Lecture-20: DTMC: Representation

1 n-step transition

Definition 1.1. For a homogeneous Markov chain $X : \Omega to \mathcal{X}^{\mathbb{Z}_+}$, we can define n-step transition probabilities for $x, y \in \mathcal{X}$ and $m, n \in \mathbb{N}$

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

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

Theorem 1.2. *The n-step transition probabilities form a semi-group. That is, for all positive integers m, n*

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

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

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

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

Corollary 1.3. *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 1. That is, for all states x, y and non-negative integers $n \in \mathbb{Z}_+$, $p_{xy}^{(n)} = P_{xy}^n$.

2 Representation

2.1 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 2.1. *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

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

□

2.2 Transition graph

We can define a collection E of possible one-step transitions indicated by the initial and the final state, as

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

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

Example 2.2 (Integer random walk). For an integer random walk $X = (X_n \in \mathbb{Z} : n \in \mathbb{N})$ with *i.i.d.* step-size sequence $Z = (Z_n \in \{-1,1\}, n \in \mathbb{N})$, we have and infinite graph $G = (\mathbb{Z}, E)$, 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.

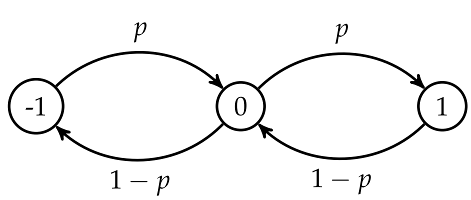

Figure 1: Sub-graph of the entire transition graph for an integer random walk with *i.i.d.* step-sizes in $\{-1,1\}$ with probability p for the positive step.

Example 2.3 (Sequence of experiments). Consider the sequence of experiments with the set of outcomes $\mathcal{X} = \{0,1\}$ with the transition matrix

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

We have plotted the corresponding transition graph for this two-state Markov chain in Figure 2.

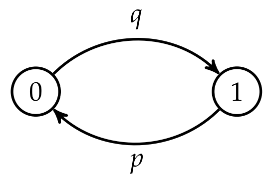

Figure 2: Markov chain for the sequence of experiments with two outcomes.

2.3 Random Mapping Theorem

We saw some example of Markov processes where $X_n = X_{n-1} + Z_n$, and $(Z_n : n \in \mathbb{N})$ is an iid 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 2.4 (Random mapping theorem). *For any DTMC X, there exists an* i.i.d. *sequence $Z \in \Lambda^{\mathbb{N}}$ and a function $f : \mathcal{X} \times \Lambda \to \mathcal{X}$ such that $X_n = f(X_{n-1}, Z_n)$ for all $n \in \mathbb{N}$.*

Remark 2. A **random mapping representation** of a transition matrix P on state space $\mathcal{X}$ is a function $f : \mathcal{X} \times \Lambda \to \mathcal{X}$, along with a Λ-valued random variable Y, satisfying

$$P\{f(x,Y) = y\} = p_{xy}, \text{ for all } x, y \in \mathcal{X}.$$

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

Let $\Lambda = [0,1]$, and we choose the *i.i.d.* sequence Z, uniformly at random from this interval. Since $\mathcal{X}$ is countable, it can be ordered. We let $\mathcal{X} = \mathbb{N}$ without any loss of generality. We set $F_{xy} \triangleq \sum_{w \leqslant y} p_{xw}$ and define

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

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

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

□

3 Strong Markov property (SMP)

We are interested in generalizing the Markov property to any random times. For a DTMC $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$, let $T : \Omega \to \mathbb{N}$ be an integer random variable, and we are interested in knowing whether for any historical event $H_{T-1} = \cap_{n=0}^{T-1} \{X_n = x_n\}$ and any state $x, y \in \mathcal{X}$, we have

$$P(\{X_{T+1} = y\} \mid H_{T-1} \cap \{X_T = x\}) = p_{xy}.$$

Example 3.1 (Two-state DTMC). For the two state Markov chain $X \in \{0,1\}^{\mathbb{Z}_+}$ such that $P_0\{X_1 = 1\} = q$ and $P_1\{X_1 = 0\} = p$ for $p, q \in [0,1]$. Let $T : \Omega \to \mathbb{N}$ be an integer random variable defined as

$$T \triangleq \sup\{n \in \mathbb{N} : X_i = 0, \text{ for all } i \leqslant n\}.$$

That is, $\{T = n\} = \{X_1 = 0, \ldots, X_n = 0, X_{n+1} = 1\}$. Hence, for the historical event $H_{T-1} = \{X_1 = \ldots, X_{T-1} = 0\}$, the conditional probability $P(\{X_{T+1} = 1\} \mid H_{T-1} \cap \{X_T = 0\}) = 1$, and not equal to q.

Definition 3.2. Let T be an integer valued stopping time with respect to a random sequence X. Then for all states $x, y \in \mathcal{X}$ and the event $H_{T-1} = \cap_{n=0}^{T-1} \{X_n = x_n\}$, the process X satisfies the **strong Markov property** if

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

Lemma 3.3. *Homogeneous Markov chains satisfy the strong Markov property.*

Proof. Let $X \in \mathcal{X}^{\mathbb{Z}_+}$ be a homogeneous DTMC with transition matrix P. We take any historical event $H_{T-1} = \cap_{n=0}^{T-1} \{X_n = x_n\}$, and $x, y \in \mathcal{X}$. Then, from the definition of conditional probability, the law of total probability, and the Markovity of the process X, we have

$$\begin{aligned}
&P(\{X_{T+1} = y\} \mid H_{T-1} \cap \{X_T = x\}) = \frac{\sum_{n \in \mathbb{Z}_+} P(\{X_{T+1} = y, X_T = x\} \cap H_{T-1} \cap \{T = n\})}{P(\{X_T = x\} \cap H_{T-1})} \\
&= \sum_{n \in \mathbb{Z}_+} P(\{X_{n+1} = y\} \mid \{X_n = x\} \cap H_{n-1} \cap \{T = n\}) P(\{T = n\} \,|\, \{X_T = x\} \cap H_{T-1}) \\
&= p_{xy} \sum_{n \in \mathbb{Z}_+} P(\{T = n\} \,|\, \{X_T = x\} \cap H_{T-1}) = p_{xy}.
\end{aligned}$$

This equality follows from the fact that the event $\{T = n\}$ is completely determined by $(X_0, \dots, X_n)$. ☐

Example 3.4 (For a non stopping time T). As an exercise, if we try to use the Markov property on arbitrary random variable T, the SMP may not hold. For example, define a non-stopping time $T \triangleq \inf\{n \in \mathbb{Z}_+ : X_{n+1} = y\}$ for $y \in \mathcal{X}$. In this case, we have

$$P(\{X_{T+1} = y\} \mid \{X_T = x, \dots, X_0 = x_0\}) = 1_{\{p_{xy} > 0\}} \neq P(\{X_1 = y\} \mid \{X_0 = x\}) = p_{xy}.$$

Remark 3. A useful application of the strong Markov property is as follows. Let $x_0 \in \mathcal{X}$ be a fixed state and $\tau_0 = 0$. Let τ_n denote the stopping times at which the Markov chain visits x_0 for the nth time. That is,

$$\tau_n \triangleq \inf\{n > \tau_{n-1} : X_n = x_0\}.$$

Then $(X_{\tau_n + m} \in \mathcal{X}^\Omega : m \in \mathbb{Z}_+)$ is a stochastic replica of $X : \Omega \to \mathcal{X}^{\mathbb{Z}_+}$ with $X_0 = x_0$.