

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

Here n *is the size of the deck and* r *the number of rising sequences in the deck after applying the permutation* π*.*

Proof. Recall that all possible ways of cutting the deck and then interleaving it have the same probability. So we need to find the number of ways to cut the deck into a packs, such that π is a possible permutation. In an a-shuffle all cards stay in their relative order of the packs, so each rising sequence in the new deck is a union of the a packs. It is therefore, that we want to find the number of ways of distributing r rising sequences into a packs.
Because π has r rising sequences, we know where $r-1$ cuts must have been: between a card that ends a rising sequence of π and another card that begin a rising sequence of π. These pairs are of course consecutive in the deck $[(1)\dots(n)]$. But, we need to make $a-1$ cuts. Hence, we remain with $(a-1)-(r-1)=a-r$ cuts in a deck of n cards. So we need to fill $n+(a-r)=n+a-r$ positions with two types: n cards and $(a-r)$ cuts. This gives the binomial coefficient. By prior explanation, the denominator a^n is trivial. □

Corollary 1. *Suppose we have a deck* $[(1)(2)\dots(n-1)(n)]$ *and we Riffle shuffle* k*-times. The probability of ending up with a certain permutation* π *on this given deck is:*

$$\frac{\binom{n+2^k-r}{n}}{2^{kn}}$$

Here n *is the size of the deck and* r *the number of rising sequences of the deck after applying the permutation* π*.*

Proof. Combine theorem 2 and theorem 4. Performing a Riffle shuffle (or 2-shuffle) k-times is equivalent to performing a 2^k-shuffle, according to theorem 2. Applying theorem 4 yields the result. □

Interesting is that we can immediately see that the identity permutation (with only one rising sequence) is more likely to appear than any other permutation. Also, suppose we have $n=r=52$ and shuffle 5 times. Then:

$$\frac{\binom{n+2^k-r}{n}}{2^{kn}} = \frac{\binom{32}{52}}{2^{52}} = 0$$

So we need to Riffle shuffle a normal deck of cards more than 5 times to even possibly achieve the deck $[(52)(51)\dots(2)(1)]$.

Knowing that the number of rising sequences determines the probability of a

certain shuffle, we might want to know how many different permutations yield a certain probability. For this, define $D(n,r)$ as the number of distinct decks after a permutation with r rising sequences. If we look at the just proved theorem, we now know that:

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

Which leads us to:

$$\sum_{r=1}^{n} D(n,r)\binom{n+a-r}{n} = a^n$$

Now define the Eulerian numbers $A(n,m)$ recursively:

$$A(n,m) = (n-m)A(n-1,m-1) + (m+1)A(n-1,m)$$

Also: $A(1,0) = A(2,0) = A(3,0) = A(2,1) = A(3,1) = A(3,2) = 1$ and $A(n,m)$ is only defined if: $n > m \geq 0$. Using the Eulerian numbers, we have the identity of Worpitzky (for a proof see [6]):

$$x^n = \sum_{m=0}^{n-1} A(n,m)\binom{x+m}{n} \text{ or } x^n = \sum_{m=1}^{n} A(n,m-1)\binom{x+m-1}{n}$$

This formula holds for the Eulerian numbers. Filling in $x = a$ and $m = n - r$ gives:

$$a^n = \sum_{r=0}^{n-1} A(n,n-r-1)\binom{n+a-r-1}{n}$$

Using the identity $A(n,m) = A(n,n-m-1)$ and making a minor substitution gives:

$$a^n = \sum_{r=0}^{n-1} A(n,r)\binom{n+a-r-1}{n} \text{ or } \sum_{r=1}^{n} A(n,r-1)\binom{n+a-r}{n}$$

We obtain: $D(n,r) = A(n,r-1)$. Fortunately, there is also an explicit formula for $D(n,r)$, using the fact that they are the Eulerian numbers $A(n,r-1)$ (see [1]):

$$D(n,r) = A(n,r-1) = \sum_{k=0}^{r}(-1)^k\binom{n+1}{k}(r-k)^n \tag{1}$$

Indeed, $D(4,1) = A(4,0) = 1$ and $D(4,2) = A(4,1) = 11$ as we saw in example 6. At last, let me remark that sometimes $A(n,r)$ is defined as the number of permutations of S_n with r rising sequences. I do not adopt this notion, but rather stick to the definition on Wikipedia. Using the closed form (or the knowledge of values of the Eulerian numbers), we can easily calculate the number of permutations in S_n which have r rising sequences.

4.1.2 Applying corollary 1

As pointed out, Riffle Shuffling a deck of 52 cards needs to be done more than 5 times to have a nontrivial probability for every permutation. However, this might not even be a problem, because the difference between $\frac{1}{52!}$ (uniform shuffle U) and 0 is not that big. Therefore it is more interesting to look at the variation distance between $R^{(k)}$ and U. Recall the variation distance between two densities (Q_1 and Q_2) on the same deck of cards:

$$\|Q_1 - Q_2\| = \frac{1}{2} \sum_{\pi \in S_n} \left| Q_1(\pi) - Q_2(\pi) \right|$$

Applying this to $R^{(k)}$ and U gives:

$$\left\| R^{(k)} - U \right\| = \frac{1}{2} \sum_{\pi \in S_n} \left| R^{(k)}(\pi) - U(\pi) \right|$$

Or, using corollary 1 and (1), we get:

$$\left\| R^{(k)} - U \right\| = \frac{1}{2} \sum_{r=1}^{n} D(n,r) \left| \frac{\binom{n+2^k-r}{n}}{2^{kn}} - \frac{1}{n!} \right|$$

$$= \frac{1}{2} \sum_{r=1}^{n} \sum_{m=0}^{r} (-1)^m \binom{n+1}{m} (r-m)^n \left| \frac{\binom{n+2^k-r}{n}}{2^{kn}} - \frac{1}{n!} \right|$$

Using $n = 52$, we can plot the value of the variation distance for $k \in \{0, 1, \ldots, 14\}$ (see Figure 1). We see that the variation distance takes a steep drop after 5 shuffles and is nearly 0 after 12 shuffles. The number 7 is normally (I do not completely agree) chosen as the number of shuffles it takes to randomize a deck (see for example [10]).

4.2 Riffle shuffle analysis using stopping time

When looking at the Riffle shuffle, we could immediately calculate the probabilities through nice properties. However, there is an alternative approach which is more strict on the variation distance between $\left\| R^{(k)} - U \right\|$. We will work this out and find out that it might indeed (as noticed before) be better to Riffle shuffle more than 7 times. To work this out, some instruments will be explained. Firstly, (recall) the definition of a discrete-time Markov Chain.

Definition 6. Suppose we have a sequence of random variables $X_1, X_2 \ldots$ and X_i can take values in a countable state space S. This sequence is called a *discrete-time Markov Chain* if it satisfies the Markov property, that is: the

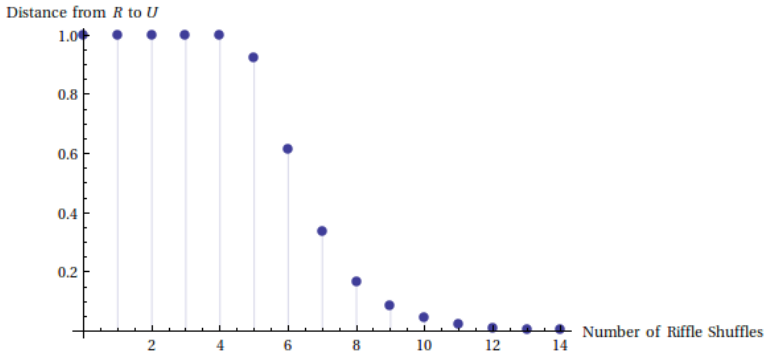

Figure 1: The variation distance between R and U (y-axis) for certain successive Riffle shuffles (x-axis). Figure created using Mathematica.

probability of what happens in state $m+1$ is only dependent on the state X_m. More formally:

$$\mathbb{P}(X_{m+1} \mid X_1, \dots X_m) = \mathbb{P}(X_{m+1} \mid X_m)$$

It is a well-known fact that if our Markov Chain has an irreducible aperiodic transition matrix on a finite state space, that there exists a unique stationary distribution. In the Riffle shuffle case, we will take X_m to be the ordering of the deck at time m. This is equivalent to some permutation $\pi \in S_n$, which determines the new positions of the numbered cards. Our initial value X_0 is the ordering $[(1) \dots (n)]$. This is obviously a Markov Chain, because the probability of a certain permutation on this deck is only dependent on the number of rising sequence at time m. From earlier computation, we know that the stationary distribution is the uniform distribution: the Riffle shuffle converges to the uniform distribution. Next, we define a random variable called the stopping time. Also, a strong uniform time T will be defined (definition from [4]).

Definition 7. Suppose we have a sequence of random variables $X_1, X_2, \dots$. A *stopping time* T is a random variable with state space $\mathbb{N}$. Moreover the probability that $T = t$ is only dependent on the values of $X_1, \dots, X_t$.

Definition 8. A *strong uniform time* T is a randomized stopping time for a Markov Chain $(X_m : m \geq 0)$ with stationary distribution π, state space S and initial state $i_0 \in S$, if:

$$\mathbb{P}(X_k = i \mid T = k) = \pi(i) \text{ for all } k \in \{0, 1, \dots\} \text{ and } i \in S.$$

Why do we need the notion of stopping time to find a good amount of Riffle shuffles? Well, we can use $\mathbb{P}(T > k)$ to estimate $\left\| R^{(k)} - U \right\|$! The proof of the following lemma is largely based on [3].

Lemma 3. *Suppose we have a probability distribution Q on a finite group G $(= S_n)$. Let T be a strong uniform time for Q. Then, for all $k \geq 0$:*

$$\left\| Q^{(k)} - U \right\| \leq \mathbb{P}(T > k)$$

Proof. Suppose $A \subset G$, then

$$Q^{(k)}(A) = \mathbb{P}(X_k \in A) = \sum_{j=1}^{k} \mathbb{P}(X_k \in A, T = j) + \mathbb{P}(X_k \in A, T > k)$$

$$= \sum_{j=1}^{k} \mathbb{P}(X_k \in A \mid T = j)\mathbb{P}(T = j) + \mathbb{P}(X_k \in A \mid T > k)\mathbb{P}(T > k)$$

But we know that T is a strong uniform time, hence $\mathbb{P}(X_k \in A \mid T = j) = \pi(A) = U(A)$. This gives, with a little re-ordering:

$$Q^{(k)}(A) = U(A) + \Big(\mathbb{P}(X_k \in A \mid T > k) - U(A)\Big)\mathbb{P}(T > k)$$

Hence:

$$\Big|Q^{(k)}(A) - U(A)\Big| = \left|\Big(\mathbb{P}(X_k \in A \mid T > k) - U(A)\Big)\mathbb{P}(T > k)\right| \leq \mathbb{P}(T > k)$$

Conclude the hypothesis, because A was arbitrarily. □

4.2.1 Stopping time for the Riffle shuffle

So, if one succeeds to find a strong uniform time for the Riffle shuffle and to obtain $\mathbb{P}(T > k)$, then it is possible to say something about $\left\| R^{(k)} - U \right\|$. Well, what stopping time T should we use? Mann describes a good stopping time T in [13]. Firstly, it is sufficient to find a stopping time for the unshuffle, because the a-unshuffle $\hat{R}_a$ is exactly the inverse of the Riffle shuffle (that is $\hat{R}_a(\pi^{-1}) = R_a(\pi)$). Secondly, we need an equivalent way of doing an a-unshuffle. This can be done by labelling all cards with a $k \in \{0, \ldots, (a-1)\}$ chosen uniformly. Place all cards with a 0 on top of the new deck, keeping their relative order. Repeat this for all numbers in ascending order, always keeping the relative order of cards with the same number k.

On the other hand, an equivalent way of describing the R_a shuffle is by taking an n digit base a number and putting bars after the number of zeros, then ones, et cetera. Here, the i-th digit of the chosen number is in the range: $\{0, \ldots, a-1\}$, chosen uniformly. After this, place the numbers in their relative order on the spots they belong, that is the first card will be the card on the position of the first 0 in the n base a number. Continue doing this for all zeros in the n base

a number. Repeat this process with all ones, et cetera. The equivalence is for example shown in [13]. I will now demonstrate with an example how this is done.

Example 9. Take $a = 3$ and $n = 8$. Take the n digit base 3 number: 10201222. That is, we have 2 zeros, 2 ones and 4 twos. We want to perform the 3-shuffle using this code. Hence, the deck is split as: 12 | 34 | 5678. From the base number, we get that the card 1 should be on place 2 and that the card 2 should be on place 4 (since the zeros are on that spot). Continuing, we end up with: [31524678], which is the permutation $\pi = (13542)$. Use the same base number to perform the 3-unshuffle. So the cards on place 2 and 4 should be in front, in that order. These are indeed the cards 1 and 2. Continuing gives the original deck [12345678]. This is the permutation (12453). Indeed $(12453) = \pi^{-1}$, as we expected.

We can do k consecutive 2-unshuffles by getting k binary numbers with n digits ($00\ldots0$ is also a n digit binary number), name them $x_1, \ldots x_k$. Then, create n binary numbers with k digits, name these $y_1, \ldots, y_n$. They are constructed as follows: the i-th digit of y_j is equal to the j-th digit of x_{k+1-i}. After this, convert $y_1, \ldots y_n$ to regular numbers $z_1, \ldots z_n$. We can then sort the z_i's in order of size (if $z_i = z_j$, then the z_k with the smallest index i or j is placest first). The order of the indices of the z_i's gives the new deck ordering. Why is this the same as performing k consecutive 2-unshuffles? Well, there is a similar theorem for unshuffles, as theorem 2, but just the other way around: a b-unshuffle followed by an a-unshuffle is equivalent to an ab-unshuffle. As Mann points out in [13], the orders are reversed: we write a b-unshuffle followed by an a-unshuffle. Mann gives the explanation: "for the same reason that one puts on socks and then shoes, but takes of shoes and then socks". I will not dive into details as with theorem 2, but assume the aforementioned as true. An example seems more illustrating to me.

Example 10. Suppose we want to perform 4 consecutive 2-unshuffles. So $k = 4$, $n = 5$ and we get the following binary numbers with $n = 5$ digits: $x_1 = 11110$, $x_2 = 11101$, $x_3 = 00001$ and $x_4 = 01010$. These n digit binary numbers are chosen by uniformly selecting each digit. We get:

card	2-unshuffle (vertically: $x_4x_3x_2x_1$)	normal numbers
1	$y_1 = 0011$	$z_1 = 3$
2	$y_2 = 1011$	$z_2 = 11$
3	$y_3 = 0011$	$z_3 = 3$
4	$y_4 = 1001$	$z_4 = 9$
5	$y_5 = 0110$	$z_5 = 6$

Firstly, y_1 is 0011. This is because the first digits of x_1, x_2, x_3 and x_4 are (respectively): 1, 1, 0 and 0. Notice how we reverse the order. Secondly, the binary numbers are easily converted to regular numbers. For example, 1011 gives: $1 \cdot 8 + 0 \cdot 4 + 1 \cdot 2 + 1 \cdot 1 = 11$. Moreover, we get: $(z_1, z_2, z_3, z_4, z_5) = (3, 11, 3, 9, 6)$. We order them ($z_1$ comes before z_3, because $1 < 3$) and get the order: $(z_1, z_3, z_5, z_4, z_2)$. So, we end up with the unshuffle of the original deck ([12345]) of: [13542].

It is finally time to define a stopping time for 2-unshuffling, which gives us a stopping time for 2-shuffling. The stopping time T is defined as the number of tries it takes to have n distinct base 2^T numbers. Looking at the previous example: $T > 4$, since the cards 1 and 3 have the same base number. This does at first glance not immediately seem like a randomized stopping time and strong uniform time, which we do need. But as it turns out, it is quite easy to prove that it in fact is.

Lemma 4. *If T is defined as the number of tries it takes to have n distinct base 2^T numbers for 2-unshuffling, it is a strong uniform time.*

Proof. First of all, it is indeed a random variable with state space $\mathbb{N}$. And the probability that $T = t$ is only dependent on the values of $X_1, \ldots, X_t$. Here X_i is the permutation of the deck at time i. This is a Markov Chain, since only the last permutation matters for the current ordering of the deck. The last thing we need to verify is if:

$$\mathbb{P}(X_k = i \mid T = k) = \pi(i) \text{ for all } k \in \{0, 1, \dots\} \text{ and } i \in S$$

This can be explained easily, since if $T = k$, then we have n distinct base 2^T numbers. But these numbers are chosen randomly, hence for two cards i and j it is equally alike that the number of card i is larger than the number of card

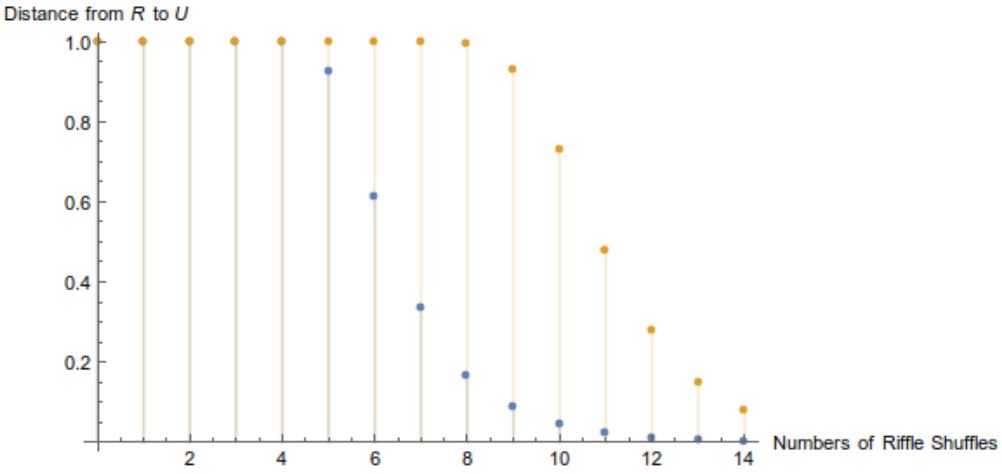

Figure 2: The blue dots represent the real variation distance. The yellow ones come from the upper bound. Figure obtained using Mathematica.

j as the other way around. So the deck is uniformly distributed. Hence given $k \in \mathbb{N}$ and $i \in S$:

$$\mathbb{P}(X_k = i \mid T = k) = \frac{1}{|S|} = \pi(i)$$

□

Now all that remains is finding $\mathbb{P}(T > k)$, then we can apply lemma 3. This probability is exactly the probability that n digit base 2^k numbers picked randomly are not all distinct. This is easily determined by its complement. That is: we need to find the probability that no number is equal. We have for the first number 2^k options. For the second $2^k - 1$, all the way to the n-th. That is: $(2^k)(2^k-1)\dots(2^k-n+1) = \frac{(2^k)!}{(2^k-n)!}$. All with equal probability $\frac{1}{2^{kn}}$. Combining gives:

$$\mathbb{P}(T > k) = 1 - \frac{(2^k)!}{(2^k - n)!}\frac{1}{2^{kn}}$$

We finally have an upper bound, but I have to disappoint you: this is not strict enough to come to the already calculated values of $\left\| R^{(k)} - U \right\|$. Take the usual case $n = 52$ and see figure 2 for the upper bound and the variation distance.

Luckily, we have another way of using $\mathbb{P}(T > k)$. We know that the shuffling is truly random if T "stops". Hence it is interesting to determine $\mathbb{E}(T)$.

Because $\mathbb{P}(T \geq 0) = 1$, we can rewrite the standard $\mathbb{E}(T)$ and behold:

$$\mathbb{E}(T) = \sum_{k=0}^{\infty} k\mathbb{P}(T = k) = \sum_{k=0}^{\infty} \mathbb{P}(T > k) = \sum_{k=0}^{\infty} \Big(1 - \frac{(2^k)!}{(2^k - n)!} \frac{1}{2^{kn}}\Big)$$

For $n = 52$, we get around 11.7. So it takes about 12 Riffle shuffles to randomize. This is almost the same as using the first value of k for which $\left\|R^{(k)} - U\right\|$ is very close to 0. This is, oppositely to the upper bound, a direct way to see that around 12 Riffle shuffles should indeed be sufficient.

5 Top-in shuffle

Another way to shuffle, which will turn out to be of less practical use, is the Top-in shuffle as in [13].

Definition 9. Define the following probability mass function $\sigma : S_n \to [0, 1]$ by:

$$\sigma(\pi) = \begin{cases} \frac{1}{n} & \text{if } \pi = \big((n)(k)(k+1)(k+2)\dots(n-1)\big) \text{ where } k \in \{1, 2, \dots, n\} \\ 0 & \text{else} \end{cases}$$

So if we apply this function, we get a permutation (with probability $\frac{1}{n}$ which sends the top card of the deck to place k and all cards on place $i \geq k$ will be send to $i + 1$.
Now the Top-in shuffle (TS). Mark the bottom card of the deck (the card 1). Then proceed iteratively:
Step 1: Find a permutation π using σ.
Step 2: If the card on place n is not equal to the marked card: apply π and return to step 1. Otherwise: apply π and quit the iteration.

What does the Top-in shuffle do? It inserts the top card in the deck at a place chosen uniformly. It continues to do so until the marked card comes on top. After that it places the marked card uniformly in the deck and the shuffle is complete. When the marked card came up top, all cards beneath where randomized and the last step randomly inserts the marked card. So this process indeed randomizes, that is: all decks are equally alike. However, as said, it is of less practical use: it is not very efficient and takes a while to do. See an example below.

Example 11. Take $n = 4$. The following k's have been generated by Mathematica (using "RandomInteger[{1, 4}]"): $\{4, 1, 3, 4, 1, 2, 1, 4, 1, 2\}$. Notice that getting at least n ones is always enough, if the last one is followed by another

k.

old deck	k	permutation in this step π	new deck
[1234]	4	e	[1234]
[1234]	1	(1432)	[4123]
[4123]	3	(23)	[4132]
[4132]	4	e	[4132]
[4132]	1	(1423)	[2413]
[2413]	2	(143)	[2341]
[2341]	1	(1432)	[1234]

So the deck is shuffled to itself.

Now define the random variable T as the number of times one needs to apply σ in total. Hence in the example above, $T = 7$. As we wanted to find out how many Riffle shuffles are needed to randomize, now we want to find out how soon the iterations quit. So we need to determine $\mathbb{E}(T)$. For this, write $T = T_n + \cdots + T_2 + 1$. Here T_j as the number of times we need to generate a number k such before the tagged card moves from position $j-1$ to position j. The probability of picking a number such that the tagged card is moved from position $j-1$ to j in a certain iteration is:

$$p := \frac{1}{n} \sum_{m=1}^{j-1} 1 = \frac{j-1}{n}$$

Using the geometric series we get $(p(1-p)^{i-1})$:

$$\mathbb{P}(T_j = i) = \begin{cases} \frac{j-1}{n}\left(\frac{n-j+1}{n}\right)^{i-1} & \text{if } i \in \mathbb{N} \\ 0 & \text{else} \end{cases}$$

Which makes determining $\mathbb{E}(T_j)$ easy, as this is just $\frac{1}{p} = \frac{n}{j-1}$. Using this, we obtain $\mathbb{E}(T)$:

$$\mathbb{E}(T) = 1 + \sum_{j=2}^{n} \mathbb{E}(T_j) = 1 + n \sum_{j=1}^{n-1} \frac{1}{j}$$

Because the Harmonic numbers can be approximated by $\log n$ for large n, we can estimate $\mathbb{E}(T)$ by $n \log n$ for large n. Specifically for $n = 52$, we get:

$$\mathbb{E}(T) = 1 + 52 \sum_{j=1}^{51} \frac{1}{j} \approx 236 \text{ or } \mathbb{E}(T) \approx 52 \log 52 \approx 205$$

This turned out to be a nice and quiet simple approach. But, as we recall from section 4.2, that $\mathbb{P}(T > k)$ an upper bound is for the variation distance. For this, the definition of T must indeed be so, that it is a strong uniform time.

Lemma 5. *If we define T as the number of permutations we need to generate for the Top-in Shuffling process, then T is a strong uniform time.*

Proof. First of all, it is indeed a random variable with state space $\mathbb{N}$. And the probability that $T = t$ is only dependent on the values of $X_1, \ldots, X_t$. Here X_i is the permutation in step i. It is therefore obvious that T is a stopping time for the Markov Chain $(X_i)_i$. The stationary distribution π is the uniform distribution, because at a certain moment all distributions of the deck are equally alike. The last thing we need to verify is if:

$$\mathbb{P}(X_k = i \mid T = k) = \pi(i) \text{ for all } k \in \{0, 1, \ldots\} \text{ and } i \in S$$

This can be explained easily, since if $T = k$, then every ordering is equally alike. This is exactly how the Top-in Shuffling process was defined. Hence given $k \in \mathbb{N}$ and $i \in S$:

$$\mathbb{P}(X_k = i \mid T = k) = \frac{1}{|S|} = \pi(i)$$

□

So from now on, a method of estimating $\mathbb{P}(T > k)$ will be presented. For this, define the coupon collecter distribution. Liang provides an extensive method to estimate $\mathbb{P}(T > k)$ in [12]. Suppose we have a coupon collector. We define the random variable C as the number of coupons the collector has when he collected all n coupons. Here the probability of getting a specific type of coupon is: $\frac{1}{n}$.

Lemma 6. *Let C be a coupon collector random variable. Given $\alpha > 0$, we have:*

$$\mathbb{P}(C > \lceil n \log n + \alpha n \rceil) \leq \frac{1}{e^{\alpha}}$$

Proof. Define A_i to be the event that the i-th coupon type is not present in the first $\lceil n \log n + \alpha n \rceil$ draws. Because the probability of not drawing coupon type i is (i.i.d.) $\frac{n-1}{n} = 1 - \frac{1}{n}$, we have:

$$\mathbb{P}(A_i) \leq (1 - \frac{1}{n})^{\lceil n \log n + \alpha n \rceil}$$

We can now estimate from above:

$$\mathbb{P}(C > \lceil n\log n + \alpha n\rceil) \leq \mathbb{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

$$\leq \sum_{i=1}^{n}(1-\frac{1}{n})^{\lceil n\log n+\alpha n\rceil} \leq n\exp\Big(\log\Big((1-\frac{1}{n})^{\lceil n\log n+\alpha n\rceil}\Big)\Big)$$

$$\leq n\exp\Big((n\log n+\alpha n)\log\big(1-\frac{1}{n}\big)\Big) \leq n\exp\Big(-\frac{n\log n+\alpha n}{n}\Big) = e^{-\alpha}$$

□

Another result will complete our estimation, combined with the previous lemma (for the proof, see [12]).

Lemma 7. *Let C and T be as before. We have, for all $k \in \mathbb{N}$:*

$$\mathbb{P}(T < k) \leq \mathbb{P}(C < k)$$

Proof. Define C_i to be the number of coupons the coupon collector has when he has collected the first i distinct coupons. Then:

$$C = (C_n - C_{n-1}) + \cdots + (C_2 - C_1) + C_1 = (C_n - C_{n-1}) + \cdots + (C_2 - C_1) + 1$$

What can we say about $(C_{i+1} - C_i)$? Well, it is for sure a geometric random variable, notation $g(p)$. But what is p? The probability of getting a new sort of coupon is $p = \frac{n-i}{n}$. Hence, we can write:

$$C = \sum_{k=1}^{n} g(\frac{k}{n})$$

On the other hand we already know that

$$T = T_n + \ldots T_2 + 1 = \sum_{k=2}^{n} g(\frac{k-1}{n}) + g(\frac{n}{n}) = \sum_{k=1}^{n} g(\frac{k}{n})$$

We can indeed conclude our hypothesis. □

Corollary 2. *The variation distance $\Big\|TS^{(k)} - U\Big\|$ is less than or equal to $\epsilon > 0$ for $k \geq \lceil n\log n + n\log\frac{1}{e}\rceil = \lceil n\log\frac{n}{e}\rceil$*

Proof. Combine lemma 3, lemma 6 and lemma 7. □

How nice this corollary might seem, it turns out to be not really good. For small values of ϵ, we need really large values of k. Even for $\epsilon = \frac{1}{3}$, we need $k \geq 263$. The approach with $\mathbb{E}(T)$ seems better. All in all, we can conclude that the Riffle shuffle is (way) more effective than the Top-in shuffle.

6 Overhand shuffle

We have already seen a frequently used shuffle, namely the Riffle shuffle. Another shuffle often used in practice is the Overhand shuffle. For this shuffle, we first define the order reversing permutation. We will denote it with σ. Applying σ to a deck is exactly reversing the order. So the original deck $[(1)\dots(n)]$ is, after applying σ, equal to $[(n)\dots(1)]$. Now the Overhand shuffle.

Definition 10. Suppose we have a deck of n cards in the original numbered order: $[(1)\dots(n)]$. The *Overhand shuffle*, denoted $O(p)$ goes as follows: define $X_1, \dots X_{n-1}$ random variables. The state space S of these random variables X_i is $\{Y, N\}$, for "Yes" and "No". We define the probability density as: $\mathbb{P}(X_i = Y) = p$. Hence $\mathbb{P}(X_i = N) = 1-p$. Of course: $p \in (0,1)$. How do we shuffle the cards? Well, if $X_i = Y$, then we separate the i-th card (counted from top, so $X_1 = Y$ means separating card 1 from card 2) from the $i+1$-th card. If $X_i = N$, we do not separate them. If some $X_i = Y$, we find $k < i$ such that $X_k = Y$ and k as large as possible. This means finding the first variable with value Y, looking back from X_i. We then place all the cards $k+1$ until i on a new pile, in their current order. If no $k \in \mathbb{N}$ can be found, do this for all cards 1 until i. Repeat this for all $n-1$ variables and apply the reversing order permutation σ.

Before I dive into the matter, let me first give an example.

Example 12. Suppose we have a deck of $n = 9$ cards. First, let $p = \frac{1}{2}$. Mathematica (using "RandomInteger[{1,2},8]". Here a 1 was a Y) generates 9 variables: $(X_1, \dots, X_9) = (Y, N, N, Y, Y, Y, Y, N)$. Since $X_1 = Y$, we should separate the top card 1 from the second card 2. So the new deck is: [.....1]. Since $X_2 = X_3 = N$ and $X_4 = Y$, we must (since X_1 is the last variable with value Y) put the cards $2, 3$ and 4 in their current order on the new pile. We get: [....2341]. Continuing, we get: [897652341]. Since we applied the Overhand shuffle once, we must reverse the order. So the new deck is:

$$[143256798]$$

This does not seem as a very good shuffle, but the case $p = \frac{1}{2}$ "feels" as our best option. Let me show anoter case: $p = \frac{4}{5}$. Mathematica generates 8 variables (using "RandomInteger[{1,5},8]". Here only a 5 gave a N.):
$(X_1, \dots, X_9) = (Y, Y, Y, N, N, Y, Y, Y)$. In the same way, this gives (before reversing): [987456321]. And after reversing, we get:

$$[123654789]$$

This indeed seems "less random".

A few questions can arise. Why do we need the reversing shuffle σ? Well, after an uneven amount of consecutive Overhand shuffles, the deck has reversed

order. However, for an analysis of the Overhand shuffle, it is more convenient to look at the original, non-reversed order. And for an analysis of how many shuffles are needed, it does not matter.
What is this Overhand shuffle exactly? Well, it is a mathematical model of taking all cards in your left hand and sliding sets of cards to your right hand. If $p = \frac{1}{2}$, this means you slide on average 2 cards into your right hand. Of course, if $(X_1, \ldots, X_{n-1}) = (Y, Y, \ldots, Y)$, then we separate every card. This exactly reverses the order of the cards, but after applying σ, we get the identity permutation on the deck. On the other hand, if $(X_1, \ldots, X_{n-1}) = (N, N, \ldots, N)$, then we do not separate a single card. Hence we put the entire deck down and after applying σ, we get the deck in reversed order.

In example (12), we recognized more of a pattern in the case $p = \frac{4}{5}$. This might of course be "dumb luck". But, it is possible to determine the variation distance between the $O(p)$ and U.

Lemma 8. *For a deck of n cards, the variation distance between $O(p)$ and U is:*

$$\big\|O(p) - U\big\| = \frac{1}{2}\left(\frac{n! - 2^{n-1}}{n!} + \sum_{k=0}^{n-1} \binom{n-1}{k} \left| p^k (1-p)^{n-1-k} - \frac{1}{n!} \right| \right)$$

Proof. We know that:

$$\big\|O(p) - U\big\| = \frac{1}{2} \sum_{\pi \in S_n} \big|O(p)(\pi) - U(\pi)\big| = \frac{1}{2} \sum_{\pi \in S_n} \left| O(p)(\pi) - \frac{1}{n!} \right|$$

Moreover, there are only 2^{n-1} permutations with nonzero probability of occurring with a Overhand shuffle. We can see this, because there are $n-1$ options of choosing Y or N, which generate all distinct permutations. This explains the term

$$\frac{n! - 2^{n-1}}{n!} = (n! - 2^{n-1}) \left| 0 - \frac{1}{n!} \right|$$

Also, for every other permutation, the probability of occurring is determined by the amount of Y's (k times). If Y occurs k times in a permutation, then N occurs $(n-1) - k = n - 1 - k$ times. However, for every k, there are $\binom{n-1}{k}$ such permutations. We have to do this for every $k \in \{0, \ldots, n-1\}$. This explains the other term:

$$\sum_{k=0}^{n-1} \binom{n-1}{k} \left| p^k (1-p)^{n-1-k} - \frac{1}{n!} \right|$$

Combining the two gives the result. □

Going back to example 12, we can fill in $n = 10$. Then we get: $\big\|O(\frac{1}{2}) - U\big\| = \big\|O(\frac{4}{5}) - U\big\| = \frac{14173}{14175} \approx 1$. So in the case $n = 10$, it does not matter if we pick

$p = \frac{1}{2}$ or $p = \frac{4}{5}$, the different Overhand shuffles seem to be evenly bad. For lower values of n, we see that $p = \frac{1}{2}$ is our best option.

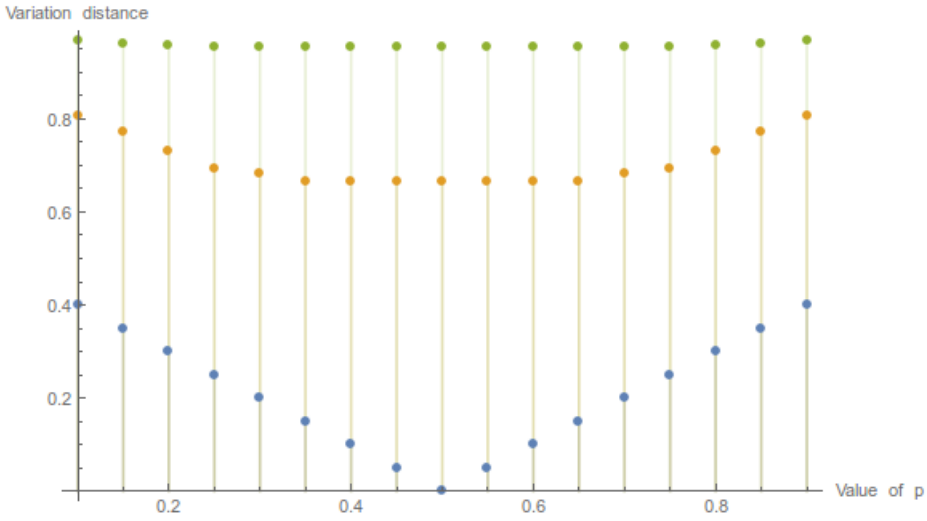

Figure 3: The variation distance between $O(p)$ and U (y-axis) for different values of deck size n. Here, the blue dots represent $n = 2$, the yellow dots represent $n = 4$ and the green dots represent $n = 6$. Figure obtained using Mathematica.

6.1 Direct calculation of variation distance for multiple shuffles

Sadly, there is no theorem for multiple shuffles as with the Riffle shuffle (theorem 2). However, we can still calculate the variation distance $\left\|O^k(p) - U\right\|$ for $k > 1$. How does this work? Well, for a deck of size n, create a $n! \times n!-$matrix for every possible permutation of the deck (of course, this takes a while, more on that at the end of this section). Then, fill in the transition probabilities for going from a certain deck ordering to another deck ordering with one Overhand shuffle. *In this section, do not apply the reversing order σ.* We can do this for every state of the deck to every state of the deck.

Example 13. Take $n = 3$. There are $3! = 6$ different decks, namely: $[123], [132], [213], [231], [312], [321]$. Hence, we have an $6 \times 6-$matrix, denote $A(p)$. From the deck $[123]$, there are 4 different permutations possible. Firstly: YY, this gives the deck $[321]$ with probability p^2. Secondly: YN, this gives the deck $[231]$ with probability $p(1-p)$. Thirdly: NY, this gives the deck $[312]$ with probability $p(1-p)$. Lastly: NN, this gives the deck $[123]$ with probability $(1-p)^2$. So,

we get:

$$\begin{array}{c} \\ [123] \end{array} \begin{array}{c} \begin{array}{cccccc} [123] & [132] & [213] & [231] & [312] & [321] \end{array} \\ \begin{pmatrix} (1-p)^2 & 0 & 0 & p(1-p) & p(1-p) & p^2 \end{pmatrix} \end{array}$$

Of course, we can do this with the other states (decks) as well and complete the matrix:

$$\begin{array}{c} \\ [123] \\ [132] \\ [213] \\ [231] \\ [312] \\ [321] \end{array} \begin{array}{c} \begin{array}{cccccc} [123] & [132] & [213] & [231] & [312] & [321] \end{array} \\ \begin{pmatrix} (1-p)^2 & 0 & 0 & p(1-p) & p(1-p) & p^2 \\ 0 & (1-p)^2 & p(1-p) & p^2 & 0 & p(1-p) \\ 0 & p(1-p) & (1-p)^2 & 0 & p^2 & p(1-p) \\ p(1-p) & p^2 & 0 & (1-p)^2 & p(1-p) & 0 \\ p(1-p) & 0 & p^2 & p(1-p) & (1-p)^2 & 0 \\ p^2 & p(1-p) & p(1-p) & 0 & 0 & (1-p)^2 \end{pmatrix} \end{array}$$

Since this forms a Markov chain, we can calculate $A(p)^k$ and from there derive the probability that a deck is in a certain state, when it began in the deck ordering [123]. For this, we only need the first row of the matrix. So, we multiply by the row vector: $(1,0,0,0,0,0)$ from the left. Then we substract a row vector with the uniform distribution values: $\frac{1}{6}(1,1,1,1,1,1)$. After that, take the absolute value of the row vector and sum its entries.

Let me illustrate this with $A(\frac{1}{2})^1 = A(\frac{1}{2})$:

$$\begin{pmatrix} 1 & 0 & 0 & \dots \end{pmatrix} \begin{pmatrix} \frac{9}{16} & 0 & 0 & \frac{3}{16} & \frac{3}{16} & \frac{1}{16} \\ 0 & \frac{9}{16} & \frac{3}{16} & \frac{1}{16} & 0 & \frac{3}{16} \\ 0 & \frac{9}{16} & \frac{3}{16} & \frac{1}{16} & 0 & \frac{3}{16} \\ \frac{3}{16} & \frac{1}{16} & 0 & \frac{9}{16} & \frac{3}{16} & 0 \\ \frac{3}{16} & 0 & \frac{1}{16} & \frac{3}{16} & \frac{9}{16} & 0 \\ \frac{1}{16} & \frac{3}{16} & \frac{3}{16} & 0 & 0 & \frac{9}{16} \end{pmatrix} = \begin{pmatrix} \frac{9}{16} & 0 & 0 & \frac{3}{16} & \frac{3}{16} & \frac{1}{16} \end{pmatrix}$$

We then substract $\frac{1}{6}(1,1,1,1,1,1)$ and take the absolute value of the entries:

$$\left| \begin{pmatrix} \frac{9}{16} & 0 & 0 & \frac{3}{16} & \frac{3}{16} & \frac{1}{16} \end{pmatrix} - \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \right| = \begin{pmatrix} \frac{19}{48} & \frac{1}{6} & \frac{1}{6} & \frac{1}{48} & \frac{1}{48} & \frac{5}{48} \end{pmatrix}$$

Summing the entries and multiplying by $\frac{1}{2}$ yields:

$$\left\| O(\frac{1}{4}) - U \right\| = \frac{1}{2}\left(\frac{19}{48} + \frac{1}{6} + \frac{1}{6} + \frac{1}{48} + \frac{1}{48} + \frac{5}{48}\right) = \frac{7}{16}$$

Because we only did one shuffle, we can verify our result by using lemma 8. Indeed:

$$\left\| O(\frac{1}{4}) - U \right\| = \frac{1}{2}\left(\frac{3! - 2^{3-1}}{3!} + \sum_{k=0}^{3-1} \binom{3-1}{k} \left| (\frac{1}{4})^k (1 - \frac{1}{4})^{3-1-k} - \frac{1}{3!} \right| \right) = \frac{7}{16}$$

Once again, let $n = 3$ and see the following figure. As you will notice, it takes five shuffles to get a small variation distance, for the value $p = \frac{1}{2}$ in the Overhand shuffle. This is quite much for a deck of 3 cards.

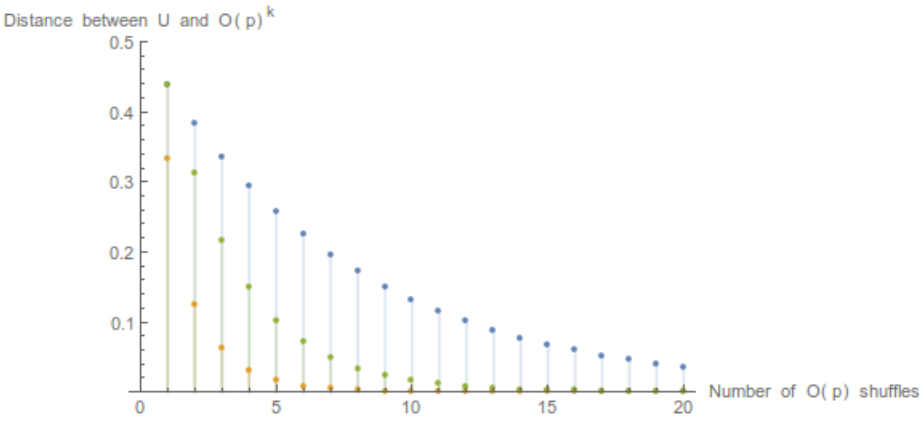

Figure 4: The variation distance between $O(p)$ and U (y-axis) for different values of consecutive shuffles k. Here, the blue dots represent $p = \frac{1}{4}$, the yellow dots represent $p = \frac{1}{2}$ and the green dots represent $p = \frac{3}{4}$. Figure obtained using Mathematica.

What about larger values of n? Well, sadly, we are not able to do the computations for a deck of 52 cards. This would take a $52! \times 52!-$matrix, something computers can not handle. So, while we are able to do multiple shuffles, this method is not really useful for large n.

6.2 Bound for Overhand shuffle

As we concluded, we can not work with large n using matrices. However, we can estimate a lower bound for the necessary amount of shuffles.

Theorem 5. *A lower bound of shuffles for the Overhand shuffle with parameter* $p \in (0, 1)$ *is given by:*

$$\frac{p^2(2-p)}{8\pi^2(1-p^2)} n^2 \log n$$

Oppositely to the Riffle shuffle bounds, I will not prove this theorem. An extensive proof for this theorem can be found in a paper of Jonasson, who came up with this lower bound theorem and its proof in [9]. Note that for smaller values of p, we need less Overhand shuffles according to the lower bound. But be careful, this does not tell us anything, since we need an upper bound as well. For example, if we use $p = \frac{1}{4}$ and $n = 3$, then the lower bound takes a value of less than 0.02. So we should not need any shuffles according to this lower bound. We already saw in figure 6.1 that this is not nearly enough. Since Pemantle already proved in [14] that the Overhand shuffle is at most of order $n^2 \log n$, we can conclude that we need about $\mathcal{O}(n^2 \log n)$ Overhand shuffles to mix up a deck of n cards.

What can we conclude from the lower bound or the mixing time of $\mathcal{O}(n^2 \log n)$? Well, if we take a standard deck of $n = 52$ cards and perform consecutive Overhand shuffles with $p = \frac{1}{2}$, the lower bound is approximately 67. Using only this lower bound, we can already conclude that Riffle Shuffling is a lot more effective to mix up a deck of cards.

7 Fisher-Yates shuffle

As with the Top-in shuffle, there is another shuffle that truly randomizes the deck. This is known as the Fisher-Yates shuffle (see [2]).

Definition 11. Suppose we have a deck of n cards in its original order: $[(1) \dots (n)]$. We proceed by doing $n - 1$ iterations:
If the size of the current deck is larger than 1, remove a random card of this current deck. Add this to the bottom of the new deck. If the size of the current deck is 1, stop. Your deck has now been completely randomized using a Fisher-Yates shuffle (denoted FS).

The only thing a card shuffler needs to do is pick $n - 1$ random numbers in different domains. The i-th random number should be an integer k with $1 \leq k \leq n - i + 1$, chosen uniformly. A proposition to prove that this shuffle indeed randomizes the deck.

Proposition 1. *The Fisher-Yates shuffle indeed randomizes a deck of n cards.*

Proof. Suppose we have a deck of n cards in its original order: $[(1) \dots (n)]$. If we can show that the probability that a card on position x ends on position y after the shuffle is $\frac{1}{n}$, then we are done. For this, denote the probability that a card on position x ends on position y after the shuffle with $\mathbb{P}_{xy}$. Firstly: $\mathbb{P}_{x1} = \frac{1}{n}$, since the card x should be chosen first. This happens with uniform probability $\frac{1}{n}$, since there are n cards left. From now on, assume $y \geq 2$. To end on position

y, the card numbered x should not be chosen in the first $y-1$ times of removing a card. To not be chosen in the i-th try has probability: $\frac{n-(i-1)-1}{n-(i-1)} = \frac{n-i}{n-i+1}$, because there are $n-(i-1)$ cards left. After that, the card should be chosen in the y time of removing a card (with probability $\frac{1}{n-(y-1)}$). This gives, for all $y \geq 2$:

$$\mathbb{P}_{xy} = \Big(\prod_{i=1}^{y-1} \frac{n-i}{n-i+1}\Big)\frac{1}{n-(y-1)} = \frac{n-y+1}{n}\frac{1}{n-y+1} = \frac{1}{n}$$

Conclude that $\mathbb{P}_{xy} = \frac{1}{n}$, for every $(x,y) \in \big(\mathbb{N}_{\leq n}\big)^2$, hence we are done. □

A small example to illustrate the Fisher-Yates shuffle. In iteration i, the random integer k, with $1 \leq k \leq n-i+1$ is generated in Mathematica by "RandomInteger[$\{1, n-i+1\}$]".

Example 14. Suppose $n = 9$, the random integers generated (in order) are: $(8, 4, 7, 3, 2, 1, 3, 1)$. I will handle all iterations.

step	k	old deck before removal	card removed	new deck after iteration
1	8	[123456789]	8	[8]
2	4	[12345679]	4	[84]
3	7	[1235679]	9	[849]
4	3	[123567]	3	[8493]
5	2	[12567]	2	[84932]
6	1	[1567]	1	[849321]
7	3	[567]	7	[8493217]
8	1	[56]	5	[849321756]

Note that in the last iteration, two cards are added to the new deck. This is because card 6 clearly must be the last one in the new deck.

One could always shuffle a deck in this way and this needs only n steps ($n-1$ iterations and adding the remaining card to the bottom of the deck). However, randomly selecting a card/number is not something most people are good at. So for a card game, it could be possible to let an outsider generate $n-1$ numbers and randomize the deck. Why an outsider? Well, since the shuffle is not as fast for the eye as for example the Riffle shuffle. A card player who pays attention, knows the order of the new deck. Therefore the Fisher-Yates shuffle is theoretically a good mixing strategy for cards, however it has its cons.

8 Conclusion

A lot of different shuffling methods came up in this thesis. Some are more theoretical and randomized a deck really easily. Others are of more practical

use, but turn out to be harder to analyse mathematically. The variation distance turned out to be a useful tool in the analysis of card shuffling. For every deck of n cards and every shuffle Q, we can analysize $Q^{(k)}$ by means of a $n! \times n!-$matrix. For some methods, especially the Riffle shuffle, better tools have been developed. As a general conclusion: the Riffle shuffle turns out to be of most practical use in casual and professional card games (i.e. in real life, with real cards). For online gambling, Fisher-Yates seems to be the best alternative. I convinced myself to stop using the Overhand shuffle, because although it is generally easier, it takes certainly more time to shuffle a deck of cards than with the Riffle shuffle.

9 Acknowledgements

I would like to thank dr. Karma Dajani, my supervisor. She has created an environment in which I had a lot of freedom of writing my bachelor's thesis in my personal style. On the other hand, she was very keen and fast when checking for improvements. Also, I would like to thank Sophie de Heus and Erica Besselink for an English grammar check, despite not being home in the field of mathematics.

References

[1] URL: `https://en.wikipedia.org/wiki/Fisher-Yates_shuffle`.

[2] URL: `https://en.wikipedia.org/wiki/Eulerian_number`.

[3] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

[4] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8(1):69–97, 1987.

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

[6] Louis Comtet. *Advanced Combinatorics: The art of finite and infinite expansions*. Springer Science & Business Media, 2012.

[7] Persi Diaconis, Ron Graham, and Ronald L Graham. *Magical mathematics: the mathematical ideas that animate great magic tricks*. Princeton University Press, 2011.

[8] Charles Miller Grinstead and James Laurie Snell. *Introduction to probability*. American Mathematical Soc., 2012.

[9] Johan Jonasson et al. The overhand shuffle mixes in θ (n2logn) steps. *The Annals of Applied Probability*, 16(1):231–243, 2006.

[10] Gina Kolata. In shuffling cards, 7 is winning number. *New York Times, Spätausgabe vom*, 9:1990, 1990.

[11] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[12] Philip Liang. Finite markov chains and the top-to-random shuffle. URL: `http://math.uchicago.edu/~may/REU2013/REUPapers/Liang.pdf`.

[13] Brad Mann. How many times should you shuffle a deck of cards. *Topics in Contemporary Probability and Its Applications*, 15:1–33, 1995.

[14] Robin Pemantle. Randomization time for the overhand shuffle. *Journal of Theoretical Probability*, 2(1):37–49, 1989.