Card Shuffling and MCMC Paradigm

2016 Stochastic Process Course Project

You Chen, Shu Wang, Ze Yang, Biaojian Feng

advised by
Prof. Jun Luo

Antai Collage of Econ & Mgmt., Shanghai Jiao Tong University

May 8, 2016

Illustration of Card Shuffle Problem

Given a deck consisting n cards.

- The **sample space** $\Omega := \{\text{All permutations of the deck.}\}$.
- The **cardinality** of sample space: $|\Omega| = n!$

We want to sample $\omega \in \Omega$, such that every x arises with probability $\mathbb{P}(\omega) = \frac{1}{n!}$. We want to consider a random variable instead of permutations.

Definition ("Index" of deck permutation)

Define $X : \Omega \to \mathbb{N}^+$ be the **lexicographic order** of deck permutation; X is a random variable over $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, $X \sim \text{Uniform}(\{1, 2, ..., n!\})$.

$$X\left(\boxed{\begin{matrix}\text{A}\\ \spadesuit\end{matrix}}\boxed{\begin{matrix}2\\ \spadesuit\end{matrix}}\boxed{\begin{matrix}3\\ \spadesuit\end{matrix}}\right) = 1 \qquad X\left(\boxed{\begin{matrix}2\\ \spadesuit\end{matrix}}\boxed{\begin{matrix}3\\ \spadesuit\end{matrix}}\boxed{\begin{matrix}\text{A}\\ \spadesuit\end{matrix}}\right) = 4$$

Sampleability of Uniform($\{1, 2, ..., n!\}$)

Sampling a random deck is equivalent to sampling X. For a standard deck, $n = 52$. How can we sample it?

(1) Roll a perfect dice with 52! facets.
  - In simulation, "dice" is actually a **random number gernerator (rng)** in computer.
  - The dice provided by GNU C library has at most $2^{31} - 1$ facets, which is the largest **unsigned int** number in 32-bit system.
  - However, $52! \approx 2^{257}$.

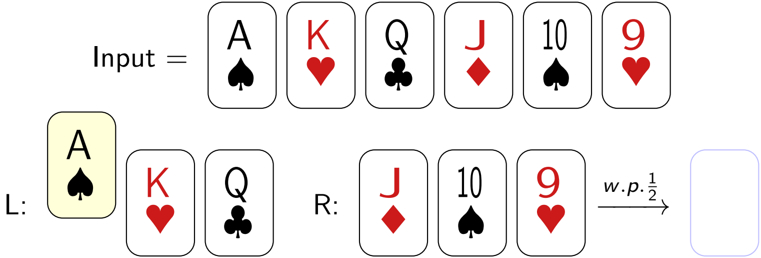
(2) Roll 52 independent dices, with $52, 51, 50, ..., 2, 1$ facets respectively.

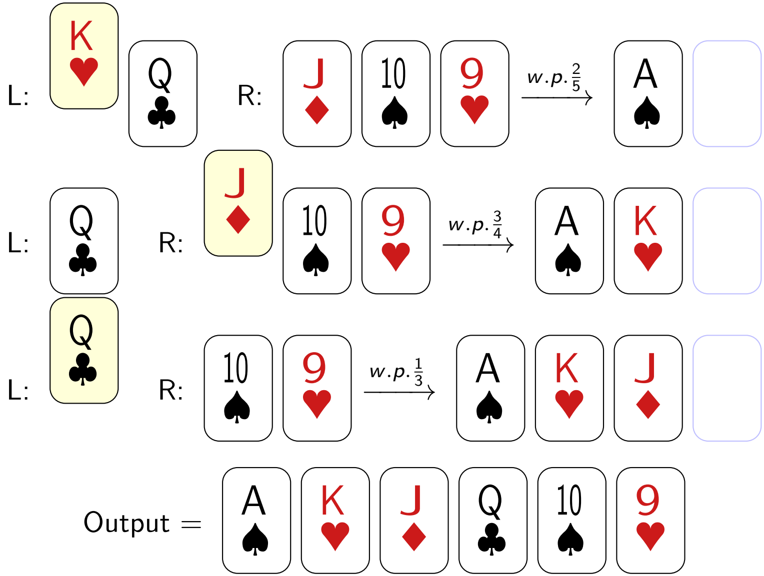
(3) Shuffle, simulate the 52! facets' dice.

Modeling Card Shuffle: Shuffling Methods

Example (Riffle Shuffle)

1. Split the deck into two parts in each hand.
2. Drop the card one after another to merge into new deck, deck card from left hand w.p. $\frac{\#\text{ of cards in left hand}}{\#\text{ of cards remaining}}$.

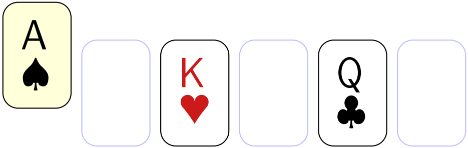



L:
R:
$w.p.\frac{2}{5}$
L:
R:
$w.p.\frac{3}{4}$
L:
R:
$w.p.\frac{1}{3}$
Output =

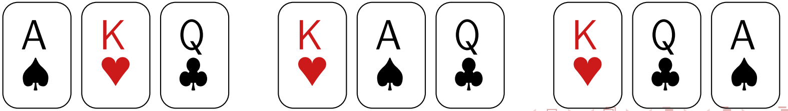
Modeling Card Shuffle: Shuffling Methods

Example (3-cards Deck Top-in Shuffle)

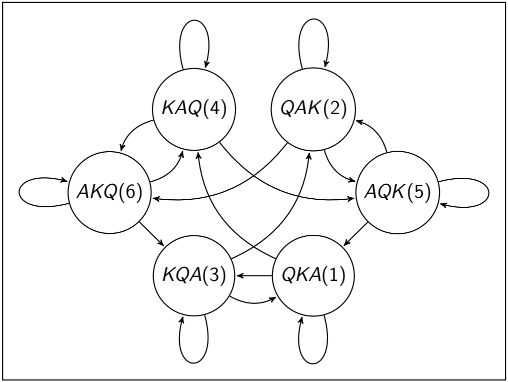
We consider a 3-cards deck shuffled by top-in method. That is, we take one card from the top and insert it at a random position in the deck. If initially, $X_0 = X(\{\text{AS, KH, QC}\}) = 6$



one out of three slots is picked to insert AS with equal probability, hence there are three outcomes w.p. $\frac{1}{3}$



Apply shuffling operation to the initial deck recursively, we obtain a process $\{X_t : t \geq 0\}$. X_t is independent of all history except for X_{t-1}. So we can visualize it as a markov process.



$$P = \begin{pmatrix} \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

General MCMC Approach

Input

1. a sample space Ω (may be large or complex).
2. a positive weight function $w : \Omega \to \mathbb{R}^+$.

Goal: We want to draw sample $x \in \Omega$ with some distribution $\pi(x)$ proportional to $w(x)$, i.e.

$$\pi(x) = \frac{w(x)}{\sum_{x \in \Omega} w(x)}$$

The MCMC algorithm is a general iterative solution to this sampleability problem. We construct a markov chain $\{X_t : t \geq 0\}$, such that X_t converges to our desired sample in (its stationary) distribution. That is:

$$\lim_{t \to \infty} P^{[t]}(x, y) = \lim_{t \to \infty} \mathbb{P}\left(X_t = y | X_0 = x\right) = \pi(y)$$

Card Shuffle as MCMC

Input

1. The sample space $\Omega = \{\text{All permutations of the deck.}\}$
2. We want to sample uniform distribution i.e., $w(k) = 1$.

Goal: Draw sample $X \sim \text{Uniform}(\{1, 2, ..., n!\})$ i.e.

$$\pi(k) = \frac{w(k)}{\sum_{k\in\Omega} w(k)} = \frac{1}{n!}$$

The constructed markov chain $\{X_t : t \geq 0\}$ is the deck obtained by repeated shuffling. For example, in the 3-cards Deck Top-in Shuffle, we want $1 \times 3!$ vector $\boldsymbol{\pi} = (\pi(1), \pi(2), ..., \pi(6))$ be the stationary distribution of $\{X_t\}$, i.e.

$$\boldsymbol{\pi} \boldsymbol{P} = \boldsymbol{\pi}$$

Convergence of Markov Chain

Theorem (Ergodicity)

If a Markov chain $\{X_t\}$ is irreducible and aperiodic, then it has a unique stationary distribution $\pi(\cdot)$. In particular

- *$\boldsymbol{\pi}$ is the unique left eigenvector of $\boldsymbol{P}$ corresponding to eigenvalue $\lambda = 1$.*
- $\lim_{t\to\infty} P^{[t]}(x,y) = \pi(y),\ \forall x, y \in \Omega.$

Theorem (Reversibility)

If Markov chain $\{X_t\}$ is reversible wrt. $\boldsymbol{\pi}$, i.e.

$$\pi(x)P(x,y) = \pi(y)P(y,x)$$

Then $\boldsymbol{\pi}$ is the stationary distribution of $\{X_t\}$.

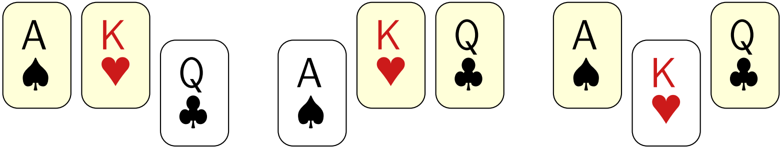
Remark

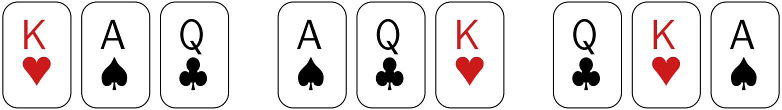
If the probability transition matrix is symmetric: $\boldsymbol{P} = \boldsymbol{P}^{\top}$, then the markov chain is reversible wrt. uniform distribution. Becasue $\pi(x) = \pi(y)$ and $P(x, y) = P(y, x)$.

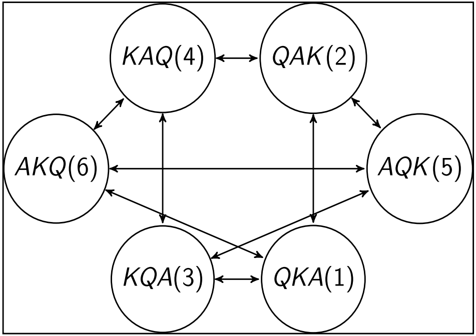
Example (3-cards Deck Random Transposition)

In each iteration, we pick uniformly at random two cards and switch them. We still consider the easy 3-cards deck [AS, KH, QC].

The first transposition has $\binom{3}{2} = 3$ possible outcomes,







$$\boldsymbol{P} = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Distance Between Distributions

We hope to construct a metric to capture how statistically "close" two distributional objects are from each other.

Definition (Total Variation Norm)

For two probability mass function π, μ, define total variation as

$$\|\mu - \pi\|_{TV} := \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \pi(x)|$$

In particular, if $\mu(\cdot) = P^{[t]}(x_0, \cdot)$ is conditional distribution of deck $\{X_t\}$, given initial deck x_0. $\pi(\cdot) = \frac{1}{n!}$, then we define

$$\Delta(t) := \left\| P^{[t]} - \pi \right\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} \left| P^{[t]}(x_0, x) - \frac{1}{n!} \right|$$

Rate of Convergence

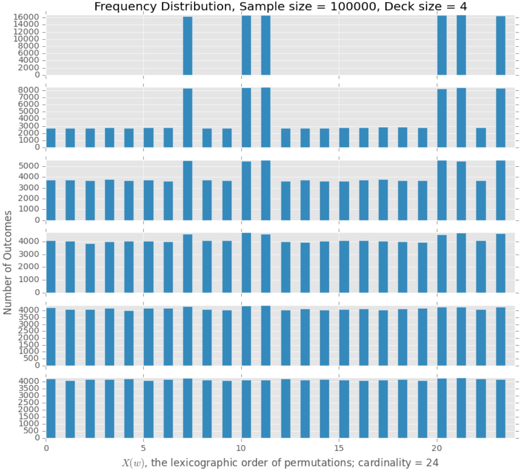
- We apply Monte Carlo simution method, draw large a sample from $P^{[t]}$ (i.e. the deck after t shuffles), then the empirical distribution of sample appoximates $P^{[t]}$.
- Then, $\Delta(t)$, as a function of t, captures the rate by which the Markov chain converges to uniform $\pi(\cdot)$.
- Moreover, we can set a **threshold** to $\Delta(t)$ and calculate time needed for the **deviation of X_t from the limit is bounded by this threshold.**

Definition (Mixing Time)

Set the threshold as constant s, the time needed is called Mixing Time of Markov chain associated with threshold s, given by

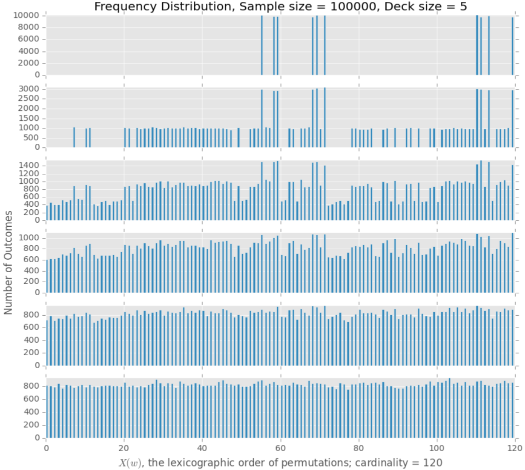
$$\tau = \min\{t : \Delta(t) \leq s\}$$

Mixing of $P^{[t]}$



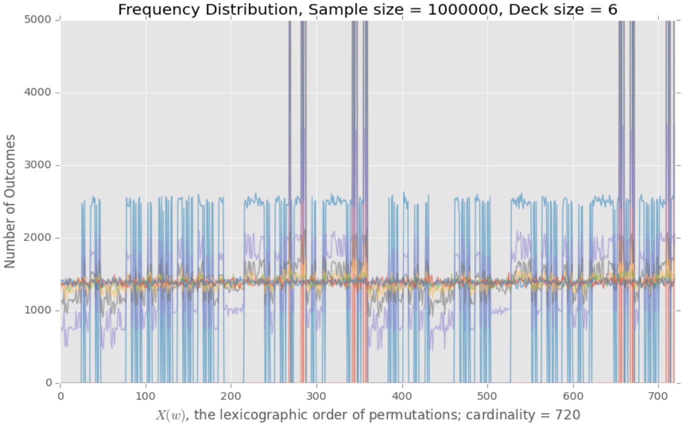
- Simulation with 4 cards.

Mixing of $P^{[t]}$

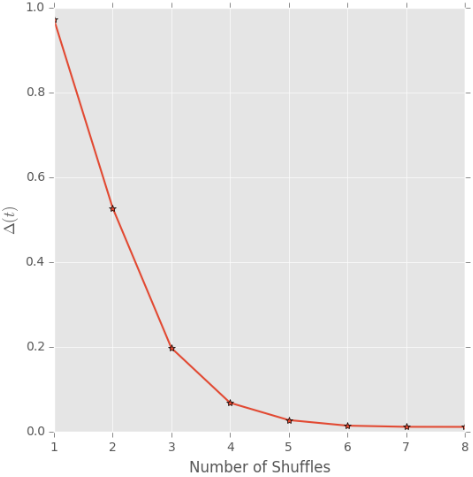


- Simulation with 5 cards.

Mixing of $P^{[t]}$: Shuffles in one Plot



Rate of Convergence



- (6 cards): $\Delta(7) \approx 0.011$

Thanks for your attention.