Part IB — Markov Chains

Based on lectures by Dr P. Sousi and notes by thirdsgames.co.uk

Michaelmas 2022

Contents

§0 History/motivation?

Markov chains are random processes (sequences of rvs) that retain no memory of the past. So the past is independent of the future

Motivation: extend the law of large numbers to the non iid setting

Kolmegorov in 1930 extended them to continuous time Markov processes.

Brownian motion: fundamental object in modern probability theory.

Why study MC? They are one of the simplest mathematical models for various random phenomena that evolve in time. They are simple as they do not rely on the past which makes them amenable to analysis, so we can use tools from probability, analysis, combinatorics.

Applications: population growth (branching processes), mathematical genetics, queuing networks, Monte carlo simulation, …

Page-Rank algorithm

Model the … as a directed graph, G: (V, E)

V: set of vertices

§1 Introduction

§1.1 Definitions

Let I be a finite or countable set. All of our random variables will be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

> **Definition 1.1** (Markov Chain)
>
> A stochastic process $(X_n)_{n \geq 0}$ is called a **Markov chain** if $\forall n \geq 0$ and for $x_1 \ldots x_{n+1} \in I$,
>
> $$\mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n, \ldots, X_1 = x_1\right) = \mathbb{P}\left(X_{n+1} = x_{n+1} \mid X_n = x_n\right)$$

We can think of n as a discrete measure of time. If $\mathbb{P}\left(X_{n+1} = y \mid X_n = x\right)$ for all x, y is independent of n, then X is called time-homogeneous. Otherwise, X is called time-inhomogeneous. In this course, we only study time-homogeneous Markov chains. If we consider time-homogeneous chains only, we may as well take $n = 0$ and we can write

$$P(x, y) = \mathbb{P}\left(X_1 = y \mid X_0 = x\right); \quad \forall x, y \in I$$

Definition 1.2 (Stochastic Matrix)

A **stochastic matrix** is a matrix where the sum of each row is equal to 1.

We call P the **transition matrix**. It is a stochastic matrix:

$$\sum_{y \in I} P(x, y) = 1$$

Remark 1. The index set does not need to be $\mathbb{N}$; it could alternatively be $\{0, 1, \dots, N\}$ for $N \in \mathbb{N}$.

We say that X is Markov (λ, P) if X_0 has distribution λ, and P is the transition matrix. Hence,

1. $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$

2. $\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0) = P(x_n, x_{n+1}) =: P_{x_n x_{n+1}}$

We usually draw a diagram of the transition matrix using a graph. Directed edges between nodes are labelled with their transition probabilities.

§1.2 Sequence definition

Theorem 1.1

The process X is Markov (λ, P) if and only if $\forall n \geq 0$ and all $x_0, \dots, x_n \in I$, we have

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \lambda_{x_0} P(x_0, x_1) P(x_1, x_2) \dots P(x_{n-1}, x_n)$$

Proof. If X is Markov, then we have

$$
\begin{aligned}
\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\
&\quad \cdot \mathbb{P}(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\
&= P(x_{n-1}, x_n) \mathbb{P}(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\
&= P(x_{n-1}, x_n) \dots P(x_0, x_1) \lambda_{x_0}
\end{aligned}
$$

as required. Conversely, $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$ satisfies (i). The transition matrix is given by

$$
\begin{aligned}
\mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) &= \frac{\lambda_{x_0} P(x_0, x_1) \dots P(x_{n-1}, x_n)}{\lambda_{x_0} P(x_0, x_1) \dots P(x_{n-2}, x_{n-1})} \\
&= P(x_{n-1}, x_n)
\end{aligned}
$$

> which is exactly the Markov property as required. □

§1.3 Point masses

> **Definition 1.3**
>
> For $i \in I$, the δ_i-mass at i is defined by
>
> $$\delta_{ij} = (i = j)$$
>
> This is a probability measure that has probability 1 at i only.

§1.4 Independence of sequences

Recall that discrete random variables (X_n) are considered independent if for all $x_1, \ldots, x_n \in I$, we have

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n)$$

A sequence (X_n) is independent if for all k, $i_1 < i_2 < \cdots < i_k$ and for all $x_1, \ldots, x_k$, we have

$$\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k) = \prod_{j=1}^{n} \mathbb{P}\left(X_{i_j} = x_j\right)$$

Let $X = (X_n), Y = (Y_n)$ be sequences of discrete random variables. They are independent if for all k, m, $i_1 < \cdots < i_k$, $j_1 < \cdots < j_m$,

$$\mathbb{P}(X_1 = x_1, \ldots, X_{i_k} = x_{i_k}, Y_{j_1} = y_{j_1}, \ldots, Y_{j_m}) = \mathbb{P}(X_1 = x_1, \ldots, X_{i_k} = x_{i_k})\,\mathbb{P}(Y_{j_1} = y_{j_1}, \ldots, Y_{j_m})$$

§1.5 Simple Markov property

> **Theorem 1.2**
>
> Suppose X is Markov (λ, P). Let $m \in \mathbb{N}$ and $i \in I$. Given that $X_m = i$, we have that the process after time m, written $(X_{m+n})_{n \geq 0}$, is Markov (δ_i, P), and it is independent of $X_0, \ldots, X_m$.

Informally, the past and the future are independent given the present.

> *Proof.* We must show that
>
> $$\mathbb{P}(X_m = x_0, \ldots, X_{m+n} = x_n \mid X_m = i) = \delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n)$$

We have

$$\mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m \mid X_m = i\right) = \frac{\mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m\right)\delta_{ix_m}}{\mathbb{P}\left(X_m = i\right)}$$

The numerator is

$$\begin{aligned}
&\mathbb{P}\left(X_{m+n}, \ldots, X_m = x_m\right) \\
&= \sum_{x_0, \ldots, x_{m-1} \in I} \mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m, X_{m-1} = x_{m-1}, \ldots, X_0 = x_0\right) \\
&= \sum_{x_0, \ldots, x_{m-1}} \lambda_{x_0} P(x_0, x_1) \ldots P(x_{m-1}, x_m) P(x_m, x_{m+1}) \ldots P(x_{m+n-1}, x_{m+n}) \\
&= P(x_m, x_{m+1}) \ldots P(x_{m+n-1}, x_{m+n}) \sum_{x_0, \ldots, x_{m-1}} \lambda_{x_0} P(x_0, x_1) \ldots P(x_{m-1}, x_m) \\
&= P(x_m, x_{m+1}) \ldots P(x_{m+n-1}, x_{m+n})\mathbb{P}\left(X_m = x_m\right)
\end{aligned}$$

Thus we have

$$\mathbb{P}\left(X_{m+n} = x_{m+n}, \ldots, X_m = x_m \mid X_m = i\right) = P(x_m, x_{m+1}) \ldots P(x_{m+n-1}, x_{m+n})\delta_{ix_m}$$

Hence $(X_{m+n})_{n \geq 0} \sim \text{Markov}\left(\delta_i, P\right)$ conditional on $X_m = i$. Now it suffices to show independence between the past and future variables. In particular, we need to show $m \leq i_1 < \cdots < i_k$ for some $k \in \mathbb{N}$ implies that

$$\begin{aligned}
&\mathbb{P}\left(X_{i_1} = x_{m+1}, \ldots, X_{i_k} = x_{m+k}, X_0 = x_0, \ldots, X_m = x_m \mid X_m = i\right) \\
&= \mathbb{P}\left(X_{i_1} = x_{m+1}, \ldots, X_{i_k} = x_{m+k} \mid X_m = i\right)\mathbb{P}\left(X_0 = x_0, \ldots, X_m = x_m \mid X_m = i\right)
\end{aligned}$$

So let $i = x_m$, and then

$$\begin{aligned}
&= \frac{\mathbb{P}\left(X_{i_1} = x_{m+1}, \ldots, X_{i_k} = x_{m+k}, X_0 = x_0, \ldots, X_m = x_m\right)}{\mathbb{P}\left(X_m = i\right)} \\
&= \frac{\lambda_{x_0} P(x_0, x_1) \ldots P(x_{m-1}, x_m)\mathbb{P}\left(X_{i_1} = x_{m+1}, \ldots, X_{i_k} = x_{m+k} \mid X_m = x_m\right)}{\mathbb{P}\left(x_m = i\right)} \\
&= \frac{\mathbb{P}\left(X_0 = x_0, \ldots, X_m = x_m\right)}{\mathbb{P}\left(X_m = x_m\right)}\mathbb{P}\left(X_{i_1} = x_{m+1}, \ldots, X_{i_k} = x_{m+k} \mid X_m = x_m\right)
\end{aligned}$$

which gives the result as required. $\qquad\square$

§1.6 Powers of the transition matrix

Suppose $X \sim \text{Markov}\left(\lambda, P\right)$ with values in I. If I is finite, then P is an $|I| \times |I|$ square matrix. In this case, we can label the states as $1, \ldots, |I|$. If I is infinite, then we label

the states using the natural numbers $\mathbb{N}$. Let $x \in I$ and $n \in \mathbb{N}$. Then,

$$\mathbb{P}\left(X_n = x\right) = \sum_{x_0, \ldots, x_{n-1} \in I} \mathbb{P}\left(X_n = x, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\right)$$

$$= \sum_{x_0, \ldots, x_{n-1} \in I} \lambda_{x_0} P(x_0, x_1) \ldots P(x_{n-1}, x)$$

We can think of λ as a row vector. So we can write this as

$$= (\lambda P^n)_x$$

By convention, we take $P^0 = I$, the identity matrix. Now, suppose $m, n \in \mathbb{N}$. By the simple Markov property,

$$\mathbb{P}\left(X_{m+n} = y \mid X_m = x\right) = \mathbb{P}\left(X_n = y \mid X_0 = x\right) = (\delta_x P^n)_y$$

We will write $\mathbb{P}_x\left(A\right) := \mathbb{P}\left(A \mid X_0 = x\right)$ as an abbreviation. Further, we write $p_{ij}(n)$ for the (i, j) element of P^n. We have therefore proven the following theorem.

> **Theorem 1.3**
>
> $$\mathbb{P}\left(X_n = x\right) = (\lambda P^n)_x;$$
>
> $$\mathbb{P}\left(X_{n+m} = y \mid X_m = x\right) = \mathbb{P}_x\left(X_n = y\right) = p_{xy}(n)$$

§1.7 Calculating powers

> **Example 1.1**
> Consider
>
> $$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}; \quad \alpha, \beta \in [0, 1]$$
>
> Note that for any stochastic matrix P, P^n is a stochastic matrix. First, we have $P^{n+1} = P^n P$. Let us begin by finding $p_{11}(n+1)$.
>
> $$p_{11}(n+1) = p_{11}(n)(1 - \alpha) + p_{12}(n)\beta$$
>
> Note that $p_{11}(n) + p_{12}(n) = 1$ since P^n is stochastic. Therefore,
>
> $$p_{11}(n+1) = p_{11}(n)(1 - \alpha - \beta) + \beta$$

We can solve this recursion relation to find

$$p_{11}(n) = \begin{cases} \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1 - \alpha - \beta)^n & \alpha + \beta > 0 \\ 1 & \alpha + \beta = 0 \end{cases}$$

The general procedure for finding P^n is as follows. Suppose that P is a $k \times k$ matrix. Then let $\lambda_1, \ldots, \lambda_k$ be its eigenvalues (which may not be all distinct).

1. All λ_i distinct. In this case, P is diagonalisable, and hence we can write $P = UDU^{-1}$ where U is a diagonal matrix, whose diagonal entries are the λ_i. Then, $P^n = UD^nU^{-1}$. Calculating D^n may be done termwise since D is diagonal. In this case, we have terms such as

$$p_{11}(n) = a_1\lambda_1^n + \cdots + a_k\lambda_k^n; \quad a_i \in \mathbb{R}$$

First, note $P^0 = I$ hence $p_{11}(0) = 1$. We can substitute small values of n and then solve the system of equations. Now, suppose λ_k is complex for some k. In this case, $\overline{\lambda_k}$ is also an eigenvalue. Then, up to reordering,

$$\lambda_k = re^{i\theta} = r(\cos\theta + i\sin\theta); \lambda_{k-1} = \overline{\lambda_k} = re^{i\theta} = r(\cos\theta - i\sin\theta)$$

We can instead write $p_{11}(n)$ as

$$p_{11}(n) = a_1\lambda_1^n + \cdots + a_{k-1}r^n\cos(n\theta) + a_kr^n\sin(n\theta)$$

Since $p_{11}(n)$ is real, all the imaginary parts disappear, so we can simply ignore them.

2. Not all λ_i distinct. In this case, λ appears with multiplicity 2, then we include also the term $(an + b)\lambda^n$ as well as $b\lambda^n$. This can be shown by considering the Jordan normal form of P.

Example 1.2

Let

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

The eigenvalues are $1, \frac{1}{2}i, -\frac{1}{2}i$. Then, writing $\frac{i}{2} = \frac{1}{2}(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2})$, we can write

$$p_{11}(n) = \alpha + \beta\left(\frac{1}{2}\right)^n \cos\frac{n\pi}{2} + \gamma\left(\frac{1}{2}\right)^n \sin\frac{n\pi}{2}$$

For $n = 0$ we have $p_{11}(0) = 1$, and for $n = 1$ we have $p_{11}(1) = 0$, and for $n = 2$ we can calculate P^2 and find $p_{11}(2) = 0$. Solving this system of equations for α, β, γ,

we can find

$$p_{11}(n) = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left(\frac{4}{5} \cos \frac{n\pi}{2} - \frac{2}{5} \sin \frac{n\pi}{2}\right)$$