Contents

1

Probability

2 Markov Chains

2.1 Markov Chains

Definition 2.1.1 ▸ Stochastic Process

A **stochastic process** is a collection of random variables $\{X(t) : t \in T\}$ where T is an **index set** and $X(t)$ is known as the **current state**. The set of all possible states is known as the **state space**.

Let Ω be a sample space, a stochastic process defined over the space can be thought of a sequence of random variables where $X(t)$ describes the distribution of an outcome $\omega \in \Omega$ at timestamp t. The state space S is simply the co-domain of the $X(t)$'s.

A stochastic process is said to be

- *discrete-time* if the index set is countable;
- *continuous-time* if the index set is a continuum;
- *discrete-state* if the state space is countable;
- *finite-state* if the state space is finite;
- *continuous-state* if the state space is a continuum.

The term "continuum" refers to a **non-empty compact connected metric space**.

In this course, we focus on discrete-time discrete-state stochastic processes. One important property we will discuss now is the *Markovian property*.

Definition 2.1.2 ▸ Markovian Property

Let $\{X_n : n \in T\}$ be a discrete-time stochastic process over some probability space $(\Omega, \mathcal{F}, P)$. The stochastic process is called **Markovian** if

$$P\big(X_{n+1} = x_{n+1} \mid X_0^n = (x_0, x_1, \cdots, x_n)\big) = P\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right)$$

for all $n \in \mathbb{N}$.

The Markovian property essentially says that, given X_n, what has happened before, i.e., X_k for all $k < n$, is independent of what happens afterwards, i.e., X_{n+m} for $m \in \mathbb{N}^+$.

As the name suggests, the Markovian property is closely related to the Markov chains, which can be defined rigorously as follows:

Definition 2.1.3 ▸ Discrete-Time Markov Chain

A **Markov chain** is a discrete-time discrete-state stochastic process satisfying the Markovian property.

Recall that the *Bayes's Theorem* states the following:

Theorem 2.1.4 ▸ Bayes's Theorem

For any random variables X and Y,

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x)\, p_X(x)}{\sum_{x' \in \mathcal{X}} p_{Y|X}(y \mid x')\, p_X(x')}.$$

Theorem 2.1.4 leads to the following important result:

Corollary 2.1.5 ▸ Bayes' Rule for Markov Chains

Let $X_1, X_2, \cdots, X_n$ be any n random variables forming a Markov chain, then

$$p_{X_1^n}(x_1, x_2, \cdots, x_n) = p_{X_1}(x_1) \prod_{i=1}^{n-1} p_{X_{i+1}|X_i}(x_{i+1} \mid x_i).$$

Proof. If $n = 2$, by Theorem 2.1.4, we know that

$$\begin{aligned} p_{X_1,X_2}(x_1, x_2) &= p_{X_1|X_2}(x_1 \mid x_2)\, p_{X_2}(x_2) \\ &= p_{X_1}(x_1)\, p_{X_2|X_1}(x_2 \mid x_1). \end{aligned}$$

Suppose that there exists some integer $k \geq 2$ such that

$$p_{X_1^k}(x_1, x_2, \cdots, x_k) = p_{X_1}(x_1) \prod_{i=1}^{k-1} p_{X_{i+1}|X_i}(x_{i+1} \mid x_i)$$

For any k random variables X_1^k forming a Markov chain. Let X_{k+1} be any random variable such that X_1^{k+1} forms a Markov chain, then

$$p_{X_{k+1}|X_1^k}(x_{k+1} \mid x_1, x_2, \cdots, x_k) = p_{X_{k+1}|X_k}(x_{k+1} \mid x_k).$$

By using Theorem 2.1.4, we have

$$\begin{aligned} p_{X_1^{k+1}}(x_1, x_2, \cdots, x_{k+1}) &= p_{X_{k+1}|X_1^k}(x_{k+1} \mid x_1, x_2, \cdots, x_k)\, p_{X_1^k}(x_1, x_2, \cdots, x_k) \\ &= p_{X_{k+1}|X_k}(x_{k+1} \mid x_k)\, p_{X_1}(x_1) \prod_{i=1}^{k-1} p_{X_{i+1}|X_i}(x_{i+1} \mid x_i) \\ &= p_{X_1}(x_1) \prod_{i=1}^{k} p_{X_{i+1}|X_i}(x_{i+1} \mid x_i). \end{aligned}$$

□

Consider a discrete-time discrete-state stochastic process $\{X_n : n \in T\}$ with state space S over some probability space $(\Omega, \mathcal{F}, P)$. Here, the σ-algebra $\mathcal{F}$ can be generated using simple events $\{\omega \in \Omega : X_n(\omega) = s\}$ for all $n \in T$ and $s \in S$. Notice that this means that we need to find the joint distribution

$$p_{X_{n_1}^{n_k}}(s_1, s_2, \cdots, s_k)$$

for any tuple of random variables $X_{n_1}^{n_k}$ in the stochastic process, where $k \in \mathbb{N}$ and $k \leq |T|$ if T is finite, and any $(s_1, s_2, \cdots, s_k) \in S^k$. In general, this joint distribution might be hard to find, but things become easier if the stochastic process is a Markov chain because by Corollary 2.1.5 we have

$$p_{X_{n_1}^{n_k}}(s_1, s_2, \cdots, s_k) = p_{X_{n_1}}(s_1) \prod_{i=1}^{k-1} p_{X_{n_{i+1}}|X_{n_i}}(s_{i+1} \mid s_i).$$

If we can find $p_{X_m|X_n}(s_m \mid s_n)$ for any $m > n$, we could simplify this expression further!

Definition 2.1.6 ▸ Transition Probability

The **transition probability** is defined as

$$p_{ij}^{n,m} := P(X_m = j \mid X_n = i).$$

In particular, $p_{ij}^{n,n+1}$ is known as the **one-step transition probability** or **jump probability**.

Take some $k \in \mathbb{N}^+$ and consider

$$p_{ij}^{n,n+k} = P(X_{n+k} = j \mid X_n = i).$$

We first marginalise $P(X_{n+k} = j \mid X_n = i)$ with respect to X_{n+1} to obtain

$$P(X_{n+k} = j \mid X_n = i) = \sum_{s \in S} P(X_{n+k} = j \mid X_n = i, X_{n+1} = s) P(X_{n+1} = s \mid X_n = i).$$

Since X_n, X_{n+1} and X_{n+k} form a Markov chain, we have

$$P(X_{n+k} = j \mid X_n = i, X_{n+1} = s) = P(X_{n+k} = j \mid X_{n+1} = s).$$

Therefore,

$$\begin{aligned} p_{ij}^{n,n+k} &= P(X_{n+k} = j \mid X_n = i) \\ &= \sum_{s \in S} P(X_{n+k} = j \mid X_{n+1} = s) P(X_{n+1} = s \mid X_n = i) \\ &= \sum_{s \in S} p_{sj}^{n+1,n+k} p_{is}^{n,n+1}. \end{aligned}$$

Notice that now we have reduced the gap by 1. By repeatedly applying this process to $p_{sj}^{n+1,n+k}$, we eventually arrive at

$$p_{ij}^{n,n+k} = \sum_{s_1, s_2, \cdots, s_{k-1} \in S} p_{is_1}^{n,n+1} \left(\prod_{r=1}^{m-n-2} p_{s_r s_{r+1}}^{n+r,n+r+1} \right) p_{s_{m-1} j}^{n+k-1,n+k}$$

It is useful to see the one-step transition probability $p_{ij}^{n,n+1}$ as a function

$$f : T \times S \times S \to \mathbb{R}.$$

Thus far, we have basically shown that to specify a Markov chain fully, we will need to define the **index set** T, the **state space** S and the **one-step transition probabilities** $p_{ij}^{n,n+1}$ for all $i, j \in S$.