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. For a countable set $\mathcal{X}$, a discrete-valued random sequence $(X_n \in \mathcal{X} : n \in \mathbb{Z}_+)$ 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 $H_{n-1} = \cap_{m=0}^{n-1} \{X_m = x_m\}$, 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 $j \in \mathcal{X}$ at time $n+1$ from a state $i \in \mathcal{X}$ 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 $P_{xy}(n) = p_{xy}(n)$. We observe that each row $P_x(n) = (p_{xy}(n) : y \in \mathcal{X})$ is the conditional distribution of X_{n+1} given $X_n = x$.

Example 1.1 (Integer random walk). Let $Z = (Z_n \in \mathbb{Z} : n \in \mathbb{N})$ be an independent (not necessarily identical) Bernoulli sequence. Let $X_0 = 0$ and $X_n \triangleq \sum_{i=1}^{n} Z_i$, then the process $X = (X_n \in \mathbb{Z} : n \in \mathbb{Z}_+)$ is called an **integer random walk**.

Theorem 1.2. *For a random walk $(S_n : n \in \mathbb{N})$ with iid step-size sequence X, the following are true.*

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

ii_ *Random walk is non-stationary with independent and stationary increments.*

iii_ *Random walk is a discrete time Markov chain.*

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 1. We make the following observations for the stochastic matrices.

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

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

iii_ 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 $x \in \mathcal{X}$

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

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

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 $\mathbf{1}^T$.

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 $(X_n : n \in \mathbb{N})$ with iid unit step size sequence $(Z_n \in \{-1,1\} : n \in \mathbb{N})$ 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

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

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 $X_0 = S$ by μ_{n+1}. That is,

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

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

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

This method of direct computation can quickly become too cumbersome.

Proposition 2.3. *Conditioned on the initial state, any finite dimensional distribution of a homogeneous Markov chain is stationary.*

Proof. To this end, we compute the transition probabilities for the path $(x_1, \dots, x_n)$ taken by the sample path $(X_1, \dots, X_n)$ when $X_0 = x_0$ and by the sample path $(X_{m+1}, \dots, X_{m+n})$ when $X_m = x_0$. From the homogeneous Markov property of the process X, we get

$$P(\{X_1 = x_1, \dots, X_n = x_n\} \mid \{X_0 = x_0\}) = \prod_{i=1}^{n} P(\{X_i = x_i)\} \mid \{X_{i-1} = x_{i-1}\}) = p_{x_0 x_1} p_{x_1 x_2} \dots p_{x_{n-1} x_n}.$$

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

$$P(\{X_{m+1} = x_1, \dots, X_{m+n} = x_n\} \mid \{X_m = x_0\}) = \prod_{i=1}^{n} P(\{X_{m+i} = x_i)\} \mid \{X_{m+i-1} = x_{i-1}\}) = p_{x_0 x_1} p_{x_1 x_2} \dots p_{x_{n-1} x_n}.$$

Corollary 2.4. *The n-step transition probabilities are stationary for any homogeneous Markov chain. That is, for any states $x, y \in \mathcal{X}$ and $n, m \in \mathbb{N}$, we have*

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

Proof. It follows from summing over intermediate steps. Hence, it follows that 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 $X_0 = x$.

Theorem 2.5. *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. It follows from the Markov property and law of total probability that for any states x, y and positive integers m, n

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

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

Corollary 2.6. *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 $P^{(n)} = P$ by induction. That is, for all states x, y and integers n,

$$p_{xy}^{(n)} = P_{xy}^{n}.$$