How many shuffles to randomize a deck of cards?

By Lloyd N. Trefethen[1] and Lloyd M. Trefethen[2]

[1] *Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK*

[2] *Department of Mechanical Engineering, Tufts University, Medford, MA 02155, USA*

A celebrated theorem of Aldous, Bayer and Diaconis asserts that it takes $\sim\frac{3}{2}\log_2 n$ riffle shuffles to randomize a deck of n cards, asymptotically as $n\to\infty$, and that the randomization occurs abruptly according to a 'cutoff phenomenon'. These results depend upon measuring randomness by a quantity known as the total variation distance. If randomness is measured by uncertainty or entropy in the sense of information theory, the behaviour is different. It takes only $\sim\log_2 n$ shuffles to reduce the information to a proportion arbitrarily close to 0, and $\sim\frac{3}{2}\log_2 n$ to reduce it to an arbitrarily small number of bits. At $\frac{3}{2}\log_2 n$ shuffles, about 0.0601 bits remain, independently of n.

1. Introduction

Wide publicity has been attracted in recent years to the question, how many riffle shuffles does it take to randomize a deck of cards? A beautiful mathematical paper by Bayer and Diaconis in 1992, building upon earlier work by Aldous and by Diaconis, proved that in a certain precise sense the answer is $\sim\frac{3}{2}\log_2 n$ for a deck of n cards in the limit $n\to\infty$ (Aldous 1983, Bayer & Diaconis 1992, Diaconis, McGrath & Pitman 1995). Moreover, the randomization arrives abruptly: after $1.4\log_2 n$ shuffles, for large enough n, the deck is nowhere near random. These conclusions have been discussed on radio talk shows and in newspapers and magazines including *The New York Times*, *The Economist*, *Newsweek*, and *Seventeen* (Kolata 1990). They do not stand in isolation but are part of the developing subject of the analysis of non-asymptotic convergence of Markov chains, with implications in condensed matter physics, computer science, and other fields (Su 1995, Diaconis 1996).

Throughout our discussion, a riffle shuffle is defined in a mathematically precise way due to Gilbert and Shannon (Gilbert 1955) and independently Reeds (J. Reeds 1981, unpublished manuscript). The deck is first cut roughly in half according to a binomial distribution: the probability that ν cards are cut is $\binom{n}{\nu}/2^n$. The two halves are then riffled together by dropping cards roughly alternately from each half onto a pile, with the probability of a card being dropped from each half proportional to the number of cards in it. There is evidence that this

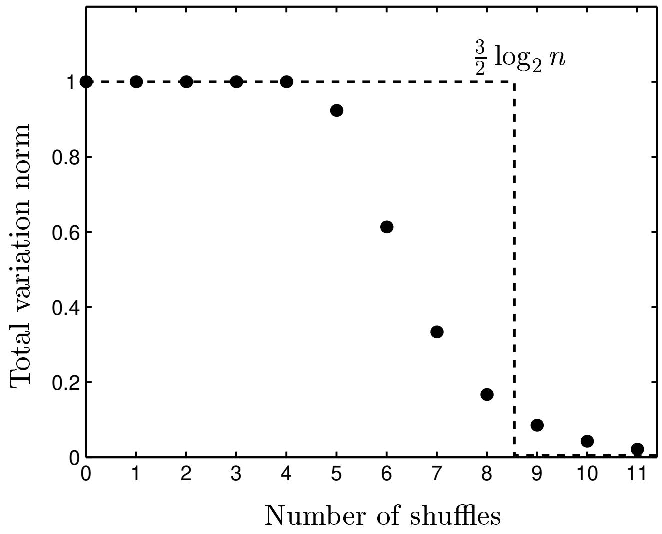

Figure 1. Randomization of a deck of n cards as measured in the total variation norm $\|P^k - P^\infty\|_{\mathrm{TV}}$ of Aldous, Bayer and Diaconis. The dots and the numerical axis labels correspond to $n = 52$ and the dashed line to the limit $n \to \infty$. In this limit a 'cutoff phenomenon' occurs, with abrupt randomization at $\sim\frac{3}{2}\log_2 n$ shuffles. For $n = 52$, $\|P^k - P^\infty\|_{\mathrm{TV}}$ falls below 0.5 at the seventh shuffle.

idealization of a shuffle is a reasonable approximation to the actual behaviour of human shufflers (Diaconis 1988).

Of course, there are other ways to achieve randomness besides shuffling. For example, "exact mixing" methods have recently been investigated by Lovász and Winkler (1995). However, we confine our attention here to randomization by riffle shuffle. Some of the wider mathematical context of our discussion, concerning ill-conditioned eigenvalues and eigenvector expansions in this and other fields, is mentioned in section 5.

2. Shuffling and total variation norm: the cutoff phenomenon

Figure 1 illustrates the theorem of Diaconis and his colleagues. The kth dot indicates the total variation distance to randomness $\|P^k - P^\infty\|_{\mathrm{TV}}$ (defined below) after k shuffles. Through step $k = 4$, virtually no reduction is achieved, and $\|P^k - P^\infty\|_{\mathrm{TV}}$ does not fall below 0.5 until step $k = 7$. This is the origin of the often-quoted conclusion that 'it takes 7 shuffles to randomize a deck of cards'. As $n \to \infty$, the dots straighten up into the sharp curve indicated by the dashed line. Specifically, if $k/\log_2 n \to \alpha$ as $n \to \infty$ for some constant α, then $\|P^k - P^\infty\|_{\mathrm{TV}} \to 1$ if $\alpha < 1.5$ and $\|P^k - P^\infty\|_{\mathrm{TV}} \to 0$ if $\alpha > 1.5$.

Mathematically, the shuffling problem is a Markov chain defined on the state space consisting of the $n!$ possible orderings of the deck (for $n = 52$, $n! \approx 8\times 10^{67}$). Suppose that at a particular moment, the probability that the deck is in ordering i is p_i, with $0 \le p_i \le 1$ and $\sum_{i=1}^{n!} p_i = 1$. If p represents the row vector of these probabilities, of length $n!$, then one step of the shuffling process replaces p by the product pP, where P is an $n! \times n!$ matrix with nonnegative entries and row sums equal to 1; the entry P_{ij} is the probability that the chain, if currently at state i, moves to state j at the next step. This much is standard material in the field of Markov chains (Feller 1968, Meyn & Tweedie 1993, Norris 1997). The total variation norm after step k is defined by the formula

$$\|P^k - P^\infty\|_{\mathrm{TV}} \;=\; \tfrac{1}{2}\,\max_i \sum_{j=1}^{n!} |(P^k - P^\infty)_{ij}|\;, \tag{2.1}$$

where P^k is the kth power of P and P^∞ is the limit of P^k as $k \to \infty$ (Jónsson & Trefethen 1998). This formula represents half the 1-norm of the matrix $P^k - P^\infty$ when viewed as acting on row vectors (Trefethen & Bau 1997), and it can be interpreted as follows. Let A be a subset containing $|A|$ elements of the set of all $n!$ permutations of the deck, and let $p^{(k)}(A)$ be the probability that the deck lies in one of the configurations of A at step k. Then $\|P^k - P^\infty\|_{\mathrm{TV}}$ is the difference $|\,p^{(k)}(A) - |A|/n!\,|$, maximized over all subsets A. This number quantifies the rate at which an infinitely competent gambler could expect to make money, on average, if permitted to place bets with payoff 1 against a fair house to the effect that the deck does or does not lie in arbitrary sets of configurations A.

3. Shuffling and uncertainty: steady randomization

In the field of probability theory, there are longstanding arguments for considering the total variation norm. On the other hand, the shuffling of a deck of cards, like the wide range of other Markov chain problems of which this may be viewed as a prototype, can also be considered from the point of view of information theory. Let the uncertainty or entropy associated with a probability vector p be defined by the familiar formula associated with Wiener and Shannon (Shannon & Weaver 1949, Kullback 1959, Renyi 1970, Barron 1986, Cover & Thomas 1991),

$$U \;=\; -\sum_{i=1}^{n!} p_i \log_2 p_i. \tag{3.1}$$

This quantity ranges from 0 if we have complete information about the system ($p_i = 1$ for a single i) to $\log_2(n!)$ if we have no information ($p_i = 1/n!$ for all i). Conversely, the information associated with p is defined by

$$I \;=\; \log_2(n!) - U. \tag{3.2}$$

According to standard results of information theory, this number quantifies the rate at which an infinitely competent coder could expect to transmit information, on average in the limit of infinitely long message lengths, if permitted to encode signals arbitrarily in shuffled decks of cards.

Shuffling a deck of n cards can thus be thought of as a process of destruction of information, in which the information content of the deck is reduced from $\log_2(n!)$ to 0 bits. The question is, how many shuffles does it take to achieve this? We have computed answers to this question numerically. An earlier analysis of alternatives to total variation for various Markov chain problems is presented by Su (1995), and our I is essentially the relative entropy distance considered by

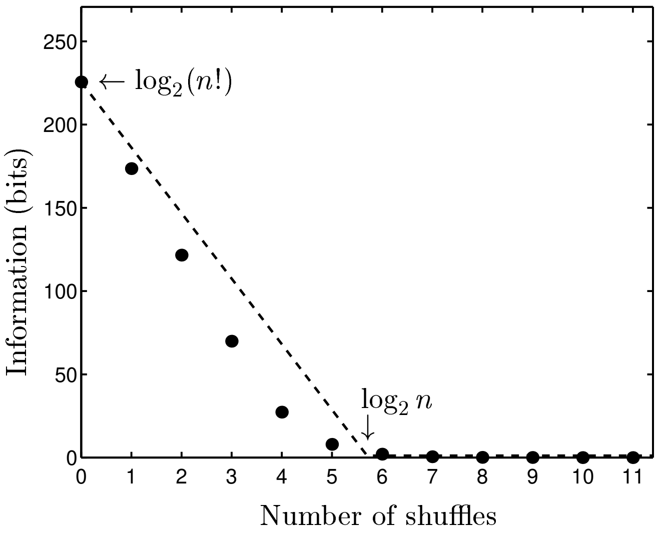

Figure 2. Randomization as measured by reduction of information from $\log_2(n!)$ to 0 bits. Again, the dots and the numerical axis labels correspond to $n = 52$ and the dashed line to $n \to \infty$. In this measure there is no cutoff effect, and randomization in the sense of reduction of the original information to a proportion arbitrarily close to 0 is achieved after only $\sim\log_2 n$ shuffles. For $n = 52$, 3.52% of the information remains after five shuffles and 0.92% after six shuffles.

him. In particular, Su observed behavior like that described here for problems related to Ehrenfest urns or random walk on a hypercube. A standard reference on the use of information-related measures in statistics is Kullback (1959).

Figure 2 shows results for both $n = 52$ and the limit $n \to \infty$, and some of the numbers for $n = 52$ are reported in table 1. The first shuffle reduces I by almost exactly n bits (about 51.999 999 999 999 93 bits, for $n = 52$). Subsequent shuffles also reduce I by approximately n bits until I reaches a level that is small relative to its initial value $\log_2(n!)$. Each further shuffle then reduces I by a factor asymptotically of $1/4$. In the measure of information, evidently, the cutoff phenomenon is absent. Shuffles remove information from the deck in a steady fashion, until asymptotically as $k \to \infty$, all the information is gone.

A quantitative analysis of the process just described sheds light on the disparity between figs. 1 and 2. Suppose we wish to reduce I from $\log_2(n!)$ to $\epsilon\log_2(n!)$ for some ϵ with $0 < \epsilon \ll 1$. At n bits per shuffle, since $\log_2(n!) \sim n\log_2(n/e) \sim n\log_2 n$, this takes $\sim\log_2 n$ shuffles. We call this the linear phase of the shuffling process. Now suppose we wish to reduce I further to some absolute level $\delta > 0$, independent of n as $n \to \infty$. With a reduction by the factor $\frac{1}{4}$ at each shuffle, this takes $\log_4(\epsilon\log_2(n!)/\delta) \sim \log_4(\log_2(n!)) \sim \log_4(n\log_2 n) \sim \log_4 n = \frac{1}{2}\log_2 n$ further shuffles. We call this the exponential phase of the shuffling process. Figure 3 illustrates these two phases.

The shuffling process is governed by powers of the $n! \times n!$ matrix $P - P^\infty$, since $(P - P^\infty)^k = P^k - P^\infty$ for $k \geq 1$ (Jónsson & Trefethen 1998), and the asymptotic convergence rate $\frac{1}{4}$ is equal to the square of the largest eigenvalue of this matrix, $\frac{1}{2}$, which is the same as the second eigenvalue of the matrix P. A

Table 1. *Information I in an initially ordered deck of 52 cards.*

shuffle no.	information (bits)
0	225.58
1	173.58
2	121.58
3	69.874
4	27.271
5	7.9452
6	2.0727
7	0.5239
8	0.1313
9	0.0329
10	0.0082

general result about this squaring of the eigenvalue appears as Corollary 5.3 of Su (1995), though it is not precisely applicable to the present case since it assumes a reversible Markov chain and the riffle shuffle chain is irreversible. The study of the second eigenvalue of Markov chains is well established; see for example Diaconis & Stroock (1991) and Fill (1991). For a different view of the gap between eigenvalues and convergence for Markov chains see Stewart (1997).

A curious observation emerges from figure 3 and related computations. After $\frac{3}{2}\log_2 n$ shuffles, we find that about 0.0601 bits of information remain in the deck. Since the shuffle number is discrete and the information is divided by approximately 4 at each step, of course, there will be no particular step k in general at which 0.0601 bits remain; one might have 0.1 bits at one step and 0.025 bits at the next. Nevertheless, the number 0.0601 emerges very cleanly when the discrete data are smoothly interpolated to $k = \frac{3}{2}\log_2 n$. Thus if one wished to communicate by encoding messages in large decks of cards, each one having been shuffled $\frac{3}{2}\log_2 n$ times, one would need to ship about 16 decks per bit.

4. Numerical methods

The results we have presented are numerical, though several of them suggest theorems that presumably could be proved. We give just a brief outline of our methods. The matrix P is of the computationally intractable dimension $n!$, but it can be reduced to an equivalent matrix problem of size n by identifying all permutations of the deck that have the same number of 'rising sequences'. The ideas that make this possible are contained in Bayer & Diaconis (1992), and the matrix entries have been worked out explicitly in Jónsson & Trefethen (1998) and Jónsson (1997). Copies of our MATLAB programs, about 100 lines in total, can be obtained from LNT.

We have also computed I vs. k for another well-known example that shows a cutoff effect in the total variation norm, the problem of Ehrenfest urns, where at each step, one of n balls located in either of two urns is selected at random and moved to the other urn (Kac 1959, Bingham 1991, Diaconis 1996, Jónsson & Trefethen 1998). As mentioned above, related theorems for this problem are reported in Su (1995). The cutoff of $\|P^k - P^\infty\|_{\text{TV}}$ for this problem is at $\sim$

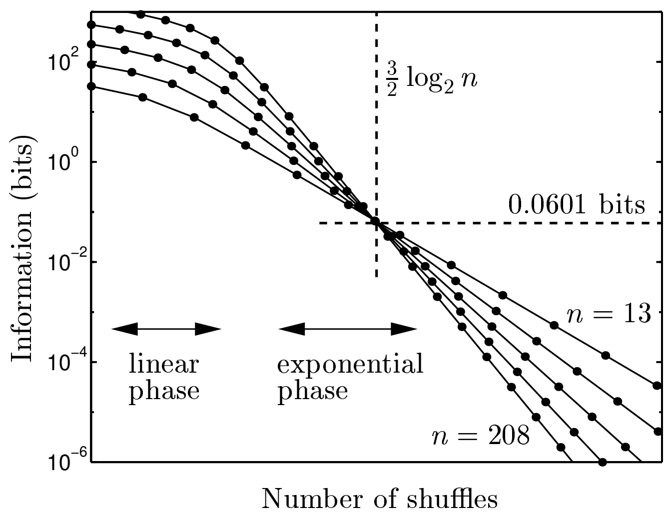

Figure 3. A different view of information for decks of sizes $n = 13, 26, 52, 104, 208$. The vertical scale is now logarithmic, facilitating consideration of the absolute as well as relative amount of information at each step, and the horizontal axis is scaled differently for each n so that $\frac{3}{2}\log_2 n$ always falls at the dashed line in the middle. Randomization is achieved in two phases: linear reduction of I for $\sim\log_2 n$ shuffles (unrelated to the eigenvalues of $P - P^\infty$) followed by exponential reduction forever (determined by the eigenvalues). At $\frac{3}{2}\log_2 n$ shuffles, approximately 0.0601 bits remain, independently of n.

$\frac{1}{4}n\log_e n$, but I decreases steadily from the start at a rate governed by the square of the largest eigenvalue, $1 - 2/(n+1)$, with no preliminary linear phase of convergence. Plots illustrating the absence of a cutoff for this problem in other senses are given in Martin-Löf (1983) and Jónsson & Trefethen (1998).

5. Mathematical context: troublesome eigenvalue problems

In exhibiting a disjunction between transient and asymptotic behavior, the shuffling problem illustrates a mathematical phenomenon that is also important in fluid mechanics, numerical analysis, and other disciplines (Trefethen, *et al.* 1993). In various problems in these fields, the eigenvectors of the matrix or operator that govern a system have no relevance to its transient behavior. For the shuffling problem, for example, let V denote the $n!\times n!$ matrix whose rows are normalized left eigenvectors of $P - P^\infty$. For a given probability distribution p, the vector pV^{-1} then consists of the coefficients of the expansion of p as a linear combination of the eigenvectors of $P - P^\infty$. For $n = 52$, the norm of V^{-1} is at least 10^{40}, indicating that the expansion coefficients may be 10^{40} times larger than p itself, or, in other words, there may be a gap as large as 10^{40} between the behavior of individual eigenmodes and the transient behavior of a vector p. It takes $\sim\log_2 n$ shuffles before this factor is breached and the asymptotic behavior governed by eigenvalues and eigenmodes becomes observable.

6. Discussion

It is not obvious, even to experts, what the full significance is of the distinction between our two measures of randomization, $\|P^k - P^\infty\|_{\mathrm{TV}}$, which shows a cutoff, and I, which does not. To shed some light on this matter, here is perhaps the simplest possible example of a Markov chain with a cutoff. (Generalizations of this chain are analyzed in Diaconis & Graham (1992).) Suppose we start with a word of n bits and modify it at each step by randomizing the last bit, then shifting the word circularly to the left. The information remaining after $k \le n$ steps is $I = n - k$ bits: the decay is exactly linear. The total variation norm, on the other hand, is $\|P^k - P^\infty\|_{\mathrm{TV}} = 1 - 2^{k-n}$: there is a cutoff, with essentially no decay until k gets close to n. (Since convergence is achieved in n steps, $P - P^\infty$ is nilpotent, with all eigenvalues equal to 0 and the largest Jordan block of dimension n.) The explanation of the formula $1 - 2^{k-n}$ is that after step k, $n - k$ bits remain untouched, so a gambler could be guaranteed to win 1 dollar on a bet for which the house, based on the assumption of randomness, would only require him or her to put up 2^{k-n} dollars. This example suggests that the difference between I and $\|P^k - P^\infty\|_{\mathrm{TV}}$ is analogous to the difference in statistics between the *magnitude* of a trend and its *statistical significance.* As a deck of cards is shuffled, the magnitude of the non-randomness decreases steadily from the start, but until $k \sim \frac{3}{2}\log_2 n$, there remains a significant pocket of non-randomness: the deck is biased in the direction of having slightly less than the asymptotically correct number $(n+1)/2$ of rising sequences. (See the theorems of Bayer & Diaconis (1992) and the figures of Jónsson & Trefethen (1998).) The question of which measure of randomization is the more important one for gamblers and card players is presumably game-dependent.

Acknowledgments

We thank David Aldous, Persi Diaconis, Peter Doyle, and Francis Edward Su for comments on a draft of this article; these four all know more about shuffling and cutoffs than we do. Doyle points out that many of our results were known to him four years ago, though not published, and even communicated to us (LNT) in January 1996, but overlooked; we thank him for graciously encouraging us to go forward nonetheless with this publication. Aldous points out that as entropy is subadditive (Aldous 1983), it is obvious that there can be no cutoff in this measure. We thank Gudbjörn Jónsson for many of the ideas that made our computations possible, and LNT thanks Diaconis for advice over the years and for an inspiring course on Markov chains in 1996 at Cornell University. It must be said, however, that not all the views expressed here are necessarily shared by Diaconis. The research of LNT has been supported by the NSF (U.S., Grant DMS-9500975CS) and by the EPSRC (U.K.).

References

Aldous, D. 1983 *Random walk on finite groups and rapidly mixing Markov chains.* Springer Lecture Notes in Mathematics no. 986, pp. 243–297.

Aldous, D. and Diaconis, P. 1986 Shuffling cards and stopping times. Amer. Math. Monthly **93**, 333–348.

Barron, A. R. 1986. Entropy and the central limit theorem. *Ann. Probab.* **14**, 336–342.

Bayer, D. and Diaconis, P. 1992 Trailing the dovetail shuffle to its lair. *Annals Appl. Probab.* **2**, 294–313.

Bingham, N. H. 1991 Fluctuation theory for the Ehrenfest urn. *Adv. Appl. Probab.* **23**, 598-611.

Cover, T. M. and Thomas, J. A. 1991 *Elements of information theory.* New York: Wiley.

Diaconis, P. 1988 *Group representations in probability and statistics.* Hayward, Calif.: IMS.

Diaconis, P. 1996 The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* **93**, 1659–1664.

Diaconis, P. and Graham, R. 1992 An affine walk on the hypercube. *J. Comp. Appl. Math.* **41**, 215–235.

Diaconis, P., McGrath, M. and Pitman, J. W. 1995. Riffle shuffles, cycles and descents. *Combinatorica* **15**, 11-29.

Diaconis, P. and Stroock, D. 1991 Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1**, 11–29.

Feller, W. 1968 *An introduction to probability theory and its applications,* v. 1, 3rd ed. New York: Wiley.

Fill, J. A. 1991 Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an appplication to the exclusion process. *Ann. Appl. Probab.* **1**, 62–87.

Gilbert, E. 1955 *Theory of shuffling.* Technical memorandum, Bell Laboratories.

Jónsson, G. F. 1997 The entries of the compressed transition matrix for riffle shuffle. Manuscript.

Jónsson, G. F. and Trefethen, L. N. 1998 A numerical analyst looks at the 'cutoff phenomenon' in card shuffling and other Markov chains. In *Numerical analysis 1997* (eds D. F. Griffiths, D. J. Higham and G. A. Watson). Harlow, Essex: Addison Wesley Longman, pp. 150–178.

Kac, M. 1959 *Probability and related topics in the physical sciences.* London: Interscience.

Kolata, G. 1990 In shuffling cards, 7 is winning number. *New York Times,* sec. C, p. 1, Jan. 9, 1990.

Kullback, S. 1959 *Information theory and statistics.* New York: Wiley.

Lovász, L. and Winkler, P. 1995 Mixing of random walks and other diffusions on a graph. In *Surveys in combinatorics, 1995* (ed P. Rowlinson). London Math. Soc. Lect. Notes 218 (1995). Cambridge: Cambridge U. Press, pp. 119-154.

Martin-Löf, A. 1983 *Statistical mechanics and the foundations of thermodynamics.* Springer Lecture Notes in Physics, no. 101.

Meyn, S. P. and Tweedie, R. L. 1993 *Markov chains and stochastic stability.* London: Springer.

Norris, J. R. 1997 *Markov chains.* Cambridge: Cambridge U. Press.

Renyi, A. 1970 *Probability theory.* Amsterdam: North-Holland.

Shannon, C. E. and Weaver, W. 1949 *The mathematical theory of communication.* Urbana, Ill.: U. of Illinois Press.

Stewart, G. W. 1997 On Markov chains with sluggish transients. *Communication in Statistics, Stochastic Models* **13**, 35–95.

Su, F. E. 1995 *Methods for quantifying rates of convergence for random walks on groups.* PhD thesis, Harvard University.

Trefethen, L. N. and Bau III, D. 1997 *Numerical linear algebra.* Philadelphia: SIAM.

Trefethen, L. N., Trefethen, A. E., Reddy, S. C. and Driscoll, T. A. 1993 Hydrodynamic stability without eigenvalues. *Science* **261**, 578–584.