1 Overview

In this lecture, we will move to our forth topic on Markov Chains after talking about basic properties of random variables, concentration inequalities, and using randomized algorithms to compress data. We will firstly introduce some examples of Random Walk problem. Then we will introduce a method Markov Chain, which is an approach to solve random walk problems. We will discuss about the properties and representations of Markov Chains. In the end, we will discuss about stationary distributions, which is an important concept to Markov Chain, and fundamental theory of Markov Chains.

2 Motivation

Markov Chain is fundamental to some sampling methods, which themselves require randomness. Instead of using a deterministic method to find one solution to optimization problem like Minimum Spanning Tree or Shortest Paths, we can uniformly randomly sample from many solutions. The benefits to do so are firstly we can then answer questions like whether a certain edge is important in finding a solution, and secondly sampling provides extra guarantees (e.g. confidence interval) during the time of uncertainty.

3 Random Walks Examples

One type of problems that Markov Chains are used to solve is random walk problem. The followings are some examples of random walk problems.

3.1 1-D Random Walk

Definition 1. The problem is to maintain a value $X$ and every time step, we have

$$X_{t+1} = \begin{cases} 
X_t + 1 & \text{w.p. } \frac{1}{2} \\
X_t - 1 & \text{w.p. } \frac{1}{2}
\end{cases} \quad (1)$$

Figure 1 (a) demonstrates possible steps. To this problem, we can ask questions like

- How many steps are needed to reach 10 starting from 0? The short answer is that the expected number of steps is infinity.

- How possible it is to reach 10 starting from 0? The short answer is with probability 1.

Another example is circular 1-D random walk as shown in Figure 1 (b), which is defined as
Definition 2. The circular 1-D random walk problem is to maintain a value $X$ and every time step, we have
\[
X_{t+1} = \begin{cases} 
X_t + 1 \mod n & \text{w.p. } \frac{1}{2} \\
X_t - 1 \mod n & \text{w.p. } \frac{1}{2} 
\end{cases},
\] (2)
where $n$ is the number of points in the circle.

We can ask questions like

- How many steps are needed to hit one specific number or visit all of them?

Remark 1. For both 1-D and 2-D random walk problems, we only discuss discrete cases here (i.e. discrete both in time and space), but it can be generalized to continuous cases.

Remark 2. $X$ is a Martingals sequence.

3.2 2-D Random Walk

Definition 3. The problem is to maintain a pair of values $(X, Y)$ and every time step, we have
\[
(X_{t+1}, Y_{t+1}) = \begin{cases} 
(X_t + 1, Y_t) & \text{w.p. } \frac{1}{4} \\
(X_t - 1, Y_t) & \text{w.p. } \frac{1}{4} \\
(X_t, Y_t + 1) & \text{w.p. } \frac{1}{4} \\
(X_t, Y_t - 1) & \text{w.p. } \frac{1}{4} 
\end{cases},
\] (3)

A demonstration of possible steps is shown in Figure 2. We can ask questions like

- What is the probability to reach a certain point from origin $(0, 0)$.
3.3 Brownian Motion (Discrete time version)

**Definition 4.** The Brownian motion problem is to maintain a real value $X$ and every time step, we have

$$X_{t+1} = X_t + V, \quad \text{where } V \sim N(0, 1).$$

3.4 Random Walk on perfect matching of a complete bipartite graph

We start from a perfect matching, and randomly swap two edges. Figure 3 demonstrates the process.

4 Markov Chain

Here is a definition for discrete-time Markov Chain that is not very precise.

**Definition 5.** Markov Chain $M$ is described by a finite set of states $S$ and a transition function $P : S \times S \to R$. For simplicity, we consider $S$ to be finite and $P$ is a $|S| \times |S|$ matrix. We define $X_t$ to be the state at time $t$ ($X_t \in S, t = 0, 1, ...$).

**Remark 1.** the output of transition function is $[0, 1]$ for discrete case and density $R$ for continuous case.

4.1 Markov Property

The Markov property says

$$Pr \left[ X_{t+1} = i_{t+1} \mid X_0 = i_0, X_1 = i_1, ..., X_t = i_t \right] = Pr \left[ X_{t+1} = i_{t+1} \mid X_t = i_t \right].$$

This means that we do not need the whole history but only the current state to know the distribution of next state. In the example of 1-D random walk problem, if we are currently in the state of number 5, we do not
need to know how we reach the current state to know that we will be in either 4 or 6 next step with equal probability.

\[ Pr \left[ X_{t+1} = j \mid X_t = i \right] = P_{i,j}, \]

which is the (i, j) entry of transition matrix \( P \).

**Remark 1.** \( P_{i,j} \) represents the probability of going to state \( j \) from state \( i \) in 1 step.

**Remark 2.** \( P_{i,j} \neq P_{j,i} \) in general. \( P_{i,j} \in [0, 1]; \forall i, \sum_j P_{i,j} = 1. \)

### 4.2 Representing Markov Chains as graphs

We represent states as vertices and transitions as edges, which are often directed. Self-loops are allowed, meaning that it remains the same state on next state. The weights of edges represent the transition probabilities. For example, we have three state 1, 2, and 3 and the transition matrix is

\[
\begin{bmatrix}
0 & 3/4 & 1/4 \\
1/4 & 0 & 3/4 \\
3/4 & 1/4 & 0 \\
\end{bmatrix}
\]

The corresponding Markov Chain graph is shown in Figure 4.

### 4.3 Probability Distribution over the states

The distribution of states

\[ p \in \mathbb{R}^{|S|}, \sum_{i \in S} p_i = 1, \quad \text{and} \quad \forall i \in S, p_i \geq 0. \]

**Claim:** if \( q \) is the probability distribution of the next state, then \( q = pP \), where \( p \) is the distribution of current state. The proof is

\[
q_j = Pr \left[ X_{t+1} = j \right] \\
= \sum_{i \in S} Pr[X_t = i] Pr \left[ X_{t+1} = j \mid X_t = i \right] \\
= \sum_{i \in S} p_i P_{i,j} = (pP)_j
\]
Figure 5: Demonstration of two bad Markov chains where graph is not strongly connected: (a) it cannot go back to vertex 1 anymore once reaching vertex 2; (b) it cannot reach vertices 3 and 4 if starting from vertices 1 or 2, which means the stationary distribution is not unique.

Remark 1. $p$ and $q$ are row vectors.

Remark 2. A lot of Markov Chain properties depend on the transition matrix $P$.

5 Stationary Distribution

Stationary distribution is used to capture the limit behavior of Markov Chain.

Definition 6. Say $\pi$ is a stationary distribution for Markov Chain $P$ if

$$\pi = \pi P.$$

This means if the current state is the stationary distribution, the probability distribution of next state does not change from the current state.

Remark 1. Stationary distribution may not be unique for a Markov Chain.

5.1 “Good” cases for Markov Chains

The “good” cases are that no matter which state (state distribution) we start with, after many steps, state distribution converges to a unique stationary distribution and we want $\pi_i > 0$ for all $i \in S$.

Remark 1. Converging to a unique stationary distribution is called “mixing”.

Remark 2. Even if it converges to a unique stationary distribution, we regard it as bad case if not all $\pi_i > 0$.

5.2 “Bad” cases

Here are the characteristics of bad cases.

1. Graph is not strongly connected. Two examples are shown in Figure 5.

2. Periodic: if the next state is deterministic by the current state. An example is shown in Figure 6.

How to avoid bad cases

1. Make sure graph is strongly connected.
Figure 6: Demonstration of a bad Markov chain that is periodic. This means that next step is deterministic given the current state. Note that there is a converging stationary distribution that is \( \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \). It just never gets any closer starting from other distributions.

2. "Lazy walk": define a new Markov Chain that stays at state \( i \) w.p. \( \frac{1}{2} \) and follows the original transition w.p. \( \frac{1}{2} \). That is

\[
P_{\text{new}} = \frac{1}{2}P + \frac{1}{2}I.
\]

Remark 1. "Lazy walk" adds self-loops to the graph, which does not make it loosely-connected.

Claim: A "lazy walk" is aperiodic because we cannot predict whether it will leave or stay in the same current state.

6 Fundamental Theory of Markov Chains

If a Markov Chain is strongly connected, finite and aperiodic, then there is a unique stationary distribution \( \pi \) for the Markov Chain. Let \( N(i,t) \) be the number of times the Markov Chain visits state \( i \) in first \( t \) steps. Then

\[
\lim_{t \to \infty} \frac{N(i,t)}{t} = \pi_i, \text{ w.p. 1}.
\]

Remark 1. No starting state is required in this theory, and \( N(i,t) \) is a random variable.