Lecture-19: Discrete Time Markov Chains

1 Introduction

We have seen that *i.i.d.* 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

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

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

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

Definition 1.2 (DTMC). For a countable set $\mathcal{X}$, a discrete-valued random sequence $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ 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

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

Remark 3. The above definition is equivalent to $P(\{X_{n+1} \leqslant x\} \mid \mathcal{F}_n) = P(\{X_{n+1} \leqslant x\} \mid \sigma(X_n))$, for discrete time discrete state space Markov chain, since $\mathcal{F}_n = \sigma(H_n(x) : x \in \mathcal{X}^n)$ and $\sigma(X_n) = \sigma(\{X_n = x\}, x \in \mathcal{X})$.

Definition 1.3. 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 **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 4. We observe that each row $P_x(n) = (p_{xy}(n) : y \in \mathcal{X})$ is the conditional distribution of X_{n+1} given $X_n = x$.

Theorem 1.4 (Random walk). *For a random walk $S : \Omega \to \mathcal{X}^{\mathbb{N}}$ with independent step-size sequence $X : \Omega \to \mathcal{X}^{\mathbb{N}}$, the following are true.*

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

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 for countable state space $\mathcal{X}$.*

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 $S_m - S_n = X_{n+1} + \dots + X_m \in \sigma(X_{n+1}, \dots, X_m)$, 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. For the countable state space $\mathcal{X}$, an given the historical event $H_{n-1}(s) \triangleq \cap_{k=1}^{n-1} \{S_k = s_k\}$ and the current state $\{S_n = s_n\}$, we can write the conditional probability

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

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 $\sigma(S_0, X_1, \dots, X_n) = \sigma(S_0, S_1, \dots, S_n)$.

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

$$(A\mathbf{1})_x = \sum_{y \in \mathcal{X}} a_{xy} \mathbf{1}_y = \sum_{y \in \mathcal{X}} a_{xy} = \mathbf{1}_x, \text{ for each } x \in \mathcal{X}.$$

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.

Definition 2.1. A DTMC with the probability transition matrix $P(n)$ that is independent of the index, is called **homogeneous**.

Example 2.2 (Integer random walk). For a one-dimensional integer valued random walk $S : \Omega \to \mathbb{Z}^{\mathbb{N}}$ with *i.i.d.* unit step size sequence $X : \Omega \to \{-1, 1\}^{\mathbb{N}}$ such that $P\{X_1 = 1\} = p$, the following are true.

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

$$p_{xy} = p1_{\{y=x+1\}} + (1-p)1_{\{y=x-1\}}.$$

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

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

Definition 2.3. Consider a homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with countable state space $\mathcal{X}$ and transition matrix P. We would respectively denote the conditional probability of events and conditional expectation of random variables, conditioned on the event $\{X_0 = x\}$, by

$$P_x(A) \triangleq P(A \mid \{X_0 = x\}), \qquad \mathbb{E}_x[Y] \triangleq \mathbb{E}\left[Y \mid \{X_0 = x\}\right].$$

Example 2.4 (Sequence of experiments). Consider a random sequence of experiments, where the success of nth outcome is indicated by X_n. That is, $X : \Omega \to \{0,1\}^{\mathbb{Z}_+}$ is a random sequence of outcomes of experiments.

Let $p, q \in [0,1]$. Given the outcome was 0, the probability of next outcome being 0 is $1-p$. Similarly, given the outcome was 1, the probability of next outcome being 1 is $1-q$. We can see that X 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 $\{X_0 = 0\}$ by ν_{n+1}, and the conditional distribution of X_{n+1} given $\{X_0 = 1\}$ by μ_{n+1}. That is,

$$\nu_n = \begin{bmatrix} P_0(\{X_n = 0\}) & P_0(\{X_n = 1\}) \end{bmatrix},$$
$$\mu_n = \begin{bmatrix} P_1(\{X_n = 0\}) & P_1(\{X_n = 1\}) \end{bmatrix}.$$

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

$$\pi_n(0) \triangleq P\{X_n = 0\} = P_0(\{X_n = 0\})\pi_0(0) + P_1(\{X_n = 0\})\pi_0(1) = \nu_n(0)\pi_0(0) + \mu_n(0)\pi_0(1).$$

Similarly, we can write $\pi_n(1) = \nu_n(1)\pi_0(0) + \mu_n(1)\pi_0(1)$. That is, we can write

$$\pi_n \triangleq \begin{bmatrix} \pi_n(0) & \pi_n(1) \end{bmatrix} = \begin{bmatrix} \pi_0(0) & \pi_0(1) \end{bmatrix} \begin{bmatrix} \nu_n(0) & \nu_n(1) \\ \mu_n(0) & \mu_n(1) \end{bmatrix} = \pi_0 \begin{bmatrix} \nu_n \\ \mu_n \end{bmatrix}.$$

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.

3 n-step transition

Proposition 3.1. *Conditioned on the initial state, any finite dimensional distribution of a homogeneous Markov chain is stationary. That is, for any finite $n, m \in \mathbb{N}$ and $x \in \mathcal{X}^{\{0,\dots,n\}}$, we have*

$$P(\cap_{i=1}^n \{X_i = x_i\} \mid \{X_0 = x_0\}) = P(\cap_{i=1}^n \{X_{m+i} = x_i\} \mid \{X_m = x_0\}) = \prod_{i=1}^n p_{x_{i-1}x_i}.$$

Proof. To this end, we compute the transition probabilities for the path $(x_1, \dots, x_n)$ taken by
(i) the sample path $(X_1, \dots, X_n)$ given the event $\{X_0 = x_0\}$ and
(ii) by the sample path $(X_{m+1}, \dots, X_{m+n})$ given the event $\{X_m = x_0\}$.
For each $i \in \{0, \dots, n\}$, we can define events $H_i \triangleq \cap_{j=0}^i \{X_j = x_j\}$. We observe that $H_i = \{X_i = x_i\} \cap H_{i-1}$ and $H_i \in \mathcal{F}_i = \sigma(X_0, \dots, X_i)$ for all $i \in \mathbb{N}$. From the definition of event H_{n-1} and the conditional probability, we can write

$$P_{x_0}(H_n) = P_{x_0}(\{X_n = x_n\} \cap H_{n-1}) = P(\{X_n = x_n\} \mid H_{n-1}) P_{x_0}(H_{n-1}).$$

Using the fact that $H_{n-1} = \{X_{n-1} = x_{n-1}\} \cap H_{n-2}$, and the Markovity and homogeneity of the process X, we obtain

$$P(\{X_n = x_n\} \mid H_{n-1}) = P(\{X_n = x_n\} \mid \{X_{n-1} = x_{n-1}\} \cap H_{n-2}) = p_{x_{n-1}x_n}.$$

Inductively, we can write the conditional joint distribution of H_n given the event $\{X_0 = x_0\}$ as

$$P_{x_0}(H_n) = p_{x_0x_1} \dots p_{x_{n-1}x_n}.$$

Similarly, we can write for the sample path $(X_{m+1}, \dots, X_{m+n})$ given $X_m = x_0$,

$$P(\{X_{m+1} = x_1, \dots, X_{m+n} = x_n\} \mid \{X_m = x_0\}) = \prod_{i=1}^n P(\{X_{m+i} = x_i)\} \mid \{X_{m+i-1} = x_{i-1}\}) = p_{x_0x_1} \dots p_{x_{n-1}x_n}.$$

□

Corollary 3.2. *The n-step transition probabilities are stationary for any homogeneous Markov chain. That is, for any states $x_0, x_n \in \mathcal{X}$ and $n, m \in \mathbb{N}$, we have*

$$P(\{X_{n+m} = x_n\} | \{X_m = x_0\}) = P(\{X_n = x_n\} | \{X_0 = x_0\}).$$

Proof. It follows from summing over intermediate steps. Let $x \triangleq (x_1, \dots, x_{n-1}) \in \mathcal{X}^{n-1}$, then we can partition the event $\{X_n = x_n\}$ in terms of disjoint events $\{H_n(x, x_n) : x \in \mathcal{X}^{n-1}\}$ defined by $H_n(x, x_n) \triangleq \cap_{i=1}^n \{X_i = x_i\}$, and partition the event $\{X_{m+n} = x_n\}$ in terms of the disjoint events $\{F_n(x, x_n) : x \in \mathcal{X}^{n-1}\}$ defined by $F_n(x, x_n) \triangleq \cap_{i=1}^n \{X_{m+i} = x_i\}$. Then, we can write

$$\{X_n = x_n\} = \cup_{x \in \mathcal{X}^{n-1}} H_n(x, x_n), \qquad \{X_{m+n} = x_n\} = \cup_{x \in \mathcal{X}^{n-1}} F_n(x, x_n).$$

From the stationarity in joint distribution conditioned on initial state for the homogeneous Markov chain X, we have

$$P(F_n(x, x_n) \mid \{X_m = x_0\}) = P(H_n(x, x_n) \mid \{X_0 = x_0\}).$$

Using the law of total probability, we can write the conditional probability

$$P_{x_0}\{X_n = x_n\} = \sum_{x \in \mathcal{X}^{n-1}} P_{x_0}(H_n(x, x_n)), \quad P(\{X_{m+n} = x_n\} \mid \{X_m = x_0\}) = \sum_{x \in \mathcal{X}^{n-1}} P(F_n(x, x_n) \mid \{X_m = x_0\}).$$

The result follows since each term in the summation is equal. □

Definition 3.3. For a homogeneous Markov chain, we can define n-step transition probabilities for $x, y \in \mathcal{X}$ and $m, n \in \mathbb{N}$

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

That is, the row $P_x^{(n)} = (p_{xy}^{(n)} : y \in \mathcal{X})$ is the conditional distribution of X_n given $X_0 = x$.