The aim of these notes is to introduce you to the mathematics of card shuffling. The notes are very sketchy but, hopefully, they will show you how various aspects of mathematics help us to understand a basic real life phenomenon: the mixing of a deck of card.

The basic question is to decide how many times a deck of cards should be shuffled in order for it to be well mixed.

The first observation is that, in order to be able to answer this question, we must make it much more precise. In particular, in order for mathematics (versus, say experiments) to be helpful, we need to make up a model that “idealizes” card shuffling. A good starting point is also to realize that what we want to understand is quite complex and that, may be, we need first to look at simpler instances of similar problems.

We are going to discuss the following basic questions:

1. What is a “deck of card”, how can one describe a “deck of cards”?
2. What does it mean for a deck of card to be mix up?
3. How can one describes the process of shuffling cards?

To answer (1), we can think of the pack of cards as an object that can be in several different states (i.e., the different arrangements of the cards). A deck is one of these states. The set of all states is the “universe” we are interested in because we want to pick one of these states “at random”.

This takes us to (2) because “at random” is not well defined yet. After a bit of thought, it may be reasonable to think that, in an ideal word, all arrangements (i.e., all states) should be equally likely when we pick a deck to play a game. The point is that, in such an ideal situation, the different players will have the least amount of information before the game begins. If all states are equally likely then, the probability of any given state must be $1/N$ where $N$ is the total number of states. This is called the uniform probability. When we mix up the cards, we are trying to pick a state (an arrangement of the pack) uniformly at random among all possible states. I say we are trying because we may not able to do it. We may only be able to approximate this idealized situation. Hopefully, the more we shuffle the pack, the closer to the idealized situation we get.

You can find a different account of some of the material in: Introduction to Probability, by Charles Grinstead and Laurie Snell, published by the American Mathematical Society (Chapter 3, Section 3, pages 119-131).
Probability

Forget about cards for a while but keep in mind the idea of a finite state space (i.e., a finite universe) that we want to explore. Say our finite universe has \( N \) elements (states) which we call 1, 2, \ldots, \( N \). A probability measure on this universe is a \( N \)-tuple \( p = (p_1, \ldots, p_N) \) of real numbers such that each \( p_i \) is greater or equal to 0 and the sum is 1:

\[
p_1 + \ldots + p_N = \sum_{1}^{N} p_i = 1.
\]

The real \( p_i \in [0,1] \) is the probability of the state \( i \) under this probability measure. Moreover, if \( A \) is a subset of our universe \( \{1, \ldots, N\} \), we define \( p(A) \), the probability of \( A \), to be

\[
p(A) = \sum_{i \in A} p_i.
\]

For instance,

\[
p(\{1, 3, 5\}) = p_1 + p_3 + p_5.
\]

# Now you can explain why, if every state must have the same probability, this probability must be \( 1/N \).#

The probability measure that gives the same probability \( 1/N \) to each state is called the uniform probability

\[
u = (1/N, \ldots, 1/N)
\]

(i.e., \( u_i = 1/N \) for \( i = 1, 2, \ldots, N \)).

# If a deck of card is perfectly mix up, what is the probability that the top card is the ace of Spades? What is the probability that the twenty-third card is the 4 of Hearts?#

As noted earlier, we will not be able to produce a perfectly mix up deck of cards but only an approximately mix up deck of cards. How do we measure the difference (discrepancy) between our ideal situation and its approximation? This is an important question with many possible answers. Our idealize situation is represented by the uniform measure \( u \), the approximation of it, by another measure \( p \). One way to measure the difference between to probability measures \( p, q \) is to use the Total Variation distance:

\[
\|p - q\|_{TV} = \sup_{A \subset \{1, \ldots, N\}} \{p(A) - q(A)\}.
\]
Given two measures \( p = (p_1, \ldots, p_N) \), \( q = (q_1, \ldots, q_n) \), describe the sets \( A_* \) such that
\[
\|p - q\|_{TV} = p(A_*) - q(A_*).
\]
Show that
\[
\|p - q\|_{TV} = \frac{1}{2} \sum_{i=1}^{N} |p_i - q_i|.
\]

The last formula gives us a convenient way to compute the total variation distance. The definition says that total variation is the largest error one can possibly make by using \( p \) to approximate the probability of any event \( A \) under \( q \) (i.e., using \( p(A) \) as an approximation of \( q(A) \)).

**Markov chains**

A Markov chain is a mechanism that can serve to explore a (finite) universe by making random moves. For each state \( i \), we are given probabilities \( k_{i,j} \) where \( k_{i,j} \) is the probability that the next step will take us to state \( j \) (hence, \( k_{i,j} \geq 0 \) and \( k_{i,1} + \ldots, k_{i,N} = 1 \)). We arrange all these numbers (\( N^2 \) of them) into a big square matrix where the \( i \)-th row is \( k_{i,1}, \ldots, k_{i,N} \). We denote this array by \( K \) (in mathematics, it is called a square matrix of dimension \( N \)).

Now, if at time 0, the probability to be at \( i \) is \( p_0^i \), at time 1, after one step of our process, we are at \( j \) with probability
\[
p_j^1 = p_0^i k_{1,j} + p_0^2 k_{2,j} + \ldots + p_0^N k_{N,j} = \sum_{i=1}^{N} p_0^i k_{i,j}.
\]
More generally, if at time \( t \), the probability to be at \( i \) is \( p_t^i \), at time \( t + 1 \), we are at \( j \) with probability
\[
p_j^{t+1} = p_t^1 k_{1,j} + p_t^2 k_{2,j} + \ldots + p_t^N k_{N,j} = \sum_{i=1}^{N} p_t^i k_{i,j}.
\]
Note that, for each time \( t \), \( p^t = (p_t^1, \ldots, p_t^N) \) is a probability measure and that the formula above tells us how to compute \( p^{t+1} \) from \( p^t \). A probability can be understood as a row matrix (1 row, \( N \) columns) and the formula for \( p^{t+1} \) can be written as
\[
p^{t+1} = p^t K
\]
where $p^t K$ means the matrix multiplication of the $(1,N)$-matrix $p^t$ by the $(N,N)$-matrix $K$. Iteration of this formula leads to

$$p^t = p^0 K^t$$

where $K^t$ is the product of the square matrix $K$ by itself, $t$ times.

# It is a nice exercise to use the important technique called induction to prove the last formula using (1) and the fact that $K^{t+1} = K^t K$. #

# Consider the universe $\{1, 2, 3\}$ with

$$K = \begin{pmatrix}
\frac{1}{2} & 1/4 & 1/4 \\
1/2 & 0 & 1/2 \\
0 & 1/2 & 1/2
\end{pmatrix}.$$ 

If we start in state 3 (that is, $p_1^0 = 0, p_2^0 = 0, p_3^0 = 1$), what is the probability we move to state 2 in 1 step? Given that we start at 3, compute the probabilities $p_1^2, p_2^2, p_3^2$ of being at 1, 2, 3, respectively, after two steps. #

# Consider the universe $\{1, 2, 3\}$ with

$$K = \begin{pmatrix}
\frac{1}{2} & 1/2 & 0 \\
1/2 & 0 & 1/2 \\
0 & 1 & 0
\end{pmatrix}.$$ 

A token starts at 1 and is moved 10 times according to $K$. You must guess where the token will be at the end. Is there a guess that is more likely to be correct? #

The theory of Markov chains study such random processes and answer questions regarding the long time behavior of such processes. One crucial property that these random processes have is that the future depends only on the present, not on the past. This property is what make these processes easier to study than other more complex processes. It is called the Markov property.

Sometimes, after a long time, the probability of being at an arbitrary given state $j$ stabilizes. More precisely,

$$p_j^t \longrightarrow p_j^\infty \quad \text{as} \quad t \longrightarrow \infty.$$ 

(2)

Taken together, the numbers $p_j^\infty$ form a probability measure $p^\infty = (p_1^\infty, \ldots, p_N^\infty)$ and we say that the probability $p^t$ converges to $p^\infty$.

# Use $\epsilon, \delta$ or theorems concerning limits to show that if for all $j = 1, \ldots, N$, (2) is satisfied, then

$$\lim_{t \rightarrow \infty} \|p^t - p^\infty\|_{TV} = 0.$$ #
When $p^t$ converges to $p^\infty$, one can think of using $p^t$ as an approximation of $p^\infty$. Here is the basic criterion that allows us to decide when $p^t$ converges to $p^\infty$.

**Theorem** For $p^t$ to converge towards a probability $p^\infty$, it suffices that there exists a large time $t$ such that it is possible to move from any state $i$ to any states $j$ in exactly $t$ steps. If that is the case then the equation $x = xK$, equivalently the system

$$
\begin{align*}
x_1 & = x_1k_{1,1} + \cdots + x_Nk_{1,N} \\
x_2 & = x_1k_{1,2} + \cdots + x_Nk_{2,N} \\
& \quad \vdots \\
x_N & = x_1k_{1,N} + \cdots + x_Nk_{N,N}
\end{align*}
$$

has a unique solution satisfying $x_1 + \cdots + x_N = 1$. Moreover, $p^\infty = (x_1, \ldots, x_N)$.

# Verify the sufficient condition given in the theorem above in the case of the two chains on three points discussed earlier. In each cases, solve the corresponding system and compute $p^\infty$. #

**Card decks, permutations**

To describe a deck, i.e., the present state of a pack of cards, we need to give the order of the cards. For instance, if the deck is made of $n$ cards numbered 1 to $n$, we can described the deck (i.e., the state of the pack) by giving the values of the cards in order from top to bottom as a string of integers. If there are 5 cards, $(3, 5, 2, 1, 4)$ describes the deck which as 3 on top, 5 in second position, 2 in third position, 1 in fourth place and 4 at the bottom. Note that the string in question contains exactly once each of the integer 1, \ldots, 5. Such a string is called a permutation of $\{1, \ldots, 5\}$ (i.e., of 5 objects). The name permutation comes from a different perspective which is also important to us. Namely, if we think of starting with the cards in order: 1 on top, 5 at the bottom, the string $(3, 5, 2, 1, 4)$ tell us that the card originally in position 1 is now in position 4, the card originally in position 2 is now in position 3, the card originally in position 3 is now in position 1, the card originally in position 4 is now in position 5, the card originally in position 5 is now in position 2. That is, the string tells us how the cards have been permuted.

# How many different decks of 3 cards are there? How many different decks of $n$ cards are there? (explain why)#
The answer is that there are
\[ n! = n \times (n - 1) \times (n - 2) \ldots 3 \times 2 \times 1 \]
different deck of \( n \) cards. For instance \( 32! \) is approximately \( 2.6 \times 10^{35} \) and \( 52! \) approximately \( 8 \times 10^{67} \).

So the little finite universe we are interested in has about \( 10^{68} \) elements. This means that the uniform probability measure on the set of all decks of 52 cards gives a probability of (very roughly) \( 10^{-68} \) to each particular deck.

Suppose you live 100 years and each day of your life you play 100 poker games (32 cards) with, each time, a perfectly mixed new deck. Then, you will get to see roughly \( 4 \times 10^6 \) decks, each of which is picked uniformly at random among the \( 2.6 \times 10^{35} \) possibilities. When you finally die (exhausted and bored!) you will have explored only a minuscule part of all possibilities and it is almost certain that you will have never seen the same deck twice.

# Try to give a convincing argument that the probability to see the same deck twice is extremely small #

# Use basic principles to compute the probability that a perfectly mixed deck has the ace of Spades on top and the ace of Hearts at the bottom #

Before we move on, let us talk about how to recognize whether or not a deck of cards is well mixed up. That is, if you are presented a deck of cards, how do you decide if it is likely to be a well mixed up deck or if it is unlikely it is a well mix deck. This is a question of great importance in many practical problems and answering this type of question is one of the occupations of statisticians. One basic idea is to compute some “characteristic” and to see how the computed value compare with the expected value of that characteristic under the hypothesis that the deck is well mix.

For instance, given a permutation, we can compute the number of fixed points of that permutation (a fixed point is an object that has not been moved). For instance, the permutation of 1, \ldots, 9 described by the string (3, 2, 5, 4, 1, 9, 6, 8, 7) has 3 fixed points (namely, 2, 4, 8).

# What is the expected number of fixed points of a permutation of \( n \) objects picked uniformly at random? What is the expected number of pairs \( i < j \) such that the objects originally at \( i, j \) have been switched? #

The cycle \([a_1, a_2, \ldots, a_k]\) is the permutation that takes \( a_1 \) to \( a_2 \), \( a_2 \) to \( a_3 \), \ldots, \( a_{i-1} \) to \( a_i \), \ldots, \( a_{k-1} \) to \( a_k \) and \( a_k \) to \( a_1 \). Here we will use square brackets to denote cycles although this is not the standard notation. The integer \( k \) is the length of the cycle. The set \( \{a_1, \ldots, a_k\} \) is the support of the cycle (objects not in the support are not moved). A cycle of length 2 is called a transposition (it transposes two objects).
We can put different cycles with disjoint supports together to form a more complex permutation. For instance, over \{1, 2, 3, 4, 5\} we can take the cycle [3, 5] and [2, 1, 4] to make [3, 5][2, 1, 4] which is the permutation (2, 4, 5, 1, 3).

Show that any permutation of \(n\) objects can be decomposed into cycles with disjoint supports.

The symmetric group

A group is a set \(G\) with a “pairing” between elements of \(G\) that produces elements of \(G\). That is, given two elements \(a, b\) in \(G\) we can compute \(ab\) which is an element of \(G\). This pairing is called the group law. It must satisfy some special properties. Namely:

1. For all \(a, b, c \in G\), \((ab)c = a(bc)\)
2. There exists an element \(e\) in \(G\) such that \(ea = ae = a\) for all \(a\) in \(G\) (this is called the identity element or neutral element).
3. For any element \(a\) in \(G\) there exists a unique element \(b\) (depending on \(a\)) such that \(ab = ba = e\). We call \(b\) the inverse of \(a\) and use the notation \(b = a^{-1}\).

Note that it may well be that \(ab \neq ba\).

Prove that, in a group \(G\), for given \(a, b\) the equation \(ax = b\) always has a unique solution.

Prove that the set \(\{0, \pm 1, \pm 2, \pm 3, \cdots\}\) with the law \(ab = a + b\) is a group. Explain why the set \(\{1, 2, 3, \cdots\}\) with the law \(ab = a \times b\) is NOT a group.

The notion of group is one of the most important notions of modern mathematics. The study of groups and other structure like this is part of algebra but groups are also central to the modern development of geometry because of the notion of group of transformations (for instance, rotations, dilations, translations are all elements of various groups of transformations that are crucial to our understanding of space), and of group of symmetries. The law of a group of transformations is always the same, that is, the \textit{composition} of transformations as in: translate by this amount in this direction then rotate by this angle which describes the composition of a translation followed by a rotation.

Fix \(n\). Permutations can be thought of as transformations of \(\{1, \ldots, n\}\). This can be visualized by looking at a rack with \(n\) slots in which \(n\) numbered balls are placed. Each permutation \textit{permutes} the balls, i.e., \textit{transforms} the
position of the balls in the rack. Now, it is clear that we can compose two permutations to get a third permutation.

#The set of all permutations of \{1, \ldots, n\} equipped with composition is a group called the symmetric group. Go through the definition to verify that this is indeed a group. #

The support of a permutation is the subset of \{1, \ldots, n\} of those elements that are moved by the permutation. If two permutations \(a, b\) have disjoint support, they commute, that is \(ab = ba\).

#Show that a cycle of length \(k\) is a product of \(k - 1\) transpositions #

# Recall that any permutation is a product of cycles with disjoint supports. Show that any permutation is a product of a finite number of transpositions.

# Show that a cycle \(c\) of length \(k\) satisfies \(c^k = Id\) where \(Id\) is the identity permutation, i.e., the permutation that leaves everything in place. If \(a = c_1 \ldots c_\ell\) with \(c_1, \ldots, c_\ell\) being cycle with disjoint supports of respective length \(k_1, \ldots, k_\ell\), show that \(a^m = Id\) where \(m\) is the smallest common multiple of \(k_1, \ldots, k_\ell\).#

**Shuffling methods**

In the previous paragraph we developed the basic material needed to describe what a shuffling methods is, in mathematical terms.

The finite universe we would like to explore is the set of all decks of cards (meaning different arrangements of a given pack of cards). This is the same as the set of all permutations of the cards (we can identify a deck of cards with the permutation (i.e., transformation) that take the pack in order to the desired deck (i.e., arrangement). It is a set with \(n!\) elements for a pack with \(n\) cards.

The reason for shuffling is that we want a deck close to a deck picked uniformly at random among all decks. To do this, the idea is to use a Markov chain. Here, this means, a fixed procedure that changes a given deck into one of a few different decks picked according to a certain random scheme. It is best to look at simple examples.

- **Top to random** The top card is picked and inserted in the pack at a uniformly chosen random position.

- **Random transpositions** Think of the deck arranged in a row on the table. Let the left hand and right hand each pick uniformly, independently
at random a position and switched the cards at the chosen position. If the hands have picked the same position, nothing changes.

- **Riffle shuffle** The deck is cut into two packs that are then rifflled into each other.

Do you notice differences in the descriptions above? I hope you do!

# What happens to the shuffling methods above for a pack of TWO cards? Draw diagrams with arrows and probabilities describing these Markov chains in the case of a pack of THREE cards. #

We start by describing results for the first two shuffling methods mentioned above. For both top to random and random transposition, if one shuffles many time, the deck becomes mix up. That is, after many shuffles the deck of cards resembles a deck picked uniformly at random. More formally, if we call $p^t$ the probability distribution (on card decks) after $t$ shuffles (say we always start with an ordered deck) and call $u$ the uniform probability, we have

$$
\|p^t - u\|_{TV} \to 0
$$
as $t$ tends to infinity.

But how many times should we repeat the basic shuffling step? The answer is different for different shuffling methods!

For top to random with a deck of $n$ cards, $n \ln n$ does the job. For random transpositions of $n$ cards, $\frac{1}{2} n \ln n$ does the job.

# Call $\phi(a)$ the number of fixed points of the deck $a$ minus 1. Thus $\phi(a) = n - 1$ if the deck $a$ is in order and $\phi(a) = -1$ is $a$ has no fixed point. Show that starting from a fixed deck $a$, the average value of $\phi$ after a random transposition step is $(1 - 2/n)\phi(a)$. #

What about riffle shuffle? Well first we need to say more precisely what a riffle shuffle is. How do we cut the deck? how do we riffle together the two packs?

To model the cut (i.e., the choice of the size $k$ of the top pack), we use the binomial probability, that is, the probability to cut $k$ cards is

$$
2^{-n} \binom{n}{k}
$$

where $\binom{n}{k}$ is the number of different ways to pick a subset of $k$ elements in a set of $n$ elements.

# Explain why there are $2^n$ different subsets of a given set with $n$ elements. Use this to explain why the sum of the numbers $2^{-n} \binom{n}{k}$ when $k$ varies from 0 to $n$ is 1#
The last statement is important to us: it says that \(2^{-n \binom{n}{k}}\) is indeed a probability distribution over the set of possible cuts \(\{0, 1, \ldots, N\}\).

Now we have to packs, of size \(k\) and \(n-k\), one in each hand. To model the rifflle part of the shuffling we say that cards fall from each hand with probability proportional to the size of the pack in that hand. In other words, if there are \(m\) cards in the left hand and \(\ell\) cards in the right hand, the next card falls from the left hand with probability \(m/(m+\ell)\). This model was invented by Gilbert and Shannon in 1955 at Bell Laboratories.

For this model, it then takes \(\frac{3}{2} \ln_2 n\) rifflle shuffles to mix up the decks. More precisely

\[
\|p^t_{\ln_2 n+c} - u\|_{TV} = 1 - \frac{2}{\sqrt{2\pi}} \int_{\infty}^{-2^{-c/4}\sqrt{3}} e^{-s^2/2} ds + O(n^{-1/4}).
\]

In fact, one can gives an exact formula for the distribution \(p^t\) of a deck of \(n\) cards after \(t\) rifflle shuffles, namely

\[
p^t(a) = 2^{-tn} \binom{n+2^t-r}{n}.
\]

Here the integer \(r\) depends on the deck \(a\) and is the number of rising se-

quences in \(a\). A rising sequence in a deck is a maximal subset of cards con-

sisting of successive face values displayed in order. For instance the deck

\((2, 4, 3, 9, 1, 6, 7, 8, 5)\) (9 cards) has five rising sequences: namely

\(1; 2, 3; 4, 5; 6, 7, 8; 9\).

This result —obtained around 1990 by Dave Bayer and Persi Diaconis—what featured in a New-York time article. It leads to the following table which gives \(\|p^t - u\|_{TV}\) for \(t = 1, \ldots, 10\) with 3 decimals for \(n = 52\) cards.

<table>
<thead>
<tr>
<th>(t)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(|p^t - u|_{TV})</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.924</td>
<td>0.614</td>
<td>0.334</td>
<td>0.167</td>
<td>0.085</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Table 1: The total variation distance for \(k\) rifflle shuffles of 52 cards.
Exercises

1. (Coupling) For two random variables $X, Y$ defined on a probability space $(\Omega, P)$ and taking values in a finite set $S$ with law $\mu$ and $\nu$, respectively (i.e., $P(X = x) = \mu(x)$, $P(Y = y) = \nu(y)$), show that $\|\mu - \nu\|_{TV} \leq P(X \neq Y)$.

2. (Repeated experiment) The result of a certain repeatable experiment is positive with probability $p$. If this experiment is repeated (infinitely many times, what is the probability that the first positive result occurs at the $k$-th trial? What is the average number of trials until the first success?

3. (Coupon collector problem) You buy cereal boxes containing baseball cards. If the complete collection has $N$ cards, what is the average number of boxes one needs to buy to obtain the complete collection?

4. (Random to top) Use the previous questions to compute the average "coupling time" for the coupling described in class for "Random to Top".

5. (Eigenfunction) Call $\phi(a)$ the number of fixed points of the deck $a$ minus 1. Thus $\phi(a) = n - 1$ if the deck $a$ is in order and $\phi(a) = -1$ is $a$ has no fixed point. Show that starting from a fixed deck $a$, the average value of $\phi$ after a random transposition step is $(1 - 2/n)\phi(a)$.

6. Show that a cycle of length $k$ is a product of $k - 1$ transpositions. Shows that any permutation is a product of disjoint cycles. Show that any permutation of $\{1, \ldots, n\}$ is a product of at most $(n - 1)$ transpositions.

7. Show that any permutation of a deck of $n$ cards can be obtained by performing a finite number of top to random moves.