Lecture-19: Discrete Time Markov Chains

1 Markovity for countable state sequences

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

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

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 $\sigma(X_n) = \sigma(\{X_n = x\}, x \in \mathcal{X})$.

2 Transition probability matrix

Definition 2.1. We denote the set of all probability mass functions over a countable state space $\mathcal{X}$ by $\mathcal{M}(\mathcal{X}) \triangleq \left\{\nu \in [0,1]^{\mathcal{X}} : \sum_{x \in \mathcal{X}} \nu_x = 1\right\}$.

Definition 2.2. The **transition probability matrix** at time n is denoted by $P(n) \in [0,1]^{\mathcal{X} \times \mathcal{X}}$, such that $P_{xy}(n) = p_{xy}(n)$ is the **transition 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, denoted by $p_{xy}(n) \triangleq P(\{X_{n+1} = y\} \mid \{X_n = x\})$.

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

Example 2.3 (Random Walk). A random walk $S : \Omega \to \mathcal{X}^{\mathbb{N}}$ with independent step-size sequence $X : \Omega \to \mathcal{X}^{\mathbb{N}}$, is a Markov sequence for countable state space $\mathcal{X}$. 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

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

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

Definition 2.4. 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 $A_x \triangleq (a_{xy} : y \in \mathcal{X})$ of the stochastic matrix $A \in \mathbb{R}_+^{\mathcal{X}\times\mathcal{X}}$ belongs to $\mathcal{M}(\mathcal{X})$.

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

3 Homogeneous Markov chains

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 3.1. A discrete time Markov chain with the probability transition matrix $P(n)$ that is independent of the index, is called **time homogeneous**.

Example 3.2 (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 transition operator $P \in [0,1]^{\mathbb{Z}\times\mathbb{Z}}$ is given by the entries $p_{xy} = p1_{\{y=x+1\}} + (1-p)1_{\{y=x-1\}}$ for all $x, y \in \mathbb{Z}$.

Example 3.3 (Sequence of experiments). Consider a random sequence of experiment outcomes $X : \Omega \to \{0,1\}_+^{\mathbb{Z}}$, such that $P(\{X_{n+1} = 0\} \mid \{X_n = 0\}) = 1 - q$ and $P(\{X_{n+1} = 1\} \mid \{X_n = 1\}) = 1 - p$ for all $n \in \mathbb{Z}_+$. Then, we can write the probability transition matrix as

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

Definition 3.4. Consider a time 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 initial state $\{X_0 = x\}$, by

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

3.1 Transition graph

A time homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{N}}$ with a probability transition matrix P, is sometimes represented by a directed weighted graph $G = (\mathcal{X}, E, w)$, where the set of nodes in the graph G is the state space

$\mathcal{X}$, and the set of directed edges is the set of possible one-step transitions indicated by the initial and the final state, as

$$E \triangleq \{[x,y\rangle \in \mathcal{X} \times \mathcal{X} : p_{xy} > 0\}.$$

In addition, this graph has a weight $w_e = p_{xy}$ on each edge $e = [x,y\rangle \in E$.

Example 3.5 (Integer random walk). The time homogeneous Markov chain in Example 3.2 can be represented by an infinite state weighted graph $G = (\mathbb{Z}, E, w)$, where the edge set is

$$E = \{(n,n+1) : n \in \mathbb{Z}\} \cup \{(n,n-1) : n \in \mathbb{Z}\}.$$

We have plotted the sub-graph of the entire transition graph for states $\{-1,0,1\}$ in Figure 1.

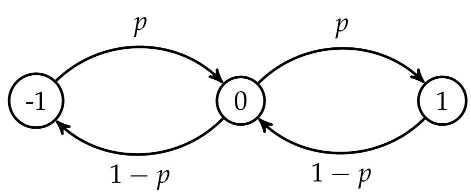

Figure 1: Sub-graph of the entire transition graph for an integer random walk with *i.i.d.* step-sizes in $\{-1,1\}$ with probability p for the positive step.

Example 3.6 (Sequence of experiments). The time homogeneous Markov chain in Example 3.3 can be represented by the following two-state weighted transition graph $G = (\{0,1\}, E, w)$, plotted in Figure 2.

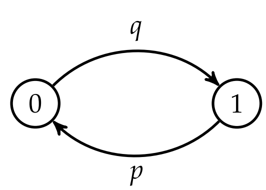

Figure 2: Markov chain for the sequence of experiments with two outcomes.

4 n-step transition

Definition 4.1. For a time homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ we denote the probability mass function of Markov chain at step n by $\pi_n \in \mathcal{M}(\mathcal{X})$.

Proposition 4.2. *Conditioned on the initial state, any finite dimensional distribution of a homogeneous Markov chain is stationary. That is, for any finite $n, m \in \mathbb{Z}_+$ and states $x_0, \dots, x_n \in \mathcal{X}$, 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 4.3. *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 $\{H_{n-1}(x) \cap \{X = x_n\}) : x \in \mathcal{X}^{n-1}\}$ defined by $H_{n-1}(x) \triangleq \cap_{i=1}^{n-1} \{X_i = x_i\}$, and partition the event $\{X_{m+n} = x_n\}$ in terms of the disjoint events $\{F_{n-1}(x) \cap \{X_{m+n} = x_n\}) : x \in \mathcal{X}^{n-1}\}$ defined by $F_{n-1}(x) \triangleq \cap_{i=1}^{n-1} \{X_{m+i} = x_i\}$. Then, we can write

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

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

$$P(F_{n-1}(x) \cap \{X_{m+n} = x_n\} \mid \{X_m = x_0\}) = P(H_{n-1}(x) \cap \{X_n = x_n\} \mid \{X_0 = x_0\}).$$

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

$$P_{x_0}\{X_n = x_n\} = \sum_{x \in \mathcal{X}^{n-1}} P_{x_0}(H_{n-1}(x) \cap \{X_n = x_n\}),$$
$$P(\{X_{m+n} = x_n\} \,|\, \{X_m = x_0\}) = \sum_{x \in \mathcal{X}^{n-1}} P(F_{n-1}(x) \cap \{X_{m+n} = x_n\} \,|\, \{X_m = x_0\}).$$

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

Definition 4.4. For a time homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, we can define ***n*-step transition probability matrix** $P^{(n)}$, with its (x,y) entry being the ***n*-step transition probability** for X_{m+n} to be in state y given the event $\{X_m = x\}$. That is, $p_{xy}^{(n)} \triangleq P(\{X_{n+m} = y\} \,|\, \{X_m = x\})$ for all $x, y \in \mathcal{X}$ and $m, n \in \mathbb{Z}_+$.

Remark 6. That is, the row $P_x^{(n)} = (p_{xy}^{(n)} : y \in \mathcal{X}) \in \mathcal{M}(\mathcal{X})$ is the conditional distribution of X_n given the initial state $\{X_0 = x\}$.

Example 4.5 (Sequence of experiments). Consider the time homogeneous Markov chain $X : \Omega \to \mathbb{Z}^{\mathbb{Z}_+}$ introduced in Example 3.3. 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.

Theorem 4.6. *The n-step transition probabilities for a homogeneous Markov chain form a semi-group. That is, for all positive integers $m, n \in \mathbb{Z}_+$*

$$P^{(m+n)} = P^{(m)} P^{(n)}.$$

Proof. The events $\{\{X_m = z\} : z \in \mathcal{X}\}$ partition the sample space Ω, and hence we can express the event $\{X_{m+n} = y\}$ as the following disjoint union

$$\{X_{m+n} = y\} = \cup_{z \in \mathcal{X}} \{X_{m+n} = y, X_m = z\}.$$

It follows from the Markov property and law of total probability that for any states x, y and positive integers m, n

$$p_{xy}^{(m+n)} = \sum_{z \in \mathcal{X}} P_x(\{X_{n+m} = y, X_m = z\}) = \sum_{z \in \mathcal{X}} P(\{X_{n+m} = y \mid X_m = z, X_0 = x\}) P_x(\{X_m = z\})$$
$$= \sum_{z \in \mathcal{X}} P(\{X_{n+m} = y \mid X_m = z\}) P_x(\{X_m = z\}) = \sum_{z \in \mathcal{X}} p_{xz}^{(m)} p_{zy}^{(n)} = (P^{(m)} P^{(n)})_{xy}.$$

Since the choice of states $x, y \in \mathcal{X}$ were arbitrary, the result follows. □

Corollary 4.7. *The n-step transition probability matrix is given by $P^{(n)} = P^n$ for any positive integer n.*

Proof. In particular, we have $P^{(n+1)} = P^{(n)} P^{(1)} = P^{(1)} P^{(n)}$. Since $P^{(1)} = P$, we have $P^{(n)} = P^n$ by induction. □

Remark 7. That is, for all states x, y and non-negative integers $n \in \mathbb{Z}_+$, $p_{xy}^{(n)} = P_{xy}^n$.

5 Chapman Kolmogorov equations

We denote by $\pi_0 \in \mathbb{R}_+^{\mathcal{X}}$ the initial distribution of the Markov chain, that is $\pi_0(x) = P\{X_0 = x\}$. The distribution of X_n is given by $\pi_n \in \mathbb{R}_+^{\mathcal{X}}$, such that for any state $x \in \mathcal{X}$.

$$\pi_n(x) = P\{X_n = x\} = \sum_{z \in \mathcal{X}} p_{zx}^{(n)} \pi_0(z) = (\pi_0 P^n)_x.$$

We can write this succinctly in terms of transition probability matrix P as $\mu_n = \mu_0 P^n$. We can alternatively derive this result by the following Lemma.

Lemma 5.1. *The right multiplication of a probability vector with the transition matrix P transforms the probability distribution of current state to probability distribution of the next state. That is,*

$$\pi_{n+1} = \pi_n P, \ \text{ for all } n \in \mathbb{N}.$$

Proof. To see this, we fix $y \in \mathcal{X}$ and from the law of total probability and the definition conditional probability, we observe that

$$\pi_{n+1}(y) = P\{X_{n+1} = y\} = \sum_{x \in \mathcal{X}} P\{X_{n+1} = y, X_n = x\} = \sum_{x \in \mathcal{X}} P\{X_n = x\} p_{xy} = (\pi_n P)_y.$$

□