HOW MANY TIMES SHOULD YOU SHUFFLE A DECK OF CARDS?[1]

Brad Mann
Department of Mathematics
Harvard University

ABSTRACT

In this paper a mathematical model of card shuffling is constructed, and used to determine how much shuffling is necessary to randomize a deck of cards. The crucial aspect of this model is rising sequences of permutations, or equivalently descents in their inverses. The probability of an arrangement of cards occuring under shuffling is a function only of the number of rising sequences in the permutation. This fact makes computation of variation distance, a measure of randomness, feasible; for in an n card deck there are at most n rising sequences but $n!$ possible arrangements. This computation is done exactly for $n = 52$, and other approximation methods are considered.

1 INTRODUCTION

How many times do you have to shuffle a deck of cards in order to mix them reasonably well? The answer is about seven for a deck of fifty-two cards, or so claims Persi Diaconis. This somewhat surprising result made the *New York Times* [5] a few years ago. It can be seen by an intriguing and yet understandable analysis of the process of shuffling. This paper is an exposition of such an analysis in Bayer and Diaconis [2], though many people have done work on shuffling. These have included E. Gilbert and Claude Shannon at Bell Labs in the 50's, and more recently Jim Reeds and David Aldous.

[1] This article was written for the Chance Project at Dartmouth College supported by the National Science Foundation and The New England Consortium for Undergraduate Education.

2 WHAT IS A SHUFFLE, REALLY?

2.1 Permutations

Let us suppose we have a deck of n cards, labeled by the integers from 1 to n. We will write the deck with the order of the cards going from left to right, so that a virgin unshuffled deck would be written $123 \cdots n$. Hereafter we will call this the natural order. The deck after complete reversal would look like $n \cdots 321$.

A concise mathematical way to think about changing orderings of the deck is given by permutations. A permutation of n things is just a one-to-one map from the set of integers, between 1 and n inclusive, to itself. Let S_n stand for the set of all such permutations. We will write the permutations in S_n by lower case Greek letters, such as π, and can associate with each permutation a way of rearranging the deck. This will be done so that the card in position i after the deck is rearranged was in position $\pi(i)$ before the deck was rearranged. For instance, consider the rearrangement of a 5 card deck by moving the first card to the end of the deck and every other card up one position. The corresponding permutation π_1 would be written

i	1	2	3	4	5
$\pi_1(i)$	2	3	4	5	1

Or consider the so-called "perfect shuffle" rearrangement of an 8 card deck, which is accomplished by cutting the deck exactly in half and then alternating cards from each half, such that the top card comes from the top half and the bottom card from the bottom half. The corresponding permutation π_2 is

i	1	2	3	4	5	6	7	8
$\pi_2(i)$	1	5	2	6	3	7	4	8

Now we don't always want to give a small table to specify permutations. So we may condense notation and just write the second line of the table, assuming the first line was the positions 1 through n in order. We will use brackets when we do this to indicate that we are talking about permutations and not orders of the deck. So in the above examples we can write $\pi_1 = [23451]$ and $\pi_2 = [15263748]$.

It is important to remember the distinction between orderings of the deck and permutations. An ordering is the specific order in which the cards lie in the deck. A permutation, on the other hand, does not say anything about the specific order of a deck. It only specifies some rearrangement, i.e. how one ordering changes to another, regardless of what the first ordering is. For example, the permutation $\pi_1 = [23451]$ changes the ordering 12345 to 23451, as well as rearranging 41325 to 13254, and 25431 to 54312. (What will be true, however, is that the numbers we write down for a permutation will always be the same as the numbers for the ordering that results when the rearrangement corresponding to this permutation is done to the naturally ordered deck.) Mathematicians say this convention gives an *action* of the group of permutations S_n on the set of orderings of the deck. (In fact, the action is a simply transitive one, which just means there is always a unique permutation that rearranges the deck from any given order to any other given order.)

Now we want to consider what happens when we perform a rearrangement corresponding to some permutation π, and then follow it by a rearrangement corresponding to some other permutation τ. This will be important later when we wish to condense several rearrangements into one, as in shuffling a deck of cards repeatedly. The card in position i after both rearrangements are done was in position $\tau(i)$ when the first but not the second rearrangement was done. But the card in position j after the first but not the second rearrangement was in position $\pi(j)$ before any rearrangements. So set $j = \tau(i)$ and get that the card in position i after both rearrangements was in position $\pi(\tau(i))$ before any rearrangements. For this reason we define the *composition* $\pi \circ \tau$ of π and τ to be the map which takes i to $\pi(\tau(i))$, and we see that doing the rearrangement corresponding to π and then the one corresponding to τ is equivalent to a single rearrangement given by $\pi \circ \tau$. (Note that we have $\pi \circ \tau$ and not $\tau \circ \pi$ when π is done first and τ second. In short, the order matters greatly when composing permutations, and mathematicians say that S_n is noncommutative.) For example, we see the complete reversal of a 5 card deck is given by $\pi_3 = [54321]$, and we can compute the composition $\pi_1 \circ \pi_3$.

i	1	2	3	4	5
$\pi_3(i)$	5	4	3	2	1
$\pi_1 \circ \pi_3(i)$	1	5	4	3	2

2.2 Shuffles

Now we must define what a shuffle, or method of shuffling, is. It's just a probability density on S_n, considering each permutation as a way of rearranging the deck. This means that each permutation is given a certain fixed probability of occuring, and that all such probabilities add up to one. A well-known example is the top-in shuffle. This is accomplished by taking the top card off the deck and reinserting it in any of the n positions between the $n-1$ cards in the remainder of the deck, doing so randomly according to a uniform choice. This means the density on S_n is given by $1/n$ for each of the cyclic permutations $[234\cdots(k-1)k1(k+1)(k+2)\cdots(n-1)n]$ for $1 \leq k \leq n$, and 0 for all other permutations. This is given for a deck of size $n=3$ in the following example:

permutation	[123]	[213]	[231]	[132]	[321]	[312]
probability under top-in	1/3	1/3	1/3	0	0	0

What this definition of shuffle leads to, when the deck is repeatedly shuffled, is a random walk on the group of permutations S_n. Suppose you are given a method of shuffling Q, meaning each permutation π is given a certain probability $Q(\pi)$ of occuring. Start at the identity of S_n, i.e. the trivial rearrangement of the deck which does not change its order at all. Now take a step in the random walk, which means choose a permutation π_1 randomly, according to the probabilities specified by the density Q. (So π_1 is really a random variable.) Rearrange the deck as directed by π_1, so that the card now in position i was in position $\pi_1(i)$ before the rearrangement. The probability of each of these various rearrangings of the deck is obviously just the density of π_1, given by Q. Now repeat the procedure for a second step in the random walk, choosing another permutation π_2, again randomly according to the density Q (i.e. π_2 is a second, independent random variable with the same density as π_1). Rearrange the deck according to π_2. We saw in the last section on permutations that the effective rearrangement of the deck including both permutations is given by $\pi_1 \circ \pi_2$.

What is the probabiltiy of any particular permutation now, i.e what is the density for $\pi_1 \circ \pi_2$? Call this density $Q^{(2)}$. To compute it, note the probability of π_1 being chosen, and then π_2, is given by $Q(\pi_1) \cdot Q(\pi_2)$, since the choices are independent of each other. So for any particular permutation π, $Q^{(2)}(\pi)$ is given by the sum of $Q(\pi_1) \cdot Q(\pi_2)$ for all pairs π_1, π_2 such that $\pi = \pi_1 \circ \pi_2$, since in general there may be many different ways of choosing π_1 and then π_2 to get the same $\pi = \pi_1 \circ \pi_2$. (For instance, completely reversing the deck and then switching the first two cards gives the same overall rearrangement as first switching the last two cards and then reversing the deck.) This way of combining Q

with itself is called a *convolution* and written $Q * Q$:

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

Here π_1^{-1} denotes the inverse of π_1, which is the permutation that "undoes" π_1, in the sense that $\pi_1 \circ \pi_1^{-1}$ and $\pi_1^{-1} \circ \pi_1$ are both equal to the identity permutation which leaves the deck unchanged. For instance, the inverse of [253641] is [613524].

So we now have a shorthand way of expressing the overall probability density on S_n after two steps of the random walk, each step determined by the same density Q. More generally, we may let each step be specified by a different density, say Q_1 and then Q_2. Then the resulting density is given by the convolution

$$Q_1 * Q_2(\pi) = \sum_{\pi_1 \circ \pi 2 = \pi} Q_1(\pi_1)Q_2(\pi_2) = \sum_{\pi_1} Q_1(\pi_1)Q_2(\pi_1^{-1} \circ \pi).$$

Further, we may run the random walk for an arbitrary number, say k, of steps, the density on S_n being given at each step i by some Q_i. Then the resulting density on S_n after these k steps will be given by $Q_1 * Q_2 * \cdots * Q_k$. Equivalently, doing the shuffle specified by Q_1, and then the shuffle specified by Q_2, and so on, up through the shuffle given by Q_k, is the same as doing the single shuffle specified by $Q_1 * Q_2 * \cdots * Q_k$. In short, repeated shuffling corresponds to convoluting densities. This method of convolutions is complicated, however, and we will see later that for a realistic type of shuffle, there is a much easier way to compute the probability of any particular permutation after any particular number of shuffles.

3 THE RIFFLE SHUFFLE

We would now like to choose a realistic model of how actual cards are physically shuffled by people. A particular one with nice mathematical properties is given by the "riffle shuffle." (Sometimes called the GSR shuffle, it was developed by Gilbert and Shannon, and independently by Reeds.) It goes as follows. First cut the deck into two packets, the first containing k cards, and the other the remaining $n - k$ cards. Choose k, the number of cards cut, according to the binomial density, meaning the probability of the cut occuring exactly after k cards is given by $\begin{pmatrix} n \\ k \end{pmatrix} / 2^n$.

Once the deck has been cut into two packets, interleave the cards from each packet in any possible way, such that the cards of each packet maintain their own relative order. This means that the cards originally in positions $1, 2, 3, \ldots k$ must still be in the same order in the deck after it is shuffled, even if there are other cards in-between; the same goes for the cards originally in positions $k+1, k+2, \ldots n$. This requirement is quite natural when you think of how a person shuffles two packets of cards, one in each hand. The cards in the left hand must still be in the same relative order in the shuffled deck, no matter how they are interleaved with the cards from the other packet, because the cards in the left hand are dropped in order when shuffling; the same goes for the cards in the right hand.

Choose among all such interleavings uniformly, meaning each is equally likely. Since there are $\left(\begin{array}{c} n \\ k \end{array} \right)$ possible interleavings (as we only need choose k spots among n places for the first packet, the spots for the cards of the other packet then being determined), this means any particular interleaving has probability $1/\left(\begin{array}{c} n \\ k \end{array} \right)$ of occuring. Hence the probability of any particular cut followed by a particular interleaving, with k the size of the cut, is $\left(\begin{array}{c} n \\ k \end{array} \right)/2^n \cdot 1/\left(\begin{array}{c} n \\ k \end{array} \right) = 1/2^n$. Note that this probability $1/2^n$ contains no information about the cut or the interleaving! In other words, the density of cuts and interleavings is uniform — every pair of a cut and a possible resulting interleaving has the same probability.

This uniform density on the set of cuts and interleavings now induces in a natural way a density on the set of permutations, i.e. a shuffle, according to our definition. We will call this the riffle shuffle and denote it by R. It is defined for π in S_n by $R(\pi) =$ the sum of the probabilities of each cut and interleaving that gives the rearrangement of the deck corresponding to π, which is $1/2^n$ times the number of ways of cutting and interleaving that give the rearrangement of the deck corresponding to π. In short, the chance of any arrangement of cards occuring under riffle shuffling is simply the proportion of ways of riffling which give that arrangement.

Here is a particular example of the riffle shuffle in the case $n = 3$, with the deck starting in natural order 123.

k = cut position	cut deck	probability of this cut	possible interleavings
0	|123	1/8	123
1	1|23	3/8	123,213,231
2	12|3	3/8	123,132,312
3	123|	1/8	123

Note that 0 or all 3 cards may be cut, in which case one packet is empty and the other is the whole deck. Now let us compute the probability of each particular ordering occurring in the above example. First, look for 213. It occurs only in the cut k=1, which has probability 3/8. There it is one of three possibilities, and hence has the conditional probability 1/3, given $k = 1$. So the overall probability for 213 is $\frac{1}{3} \cdot \frac{3}{8} = \frac{1}{8}$, where of course $\frac{1}{8} = \frac{1}{2^3}$ is the probability of any particular cut and interleaving pair. Similar analyses hold for 312, 132, and 231, since they all occur only through a single cut and interleaving. For 123, it is different; there are four cuts and interleavings which give rise to it. It occurs for $k = 0, 1$,2, and 3, these situations having probabilities 1/8, 3/8, 3/8, and 1/8, respectively. In these cases, the conditional probability of 123, given the cut, is 1, 1/3, 1/3, and 1. So the overall probability of the ordering is $\frac{1}{8} \cdot 1 + \frac{3}{8} \cdot \frac{1}{3} + \frac{3}{8} \cdot \frac{1}{3} + \frac{1}{8} \cdot 1 = \frac{1}{2}$, which also equals $4 \cdot \frac{1}{2^3}$, the number of ways of cutting and interleaving that give rise to the ordering times the probability of any particular cut and interleaving. We may write down the entire density, now dropping the assumption that the deck started in the natural order, which means we must use permutations instead of orderings.

permutation π	[123]	[213]	[231]	[132]	[312]	[321]
probability $R(\pi)$ under riffle	1/2	1/8	1/8	1/8	1/8	0

It is worth making obvious a point which should be apparent. The information specified by a cut and an interleaving is richer than the information specified by the resulting permutation. In other words, there may be several different ways of cutting and interleaving that give rise to the same permutation, but different permutations necessarily arise from distinct cut/interleaving pairs. (An exercise for the reader is to show that for the riffle shuffle, this distinction is nontrivial only when the permutation is the identity, i.e. the only time distinct cut/interleaving pairs give rise to the same permutation is when the permutation is the identity.)

There is a second, equivalent way of describing the riffle shuffle. Start the same way, by cutting the deck according to the binomial density into two packets of size k and $n-k$. Now we are going to drop a card from the bottom of one of the two packets onto a table, face down. Choose between the packets with probability proportional to packet size, meaning if the two packets are of size p_1 and p_2, then the probability of the card dropping from the first is $\frac{p_1}{p_1+p_2}$, and $\frac{p_2}{p_1+p_2}$ from the second. So this first time, the probabilities would be $\frac{k}{n}$ and $\frac{n-k}{n}$. Now repeat the process, with the numbers p_1 and p_2 being updated to reflect the actual packet sizes by subtracting one from the size of whichever packet had the card dropped last time. For instance, if the first card was dropped from the first packet, then the probabilities for the next drop would be $\frac{k-1}{n-1}$ and $\frac{n-k}{n-1}$. Keep going until all cards are dropped. This method is equivalent to the first description of the riffle in that this process also assigns uniform probability $1/\left(\begin{array}{c} n \\ k \end{array}\right)$ to each possible resulting interleaving of the cards.

To see this, let us figure out the probability for some particular way of dropping the cards, say, for the sake of definiteness, from the first packet and then from the first, second, second, second, first, and so on. The probability of the drops occuring this way is

$$\frac{k}{n} \cdot \frac{k-1}{n-1} \cdot \frac{n-k}{n-2} \cdot \frac{n-k-1}{n-3} \cdot \frac{n-k-2}{n-4} \cdot \frac{k-2}{n-5} \cdots,$$

where we have multiplied probabilities since each drop decision is independent of the others once the packet sizes have been readjusted. Now the product of the denominators of these fractions is $n!$, since it is just the product of the total number of cards left in both packets before each drop, and this number decreases by one each time. What is the product of the numerators? Well, we get one factor every time a card is dropped from one of the packets, this factor being the size of the packet at that time. But then we get all the numbers $k, k-1, \ldots, 1$ and $n-k, n-k-1, \ldots, 1$ as factors in some order, since each packet passes through all of the sizes in its respective list as the cards are dropped from the two packets. So the numerator is $k!(n-k)!$, which makes the overall probability $k!(n-k)!/n! = 1/\left(\begin{array}{c} n \\ k \end{array}\right)$, which is obviously valid for any particular sequence of drops, and not just the above example. So we have now shown the two descriptions of the riffle shuffle are equivalent, as they have the same uniform probability of interleaving after a binomial cut.

Now let $R^{(k)}$ stand for convoluting R with itself k times. This corresponds to the density after k riffle shuffles. For which k does $R^{(k)}$ produce a randomized deck? The next section begins to answer this question.

4 HOW FAR AWAY FROM RANDOMNESS?

Before we consider the question of how many times we need to shuffle, we must decide what we want to achieve by shuffling. The answer should be randomness of some sort. What does randomness mean? Simply put, any arrangement of cards is equaly likely; no one ordering should be favored over another. This means the uniform density U on S_n, each permutation having probability $U(\pi) = 1/|S_n| = 1/n!$.

Now it turns out that for any fixed number of shuffles, no matter how large, riffle shuffling does not produce complete randomness in this sense. (We will, in fact, give an explicit formula which shows that after any number of riffle shuffles, the identity permutation is always more likely than any other to occur.) So when we ask how many times we need to shuffle, we are not asking how far to go in order to achieve randomness, but rather to get close to randomness. So we must define what we mean by close, or far, i.e. we need a distance between densities.

The concept we will use is called variation distance (which is essentially the L^1 metric on the space of densities). Suppose we are given two probability densities, Q_1 and Q_2, on S_n. Then the variation distance between Q_1 and Q_2 is defined to be

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

The $\frac{1}{2}$ normalizes the result to always be between 0 and 1.

Here is an example. Let $Q_1 = R$ be the density calculated above for the three card riffle shuffle. Let Q_2 be the complete reversal — the density that gives probability 1 for [321], i.e. certainty, and 0 for all other permutations, i.e. nonoccurence.

π	$Q_1(\pi)$	$Q_2(\pi)$	$\lvert Q_1(\pi) - Q_2(\pi)\rvert$
[123]	1/2	0	1/2
[213]	1/8	0	1/8
[312]	1/8	0	1/8
[132]	1/8	0	1/8
[231]	1/8	0	1/8
[321]	0	1	1
		Total	2

So here $\|Q_1 - Q_2\| = 2/2 = 1$, and the densities are as far apart as possible.

Now the question we really want to ask is: how big must we take k to make the variation distance $||R^{(k)} - U||$ between the riffle and uniform small? This can be best answered by a graph of $||R^{(k)} - U||$ versus k. The following theory is directed towards constructing this graph.

5 RISING SEQUENCES

To begin to determine what the density $R^{(k)}$ is, we need to consider a fundamental concept, that of a rising sequence. A rising sequence of a permutation is a maximal consecutively increasing subsequence. What does this really mean for cards? Well, perform the rearrangement corresponding to the permutation on a naturally ordered deck. Pick any card, labeled x say, and look after it in the deck for the card labeled $x + 1$. If you find it, repeat the procedure, now looking after the $x + 1$ card for the $x + 2$ card. Keep going in this manner until you have to stop because you can't find the next card after a given card. Now go back to your original card x and reverse the procedure, looking before the original card for the $x - 1$ card, and so on. When you are done, you have a rising sequence. It turns out that a deck breaks down as a disjoint union of its rising sequences, since the union of any two consecutively increasing subsequences containing a given element is also a consecutively increasing subsequence that contains that element.

Let's look at an example. Suppose we know that the order of an eight card deck after shuffling the natural order is 45162378. Start with any card, say 3. We look for the next card in value after it, 4, and do not find it. So we stop looking after and look before the 3. We find 2, and then we look for 1 before 2 and find it. So one of the rising sequences is given by 123. Now start again with 6. We find 7 and then

8 after it, and 5 and then 4 before it. So another rising sequence is 45678. We have accounted for all the cards, and are therefore done. Thus this deck has only two rising sequences. This is immediately clear if we write the order of the deck this way, $45_{1}6_{23}78$, offsetting the two rising sequences.

It is clear that a trained eye may pick out rising sequences immediately, and this forms the basis for some card tricks. Suppose a brand new deck of cards is riffle shuffled three times by a spectator, who then takes the top card, looks at it without showing it to a magician, and places it back in the deck at random. The magician then tries to identify the reinserted card. He is often able to do so because the reinserted card will often form a singleton rising sequence, consisting of just itself. Most likely, all the other cards will fall into $2^3 = 8$ rising sequences of length 6 to 7, since repeated riffle shuffling, at least the first few times, roughly tends to double the number of the rising sequences and halve the length of each one each time. Diaconis, himself a magician, and Bayer [2] describe variants of this trick that magicians have actually used.

It is interesting to note that the order of the deck in our example, $45_{1}6_{23}78$, is a possible result of a riffle shuffle with a cut after 3 cards. In fact, any ordering with just two rising sequences is a possible result of a riffle shuffle. Here the cut must divide the deck into two packets such that the length of each is the same as the length of the corresponding rising sequence. So if we started in the natural order 12345678 and cut the deck into 123 and 45678, we would interleave by taking 4, then 5, then 1, then 6, then 2, then 3, then 7, then 8, thus obtaining the given order through riffling. The converse of this result is that the riffle shuffle always gives decks with either one or two rising sequences.

6 BIGGER AND BETTER: a-SHUFFLES

The result that a permutation has nonzero probability under the riffle shuffle if and only if it has exactly one or two rising sequences is true, but it only holds for a single riffle shuffle. We would like similar results on what happens after multiple riffle suffles. This can ingeniously be accomplished by considering a-shuffles, a generalization of the riffle shuffle. An a-shuffle is another probability density on S_n, achieved as follows. Let a stand for any positive integer. Cut the deck into a packets, of nonnegative sizes $p_1, p_2, \ldots, p_a$, with the probability of this particular packet structure given by the multinomial density: $\binom{n}{p_1, p_2, \ldots, p_a}/a^n$. Note we must have $p_1 + \cdots + p_a = n$, but some of the p_i may be zero. Now interleave the cards from each packet in any way, so long as the cards from each packet maintain their relative order among themselves. With a fixed packet structure, consider all interleavings equally likely. Let us count the number of such interleavings. We simply want the number of different ways of choosing, among n positions in the deck, p_1 places for things of one type, p_2 places for things of another type, etc. This is given by the multinomial coefficient $\binom{n}{p_1, p_2, \ldots, p_a}$. Hence the probability of a particular rearrangement, i.e. a cut of the deck and an interleaving, is

$$\binom{n}{p_1, p_2, \ldots, p_a}/a^n \cdot \binom{n}{p_1, p_2, \ldots, p_a} = \frac{1}{a^n}.$$

So it turns out that each combination of a particular cut into a packets and a particular interleaving is equally likely, just as in the riffle shuffle. The induced density on the permutations corresponding to the cuts and interleavings is then called the a-shuffle. We will denote it by R_a. It is apparent that the riffle is just the 2-shuffle, so $R = R_2$.

An equivalent description of the a-shuffle begins the same way, by cutting the deck into packets multinomially. But then drop cards from the bottom of the packets, one at a time, such that the probability of choosing a particular packet to drop from is proportional to the relative size of that packet compared to the number of cards left in all the packets. The proof that this description is indeed equivalent is exactly analogous to the $a = 2$ case. A third equivalent description is given by cutting multinomially into $p_1, p_2, \ldots, p_a$ and riffling p_1 and p_2 together (meaning choose uniformly among all interleavings which maintain the relative order of each packet), then riffling the resulting pile with p_3, then riffling that resulting pile with p_4, and so on.

There is a useful code that we can construct to specify how a particular a-shuffle is done. (Note that we are abusing terminology slightly and using shuffle here to indicate a particular way of rearranging the deck, and not the density on all such rearrangements.) This is done through n digit base a numbers. Let A be any one of these n digit numbers. Count the number of 0's in A. This will be the size of the first packet in the a-shuffle, p_1. Then p_2 is the number of 1's in A, and so on, up through $p_a =$ the number of $(a-1)$'s. This cuts the deck cut into a packets. Now take the beginning packet of cards, of size p_1.

Envision placing these cards on top of all the 0 digits of A, maintaining their relative order as a rising sequence. Do the same for the next packet, p_2, except placing them on the 1's. Again, continue up through the $(a-1)$'s. This particular way of rearranging the cards will then be the particular cut and interleaving corresponding to A.

Here is an example, with the deck starting in natural order. Let $A = 23004103$ be the code for a particular 5-shuffle of the 8 card deck. There are three 0's, one 1, one 2, two 3's, and one 4. Thus $p_1 = 3$, $p_2 = 1$, $p_3 = 1$, $p_4 = 2$, and $p_5 = 1$. So the deck is cut into 123 | 4 | 5 | 67 | 8. So we place 123 where the 0's are in A, 4 where the 1 is, 5 where the 2 is, 67 where the 3's are, and 8 where the 4 is. We then get a shuffled deck of 56128437 when A is applied to the natural order.

Reflection shows that this code gives a bijective correspondence between n digit base a numbers and the set of all ways of cutting and interleaving an n card deck according to the a-shuffle. In fact, if we put the uniform density on the set of n digit base a numbers, this transfers to the correct uniform probability for cutting and interleaving in an a-shuffle, which means the correct density is induced on S_n, i.e. we get the right probabilities for an a-shuffle. This code will prove useful later on.

7 VIRTUES OF THE a-SHUFFLE

7.1 Relation to rising sequences

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

The probablity of achieving a permutation π when doing an a-shuffle is given by $\binom{n+a-r}{n}/a^n$, where r is the number of rising sequences in π.

Proof: First note that if we establish and fix where the $a-1$ cuts occur in an a-shuffle, then whatever permutations can actually be achieved by interleaving the cards from this cut/packet structure are achieved in exactly one way; namely, just drop the cards in exactly the order of the permutation. Thus the probability of achieving a particular permutation is the number of possible ways of making cuts that could actually give rise to that permutation, divided by the total number of ways of making cuts and interleaving for an a-shuffle.

So let us count the ways of making cuts in the naturally ordered deck that could give the ordering that results when π is applied. If we have r rising sequences in π, we know exactly where $r-1$ of the cuts have to have been; they must have occurred between pairs of consecutive cards in the naturally ordered deck such that the first card ends one rising sequence of π and the second begins another rising sequence of π. This means we have $a-1-(r-1)=a-r$ unspecified, or free, cuts. These are free in the sense that they can in fact go anywhere. So we must count the number of ways of putting $a-r$ cuts among n cards. This can easily be done by considering a sequence of $(a-r)+n$ blank spots which must be filled by $(a-r)$ things of one type (cuts) and n things of another type (cards). There are $\left(\begin{array}{c} (a-r)+n \\ n \end{array}\right)$ ways to do this, i.e. choosing n places among $(a-r)+n$.

This is the numerator for our probability expressed as a fraction; the denominator is the number of possible ways to cut and interleave for an a-shuffle. By considering the encoding of shuffles we see there are a^n ways to do this, as there are this many n digit base a numbers. Hence our result is true.

This allows us to envision the probability density associated with an a-shufle in a nice way. Order all the permutation in S_n in any way such that the number of rising sequences is non-decreasing. If we label these permutations as points on a horizontal axis, we may take the vertical axis to be the numbers between 0 and 1, and at each permutation place a point whose vertical coordinate is the probability of the permutation. Obviously, the above result means we will have sets of points of the same height. Here is an example for a 7-shuffle of the five card deck (solid line), along with the uniform density $U \equiv 1/5! = 1/120$ (dashed line).

Notice the probability $\left(\begin{array}{c} n+a-r \\ n \end{array}\right)/a^n$ is a monotone decreasing function of r. This means if $1 \leq r_1 < r_2 \leq n$, then a particular permutations with r_1 rising sequences is always more likely than a permutation with r_2 rising sequences under any a-shuffle. Hence the graph of the density for an a-shuffle, if the permutations are ordered as above, will always be nonincreasing. In particular, the probability starts above uniform for the identity, the only permutation with $r = 1$. (In our example $R_7(\text{identity}) = \binom{5+7-1}{5}/7^5 = .0275$.) It then decreases for increasing r, at some point crossing below uniform (from $r = 2$ to 3 in the example). The greatest r value such that the probability is above uniform is called the *crossover point*. Eventually at $r = n$, which occurs only for the permutation corresponding to complete reversal of the deck, the probability is at its lowest value. (In the example $\binom{5+7-5}{5}/7^5 = .0012$.) All this explains the earlier statement that after an a-shuffle, the identity is always more likely than it would be under a truly random density, and is always more likely than any other particular permutation after the same a-shuffle.

For a fixed deck size n, it is interesting to note the behavior of the crossover point as a increases. By analyzing the inequality

$$\binom{n+a-r}{n}/a^n \geq \frac{1}{n!},$$

the reader may prove that the crossover point never moves to the left, i.e. it is a nondecreasing function of a, and that it eventually moves to the right, up to $n/2$ for n even and $(n-1)/2$ for n odd, but never beyond. Furthermore, it will reach this halfway point for a approximately the size of $n^2/12$. Combining with the results of the next section, this means roughly $2\log_2 n$ riffle shuffles are needed to bring the crossover point to halfway.

7.2 The multiplication theorem

Why bother with an a-shuffle? In spite of the nice formula for a density dependent only on the number of rising sequences, a-shuffles seem of little practical use to any creature that is not a-handed. This turns out to be false. After we establish another major result that addresses this question, we will be in business to construct our variation distance graph.

This result concerns multiple shuffles. Suppose you do a riffle shuffle twice. Is there any simple way to describe what happens, all in one step, other than the convolution of densities described in section 2.2?

Or more generally, if you do an a-shuffle and then do a b-shuffle, how can you describe the result? The answer is the following:

An a-shuffle followed by a b-shuffle is equivalent to a single ab-shuffle, in the sense that both processes give exactly the same resulting probability density on the set of permutations.

Proof: Let us use the previously described code for shuffles. Suppose that A is an n digit base a number, and B is an n digit base b number. Then first doing the cut and interleaving encoded by A and then doing the cut and interleaving encoded by B gives the same permutation as the one resulting from the cut and interleaving encoded by the n digit base ab number given by $A^B \& B$, as John Finn figured out. (The proof for this formula will be deferred until section 9.4, where the inverse shuffle is discussed.) This formula needs some explanation. A^B is defined to be the code that has the same base a digits as A, but rearranged according to the permutation specified by B. The symbol & in $A^B \& B$ stands for digit-wise concatenation of two numbers, meaning treat the base a digit A_i^B in the ith place of A^B together with the base b digit B_i in the ith place of B as the base ab digit given by $A_i^B \cdot b + B_i$. In other words, treat the combination $A_i^B \& B_i$ as a two digit number, the right-most place having value 1, and the left-most place having value b, and then treat the result as a one digit base ab number.

Why this formula holds is better shown by an example than by general formulas. Suppose $A = 012210$ is the code for a particular 3-shuffle, and $B = 310100$ is the code for a particular 4-shuffle. (Again we are abusing terminology slightly.) Let π_A and π_B be the respective permutations. Then in the tables below note that $\pi_A \circ \pi_B$, the result of a particular 3-shuffle followed by a particular 4-shuffle, and $\pi_{A^B \& B}$, the result of a particular 12-shuffle, are the same permutation.

i	1	2	3	4	5	6
$\pi_A(i)$	1	3	5	6	4	2
$\pi_B(i)$	6	4	1	5	2	3
$\pi_A \circ \pi_B(i)$	2	6	1	4	3	5

A	0	1	2	2	1	0
B	3	1	0	1	0	0
A^B	0	2	0	1	1	2
B	3	1	0	1	0	0
$A^B \& B$	3	9	0	5	4	8

i	1	2	3	4	5	6
$\pi_{A^B \& B}(i)$	2	6	1	4	3	5

We now have a formula $A^B\&B$ that is really a one-to-one correspondence between the set of pairs, consisting of one n digit base a number and one n digit base b number, and the set of n digit base ab numbers; further this formula has the property that the cut and interleaving specified by A, followed by the cut and interleaving specified by B, result in the same permutation of the deck as that resulting from the cut and interleaving specified by $A^B\&B$. Since the probability densities for a, b, and ab-shuffles are induced by the uniform densities on the sets of n digit base a, b, or ab codes, respectively, the properties of the one-to-one correspondence imply the induced densities on S_n of an a-shuffle followed by a b-shuffle and an ab-shuffle are the same. Hence our result is true.

7.3 Expected happenings after an a-shuffle

It is of theoretical interest to measure the expected value of various quantities after an a-shuffle of the deck. For instance, we may ask what is the expected number of rising sequences after an a-shuffle? I've found an approach to this question which has too much computation to be presented here, but gives the answer as

$$a - \frac{n+1}{a^n} \sum_{r=0}^{a-1} r^n.$$

As $a \to \infty$, this expression tends to $\frac{n+1}{2}$, which is the expected number of rising sequences for a random permutation. When $n \to \infty$, the expression goes to a. This makes sense, since when the number of packets is much less than the size of the deck, the expected number of rising sequences is the same as the number of packets.

The expected number of fixed points of a permutation after an a-shuffle is given by $\sum_{i=0}^{n-1} a^{-i}$, as mentioned in [2]. As $n \to \infty$, this expression tends to $\frac{1}{1-1/a} = \frac{a}{a-1}$, which is between 1 and 2. As $a \to \infty$, the expected number of fixed points goes to 1, which is the expected number of fixed points for a random permutation.

8 PUTTING IT ALL TOGETHER

Let us now combine our two major results of the last section to get a formula for $R^{(k)}$, the probability density for the riffle shuffle done k times. This is just k 2-shuffles, one after another. So by the multiplication theorem, this is equivalent to a single $2 \cdot 2 \cdot 2 \cdots 2 = 2^k$-shuffle.

Hence in the $R^{(k)}$ density, there is a $\left(\begin{array}{c} 2^k + n - r \\ n \end{array} \right) / 2^{nk}$ chance of a permutation with r rising sequences occurring, by our rising sequence formula. This now allows us to work on the variation distance $\|R^k - U\|$. For a permutation π with r rising sequences, we see that

$$|R^k(\pi) - U(\pi)| = \left| \left(\begin{array}{c} 2^k + n - r \\ n \end{array} \right) / 2^{nk} - \frac{1}{n!} \right|.$$

We must now add up all the terms like this, one for each permutation. We can group terms in our sum according to the number of rising sequences. If we let $A_{n,r}$ stand for the number of permutations of n cards that have r rising sequences, each of which have the same probabilities, then the variation distance is given by

$$\|R^k - U\| = \frac{1}{2} \sum_{r=1}^{n} A_{n,r} \left| \left(\begin{array}{c} 2^k + n - r \\ n \end{array} \right) / 2^{nk} - \frac{1}{n!} \right|.$$

The only thing unexplained is how to calculate the $A_{n,r}$. These are called the Eulerian numbers, and various formulas are given for them (e.g. see [8]). One recursive one is $A_{n,1} = 1$ and $A_{n,r} = r^n - \sum_{j=1}^{r-1} \left(\begin{array}{c} n + r - j \\ n \end{array} \right) A_{n,j}$. (It is interesting to note that the Eulerian numbers are symmetric in the sense that $A_{n,r} = A_{n,n-r+1}$. So there are just as many permutations with r rising sequences as there are with $n - r + 1$ rising sequences, which the reader is invited to prove directly.)

Now the expression for variation distance may seem formidable, and it is. But it is easy and quick for a computer program to calculate and graph $\|R^k - U\|$ versus k for any specific, moderately sized n. Even on the computer, however, this computation is tractable because we only have n terms, corresponding to each possible number of rising sequences. If we did not have the result on the invariance of the probability when the number of rising sequences is constant, we would have $|S_n| = n!$ terms in the sum. For $n = 52$, this is approximately 10^{68}, which is much larger than any computer could handle. Here is the graphical result of a short Mathematica program that does the calculations for $n = 52$. The horizontal axis is the number of riffle shuffles, and the vertical axis is the variation distance to uniform.

The answer is finally at hand. It is clear that the graph makes a sharp cutoff at $k = 5$, and gets reasonably close to 0 by $k = 11$. A good middle point for the cutoff seems to $k = 7$, and this is why seven shuffles are said to be enough for the usual deck of 52 cards.

Additionally, asymptotic analysis in [2] shows that when n, the number of cards, is large, approximately $k = \frac{3}{2}\log n$ shuffles suffice to get the variation distance through the cutoff and close to 0.

We have now achieved our goal of constructing the variation distance graph, which explains why seven shuffles are "enough". In the remaining sections we present some other aspects to shuffling, as well as some other ways of approaching the question of how many shuffles should be done to deck.

9 THE INVERSE SHUFFLE

There is an unshuffling procedure which is in some sense the reverse of the riffle shuffle. It is actually simpler to describe, and some of the theorems are more evident in the reverse direction. Take a face-down deck, and deal cards from the bottom of the deck one at a time, placing the cards face-down into one of two piles. Make all the choices of which pile independently and uniformly, i.e. go 50/50 each way each time. Then simply put one pile on top of the other. This may be called the riffle unshuffle, and the induced density on S_n may be labeled $\hat{R}$. An equivalent process is generated by labeling the backs of all the cards with 0's and 1's independently and uniformly, and then pulling all the 0's to the front of the deck, maintaining their relative order, and pulling all the 1's the back of the deck, maintaining their relative order. This may quickly be generalized to an a-unshuffle, which is described by labeling the back of each card independently with a base a digit chosen uniformly. Now place all the cards labeled 0 at the front of the deck, maintaining their relative order, then all the 1's, and so on, up through the $(a-1)$'s. This is the a-unshuffle, denoted by $\hat{R}_a$.

We really have a reverse or inverse operation in the sense that $\hat{R}_a(\pi) = R_a(\pi^{-1})$ holds. This is seen most easily by looking at n digit base a numbers. We have already seen in section 6 that each such n digit base a number may be treated as a code for a particular cut and interleaving in an a-shuffle; the above paragraph in effect gives a way of also treating each n digit base a numbers as code for a particular way of achieving an a-unshuffle. The two induced permutations we get when looking at a given n digit base a number in these two ways are inverse to one another, and this proves $\hat{R}_a(\pi) = R_a(\pi^{-1})$ since the uniform density on n digit base a numbers induces the right density on S_n.

We give a particular example which makes the general case clear. Take the 9 digit base 3 code 122020110 and apply it in the forward direction, i.e. treat it as directions for a particular 3-shuffle of the deck 123456789 in natural order. We get the cut structure 123|456|789 and hence the shuffled deck 478192563. Now apply the code to this deck order, but backwards, i.e. treat it as directions for a 3-unshuffle of 478192563. We get the cards where the 0's are, 123, pulled forward; then the 1's, 456; and then the 2's, 789, to get back to the naturally ordered deck 123456789. It is clear from this example that, in general, the a-unshuffle directions for a given n digit base a number pull back the cards in a way exactly opposite to the way the a-shuffle directions from that code distributed them. This may be checked by applying the code both forwards and backwards to the unshuffled deck 123456789 and getting

$$\begin{pmatrix} 123456789 \\ 478192563 \end{pmatrix} \qquad \begin{pmatrix} 123456789 \\ 469178235 \end{pmatrix},$$

which inspection shows are indeed inverse to one another.

The advantage to using unshuffles is that they motivate the $A^B\&B$ formula in the proof of the multiplication theorem for an a-shuffle followed by a b-shuffle. Suppose you do a 2-unshuffle by labeling the cards with 0's and 1's in the upper right corner according to a uniform and independent random choice each time, and then sorting the 0's before the 1's. Then do a second 2-unshuffle by labeling the cards again with 0's and 1's, placed just to the left of the digit already on each card, and sorting these left-most 0's before the left-most 1's. Reflection shows that doing these two processes is equivalent to doing a single process: label each card with a 00, 01, 10, or 11 according to uniform and independent choices, sort all cards labeled 00 and 10 before all those labeled 01 and 11, and then sort all cards labeled 00 and 01 before all those labeled 10 and 11. In other words, sort according to the right-most digit, and then according to the left-most digit. But this is the same as sorting the 00's before the 01's, the 01's before the 10's, and the 10's before the 11's all at once. So this single process is equivalent to the following: label each card with a 0, 1, 2, or 3 according to uniform and independent choices, and sort the 0's before the 1's before the 2's before the 3's. But this is exactly a 4-unshuffle!

So two 2-unshuffles are equivalent to a $2 \cdot 2 = 4$-unshuffle, and generalizing in the obvious way, a b-unshuffle followed by an a-unshuffle is equivalent to an ab-unshuffle. (In the case of unshuffles we have orders reversed and write a b-unshuffle followed by an a-unshuffle, rather than vice-versa, for the same reason that one puts on socks and then shoes, but takes off shoes and then socks.) Since the density for unshuffles is the inverse of the density for shuffles (in the sense that $\hat{R}_a(\pi) = R_a(\pi^{-1})$), this means an a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle. Furthermore, we are tempted to believe that combining the codes for unshuffles should be given by $A\&B$, where A and B are the sequences of 0's and 1's put on the cards, encapsulated as n digit base 2 numbers, and & is the already described symbol for digitary concatenation. This $A\&B$ is not quite right, however; for when two 2-unshuffles are done, the second group of 0's and 1's will not be put on the cards in their original order, but will be put on the cards in the order they are in after the first unshuffle. Thus we must compensate in the formula if we wish to treat the 00's, 01's, 10's, and 11's as being written down on the cards in their original order at the beginning, before any unshuffling. We can do this by by having the second sequence of 0's and 1's permuted, according to the inverse of the permutation described by the first sequence of 0's and 1's. So we must use A^B instead of A. Clearly this works for all a and b and not just $a = b = 2$. This is why the formula for combined unshuffles, and hence shuffles, is $A^B\&B$ and not just $A\&B$. (The fact that it is actually $A^B\&B$ and not $A\&B^A$ or some such variant is best checked by looking at particular examples, as in section 7.2.)

10 ANOTHER APPROACH TO SUFFICIENT SHUFFLING

10.1 Seven is not enough

A footnote must be added to the choosing of any specific number, such as seven, as making the variation distance small enough. There are examples where this does not randomize the deck enough. Peter Doyle has invented a game of solitaire that shows this quite nicely. A simplified, albeit less colorful version is given here. Take a deck of 52 cards, turned face-down, that is labeled in top to bottom order $123\cdots(25)(26)(52)(51)\cdots(28)(27)$. Riffle shuffle seven times. Then deal the cards one at a time from the top of the deck. If the 1 comes up, place it face up on the table. Call this pile A. If the 27 comes up, place it face up on the table in a separate pile, calling this B. If any other card comes up that it is not the immediate successor of the top card in either A or B, then place it face up in the pile C. If the immediate successor of the top card of A comes up, place it face up on top of A, and the same for B. Go through the whole deck this way. When done, pick up pile C, turn it face down, and repeat the procedure. Keep doing so. End the game when either pile A or pile B is full, i.e. has twenty-six cards in it. Let us say the game has been won if A is filled up first, and lost if B is.

It turns out that the game will end much more than half the time with pile A being full, i.e the deck is not randomized 'enough.' Computer simulations indicate that we win about 81% of the time. Heuristically, this is because the rising sequences in the permuted deck after a $2^7 = 128$-shuffle can be expected to come from both the first and second halves of the original deck in roughly the same numbers and length. However, the rising sequences from the first half will be 'forward' in order and the ones from the second half will be 'backward.' The forward ones require only one pass through the deck to be placed in pile A, but the backward ones require as many passes through the deck as their length, since only the last card can be picked off and put into pile B each time. Thus pile A should be filled more quickly; what really makes this go is that a 128-shuffle still has some rising sequences of length 2 or longer, and it is faster to get these longer rising sequences into A than it is into to get sequences of the same length into B.

In a sense, this game is almost a worst case scenario. This is because of the following definition of variation distance, which is equivalent to the one given in section 4. (The reader is invited to prove this.) Given two densities Q_1 and Q_2 on S_n,

$$\|Q_1 - Q_2\| = \max_{S \subset S_n} |Q_1(S) - Q_2(S)|,$$

where the maximum on the r.h.s. is taken over all subsets S of S_n, and the $Q_i(S)$ are defined to be $\sum_{\pi \in S} Q_i(\pi)$. What this really means is that the variation distance is an upper bound (in fact a least upper bound) for the difference of the probabilities of an event given by the two densities. This can be directly applied to our game. Let S be the set of all permutations for which pile A is filled up first, i.e. the event that we win. Then the variation distance $\|R^{(7)} - U\|$ is an upper bound for the difference between the probability of a permutation in S occuring after 7 riffles, and the probability of such a permutation occuring truly randomly. Now such winning permutations should occur truly randomly only half the time (by symmetry), but the simulations indicate that they occur 81% percent of the time after 7 riffle shuffles. So the probability difference is $|.81 - .50| = .31$. On the other hand, the variation distance $\|R^{(7)} - U\|$ as calculated in section 8 is .334, which is indeed greater than .31, but not by much. So Doyle's game of solitaire is nearly as far away from being a fair game as possible.

10.2 Is variation distance the right thing to use?

The variation distance has been chosen to be the measure of how far apart two densities are. It seems intutively reasonable as a measure of distance, just taking the differences of the probabilities for each permutation, and adding them all up. But the game of the last section might indicate that it is too forgiving a measure, rating a shuffling method as nearly randomizing, even though in some ways it clearly is not. At the other end of the spectrum, however, some examples, as modified from [1] and [4], suggest that variation distance may be too harsh a measure of distance. Suppose that you are presented with a face-down deck, with n even, and told that it has been perfectly randomized, so that as far as you know, any ordering is equally as likely as any other. So you simply have the uniform density $U(\pi) = 1/n!$ for all $\pi \in S_n$. But now suppose that the top card falls off, and you see what it is. You realize that to put the card back on the top of the deck would destroy the complete randomization by restricting the possible permutations, namely to those that have this paricular card at the first position. So you decide to place the card back at random in the deck. Doing this would have restored complete randomization and hence the uniform density. Suppose, however, that you realize this, but also figure superstitiously that you shouldn't move this original top card too far from the top. So instead of placing it back in the deck at random, you place it back at random subject to being in the *top half* of the deck.

How much does this fudging of randomization cost you in terms of variation distance? Well, the number of restricted possible orderings of the deck, each equally likely, is exactly half the possible total, since we want those orderings where a given card is in the first half, and not those where it is in the second half. So this density is given by $\bar{U}$, which is $2/n!$ for half of the permutations and 0 for the other half. So the variation distance is

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

This seems a high value, given the range between 0 and 1. Should a good notion of distance place this density $\bar{U}$, which most everyone would agree is very nearly random, half as far away from complete randomness as possible?

10.3 The birthday bound

Because of some of the counterintuitive aspects of the variation distance presented in the last two subsections, we present another idea of how to measure how far away repeated riffle shuffling is from randomness. It turns out that this idea will give an upper bound on the variation distance, and it is tied up with the well-known birthday problem as well.

We begin by first looking at a simpler case, that of the top-in shuffle, where the top card is taken off and reinserted randomly anywhere into the deck, choosing among each of the n possible places between cards uniformly. Before any top-in shuffling is done, place a tag on the bottom card of the deck, so that it can be identified. Now start top-in shuffling repeatedly. What happens to the tagged card? Well, the first time a card, say a, is inserted below the tagged card, and hence on the bottom of the deck, the tagged card will move up to the penultimate position in the deck. The next time a card, say b, is inserted below the tagged card, the tagged card will move up to the antepenultimate position. Note that all possible orderings of a and b below the tagged card are equally likely, since it was equally likely that b went above or below a, given only that it went below the tagged card. The next time a card, say c, is put below the tagged card, its equal likeliness of being put anywhere among the order of a and b already there, which comes from a uniform choice among all orderings of a and b, means that all orders of a, b, and c are equally likely. Clearly as this process continues the tagged card either stays in its position in the deck, or it moves up one position; and when this happens, all orderings of the cards below the tagged card are equally likely. Eventually the tagged card gets moved up to the top of the deck by having another card inserted underneath it. Say this happens on the $T'-1$st top-in shuffle. All the cards below the tagged card, i.e. all the cards but the tagged card, are now randomized, in the sense that any order of them is equally likely. Now take the tag off the top card and top-in shuffle for the T'th time. The deck is now completely randomized, since the formerly tagged card has been reinserted uniformly into an ordering that is a uniform choice of all ones possible for the remaining $n-1$ cards.

Now $\mathbf{T}'$ is really a random variable, i.e. there are probabilities that $\mathbf{T}' = 1, 2, \ldots$, and by convention we write it in boldface. It is a particular example of a *stopping time*, when all orderings of the deck are equally likely. We may consider its expected value $E(\mathbf{T}')$, which clearly serves well as an intuitive idea of how randomizing a shuffle is, for

$E(\mathbf{T}')$ is just the average number of top-in shuffles needed to guarantee randomness by this method. The reader may wish to show that $E(\mathbf{T}')$ is asymptotic to $n \log n$. This is sketched in the following: Create random variables $\mathbf{T}_j$ for $2 \leq j \leq n$, which stand for the difference in time between when the tagged card first moves up to the jth position from the bottom and when it first moves up to the $j-1$st position. (The tagged card is said to have moved up to position 1 at step 0.) Then $\mathbf{T}' = \mathbf{T}_2 + \mathbf{T}_3 + \cdots + \mathbf{T}_n + 1$. Now the $\mathbf{T}_j$ are all independent and have densities

$$P[\mathbf{T}_j = i] = \frac{j-1}{n}\left(\frac{n-j+1}{n}\right)^{i-1}.$$

Calculating the expected values of these geometric densities gives $E(\mathbf{T}_j) = n/(j-1)$. Summing over j and adding one shows $E(\mathbf{T}') = 1 + n\sum_{j=1}^{n-1} j^{-1}$, which, with a little calculus, gives the result.

$\mathbf{T}'$ is good for other things as well. It is a theorem of Aldous and Diaconis [1] that $P[\mathbf{T}' > k]$ is an upper bound for the variation distance between the density on S_n after k top-in shuffles and the uniform density corresponding to true randomness. This is because $\mathbf{T}'$ is what's known as a strong uniform time.

Now we would like to make a similar construction of a stopping time for the riffle shuffle. It turns out that this is actually easier to do for the 2-unshuffle; but the property of being a stopping time will hold for both processes since they are exactly inverse in the sense that $\hat{R}_a(\pi) = R_a(\pi^{-1})$. To begin, recall from section 9 that an equivalent way of doing a 2-unshuffle is to place a sequence of n 0's and 1's on the deck, one on each card. Subsequent 2-unshuffles are done by placing additional sequences of 0's and 1's on the deck, one on each card, each time placing a new 0 or 1 to left of the 0's and 1's already on the card. Here is an example of the directions for 5 particular 2-unshuffles, as written on the cards of a size $n = 7$ deck before any shuffling is done:

card#	unshuffle# 54321	base 32
1	01001	9
2	10101	21
3	11111	31
4	00110	6
5	10101	21
6	11000	24
7	00101	5

The numbers in the last column are obtain by using the digitary concatenation operator & on the five 0's and 1's on each card, i.e. they are obtained by treating the sequence of five 0's and 1's as a base $2^5 = 32$ number. Now we know that doing these 5 particular 2-unshuffles is equivalent to doing one particular 32-unshuffle by sorting the cards so that the base 32 labels are in the order 5, 6, 9, 21, 21, 24. Thus we get the deck ordering 741256.

Now we are ready to define a stopping time for 2-unshuffling. We will stop after $\mathbf{T}$ 2-unshuffles if $\mathbf{T}$ is the first time that the base $2^{\mathbf{T}}$ numbers, one on each card, are all distinct. Why in the world should this be a guarantee that the deck is randomized? Well, consider all orderings of the deck resulting from randomly and uniformly labeling the cards, each with a base $2^{\mathbf{T}}$ number, conditional on all the numbers being distinct. Any two cards in the deck before shuffling, say i and j, having received different base $2^{\mathbf{T}}$ numbers, are equally as likely to have gotten numbers such that i's is greater than j's as they are to have gotten numbers such that j's is greater than i's. This means after $2^{\mathbf{T}}$-unshuffling, i is equally as likely to come after j as to come before j. Since this holds for any pair of cards i and j, it means the deck is entirely randomized!

John Finn has contructed a counting argument which directly shows the same thing for 2-shuffling. Assume $2^{\mathbf{T}}$ is bigger than n, which is obviously necessary to get distinct numbers. There are $2^{\mathbf{T}}!/(2^{\mathbf{T}} - n)!$ ways to make a list of n distict $\mathbf{T}$ digit base 2 numbers, i.e. there are that many ways to 2-shuffle using distinct numbers, each equally likely. But every permutation can be achieved by $\left(\begin{array}{c} 2^{\mathbf{T}} \\ n \end{array} \right)$ such ways, since we need only choose n different numbers from the $2^{\mathbf{T}}$ ones possible (so we have n nonempty packets of size 1) and arrange them in the necessary order to achieve the permutation. So the probability of any permutation under 2-shuffling with distinct numbers is

$$\left(\begin{array}{c} 2^{\mathbf{T}} \\ n \end{array} \right) / \left[\frac{2^{\mathbf{T}}!}{(2^{\mathbf{T}} - n)!} \right] = \frac{1}{n!},$$

which shows we have the uniform density, and hence that $\mathbf{T}$ actually is a stopping time.

Looking at the particular example above, we see that $T > 5$, since all the base 32 numbers are not distinct. The 2 and 5 cards both have the base 32 number 21 on them. This means that no matter how the rest of the deck is labeled, the 2 card will always come before the 5, since all the 21's in the deck will get pulled somewhere, but maintaining their relative order. Suppose, however, that we do a 6th 2-unshuffle by putting the numbers 0100000 on the naturally ordered deck at the beginning before any shuffling. Then we have $T = 6$ since all the base 64 numbers are distinct:

card#	unshuffle#	base 64
	654321	
1	001001	9
2	110101	53
3	011111	31
4	000110	6
5	010101	21
6	011000	24
7	000101	5

Again, $\mathbf{T}$ is really a random variable, as was $\mathbf{T}'$. Intuitively $\mathbf{T}$ really gives a necessary number of shuffles to get randomness; for if we have not reached the time when all the base $2^{\mathbf{T}}$ numbers are distinct, then those cards having the same numbers will necessarily always be in their original relative order, and hence the deck could not be randomized. Also analogous to $\mathbf{T}'$ for the top-in shuffle is the fact that $P[\mathbf{T} > k]$ is an upper bound for the variation distance between the density after k 2-unshuffles and true randomness, and hence between k riffle shuffles and true randomness. So let us calculate $P[\mathbf{T} > k]$.

The probability that $\mathbf{T} > k$ is the probability that an n digit base 2^k number picked at random does not have distinct digits. Essentially this is just the birthday problem: given n people who live in a world that has a year of m days, what is the probability that two or more people have the same birthday? (Our case corresponds to $m = 2^k$ possible base 2^k digits/days.) It is easier to look at the complement of this event, namely that no two people have the same birthday. There are clearly m^n different and equally likely ways to choose birthdays for everybody. If we wish to choose distinct ones for everyone, the first person's may be chosen in m ways (any day), the second's in $m-1$ ways (any but the day chosen for the first person), the third's in $m-2$ ways (any but the days chosen for the first two people), and so on. Thus the probability of distinct birthdays being chosen is

$$\frac{\prod_{i=0}^{n-1}(m-i)}{m^n} = \frac{m!}{(m-n)!m^n} = \binom{m}{n}\frac{n!}{m^n},$$

and hence the probability of two people having the same birthday is one minus this number. (It is is interesting to note that for $m = 365$, the probability of matching birthdays is about 50% for $n = 23$ and about 70% for $n = 30$. So for a class of more than 23 students, it's a better than fair bet that two or more students have the same birthday.) Transferring to the setting of stopping times for 2-unshuffles, we have

$$P[\mathbf{T} > k] = 1 - \binom{2^k}{n} \frac{n!}{2^{kn}}$$

by taking $m = 2^k$. Here is a graph of $P[\mathbf{T} > k]$ (solid line), along with the variation distance $\|R^k - U\|$ (points) that it is an upper bound for.

It is interesting to calculate $E(\mathbf{T})$. This is given by

$$E(\mathbf{T}) = \sum_{k=0}^{\infty} P[\mathbf{T} > k] = \sum_{k=0}^{\infty} \left[1 - \binom{2^k}{n} \frac{n!}{2^{kn}} \right].$$

This is approximately 11.7 for $n = 52$, which means that, according to this viewpoint, we expect on average 11 or 12 shuffles to be necessary for randomizing a real deck of cards. Note that this is substantially larger than 7.

11 STILL ANOTHER VIEWPOINT: MARKOV CHAINS

An equivalent way of looking at the whole business of shuffling is through Markov chains. A Markov chain is a stochastic process (meaning that the steps in the process are governed by some element of randomness) that consists of bouncing around among some finite set of states S, subject to certain restrictions. This is described exactly by a sequence of random variables $\{\mathbf{X}_t\}|_{t=0}^{\infty}$, each taking values in S, where $\mathbf{X}_t = i$ corresponds to the process being in state $i \in S$ at discrete time t. The density for $\mathbf{X}_0$ is arbitrary, meaning you can start the process off any way you wish. It is often called the *initial density.* In order to be a Markov chain, the subsequent densities are subject to a strong restriction: the probability of going to any particular state on the next step only depends on the current state, not on the time or the past history of states occupied. In particular, for each i and j in S there exists a fixed *transition probability* p_{ij} independent of t, such that $P[\mathbf{X}_t = j \mid \mathbf{X}_{t-1} = i] = p_{ij}$ for all $t \geq 1$. The only requirements on the p_{ij} are that they can actually be probabilities, i.e. they are nonegative and $\sum_j p_{ij} = 1$ for all $i \in S$. We may write the p_{ij} as a *transition matrix* $p = (p_{ij})$ indexed by i and j, and the densities of the $\mathbf{X}_t$ as row vectors $(P[\mathbf{X}_t = j])$ indexed by j.

It turns out that once the initial density is known, the densities at any subsequent time can be exactly calculated (in theory), using the transition probabilities. This is accomplished inductively by conditioning on the previous state. For $t \geq 1$,

$$P[\mathbf{X}_t = j] = \sum_{i \in S} P[\mathbf{X}_t = j \mid \mathbf{X}_{t-1} = i] \cdot P[\mathbf{X}_{t-1} = i].$$

There is a concise way to write this equation, if we treat $(P[\mathbf{X}_t = j])$ as a row vector. Then we get a matrix form for the above equation:

$$(P[\mathbf{X}_t = j]) = (P[\mathbf{X}_{t-1} = j]) \cdot p,$$

where the $\cdot$ on the r.h.s. stands for matrix multiplication of a row vector times a square matrix. We may of course iterate this equation to get

$$(P[\mathbf{X}_t = j]) = (P[\mathbf{X}_0 = j]) \cdot p^t,$$

where p^t is the tth power of the transition matrix. So the distribution at time t is essentially determined by the tth power of the transition matrix.

For a large class of Markov chains, called *regular*, there is a theorem that as $t \to \infty$, the powers p^t will approach a limit matrix, and this limit matrix has all rows the same. This row (i.e. any one of the rows) gives a density on S, and it is known as the *stationary density.* For these regular Markov chains, the stationary density is a unique limit for the densities of $\mathbf{X}_t$ as $t \to \infty$, regardless of the initial density. Furthermore, the stationary density is aptly named in the sense that if the initial density $\mathbf{X}_0$ is taken to be the stationary one, then the subsequent densities for $\mathbf{X}_t$ for all t are all the same as the initial stationary density. In short, the stationary density is an equilibrium density for the process. We still need to define a regular chain. It is a Markov chain whose transition matrix raised to some power consists of all strictly positive probabilities. This is equivalent to the existence of some finite number t_0 for the Markov chain such that one can go from any state to any other state in exactly t_0 steps.

To apply all this to shuffling, let S be S_n, the set of permutations on n cards, and let Q be the type of shuffle we are doing (so Q is a density on S). Set $\mathbf{X}_0$ to be the identity with probability one. In other words, we are choosing the intial density to reflect not having done anything to the deck yet. The transition probabilities are given by $p_{\pi\tau} = P[\mathbf{X}_t = \tau \mid \mathbf{X}_{t-1} = \pi] = Q(\pi^{-1} \circ \tau)$, since going from π to τ is accomplished by composing π with the permutation $\pi^{-1} \circ \tau$ to get τ. (An immediate consequence of this is that the transition matrix for unshuffling is the transpose of the transition matrix for shuffling, since $\hat{p}_{\pi\tau} = \hat{R}(\pi^{-1} \circ \tau) = R((\pi^{-1} \circ \tau)^{-1}) = R(\tau^{-1} \circ \pi) = p_{\tau\pi}$.)

Let us look at the example of the riffle shuffle with $n = 3$ from section 3 again, this time as a Markov chain. For $Q = R$ we had

π	[123]	[213]	[231]	[132]	[312]	[321]
$Q(\pi)$	1/2	1/8	1/8	1/8	1/8	0

So the transition matrix p, under this ordering of the permutations, is

$$\begin{array}{c} \\ {[123]} \\ {[213]} \\ {[231]} \\ {[132]} \\ {[312]} \\ {[321]} \end{array} \begin{array}{c} \begin{array}{cccccc} [123] & [213] & [231] & [132] & [312] & [321] \end{array} \\ \begin{pmatrix} 1/2 & 1/8 & 1/8 & 1/8 & 1/8 & 0 \\ 1/8 & 1/2 & 1/8 & 1/8 & 0 & 1/8 \\ 1/8 & 1/8 & 1/2 & 0 & 1/8 & 1/8 \\ 1/8 & 1/8 & 0 & 1/2 & 1/8 & 1/8 \\ 1/8 & 0 & 1/8 & 1/8 & 1/2 & 1/8 \\ 0 & 1/8 & 1/8 & 1/8 & 1/8 & 1/2 \end{pmatrix} \end{array}$$

Let us do the computation for a typical element of this matrix, say $p_{\pi\tau}$ with $\pi = [213]$ and $\tau = [132]$. Then $\pi^{-1} = [213]$ and $\pi^{-1} \circ \tau = [231]$ and $R([231]) = 1/8$, giving us $p_{[213][132]} = 1/8$ in the transition matrix. Although in this case, the $n = 3$ riffle shuffle, the matrix is symmetric, this is not in general true; the transition matrix for the riffle shuffle with deck sizes greater than 3 is always nonsymmetric.

The reader may wish to verify the following transition matrix for the top-in shuffle:

$$\begin{array}{c} \\ {[123]} \\ {[213]} \\ {[231]} \\ {[132]} \\ {[312]} \\ {[321]} \end{array} \begin{array}{c} \begin{array}{cccccc} [123] & [213] & [231] & [132] & [312] & [321] \end{array} \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 1/3 \end{pmatrix} \end{array}$$

The advantage now is that riffle shuffling k times is equivalent to simply taking the kth power of the riffle transition matrix, which for a matrix of size 6-by-6 can be done almost immediately on a computer for reasonable k. By virtue of the formula

$$(P[\mathbf{X}_t = j]) = (P[\mathbf{X}_0 = j]) \cdot p^t$$

for Markov chains and that fact that in our example $(P[\mathbf{X}_0 = j]) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$, we may read off the density of the permutations after k shuffles simply as the first row of the kth power of the transition matrix. For instance, Mathematica gives p^7 approximately:

$$\begin{array}{c c} & \begin{array}{cccccc} [123] & [123] & [123] & [123] & [123] & [123] \end{array} \\ \begin{array}{c} [123] \\ [213] \\ [231] \\ [132] \\ [312] \\ [321] \end{array} & \begin{pmatrix} .170593 & .166656 & .166656 & .166656 & .166656 & .162781 \\ .166656 & .170593 & .166656 & .166656 & .162781 & .166656 \\ .166656 & .166656 & .170593 & .162781 & .166656 & .166656 \\ .166656 & .166656 & .162781 & .170593 & .166656 & .166656 \\ .166656 & .162781 & .166656 & .166656 & .170593 & .166656 \\ .162781 & .166656 & .166656 & .166656 & .166656 & .170593 \end{pmatrix} \end{array}$$

and therefore the density after 7 shuffles is the first row:

π	[123]	[213]	[231]	[132]	[312]	[321]
$Q(\pi)$	.170593	.166656	.166656	.166656	.166656	.162781

It is clear that seven shuffles of the three card deck gets us very close to the uniform density (noting, as always, that the identity is still the most likely permutation), which turns out to be the stationary density. We first must note, not surprisingly, that the Markov chains for riffle shuffling are regular, i.e. there is some number of shuffles after which any permutation has a positive probability of being achieved. (In fact we know, from the formula $\begin{pmatrix} 2^k + n - r \\ n \end{pmatrix} / 2^{nk}$ for the probability of a permutation with r rising sequences being achieved after k riffle shuffles, that any number of shuffles greater than $\log_2 n$ will do.) Since the riffle shuffle Markov chains are regular, we know they have a unique stationary density, and this is clearly the uniform density on S_n.

From the Markov chain point of view, the rate of convergence of the $\mathbf{X}_t$ to the stationary density, measured by variation distance or some other metric, is often asymptotically determined by the eigenvalues of the transition matrix. We will not go into this in detail, but rather will be content to determine the eigenvalues for the transition matrix p for riffle shuffling. We know that the entries of p^k are the probabilities of certain permutations being achieved under k riffle shuffles. These are of the form $\left(\begin{array}{c} 2^k + n - r \\ n \end{array} \right) /2^{nk}$. Now we may explicitly write out

$$\left(\begin{array}{c} x + n - r \\ n \end{array} \right) = \sum_{i=0}^{n} c_{n,r,i} x^i,$$

an nth degree polynomial in x, with coefficients a function of n and r. It doesn't really matter exactly what the coefficients are, only that we can write a polynomial in x. Substituting 2^k for x, we see the entries of p^k are of the form

$$[\sum_{i=0}^{n} c_{n,r,i} (2^k)^i]/2^{nk} = \sum_{i=0}^{n} c_{n,r,n-i} (\frac{1}{2^i})^k.$$

This means the entries of the kth power of p are given by fixed linear combinations of kth powers of 1, 1/2, 1/4, . . ., and $1/2^n$. It follows from some linear algebra the set of all eigenvalues of p is exactly 1, 1/2, 1/4, . . ., and $1/2^n$. Their multiplicities are given by the Stirling numbers of the first kind, up to sign: multiplicity$(1/2^i) = (-1)^(n - i) s1(n, i)$. This is a challenge to prove, however. The second highest eigenvalue is the most important in determining the rate of convergence of the Markov chain. For riffle shuffling, this eigenvalue is 1/2, and it is interesting to note in the variation distance graph of section 8 that once the distance gets to the cutoff, it decreases approximately by a factor of 1/2 each shuffle.

References

[1] **Aldous, David and Diaconis, Persi,** Strong Uniform Times and Finite Random Walks, *Advances in Applied Mathematics*, **8**, 69-97, 1987.

[2] **Bayer, Dave and Diaconis, Persi,** Trailing the Dovetail Shuffle to its Lair, *Annals of Applied Probability*, **2(2)**, 294-313, 1992.

[3] **Diaconis, Persi,** *Group Representations in Probability and Statistics*, Hayward, Calif: IMS, 1988.

[4] **Harris, C., Peter,** The Mathematics of Card Shuffling, senior thesis, Middlebury College, 1992.

[5] **Kolata, Gina,** In Shuffling Cards, Seven is Winning Number, *New York Times*, Jan. 9, 1990.

[6] **Reeds, Jim,** unpublished manuscript, 1981.

[7] **Snell, Laurie,** *Introduction to Probability*, New York: Random House Press, 1988.

[8] **Tanny, S.,** A Probabilistic Interpretation of the Eulerian Numbers, *Duke Mathematical Journal*, **40**, 717-722, 1973.