Shuffling Cards and Stopping Times

David Aldous; Persi Diaconis

The American Mathematical Monthly, Vol. 93, No. 5. (May, 1986), pp. 333-348.

Stable URL:
http://links.jstor.org/sici?sici=0002-9890%28198605%2993%3A5%3C333%3ASCAST%3E2.0.CO%3B2-9

The American Mathematical Monthly is currently published by Mathematical Association of America.

SHUFFLING CARDS AND STOPPING TIMES

DAVID ALDOUS*
Department of Statistics, University of California, Berkeley, CA 94720

PERSI DIACONIS**
Department of Statistics, Stanford University, Stanford, CA 94305

1. Introduction. How many times must a deck of cards be shuffled until it is close to random? There is an elementary technique which often yields sharp estimates in such problems. The method is best understood through a simple example.

EXAMPLE 1. *Top in at random shuffle.* Consider the following method of mixing a deck of cards: the top card is removed and inserted into the deck at a random position. This procedure is repeated a number of times. The following argument should convince the reader that about $n \log n$ shuffles suffice to mix up n cards. The argument depends on following the bottom card of the deck. This card stays at the bottom until the first time (T_1) a card is inserted below it. Standard calculations, reviewed below, imply this takes about n shuffles. As the shuffles continue, eventually a second card is inserted below the original bottom card (this takes about $n/2$ further shuffles). Consider the instant (T_2) that a second card is inserted below the original bottom card. The two cards under the original bottom card are equally likely to be in relative order low-high or high-low.

Similarly, the first time a third card is inserted below the original bottom card, each of the 6 possible orders of the 3 bottom cards is equally likely. Now consider the first time T_{n-1} that the original bottom card comes up to the top. By an inductive argument, all $(n-1)!$ arrangements of the lower cards are equally likely. When the original bottom card is inserted at random, at time $T = T_{n-1} + 1$, then all $n!$ possible arrangements of the deck are equally likely.

	1	2	3	4	5	6	7	8	9
a	*b*	*c*	*c*	*a*	*b*	*a*	*a*	*d*	*c*
b	*c*	*a*	*a*	*b*	*a*	*d*	*d*	*c*	*d*
c	*a*	*b*	*b*	*d*	*d*	*c*	*c*	*a*	*a*
d	*d*	*d*	*d*	*c*	*c*	*b*	*b*	*b*	*b*
				T_1		T_2		T_3	T

FIG. 1. Example of repeated top in at random shuffles of a 4-card deck.

When the original bottom card is at position k from the bottom, the waiting time for a new card to be inserted below it is about n/k. Thus the waiting time T for the bottom card to come to

---

David Aldous confesses to a conventional career, going from a Ph.D. and Research Fellowship at Cambridge University to the Statistics Department at Berkeley, where he is now Associate Professor. He does research in theoretical and applied probability theory, and for recreation he plays volleyball (well), bridge (badly) and watches Monty Python reruns.

Persi Diaconis left High School at an early age to earn a living as a magician and gambler, only later to become interested in mathematics and earn a Ph.D. at Harvard. After a spell at Bell Labs, he is now Professor in the Statistics Department at Stanford. He was an early recipient of a MacArthur Foundation award, and his wide range of mathematical interests is partly reflected in his first book *Group Theory in Statistics*. He retains an interest in magic and the exposure of fraudulent psychics.

*Research supported by National Science Foundation Grant MCS80-02698.
**Research supported by National Science Foundation Grant MCS80-24649.

the top and be inserted is about

$$n + \frac{n}{2} + \frac{n}{3} + \cdots + \frac{n}{n} \doteq n \log n.$$

This paper presents a rigorous version of the argument and illustrates its use in a variety of random walk problems. The next section introduces the basic mathematical set up. Section 3 details a number of examples drawn from applications such as computer generated pseudo random numbers. Section 4 treats ordinary riffle shuffling, analyzing a model introduced by Gilbert, Shannon, and Reeds. Section 5 explains a sense in which the method of stopping times always works and compares this to two other techniques (Fourier analysis and coupling). Some open problems are listed.

2. The Basic Set-Up. Repeated shuffling is best treated as random walk on the permutation group S_n. For later applications, we treat an arbitrary finite group G. Given some scheme for randomly picking elements of G, let $Q(g)$ be the probability that g is picked. The numbers $\{Q(g): g \in G\}$ are a (probability) *distribution*: $Q(g) \geqslant 0$ and $\Sigma Q(g) = 1$. Repeated independent picks according to the same scheme yield random elements $\xi_1, \xi_2, \xi_3, \ldots$, of G. Define the products

$$\begin{aligned} X_0 &= \text{identity} \\ X_1 &= \xi_1 \\ &\vdots \\ X_k &= \xi_k X_{k-1} = \xi_k \xi_{k-1} \cdots \xi_1. \end{aligned}$$

The random variables $X_0, X_1, X_2, \ldots$, are the *random walk* on G with *step distribution* Q. Think of X_k as the position at time k of a randomly-moving particle. The distribution of X_2, that is the set of probabilities $P(X_2 = g)$, $g \in G$, is given by convolution

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

For $Q(h)Q(gh^{-1})$ is the chance that element h was picked first and gh^{-1} was picked second; for any h, this makes the product equal to g. Similarly, $P(X_k = g) = Q^{k*}(g)$, where Q^{k*} is the repeated convolution

$$Q^{k*} = Q^*Q^{(k-1)*} = \sum_{h \in G} Q(h) Q^{(k-1)*}(gh^{-1}). \tag{2.1}$$

In modelling shuffling of an n-card deck, the state of the deck is represented as a permutation $\pi \in S_n$, meaning that the card originally at position i is now at position $\pi(i)$.

In Example 1, $G = S_n$, and using cycle notation for permutations π,

$$\begin{aligned} Q(i, i-1, \ldots, 1) &= 1/n, \quad 1 \leqslant i \leqslant n, \\ Q(\pi) &= 0, \text{ else.} \end{aligned}$$

Here ξ_k is a randomly chosen cycle, X_k is the state of the deck after k shuffles, and $Q^{k*}(\pi)$ is the chance that the state after k shuffles is π. In Fig. 1, $\xi_1 = (3,2,1)$, $\xi_2 = (3,2,1)$ and $X_2 = \xi_1^*\xi_2 = (1,2,3)$.

We shall study the distribution Q^{k*}. Note that Q^{k*} can be defined by (2.1) without using the richer structure of the random walk (X_k); however, this richer structure is essential for our method of analysis.

A fundamental result is that repeated convolutions converge to the uniform distribution U:

$$Q^{k*}(g) \to U(g) = 1/|G| \quad \text{as } k \to \infty, \tag{2.2}$$

unless Q is concentrated on some coset of some subgroup. This was first proved by Markov (1906)—see Feller (1968), Section 15.10 for a clear discussion—and can nowadays be regarded as a special case of the basic limit theory of finite Markov chains. Poincaré (1912) gave a Fourier

analytic proof, and subsequent workers have extended (2.2) to general compact groups—see Grenander (1963), Heyer (1977), Diaconis (1982) for surveys. A version of this result is given here as Theorem 3 of Section 3.

Despite this work on abstracting the asymptotic result (2.2), little attention has been paid until recently to the kind of non-asymptotic questions which are the subject of this paper.

A natural way to measure the difference between two probability distributions Q_1, Q_2 on G is by *variation distance*

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

There are equivalent definitions

$$\|Q_1 - Q_2\| = \max_{A \subset G}|Q_1(A) - Q_2(A)| = \frac{1}{2}\max_{\|f\|=1}|Q_1(f) - Q_2(f)|, \tag{2.3}$$

where $Q(A) = \Sigma_{g \in A}Q(g)$, $Q(f) = \Sigma f(g)Q(g)$, and $\|f\| = \max|f(g)|$. The string of equalities is proved by noting that the maxima occur for $A = \{g: Q_1(g) > Q_2(g)\}$ and for $f = 1_A - 1_{\bar{A}}$. Thus, two distributions are close in variation distance if and only if they are uniformly close on all subsets. Plainly $0 \leqslant \|Q_1 - Q_2\| \leqslant 1$.

An example may be useful. Suppose, after well-shuffling a deck of n cards, that you happen to see the bottom card, c. Then your distribution Q on S_n is uniform on the set of permutations π for which $\pi(c) = n$, and $\|Q - U\| = 1 - 1/n$. This shows the variation distance can be very "unforgiving" of small deviations from uniformity.

Given a distribution Q on a group G, (2.2) says

$$d_Q(k) \overset{\text{def}}{\equiv} \|Q^{k*} - U\| \to 0 \quad \text{as } k \to \infty. \tag{2.4}$$

Where Q models a random shuffle, $d(k)$ measures how close k repeated shuffles get the deck to being perfectly (uniformly) shuffled. One might suppose $d(k)$ decreases smoothly from (near) 1 to 0; and it is not hard to show $d(k)$ is decreasing. However,

THEOREM 1. *For the "top in at random" shuffle*, Example 1,
(a) $d(n \log n + cn) \leqslant e^{-c}$; *all* $c \geqslant 0$, $n \geqslant 2$.
(b) $d(n \log n - c_n n) \to 1$ *as* $n \to \infty$; *all* $c_n \to \infty$.

This gives a sharp sense to the assertion that $n \log n$ shuffles are enough. This is a particular case of a general *cut-off phenomenon*, which occurs in all shuffling models we have been able to analyze; there is a critical number k_0 of shuffles such that $d(k_0 + o(k_0)) \approx 0$ but $d(k_0 - o(k_0)) \approx 1$. (See Fig. 2.)

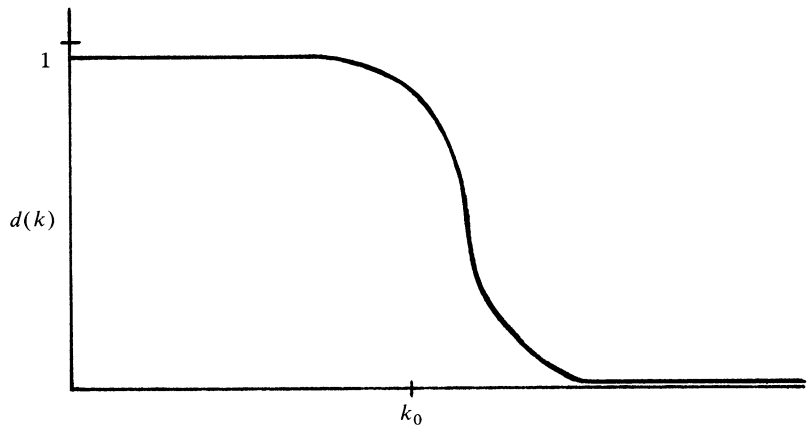


FIG. 2

Our aim is to determine k_0 in particular cases. This is quite different from looking at asymptotics in (2.4): it is elementary that $d(k) \to 0$ geometrically fast, and Perron-Frobenius theory says $d(k) \sim a\lambda^k$, where a, λ have eigenvalue/eigenvector interpretation, but these asymptotics miss the cut-off phenomenon. For card players, the question is not "exactly how close to uniform is the deck after a million riffle shuffles?", but "is 7 shuffles enough?".

The main purpose of this paper is to show how upper bounds on $d(k)$, like (a) in Theorem 1, can be obtained using the notion of strong uniform times, which we now define in two steps.

DEFINITION 1. Let G be a finite group, and G^∞ the set of all G-valued infinite sequences $\mathbf{g} = (g_1, g_2, \ldots)$. A *stopping rule* $\hat{T}$ is a function $\hat{T}\colon G^\infty \to \{1, 2, 3, \ldots; \infty\}$ such that if $\hat{T}(\mathbf{g}) = j$, then $\hat{T}(\hat{\mathbf{g}}) = j$ for all $\hat{\mathbf{g}}$ with $\hat{g}_i = g_i$, $i \leqslant j$.

DEFINITION 2. Let Q be a distribution on G, and let (X_k) be the associated random walk. Given a stopping rule $\hat{T}$, the random time $T = \hat{T}(X_1, X_2, \ldots)$ is a *stopping time*. Call T a strong uniform time (for U) if for each $k < \infty$

(a) $P(T = k,\ X_k = g)$ does not depend on g.

REMARK (i). Note that (a) is equivalent to

(b) $P(X_k = g | T = k) = 1/|G|,\ g \in G$

and to

(c) $P(X_k = g | T \leqslant k) = 1/|G|;\ g \in G$.

REMARK (ii). Picture the process of picking group elements and multiplying. A stopping time is a rule which tells you when to "stop" with the current value of the product. The time is strong uniform if, conditional on stopping after k steps, the value of the product is uniform on G.

REMARK (iii). In Example 1, we defined a time T as the first time that the original bottom card has come to the top and been inserted into the deck. This is certainly a stopping time, and the inductive argument in Section 1 shows that, given $T = k$, all arrangements of the deck are equally likely.

REMARK (iv). In practice it is often useful to have a slightly more general notion of stopping time, which allows the decision on whether or not to stop at n to depend not only on $(X_1, \ldots, X_n)$ but also on the value of some random quantity Y independent of the X process. Such a time T is called a *randomized* stopping time T; our results extend to this case without essential change.

Here is a basic upper bound lemma which relates strong uniform times to the distance between Q^{k*} and the uniform distribution.

LEMMA 1. *Let Q be a probability distribution on a finite group G. Let T be a strong uniform time for Q. Then*

$$d(k) \equiv \|Q^{k*} - U\| \leqslant P(T > k), \quad \text{all } k \geqslant 0.$$

Proof. For any $A \subset G$

$$\begin{aligned} Q^{k*}(A) &= P(X_k \in A) \\ &= \sum_{j \leqslant k} P(X_k \in A, T = j) + P(X_k \in A, T > k) \\ &= \sum_{j \leqslant k} U(A) P(T = j) + P(X_k \in A | T > k) P(T > k) \\ &= U(A) + \{P(X_k \in A | T > k) - U(A)\} P(T > k) \end{aligned}$$

and so

$$|Q^{k*}(A) - U(A)| \leqslant P(T > k). \quad \square$$

We conclude this section by using Lemma 1 and elementary probability concepts to prove Theorem 1. Here is one elementary result we shall use in several examples.

LEMMA 2. *Sample uniformly with replacement from an urn with n balls. Let V be the number of draws required until each ball has been drawn at least once. Then*

$$P(V > n \log n + cn) \leqslant e^{-c}; \quad c \geqslant 0, n \geqslant 1.$$

Proof. Let $m = n \log n + cn$. For each ball b let A_b be the event "ball b not drawn in the first m draws". Then

$$P(V > m) = P\Big(\bigcup_b A_b\Big) \leqslant \sum_b P(A_b) = n\Big(1 - \frac{1}{n}\Big)^m$$
$$\leqslant n \exp(-m/n) = e^{-c}. \quad \square$$

REMARK. This is the famous "coupon-collector's problem", discussed in Feller (1968). The asymptotics are $P(V > n \log n + cn) \to 1 - \exp(-e^{-c})$ as $n \to \infty$, c fixed. So for c not small the bound in Lemma 2 is close to sharp.

Proof of Theorem 1. Recall we have argued that T, the first time that the original bottom card has come to the top and been inserted into the deck, is a strong uniform time for this shuffling scheme. We shall prove that T has the same distribution as V in Lemma 2; then assertion (a) is a consequence of Lemmas 1 and 2.

We can write

$$T = T_1 + (T_2 - T_1) + \cdots + (T_{n-1} - T_{n-2}) + (T - T_{n-1}), \tag{2.5}$$

where T_i is the time until the ith card is placed under the original bottom card. When exactly i cards are under the original bottom card b, the chance that the current top card is inserted below b is $\dfrac{i+1}{n}$, and hence the random variable $T_{i+1} - T_i$ has geometric distribution

$$P(T_{i+1} - T_i = j) = \frac{i+1}{n}\Big(1 - \frac{i+1}{n}\Big)^{j-1}; \quad j \geqslant 1. \tag{2.6}$$

The random variable V in Lemma 2 can be written as

$$V = (V - V_{n-1}) + (V_{n-1} - V_{n-2}) + \cdots + (V_2 - V_1) + V_1, \tag{2.7}$$

where V_i is the number of draws required until i distinct balls have been drawn at least once. After i distinct balls have been drawn, the chance that a draw produces a not-previously-drawn ball is $\dfrac{n-i}{n}$. So $V_i - V_{i-1}$ has distribution

$$P(V_i - V_{i-1} = j) = \frac{n-i}{n}\Big(1 - \frac{n-i}{n}\Big)^{j-1}; \quad j \geqslant 1.$$

Comparing with (2.6), we see that corresponding terms $(T_{i+1} - T_i)$ and $(V_{n-i} - V_{n-i-1})$ have the same distribution; since the summands within each of (2.5) and (2.7) are independent, it follows that the sums T and V have the same distribution, as required.

To prove (b), fix j and let A_j be the set of configurations of the deck such that the bottom j original cards remain in their original relative order. Plainly $U(A_j) = 1/j!$ Let $k = k(n)$ be of the form $n \log n - c_n n$, $c_n \to \infty$. We shall show

$$Q^{k^*}(A_j) \to 1 \quad \text{as } n \to \infty; \quad j \text{ fixed}. \tag{2.8}$$

Then $d(k) \geqslant \max_j \{Q^{k^*}(A_j) - U(A_j)\} \to 1$ as $n \to \infty$, establishing part (b).

To prove (2.8), observe that $Q^{k^*}(A_j) \geqslant P(T - T_{j-1} > k)$. For $T - T_{j-1}$ is distributed as the

time for the card initially jth from bottom to come to the top and be inserted; and if this has not occurred by time k, then the original bottom j cards must still be in their original relative order at time k. Thus it suffices to show

$$P(T - T_{j-1} \leqslant k) \to 0 \quad \text{as } n \to \infty; \quad j \text{ fixed.} \tag{2.9}$$

We shall prove this using *Chebyshev's inequality*:

$$P(|Z - EZ| \geqslant a) \leqslant \frac{\operatorname{var}(Z)}{a^2}, \quad \text{where } a \geqslant 0, \text{ and } Z \text{ is any random variable.}$$

From (2.6),

$$E(T_{i+1} - T_i) = \frac{n}{i+1}, \quad \operatorname{var}(T_{i+1} - T_i) = \left(\frac{n}{i+1}\right)^2\left(1 - \frac{i+1}{n}\right),$$

and so from (2.5)

$$E(T - T_j) = \sum_{i=j}^{n-1} \frac{n}{i+1} = n \log n + O(n),$$

$$\operatorname{var}(T - T_j) = \sum_{i=j}^{n-1} \left(\frac{n}{i+1}\right)^2\left(1 - \frac{i+1}{n}\right) = O(n^2),$$

and Chebyshev's inequality applied to $Z = T - T_{j-1}$ readily yields (2.9). □

REMARK. Note that the "strong uniform time" property of T played no role in establishing the *lower* bound (b). Essentially, we get lower bounds by guessing some set A for which $|Q^{k*}(A) - U(A)|$ should be large, and using the obvious (from (2.3)) inequality

$$d(k) = \|Q^{k*} - U\| \geqslant |Q^{k*}(A) - U(A)|.$$

3. Examples. We present constructions of strong uniform times for a variety of random walks: simple random walk on the circle, general random walks on finite groups, and a random walk arising in random number generation. Sometimes our arguments give the optimal rate, often they give the correct order of magnitude.

EXAMPLE 2. *Simple random walk on the integers* mod n. Let n be a positive odd integer. Let Z_n be the integers mod n, thought of as n points on a circle. Imagine a particle which moves by steps, each step being equally likely to move 1 to the right or 1 to the left. This random walk has step distribution Q on Z_n;

$$Q(1) = Q(-1) = \frac{1}{2}. \tag{3.1}$$

The following theorem shows that the number of steps k required to become uniform is slightly more than n^2.

THEOREM 2. *Let $n \geqslant 3$ be an odd integer. For simple random walk on the integers* mod n *defined by* (3.1), *for $k \geqslant n^2$,*

$$d(k) \leqslant 6e^{-ak/n^2}$$

with $a = 4\pi^2/3$.

Proof. First consider $n = 5$ and the following 5 patterns

$$++--, +---, -+++, ++++, ----.$$

A sequence of successive steps of the walk on Z_5 yields a sequence of $\pm$ symbols. Consider the sequence in disjoint blocks of 4. Stop the first time T that a block of 4 equals one of the above 5 patterns. Thus, if the sequence starts $++-+, +++-, ++--$, $T = 12$.

This stopping time is clearly a strong uniform time; given that $T = 12$, all 5 final positions in Z_5 are equally likely. Such sets of k-tuples can be chosen for any odd n. It turns out that to get the correct rate of convergence, k should be chosen as a large multiple of n^2. Here are some details.

For fixed integers n and k, with n odd, let B_j be the set of binary k-tuples with j pluses (mod n).

Let j^* be the index corresponding to the smallest $|B_{j^*}|$. Partition the set of binary k-tuples into n groups of size $|B_{j^*}|$, the jth group being chosen arbitrarily from B_j. The random walk generates a sequence of $\pm$ symbols. Consider these in disjoint blocks of length k. Define T as the first time a block equals one of the chosen group. This clearly yields a strong uniform time. The following lemma gives an explicit upper bound for $d(k)$.

LEMMA 3. *Let T be as defined above. For $n \geqslant 3$ and $k \geqslant n^2$,*

$$P(T > k) \leqslant 6e^{-ak/n^2}$$

with $a = 4\pi^2/3$.

Proof. The number of elements in B_j is

$$\sum_{l \geqslant 0} \binom{k}{ln + j} = \frac{2^k}{n} \sum_{l=0}^{n-1} e^{\frac{-2\pi i l j^*}{n}} \left(\cos \frac{2\pi l}{n}\right)^k,$$

this being a classical identity due to C. Ramus (see Knuth (1973, p. 70)). The chance of a given block falling in the chosen group equals

$$p^{\text{def}} = \frac{n}{2^k}|B_{j^*}| = \sum_{l=0}^{n-1} e^{\frac{-2\pi i l j}{n}} \left(\cos \frac{2\pi l}{n}\right)^k.$$

Now

$$P(T > k) = p(1-p) + p(1-p)^2 + p(1-p)^3 + \cdots = 1 - p \leqslant \sum_{l=1}^{n-1} \left|\cos \frac{2\pi l}{n}\right|^k.$$

Straightforward calculus using quadratic approximations to cosine such as $\cos x \leqslant 1 - \frac{x^2}{3} \leqslant e^{-x^2/3}$ for $0 \leqslant x \leqslant \pi/2$ leads to the stated result. Further details may be found in Chung, Diaconis, and Graham (1986). □

REMARK. There is a lower bound for $d(k)$ of the form $\alpha e^{-\beta k/n^2}$ for positive α and β, so somewhat more than n^2 steps really are required. One way to prove this is to use the central limit theorem; this implies that after k steps the walk has moved a net distance of order $k^{1/2}$. Hence we need k of order n^2 at least in order that the distribution after k steps is close to uniform. Further details are in Chung, Diaconis and Graham (1986).

There is a sense in which the cutoff phenomenon does not occur for this example. It is possible to show there is a continuous function $d^*(t)$, with $d^*(t) \to 0$ as $t \to \infty$, such that for simple random walk on Z_n,

$$\max_k |d(k) - d^*(k/n^2)| \to 0 \quad \text{as } n \to \infty.$$

Indeed, as $n \to \infty$, a rescaled version of the random walk tends to Brownian motion on the circle. The function $d^*(t)$ is the variation distance to uniformity for Brownian motion at time t.

EXAMPLE 3. *A bound for general problems.* Let G be a finite group and Q a probability on G. The following result shows that Q^{*k} converges to the uniform distribution geometrically fast provided Q is not concentrated on a subgroup or a translate of a subgroup. To see the need for this condition, consider Example 2 above (simple random walk on Z_n). If n is even, then the

particle is at an even position after an even number of steps—the distribution never converges to uniform.

A simple way to force convergence is the following:

(3.2) Suppose for some k_0 and $0 < c < 1$, $Q^{*k_0}(g) \geq cU(g)$ for all $g \in G$.

THEOREM 3. *Condition* (3.2) *implies*

$$d(k) \leq (1 - c)^{[k/k_0]} \quad \text{for all } k \geq k_0.$$

Proof. The argument proceeds by constructing another process which behaves like the original random walk but easily exhibits a strong uniform time. Suppose first that $k_0 = 1$, so $Q(g) \geq cU(g)$ for all g. Define

$$R(g) = [Q(g) - cU(g)]/[1 - c].$$

Observe that $R(g)$ is a probability and

$$Q(g) = (1 - c)R(g) + cU(g). \tag{3.3}$$

Consider a new random walk defined as follows. For each step, flip a coin with probability of heads c. If the coin comes up heads, take the step according to $U(g)$. If the coin comes up tails, take the step according to $R(g)$. Because of (3.3), each step is taken according to Q overall. Let T be the first time that the coin comes up heads. Then T is a (randomized) stopping time and because the convolution of the uniform distribution with any distribution is uniform, T is a strong uniform time.

Clearly,

$$P\{T > k\} = (1 - c)^k.$$

For $k_0 > 1$, apply the argument to the probability Q^{*k_0}. □

REMARK (i). The argument given is valid for a probability on a general compact group. In this form, Theorem 3 is due to Kloss (1959). The proof we give is very close to techniques exploited by Athreya and Ney (1977) for general state space Markov processes.

REMARK (ii). While Theorem 3 seems quantitative, the simplicity of the argument should make one suspicious. The reader can see the difficulty by trying to get a rate of convergence for simple random walk on Z_n. Estimating c and k_0 is not an easy problem, we do not know how to use Theorem 3 to get the correct rate of convergence for any non-trivial problem.

EXAMPLE 4. *A random walk on Z_n arising in random number generation.* Random number generators often work by recursively computing $X_{k+1} = aX_k + b$ (modulo n), where a, b and n are chosen carefully—see Knuth (1981). Of course the sequence X_k is really deterministic and exhibits many regularities. To improve things, various schemes have been suggested involving combining several generators. In one scheme, a and b are chosen each step from another generator. If this second source is considered truly random (it may be the result of a physical generator using a radioactive source) one may inquire how long it takes X_k to become random.

For example, if $a = 1$ and $b = 0, +1$, or -1 each with probability $1/3$, the process becomes simple random walk on Z_n: $X_k = X_{k-1} + b_k (\text{mod } n)$ with a slightly different step size than considered in Example 2. The argument given there can easily be adapted to show that slightly more than n^2 steps are required to become random.

We now consider the effect of a deterministic doubling:

$$X_k = 2X_{k-1} + b_k (\text{mod } n), \quad b_k = 0, \pm 1 \text{ with probability } \frac{1}{3}. \tag{3.4}$$

We will show that this dramatically speeds things up: from n^2 down to $\log n \log\log n$. The argument is presented as a non-trivial illustration of the method of strong uniform times. It

involves a novel construction of an almost uniform time. For simplicity, we take $n = 2^l - 1$ (a common choice in the application).

THEOREM 4. *Let Q_k be the probability distribution of X_k defined by* (3.4) *with $n = 2^l - 1$. Let $d(k) = \|Q_k - U\|$. Then*

$$d(cl \log l) \to 0 \quad \text{as } l \to \infty, \quad \text{for } c > \frac{1}{\log 3}.$$

Proof. Observe first that if δ_i takes values ± 1 with probability $1/2$, then

$$U^* = 2^{l-1}\delta_1 + 2^{l-2}\delta_2 + \cdots + \delta_l$$

is very close to uniformly distributed mod $2^l - 1$. Indeed,

$$P\big(U^* = j(\bmod 2^l - 1)\big) = \begin{cases} \dfrac{1}{2^l}, & j \neq 0, \\ \dfrac{2}{2^l}, & j = 0. \end{cases}$$

Thus

$$\|U^* - U\| = \frac{2}{2^{l-1}} - \frac{1}{2^l - 1}. \tag{3.5}$$

The argument proceeds by finding a stopping time T such that the process stopped at time T has distribution at least as close to uniform as U^*. An appropriate modification of the upper bound lemma will complete the proof. We isolate the steps as a sequence of lemmas. The first and second lemmas are elementary with proofs omitted.

LEMMA 4. *Let $X_1, X_2, \ldots$ be a process with values in a finite group G. Write Q_k for the probability distribution of X_k. Let T be a stopping time with the property that for some $\varepsilon > 0$,*

$$\|Q_k(\cdot | T = j) - U\| \leqslant \varepsilon; \quad \text{all } j \leqslant k.$$

Then

$$\|Q_k - U\| \leqslant \varepsilon + P(T > k).$$

LEMMA 5. *Let Q_1 and Q_2 be probability distributions on a finite group G. Then*

$$\|Q_1^* Q_2 - U\| \leqslant \|Q_1 - U\|.$$

To state the third lemma, an appropriate stopping time T must be defined. Using the defining recurrence $X_k = 2X_{k-1} + b_k (\bmod n)$,

$$X_k = 2^{k-1}b_1 + 2^{k-2}b_2 + \cdots + b_k (\bmod n). \tag{3.6}$$

Since $n = 2^l - 1$, $2^l = 1 (\bmod n)$. Group the terms on the right side of (3.6) by distinct powers of 2:

$$X_k = 2^{l-1}A_1 + 2^{l-2}A_2 + \cdots + A_l (\bmod n)$$

with

$$A_1 = b_1 + b_{l+1} + b_{2l+1} \ldots, \quad A_2 = b_2 + b_{l+2} + \cdots, \text{ etc.}$$

Define T as the first time each of the sums $A_1, A_2, \ldots, A_l$ contains at least one non-zero summand.

LEMMA 6. *The probability distribution of X_k given $T = j < k$ is the convolution of U^* defined above with an independent random variable.*

Proof. Let δ_i^* be the first non-zero summand in A_i. Write

$$X_k = \left[2^{l-1}\delta_1^* + 2^{l-2}\delta_2^* + \cdots + \delta_l^*\right] + \left[2^{l-1}(A_1 - \delta_1^*) + \cdots + (A_l - \delta_l^*)\right].$$

Clearly the first term on the right has distribution U^*. Further, given all the remaining values of b_k, and the labels of δ_i^*, all 2^l values of $\delta_1^*, \ldots, \delta_l^*$ are equally likely, so the decomposition of X_k is into independent parts. □

Using Lemmas 5, 6, along with the bound (3.5) allows us to take $\varepsilon = 2/2^l$ in Lemma 4 for this stopping time T. To complete the proof of Theorem 4, it only remains to estimate $P(T > k)$.

Toward this end, consider $k = al$ for integer a,

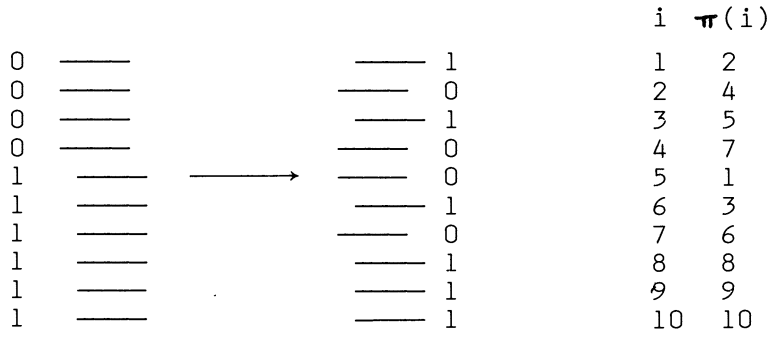
$$P(T > al) = 1 - \left(1 - \left(\frac{1}{3}\right)^a\right)^l.$$

For large l, this is approximately $1 - \exp\{-le^{-a\log 3}\}$. If $a = \dfrac{\log l + c}{\log 3}$ for some value of c, this becomes $1 - \exp\{-e^{-c}\}$ which is well approximated by e^{-c} for large c. It follows that for c large, $\dfrac{l \log l}{\log 3} + cl$ steps suffice to be close to uniform. This is more than was claimed in Theorem 4. □

REMARK. Chung, Diaconis and Graham (1986) give a more detailed analysis, showing that $l \log l$ is the correct order of magnitude.

4. An Analysis of Riffle Shuffles. In this section we analyze the most commonly used method of shuffling cards—the ordinary riffle shuffle. This involves cutting the deck approximately in half, and interleaving (or riffling) the two halves together. We begin by introducing a mathematical model for shuffling suggested by Gilbert, Shannon and Reeds. Following Reeds, we introduce a strong uniform time for this model and show how the calculations reduce to simple facts about the birthday problem.

The diagram gives the result of a single riffle shuffle of a 10 card deck in the usual $i \to \pi(i)$ format



This shuffle is the result of cutting 4 cards off the top of a 10 card "deck" and riffling the packets together, first dropping cards 10, 9, 8, then card 4, then 7, and so on.

This permutation has two rising sequences

$$\pi(1) < \pi(2) < \pi(3) < \pi(4) \quad \text{and} \quad \pi(5) < \pi(6) < \pi(7) < \pi(8) < \pi(9) < \pi(10).$$

In general, a permutation π of n cards made by a riffle shuffle will have exactly 2 rising sequences (unless it is the identity, which has 1). Conversely, any permutation of n cards with 1 or 2 rising sequences can be obtained by a physical riffle. Thus the mathematical definition of *a riffle shuffle*

is "a permutation with 1 or 2 rising sequences". Suppose c cards are initially cut off the top. Then there are $\binom{n}{c}$ possible riffle shuffles (1 of which is the identity). To see why, label each of the c cards cut with "0" and the others with "1". After the shuffle, the labels form a binary n-tuple with c "0"s: there are $\binom{n}{c}$ such n-tuples and each corresponds to a unique riffle shuffle. Finally, the total number of possible riffle shuffles is

$$1 + \sum_{c=0}^{n} \left\{ \binom{n}{c} - 1 \right\} = 2^n - n.$$

Some stage magicians can perform "perfect" shuffles, but for most of us the result of a shuffle is somewhat random. The actual distribution of one shuffle (that is, the set of probabilities of each of the $2^n - n$ possible riffle shuffles) will depend on the skill of the individual shuffler. The following model for random riffle shuffle, suggested by Gilbert and Shannon (1955) and Reeds (1981), is mathematically tractable and qualitatively similar to shuffles done by amateur card players.

1st description. Begin by choosing an integer c from $0, 1, \ldots, n$ according to the binomial distribution $P\{C = c\} = \frac{1}{2^n}\binom{n}{c}$. Then, c cards are cut off and held in the left hand, and $n - c$ cards are held in the right hand. The cards are dropped 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 dropped from the left packet is $(c - 1)/(n - 1)$.

There are two other descriptions of this shuffling mechanism that are useful.

2nd description. Cut an n card deck according to a binomial distribution. If c cards are cut off, pick one of the $\binom{n}{c}$ possible shuffles uniformly.

3rd description. This generates π^{-1} with the correct probability. Label the back of each card with the result of an independent, fair coin flip as 0 or 1. Remove all cards labelled 0 and place them on top of the deck, keeping them in the same relative order.

LEMMA 7. *The three descriptions yield the same probability distribution.*

Proof. The second and third descriptions are equivalent. Indeed, the binary labelling chooses a binomially distributed number of zeros, and conditional on this choice, all possible placements of the zeros are equally likely.

The first and second descriptions are equivalent. Suppose c cards have been cut off. For the first description, a given shuffle is specified by a sequence $D_1, D_2, \ldots, D_n$, where each D_i can be L or R and c of the D_i's must be L. Under the given model, the chance of all such sequences, determined by multiplying the chance at each stage, is $c!(n - c)!/n!$ □

The argument to follow analyzes the repeated inverse shuffle. This has the same distance to uniform as repeated shuffling because of the following lemma.

LEMMA 8. *Let G be a finite group, $T: G \to G$ one-to-one, and Q a probability on G. Then*

$$\|Q - U\| = \|QT^{-1} - U\|,$$

where $QT^{-1}(g) = Q(T^{-1}(g))$ is the probability induced by T. □

The results of repeated inverse shuffles of n cards can be recorded by forming a binary matrix with n rows. The first column records the zeros and ones that determine the first shuffle, and so on. The ith row of the matrix is associated to the ith card in the original ordering of the deck, recording in coordinate j the behavior of this card on the jth shuffle.

	1	2	3	4
a 1101	*c* 0010	*c* 0010	*f* 1000	*f* 1000
b 1100	*e* 0110	*d* 1011	*a* 1101	*b* 1100
c 0010	*a* 1101	*f* 1000	*b* 1100	*c* 0010
d 1011	*b* 1100	*e* 0110	*c* 0010	*e* 0110
e 0110	*d* 1011	*a* 1101	*d* 1011	*a* 1101
f 1000	*f* 1000	*b* 1100	*e* 0110	*d* 1011

LEMMA 9 (Reeds). *Let T be the first time that the binary matrix formed from inverse shuffling has distinct rows. Then T is a strong uniform time.*

Proof. The matrix can be considered as formed by flipping a fair coin to fill out the i, j entry. At every stage, the rows are independent binary vectors. The joint distribution of the rows, conditional on being all distinct, is invariant under permutations.

After the first inverse shuffle, all cards associated to binary vectors starting with 0 are above cards with binary vectors starting with 1. After two shuffles, cards associated with binary vectors starting $(0,0)$ are on top followed by cards associated to vectors beginning $(1,0)$, followed by $(0,1)$, followed by $(1,1)$ at the bottom of the deck.

Inductively, the inverse shuffles sort the binary vectors (from right to left) in lexicographic order. At time T the vectors are all distinct, and all sorted. By permutation invariance, any of the n cards is equally likely to have been associated with the smallest row of the matrix (and so be on top). Similarly, at time T, all $n!$ orders are equally likely. □

To complete this analysis, the chance that $T > k$ must be computed. This is simply the probability that if n balls are dropped into 2^k boxes there are not two or more balls in a box. If the balls are thought of as people, and the boxes as birthdays, we have the familiar question of the birthday problem and its well-known answer. This yields:

THEOREM 5. *For Q the Gilbert-Shannon-Reeds distribution defined in Lemma 7,*

$$\|Q^{*k} - U\| \leqslant P(T > k) = 1 - \prod_{i=1}^{n-1}\left(1 - \frac{i}{2^k}\right). \tag{4.1}$$

Standard calculus shows that if $k = 2\log_2(n/c)$,

$$P(T > k) \overset{n\to\infty}{\sim} 1 - e^{-\frac{c^2}{2}} \overset{c\to 0}{\sim} \frac{1}{2}c^2.$$

In this sense, $2\log_2 n$ is the cut off point for this bound. Exact computation of the right side of (4.1) when $n = 52$ gives the bounds

k	upper bound
10	.73
11	.48
12	.28
13	.15
14	.08

REMARK (a). The lovely new idea here is to consider shuffling as inverse sorting. The argument works for any symmetric method of labelling the cards. For example, biased cuts can be modeled by flipping an unfair coin. To model cutting off exactly j cards each time, fill the columns of the matrix with the results of n draws without replacement from an urn containing j balls labelled zero and $n - j$ balls labelled one. These lead to slightly unorthodox birthday problems which turn out to be easy to work with.

REMARK (b). The argument can be refined. Suppose shuffling is stopped slightly before all rows of the matrix are distinct—e.g., stop after $2 \log n$ shuffles. Cards associated to identical binary rows correspond to cards in their original relative positions. It is possible to bound how far such permutations are from uniform and get bounds on $\|Q^{*k} - U\|$. Reeds (1981) has used such arguments to show that 9 or fewer shuffles make the variation distance small for 52 cards.

REMARK (c). A variety of ad hoc techniques have been used to get lower bounds. One simple method that works well is to simply follow the top card after repeated shuffles. This executes a Markov chain on n states with a simple transition matrix. For n in the range of real deck sizes, $n \times n$ matrices can be numerically multiplied and then the variation distance to uniform computed. Reeds (1981) has carried this out for decks of size 52 and shown that $\|Q^{*6} - U\| \geq .1$. Techniques which allow asymptotic verification that $k = (3/2)\log_2 n$ is the right cutoff for large n are described in Aldous (1983a). These analyses, and the results quoted above, suggest that seven riffle shuffles are needed to get close to random.

REMARK (d). Other mathematical models for riffle shuffling are suggested in Donner and Uppulini (1970), Epstein (1977), and Thorp (1973). Borel and Cheron (1955) and Kosambi and Rao (1958) discuss the problem in a less formal way. Where conclusions are drawn, 6 to 7 shuffles are recommended to randomize 52 cards.

REMARK (e). Of course, our ability to shuffle cards depends on practice and agility. The model produces shuffles with single cards being dropped about $1/2$ of the time, pairs of cards being dropped about $1/4$ of the time, and i cards blocks being dropped about $1/2^i$ of the time. Professional dealers drop single cards 80% of the time, pairs about 18% of the time and hardly ever drop 3 or more cards. Less sophisticated card handlers drop single cards about 60% of the time. Further discussion is in Diaconis (1982) or Epstein (1977).

It is not clear if neater shuffling makes for a better randomization mechanism. After all, eight perfect shuffles bring a deck back to order. Diaconis, Kantor, and Graham (1983) contains an extensive discussion of the mathematics of perfect shuffles, giving history and applications to gambling, computer science and group theory.

The shuffle analyzed above is the most random among all single shuffles with a given distribution of cut size, being uniform among the possible outcomes. It may therefore serve as a lower bound; any less uniform shuffle might take at least as long to randomize things. Further discussion is in Mellish (1973).

REMARK (f). One may ask, "Does it matter?" It seems to many people that if a deck of cards is shuffled 3 or 4 times, it will be quite mixed up for practical purposes with none of the esoteric patterns involved in the above analysis coming in.

Magicians and card cheats have long taken advantage of such patterns. Suppose a deck of 52 cards in known order is shuffled 3 times and cut arbitrarily in between these shuffles. Then a card is taken out, noted and replaced in a different position. The noted card can be determined with near certainty! Gardner (1977) describes card tricks based on the inefficiency of too few riffle shuffles.

Berger (1973) describes a different appearance of pattern. He compared the distribution of hands at tournament bridge before and after computers were used to randomize the order of the deck. The earlier, hand shuffled, distribution showed noticeable patterns (the suit distributions were too near "even" 4333) that a knowledgeable expert could use.

It is worth noting that it is not totally trivial to shuffle cards on a computer. The usual method, described in Knuth (1981), goes as follows. Imagine the n cards in a row. At stage i, pick a random position between i and n and switch the card at the chosen position with the card at position i. Carried out for $1 \leq i \leq n - 1$, this results in a uniform permutation. In the early days of computer randomization, we are told that Bridge Clubs randomized by choosing about 60 random transpositions (as opposed to 51 carefully randomized transpositions). As the analysis of

Diaconis and Shahshahani (1981) shows, 60 is not enough.

REMARK (g). While revising this paper we noted the following question and answer in a newspaper bridge column ("The Aces", by Bobby Wolff).

Q: How many times should a deck be shuffled before it is dealt? My fellow players insist on at least seven or eight shuffles. Isn't this overdoing it?

A: The laws stipulate that the deck must be "thoroughly shuffled". While no specific number is stated, I would guess that five or six shuffles would be about right; seven or eight would not be out of order.

5. Other Techniques and Open Problems. A number of other natural random walks admit elegant analyses with strong uniform times. For example, Andre Broder (1985) has given stopping times for simple random walk on the "cube" Z_2^d, and for the problem of randomizing n cards by random transpositions. We can similarly analyze nearest neighbor random walks on a variety of 2 point homogeneous spaces. It is natural to inquire if a suitable stopping time can always be found. This problem is analyzed in Aldous and Diaconis (1985): let us merely state two results.

We need to introduce a second notion of distance to the uniform distribution. Let Q be a probability on a finite group G. The *separation of* Q^{k*} to the uniform distribution U after k steps is defined as

$$s(k) = |G| \max_g \{ U(g) - Q^{k*}(g) \}.$$

Clearly $0 \leqslant s(k) \leqslant 1$ with $s(k) = 0$ if and only if $Q^{k*} = U$. The separation is an upper bound for the variation distance:

$$d(k) \leqslant s(k)$$

because

$$\|Q^{k*} - U\| = \sum_{g: Q^{k*}(g) < U(g)} \{ U(g) - Q^{k*}(g) \}.$$

The following result generalizes Lemma 1.

THEOREM 6. *If T is a strong uniform time for the random walk generated by Q on G, then*

$$s(k) \leqslant P(T > k); \quad \text{all } k \geqslant 0. \tag{5.1}$$

Conversely, for every random walk there exists a randomized strong uniform time T such that (5.1) holds with equality.

While separation and variation distance can differ, for random walk problems there is a sense in which they only differ by a factor of 2. For $0 < \varepsilon < \frac{1}{4}$, define

$$\phi(\varepsilon) = 1 - (1 - 2\varepsilon^{1/2})(1 - \varepsilon^{1/2})^2$$

and observe that $\phi(\varepsilon)$ decreases as ε decreases, and $\phi(\varepsilon) \sim 4\varepsilon^{1/2}$ as $\varepsilon \to 0$.

THEOREM 7. *For any distribution Q on any finite group G,*

$$s(2k) \leqslant \phi(2d(k)): k \geqslant 1, \text{ provided } d(k) < \frac{1}{8}.$$

Thus, if k steps suffice to make the variation distance small, at most $2k$ steps are required to make the separation small.

Coupling is a probabilistic technique closely related to strong uniform times which achieves the exact variation distance. The coupling technique applies to Markov chains far more general than random walks on groups: see Griffeath (1975, 1978), Pitman (1976), Athreya and Ney (1977).

Random walk involves repeated convolution and it is natural to try to use Fourier analysis or its non-commutative analog, group representation. Such techniques can sometimes give very sharp

bounds. Letac (1981) and Takács (1982) are readable surveys. Diaconis and Shahshahani (1981, 1984) present further examples. Robbins and Bolker (1981) use other techniques.

Despite this range of available techniques, there are some shuffling methods for which we do not have good results on how many shuffles are needed; for example:

(i) Riffle shuffles where there is a tendency for successive cards to be dropped from opposite hands.

(ii) *Overhand shuffle.* The deck is divided into K blocks in some random way, and the order of the blocks is reversed.

(iii) *Semi-random transposition.* At the kth shuffle, transpose the kth card (counting modulo n) with a uniform random card.

From a theoretical viewpoint, there are interesting questions concerning the cut-off phenomenon. This occurs in all the examples we can explicitly calculate, but we know no general result which says that the phenomenon must happen for all "reasonable" shuffling methods.

Acknowledgment. We thank Brad Efron, Leo Flatto, and Larry Shepp for help with Example 1, and Jim Reeds for help with Section 4.

References

D. Aldous, Markov chains with almost exponential hitting times, Stochastic Proc. Appl., 13 (1982) 305–310.

_____, Random walks on finite groups and rapidly mixing Markov chains, Séminaire de Probabilités, XVII (1983a) 243–297.

_____, Minimization algorithms and random walk on the d-cube, Ann. Probab., 11 (1983b) 403–413.

D. J. Aldous and P. Diaconis, Uniform stopping times for random walks on groups, 1985. In preparation.

K. B. Athreya and P. Ney, A new approach to the limit theory of recurrent Markov chains, Trans. Amer. Math. Soc., 1977.

K. B. Athreya, D. McDonald, and P. Ney, Limit theorems for semi-Markov processes and renewal theory for Markov chains, Ann. Probab., 6 (1978) 788–797.

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

E. Borel and A. Cheron, Théorie Mathématique du Bridge, 2nd ed., Gauthier Villars, Paris, 1955.

A. Broder, unpublished, Stanford University, 1985.

F. K. Chung, P. Diaconis, and R. L. Graham, Random walks arising from random number generation, Technical Report #212, Stanford University, 1983 (to appear, Ann. Probab.).

P. Diaconis (1982), On the use of group representations in probability and statistics, Typed Lecture Notes, Department of Statistics, Harvard University. To appear, Institute of Mathematical Statistics.

P. Diaconis, W. Kantor, and R. L. Graham, The mathematics of perfect shuffles, Advances in Applied Math., 4 (1983) 175–196.

P. Diaconis and M. Shahshahani, Generating a random permutation with random transpositions, Z. Wahrsch. Verw. Gebiete, 57 (1981) 159–179.

_____, Factoring probabilities on compact groups, Technical Report #178, Stanford University (also TR#PH-9, Harvard Univ.) 1986 (to appear, Proc. Amer. Math. Soc.).

_____, Products of random matrices and random walks on groups. To appear, Proc. Conference on Random Matrices (J. Cohen, H. Kesten, C. Newman, eds.), Amer. Math. Soc., Providence, RI, 1984.

J. R. Donner and V. R. R. Uppulini, A Markov chain structure for riffle shuffling, SIAM J. Appl. Math., 18 (1970) 191–209.

R. Epstein, The Theory of Gambling and Statistical Logic, Revised ed., Academic Press, New York, 1977.

W. Feller, An Introduction to Probability and Its Applications, Vol. I, 3rd ed., Wiley, New York, 1968.

L. Flatto, A. Odlyzko, and D. Wales, Random shuffles and group representations, Ann. Probab., 13 (1985) 181–193.

M. Gardner, Mathematical Magic Show, Knopf, New York, 1977.

E. W. Gilbert, Theory of Shuffling, Bell Laboratories Technical Memorandum, Murray Hill, NJ, 1955.

U. Grenander, Probability on Algebraic Structures, Wiley, New York, 1963.

D. Griffeath, A maximal coupling for Markov chains, Z. Wahrsch. Verw. Gebiete, 31 (1975) 95–106.

_____, Coupling methods for Markov processes, in Studies in Probability and Ergodic Theory, Adv. in Math., Supplementary Studies Vol. 2 (J. C. Rota ed.), (1978) 1–43.

H. Heyer, Probability Measures on Locally Compact Groups, Springer, Berlin, 1977.

M. Iosifescu, Finite Markov Chains and Their Applications, Wiley, New York, 1980.

M. Kac, Random walk and the theory of Brownian motion, this MONTHLY, 54 (1947) 369–391.

B. M. Kloss, Probability distributions on bicompact topological groups, Theory Probab. Appl., 4 (1959) 237–270.

D. Knuth, The Art of Computer Programming, Vol. 1, 2nd ed., Addison-Wesley, Reading, MA, 1973.

_____, The Art of Computer Programming, Vol. 2, 2nd ed., Addison-Wesley, Reading, MA, 1981.

D. D. Kosambi and U. V. R. Rao, The efficiency of randomization by card shuffling, J. Roy. Statist. Soc. Ser. A, 128 (1958) 223–233.

G. Letac, Problèmes Classiques de Probabilité sur un couple de Gelfand, Springer Lecture Notes 861, Springer, New York, 1981.

A. A. Markov, Extension of the law of large numbers to dependent events (Russian), Bull. Soc. Phys. Math. Kazan (2) 15 (1906) 135–156.

P. Matthews, Covering problems and random walks on groups, Ph.D. dissertation, Department of Statistics, Stanford University, 1985.

M. J. Mellish, Optimal Card-Shuffling, Eureka, No. 36 (1973) 9–10.

J. Pitman, On coupling of Markov chains, Z. Wahrsch. Verw. Gebiete, 35 (1976) 315–322.

H. Poincaré, Calcul Probabilités, Gauthier-Villars, Paris, 1912.

J. Reeds, Unpublished manuscript, 1981.

D. P. Robbins and E. D. Bolker, The bias of three pseudo random shuffles, Aequationes Math., 22 (1981) 268–292.

L. Takács, Random walks on groups, Linear Algebra Appl., 43 (1982) 49–67.

E. Thorp, Nonrandom shuffling with applications to the game of Faro, J. Amer. Statist. Assoc., 68 (1973) 842–847.

WHAT IS A DIFFERENTIAL? A NEW ANSWER FROM THE GENERALIZED RIEMANN INTEGRAL

SOLOMON LEADER

Mathematics Department, Hill Center, Busch Campus, Rutgers University, New Brunswick, NJ 08903

Unlike derivatives which gained a solid basis in Cauchy's theory of limits, differentials found no effective accommodation with the rising level of rigor in calculus. Justly castigated by Berkeley as "ghosts of departed quantities", differentials clung fortuitously to the notational niche in calculus created for them by Leibniz. In this century they came to be presented as functionals on tangent spaces, a constricted role that made them respectable but evaded the issue of their wider role in integration. The resurrection of infinitesimals by nonstandard analysis rekindled interest in Leibniz' original concept of differential.

We present here a completely new approach to differentials in one dimension. This approach is motivated by the following considerations: (i) differentials spring directly from the integration process, (ii) the utility of differentials in integration extends beyond conventional differential forms, (iii) a viable theory of differentials is readily attainable by standard analysis, and (iv) the generalized Riemann integral fills a vital gap in analysis and should have an innovative impact on our calculus and real variables courses. In the theory expounded here differentials on a 1-cell K form a Riesz space (lattice-ordered linear space). So for each differential σ we have the differentials

$$|\sigma| = \sigma \vee -\sigma, \quad \sigma^+ = \sigma \vee 0, \quad \text{and} \quad \sigma^- = (-\sigma)^+ = -(\sigma \wedge 0)$$

Solomon Leader: I wrote my Ph.D. thesis in analysis at Princeton in 1952 under the late Salomon Bochner. For the past 33 years I have been at Rutgers figuring out how calculus should be taught. My main interests have been in measure theory, integration, proximity spaces, and fixed points. In warm weather my favorite diversion is body-surfing off Long Beach Island. My wife and I enjoy snorkeling in the Virgin Islands and welcome any excuse to visit Switzerland.