Markov Chains

Cambridge University Mathematical Tripos: Part IB

4th May 2024

Contents

1 Introduction

1.1 Definition

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. A stochastic process $(X_n)_{n\geq 0}$ is called a *Markov chain* if for all $n \geq 0$ and for all $x_1 \dots x_{n+1} \in I$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_1 = x_1) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

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

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

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

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

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

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. 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

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

which is exactly the Markov property as required. □

1.3 Point masses

Definition. For $i \in I$, the δ_i-mass at i is defined by

$$\delta_{ij} = \mathbb{1}(i = j)$$

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 $x_1, \dots, x_n \in I$, we have

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

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

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

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

$$\begin{aligned} &probX_1 = x_1, \dots, X_{i_k} = x_{i_k}, Y_{j_1} = y_{j_1}, \dots, Y_{j_m} \\ &= \mathbb{P}\left(X_1 = x_1, \dots, X_{i_k} = x_{i_k}\right) \mathbb{P}\left(Y_{j_1} = y_{j_1}, \dots, Y_{j_m}\right) \end{aligned}$$

1.5 Simple Markov property

Theorem. 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 $X_0, \dots, X_m$.

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

Proof. We must show that

$$\mathbb{P}(X_m = x_0, \dots, X_{m+n} = x_n \mid X_m = i) = \delta_{ix_0} P(x_0, x_1) \dots P(x_{n-1}, x_n)$$

We have

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

The numerator is

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

Thus we have

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

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

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

So let $i = x_m$, and then

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

which gives the result as required. □

1.6 Powers of the transition matrix

Suppose $X \sim \operatorname{Markov}(\lambda, P)$ 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 $1,\dots,|I|$. 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,

$$\begin{aligned}
\mathbb{P}\left(X_n = x\right) &= \sum_{x_0,\dots,x_{n-1}\in I} \mathbb{P}\left(X_n = x, X_{n-1} = x_{n-1},\dots,X_0 = x_0\right)\\
&= \sum_{x_0,\dots,x_{n-1}\in I} \lambda_{x_0} P(x_0,x_1)\dots P(x_{n-1},x)
\end{aligned}$$

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,

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

We will write $\mathbb{P}_x(A) \coloneqq \mathbb{P}(A \mid X_0 = x)$ 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.

$$\mathbb{P}(X_n = x) = (\lambda P^n)_x;$$

$$\mathbb{P}(X_{n+m} = y \mid X_m = x) = \mathbb{P}_x(X_n = y) = p_{xy}(n)$$

1.7 Calculating powers

Example. 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 $\lambda_1, \dots, \lambda_k$ 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

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

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

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

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. 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

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

2 Elementary properties

2.1 Communicating classes

Definition. 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. The following are equivalent:
(i) $x \to y$
(ii) There exists a sequence of states $x = x_0, x_1, \dots, x_k = y$ such that

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

(iii) 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

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

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

Corollary. Communication is an equivalence relation on I.

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

Definition. The equivalence classes induced on I by the communication equivalence relation are called *communicating classes*. A communicating class C is *closed* if $x \in C, x \to y \implies y \in C$.

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

2.2 Hitting times

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

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

with the convention that $\inf \varnothing = \infty$. The *hitting probability* of A is $h^A : I \to [0, 1]$, defined by

$$h_i^A = \mathbb{P}_i\left(T_A < \infty\right)$$

The *mean hitting time* of A is $k^A : I \to [0, \infty]$, defined by

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

Example. 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 $B = \{1, 4\}$.

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

$$k_2^B = 1 + \frac{1}{2}k_1^B + \frac{1}{2}k_3^B$$
$$k_3^B = 1 + \frac{1}{2}k_4^B + \frac{1}{2}k_2^B$$

Hence $k_2^B = 2$.

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

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

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 is typically not minimal.

Proof. First, we will show that $(h_i)_{i \in A}$ 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:

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

By countable additivity,

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

By the definition of the Markov chain, we can drop the condition on X_0, and subtract one from all indices.

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

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

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

Substituting again,

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

Then

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

The last term is non-negative since x is non-negative. So

$$
x_i \geq \mathbb{P}_i\left(T_A = 1\right) + \mathbb{P}_i\left(T_A = 2\right) + \cdots + \mathbb{P}_i\left(T_A = n\right) \geq \mathbb{P}_i\left(T_A \leq n\right),\ \forall n \in \mathbb{N}
$$

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

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

Then $h_i = \mathbb{P}_i(T_0 < \infty)$ 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. The vector $(k_i^A)_{i \in I}$ is the minimal non-negative solution to the system of equations

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

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

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

Because P is stochastic,

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

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

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

Then, using the same method as the previous theorem,

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

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.

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

Hence

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

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

Example. The hitting time $T_A = \inf\{n \geq 0 : X_n \in A\}$ 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. The time $L_A = \sup\{n \geq 0 : X_n \in A\}$ 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 (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 $X_0, \dots, X_T$.

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

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

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

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

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

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

Now,

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

as required. $\square$

Example. 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

$$\begin{aligned}\mathbb{P}_2(T_0 < \infty \mid T_1 < \infty) &= \mathbb{P}_2\left(T_1 + \widetilde{T}_0 < \infty \mid T_1 < \infty\right) \\ &= \mathbb{P}_2\left(\widetilde{T}_0 < \infty \mid T_1 < \infty\right) \\ &= \mathbb{P}_1(T_0 < \infty)\end{aligned}$$

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 $n \in \mathbb{N}$,

$$h_n = h_1^n$$

3 Transience and recurrence

3.1 Definitions

Definition. Let X be a Markov chain, and let $i \in I$. i is called *recurrent* if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1$$

i is called *transient* if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 0$$

We will prove that any i is either recurrent or transient.

3.2 Probability of visits

Definition. Let $T_i^{(0)} = 0$ and inductively define

$$T_i^{(r+1)} = \inf\left\{n \geq T_i^{(r)} + 1 : X_n = i\right\}$$

We write $T_i^{(1)} = T_i$, called the first return time (or first passage time) to i. Further, let

$$f_i = \mathbb{P}_i(T_i < \infty)$$

and let the number of visits to i be defined by

$$V_i = \sum_{n=0}^{\infty} 1(X_n = i)$$

Lemma. For all $r \in \mathbb{N}, i \in I$, $\mathbb{P}_i(V_i > r) = f_i^r$.

Proof. For $r = 0$, this is trivially true. Now, suppose that the statement is true for r, and we will show that it is true for $r + 1$.

$$\begin{aligned}
\mathbb{P}_i(V_i > r+1) &= \mathbb{P}_i\left(T_i^{(r+1)} < \infty\right) \\
&= \mathbb{P}_i\left(T_i^{(r+1)} < \infty, T_i^{(r)} < \infty\right) \\
&= \mathbb{P}_i\left(T_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty\right)\mathbb{P}_i\left(T_i^{(r)} < \infty\right) \\
&= \mathbb{P}_i\left(T_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty\right)\mathbb{P}_i(V_i > r) \\
&= \mathbb{P}_i\left(T_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty\right) f_i^r
\end{aligned}$$

By the strong Markov property applied to the stopping time $T_i^{(r)}$,

$$\begin{aligned}
&= \mathbb{P}_i(T_i < \infty) f_i^r \\
&= f_i f_i^r \\
&= f_i^{r+1}
\end{aligned}$$

□

3.3 Duality of transience and recurrence

Theorem. Let X be a Markov chain with transition matrix P, and let $i \in I$. Then, exactly one of the following is true.

(i) If $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent, and

$$\sum_{n=0}^{\infty} p_{ii}(n) = \infty$$

(ii) If $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient, and

$$\sum_{n=0}^{\infty} p_{ii}(n) < \infty$$

Proof.

$$\begin{aligned}
\mathbb{E}_i\left[V_i\right] &= \mathbb{E}_i\left[\sum_{n=0}^{\infty} 1(X_n = i)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}_i\left[1(X_n = i)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i\left(X_n = i\right) \\
&= \sum_{n=0}^{\infty} p_{ii}(n)
\end{aligned}$$

First, suppose $\mathbb{P}_i\left(T_i < \infty\right) = 1$. Then, for all r, $\mathbb{P}_i\left(V_i > r\right) = 1$, so $\mathbb{P}_i\left(V_i = \infty\right) = 1$. Hence, i is recurrent. Further, $\mathbb{E}_i\left[V_i\right] = \infty$ so $\sum_{n=0}^{\infty} p_{ii}(n) = \infty$.

Now, if $f_i < 1$, by the previous lemma we see that $\mathbb{E}_i\left[V_i\right] = \frac{1}{1-f_i} < \infty$ hence $\mathbb{P}_i\left(V_i < \infty\right) = 1$. Thus, i is transient. Further, $\mathbb{E}_i\left[V_i\right] < \infty$ so $\sum_{n=0}^{\infty} p_{ii}(n) < \infty$. □

Theorem. Let x, y be states that communicate. Then, either x and y are both recurrent, or they are both transient.

Proof. Suppose x is recurrent. Then, since x and y communicate, $\exists m, \ell \in \mathbb{N}$ such that

$$p_{xy}(m) > 0; \quad p_{yx}(\ell) > 0$$

Note, $\sum_n p_{xx}(n) = \infty$. Then,

$$p_{yy}(n) \geq \sum_n p_{yy}(n + m + \ell) \geq \sum_n p_{yx}(\ell) p_{xx}(n) p_{xy}(m) \geq p_{yx}(\ell) p_{xy}(m) p_{xx}(n) = \infty$$

□

Corollary. Either all states in a communicating class are recurrent or they are all transient.

3.4 Recurrent communicating classes

Theorem. Any recurrent communicating class is closed.

Proof. Suppose a communicating class C is not closed. Then there exists $x \in C$ and $y \notin C$ such that $x \to y$. Let m be such that $p_{xy}(m) > 0$. If, starting from x, we hit y which is outside the communicating class, then we can never return to the communicating class (including x) again. In particular,

$$\mathbb{P}_x\left(V_x < \infty\right) \geq \mathbb{P}_x\left(X_m = y\right) = p_{xy}(m) > 0$$

Hence x is not recurrent, which is a contradiction. □

Theorem. Any finite closed communicating class is recurrent.

Proof. Let C be a finite closed communicating class. Let $x \in C$. Then, by the pigeonhole principle, there must exist $y \in C$ such that

$$\mathbb{P}_x(X_n = y \text{ for infinitely many } n) > 0$$

Since x and y communicate, there exists $m \in \mathbb{N}$ such that $p_{yx}(m) > 0$. Now,

$$\begin{aligned}\mathbb{P}_y(X_m = y \text{ for infinitely many } n) &\geq \mathbb{P}_x(X_m = x, X_n = y \text{ for infinitely many } n \geq m) \\ &= \mathbb{P}_x(X_n = y \text{ for infinitely many } n \geq m \mid X_m = x)\,\mathbb{P}_y(X_m = x) \\ &= \mathbb{P}_x(X_n = y \text{ for infinitely many } n)\,\mathbb{P}_y(X_m = x) > 0\end{aligned}$$

Thus y is recurrent. Since recurrence is a class property, C is recurrent. □

Theorem. Let P be irreducible and recurrent. Then, for all x, y,

$$\mathbb{P}_x\left(T_y < \infty\right) = 1$$

Proof. Since y is recurrent,

$$1 = \mathbb{P}_y(X_n = y \text{ for infinitely many } n)$$

Let m such that $p_{yx}(m) > 0$. Now,

$$\begin{aligned}1 &= \mathbb{P}_y(X_n = y \text{ infinitely often}) \\ &= \sum_z \mathbb{P}_y(X_m = z, X_n = y \text{ for infinitely many } n \geq m) \\ &= \sum_z \mathbb{P}_y(X_n = y \text{ for infinitely many } n \geq m \mid X_m = z)\,\mathbb{P}_y(X_m = z) \\ &= \sum_z \mathbb{P}_z(X_n = y \text{ for infinitely many } n)\,\mathbb{P}_y(X_m = z)\end{aligned}$$

By the strong Markov property,

$$= \sum_z \mathbb{P}_z\left(T_y < \infty\right)\mathbb{P}_y(X_n = y \text{ for infinitely many } n)\,\mathbb{P}_y(X_m = z)$$

Since y is recurrent,

$$\begin{aligned}&= \sum_z \mathbb{P}_z\left(T_y < \infty\right)\mathbb{P}_y(X_m = z) \\ &= \sum_z \mathbb{P}_z\left(T_y < \infty\right) p_{yz}(m)\end{aligned}$$

Since $p_{yz}(m) > 0$ and $\sum_z p_{yz}(m) = 1$, $\mathbb{P}_x\left(T_y < \infty\right) = 1$. □

4 Pólya's recurrence theorem

4.1 Statement of theorem

Definition. The simple random walk in $\mathbb{Z}^d$ is the Markov chain defined by

$$P(x, x+e_i) = P(x, x-e_i) = \frac{1}{2d}$$

where e_i is the standard basis.

Theorem. The simple random walk in $\mathbb{Z}^d$ is recurrent for $d = 1, d = 2$ and transient for $d \geq 3$.

4.2 One-dimensional proof

Consider $d = 1$. In this case, $P(x, x+1) = P(x, x-1) = \frac{1}{2}$. We will show that $\sum_n p_{00}(n) = \infty$, then recurrence will hold. We have $p_{00}(n) = \mathbb{P}_0(X_n = 0)$. Note that if n is odd, X_n is odd, so $\mathbb{P}_0(X_{2k+1} = 0) = 0$. So we will consider only even numbers. In order to be back at zero after $2n$ steps, we must make n steps 'up' away from the origin and make n steps 'down'. There are $\binom{2n}{n}$ ways of choosing which steps are 'up' steps. The probability of a specific choice of n 'up' and n 'down' is $\left(\frac{1}{2}\right)^{2n}$. Hence,

$$p_{00}(2n) = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2} \cdot \frac{1}{2^{2n}}$$

Recall Stirling's formula:

$$n! \sim n^n e^{-n}\sqrt{2\pi n}$$

Substituting in,

$$\frac{(2n)!}{(n!)^2} \cdot \frac{1}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}} = \frac{A}{\sqrt{n}}$$

for $A > 0$; the precise value of A is unnecessary. Hence, for some large n_0, $\forall n \geq n_0$, $p_{00}(2n) \geq \frac{A}{2\sqrt{n}}$. So

$$\sum_n p_{00}(2n) \geq \sum_{n \geq n_0} \frac{A}{2\sqrt{n}} = \infty$$

Now, let us consider the asymmetric random walk for $d = 1$, defined by $P(x, x+1) = p$ and $P(x, x-1) = q$. We can compute $p_{00}(2n) = \binom{2n}{n}(pq)^n \sim A\frac{(4pq)^n}{\sqrt{n}}$. If $p \neq q$, then $4pq < 1$ so by the geometric series we have

$$\sum_{n \geq n_0} p_{00}(2n) \leq \sum_{n \geq n_0} 2A(4pq)^n < \infty$$

So the asymmetric random walk is transient.

4.3 Two-dimensional proof

Now, let us consider the simple random walk on $\mathbb{Z}^2$. For each point $(x, y) \in \mathbb{Z}^2$, we will project this coordinate onto the lines $y = x$ and $y = -x$. In particular, we define

$$f(x, y) = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right)$$

If X_n is the simple random walk on $\mathbb{Z}^2$, we consider $f(X_n) = (X_n^+, X_n^-)$.

Lemma. $(X_n^+), (X_n^-)$ are independent simple random walks on $\frac{1}{\sqrt{2}}\mathbb{Z}$.

Proof. We can write X_n as

$$X_n = \sum_{i=1}^n \xi_i$$

where ξ_i are independent and identically distributed by

$$\mathbb{P}(\xi_1 = (1,0)) = \mathbb{P}(\xi_1 = (-1,0)) = \mathbb{P}(\xi_1 = (0,1)) = \mathbb{P}(\xi_1 = (0,-1)) = \frac{1}{4}$$

and we write $\xi_i = (\xi_i^1, \xi_i^2)$. We can then see that

$$X_n^+ = \sum_{i=1}^n \frac{\xi_i^1 + \xi_i^2}{\sqrt{2}}; \quad X_n^- = \sum_{i=1}^n \frac{\xi_i^1 - \xi_i^2}{\sqrt{2}}$$

We can check that $(X_n^+), (X_n^-)$ are simple random walks on $\frac{1}{\sqrt{2}}\mathbb{Z}$. It now suffices to prove the independence property. Note that it suffices to show that $\xi_i^1 + \xi_1^2$ and $\xi_i^1 - \xi_i^2$ are independent, since the X_n^+, X_n^- are sums of independent and identically distributed copies of these random variables. We can simply enumerate all possible values of ξ_i^1, ξ_i^2 and the result follows. □

We know that $p_{00}(n) = 0$ if n is odd. We want to find $p_{00}(2n) = \mathbb{P}_0(X_{2n} = 0)$. Note, $X_n = 0 \iff X_n^+ = X_n^- = 0$. Using the lemma above,

$$\mathbb{P}_0(X_{2n} = 0) = \mathbb{P}_0(X_n^+ = 0, X_n^- = 0) = \mathbb{P}_0(X_n^+ = 0)\,\mathbb{P}_0(X_n^- = 0) \sim \frac{A}{\sqrt{n}}\frac{A}{\sqrt{n}} = \frac{A^2}{n}$$

Hence,

$$\sum_{n \geq n_0} \mathbb{P}_0(X_{2n} = 0) \geq \sum_{n \geq n_0} = \frac{A^2}{2n} = \infty$$

which gives recurrence as required.

4.4 Three-dimensional proof

Consider $d = 3$. Again, $p_{00}(n) = 0$ if n odd. In order to return to zero after $2n$ steps, we must make i steps both up and down, j steps north and south, and k steps east and west, with $i + j + k = n$. There are $\binom{2n}{i,i,j,j,k,k}$ ways of choosing which steps in each direction we take. Each combination has probability $\left(\frac{1}{6}\right)^{2n}$ of happening. Hence,

$$p_{00}(2n) = \sum_{i,j,k \geq 0, i+j+k=n} \binom{2n}{i,i,j,j,k,k}\left(\frac{1}{6}\right)^{2n} = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \sum_{i,j,k \geq 0, i+j+k=n} \binom{n}{i,j,k}^2 \left(\frac{1}{3}\right)^{2n}$$

The sum on the right hand side is the total probability of the number of ways of placing n balls in three boxes uniformly at random, so equals one. Suppose $n = 3m$. Then we can show that $\binom{n}{i,j,k} \leq \binom{n}{m,m,m}$.

$$p_{00}(6m) \geq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{n}{m,m,m}\left(\frac{1}{3}\right)^{n}$$

Applying Stirling's formula again, we have

$$p_{00}(6m) \sim \frac{A}{n^{3/2}}$$

It is sufficient to consider $n = 3m$:

$$p_{00}(6m) \geq \frac{1}{6^2}p_{00}(6m-2); \quad p_{00}(6m) \geq \frac{1}{6^4}p_{00}(6m-4)$$

Hence

$$\sum_n p_{00}(n) < \infty$$

So the Markov chain is transient.

5 Invariant distributions

5.1 Invariant distributions

Let I be a countable set. (λ_i) is a probability distribution if $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$.

Example. Consider a Markov chain with two elements, and $P(1,1) = P(1,2) = P(2,1) = P(2,2) = \frac{1}{2}$. As $n \to \infty$, it is easy to see here that both states should be equally likely to occur. In fact, $p_{11}(n) = p_{12}(n) = p_{21}(n) = p_{22}(n) = \frac{1}{2}$. In this case, the row vector $\left(\frac{1}{2}, \frac{1}{2}\right)$ is an equilibrium probability distribution.

In general, we want to find a distribution π such that if $X_0 \sim \pi$, we have $X_n \sim \pi$ for all n. Suppose $X_0 \sim \pi$. Then,

$$\begin{aligned}
\mathbb{P}(X_1 = j) &= \sum_{i\in I} \mathbb{P}(X_0 = i, X_1 = j) \\
&= \sum_{i\in I} \mathbb{P}(X_1 = j \mid X_0 = i)\,\mathbb{P}(X_0 = i) \\
&= \sum_{i\in I} \pi(i)P(i,j)
\end{aligned}$$

Since we want $X_1 \sim \pi$, we must have $\pi(j) = \sum_{i\in I}\pi(i)P(i,j)$ for all j. In matrix form, $\pi = \pi P$.

Definition. An *invariant* (or *equilibrium*, or *stationary*) distribution for P is a probability distribution π such that $\pi = \pi P$.

Theorem. Let π be invariant. Then, if $X_0 \sim \pi$, for all n we have $X_n \sim \pi$.

Proof. If $X_0 \sim \pi$, then $X_n \sim \pi P^n = \pi$. □

Theorem. Suppose I is finite, and there exists $i \in I$ such that $p_{ij}(n) \to \pi_j$ as $n \to \infty$ for all j. Then $\pi = (\pi_j)$ is an invariant distribution.

Proof. First, we check that the sum of π_j is one. Since I is finite, we can interchange the sum and limit.

$$\sum_{j\in I} \pi_j = \sum_{j\in I} \lim_{n\to\infty} p_{ij}(n) = \lim_{n\to\infty} \sum_{j\in I} p_{ij}(n) = \lim_{n\to\infty} 1 = 1$$

So π_j is a probability distribution. We now must show $\pi = \pi P$.

$$\pi_j = \lim_{n\to\infty} p_{ij}(n) = \lim_{n\to\infty} \sum_{k\in I} p_{ik}(n-1)P(k,j) = \sum_{k\in I} \lim_{n\to\infty} p_{ik}(n-1)P(k,j) = \sum_{k\in I} \pi_k P(k,j)$$

as required. □

Remark. If I is infinite, the theorem does not necessarily hold. For example, let $I = \mathbb{Z}$, X be a simple symmetric random walk. We know that $p_{00}(n) \sim \frac{c}{\sqrt{n}}$, and $p_{0x}(n) \to 0$ as $n \to \infty$ for all $x \in \mathbb{Z}$. So zero is given by the limit but this is not a distribution.

5.2 Conditions for unique invariant distribution

In this section, we restrict our analysis to irreducible chains. If P is finite and irreducible, then 1 is an eigenvalue, since P is stochastic. The corresponding right eigenvector is $(1, \dots, 1)^\intercal$. We know that 1 is an eigenvalue of $P^\intercal$, so $P^\intercal$ has a right eigenvector corresponding to the eigenvalue of 1, which can be transposed to find a left eigenvector for P. It is possible to show using the Perron–Frobenius theorem that the eigenvector has non-negative components since P is irreducible. Since I is finite, we can normalise the left eigenvector such that its components sum to 1, giving an invariant distribution.

Definition. Let $k \in I$. Recall that T_k is the first return time to k. For every $i \in I$, we define

$$\nu_k(i) = \mathbb{E}_k\left[\sum_{n=0}^{T_k-1} 1(X_n = i)\right]$$

which is the expected number of times that we hit i while on an excursion from k (returning back to k).

Theorem. If P is irreducible and recurrent, then ν_k is an invariant measure: $\nu_k = \nu_k P$. Further, ν_k satisfies $\nu_k(k) = 1$ and in general $\nu_k(i) \in (0, \infty)$ for all i.

Proof. It is clear from the definition that $\nu_k(k) = 1$, since we must hit k exactly once on the outset, and we do not count the return. We will now prove that $\nu_k = \nu_k P$. $T_k < \infty$ with probability 1 by recurrence, and $X_{T_k} = k$. Then,

$$\begin{aligned}
\nu_k(i) &= \mathbb{E}_k\left[\sum_{n=0}^{T_k-1} 1(X_n = i)\right] \\
&= \mathbb{E}_k\left[\sum_{n=1}^{T_k} 1(X_n = i)\right] \\
&= \mathbb{E}_k\left[\sum_{n=1}^{\infty} 1(X_n = i, T_k \geq n)\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_k\left[1(X_n = i, T_k \geq n)\right] \\
&= \sum_{n=1}^{\infty} \mathbb{P}_k\left(X_n = i, T_k \geq n\right) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{P}_k\left(X_n = i, X_{n-1} = j, T_k \geq n\right) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{P}_k\left(X_n = i \mid X_{n-1} = j, T_k \geq n\right) \mathbb{P}_k\left(X_{n-1} = j, T_k \geq n\right)
\end{aligned}$$

T_k is a stopping time, so the event $\{T_k \geq n\} = \{T_k \leq n-1\}^c$ depends only on values we already know or don't care about. Hence, we can remove it.

$$\begin{aligned}
&= \sum_{n=1}^{\infty} \sum_{j \in I} \mathbb{P}_k\left(X_n = i \mid X_{n-1} = j\right) \mathbb{P}_k\left(X_{n-1} = j, T_k \geq n\right) \\
&= \sum_{n=1}^{\infty} \sum_{j \in I} P(j,i) \mathbb{P}_k\left(X_{n-1} = j, T_k \geq n\right) \\
&= \sum_{j \in I} \sum_{n=1}^{\infty} P(j,i) \mathbb{P}_k\left(X_{n-1} = j, T_k \geq n\right) \\
&= \sum_{j \in I} \sum_{n=0}^{\infty} P(j,i) \mathbb{P}_k\left(X_n = j, T_k \geq n+1\right) \\
&= \sum_{j \in I} P(j,i) \mathbb{E}_k\left[\sum_{n=0}^{T_k-1} 1(X_n = j)\right] \\
&= \sum_{j \in I} P(j,i) \nu_k(j)
\end{aligned}$$

Hence $\nu_k = \nu_k P$. We must show $\nu_k > 0$. P is irreducible, hence there exists n such that $p_{ki}(n) > 0$. Then

$$\nu_k(i) = \sum_{j \in I} \nu_k(j) P^n(j,i) \geq \nu_k(k) p_{ki}(n) > 0$$

To show $\nu_k < \infty$, let m such that $p_{ik}(m) > 0$.

$$1 = \nu_k(k) = \sum_{j \in I} \nu_k(j) P^m(j,k) \geq \nu_k(i) P^m(i,k) \implies \nu_k(i) \leq \frac{1}{P^m(i,k)} < \infty$$

□

5.3 Uniqueness of invariant distributions

Theorem. Let P be irreducible. Let λ be an invariant measure ($\lambda = \lambda P$) with $\lambda_k = 1$. Then $\lambda \geq \nu_k$. If P is recurrent, then $\lambda = \nu_k$.

Proof. Let λ be an invariant measure with $\lambda_k = 1$. Then,

$$\begin{aligned}
\lambda_i &= \sum_{j_1} \lambda_{j_1} P(j_1, i) \\
&= P(k,i) + \sum_{j_1 \neq k} \lambda_{j_1} P(j_1, i) \\
&= P(k,i) + \sum_{j_1 \neq k} P(k, j_1) P(j_1, i) + \sum_{j_1, j_2 \neq k} P(j_2, j_1) P(j_1, i) \lambda_{j_2} \\
&= P(k,i) + \sum_{j_1 \neq k} P(k, j_1) P(j_1, i) + \dots \\
&+ \sum_{j_1, \dots j_{n-1} \neq k} P(k, j_{n-1}) P(j_{n-1}, j_{n-2}) \dots P(j_2, j_1) P(j_1 i) + \underbrace{\sum_{j_1, \dots, j_n \neq k} P(j_n, j_{n-1}) \dots P(j_n, i) \lambda_{j_n}}_{\geq 0} \\
&\geq \mathbb{P}_k\left(X_1 = i, T_k \geq 1\right) + \mathbb{P}_k\left(X_2 = i, T_k \geq 2\right) + \dots + \mathbb{P}_k\left(X_n = i, T_k \geq n\right) \\
&\geq \sum_{i=1}^{n} \mathbb{P}_k\left(X_n = i, T_k \geq n\right) \\
&\to \nu_k(i)
\end{aligned}$$

as $n \to \infty$. Now, suppose P is recurrent, so ν_k is invariant. We define $\mu = \lambda - \nu_k$. Then $\mu \geq 0$ is an invariant measure satisfying $\mu_k = 0$. We need to show $\mu_i = 0$ for all i. By invariance, for all n,

$$\mu_k = \sum_j \mu_j P^n(j,k)$$

By irreducibility, we can choose n such that $P^n(i,k) > 0$.

$$\mu_k \geq P^n(i,k) \mu_i \implies \mu_i = 0$$

□

Remark. In the irreducible and recurrent case, all invariant measures are equal up to a scaling factor.

Let k be fixed. Then, ν_k is invariant, and unique in the above sense. If $\sum_i \nu_k(i)$ is finite, we can take

$$\pi_i = \frac{\nu_k(i)}{\sum_j \nu_k(j)}$$

which is an invariant distribution. The sum as required is

$$\begin{aligned}\sum_{i\in I}\nu_k(i) &= \sum_{i\in I}\mathbb{E}_k\left[\sum_{n=0}^{T_k-1}1(X_n=i)\right]\\ &= \mathbb{E}_k\left[\sum_{n=0}^{T_k-1}\sum_{i\in I}1(X_n=i)\right]\\ &= \mathbb{E}_k\left[\sum_{n=0}^{T_k-1}1\right]\\ &= \mathbb{E}_k\left[T_k\right]\end{aligned}$$

So we require that the expectation of the first return time is finite. If $\mathbb{E}_k\left[T_k\right]$ is finite, we can normalise ν_k into a (unique) invariant distribution.

5.4 Positive and null recurrence

Definition. Let $k \in I$ be a recurrent state (so $\mathbb{P}_k\left(T_k < \infty\right) = 1$). k is *positive recurrent* if $\mathbb{E}_k\left[T_k\right] < \infty$. k is called *null recurrent* otherwise; so if $\mathbb{E}_k\left[T_k\right] = \infty$.

Theorem. Let P be irreducible. Then the following are equivalent.
(i) every state is positive recurrent;
(ii) some state is positive recurrent;
(iii) P has an invariant distribution π.
If any of these conditions hold, we have

$$\pi_i = \frac{1}{\mathbb{E}_i\left[T_i\right]}$$

for all i.

Proof. First, (i) clearly implies (ii). We now show (ii) implies (iii). Let k be the a positive recurrent state, and consider ν_k. Since k is recurrent, we know that ν_k is an invariant measure. Then,

$$\sum_{i\in I}\nu_k(i) = \mathbb{E}_k\left[T_k\right] < \infty$$

since k is positive recurrent. If we define

$$\pi_i = \frac{\nu_k(i)}{\mathbb{E}_k\left[T_k\right]}$$

we have that π is an invariant distribution.

Now we show that (iii) implies (i). Let k be a state, which we will prove is positive recurrent. First, we show that $\pi_k > 0$. There exists i such that $\pi_i > 0$, and we will choose n such that $P^n(i,k) > 0$ by irreducibility. Then,

$$\pi_k = \sum_j \pi_j P^n(j,k) \geq \pi_i P^n(i,k) > 0$$

Now, we define $\lambda_i = \frac{\pi_i}{\pi_k}$. This is an invariant measure with $\lambda_k = 1$. So from the above theorem, $\lambda \geq \nu_k$. Now, since π is a distribution,

$$\mathbb{E}_k[T_k] = \sum_i \nu_k(i) \leq \sum_i \lambda_i = \sum_i \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k}\sum_i \pi_i = \frac{1}{\pi_k}$$

Hence $\mathbb{E}_k[T_k] < \infty$, so k is positive recurrent.

For the last part, we know that P is recurrent and $\lambda_i = \frac{\pi_i}{\pi_k}$ is an invariant measure with $\lambda_k = 1$. From the previous theorem, $\lambda_i = \nu_k(i)$. Hence, $\frac{\pi_i}{\pi_k} = \nu_k(i)$. Taking the sum over all i,

$$\frac{1}{\pi_k} = \mathbb{E}_k[T_k]$$

which proves the last part. □

Corollary. If P is irreducible and π is an invariant distribution, then for all x, y, the expected number of visits to y starting from x is given by

$$\nu_x(y) = \frac{\pi(y)}{\pi(x)}$$

Example. Consider the simple symmetric random walk on $\mathbb{Z}$. We have proven that this is recurrent. Suppose π is an invariant measure. So $\pi = \pi P$, giving

$$\pi_i = \frac{1}{2}\pi_{i-1} + \frac{1}{2}\pi_{i+1}$$

So $\pi_i = 1$ is an invariant measure. So all invariant measures are multiples of this. But since this is not normalisable, there exists no invariant distribution. So this walk is null recurrent.

Remark. If I is finite, we can always normalise the distribution, since we have only a finite sum.

Remark. Consider a simple random walk on $\mathbb{Z}^3$. This is transient. However, $\lambda_i = 1$ for all $i \in \mathbb{Z}^3$, this is clearly an invariant measure, so existence of an invariant measure does not imply recurrence.

Example. Consider a random walk on $\mathbb{Z}$ with transition probabilities $P(i, i+1) = p, P(i, i-1) = q$ such that $1 > p > q > 0$ and $p + q = 1$. This random walk is transient. Suppose there is an invariant distribution π, so $\pi = \pi P$. Then

$$\pi_i = \pi_{i-1}q + \pi_{i+1}p$$

Solving the recursion gives

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

This is not unique up to a multiplicative constant, due to the constant a.

Example. Consider a random walk on $\mathbb{Z}^+$ with transition probabilities $P(i, i+1) = p, P(i, i-1) = q, P(0, 0) = q$, and $p < q$ so there is a drift towards zero. We can check that this is recurrent. We will look for a solution to $\pi = \pi P$.

$$\pi_0 = q\pi_0 + q\pi_1; \quad \pi_i = p\pi_{i-1} + q\pi_{i+1}$$

Solving this system yields

$$\pi_1 = \frac{p}{q}\pi_0; \quad \pi_i = \left(\frac{p}{q}\right)^i \pi_0$$

This is unique up to a multiplicative constant. Since $p < q$, we can normalise this to reach an invariant distribution. Let $\pi_0 = 1 - \frac{p}{q}$. Then,

$$\pi_i = \left(\frac{p}{q}\right)^i \left(1 - \frac{p}{q}\right)$$

Hence the walk is positive recurrent.

5.5 Time reversibility

Theorem. Let P be irreducible, and π be an invariant distribution. Let $N \in \mathbb{N}$ and let $Y_n = X_{N-n}$ for $0 \leq n \leq N$. If $X_0 \sim \pi$, then $(Y_n)_{0 \leq n \leq N}$ is a Markov chain with transition matrix

$$\hat{P}(x, y) = \frac{\pi(y)}{\pi(x)} P(y, x)$$

and has invariant distribution π, so $\pi\hat{P} = \pi$. Further, $\hat{P}$ is also irreducible.

Proof. First, note that $\hat{P}$ is stochastic. Since $\pi = \pi P$,

$$\sum_y \hat{P}(x, y) = \sum_y \frac{\pi(y)P(y, x)}{\pi(x)} = \frac{\pi(x)}{\pi(x)} = 1$$

Now we show Y is a Markov chain.

$$\begin{aligned}
\mathbb{P}(Y_0 = y_0, \dots, Y_N = y_N) &= \mathbb{P}(X_N = y_0, \dots, X_0 = y_n) \\
&= \pi(y_N)P(y_N, y_{N-1}) \dots P(y_1, y_0) \\
&= \hat{P}(y_{N-1}, y_N)\pi(y_{N-1})P(y_{N-1}, y_{N-2}) \dots P(y_1, y_0) \\
&= \dots \\
&= \pi(y_0)\hat{P}(y_0, y_1) \dots P(y_{N-1}, y_N)
\end{aligned}$$

Hence $Y \sim \text{Markov}\left(\pi, \hat{P}\right)$. Now, we must show $\pi = \pi\hat{P}$.

$$\sum_x \pi(x)\hat{P}(x, y) = \sum_x \pi(x)\frac{P(y, x)\pi(y)}{\pi(x)} = \pi(y)\sum_x P(y, x) = \pi(y)$$

Hence π is invariant for $\hat{P}$. Now we show $\hat{P}$ is irreducible. Let $x, y \in I$. Then there exists $x = x_0, x_1, \dots, x_k = y$ such that

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

Hence

$$\hat{P}(x_k, x_{k-1}) \dots \hat{P}(x_1, x_0) = \pi(x_0)P(x_0, x_1) \dots \frac{P(x_{k-1}, x_k)}{\pi(x_k)} > 0$$

So $\hat{P}$ is irreducible. □

Definition. A Markov chain X with transition matrix P and invariant distribution π is called *reversible* or time reversible if $\hat{P} = P$. Equivalently, for all x, y,

$$\pi(x)P(x,y) = \pi(y)P(y,x)$$

These equations are called the *detailed balance equations*. Equivalently, X is reversible if, for any fixed $N \in \mathbb{N}$, $X_0 \sim \pi$ implies

$$(X_0, \dots, X_N) \overset{d}{=} (X_N, \dots, X_0)$$

which means that they are equal in distribution.

Remark. Intuitively, X is reversible if, starting from π, we cannot tell if we are watching X evolve forwards in time or backwards in time.

Lemma. Let P be a transition matrix, and μ a distribution satisfying the detailed balance equations.

$$\mu(x)P(x,y) = \mu(y)P(y,x)$$

Then μ is invariant for P.

Proof.

$$\sum_x \mu(x)P(x,y) = \sum_x \mu(y)P(y,x) = \mu(y)$$

□

Remark. If we can find a solution to the detailed balance equations which is a distribution, it must be an invariant distribution. It is simpler to solve this set of equations than to solve $\pi = \pi P$. If there is no solution to the detailed balance equations, then even if there exists an invariant distribution, the Markov chain is not reversible.

Example. Consider a random walk on the integers modulo n, with $P(i, i+1) = \frac{2}{3}$ and $P(i, i-1) = \frac{1}{3}$. We can check $\pi_i = \frac{1}{n}$ is an invariant distribution. This does not satisfy the detailed balance equations. Hence the Markov chain is not reversible.

Example. Consider a random walk on $\{0, \dots, n-1\}$ with $P(i, i+1) = \frac{2}{3}$, $P(i, i-1) = \frac{1}{3}$ and $P(0,0) = \frac{1}{3}$, $P(n-1, n-1) = \frac{2}{3}$. This is an 'opened up' version of the previous example; the circle is 'cut' open into a line at zero. The detailed balance equations give

$$\pi_i P(i, i+1) = \pi_{i+1} P(i+1, i) \implies \pi_i = k2^i$$

We can normalise this by setting k such that π is a distribution. Hence the chain is reversible.

Example. Consider a random walk on a graph. Let $G = (V, E)$ be a finite connected graph, where V is a set of vertices and E is a set of edges. The simple random walk on G has the transition matrix

$$P(x,y) = \begin{cases} \frac{1}{d(x)} & (x,y) \in E \\ 0 & (x,y) \notin E \end{cases}$$

where $d(x) = \sum_y 1((x,y) \in E)$ is the degree of x. The detailed balance equations give, for $(x,y) \in E$,

$$\pi(x)P(x,y) = \pi(y)P(y,x) \implies \frac{\pi(x)}{d(x)} = \frac{\pi(y)}{d(y)}$$

Let $\pi(x) \propto d(x)$. Then this is an invariant distribution with normalising constant $\frac{1}{\sum_y d(y)} = \frac{1}{2|E|}$. So the simple random walk on a finite connected graph is always reversible.

5.6 Aperiodicity

Definition. Let P be a transition matrix. For all i, we write

$$d_i = \gcd\{n \geq 1 : P^n(i,i) > 0\}$$

This is called the *period* of i. If $d_i = 1$, we say that i is aperiodic.

Lemma. $d_i = 1$ if and only if $P^n(i,i) > 0$ for all n sufficiently large. More rigorously, there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, $P^n(i,i) > 0$.

Proof. First, if $P^n(i,i) > 0$ for all n sufficiently large, the greatest common divisor of all sufficiently large numbers is one so this direction is trivial. Conversely, let

$$D(i) = \{n \geq 1 : P^n(i,i) > 0\}$$

Observe that if $a, b \in D(i)$ then $a + b \in D(i)$.

We claim that $D(i)$ contains two consecutive integers. Suppose that it does not, so for all $a, b \in D(i)$ we must have $|a - b| > 1$. Let r be the minimal distance between two integers in $D(i)$, so $r \geq 2$. Let n, m be numbers in $D(i)$ separated by r, so $n = m + r$. Then we can show there exists $k \in D(i)$ which can be written as $\ell r + s$ with $0 < s < r$. Indeed, if there were not such a k, we would have $d_i = 1$, since all elements would be multiples of r. Now, let $a = (\ell + 1)n$ and $b = (\ell + 1)m + k$. Then $a, b \in D(i)$, and $a - b = r - s < r$. This is a contradiction, since we have found two points in $D(i)$ with a distance smaller than the minimal distance.

Now, let $n_1, n_1 + 1$ be elements of $D(i)$. Then

$$\{xn_1 + y(n_1 + 1) : x, y \in \mathbb{N}\} \subseteq D(i)$$

It is then easy to check that $D(i) \supseteq \{n : n \geq n_1^2\}$. □

Lemma. Suppose P is irreducible and i is aperiodic. Then for all $j \in I$, j is aperiodic. Hence, aperiodicity is a class property.

Proof. There exist n, m such that $P^n(i,j) > 0, P^m(i,j) > 0$. Hence,

$$P^{n+m+r}(j,j) \geq P^n(j,i)P^r(i,i)P^n(i,j)$$

The first and last terms are positive, and the middle term is positive for sufficiently large r. □

5.7 Positive recurrent limiting behaviour

Theorem. Let P be irreducible and aperiodic with invariant distribution π, and further let $X \sim \text{Markov}(\lambda, P)$. Then for all $y \in I$, $\mathbb{P}(X_n = y) \to \pi_y$ as $n \to \infty$. Taking $\lambda = \delta_x$, we get $p_{xy}(n) \to \pi(y)$ as $n \to \infty$.

Proof. This proof will use the idea of 'coupling' of Markov chains. Let $Y \sim \text{Markov}(\pi, P)$ be independent of X. Consider the pair $((X_n, Y_n))_{n \geq 0}$. This is a Markov chain on the state space $I \times I$, because X and Y are independent. The initial distribution is $\lambda \times \pi$. We have $\mathbb{P}((X_0, Y_0) = (x, y)) = \lambda(x)\pi(y)$ and transition matrix $\widetilde{P}$ given by

$$\widetilde{P}((x, y), (x', y')) = P(x, x')P(y, y')$$

This product chain has invariant distribution $\widetilde{\pi}$ given by

$$\widetilde{\pi}(x, y) = \pi(x)\pi(y)$$

Let $a \in I$, and let $T = \inf n \geq 1 : (X_n, Y_n) = (a, a)$ be the hitting time of (a, a).

First, we want to show that $\mathbb{P}(T < \infty) = 1$. We show that $\widetilde{P}$ is irreducible. Let $(x, y), (x', y') \in I \times I$. By irreducibility of P, there exist ℓ, m such that $P^\ell(x, x') > 0$ and $P^m(y, y') > 0$. Now,

$$\widetilde{P}^{\ell+m+n}((x, y), (x', y')) = P^{\ell+m+n}(x, x')P^{\ell+m+n}(y, y')$$

Note that

$$P^{\ell+m+n}(x, x') \geq P^\ell(x, x')P^{m+n}(x', x')$$

By taking n large, by aperiodicity the product is positive. Therefore, for sufficiently large n, $P^n(x, x') > 0$. So $\widetilde{P}$ is irreducible, and there exists an invariant distribution $\widetilde{\pi}$. Hence $\widetilde{P}$ is positive recurrent. So $\mathbb{P}(T < \infty) = 1$.

Now, we define

$$Z_n = \begin{cases} X_n & n < T \\ Y_n & n \geq T \end{cases}$$

We wish to show $Z = (Z_n)n \geq 0$ has the same distribution as X, that is, $Z \sim \text{Markov}(\lambda, P)$. Now,

$$\mathbb{P}(Z_0 = x) = \mathbb{P}(X_0 = x) = \lambda(x)$$

so the initial distribution is the same. Now, we will check that Z evolves with transition matrix P. Let $A = \{Z_{n-1} = z_{n-1}, \dots, Z_0 = z_0\}$. We need to show $\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) = P(x, y)$.

$$\begin{aligned}
\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) &= \mathbb{P}(Z_{n+1} = y, T > n \mid Z_n = x, A) \\
&\quad + \mathbb{P}(Z_{n+1} = y, T \leq n \mid Z_n = x, A) \\
&= \mathbb{P}(X_{n+1} = y \mid T > n, Z_n = x, A)\,\mathbb{P}(T > n \mid Z_n = x, A) \\
&\quad + \mathbb{P}(Y_n + 1 = y \mid T \leq n, Z_n = x, A)\,\mathbb{P}(T \leq n \mid Z_n = x, A)
\end{aligned}$$

Now,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = y \mid T > n, Z_n = x, A) \\
&= \sum_z \mathbb{P}(X_{n+1} = y \mid T > n, Z_n = x, Y_n = z, A)\,\mathbb{P}(Y_n = z \mid T > n, Z - n = x, A)
\end{aligned}$$

Note, $\{T > n\}$ depends only on $(X_0, Y_0), \dots, (X_n, Y_n)$ since it is the complement of $\{T \leq n\}$, so it is a stopping time. Hence,

$$\mathbb{P}(X_{n+1} = y \mid T > n, Z_n = x, A) = \sum_z P(x,y)\mathbb{P}(Y_n = z \mid T > n, Z - n = x, A) = P(x,y)$$

Similarly,

$$\mathbb{P}(Y_{n+1} = y \mid T > n, Z_n = x, A) = P(x,y)$$

Hence,

$$\begin{aligned}\mathbb{P}(Z_{n+1} = y \mid Z_n = x, A) &= P(x,y)\mathbb{P}(T > n \mid Z_n = x, A) + P(x,y)\mathbb{P}(T \leq n \mid Z_n = x, A) \\ &= P(x,y)[\mathbb{P}(T > n \mid Z_n = x, A) + \mathbb{P}(T \leq n \mid Z_n = x, A)] \\ &= P(x,y)\end{aligned}$$

as required. Hence $Z \sim \text{Markov}(\lambda, P)$. Thus,

$$\begin{aligned}|\mathbb{P}(X_n = y) - \pi(y)| &= |\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)| \\ &= |\mathbb{P}(X_n = y, n < T) + \mathbb{P}(Y_n = y, n \geq T) \\ &- Y_n = y, n < T - \mathbb{P}(Y_n = y, n \geq T)| \\ &= |\mathbb{P}(X_n = y, n < T) - \mathbb{P}(Y_n = y, n < T)| \\ &\leq \mathbb{P}(n < T)\end{aligned}$$

As $n \to \infty$, this upper bound becomes zero, since $\mathbb{P}(T < \infty) = 1$. □

5.8 Null recurrent limiting behaviour

Theorem. Let P be irreducible, aperiodic, and null recurrent. Then, for all x, y,

$$\lim_{n \to \infty} P^n(x,y) = 0$$

Proof. Let $\widetilde{P}((x,y),(x',y')) = P(x,x')P(y,y')$ as before. We have shown previously that $\widetilde{P}$ is also irreducible. Suppose first that $\widetilde{P}$ is transient. Then,

$$\sum_n \widetilde{P}^n((x,y),(x,y)) < \infty$$

This sum is equal to

$$\sum_n (P^n(x,y))^2 < \infty$$

Hence,

$$P^n(x,y) \to 0$$

Now, conversely suppose that $\widetilde{P}$ is recurrent. Let $y \in I$. Define as before

$$\nu_y(x) = \mathbb{E}_y\left[\sum_{i=0}^{T_y - 1} 1(X_i = x)\right]$$

This measure is invariant for P since P is recurrent. Since P is null recurrent in particular, $\mathbb{E}_y\left[T_y\right] = \infty$. Hence,

$$\nu_y(I) = \sum_{x \in I} \nu_y(x) = \mathbb{E}_y\left[\sum_{i=0}^{T_y - 1} 1\right] = \mathbb{E}_y\left[T_y\right] = \infty$$

Because $\nu_y(I)$ is infinite, for all $M > 0$ there exists a finite set $A \subset I$ with $\nu_y(A) > M$. Now, we define a probability measure

$$\mu(z) = \frac{\nu_y(z)}{\nu_y(A)} 1(z \in A)$$

Now, for all $z \in I$,

$$\mu P^n(z) = \sum_x \mu(x) P^n(x, z) = \sum_x \frac{\nu_y(x)}{\nu_y(A)} 1(z \in A) P^n(x, z) \leq \frac{1}{\nu_y(A)} \sum_x \nu_y(x) P^n(x, z)$$

Since ν_y is invariant,

$$\mu P^n(z) \leq \frac{1}{\nu_y(A)} \nu_y(z) = \frac{\nu_y(z)}{\nu_y(A)}$$

Let (X, Y) be a Markov chain with matrix $\widetilde{P}$, started according to $\mu \times \delta_x$, so

$$\mathbb{P}\left(X_0 = z, Y_0 = w\right) = \mu(z) \delta_x(w)$$

Now, let

$$T = \inf\{n \geq 1 : (X_n, Y_n) = (x, x)\}$$

Since $\widetilde{P}$ is recurrent, T is finite with probability 1. Let

$$Z_n = \begin{cases} X_n & n < T \\ Y_n & n \geq T \end{cases}$$

We have already proven that Z is a Markov chain with transition matrix P, started according to μ; it has the same distribution as X. Hence,

$$\mathbb{P}\left(Z_n = y\right) = \mu P^n(y) \leq \frac{\nu_y(y)}{\nu_y(A)} = \frac{1}{\nu_y(A)}$$

Note,

$$\mathbb{P}_x\left(Y_n = y\right) \leq \mathbb{P}_x\left(Y_n = y, n \geq T\right) + \mathbb{P}_x\left(T > n\right) = \mathbb{P}_x\left(Z_n = y\right) + \mathbb{P}_x\left(T > n\right)$$

Hence,

$$\limsup_{n \to \infty} \mathbb{P}_x\left(Y_n = y\right) \leq \frac{1}{M} + 0 = \frac{1}{M}$$

Since this is true for all M, $P^n(x, y) \to 0$ as $n \to \infty$. □