Card Shuffling and MCMC Paradigm

Zed

May 8, 2016

1 Introduction

This project is inspired by course 6.896, 2011 Spring in MIT.[1] This project aims to study the theory and implementation of *Markov Chain Monte Carlo* algorithm. We discuss general MCMC paradigm through the famous card-shuffling problem arisen in combinatorics. We will model the card shuffle by markov chain, and further discuss its stationary distribution, viariation distance and mixing time. Numerical computations will also be performed to validate the theory.

2 Illustration of the Problam and Related Theories

2.1 The MCMC Paradigm

Consider probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define a real, positive *weight function* $w : \Omega \to \mathbb{R}^+$. The goal is to sample an element $x \in \Omega$, with probability distribution $\pi(x) = \mathbb{P}(X = x)$ defined as

$$\pi(x) := \frac{w(x)}{Z_w}; \quad Z_w := \sum_{x \in \Omega} w(x)$$

Where Z_w is called the partition function of weight function w in Omega.
In MCMC paradigm, we establish a general iterative solution approach to this sampleability problem, where we construct a sequence of RV $\mathcal{X} = \{X_t\}$ that converges to the target probability, in the sense that

$$\lim_{t \to \infty} \mathbb{P}(X_t = y | X_0 = x) = \pi(y)$$

Example 1. We now illustrate the **Card Shuffling Problem** as an example of MCMC paradigm. Namely, we set

- The probability space $\Omega :=$ {All **Permutations** of the deck.}, $x \in \Omega$ is a permutation of the deck. In a 3-cards deck [AS, KH, QC][1], the probability space has cardinality $|\Omega^3| = 3! = 6$, namely $\Omega^3 =$ {AKQ, AQK, KAQ, KQA, QAK, QKA}.
- $w(x) = 1$ for every permutation $x \in \Omega$. Hence $\pi(x) = \frac{1}{|\Omega|}$. That is, by assigning equal probability mass to each permuation, we want to sample a uniformly distributed $x \sim U(\Omega)$.
- Denote X_0 the initial configuration of deck, which is arbitrary; and X_t is the state of deck after t^{th} shuffle. Note that the state of deck at t only depends on that of $t-1$ and how it is shuffled, we conclude that $\{X_t\}$ is a **markov chain**.

2.2 Markov Chains as Graphs

Definition 1. (*Transition Matrix*) $\forall x, y \in \Omega$, we denote one-step transition probability, t-step transition probability and corresponding transition matrix as $P(x, y), P^{[t]}(x, y), \boldsymbol{P}, \boldsymbol{P}^{[t]}$.[2] We refer to matrices like $\boldsymbol{P}$ as *stochastic matrix* if it satisfies

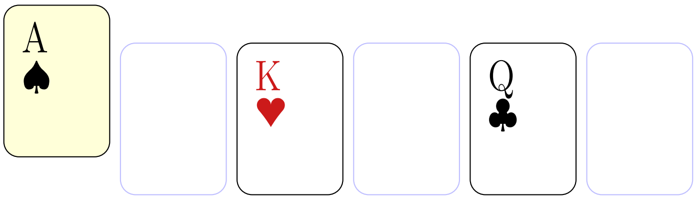
[1]We use two letters to represent a card, for example, AS means spade A.
[2]Slightly deviates from the notations in textbook [5], because the x-es are not indexed by integers.

- *Non-negativity*: $P(x, y) \geq 0$, $\forall x, y \in \Omega$. And
- *Stochasticity*: $\sum_{y \in \Omega} P(x, y) = 1$, for any fixed $x \in \Omega$.

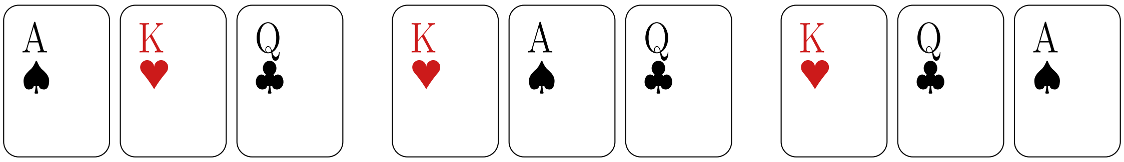
We can establish the combinatorial flavored thoery of Markov chains by viewing them as graphs. In particular, for a given markov chain $\{X_t\}$, we define a weighed, directed graph $G(\{X_t\}) = (V, E)$ as follows

- $V = \Omega$, i.e. the vertices are the states in Ω.
- $e_{x \to y} \in E$ if $P(x, y) > 0$, with weight $w(e_{x \to y}) = P(x, y)$. That is, $\boldsymbol{P}$ it the *adjacency matrix* of graph G.

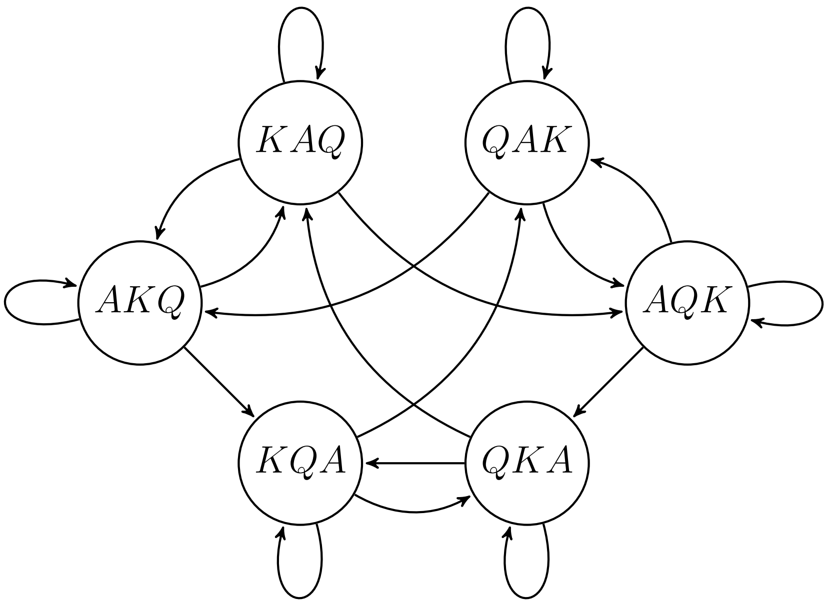
Example 2. (*3-cards Deck Top-in Shuffle*) We consider a 3-cards deck [AS, KH, QC] and shuffled by top-in method. That is, we take one card from the top and insert it at a random position in the deck. If initially, $X_0 = (\text{AS, KH, QC})$, we have



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



And $G(\{X_t\}) = (V, E)$:



The graph above represents the markov chain associated with top-in shuffle. All edges have weight $\frac{1}{3}$, and its adjacency matrix gives the one-step transition matrix of this markov process.

Definition 2. (*Undirected Graph*) Graph of markov chain $G(\mathcal{X})$ is called undirected if $P(x, y) > 0 \iff P(y, x) > 0$. The exact value of weight doesn't matter, we only care about positivity.

2.3 Basic Properties

Definition 3. (*Irreducible*) $\mathcal{X}$ is called irreducible if (TFAE)

- $\forall x, y \in \Omega$, $\exists t \in \mathbb{N}$ such that $P^{[t]}(x,y) > 0$.
- The graph $G(\mathcal{X})$ is strongly connected.

Definition 4. (*Aperiodic*, *Period*) $\mathcal{X}$ is called aperiodic if $\forall x, y \in \Omega$, we have

$$\gcd\{t : P^{[t]}(x,y) > 0\} = 1$$

For any vertex $x \in \Omega$, define the period of x as $\gcd\{t : P^{[t]}(x,x) > 0\}$.

Lemma 1. If $\mathcal{X}$ is irreducible, the period of every $x \in \Omega$ is the same. Further, if $G(\mathcal{X})$ is undirected, the period is either 1 or 2.

Lemma 2. If $\mathcal{X}$ is irreducible and $G(\mathcal{X})$ is undirected, then $\mathcal{X}$ is aperiodic $\iff$ $G(\mathbf{X})$ is *non-bipartite*. I.e. there is **Not** a patition $V = V \cup V'$ ($V \cap V' = \emptyset$), such that every edge connects a vertex in V to one in V'.

Proof. ($\Rightarrow$) it suffices to show its converse-negative, i.e. bipartite $\Rightarrow$ periodic. Suppose $G(\mathcal{X})$ is bipartite, pick $x, y \in V$ belongs to same side, then every path from x to y requires even # of steps. So $\gcd\{t : P^{[t]}(x,y) > 0\} \geq 2 > 1$.
($\Leftarrow$) Suppose $G(\mathcal{X})$ is bipartite $\iff$ $\exists C \subset G(\mathcal{X})$ is an odd cycle. Pick any $x, y \in \Omega$ and $c \in C$, we can always go through $x \to c \to y$ because the chain is irreducible. Moreover, denote $l(x,y)$ be the path from x to y, we can go either

- $l(x,c) \to C \to l(c,y)$. $t = |l(x,c)| + |C| + |l(c,y)|$
- $l(x,c) \to$ go on C until the $\frac{|C|-1}{2}$ vertex on C, then go back to $c \to l(c,y)$. $t' = |l(x,c)| + |C| - 1 + |l(c,y)|$

We see $\gcd(t,t')$ is already 1, because they are consecutive integers. So it is aperiodic. □

Lemma 3. If $\mathcal{X}$ is irreducible, then $\mathcal{X}$ is aperiodic $\iff$ there exists some t such that $P^{[t]}(x,y) > 0$ for all $x, y \in \Omega$.

Lemma 4. If $\mathcal{X}$ is irreducible and has at least one self-loop, i.e. $P(x,x) > 0$ for some x; then $\mathcal{X}$ is aperiodic.

Proof. Suppose this loop is on z. From x to y we can either

- $l(x,z) \to l(z,y)$. $t = |l(x,z)| + |l(z,y)|$
- $l(x,z) \to loop(z) \to l(z,y)$. $t' = |l(x,z)| + 1 + |l(z,y)|$

Same argument as lemma 2. □

2.4 Mixing of Markov Chains

Definition 5. (*Stationary Distribution*) Let $\mathcal{X} = \{X_t\}$ be a Markov chain. A probability distribution $\pi(\cdot) : \Omega \to \mathbb{R}$ over Ω is a stationary distribution of $\mathcal{X}$ if

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

Where $\boldsymbol{\pi}$ is $1 \times |\Omega|$ vector that stacks the values of $\pi(\cdot)$ together, i.e. $\boldsymbol{\pi} = (\pi(x))_{x \in \Omega}$.

Theorem 1. (*Ergodic Thm.*) If a Markov chain $\mathcal{X}$ 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 $\lambda = 1$.

- $\lim_{t \to \infty} P^{[t]}(x, y) = \pi(y)$, $\forall x, y \in \Omega$.

Proof. *TODO* $\square$

Definition 6. (*Reversibility*) Let $\mathcal{X} = \{X_t\}$ be a Markov chain. And $\pi(\cdot) : \Omega \to \mathbb{R}$ is a probability distribution over Ω. $\mathcal{X}$ is called reversible with respect to π if, for any $x, y \in \Omega$

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

Remark.(i) Intuitively, a markov chain is reversible w.r.t π means that the probability mass that flows from note $x \to y$ is same as that from $y \to x$.
Remark.(ii) If $\boldsymbol{P}$ is symmetric, i.e. $P(x, y) = P(y, x)$, then $\mathcal{X}$ is trivially reversible with respect to uniform distribution on Ω, $\pi(\cdot) \sim U(\Omega)$.

Lemma 5. If $\mathcal{X}$ is reversible wrt. π, then π is a stationary distribution of $\mathcal{X}$.

Proof. For any fixed $y \in \Omega$:

$$\sum_{x \in \Omega} \pi(x)P(x, y) = \sum_{x \in \Omega} \pi(y)P(y, x) = \pi(y)$$

Implies $\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$. $\square$

Definition 7. (*Ergodic Flow*) The ergodic flow between nodes x and y wrt. a probability distribution π is defined to be the amount of probability mass flowing between x, y:

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

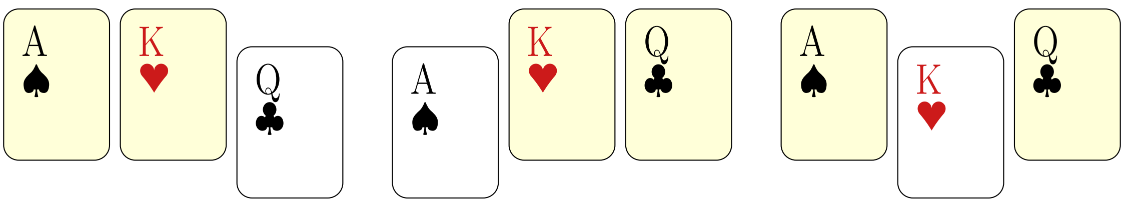
Enables us to write transition probability of reversible markov chain

3 Types of Shuffles

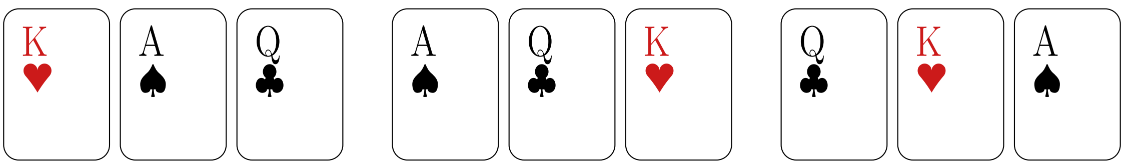
3.1 Random Transposition

In the method of random transposition, in each iteration, we pick uniformly at random two cards and switch them.

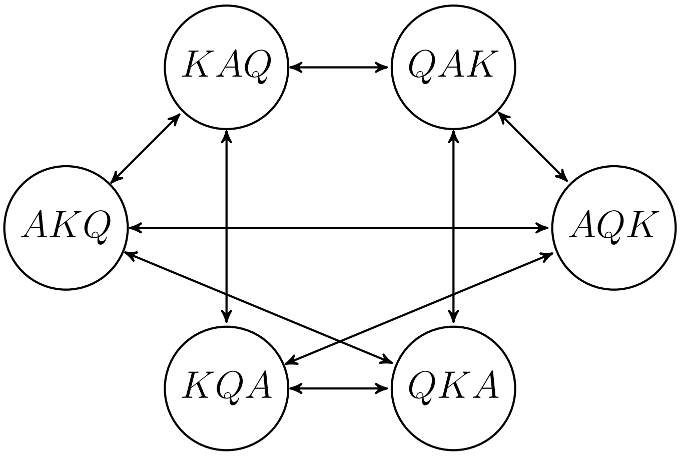
Example 3. (*3-cards Deck Random Transposition*) We still consider our 3-cards deck [AS, KH, QC] and shuffled Random Transposition. Initially $X_0 =$ (AS, KH, QC), the first transposition has $\binom{3}{2} = 3$ possible cases,



And after the first transposition, the deck becomes



$G(\mathcal{X})$ is undirect, given by



3.2 Top-in Shuffle

3.3 Overhand Shuffle

3.4 Riffle Shuffle

References

[1] Constantinos Daskalakis: 6.896: Probability and Computation, MIT http://people.csail.mit.edu/costis/6896sp11/

[2] Harald Hammarström. *Card-Shuffling Analysis with Markov Chains*. 2005.

[3] Bayer, Dave and Diaconis, Persi, *Trailing the Dovetail Shuffle to its Lair*. Annals of Applied Probability, 2(2), 294-313, 1992.

[4] Brad Mann. *How many times should you shuffle a deck of cards*. Topics in Contemporary Probability and Its Applications, Harvard University, 1995.

[5] Sheldon M. Ross. *Introduction to Probability Models, Eleventh Edition*. ISBN: 978-0-12-407948-9, 2015