Lecture-15: Discrete Time Markov Chains

1 Introduction

Definition 1.1. Consider a discrete random process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, where X_n is the state of the process at time n. The set $\mathcal{X}$ is called the *state space* of discrete time process X. The history of the process 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 $\mathcal{F}_\bullet \triangleq (\mathcal{F}_n : n \in \mathbb{Z}_+)$.

Remark 1. Let $Z : \Omega \to \mathcal{Z}^{\mathbb{N}}$ be an *i.i.d.* sequence. 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 is independent of initial state $X_0 \in \mathcal{X}$, and $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ is a measurable function. For process X, the history until time n is $\mathcal{F}_n \subseteq \sigma(X_0, Z_1, \dots, Z_n)$.

Definition 1.2. A discrete random process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ adapted to its natural filtration $\mathcal{F}_\bullet$ is said to have the *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. For a countable set $\mathcal{X}$, a stochastic process $X : \Omega \to \in \mathcal{X}^{\mathbb{Z}_+}$ is called a *discrete time Markov chain (discrete time Markov chain)* if it satisfies the Markov property.

Remark 2. For a discrete Markov process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we have for all $n \in \mathbb{Z}_+$ and states $x_0, x_1, \dots, x_{n-1}, x, y \in \mathcal{X}$,

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

Definition 1.4. For a countable state space $\mathcal{X}$, we define the set of probability measures on $\mathcal{X}$ as

$$\mathcal{M}(\mathcal{X}) \triangleq \left\{ \nu \in [0,1]^{\mathcal{X}} : \sum_{x \in \mathcal{X}} \nu_x = 1 \right\}.$$

1.1 Homogeneous Markov chain

Definition 1.5. We can define the transition probability $p_{xy}(n) \triangleq P(\{X_{n+1} = y\} \mid \{X_n = x\})$, for each time $n \in \mathbb{Z}_+$. When the transition probability does not depend on n, the discrete time Markov chain is called *homogeneous*. The matrix $p \in \mathcal{M}(\mathcal{X})^{\mathcal{X}}$ is called the *transition matrix*.

Example 1.6 (Random walk on lattice). For the random *i.i.d.* step-size sequence $Z : \Omega \to (\mathbb{Z}^d)^{\mathbb{N}}$ having common probability mass function $p \in \mathcal{M}(\mathbb{Z}^d)$, we denote the random particle location on a d-dimensional lattice after n steps as $X_n \triangleq \sum_{i=1}^n Z_i$. Let $\mathcal{F}_\bullet$ be the natural filtration associated with process X. We will show that X is a homogeneous discrete time Markov chain. For a lattice point $x \in \mathbb{Z}^d$, we can write the conditional expectation

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

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

Definition 1.7. 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 A is called a *sub-stochastic* matrix. If $\sum_{y\in\mathcal{X}} a_{xy} = 1$ for all $x \in \mathcal{X}$, then A is called a *stochastic* matrix. If A and $A^\top$ are stochastic matrices, then A is called *doubly stochastic* matrix.

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

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

Remark 5. For a doubly stochastic matrix A, the all one row vector $\mathbf{1}$ is a left eigenvector and $\mathbf{1}^\top$ is a right eigenvector, both with eigenvalue unity. To see this we observe that $\mathbf{1}A^\top = (A\mathbf{1}^\top)^\top = \mathbf{1}$.

1.2 Transition graph

Consider a discrete time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with probability transition matrix $p \in \mathcal{M}(\mathcal{X})^{\mathcal{X}}$.

Definition 1.8. We define edge set E to be the collection of ordered pairs of states $(x,y) \in \mathcal{X}\times\mathcal{X}$ such that $p_{xy} > 0$. That is, $E \triangleq \{(x,y) \in \mathcal{X}\times\mathcal{X} : p_{xy} > 0\}$. 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 $e = (x,y) \in E$. Then a transition matrix p can be represented by a directed edge-weighted graph $G \triangleq (\mathcal{X}, E, w)$.

Definition 1.9. We say that x is a neighbor of y, when $(x,y) \in E$ and denote it by $x \sim y$. The out neighborhood of x is defined as $\mathrm{nbr}_{\mathrm{out}}(x) \triangleq \{y \in \mathcal{X} : (x,y) \in E\}$ and has cardinality $\deg_{\mathrm{out}}(x) \triangleq |\mathrm{nbr}_{\mathrm{out}}|$. The in neighborhood of x is defined as $\mathrm{nbr}_{\mathrm{in}}(x) \triangleq \{y \in \mathcal{X} : (y,x) \in E\}$ and has cardinality $\deg_{\mathrm{in}}(x) \triangleq |\mathrm{nbr}_{\mathrm{in}}|$.

Remark 6. We observe that for a fixed vertex $x \in \mathcal{X}$, we have $\sum_{y\in\mathrm{nbr}_{\mathrm{out}}(x)} w_{xy} = 1$.

Remark 7. 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 $G = (\mathcal{X}, E, w)$. The location of a single particle on this graph after n random steps is denoted by X_n, where particle can jump from one location to another if it is connected by an edge and with the jump probability being equal to the edge weight. That is,

$$P(\{X_{n+1} = y\} \mid \{X_n = x\}) = w_{xy}\mathbb{1}_{\{(x,y)\in E\}}.$$

1.3 Chapman Kolmogorov equations

Definition 1.10. 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 1.11. 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

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

Remark 8. It follows from the Markov property and the 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 9. 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 10. Let $f \in \mathbb{R}^{\mathcal{X}}$ be a vector then we define its inner product with a real-valued matrix $P \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ as a vector $Pf \triangleq \langle P, f\rangle \in \mathbb{R}^{\mathcal{X}}$, where $(Pf)_x \triangleq \langle p_x, f\rangle \triangleq \sum_{y\in\mathcal{X}} p_{xy} f_y$, for all $x \in \mathcal{X}$. It follows that we can write $(pf)_x = \mathbb{E}[f(X_1) \mid \{X_0 = x\}] = \mathbb{E}_x f(X_1)$ for a time homogeneous discrete time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with transition probability matrix p.

1.4 Strong Markov property (SMP)

Definition 1.12. Let $\tau : \Omega \to \mathbb{Z}_+$ be an almost surely finite stopping time adapted to the natural filtration $\mathcal{F}_\bullet$ of the stochastic process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$. Then the process X satisfies the *strong Markov property* if $P(\{X_{\tau+1} = y\} \mid \{X_\tau = x, \dots, X_0 = x_0\}) = p_{xy}$ for all states $x_0, \dots, x_{n-1}, x, y \in \mathcal{X}$ and stopping time τ.

Lemma 1.13. *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 1.14 (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 $\mathcal{F}_\bullet$. We define a non-stopping time random variable $\tau_y : \Omega \to \mathbb{Z}_+$ 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 $x \in \mathcal{X} \setminus \{y\}$ such that $p_{xy} > 0$

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

Example 1.15 (Regeneration points of discrete time Markov chain). Let $x_0 \in \mathcal{X}$ be a fixed state and $\tau^+_{x_0}(0) \triangleq 0$. Let $\tau^+_{x_0}(n)$ denote the stopping times at which the Markov chain visits state x_0 for the nth time. That is,

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

If $\tau^+_{x_0}$ is almost surely finite, 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.

1.5 Random mapping representation

Proposition 1.16. *Any homogeneous discrete time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ 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 xth row of the transition matrix $p \in \mathcal{M}(\mathcal{X})^{\mathcal{X}}$, we can define

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

We define $\mathcal{Z} \triangleq [0,1]$ and assume $Z : \Omega \to \mathcal{Z}^{\mathbb{N}}$ to be a sequence of *i.i.d.* uniform random variables. We define a function $f : [N] \times \mathcal{Z} \to [N]$ for each $x \in \mathcal{X}$ and $z \in \mathcal{Z}$ as

$$f(x,z) \triangleq \sum_{k=1}^{N} k \mathbb{1}_{\{F_{x,k-1} < z \leqslant F_{x,k}\}} = \sum_{k=1}^{N} k \mathbb{1}_{(F_{x,k-1}, F_{x,k}]}(z).$$

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

$$P\{f(x,Z_n) = y\} = \mathbb{E}\mathbb{1}_{\{f(x,Z_n)=y\}} = \mathbb{E}\mathbb{1}_{(F_{x,y-1}, F_{x,y}]}(Z_n) = F_{x,y} - F_{x,y-1} = p_{xy}.$$

$\square$