Markov chains

Perla Sousi*

October 4, 2022

Contents

*University of Cambridge

These notes are largely based on [2].

1 Definition of a Markov chain

For this course we will always take I to be a countable set and all our random variables will be defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.1. A stochastic process $(X_n)_{n\geq 0}$ with values in I is called a Markov chain if for all $n \geq 0$ and for all $x_0, \ldots, x_n, x_{n+1} \in I$ we have

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

If $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ for all $x, y \in I$ is independent of n, then we call X a (time)-homogeneous Markov chain. Otherwise, it is called time-inhomogeneous.

In this course we will focus on time-homogeneous Markov chains. In this case, we will work with the transition matrix P which is defined via

$$P(x,y) = \mathbb{P}(X_1 = y \mid X_0 = x) \quad \text{for all} \quad x, y \in I.$$

The matrix P is called a stochastic matrix, because it satisfies

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

Remark 1.2. We note that we will not always work with Markov chains indexed by $\mathbb{N}$, but they could sometimes have a finite time index set.

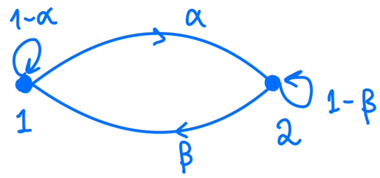
Definition 1.3. We say that the stochastic process $(X_n)_{n\geq 0}$ with values in I is Markov(λ, P) if it has initial distribution λ and transition matrix P, i.e. if for all $n \geq 0$ and for all $x_0, \ldots, x_n, x_{n+1} \in I$ we have

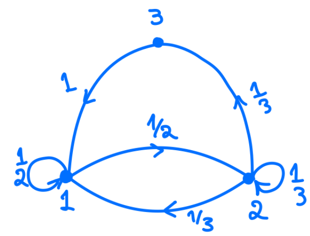
(i) $\mathbb{P}(X_0 = x_0) = \lambda_{x_0}$ and

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

We usually represent a Markov chain by its diagram corresponding to the allowed transitions. More precisely, we place a directed edge from state x to state y if there is a positive probability of jumping from x to y and on top of the edge we write the probability of this transition.

Example 1.4. Let $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$. This is a transition matrix on 2 states that we call 1 and 2. Let $P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 \end{pmatrix}$. This is a transition matrix on 3 states that we call 1, 2 and 3. We represent them as in the picture below.





2 Basic properties

The following theorem will be very useful in this course and its proof follows from the definitions above.

Theorem 2.1. *The process* X *is* Markov(λ, P) *if and only if for all* $n \geq 0$ *and all* $x_0, \dots, x_n \in I$ *we have*

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \lambda_{x_0} p_{x_0 x_1} \cdots p_{x_{n-1} x_n}. \tag{2.1}$$

Proof. Suppose first that X is Markov(λ, P) and let $x_0, \dots, x_n \in I$. Then we have

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

Iterating, we obtain

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

which completes the proof of this direction.

The other direction, i.e. assuming that X satisfies (2.1) it is immediate to prove that X is Markov(λ, P), since taking $n = 0$ gives

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

and for the Markov property we have

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

which proves (ii) of the definition. $\square$

Definition 2.2. *Let $i \in I$. The δ_i-mass at i is defined to be*

$$\delta_{ij} = \mathbf{1}(i = j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} .$$

We recall from Probability 1A the notion of independence for discrete random variables.

Definition 2.3. Let $X_1, \ldots, X_n$ be n discrete random variables taking values in I. They are called independent if for all $x_1, \ldots, x_n \in I$ we have

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) .$$

Suppose now that $(X_n)_{n \in \mathbb{N}}$ is a sequence of random variables with values in I. They are called independent if for all $k \geq 0$ and all $i_1, \ldots, i_k$ distinct and all $x_1, \ldots, x_k$ we have

$$\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k) = \prod_{\ell=1}^{k} \mathbb{P}(X_{i_\ell} = x_\ell) .$$

Suppose that $X = (X_n)_{n \geq 0}$ and $Y = (Y_n)_{n \geq 0}$ are two sequences of random variables taking values in I. Then X and Y are called independent, if for all $k, m \in \mathbb{N}$, all $i_1, \ldots, i_k \in \mathbb{N}$ and $j_1, \ldots, j_m \in \mathbb{N}$ and $x_1, \ldots, x_k, y_1, \ldots, y_m \in I$ we have

$$\begin{aligned} &\mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k, Y_{j_1} = y_1, \ldots, Y_{j_m} = y_m) \\ &= \mathbb{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k) \, \mathbb{P}(Y_{j_1} = y_1, \ldots, Y_{j_m} = y_m) . \end{aligned}$$

The next theorem formalises the statement that for a Markov chain the past and the future are independent given the present.

Theorem 2.4 ((Simple) Markov property)**.** *Suppose that X is* Markov(λ, P) *with values in I and let $m \in \mathbb{N}$. Then conditional on $X_m = i$ the process $(X_{m+n})_{n \geq 0}$ is* Markov(δ_i, P) *and it is independent of $X_0, \ldots, X_m$.*

Proof. We first show that conditional on $X_m = i$ the process $(X_{m+n})_{n \geq 0}$ is Markov(δ_i, P). We first write

$$\mathbb{P}(X_{m+n} = x_{m+n}, \ldots, X_m = x_m \mid X_m = i) = \mathbf{1}(x_m = i) \cdot \frac{\mathbb{P}(X_{m+n} = x_{m+n}, \ldots, X_m = x_m)}{\mathbb{P}(X_m = i)} . \quad (2.2)$$

By the law of total probability we get

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

where we used Theorem 2.1 in the second and final equalities. Plugging this into (2.2) we obtain

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

which again by Theorem 2.1 proves that conditional on $X_m = i$, the process $(X_{m+n})_{n \geq 0}$ is Markov(δ_i, P).

We now prove that conditional on $X_m = i$, the process $(X_{m+n})_{n \geq 0}$ is independent of $X_0, \ldots, X_m$. Let $k \geq 0$ and $x_1, \ldots, x_{m-1}, x_{m+1}, \ldots, x_{k+m} \in I$ and let $m \leq i_1 \leq \ldots \leq i_k$. We will prove that

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

First of all the probability on the left hand side above is nonzero when $x_m = i$. So with $x_m = i$ we now have

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

where the second and last equality follow from Theorem 2.1 again. This completes the proof. □

3 Powers of the transition matrix

Suppose that X is Markov(λ, P) taking values in I. We think of the rows and columns of P as indexed by $1, \ldots, |I|$ (if I is infinite, we index P by $\mathbb{N}$).

We want to understand the probability that after running the Markov chain for n steps it is in a given state $x \in I$. Let's calculate this

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

where we think of the initial distribution λ as a row vector and P^n denotes the n-th power of the transition matrix P. By convention we always take $P^0 = I$.

Let $m, n \in \mathbb{N}$. We want to calculate $\mathbb{P}(X_{n+m} = y \mid X_m = x)$. Given $X_m = x$ we know from Theorem 2.4 that $(X_{m+n})_{n\geq 0}$ is Markov(δ_x, P). Hence we have

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

It will become handy to use the following notation: for an event A we will write $\mathbb{P}_i(A)$ for the probability $\mathbb{P}(A \mid X_0 = i)$. For $i, j \in I$ we will write $p_{ij}(n)$ for the (i, j) element of the n-th power of P, i.e. $p_{ij}(n) = (P^n)_{ij}$.

We have thus proved the following theorem.

Theorem 3.1. *Suppose that X is* Markov(λ, P)*. Then for all $n, m \geq 0$ and all x, y we have*

$$\begin{aligned}&\mathbb{P}(X_n = x) = (\lambda P^n)_x\\ &\mathbb{P}(X_{n+m} = y \mid X_m = x) = p_{xy}(n).\end{aligned}$$

We now present some examples where we calculate the n-th power of the transition matrix.

Example 3.2. Let $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$ for $\alpha, \beta \in [0, 1]$. Then exploiting $P^{n+1} = P^n \cdot P$ we get

$$p_{11}(n + 1) = (1 - \alpha)p_{11}(n) + \beta p_{12}(n).$$

Also $p_{11}(n) + p_{12}(n) = 1$, which substituting above gives

$$p_{11}(n + 1) = \beta + (1 - \alpha - \beta)p_{11}(n).$$

Since $p_{11}(0) = 1$ we can solve this recursion and get the unique solution

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

General method for finding the powers of a transition matrix. Let P be a $k \times k$ stochastic matrix. We want to calculate $p_{11}(n)$. In order to do so we first find the eigenvalues of P, $\lambda_1, \dots, \lambda_k$.

$\boxed{\text{If all the eigenvalues are distinct,}}$ then P is diagonalisable, i.e. we can write

$$P = U \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k \end{pmatrix} U^{-1},$$

where U is the change of basis matrix and is invertible. Thus taking the n-th power we obtain

$$P^n = U \begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_k^n \end{pmatrix} U^{-1}.$$

Therefore we can write

$$p_{11}(n) = a_1\lambda_1^n + \ldots + a_k\lambda_k^n$$

and plugging in values for small n and using that $p_{11}(0) = 1$ we can solve for $a_1, \ldots, a_k$. If there are complex eigenvalues, then they always appear in conjugate pairs. Suppose that there are only two complex eigenvalues, λ_{k-1} and λ_k. We write them in trigonometric form, i.e.

$$\lambda_{k-1} = re^{i\theta} = r(\cos\theta + i\sin\theta) \quad \text{and} \quad \lambda_k = \overline{\lambda}_{k-1} = re^{-i\theta} = r(\cos\theta - i\sin\theta).$$

Since $p_{11}(n)$ is a real number, we can safely write the general expression for $p_{11}(n)$ as follows

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

Then as before we plug in small values of n to solve for the coefficients $a_1, \ldots, a_k$.

$\boxed{\text{If the eigenvalues are not all distinct,}}$ then if eigenvalue λ is repeated once, we include the term $(an + b)\lambda^n$ in the sum above. This follows from the Jordan normal form of P. For higher multiplicities we include the analogous terms.

Example 3.3. Let $P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$. We want to calculate $p_{11}(n)$. Following the discussion above we find the eigenvalues that are $1, i/2, -i/2$. Since

$$\frac{i}{2} = \frac{1}{2}(\cos(\pi/2) + i\sin(\pi/2)),$$

we can write

$$p_{11}(n) = a + b\left(\frac{1}{2}\right)^n \cos(n\pi/2) + c\left(\frac{1}{2}\right)^n \sin(n\pi/2).$$

Since $p_{11}(0) = 1$, $p_{11}(1) = 0$ and $p_{11}(2) = 0$ we can write down the system of equations that a, b and c satisfy to get

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

4 Communicating classes

We want to define a notion of connectivity for a Markov chain analogous to the definition of connectivity for a graph.

Definition 4.1. Let X be a Markov chain with transition matrix P and values in I. For two states x and y in I we say that x leads to y and write it $x \to y$ if

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

We say that x communicates with y if $x \to y$ and $y \to x$ and we denote it by $x \leftrightarrow y$.

We now give some equivalent conditions for two states to communicate.

Theorem 4.2. *Let x and y be two states. Then the following are equivalent:*

(1) $x \to y$*;*

(2) *there exists a sequence of states $x_0 = x, x_1, \ldots, x_k = y$ such that $P(x_0, x_1) \cdots P(x_{k-1}, x_k) > 0$;*

(3) *there exists $n \geq 0$ such that $p_{xy}(n) > 0$.*

Proof. (1)⇔(3): The event $\{X_n = y \text{ for some } n \geq 0\}$ is equal to

$$\{X_n = y \text{ for some } n \geq 0\} = \cup_{n \geq 0}\{X_n = y\}.$$

So if $\mathbb{P}_x(X_n = y \text{ for some } n \geq 0) > 0$, then there exists some $n \geq 0$ such that $\mathbb{P}_x(X_n = y) > 0$. Clearly, if $\mathbb{P}_x(X_n = y) > 0$, then $x \to y$.

(2)⇔(3): Since $p_{xy}(n)$ is the (x, y) element of P^n we have

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

Therefore the equivalence of (2) and (3) follows. □

Corollary 4.3. *The relation $\leftrightarrow$ defines an equivalence relation on I.*

Proof. We only need to check that for every $x, y, z \in I$ we have $x \leftrightarrow x$ and if $x \leftrightarrow y$ and $y \leftrightarrow z$, then $x \leftrightarrow z$.

By (3) for $n = 0$ we see that $x \leftrightarrow x$. Using (2) we get the transitivity property. □

Definition 4.4. The equivalence classes induced by $\leftrightarrow$ on I are called **communicating classes**. We say that a class C is **closed** if whenever $x \in C$ and $x \to y$, then $y \in C$.

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

A state x is called **absorbing** if $\{x\}$ is a closed class. In other words, the Markov chain starting from x always remains at x.

5 Hitting times

Suppose that X is a Markov chain with transition matrix P with values in I and let $A \subseteq I$.

Definition 5.1. We define T_A to be the **first hitting time** of A, i.e. T_A is a random variable $T_A : \Omega \to \{0, 1, \dots\} \cup \{\infty\}$ given by

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

We use the convention that the infimum of the empty set is equal to ∞.

The **hitting probability** of A is defined to be the function $h^A : I \to [0, 1]$ given by

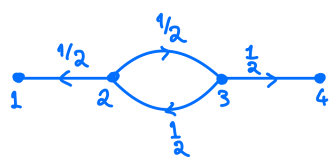
$$h_i^A = \mathbb{P}_i(T_A < \infty).$$

The **mean hitting time** of A is defined to be the function $k^A : I \to \mathbb{R}_+ \cup \{\infty\}$ given by

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

In this section we are going to show that the vectors h^A and k^A are the minimal nonnegative solutions to certain systems of linear equations. Before stating the theorems, we will look at an example to gain some intuition.

Example 5.2. Let X be the Markov chain with the following diagram:



Let $A = \{4\}$. What is h_2^A?

Starting from 2 after one step, the Markov chain could be at 1 with probability 1/2 or at 3 with probability 1/2. If it hits 1, then it gets absorbed there. If instead it hits 3, then we can forget where we started from by the Markov property (we will make this rigorous in the next theorem) and start afresh from 3. So we would have

$$h_2^A = \frac{1}{2}\mathbb{P}_1(T_A < \infty) + \frac{1}{2}h_3^A = \frac{1}{2}h_3^A$$
$$h_3^A = \frac{1}{2}h_2^A + \frac{1}{2}\mathbb{P}_4(T_A < \infty) = \frac{1}{2}h_2^A + \frac{1}{2}.$$

We can then solve the system to get $h_2^A = 1/3$.

We next set $B = \{1, 4\}$ and want to calculate k_2^B. Starting from 2 we make one step and either get absorbed at 1 or we go to 3. Then by the Markov property from 3 we start afresh. So similarly to above we have

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

Solving this system we obtain $k_2^B = 2$.

Theorem 5.3. *Let $A \subseteq I$. The vector of hitting probabilities $(h_i^A : i \in I)$ solves the following system of linear equations*

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

Moreover, $(h_i^A)_{i\in I}$ is the minimal non-negative solution ($h_i^A \geq 0$ for all i) to this system of equations, in the sense that if $(x_i)_{i\in I}$ is another non-negative solution, then $h_i^A \leq x_i$ for all $i \in I$.

Proof. We start by proving that $(h_i^A)_{i\in I}$ solves the system of linear equations. Clearly if $i \in A$, then $h_i^A = 1$. So suppose now that $i \notin A$. We can write the event $\{T_A < \infty\}$ as a disjoint countable union of events as follows

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

So taking probabilities of both sides and using the countable additivity property of probability measures when the sets are disjoint we get

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

By the definition of a Markov chain we have for every $j \in I$

$$
\mathbb{P}(X_1 \in A \mid X_0 = i, X_1 = j) = \mathbb{P}(X_1 \in A \mid X_1 = j) = \mathbb{P}(X_0 \in A \mid X_0 = j).
$$

By the simple Markov property for every $n \geq 2$ and $j \in I$ we have

$$
\begin{aligned}
\mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A \mid X_0 = i, X_1 = j) &= \mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A \mid X_1 = j) \\
&= \mathbb{P}(X_0 \notin A, \ldots, X_{n-2} \notin A, X_{n-1} \in A \mid X_0 = j).
\end{aligned}
$$

Substituting these equalities above we obtain

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

and this completes the proof of the first part of the theorem.

We next turn to prove the minimality property of $(h_i^A)_{i\in I}$. Let $(x_i)_{i\in I}$ be another non-negative solution of the linear system. We will prove that for all $i \in I$ we have $h_i^A \leq x_i$. For $i \in A$ this clearly holds, so we assume that $i \notin A$. Then we have

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

Substituting $(x_j)_{j\notin A}$ above from the linear equations we obtain

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

Iterating we get

$$\begin{aligned} x_i = \sum_{j_1\in A} P(i,j_1) + \sum_{j_1\notin A}\sum_{j_2\in A} P(i,j_1)P(j_1,j_2) + \ldots + \sum_{j_1\notin A,\ldots,j_{n-1}\notin A, j_n\in A} P(i,j_1)\cdots P(j_{n-1},j_n) \\ + \sum_{j_1\notin A,\ldots,j_{n-1}\notin A, j_n\notin A} P(i,j_1)\cdots P(j_{n-1},j_n) x_{j_n} \\ = \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \ldots + \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \\ + \sum_{j_1\notin A,\ldots,j_{n-1}\notin A, j_n\notin A} P(i,j_1)\cdots P(j_{n-1},j_n) x_{j_n} \\ = \mathbb{P}_i(T_A \leq n) + \sum_{j_1\notin A,\ldots,j_{n-1}\notin A, j_n\notin A} P(i,j_1)\cdots P(j_{n-1},j_n) x_{j_n}. \end{aligned}$$

Since $x_j \geq 0$ for all j we get that for all $n \in \mathbb{N}$

$$x_i \geq \mathbb{P}_i(T_A \leq n).$$

Taking the limit as $n \to \infty$ and using that $\{T_A \leq n\}$ are increasing events with $\cup_n \{T_A \leq n\} = \{T_A < \infty\}$ we obtain

$$x_i \geq \lim_{n\to\infty} \mathbb{P}_i(T_A \leq n) = h_i^A$$

and this finishes the proof. □

Going back to Example 5.2 we want to find h_2^A using the theorem above. We then have $h_4^A = 1$

and

$$h_2^A = \frac{1}{2}h_1^A + \frac{1}{2}h_3^A$$
$$h_3^A = \frac{1}{2}h_2^A + \frac{1}{2}h_4^A = \frac{1}{2}h_2^A + \frac{1}{2}.$$

Solving the system we get

$$h_2^A = \frac{2}{3}h_1^A + \frac{1}{3}$$
$$h_3^A = \frac{1}{3}h_1^A + \frac{2}{3}.$$

So we see that we cannot determine the value of h_1^A simply from these equations. But using the minimality condition we can deduce that $h_1^A = 0$. Of course from the formulation of the problem it was clear that $h_1^A = 0$, since there is no route from 1 to 4. There could be more complicated examples though (e.g. for infinite state spaces) where it is not possible to determine the values of some components other than using the minimality condition.

Example 5.4. We consider a simple random walk on $\mathbb{Z}_+$ with transition probabilities

$$P(0,1) = 1, \; P(i,i+1) = p \;\text{ and }\; P(i,i-1) = q \;\text{ for }\; i \geq 1$$

with $p+q=1$ and $p,q \in (0,1)$. Let $h_i = \mathbb{P}_i(T_0 < \infty)$ for $i \in \mathbb{N}$. Then $h_0 = 1$ and for every $i \geq 1$ we have

$$h_i = ph_{i+1} + qh_{i-1}.$$

If $p \neq q$, then the general solution will be of the form

$$h_i = a + b\left(\frac{q}{p}\right)^i \;\text{ for }\; i \geq 1.$$

Since h has to be non-negative and minimal, it follows that if $p < q$, then $a = 1$ and $b = 0$, i.e. $h_i = 1$ for all $i \geq 1$. On the other hand, if $q < p$, then since $h_0 = 1$, we get $a = 0$ and $b = 1$, which means that $h_i = (q/p)^i$.

In the symmetric case, i.e. when $p = q$, then the general solution is of the form

$$h_i = a + bi.$$

Non-negativity and minimality impose the conditions $a = 1$ and $b = 0$, which means that $h_i = 1$ for all i.

Example 5.5 (Birth and death chain)**.** A birth and death chain is a Markov chain on $\mathbb{N}$ with transition probabilities

$$P(0,0) = 1, P(i,i+1) = p_i \;\text{ and }\; P(i,i-1) = q_i = 1 - p_i \;\text{ for } i \geq 1,$$

with $q_i, p_i \in (0,1)$. This is a Markov chain that models the evolution of the population size, where we assume that at every step if the current population size is i, then there is probability p_i of having a birth and probability q_i of having a death. We are interested in the probability that 0 is hit, which in other words means that the population goes extinct. To this end we set $h_i = \mathbb{P}_i(T_0 < \infty)$. Then $h_0 = 1$ and

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

Rearranging the above equality we get

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

Setting $u_i = h_i - h_{i-1}$ for $i \geq 1$ we get from the above equality that

$$u_{i+1} = \frac{q_i}{p_i} u_i = \ldots = u_1 \cdot \prod_{j=1}^{i} \frac{q_j}{p_j}.$$

Note that $u_1 = h_1 - 1$, and hence

$$h_i = \sum_{j=1}^{i} u_j + 1 = 1 - (1 - h_1) \cdot \sum_{j=2}^{i} \left(\prod_{k=1}^{j-1} \frac{q_k}{p_k} \right) + (1 - h_1).$$

Writing

$$\gamma_i = \prod_{j=1}^{i} \frac{q_j}{p_j} \quad \text{and} \quad \gamma_0 = 1,$$

we have

$$h_i = 1 - (1 - h_1) \sum_{j=0}^{i-1} \gamma_j$$

and requiring h to be a non-negative solution we obtain

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

and to achieve minimality we need to take

$$h_1 = 1 - \frac{1}{\sum_{j=0}^{\infty} \gamma_j}.$$

So if $\sum_{j=0}^{\infty} \gamma_j < \infty$, then we get

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

while if $\sum_{j=0}^{\infty} \gamma_j = \infty$, then

$$h_i = 1 \quad \text{for all} \quad i \geq 1.$$

Recall that $k_i^A = \mathbb{E}_i[T_A]$ is the mean hitting time of the set A.

Theorem 5.6. *The vector $(k_i^A)_{i\in I}$ is the minimal non-negative solution to the linear system*

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

Proof. It is clear that $k_i^A = 0$ for $i \in A$, so from now on we assume that $i \notin A$.

Using the simple Markov property as in the proof of Theorem 5.3 we then have

$$\begin{aligned} k_i^A &= \sum_{n=0}^{\infty} \mathbb{P}_i(T_A > n) = \sum_{n=0}^{\infty} n\mathbb{P}(X_0 \notin A, X_1 \notin A, \ldots, X_n \notin A \mid X_0 = i) \\ &= 1 + \sum_{n=1}^{\infty} \mathbb{P}(X_1 \notin A, \ldots, X_n \notin A \mid X_0 = i) \\ &= 1 + \sum_{n=1}^{\infty}\sum_j P(i,j)\mathbb{P}(X_1 \notin A, \ldots, X_n \notin A \mid X_1 = j, X_0 = i) \\ &= 1 + \sum_{n=1}^{\infty}\sum_j P(i,j)\mathbb{P}_j(X_0 \notin A, \ldots, X_{n-1} \notin A) \\ &= 1 + \sum_{n=0}^{\infty}\sum_j P(i,j)\mathbb{P}_j(T_A > n) = 1 + \sum_j P(i,j)k_j^A = 1 + \sum_{j\notin A} P(i,j)k_j^A, \end{aligned}$$

where the last equality follows since $k_j^A = 0$ for $j \in A$.

Next we prove the minimality property of $(k_i^A)_{i\in I}$. Let $(x_i)_{i\in I}$ be another non-negative solution to the system of linear equations. Then $x_i = 0$ for $i \in A$ and iterating the linear equations we get

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

Since $x_j \geq 0$ for all j, we get

$$\begin{aligned} x_i \geq 1 + \sum_{j_1\notin A} P(i,j_1) + \sum_{j_1\notin A}\sum_{j_2\notin A} P(i,j_1)P(j_1,j_2) + \ldots + \sum_{j_1\notin A,\ldots,j_n\notin A} P(i,j_1)\cdots P(j_{n-1},j_n) \\ = 1 + \mathbb{P}_i(T_A > 1) + \ldots + \mathbb{P}_i(T_A > n) = \mathbb{P}_i(T_A > 0) + \ldots + \mathbb{P}_i(T_A > n)\,, \end{aligned}$$

which means that for all $n \in \mathbb{N}$ we have

$$x_i \geq \sum_{k=0}^{n} \mathbb{P}_i(T_A > k)\,.$$

Sending $n \to \infty$ we immediately then see that $x_i \geq k_i^A$ and this completes the proof. □

6 Strong Markov property

For a Markov chain we have proved that if we condition on $\{X_m = i\}$, where $m \in \mathbb{N}$, then the future of the process is a Markov chain starting from i with transition matrix P and it is independent of $X_0, \ldots, X_m$ (see Theorem 2.4). The question is whether we could replace the deterministic time m by a random time and still preserve the Markov property. It is quite clear that not any random time could work, simply because it could contain some information about the future of the process, in which case it would ruin the Markov property. It turns out that we can extend Theorem 2.4 to the case of a random time which is a stopping time as we now define.

Definition 6.1. A stopping time T is a random variable $T : \Omega \to \{0, 1, \ldots\} \cup \{\infty\}$ with the property that for every $n \in \mathbb{N}$ the event $\{T = n\}$ only depends on $X_0, \ldots, X_n$.

Let us now look at some examples.

Hitting times of sets are stopping times Let $A \subseteq I$ and let $T_A = \inf\{n \geq 0 : X_n \in A\}$. Then T_A is a stopping time, since for all n we have

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

Last exit times are NOT stopping times Let $A \subseteq I$ and let $L_A = \sup\{n \geq 0 : X_n \in A\}$. Then L_A is not a stopping time, since for any n we have that the event $\{L_A = n\}$ depends on $(X_{m+n})_{m \geq 0}$.

Theorem 6.2 (Strong Markov property)**.** *Let X be* Markov(λ, P) *and let T be a stopping time. Then conditional on $T < \infty$ and $X_T = i$, we have that $(X_{T+n})_{n \geq 0}$ is* Markov(δ_i, P) *and it is independent of $X_0, \ldots, X_T$.*

Proof. We need to show that for every $n \geq 0$, all $x_0, \ldots, x_n \in I$ and every $w \in \cup_k I^k$ we have

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

Suppose that w has length k. Then

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

Since the event $\{T = k\}$ only depends on $X_0, \ldots, X_k$ we have by the simple Markov property, Theorem 2.4,

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

Therefore, we deduce

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

which completes the proof the theorem. □

Example 6.3. We now want to apply the strong Markov property to derive in a slightly different way the hitting probabilities of 0 for a simple random walk on $\mathbb{N}$ with transition probabilities

$$P(0,1) = 1, P(i, i-1) = 1/2 \text{ and } P(i, i+1) = 1/2 \text{ for } i \geq 1.$$

Let $h_i = \mathbb{P}_i(T_0 < \infty)$. Then writing the hitting equations as before we get

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

Now notice that in order to hit 0 starting from 2 we first need to hit 1 and then 0. We also notice that by the strong Markov property, under $\mathbb{P}_2$ conditional on $T_1 < \infty$ we can express $T_0 = T_1 + \widetilde{T}_0$, where $\widetilde{T}_0$ is independent of T_1 and has the same distribution as T_0 under $\mathbb{P}_1$. Therefore we have

$$\begin{aligned}
\mathbb{P}_2(T_0 < \infty) = \mathbb{P}_2(T_1 < \infty, T_0 < \infty) &= \mathbb{P}_2\left(T_1 + \widetilde{T}_0 < \infty \;\middle|\; T_1 < \infty\right) \mathbb{P}_2(T_1 < \infty) \\
&= \mathbb{P}_1(T_0 < \infty)\, \mathbb{P}_2(T_1 < \infty) = h_1^2.
\end{aligned}$$

Substituting this above we deduce

$$h_1 = \frac{1}{2} + \frac{1}{2} h_1^2 \Rightarrow h_1 = 1.$$

7 Transience and recurrence

Let X be a Markov chain with state space I and transition matrix P.

Definition 7.1. A state $i \in I$ is called **recurrent** if

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

A state $i \in I$ is called **transient** if

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

Later in this section we are going to establish a dichotomy for transience and recurrence, i.e. we will show that every state is either transient or recurrent.

Definition 7.2. We define the successive times at which X visits i as follows: first we set $T_i^{(0)} = 0$

and for $k \geq 1$ inductively we set

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

We usually call $T_i = T_i^{(1)}$ the first return time to i. We also define the successive excursion lengths from i by setting

$$S_i^{(k)} = \begin{cases} T_i^{(k)} - T_i^{(k-1)} & \text{if } T_i^{(k-1)} < \infty \\ 0 & \text{otherwise} \end{cases}.$$

We set $f_i = \mathbb{P}_i(T_i < \infty)$ and write V_i for the total number of visits to i, i.e.

$$V_i = \sum_{\ell=0}^{\infty} \mathbf{1}(X_\ell = i).$$

Lemma 7.3. *For every $r \in \mathbb{N}$ we have*

$$\mathbb{P}_i(V_i > r) = f_i^r.$$

Proof. We prove the lemma by induction. For $r = 1$ we have

$$\mathbb{P}_i(V_i > 1) = \mathbb{P}_i(T_i < \infty) = f_i.$$

Suppose the claim holds for all $r \leq k$. We will prove it also holds for $r = k+1$. Indeed, we have

$$\begin{aligned} \mathbb{P}_i(V_i > k+1) &= \mathbb{P}_i\left(T_i^{(k+1)} < \infty\right) = \mathbb{P}_i\left(T_i^{(k)} < \infty, T_i^{(k+1)} < \infty\right) \\ &= \mathbb{P}_i\left(T_i^{(k+1)} < \infty \;\middle|\; T_i^{(k)} < \infty\right) \mathbb{P}_i\left(T_i^{(k)} < \infty\right). \end{aligned}$$

We now use the strong Markov property applied to the stopping time $T_i^{(k)}$. Conditional on $T_i^{(k)} < \infty$, since at this time the Markov chain is at i, it follows that $T_i^{(k+1)}$ has the same distribution as T_i under $\mathbb{P}_i$. Therefore we get

$$\mathbb{P}_i(V_i > k+1) = \mathbb{P}_i\left(T_i^{(1)} < \infty\right) \mathbb{P}_i\left(T_i^{(k)} < \infty\right) = f_i^{k+1}$$

and this finishes the proof. □

We now present a criterion for a state to be either recurrent or transient.

Theorem 7.4. *Let X be a Markov chain and $i \in I$. Then we have the following dichotomy:*

(a) *If $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\sum_n p_{ii}(n) = \infty$.*

(b) *If $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient and $\sum_n p_{ii}(n) < \infty$.*

Proof. First we write the expectation of V_i starting from i as follows

$$\mathbb{E}_i[V_i] = \mathbb{E}_i\left[\sum_{n=0}^{\infty} \mathbf{1}(X_n = i)\right] = \sum_{n=0}^{\infty} p_{ii}(n) \tag{7.1}$$

(a) In this case we have $f_i = 1$, and hence from Lemma 7.3 we get that $V_i = \infty$ with probability 1, which means that i is recurrent. Moreover, $\mathbb{E}_i[V_i] = \infty$, which implies using (7.1) that

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

(b) If $f_i < 1$, then see from Lemma 7.3 that V_i is a geometric random variable of mean $1/(1-f_i) < \infty$, which implies that V_i is finite with probability 1, i.e. that i is transient and moreover $\mathbb{E}_i[V_i] < \infty$, which means

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

This completes the proof. $\square$

Theorem 7.5. *Let $x, y \in I$ be two states that communicate, i.e. $x \leftrightarrow y$. Then either they are both recurrent or both transient.*

Proof. Suppose that x is recurrent. Then

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

Since $x \leftrightarrow y$ there exist $m, \ell \in \mathbb{N}$ such that $p_{xy}(m) > 0$ and $p_{yx}(\ell) > 0$. So then we have

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

which means that y is also recurrent and finishes the proof. $\square$

Corollary 7.6. *Either all states in a communicating class are transient or they are all recurrent.*

Theorem 7.7. *If C is a recurrent communicating class, then it must be closed.*

Proof. Suppose C is not closed. Then there must exist $x \in C$ and $y \notin C$ such that $x \to y$. ~~Let~~ As $x \to y$ $\exists$ m be such that $p_{xy}(m) > 0$. Then since once $\cancel{X}$ x hits y it cannot come back to x, as $y \notin C$ we get

$$\mathbb{P}_x(V_x < \infty) \geq \mathbb{P}_x(X_m = y) > 0,$$

implying that x is transient, which is a contradiction. $\square$

Theorem 7.8. *A finite closed class is recurrent.*

Proof. Let C be a finite closed communicating class and let $x \in C$. Then by the pigeonhole principle there exists $y \in C$ such that

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

Since $x \leftrightarrow y$, there exists $m \geq 0$ such that $\mathbb{P}_y(X_m = x) > 0$. Therefore by the simple Markov property we get

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

This now implies from Theorem 7.4 that y is recurrent, and hence by Theorem 7.5 we conclude that C is a recurrent class. $\square$

Theorem 7.9. *Let P be irreducible and recurrent. Then for all x, y we have*

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

Proof. Since P is irreducible, there exists $m \geq 0$ such that $p_{yx}(m) > 0$. Also since y is recurrent and by the simple Markov property, we get

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

Now notice that by the strong Markov property at time T_y we get

$$\mathbb{P}_z(X_n = y \text{ for infinitely many } n) = \mathbb{P}_z(T_y < \infty)\, \mathbb{P}_y(X_n = y \text{ for infinitely many } n) = \mathbb{P}_z(T_y < \infty)\,.$$

Substituting this above yields

$$1 = \mathbb{P}_y(X_n = y \text{ for infinitely many } n) = \sum_z p_{yz}(m) \mathbb{P}_z(T_y < \infty)\,.$$

Since $p_{yx}(m) > 0$ and $\sum_z p_{yz}(m) = 1$, we get from the above equality that

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

and this concludes the proof. $\square$

7.1 Random walks on $\mathbb{Z}^d$

In this section we are going to show that simple random walk on $\mathbb{Z}$ and $\mathbb{Z}^2$ is recurrent, while it is transient in $\mathbb{Z}^d$ for $d \geq 3$.

Definition 7.10. A simple random walk in $\mathbb{Z}^d$ is a Markov chain that has transition probabilities

$$P(x, x + e_i) = P(x, x - e_i) = \frac{1}{2d} \quad \text{for all } x \in \mathbb{Z}^d,$$

where $(e_i)_{i \leq d}$ is the standard basis of $\mathbb{R}^d$.

Theorem 7.11 (Polya). *Simple random walk in $\mathbb{Z}^d$ is recurrent when $d \leq 2$ and transient for $d \geq 3$.*

Random walk on $\mathbb{Z}$

Let X be a simple symmetric random walk on $\mathbb{Z}$, i.e. it has transition probabilities

$$P(x, x-i) = P(x, x+1) = \frac{1}{2} \quad \forall\, x \in \mathbb{Z}.$$

We will prove that it is recurrent by showing that

$$\sum_n p_{00}(n) = \infty.$$

First of all we notice that $p_{00}(n)$ is non-zero when n is even. So let's calculate $\mathbb{P}_0(X_{2n} = 0)$. In order to be at 0 at time $2n$ when the walk starts from 0, it needs to make n steps to the right and n steps to the left. There are $\binom{2n}{n}$ ways to pick the steps that will be to the right and then the remaining ones will be to the left. Each choice of right/left steps has probability $1/2^{2n}$ of occurring. Therefore we get

$$\mathbb{P}_0(X_{2n} = 0) = \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n!n!} \cdot \frac{1}{2^{2n}}.$$

Using Stirling's formula, i.e. that as $n \to \infty$

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

we get

$$\mathbb{P}_0(X_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}.$$

Therefore, for n_0 sufficiently large and $n \geq n_0$ we get

$$p_{00}(2n) \geq \frac{1}{2\sqrt{\pi n}},$$

and hence

$$\sum_n p_{00}(2n) \geq \sum_{n \geq n_0} \frac{1}{2\sqrt{\pi n}} = \infty,$$

which implies that the walk is recurrent.

Simple asymmetric random walk on $\mathbb{Z}$

Suppose that X is a simple random walk on $\mathbb{Z}$ with transition probabilities

$$P(i, i+1) = p \;\text{ and }\; P(i, i-1) = q$$

with $p + q = 1$ and $p, q \in (0, 1)$. Then the same reasoning as above gives

$$p_{00}(2n) = \binom{2n}{n} p^n q^n \sim \frac{(4pq)^n}{\sqrt{\pi n}},$$

where in the last step we used again Stirling's formula as $n \to \infty$. If $p \neq q$, then $4pq < 1$, and hence

$$\sum_n p_{00}(2n) \le \sum_{n \ge n_0} 2(4pq)^n < \infty,$$

thus proving transience.

Random walk on $\mathbb{Z}^2$

We will now prove recurrence of simple random walk on $\mathbb{Z}^2$. We will use a very nice trick that works in two dimensions by projecting the walk on the two diagonal lines $y = x$ and $y = -x$. More precisely, let X_n be the simple random walk in $\mathbb{Z}^2$and let f be the transformation

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

Then we write $f(X_n)$ as $f(X_n) = (X_n^+, X_n^-)$.

The reason we project on the two diagonals is because this results into two independent simple random walks in $\mathbb{Z}/\sqrt{2}$ as we will explain now.

Lemma 7.12. *Both (X_n^+) and (X_n^-) are simple random walks in $\mathbb{Z}/\sqrt{2}$ and they are independent.*

Proof. We can write $X_n = \sum_{i=1}^n \xi_i$ where $(\xi_i)_i$ are i.i.d. random variables distributed as follows

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

Then writing $\xi_i = (\xi_i^1, \xi_i^2)$, we have $X_n^+ = \sum_{i=1}^n (\xi_i^1 + \xi_i^2)/\sqrt{2}$ and $X_n^- = \sum_{i=1}^n (\xi_i^1 - \xi_i^2)/\sqrt{2}$. One can then check that both X_n^+ and X_n^- are simple random walks in $\mathbb{Z}/\sqrt{2}$. To prove the independence, since the (ξ_i) are independent, it is enough to show that $\xi_i^1 + \xi_i^2$ is independent of $\xi_i^1 - \xi_i^2$. This follows from calculating all possible probabilities, i.e.

$$\mathbb{P}\big(\xi_i^1 + \xi_i^2 = 1, \xi_i^1 - \xi_i^2 = -1\big) = \mathbb{P}(\xi_i = (0, 1)) = \frac{1}{4} = \mathbb{P}\big(\xi_i^1 + \xi_i^2 = 0\big)\, \mathbb{P}\big(\xi_i^1 - \xi_i^2 = -1\big)$$

and similarly for all the other possible events. □

We now notice that $X_n = 0$ if and only if both $X_n^+ = 0$ and $X_n^- = 0$. Using the independence we proved above we obtain

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

using the estimates we established in the 1-dimensional case. Therefore, we conclude

$$\sum_n p_{00}(2n) = \infty,$$

showing that the random walk in $\mathbb{Z}^2$ is recurrent.

Simple random walk in $\mathbb{Z}^3$

As in the previous cases, X can be back at 0 only at even times. In order for X to be at 0 at time $2n$ it must make i steps up, i steps down, j steps north and j south and k east and k west for some $i,j,k \geq 0$ with $i+j+k=n$. There are $\binom{2n}{i,i,j,j,k,k}$ ways of choosing which steps will be done in each direction. So we get

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

We now see that

$$\sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i,j,k} \cdot \left(\frac{1}{3}\right)^{n} = 1,$$

since the sum corresponds to the total probability of all the ways of placing n balls into 3 boxes uniformly at random. Let us now take $n = 3m$. Then by direct comparison, it is not hard to see that

$$\binom{n}{i,j,k} \leq \binom{n}{m,m,m}.$$

Therefore, we deduce

$$p_{00}(2n) \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{n}{m,m,m}\left(\frac{1}{3}\right)^{n} \sim \frac{C}{n^{3/2}},$$

where C is a positive constant and this asymptotic follows again from Stirling's formula. We thus get

$$\sum_m p_{00}(6m) < \infty.$$

Using that $p_{00}(6m) \geq (1/6)^2 p_{00}(6m-2)$ and $p_{00}(6m) \geq (1/6)^4 p_{00}(6m-4)$ we get that

$$\sum_n p_{00}(2n) < \infty$$

and this concludes the proof that simple random walk on $\mathbb{Z}^3$ is transient.

8 Invariant distribution

Let I be a discrete set. Recall that $\lambda = (\lambda_i : i \in I)$ is called a (probability) distribution if $\lambda_i \geq 0$ for all i and $\sum_{i\in I} \lambda_i = 1$.

Before defining the notion of an invariant distribution (or otherwise called equilibrium or stationary) let us start with an example.

Let us consider a 2 state Markov chain with states called 1 and 2 and let

$$P(1,2) = P(1,1) = P(2,1) = P(2,2) = \frac{1}{2}.$$

As $n \to \infty$ where do we expect the chain to be? Will state 1 be more likely than state 2? Clearly, by symmetry, we expect both of them to be equally likely. So we expect, $p_{11}(n) \to 1/2$ and $p_{12}(n) \to 1/2$ as $n \to \infty$ and similarly starting from 2.

We can think of $(1/2, 1/2)$ as the equilibrium distribution of the Markov chain, since running it for a long time we expect it to equilibrate at $(1/2, 1/2)$.

Suppose next that we want to define a probability distribution π on the state space I, so that if $X_0 \sim \pi$, then $X_n \sim \pi$ for all n. Let's see what π should satisfy in this case. Since $X_0 \sim \pi$, we get

$$\mathbb{P}(X_1 = j) = \sum_i \mathbb{P}(X_0 = i, X_1 = j) = \sum_i \mathbb{P}(X_0 = i)\, \mathbb{P}(X_1 = j \mid X_0 = i) = \sum_i \pi(i) P(i, j).$$

So if we want $X_n \sim \pi$ for all n, it follows that

$$\pi(j) = \sum_i \pi(i) P(i, j),$$

or in matrix form

$$\pi = \pi P,$$

where π is a row vector.

Definition 8.1. A probability distribution $\pi = (\pi_i : i \in I)$ is called an invariant/equilibrium/stationary distribution for the Markov chain with transition matrix P if $\pi = \pi P$.

Theorem 8.2. *Let X be a Markov chain with transition matrix P and invariant distribution π. If $X_0 \sim \pi$, then $X_n \sim \pi$ for all n.*

Proof. We will prove it by induction. For $n = 0$ it holds, since $X_0 \sim \pi$. Suppose it holds for n, then

$$\mathbb{P}(X_{n+1} = j) = \sum_i \mathbb{P}(X_n = i, X_{n+1} = j) = \sum_i \mathbb{P}(X_n = i)\, P(i, j) = \sum_i \pi_i P(i, j) = \pi_j,$$

where for the penultimate equality we used the induction hypothesis and for the last one we used that $\pi = \pi P$. □

Theorem 8.3. *Let I be a finite state space and suppose that there exists $i \in I$ such that for all j*

$$p_{ij}(n) \to \pi_j \quad \text{as } n \to \infty.$$

Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

Proof. We first show that π is a probability distribution. Indeed, we have

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

where we used that I is a finite set, so we can safely interchange the sum and the limit.

We next prove that $\pi = \pi P$. Let $j \in I$. we have

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

where again we were able to interchange sum and limit because I is finite. $\square$

Remark 8.4. Note that the assumption that I is finite is essential in the theorem above. Take for instance the simple symmetric random walk on $\mathbb{Z}$ as we studied previously. Then we showed that

$$p_{00}(n) \sim \frac{C}{\sqrt{n}} \quad \text{as } n \to \infty,$$

where C is a positive constant. It is not hard to show that also $p_{0x}(n) \to 0$ as $n \to \infty$ for all x. So in this case, the limit is not a probability distribution.

Example 8.5. Let us consider again the two state Markov chain with

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

By calculating the eigenvalues or otherwise we get

$$p_{11}(n) \to \frac{1}{2} \text{ and } p_{12}(n) \to \frac{1}{2} \quad \text{as } n \to \infty$$

So $\pi = (1/2, 1/2)$ is an invariant distribution. Solving $\pi = \pi P$ we also get $\pi = (1/2, 1/2)$.

We now want to understand whether a transition matrix has an invariant distribution and whether this is unique. We will only talk about irreducible Markov chains, since if the state space consists of several communicating classes, then the invariant distribution might not be unique.

Remark 8.6. Note that for an irreducible transition matrix P on a finite state space I one can deduce the existence of an invariant distribution using the Perron-Frobenius theorem, a result in linear algebra. The following theorem works both in the finite and infinite setting using probabilistic arguments.

We next define a measure in terms of the numbers of visits to a vertex during an excursion from another vertex. Afterwards we will show that when P is irreducible and recurrent, this is always an invariant measure.

Definition 8.7. Let $k \in I$. Recall that T_k is the first return time to k, i.e.

$$T_k = \inf\{n \geq 1 : X_n = k\}.$$

We define $\nu_k(i)$ to be the expected number of visits to i during an excursion from k, i.e.

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

Theorem 8.8. *Suppose that P is irreducible and recurrent. Then ν_k is an invariant measure, i.e. $\nu_k = \nu_k P$, satisfying $0 < \nu_k(i) < \infty$ and $\nu_k(k) = 1$.*

Proof. By the definition of ν_k it is clear that $\nu_k(k) = 1$. We next prove that ν_k is invariant.

First we note that since the chain is recurrent, then $T_k < \infty$ with probability 1 and since $X_{T_k} = 1$ by definition of T_k, we obtain

$$\nu_k(i) = \mathbb{E}_k\left[\sum_{n=1}^{T_k} \mathbf{1}(X_n = i)\right] = \mathbb{E}_k\left[\sum_{n=1}^{\infty} \mathbf{1}(n \le T_k) \cdot \mathbf{1}(X_n = i)\right] = \sum_{n=1}^{\infty} \mathbb{P}_k(X_n = i, n \le T_k) \,. \quad (8.1)$$

By the law of total probability we get for all $n \ge 1$

$$\mathbb{P}_k(X_n = i, n \le T_k) = \sum_j \mathbb{P}_k(X_n = i, X_{n-1} = j, n \le T_k) \,.$$

We now claim that the event $\{T_k \ge n\}$ only depends on $X_0, \ldots, X_{n-1}$. Indeed, $\{T_k \ge n\}$ is the complement of $\{T_k \le n-1\}$ which is the union

$$\{T_k \le n-1\} = \bigcup_{\ell \le n-1} \{T_k = \ell\}.$$

Using that T_k is a stopping time, each of the events of the union only depends on $X_0, \ldots, X_k$, and hence the claim follows. Therefore, we get

$$\begin{aligned}\mathbb{P}_k(X_n = i, X_{n-1} = j, n \le T_k) &= \mathbb{P}_k(X_n = i \mid X_{n-1} = j, n \le T_j)\, \mathbb{P}_k(X_{n-1} = j, n \le T_k) \\ &= P(j,i)\mathbb{P}_k(X_{n-1} = j, n \le T_k) \,,\end{aligned}$$

where for the second equality we used the Markov property. Plugging this back into (8.1) we deduce

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

thus proving that ν_k is an invariant measure.

To show that $0 < \nu_k(i) < \infty$, let m, n be such that $p_{ki}(n) > 0$ and $p_{ik}(m) > 0$ (which exist by irreducibility). Then using the invariance of ν_k we get

$$\nu_k(i) \ge \nu_k(k) p_{ki}(n) = p_{ki}(n) > 0,$$

since $\nu_k(k) = 1$. To prove the finiteness, we use the invariance of ν_k at k, i.e.

$$1 = \nu_k(k) \ge p_{ik}(m)\nu_k(i) \Rightarrow \nu_k(i) \le \frac{1}{p_{ik}(m)} < \infty$$

and this concludes the proof. $\square$

Theorem 8.9. *Suppose that P is an irreducible transition matrix and let λ be an invariant measure satisfying $\lambda_k = 1$. Then $\lambda \geq \nu_k$.*

If P is also recurrent, then $\lambda = \nu_k$.

Proof. Since λ is invariant we get

$$\begin{aligned}
\lambda_i &= P(k,i) + \sum_{j_1 \neq k} \lambda_j P(j,i) = P(k,i) + \sum_{j_1 \neq k} P(k,j_1)P(j_1,i) + \sum_{j_1,j_2 \neq k} P(j_2,j_1)P(j_1,i)\lambda_{j_2} \\
&= P(k,i) + \sum_{j_1 \neq k} P(k,j_1)P(j_1,i) + \cdots + \sum_{j_1,\dots,j_{n-1} \neq k} P(k,j_{n-1}) \cdots P(j_2,j_1)P(j_1,i) \\
&\qquad + \sum_{j_1,\dots,j_n \neq k} \lambda_{j_n} P(j_n,j_{n-1}) \cdots P(j_2,j_1)P(j_1,i).
\end{aligned}$$

Since λ is a measure, it follows that $\lambda_x \geq 0$ for all x, and hence for all $i \neq k$

$$\begin{aligned}
\lambda_i &\geq P(k,i) + \sum_{j_1 \neq k} P(k,j_1)P(j_1,i) + \cdots + \sum_{j_1,\dots,j_{n-1} \neq k} P(k,j_{n-1}) \cdots P(j_2,j_1)P(j_1,i) \\
&= \mathbb{P}_k(X_1 = i, T_k \geq 1) + \mathbb{P}_k(X_2 = i, T_k \geq 2) + \cdots + \mathbb{P}_k(X_n = i, T_k \geq n) \\
&= \sum_{\ell=1}^{n} \mathbb{P}_k(X_\ell = i, T_k \geq \ell) \to \sum_{\ell=1}^{\infty} \mathbb{P}_k(X_\ell = i, T_k \geq \ell) = \nu_k(i),
\end{aligned}$$

thus proving that $\lambda_i \geq \nu_k(i)$ for all i.

Suppose next that P is also recurrent. Then ν_k is an invariant measure by Theorem 8.8. So $\lambda - \nu_k$ is an invariant measure, since $\lambda \geq \nu_k$. It also satisfies that $\lambda_k - \nu_k(k) = 0$. By irreducibility there exists $m > 0$ such that $p_{ik}(m) > 0$. Using the invariance property we get

$$0 = \lambda_k - \nu_k(k) = \sum_j (\lambda_j - \nu_k(j)) p_{jk}(m) \geq (\lambda_i - \nu_k(i)) p_{ik}(m),$$

which implies that $\lambda_i = \nu_k(i)$ and this finishes the proof. □

So far we have established that if P is irreducible and recurrent, then it has a unique invariant measure up to multiplicative constants. The question is when we can get an invariant distribution out of an invariant measure. In order to get a distribution, the total mass of the invariant measure has to be finite. So let us fix $k \in I$ and consider the invariant measure ν_k. Then we have

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

We thus see that in order to be able to normalise and get an invariant distribution, we require $\mathbb{E}_k[T_k] < \infty$. This leads us to the following definition.

Definition 8.10. Let $i \in I$ be a recurrent state i.e. if $T_i = \inf\{n \geq 1 : X_n = i\}$ is the first return time to i, then $\mathbb{P}_i(T_i < \infty) = 1$. We call i **positive recurrent** if T_i also has finite expectation, i.e.

$$\mathbb{E}_i[T_i] < \infty.$$

If $\mathbb{E}_i[T_i] = \infty$, then i is called **null recurrent**.

Theorem 8.11. *Suppose that P is an irreducible transition matrix. 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 the above holds, then $\pi_i = \frac{1}{\mathbb{E}_i[T_i]}$ for every i.

Proof. Obviously (i) implies (ii). Suppose now that (ii) holds. We will prove that (iii) holds too. Suppose that k is the positive recurrent state and consider the invariant measure ν_k. Then

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

since k is positive recurrent. So if we define for every i

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

then π is an invariant distribution.

Suppose next that (iii) holds and let k be a state. We want to show that k is positive recurrent. First we show that $\pi_k > 0$. Let $i \in I$ be such that $\pi_i > 0$ and by irreducibility there exists $n \geq 0$ such that $p_{ik}(n) > 0$. By stationarity of π we get

$$\pi_k = \sum_{j \in I} \pi_j p_{jk}(n) \geq \pi_i p_{ik}(n) > 0.$$

So we can define a new invariant measure $\lambda_i = \pi_i / \pi_k$ for every $i \in I$. Also $\lambda_k = 1$, and hence by Theorem 8.9 we get that $\lambda \geq \nu_k$, which implies

$$\mathbb{E}_k[T_k] = \sum_{i \in I} \nu_k(i) \leq \sum_{i \in I} \lambda_i = \frac{1}{\pi_k} < \infty,$$

since $\pi_k > 0$. This proves that k is positive recurrent.

For the last part, let $k \in I$. Then k is positive recurrent by (i), and therefore also recurrent. So the measure λ that we defined above must be equal to ν_k by Theorem 8.9. We thus obtain that

$$\sum_{i \in I} \frac{\pi_i}{\pi_k} = \sum_{i \in I} \nu_k(i) = \mathbb{E}_k[T_k],$$

which using that π is a distribution gives

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

and completes the proof. $\square$

Corollary 8.12. *Suppose that P is an irreducible matrix and it has an invariant distribution π. Then for all x, y we have*

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

Example 8.13. Let X be a simple symmetric random walk on $\mathbb{Z}$, i.e.

$$P(x, x+1) = P(x, x-1) = \frac{1}{2} \quad \text{for all } x \in \mathbb{Z}.$$

It is immediate to check that $\pi_i = 1$ for all $i \in \mathbb{Z}$ is an invariant measure, since

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

By Theorem 8.9, since P is a recurrent walk, we get that all invariant measures are multiples of π. We thus deduce that X is not positive recurrent, since $\sum_{i \in \mathbb{Z}} \pi_i = \infty$, and hence we cannot normalise to obtain a probability distribution.

Remark 8.14. We note that the existence of an invariant measure does not imply recurrence. Indeed, let X be a simple symmetric random walk on $\mathbb{Z}^3$. Then $\pi_i = 1$ for all $i \in \mathbb{Z}^3$ is an invariant measure, but X is transient as we have already showed.

Example 8.15. Let X be an asymmetric random walk on $\mathbb{Z}$ with transition probabilities

$$P(x, x-1) = q \quad \text{and} \quad P(x, x+1) = p$$

with $p > q$ and $p + q = 1$. Writing down the equations that an invariant measure π should satisfy we get

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

which we can solve to get

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

So here uniqueness up to multiplicative constants does not hold.

Example 8.16. Let X be a random walk on $\mathbb{Z}_+$ with transition probabilities

$$P(x, x-1) = q > p = P(x, x+1) \quad \text{for } x \geq 1$$

and $p + q = 1$. Suppose also that $P(0,1) = p$ and $P(0,0) = q$. We look for an invariant distribution by solving $\pi = \pi P$. We have

$$\begin{aligned}
&\pi_0 = q\pi_1 + q\pi_0 \\
&\pi_k = p\pi_{k-1} + q\pi_{k+1} \quad \text{for all } k \geq 1.
\end{aligned}$$

Solving the above system we get

$$\pi_1 = \pi_0 \cdot \frac{p}{q} \quad \text{and} \quad \pi_k = \left(\frac{p}{q}\right)^k \cdot \pi_0.$$

So we can normalise π in order to get a probability distribution by taking $\pi_0 = 1 - p/q$. We then get

$$\pi_k = \left(\frac{p}{q}\right)^k \cdot \left(1 - \frac{p}{q}\right).$$

Since we found an invariant distribution, it follows that X is positive recurrent.

9 Time reversibility

We start with a lemma which says that if a Markov chain is started from stationarity and we reverse time, then we obtain another Markov chain with the same invariant distribution.

Proposition 9.1. *Let X be a Markov chain with transition matrix P which is irreducible and has invariant distribution π. Fix $N \in \mathbb{N}$ and suppose that $X_0 \sim \pi$. Then the process $(Y_n)_{0 \le n \le N}$ defined via $Y_n = X_{N-n}$ is a Markov chain with transition matrix $\widehat{P}$ given by*

$$\widehat{P}(x, y) = \frac{\pi(y)}{\pi(x)} P(y, x) \quad \text{for all } x, y.$$

Moreover, $\widehat{P}$ is irreducible and has invariant distribution π.

Proof. First we check that $\widehat{P}$ is indeed a transition matrix. We have

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

where for the second equality we used that π is invariant, i.e. that $\pi = \pi P$.

We next show that Y is a Markov chain. Let $y_0, \ldots, y_N \in I$. Then

$$\begin{aligned} \mathbb{P}(Y_0 = y_0, \ldots, Y_n = y_N) &= \mathbb{P}(X_0 = y_N, \ldots, X_N = y_0) \\ &= \pi(y_N) P(y_N, y_{N-1}) \cdots P(y_1, y_0) \\ &= \pi(y_0) \widehat{P}(y_0, y_1) \cdots \widehat{P}(y_{N-1}, y_N), \end{aligned}$$

which shows that Y is Markov$(\pi, \widehat{P})$.

We check now that π is invariant for $\widehat{P}$. Indeed, we have

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

since P is a stochastic matrix.

Finally, to prove that $\widehat{P}$ is irreducible, let x, y be two states. Then there exists a sequence of states $x_0 = x, x_1, \ldots, x_k = y$ such that $P(x_0, x_1) \cdots P(x_{k-1}, x_k) > 0$. But

$$\widehat{P}(x_k, x_{k-1}) \cdots \widehat{P}(x_1, x_0) = \pi(x_0) P(x_0, x_1) \cdots P(x_{k-1}, x_k) / \pi(x_k)$$

$\widehat{P}(x_k, x_{k-1}) \cdots \widehat{P}(x_1, x_0) > 0$, which shows that $\widehat{P}$ is also irreducible. $\square$

In the previous proposition we saw that if $X_0 \sim \pi$ and we reverse time, then we obtain a Markov chain again with a different transition matrix but the same invariant distribution. So in general, the matrices P and $\widehat{P}$ are not equal. In the particular case that they are equal, then we say that X is time reversible.

Definition 9.2. We say that X with transition matrix P and invariant distribution π is **time reversible**, if $\widehat{P} = P$. By the definition of $\widehat{P}$, we see that X is time reversible, if for all x, y we have

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

These equations are called **the detailed balance equations**.

Equivalently, X is called time-reversible if for any fixed $N \in \mathbb{N}$ whenever $X_0 \sim \pi$ then

$$(X_0, \ldots, X_N) \stackrel{d}{=} (X_N, \ldots, X_0).$$

Remark 9.3. It is important to emphasise that in the definition of reversibility we start the chain from π. Indeed, since we want the two vectors $(X_0, \ldots, X_N)$ and $(X_N, \ldots, X_0)$ to have the same distribution, this would not be possible if say $X_0 \sim \delta_x$.

Remark 9.4. An intuitive way of thinking of time reversibility is by imagining watching a video of the Markov chain started from stationarity. Then reversibility means that whether we watch the video forwards or backwards in time, what we see is indistinguishable statistically, i.e. we cannot tell the two movies apart.

Lemma 9.5. *Suppose that μ is a distribution satisfying*

$$\mu(x)P(x,y) = \mu(y)P(y,x) \quad \textit{for all } x, y.$$

Then μ is an invariant distribution for P.

Proof. Taking the sum over all x of both sides of the equality of the statement we get

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

where we used that P is a stochastic matrix. Therefore, $\mu = \mu P$, which means that μ is an invariant distribution. □

Remark 9.6. From the lemma above we see that if we find a solution to the detailed balance equations, i.e. we find a distribution μ satisfying

$$\mu(x)P(x,y) = \mu(y)P(y,x) \quad \text{for all } x, y,$$

then μ is necessarily invariant for P, and in particular this implies that P is time reversible. So when looking for an invariant distribution, we should first check whether there is a solution to the detailed balance equations, since this is much easier than trying to solve $\pi = \pi P$. Of course, if there is no solution to detailed balance, it does not mean that P does not have an invariant distribution. All it means is that even if P has an invariant distribution, then it is not time-reversible.

Example 9.7. Consider a biased random walk X on the cycle $\mathbb{Z}_n$ with transition probabilities

$$P(i,(i+1) \bmod n) = \frac{2}{3} = 1 - P(i,(i-1) \bmod n), \quad \text{for all } i \in \{0,\dots,n-1\}.$$

Then $\pi_i = 1/n$ for all i is clearly an invariant distribution, but P is not reversible, because the detailed balance equations are not satisfied. Thinking of the intuitive explanation of reversibility, we see that if we start X from π, then all states are equally likely, and if we run the chain forwards in time we will observe a rightwards drift, while if we run it backwards in time, we will observe a leftwards drift. This means that the two are not indistinguishable.

Example 9.8. Consider a biased random walk X on $\{0,1,\dots,n-1\}$ with transition probabilities

$$P(i,i+1) = \frac{2}{3} = 1 - P(i,i-1) \quad \text{for } i \in \{1,\dots,n-2\}$$

and $P(0,1) = 2/3 = 1 - P(0,0)$ and $P(n,n-1) = 1/3 = 1 - P(n,n)$. Then solving the detailed balance equations one gets that $\lambda_i = 2^i$ is an invariant measure which can be normalised, and hence that X is reversible. The difference with the previous example, is that the invariant distribution here is concentrated on the right side, and hence starting the chain according to π we will observe the chain bouncing off the endpoint $n-1$, and this will not be indistinguishable between forwards or backwards in time.

Example 9.9 (Random walk on a graph)**.** Let $G = (V,E)$ be a finite connected graph with V being the set of vertices and E the set of edges. A simple random walk on G is a Markov chain with transition matrix

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

where $d(x)$ is the degree of x, i.e. the total number of edges having x as an endpoint. Since G is connected, it follows that P is irreducible. We now look for an invariant distribution by solving the detailed balance equations:

$$\pi(x)P(x,y) = \pi(y)P(y,x) \quad \text{for all } x,y.$$

For $(x,y) \in E$ we get

$$\pi(x) \cdot \frac{1}{d(x)} = \pi(y) \cdot \frac{1}{d(y)}.$$

So we see that taking $\nu(x) = d(x)$ for every x gives an invariant measure. Normalising it (since the graph is finite), we get an invariant distribution π given by

$$\pi(x) = \frac{d(x)}{\sum_{y \in V} d(y)} = \frac{d(x)}{2|E|} \quad \text{for all } x \in V.$$

10 Convergence to equilibrium

Recall the theorem below from Section 8.

Theorem 10.1. *Let I be a finite state space and suppose that there exists $i \in I$ such that for all j*

$$p_{ij}(n) \to \pi_j \quad \textit{as } n \to \infty.$$

Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

The question we are interested in is the conditions under which the n-th power of a transition matrix converges to the invariant distribution. Let us start with an example.

Consider a simple random walk on $\mathbb{Z}_n$ with transition probabilities

$$P(i, (i+1) \bmod n) = \frac{1}{2} = P(i, (i-1) \bmod n) \quad \text{for all } i \in \mathbb{Z}_n.$$

The invariant distribution is $\pi(i) = 1/n$ for all i. We see that starting from 0, the random walk is back at 0 only at even times. So if we look at the limit $P^k(0,0)$ as $k \to \infty$, then it would not converge. This leads us to the definition of the period of a state.

Definition 10.2. Let P be a transition matrix and i a state. We define the period of i to be

$$d_i = \text{g.c.d.}\{n \geq 1 : P^n(i,i) > 0\}.$$

We say that i is **aperiodic** if $d_i = 1$.

Lemma 10.3. *Let P be a transition matrix and i a state. Then $d_i = 1$ if and only if $P^n(i,i) > 0$ for all n large enough.*

Proof. It is clear that if $P^n(i,i) > 0$ for all n sufficiently large, then $d_i = 1$.

Suppose next that $d_i = 1$ and set $D(i) = \{n \geq 1 : P^n(i,i) > 0\}$. We first show that $D(i)$ contains two consecutive integers. First of all, if $m, n \in D(i)$, then also $m + n \in D(i)$. Suppose now that the minimum distance between any two elements of $D(i)$ is $r \geq 2$ and let $n, m \in D(i)$ with $n = m + r$. Let also $k \in D(i)$ with $k = \ell r + s$ with $\ell \in \mathbb{N}$ and $0 < s < r$, which exists since otherwise only multiples of r would be contained in $D(i)$ and this would contradict that g.c.d.$(D(i)) = 1$. Letting $a = (\ell + 1)n$ and $b = (\ell + 1)m + k$ we see that both a and b are contained in $D(i)$ and

$$a - b = r - s \in (0, r)$$

and this contradicts the definition of r as the minimum distance between any two elements of $D(i)$. Therefore, $D(i)$ contains two consecutive numbers $n_1, n_1 + 1$. Since $an_1 + b(n_1 + 1) \in D(i)$ for all a, b it easily follows that $D(i)$ contains all n sufficiently large ($n \geq n_1^2$) and this completes the proof. □

Lemma 10.4. *Let P be an irreducible transition matrix and let i be an aperiodic state. Then all states are aperiodic.*

Proof. Let j be another state. Since i and j communicate there exist $m, n \geq 0$ such that $P^n(j,i) > 0$ and $P^m(i,j) > 0$. Then for all $s \geq 0$ sufficiently large we have

$$P^{s+n+m}(j,j) \geq P^n(j,i)P^s(i,i)P^m(i,j) > 0,$$

which shows that $P^\ell(j,j)$ is positive for all ℓ sufficiently large, and hence j is also aperiodic. $\square$

Theorem 10.5 (Convergence to equilibrium)**.** *Suppose that P is irreducible and aperiodic and has an invariant distribution π. Let $X \sim \text{Markov}(\lambda, P)$, where λ is a distribution. Then for all y we have*

$$\lim_{n\to\infty} \mathbb{P}(X_n = y) = \pi(y).$$

In particular, for all x, y taking $\lambda = \delta_x$ we have

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

Proof. Let $(Y_n)_{n\geq 0} \sim \text{Markov}(\pi, P)$ be independent of X. We now consider the process $((X_n, Y_n))_{n\geq 0}$ which is clearly a Markov chain on $I \times I$ with initial distribution $\lambda \times \pi$ and transition matrix $\widetilde{P}$ given by

$$\widetilde{P}((x_1,x_2),(y_1,y_2)) = P(x_1,y_1)P(x_2,y_2).$$

We first claim that $\widetilde{P}$ is irreducible. Let (x,y) and (x',y') be two states in $I \times I$. By irreducibility of P we have that there exist m and ℓ such that $P^m(x,x') > 0$ and $P^\ell(y,y') > 0$. By aperiodicity of P for all n large enough we have that

$$P^n(x,x') \geq P^m(x,x')P^{n-m}(x',x') > 0 \quad \text{and} \quad P^n(y,y') \geq P^\ell(y,y')P^{n-\ell}(y',y') > 0.$$

Therefore, this proves that for all n large enough $\widetilde{P}^n((x,y),(x',y')) > 0$, which means that $\widetilde{P}$ is irreducible.

Let $a \in I$ and we define

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

Then T is a stopping time for the Markov chain (X, Y). We now show that $\mathbb{P}(T < \infty) = 1$. It is immediate to check that $\widetilde{\pi}$ given by

$$\widetilde{\pi}(x,y) = \pi(x)\pi(y) \quad \text{for all } x, y$$

is an invariant distribution for $\widetilde{P}$. Theorem 8.11 now implies that $\widetilde{P}$ is positive recurrent, which means in particular that the state (a,a) is recurrent, and hence using also Theorem 7.9 we get $\mathbb{P}(T < \infty) = 1$.

We next define a new process $(Z_n)_{n\geq 0}$ as follows

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

We claim that Z is Markov(λ, P). Since $T \geq 1$ by definition, for all $x \in I$ we have

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

We now show that Z is a Markov chain with transition matrix P. Let $A = \{Z_{n-1} = z_{n-1}, \ldots, Z_0 =$

$z_0\}$. We have

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

The event $\{T>n\}$ is the complement of $\{T\le n\}$ which since T is a stopping time only depends on $(X_0,Y_0),\ldots,(X_n,Y_n)$. Therefore, we obtain

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

Similarly,

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

Therefore, we obtain

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

which shows that $Z\sim \mathrm{Markov}(\lambda,P)$.

We can now finish the proof. Let $y\in I$. Since X and Z have the same distribution and Y is a stationary chain we have

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

Since $\mathbb{P}(T<\infty)=1$, we have that $\mathbb{P}(T>n)\to 0$ as $n\to\infty$ and this completes the proof. □

Theorem 10.6. *Suppose that P is irreducible, null-recurrent and aperiodic transition matrix. Then for all x,y we have*

$$P^n(x,y)\to 0 \quad \text{as } n\to\infty.$$

Proof. This proof follows [1, Theorem 21.29].

Consider the transition matrix $\widetilde{P}((x,y),(x',y'))=P(x,x')P(y,y')$ as in the proof of Theorem 10.5. As in the proof of Theorem 10.5 we get that $\widetilde{P}$ is irreducible, since P is assumed to be irreducible and aperiodic. If $\widetilde{P}$ is transient, then the statement of the theorem follows, since in this case for every $(x,y)\in I\times I$ we get

$$\sum_n \widetilde{P}^n((x,x),(y,y))=\sum_n (P^n(x,y))^2<\infty,$$

which implies that $P^n(x,y)\to 0$ as $n\to\infty$.

So we suppose now that $\widetilde{P}$ is recurrent. Fix $y \in I$ and consider the measure ν_y defined by

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

Then since P is irreducible and recurrent, it follows from Theorem 8.8 that ν_y is invariant for P. Since P is null-recurrent, it means that $\nu_y(I) = \mathbb{E}_y[T_y] = \infty$. Fix $M > 0$. Then there exists a finite set A such that $\nu_y(A) > M$, which follows from the fact that $\nu_y(I) = \infty$. Consider now a new measure μ defined as follows

$$\mu(x) = \frac{\nu_y(x)}{\nu_y(A)}\mathbf{1}(x \in A).$$

Then μ is a probability measure and by the invariance of ν_y ($\nu_y = \nu_y P^n$ for all n) we have

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

We now let (X, Y) be a Markov chain started according to $\mu \times \delta_x$ and with transition matrix $\widetilde{P}$. As in the proof of Theorem 10.5 we let

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

Then T is a finite stopping time with probability 1 (since $\widetilde{P}$ is recurrent) and defining

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

we see as in Theorem 10.5 that Z is a Markov chain started according to μ and with transition matrix P. Therefore, we obtain

$$\mathbb{P}(Z_n = y) = \mu P^n(y) \leq \frac{\nu_y(y)}{\nu_y(A)} = \frac{1}{\nu_y(A)} < \frac{1}{M}. \tag{10.1}$$

We can now finish the proof, since

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

Using the bound (10.1) and taking the limit as $n \to \infty$ we get

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

since $T < \infty$ with probability 1. Since this holds for any $M > 0$, the statement of the theorem follows. $\square$

11 Ergodic theorem

Theorem 11.1. *Let P be an irreducible and positive recurrent matrix with invariant distribution π. Suppose that X starts according to a distribution λ. Then for all x we have almost surely*

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{1}(X_i = x) \to \pi(x) \qquad \textit{as } n \to \infty.$$

Proof. We write $V_n(x) = \sum_{i=0}^{n-1}\mathbb{1}(X_i = x)$ for the total number of visits to x up to time $n-1$.

Since P is recurrent, it follows that $T_x < \infty$ with probability 1 and by the strong Markov property the process $(X_{T_x+n})_{n\geq 0}$ is Markov(δ_x, P) independent of $X_0, \ldots, X_T$. Since $T_x < \infty$, we get that the limit $\lim V_n(x)/n$ is the same when the initial distribution is λ and δ_x. So it suffices to consider the case when $\lambda = \delta_x$.

We recall the definition of the successive times at which X visits x: first we set $T_x^{(0)} = 0$ and for $k \geq 1$ inductively we set

$$T_x^{(k+1)} = \inf\{n \geq T_x^{(k)} + 1 : X_n = x\}.$$

We also define the successive excursion lengths from x by setting

$$S_x^{(k)} = \begin{cases} T_x^{(k)} - T_x^{(k-1)} & \text{if } T_x^{(k-1)} < \infty \\ 0 & \text{otherwise} \end{cases}.$$

By the definition of the return times we have

$$T_x^{(V_n(x)-1)} \leq n-1$$

and equivalently

$$S_x^{(1)} + \ldots + S_x^{(V_n(x)-1)} \leq n-1.$$

Similarly,

$$T_x^{(V_n(x))} \geq n$$

and equivalently

$$S_x^{(1)} + \ldots + S_x^{(V_n(x))} \geq n.$$

By the strong Markov property, the excursion lengths $(S_x^{(k)})_{k\geq 1}$ are i.i.d. with expectation given by $m_x = \mathbb{E}_x[T_x] = 1/\pi(x)$. The strong law of large numbers then asserts that almost surely

$$\frac{S_x^{(1)} + \ldots + S_x^{(k)}}{k} \to m_x \qquad \text{as } k \to \infty.$$

We now have

$$S_x^{(1)} + \ldots + S_x^{(V_n(x)-1)} \leq n \leq S_x^{(1)} + \ldots + S_x^{(V_n(x))}.$$

Dividing through by $V_n(x)$ we obtain

$$\frac{S_x^{(1)} + \ldots + S_x^{(V_n(x)-1)}}{V_n(x)} \leq \frac{n}{V_n(x)} \leq \frac{S_x^{(1)} + \ldots + S_x^{(V_n(x))}}{V_n(x)},$$

and hence taking the limit as $n \to \infty$ we deduce

$$\lim_{n\to\infty} \frac{V_n(x)}{n} = \frac{1}{m_x} = \pi(x)$$

and this concludes the proof. □

References

[1] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[2] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.