Lecture-14: 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 discrete time process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$. 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 $\mathcal{F}_\bullet \triangleq (\mathcal{F}_n : n \in \mathbb{Z}_+)$.

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. 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 (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 $x_0, x_1, \dots, x_{n-1}, x, y \in \mathcal{X}$.

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 DTMC is called **homogeneous**. The matrix $P \in [0,1]^{\mathcal{X} \times \mathcal{X}}$ 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 $p \in [0,1]^{\mathbb{Z}^d}$, we denote the random particle location on a d-dimensional lattice after n steps by $X_n \in \mathbb{Z}^d$ defined at time n as $X_n \triangleq \sum_{i=1}^n Z_i$. 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.

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 $\sum_{y\in\mathcal{X}} a_{xy} = 1$, then A is called a **stochastic** matrix. If A and A^T are stochastic matrices, then A is called **doubly stochastic** matrix.

Remark 3. Let $\mathbf{1} : \mathcal{X} \to \{1\}$ be the all one vector. For a stochastic matrix, the all one column vector $\mathbf{1}^T$ 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 $p_x \triangleq (p_{xy} : y \in \mathcal{X}) \in \mathcal{M}(\mathcal{X})$ 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 right eigenvector and $\mathbf{1}^T$ is a left eigenvector, both with eigenvalue unity. To see this we observe that $\mathbf{1}A^T = (A\mathbf{1}^T)^T = \mathbf{1}$.

1.2 Transition graph

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

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

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 $x \in \mathcal{X}$ are defined as

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

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 $x \in \mathcal{X}$, 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 $G \triangleq (\mathcal{X}, E, w)$.

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 $G = (\mathcal{X}, E, w)$. The location of a single particle on this graph after n random steps is denoted by $X_n : \Omega \to \mathcal{X}$, 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,

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

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 1.8. 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 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 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 with a matrix $P : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ as a vector $\langle P, f\rangle : \mathcal{X} \to \mathbb{R}$, where $(Pf)_x \triangleq \langle p_x, f\rangle = \sum_{y\in\mathcal{X}} p_{xy} f_y$, for all $x \in \mathcal{X}$. It follows that, we can write $(Pf)_x = \mathbb{E}[f(X_1) | \{X_0 = x\}] = \mathbb{E}_x f(X_1)$ for a time homogeneous discrete time Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with transition probability matrix P.

1.5 Strong Markov property (SMP)

Definition 1.9. Let $\tau : \Omega \to \mathbb{Z}_+$ be an almost surely finite integer valued stopping time adapted to the natural filtration $\mathcal{F}_\bullet$ of the stochastic process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$. Then for all states $x_0, \dots, x_{n-1}, x, y \in \mathcal{X}$, 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 1.10. *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 $\{X_0, \dots, X_n\}$ □

Example 1.11 (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 1.12 (Regeneration points of DTMC). Let $x_0 \in \mathcal{X}$ be a fixed state and $\tau^+_{x_0}(0) = 0$. Let $\tau^+_{x_0}(n)$ denote the stopping times at which the Markov chain visits state x_0 for the nth time. That is,

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

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 1.13. *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 $f : [n] \times [0,1] \to [n]$ for each $i \in [n]$ and $z \in [0,1]$ as $f(i,z) \triangleq \sum_{k=1}^{n} k \mathbb{1}_{\{F_{i,k-1} \leqslant z < F_{i,k}\}}$. 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

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

□