Exploration to Markov Chain: Motivated by an Exercise

Guo Linsong 518030910419

June 28, 2020

Contents

1 Why I'm Interested In Markov Chain 2

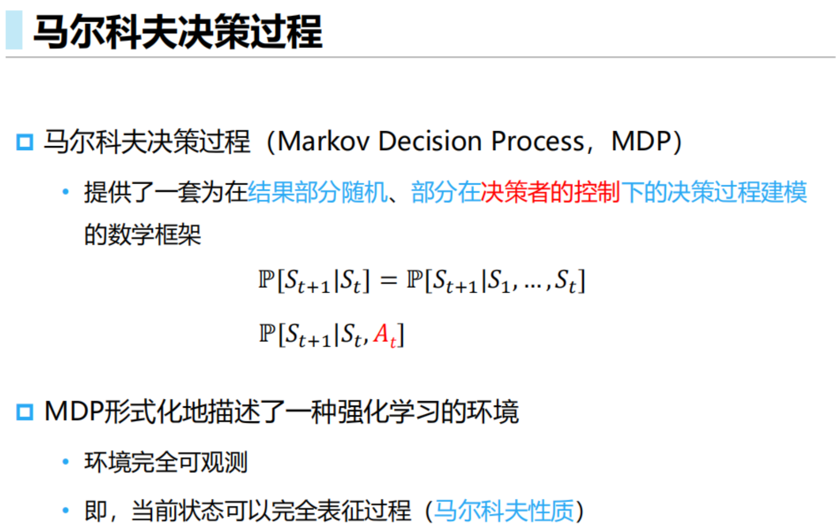
2 Basic Theory 3
2.1 Discrete-Time Markov Chain 3
2.2 Transition Matrix of Markov Chain 4
2.3 Initial Distribution of Markov Chain 4
2.4 Topological Structure 5

3 A Personal Solution to Question 1 8

4 Another Interesting Discussion: Stability 9

1 Why I'm Interested In Markov Chain

I first saw the Markov Chain when I was working on a machine learning course, as shown in the following figure. *Markov Decision Process* is an important model in machine learning. But at that time I was just beginning to study probability theory, so I can't understand Markov Decision Process well.



Later I saw the chapter 10.13 *Non-negative superharmonic functions for Markov chains* in the textbook. The chapter briefly introduces the Markov Chain and gives a question:

> **Question 1.** *Show that a Markov Chain is irreducible and recurrent if and only if every nonnegative superharmonic function is constant.*

The textbook has proved 'only if' but left 'if' as an exercise. What's *irreducible*, *recurrent* and *superharmonic*? These concepts didn't appear in the previous chapters, so I was very confused about this question. Although the textbook does not cover Markov Chain in detail, Markov Chain appears a lot in the textbook, which arouses my curiosity. I looked up some articles and tried to solve the above question. **In the article, I will summarize the basic theory of Markov Chain, give a personal solution of the above question and discuss another interesting problem of Markov Chains.**

2 Basic Theory

The section will introduce some basic definitions and notations which are necessary to solve question 1. So the section may be a little boring, you can skip into the next section if you are familiar with the basic theory.

2.1 Discrete-Time Markov Chain

In probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $X = (X_n : n \geq 0)$ is a stochastic process with time space $\mathbb{N}$ and state space S. This means that X_n takes values in the state space S. For $n \geq 0$, let $\mathcal{F}_n = \{X_0, X_1, \cdots, X_n\}$ be the σ-algebra generated by the process up to time n and $\mathcal{T}_n = \{X_n, X_{n+1}, \cdots\}$ be the σ-algebra generated by the process from time n on.

Definition 2. *The process X is* Markov Chain *if the future is independent of the past, given the present. In a formal way,*

$$\mathbb{P}(F \cap T | X_n) = \mathbb{P}(F|X_n)\mathbb{P}(T|X_n) \quad \forall n \in \mathbb{N}, F \in \mathcal{F}_n, T \in \mathcal{T}_n$$

The above formulation reveals the essence of Markov Chain. However, in most materials I have seen, the following formulations are more common:

$$\mathbb{P}(X_{n+1} = i \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} = i \mid X_n) \quad \forall n \in \mathbb{N}, i \in S$$

or

$$\mathbb{P}(X_{n+1} = i \mid X_0 = i_0, \cdots, X_n = i_n) = \mathbb{P}(X_{n+1} = i | X_n = i_n) \quad \forall n \in \mathbb{N}, i, i_0, \cdots, i_n \in S$$

Time Homogeneous

Definition 3. *A Markov Chain X is* time homogeneous *if*

$$\mathbb{P}(X_{n+m} = j | X_n = i) = \mathbb{P}(X_m = j | X_0 = i) \quad \forall n, m \in \mathbb{N}, i, j \in S$$

This means that if the process is in the state i at the time n, then it doesn't matter how the process got to the current state and it is as if the chain starts over again. This concept is useful in the following section.

2.2 Transition Matrix of Markov Chain

We use $p_{i,j}$ to denote the probability from state i to j in one time. Thus p is a probability matrix of size $|S| \times |S|$. Similarly, $p^n_{i,j}$ is the probability from state i to state j in n times.

Lemma 4. *If a Markov Chain is time homogeneous, then* $p^{n+m} = p^n p^m$.

Proof. For $i, k \in S$, we have

$$p^{n+m}_{i,j} = \mathbb{P}(X_{m+n} = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_n = k | X_0 = i) \mathbb{P}(X_{m+n} = j | X_n = k)$$

By time homogeneous property

$$\mathbb{P}(X_{m+n} = j | X_n = k) = \mathbb{P}(X_m = j | X_0 = k) = p^m_{k,j}$$

Hence

$$p^{n+m}_{i,j} = \sum_{k \in S} p^n_{i,k} p^m_{k,j}$$

$\square$

This lemma says that the probability matrix from time n to time m is determined only by the difference between n and m, hence is p^{m-n}. We assume that the following Markov chains all conform to time homogeneous property.

2.3 Initial Distribution of Markov Chain

The *initial distribution* π of a Markov Chain is the distribution of X_0 on the state space S.

Unit mass

Unit mass is the distribution that assigns value 1 to state i and 0 to all other states. We call δ_i the unit mass at i:

$$\delta_i(j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

If the initial distribution of Markov Chain is δ_i, then this corresponds to picking state i as initial state with probability 1. We use P^i to denote the law of chain when the initial distribution is δ_i:

$$P^i(A) = \mathbb{P}(A \mid X_0 = i)$$

Similarly, the expectation with the initial distribution δ_i is

$$E^i(Y) = \sum_{y \in S} y P^i(Y = y)$$

Stationary Distribution

The 'stationary' means the distribution does not change with time: $\mathbb{P}(X_n = i) = \pi(i)$ for all $n \in \mathbb{N}$. An initial distribution π is *stationary* if $\pi = \pi p$. This is to say

$$\pi(i) = \sum_{j \in S} \pi(j) p_{j,i} \quad \forall i \in S$$

It follows that $\pi = \pi p^n$ for every $n \in \mathbb{N}$.

2.4 Topological Structure

The Relation *leads to*

We say that state i leads to state j if, starting from i, the chain will visit j at finite time. In a formal way,

$$i \to j \iff p^n_{i,j} > 0 \text{ for some } n.$$

The Relation *communicates with*

We say that state i communicates with state j (written as $i \leftrightarrow j$) if $i \to j$ and $j \to i$. Obviously, the relation $\leftrightarrow$ conforms to the symmetric, reflexive and transitive property, hence is an equivalence relation. The equivalence class of state i is the set of all states that communicate with i:

$$[i] = \{j \in S : i \leftrightarrow j\}$$

Irreducible Property

We call a Markov Chain is *irreducible* if

$$i \leftrightarrow j \quad \forall i, j \in S$$

Recurrent Property

We define the *hitting time* to state j is

$$T_j = \inf\{n \geq 1 : X_n = j\}$$

Note that T_j is a stopping time because we can tell if $T_j = n$ by observing the chain up to time n:

$$\{T_j = n\} = \bigcap_{i=1}^{n-1} \{X_i \neq j\} \cap \{X_n = j\} \in \mathcal{F}_n$$

We define $f_{i,j}$ is the probability from state i to j in a finite number of times:

$$f_{i,j} = P^i(T_j < \infty)$$

A state i is called *recurrent* if $f_{i,i} = 1$ and *transient* otherwise. By the Markov property, once the chain revisits state i, the future is independent of the past, and it is as if the chain starts over again from state i. Thus the state i will be revisited with the same probability $f_{i,i}$. Note that $f_{i,i} = 1$, the chain will visit state i over and over again, an infinite number of times. Maybe this is why the property is called recurrent. On the contrary, if state i is transient, then it will only be visited a finite number of times.

Lemma 5. *If a Markov Chain is irreducible recurrent, then*

$$f_{i,j} = P^i(T_j < \infty) = 1 \quad \forall i, j \in S$$

Positive vs Null Recurrent

While a recurrent state is expected to be revisited an infinite number of times, the state is not necessarily expected to be revisited even once within a finite number of steps. So there is a further classification of recurrent property.

- A state i is known as *positive recurrent* if $\mathbb{E}\left(T_i \mid X_0 = i\right) < \infty$.
- A state i is known as *null recurrent* if $\mathbb{E}\left(T_i \mid X_0 = i\right) = \infty$.

Aperiodic Property

Consider the Markov chain in figure 1 with $S = \{1, 2, 3, 4\}$, all states form a single equivalence class, so the chain is irreducible. Let $A = \{1, 3\}$ and $B = \{2, 4\}$. If $X_0 \in A$, then $X_1 \in B$, $X_2 \in A$, $X_3 \in A$, $\cdots$. Thus it's meaningful to introduce *period*.

Definition 6. *Let $D_i = \{n \in \mathbb{N} : p_{i,i}^n > 0\}$ be the set of all $n \in \mathbb{N}$ such that it is possible to revisit i at time n. The* period *of a state is defined as the greatest common divisor of all such n: $d_i := \gcd\{n \in \mathbb{N} : n \in D_i\}$. A state is aperiodic if its period is* 1.

For example, in figure 1, the period of all states is 2. Hence the chain is not aperiodic.

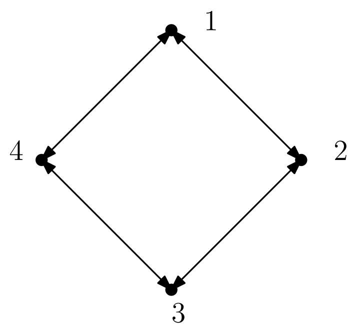


Figure 1: A Markov Chain with period 2

Lemma 7. *If $i \leftrightarrow j$, then $d_i = d_j$.*

Proof. If $i \leftrightarrow j$, then there exists $n_1 \in \mathbb{N}$ such that $p_{i,j}^{n_1} > 0$ and $n_2 \in \mathbb{N}$ such that $p_{j,i}^{n_2} > 0$. So $p_{i,i}^{n_1+n_2} \geq p_{i,j}^{n_1} \cdot p_{j,i}^{n_2}$, which implies that

$$d_i \mid n_1 + n_2$$

For all $n \in D_j$, we have $p_{j,j}^n > 0$. So $p_{i,i}^{n_1+n+n_2} \geq p_{i,j}^{n_1} \cdot p_{j,j}^n \cdot p_{j,i}^{n_2}$, which implies that

$$d_i \mid n_1 + n + n_2 \quad \forall n \in D_j$$

Combining the two equations, we have

$$d_i \mid n \quad \forall n \in D_j$$

This implies $d_i | d_j$. In a similar way, we can prove that $d_j | d_i$. Therefore, $d_i = d_j$. $\square$

Lemma 8. *If i is an aperiodic state, then there exists $N \in \mathbb{N}$ such that $p_{i,i}^n > 0$ for all $n \geq N$.*

Proof. If i is an aperiodic state, then $d_i = 1$. So there exists $n_1, n_2 \in D_i$ such that $n_2 - n_1 = 1$. If n is sufficiently large, we can divide n by n_1 to obtain $n = qn_1 + r$ with $q > r$. Hence

$$n = qn_1 + r \cdot 1 = qn_1 + r(n_2 - n_1) = (q - r)n_1 + rn_2$$

Note that $d_i | n_1, d_i | n_2$, we have $d_i | n$, hence $n \in D_i$. $\square$

3 A Personal Solution to Question 1

Definition 9. *Let h be a function defined on state space S. h is superharmonic if $h \geq ph$. This is to say*

$$h(i) \geq \sum_{j \in S} p_{i,j} h(j)$$

Lemma 10. *If the nonnegative function h superharmonic, then $(h(X_n) : n \geq 0)$ is supermartingale.*

Proof.

$$\mathbb{E}[h(X_{n+1}) \mid \mathcal{F}_n] = \sum_{j \in S} p_{X_n,j} h(j) = (ph)(X_n) \leq h(X_n)$$

Thus $(h(X_n) : n \geq 0)$ is a supermartingale. □

To prove ⇔ in question 1, we firstly prove ⇒.

Lemma 11. *If a Markov chain X is irreducible recurrent, then every nonnegative superharmonic function h is constant.*

Proof. This lemma has been proved in the textbook. If X is irreducible recurrent, then by lemma 5, we have

$$f_{i,j} = P^i(T_j < \infty) = 1 \quad \forall i, j \in S$$

where

$$T_j = \inf\{n \geq 1 : X_n = j\}$$

If h is nonnegative and superharmonic, then by lemma 10, $h(X_n)$ is a supermartingale. Thus for all $i, j \in S$, we have

$$h(j) \leq E^i h(X_{T_j}) \leq E^i h(Z_0) = h(i)$$

By reversing i and j, we get $h(i) \leq h(j)$. Therefore, $h(i) = h(j)$ for all $i, j \in S$ and h is constant. □

Next we prove ⇐ in question 1.

Lemma 12. *If every nonnegative superharmonic function h is constant, then the Markov chain X is irreducible recurrent.*

Proof. The textbook doesn't give the proof and makes it as an exercise. We introduce a new hitting time

$$\hat{T}_j = \inf\{n \geq 0 : X_n = j\}$$

Note that the difference of T and $\hat{T}$ is only $n \geq 1$ and $n \geq 0$. Similarly, we define $\hat{f}$:

$$\hat{f}_{i,j} = P^i(\hat{T}_j < \infty) = 1 \quad \forall i, j \in S$$

Next we prove that $\hat{f}_{ij} \geq \sum_{k \in S} p_{ik} \hat{f}_{kj}$ for all $i, j \in S$:

- $i = j$:

$$\hat{f}_{i,j} = \hat{f}_{j,j} = P_j(\hat{T}_j = 0) = 1 \geq \sum_k p_{ik} \hat{f}_{kj}$$

- $i \neq j$:

$$\begin{aligned}
\hat{f}_{ij} &= P^i(\hat{T}_j < \infty) \\
&= \sum_{k \in S} P^i(\hat{T}_j < \infty \mid X_1 = k) P^i(X_1 = k) \\
&= \sum_{k \in S, k \neq j} P^i(\hat{T}_j < \infty \mid X_1 = k) P^i(X_1 = k) + P^i(X_1 = j) \\
&= \sum_{k \in S, k \neq j} \hat{f}_{k,j} p_{i,k} + p_{i,j} \\
&\geq \sum_{k \in S} p_{ik} \hat{f}_{kj}
\end{aligned}$$

This implies that $i \to f_{i,j}$ is superharmonic when fixing j, and hence constant by assumption. Note that $f_{j,j} = 1$, by constant, we have $f_{i,j} = f_{j,j} = 1$ for all $i \in S$. Thus we have

$$i \to j \quad \forall i, j \in S$$

This implies that the Markov Chain is irreducible. As for recurrent property, Note that

$$f_{i,i} \geq \hat{f}_{i,j} \cdot \hat{f}_{j,i} = 1 \cdot 1 = 1 \quad \forall i \in S$$

Therefore, we have prove that X conforms to irreducible and recurrent property. $\square$

Combining lemma 11 and 12, we have finished the proof of the question 1.

4 Another Interesting Discussion: Stability

Stability refers to the convergence of the probability distribution $P(X_n = x)$ as $n \to \infty$. For example, consider a two-state transition matrix

$$p = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Note that for any initial distribution μ, the distribution will convergence to

$$\lim_{n\to\infty} \mu p^n = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} \end{pmatrix}$$

Another interesting fact is that $\left(\frac{2}{5}, \frac{3}{5}\right)$ is a *stationary distribution* of X. In the following, we will show that X has a unique stationary distribution and the chain X with any initial distribution will convergence to the unique stationary distribution. Clearly, not all Markov Chains have so nice properties. But for Markov Chains with *stationary distribution*, stability is obvious.

Lemma 13. *Every irreducible and positive recurrent Markov chain has a unique stationary distribution.*

The lemma is not obvious, and it's not easy to prove it rigorously. But constructing a stationary distribution is easy, because $\pi = \pi p$ is just a linear question. Next, we focus on the interesting stability.

Theorem 14. *Suppose that the Markov Chain X is irreducible, positive recurrent and aperiodic with transition matrix $p_{i,j}$. By theorem 13, X has a unique stationary distribution π. Then for any initial distribution μ, we have*

$$\lim_{n\to\infty} \mathbb{P}(X_n = i) = \pi(i) \quad \forall i \in S$$

Lemma 15. *Suppose that X and Y are two Markov Chains. T is called meeting time of two chains if*

$$X_n = Y_n \quad \forall n \geq T$$

Then for all $n \in \mathbb{N}$ and $i \in S$, we have

$$|\mathbb{P}(X_n = i) - \mathbb{P}(Y_n = i)| \leq \mathbb{P}(T > n)$$

Proof. For all $n \in \mathbb{N}$ and $i \in S$, we have

$$\begin{aligned}
\mathbb{P}(X_n = i) = & \quad \mathbb{P}(X_n = i, n < T) + \mathbb{P}(X_n = i, n \geq T) \\
= & \quad \mathbb{P}(X_n = i, n < T) + \mathbb{P}(Y_n = i, n \geq T) \\
= & \quad \mathbb{P}(X_n = i | n < T)\mathbb{P}(n < T) + \mathbb{P}(n \geq T | Y_n = i)\mathbb{P}(Y_n = i) \\
\leq & \quad \mathbb{P}(n < T) + \mathbb{P}(Y_n = i)
\end{aligned}$$

Thus we have

$$\mathbb{P}(X_n = i) - \mathbb{P}(Y_n = i) \leq \mathbb{P}(T > n) \quad \forall n \in \mathbb{N}, i \in S$$

Reversing X_n and Y_n, we have

$$\mathbb{P}(Y_n = i) - \mathbb{P}(X_n = i) \leq \mathbb{P}(T > n) \ \ \forall n \in \mathbb{N}, i \in S$$

Combining the two above inequalities, we have

$$|\mathbb{P}(Y_n = i) - \mathbb{P}(X_n = i)| \leq \mathbb{P}(T > n) \ \ \forall n \in \mathbb{N}, i \in S$$

□

We will prove theorem 14 using the *coupling* technique. Firstly we construct two independent Markov Chains. The first chain, denoted by $X = (X_n : n \geq 0)$, with transition matrix $p_{i,j}$ and any initial distribution μ. Another chain, denoted by $Y = (Y_n : n \geq 0)$, with transition matrix $p_{i,j}$ and the unique stationary distribution π. Because the initial distributions of the two processes are both $p_{i,j}$, by assumption of the theorem, X and Y are both irreducible, positive recurrent and aperiodic. We couple them into a Markov Chain $W = (W_n : n \geq 0)$ with state space $S \times S$. Its initial distribution is

$$w(i,j) = \mu(i)\pi(j) \ \ \forall i, j \in S$$

As for transition matrix, the transition probability from (i_1, j_1) to (i_2, j_2) is

$$\begin{aligned} p_{(i_1,j_1),(i_2,j_2)} &= \mathbb{P}(W_{n+1} = (i_2, j_2) \mid W_n = (i_1, j_1)) \\ &= \mathbb{P}(X_{n+1} = i_2 | X_n = i_1)\mathbb{P}(Y_{n+1} = j_2 \mid Y_n = j_1) \\ &= p_{i_1,i_2} p_{j_1,j_2} \end{aligned}$$

X and Y are both aperiodic, by lemma 8, we have

$$p^n_{(i,j),(i,j)} = p^n_{i,i} p^n_{j,j} > 0 \ \text{ for sufficiently large } n, \forall i, j \in S$$

X and Y are both irreducible and positive recurrent, by lemma 5, we have

$$f_{(i_1,j_1),(i_2,j_2)} = f_{i_1,i_2} f_{j_1,j_2} = 1 \ \ \forall i_1, j_1, i_2, j_2 \in S$$

Combining the two above properties, we have

$$\mathbb{P}(\hat{T} < \infty) = 1$$

where $\hat{T} = \inf\{n \in \mathbb{N} : X_n = Y_n\}$ be the *first meeting time*.

In the figure 2, we define a new process $Z = (Z_n : n \geq 0)$ where $Z_n = X_n \mathbf{1}_{n<\hat{T}} + Y_n \mathbf{1}_{n \geq \hat{T}}$. Note

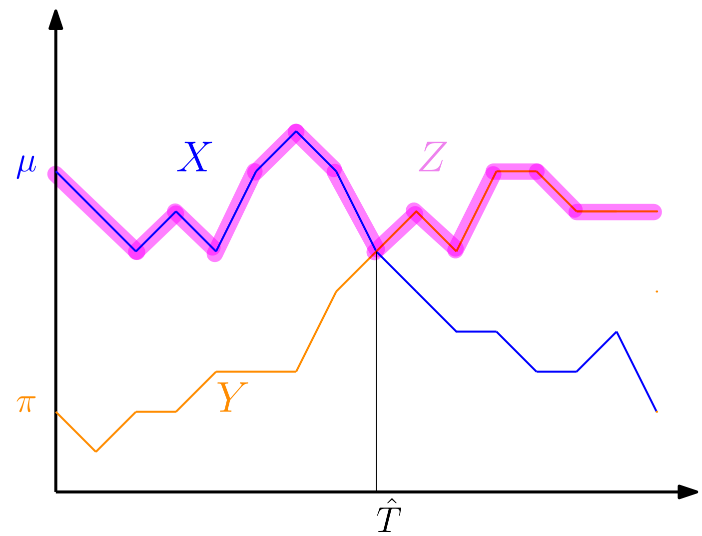


Figure 2: Three Markov Chains: X, Y and Z

that Z_n and X_n have the same initial distribution μ and transition matrix $p_{i,j}$. Thus $(Z_n : n \geq 0)$ is identical in law to $(X_n : n \geq 0)$. This means that

$$\mathbb{P}(X_n = i) = P(Z_n = i) \ \ \forall n \in \mathbb{N}, i \in S$$

Note that $Z_n = Y_n$ for $n \geq \hat{T}$, $\hat{T}$ is also the *meeting time* of Y and Z. By lemma 15, we have

$$|\mathbb{P}(Y_n = i) - \mathbb{P}(Z_n = i)| \leq \mathbb{P}(\hat{T} > n) \ \ \forall n \in \mathbb{N}, i \in S$$

Since $\mathbb{P}(Z_n = i) = \mathbb{P}(X_n = i)$ and $\mathbb{P}(Y_n = i) = \pi(i)$, we have

$$|\pi(i) - \mathbb{P}(X_n = i)| \leq \mathbb{P}(\hat{T} > n) \ \ \forall n \in \mathbb{N}, i \in S$$

Note that $\mathbb{P}(\hat{T} < \infty) = 1$, we have

$$\lim_{n \to \infty} |\pi(i) - \mathbb{P}(X_n = i)| \leq \lim_{n \to \infty} \mathbb{P}(\hat{T} > n) = \mathbb{P}(\hat{T} = \infty) = 0 \ \ \forall n \in \mathbb{N}, i \in S$$

Therefore, when $n \to \infty$, $\mathbb{P}(X_n = i) \to \pi(i)$ for all $i \in S$. We have finished the proof of theorem 14.

Reference

- https://www.randomservices.org/random/markov/index.html
- http://www.math.hkbu.edu.hk/~zeng/Teaching/mffm7010/StochProc.pdf
- https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/Chapter11.pdf
- http://staff.utia.cas.cz/swart/chain10.pdf
- https://brilliant.org/wiki/transience-and-recurrence/
- http://www.columbia.edu/~ww2040/4701Sum07/4701-06-Notes-MCII.pdf
- http://www.bioinfo.org.cn/~wangchao/maa/mcrw.pdf
- http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf
- https://resources.mpi-inf.mpg.de/departments/d1/teaching/ws11/SGT/Lecture5.pdf
- https://www.cc.gatech.edu/~vigoda/MCMC_Course/MC-basics.pdf