Markov Chains

Adam Kelly (ak2316@cam.ac.uk)

October 15, 2021

A stochastic process is said to have the 'Markov property' if, conditional on its present value, the future is independent of the past. This is a *restrictive* assumption, but we do end up with a useful model with a rich mathematical theory, which we shall study in this course.

This article constitutes my notes for the 'Markov Chains' course, held in Michaelmas 2021 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered.

Contents

1 The Markov Property **1**
- 1.1 What is a Markov Chain? . . . 1
- 1.2 Simple Markov Property . . . 3
- 1.3 Transition Probabilities . . . 3

2 Class Structure **6**
- 2.1 Communicating Classes . . . 6
- 2.2 Hitting Times . . . 7

§1 The Markov Property

§1.1 What is a Markov Chain?

Let S be a countable set (the set of possible 'states'), and let X_n be a sequence of random variables taking values in S.

Definition 1.1 (Markov Chain)

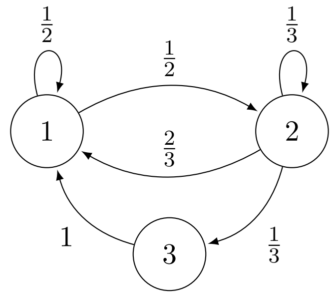
The sequence of random variables X_n is a **Markov chain** if it satisfies the **Markov property**

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

The Markov chain is said to be **homogeneous** if for all $i, j \in S$ the conditional probability $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ is independent of n.

In this course we are only going to study homogeneous Markov chains.

Markov chains are often best described by diagrams[1] which show the probability of moving from one state to another. For example, the Markov chain in the diagram below has three states which we label $\{1, 2, 3\}$, and the probability of moving from state 1 to state 2 is $1/2$, and the probability of moving from state 2 to state 3 is $1/3$, and so on.



In general, to calculate the probabilities associated with a Markov chain, we need to know two quantities.

- *The initial distribution.* We first need to know about the starting state of a Markov chain. This is described by the initial distribution $\lambda = (\lambda_i \mid i \in S)$, where $\lambda_i = \mathbb{P}(X_0 = i)$.
- *The transition probabilities.* We also need to know the probability of moving from a state $i \in S$ to a state $j \in S$. This is typically given by a *transition matrix* $P = (P_{i,j} \mid i, j \in S)$ with $p_{i,j} = \mathbb{P}(X_1 = j \mid X_0 = i)$.

These quantities are of course subject to some constraints, in that we require $\sum_{i \in S} \lambda_i = 1$ and the transition matrix P must be **stochastic**, in that $\sum_{j \in S} P_{i,j} = 1$ for all $i \in S$.

If a Markov chain X_n has initial distribution λ and transition matrix P, we say that it is Markov(λ, P).

Once we have these quantities, we can begin to actually establish various properties about the Markov chain, for example the probability that it goes through a given sequence of states.

Theorem 1.2 (Probability of a Sequence of States)
The sequence of random variables X_n is Markov(λ, P) if and only if

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \lambda_{i_0} P_{i_0,i_1} \cdots P_{i_{n-1},i_n},$$

for all $n \geq 0$ and $i_0, \ldots, i_n \in S$.

Proof. Let A_k denote the event $\{X_k = i_k\}$. First suppose that X_n is Markov(λ, P). We prove the result holds by induction. For $n = 0$ this is true by definition. Then if it holds up to n, we have

$$\begin{aligned}
\mathbb{P}(A_0 \cap \cdots \cap A_n) &= \mathbb{P}(A_0 \cap \cdots \cap A_{n-1}) \mathbb{P}(A_n \mid A_0 \cap \cdots \cap A_{n-1}) \\
&= \mathbb{P}(A_0 \cap \cdots \cap A_{n-1}) \mathbb{P}(A_n \mid A_{n-1}) \\
&= \left(\lambda_{i_0} P_{i_0,i_1} \cdots P_{i_{n-2},i_{n-1}}\right) P_{i_{n-1},i_n},
\end{aligned}$$

[1] You might notice that these diagrams are labelled directed graphs, and indeed you will see concepts from graph theory such as connectivity reoccur when we later talk about communicating classes.

which completes our induction.

Conversely, suppose that the result holds. Then with $n = 0$ we get that the initial distribution of X_n is λ. Then

$$\mathbb{P}(A_{n+1} \mid A_0 \cap \cdots \cap A_n) = \frac{\mathbb{P}(A_0 \cap \cdots \cap A_{n+1})}{\mathbb{P}(A_0 \cap \cdots \cap A_n)} = P_{i_n, i_{n+1}}.$$

Since this does not depend on $i_0, \ldots, i_{n-1}$, X_n is a homogeneous Markov chain with transition matrix P, as required. □

§1.2 Simple Markov Property

An important aspect of Markov chains is that they are *memoryless*, in the future is independent of the past, conditional on the present. This is encapsulated in the **simple Markov property**.

Theorem 1.3 (Simple Markov Property)

Let X_n be a Markov chain. Then conditional on $X_m = i$, the sequence of random variables $(X_{m+n})_{n \geq 0}$ is[a] Markov(δ_i, P), and is independent of $X_0, \ldots, X_m$.

[a] Here δ_{ij} is 1 if $i = j$ and 0 otherwise.

Proof. Given any event H determined by $X_0, \ldots, X_m$, we want to show that for an event $F = \{X_m = i_m, \ldots X_{m+n} = i_{m+n}\}$ we have

$$\mathbb{P}(H \cap F \mid X_m = i) = \delta_{i i_m} P_{i_m, i_{m+1}} \cdots P_{i_{m+n-1}, i_{m+n}} \mathbb{P}(H \mid X_m = i).$$

Indeed, considering the case of $H = \{X_0 = i_0, \ldots, X_m = i_m\}$ we have

$$\begin{aligned}\mathbb{P}(H \cap F \mid X_m = i) &= \frac{\lambda_{i_0} P_{i_0, i_1} \cdots P_{i_{m-1}, i} P_{i, i_{m+1}} \cdots P_{i_{m+n-1}, i_{m+n}}}{\mathbb{P}(X_m = i)} \\ &= \delta_{i i_m} P_{i, i_{m+1}} \cdots P_{i_{m+n-1}, i_{m+n}} \mathbb{P}(H \mid X_m = i),\end{aligned}$$

as required.

Then for a general H, we can write it as the disjoint union $H = \bigcup_{k=1}^{\infty} H_k$, and then the overall result follows by summing the above result for relevant H_k. □

§1.3 Transition Probabilities

We are now going to address how to find the probability that a Markov chain is in a given state after n steps. The core idea of this section is that we will be able to reduce such questions into questions about the transition matrix that we introduced earlier.

Recall that if X_n is a Markov chain with transition matrix P, then $P_{i,j}$ was the probability of moving from the state i to the state j. We call these the **one-step transition probabilities**. We can generalize this slightly.

Definition 1.4 (n-Step Transition Probabilities)

For a Markov chain X_n, the **n-step transition probabilities** are given by

$$p_{i,j}(n) = \mathbb{P}(X_n = j \mid X_0 = i).$$

These n-step transition probabilities naturally form the **n-step transition matrix** $P(n) = (p_{i,j}(n) \mid i, j \in S)$. The nice thing about writing these transition probabilities as a matrix is that they satisfy a lovely set of equations that relate extremely well to matrix algebra.

Theorem 1.5 (Chapman-Kolmogorov Equations)
We have that

$$p_{i,j}(n + m) = \sum_{k \in S} p_{i,k}(n) p_{k,j}(m),$$

where $i, j \in S$ and $m, n \geq 0$. In particular, $P(m + n) = P(m)P(n)$.

Proof. Using the partition theorem and simple Markov property,

$$\begin{aligned}
p_{i,j}(n + m) &= \mathbb{P}(X_{n+m} = j \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{m+n} = j \mid X_n = k, X_0 = i)\mathbb{P}(X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{m+n} = j \mid X_n = k)\mathbb{P}(X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} p_{i,k}(n) p_{k,j}(m).
\end{aligned}$$

□

So, if we have some n-step transition matrix $P(n)$, the above result shows us that it satisfies $P(n) = P^n$. This reduces our problem to just this:

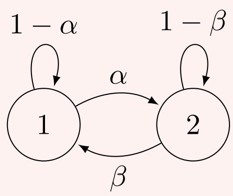
To compute $p_{i,j}(n)$, we can compute powers of
the transition matrix, and take $(P^n)_{i,j}$.

In general, this makes our problem significantly easier, and if the state space is finite, we can use tools from linear algebra such as diagonalisation.

Example 1.6 (Computing Transition Probabilities)
Let $\alpha, \beta \in (0, 1)$. Consider the Markov chain X_n with states $S = \{1, 2\}$, with transition matrix

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

A diagram of this markov chain is shown below.



We want to find the n-step transition probabilities.

Method 1 (Difference Equations). We are first going to find $p_{1,1}(n)$ using a 'bare-hands' method. By the Chapman-Kolmogorov equations (that is, conditioning on X_n) we have

$$\begin{aligned} p_{1,1}(n+1) &= p_{1,1}(n)p_{1,1}(1) + p_{1,2}(n)p_{2,1}(1) \\ &= (1-\alpha)p_{1,1}(n) + \beta p_{1,2}(n) \\ &= (1-\alpha)p_{1,1}(n) + \beta(1-p_{1,1}(n)) \\ &= p_{1,1}(n)(1-\alpha-\beta) + \beta. \end{aligned}$$

This is a recurrence relation which we can then solve using the boundary condition $p_{1,1}(0) = 1$ to get

$$p_{1,1}(n) = \begin{cases} \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} \cdot (1-\alpha-\beta)^n & \text{if } \alpha+\beta > 0, \\ 1 & \text{if } \alpha+\beta = 0. \end{cases}$$

Method 2 (Diagonalisation). An alternative solution uses some tools from matrix algebra. To calculate the n-step transition matrix P^n, we are going to diagonalise P.

The eigenvalues of P are given by the solutions to $\det(P - \mu I) = 0$, which are $\{1, 1-\alpha-\beta\}$. Thus for some invertible matrix U we have

$$P = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix} U \quad \text{and} \quad P^n = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} U.$$

Thus $p_{1,1}(n) = A + B(1-\alpha-\beta)^n$, for constants A and B. We can find these by noting the boundary conditions $p_{1,1}(0) = 1$ and $p_{1,1}(1) = 1-\alpha$, giving

$$p_{1,1}(n) = \begin{cases} \frac{\alpha}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} \cdot (1-\alpha-\beta)^n & \text{if } \alpha+\beta > 0, \\ 1 & \text{if } \alpha+\beta = 0. \end{cases}$$

In general, if the state space of a Markov chain is finite with $|S| = k$, then P is a $k \times k$ matrix with eigenvalues $\mu_1, \dots, \mu_k$.

If *all of the eigenvalues are distinct*, then P is diagonalisable, and we can write

$$P = U^{-1} \begin{pmatrix} \mu_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \mu_k \end{pmatrix} U \quad \text{and} \quad P^n = U^{-1} \begin{pmatrix} \mu_1^n & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \mu_k^n \end{pmatrix} U,$$

and $p_{i,j} = a_1\mu_1^n + \cdots + a_k\mu_k^n$, for some constants $a_1, \dots, a_k$, which are determined by the boundary conditions.

If some eigenvalue μ_k is complex, then it's conjugate is also an eigenvalue, so if $\mu_k = re^{i\theta}$ we have also the eigenvalue $\overline{\mu_k} = re^{-i\theta}$, and we can write

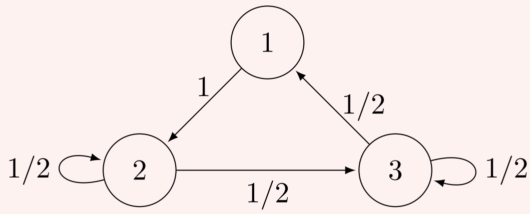
$$p_{i,j} = a_1\mu_1^n + \cdots + a_{k-2}\mu_{k-2}^n + a_{k-1}r^n\cos(n\theta) + a_k r^n \sin(n\theta),$$

since $p_{i,j}$ is real and so all of the imaginary parts must cancel out.

If *some eigenvalues repeat*, then the situation is slightly more complicated. If an eigenvalue μ_k has multiplicity m, we can simply replace a_k by a degree m polynomial in n[2].

Example 1.7 (Transition Matrix with Complex Eigenvalues)

Consider the markov chain X_s with states $S = \{1, 2, 3\}$ as shown in the diagram below. We want to find the n-step transition probability $p_{i,i}(n)$.



This Markov chain has the transition matrix

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

which has distinct eigenvalues $\{1, i/2, -i/2\}$.

We can rewrite the complex eigenvalues using trigonometric functions as

$$\begin{aligned} \frac{i}{2} &= \cos\frac{\pi}{2} + i\sin\frac{\pi}{2}, \\ -\frac{i}{2} &= \cos\frac{\pi}{2} - i\sin\frac{\pi}{2}. \end{aligned}$$

The general form for $p_{1,1}(n)$ is then given by

$$p_{1n1}(n) = A + B \cdot \left(\frac{1}{2}\right)^n \cos\left(\frac{n\pi}{2}\right) + C \cdot \left(\frac{1}{2}\right)^n \sin\left(\frac{n\pi}{2}\right).$$

The boundary conditions can be computed by hand as $p_{1,1}(0) = 1$, $p_{1,1}(1) = 0$ and $p_{1,1}(2) = 0$. This allows us to solve for A, B and C to get

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

§2 Class Structure

Test

§2.1 Communicating Classes

Definition 2.1

X is a Markov chain with transition matrix P and values in I. For $x, y \in I$ we say

[2]This comes from considering the Jordan Normal form of the transition matrix

that x **leads to** y and write it $x \to y$ if

$$\mathbb{P}(X_m = y \text{ for some } n \geq 0) > 0.$$

We say that x **communicates** with y and write $x \longleftrightarrow y$ if both $x \to y$ and $y \to x$.

Theorem 2.2

The following are equivalent:

(i) $x \to y$;

(ii) There exists a sequence of states $x = x_0, x_1, \ldots, x_k = y$ such that

$$P(x_0, x_1)P(x_1, x_2)\cdots P(x_{k-1}, x_k) > 0;$$

(iii) There exists $n \geq 0$ such that $p_{xy}(n) > 0$.

Proof. Trivial. □

Corollary 2.3

$\longleftrightarrow$ is an equivalence relation on I.

Proof. Trivial. □

Definition 2.4 (Communicating Classes)

The equivalence classes induced by $\longleftrightarrow$ on I are called **communicating classes**.

A communicating class C is **closed** if whenever $x \in C$ and $x \to y$ then $y \in C$.

A matrix P is called **irreducible** if it has a single communicating class, that is, for all $x, y \in I$ we have $x \longleftrightarrow y$.

A state x is called **absorbing** if $\{x\}$ is a closed class.

§2.2 Hitting Times

Definition 2.5

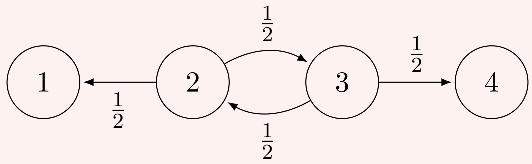
For $A \subseteq I$, we define T_A to be the **hitting time** of A, $T_A : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$, defined by $T_A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$, where we take $\inf \emptyset = \infty$.

The **hitting probability** of A is $h^A : I \to [0, 1]$ such that $h_i^A = \mathbb{P}_i(T_A < \infty)$.

The **mean hitting time** of A is $k^A : I \to \mathbb{R}$ with $k_i^A = \mathbb{E}_i[T_A] = \sum_{n=0}^{\infty} n \cdot \mathbb{P}_i(T_a = n) + \infty \cdot \mathbb{P}_i(T_a = \infty)$.

Example 2.6

Consider the Markov chain in the diagram below.



We take $A = \{4\}$, and want to find $h_2^A = \mathbb{P}_2(T_A < \infty)$. We have

$$\begin{aligned} h_2^A &= \frac{1}{2} h_3^A \\ h_3^A &= \frac{1}{2} \cdot 1 + \frac{1}{2} h_2^A \\ \implies h_2^A &= \frac{1}{3}. \end{aligned}$$

If instead we took $B = \{1, 4\}$ and wanted to find k_2^B, we would get

$$\begin{aligned} k_2^B &= 1 + \frac{1}{2} k_3^B \\ k_3^B &= 1 + \frac{1}{2} k_2^B \\ \implies k_2^B &= 2. \end{aligned}$$

In the computations above, we really should check that this is a valid method (though it is quite intuitive).

Theorem 2.7

Let $A \subseteq I$. The vector $(h_i^A)_{i \in A}$ is the minimal non-negative solution to

$$h_i^A = \begin{cases} 1 & \text{if } i \in A, \\ \sum_j P(i,j) h_j^A & \text{if } i \notin A, \end{cases}$$

where minimality means that if $(x_i)_{i \in A}$ is another solution to the linear system, then $x_i \geq h_i^A$ for all i.

Proof. We first check that h_i does indeed solve this system. □