CARD SHUFFLING & MCMC SIMULATION

2016 STOCHASTIC PROCESSES (MA335) COURSE PROJECT

CHEN You, WANG Shu, YANG Ze, FENG Biaojian

Advised by Prof. LUO Jun

Inspired by *Course 6.896,2001 Spring, MIT*

ACEM, SJTU

Shuffling Intro

- Main purpose: Random Permutations $\Rightarrow$ Uniform Distribution.
- A stack of 52 cards $\Rightarrow$ Prob $= 1/52!$.

Shuffling Intro

- Main purpose: Random Permutations $\Rightarrow$ Uniform Distribution.
- A stack of 52 cards $\Rightarrow$ Prob $= 1/52!$.
- Prevailing Methods:
  (a) Random Transpositions.
  (b) Top-in at Random.

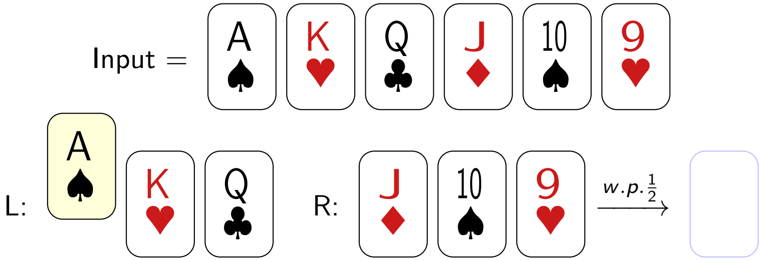
Shuffling Intro

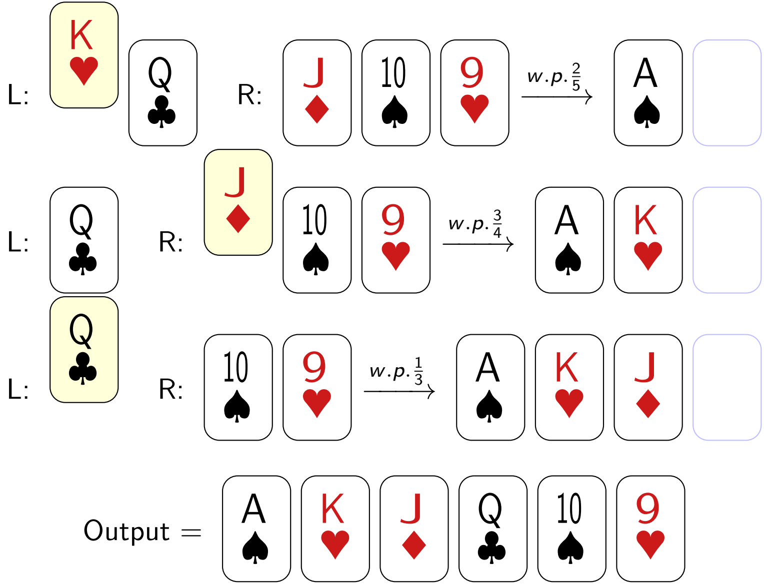
- Main purpose: Random Permutations $\Rightarrow$ Uniform Distribution.
- A stack of 52 cards $\Rightarrow$ Prob $= 1/52!$.
- Prevailing Methods:
  (a) Random Transpositions.
  (b) Top-in at Random.
  (c) Riffle Shuffle.
    - Split: $\Rightarrow Bin(n, \frac{1}{2}) \Rightarrow \mathbb{E}R = \mathbb{E}L = \frac{n}{2}$
    - Drop: $\Rightarrow \mathbb{P}\{\text{The next card is from the left}\} = \frac{L}{L+R}$, so as the right hand side situation.

The Model

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. $\frac{\#\text{ of cards in left hand}}{\#\text{ of cards remaining}}$.

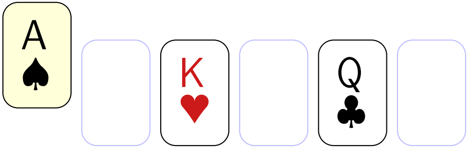



L: K Q R: J 10 9 $w.p.\frac{2}{5}$ A
L: Q R: J 10 9 $w.p.\frac{3}{4}$ A K
L: Q R: 10 9 $w.p.\frac{1}{3}$ A K J
Output = A K J Q 10 9

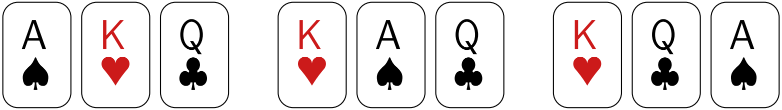
The Model

Example (3-cards Deck Top-in Shuffle)

We consider a 3-cards deck [AS, KH, QC] and shuffled by top-in method. If initially, $X_0 = (\text{AS, KH, QC})$, we have



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

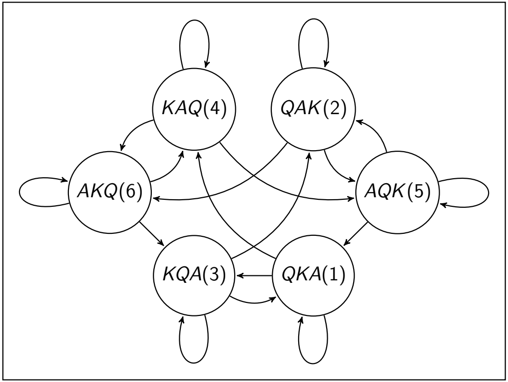


Visualization

- 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.

Visualization

- 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.



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

The Markov Chain Monte Carlo Method

- MCMC Paradigm:
  (a) Input a large finite: Ω.
  (b) *Define* a *positive* weight function: $w : \Omega \to \mathbb{R}^+$
  (c) Sample $x \in \Omega$ w.p. $\pi(x) \propto w(x)$,
  or $\pi(x) = w(x)/\sum_{x\in\Omega} w(x)$ by $\sum_{x\in\Omega} \pi(x) = 1$.

The Markov Chain Monte Carlo Method

- MCMC Paradigm:
  (a) Input a large finite: Ω.
  (b) *Define* a *positive* weight function: $w : \Omega \to \mathbb{R}^+$
  (c) Sample $x \in \Omega$ w.p. $\pi(x) \propto w(x)$,
  or $\pi(x) = w(x)/\sum_{x\in\Omega} w(x)$ by $\sum_{x\in\Omega} \pi(x) = 1$.
- MCMC Approach:
  (a) Construct a DTMC $(X_t)_t := \{X(t), t \geq 0\}$.
  (b) Let it converge to $\pi(x)$:

$$\lim_{t\to\infty} \mathbb{P}\{X(t) = y|X(0) = x\} = \pi(y)$$

which is independent of the initial state $X(0)$.

The Markov Chain Monte Carlo Method

- MCMC Paradigm:
  (a) Input a large finite: Ω.
  (b) *Define* a *positive* weight function: $w : \Omega \to \mathbb{R}^+$
  (c) Sample $x \in \Omega$ w.p. $\pi(x) \propto w(x)$,
  or $\pi(x) = w(x)/\sum_{x\in\Omega} w(x)$ by $\sum_{x\in\Omega} \pi(x) = 1$.
- MCMC Approach:
  (a) Construct a DTMC $(X_t)_t := \{X(t), t \geq 0\}$.
  (b) Let it converge to $\pi(x)$:

$$\lim_{t\to\infty} \mathbb{P}\{X(t) = y | X(0) = x\} = \pi(y)$$

  which is independent of the initial state $X(0)$.
- Model Formulation:
  (a) State Space: $\{$All possible permutations of cards$\}$.
  (b) Weight Function: $w(x) = 1$. (Sample from an uniform distribution)

The DTMC's Properties

Theorem (Ergodicity)

If a Markov Chain $\mathcal{X}$ is irreducible and aperiodic, then it has a unique stationary distribution π and

$$\lim_{t\to\infty} \mathbb{P}\{X(t) = y | X(0) = x\} = \pi(y)$$

The DTMC's Properties

Theorem (Ergodicity)

If a Markov Chain $\mathcal{X}$ is irreducible and aperiodic, then it has a unique stationary distribution π and

$$\lim_{t\to\infty} \mathbb{P}\{X(t) = y | X(0) = x\} = \pi(y)$$

Theorem (Reversibility)

If Markov chain $\mathcal{X}$ 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 $\{X_t\}$.

Convergence

- Random Transpositions.
  (a) Irreducible.
  (b) Aperiodic. (Self-loop by choosing the same card.)
  (c) Reversible with respect to the Uniform Distribution.
- Top-in at Random.
  (a) Irreducible.
  (b) Aperiodic. (By putting the card back to where it comes from.)
  (c) Rev. (By doubly stochastic from the symmetric trans. matrix.)
- *Riffle Shuffle*.
  (a) Irreductible. (By simulating "Top-in Shuffle" with positive prob.)
  (b) Aperiodic. (Self-loop by setting one of the stacks be empty.)
  (c) Rev. (Symmetric transitions.)

Convergence

- Random Transpositions.
  (a) Irreducible.
  (b) Aperiodic. (Self-loop by choosing the same card.)
  (c) Reversible with respect to the Uniform Distribution.
- Top-in at Random.
  (a) Irreducible.
  (b) Aperiodic. (By putting the card back to where it comes from.)
  (c) Rev. (By doubly stochastic from the symmetric trans. matrix.)
- *Riffle Shuffle*.
  (a) Irreductible. (By simulating "Top-in Shuffle" with positive prob.)
  (b) Aperiodic. (Self-loop by setting one of the stacks be empty.)
  (c) Rev. (Symmetric transitions.)
- *ALL* of 3 shuffling methods converge to the *uniform distribution*.

Which One Is Better?

Definition (**Closeness**: Total Variation Norm)

$$||\pi - p_x^t||_{TV} := \frac{1}{2}\sum_{y\in\Omega}|\pi(y) - p_x^t(y)|$$

Used to measure the “distance” of 2 different *Distribution Measures*.

Which One Is Better?

Definition (Closeness: Total Variation Norm)

$$||\pi - p_x^t||_{TV} := \frac{1}{2}\sum_{y\in\Omega} |\pi(y) - p_x^t(y)|$$

Used to measure the "distance" of 2 different *Distribution Measures*.

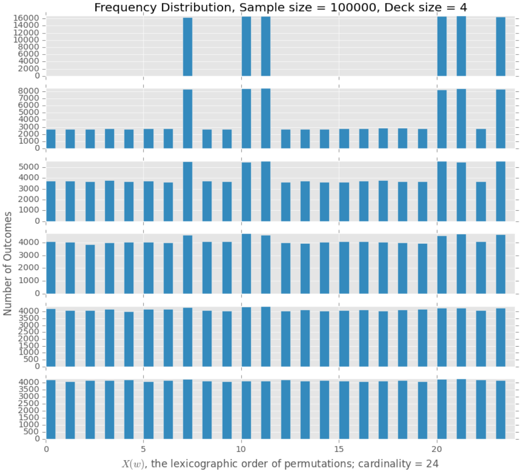
Definition (Effectiveness: Mixing Time)

Time required for the chain to come within $1/2e$ of π in *Total Variation Distance*: ($\frac{1}{2e}$ is a arbitrary choice to capture mixing.)

$$\tau_{min} := \min_t \{||\pi - p_x^t||_{TV} \leq \frac{1}{2e}, \forall x\}$$

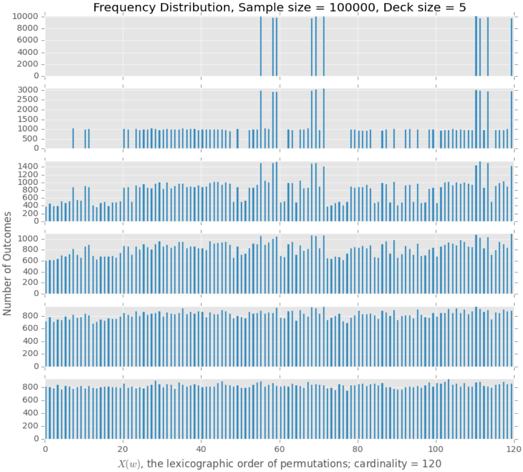
$$p_x^t = \mathbb{P}\{X_t = y | X_0 = x\}$$

Mixing of $P^{[t]}$



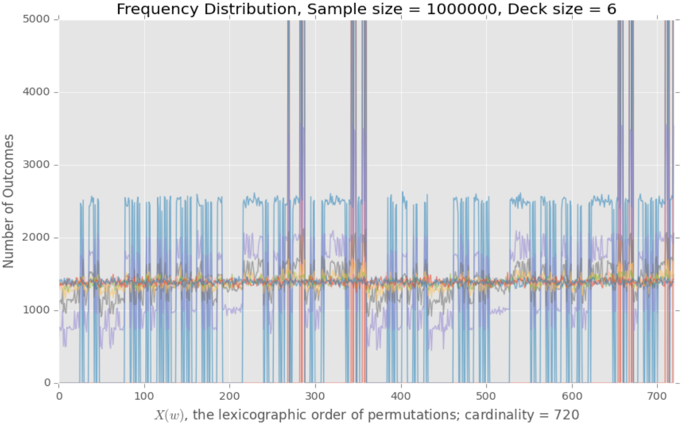
- 4 cards.

Mixing of $P^{[t]}$

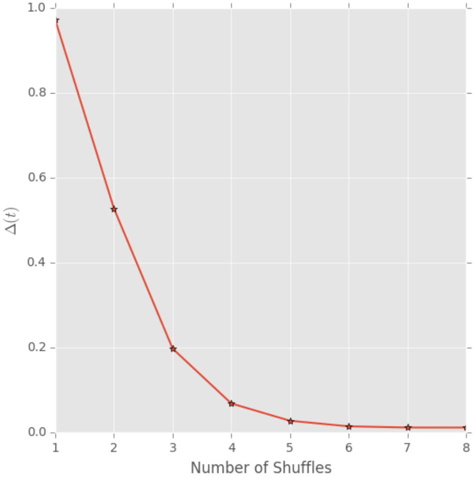


- 5 cards.

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



Rate of Convergence



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

CARD SHUFFLING & MCMC SIMULATION

2016 STOCHASTIC PROCESSES (MA335) COURSE PROJECT

CHEN You, WANG Shu, YANG Ze, FENG Biaojian

Advised by Prof. LUO Jun

Inspired by *Course 6.896,2001 Spring, MIT*

ACEM, SJTU