The Mathematics of Card Shuffling

Cameron Martin

University of Toronto

December 8, 2020

Introduction - a simple card game

- Guess which card I'm going to flip over next!
- 2 extremes: 1) you know the exact order, and 2) you know nothing about the order of the deck.
- 1) Expected number of correct guesses: 52.
- 2) Expected number of correct guesses: $\frac{1}{52} + \frac{1}{51} + \ldots + 1 \approx 4.54$.
- How much shuffling does it take to get the expected number of correct guesses down from 52 to 4.54 (using some optimal guessing strategy)?

1	2	3	4	5	6	7	8	9	10
31.17	19.69	12.92	8.80	6.56	5.51	5.01	4.76	4.65	4.60

Table 1: Number of cards guessed correctly after k shuffles of 52 cards. See Table 5 in Bayer & Diaconis (1992) [1].

Modeling shuffling techniques

Suppose a deck has n cards, labeled $1, 2, \ldots, n$.

Gilbert-Shannon-Reeds (GSR) model

- Cut the deck, with the probability of removing j cards being $\binom{n}{j}2^{-n}$.
    - In other words, the cut point C is *binomially distributed*: $C \sim \text{Binomial}(n, 1/2)$.
- If A cards remain in pile 1 and B cards remain in pile 2, the probability of dropping the next card from pile 1 is $\frac{A}{A+B}$.

Top-in-at-random

- Remove the top card from the deck and place it at a uniformly distributed random point in the deck.

A crash course on Markov chains (1/4)

Definition 1

A (discrete-time) *Markov chain* on (countable/finite) state space Ω is a random sequence of elements $X_0, X_1, X_2, \ldots$ in Ω such that X_k depends only on X_{k-1} (the *Markov property*).

Examples

1. Discrete random walk on the circle: $\Omega = \{0, 1, 2, \ldots, n-1\}$, $X_0 = 0$, and
$$X_k = \begin{cases} X_{k-1} - 1 (\text{mod } n), \\ X_{k-1}, \\ X_{k-1} + 1 (\text{mod } n), \end{cases} \qquad \text{each with equal}$$
probability $\frac{1}{3}$.
2. Discrete random walk on the circle with skips: $\Omega = \{0, 1, 2, \ldots, n-1\}$, $X_0 = 0$, and $X_k = (X_{k-1} \pm 2) (\text{mod } n)$ with equal probability $\frac{1}{2}$.
3. Discrete not-so-random walk: $\Omega = \{0, 1, 2, \ldots, n-1\}$, $X_0 = 0$, and $X_k = (X_{k-1} + 1) (\text{mod } n)$.

A crash course on Markov chains (2/4)

Definition 2

A Markov chain $(X_k)_{k\geq 0}$ on $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ is *irreducible* if, given arbitrary states $\omega_i, \omega_j \in \Omega$, there is a positive probability of reaching ω_i from ω_j.

Let $p_{ii}^{(k)}$ be the probability of the Markov chain returning to state ω_i in k steps.

Definition 3

The *period* d_i of a state ω_i is given by $d_i = \gcd\{k \geq 1 : p_{ii}^{(k)} > 0\}$. If $p_{ii}^{(k)} = 0$ for every $k \geq 1$, we set $d_i = \infty$.

Definition 4

A state ω_i is said to be *aperiodic* if $d_i = 1$, and a Markov chain $(X_k)_{k\geq 0}$ is said to be *aperiodic* if all states in Ω are aperiodic.

Fact: If $p_{ii}^{(1)} > 0$ (i.e. you can remain at state ω_i), then ω_i is aperiodic.

A crash course on Markov chains (3/4)

Definition 5

A Markov chain $(X_k)_{k\geq 0}$ on $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ has stationary distribution Q if it is the case that if we start with the distribution Q and take one step of the Markov chain, we still have the distribution Q.

Examples

- 1) Discrete random walk on the circle with skips: suppose $X_k = i$ with probability $Q(i) = \frac{1}{n}$ for each i. Then the probability of X_{k+1} being in state i is:

$$\mathbb{P}\left(X_{k+1} = i\right) = \sum_{j=0}^{n-1} \mathbb{P}\left(X_{k+1} = i | X_k = j\right) \mathbb{P}\left(X_k = j\right)$$

$$= \mathbb{P}\left(X_{k+1} = i | X_k = i-2\right) Q(i-2) + \mathbb{P}\left(X_{k+1} = i | X_k = i+2\right) Q(i+2)$$

$$= \frac{1}{2}\frac{1}{n} + \frac{1}{2}\frac{1}{n} = \frac{1}{n} = Q(i).$$

  The uniform distribution on $\{0, 1, \ldots, n-1\}$ is a stationary distribution of this random walk.

A crash course on Markov chains (4/4)

Theorem 6

If a Markov chain is irreducible and aperiodic with stationary probability distribution Q, then for every initial value $X_0 = \omega$ and $S \subset \Omega$,

$$\lim_{n\to\infty} \mathbb{P}(X_n \in S) = \sum_{\omega\in S} Q(\omega).$$

Examples

- 1) Discrete random walk on the circle: irreducible, aperiodic, stationary uniform distribution. For any starting point, we have $\lim_{k\to\infty} \mathbb{P}(X_k = i) = \frac{1}{n}$ for every i.
- 2) Discrete random walk with skips, $n = 6$: aperiodic, stationary uniform distribution, but *not* irreducible.
- 3) Discrete not-so-random-walk: irreducible, stationary uniform distribution, but *not* aperiodic.

What is a shuffle?

- With a deck of n cards, consider S_n, the set of all permutations of $\{1, 2, \ldots, n\}$. There are $n!$ possible arrangements of cards.
- A *shuffle* of a deck of n cards is a random walk on S_n, which can be described by a *Markov chain* $X_0, X_1, X_2, \ldots$, where each X_k is some arrangement of the deck obtained in a random fashion from X_{k-1}.
- For a shuffling technique to work for its intended purpose, we need that for sufficiently large k, X_k is (approximately) uniformly distributed on S_n—that is, if you start with some known arrangement X_0, after k steps every arrangement of the n cards is (approximately) equally likely.
- We want to apply the main result from earlier to our shuffling algorithms.

Top-in-at-random shuffle as a Markov chain

The top-in-at-random shuffle on a deck with n cards can be represented as a Markov chain on S_n (the set of permutations of n elements). Consider $X_0 = \{1, 2, \ldots, n\}$, the "new deck order". Then, to obtain X_{k+1} from X_k, we simply remove the first element and place it somewhere uniformly at random.

- This Markov chain is *irreducible* because we can get to any arrangement by swapping with the first element (just a fact about permutations, we could say "S_n is generated by transpositions of 1 and i for $2 \leq i \leq n$");
- It is *aperiodic* because it is possible for the card to be placed back in the same spot;
- It has stationary uniform distribution.

By the main theorem, this Markov chain will converge to the uniform distribution on S_n—it will genuinely mix up a deck of cards!

GSR riffle shuffle as a Markov chain

- It is not entirely obvious that this Markov chain is irreducible—if time permits we'll prove this later;
- It is aperiodic, which can be seen from the fact that under this model, it is possible for 0 or n cards to be removed from the top;
- It has stationary uniform distribution (similar computation to ones we did earlier, just with a lot more terms).

Total Variation Distance (TVD)

- We know these Markov chains will converge to the uniform distribution, but how long will that take?
- How can we measure how far we are from the uniform distribution after k shuffles?
- The number of correct guesses was a metric on probability measures
- A more canonical metric is the *total variation distance*

Definition 7

The *total variation distance* between two distributions P and Q on a state space Ω is defined by

$$d_{TV}(P, Q) = \frac{1}{2} \sum_{\omega \in \Omega} |P(\omega) - Q(\omega)| .$$

The factor of $\frac{1}{2}$ is there so that $0 \leq d_{TV}(P, Q) \leq 1$.

TVD and Shuffling

For a given shuffling algorithm and its associated Markov chain, we aim to construct a graph of total variation distance between the shuffling and uniform distributions vs. the number of shuffles k. The naïve way to do this is as follows:

- Start with $N \gg n!$ copies of the same deck of n cards;
- Simulate the shuffling algorithm on each deck k times;
- Count the number of copies of each possible arrangement of cards to obtain an empirical distribution $Q^{(k)}$ on S_n;
- Compute the total variation distance between $Q^{(k)}$ and U, the uniform distribution: $d_{TV}(Q^{(k)}, U) = \sum_{i=1}^{n!} \left| Q^{(k)}(i) - \frac{1}{n!} \right|$.

This can only be done for small decks.

Results

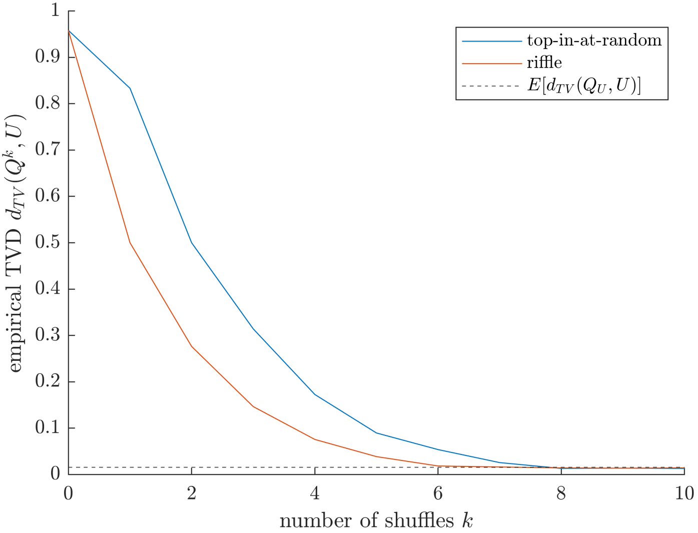


Figure 1: Empirical TVD vs number of shuffles for each

Justification for results

Top-in-at-random

- Consider the bottom card of the deck—under this shuffling scheme, it will slowly (randomly) rise in the deck until it reaches the top, and then it will be removed and placed somewhere at random
- When this happens at random time T, the deck is truly randomized
- When do we expect this to happen?
- $\mathbb{E}T = n + \frac{n}{2} + \frac{n}{3} + \ldots + \frac{n}{n} \approx n \log n$.
- For $n = 4$, $\mathbb{E}T \approx 8.3$.

Justification for results

Riffle

Definition 8

A *rising sequence* is a "maximal subset of an arrangement of cards, consisting of successive face values displayed in order."[1]

Examples

{3,1,5,2,4,6,7} consists of rising sequences $\{1, 2\}$, $\{3, 4\}$, $\{5, 6, 7\}$; $\{4, 1, 2, 5, 6, 3\}$ consists of rising sequences $\{1, 2, 3\}$ and $\{4, 5, 6\}$.

- There are at most n rising sequences in an arrangement of n cards.

Theorem 9

If n cards are riffle-shuffled k times, then the probability that the deck is in arrangement π is $\binom{2^k+n-r}{n}/2^{kn}$, where r is the number of rising sequences in π.

Proof of theorem

Theorem 10

If n cards are riffle-shuffled k times, then the probability that the deck is in arrangement π is $\binom{2^k+n-r}{n}/2^{kn}$, where r is the number of rising sequences in π.

We'll use the following fact without proof: riffle shuffling a deck k times is equivalent to splitting the original deck into 2^k piles and riffling them together.

Proof.

Suppose some arrangement π has r rising sequences. To obtain these rising sequences in the shuffled deck, we must cut the deck at the r points separating those rising sequences from each other. The remaining $2^k - r$ cuts can be placed arbitrarily. The number of ways to place $2^k - r$ cuts among $n + 1$ spots (n cards yield $n + 1$ possible cut locations) is $\binom{2^k+n-r}{n}$ (see the Wikipedia page on "stars and bars" arguments). There are 2^{kn} possible shuffles, so we get the desired probability. □

Revisiting the GSR model (time permitting)

- Under the GSR model, the probability of dropping alternate cards (i.e. one from pile 1 and then one from pile 2 or vice versa) is about 50%
- However, some research has shown that professional dealers do this about 80% of the time
- Can we modify the GSR model to more closely resemble what professional dealers do? And does it actually improve the mixing properties?

Revisiting the GSR model

Let $0 \leq \alpha \leq 1$. After cutting the deck, drop the first card according to the same rule as the GSR model. For any subsequent drop, suppose A cards remain in the first pile and B remain in the second. Let the outcome of any particular drop be

$$D_j = \begin{cases} 1, & \text{the } j\text{th drop was from pile 1} \\ 2, & \text{the } j\text{th drop was from pile 2.} \end{cases}$$

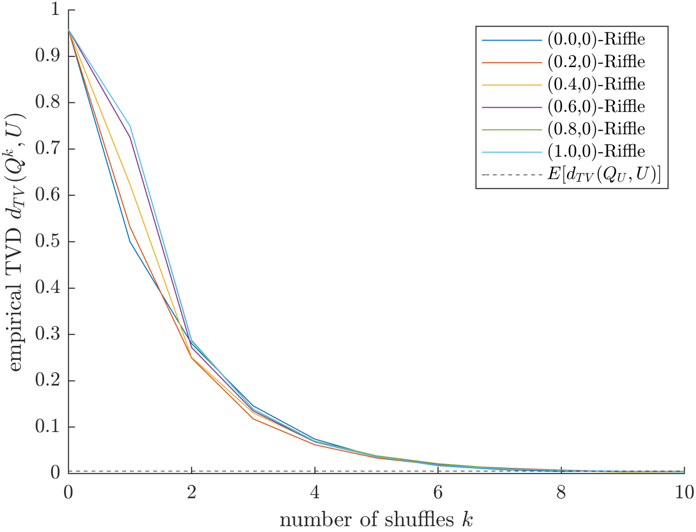
If the previous drop was from the first pile, then drop from the first pile with probability

$$\mathbb{P}(D_j = 1 | D_{j-1} = 1) = \max\left\{0, \left(1 - \alpha\left(\frac{A+B}{A}\right)\right)\left(\frac{A}{A+B}\right)\right\}.$$

If the previous drop was from the second pile, then drop from the first pile with probability

$$\mathbb{P}(D_j = 1 | D_{j-1} = 2) = \min\left\{1, \left(1 + \alpha\left(\frac{A+B}{A}\right)\right)\left(\frac{A}{A+B}\right)\right\}.$$

Revisiting the GSR model results



Conclusions/Future Work

- Shuffling is conveniently represented as a random walk/Markov chain on S_n;
- There are explicit results for only some shuffling algorithms—a more convenient approach is often to just create a graph of some distance vs. number of shuffles;
- Using rising sequences and other tricks, it is possible to write much more efficient code than the naïve approach I took;
- It would be nice to find an explicit result like that of the GSR model for the modified GSR.

Thank you!!!

Thanks for coming to my talk. I hope you enjoyed it and learned something!

References

[1] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313, 1992.

[2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

[3] Brad Mann. How many times should you shuffle a deck of cards?

[4] Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, 1988.

[5] E. Gilbert. Theory of shuffling, technical memorandum. *Bell Labs*, 1955.

[6] J. Reeds. Unpublished manuscript. 1981.

[7] Richard A. Epstein. *The Theory of Gambling and Statistical Logic, Revised ed.* Academic Press, 1977.