Probability and Algorithms **Caltech CS150, Winter 2003**
Leonard J. Schulman **Scribe: Robert Forster**
Notes for lecture 11, February 19. Strong stationary time for riffle shuffle.

Example: *Random Walk on a grid*
Consider a random walk on in $\mathbb{Z}^d$ for d fixed. The distribution will be fairly smooth across a ball of radius $r = \sqrt{n}$ after time n. The real distribution will be some multi-dimensional binomial, which has a quadratic Taylor expansion about the starting position; the distributions conditional on various starting points that are within distance $\sqrt{n}$ of each other, are very similar. We can therefore say that the walk has "forgotten about it's origin to within a ball of radius $\sqrt{n}$". To make this rigorous we use a coupling argument.

Coupling
Consider starting two walks, X and Y, from two nearby points on the grid. Let K_i be the offset in the i^{th} direction between these two starting points. Each walk follows the Markov process:

$$\begin{cases} p = 1/2 & \text{stay put} \\ p = 1/4d & \text{take each edge} \end{cases}$$

Couple the walks, by considering X and Y as independent particles, except that whenever X moves along the i^{th} axis, Y will also. Once their positions agree in a coordinate, they will subsequently use the "same coin" for determining movement within that coordinate. Thus X and Y will continue to agree on any coordinate after they first agree.

In dimension d, coupling time is related to K_d, a 1-D problem. Consider each 1-D walk, and the time for X and Y to agree as a function of their initial displacement. Given the coupling described above, the 1-D walk starts at K_d, stops at 0, and moves according to:

$$\begin{cases} p = 3/8 & \text{stay put} \\ p = 1/4 & \text{move 1 right} \\ p = 1/4 & \text{move 1 left} \\ p = 1/16 & \text{move 2 right} \\ p = 1/16 & \text{move 2 left} \end{cases}$$

We know that for random walk without drift, there is a function c_1 such that

$$\forall \delta \ : \ P(\text{doesn't hit 0 by time } c_1(\delta) K_d^2) < \delta$$

applying a union bound gives, for some $c_2(d)$,

$$P(\text{particles don't couple by time } c_2(d) \cdot \max_i \{K_i^2\}) < 1/10$$

This gives a bound on the variation distance between the distributions.

Example: *A Real Shuffle?*
Consider the following model of an "imperfect riffle shuffle". First split the deck of n cards into left and right halves, with the number of cards in one "half" being distributed as Binomial$(n, 1/2)$. Model the random interleaving of cards by a shuffle, which consists of dropping the bottom card from one side, with the side chosen with relative probability given by the number of cards currently in that side.

How fast does the deck mix? There are $52!$ arrangements, and shuffles are a stochastic process on this state space. We can think of a particular shuffle as an element $g \in S_n$, as S_n acts on the deck configurations. How long to reach a uniform distribution? Initially we begin with a distribution supported entirely on the initial configuration of the deck; eventually we tend exponentially tend towards the uniform distribution. We can define a stochastic transition matrix A, where

$$\vec{x}_{t+1} = \vec{x}_t A$$

describes how the initial distribution vector is changed by the action of one shuffle. Here

$$A_{ij} = \sum_{g \in S_n, g(i)=j} p(g)$$

where $p(g)$ for the specified shuffle algorithm. (In this example there is only one g such that $g(i) = j$). Repeated shuffles result in powers of A. We achieve a well-shuffled deck after t shuffles if:

$$\forall i, j \ (A^t)_{ij} \approx \frac{1}{n!}$$

Instead of directly analyzing this riffle shuffle, consider the "backward shuffle" g^{-1}, where $g \in S_n$ corresponds to the "forward" riffle shuffle. We'll show that the convergence of these walks is the same in "L_∞". To be specific, if $T_{\max}$ is the time until every A^t_{ij} is bounded by $3/2$ times π_j, then $T_{\max}$ is an upper bound on the mixing time, which in turn is no more than $O(\log \max_j 1/\pi_j)$ times $T_{\max}$.

Claim: *The backward shuffle has the same $T_{\max}$ as the forward one.*
The claim applies to any Markov chain given by the action of a group on a set, not just for the riffle shuffle.

Proof: This is because the matrix B to describe the backward shuffle is just a reordering of the matrix A for the forward one. Specifically,

$$B_{ij} = \sum_{g \in S_n, g(i)=j} p(g^{-1}) = \sum_{g \in S_n, g(j)=i} p(g) = A_{ji}.$$

Therefore also $(B^t)_{ij} = (A^t)_{ji}$.

This argument is good enough to show that poly-time convergence of one chain implies the same for its backward chain, but its not quite good enough to show that their mixing times are asymptotically equal, due to the gap between the L_1 and L_∞ formulations. So for example last time we showed the same time bound (to within constants) of $O(n \log n)$ on the random-to-top and top-to-random shuffles; this argument wouldn't automatically give us that conclusion, though it shows that, given the bound on one process, the other can't be more than $O(n \log^2 n)$.

Of course it is a matter of discretion whether one is interested in L_∞ or L_1 convergence; in the former, this argument shows that a precise equivalence between the forward and backward analyses.

Analyzing the Backward Shuffle:
The backward shuffle can be thought of as follows. First, assign a random bit (0,1) to each card in the deck uniformly. Now, "uncut" the deck by putting all the 0's on the bottom (but preserving their relative order), and similarly putting all the 1's on top. Note that the random bit assignment gives the required binomial distribution of cut sizes. Verify that this shuffle is indeed the "backwards shuffle" of the riffle shuffle.

We will now associate a label to each card. The label consists of several registers. The "low order" register contains the original position (between 1 to n) of the card. The next register contains the random bit that the card receives in the first shuffle. The next higher register contains the random bit that the card receives in the second shuffle, and so on. Continue this process, where subsequent shuffles extend the length of each card's label by one more bit. Note that the deck is at all times in lexicographical order with respect to these strings (i.e., the highest string values are on top).

Claim: A strong stationary time for the shuffle is when all cards get a distinct binary string (so that the initial index between 1 and n in the low-order register is no longer relevant). Analysis of this SST comes down to the "birthday problem:" for a "year" of length T and for K people with randomly chosen birthdays, what is the probability that no two share a birthday? The answer is, "high" for $K << \sqrt{T}$ and "low" for $K >> \sqrt{T}$. We conclude that

$$T_{mix} \approx 2 \log n$$

(see Aldous for a more accurate analysis). Diaconis and Baker analyzed this in the specific case of $n = 52$, and found that the variation distance from uniform showed a relatively sharp drop between 6 and 7 shuffles. So fairly good randomization of a deck requires at least 7 shuffles.