Chapter 4. Probabilistic Arguments

A. Introduction — Strong uniform times.

There are a number of other arguments available for bounding the rate of convergence to the uniform distribution. This chapter discusses the method of strong uniform times and coupling. Let's begin with a simple example, drawn from Aldous and Diaconis (1986).

Example 1. Top in at random. Consider mixing a deck of n cards by repeatedly removing the top card and inserting it at a random position. This corresponds to choosing a random cycle:

$$(1) \qquad P(\mathrm{id}) = P(21) = P(321) = P(4321) = \ldots = P(nn-1\ldots1) = \frac{1}{n}.$$

The following argument will be used to show that $n \log n$ shuffles suffice to mix up the cards. Consider the bottom card of the deck. This card stays at the bottom until the first time a card is inserted below it. This is a geometric waiting time with mean n. As the shuffles continue, eventually a second card is inserted below the original bottom card (this takes about $n/2$ further shuffles). 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 six possible orders of the three bottom cards is equally likely. Now consider the first time T that the original bottom card comes 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, all $n!$ possible arrangements of the deck are equally likely.

When the original bottom card is at position k from the bottom, the waiting time for a new card to be inserted is geometric with mean n/k. Thus the waiting time T has mean $n + \frac{n}{2} + \frac{n}{3} + \ldots + \frac{n}{n} \doteq n \log n$.

To make this argument rigorous, introduce strong uniform times. Let G be a finite group. Intuitively, a stopping time is a rule which looks at a sequence of elements in G and says "stop at the jth one." The rule is allowed to depend on what appears up to time j, but not to look in the future. Formally, *a stopping time* is a function $T: G^\infty \to \{1, 2, \ldots, \infty\}$ such that if $T(\underset{\sim}{s}) = j$ then $T(\underset{\sim}{s}') = j$ for all $\underset{\sim}{s}'$ with $s_i' = s_i$ for $1 \le i \le j$. Let Q be a probability on G, X_k the associated random walk, P the associated probability on G^∞. A *strong uniform time* T is a stopping time T such that for each $k < \infty$,

$$(2) \qquad P\{T = k,\ X_k = s\} \quad \text{is constant in } s.$$

Note that (2) is equivalent to independence of the stopping time and the stopped process:

$$P\{X_k = s|T = k\} = 1/|G| \tag{3}$$

or to

$$P\{X_k = s|T \leq k\} = 1/|G|. \tag{4}$$

In Example 1, the time T that the first card takes to reach the top and has been inserted into the deck is certainly a stopping time. The inductive argument given shows that, given $T = k$, all arrangements of the deck are equally likely, so T is a strong uniform time. Many other examples will be given in the remainder of this chapter. The following lemma relates strong uniform times to the distance between Q^{*k} and the uniform distribution U.

LEMMA 1. *Let Q be a probability on the finite group G. Let T be a strong uniform time for Q. Then for all $k \geq 0$*

$$||Q^{*k} - U|| \leq P\{T > k\}.$$

Proof. For any $A \subset G$,

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

Thus,

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

□

Using this result we can deduce a sharp bound for the first example: $n \log n$ steps are both necessary and sufficient to drive the variation distance to zero.

Theorem 1. *For the top in at random shuffle defined in (1), let $k = n \log n + cn$. Then,*

$$||P^{*k} - U|| \leq e^{-c} \text{ for } c \geq 0,\ n \geq 2, \tag{5}$$

$$||P^{*k} - U|| \to 1 \text{ as } n \to \infty, \text{ for } c = c(n) \to -\infty. \tag{6}$$

Proof. As argued above, $T = T_1 + (T_2 - T_1) + \ldots + (T_{n-1} - T_{n-2}) + (T - T_{n-1})$ where T_1 is the time until the 1st card is placed under the bottom card and $T_{i+1} - T_i$ has a geometric distribution $P\{T_{i+1} - T_i = j\} = \frac{i+1}{n}(1 - \frac{i+1}{n})^{j-1}$; $j \geq 1$. Further, these differences are independent.

The time T has the same distribution as the waiting time in the coupon collector's problem; to define this, consider a random sample with replacement from an urn with n balls. Let V be the number of balls required until each ball has been drawn at least once. Let $m = n \log n + cn$. For each ball b, let A_b be the event "ball b is not drawn in the first m draws." Then,

$$(7) \quad P\{V > m\} = P\{\cup_b A_b\} \leq \Sigma_b \, P\{A_b\} = n(1 - \frac{1}{n})^m \leq n \, e^{-m/n} = e^{-c}.$$

Now V can be written

$$V = (V - V_{n-1}) + (V_{n-1} - V_{n-2}) + \ldots + (V_2 - V_1) + V_1$$

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 new ball is $\frac{n-i}{n}$, so $V_{i+1} - V_i$ is geometric,

$$P\{V_{i+1} - V_i = j\} = \frac{n-i}{n}(1 - \frac{n-i}{n})^{j-1}, \; j \geq 1.$$

It follows that the laws of T and V are the same. So (7) and lemma 1 (the upper bound lemma) combine to give a proof of (5).

To prove (6), 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 $U(A_j) = 1/j!$. For $k = n \log n + c_n n$, $c_n \to -\infty$, we argue that

$$(8) \qquad P^{*k}(A_j) \to 1 \text{ as } n \to \infty, \; j \text{ fixed.}$$

Then $||P^{*k} - U|| \geq \max_j \{P^{*k}(A_j) - U(A_j)\} \to 1$ as $n \to \infty$, establishing (6).

To prove (8), observe that $P^{*k}(A_j) \geq P(T - T_{j-1} > k)$. 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

$$(9) \qquad P(T - T_{j-1} \leq k) \to 0 \text{ as } n \to \infty; \; j \text{ fixed} \; .$$

We shall prove this using Chebyshev's inequality:

$$P(|Z - EZ| \geq a) \leq \frac{\text{var}(Z)}{a^2}, \text{ where } a \geq 0, \text{ and } Z \text{ is any random variable} \; .$$

For a geometric variable

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

and so

$$E(T - T_j) = \sum_{i=j}^{n-1} \frac{n}{i+1} = n \log n + 0(n),$$

$$\text{var}(T - T_j) = \sum_{i=j}^{n-1} (\frac{n}{i+1})^2(1 - \frac{i+1}{n}) = 0(n^2),$$

and Chebyshev's inequality applied to $Z = T - T_{j-1}$ readily yields (9). □

B. EXAMPLES OF STRONG UNIFORM TIMES.

Example 2. Simple random walk on Z_2^d. For simplicity we work with the following probability

$$Q(0\ldots0) = \frac{1}{2},\ Q(10\ldots0) = Q(01\ldots0) = \ldots = Q(0\ldots1) = \frac{1}{2d},$$
$$Q = 0 \quad \text{otherwise}. \tag{10}$$

The following simple stopping time has been developed by Andre Broder. It involves "checking off coordinates" according to the following scheme: at each time, pick one of the d coordinates at random and check it off. Then flip a fair coin. If the coin comes up heads, take a step in the direction of the chosen coordinate. If the coin comes up tails, the random walk stays where it is. Stop at time T when all coordinates have been checked.

Clearly the particle evolves according to the probability (10). To see that T is a strong uniform time, observe that because of the randomized coin toss, the particle is equally likely to have a zero or one in each checked coordinate.

Theorem 2. *For simple random walk on Z_2^d (10), and $k = n \log n + cn$,*

$$||P^{*k} - U|| \leq e^{-c}.$$

Proof. This follows from the upper bound lemma and the bound from the coupon collector's waiting time (7). □

Remark 1. Fourier analysis and the lower bound arguments of Chapter 3 show that $\frac{1}{2}n \log n + cn$ steps is the right answer for this version of random walk. The discrepancy is explained in Section C (exercise 4) below.

Remark 2. In Example 2, the uniform time depends on added, external, randomization. It was not constructed just by looking at the past of the process. The

upper bound, lemma 1, holds for such randomized strong uniform times without change.

EXERCISE 1. Give a strong uniform time for random walk on Z_2^d determined by $Q(00\ldots0) = Q(10\ldots0) = \ldots = Q(0\ldots01) = \frac{1}{d+1}$.

Example 3. General random walk on a finite group.

Theorem 3. *Let G be a finite group and Q a probability on G such that for some $c(0 < c < 1)$ and k_0,*

$$Q^{*k}\{A\} \geq cU(A) \tag{11}$$

for all $A \subset G$, $k \geq k_0$. Then

$$||Q^{*k} - U|| \leq (1-c)^{\lfloor k/k_0 \rfloor}.$$

Proof. Suppose first that $k_0 = 1$. Define a probability Q_1 on G by

$$Q_1(s) = \frac{Q(s) - cU(s)}{1-c}.$$

Thus

$$Q(s) = cU(s) + (1-c)Q_1(s).$$

This gives the following recipe for choosing steps according to Q: flip a coin with probability of heads equal to c. If the coin comes up heads, step according to U, if tails, step according to Q_1. Let T be the first time a head occurs. This T is clearly a strong uniform time and

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

For general k_0, apply the argument to Q^{*k_0}. □

Remarks. The argument above extends easily to compact groups with condition (11) required to hold for all open sets. In this generality, the theorem appears in Kloss (1959) whose proof is a Fourier version of the same argument Athreya and Ney (1978) apply this idea to prove convergence to stationarity for general state space Markov chains.

The simplicity of the proof, coupled with the generality of the argument, should make the reader suspicious. While the result seems quantitative, all depends on estimating c and k_0. I do not know how to use this theorem to get the right rate of convergence in a single example.

Example 4. Random transpositions. This problem was discussed at some length in Chapter 3. Here are two constructions of strong uniform times. Both involve the notion of "checking" the backs of certain cards as they are chosen successively

in pairs. This argument is capable of use in a variety of other random walks. It is due to Andre Broder.

Construction A (Broder). The basic mixing procedure involves switching pairs of cards (L_i, R_i). If either

a) both hands touch the same unchecked card; or
b) the card touched by the left hand is unchecked but the card touched by the right hand is checked,

then check the card touched by the left hand. Stop at time T when all cards are checked.

Construction B (Matthews). If both hands touch unchecked cards, then check the card touched by the left hand.

In each construction stop at the time T that only one card remains unchecked.

Proof. Construction A. First consider the situation informally. The procedure starts when both hands hit the same card (say card 1). This is checked. Nothing happens until either both hands hit a different card (say 2) or the left hand hits an unchecked card (say 2) and the right hand hits card 1 whereupon these cards are switched. At this stage, conditional on the positions of the two checked cards A_1, A_2 say, and the labels 1, 2, the positions are equally likely to correspond $(A_1, 1)(A_2, 2)$ or $(A_1, 2)(A_2, 1)$. This is because the chance of choosing card 2 is $\frac{1}{n^2}$ for both possibilities.

In general, the position may be described as follows:

$$\{L, \{A_1 \dots A_L\}, \{C_1 \dots C_L\}, \Pi_L\}.$$

Where

$$\begin{aligned} L &= \text{number of checked cards} \\ \{A_1 \dots A_L\} &= \text{set of positions of the checked cards} \\ \{C_1 \dots C_L\} &= \text{labels (or names) of the checked cards.} \\ \Pi_L\colon \{A_1 \dots A_L\} &\to \{C_1 \dots C_L\} \text{ records the card at each position.} \end{aligned}$$

□

Claim. *At each time, conditional on* L, $\{A_1 \dots A_L\}$, $\{C_1 \dots C_L\}$, *the permutation* Π_L *is uniform.*

The claim is proved by induction. It is clearly true for $L = 0$ and 1. Assume it for $L = \rho$. The claim remains true until a new card c is checked. This can occur by both hands hitting the same new card or by the left hand hitting c and the right hand hitting one of the ρ checked cards. For any new card c, each of these $\rho + 1$ possibilities has the same chance $\frac{1}{n^2}$. It follows that Π_{L+1} is uniform.

The proof for Construction B is similar. The state at any time can again be taken as above. This time the inductive step is that, given L,

a) $\{A_1 \dots A_L\}$ and $\{C_1 \dots C_L\}$ are independent and uniformly distributed
b) Given L, $\{A_1 \dots A_L\}$ and $\{C_1 \dots C_L\}$, the permutation Π_L is uniform.

It can be verified that both a and b hold at each time for Construction B. Here, (a) is needed to check (b) in the argument. Note that (a) is not valid (or needed) for Construction A. ($\{A_1 \dots A_L\}$ and $\{C_1 \dots C_L\}$ are marginally uniform but not independent.)

The analysis of T in Construction A is similar to that in Example 2. Write $T = \sum_{i=1}^{n}(T_i - T_{i-1})$ where T_i is the number of transpositions until i cards are checked. The random variables $(T_i - T_i)$ are independent with geometric distributions of mean $n^2/[(i+1)(n-i)]$. Thus

$$E(T) = \sum_{i=0}^{n-1} n^2/[(i+1)(n-i)] = (2 + 0(\frac{1}{n}))n \log n$$
$$\mathrm{Var}(T) = 0(n^2).$$

Now the central limit theorem implies for $k = 2n \log n + c(n)n$, with $c(n) \to \infty$.

$$||P^{*k} - U|| \to 0 \text{ as } n \to \infty.$$

The T given by Construction B turns out to give $k = 0(n^2)$ as required. However, Construction B starts out by checking cards rapidly. Peter Matthews (1986b) observes that the two constructions can be combined: use Construction B until m cards have been checked (for fixed m, say $m = \frac{n}{2}$). Then use Construction A. Because (a) and (b) are valid throughout the time involved for Construction B, when A takes over, (b) remains valid until the time T that all cards are checked. This time gives $k = n \log n$ sufficient to drive the variance distance to zero. Matthews has suggested variants which give the correct number of steps $\frac{1}{2}n \log n$.

EXERCISE 2. To emphasize the need for careful proof, show that checking each card as it is touched or checking each card the left hand touches do not yield strong uniform times in the random transposition problem. (Hint: consider a three-card deck, and see what the distribution is given $T = 3$.)

Further examples (simple random walk on Z_p or $X_{k+1} = a_k X_k + b_k$) are given in Aldous and Diaconis (1986, 1987a, 1987b) and Matthews (1986a,b).

C. A CLOSER LOOK AT STRONG UNIFORM TIMES.

The success of strong uniform times in the examples above and a variety of other examples given below prompts obvious questions: can one always find a useful strong uniform time? Are there strong uniform times that achieve the variation distance? To answer these questions it is useful to introduce a different notion of distance from uniformity.

Definition. Let Q be a probability on the finite group G. Define the n step *separation* by

$$s(n) = |G| \max_s \{\frac{1}{|G|} - Q^{*n}(s)\}.$$

Clearly, $0 \le s(n) \le 1$ with $s(n) = 0$ iff $Q^{*n} = U$, $s(n) = 1$ iff $Q^{*n}(s) = 0$ for some s. The separation is an upper bound for the variation distance:

$$||Q^{*n} - U|| \le s(n)$$

because

$$||Q^{*n} - U|| = \sum_{s:Q^{*n}(s)<1/|G|} \Big\{\frac{1}{|G|} - Q^{*n}(s)\Big\}.$$

Note that the two distances can be very different: If Q is uniform on $G - \{\mathrm{id}\}$ then $||Q - U|| = 1/|G|$ but $s(1) = 1$. The following theorem improves the upper bound of lemma 1.

Theorem 4. *If T is a strong uniform time for the random walk generated by Q on G, then for all k*

$$s(k) \le P\{T > k\}. \tag{12}$$

Conversely, for every random walk there is a strong uniform time such that (12) holds with equality.

Proof. Let k_0 be the smallest value of k such that $P\{T \le k_0\} > 0$. The result holds vacuously if $k_0 = \infty$ and for $k < k_0$. For $k \ge k_0$, $s \in G$

$$\begin{aligned} |G|\{\frac{1}{|G|} - Q^{*k}(s)\} &= 1 - |G|Q^{*k}(s) \le 1 - |G|P\{X_k = s \text{ and } T \le k\} \\ &= 1 - |G|P\{X_k = s | T \le k\} \cdot P\{T \le k\} \\ &= 1 - P\{T \le k\} = P\{T > k\}. \end{aligned}$$

This proves (12).

For the converse, the random time T will be defined as follows: at time k, given that the random walk is at t, flip a coin with probability of heads

$$p_k(t) = \frac{\alpha_k - \alpha_{k-1}}{Q^{*k}(t) - \alpha_{k-1}}$$

where $\alpha_k = \min_s Q^{*k}(s)$. If heads comes up, stop. If tails comes up, take another step and flip again with probability p_{k+1}. Observe that $p_k(t) \ge 0$. Let k_0 be the smallest integer such that $\alpha_{k_0} > 0$. Clearly $p_k = 0$ for $k < k_0$, and

$$p_{k_0}(t) = P\{T = k_0 | X_{k_0} = t\} = \frac{\alpha_{k_0}}{P\{X_{k_0} = t\}}.$$

Thus, $P\{T = k_0\} = \Sigma_t\, P\{T = k_0 | X_{k_0} = t\} \cdot P\{X_{k_0} = t\} = |G|\alpha_{k_0}$. Further,

$$P\{X_{k_0} = s | T = k_0\} = P\{T = k_0 | X_{k_0} = s\} \cdot \frac{P\{X_{k_0} = s\}}{P\{T = k_0\}} = \frac{1}{|G|}.$$

This is the first step in an inductive argument to show that T is a strong uniform time. For general k,

$$P\{X_k = s, T = k\} = \alpha_k - \alpha_{k-1}. \tag{13}$$

This follows because

$$\begin{aligned} P\{X_k = s, T = k\} &= P\{T = k | X_k = s, T \geq k\} \cdot P\{X_k = s, T \geq k\} \\ &= \frac{\alpha_k - \alpha_{k-1}}{P\{X_k = s\} - \alpha_{k-1}} \cdot [P\{X_k = s\} - P\{X_k = s; \\ &\qquad T \leq k-1\}] \end{aligned}$$

If (13) holds for all integers smaller than k, then

$$P\{X_k = s, T \leq k-1\} = \alpha_{k-1}.$$

This shows T is strong uniform. □

EXERCISE 3. Prove that the strong uniform time T^* constructed in the course of proving Theorem 4 is the stochastically fastest strong uniform: $P\{T^* > k\} \leq P\{T > k\}$ for all k. Now consider Example 1 (top in at random). The stopping time defined there can be improved: consider T^* — the first time that the card originally *second* from the bottom comes up to the top. Show that T^* is a fastest strong uniform time.

EXERCISE 4. As an example of Theorem 4, consider the model for random walk on the d-cube treated in Section B. The cutoff point for variation distance is $\frac{1}{2}\, d \log d$, and the stopping time argument gives d log d. Show that this is sharp: it takes $d \log d + cd$ steps to have a reasonable probability of reaching the vertex opposite $\underline{0}$, namely $(1 \ldots 1)$. Hint: try Fourier analysis.

The following result, proved in Aldous and Diaconis (1987) shows that the factor of 2 found above is no accident. Roughly, if the variation distance becomes small after k steps, the separation becomes small after at most $2k$ steps. To make this precise, let $\phi(\varepsilon) = 1 - (1 - 2\varepsilon^{\frac{1}{2}})(1 - \varepsilon^{\frac{1}{2}})^2$. Observe that $\phi(\varepsilon)$ decreases with ε and $\phi(\varepsilon) \sim 4\varepsilon^{\frac{1}{2}}$ as $\varepsilon \to 0$.

Theorem 5. *For any probability Q on any finite group G, and all $k \geq 1$,*

$$s(2k) \leq \phi(2||Q^{*k} - U||) \text{ provided } ||Q^{*k} - U|| < \frac{1}{8}.$$

Further discussion of separation can be found in Aldous and Diaconis (1986, 1987) or Diaconis and Fill (1988).

D. AN ANALYSIS OF REAL RIFFLE SHUFFLES.

How many ordinary riffle shuffles are required to bring a deck of cards close to random? We will show that the answer is 7. The discussion proceeds in two sections: (1) practical discussion and data analysis, (2) a model for riffle shuffling.

(1) *Practical shuffling.* Of course, people shuffle cards all the time for card games. We begin by asking "Does it matter?" That is, even if people don't shuffle really well, will it make any practical difference? One answer to this question is in Berger (1973). Berger uses the fact that tournament bridge went from hand shuffling to computer shuffling in the late 1960's. Berger obtained records of the suit distribution of the south hand in 2000 deals, one thousand before the computer, one thousand after the computer. A summary is in Table 1.

Inspection of the table shows that hands with an even suit distribution occur with higher than the expected frequency in hand shuffling. A chi-squared test rejects uniformity of the suit distribution in hand shuffling. Uniformity is accepted for computer shuffling. Something is going on that *does* make a practical, observable difference. Here is a first explanation: The way bridge tends to be played, cards are collected in groups of 4, by suit. If the riffle shuffling was clumpy and clustery, cards of the same suit would tend to clump together and then, when the deck was dealt into 4 hands, tend to be separated.

Table 1
Frequency of Computer-dealt Hands Versus Theoretical
Expected Frequencies from Berger (1973)

Distribution of the 4 suits	Expected Frequencies	Actual Frequencies of Computer-dealt Hands	Actual Frequencies of Man-dealt Hands
4,4,3,2*	216	198	241
5,3,3,2	155	160	172
5,4,3,1	129	116	124
5,4,2,2	106	92	105
4,3,3,3	105	103	129
6,3,2,2	56	64	46
6,4,2,1	47	53	36
6,3,3,1	34	40	41
5,5,2,1	32	40	19
4,4,4,1	30	35	25
7,3,2,1 and others	90	99	62
	1,000	1,000	1,000

* by "4,4,3,2" we mean that the thirteen cards contained 4 cards in one suit, 4 cards in another suit, 3 cards in another suit, and 2 cards in the remaining suit.

This would make "even splits" like 4 3 3 3 occur more often than they should. One objection to this is that the cards in duplicate bridge are not usually collected in groups of 4 (they are in non-duplicate games). In duplicate, the cards are collected into 4 piles of 13, each pile being roughly in the same suit order. If these piles were placed on top of one another and riffle shuffled twice, the cards would

tend to clump in suit groups of 4 and we are back to the previous case.

Ely Culbertson (1934) discusses ways of taking advantage of poor shuffling in Bridge. Thorp (1973) discusses other card games.

A second practical view comes from considering the results of a single shuffle. An example is presented in Table 2 below. This records a shuffle for which the deck was cut 1 through 29, and 30 through 52. Card 29 was dropped first, then 28, then 52, then, ..., then 1, finally 30.

Table 2
A Single Riffle Shuffle

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
$\pi(i)$	2	3	5	7	9	11	13	14	16	18	20	22	24	26	27	29	31	33	35	36	38	39	41	42	45	46
	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
	48	51	52	1	4	6	8	10	12	15	17	19	21	23	25	28	30	32	34	37	40	43	44	47	49	50

A single riffle shuffle can have at most 2 rising sequences. There are $2^n - n$ possible arrangements of n cards after a single riffle shuffle. Similarly, there are at most 4 rising sequences after 2 riffle shuffles. This generalizes:

Theorem 6 (Shannon).
(1) Let π be a permutation with R rising sequences. π is the outcome of k riffle shuffles if and only if $R \leq 2^k$.
(2) Each π with exactly 2^k rising sequences can be obtained by k riffle shuffles in only one way.

This theorem appears in E. Gilbert (1955), "Theory of Shuffling," Bell Laboratories Technical Memorandum. Part (1) has been used as the basis of a card trick for many years. In this trick, a deck of cards is mailed to a spectator who is instructed to riffle shuffle the deck 3 times, giving the deck any number of straight cuts during the shuffling. Then the top card is removed, noted, and placed into the center of the pack. This is followed by more cuts. The pack is mailed back to the magician who unerringly finds the card. The secret is that there will be eight rising sequences and 1 card in its own rising sequence. It is not hard to show that a random permutation has about $\frac{n}{2}$ rising sequences so a few shuffles (on order $\log_2 \frac{n}{2}$) will not be enough to randomize n cards.

These arguments yield a lower bound. In a bit more generaltiy, if P is a probability on a finite group G supported and uniform on the set $A \subset G$, then

$$||P - U|| \geq P(A) - U(A) = 1 - \frac{|A|}{|G|}.$$

This can be combined with the observations on rising sequences to give a lower bound that works for a few shuffles. Let $F_n(R)$ be the number of permutations of n items with exactly R rising sequences. Thus $F_n(1) = 1$ and $F_n(2) = 2^n - (n+1)$. A formula for $F_n(R)$ is derived in Sade (1949); see also Riordan (1950):

$$F_n(R) = \sum_{j=0}^{R} (-1)^j \binom{n+1}{j} (R-j)^n.$$

In k shuffles, the total number of permutations that can be achieved is $T_n(k) = \sum_{R=1}^{2^k} F_n(R)$. Thus $1 - T_n(k)/n!$ is a lower bound for the variation distance. For $n = 52$, the lower bound is larger than .99 for $1 \leq k \leq 4$. For $k = 5$ it is .38, for $k = 6$ it is zero.

Observe that this approach makes *no* assumptions about the stochastic mechanism for shuffling, but the argument breaks down at 6 shuffles.

(2) *A probability model.* The following model for riffle shuffling was suggested by Shannon and Gilbert, and Reeds.

1st description: Cut the n card deck according to a binomial distribution with parameters $\frac{1}{2}, n$. Suppose k cards are cut off. Pick one of the $\binom{n}{k}$ possible riffle shuffles uniformly.

2nd description: Cut the n card deck according to a binomial distribution with parameters $\frac{1}{2}, n$. Suppose k cards are cut off and held in the left hand and $n - k$ held in the right hand. Drop cards with probability proportional to packet size. Thus the chance that a card is dropped first from the left hand is $\frac{k}{n}$. If this happens, the chance that the left hand drops a second card is $\frac{k-1}{n-1}$; and so on.

3rd description: To generate the inverse shuffle, label the back of each card with the result of an independent fair coin flip: $\{0, 1\}$. Remove all cards labeled 0 and place them on top of the deck, keeping the cards otherwise in the same relative order.

LEMMA 2. *The three descriptions yield the same probability distribution.*

Proof. The 1st and 3rd descriptions are equivalent: indeed, the binary labeling chooses a binomial number of zeros and conditional on this choice, all possible placements of the zeros are equally likely. The lst and 2nd descriptions are equivalent: Suppose k cards have been cut off. Under the 2nd description, the chance of a shuffle is the chance of the sequence of drops $D_1, D_2, \ldots, D_n$, where each D_i can be L or R and k D_i's must be L and $n - k$ D_i's must be R. The chance of any such sequence is $k!(n-k)!/n!$. □

Remarks. This shuffling mechanism has some claim to being the "most random" subject to the binomial cutting. It has the largest entropy, for example. As a model for shuffling, it yields shuffles a bit "clumpier" than either the shuffles of Diaconis or Reeds discussed in remark (e) below. Only half the packets are expected to be of size 1, a quarter of size 2, etc. Of course, extremely neat shuffles are not necessarily good for randomization. A perfect shuffle is completely non-random for example, eight perfect shuffles bring the deck back to order. See Diaconis, Graham, and Kantor (1983). Mellish (1973) discusses these issues.

To proceed further, we construct a strong uniform time for this model of shuffling. To begin, observe that the variation distance is invariant under 1-1 transformations, so it is the same problem to bound the number of inverse shuffles required to get close to random.

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.

LEMMA 3 *(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 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 lexographic 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 7. *For Q the Gilbert-Shannon-Reeds distribution defined in Lemma 2,*

$$||Q^{*k} - U|| \leq P\{T > k\} = 1 - \prod_{i=1}^{n-1}(1 - \frac{i}{2^k}). \tag{14}$$

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

$$P\{T > k\} \underset{\infty}{\overset{n}{\approx}} 1 - e^{-\frac{c^2}{2}} \underset{0}{\overset{c}{\approx}} \frac{c^2}{2}.$$

In this sense, 2 log n is the cut off point for this bound. Exact computation of the right side of (14) 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. The first time all vectors are different is a strong uniform time. These lead to slightly unorthodox birthday problems which turn out to be easy to work with.

Observe that the shuffle in which only 1 card is cut off and randomly riffled into the deck is the "top in at random" shuffle of example 1. The two stopping times are the same!

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 simply to 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 (1983). 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 Uppuluri (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) some data analysis. 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 card blocks being dropped about $1/2^i$ of the time.

To get a feeling for the difference between shufflers, the following experiment was performed: Diaconis and Reeds each shuffled a deck of 52 cards about 100

times; every permutation was recorded. The following summary statistics are relevant.

Diaconis — 103 Shuffles

# cut off top	23	24	25	26	27	28	29
	2	4	22	32	33	9	1

In shuffling, the left hand dropped first 44 times. In all there were 4,376 "packets" dropped. The counts and proportions were

	1	2	3	4	5
	3501	793	63	15	4
	.80	.18	.01	.00	.00

The packet size distribution of first dropped packets was

1	2	3	4	5
.37	.37	.17	.1	.01

Reeds — 100 Shuffles

# cut off top	23	24	25	26	27	28	29	30	31
	2	2	8	16	23	26	16	5	2

In shuffling, the left hand dropped first 18 times. In all there were 3,375 "packets" dropped. The counts and proportions were

1	2	3	4	5	6	7	8	9	10	11	12	13
2102	931	228	68	24	12	3	3	2	0	0	1	1
.62	.28	.07	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00

Diaconis does very neat shuffles and can be compared to a Las Vegas dealer. Reeds shuffles like an "ordinary person." Observe that the first drops for Diaconis are quite different from the average drop. Even though the two types of shufflers are fairly different, to a first approximation they are quite similar, both dropping 1, 2, or 3 cards most of the time.

Remark (f). There is another equivalent way to describe repeated riffle shuffles under the Gilbert-Shannon-Reeds model that suggests much further research. The following evolved in conversations with Izzy Katznelson and Jim Reeds. Begin by dropping n points at random into the unit interval and labeling them, left to right, as $1, 2, \ldots, n$. The transformation $T(x) = 2x (\mathrm{mod}\ 1)$ (sometimes called the Baker's transformation) maps the unit interval into itself and so permutes the points. T takes each of the two half intervals and stretches it out to cover $[0, 1]$. There are a binomial number of points in each half and T shuffles them together. Arguing as in Lemma 1 above, it is easy to see that the induced permutation is precisely a basic riffle shuffle. Further, successive shuffles are independent (they depend on successive bits of the underlying uniform variables).

To complete the argument, consider k chosen so large that n points dropped at random into $[0, 1]$ fall into disjoint pieces of a partition with pieces of length $1/2^k$, with high probability. Picture a point in a piece of the partition. After k shuffles, the piece is stretched out to cover $[0, 1]$. The point is randomly distributed in the piece of the partition. After k shuffles it's randomly distributed in $[0, 1]$. Further, points in disjoint pieces are independent. After k shuffles, the n points are in random relative arrangement (given that they fall into disjoint pieces).

This argument generalizes to some extent ($x \to k_j x (\text{mod } 1)$ on shuffle j for integer k_j). It should be possible to take other measure preserving transformations (toral endomorphisms of the unit square (see Walters (1982)) and convert them to other shuffles.

E. Coupling.

There is a more widely known purely probabilistic argument called coupling. Again, perhaps it is best to begin with an example, this one is due to David Aldous.

Example 1-Borel's shuffle. Borel and Cheron (1955) discuss several methods of mixing up cards. They give the following as an open problem: begin with n cards. Take the top card off and put it into the deck at a random position. Take the bottom card off and put it into the deck at a random position. Continue alternately placing top in at random, bottom in at random.

Observe that there is no longer an obvious stopping time. The following elegant coupling argument has been suggested by David Aldous. It is better to work with the inverse shuffle that removes a random card and places it alternately on top or bottom. Because of the invariance of variation distance under 1-1 maps ($||Q - U|| = ||Qh^{-1} - Uh^{-1}||$ for any 1-1 function $G \to G$) the two shuffles have the same rates of convergence (see exercise 3 of Chapter 3).

To describe a "coupling", consider a second deck of cards. The first deck starts in order $\{1, 2, 3, \ldots, n\}$. The second deck starts in a random order. A card is determined at each stage by shuffling a *third* deck and choosing a card at random. Say the first card chosen is the six of hearts. Remove the six from deck 1 and place it on top. Remove the six from deck two and place it on top. Note that from each deck's marginal vantage point, a card was removed at random and placed on top.

The second step is to reshuffle the 3rd deck and choose a second card, say the Ace of spades. This is removed and placed at the bottom of each deck. Continue in this way, each time choosing a card at random from the third deck, removing the card from decks one and two, and placing the card alternately on top and bottom.

As this process continues, decks one and two match up. The same cards being in the same order at top and bottom. If the same card is chosen again, the procedure keeps the same number of matches. A new match is created for each new card touched. Let T be the first time that each card has been touched. Clearly, the two decks are in the same order, but deck two started at random, and so remains random. It follows that deck one is random at time T. The bound on the coupon collector's problem yields

Theorem 8. *For Borel's shuffle, if $k = n \log n + cn$ for $c > 0$,*

$$||P^{*k} - U|| \leq e^{-c}.$$

EXERCISE 5. Find a strong uniform time to get a bound in Borel's problem.

To discuss coupling more carefully, we need the following fact about variation distance:

LEMMA 4. *Let S be a finite set. Let P_1 and P_2 be probability measures on S. Let Q be a probability on $S \times S$ with margins P_1 and P_2. Let Δ be the diagonal: $\Delta = \{(s,s): s \in S\}$ then,*

$$||P_1 - P_2|| \leq Q(\Delta^c).$$

Proof.

$$\begin{aligned} |P_1(A) - P_2(A)| &= |Q(A \times S) - Q(S \times A)| \\ &= |Q(A \times S \cap \Delta) + Q(A \times S \cap \Delta^c) - Q(S \times A \cap \Delta) \\ &\qquad - Q(S \times A \cap \Delta^c)|. \end{aligned}$$

The first and third numbers in the absolute value sign are equal. The second and fourth give a difference between two numbers, both non-negative and smaller than $Q(\Delta^c)$. □

Remarks. The inequality is sharp in the sense that there is a Q which achieves equality. A proof and discussion may be found in V. Strassen (1965). This Q allows the following interpretation of variation distance: $||P_1 - P_2|| = \varepsilon$ if and only if there are two random variables, X_1 distributed as P_1 and X_2 distributed as P_2, such that $X_1 = X_2$ with probability $1 - \varepsilon$; X_1 and X_2 may be arbitrarily different with probability ε. Another interpretation: the optimal Q is most concentrated about the diagonal with these fixed margins.

Let us define a coupling for a Markov chain on state space I, with transition probability $P_i(j)$, and stationary distribution π. We will work with Markovian couplings. These are processes on $I \times I$ with transition probability Q satisfying

$$\sum_t Q_{i,j}(s,t) = P_i(s) \text{ for all } j$$

$$\sum_s Q_{i,j}(s,t) = P_j(t) \text{ for all } i.$$

These conditions just say that the transition mechanism of each component of the vector process is $P_i(j)$. Call the vector process (X_k^1, X_k^2). Suppose that X^1 starts in i and X^2 starts according to π. Let $T = \min\{k: X_k^1 = X_k^2\}$. This T is a stopping time. Suppose that T is finite with probability 1. Let

$$X_k^3 = \begin{cases} X_k^2 & k \leq T \\ X_k^1 & k > T. \end{cases}$$

The process (X_k^1, X_k^3) is called a coupling, the interpretation being that the two processes evolve until they are equal, at which time they couple, and thereafter

remain equal. The usefulness of coupling depends on being able to get our hands on T: Let P_i^k be the law of the process after k steps started from i.

LEMMA 5. *(Coupling inequality).* $||P_i^k - \pi|| \leq P(T > k)$.

Proof. Take Q to be the distribution of (X_k^1, X_k^3). This Q has marginal distributions $P_i^k(\ )$ and π. Lemma 4 implies that

$$||P_i^k - \pi|| \leq Q(\Delta^c) = P(T > k).$$

□

Remarks. It is instructive to note that while the distribution of X_T is stationary (and so uniform in our examples) the time T is *not* a strong uniform time as we have defined it; for this requires $P(X_k \in A | T = k) = U(A)$ for all k.

Remarks. Example 1 gives an actual construction of $Q_{ij}(k\ell)$. It might be instructive to write down what $Q_{ij}(k\ell)$ is for this example. Griffeath (1975), Pitman (1976) or Goldstein (1979) show that the argument is tight in the sense that there is a coupling that achieves the total variation distance. This coupling cannot be taken as Markovian in general (that is, the bivariate process needn't be Markov).

Example 2. Random walk on the d-cube. Here $G = Z_2^d$. Take

$$P(\underline{0}) = p,\ P(1\ 0 \ldots 0) = P(0\ 1\ 00 \ldots 0) \ldots = P(0 \ldots 1) = (1-p)/d.$$

Here is a coupling argument, due to David Aldous, for bounding convergence to uniform. Consider two cubes. The process X_0^1 starts at zero, X_0^2 starts in a uniformly distributed position. The pair (X_i^1, X_i^2) evolves as follows: if X_i^1 and X_i^2 differ in an odd number of places, the two processes take independent steps according to P. If X_i^1 and X_i^2 differ in an even number of places, then with probability p each remains unchanged. If they don't stay the same, then pick an index j at random in $\{1, 2, \ldots, d\}$. If the jth component of X_i^1 and X_i^2 agree, change that component to its opposite (mod 2) in both processes. If the jth component of X_i^1 and X_i^2 do not agree, complement the jth component of X_i^1 and the *next* non-agreeing component of X_i^2 from j counting cyclically. This forces X_i^1 and X_i^2 to agree in two more coordinates. Once the number of disagreeing places is even, it stays even, so the coupled process "gets together" very rapidly. Of course, once $X_i^1 = X_i^2$, they stay coupled.

EXERCISE 6. Analyze the coupling time T and get a bound on the rate of convergence for Example 2 . Compare this with the right rate derived from Fourier analysis.

Matthews (1986b) has constructed non-Markovian couplings that give the right rate of convergence for the cube.

Coupling is a very widely used tool which has many success stories to its credit. Aldous (1983a) gives a number of card shuffling examples. Robin Pemantle

(1989) has given a marvelous coupling analysis of the familiar over-hand shuffle. For a range of reasonable models he shows that order n^2 shuffles are required to mix up n cards. Thus about 2,500 shuffles are required for 52 cards. This should be compared with 7 or 8 riffle shuffles, and the computation that a single riffle and single over-hand shuffle produce the same number of distinct permutations.

Aldous and Diaconis (1987a) and Thorisson (1987) study the relation between coupling and strong uniform times. Briefly, for any strong uniform time there is a coupling with the same time. Thus couplings can occur faster in principle.

Theorem 5 of Section C shows that couplings can only speed things up by a factor of at most 2. The example of simple random walk on the cube shows that this actually happens: it takes $\frac{1}{4}$ n log n $+cn$ steps to make the variation distance small; $\frac{1}{2}$ n log n $+cn$ steps are needed to make the separation small.

Despite the similarities, the connection is fairly formal. The way of thinking, and basic examples, can be very different. There is no known direct coupling argument to get anything better than n^2 for random transpositions, while strong uniform times or Fourier analysis show the right rate is order n log n. Similarly, there is no strong uniform time for the over-hand shuffle, or the shuffle that picks a card at random and switches it with a neighbor. Coupling can handle these problems.

F. First hits and first time to cover all.

(1) *Introduction.* Most of the work in this and the previous chapter has been devoted to estimating rates of convergence to uniformity. There are many other natural questions connectd to random walk. One may ask

- How long does it take to hit a fixed state (or set of states) from a given (or random) start?
- How long does it take to hit every state?
- How long until the first return to the starting state? How far away is the walk likely to get before first returning? How many states does the walk hit before first returning? What is the maximum number of times any state has been visited at first return?
- How long does a walk take before it hits a point previously hit (the birthday problem for random walk)?

David Aldous has introduced an important heuristic which suggests and explains answers to such questions, and sometimes allows a proof using only bounds on convergence to stationarity.

The idea is as follows. Suppose a specific random walk on a group G is rapidly mixing in the sense that the variation distance is less than $\frac{1}{2}$ after k steps with log k of order a polynomial in $\log|G|$. Then, the random walk forgets where it is rapidly, and successive steps may be thought about as the position of balls, dropped at random, into $|G|$ boxes.

Questions about balls in boxes are well understood. For example, the mean waiting time T until a ball is dropped into a fixed box is $|G|$ and

$$P\{\frac{T}{|G|} > t\} \to e^{-t} \text{ as } G \to \infty.$$

This suggests that a rapidly mixing random walk takes about $|G|$ steps to hit a fixed point and the waiting time is approximately exponential. A precise version is given in (2) below.

As a second example, the waiting time V for all boxes to have at least one ball is well studied as the coupon collector's problem. For balls dropped at random into $|G|$ boxes, it takes about $|G| \log|G|$ balls to have a good chance of filling all boxes. Results in Feller (1968, pg. 106) yield

$$P\{\frac{V - |G| \log|G|}{|G|} \leq x\} \to e^{-e^{-x}} \text{ as } |G| \to \infty.$$

This suggests that a rapidly mixing random walk takes about $|G| \log|G|$ steps to cover all points. (3) below gives some precise results due to Aldous and Matthews.

Section 4 points to what little is known about other problems on the list above.

(2) *First hit distributions.* The heuristics above are right "up to constants." One remarkable finding of Aldous (1982, 1983b) is that only one other feature of the walk enters. This is a measure of the amount of time the walk spends in its starting state in a short time period. Consider throughout a random walk on a finite group G. The transition mechanism is assumed to be aperiodic, and the uniform distribution on G is the stationary distribution.

Standard renewal theory implies that $R(s;t)$, the amount of time the walk spends in a fixed state s up to time t, is asymptotically $t/|G|$. Moreover

$$R = \lim_{t\to\infty} E\{R(s;t)\} - t/|G|$$

exists and is finite. By homogeneity R doesn't depend on s. Aldous (1983b) argues that for rapidly mixing walks, R can be interpreted as the mean number of visits the random walk spends in its initial state in a short time. For most of the examples in this and the previous chapter, $R = 1$.

With this notation, some careful results can be stated.

Theorem 9. *(Aldous). Let T_s be the first time a random walk starting in a uniformly chosen position hits state s. Then*

$$E(T_s) = R|G| \text{ for all } s \in G. \tag{1}$$

Let $\tau = \inf_k\{||P^{*k} - U|| \leq 1/2e\}$. Then

$$\sup_{t\geq 0} |P\{T_s > t\} - e^{-t/R|G|}| \leq \psi(\tau/R|G|), \tag{2}$$

with $\psi(x)$ tending monotonically to zero as x tends to 0.

Remarks. Part (2) makes precise the heuristics of the previous section. Consider a process like simple random walk on the d-cube. Then $|G| = 2^d$, and $\tau \doteq \frac{1}{4}d \log d$, $R = 1$, and (1) and (2) recapture the limiting results derived in Chapter

3H. Aldous (1983b) gives similar results for the first hitting time to arbitrary sets with any starting distribution.

Most random walks considered above have $\tau/|G| \to 0$. An exception is simple random walk on Z_n, where τ is of order n^2. The wait to first hit a point has a rather complicated distribution (see Chapter 3H). Flatto, Odlyzko, and Wales (1985) use Fourier analytic methods to get higher order correction terms.

(3) *Time to cover all.* Let G be a finite group and P a probability with convolutions that converge to uniform aperiodically. Let V be the first time that a random walk hits every point in G. Note that the distribution of V doesn't depend on the starting state. Let τ and R be as defined in Section 2. Aldous (1983a) proves

Theorem 10.

$$E\left|\frac{V}{R|G|\ log|G|} - 1\right| \leq \psi\left(\frac{log(1+\tau)}{log|G|}\right)$$

with $\psi(x)$ tending monotonically to zero as x tends to 0.

Remark. In the case of the cube, $\log(1+\tau) \sim \log$ d, $\log|G| \sim$ d log 2, so the ratio tends to zero. Analogs of the extreme value limits for the coupon collector's problem are not established in this generality. However, Matthews (1985) has established limit theorems for many of the examples where Fourier analysis can be successfully applied.

Usually the results follow the heuristic. For the cube there is an extra factor of 2. Matthews shows

$$P\left\{\frac{V - 2^n \log 2^{n+1}}{2^n} \leq x\right\} \to e^{-e^{-x}}$$

for all fixed x as n tends to infinity. Here $R \sim 1 + \frac{1}{n}$ which explains the 2.

Matthews' argument works by getting upper and lower bounds on the required probability. These apply to problems like first time for a Markov chain to hit every point in a finite state space or first time for Brownian motion to come within ε of every point on a high-dimensional sphere. The bounds merge as $|G| \to \infty$ for random walk problems.

(4) *Other problems.* There has been some work on special cases of the problems listed in (1) above. Aldous (1985) started to classify the kind of limiting behavior that can occur in the birthday problem for random walk. Diaconis and Smith (1988) have begun to develop a fluctuation theory (as in Chapter 3 of Feller (1968)). Some neat results emerge for nearest neighbor random walk on a 2-point homogeneous space. For example, on the n-cube, the probability that random walk starting at $(00\ldots0)$ hits a given point at any specified distance less than n before returning to zero tends to 1 as n tends to ∞. The probability tends to $1/2$ for $(1,\ldots,1)$.

This seems like a rich collection of reasonably tractable problems. Passing to the limit should give results for the approximating diffusions (e.g., Ornstein-Uhlenbeck process for the cube) in much the same way as results about simple random walk lead to results for Brownian motion.

G. Some open problems on random walk and strong uniform times.

Here is a small list of problems that seem worth careful work.

(1) *The slowest shuffle.* Arunas Rudvalis has suggested the following candidate for the slowest shuffle: At each time, the top card is placed either at the bottom, or second from the bottom, each with chance $\frac{1}{2}$. How long does it take to get random? Is this the slowest shuffle equally supported at 2 generating permutations?

(2) Let $G = Z_n$. Pick k points in G, and repeatedly choose one of them at random. This determines a random walk. What are the slowest k points (given no paritiy problems) — a "arc" near zero? (i.e. the set of points j with $|j| < k/2$.) What are the fastest k points? Andy Greenhalgh has shown how to get rate $n^{1/k}$ by an appropriate choice. What's the rate for "most" sets of k points? These questions are already non-trivial for $k = 3$. They are also worth studying when k grows with n.

(3) Moving on to other groups, Aldous and Diaconis showed that for most measures P on a finite group G, $||P * P - U|| \leq \frac{1}{|G|}$, so for G large, most measures are random after two steps. To get an interesting theory, constraints must be put on the support. Andre Broder asked the following: pick a pair of elements in S_n. Consider the walk generated by choosing both of these elements at random. It can be shown that such a pair generates S_n with probability 3/4 asymptotically. Is the walk random after a polynomial number of steps? Similar problems are worth investigating for any of the classical infinite families of finite simple groups (I'd try $PGL_n(q)$). Back on S_n; it seems that any "reasonable" shuffle gets random in at most a polynomial number of steps.

(4) *The 15 puzzle.* This familiar puzzle has 15 blocks arranged in a 4×4 grid. At each state, any of the blocks can be slid into the blank. Suppose uniform choices are made among the current possibilities.

Here is a simplified version: Consider the blank as a 16th block, and consider the puzzle on a "torus." An allowable move now involves picking one of the 16 squares at random, and then a direction (North, South, East, West) and "cycling" that square in the chosen direction. For example, the bottom row might change from 13, 14, 15, 16 to 16, 13, 14, 15 or to 14, 15, 16, 13. It is not hard to show that it takes order n^3 steps to randomize a single square (on an $n \times n$ grid). I presume that order n^3 log n steps suffice to randomize everything. For a 4×4, this gives about 90 "moves" to randomize. I presume this simplified version converges to uniform faster than the original 15 puzzle. Similar questions can be asked for other puzzles such as Rubic's cube.

(5) *The affine group.* Consider random walks of form $X_n = a_n X_{n-1} + b_n (\text{mod } p)$. Here p is a fixed number (perhaps a prime) and (a_n, b_n) are chosen at random: e.g., $a_n = 2$ or $\frac{1}{2} (\text{mod } p), b_n = \pm 1$. It seems that the right answer for these is $(\log p)^a$ for $a = 1$ or 2. The best that has been proved at present is order p^2 (see Diaconis and Shahshahani (1986a).

(6) *Thorp's shuffle.* A simple model for a random riffle shuffle has been described by Thorp (1973). Cut the deck exactly in half. Start to drop the cards from left or right hand as with an ordinary shuffle. At each time, choose left or

right with chance $\frac{1}{2}$. Whatever is chosen, drop that card, and then a card from the opposite half. Continue inductively. I think use of the mathematics of shuffle nets (or work on sorting in parallel) will allow an elegant solution to this problem.

(7) *Continuous groups.* We have no examples of a strong uniform time argument being used to get rates of convergence for a random walk on a compact, infinite group. It may be necessary to change the distance to the Prohorov metric. For problems like random reflections (see Diaconis and Shahshahani (1968a)) or random walk on the circle determined by repeatedly choosing a point in a small arc uniformly, there is convergence in total variation metric.

(8) *The cutoff phenomenon.* The most striking finding is the existence of sharp phase transition, $||P^{*k} - U||$ cutting down from 1 to zero in a relatively short time. It would be great to understand if this usually happens. As explained in problem (3) above, restrictions will have to be put on the support.

(9) *Relations between various approaches.* A curious feature of the examples is that *usually* if one method of attack works (e. g., Fourier analysis, or coupling, or strong uniform times), then all the methods work. There must be a reason. The greatest mystery is to understand the connections between the analytic and probabilistic methods. One place to start is "top in at random," the first example of Chapter 4. This can be done by strong uniform times and coupling. There *must* be a way to do it Fourier analytically.