Lecture-20: Discrete Time Markov Chains

1 Introduction

We have seen that *iid* sequences are easiest discrete time random processes. However, they don't capture correlation well.

Definition 1.1. For a state space $\mathcal{X} \subseteq \mathbb{R}$ and the random sequence $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we define the history until time $n \in \mathbb{Z}_+$ as $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.

Remark 1. Recall that the event space $\mathcal{F}_n$ is generated by the historical events of the form

$$H_n = \cap_{i=1}^n \{X_i \leqslant x_i\}, \text{ where } x \in \mathbb{R}^n.$$

Remark 2. When the state space $\mathcal{X}$ is countable, the event space $\mathcal{F}_n$ is generated by the historical events of the form

$$H_n = \cap_{i=1}^n \{X_i = x_i\}, \text{ where } x \in \mathcal{X}^n.$$

Definition 1.2 (DTMC). For a countable set $\mathcal{X}$, a discrete-valued random sequence $(X_n \in \mathcal{X}^\Omega : n \in \mathbb{Z}_+)$ is called a **discrete time Markov chain (DTMC)** if for all positive integers $n \in \mathbb{Z}_+$, all states $x, y \in \mathcal{X}$, and any historical event $H_{n-1} = \cap_{m=0}^{n-1} \{X_m = x_m\} \in \mathcal{F}_n$ for $(x_0, \dots, x_{n-1}) \in \mathcal{X}^n$, the process X satisfies the Markov property

$$P(\{X_{n+1} = y\} \mid H_{n-1} \cap \{X_n = x\}) = P(\{X_{n+1} = y\} \mid \{X_n = x\}).$$

The probability of a discrete time Markov chain X being in state $y \in \mathcal{X}$ at time $n+1$ from a state $x \in \mathcal{X}$ at time n, is determined by the **transition probability** denoted by

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

The set $\mathcal{X}$ is called the state space of the Markov chain. The **transition probability matrix** at time n is denoted by $P(n) \in [0,1]^{\mathcal{X} \times \mathcal{X}}$, such that $P_{xy}(n) = p_{xy}(n)$.

Remark 3. We observe that each row $P_x(n) = (p_{xy}(n) : y \in \mathcal{X})$ is the conditional distribution of X_{n+1} given $X_n = x$.

Definition 1.3 (Random walk). Consider the state space $\mathcal{X} = \mathbb{R}^d$, and the countable index set $T = \mathbb{Z}_+$. Consider the random process $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ be an be an independent (not necessarily identical) sequence. Let $S_0 = 0$ and $S_n \triangleq \sum_{i=1}^n X_i$ for all $n \in \mathbb{Z}_+$, then the process $S : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ is called a **random walk**.

Remark 4. We can think of S_n as the random location of a particle after n steps, where the particle starts from origin and takes steps of size X_i at the ith step.

Theorem 1.4 (Random walk). *For a random walk $(S_n : n \in \mathbb{N})$ with independent step-size sequence X, the following are true.*

i_ *The first two moments are $\mathbb{E}S_n = \sum_{i=1}^n \mathbb{E}X_i$ and $\mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_i]$.*

ii_ *Random walk is non-stationary with independent increments. The disjoint increments are stationary if the step-size sequence X is identically distributed.*

iii_ Random walk is a Markov sequence.

Proof. Results follow from the independence of the step-size sequence X.

i_ Follows from the linearity of expectation and independence of step sizes.

ii_ Since the mean is time dependent, random walk is non-stationary process. Independence of increments follows from the independence of step sizes. That is, since $\mathcal{F}_n = \sigma(S_0, S_1, \dots, S_n) = \sigma(S_0, X_1, \dots, X_n)$ and the collection $(X_{n+1}, \dots, X_m)$ is independent of $\sigma(S_0, X_1, \dots, X_n)$ for all $m > n$. Since $S_m - S_n = X_{n+1} + \cdots + X_m \in \sigma(X_{n+1}, \dots, X_m)$, we have the independent increments. When the step-sizes are also identically distributed, the joint distributions of $(X_1, \dots, X_{m-n})$ and $(X_{n+1}, \dots, X_m)$ are identical. This implies the stationarity of increments for *i.i.d.* step-sizes.

iii_ Given the historical event $H_{n-1} \triangleq \cap_{k=1}^{n-1} \{S_k \leqslant s_k\}$ and the current state $\{S_n \leqslant s_n\}$, we can write the conditional probability

$$\begin{aligned} P(\{S_{n+1} \leqslant s_{n+1}\} \mid H_{n-1} \cap \{S_n \leqslant s_n\}) &= P(\{X_{n+1} \leqslant s_{n+1} - S_n\} \mid H_{n-1} \cap \{S_n \leqslant s_n\}) \\ &= P(\{S_{n+1} \leqslant s_{n+1}\} \mid \{S_n \leqslant s_n\}). \end{aligned}$$

The equality in the second line follows from the independence of the step-size sequence. In particular, from the independence of X_{n+1} from the collection $\sigma(S_0, X_1, \dots, X_n) = \sigma(S_0, S_1, \dots, S_n)$. For the countable state space $\mathcal{X}$, an given the historical event $H_{n-1} \triangleq \cap_{k=1}^{n-1} \{S_k = s_k\}$ and the current state $\{S_n = s_n\}$, we can write the conditional probability

$$\begin{aligned} P(\{S_{n+1} = s_{n+1}\} \mid H_{n-1} \cap \{S_n = s_n\}) &= P(\{X_{n+1} = s_{n+1} - S_n\} \mid H_{n-1} \cap \{S_n = s_n\}) \\ &= P(\{S_{n+1} = s_{n+1}\} \mid \{S_n = s_n\}) = P\{X_{n+1} = s_{n+1} - s_n\}. \end{aligned}$$

Definition 1.5. For all states $x, y \in \mathcal{X}$, a matrix $A \in \mathbb{R}_+^{\mathcal{X} \times \mathcal{X}}$ with non-negative entries is called **sub-stochastic** if the row-sum $\sum_{y \in \mathcal{X}} a_{xy} \leqslant 1$ for all rows $x \in \mathcal{X}$. If the above property holds with equality for all rows, then it is called a **stochastic** matrix. If matrices A and A^T are both stochastic, then the matrix A is called **doubly stochastic**.

Remark 5. We make the following observations for the stochastic matrices.

i_ Every probability transition matrix $P(n)$ is a stochastic matrix.

ii_ All the entries of a sub-stochastic matrix lie in $[0,1]$.

iii_ Each row of the stochastic matrix $A \in \mathbb{R}_+^{\mathcal{X} \times \mathcal{X}}$ is probability mass function over the state space $\mathcal{X}$.

iv_ Every finite stochastic matrix has a right eigenvector with unit eigenvalue. This can be observed by taking $\mathbf{1}^T = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}$ to be an all-one vector of length $|\mathcal{X}|$. Then we see that $A\mathbf{1} = \mathbf{1}$, since for each $x \in \mathcal{X}$

$$(A\mathbf{1})_x = \sum_{y \in \mathcal{X}} a_{xy} \mathbf{1}_y = \frac{1}{n} \sum_{y \in \mathcal{X}} a_{xy} = \mathbf{1}_x.$$

v_ Every finite doubly stochastic matrix has a left and right eigenvector with unit eigenvalue. This follows from the fact that finite stochastic matrices A and A^T have a common right eigenvector $\mathbf{1}$. It follows that A has a left eigenvector $\mathbf{1}^T$.

vi_ For a probability transition matrix $P(n)$, we have

$$\sum_{y \in \mathcal{X}} f(y) p_{xy}(n) = \mathbb{E}[f(X_{n+1}) \mid X_n = x].$$

2 Homogeneous Markov chain

In general, not much can be said about Markov chains with index dependent transition probabilities. Hence, we consider the simpler case where the transition probabilities $p_{xy}(n) = p_{xy}$ are independent of the index. We call such DTMC as **homogeneous** and call the linear operator $P = (p_{xy} : x, y \in \mathcal{X})$ the **transition matrix**.

Example 2.1 (Integer random walk). For a one-dimensional integer valued random walk $X : \Omega \to \mathbb{Z}^{\mathbb{N}}$ with *i.i.d.* unit step size sequence $Z : \Omega \to \{-1, 1\}^{\mathbb{N}}$ such that $P\{Z_1 = 1\} = p$, the following are true.

i_ The transition operator $P \in [0, 1]^{\mathbb{Z}_+ \times \mathbb{Z}_+}$ is given by the entries

$$p_{xy} = p 1_{\{y=x+1\}} + (1-p) 1_{\{y=x-1\}}.$$

ii_ Number of positive steps after n steps is Binomial (n, p).

iii_ $P\{X_n = k\} = \binom{n}{(n+k)/2} p^{(n+k)/2} q^{(n-k)/2}$ for $n + k$ even, and 0 otherwise.

Example 2.2 (Sequence of experiments). Consider a random sequence of experiments, where the nth outcome is denoted by X_n, such that each experiment has two possible outcomes in $\mathcal{X} = \{S, F\}$. We assume that it takes unit time to perform each experiment.

Let $p, q \in [0, 1]$. Given the outcome was S, the probability of next outcome being S is $1 - p$. Similarly, given the outcome was F, the probability of next outcome being F is $1 - q$. We can see that $X = (X_n : n \in \mathbb{Z}_+)$ is homogeneous Markov chain, with probability transition matrix

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

We denote the conditional distribution of X_{n+1} given $X_0 = S$ by ν_{n+1}, and the conditional distribution of X_{n+1} given $X_0 = F$ by μ_{n+1}. That is,

$$\nu_n = \begin{bmatrix} P(\{X_n = S\} \mid \{X_0 = S\}) & P(\{X_n = F\} \mid \{X_0 = S\}) \end{bmatrix},$$
$$\mu_n = \begin{bmatrix} P(\{X_n = S\} \mid \{X_0 = F\}) & P(\{X_n = F\} \mid \{X_0 = F\}) \end{bmatrix}.$$

Let π_0 be the initial distribution on the experiment outcome, and π_n be the distribution of the experiment outcome at time n. Then, we can write

$$\begin{aligned} \pi_n(S) &\triangleq P\{X_n = S\} = P(\{X_n = S\} \mid \{X_0 = S\})\pi_0(S) + P(\{X_n = S\} \mid \{X_0 = F\})\pi_0(F) \\ &= \nu_n(S)\pi_0(S) + \mu_n(S)\pi_0(F). \end{aligned}$$

Similarly, we can write $\pi_n(F) = \nu_n(F)\pi_0(S) + \mu_n(F)\pi_0(F)$. That is, we can write

$$\pi_n \triangleq \begin{bmatrix} \pi_n(S) & \pi_n(F) \end{bmatrix} = \begin{bmatrix} \pi_0(S) & \pi_0(F) \end{bmatrix} \begin{bmatrix} \nu_n(S) & \nu_n(F) \\ \mu_n(S) & \mu_n(F) \end{bmatrix} = \pi_0 \begin{bmatrix} \nu_n \\ \mu_n \end{bmatrix}.$$

That is to compute the unconditional distribution of X_n, given initial distribution π_0, we need to compute conditional distributions ν_n and μ_n. We can see that

$$\nu_1 = \begin{bmatrix} 1-p & p \end{bmatrix}, \qquad \nu_2 = \begin{bmatrix} (1-p)^2 + pq & (1-p)p + p(1-q) \end{bmatrix},$$
$$\mu_1 = \begin{bmatrix} q & 1-q \end{bmatrix}, \qquad \mu_2 = \begin{bmatrix} q(1-p) + (1-q)q & (1-q)^2 + qp \end{bmatrix}.$$

This method of direct computation can quickly become too cumbersome.