Markov Chains

Notes made by Finley Cooper

19th October 2025

Contents

1 Markov Chains

1.1 The Markov property

Throughout all our random variables and random processes will be assumed to be defined on an appropiate underlying probablity space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition. (Markov chain) A discrete-time Markov chain is a sequence $\underline{X} = (X_n)_{n\geq 0}$ of random variables taking values in the same discrete countable state space I, such that:

$$\mathbb{P}\left(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n\right) = \mathbb{P}\left(X_{n+1} = x_{n+1} | X_n = x_n\right) \quad \forall n \geq 0.$$

If $\mathbb{P}\left(X_{n+1} = y | X_n = x\right)$ is indepedent of n for all x, y then we call $\underline{X}$ a time-homogeneous Markov chain. For this course all Markov chains are time-homogeneous with a countable state space.

Definition. (Transition matrix) We define the transition matrix P as the matrix

$$P(x, y) = P_{xy} = \mathbb{P}\left(X_{n+1} = y | X_n = x\right).$$

Note that P is a stochastic matrix i.e. $P_{xy} \geq 0$ for all x, y and the sum of each row is 1. For example take the simple Markov chain with $I = \{0, 1\}$ moving from 0 to 1 w.p. α and moving from 1 to 0 w.p. β, so

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

We say that $\underline{X} = (X_n)$ is a Markov chain with transition matrix P with initial distribution λ if $\lambda = (\lambda_n)$ is a distribution and I is such that $\mathbb{P}\left(X_0 = x\right) = \lambda_i$, for all $x \in I$, P is the transition matrix of $\underline{X}$ i.e.

$$\mathbb{P}\left(X_{n+1} = y | X_n = x, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) = P_{xy}$$

for all $i_0, \ldots, i_{n-1} \in I$. Then $\underline{X} \sim \text{Markov}(\lambda, P)$

Theorem. $\underline{X} = (X_n)$ is Markov(λ, P) on I if and only if

$$\mathbb{P}\left(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right) = \lambda_{x_0} p_{x_0 x_1}, \ldots p_{x_{n-1} x_n}$$

for all $n \geq 0$ and all $x_0, x_1, \ldots, x_n \in I$.

Proof. First let's prove the forward direction. Suppose that $\underline{X}$ is Markov. Then

$$\mathbb{P}\left(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\right) = \quad \mathbb{P}\left(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\right) \mathbb{P}\left(X_n = x_n | X_{n-1} = x_{n-1} \ldots, X_0 = x_0\right)$$

which iterating over n gives that

$$= \mathbb{P}\left(X_0 = x_0\right) P_{x_0 x_1} \ldots P_{x_{n-1} x_n}$$

proving the foward direction. For the converse

$$\mathbb{P}\left(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\right)$$
$$= \frac{\mathbb{P}\left(X_0 = x_0, \ldots, X_n = x_n\right)}{\mathbb{P}\left(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\right)} = \frac{\lambda_{x_0} P_{x_0 x_1} \cdots}{\lambda_{x_0} P_{x_0 x_1} \cdots} = P_{x_{n-1} x_n}$$

and with $n = 0$ we get our $\mathbb{P}(X_0 = x_0)) = \lambda_{x_0}$

Definition. For $i \in I$ the δ_i-mass at i denotes the probability mass function at i

$$\delta_{ij} = \begin{cases} 1 & j = i \\ 0 & j \neq 1 \end{cases}$$

Recall that form a finite collection of random variables $(X_0, \dots, X_n)$ are indepedent if and only if

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)) = \prod_{i=0}^{n} \mathbb{P}(X_i = x_i)$$

for all $x_0, \dots, x_n \in I$.

A process (X_n) consistant of indepedent RVS ifand only if for any collection of indices $\{t_1, \dots, t_k\}$ in $\mathbb{N}$ we have that

$$\mathbb{P}(X_{t_1} = x_{t_1}, \dots, X_{t_k} = x_{t_k}) = \prod_{i=1}^{k} \mathbb{P}(X_{t_i} = x_{t_i})$$

The process (X_i) is indepedent from the process (Y_i) iff for any $\{t_1, t_2, \dots, t_k\}$ and $\{s_1, \dots, s_m\}$ for any $k, m \geq \mathbb{N}$ we have that

$$\mathbb{P}(X_{t_1} = x_{t_1}, \cdots, Y_{s_1} = y_{s_1}, \cdots) = \mathbb{P}(X_{t_1} = x_{t_1}, \cdots) \mathbb{P}(Y_{s_1} = y_{s_1}, \cdots)$$

Note that for a Markov chain $\underline{\overline{X}}$ it is always the case that X_{n+1} is conditional independent of X_{n-1} given X_n. But typically X_{n+1} is not indepedent of X_{n-1}. Let's see an example of this.

If (X_n) are IID then $\underline{\overline{X}} = (X_n)$ is a Markov chain. What is λ and P.

Theorem. (Markov property) If $\overline{X} \sim \text{Markov}(\lambda, P)$. Then for any $m \geq 1$ and $i \in I$ conditional on $X_m = i$ the process (X_{m+n}) is Markov(δ_i, P) and it is indepdent of $X_0, \dots, X_m$.

Proof. Clearly, $\mathbb{P}(X_m = j | X_m = i) = \delta_{ij}$,

$$\begin{aligned} &\mathbb{P}(X_{m+n} = x_{m+n} | X_m = x_m \dots, X_{m+n-1} = x_{m+n-1}) \\ &= \mathbb{P}(X_{m+n} = x_{m+n} | X_{m+n-1} = x_{m+n-1}) = P_{x_{m+n-1} x_{m_n}} \end{aligned}$$

so we have that (X_{m+n}) is Markov(δ_i, P).

Now to show independence, is just an application of the law of total probability and is a lot and lot of indices. $\square$

2 Powers of the transition matrix

Suppose that $\underline{X} \sim \text{Markov}(\lambda, P)$. Where is $\mathbb{P}(X_n = x_n)$ for large n?

$$\begin{aligned}\mathbb{P}(X_n = x) &= \sum_{x_0,\ldots,x_{n-1}} \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \\ &= \sum_{x_0,\ldots,x_{n-1}} \lambda_{x_0} P_{x_0 x_1} \ldots P_{x_{n-1}x_n} \\ &= (\lambda P^n)_{x_n}\end{aligned}$$

So to understand the long time distribution of $\underline{X}$ it suffices understand the behaviour of P^n for stochastic matrices. Recall that P is stochastic if $P_{xy} \geq 0$ and each row is a PMF.

Theorem. Suppose that $\underline{X} \sim \text{Markov}(\lambda, P)$. Then
(i) $\mathbb{P}(X_n = x) = (\lambda P^n)_x$ for all $x \in I, n \geq 1$.
(ii) $\mathbb{P}(X_{n+m} = y | X_m = x) = (\delta_x P^n)_y = (P^n)_{xy}$.

Proof. We've proved the first part, let's prove the second statement. Let (X_{n+m}) be Markov with initial distribution δ_m conditional on $X_m = x$. So by the first statement

$$\mathbb{P}(X_{m+n} = y | X_n = x) = (\delta_x P^n)_y = (P^n)_{xy}$$

□

We will use the notation that

$$\begin{aligned}\mathbb{P}_x(\cdots) &= \mathbb{P}(\cdots \mid X_0 = x) \\ \mathbb{E}_x[\cdots] &= \mathbb{E}[\cdots \mid X_0 = x]\end{aligned}$$

Let's look how to calculate $P_{ij}(n)$. Suppose that I is finite, say that $I = \{1, \ldots, k\}$. How do we compute $P_{11}(n)$? If the matrix P has k distinct real eigenvalues, then it is diagonalisable. So we can write

$$P = U\text{diag}(\lambda_1, \ldots, \lambda_k)U^{-1}$$

using a change of basis matrix U. Then

$$P^n = U\text{diag}(\lambda_1^n, \ldots, \lambda_k^n)U^{-1}$$

. So $P_{11}(x) = (P^n)_{11} = a_1\lambda_1^n + \cdots + a_k\lambda_k^n$. Then we can find $P_{11}(n)$ for small values of n, substitute them to find $a_1, \ldots, a_k$.

If P has some complex eigenvalues, since P is a real-valued matrix, they necessarily come in complex conjugate pairs. So if $\lambda_1, \ldots, \lambda_{k-2}$ are real and distinct, then $\lambda_{k-1} = re^{i\theta}$ and $\lambda_k = re^{-i\theta}$. In this case since all $P_{ij}(n)$ are real

$$P_{11}(n) = \sum_{i=1}^{k-2} a_i\lambda_i^n + a_{k-1}r^n\cos(n\theta) + a_k r^n \sin(n\theta).$$

If the are repeated eigenvalues, e.g. if $\lambda_1, \ldots, \lambda_{k-2}$ are distinct and $\lambda_{k-1} = \lambda_k$ then we can use the Jordan normal form of P to get that the same expansion holds for $P_{11}(n)$ except that we need to include a term of the form $(a + bn)\lambda_{k-1}^n$.

Definition. We say that state i *leads to* j, denoted as $i \to j$ if

$$\mathbb{P}\left(X_n = j \text{ for some } n \geq 0 \mid X_0 = i\right)$$

and we say that i and j *communicate* if $i \to j$ and $j \to i$ we denote this as $i \longleftrightarrow j$.

Theorem. The following statements are equivalent.
(i) $i \to j$.
(ii) There is a path $x_0 = i, x_1, \ldots, x_n = j$ such that $p_{x_0 x_1}, \ldots, p_{x_{n-1} x_n}$ are all positive.
(iii) $P_{ij}(n) \geq 0$ for some n.

Proof. We have equality in the events

$$\{x_0 = j \text{ for some } n \geq 0\} = \bigcup_{n \geq 0} \{X_n = j\}$$

hence (i) $\iff$ (iii)

Also

$$\begin{aligned} P_{ij}(n) &= \mathbb{P}_i(X_n = j) \\ &= \sum_{\text{all } x_1, \ldots, x_{n-1}} P_{ix_1} P_{x_1 x_2} \cdots P_{x_{n-1} j} \end{aligned}$$

hence we have that (ii) $\iff$ (iii). ☐

Corollary. Communication defines a equvalence relation on the state space.

Proof. By definition $x \longleftrightarrow x$ and $x \longleftrightarrow y \iff y \longleftrightarrow x$ are obvious. Suppose that $x \longleftrightarrow y$ and $y \longleftrightarrow z$. Then by (ii) in the theorem we have a path from x to y to z so $x \longleftrightarrow z$. ☐

Definition. (Communicating class) The induced equivalences classes are called *communicating classes*. A communicating class $C \subseteq I$ is *closed* if $x \to y$ for some $x \in C$ and $y \in I$ then we have that $y \in C$.

Definition. (Absorbing) A state x is absorbing if $\{x\}$ is closed.

This is equivalent to $P_{xx} = 1$.

Definition. (Irreducible) A transition matrix P is called *irreducible* if I is a communicating class. i.e. $x \longleftrightarrow y$ for all $x, y \in I$.

Definition. (First hitting time) Let $A \subseteq I$. Then the *first hitting time* T_A for A is

$$T_A = \inf\{x \geq 0 : X_n \in A\}$$

which can be infinite if the set empty. The *hitting probability* of A is the function

$$h^A : I \to [0,1]$$

defined by

$$h_i^A = \mathbb{P}_i(T_A \leq \infty)$$

and the mean hitting time is the function

$$k^A : I \to (0,\infty]$$

is

$$k_i^A = \mathbb{E}_i(T_A) = \sum_{n=0}^{\infty} n\mathbb{P}_i(T_A = n) + \infty \cdot \mathbb{P}\left(T_A = \infty\right).$$