Shuffling and Dealing Cards

Steven Portzer

1 Introduction

Card shuffling is an example of a mixing problem: how quickly is a deck randomized by repeated shuffling? Given some assumptions about the kinds of shuffles we are making, it is possible to determine how far the deck is from the uniform distribution. However, this analysis assumes that we care about the entire ordering of the sorted deck. For many card games, like bridge, the cards are dealt into hands with no inherent order. That is, we care about which player receives a given card but not about the order in which a player receives their cards. Hands of cards are not as sensitive to the ordering of the shuffled deck as full permutations. However, since the deck is not fully randomized, the scheme by which cards are dealt to players greatly affects the randomness of the resulting hands. Some methods perform substantially worse than others, and the commonly used method of going one direction around the table and cyclically dealing each player one card at a time turns out to be suboptimal.

2 Shuffling

While any reasonable method of shuffling allows for an arbitrarily well-mixed deck of cards, we will just consider the riffle shuffle. Commonly used by card players, riffle shuffling is simply cutting the deck into two piles and then interleaving, or riffling, the two piles back together. By assuming that all possible cut/riffle combinations are equally likely, we get the Gilbert-Shannon-Reeds model of card shuffling.

2.1 Generalizing Repeated Shuffles

The riffle shuffle naturally generalizes to the $a$-shuffle, as defined by Bayer and Diaconis [1]. The $a$-shuffle cuts the deck into $a$ distinct packets, and then riffles the packets together. Under this scheme, ordinary riffle shuffles are 2-shuffles. As we will prove, a sequence of 2-shuffles can also be represented as an $a$-shuffle. To visualize how an $a$-shuffle works, we first consider two alternate models of shuffling.

**Geometric Description.** For a deck of $n$ cards, take $n$ points chosen uniformly and independently in the unit interval. These points are labeled in order $x_1 < x_2 < \cdots < x_n$. An $a$-shuffle is defined by the map $x \rightarrow ax \mod 1$, taking the points $x_i$ in the unit interval to the unit interval. This mapping rearranges the points, which we take to be the permutation induced by the $a$-shuffle. The $x_i$’s in an interval $[(k-1)/a, k/a[$, $k \in \{1, 2, \ldots, a\}$ correspond to a packet produced by cutting the deck, and are interleaved with the points from other packets.

**Sequential Description.** Choose integers $j_1, j_2, \ldots, j_a$ with a multinomial distribution, that is, the $j_i$’s have the same distribution as the number of balls in each box $i$ if you drop $n$ balls independently at random into $a$ boxes. Cut the deck by taking the first $j_1$ cards off the top of the deck, then the next $j_2$ cards, and so on to divide the deck into $a$ packets. To recombine the packets back into one deck, riffle the first two packets together like you would for a 2-shuffle (i.e., with all possible interleaving being equally likely). Then riffle this new pile of cards with the third packet, and so on until every packet has been riffled back into the deck. This is equivalent to riffling all $a$ packets together at once by taking the next card from the $i$th packet with probability $A_i/(A_1 + A_2 + \cdots + A_a)$, where $A_i$ is the number of cards remaining in the $i$th packet.

**Lemma 1.** Both of these descriptions are equivalent to our assumption that all cut/riffle combinations are equally likely.

**Proof.** The number of possible interleavings of a given cut is a multinomial distribution because an interleaving is equivalent to a way of dividing the deck that results from the shuffle into $a$ subsets with the given
packets sizes, and the geometric description also follows a multinomial distribution for the packet sizes since the packet of each point is decided uniformly and independently at random. Given the packet sizes, we want all possible interleavings to be equally likely.

For the sequential description all interleavings of the first two packets are equally likely, given that interleaving all possible interleavings with the third packet are equally likely, and for the ith packet, all interleaving with the shuffled first i−1 packets are equally likely. If we consider the probability of some final interleaving, that interleaving uniquely determines all the intermediate interleavings. At each step the probability of a possible intermediate interleaving of \( j_i \) cards into the pile of \( j_1 + \cdots + j_i \) cards is, given the previous interleavings,

\[
\left( \frac{j_1 + \cdots + j_i - 1}{j_i} \right)
\]

which is independent of the previous interleavings for any legal interleaving. So, if we multiply \( a \) of these probabilities together to get the overall probability of some series of interleavings, then the result is the same for all series of interleavings. Therefore, the sequential description produces all combinations of cuts/riffles with equal likelihood.

For the geometric description, choosing \( n \) points in \([0, 1]\), labeling them with their leading digit in base \( a \), and then applying the map \( x \to ax (\mod 1) \) is equivalent to choosing the \( n \) points and arbitrarily labeling them with an integer in \([0, \ldots, a-1]\). Therefore, for given packet sizes, all interleavings are equally likely. Furthermore, the mapping produces points distributed uniformly in \([0, 1]\) and distributed independently of packet, so these mapped points can be reused for successive shuffles.

**Lemma 2.** An a-shuffle followed by a b-shuffle is equivalent to a single ab-shuffle.

**Proof.** Since the geometric description allows us to reuse the points for successive shuffles, this follows from the identity

\[
b(ax (\mod 1))(\mod 1) = abx (\mod 1).
\]

This result allow us to treat any sequence of successive a-shuffles as a single shuffle, most notably a sequence of \( n \) 2-shuffles is equivalent to a \( 2^n \)-shuffle.

**Theorem 1.** The probability that an a-shuffle will result in the permutation \( \pi \) is

\[
P_a(\pi) = \frac{\binom{a+n-r}{n}}{a^n}, \text{ where } r \text{ is the number of rising sequences in } \pi.
\]

Here, a rising sequence is defined as a maximal subsequence such that the subsequence was a sequence of successive cards in the original deck. For example, if a 2-shuffle takes 2, 5, 1, 4, 6, 3 to 2, 5, 6, 1, 3, 4 then the rising sequences are 2, 5, 1, 4 and 6, 3. Two rising sequences can not intersect, because then they would form a single larger rising sequence and rising sequences are maximal, so every card in the deck is part of exactly one rising sequences in \( \pi \).

**Proof.** This probability is proportional to the number of ways we can produce \( \pi \) by cutting and riffling the deck. Given a cut of the deck there is at most one way to produce a given permutation, so we just need to count the number of cuts such that it is possible to produce \( \pi \) by interleaving the packets. Each rising sequence in the shuffled deck must be a concatenation of packets from the original deck, so to count the number of ways we can produce \( a \) packets from \( r \) rising sequences we take the original deck and cut it into packets. The way we compute this is similar to the stars and bars argument from combinatorics. There are \( a-1 \) cuts, but \( r-1 \) of these occur between the successive pairs of rising sequences. The remaining \( a-r \) of these cuts can occur anywhere. If we consider a sequence of \( a+n-r \) objects, \( n \) of which are cards and the remaining \( a-r \) of which are the cuts between packets that occur inside rising sequences, then there is a one to one correspondence between ways we can pick the \( n \) which are cards and the ways we can divide the deck into \( a \) packets such the \( \pi \) is a possible interleaving of the packets. Thus, there are \( n+a-r \) choose \( n \) ways to produce \( \pi \). The total number of cut/riffle combinations is \( a^n \) since each such combination can be uniquely defined by assigning each card to a random packet, so the stated result is the probability of an a-shuffle resulting in \( \pi \).
2.2 Variation Distance

Now that we have a simple formula for the probability of a given permutation resulting from an \( a \)-shuffle, it’s reasonable to ask how well-mixed the deck is after shuffling. The metric Bayer and Diaconis use is variation distance from uniform, which is defined as

\[
||P_a - U|| = \frac{1}{2} \sum_{\pi \in S_n} |P_a(\pi) - U(\pi)|
\]

where \( U \) is the uniform distribution \( U(\pi) = \frac{1}{n!} \). Using Theorem 1 and the fact that the Eulerian number, \( A(n, r) \), equals the number of permutations of a deck of \( n \) cards such that the permutation has \( r \) rising sequences, we get

\[
||P_a - U|| = \frac{1}{2} \sum_{r=1}^{n} A(n, r) \left| \frac{1}{a^n} \binom{a+n-r}{n} - \frac{1}{n!} \right|
\]

It’s not too difficult to see that this goes to zero as \( a \) goes to infinity. Specifically though, the variation distance becomes roughly inversely proportional to \( a \) \([2]\). For repeated 2-shuffles, this means that eventually each successive shuffle halves the variation distance.

3 Dealing

The previous analysis assumes that we care about the exact ordering of the shuffled deck, but this is typically not the case. A more typical case is dealing cards into hands, as is the case in poker or bridge. The ordering of cards within a hand is unimportant. Consequently, the equation for variation distance from uniform is different from the basic case. Unfortunately, dealing into hand introduces complications that make it difficult to calculate exact probabilities.

Consider what happens when we do an \( a \)-shuffle on a deck of \( n \) cards. If \( a \) is much larger than \( n \), then we would expect all the packets to contain either one or zero cards. If this is the case, then when the packets are riffled together the deck will be perfectly randomized. However, if \( a \) is only fairly large and there are packets containing 2 cards, then after riffling these pairs of cards are guaranteed to be in the same order they were before shuffling. This effect biases cards that start near the top of the deck to remain closer to the top of the deck, and similarly cards near to the bottom of the deck will be more likely to end up near the bottom of the deck.

The common method of dealing bridge hands is cyclic dealing, for example dealing the first card to North, the second to East, the third to South, the fourth card to West, and then repeating the pattern to produce the sequence of deals \( NESWNESW \cdots NESW = (NESW)^{13} \). However, let us consider for a moment cutting the deck into hands, that is, dealing the first 13 cards to North, the next 13 to South, and so on \((N^{13}E^{13}S^{13}W^{13})\). Due to the bias of the deck, North is more likely to get cards that were originally at the top of the deck. The bias goes to zero as \( a \) gets very large, but it’s still far from ideal. To approximate the extent of the bias, consider \( u-v \) pairs, which we define as pairs \( (i, j) \) such that player \( u \) gets dealt the \( i \)th card, player \( v \) gets dealt the \( j \)th card, and \( i < j \). The difference between the number of \( u-v \) pairs and the number of \( v-u \) pairs will be proportional to how biased the dealing method is in terms of player \( u \) being more likely to receive cards near the start of the deck than player \( v \). In the case of cutting straight into hands we get 169 \( N-E \) pairs and no \( E-N \) pairs, and we’ll get the same numbers for other pairs of players. Cyclic dealing is a substantial improvement, with 91 \( N-E \) pairs and 78 \( E-N \) pairs, which gives a difference of 13. This isn’t the only source of bias, but Conger and Howald \([2]\) determined that it dominates the variation distance from uniform for \( a \)-shuffling with large values of \( a \). So, in the long run cyclic dealing is 13 times better than cutting into hands, which is worth \( \log_2(13) \approx 3.7 \) extra 2-shuffles.

Can we do better, though? As it turns out, yes. Notice that every time we cycle through the players, North gets dealt a card first. We can balance this by doing back-and-forth dealing, \((NESWSEN)^9(NESW)\). This method of dealing corresponds to dealing everyone one card going clockwise around the table, then
dealing a card to everyone going counterclockwise, then clockwise again, counterclockwise, and so on. This gives us 85 $N$-$E$ pairs and 84 $E$-$N$ pairs, which is as close as we can make them. This is 13 times better than cyclic dealing in the long run. For smaller values of $a$, Conger and Howald found that back-and-forth dealing surpassed cyclic dealing for 5 or more 2-shuffles. It’s not clear how well this applies to shuffling by actual humans, but the results would seem to indicate that the traditional method of cyclic dealing is not optimal.

References
