Card Shuffling Report

STAT 157 Project

Feng Cheng Lucy Meng

Top-to-random shuffle

It is common to model card shuffles by random walks on groups. Given a probability distribution on a finite group G, we define a *left random walk on G* (with increment distribution μ) if it is a Markov chain with state space G and transition probability given by

$$P(g, hg) = \mu(h)$$

for all $g, h \in G$.

It is easy to check that the uniform distribution U is stationary by definition: for all $h \in G$, we have

$$\sum_{g\in G} U(g)P(g,h) = \sum_{k\in G} U(k^{-1}h)P(k^{-1}h,h) = \sum_{k\in G} U(k^{-1}h)\mu(h) = \frac{1}{|G|} = U(h),$$

where the second equality is justified by the fact that the right multiplication map $\rho_h \colon G \to G$ is one-to-one and onto.

Card shuffling is basically a random walk on S_n based on a given increment distribution μ. An effective shuffling technique should be able to reach every possible permutation, and thus usually the shuffle chain is irreducible. This means the uniform measure U over G is the unique stationary distribution of a shuffle chain.

The top-to-random shuffle is performed by taking the top card and putting it back into the deck at random. In the top-to-random shuffle case, the increment distribution μ is given by

$$\mu(\sigma) = \begin{cases} 1/n & \text{if } \sigma = (k\ \cdots\ 2\ 1) \text{ for } 1 \le k \le n; \\ 0 & \text{otherwise.} \end{cases}$$

Since here $\mu(\mathrm{id}) > 0$, the top-to-random chain is aperiodic.

By the aperiodicity and irreducibility of the chain, it becomes meaningful to use the total variation distance as a measure between the t-step transition probability measure $P^t(\sigma, \cdot\,)$ and U. Recall

$$d(t) = \max_{\sigma\in S_n} \|P^t(\sigma, \cdot\,) - U\|_{\mathrm{TV}} = \frac{1}{2} \max_{\sigma\in S_n} \sum_{\omega\in S_n} \bigl|P^t(\sigma,\omega) - U(\omega)\bigr|.$$

Note that fix σ, we have

$$\begin{aligned} \sum_{\omega\in S_n} \bigl|P^t(\sigma,\omega) - U(\omega)\bigr| &= \sum_{\omega\in S_n} \bigl|P^t(\sigma,\omega\sigma) - U(\omega\sigma)\bigr| \\ &= \sum_{\omega\in S_n} \bigl|P^t(\mathrm{id},\omega\cdot\mathrm{id}) - U(\omega)\bigr|, \end{aligned}$$

and therefore

$$\begin{aligned} d(t) &= \frac{1}{2} \max_{\sigma\in S_n} \sum_{\omega\in S_n} \bigl|P^t(\mathrm{id},\omega) - U(\omega)\bigr| \\ &= \|P^t(\mathrm{id}, \cdot\,) - U\|_{\mathrm{TV}}. \end{aligned}$$

(Clearly the above holds in general for any group. We may omit the id as well because the starting state does not matter.) To find the mixing time $t_{\mathrm{mix}}(\epsilon) = \min\{t : d(t) \le \epsilon\}$ is to bound $d(t) = \|P^t(\mathrm{id}, \cdot\,) - U\|_{\mathrm{TV}}$.

It was proved in [AD86] that for $\alpha \ge 0$, it holds that

$$d(n\log n + \alpha n) \le e^{-\alpha} \quad \text{and} \quad \liminf_{n\to\infty} d(n\log n - \alpha n) \ge 1 - 2e^{2-\alpha}. \tag{1}$$

This means that $t_{\mathrm{mix}}^{(n)} = n\log n$. We will give a sketch proof of the upper bound here. A proof of the lower bound can be found in the original paper or [LPW17] section 7.4.1.

We first note that the top-to-random chain (X_t) has the following property. If t is one shuffle after the first time the original bottom card (say ♠K) reaches the top, then the deck of cards is completely random. We claim that the orderings of cards under ♠K are all equally likely.

This is easy to show by induction. At $t = 0$ this is trivial. Suppose at time t this holds. If the top card D_t is inserted above ♠K, then the inductive hypothesis still holds because the cards below ♠K remain unchanged. If D_t is inserted below ♠K, since D_t can be in any position, the orderings of cards below ♠K are still equiprobable. This property will turn out to be very useful soon, and we will call τ_{top} the first time the original bottom card reaches the top plus one.

Recall the coupon collector random variable τ_{coupon} is the the total number of coupons collected when the set first contains all n types of coupons. It is not hard to see that

$$\tau_{\mathrm{top}} = G_1 + G_2 + \cdots + G_n = \tau_{\mathrm{coupon}},$$

where the G_j's are independent, and each $G_j \sim \mathrm{Geometric}(1/j)$.

Let (X_t) be an irreducible Markov chain with stationary distribution π. A *stationary time* τ for (X_t) started at x is a stopping time such that for all state y,

$$\mathbf{P}_x(X_\tau = y) = \pi(y).$$

Colloquially this is the time when (X_t) reaches stationarity. We further define τ to be a *strong stationary time* if it is a stationary time and X_τ is independent of τ. This is equivalent to saying that for all t and y,

$$\mathbf{P}_x(\tau = t, X_\tau = y) = \mathbf{P}_x(\tau = t)\pi(y).$$

Our τ_{top} is exactly a strong stationary time for the top-to-random chain.

We now cite two theorems from [LPW17] to conclude this part.

Proposition (6.11). Given a strong stationary τ for (X_n) with starting state x, we have the inequality

$$\|P^t(x, \cdot\,) - \pi\|_{\mathrm{TV}} \le \mathbf{P}_x(\tau > t).$$

Hence for τ_{top},

$$d(t) = \|P^t - U\|_{\mathrm{TV}} \le \mathbf{P}(\tau_{\mathrm{top}} > t). \tag{2}$$

Proposition (2.4). For any $\alpha \ge 0$,

$$\mathbf{P}(\tau_{\mathrm{coupon}} > \lceil n\log n + \alpha n\rceil) \le e^{-\alpha}. \tag{3}$$

Combining (2) and (3) with $\tau_{\mathrm{top}} = \tau_{\mathrm{coupon}}$, and we conclude that $d(n\log n + \alpha n) \le e^{-\alpha}$, as desired.

Riffle Shuffle

The riffle shuffle is performed by dividing the deck into two stacks and interleaving them. It can be mathematically modeled in a variety of ways. We give the two most straightforward ways below:

1. Let $M \sim \text{Binomial}(n, 1/2)$ be the number of cards in the left deck and $n - M$ be the number of cards in the right deck. There would be in total $\binom{n}{M}$ ways to riffle the two together.
2. Let the two decks still be of M cards and $n - M$ cards each. At time t suppose the left deck has a remaining cards and the right deck has b remaining cards, we drop the left (resp. right) bottom card with probability $\frac{a}{a+b}$ (resp. $\frac{b}{a+b}$).

A simple exercise with binomial coefficients shows that the two are equivalent formulations. The increment distribution μ here is given by

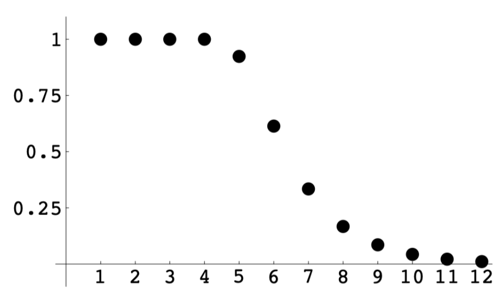
$$\mu(\sigma) = \begin{cases} (n+1)/2^n & \text{if } \sigma = \text{id}; \\ 1/2^n & \text{if } \sigma \text{ has two rising sequences;} \\ 0 & \text{otherwise.} \end{cases}$$

This model is called the *Gilbert-Shannon-Reeds (GSR) model*.

The famous [BD92] paper presented a explicit formula for $d(t)$:

$$\|P^k - U\|_{\text{TV}} = \frac{1}{2} \sum_{j=0}^{n-1} A(n,j) \left| \binom{n + 2^k - j - 1}{2^{kn}} - \frac{1}{n!} \right|,$$

where $A(n, j)$ is the *Eulerian number*, which computes the number of permutations on n symbols with j descents. The important thing that it can be recursively computed, and hence we may plug in $n = 52$ and different k's, and get figure 1.



1	2	3	4	5	6
1.0000	1.0000	1.0000	1.0000	0.9237	0.6135
7	8	9	10	11	12
0.3341	0.1672	0.0854	0.0429	0.0215	0.0108

Figure 1: $n = 52, 1 \leq k \leq 12$

[BD92] also gave the asymptotic result: if n cards are riffle shuffled $k = \frac{3}{2}\log_2(n) + c$ times, then for large n,

$$\|P^k - U\|_{\text{TV}} = 1 - 2\Phi\Big(\frac{-2^{-c}}{4\sqrt{3}}\Big) + O(n^{-1/4}), \tag{4}$$

where n is the normal CDF. Hence $t_{\text{mix}}^{(n)} = \frac{3}{2}\log_2(n)$.

The cutoff phenomenon

The main results (1) and (4) are very similar in nature, which is known as the cutoff phenomenon. A bit informally, we call a sequence of Markov chains indexed by the state space size n has a *cutoff* at k_0 if $d\big(k_0 + o(k_0)\big) \approx 0$ and $d\big(k_0 - o(k_0)\big) \approx 1$. See figure 2. Asymptotically as $n \to \infty$ we should see the plot becoming a step function at k_0. Of course k_0 is just $t_{\text{mix}}^{(n)}$.[1]

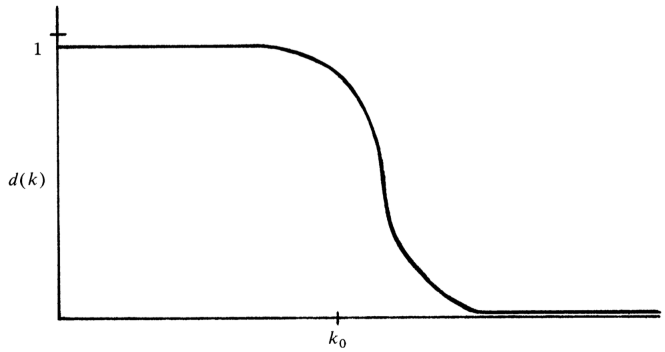


Figure 2: cutoff at k_0

It is clear that (1) and (4) are examples of the cutoff phenomenon, and in fact most shuffle models have this phenomenon. Note that even for a relatively small $n = 52$, the cutoff phenomenon in figure 1 is already quite evident.

[1]We refer to chapter 18 of [LPW17] for a more general and precise definition of the cutoff phenomenon.

References

[AD86] David Aldous and Persi Diaconis. "Shuffling Cards and Stopping Times". *The American Mathematical Monthly* 93.5 (May 1986), pp. 333–348. ISSN: 1930-0972. DOI: 10.1080/00029890.1986.11971821. URL: http://dx.doi.org/10.1080/00029890.1986.11971821.

[BD92] Dave Bayer and Persi Diaconis. "Trailing the Dovetail Shuffle to its Lair". *The Annals of Applied Probability* 2.2 (May 1992). ISSN: 1050-5164. DOI: 10.1214/aoap/1177005705. URL: http://dx.doi.org/10.1214/aoap/1177005705.

[DF23] Persi Diaconis and Jason Fulman. *The mathematics of shuffling cards*. American Mathematical Society, 2023.

[LPW17] David Levin, Yuval Peres, and Elizabeth Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2017.

Our treatment mostly follows [LPW17], with inspirations from the other sources as well.