Card-Shuffling Analysis with Markov Chains

Harald Hammarström

January 14, 2005

1 Introduction

In this essay we shall discuss mathematical models of card-shuffling. The basic question to answer is "how many times do you need to shuffle a deck of cards for it to become sufficiently randomized?". This obviously depends on what we mean by *shuffling* and *sufficiently randomized* so we shall dwell quite a bit on these points too.

In section 2 we will name and describe some popular shuffles. Then, in section 3, we describe the general framework for analyzing shuffles as random walks on Markov chains with the set of deck permutations as the state space. The payoff is section 4 that reviews some quite intriguing results on specific shuffles. Section 5 concludes with some general thoughts and comments.

2 Types of Shuffles

On can scramble the order of a deck of n cards in any manner at all and call it a shuffle. The following is a description of the most popular ones:

Riffle Shuffle Split the cards into two halves and interlace them (Aldous and Diaconis 1986).

Overhand Shuffle Hold the deck of cards in your right hand. Slide a small pack of cards off the top into your left hand. Repeat this process, putting each successive packet on top of the previous one so, until all the cards are in your left hand (Pemantle 1989).

Transposition Shuffle For $1 \leq i \leq n-1$, pick a random card at position j in $[i+1, n]$ and transpose the cards at i and j (Knuth 1981).

Top-in Shuffle Take a card from the top and insert it at a random position in the deck (Aldous and Diaconis 1986).

In each case above, what is described constitutes *one* shuffle. One occasionally encounters sloppy usage of the terminology in that a series of shuffles is also referred to as "a shuffle". As shall be seen later, we will also interpret a

shuffle more technically as simply a probability measure (involving no hands, algorithms, transformations or other processes).

The top-in shuffle is important in the theoretical analysis of randomization (see below). The transposition shuffle is a strong scrambling algorithm guaranteed to produce a uniform distribution with only n random bits (thus the preferred choice for computer card-shuffling). In the overhand shuffle, the order of the cards only gets reversed in clumps. Thus, understandably, by far the most popular among humans is the riffle shuffle.

The riffle-shuffle was given an exact interpretation by Gilbert and Shannon(Gilbert 1955) and independently by Reeds (Reeds 1981) as follows:

GSR-model Begin by choosing an c from $0, 1, \ldots, n$ according to the binominal distribution i.e $P(X = c) = \binom{n}{c}/2^n$. Holding c cards in the left hand and $n - c$ cards in the right, drop a card from a given hand with probability proportional to packet size. Thus, the chance that a card is first dropped from the left hand packet is c/n. If this happens, the chance that the next card is is dropped from the left packet is $(c - 1)/(n - 1)$ and so on.

The chance factor in the outcome is strong. In contrast, series of perfect riffle shuffles, where the deck is cut exactly in the middle and cards are perfectly interleaved, are not interesting. This is because 8 consecutive perfect riffle shuffles will restore a standard 52-deck to its original order!

In the GSR-model, we will cut a 52-card deck in the middle in only $\binom{52}{26}/2^{52} \approx 11\%$ of the time, and even if so we can expect two consecutive cards from the same hand to be dropped at almost every second drop. Because drops are probabilistic, proportional to packet size.

Moreover, the GSR is a realistic model for how professional card players actually shuffle cards. Empirical data on how often pairs of cards are dropped in practice can be found in (Epstein 1995). But it must be acknowledged that some realism has been sacrificed in order to promote a mathematically tractable formula. For instance, it would perhaps be more realistic to have a tendency for successive cards to be dropped from opposite hands, independent of packet size.

3 Modelling Shuffling with Markov Chains

3.1 Orders, Permutations and Shuffles

A deck of n cards can be ordered in $n!$ ways. The deck comes in some order, we (perhaps repeatedly) shuffle it, and then it comes out in some order. The outcome order also is dependent on the type of shuffle, the random data fed into it, and number of shuffles. In fact, the aim is to look at what happens when we vary shuffle methods and iterations. Since we are not interested in looking at shuffles for some special set of initial deck orders, we shall hold the initial order arbitrary. A reordering of an arbitrary sequence is a permutation, i.e a bijective function from an n-length sequence to an n-length sequence. But, as is

customary, permutations are written in bracket notation. So, for example, the 5-element permutation:

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

that changes an ordering 12345 to 23451 as well as 54231 to 42315 will be written $\pi = [23451]$.

As shuffle can now be seen as a probability measure on the set of permutations. Thus, a shuffle is not a stochastic mapping between orders, but plain and simple a probability measure on a set of permutations. If you have a deck in some order, to shuffle it is to pick a permutation according to the measure and permute your deck with it.

Any shuffle will induce a probability distribution on the $n!$ permutations of the deck. For exact shuffles the distribution will trivially be 1 for some permutation and 0 for the rest. But for shuffles which have a stochastic element in them, this distribution will be more interesting. For example, one riffle shuffle will introduce non-zero probabilities for some permutations whereas some permutations cannot occur after only one riffle shuffle. A permutation that cannot occur after only one riffle shuffle on 6 cards is [563412], because the original order, called [123456] would be split in two packets. It's clear from the interleaving that the relative order within each packet cannot be changed – there can only be cards inserted in-between. The [563412]-permutation does not correspond to any split of [123456] into two packets, where each preserves internal packet order, because it has scrambled order in three chunks.

Repeated riffle shuffles will intuitively make all orders possible at some point. Also intuitively, there is no point when the probabilities will be exactly uniform. The initial order will always be slightly favoured (cf. end of section 4.2.2). But we will get arbitrarily close, so close that even if you are a computer, you know how our riffle shuffle works in principle, you can hardly exploit the advantage.

3.2 Defining the Markov Chain

So a model of repeated shuffling is a Markov chain, i.e a sequence of random variables X_t for integer $t \geq 0$ that take values on the (finite) state space S_n (the $n!$ permutations on n elements). The transition probabilities, which are indeed independent of t, are described below.

Let Q be the type of shuffling we are using, so Q is a probability density on S_n. $Q(g) \geq 0$ and $\sum Q(g) = 0$ for $g \in S_n$. Set X_0 to the identity permutation.

$$P(X_1 = g) = Q(g)$$
$$P(X_2 = g) = Q * Q(g) = \textstyle\sum_{h \in G} Q(h)Q(gh^{-1})$$

Similarly $P(X_k = g) = Q^{k^*}(g)$ where $Q^{k^*}(g)$ is the repeated convolution:

$$Q^{k^*}(g) = Q * Q^{(k-1)^*}(g) = \sum_{h \in G} Q(h)Q^{(k-1)^*}(gh^{-1}) \tag{1}$$

For example, over S_3 and riffle shuffles we get the transition matrix:

$p(i,j)$	[123]	[213]	[231]	[132]	[312]	[321]
[123]	1/2	1/8	1/8	1/8	1/8	0
[213]	1/8	1/2	1/8	1/8	0	1/8
[231]	1/8	1/8	1/2	0	1/8	1/8
[132]	1/8	1/8	0	1/2	1/8	1/8
[312]	1/8	0	1/8	1/8	1/2	1/8
[321]	0	1/8	1/8	1/8	1/8	1/2

And, of course:

$$P(X_0 = j) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

A computer simulation gives e.g:

$$P(X_7 = j) = Q^{7^*} = \begin{pmatrix} 0.170593 & 0.166656 & 0.166656 & 0.166656 & 0.166656 & 0.162781 \end{pmatrix}$$

So after 7 riffle shuffles we are very close to the uniform distribution, but the identity is still the most likely permutation.

Riffle shuffle, and all other interesting shuffles, obviously yield regular Markov chains so we have:

$$Q^{k^*}(g) \to U(g) = 1/|S_n| \qquad \text{as k} \to \infty \tag{2}$$

A theorem first proved by Markov (Markov 1906).

The asymptotic result says nothing on how fast we approach uniformity. One might perhaps think that the decrease to uniformity is smooth and uneventful, but as we shall see below, what actually happens is that there is a sudden jump towards uniformity.

In probability theory the traditional measure of how far two probability measures are from each other is the *variation distance*, defined as:

$$\|Q_1 - Q_2\| = \frac{1}{2}\sum |Q_1(g) - Q_2(g)| \tag{3}$$

The $\frac{1}{2}$ is so that $0 \leq \|Q_1 - Q_2\| \leq 1$.

Suppose I know where in the deck a certain card i is, or equivalently, I know which card is at a certain place i. The rest of the deck is randomized. This defines a probability distribution Z over the permutations, where the $(n-1)!$ permutations satisfying my requirement on i have equal probability but the rest, $(n-1)(n-1)!$ have 0. Now $\|Z - U\| = 1 - 1/n$ which is almost maximal. Some may find it unintuitive that the variation distance is big yet we only knew one out of the possibly very many n cards. I consider this view mistaken given that knowing the place one card (or two etc) is very much information relative to other constraints one may have on the giant set of permutations. Knowing the place of a card is certainly more useful than knowing that, say, 17 cards (don't know which) are not part of any cycles of length 26.

We mentioned above that the convergence to uniformity happens abruptly. We can now make this precise by looking at the variation distance to the uniform distribution as the number of shuffles increase. Define

$$d_Q(k) \stackrel{\text{def}}{=} \|Q^{k^*} - U\| \tag{4}$$

for some shuffling method Q. From asymptotics we know that $d(k) < \epsilon$ for some large enough k, in fact it decreases to 0 geometrically fast. But the relevant question for card players is not "exactly how close to random does one million shuffles get you?", but rather "how many shuffles are enough?".

We shall now proceed to show the "abrupt decrease in variation distance", also known as a cut-off phenomenon, for some particular shuffles.

4 Analysis of Shuffles

4.1 The Top-in Shuffle

Deck has n cards. If we shuffle at least $n \log n$ times, then variation distance will be exponentially small in c for another cn shuffles.

$$d(n \log n + cn) \leq e^{-c}; \quad \text{all } c \geq 0, n \geq 2 \tag{5}$$

Proof: First we note that, if the card that was originally at the bottom, has moved to the top and we do one more top-in shuffle, then the deck is perfectly shuffled. All orders are equally possible. Define a random variable T as the number of shuffles until the first time this happens. What the the probability that this hasn't happened after $m = n \log n + cn$ shuffles? To put it another way, what is $P(T > n \log n + cn)$?

It is easier to look at the negation, that is, what is the probability that the top card has been at the top (at least once)? We can think of the process as an m-length sequence of randomly picked places in the deck, that is a an m-length base-n digit. We want to count the number of such sequences for which the top card has been at the top at least once. A combinatorial argument shows that this is $n! \left\{ {m \atop n} \right\}$ where $\left\{ {m \atop n} \right\}$ are the Sterling numbers of the second kind in Knuth-notation (Graham, Knuth, and Patashnik 1994). The Sterling numbers of the second kind yield the number of ways of arranging m different objects into n-nonempty (unlabeled) subsets. The total number of sequences is of course given by n^m, so the probability that the initial bottom card has reached the top and been re-inserted is:

$$\frac{n! \left\{ {m \atop n} \right\}}{n^m} \tag{6}$$

There is a combinatorial identity for the expression in the numerator ((6.19) p. 265 in (Graham, Knuth, and Patashnik 1994)):

$$n! \left\{ {m \atop n} \right\} = \sum_{k=0}^{n} \binom{n}{k} k^m (-1)^{n-k} \tag{7}$$

Plugging it in 6 we get

$$\frac{\sum_{k=0}^{n}\binom{n}{k}k^m(-1)^{n-k}}{n^m} \tag{8}$$

The terms in the sum can be re-ordered:

$$\frac{\sum_{k=0}^{n}\binom{n}{k}(n-k)^m(-1)^{k}}{n^m}$$

What we were really considering was the probability of this not occuring, which is:

$$1-\frac{\sum_{k=0}^{n}\binom{n}{k}(n-k)^m(-1)^{k}}{n^m}$$

By putting them on the same denominator, canceling out the first term in the sum, and switching signs, it simplifies to:

$$\frac{\sum_{k=1}^{n}\binom{n}{k}(n-k)^m(-1)^{k-1}}{n^m}$$

Now the terms in the sum decrease in value and alternate in sign, so of course $S_k \geq S_n \leq S_{k+1}$ (odd $k \geq 1$) where S_k if the sum of k first terms. We need only stop at S_1 to get the bound:

$$\begin{aligned} P(T > n\log n + cn) &= \frac{S_n}{n^m} \leq \frac{\binom{n}{1}(n-1)^m}{n^m} = \frac{n(n-1)^m}{n^m} = n\left(\frac{n-1}{n}\right)^m \\ &\leq n\exp(-m/n) = e^{-c} \end{aligned} \tag{9}$$

The next-to-last inequality uses the fact that $1-x \leq e^{-x}$ for all numbers x.

There is a shorter, perhaps simpler, proof of $P(T > n\log n + cn) \leq e^{-c}$ in (Aldous and Diaconis 1986). The idea is this: consider the waiting times for each T_i, the (random) number of shuffles it takes from the time there are $i-1$ cards below the original bottom card until there are i cards below it. Each T_i is labeled and has a different geometric probability. Distribute shuffles over these times (imagine for $m = 5$ shuffles, $n = 3$ times you put e.g 2 in T_1 and 0 in T_2, 3 in T_3). The problem can then be seen as distributing m units of shuffles into n labeled boxes of step times. Failure of the top card to ever come to the top happens when there is some T_i that doesn't get any draws, that is, all the other T:s used up all m draws. This probability can then be easily intutively bounded by $n(1-1/n)^m$.

But wait a minute! Now we have only proved that $m = n\log n + cn$ shuffles is enough, with high probability. Weren't we going to say something about the variation distance $d_Q(m)$? We need a lemma saying that the probability that m shuffles aren't enough is greater than or equal to the variation distance after m steps. This lemma holds if m, or rather the random variable $T =$ *the first time the original bottom card comes to the top and is inserted*, is a so-called strong stationary time. Formally:

Stopping Time T is a *stopping time* if, for each n, one can determine whether or not $T = n$ just by looking at the values of $X_0, \ldots, X_n$. In particular, to determine whether or not $T = n$ it is not necessary to know any "future" values $X_{n+1}, X_{n+2}, \ldots$.

Strong Stationary Time A Random Variable T is a *strong stationary time* if (i) T is a stopping time (ii) X_T is distributed as U, and (iii) X_T is independent of T.

Lemma: If T is a strong stationary time for the Markov chain X_n, then

$$\|\pi_n - U\| \leq P(T > n) \text{ for all } n \tag{10}$$

This lemma is proved in (Aldous and Diaconis 1986). It is not lenghty, but it is boring symbol manipulation. So with the above lemma, finally we get the desired (5):

$$d(n \log n + cn) \leq e^{-c}; \quad \text{all } c \geq 0, n \geq 2 \tag{11}$$

Similarly, one can show that $n \log n$ is not overkill, that many are necessary, because:

$$d(n \log n - c_n n) \to 1 \text{ as } n \to \infty; \text{ all } c_n \to \infty \tag{12}$$

This means that if you shuffle only $n \log n - c_n n$ times then, as n goes to infinity, the probability that we are not finished, that is there are still j cards above the original bottom cards, goes to one. When there are j cards that have not been shuffled properly the variation distance can obviously be arbitrarily high i.e arbitraily close to it's maximum of 1. Remember that if I know one card, then the variation distance is only off $1/n$ to its max. If I know two then it's only off by $1/n(n-1)$ etc. A proof of (12) is beyond my present capability, and the reader is referred to (Aldous and Diaconis 1986) for one that uses Chebyshev's inequality.

4.2 The Riffle Shuffle

Now that we have seen an example with the top-in shuffle, let's look at the riffle shuffle, which is the most popular. The asymptotic result, analoguous to that of the top-in shuffle, is that roughly $\frac{3 \log n}{2}$ shuffles are necessary and sufficient. The proof (Bayer and Diaconis 1992) is lengthy.

But what about the exact result for $n = 52$? Until a paper by (Bayer and Diaconis 1992) an exact computation used to be intractable, even for computers, due to the sheer size of the state space. The insight is that:

> The probability of achieving a permutation π depends not on all information in π but only on the number of rising sequences that π has.

4.2.1 Rising Sequences

So what is a rising sequence? A rising sequence of a permutation is a maximal consecutively increasing subsequence. For cards, assume you have a deck in some order labeled in increasing order, and perform a chosen permutation on them. Pick any card, call it say x, and look for card $x+1$ after it in the deck. If you find it, repeat the procedure looking for $x+2$ after $x+1$ and so on until you can't find the next card. Then, go back to card x and look for card $x-1$ before it, and so on. When finished you have a rising sequence $x-i, x-i+1, \ldots, x, \ldots, x+j-1, x+j$. It turns out that a deck breaks down as a disjoint union of its rising sequences.

As an example, let's say we have a deck labeled 12345678. We permute it with the permutation [45162378]. The deck is now in order 45162378. We start with a card, say 3, and look for 4 after it, but we don't find it. However, we find 2, and then 1 below it. So 123 is one rising sequence. Then suppose we start with 6, we then get the rising sequence 45678 and that exhausts the cards. So the permutation [45162378] has exactly two rising sequences. Note that this permutation could be the result of a riffle shuffle with cut 3. It is also easily seen that the result of one riffle must be a permutation with 2 rising sequences, or the identity, with 1 rising sequence.

4.2.2 Introducing a-shuffles

The generalization of the riffle shuffle, called the a-shuffle is achieved as follows: Cut the deck into a packets of (nonnegative) sizes $b_1, b_2, \ldots, b_a$, with the probability of this particular packet structure given by the multinominal density: $\binom{n}{b_1,b_2,\ldots,b_a}/a^n$. Note that $b_1 + b_2 + \ldots + b_a = n$ but some of the b_i:s may be zero. Interleave by dropping cards from each packet, one at a time, with probability proportional to packet size (relative to the total number of cards still left to drop). This interleaving is also equivalent to riffling first packets b_1 and b_2 together, then that packet with b_3, the resulting with b_4 and so on. It is obvious that the GSR-riffle shuffle is the a-shuffle with $a = 2$.

So, what is the relevance of the a-shuffle for a creature that is not a-handed? The answer is that k repeated 2-shuffles are equivalent to one single 2^k-shuffle. In fact, the following is a theorem that can be proved with elementary combinatorics (can be found in e.g (Mann 1995)):

> 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.

We are now ready to state the exact result for achieving any permutation by an a-shuffle:

> The probability 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.

The proof, also using only elementary combinatorics, is given in (Mann 1995) or (Bayer and Diaconis 1992) where it is stated that it is a generalization of earlier work by Shannon. It should be noted that the probability $\binom{n+a-r}{n}/a^n$ is a monotone descreasing function of r. So if $1 \leq r_1 \leq r_2 \leq n$, then the probability of a permutation with r_1 rising sequences is always more probable than one with r_2.

4.2.3 The Final Result

Now we can get an exact formula for the distance of k riffle shuffles to the uniform distribution:

$$d_R(k) = \|R^{k*} - U\| = \frac{1}{2}\sum_{r=1}^{n} \left\langle {n \atop r} \right\rangle \left| \binom{2^k + n - r}{n} / 2^{nk} - \frac{1}{n!} \right| \tag{13}$$

Where $\left\langle {n \atop r} \right\rangle$ is the Eulerian numbers, using the notation from (Graham, Knuth, and Patashnik 1994). $\left\langle {n \atop r} \right\rangle$ counts the number of permutations of n elements with exactly r rising sequences, and there are various recursive formulas to compute them.

The insight of the invariance of probabilities for different permutations with the same number of rising sequences is what reduces the number of terms in the sum from $n!$ ($\approx 10^{68}$ for $n = 52$) to n. Now it is tractable. Figure 1 has a graph for $d_R(k)$ for k up to 10:

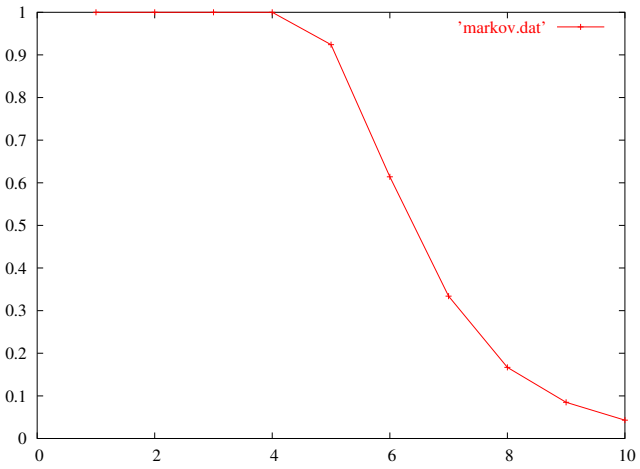

Figure 1: $d_R(k)$ for k in $[1, 10]$. R is the riffle shuffle.

It is apparent that the graph makes a sharp cutoff, and is reasonably low at $k = 7$ which is why "seven shuffles are enough to randomize a deck of cards". Seven shuffles are indeed enough for most kinds of games where you cannot really exploit the kind of non-randomness that may still remain. The variation

distance after 7 shuffle is ≤ 0.334 and Peter Doyle has invented a game specifically designed to make use of the more-than-random number of length-2 rising sequences that still remain (described in (Mann 1995)).

On the other hand, if you do less than 7 shuffles, say 3 or 4, and think that this is enough for most purposes you are on thin ice. There is a wonderful card trick described in (Bayer and Diaconis 1992) that magicians have been using since the beginning of the century, which exploits the highly predictable outcome of only 3 or 4 shuffles.

4.3 Other Shuffles

In (Jonasson 1995) some generalizations on the riffle shuffle are discussed; such as results on *biased riffle shuffles*. Then the cut is not binominal, but can have an arbitrary distribution. Lower and upper bounds are given.

Since (Pemantle 1989), it has been known that at least $O(n^2)$ and at most $O(n^2 \log n)$ shuffles is necessary to randomize using the overhand shuffle (note that this amount to thousands of shuffles for $n = 52$). But (Jonasson 2004) has just recently shown, following a technique introduced by D. B. Wilson, that the upper bound is also tight – so $\Theta(n^2 \log n)$ overhand shuffles are required.

The (Jonasson 2004) paper has references to studies of yet other shuffle types that have appeared in the literature.

5 Comments

The cut-off point of the riffle shuffle for $n = 52$ is indisputable, but it is not completely sharp. As n goes to infinity however, the variation distance curve will have a square drop at the cut-off!

There are many other Markov chains that exhibit the peculiar cut-off phenomenon in variation distance (also called threshold phenomenon), in their approach to stationarity, but there also those who do not. Why some do and some don't was still not fully understood according to Diaconis in 1996 (Diaconis 1996). Examples and further discussion are in that paper and (Aldous and Diaconis 1986).

It should also be mentioned that riffle-shuffling does not exhibit a threshold phenomenon when looking at entropy rather than variation distance (Trefethen and Trefethen 2000). That is, if we look at how the entropy decreases in the number of shuffles, it decreases smoothly without any cut-off. Entropy is, as usual, the number of decks an infinitely competent coder will have to transmit on average, to encode one bit. This difference between variation distance and entropy has been known to the card-shuffling community since the beginning. I know of no argument as to why entropy should be the relevant measure rather than variation distance[1].

[1] The explanation in (Trefethen and Trefethen 2000) is hard to follow.

References

Aldous, D. and P. Diaconis (1986, May). Shuffling cards and stopping times. *American Mathematical Monthly 93*(5), 333–348.

Bayer, D. and P. Diaconis (1992). Trailing the dovetail shuffle to its lair. *Annals of Applied Probability 2*, 294–313.

Diaconis, P. (1996, Feb. 8). The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences 93*, 1659–1664.

Epstein, R. (1995). *The Theory of Gambling and Statistical Logic*. Academic Press. Reprint of the revised 1977 edition.

Gilbert, E. W. (1955). Theory of shuffling. Technical report, Bell Laboratories Technical Memorandum, Murray Hill, NJ. Cited in Aldous & Diaconis, 1986. Not consulted by the present author.

Graham, R. L., D. E. Knuth, and O. Patashnik (1994). *Concrete Mathematics* (2 ed.). Addison-Wesley, Reading, Massachusetts.

Jonasson, J. (1995). Some generalizations of the riffle shuffle. Technical report, Chalmers University.

Jonasson, J. (2004). The overhand shuffle mixes in order $n^2 \log n$ steps. *Submitted*.

Knuth, D. E. (1981). *Seminumerical Algorithms*, Volume 2 of *The Art of Computer Programming*. Addison-Wesley, Reading, MA.

Mann, B. (1995). How many times should you shuffle a deck of cards? In L. J. Snell (Ed.), *Topics in Contemporary Probability and its Applications*, Probability and Stochastics Series, pp. 261–289. CRC Press, Boca Raton, FL.

Markov, A. A. (1906). Rasprostranenie zakona bol'shix chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya 15*, 135–156. Cited in Aldous & Diaconis, 1986. Not consulted by the present author.

Pemantle, R. (1989). Randomization time for the overhand shuffle. *Journal of Theoretical Probability 2*(1), 37–49.

Reeds, J. (1981). Unpublished manuscript. Cited in Aldous & Diaconis, 1986. Not consulted by the present author.

Trefethen, L. N. and L. M. Trefethen (2000, Oct. 8). How many shuffles to randomize a deck of cards? *Proceedings of the Royal Society, London A 456*, 2561–2568.