Lecture-19: Discrete Time Markov Chains

1 Markov processes

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

Definition 1.1. A stochastic process $X : \Omega \to \mathcal{X}^T$ with state space $\mathcal{X}$ and ordered index set T is **Markov** if conditioned on the present state X_t, future $\sigma(X_u, u > t)$ is independent of the past $\sigma(X_s, s < t)$. We denote the history of the process until time t as $\mathcal{F}_t \triangleq \sigma(X_s, s \leqslant t)$. That is, for any Borel measurable set $B \in \mathcal{B}(\mathcal{X})$ any two indices $u > t$, we have

$$P(\{X_u \in B\} \mid \mathcal{F}_t) = P(\{X_u \in B\} \mid \sigma(X_t)).$$

Remark 1. We next re-write the Markov property more explicitly for the process $X : \Omega \to \mathbb{R}^{\mathbb{R}}$. For all $x, y \in \mathcal{X}$, finite set $S \subseteq \mathbb{R}$ such that $\max S < t < u$, and $H_S(x_S) \triangleq \cap_{s \in S} \{X_s \leqslant x_s\} \in \mathcal{F}_t$, we have

$$P(\{X_u \leqslant y\} \mid H_S(x_S) \cap \{X_t \leqslant x\}) = P(\{X_u \leqslant y\} \mid \{X_t \leqslant x\}).$$

1.1 Discrete time Markov chains

Definition 1.2. 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 $\mathcal{F}_n \triangleq \sigma(X_1, \dots, X_n)$.

Remark 2. Recall that the event space $\mathcal{F}_n$ is generated by the historical events $A_X(x) \triangleq \cap_{i=0}^n \{X_i \leqslant x_i\}$ where $x \in \mathbb{R}^{n+1}$.

Remark 3. When the state space $\mathcal{X}$ is countable, the event space $\mathcal{F}_n$ is generated by the historical events $H_n(x) \triangleq \cap_{i=0}^n \{X_i = x_i\}$, where $x \in \mathcal{X}^{n+1}$. That is, $\mathcal{F}_n = \sigma(H_n(x) : x \in \mathcal{X}^{n+1})$

Definition 1.3. 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 4. 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 X, since $\mathcal{F}_n = \sigma(H_n(x) : x \in \mathcal{X}^n)$ and $\sigma(X_n) = \sigma(\{X_n = x\}, x \in \mathcal{X})$.

Example 1.4 (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 Markov for a countable state space $\mathcal{X}$ that is closed under addition. Given a 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 $\sigma(S_0, X_1, \dots, X_n) = \sigma(S_0, S_1, \dots, S_n)$.

1.2 Transition probability matrix

Definition 1.5. 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 1.6. The **transition probability matrix** at time n is denoted by $P(n) \in [0,1]^{\mathcal{X}\times\mathcal{X}}$, such that its (x,y)th entry is denoted by $p_{xy}(n) \triangleq P(\{X_{n+1} = y\} \mid \{X_n = x\})$, that is the **transition probability** of a discrete time Markov chain X from a state $x \in \mathcal{X}$ at time n to state $y \in \mathcal{X}$ at time $n+1$.

Remark 5. We observe that each row $P_x(n) \triangleq (p_{xy}(n) : y \in \mathcal{X}) \in \mathcal{M}(\mathcal{X})$ is the conditional distribution of X_{n+1} given the event $\{X_n = x\}$.

Definition 1.7. 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 6. 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 $\sum_{y\in\mathcal{X}} f(y)p_{xy}(n) = \mathbb{E}[f(X_{n+1}) \mid \{X_n = x\}]$.

1.3 Homogeneous Markov chains

In general, not much can be said about Markov chains with index dependent transition probabilities. We consider the simpler case where the transition probabilities $p_{xy}(n) = p_{xy}$ are independent of the index.

Definition 1.8. A discrete time Markov chain with the probability transition matrix $P(n)$ that is independent of the index, is called **time homogeneous**.

Example 1.9 (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 1.10 (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 1.11. 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

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

Proposition 1.12. *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. Consider a homogeneous Markov chain $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with natural filtration $\mathcal{F}_\bullet$ such that $\mathcal{F}_n \triangleq \sigma(X_0, \dots, X_n)$ for each $n \in \mathbb{N}$. Using the property of conditional probabilities and Markovity of X, we can write the conditional probability of sample path $(X_1, \dots, X_n)$ given the event $\{X_0 = x_0\}$ as

$$P\Big(\cap_{i=1}^n \{X_i = x_i\} \mid \{X_0 = x_0\} \Big) = \prod_{i=1}^n P\Big(\{X_i = x_i\} \mid \cap_{j=0}^{i-1} \{X_j = x_j\} \Big) = \prod_{i=1}^n P\Big(\{X_i = x_i\} \mid \{X_{i-1} = x_{i-1}\} \Big).$$

Using the property of conditional probabilities and Markovity of X, we can write the conditional probability of sample path $(X_{m+1}, \dots, X_{m+n})$ given the event $\{X_m = x_0\}$ as

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

From time-homogeneity of transition probabilities of Markov chain X, it follows that both the transition probabilities are identical and equal to $\prod_{i=1}^n p_{x_{i-1},x_i}$. ☐

Corollary 1.13. *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 $P(\{X_{n+m} = x_n\} \mid \{X_m = x_0\}) = P(\{X_n = x_n\} \mid \{X_0 = x_0\})$.*

Proof. It follows from summing over all possible paths $(X_0, \dots, X_n)$ and $(X_m, \dots, X_{m+n})$. In particular, we can partition events $\{X_n = x_n\}$ and $\{X_{m+n} = x_n\}$ in terms of unions over disjoint paths

$$\{X_n = x_n\} = \bigcup_{x \in \mathcal{X}^{n-1}} \Big(\cap_{i=1}^n \{X_i = x_i\} \Big), \qquad \{X_{m+n} = x_n\} = \bigcup_{x \in \mathcal{X}^{n-1}} \Big(\cap_{i=1}^n \{X_{m+i} = x_i\} \Big).$$

The result follows from the countable additivity of conditional probability for disjoint events of taking distinct paths, and the fact that probability of taking same path is identical for both sums. ☐

1.4 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 1.14 (Integer random walk). The time homogeneous Markov chain in Example 1.9 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.

Example 1.15 (Sequence of experiments). The time homogeneous Markov chain in Example 1.10 can be represented by the following two-state weighted transition graph $G = (\{0, 1\}, E, w)$, plotted in Figure 2.

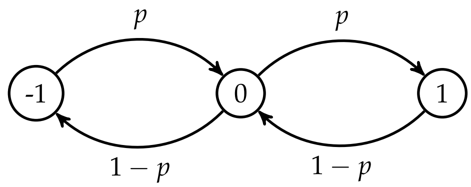

Figure 1: Sub-graph of the entire transition graph for an integer random walk with *i.i.d.* step-sizes in $\{-1,1\}$ with probability p for the positive step.

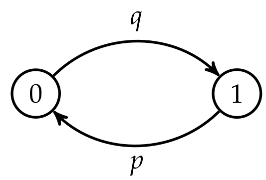

Figure 2: Markov chain for the sequence of experiments with two outcomes.

1.5 Random mapping theorem

We saw some example of Markov processes where $X_n = X_{n-1} + Z_n$, and $Z : \Omega \to \Lambda^{\mathbb{N}}$ is an *i.i.d.* sequence, independent of the initial state X_0. We will show that any discrete time Markov chain is of this form, where the sum is replaced by arbitrary functions.

Theorem 1.16 (Random mapping theorem). *For any DTMC $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, there exists an* i.i.d. *sequence $Z \in \Lambda^{\mathbb{N}}$ and functions $f_n : \mathcal{X} \times \Lambda \to \mathcal{X}$ such that $X_n = f_n(X_{n-1}, Z_n)$ for all $n \in \mathbb{N}$.*

Remark 7. A **random mapping representation** of a transition matrix $P(n)$ on state space $\mathcal{X}$ is a function $f_n : \mathcal{X} \times \Lambda \to \mathcal{X}$, along with a random variable $Z_n : \Omega \to \Lambda$, satisfying for all $x, y \in \mathcal{X}$,

$$P\{f_n(x, Z_n) = y\} = p_{xy}(n).$$

Proof. It suffices to show that every transition matrix $P(n)$ has a random mapping representation. Then, for the mapping f_n and the *i.i.d.* sequence $Z : \Omega \to \Lambda^{\mathbb{N}}$, we would have $X_n = f_n(X_{n-1}, Z_n)$ for all $n \in \mathbb{N}$.

Let $\Lambda \triangleq [0,1]$, and we choose the *i.i.d.* uniform sequence $Z : \Omega \to \Lambda^{\mathbb{N}}$. Since $\mathcal{X}$ is countable, it can be ordered. We let $\mathcal{X} = \mathbb{N}$ without any loss of generality. We set $F_{xy}(n) \triangleq \sum_{w \leqslant y} p_{xw}(n)$ and define function $f_n : \mathcal{X} \times \Lambda \to \mathcal{X}$ for all pairs $(x, z) \in \mathcal{X} \times \Lambda$ by

$$f_n(x,z) \triangleq \sum_{y \in \mathbb{N}} y \mathbb{1}_{\left\{F_{x,y-1}(n) < z \leqslant F_{x,y}(n)\right\}} = \inf\left\{y \in \mathcal{X} : z \leqslant F_{x,y}(n)\right\}.$$

Since $f_n(x, Z_n)$ is a discrete random variable taking value $y \in \mathcal{X}$, iff the uniform random variable Z_n lies in the interval $(F_{x,y-1}(n), F_{x,y}(n)]$. That is, the event $\{f_n(x, Z_n) = y\} = \left\{Z_n \in (F_{x,y-1}(n), F_{x,y}(n)]\right\}$ for all $y \in \mathcal{X}$. It follows that

$$P\{f_n(x, Z_n) = y\} = P\left\{F_{x,y-1}(n) < Z_n \leqslant F_{x,y}(n)\right\} = F_{x,y}(n) - F_{x,y-1}(n) = p_{xy}(n).$$

□