Lecture-15: Discrete Time Markov Chains

1 Introduction

We have seen that *i.i.d.* sequences are easiest discrete time processes. However, they don't capture correlation well. Hence, we look at the discrete time stochastic processes of the form

$$X_{n+1} = f(X_n, Z_{n+1}),$$

where $Z : \Omega \to \mathcal{Z}^{\mathbb{N}}$ is an *i.i.d.* sequence independent of initial state $X_0 \in \mathcal{X}$, and $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ is a measurable function. The set $\mathcal{X}$ is called the **state space** of process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$. We consider a countable state space $\mathcal{X}$, and if $X_n = x \in \mathcal{X}$, then we say that the process X is in state x at time n.

Definition 1.1. For the discrete random process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, the history until time n is denoted by

$$\mathcal{F}_n \triangleq \sigma(X_0, \dots, X_n).$$

The natural filtration of process X is denoted by $\mathcal{F}_\bullet = (\mathcal{F}_n : n \in \mathbb{Z}_+)$.

Remark 1. We observe that for a process of the form $X_{n+1} = f(X_n, Z_{n+1})$, the event space at time n is $\mathcal{F}_n \subseteq \sigma(X_0, Z_1, \dots, Z_n)$.

Definition 1.2 (Markov property). A discrete random process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ adapted to a filtration $\mathcal{F}_\bullet$ is said to have **Markov property** if

$$P(\{X_{n+1} \leqslant x\} \mid \mathcal{F}_n) = P(\{X_{n+1} \leqslant x\} \mid \sigma(X_n)), \quad n \in \mathbb{Z}_+.$$

Definition 1.3 (DTMC). For a countable set $\mathcal{X}$, a stochastic process $X : \Omega \to \in \mathcal{X}^{\mathbb{Z}_+}$ is called a **discrete time Markov chain (DTMC)** if it satisfies the Markov property.

Remark 2. For a discrete Markov process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we have

$$P(\{X_{n+1} = y\} \mid \{X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\}) = P(\{X_{n+1} = y\} \mid \{X_n = x\}),$$

for all non-negative integers $n \in \mathbb{Z}_+$ and all states $x_0, x_1, \dots, x_{n-1}, x, y \in \mathcal{X}$.

1.1 Homogeneous Markov chain

Definition 1.4. For each time $n \in \mathbb{Z}_+$, we can define the transition probability

$$p_{xy}(n) \triangleq P(\{X_{n+1} = y\} \mid \{X_n = x\}).$$

When the transition probability does not depend on n, the DTMC is called **homogeneous**. The matrix $P \in [0,1]^{\mathcal{X} \times \mathcal{X}}$ is called the **transition matrix**.

Definition 1.5. If a non-negative matrix $A \in \mathbb{R}_+^{\mathcal{X} \times \mathcal{X}}$ satisfies $\sum_{y \in \mathcal{X}} a_{xy} \leqslant 1$ for all $x \in \mathcal{X}$, then it is called a **sub-stochastic** matrix.

Definition 1.6. If the second property holds with equality, then it is called a **stochastic** matrix.

Remark 3. For a stochastic matrix, the all one column vector $\mathbf{1}^T \in \{1\}^{\mathcal{X}}$ is a right eigenvector with eigenvalue unity, i.e. $A\mathbf{1}^T = \mathbf{1}^T$.

Remark 4. The transition matrix P is stochastic matrix. Each row $p_x = (p_{xy} : y \in \mathcal{X})$ of the stochastic matrix P is a distribution on the state space $\mathcal{X}$. This is the conditional distribution of X_{n+1} given $X_n = x$.

Definition 1.7. If in addition A^T is stochastic, then A is called **doubly stochastic**.

Remark 5. For a doubly stochastic matrix $A = A^T$, and hence $\mathbf{1}A = \mathbf{1}A^T = (A\mathbf{1}^T)^T = \mathbf{1}$. For a doubly stochastic matrix, the all one vector $\mathbf{1} \in \{1\}^{\mathcal{X}}$ is both a left and right eigenvector with eigenvalue unity.

1.2 Transition graph

Let E be the collection of ordered pairs of states $(x,y) \in \mathcal{X} \times \mathcal{X}$ such that $p_{xy} > 0$, i.e.

$$E = \left\{(x,y) \in \mathcal{X} \times \mathcal{X} : p_{xy} > 0\right\}.$$

We say that x is a neighbor of y, when $(x,y) \in E$ and denote it by $x \sim y$. The out and in degrees of a vertex $x \in V$ are defined as

$$\deg_{\text{out}}(x) = |\{y \in V : x \sim y\}| = \sum_{y \in V} \mathbb{1}_{\{(x,y) \in E\}}, \qquad \deg_{\text{in}}(x) = |\{y \in V : y \sim x\}| = \sum_{y \in V} \mathbb{1}_{\{(y,x) \in E\}}.$$

For each edge $e \in E$, we define the weight function $w : E \to [0,1]$ such that $w(e) \triangleq p_{xy}$ for each edge $e = (x,y) \in E$. We observe that for a fixed vertex x, we have $\sum_{e=(x,y) \in E} w(e) = 1$. Then a transition matrix P can be represented by a directed edge-weighted graph $G = (\mathcal{X}, E, w)$.

Example 1.8 (Random walk on lattice). We denote the random particle location on a d-dimensional lattice after n steps by $X_n \in \mathbb{Z}^d$, where the random *i.i.d.* step-size sequence is denoted by $Z : \Omega \to (\mathbb{Z}^d)^{\mathbb{N}}$ having common probability mass function $p \in [0,1]^{\mathbb{Z}^d}$. The particle location at time n is $X_n = \sum_{i=1}^n Z_i$. We will show that X is a homogeneous DTMC.

For a lattice point $x \in \mathbb{Z}^d$, we can write the conditional expectation

$$\mathbb{E}[\mathbb{1}_{\{X_n = x\}} | \mathcal{F}_{n-1}] = \sum_{y \in \mathbb{Z}^d} \mathbb{E}[\mathbb{1}_{\{X_{n-1} = x-y\}} \mathbb{1}_{\{Z_n = y\}} | \mathcal{F}_{n-1}] = \sum_{y \in \mathbb{Z}^d} p(y) \mathbb{1}_{\{X_{n-1} = x-y\}} = \mathbb{E}[\mathbb{1}_{\{X_n = x\}} | \sigma(X_{n-1})].$$

Markov property of the random walk follows from the independence of random step-sizes. Homogeneity follows from the identical distribution of random step-sizes.

1.3 Random walks on graphs

Any homogeneous finite state Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ can be thought of as a random walk on the directed edge weighted transition graph $G = (\mathcal{X}, E, w)$, where $E \subseteq V \times V \triangleq \left\{(x,y) : p_{xy} > 0\right\}$ and $w : E \to [0,1]$ defined as $w_e \triangleq p_e$ for all $e \in E$. Random walk on this graph is denoted by the random location $X_n : \Omega \to \mathcal{X}$ of a particle on this graph after n random steps, where each step is random such that

$$P(\{X_{n+1} = y\} \,|\, \{X_n = x\}) = p_e \mathbb{1}_{\{e=(x,y) \in E\}}.$$

1.4 Chapman Kolmogorov equations

Let $\nu(n) \in \mathcal{M}(\mathcal{X})$ denote the marginal distribution of the process X at time $n \in \mathbb{Z}_+$, i.e. $\nu_x(n) \triangleq P\{X_n = x\}$ for all $x \in \mathcal{X}$.

Definition 1.9. We can define n-step transition probabilities for a homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ for states $x, y \in \mathcal{X}$ and non-negative integers $m, n \in \mathbb{Z}_+$ as

$$p_{xy}^{(n)} \triangleq P(\{X_{n+m} = y\} \,|\, \{X_m = x\}).$$

Remark 6. It follows from the Markov property and law of total probability that

$$p_{xy}^{(m+n)} = \sum_{z \in \mathcal{X}} p_{xz}^{(m)} p_{zy}^{(n)}.$$

We can write this result compactly in terms of transition probability matrix P as $P^{(n)} = P^n$.

Remark 7. We can write this vector $\nu(n)$ in terms of initial probability vector $\nu(0)$ and the transition matrix P as $\nu(n) = \nu(0) P^n$.

Remark 8. Let $f : \mathcal{X} \to \mathbb{R}$ be a vector then we define its inner product of matrix $P : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ as a vector $\langle P, f \rangle : \mathcal{X} \to \mathbb{R}$, where $(Pf)_x \triangleq \langle p_x, f \rangle = \sum_{y \in \mathcal{X}} P_{xy} f_y, \quad x \in \mathcal{X}$. It follows that, we can write $(Pf)_x = \mathbb{E}[f(X_1) | \{X_0 = x\}] = \mathbb{E}_x f(X_1)$.

1.5 Strong Markov property (SMP)

Definition 1.10. Let $\tau : \Omega \to \mathbb{Z}_+$ be an almost surely finite integer valued stopping time adapted to the natural filtration of the stochastic process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$. Then for all $x_0, \dots, x_{n-1}, x, y \in \mathcal{X}$, the process X satisfies the **strong Markov property** if

$$P(\{X_{\tau+1} = y\} \mid \{X_\tau = x, \dots, X_0 = x_0\}) = p_{xy}.$$

Lemma 1.11. *Discrete time Markov chains satisfy the strong Markov property.*

Proof. Let X be a Markov chain and an event $A = \{X_\tau = x, \dots, X_0 = x_0\} \in \mathcal{F}_\tau$. Then, we have

$$P(\{X_{\tau+1} = y\} \cap A) = \sum_{n \in \mathbb{Z}_+} P(\{X_{\tau+1} = y, \tau = n\} \cap A) = \sum_{n \in \mathbb{Z}_+} p_{xy} P(A \cap \{\tau = n\}) = p_{xy} P(A).$$

This equality follows from the fact that the event $\{\tau = n\}$ is completely determined by $\{X_0, \dots, X_n\}$ $\square$

Example 1.12 (Non-stopping time). As an exercise, if we try to use the Markov property on arbitrary random variable τ, the SMP may not hold. Consider a Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with natural filtration $\mathcal{F}_\bullet$. For example, define a non-stopping time $\tau_y : \Omega \to \mathbb{Z}_+$ for some state $y \in \mathcal{X}$

$$\tau_y \triangleq \inf\{n \in \mathbb{Z}_+ : X_{n+1} = y\}.$$

We can verify that τ_y is not a stopping time for the process X. From the definition of τ_y, we have $X_{\tau_y+1} = y$, and for $x \in \mathcal{X}$ such that $p_{xy} > 0$

$$P(\left\{X_{\tau_y+1} = y\right\} \mid \left\{X_{\tau_y} = x, \dots, X_0 = x_0\right\}) = 1 \neq P(\{X_1 = y\} \mid \{X_0 = x\}) = p_{xy}.$$

Example 1.13 (Regeneration points of DTMC). Let $x_0 \in \mathcal{X}$ be a fixed state and $\tau_0 = 0$. Let $\tau^+_{x_0}(n)$ denote the stopping times at which the Markov chain visits state x_0 for the nth time. That is,

$$\tau^+_{x_0}(n) \triangleq \inf\left\{n > \tau^+_{x_0}(n-1) : X_n = x_0\right\}.$$

Then $(X_{\tau^+_{x_0}+m} : m \in \mathbb{Z}_+)$ is a stochastic replica of X with $X_0 = x_0$ and can be studied as a regenerative process.

1.6 Random mapping representation

Proposition 1.14. *Any homogeneous DTMC $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ on finite state space $\mathcal{X}$ has a random mapping representation. That is, there exists an* i.i.d. *sequence $Z : \Omega \to \mathcal{Z}^{\mathbb{N}}$ and a measurable function $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ such that $X_n = f(X_{n-1}, Z_n)$ for each $n \in \mathbb{N}$.*

Proof. We can order any finite set, and hence we can assume the finite state space $\mathcal{X} = [n]$, without any loss of generality. For ith row of the transition matrix P, we can define

$$F_{i,k} \triangleq \sum_{j=1}^{k} p_{ij} = P(\{X_{n+1} \leqslant k\} \mid \{X_n = i\}).$$

We assume $Z : \Omega \to [0,1]^{\mathbb{N}}$ to be a sequence of *i.i.d.* uniform random variables. We define a function $f : [n] \times [0,1] \to [n]$ as

$$f(i,z) \triangleq \sum_{k=1}^{n} k \mathbb{1}_{\left\{F_{i,k-1} \leqslant z < F_{i,k}\right\}}, \quad i \in [n], z \in [0,1].$$

To show that this choice of function f and *i.i.d.* sequence Z works, it suffices to show that $p_{ij} = P\{f(i, Z_n) = j\}$. Indeed, we can write

$$P\{f(i,Z_n) = j\} = \mathbb{E}\mathbb{1}_{\{f(i,Z_n)=j\}} = \mathbb{E}\mathbb{1}_{\left\{F_{i,j-1} \leqslant Z_n < F_{i,j}\right\}} = F_{i,j} - F_{i,j-1} = p_{ij}.$$

$\square$

2 Communicating classes

Definition 2.1. Let $x, y \in \mathcal{X}$. If $p_{xy}^{(n)} > 0$ for some $n \in \mathbb{Z}_+$, then we say that state y is **accessible** from state x and denote it by $x \to y$. If two states $x, y \in \mathcal{X}$ are accessible to each other, they are said to **communicate** with each other and denoted by $x \leftrightarrow y$. A set of states that communicate are called a **communicating class**.

Definition 2.2. A relation R on a set $\mathcal{X}$ is a subset of $\mathcal{X} \times \mathcal{X}$.

Definition 2.3. An equivalence relation $R \subseteq \mathcal{X} \times \mathcal{X}$ has following three properties.

Reflexivity: If $x \in \mathcal{X}$, then $(x, x) \in R$.

Symmetry: If $(x, y) \in R$, then $(y, x) \in R$.

Transitivity: If $(x, y), (y, z) \in R$, then $(x, z) \in R$.

Remark 9. Equivalence relations partition a set $\mathcal{X}$.

Proposition 2.4. *Communication is an equivalence relation.*

Proof. Reflexivity follows from zero-step transition, and symmetry follows from the definition of communicating class. For transitivity, suppose $x \leftrightarrow y$ and $y \leftrightarrow z$. Then we can find $m, n \in \mathbb{N}$ such that $p_{xy}^{(m)} > 0$ and $p_{yz}^{(n)} > 0$. From Chapman Kolmogorov equations, we have $m + n \in \mathbb{N}$ such that

$$p_{xz}^{(m+n)} = \sum_{w \in \mathbb{Z}_+} p_{xw}^{(m)} p_{wz}^{(n)} \geqslant p_{xy}^{(m)} p_{yz}^{(n)} > 0.$$

□

2.1 Irreducibility and periodicity

A consequence of the previous result is that communicating classes are disjoint or identical.

Definition 2.5. A Markov chain with a single communicating class is called an **irreducible** Markov chain.

Definition 2.6. A **class property** is the one that is satisfied by all states in the communicating class.

Remark 10. We will see many examples of class properties. Once we have shown that a property is a class property, then one only needs to check that one of the states in the communicating class has the property for the entire class to have that.

Definition 2.7. We denote the set of recurrence times for a Markov chain with transition probability matrix $P : \mathcal{X} \times \mathcal{X} \to [0, 1]$ to re-visit a state $x \in \mathcal{X}$ as

$$A_x \triangleq \left\{ n \in \mathbb{N} : p_{xx}^{(n)} > 0 \right\} \subseteq \mathbb{N}.$$

Remark 11. If one can re-visit a state x in m and n steps, then also in $m + n$ steps, since $p_{xx}^{(m+n)} \geqslant p_{xx}^{(m)} p_{xx}^{(n)}$. It follows that this set is closed under addition.

Definition 2.8. The **period** of state x is defined as $d(x) \triangleq \gcd(A_x)$. If the period is 1, we say the state is **aperiodic**.

Proposition 2.9. *Periodicity is a class property.*

Proof. We will show that for two communicating states $x \leftrightarrow y$, the periodicities are identical. We will show that $d(x) | d(y)$ and $d(y) | d(x)$. We choose $m, n \in \mathbb{N}$ such that

$$p_{xx}^{(m+n)} \geqslant p_{xy}^{(m)} p_{yx}^{(n)} > 0, \qquad p_{yy}^{(m+n)} \geqslant p_{yx}^{(n)} p_{xy}^{(m)} > 0.$$

It follows that $m + n \in A_x \cap A_y$. Let $s \in A_x$, then it follows that $m + n + s \in A_y$, since

$$p_{yy}^{(n+s+m)} \geqslant p_{yx}^{(n)} p_{xx}^{(s)} p_{xy}^{(m)} > 0.$$

Hence $d(y) | n + m$ and $d(y) | n + s + m$ which implies $d(y) | s$. Since the choice of $s \in A_x$ was arbitrary, it follows that $d(y) | d(x)$. Similarly, we can show that $d(x) | d(y)$. □

Lemma 2.10. *If A is a set closed under addition and $\gcd(A) = 1$, then there exists $m_0 \in A$ such that $m \in A$ for all $m \geqslant m_0$.*

Proof. Since the set of recurrence A is closed under addition, for any $a \in A$, we have $na \in A$ for all $n \in \mathbb{N}$. If the minimal element of A is 1, then there is nothing to prove.

Let $|A| = N$. By Bezout's Lemma, there exists $c \in \mathbb{Z}^N$ such that,

$$\sum_{i \in [N]} c_i a_i = 1.$$

Hence,

$$\sum_{i \in [N]} c_i a_i \mod a_1 \equiv 1 \mod a_1$$
$$\sum_{i \in [N]} (c_i \mod a_1) a_i \mod a_1 \equiv 1 \mod a_1.$$

Let $f_i \triangleq c_i \mod a_1 \geqslant 0$. Also, define $d = \sum_{i \in [N]} f_i a_i$. Clearly, $d \mod a_1 \equiv 1 \mod a_1$. Hence, any number $m > a_1 d$ can be represented as

$$m = kd + la_1, \ m \mod a_1 = k, \ l = (m - kd)/a_1.$$

$\square$

Proposition 2.11. *If a Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ on a finite state space $\mathcal{X}$ is irreducible and aperiodic, then there exists an integer n_0 such that $p_{xy}^{(n)} > 0$ for all $x, y \in \mathcal{X}$ and $n \geqslant n_0$.*

Proof. Since periodicity is a class property, it follows that $\gcd(A_x) = 1$ for all states $x \in \mathcal{X}$. Further, we have $m_x \in A_x$ such that $n \in A_x$ for all $n \geqslant m_x$. Further for any pair of states $x, y \in \mathcal{X}$, we can find $n_{xy} \in \mathbb{N}$ such that $p_{xy}^{(n_{xy})} > 0$ from the irreducibility of the Markov chain. It follows that $p_{xy}^{(n)} > 0$ for all $n \geqslant n_{xy} + m_y \in \mathbb{N}$. Since the state space $\mathcal{X}$ is finite, we have a finite $n_0 \triangleq \sup_{x \in \mathcal{X}} m_x + \sup_{x,y \in \mathcal{X}} n_{xy} \in \mathbb{N}$ such that $p_{xy}^n > 0$ for any state $x, y \in \mathcal{X}$ for all $n \geqslant a$. $\square$

Example 2.12 (Random walk on a ring). Let $G = (V, E)$ be a finite graph where $V = \{0, \dots, n-1\}$ and $E = \{(i, i+1) : i \in V\}$ where addition is modulo n. Let $\xi : \Omega \to \{-1, 1\}^{\mathbb{N}}$ be a random *i.i.d.* sequence of step-sizes with $\mathbb{E}\xi_n = 2p - 1$. We denote the location of particle after n random steps by $X_n \triangleq X_0 + \sum_{i=1}^n \xi_i$. It follows that the random walk $X : \Omega \to V^{\mathbb{N}}$ is an irreducible homogeneous Markov chain with period 2 if n is even, else it is aperiodic as well.

A Bézout's identity

Lemma A.1 (Bézout). *Let a and b be integers with greatest common divisor d. Then there exist integers x and y such that $ax + by = d$. Moreover, the integers of the form $az + bt$ are exactly the multiples of d.*

Proof. $\square$