Far East Journal of Applied Mathematics

http://www.pphmj.com
http://dx.doi.org/10.17654/AM099040275
Volume 99, Number 4, 2018, Pages 275-288 **ISSN: 0972-0960**

CARD SHUFFLING FOR YOU AND ME

Alan Washburn

Operations Research
Naval Postgraduate School
1411 Cunningham Drive
Monterey, CA 93943
U. S. A.

Abstract

Most card games begin by shuffling the cards, ideally producing a deck where every possible permutation of cards occurs with equal probability. There is a popular notion that 7 shuffle repetitions will produce a sufficiently random deck, but that number is based on a theoretical analysis of an abstract kind of shuffle. Is 7 sufficient as a practical matter? The answer depends on who you are.

Introduction

There are shuffling machines capable of producing a perfect shuffle, and it is not difficult to program a computer to do the same thing as long as the computer has a random number generator. However, most card shuffling is still done by humans, often by humans who are impatient to get on with the game, so the question arises as to how much time should be spent shuffling, and how that time should be spent. There is a popular notion (Kolata [7]) that the right amount of time is whatever it takes to accomplish seven repetitions of the riffle, an operation where the deck is cut in half and then the two halves are whizzed together. This paper questions that notion.

Received: February 27, 2018; Accepted: April 3, 2018
Keywords and phrases: card shuffling.

There are manual methods in use other than riffling. One could, of course, simply cut the deck by placing the bottom part on the top, with the division between the two parts being random. One can also place the deck in one hand, using the other to repeatedly move a small part of the bottom to the top in a repeated cut until the remainder of the bottom is finally placed on top. The associated sound is of repeated chops, so we will refer to this shuffling method as “chopping”.

One could also put the cards all face down on the table and then just move them around for a while in close but not perfect proximity, combining the cards and then taking them apart again in a continuous motion. We will call this “messing”. Messing has the virtue that the shuffler cannot see the card faces.

The only manual methods referred to in the sequel are riffling, chopping and messing. Of these three, the most important is riffling.

Theoretical Shuffling Models

Most shuffling analyses begin by positing a particular type of theoretical shuffle, thus avoiding the physical details of manual shuffling. That theoretical shuffle is invariably chosen to permit the application of probability theory to the question of how many times the shuffle must be repeated to achieve near-perfect randomness.

Perhaps the simplest theoretical shuffle is introduced by Aldous and Diaconis [1]: the top card is removed and reinserted at a uniformly random place in the deck (one of 52 places). They argue that about $52\ln 52 \simeq 205$ repetitions would be required to achieve randomness. Fortunately there are more efficient methods of shuffling, the main one of interest here being the riffle.

An analysis of the question of how many times a riffle must be repeated to achieve randomness must proceed in two stages. The first is to find some sufficiently accurate abstract model of riffling, and the second is to find how many times that theoretical shuffle must be repeated. A model of some kind is required because no one (certainly not the author) has enough time to investigate whether riffling is sufficiently random by repeatedly riffling the cards over and over again.

The Riffle (note the uppercase R) is the most widely analyzed theoretical model of a riffle. In a Riffle, let X be a binomial random variable that counts the number of heads in 52 fair coin flips, and let the left hand hold the first X cards from the top of the deck while the right hand holds all the rest. Given the left and right hand holdings, the cards are then interleaved by selecting the next card for the shuffled pile to be the bottom card held in each hand with probability proportional to the number of cards left in that hand. This two-stage procedure is statistically equivalent (Aldous and Diaconis [1], and Levin et. al [8]) to flipping a coin to label each card with either 0 or 1, and then putting all of the 0 cards on top of the deck without changing the order of either the 0 cards or the 1 cards. It is the study of the Riffle (Bayer and Diaconis [2] call it the "dovetail") that have popularized 7 as the right number of repetitions of the riffle.

The physics of riffling have got to be interesting. One might expect there to be a hopeless card jam, with cards scattering all over the place when interleaving is attempted. Beginners sometimes suffer this fate, but most people learn the proper grip and bending and pressure that allows the cards to interleave smoothly. The shuffled deck will then consist of alternating "clots", each clot being a sequence of contiguous cards from one hand or the other. The last clot will consist of all remaining cards from whichever hand still has cards in it. Clot size no doubt depends in some subtle manner on the condition of the cards, the nature of the skin against which they are held and the pressures exerted by the riffler, both longitudinal and torsional. I am unaware of any study of riffling that delves into these physical questions, but Gilbert [6] apparently did some experiments to the effect that the Riffle is not a bad model of riffling, statistically speaking, at least for him. A semi-theoretical model of the author's riffles will be described below, but it is not the Riffle – the Riffle is not a good model of my riffle.

To investigate how many times a riffle must be repeated, whether mine or yours, we must first establish a measure to use in deciding whether a shuffled deck is "sufficiently random". That is the object of the next section.

Measuring Randomness

If M is a method of shuffling, let $M(x)$ be the probability of permutation x. A perfect method would have $M(x) = 1/N$ for all $N \equiv 52!$ permutations. Most theoretical studies of shuffling involve the *variation distance* between the studied method and the perfect method. This measure is bounded above by 1, so the object is to find a method where the measure is much closer to 0 than to 1. For example, Diaconis [4] (or see Snell [9]) employ the useful analytic properties of the Riffle to show that repeating it 7 times comes close enough to perfection in the sense that the variation distance from perfection is only about $1/3$. With six Riffles, the distance would be almost $2/3$, substantially larger. As the number of repetitions increases beyond 6, the variation distance decreases by a factor of about 2 with each one, so of course 8 would be better than 7, etc.

In spite of its useful analytic properties, we will not use variation distance as our measure of randomness here. This is mainly because it does not take well to the kind of simulation experiments we have in mind. For example, suppose we shuffled the cards n times, obtaining n distinct permutations in the process, and used the empirical distribution over permutations in measuring the variation distance. The variation distance from perfection would then be essentially 1 as long as n is much smaller than N. Since N exceeds 10^{67}, we cannot repeat any experiment enough times to make the sample variation distance be anything other than 1, even on a computer. We must find a different measure of randomness.

A good measure of randomness will depend on the subsequent use of the shuffled deck. Our measure here is motivated by the kind of games that people play with cards. At the conclusion of most card games, the cards will have been arranged in some order that would be significant for the next deal. In Bridge, each trick is likely to have four cards of the same suit. If those four cards stay together for the next deal, then each hand will get one card of that suit, leading to suit distributions that are more even than ought to be the case. Berger [3] examines statistics from 1000 tournament bridge deals to show that the suit distribution is in fact significantly more even in practice than it ought to be when cards are shuffled manually. Hands that have voids (0 cards in some suit) are not common enough.

In Poker games such as Hold'em, the last exposed card might be a clue as to the next one. If the last card were the queen of clubs, for example, then the next card could very well be a queen (if the previous game had involved collecting queens) or a jack or king (previous straight) or a club (previous flush).

The goal of shuffling, then, should be to break up this "stickiness" – the tendency for a shuffled card to be followed by the same card that followed it in the unshuffled deck. To test the success of a shuffle, we can conceptually number the cards and then go through the top 51 cards of the shuffled deck, counting the number of cases where that card's number is followed by a card with the next higher number. Only the first 51 cards are considered for the count – it is analytically tempting to make the first card "follow" the 52nd, but in fact that never happens when cards are actually dealt. We can generalize a bit by letting X_k be the number of times one of the first $52-k$ cards is followed by a card *k* positions down whose number is *k* larger; that is, the number of *k*-separated card pairs that survive the shuffling or more briefly the number of "*k*-matches". The idea is to compare X_k with what it would be, on the average, if shuffling were perfect.

Let m_k be the expected value of X_k under the hypothesis that shuffling is perfect. To derive an expression for m_k, first let random variable I_{ik} indicate whether the card initially in position *i* is, after shuffling, followed *k* positions down from the shuffled position of that card by the card initially in position $i+k$. For example, suppose card 3 of the unshuffled deck is followed 7 positions down by card 10. If the shuffled deck has card 3 in position 24 and card 10 in position 31, then $I_{3,7} = 1.$ There are in total $52 - 7$ possibilities for the shuffled deck where $I_{3,7} = 1,$ with card 3 occupying any position from 1 to 45. Each of those 45 possibilities has probability $1/(52 \times 51)$ when shuffling is perfect, and, except for those 45 possibilities, $I_{3,7} = 0.$ Letting $p_{i,k} \equiv P(I_{ik} = 1)$ and generalizing this argument, we have

$$p_{i,k} = (52 - k)/(52 \times 51); \quad i + k \leq 52.$$

Since $X_k = \sum_{i=1}^{52-k} I_{i,k}$ and $E(I_{ik}) = p_{ik}$, it follows that

$$m_k = E(X_k) = \sum_{i=1}^{52-k} p_{i,k} = (52 - k)^2/(52 \times 51); \quad 1 \leq k \leq 51.$$

Thus $m_1 = 51/52,$ slightly less than 1 because the card initially in position 1 could be in position 52 after shuffling, in which case its chance to be part of a 1-match disappears. Note that m_{51} is very small. This is because the only way for X_{51} to be nonzero is if both the first and last cards remain in their original positions after the deck is shuffled.

If we now define $\overline{X}_k$ to be the average value of X_k under whatever method of shuffling is under consideration, we have a natural statistic for the quality of the method, namely

$$Q_k \equiv \overline{X}_k / m_k.$$

Roughly speaking, the chance of a *k*-match is Q_k times as large as it ought to be. Most of our interest will be in $Q_1,$ and we will mostly insist that a shuffling method is satisfactory if and only if that number is smaller than 1.22. This is the value of Q_1 for 7 repetitions of the Riffle, the situation for which the variation distance is about $1/3.$

How many times should I riffle the cards?

Without an accurate understanding of what is happening physically when I riffle the cards, which I admit to not having, the only alternative is to simply riffle the cards *n* times and take note of the permutations that result. I did that for $n = 30$, and then investigated the effect of applying the 30 riffles sequentially on the number of 1-matches. That number starts out at 51 for the unshuffled deck. After each of the sequential riffles, in order, the number of 1-matches is (35, 28, 23, 17, 8, 8, 6, 4, 4, 3, 3, 2, 2, 2, 3, 4, 4, 2, 2, 2, 2, 2, 2, 2, 3, 3, 2, 3, 2, 2). The reason for the plunge from 17 to 8 after the fifth riffle is that the fifth riffle happens to be the first one where the shuffled bottom card is not the same as the unshuffled bottom card, thus for the first time preventing several matches at the bottom of the shuffled deck. For some reason my right hand, which holds the bottom half of the deck, almost invariably releases the first cards. After riffling the cards 30 times, I would still have two 1-matches, so $X_1 = 2$ for this replication of riffling the cards 30 times.

Not much can be concluded from this because it is only one replication of riffling the cards 30 times. Without making any assumptions about my riffling, the only alternative would be to repeat this experiment (say) 1,000 times, obtaining 1,000 1-match vectors, and then use those 1,000 vectors to measure Q_1 for up to 30 sequential riffles. Doing that would require 30,000 riffles. I am not willing to do that much riffling – 30 is my limit. Therefore, I will have to make enough assumptions about my riffling to permit computer replication by Monte Carlo simulation.

One possibility is to assume that each of my riffles is an independent random sample from the measured set of 30 riffles. I will subsequently refer to that theoretical riffler as “Alan”. Alan’s shuffling is a sequence of riffles, with each riffle being an independent random sample from the set of 30. Alan’s Q_1 can now be measured by Monte Carlo simulation, using a random number generator to select one of the 30 possible riffles on each occasion. Figure 1 shows the result of doing that (the dashed approximation will be introduced later). In Figure 1 and from here on, each of the riffles in a shuffle will be called a "repetition", whereas the Monte Carlo simulation will be indexed by "replications".

It turns out that Alan must riffle 17 times to make Q_1 smaller than 1.22. If Alan riffles the cards only 7 times, then Q_1 is almost 5. With only seven riffles, a poker player dealing with Alan's shuffled deck might benefit by taking careful notice of the unshuffled cards.

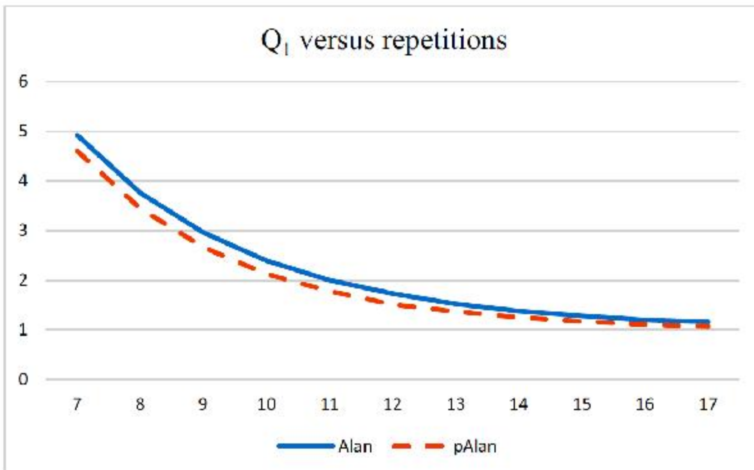

Figure 1. The Q_1 statistic versus repetitions for Alan (solid) as well as for an optimistic approximation called *pAlan* (dashed). Each number is based on 30,000 replications of a Monte Carlo simulation.

While Figure 1 is based on randomizing among all 30 of my recorded riffles, it is informative to consider randomizing among a smaller set. If I randomize among the odd numbered riffles, I get a Q_1 statistic that is consistently about 7% larger than when I randomize among the even numbered riffles – my even riffles are for some reason a better set to randomize over than my odd ones. This difference is sufficient to change the number of riffles required by about 1. Presumably the difference would disappear if there were 1,000 recorded riffles instead of only 30 – I have no reason to suppose that there is any fundamental difference between my even riffles and my odd riffles. Even smaller sets of riffles could be considered, but a set of one riffle is definitely not large enough. Every individual riffle will eventually result in replicating the unshuffled deck if it is repeated often enough, at which point the number of 1-matches will be 51 and a cycle will be initiated. The cycle length for my first riffle happens to be 30, so repeating that riffle 30 times is equivalent to doing nothing.

Instead of randomizing over a fixed set of 30 measured riffles, consider a different theoretical model called the *p-riffle*, the probability p being a parameter. To p-riffle the cards, first construct left and right piles as follows: put the top card in the left pile, and then continue putting cards in the same pile until a switch occurs, with a switch occurring independently with probability p for each card laid. Every time a switch occurs, change piles. The cards laid in each pile between switches are "clots", each of which contains 1 or more cards. When the two piles are finally complete, flip a coin to decide which one to put on top of the other to form an assembled deck.

The Riffle is equivalent to a $1/2$-riffle in the sense of producing each permutation x with the same probability $M(x)$. To show this, first observe that the identity permutation e is a special case that occurs in the $1/2$-riffle only if there are no switches (all heads or all tails) or if the first coin flip chooses the left pile and the other 51 flips produce exactly one switch. Each of these 53 flipping sequences has probability $1/2^{52}$, so $M(e) = 53/2^{52}$. The sequence of 52 coin flips otherwise determines a unique permutation that is not e, so for each of these we have $M(x) = 1/2^{52}$. Levin et al. [8] give this same distribution for the Riffle's permutations, so the two procedures are equivalent. This is convenient for comparisons, and the parameter p makes it possible to adapt the p-riffle to any individual's riffling.

In a p-riffle, there are 51 chances to start a new clot, each of which occurs with probability p, so there will on the average be $\mu \equiv 1 + 51p$ clots. A Riffle will therefore have 26.5 clots, on the average. Among my 30 measured riffles, the average number of clots is $\mu = 16.5$, considerably smaller than 26.5. I tend to have a few large clots, rather than lots of smaller ones. My associated value of p is then 0.305. The dashed curve in Figure 1 corresponds to a p-riffle with that parameter, hereafter called the *pAlan riffle*.

We now have two theoretical models of my riffling, Alan and pAlan. As should be evident from Figure 1, while both theoretical models of my riffling have the same average number of clots, the p-riffle is the more effective of the two. The basic reason for this is that variety is good in shuffling, and Alan's riffle does not have as much variety as the pAlan riffle. The probability that a clot will be of size 1 is always at least p in a p-riffle, whereas Alan's probability is only 0.24. Alan is also restricted to always using one of the 30 measured riffles, whereas a p-riffle is not so restricted. Figure 1 makes it clear that the difference is significant, but note that even pAlan is nowhere near as effective as a Riffle. This is because p-riffles improve strongly and monotonically with p in the interval $(0, 1/2]$, and a Riffle is the same as a $1/2$-riffle. This improvement continues for a while for $p > 0.5$. For example, Q_1 is only 1.04 for a 0.6-riffle. However, a 1-riffle would be a bad idea, and it should not be forgotten that Q_1 can be too small, as well as too large. The question of the best value for p is academic for me because I am unable to produce enough clots to be dangerous. For me, the best way to improve my riffling is to practice making more clots per riffle. I am working on it.

Another way for me to improve my riffling would be to alternately turn the deck upside down between riffles. Turning the deck upside down reverses bottom and top, thus fixing my problem with right-handedness, and has the additional physical benefit of keeping the shuffled cards more or less flat, rather than bowed in the middle. The downside is that my fellow players will notice that I am alternately examining and perhaps memorizing the cards as I riffle them, and no doubt make remarks about it.

How many times should you riffle the cards?

To find out how good a riffler you are, first measure the average clot count μ. Riffle the cards a few times. On each replication count the number of clots before you push the cards together. Average the clot counts to find μ and compare it to Alan's average clot count of 16.5. You will probably not get the same number, but, if you do, the results of the previous section apply. Diaconis [4] describes an experiment where he and also David Reeds riffled cards repeatedly and counted clots in that manner. Diaconis describes himself as similar to a professional dealer, and measures $\mu = 4376/103 = 42.5$, many of his clots consisting of single cards. Reeds shuffles like an "ordinary person" and gets $\mu = 3375/100 = 33.8$. Both of these numbers are considerably larger than mine, and may explain those authors' comfort in using Riffling as a model for riffling. I have also found people who get numbers significantly smaller than mine.

If your μ is significantly different from mine, you might download Excel™ workbook *riffle.xlsm* and do some further experimentation. Instructions are in the appendix.

What should we do about this?

If you are at all like me, riffling the cards 7 times is insufficient; the "stickiness" of cards for their neighbors is not well enough interrupted. Your fellow players are not going to wait patiently for you to riffle 17 times or whatever number it takes to make $Q_1 \le 1.22$, but an insufficiently shuffled deck may bias the game or even permit exploitation by players with a good memory. What should we do about this apparent crisis?

Whether there is truly a crisis depends on what game is being played. In the simplest form of Poker where every player gets five cards and the best hand wins, the stickiness of cards in the shuffled deck is of little import because the very process of dealing keeps any player from getting multiple consecutive cards. There is more to worry about in draw poker where each draw consists of consecutive cards. Texas Hold'em is even more worrisome, since so much depends on predicting the next card based on previous cards. The last player to receive a card face down has an enhanced probability of seeing one of the same rank on the Flop, and the Turn and the River cards are to some extent predictable from the last exposed card. The practice of "burning" a card before the Flop, Turn and River is usually justified as partially defeating the advantage of marking the backs of the cards, but could also be justified as partially defeating the effects of insufficient shuffling. When Alan riffles the cards seven times, for example, Q_2 is only 1.98, still larger than 1.22 but far smaller than Q_1. The practice of burning cards is thus one practical thing that can be done to compensate for imperfect shuffling.

Chopping instead of riffling is unlikely to be the solution to the crisis. It takes me about 3 seconds to riffle the cards once, so I can generate about $16.5/3 = 5.5$ clots per second by riffling. My chopping generates only about 1.2 clots per second, far smaller. Both procedures are basically clot generators, so riffling is over four times as effective a use of my time as chopping. An exception to this argument against chopping is that chopping, unlike riffling, is pretty much guaranteed to get the top card off the top and (more importantly, for me) the bottom card off the bottom. It makes sense to spend a small part of your shuffling time chopping.

Messing is not a clot generator like riffling and chopping, and can only be measured by the amount of time spent stirring the cards around. The identity of the messer is important, as is the state of the cards (new cards work better) and the nature of the table on which the messing is done. Very limited experiments with messing have been encouraging, but I am not in a position to compare messing with riffling. Perhaps new cards should be messed while old cards should be riffled.

Another solution, of course, would be to buy a shuffling machine. In gambling games where the stakes are significant, this may be a case where robots should replace humans.

References

[1] D. Aldous and P. Diaconis, Shuffling cards and stopping times, Amer. Math. Monthly 93 (1986), 333-348.

[2] D. Bayer and P. Diaconis, Trailing the dovetail shuffle to its lair, Ann. Appl. Probab. 2 (1992), 294-313.

[3] P. Berger, On the distribution of hand patterns in bridge: man-dealt versus computer-dealt, Canad. J. Statist. 1 (1973), 261-266.

[4] P. Diaconis, Group Representations in Probability and Statistics, IMS, Hayward, CA, Chapter 4, 1988.

[5] P. Diaconis, M. Grath and J. Pitman, Riffle shuffles, cycles and descents, Combinatorica 15 (1995), 11-29.

[6] E. Gilbert, Theory of Shuffling, Bell Laboratories Technical Memorandum, Murray Hill, NJ, 1955.

[7] G. Kolata, In Shuffling Cards, 7 is Winning Number, New York Times, January 9, 1990. http://www.nytimes.com/1990/01/09/science/in-shuffling-cards-7-is-winning-number.html?pagewanted=all

[8] D. Levin, Y. Peres and E. Wilmer, Markov Chains and Mixing Times, Chapter 8, American Mathematical Society, 2008 or see http://pages.uoregon.edu/dlevin/MARKOV/.

[9] J. L. Snell, Topics in Contemporary Probability and its Applications, J. L. Snell, ed., CRC Press, Chapter 9, 1995.

Appendix: Experiments with Riffle.xlsm

After measuring your value for μ, download Riffle.xlsm from http://faculty.nps.edu/awashburn/. That Excel™ workbook uses macros, so macros will have to be enabled to use it. It has two pages named "Simulation" and "Shuffle Record". The simplest experiments involve only the Simulation page. There are four input cells on that page. First, solve $\mu = 1 + 51p$ for *p*, input that value in the "Switch" cell, put 1 in the "Offset" cell, and then press the Simulate button, which runs a subroutine called *Shuffle*(). You will then see the Q_1 value for the input number of replications of the input number of *p*-riffle repetitions. Do not hesitate to do thousands of replications. You may wish to experiment a bit with the number of repetitions to see what the impact is on Q_1, perhaps seeking to make $Q \leq 1.22$. If you want to know Q_2, then repeat the experiment with 2 in the Offset cell. If you input a negative number in the Switch cell, shuffling will be perfect (even with only one repetition) and Q_1 should be approximately 1.

If you want to get a better idea of the effect of your own personal riffling, you will have to do what I did – riffle the cards 30 times and record the detailed results on the Shuffle Record page. Be guided by my own experience to minimize the time required to do this. Arrange the deck of cards in manufactured order with the ace of spades on top and the two of clubs on the bottom, and then number the cards from 1 to 52 by writing on them. Cut and interleave the cards but *do not* complete the riffle by pushing them together, lest you will have to arrange them all over again. Note the cards in each clot, alternately putting the clots in the right or left pile that they started in once you have done so. Overwrite my numbers with your own in each of the first 30 columns (first card number goes in the first row, second card number in the second row, …) as you repeat the 30 riffles. After you have recorded the results, put the two piles on top of one another so that card 1 is on top. The cards will then once again be in manufactured order and you will be ready for the next riffle. Each riffle should start with the cards in manufactured order.

After all 30 columns are input, push the Readem button on the Shuffle Record page to see various outputs. Look first at the “cycle length” outputs and find any input errors by finding columns where the cycle length is stated to be 100,000; in my case, the usual error is that some card number is mentioned more than once in that riffle’s column, thus turning it into something that is not a permutation of (1, ..., 52). Once you are satisfied that the input columns are all permutations, you can return to the Simulation page. If you input 0 in the Switch cell on that page, you will be simulating a riffle where each replication is a random sample from your set of 30 riffles. The value of p that corresponds to the mean clot number for your 30 riffles is reflected from the Shuffle Record page to the Simulation page, so you could also put that probability into the Switch cell as in the first paragraph.