Part IB — Markov Chains

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

Michaelmas 2022

Contents

§2 Elementary Properties

§2.1 Communicating classes

Definition 2.1 (Communication)

Let X be a Markov chain with transition matrix P and values in I. For $x, y \in I$, we say that x *leads to* y, written $x \to y$, if

$$\mathbb{P}_x(\exists n \geq 0, X_n = y) > 0$$

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

Theorem 2.1

The following are equivalent:

1. $x \to y$
2. There exists a sequence of states $x = x_0, x_1, \ldots, x_k = y$ such that
$$P(x_0, x_1)P(x_1, x_2)\ldots P(x_{k-1}, x_k) > 0$$
3. There exists $n \geq 0$ such that $p_{xy}(n) > 0$.

Proof. First, we show (i) and (iii) are equivalent. If $x \to y$, then $\mathbb{P}_x(\exists n \geq 0, X_n = y) > 0$. Then if $\mathbb{P}_x(\exists n \geq 0, X_n = y) > 0$ we must have some $n \geq 0$ such that $\mathbb{P}_x(X_n = y) = p_{xy}(n) > 0$. Note that we can write (i) as $\mathbb{P}_x\left(\bigcup_{n=0}^{\infty} X_n = y\right) > 0$. If there exists $n \geq 0$ such that $p_{xy}(n) > 0$, then certainly the probability of the union is also positive.

Now we show (ii) and (iii) are equivalent. We can write

$$p_{xy}(n) = \sum_{x_1,\ldots,x_{n-1}} P(x, x_1)\ldots P(x_{n-1}, y)$$

which leads directly to the equivalence of (ii) with (iii). □

Corollary 2.1

Communication. $\leftrightarrow$, is an equivalence relation on I.

Proof. Reflexivity: $x \leftrightarrow x$ since $p_{xx}(0) = 1$.
Transitivity: If $x \to y$ and $y \to z$ then by (ii) above, $x \to z$.

Symmetric by definition. ☐

Definition 2.2 (Communicating Classes)
The equivalence classes induced on I by the communication equivalence relation are called **communicating classes**.

Definition 2.3 (Closed Communicating Class)
A communicating class C is **closed** if $x \in C, x \to y \implies y \in C$.

Definition 2.4 (Irreducibility)
A transition matrix P is called **irreducible** if it has a single communicating class. In other words, $\forall x, y \in I, x \leftrightarrow y$.

Definition 2.5 (Absorption)
A state x is called **absorbing** if $\{x\}$ is a closed (communicating) class. Equivalently if the markov chain started from x it stays at x forever.

§2.2 Hitting times

Definition 2.6 (Hitting Time)
For $A \subseteq I$, we define the *hitting time* of A to be a random variable $T_A \colon \Omega \to \{0, 1, 2 \dots\} \cup \{\infty\}$, defined by

$$T_A(\omega) = \inf \{n \geq 0 \colon X_n(\omega) \in A\}$$

with the convention that $\inf \varnothing = \infty$.

Definition 2.7 (Hitting Probability)
The *hitting probability* of A starting at $i \in I$ is $h^A \colon I \to [0, 1]$, defined by

$$h^A(i) = h_i^A = \mathbb{P}_i \left(T_A < \infty \right), \ i \in I$$

Definition 2.8 (Mean Hitting Time)

The *mean hitting time* of A starting at $i \in I$ is $k^A \colon I \to [0, \infty]$, defined by

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

Example 2.1

Consider

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

Consider $A = \{4\}$.

$$h_1^A = 0$$

$$h_2^A = \mathbb{P}_2\left(T_A < \infty\right) = \frac{1}{2}h_1^A + \frac{1}{2}h_3^A$$

$$h_3^A = \frac{1}{2} \cdot 1 + \frac{1}{2}h_2^A$$

Hence $h_2^A = \frac{1}{3}$.

Now, consider $B = \{1, 4\}$. $k_2^B = \mathbb{E}[T_B]$

$$k_1^B = k_4^B = 0$$

$$k_2^B = 1 + \frac{1}{2}k_1^B + \frac{1}{2}k_3^B$$

$$k_3^B = 1 + \frac{1}{2}k_4^B + \frac{1}{2}k_2^B$$

Hence $k_2^B = 2$.

Theorem 2.2

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

$$x_i = \begin{cases} 1 & i \in A \\ \sum_j P(i,j)x_j & i \notin A \end{cases}$$

Minimality here means that if $(x_i)_{i \in I}$ is another non-negative solution, then $\forall i, h_i^A \leq x_i$.

Note. The vector $h_i^A = 1$ always satisfies the equation, since P is stochastic, but this solution is typically not minimal.

Proof. First, we will show that $(h_i^A)_{i \in I}$ solves the system of equations.

Certainly if $i \in A$ then $h_i^A = 1$. Suppose $i \notin A$. Consider the event $\{T_A < \infty\}$. We can write this event as a disjoint union of the following events:

$$\{T_A < \infty\} = \bigcup_{n=0}^{\infty} \{T_A = n\} = \bigcup_{n=0}^{\infty} \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$$

By countable additivity,

$$\begin{aligned}
\mathbb{P}_i\left(T_A < \infty\right) &= \underbrace{\mathbb{P}_i\left(X_0 \in A\right)}_{=0} + \sum_{n=1}^{\infty} \mathbb{P}_i\left(X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\right) \\
&= \mathbb{P}(X_1 \in A) + \sum_{n=2}^{\infty} \mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) \\
\mathbb{P}_i(X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A) &= \sum_{j \notin A} \mathbb{P}_i(X_1 = j, X_2 \notin A, \dots, X_{n-1} \in A, X_n \in A) \\
&= \sum_{j \notin A} \mathbb{P}_i(X_2 \notin A, \dots, X_{n-1} \notin A, X_n \in A \mid X_0 = i, X_1 = j)P(i,j) \\
&= \sum_{j \notin A} P(i,j)\mathbb{P}_j(X_1 \notin A, \dots, X_{n-2} \notin A, X_{n-1} \in A) \\
h_i^A &= \mathbb{P}_i(X_1 \in A) + \sum_{n=1}^{\infty} \sum_{j \notin A} P(i,j)\mathbb{P}_j(X_1 \notin A, \dots, X_n \in A) \\
&= \sum_{j \in A} h_j^A + \sum_{j \notin A} P(i,j)h_j^A \\
&= \sum_j P(i,j)h_j^A
\end{aligned}$$

So $(h_i^A)_{i \in I}$ satisfies the equation.

Now we must show minimality. If (x_i) is another non-negative solution, we must show that $h_i^A \leq x_i \; \forall \, i$. We have for $i \notin A$

$$x_i = \sum_j P(i,j)x_j = \sum_{j \in A} P(i,j) + \sum_{j \notin A} P(i,j)x_j$$

Substituting again,

$$x_i = \sum_{j \in A} P(i,j)x_j + \sum_{j \notin A} P(i,j)\left(\sum_{k \in A} P(j,k) + \sum_{k \notin A} P(j,k)x_k\right)$$
$$x_i = \mathbb{P}_i(X_1 \in A) + \mathbb{P}(X_1 \notin A, X_2 \in A) + \sum_{j \notin A}\sum_{k \notin A} P(i,j)P(j,k)x_k$$
$$x_i \geq \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \cdots + \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \text{ as } x_k \geq 0.$$
$$x_i \geq \mathbb{P}_i(T_a \leq n) \; \forall \, n \in \mathbb{N}$$

Now, note $\{T_A \leq n\}$ are a set of increasing functions of n, so by continuity of the probability measure, the probability increases to that of the union, $\{T_A < \infty\} = h_i^A$. □

Example 2.2

Consider the Markov chain previously explored:

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

Let $A = \{4\}$. Then $h_1^A = 0$ since there is no route from 1 to 4. From the theorem above, the system of linear equations is

$$h_2 = \frac{1}{2}h_1 + \frac{1}{2}h_3$$

$$h_3 = \frac{1}{2}h_4 + \frac{1}{2}h_2$$

$$h_4 = 1$$

Hence,

$$h_2 = \frac{2}{3}h_1 + \frac{1}{3}$$

$$h_3 = \frac{1}{3}h_1 + \frac{2}{3}$$

So the minimal solution arises at $h_1 = 0$.

Example 2.3

Consider $I = \mathbb{N}$, and

$$P(i, i+1) = p \in (0,1); \quad P(i, i-1) = 1 - p = q$$

Then $h_i = \mathbb{P}_i\left(T_0 < \infty\right)$ hence $h_0 = 1$. The linear equations are

$$p \neq q \implies h_i = ph_{i+1} + qh_{i-1}$$

$$p(h_{i+1} - h_i) = q(h_i - h_{i-1})$$

Let $u_i = h_i - h_{i-1}$. Then,

$$\frac{q}{p}u_i = \cdots = \left(\frac{q}{p}\right)^i u_1$$

Hence

$$h_i = \sum_{j=1}^{i}(h_j - h_{j-1}) + 1 = 1 - (1 - h_1)\sum_{j=1}^{i}\left(\frac{q}{p}\right)^j$$

The general solution is therefore

$$h_i = a + b\left(\frac{q}{p}\right)^i$$

If $q > p$, then minimality of h_i implies $b = 0$, $a = 1$. Hence,

$$h_i = 1$$

Otherwise, if $p > q$, minimality of h_i implies $a = 0$, $b = 1$. Hence,

$$h_i = \left(\frac{q}{p}\right)^i$$

If $p = q = \frac{1}{2}$, then

$$h_i = \frac{1}{2}h_{i+1} + \frac{1}{2}h_{i-1}$$

Hence, $h_i = a + bi$. Minimality implies $a = 1$ and $b = 0$.

$$h_i = 1$$

§2.3 Birth and death chain

Consider a Markov chain on $\mathbb{N}$ with

$$P(i, i+1) = p_i; \quad P(i, i-1) = q_i; \quad \forall i,\ p_i + q_i = 1$$

Now, consider $h_i = \mathbb{P}_i\left(T_0 < \infty\right)$. $h_0 = 1$, and $h_i = p_i h_{i+1} + q_i h_{i-1}$.

$$p_i(h_{i+1} - h_i) = q_i(h_i - h_{i-1})$$

Let $u_i = h_i - h_{i-1}$ to give

$$u_{i+1} = \frac{q_i}{p_i}u_i = \underbrace{\prod j = 1^i \frac{q_i}{p_i}}_{\gamma_i} u_i$$

Then

$$h_i = 1 - (1 - h_1)(\gamma_0 + \gamma_1 + \dots + \gamma_{i-1})$$

where we let $\gamma_0 = 1$. Since h_i is the minimal non-negative solution,

$$h_i \geq 0 \implies 1 - h_1 \leq \frac{1}{\sum_{j=0}^{i-1} \gamma_j} \leq \frac{1}{\sum_{j=0}^{\infty} \gamma_j}$$

By minimality, we must have exactly this bound. If $\sum_{j=0}^{\infty} \gamma_j = \infty$ then $1 - h_1 = 0 \implies h_i = 1$ for all i. If $\sum_{j=0}^{\infty} \gamma_j < \infty$ then

$$h_i = \frac{\sum_{j=i}^{\infty} \gamma_j}{\sum_{j=0}^{\infty} \gamma_j}$$

§2.4 Mean hitting times

Recall that

$$k_i^A = \mathbb{E}_i\left[T_A\right] = \sum_n n\mathbb{P}_i\left(T_A = n\right) + \infty\mathbb{P}_i\left(T_A = \infty\right)$$

Theorem 2.3

The vector $(k_i^A)_{i\in I}$ is the minimal non-negative solution to the system of equations

$$\begin{cases} 0 & \text{if } i \in A \\ 1 + \sum_{j\notin A} P(i,j)k_j^A & \text{if } i \notin A \end{cases}$$

Proof. Suppose $i \in A$. Then $k_i = 0$. Now suppose $i \notin A$. Further, we may assume that $\mathbb{P}_i(T_A = \infty) = 0$, since if that probability is positive then the claim is trivial. Indeed, if $\mathbb{P}_i(T_A = \infty) > 0$, then there must exist j such that $P(i,j) > 0$ and $\mathbb{P}_j(T_A = \infty) > 0$ since

$$\mathbb{P}_i(T_A < \infty) = \sum_j P(i,j)h_j^A \implies 1 - \mathbb{P}_i(T_A = \infty) = \sum_j P(i,j)(1 - \mathbb{P}_j(T_A = \infty))$$

Because P is stochastic,

$$\mathbb{P}_i(T_A = \infty) = \sum_j P(i,j)\mathbb{P}_j(T_A = \infty)$$

so since the left hand side is positive, there must exist j with $P(i,j) > 0$ and $\mathbb{P}_j(T_A = \infty > 0)$. For this j, we also have $k_j^A = \infty$. Now we only need to compute $\sum_n n\mathbb{P}_i(T_A = n)$.

$$\mathbb{P}_i(T_A = n) = \mathbb{P}_i(X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A)$$

Then, using the same method as the previous theorem,

$$k_i^A = \sum_n n\mathbb{P}_i(T_A = n) = 1 + \sum_{j\notin A} P(i,j)k_j^A$$

It now suffices to prove minimality. Suppose (x_i) is another solution to this system of equations. We need to show that $x_i \geq k_i^A$ for all i. Suppose $i \notin A$. Then

$$x_i = 1 + \sum_{j\notin A} P(i,j)x_j = 1 + \sum_{j\notin A} P(i,j)\left(1 + \sum_{k\notin A} P(j,k)x_k\right)$$

Expanding inductively,

$$\begin{aligned} x_i = 1 &+ \sum_{j_1\notin A} P(i,j_1) + \sum_{j_1\notin A, j_2\notin A} P(i,j_1)P(j_1,j_2) + \cdots \\ &+ \sum_{j_1\notin A,\dots,j_n\notin A} P(i,j_1)\dots P(j_{n-1},j_n) + \sum_{j_1\notin A,\dots,j_{n+1}\notin A} P(i,j)\dots P(j_n,j_{n+1})x_{j_{n+1}} \end{aligned}$$

Since x is non-negative, we can remove the last term and reach an inequality.

$$x_i \geq 1 + \sum_{j_1 \notin A} P(i, j_1) + \sum_{j_1 \notin A, j_2 \notin A} P(i, j_1) P(j_1, j_2) + \dots + \sum_{j_1 \notin A, \dots, j_n \notin A} P(i, j_1) \dots P(j_{n-1}, j_n)$$

Hence

$$\begin{aligned} x_i &\geq 1 + \mathbb{P}_i\left(T_A > 1\right) + \mathbb{P}_i\left(T_A > 2\right) + \dots + \mathbb{P}_i\left(T_A > n\right) \\ &= \mathbb{P}_i\left(T_A > 0\right) + \mathbb{P}_i\left(T_A > 1\right) + \mathbb{P}_i\left(T_A > 2\right) + \dots + \mathbb{P}_i\left(T_A > n\right) \\ &= \sum_{k=0}^{n} \mathbb{P}_i\left(T_A > k\right) \end{aligned}$$

for all n. Hence, the limit of this sum is

$$x_i \geq \sum_{k=0}^{\infty} \mathbb{P}_i\left(T_A > k\right) = \mathbb{E}_i\left[T_A\right]$$

which gives minimality as required. □

§2.5 Strong Markov property

The simple Markov property shows that, if $X_m = i$,

$$X_{m+n} \sim \text{Markov}\left(\delta_i, P\right)$$

and this is independent of $X_0, \dots, X_m$. The strong Markov property will show that the same property holds when we replace m with a finite random 'time' variable. It is not the case that *any* random variable will work; indeed, an m very dependent on the Markov chain itself might not satisfy this property.

Definition 2.9

A random time $T \colon \Omega \to \{0, 1, \dots\} \cup \{\infty\}$ is called a *stopping time* if, for all $n \in \mathbb{N}$, $\{T = n\}$ depends only on $X_0, \dots, X_n$.

Example 2.4

The hitting time $T_A = \inf\left\{n \geq 0 \colon X_n \in A\right\}$ is a stopping time. This is because we can write

$$\{T_A = n\} = \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$$

Example 2.5

The time $L_A = \sup\{n \geq 0 \colon X_n \in A\}$ is not a stopping time. This is because we need to know information about the future behaviour of X_n in order to guarantee that we are at the supremum of such events.

Theorem 2.4 (Strong Markov Property)

Let $X \sim \text{Markov}(\lambda, P)$ and T be a stopping time. Conditional on $T < \infty$ and $X_T = i$,

$$(X_{n+T})_{n\geq 0} \sim \text{Markov}(\delta_i, P)$$

and this distribution is independent of $X_0, \ldots, X_T$.

Proof. We need to show that, for all $x_0, \ldots, x_n$ and for all vectors w of any length,

$$\begin{aligned}
&\mathbb{P}(X_T = x_0, \ldots, X_{T+n} = x_n, (X_0, \ldots, X_T) = w \mid T < \infty, X_T = i) \\
&= \delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n) \mathbb{P}((X_0, \ldots, X_T) = w \colon T < \infty, X_T = i)
\end{aligned}$$

Suppose that w is of the form $w = (w_0, \ldots, w_k)$. Then,

$$\begin{aligned}
&\mathbb{P}(X_T = X_0, \ldots, X_{T+n} = x_n, (X_0, \ldots, X_T) = w \mid T < \infty, X_T = i) \\
&= \frac{\mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n, (X_0, \ldots, X_k) = w, T = k, X_k = i)}{\mathbb{P}(T < \infty, X_T = i)}
\end{aligned}$$

Now, since $\{T = k\}$ depends only on $X_0, \ldots, X_k$, by the simple Markov property we have

$$\begin{aligned}
&\mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n \mid (X_0, \ldots, X_k) = w, T = k, X_k = i) \\
&= \mathbb{P}(X_k = x_0, \ldots, X_{k+n} = x_n \mid X_k = i) = \delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n)
\end{aligned}$$

Now,

$$\begin{aligned}
&\mathbb{P}(X_T = x_0, \ldots, X_{T+n} = x_n, (X_0, \ldots, X_T) = w \mid T < \infty, X_T = i) \\
&= \frac{\delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n) \mathbb{P}((X_0, \ldots, X_k) = w \colon T = k, X_k = i)}{\mathbb{P}(T < \infty, X_T = i)} \\
&= \delta_{ix_0} P(x_0, x_1) \ldots P(x_{n-1}, x_n) \mathbb{P}((X_0, \ldots, X_T) = w \colon T < \infty, X_T = i)
\end{aligned}$$

as required. □

Example 2.6

Consider a simple random walk on $I = \mathbb{N}$, where $P(x, x \pm 1) = \frac{1}{2}$ for $x \neq 0$, and $P(0, 1) = 1$. Now, let $h_i = \mathbb{P}_i(T_0 < \infty)$. We want to calculate h_1. We can write

$$h_1 = \frac{1}{2} + \frac{1}{2}h_2$$

but the system of recursion relations this generates is difficult to solve. Instead, we will write

$$h_2 = \mathbb{P}_2(T_0 < \infty)$$

Note that in order to hit 0, we must first hit 1. So conditioning on the first hitting time of 1 being finite, after this time the process starts again from 1. We can write $T_0 = T_1 + \widetilde{T_0}$, where $\widetilde{T_0}$ is independent of T_1, with the same distribution as T_0 under $\mathbb{P}_1$. Now,

$$h_2 = \mathbb{P}_2(T_0 < \infty, T_1 < \infty) = \mathbb{P}_2(T_0 < \infty \mid T_1 < \infty)\,\mathbb{P}_2(T_2 < \infty)$$

Note that

$$\mathbb{P}_2(T_0 < \infty \mid T_1 < \infty) = \mathbb{P}_2\left(T_1 + \widetilde{T_0} < \infty \mid T_1 < \infty\right) = \mathbb{P}_2\left(\widetilde{T_0} < \infty \mid T_1 < \infty\right) = \mathbb{P}_1(T_0 < \infty)$$

But $\mathbb{P}_2(T_1 < \infty) = \mathbb{P}_1(T_0 < \infty)$, so

$$h_2 = \mathbb{P}_2(T_1 < \infty)\,\mathbb{P}_1(T_0 < \infty)$$

By translation invariance,

$$h_2 = h_1^2$$

In general, therefore, for any $n \in \mathbb{N}$,

$$h_n = h_1^n$$