How many shuffles to randomize a deck of cards?

Lloyd N. Trefethen* and Lloyd M. Trefethen

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 for large n, 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, there is no cutoff. 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 level in an absolute sense. At $\frac{3}{2}\log_2 n$ shuffles, about 0.0601 bits remain, independently of n.

L. N. Trefethen, Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, UK.
L. M. Trefethen, Department of Mechanical Engineering, Tufts University, Medford, MA 02155, USA.

*To whom correspondence should be addressed. E-mail: LNT@comlab.ox.ac.uk

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 $n \to \infty$ (1,2,3,4). 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* (5). They do not stand in isolation but are part of the rapidly developing subject of the analysis of non-asymptotic convergence of Markov chains, with implications in condensed matter physics, computer science, and other fields (6,7).

Throughout our discussion, a riffle shuffle is defined in a mathematically precise way due to Gilbert and Shannon and independently Reeds (8). 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 idealization of a shuffle is a reasonable approximation to the actual behavior of human shufflers (9).

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 $\| P^k - P^\infty \|_{\rm TV} \to 1$ if $\alpha < 1.5$ and $\| P^k - P^\infty \|_{\rm TV} \to 0$ 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. This much is standard material in the field of Markov chains (10). The total variation norm after step k is defined by the formula

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

where P^k is the kth power of P and P^∞ is the limit of P^k as $k \to \infty$ (11,12). This norm 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 $\| P^k - P^\infty \|_{\rm TV}$ 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.

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 Fisher, Shannon, and Wiener (13,14),

$$U \;=\; -\sum_{i=1}^{n!} p_i \log_2 p_i . \qquad (2)$$

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

$$I \;=\; \log_2(n!) - U. \tag{3}$$

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 by methods based on explicit manipulation of $n \times n$ matrices adapted from earlier joint work with Jónsson (Table 1) (15,16,17).

Figure 2 shows results for both $n = 52$ and the limit $n \to \infty$. The first shuffle reduces I by almost exactly n bits (18). 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 $k \geq 1$ (11), and the asymptotic convergence rate $\frac{1}{4}$ is equal to the square of the largest eigenvalue of this matrix, $\frac{1}{2}$ (19,20,21,22).

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. 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 $k \leq n$ 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 (23). 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 pocket of non-randomness that is significant (24). The question of which measure of randomization is the more important one for card players is presumably game-dependent.

How many riffle shuffles, then, to randomize a deck? The best answer from the point of view of information theory seems to be $\sim \log_2 n$ as $n \to \infty$, or for $n = 52$, 5 shuffles.

REFERENCES AND NOTES

1. The symbol $\sim$ is defined by ratios: $f(n) \sim g(n)$ as $n \to \infty$ if $f(n)/g(n) \to 1$.
2. D. Aldous, *Random Walk on Finite Groups and Rapidly Mixing Markov Chains* (Lecture Notes in Math. v. **986**, Springer, Berlin, 1983), p. 243–297.
3. D. Aldous and P. Diaconis, *Amer. Math. Monthly* **93**, 933 (1986).
4. D. Bayer and P. Diaconis, *Annals Appl. Prob.* **2**, 294 (1992).
5. G. Kolata, In shuffling cards, 7 is winning number, *New York Times,* sec. C, p. 1, Jan. 9, 1990.
6. P. Diaconis, *Proc. Natl. Acad. Sci. USA* **93**, 1659 (1996).
7. F. E. Su, *Methods for Quantifying Rates of Convergence for Random Walks on Groups* (PhD diss., Harvard U., 1995).
8. E. Gilbert, *Theory of shuffling,* tech. memo., Bell Laboratories, 1955; J. Reeds, unpublished manuscript, 1981.
9. P. Diaconis, *Group Representations in Probability and Statistics* (IMS, Hayward, Calif., 1988).
10. W. Feller, *An Introduction to Probability Theory and its Applications,* v. 1, 3rd ed. (Wiley, New York, 1968); S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability* (Springer, London, 1993).
11. G. F. Jónsson and L. N. Trefethen, A numerical analyst looks at the "cutoff phenomenon" in card shuffling and other Markov chains, in D. F. Griffiths, D. J. Higham, and G. A. Watson, Eds., *Numerical Analysis 1997* (Addison Wesley Longman, Harlow, Essex, UK, 1998), pp. 150–178.
12. Formula (1) represents half the 1-norm of the matrix $P^k - P^\infty$ when viewed as acting on row vectors. See e.g. L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra* (SIAM, Philadelphia, 1997).
13. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (U. Illinois, Urbana, Ill., 1949); T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, New

York, 1991).

14. S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

15. Though the matrix P is of the computationally intractable dimension $n!$, 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 (4), and the matrix entries have been worked out explicitly in (11) and in G. F. Jónsson, The entries of the compressed transition matrix for riffle shuffle, manuscript. Our computations could have been based on Theorem 1 of (4), but we used the matrices instead. Copies of our Matlab programs, about 100 lines in total, can be obtained from the first author.

16. All the results presented in this report are numerical, though several of them suggest theorems that presumably could be proved.

17. An analysis of alternatives to total variation for various Markov chain problems is presented in (7), and our I is essentially the relative entropy distance considered there. A standard reference on the use of such measures in statistics is (14). Uncertainties have apparently not been calculated previously for the riffle shuffling problem, however.

18. For $n = 52$, the first shuffle destroys about $52 - 7 \times 10^{-14}$ bits. An elementary argument explains why the first shuffle destroys about n bits: according to the inverse interpretation of riffle shuffling described in (4), one shuffle is equivalent to separating the n cards into two subsequences at random, then concatenating the subsequences, a process that comes very close to moving the deck to one of 2^n possible configurations with equal probability.

19. See Corollary 5.3 of (17).

20. In exhibiting a disjunction between transient and asymptotic behavior, the shuffling problem exemplifies a mathematical phenomenon that is also important in fluid mechanics, numerical analysis, and other disciplines; see L. N. Trefethen *et al., Science* **261**, 578 (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 about 10^{52}, indicating that the expansion coefficient may be 10^{52} times larger than p itself, or in other words, there may be a gap as large as 10^{52} 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.

21. For a different view of the gap between eigenvalues and convergence for Markov chains see G. W. Stewart, *Communication in Statistics, Stochastic Models*, to appear.

22. 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 (6,11); closely related theorems are reported in (7). The cutoff of $\| P^k - P^\infty \|_{\mathrm{TV}}$ for this problem is at $\sim \frac{1}{4} \log_e n$, 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 (11) and in A. Martin-Löf, *Statistical Mechanics and the Foundations of Thermodynamics* (Lecture Notes in Phys. v. **101**, Springer, Berlin, 1983).

23. 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.

24. That nature of the pocket of non-randomness is understood. Until $k \sim \frac{3}{2} \log_2 n$, the shuffled 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 (4) and the figures of (11).

25. We thank Gudbjörn Jónsson for developing the formulas that this work is based upon and Persi Diaconis for advice on various points and for an inspiring course on Markov chains taught 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 the first author was supported by NSF Grant DMS-9500975CS.

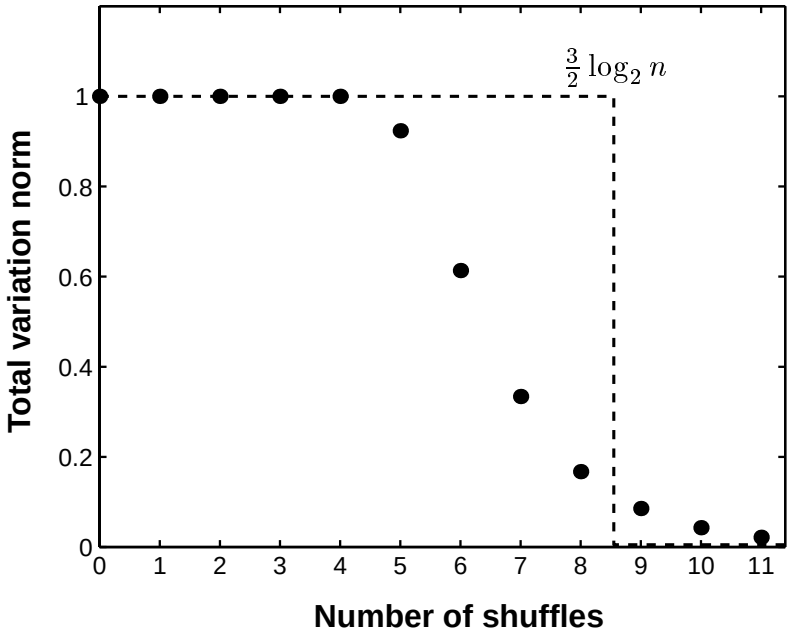

Fig. 1. Randomization of a deck of n cards as measured in the total variation norm $\|P^k - P^\infty\|_{\rm TV}$ of Aldous, Bayer and Diaconis. The dots and the numerical axis labels correspond to $n = 52$ and the dashed line to the limit $n \to \infty$. In this limit a "cutoff phenomenon" occurs, with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

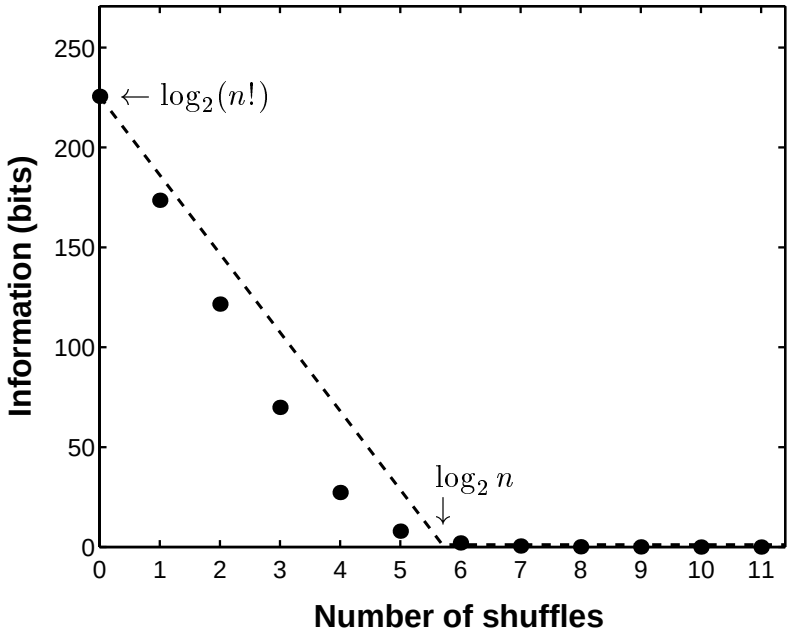

Fig. 2. Randomization as measured by reduction of information from $\log_2(n!)$ to 0 bits. Again, the dots and the numerical axis labels correspond to $n = 52$ and the dashed line to $n \to \infty$. In this measure there is no cutoff effect, and randomization in the sense of reduction of the original 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.

Table 1. Information I (bits) remaining in an initially ordered deck of 52 cards after 0,1,. . .,10 riffle shuffles.

shuffle no.	information
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

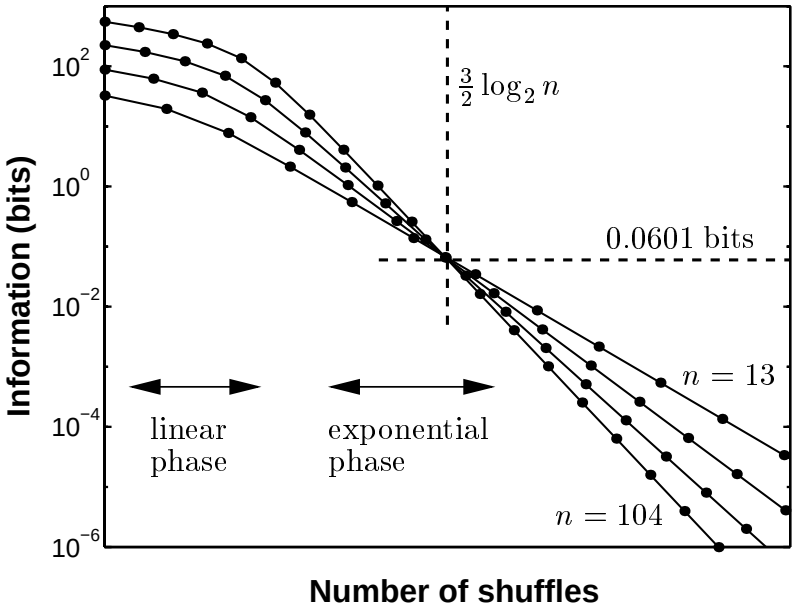

Fig. 3. A different view of information for decks of sizes $n = 13, 26, 52, 104$. The vertical scale is now logarithmic, facilitating consideration of the absolute as well as relative amount of information at each step, and the horizontal axis is scaled differently for each n so that $\frac{3}{2}\log_2 n$ 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.