ADVANCED PROBABILITY THEORY

Part 2: Markov chains

Arvind Singh[1]

Preliminary Draft.

Tsinghua University

April 2017

[1]LMO. Université Paris-Sud. F-91405 Orsay Cedex.
arvind.singh@math.u-psud.fr
https://www.math.u-psud.fr/~singh/

Contents

Chapter 1

Markov Chains

This chapter gives an introduction to time-homogeneous Markov chains on a countable state space. Informally speaking, these are discrete time stochastic processes that enjoy the important property that "the future of the process depends on its past only through its present state". Together with martingales, they form the two most important classes of stochastic processes and are ubiquitous throughout Probability theory.

References

[1] Rick Durrett. *Probability Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics 31, 4th edition, 2010.

[2] Jean-François Le Gall. *Intégration, Probabilité et Processus Aléatoires.* Lecture notes (in French!) Available at https://www.math.u-psud.fr/~jflegall/IPPA2.pdf.

1. Definition and first properties

In everything that follows, E will denote a **finite or countably infinite** set endowed with the discrete topology. We denote by $\mathcal{M}(E)$ the set of finite measures on E and by $\mathcal{P}(E)$ the subset of probability measures.

Definition 1.1: transition kernel

A function $p : E \times E \to \mathbb{R}$, which we write $(p(x, y),\ x, y \in E)$, is said to be a **transition kernel** (also called a transition matrix) if

1. $0 \leq p(x, y) \leq 1$ for all $x, y \in E$;
2. $\sum_{y \in E} p(x, y) = 1$ for all $x \in E$.

We say that $p(x, y)$ is the **probability of going from state** x **to state** y.

When the set E is finite, p is simply a **stochastic matrix** *i.e.* a square matrix with non-negative entries such that each row sums up to 1. Just as for matrices, one can define the multiplication of a kernel by a left row vector (interpreted as a measure), a right column vector (interpreted as a function) and multiplications of kernels together (composition).

Right multiplication. The definition above is equivalent to the assertion that for each $x \in E$, the function $y \mapsto p(x, y)$ defines a probability density on $(E, \mathcal{P}(E))$. Transition kernel acts on real-valued function via integration: for any $f : E \to \mathbb{R}_+$, we define $p.f : E \mapsto \mathbb{R}_+$ by

$$(p.f)(x) := \sum_{y \in E} p(x, y) f(y) \quad \text{for all } x \in E. \tag{1.1}$$

We extend definition (1.1) to any $f : E \to \mathbb{R}$ such that $\sum_y p(x, y)|f(y)| < \infty$ for all $x \in E$.

Left multiplication. Given a measure $\mu = (\mu(x), x \in E) \in \mathcal{M}(E)$, we define the measure $\mu.p \in \mathcal{M}(E)$ to be the image of μ via the transition kernel p:

$$(\mu.p)(y) := \sum_{x \in E} \mu(x) p(x, y) \quad \text{for all } y \in E$$

where, for the sake of simplicity, we use the notation $\mu(y)$ in place of $\mu(\{y\})$. The total mass of the measure is preserved:

$$(\mu.p)(E) = \sum_{y \in E} (\mu.p)(y) = \sum_y \sum_x \mu(x) p(x, y) = \sum_x \mu(x) \sum_y p(x, y) = \sum_x \mu(x) = \mu(E).$$

In particular, this operation induces a mapping $\mu \mapsto \mu.p$ from $\mathcal{P}(E)$ onto itself.

Composition of kernels. Given two transition kernels p and q, we can define another kernel denoted pq by

$$(pq)(x, y) := \sum_{z \in E} p(x, z) q(z, y) \quad \text{for all } x, y \in E. \tag{1.2}$$

To see that pq is indeed a kernel, we check that $(pq)(x, y) \geq 0$ (trivial) and that

$$\sum_{y \in E} (pq)(x, y) = \sum_{y \in E} \sum_{z \in E} p(x, z) q(z, y) = \sum_{z \in E} p(x, z) \sum_{y \in E} q(z, y) = \sum_{z \in E} p(x, z) = 1.$$

The identity transition kernel I is defined by $I(x, y) = \mathbf{1}_{\{x=y\}}$. For $n \in \mathbb{N}$, we define p^n by induction:

$$\begin{cases} p^0 & := I, \\ p^{n+1} & := p^n p = p p^n. \end{cases}$$

For any $n, m \geq 0$, we have $p^{n+m} = p^n p^m = p^m p^n$ which yields the so-called **Chapman–Kolmogorov equation** for transition kernels:

$$p^{n+m}(x, y) = \sum_{z \in E} p^n(x, z)\, p^m(z, y).$$

Moreover, it follows by a trivial induction that

$$p^n(x, y) = \sum_{z_1, \ldots, z_{n-1} \in E} p(x, z_1)\, p(z_1, z_2) \ldots p(z_{n-1}, y). \tag{1.3}$$

Definition 1.2: Markov chain

Let p be a transition kernel on E. A sequence of random variables $(X_n)_{n\geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and taking values in E is said to be a **Markov chain with transition probability** p if, for any integer $n \geq 0$ and any $y \in E$, we have

$$\mathbf{P}(X_{n+1} = y \,|\, X_0, \dots, X_n) = p(X_n, y), \qquad \text{a.s.} \tag{1.4}$$

or equivalently,

$$\mathbf{P}(X_{n+1} = y \,|\, X_0 = x_0, \dots, X_n = x_n) = p(x_n, y), \tag{1.5}$$

for all $x_0, x_1, \dots, x_n \in E$ such that $\mathbf{P}(X_0 = x_0, \dots, X_n = x_n) > 0$.

Remark 1.3

(a) Formula (1.5) involves conditional probability whereas the Formula (1.4) involves a conditional expectation: $\mathbf{P}(A|X_0, \dots, X_n) := \mathbf{E}[\mathbf{1}_A|X_0, \dots, X_n]$.

(b) The definition above is that of a **time-homogeneous** Markov chain. One can also define **time-inhomogeneous** Markov chains by allowing the kernel to depend also on n but we will not consider this case here.

(c) For a general process, the conditional law of X_{n+1} with respect to $X_0, \dots, X_n$ depends, a priori, on all these variables. But, when X is a Markov chain, this conditional law depends, in fact, only on X_n. In particular, we can write

$$\begin{aligned} \mathbf{P}(X_{n+1} = y \,|\, X_n) &= \mathbf{E}[\mathbf{P}(X_{n+1} = y \,|\, X_0, \dots, X_n) \,|\, X_n] && (1.6) \\ &= \mathbf{E}[p(X_n, y) \,|\, X_n] && (1.7) \\ &= p(X_n, y) && (1.8) \\ &= \mathbf{P}(X_{n+1} = y \,|\, X_0, \dots, X_n). && (1.9) \end{aligned}$$

This shows that the knowledge of the whole trajectory up to time n provides no useful information over knowing only the current position in order to predict the next value of the chain. This formalizes the idea that **the future of the process depends only on its past through its present position**.

(d) Definition 1.2 makes no assumption on X_0 which may be deterministic or random.

Proposition 1.4

A process $(X_n)_{n\geq 0}$ taking values in E is a Markov chain with transition kernel p if and only if for all $n \geq 0$ and all $x_0, x_1, \cdots, x_n \in E$, we have

$$\mathbf{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n) \;=\; \mathbf{P}(X_0 = x_0)\, p(x_0, x_1)\, p(x_1, x_2) \cdots p(x_{n-1}, x_n). \tag{1.10}$$

In particular, if $\mathbf{P}(X_0 = x_0) > 0$, then

$$\mathbf{P}(X_n = x_n \,|\, X_0 = x_0) = p^n(x_0,\, x_n).$$

Proof. Assume that X is a Markov chain. We have

$$\begin{aligned}
&\mathbf{P}(X_0 = x_0, \cdots, X_n = x_n,\, X_{n+1} = x_{n+1}) \\
&\qquad = \mathbf{P}(X_0 = x_0, \cdots, X_n = x_n)\, \mathbf{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0, \cdots, X_n = x_n). \\
&\qquad = \mathbf{P}(X_0 = x_0, \cdots, X_n = x_n)\, p(x_x, x_{n+1}).
\end{aligned}$$

and formula (1.10) follows by induction from this recurrence relation. Conversely assume that (1.10) holds and that $\mathbf{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n) > 0$. Then, for any y,

$$\begin{aligned}
&\mathbf{P}(X_{n+1} = y \,|\, X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n) \\
&\qquad = \frac{\mathbf{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n,\, X_{n+1} = y)}{\mathbf{P}(X_0 = x_0,\, X_1 = x_1,\, \cdots, X_n = x_n)} \\
&\qquad = \frac{\mathbf{P}(X_0 = x_0)\, p(x_0,\, x_1)\, p(x_1,\, x_2) \cdots P(x_{n-1},\, x_n)\, p(x_n,\, y)}{\mathbf{P}(X_0 = x_0)\, p(x_0,\, x_1)\, p(x_1,\, x_2) \cdots p(x_{n-1},\, x_n)} \\
&\qquad = p(x_n,\, y),
\end{aligned}$$

which proves that X is a Markov chain. The last assertion follows from (1.3) and (1.10) because

$$\begin{aligned}
\mathbf{P}(X_n = x_n \,|\, X_0 = x_0) &= \frac{\mathbf{P}(X_0 = x_0,\, X_n = x_n)}{\mathbf{P}(X_0 = x_0)} \\
&= \sum_{x_1, \ldots, x_{n-1} \in E} \frac{\mathbf{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n)}{\mathbf{P}(X_0 = x_0)} \\
&= \sum_{x_1, \ldots, x_{n-1} \in E} p(x_0, x_1) \ldots p(x_{n-1}, x_n) \\
&= p^n(x_0, x_n).
\end{aligned}$$

□

Remark 1.5

The proposition shows that the law of the finite dimensional marginals of a Markov chain are uniquely determined by the transition kernel p together with the law of its starting point X_0. We will see later that, in fact, the whole trajectory of the Markov chain is characterized by p and the law of X_0.

Proposition 1.6

Let $(X_n,\, n \geq 0)$ be a Markov chain with transition kernel p.

(a) For any function $f : E \to \mathbb{R}_+$, we have

$$\mathbf{E}[f(X_{n+1}) \,|\, X_0, \ldots, X_n] = (p.f)(X_n).$$

(b) Let μ_n denote the law of X_n. We have, for all $n, m \geq 0$

$$\mu_{n+m} = \mu_n . p^m.$$

(c) Let $n \geq 0$ and $m \geq 1$. For any $y_1, \cdots, y_m \in E$,

$$\mathbf{P}(X_{n+1} = y_1, \ldots, X_{n+m} = y_m \,|\, X_0, \ldots, X_n) = p(X_n,\, y_1)\, p(y_1,\, y_2) \ldots p(y_{m-1},\, y_m).$$

In particular, if we write $Y_i := X_{n+i}$, then $(Y_i,\, i \geq 0)$ is again Markov chain with transition probability p.

Proof. Assertion (a) is straightforward since, by definition

$$\mathbf{E}[f(X_{n+1}) \,|\, X_0, \ldots, X_n] = \sum_{y \in E} f(x)\mathbf{P}[X_{n+1} = y \,|\, X_0, \ldots, X_n] = \sum_{x \in E} f(x)p(X_n, y) = (p.f)(X_n).$$

To check Assertion (b), we write that, for any $y \in E$,

$$\mu_{n+1}(y) = \mathbf{P}(X_{n+1} = y) = \mathbf{E}[\mathbf{P}(X_{n+1} = y \,|\, X_0, \ldots, X_n)] = \mathbf{E}[p(X_n, y)] = \sum_{x \in E} \mu_n(x)p(x, y) = (\mu_n.p)(y)$$

which proves that $\mu_{n+1} = \mu_n.p$ so the formula holds for $m = 1$ hence for any m by a trivial induction. In view of (1.10), we have

$$\mathbf{P}(X_{n+1} = y_1, \cdots X_{n+m} = y_m \,|\, X_0 = x_0, \cdots, X_n = x_n) = p(x_n,\, y_1)\, p(y_1,\, y_2) \cdots p(y_{m-1},\, y_m),$$

which proves the first part of Assertion (c). Finally, to check that Y is a Markov chain with transition kernel p, we observe that

$$\mathbf{P}(Y_0 = y_0,\, Y_1 = y_1, \cdots, Y_m = y_m) = \mathbf{P}(X_n = y_0)\, p(y_0,\, y_1) \cdots P(y_{m-1},\, y_m),$$

and use the characterization of Markov chains of Proposition 1.4. □

2. Examples of Markov chains

We now give some classical examples of Markov chains.

• **Independent random variables.** Let $(X_n,\, n \geq 0)$ be independent random variables taking values in E, having the same distribution μ. It is easily checked that $(X_n,\, n \geq 0)$ is a Markov chain with transition probability

$$p(x,\, y) := \mu(y), \qquad \forall x,\, y \in E.$$

This is not the most interesting example...

• **Random walks on $\mathbb{Z}^d$.** Here $E := \mathbb{Z}^d$. Let $(\xi_i)_{i \geq 0}$ denote a sequence of i.i.d. random variables with distribution μ. Let also $x_0 \in \mathbb{Z}^d$. We define the **random walk X with step distribution** μ by

$$X_n := x_0 + \sum_{i=1}^{n} \xi_i$$

(with the convention that $\sum_1^0 = 0$ which we shall enforce throughout these lecture notes). This is a Markov chain with transition probability

$$p(x, y) := \mu(y - x), \qquad \forall x,\, y \in \mathbb{R}^d.$$

Indeed, using that ξ_{n+1} is independent of $\sigma(\xi_1, \dots, \xi_n) \supset \sigma(X_0, \dots, X_n)$, we get

$$\begin{aligned} \mathbf{P}(X_{n+1} = y \,|\, X_0 = x_0, \dots, X_n = x_n) &= \mathbf{P}(\xi_{n+1} = y - x_n \,|\, X_0 = x_0, \dots, X_n = x_n) \\ &= \mathbf{P}(\xi_{n+1} = y - x_n) \\ &= \mu(y - x_n). \end{aligned}$$

Special case. Let $e_1, \cdots, e_d$ be unit vectors of $\mathbb{R}^d$. In case $\mu(e_i) = \mu(-e_i) = \frac{1}{2d}$ for $1 \le i \le d$, the corresponding Markov chain is called the **simple random walk** on $\mathbb{Z}^d$. $\square$

• **Random walk on a graph/electrical network.** We assume here that E is equipped with a graph structure.[1] This means that we have a set $\mathcal{E}$ of pairs of sites of E that describes the (unoriented) edges connecting pairs of points in E:

$$(x, y) \in \mathcal{E} \quad \Longleftrightarrow \quad (y, x) \in \mathcal{E} \quad \Longleftrightarrow \quad x \text{ and } y \text{ are connected by an edge in the graph.}$$

For $x \in E$, the degree $\deg(x)$ is the number of site adjacent to x *i.e.*

$$\deg(x) := \sharp\{y \in E,\, (x, y) \in \mathcal{E}\}.$$

We assume that $\deg(x) < \infty$ for all $x \in E$. We define the transition kernel p on E by

$$p(x, y) := \begin{cases} \frac{1}{\deg(x)} & \text{if } (x, y) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

A Markov chain with transition kernel p is called a **simple random walk on the graph** $(E, \mathcal{E})$.

A slightly more general setting is obtained extending the graph structure to an **electrical network**. Suppose that we are given weights (also called conductances) for every edge of the graph. This means that we have a function $c : E \times E \to \mathbb{R}_+$ satisfying

(i) It is symmetric: $c(x, y) = c(y, x)$ for all $x, y \in E$.

(ii) It is supported by the edges of the graph: $c(x, y) > 0$ if and only if $(x, y) \in \mathcal{E}$.

We define $\pi(x) = \sum_{y/(x,y) \in \mathcal{E}} c(x, y)$ the sum of the conductances of the edges around $x \in E$. Assuming that $0 < \pi(x) < \infty$ for all $x \in E$, we can define a transition kernel

$$p(x, y) := \begin{cases} \frac{c(x,y)}{\pi(x)} & \text{if } (x, y) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

A Markov chain with transition kernel p is called a **random walk on the electrical network** $(E, \mathcal{E}, c)$. When c is constant, we recover the simple random walk on the graph.

[1] In this definition, double edges are forbidden but loops are allowed: (x, x) mean that there is an edge joining site x to itself.

Random walks on electrical networks form an important class of Markov chain. We will see later that they are the Markov chains that are **reversible** (informally, this means that running time backwards yields the same Markov chain when started from equilibrium).

• **Branching processes.** These processes model the evolution in time of a population where each individual give birth to a certain number of offspring according to a given reproduction law. Formally, we set $E := \mathbb{N}$ and let μ be a probability measure on E. We define a sequence of random variables $(X_n,\, n \geq 0)$ with $X_0 = x_0 \in \mathbb{N}$ and by induction

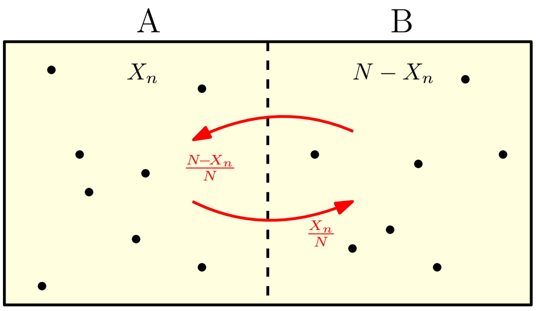
$$X_{n+1} := \sum_{j=1}^{X_n} \xi_{n,j},$$

where $(\xi_{n,j}, n \geq 0, j \geq 1)$ is a family of i.i.d. random variables having the distribution μ. Then $(X_n)_{n\geq 0}$ is a Markov chain on $\mathbb{N}$ with transition kernel

$$p(x,y) := \mu^{*x}(y), \qquad \forall x,\, y \in \mathbb{N},$$

where μ^{*x} denotes the x-th fold convolution of μ, or, in the probabilistic language, the law of the sum of x i.i.d. random variables having the law μ.

• **The Ehrenfest chain.**



This model is of historical relevance, it was proposed by Paul and Tatiana Ehrenfest to explain the irreversibility stated in the second law of thermodynamics. It models the diffusion of a gaz between two containers, say A and B. Here the state space is $E = \{0, \dots, N\}$ where N represents the total number of molecules in the gas. We consider a Markov chain $(X_n)_{n\geq 0}$ taking value in E with transition kernel

$$p(i,j) := \begin{cases} \frac{N-i}{N} & \text{if } j = i+1, \\ \frac{i}{N} & \text{if } j = i-1. \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Then, we can interpret X_n as the number of molecules in the box A at time n (hence $N - X_n$ is the number of molecule in the box B). At each unit of time, we pick a molecule at random uniformly and put it in the opposite box.

• **Pólya urn.** We have an urn that contains balls of two colors: black or white. At each unit of time, we pick a ball inside the urn, uniformly among all balls, we look at its color and then replace it inside the urn together with a new ball of the same color. Mathematically speaking, the state space of the process is $E = \mathbb{N}^{*2}$ and we consider a Markov chain (B_n, W_n) with transition kernel

$$p\big((b,w),\,(b',w')\big) := \begin{cases} \frac{b}{b+w} & \text{if } b' = b+1 \text{ and } w' = w, \\ \frac{w}{b+w} & \text{if } w' = w+1 \text{ and } b' = b, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

Exercice 2.1

Consider a Pólya urn decribed in the example above. Suppose that it starts at time $n = 0$ with one ball of each color *i.e.* $(B_0, W_0) = (1, 1)$. Show, by induction that, for each time $n \geq 0$, the number B_n of black balls is uniformly distributed in $\{1, \dots, n+1\}$. What is the limit of $(\frac{B_n}{n+1}, \frac{W_n}{n+1})$?

Exercice 2.2

Let $(\xi_i)_{i\geq 0}$ denote a sequence of i.i.d. Bernoulli random variables with parameter $p \in (0,1)$.

$$\mathbf{P}(\xi_n = 1) = 1 - \mathbf{P}(\xi_n = 0) = p.$$

We consider the walk X on $\mathbb{Z}$ defined by

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

1. Show that the Markov chain X may be interpreted as a random walk on an electrical network and compute the conductances.
2. Fix $A \in \mathbb{N}^*$ and set $Z_n = X_n \bmod A$ for any $n \geq 0$. Show that Z is again a Markov chain and compute its transition kernel. For which values of p can Z be interpreted as a random walk on some electrical network?

Exercice 2.3

Let $(X_n)_{n\geq 0}$ be a Markov chain on E.

1. Show that, for any $k \in \mathbb{N}^*$, the sub-sequence $(X_{kn})_{n\geq 0}$ is a Markov chain and determine its transition kernel.
2. Let $(Y_i)_{i\geq 1}$ denote a sequence of i.i.d. random variables taking values in $\mathbb{N}^*$. Set $Z_0 := X_0$ and $Z_n := X_{Y_1+\ldots+Y_n}$ for $n \geq 1$. Show that $(Z_n)_{n\geq 0}$ is a Markov chain and determine its transition kernel.

Exercice 2.4

Let $(Y_n)_{n\geq 0}$ be a sequence of i.i.d. Bernoulli random variables with parameter $p \in [0,1]$.

Define by induction

$$\begin{cases} X_0 := 0, \\ X_{n+1} := |Y_n - X_{n-X_n}|. \end{cases}$$

1. Show that $(X_n)_{n\geq 0}$ a Markov chain if and only if $p = 0$ or $p = \frac{1}{2}$.
2. Let $Z_n = (X_n, X_{n+1})$. Show that $Z = (Z_n)_{n\geq 0}$ is a Markov chain for any value of p.

Exercice 2.5

Let $(Y_n)_{n\geq 0}$ be a sequence of i.i.d. Bernoulli random variables with parameter $p \in [0,1]$ *i.e.*

$$\mathbf{P}(Y_n = 1) = 1 - \mathbf{P}(Y_n = 0) = p.$$

Define $X_n := \mathbf{1}_{\{Y_n = Y_{n+1}\}}$. For which value of p is X a Markov chain ?

Exercice 2.6

1. Let $(X_n)_{n\geq 0}$ be a Markov chain taking value in $\mathbb{Z}$. Show that the sequence $(Z_n)_{n\geq 0}$ defined by $Z_{n+1} = |Z_n|$ may, or may not, be a Markov chain.
2. More generally. Let E, F denote two finite or countably infinite sets. Let also $f : E \to F$. Show the equivalence

   For any Markov chain $X = (X_n)_{n\geq 0}$ on E, the random sequence $(f(X_n))_{n\geq 0}$ is a Markov chain on F $\iff$ f is injective or constant.

Exercice 2.7

Let $p \geq 1$. For each $1 \leq i \leq p$, let $(X_n^i)_{n\geq 0}$ denote a Markov chain taking value in some space E_i (all these processes being defined on the same abstract probability space). Show that the sequence $(X_n^1, \ldots, X_n^p)_{n\geq 0}$ is a Markov chain on $E_1 \times \ldots \times E_p$ and compute its kernel.

Exercice 2.8

Let $(Y_n)_{n\geq 0}$ denote a sequence of i.i.d. random variables taking values in some measurable space $(S, \mathcal{S})$. Let $\Phi : E \times S \to E$ be a measurable mapping. Let $x \in E$ and define by induction

$$\begin{cases} X_0 := x, \\ X_{n+1} := \Phi(X_n, Y_{n+1}). \end{cases}$$

Prove that $(X_n)_{n\geq 0}$ is a Markov chain and determine its transition kernel.

▶ **Suggested additional exercises:** Durrett **[2]**, Exercises 6.22 - 6.29.

3. Construction of a Markov chain

We now show that there exists a Markov chain for any transition kernel and any initial distribution. This is a direct consequence of Kolmogorov's extension theorem but we will here give a constructive proof instead.

Proposition 3.1

Let p be a transition kernel on E. There exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ such that

1. For any $x \in E$, there exists a process $(X_n^x,\, n \geq 0)$ defined on $\tilde{\Omega}$ which is a Markov chain with transition kernel p starting from $X_0^x = x$.

2. For any two $x, y \in E$, the following coupling property holds:
$$X_n^x = X_n^y \quad \Longrightarrow \quad (X_m^x = X_m^y \quad \text{a.s. for all } m \geq n.)$$

The proof of the proposition is based on a general result asserting the existence of arbitrary sequences of independent random variables. The proof of the lemma is given in the appendix.

Lemma 3.2

Let $(\nu_i)_{i\in\mathbb{N}}$ be a family of probability measure on E. Consider the probability space $\tilde{\Omega} = [0,1]$ endowed with the Borel σ-field and the Lesbegue measure. There exists a sequence on random variables $\xi_i : \tilde{\Omega} \to E$ such that:

1. The random variables $(\xi_i)_{i\in\mathbb{N}}$ are independent.

2. The law of ξ_i is ν_i.

Proof of Proposition 3.1. Since E is countable, so is $E \times \mathbb{N}$. According to the lemma above, we can construct a sequence $(\xi_{x,i})_{(x,i)\in E\times\mathbb{N}}$ of independent random variables such that $\xi_{x,i}$ is distributed as $p(x,\cdot)$ for any $(x,i) \in E\times\mathbb{N}$. For each $x \in E$, we define the sequence $X^x = (X_n^x)_{n\geq 0}$ by induction
$$\begin{cases} X_0^x & := \quad x \\ X_{n+1}^x & := \quad \xi_{X_n^x, n}. \end{cases} \tag{3.1}$$
In words, this means that, at time n, if the chain is at position x, then we use the random variable $\xi_{x,n}$ to choose the new position at time $n+1$. Thus, if $X_n^x = X_n^y$ for some n, then both processes will use the same ξ's at all later times. This shows that the coupling property 2 holds. Let us check that X^x is a Markov chain. Let $x = x_0, x_1, \ldots, x_n \in E$ such that $\tilde{\mathbf{P}}(X_0^x = x_0, \ldots, X_n^x = x_n) > 0$. Notice that the random variables $X_0, \ldots, X_n$ are $\sigma(\xi_{x,i}, x \in E, i < n)$-measurable. In particular, they are independent of the sequence $(\xi_{x,n})_{x\in E}$ hence
$$\begin{aligned} \tilde{\mathbf{P}}(X_{n+1}^x = y \,|\, X_0^x = x_0, \ldots, X_n^x = x_n) &= \tilde{\mathbf{P}}(\xi_{x_n,n} = y \,|\, X_0^x = x_0, \ldots, X_n^x = x_n) \\ &= \tilde{\mathbf{P}}(\xi_{x_n,n} = y) \\ &= p(x_n, y) \end{aligned}$$

which shows that X^x is indeed a Markov chain with transition kernel p starting, by construction, from $X_0^x = x$. $\square$

Corollary 3.3

For any probability distribution $\nu \in \mathcal{P}(E)$, we can create a Markov chain Y with transition kernel p and initial position Y_0 distributed as ν.

Proof. We can construct another random variable ζ with distribution ν on the same probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ as before which is independent of the Markov chains $(X^x, x \in E)$. Then, we define $Y_n := X_n^\zeta$. The random variable Y_n is measurable for any n and, by construction, Y_0 is distributed according to ν. Let $y_0, \ldots, y_n \in E$. By independence of ζ and X^{y_0} and using the characterization of Proposition 1.4, we find that

$$\begin{aligned}
\tilde{\mathbf{P}}(Y_0 = y_0, \ldots, Y_n = y_n) &= \tilde{\mathbf{P}}(\zeta = y_0, X_0^{y_0} = y_0, X_1^{y_0} = y_1, \ldots, X_n^{y_0} = y_n) \\
&= \tilde{\mathbf{P}}(\zeta = y_0)\tilde{\mathbf{P}}(X_0^{y_0} = y_0, X_1^{y_0} = y_1, \ldots, X_n^{y_0} = y_n) \\
&= \tilde{\mathbf{P}}(Y_0 = y_0)p(y_0, y_1) \ldots p(y_{n-1}, y_n)
\end{aligned}$$

which proves that Y is indeed a Markov chain. $\square$

Exercice 3.4

Modify (3.1) in the proof of Proposition 3.1 in such way that the Markov chains constructed from the ξ's satisfy the following coupling property instead of 2.

- For any $x, y \in E$, we set $T := \inf\{i \in \mathbb{N} \,:\, X_i^x = y\}$ with the convention $\inf \emptyset = \infty$. On the event $\{T < \infty\}$, we have

$$X_{T+n}^x = X_n^y \quad \text{a.s. for all } n \geq 0.$$

Hint: choose the index i of $\xi_{x,i}$ according to the number of previous visits to the current position x.

4. The canonical Markov chain

It is often useful to consider the whole path of a Markov chain *i.e.* consider $(X_n)_{n\geq 0}$ as a single random variable X taking values in the space of trajectories $E^{\mathbb{N}}$. This enables to study functionals of the chain that depend on an unbounded number of steps. Two particularly important examples are return times quantities of the form $T(X) = \inf\{n, X_n \in A\}$ and number of visits of a set such as $N(X) = \sum_n \mathbf{1}_{\{X_n \in A\}}$.

In order to do this, we will apply the usual trick which states that, for any random variable $Z : \tilde{\Omega} \to \Omega$, the identity function $\mathrm{Id} : \Omega \to \Omega$ under the image measure $\mathbf{P}(\cdot) := \tilde{\mathbf{P}}(Z \in \cdot)$ has the same law as Z under $\tilde{\mathbf{P}}$. Our our setting, we are working with the space of trajectories $E^{\mathbb{N}}$ so we must specify the σ-field we will use.

Definition 4.1: canonical set-up

Set $\boldsymbol{\Omega} := E^{\mathbb{N}}$. An element $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ is written in the form $\boldsymbol{\omega} = (\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \ldots)$. The coordinate projections $\mathbf{X}_n$ are defined by

$$\begin{array}{rccc} \mathbf{X}_n : & \boldsymbol{\Omega} & \to & E \\ & \boldsymbol{\omega} & \mapsto & \boldsymbol{\omega}_n \end{array}$$

Let $\boldsymbol{\mathcal{F}}$ be the σ-field generated by the cylinder sets, *i.e.* the subsets $C \subset \boldsymbol{\Omega}$ of the form

$$C = \{\boldsymbol{\omega} \in \boldsymbol{\Omega} : \boldsymbol{\omega}_0 = x_0, \ldots, \boldsymbol{\omega}_n = x_n\} = \bigcap_{i=0}^{n} \mathbf{X}_i^{-1}(x_i)$$

for some $n \geq 0$ and $x_0, \ldots, x_n \in E$. Thus, $\boldsymbol{\mathcal{F}}$ it is the smallest σ-field that makes all the coordinate projection measurable. We also define the canonical filtration $(\boldsymbol{\mathcal{F}}_n)_{n\geq 0}$ of $\mathbf{X}$ by

$$\boldsymbol{\mathcal{F}}_n := \sigma(\mathbf{X}_0, \ldots, \mathbf{X}_n).$$

Lemma 4.2

Let $(G, \mathcal{G})$ denote a measurable space and consider a function $F : G \to \boldsymbol{\Omega}$. Then, F is measurable if and only if $\mathbf{X}_n \circ F$ is measurable for all n.

Proof. If F is measurable then $\mathbf{X}_n \circ F$ is also measurable as a composition of measurable functions. Conversely, consider the sigma-field $\widehat{\boldsymbol{\mathcal{F}}} = \{A \in \boldsymbol{\mathcal{F}},\ F^{-1}(A) \in \mathcal{G}\}$. By hypothesis, it contains all the sets of the form $\mathbf{X}_n^{-1}(x)$ for all n and all x. Hence, it contains $\boldsymbol{\mathcal{F}}$ and therefore $\widehat{\boldsymbol{\mathcal{F}}} = \boldsymbol{\mathcal{F}}$. $\square$

Proposition 4.3: The canonical Markov Chain

Let p be a transition kernel on E. For any $x \in E$, there exists a unique probability measure $\mathbf{P}_x$ on $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}})$ such that the canonical process $\mathbf{X}$ under $\mathbf{P}_x$ is a Markov chain starting from x a.s. and with transition kernel p.

Proof. According to Proposition 3.1, there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ on which we can construct a Markov chain $(X_n^x)_{n\geq 0}$ with transition kernel p starting from $X_0^x = x$ a.s. Consider the mapping

$$\begin{array}{rccc} F_x : & \tilde{\Omega} & \to & \boldsymbol{\Omega} \\ & \tilde{\omega} & \mapsto & \boldsymbol{\omega} = (X_n^x(\tilde{\omega}))_{n\geq 0}. \end{array}$$

This application is measurable thanks to the previous lemma so we can define $\mathbf{P}_x$ as the image of the measure $\tilde{\mathbf{P}}$ by F_x:

$$\mathbf{P}_x(A) := \tilde{\mathbf{P}}(F_x^{-1}(A)) = \tilde{\mathbf{P}}(X^x \in A)$$

By construction, we have $\mathbf{P}_x(\mathbf{X}_0 = x) = \tilde{\mathbf{P}}(X_0^x = x) = 1$ and

$$\begin{aligned} \mathbf{P}_x(\mathbf{X}_0 = x_0, \dots, \mathbf{X}_n = x_n) &= \tilde{\mathbf{P}}(X_0^x = x_0, \dots, X_n^x = x_n) \\ &= \tilde{\mathbf{P}}(X_0^x = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) \\ &= \mathbf{P}_x(\mathbf{X}_0 = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) \end{aligned} \tag{4.1}$$

hence $\mathbf{X}$ is indeed a Markov chain with transition kernel p. It remains to check that this measure is unique. Notice that (4.1) fixes the value of $\mathbf{P}_x$ on the all the cylinder sets. Those sets form a π-system which generate the σ-field $\mathcal{F}$. Thus, Dynkin's $\pi - \lambda$ theorem implies that $\mathbf{P}_x$ is uniquely determined by its value on these cylinder set, hence it is unique. □

Corollary 4.4

Let p be a transition kernel on E and let $\nu \in \mathcal{P}(E)$. There exists a unique probability measure $\mathbf{P}_\nu$ on $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}})$ such that the canonical process $\mathbf{X}$ under $\mathbf{P}_\nu$ is a Markov chain with transition kernel p starting from $\mathbf{X}_0$ distributed as ν.

Proof. We simply set

$$\mathbf{P}_\nu(\cdot) := \sum_{x \in E} \nu(x) \mathbf{P}_x(\cdot) \tag{4.2}$$

which defines a probability on $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}})$. Then, under $\mathbf{P}_\nu$, the process $\mathbf{X}$ has its initial position $\mathbf{X}_0$ distributed as ν and it is a Markov chain with kernel p since

$$\begin{aligned} \mathbf{P}_\nu(\mathbf{X}_0 = x_0, \dots, \mathbf{X}_n = x_n) &= \sum_{x \in E} \nu(x) \mathbf{P}_x(\mathbf{X}_0 = x_0, \dots, \mathbf{X}_n = x_n) \\ &= \sum_{x \in E} \nu(x) \mathbf{P}_x(\mathbf{X}_0 = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n) \\ &= \mathbf{P}_\nu(\mathbf{X}_0 = x_0) p(x_0, x_1) \dots p(x_{n-1}, x_n). \end{aligned}$$

The uniqueness of the measure follows from the exact same argument as in the previous proposition. □

Let us remark that if $(\widehat{X}_n)_{n \geq 0}$ is another Markov chain defined on some other probability space $(\widehat{\Omega}, \widehat{\mathscr{F}}, \widehat{P})$, with transition kernel p and initial law ν, then for any measurable set $A \in \boldsymbol{\mathcal{F}}$, we have

$$\widehat{P}\{(\widehat{X}_n)_{n \geq 0} \in A\} = \mathbf{P}_\nu(A).$$

This follows from the uniqueness part of Proposition 4.3. Thus, from now on, we will only work with the canonical Markov chain $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P}, \mathbf{X})$ yet **all the results stated below remain valid for any Markov chain**.

5. The Markov property

One of the advantage of working with the canonical representation on the space of trajectories is that we can consider the **shift operator**:

$$\begin{array}{cccc} \boldsymbol{\theta}: & \boldsymbol{\Omega} & \to & \boldsymbol{\Omega} \\ & (\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \ldots) & \mapsto & (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots). \end{array}$$

This operator is measurable thanks to Lemma 4.2. By composition, we define, the k-shift operators by $\theta^0 = \mathrm{Id}_{\boldsymbol{\Omega}}$ and, for $k \geq 1$,

$$\begin{array}{cccc} \boldsymbol{\theta}^k: & \boldsymbol{\Omega} & \to & \boldsymbol{\Omega} \\ & (\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \ldots) & \mapsto & (\boldsymbol{\omega}_k, \boldsymbol{\omega}_{k+1}, \ldots). \end{array}$$

Making use of the shift operator enables to provide a concise statement of the Markov property.

Theorem 5.1: (weak) Markov property

Let $F : \boldsymbol{\Omega} \to \mathbb{R}_+$ be a measurable function. For any $n \geq 0$ and any $x \in E$, we have [a]

$$\mathbf{E}_x[F(\boldsymbol{\theta}^n \mathbf{X}) \,|\, \mathcal{F}_n] = \mathbf{E}_{\mathbf{X}_n}[F(\mathbf{X})] \tag{5.1}$$

or equivalently, for any $y \in E$,

$$\mathbf{E}_x[F(\boldsymbol{\theta}^n \mathbf{X}) \mathbf{1}_{\{\mathbf{X}_n = y\}} \,|\, \mathcal{F}_n] = \mathbf{E}_y[F(\mathbf{X})] \mathbf{1}_{\{\mathbf{X}_n = y\}}.$$

In words: *Conditionally on $\mathbf{X}_n = y$, the process $\mathbf{X}$ after time n is independent of the events in $\mathcal{F}_n$ and has the same law as the initial Markov chain starting from y.*

[a] The use of $\mathbf{X}$ in equation (5.1) is redundant and could simply be written $\mathbf{E}_x[F \circ \boldsymbol{\theta}^n \,|\, \mathcal{F}_n] = \mathbf{E}_{\mathbf{X}_n}[F]$. Yet we use $\mathbf{X}$ here so that the formula still makes sense when X is not necessarily the canonical Markov chain.

Proof. By definition of the conditional expectation, we need to prove that, for any $B \in \mathcal{F}_n$,

$$\mathbf{E}_x\left[\mathbf{1}_B F(\boldsymbol{\theta}^n \mathbf{X})\right] = \mathbf{E}_x\left[\mathbf{1}_B \mathbf{E}_{\mathbf{X}_n}[F(\mathbf{X})]\right]. \tag{5.2}$$

Any set $B \in \mathcal{F}_n$ can be written as a (countable) disjoint union of elementary sets of the form $\{\mathbf{X}_0 = x_0, \ldots, \mathbf{X}_n = x_n\}$ so we just need to prove the result for these elementary sets. Similarly, in view of the monotone convergence theorem, we only need to prove the result when F is an indicator function $F = \mathbf{1}_A$ and, by application Dynkin's $\pi - \lambda$ theorem, we can furthermore assume that A takes the form $A = \{\mathbf{X}_0 = y_0, \ldots, \mathbf{X}_m = y_m\}$. Thus, we first compute

$$\begin{aligned} \mathbf{E}_x[\mathbf{1}_B F(\boldsymbol{\theta}^n \mathbf{X})] &= \mathbf{P}_x(\mathbf{X}_0 = x_0, \ldots, \mathbf{X}_n = x_n, \mathbf{X}_n = y_0, \ldots \mathbf{X}_{n+m} = y_m) \\ &= \mathbf{1}_{\{x_n = y_0\}} \mathbf{P}_x(\mathbf{X}_0 = x_0) p(x_0, x_1) \ldots p(x_{n-1}, x_n) p(y_0, y_1) \ldots p(y_{m-1}, y_m). \end{aligned}$$

On the other hand, we have, for $z \in E$,

$$\mathbf{E}_z[F(\mathbf{X})] = \mathbf{P}_z(X_0 = y_0, \ldots, \mathbf{X}_m = y_m) = \mathbf{1}_{\{z = y_0\}} p(y_0, y_1) \ldots p(y_{m-1}, y_m)$$

therefore

$$\begin{aligned}\mathbf{E}_x\big[\mathbf{1}_B\mathbf{E}_{\mathbf{X}_n}[F(\mathbf{X})]\big] &= \mathbf{E}_x[\mathbf{1}_{\mathbf{X}_0=x_0,\dots,\mathbf{X}_n=x_n}\mathbf{1}_{\mathbf{X}_n=y_0}p(y_0,y_1)\dots p(y_{m-1},y_m)]\\ &= \mathbf{1}_{\{x_n=y_0\}}\mathbf{P}_x(\mathbf{X}_0=x_0,\dots,\mathbf{X}_n=x_n)p(y_0,y_1)\dots p(y_{m-1},y_m)\\ &= \mathbf{1}_{\{x_n=y_0\}}\mathbf{P}_x(\mathbf{X}_0=x_0)p(x_0,x_1)\dots p(x_{n-1},x_n)p(y_0,y_1)\dots p(y_{m-1},y_m)\end{aligned}$$

which establishes (5.2). $\square$

We can now build on this theorem to obtain a stronger statement which shows that the Markov property holds, not only when the process is shifted by a deterministic time, but also when it is shifted by a random stopping time.

Recall that a stopping time with respect to the canonical filtration $(\mathcal{F}_n)_{n\geq 0}$ is a random variable $T:\mathbf{\Omega}\to\mathbb{N}\cup\{\infty\}$ such that $\{T\leq n\}\in\mathcal{F}_n$ (or equivalently $\{T=n\}\in\mathcal{F}_n$) for all $n\in\mathbb{N}$. The σ-field of "past events" generated by T is defined by

$$\mathcal{F}_T := \{A\subset\mathcal{F}:\ A\cap\{T\leq n\}\in\mathcal{F}_n,\ \forall n\in\mathbb{N}\}$$

Theorem 5.2: (strong) Markov property

Let T be a $(\mathcal{F}_n)$-stopping time. Let $F:\mathbf{\Omega}\to\mathbb{R}_+$ be a measurable function. For any $x\in E$, we have

$$\mathbf{E}_x[\mathbf{1}_{\{T<\infty\}}F(\boldsymbol{\theta}^T\mathbf{X})\,|\,\mathcal{F}_T] = \mathbf{1}_{\{T<\infty\}}\mathbf{E}_{\mathbf{X}_T}[F(\mathbf{X})]$$

or equivalently, for any $y\in E$,

$$\mathbf{E}_x[\mathbf{1}_{\{T<\infty\text{ and }\mathbf{X}_T=y\}}F(\boldsymbol{\theta}^T\mathbf{X})\,|\,\mathcal{F}_T] = \mathbf{1}_{\{T<\infty\text{ and }\mathbf{X}_T=y\}}\mathbf{E}_y[F(\mathbf{X})].$$

In words: *conditionally on $\mathbf{X}_T=y$, the process $\mathbf{X}$ after time T is independent of the past events in $\mathcal{F}_T$ and has the same law as the initial Markov chain starting from y.*

Proof. We need to show that

$$\mathbf{E}_x\big[\mathbf{1}_B\mathbf{1}_{\{T<\infty\}}F(\boldsymbol{\theta}^T\mathbf{X})\big] = \mathbf{E}_x\big[\mathbf{1}_B\mathbf{1}_{\{T<\infty\}}\mathbf{E}_{\mathbf{X}_T}[F(\mathbf{X})]\big].$$

for any $B\in\mathcal{F}_T$. Let $n\in\mathbb{N}$, the event $\{T=n\}\cap B$ belong to $\mathcal{F}_n$ by definition of a stopping time so the weak Markov property states that

$$\mathbf{E}_x\big[\mathbf{1}_B\mathbf{1}_{\{T=n\}}F(\boldsymbol{\theta}^n\mathbf{X})\big] = \mathbf{E}_x\big[\mathbf{1}_B\mathbf{1}_{\{T=n\}}\mathbf{E}_{\mathbf{X}_n}[F(\mathbf{X})]\big].$$

Summing over n gives the required equality. $\square$

Remark 5.3

The previous Markov property (in both weak and strong form) remains valid if we replace $\mathbf{E}_x$

by $\mathbf{E}_\nu$ for any initial law $\nu \in \mathcal{P}(E)$. This follows directly from equality (4.2).

Corollary 5.4

Let $x, y \in E$. Let T be a $(\mathcal{F}_n)$-stopping time such that $\mathbf{X}_T = y$ $\mathbf{P}_x$-a.s. (in particular T is finite). Then, under $\mathbf{P}_x$, the process $\theta^T \mathbf{X}$ if independent of $\mathcal{F}_T$ and has law $\mathbf{P}_y$.

Example 5.5: The reflection principle

Consider a random walk $X_n = x_0 + \xi_1 + \ldots + \xi_n$ where $(\xi_n)_{n\geq 0}$ denotes a sequence of i.i.d. random variables taking value in $\mathbb{Z}$. Suppose that the law of the increments is symmetric *i.e.* $\mathbf{P}(\xi_i = x) = \mathbf{P}(\xi_i = -x)$, then for any $a \geq 0$ and any $n \geq 1$,

$$\mathbf{P}_0(\sup_{i\leq n} X_i \geq a) \leq 2\mathbf{P}_0(X_n \geq a).$$

Proof. Fix $n \geq 1$ and define the stopping $T = \inf(i \geq 0, X_i \geq a)$. We have

$$\begin{aligned}\mathbf{P}_0(X_n \geq a) &= \mathbf{P}_0(T \leq n,\ X_n \geq a)\\ &= \sum_{k=0}^{n} \mathbf{P}_0(T = k,\ X_{T+(n-k)} \geq a)\\ &= \sum_{k=0}^{n} \mathbf{E}_0[\mathbf{1}_{\{T=k\}} \mathbf{P}_{X_T}(X_{n-k} \geq a)]\end{aligned}$$

where we applied the Markov property for the last equality. Noticing that $\mathbf{P}_b(X_m \geq a) \geq \frac{1}{2}$ for any $b \geq a$ and any $m \geq 0$ because the walk is symmetric, we conclude that

$$\mathbf{P}_0(X_n \geq a) \geq \frac{1}{2}\sum_{k=0}^{n} \mathbf{E}_0[\mathbf{1}_{\{T=k\}}] = \frac{1}{2}\mathbf{P}_0(T \leq n) = \frac{1}{2}\mathbf{P}_0(\sup_{i\leq n} X_n \geq a). \qquad \square$$

Comment. The proof is the same for a random walk with increments in $\mathbb{R}$. The symmetry assumption can also be weakened, requiring instead that $\mathbf{P}_0(X_m \geq 0)$ is uniformly bounded below in m by some constant $\alpha > 0$. This happens, for instance, whenever the r.v. ξ_i's are centered and admit a finite second moment. In this case, the same argument now shows that, for all $n \geq 0$,

$$\mathbf{P}_0(\sup_{i\leq n} X_i \geq a) \leq \frac{1}{\alpha}\mathbf{P}_0(X_n \geq a).$$

6. Classification of states: recurrence and transience

We now study the asymptotic behavior of a Markov chain. In particular, we want to understand which states of the space E are visited infinitely often and which one are only visited finitely many times. Two quantities are of particular interest: for any site $x \in E$, we define

$$N(x) := \sum_{n=0}^{\infty} \mathbf{1}_{\{\mathbf{X}_n = x\}} \qquad \text{(number of visits of site } x\text{)},$$

and

$$T_x := \inf(n \geq 1 : \mathbf{X}_n = x) \qquad \text{(first hitting time of site } x\text{)}$$

Here and everywhere else, we enforce the convention $\inf \emptyset = \infty$.

Proposition 6.1

Let $x \in E$. We have the dichotomy:

(a) either $\mathbf{P}_x(T_x < \infty) = 1$, in which case

$$N(x) = \infty, \qquad \mathbf{P}_x\text{-a.s.,}$$

(b) or $\mathbf{P}_x(T_x < \infty) < 1$, in which case

$$N(x) < \infty, \qquad \mathbf{P}_x\text{-a.s.,}$$

and $N(x)$ has a geometric law with parameter $\mathbf{P}_x(T_x = \infty)$. In particular, we have

$$\mathbf{E}_x[N(x)] = \frac{1}{\mathbf{P}_x(T_x = \infty)} < \infty.$$

Definition 6.2: Recurrent and transient states

We say that a site $x \in E$ is **recurrent** in case (a) and **transient** in case (b). The set of all recurrent sites is denoted by $\mathcal{R}$ and the the set of transient sites is denoted by $\mathcal{T}$. With this notations, we get a partitioning of the state space:

$$E = \mathcal{R} \sqcup \mathcal{T}.$$

Proof of Proposition 6.1. Let $k \geq 1$. Applying the strong Markov property for the stopping time T_x,

$$\begin{aligned}
\mathbf{P}_x(N(x) \geq k+1) &= \mathbf{E}_x\left[\mathbf{1}_{\{T_x<\infty\}}\left(\mathbf{1}_{\{N(x)\geq k\}} \circ \boldsymbol{\theta}^{T_x}\right)\right] \\
&= \mathbf{E}_x\left[\mathbf{1}_{\{T_x<\infty\}}\, \mathbf{E}_{\mathbf{X}_{T_x}}\left[\mathbf{1}_{\{N(x)\geq k\}}\right]\right] \qquad \text{(strong Markov property)} \\
&= \mathbf{E}_x\left[\mathbf{1}_{\{T_x<\infty\}}\, \mathbf{P}_x(N(x) \geq k)\right] \qquad (\mathbf{X}_{T_x} = x \text{ on the set } \{T_x < \infty\}) \\
&= \mathbf{P}_x(T_x < \infty)\, \mathbf{P}_x(N(x) \geq k).
\end{aligned}$$

Noticing that $\mathbf{P}_x(N(x) \geq 1) = 1$, we get by induction that, for $k \geq 1$,

$$\mathbf{P}_x(N(x) \geq k) = \mathbf{P}_x(T_x < \infty)^{k-1}. \tag{6.1}$$

If $\mathbf{P}_x\{T_x < \infty\} = 1$, then $\mathbf{P}_x\{N(x) \geq k\} = 1$ for all $k \geq 1$, which means $\mathbf{P}_x\{N(x) = \infty\} = 1$. This proves (a). On the other hand, if $\mathbf{P}_x\{T_x < \infty\} < 1$, then (6.1) shows that $N(x)$ has a geometric law with parameter $1 - \mathbf{P}_x(T_x < \infty) = \mathbf{P}_x(T_x = \infty)$. In particular, it is finite and its expectation is $\frac{1}{\mathbf{P}_x(T_x=\infty)}$ which proves (b). □

Definition 6.3: Green function

The **Green function** (also called **potential kernel**) of the Markov chain $\mathbf{X}$ is the function $G : E \times E \to [0, \infty]$ defined by

$$G(x, y) := \mathbf{E}_x[N(y)].$$

Proposition 6.4

1. For any $x, y \in E$,
$$G(x, y) = \sum_{n=0}^{\infty} p^n(x, y).$$
In particular,
$$G(x, y) > 0 \iff \mathbf{P}_x(T_y < \infty) > 0.$$

2. For any $x \in E$, we have the equivalence
$$G(x,\, x) = \infty \iff x \text{ is recurrent.}$$

3. If $x \neq y$, then we have
$$G(x, y) = \mathbf{P}_x(T_y < \infty)\, G(y, y).$$
In particular,
$$G(x,\, y) \leq G(y,\, y).$$

Proof. By definition of $N(y)$ and using the linearity of the expectation, we can write

$$G(x, y) = \mathbf{E}_x\Big(\sum_{n=0}^{\infty} \mathbf{1}_{\{\mathbf{X}_n = y\}} \Big) = \sum_{n=0}^{\infty} \mathbf{P}_x(\mathbf{X}_n = y) = \sum_{n=0}^{\infty} p^n(x,\, y)$$

which proves the first identity. To see why the equivalence is true, we notice that, $\mathbf{P}_x(T_y < \infty) > 0$ if and only if $p^n(x, y) = \mathbf{P}_x(\mathbf{X}_n = y) > 0$ for at least one n, which is equivalent to the Green function being non-zero since all the terms in the sum are non-negative. This proves 1. Assertion 2. is a direct consequence of Proposition 6.1. For assertion 3., we use the strong Markov property at time T_y to conclude that

$$G(x, y) = \mathbf{E}_x[N(y)] = \mathbf{E}_x\big[\mathbf{1}_{\{T_y < \infty\}}\, (N(y) \circ \boldsymbol{\theta}^{T_y})\big] = \mathbf{E}_x\big[\mathbf{1}_{\{T_y < \infty\}}\, \mathbf{E}_y[N(y)]\big] = \mathbf{P}_x(T_y < \infty)\, G(y, y)$$

□

Example 6.5

Consider the Markov chain in $\mathbb{Z}^d$ with transition kernel

$$p(x, y) = \frac{1}{2^d} \prod_{i=1}^{d} \mathbf{1}_{\{|y_i - x_i| = 1\}}, \qquad x := (x_1, \cdots, x_d),\ y := (y_1, \cdots, y_d) \in \mathbb{Z}^d. \tag{6.2}$$

This defines a random walk which is not the simple random walk on the graph $\mathbb{Z}^d$ (but it is the simple random for some other graph obtained modifying locally the edges around each vertex of $\mathbb{Z}^d$). We write $X_n = (X_n^1, \ldots, X_n^d)$. Then, if follows directly from the definition (6.2) written as of product along each dimension that the processes $X^1, \cdots, X^d$ are independent simple symmetric random walks on $\mathbb{Z}$. In particular, we find that

$$p^n(0,\,0) = \mathbf{P}(X_n^1 = 0,\, \cdots,\, X_n^d = 0) = \mathbf{P}(X_n^1 = 0)^d.$$

The probability $\mathbf{P}(X_n^1 = 0)$ is the probability of having the same number of heads and tails in the first n toss of an unbiased coins hence

$$\mathbf{P}(X_n^1 = 0) = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \frac{\binom{n}{n/2}}{2^n} & \text{if } n \text{ is even.} \end{cases}$$

Therefore,

$$G(0,0) = \sum_{k=0}^{\infty} p^{2k}(0,\,0) = \sum_{k=0}^{\infty} \left(\frac{\binom{2k}{k}}{2^{2k}} \right)^d.$$

Making use of the Stirling formula, we have the asymptotics when $k \to \infty$:

$$\frac{\binom{2k}{k}}{2^{2k}} = \frac{(2k)!}{2^{2k}(k!)^2} \sim \frac{\sqrt{4\pi k}\left(\frac{2k}{e}\right)^{2k}}{2^{2k}\left(\sqrt{2\pi k}\left(\frac{k}{e}\right)^k\right)^2} = \frac{1}{\sqrt{\pi k}}$$

Thus, we conclude that

$$G(0,\,0) \begin{cases} = \infty & \text{for } d = 1, 2, \\ < \infty & \text{for } d \geq 3. \end{cases}$$

This means that, the origin (and by translation invariance any $x \in \mathbb{Z}^d$) is recurrent if $d \leq 2$ and transient for $d \geq 3$.

Remark 6.6

Let X denote the Markov chain in $\mathbb{Z}^2$ with transition kernel given by (6.2). Consider the process Y defined by $Y_n = AX_n$ where A is the linear transformation obtained by a rotation of 45 degree followed by a scaling of factor $1/\sqrt{2}$. It is easy to check that Y is a simple random walk on $\mathbb{Z}^2$. Thus, it is recurrent since X is recurrent. Unfortunately, this argument cannot be use study simple random walks in higher dimensions because the walk with kernel (6.2) has 2^d neighbours in dimension d whereas the simple random walk only has $2d$ neighours in $\mathbb{Z}^d$.

Lemma 6.7

Let $x \in E$ be a recurrent state. Let $y \neq x$ be such that $G(x, y) > 0$. Then y is also recurrent and $\mathbf{P}_y(T_x < \infty) = 1$, in particular $G(y, x) > 0$.

Proof. Since x is recurrent, we have

$$\begin{aligned} 0 = \mathbf{P}_x(N(x) < \infty) &\geq \mathbf{P}_x(T_y < \infty,\ T_x \circ \theta^{T_y} = \infty) \\ &= \mathbf{E}_x\big[\mathbf{1}_{\{T_y<\infty\}}\,\mathbf{1}_{\{T_x=\infty\}} \circ \theta^{T_y}\big] \\ &= \mathbf{E}_x\big[\mathbf{1}_{\{T_y<\infty\}}\,\mathbf{P}_y(T_x = \infty)\big] \qquad \text{(strong Markov property)} \\ &= \mathbf{P}_x(T_y < \infty)\,\mathbf{P}_y(T_x = \infty). \end{aligned}$$

By assumption, $G(x, y) > 0$ hence $\mathbf{P}_x(T_y < \infty) > 0$ which implies that $\mathbf{P}_y(T_x = \infty) = 0$ therefore $\mathbf{P}_y(T_x < \infty) = 1$ as requested. It remains to prove that y is recurrent. We already know $G(x, y) > 0$ and $G(y,\ x) > 0$. So there exist $n, m \geq 1$ such that $p^n(x, y) > 0$ and $p^m(y, x) > 0$. For any $i \geq 0$, we have

$$p^{m+i+n}(y,\ y) \geq p^m(y,\ x)\,p^i(x,\ x)\,p^n(x,\ y).$$

Thus, we get that

$$G(y, y) \ \geq\ \sum_{i=0}^{\infty} p^{m+i+n}(y, y) \geq p^m(y, x)\,p^n(x, y) \sum_{i=0}^{\infty} p^i(x, x) \ =\ p^m(y, x)\,p^n(x, y)\,G(x, x).$$

By assumption, $G(x, x) = \infty$ because x is recurrent, whereas $p^m(y,\ x)\,p^n(x,\ y) > 0$, so $G(y,\ y) = \infty$ which means that y is recurrent. □

Remark 6.8

The previous lemma states that if x is a recurrent state and y is a transient state, then necessarily $G(x, y) = 0$. This means there is no path (that has positive probability) that connects a recurrent state to a transient state (but the opposite may happen).

Theorem 6.9: Decomposition of a Markov chain

The relation $x \sim y$ defined by

$$x \sim y \quad \Longleftrightarrow \quad G(x, y) > 0$$

is an equivalence relation on the set of recurrent sites $\mathcal{R}$. Thus, this set can be partitioned:

$$\mathcal{R} = \bigsqcup_{i \in I} \mathcal{R}_i$$

where $(\mathcal{R}_i, i \in I)$ denote the equivalent classes of $\sim$ on $\mathcal{R}$ called **recurrence classes**.

The following properties hold.

1. Let $x \in \mathcal{R}_i$, then
$$N(y) = \begin{cases} \infty & \text{for all } y \in \mathcal{R}_i, \\ 0 & \text{for all } y \in E \setminus \mathcal{R}_i, \end{cases} \qquad \mathbf{P}_x\text{-a.s.}$$

2. Let $x \in \mathcal{T}$ and define $\tau = \inf(n \geq 1 : \mathbf{X}_n \in \mathcal{R})$,

- On the event $\{\tau = \infty\}$, we have $N(y) < \infty$ for all $y \in E$, $\mathbf{P}_x$-a.s.
- On the event $\{\tau < \infty\}$, there exists a (possibly random) index $i \in I$ such that $\mathbf{X}_n \in \mathcal{R}_i$ for all $n \geq \tau$, $\mathbf{P}_x$-a.s.

Proof. By definition, $G(x, x) \geq 1$ hence $x \sim x$ for all $x \in \mathcal{R}$ (reflexivity property). The previous lemma shows that, if $x \sim y$ and $x \in \mathcal{R}$, then $y \sim x$ (symmetry property). Finally, if $x, y, z \in \mathcal{R}$ with $x \sim y$ and $y \sim z$, then there exists n, m such that $p^n(x, y) > 0$ and $p^m(y, z) > 0$, therefore

$$p^{n+m}(x, z) \geq p^n(x, y)p^m(y, z) > 0.$$

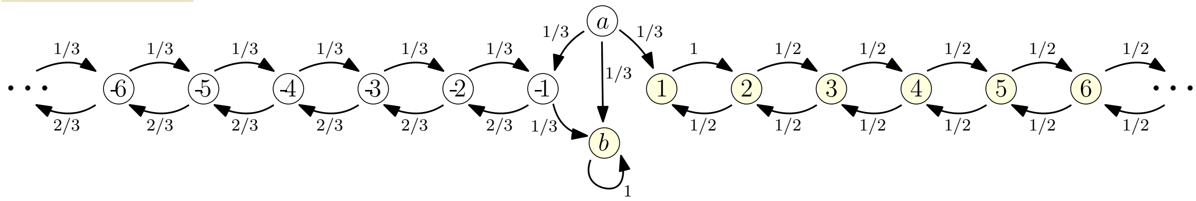
This implies $x \sim z$ (transitivity property) so $\sim$ is indeed an equivalence relation on $\mathcal{R}$ (but not necessarily on E).

Let $x \in \mathcal{R}_i$ for some $i \in I$. Let $y \in E \setminus \mathcal{R}_i$. Either $y \in \mathcal{T}$ but then $G(x, y) = 0$ thanks to the previous lemma or $y \in \mathcal{R}_j$ for some $j \neq i$ but then again $G(x, y) = 0$ by definition of the equivalence relation. Thus in any case $G(x, y) = 0$ which means that $N(y) = 0$ $\mathbf{P}_x$-a.s. On the other hand, if $y \in \mathcal{R}_i$, then, the previous lemma states that $\mathbf{P}_x(T_y < \infty) = 1$ and, using (yet again) the strong Markov property, we find that

$$\mathbf{P}_x(N(y) = \infty) = \mathbf{E}_x\big[\mathbf{1}_{\{T_y<\infty\}}\,(\mathbf{1}_{\{N(y)=\infty\}} \circ \boldsymbol{\theta}^{T_y})\big] = \mathbf{P}_x(T_y < \infty)\,\mathbf{P}_y(N(y) = \infty). \tag{6.3}$$

Since both probabilities on the right hand side are equal to 1, we get that $\mathbf{P}_x(N(y) = \infty) = 1$ which complete the proof of Assertion 1. Let now $x \in \mathcal{T}$. Note that, according to (6.3), $N(y) < \infty$ $\mathbf{P}_x$ -p.s. whenever $y \in \mathcal{T}$ is transient. In particular, on the event $\{\tau = \infty\}$, we have $N(y) < +\infty$ for all $y \in E$. Conversely, on $\{\tau < \infty\}$, there exists, by definition, some index j such that $\mathbf{X}_\tau \in \mathcal{R}_j$. Using the strong Markov property with the stopping time τ combined with Assertion 1., we see that $\mathbf{X}$ remains inside $\mathcal{R}_j$ for all times after τ which proves Assertion 2. $\square$

Example 6.10



Consider the Markov chain with transition kernel described by the graph above.

1. The set $\mathcal{R}$ of recurrent sites is composed of two recurrence classes:
   - $\{b\}$ which is an absorbing state.
   - $\{1, 2, 3, \ldots\}$ because the chain on this set is a (reflected) symmetric simple one dimensional random walk hence recurrent according to Example 6.5.
2. All other sites are transient
   - $\{a\}$ is transient because it can never be visited twice.
   - $\{-1, -2, -3, \ldots\}$ are transient because a biased random walk drifts to ∞ thanks to the law of large numbers.

Notice also that, starting from a transient site, the walk can ultimately end inside either of two the recurrence classes with positive probability or, also with positive probability, remain inside the transient set for all time.

Definition 6.11

A Markov chain is said to be **irreducible** if $G(x,\, y) > 0$ for all $x, y \in E$.

Corollary 6.12

Assume that $\mathbf{X}$ is an irreducible Markov chain on E. Then

- either all states are recurrent. Then, there is only one recurrence class, and for any $x \in E$,

$$\mathbf{P}_x(N(y) = \infty,\ \forall y \in E) = 1.$$

  In this case, $\mathbf{X}$ is called an **irreducible recurrent Markov chain**.

- or all states are transient. Then, for any $x \in E$,

$$\mathbf{P}_x(N(y) < \infty,\ \forall y \in E) = 1.$$

  In this case, $\mathbf{X}$ is called an **irreducible transient Markov chain**.

When E is finite, only the first situation occurs.

Proof. Assume first there exists a recurrent state. Then, by irreducibility and according to Lemma 6.7, all states are recurrent and there exists a unique recurrence class. Thus, Assertion 1. of Theorem 6.9 insures that all sites are visited infinitely often a.s. Conversely, suppose that all sites are transient, then, of course $\tau = \inf(n \geq 1, \mathbf{X}_n \in \mathcal{R}) = \infty$ a.s. which means that we are in the second case of Assertion 2. of Theorem 6.9 hence each site is visited only finitely many times. Finally, when E is finite, then at least one site must be visited infinitely often therefore the second case cannot happen. □

Remark 6.13

If $\mathbf{X}$ if a (non) irreducible Markov chain with recurrence classes $(\mathcal{R}_i)_{i \in I}$, then, for each $i \in I$, we can define the restriction of $\mathbf{X}$ on $\mathcal{R}_i$ which it the Markov chain $\mathbf{X}^i$ with kernel

$$p^i(x, y) = p(x, y) \text{ for all } x, y \in \mathcal{R}_i.$$

This equation defines a kernel since $p(x, y) = 0$ whenever $x \in \mathcal{R}_i$ and $y \notin \mathcal{R}_i$. Then, X^i is an irreducible recurrent Markov chain.

We end this section by discussing the classification of states for the Markov chains examples presented in Section 2.

• **i.i.d. random variables with law μ.** We have $p(x,\, y) = \mu(y)$. It is trivial that y is recurrent if and only if $\mu(y) > 0$, and that there is only one recurrence class. The chain is irreducible if and only if $\mu(y) > 0$ for all $y \in E$.

• **Random walks on $\mathbb{Z}$.** In this example, $E = \mathbb{Z}$, and the transition kernel is given by

$$p(x,\, y) = \mu(y - x).$$

By translation invariance, the Green function $G(x, y)$ depends only of $y - x$, so $G(x,\, x) = G(y,\, y)$ and all states are of same type: they are either all recurrent, or all transient. Let ξ denote a random variable whose law is μ.

Theorem 6.14

Assume that $\mathbf{E}[|\xi|] < \infty$.

1. If $\mathbf{E}[\xi] \neq 0$, then all states are transient.
2. If $\mathbf{E}[\xi] = 0$, all states are recurrent. Moreover, the chain is irreducible if and only if the sub-group generated by $\{y \in \mathbb{Z} :\ \mu(y) > 0\}$ is $\mathbb{Z}$ itself.

Proof. If $\mathbf{E}(\xi) \neq 0$, the law of large numbers shows us that $|X_n| \to \infty$ a.s., so that all states are transient. We now assume that $\mathbf{E}[\xi] = 0$ and prove that 0 is recurrent. Suppose, by contradiction, that 0 is transient hence $G(0, 0) < \infty$. Then according to Proposition 6.4, for all $x \in \mathbb{Z}$, we have

$$G(0, x) \leq G(x, x) = G(0, 0).$$

Therefore, for any $n \geq 1$,

$$\sum_{|x| \leq n} G(0,\, x) \leq (2n + 1)G(0,\, 0) \leq cn,$$

where $c := 3G(0, 0) < \infty$. On the other hand, by the weak law of large numbers, $\frac{X_n}{n} \to 0$ in probability, so that for any $\varepsilon > 0$, and all sufficiently large n (say, $n > n_0$),

$$\mathbf{P}(|X_n| \leq \varepsilon n) > \frac{1}{2},$$

which is equivalent to saying that

$$\sum_{|x| \leq \varepsilon n} p^n(0,\, x) > \frac{1}{2}.$$

For $n \geq i > n_0$, we have

$$\sum_{|x| \leq \varepsilon n} p^i(0, x) \geq \sum_{|x| \leq \varepsilon i} p^i(0, x) > \frac{1}{2},$$

so that for $n \geq n_0$,

$$\sum_{|x| \leq \varepsilon n} G(0,\, x) \geq \sum_{i=n_0}^{n} \sum_{|x| \leq \varepsilon n} p^i(0,\, x) > \frac{n - n_0}{2}$$

which contradicts the inequality

$$\sum_{|x|\leq \varepsilon n} G(0,\,x) \leq c\varepsilon n,$$

when $\varepsilon < \frac{1}{2c}$ and n is large enough. As a consequence, 0 is recurrent. It remains to characterize irreducibility. Let $\mathbb{G}$ be the sub-group generated by $\{x \in \mathbb{Z}:\ \mu(x) > 0\}$. Clearly

$$\mathbf{P}_0(X_n \in \mathbb{G},\ \forall n \in \mathbb{N}) = 1.$$

Thus, if $\mathbb{G} \neq \mathbb{Z}$, the chain is not irreducible. Conversely, assume now $\mathbb{G} = \mathbb{Z}$ and define

$$\mathbb{G}_1 := \{x \in \mathbb{Z}:\ G(0,\,x) > 0\}.$$

We claim that $\mathbb{G}_1$ is a sub-group of $\mathbb{Z}$.

- If x, $y \in \mathbb{G}_1$, we can find n, m such that $p^n(0,x) > 0$ and $p^m(0,y) > 0$ and then

  $$p^{n+m}(0, x+y) \geq p^n(0,x)\, p^m(x, x+y) = p^n(0,x)\, p^m(0,y) > 0,$$

  shows that $x + y \in \mathbb{G}_1$,

- If $x \in \mathbb{G}_1$, and since 0 is recurrent, $G(0,\,x) > 0$ implies $G(x,\,0) = G(0, -x) > 0$ (Lemma 6.7), which shows that $-x \in \mathbb{G}_1$.

Finally, since $\mathbb{G}_1 \supset \{x \in \mathbb{Z}:\ \mu(x) > 0\}$, it contains the sub-group generated by the latter which by assumption is $\mathbb{Z}$. This shows that the chain is irreducible. □

Remark 6.15

If the walk is not irreducible, then the group generated by the increments of the walk is of the form $a\mathbb{Z}$ for some $a \geq 2$ and there are exactly a recurrence classes.

• **Simple random walk on a graph and electrical network.** We will discuss this model in greater details later. For the time being, we only consider the case when E is finite.

Proposition 6.16

Consider a random walk on an electrical network $(E, \mathcal{E}, c)$. Assume that $\sharp E < +\infty$. Then, every site is recurrent and the recurrence classes are exactly the connected component of the graph.

Proof. By definition of the electrical network, the weight on each edge is strictly positive. Thus, any finite path inside the graph has positive probability. This implies that the equivalence relation "x and y are connected in the graph" and "$G(x,y)G(y,x) > 0$" are identical and thus define the same equivalence classes. Finally, the walk on each class is clearly irreducible and recurrent since the graph is finite. □

• **Branching processes.** In this example, $E = \mathbb{N}$ and $p(x, y) = \mu^{*x}(y)$ (with $\mu^{0*} := \delta_0$). We observe that state 0 is absorbing:

$$\mathbf{P}_0(X_n = 0, \ \forall n \in \mathbb{N}) = 1.$$

A fortiori, 0 is also recurrent. In the following proposition, we exclude the trivial case $\mu = \delta_1$, where all states are absorbing.

Proposition 6.17

Assume that $\mu \neq \delta_1$, then 0 is the only recurrent state for the branching process with reproduction law μ. As a consequence, we have, almost surely, for any starting point.

- either there exists $\tau < \infty$ such that $X_n = 0$, for all $n \geq \tau$;
- or $X_n \to \infty$ as $n \to \infty$.

Remark 6.1. In the chapter on martingales, we have seen that the first situation occurs a.s. if $m := \sum_{k \in \mathbb{N}} k\mu(k) \leq 1$ (critical or sub-critical case), whereas the second situation is produced with positive probability if $m > 1$ and the process starts from $x \neq 0$ (super-critical case).

Proof. We only need to check that 0 is the only recurrent state, the rest of the theorem is a consequence of the decomposition theorem. Let $x \geq 1$. We want to prove that x is transient. We consider two possible situations.

- $\mu(0) > 0$. Then $G(x, 0) \geq \mathbf{P}_x(X_1 = 0) = \mu(0)^x > 0$, whereas $G(0, x) = 0$. This is possible only if x is transient (Lemma 6.7).
- $\mu(0) = 0$. Since $\mu \neq \delta_1$, there exists $k \geq 2$ such that $\mu(k) > 0$. Since $\mathbf{P}_x(X_1 > x) \geq \mu(k)^x > 0$, there exists $y > x$ such that $p(x, y) > 0$, and a fortiori, $G(x, y) > 0$. But $G(y, x) = 0$ (since $\mu(0) = 0$), which, again, is possible only if x is transient.

□

• **Ehrenfest Urn.** Recall that we consider Markov chain on the finite state space $E = \{0, \ldots, N\}$ with transition kernel given by (2.1). Then, it is clear that the chain is irreducible since

$$p^{|i-j|}(i, j) \geq (1/N)^{|i-j|} > 0$$

for any $i, j \in E$. Thus, according to Corollary 6.12, the chain is irreducible recurrent.

• **Pólya Urn.** Here the state space is $E = \mathbb{N}^{*2}$ and we consider a Markov chain (B_n, W_n) with kernel given by (2.2). Clearly, every state is transient since $B_n + W_n \to \infty$ as $n \to \infty$.

7. Stationary measures

We have seen in the previous section that an irreducible Markov chain is either recurrent or transient *i.e.* either all states are visited infinitely often a.s. or they are visited only finitely many times a.s. When we have a recurrent Markov chain, a natural question is to quantity "how often does the chain return to a given state x" or "is a state x visited more often than some other state y" ? The study of the invariant measures of the chain gives simple answers to these questions and provides a powerful framework for studying its asymptotic behavior.

Definition 7.1: invariant measure

A measure μ on E is said to be **invariant** (or **stationary**) for a Markov chain with transition kernel p if it is not identically 0, it is locally finite ($\mu(y) < \infty$ for all $y \in E$) and

$$\mu.p = \mu \qquad \text{which means} \qquad \mu(y) = \sum_{x \in E} \mu(x)p(x,y) \quad \text{for all } y \in E$$

i.e. μ is a left eigenvector for the eigenvalue 1 of p.

By definition, if μ is an invariant measure, then $\mu.p^n = \mu$ for any $n \geq 0$. In particular, suppose that μ is a probability measure *i.e.* $\mu(E) = 1$. Then for any $f : E \to \mathbb{R}_+$ and any $n \in \mathbb{N}$, we have

$$\mathbf{E}_\mu[f(X_n)] = \mu.p^n.f = \mu.f = \mathbf{E}_\mu[f(X_0)].$$

This means that, **starting from an invariant probability distribution μ, the law of X_n is equal to μ at all times n. In particular, it does not depend on n.**

Remark 7.2

1. If μ is an invariant probability distribution, then under $\mathbf{P}_\mu$, the chains $(X_k)_{k\geq 0}$ and $(X_{n+k})_{k\geq 0}$ have the same law for all n. This follows directly from the fact that X_0 and X_n have the same law combined with the weak Markov property.
2. If we have an invariant measure μ with finite weight $\mu(E) < \infty$, then we can define an invariant probability measure π via $\pi(x) := \mu(x)/\mu(E)$.
3. When the measure μ has infinite weight $\mu(E) = \infty$, there does not exist a probability distribution proportional to μ anymore. However, we can still give a probabilistic interpretation of stationarity. To do so, consider a family of independent random variable $(U^x,\ x \in E)$ such that U^x has a Poisson distribution with parameter $\mu(x)$. Then, from each site $x \in E$, we start U^x Markov chains with transition kernel p. We assume that all these chains are independent. Let U^x_n denote the number of chains that are at position x at time n. Then, for all n, the law of $(U^x_n,\ x \in E)$ is the same as $(U^x,\ x \in E)$. In particular, it does not depend on n.

Definition 7.3

A measure μ on E is said to be **reversible** for a Markov chain with transition kernel p if it is not identically 0, locally finite ($\mu(y) < \infty$ for all $y \in E$) and

$$\mu(x)p(x,y) = \mu(y)p(y,x) \qquad \text{for all } x, y \in E.$$

Proposition 7.4

A reversible measure is invariant.

Proof. Suppose that μ is reversible, then

$$\sum_{x \in E} \mu(x)p(x,y) = \sum_{x \in E} \mu(y)p(y,x) = \mu(y).$$

which shows that μ is invariant. $\square$

Remark 7.5

1. There exist invariant measures that are not reversible: consider the markov chain on $\{0,1,2\}$ such that $p(0,1) = p(1,2) = p(2,0) = 1$ and $p(x,y) = 0$ otherwise. Then, the uniform measure on $\{0,1,2\}$ is invariant but it is not reversible.
2. One advantage of reversible measure over generic invariant measure is that they are usually much easier to find and compute. Advice: when looking for an invariant measure, always search for a reversible one first!

Some examples

(a) i.i.d random variables. Suppose that $(X_n)_{n \geq 1}$ is a family of i.i.d. random variables with law ν, then the unique invariant measure for the chain is ν.

(b) Random walks. Consider a random walk on $\mathbb{Z}^d$,

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

where the $(\xi_i)_{i \geq 1}$ is a family of i.i.d. random variables with law ν. The kernel of the chain is given by $p(x,y) = \nu(y-x)$ hence it is invariant by translation which implies that the counting measure $\mu(x) = 1$ for all $x \in \mathbb{Z}^d$ is invariant. However, this measure is reversible if and only if the law of μ is symmetric *i.e.* $\mu(x) = \mu(-x)$ for all $x \in \mathbb{Z}^d$.

Consider the particular case of the biased random walk on $\mathbb{Z}$. Let $p \in (0,1)$ and suppose that

$$\mathbf{P}(\xi_i = 1) = p = 1 - \mathbf{P}(\xi_i = -1)$$

so we have a Markov chain on $\mathbb{Z}$ with transition kernel

$$p(i,\, i+1) = p, \qquad p(i,\, i-1) = 1-p,$$

It is straightforward to check that the measure

$$\mu(i) := \left(\frac{p}{1-p}\right)^i \quad \text{for all } i \in \mathbb{Z}$$

is reversible (hence invariant) for the chain. We observe that μ is different from the counting measure (which is also invariant), except in the case $p = \frac{1}{2}$.

(c) Random walks on electrical networks. Consider an electrical network $(E, \mathcal{E}, c)$. Define π to be the sum of the conductances around each edges of the graph:

$$\pi(x) := \sum_{y/\{x,y\}\in\mathcal{E}} c(x,y) \tag{7.1}$$

Assume that $0 < \pi(x) < \infty$ for all $x \in E$. The random walk on the electrical network $(E, \mathcal{E}, c)$ is given by the transition kernel

$$p(x,y) := \frac{c(x,y)}{\pi(x)}. \tag{7.2}$$

The following proposition shows that random walks on electrical networks are exactly Markov chains that admit a reversible measure (everywhere positive).

Proposition 7.6

1. The measure π is reversible for the random walk on the electrical network $(E, \mathcal{E}, c)$.
2. Conversely, if X is a Markov on some state space E that admits a reversible measure μ such that $\mu(x) > 0$ for all $x \in E$, then, we can construct an electrical network on E that represents X.

Proof. 1. Recall that, by definition, the conductance are symmetric: $c(x,y) = c(y,x)$ for all $x, y \in E$. Therefore, according to (7.2),

$$\pi(x)p(x,y) = c(x,y) = c(y,x) = \pi(y)p(y,x)$$

which shows that π is indeed reversible for the random walk with kernel (7.2).
2. Conversely, suppose that we have a Markov chain with some transition kernel p on some state space E that admits a reversible measure μ such that $\mu(x) > 0$ for all $x \in E$. We remark that the reversibility equation

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

shows that $p(x,y) > 0$ if and only if $p(y,x) > 0$. Thus, we can define a graph structure $\mathcal{E}$ on E by setting

$$\mathcal{E} := \{(x,y) \in E \times E \,:\, p(x,y) > 0\}.$$

We also set

$$c(x,y) := \mu(x)p(x,y).$$

This defines a set of conductances on $(E, \mathcal{E})$. Indeed, by construction, $c(x,y) > 0$ if and only if $(x,y) \in \mathcal{E}$ and the symmetry properties follows from the reversibility of μ:

$$c(x,y) = \mu(x)p(x,y) = \mu(y)p(y,x) = c(y,x).$$

Again, we define π by (7.1). We observe that

$$\pi(x) = \sum_{y/\{x,y\}\in\mathcal{E}} c(x,y) = \mu(x)\sum_{y} p(x,y) = \mu(x).$$

so we conclude that the Markov chain on the electrical network $(E, \mathcal{E}, c)$ has transition kernel q given by

$$q(x,y) := \frac{c(x,y)}{\pi(x)} = \frac{\mu(x)p(x,y)}{\mu(x)} = p(x,y).$$

as required. □

(c) Ehrenfest urn. Recall that, $E := \{0,\, 1,\, \cdots,\, N\}$ and

$$p(i,j) := \begin{cases} \frac{N-i}{N} & \text{if } j = i+1, \\ \frac{i}{N} & \text{if } j = i-1. \\ 0 & \text{otherwise.} \end{cases} \tag{7.3}$$

which corresponds to the Markov chain that random moves balls from one urn to the other and counts the number of balls in the (say first) urn. Then, a measure μ is reversible for this chain if and only if it satisfies

$$\mu(i)\,\frac{N-i}{N} = \mu(i+1)\,\frac{i+1}{N} \quad \text{for all } 0 \le i \le N-1.$$

It is easy to check that any solution is of the form

$$\mu(i) = c\binom{N}{i}$$

for some constant $c > 0$. In particular, the Binomial(N,1/2) distribution is a reversible probability measure for the chain.

Construction of invariant measures

The following result shows that there always exists an invariant measure supported on the set $\mathcal{R}$ of recurrent sites. Moreover, it provides an explicit form for measure.

Theorem 7.7

Let $\mathbf{X}$ be a Markov chain on E and suppose that $x \in E$ is a recurrent state. The formula

$$\mu_x(y) := \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1} \mathbf{1}_{\{\mathbf{X}_i = y\}}\Big], \qquad y \in E, \tag{7.4}$$

defines an invariant measure. Moreover, $\mu_x(y) > 0$ if and only if y is in the same recurrence class as x.

Remark 7.8

1. Part of the conclusion of the theorem is that $\mu_x(y) < \infty$, $\forall y \in E$.
2. In case there are several recurrence classes R_i, $i \in I$, we can choose, for each $i \in I$, a state $x_i \in R_i$, and then define

$$\mu_{x_i}(y) := \mathbf{E}_{x_i}\Big(\sum_{k=0}^{T_{x_i}-1} \mathbf{1}_{\{\mathbf{X}_k = y\}} \Big), \qquad y \in E.$$

   In this way, we obtain invariant measures supported in disjoint subsets.

Proof of Theorem 7.7. First, we observe that if y is not in the same recurrence class as x (for instance if y is transient), then $\mathbf{E}_x[N(y)] = G(x,y) = 0$ so that $\mu_x(y) = 0$. We now compute μ_x. For any $y \in E$, we write

$$\begin{aligned}
\mu_x(y) &= \mathbf{E}_x\Big[\sum_{i=1}^{T_x} \mathbf{1}_{\{\mathbf{X}_i = y\}}\Big] && \text{(because } \mathbf{X}_0 = \mathbf{X}_{T_x}\ \mathbf{P}_x\text{-a.s.)}\\
&= \sum_{z\in E} \mathbf{E}_x\Big[\sum_{i=1}^{T_x} \mathbf{1}_{\{\mathbf{X}_{i-1}=z,\ \mathbf{X}_i = y\}}\Big]\\
&= \sum_{z\in E}\sum_{i=1}^{\infty} \mathbf{E}_x\Big[\mathbf{1}_{\{T_x \geq i,\ \mathbf{X}_{i-1}=z\}}\mathbf{1}_{\{\mathbf{X}_i = y\}}\Big] && \text{(Fubini)}\\
&= \sum_{z\in E}\sum_{i=1}^{\infty} \mathbf{E}_x\Big[\mathbf{1}_{\{T_x \geq i,\ \mathbf{X}_{i-1}=z\}}\Big] p(z,y) && \text{(Markov prop. using } \{T_x \geq i\} = \{T_x \leq i-1\}^c \in \mathscr{F}_{i-1})\\
&= \sum_{z\in E} \mathbf{E}_x\Big[\sum_{i=1}^{T_x} \mathbf{1}_{\{\mathbf{X}_{i-1}=z\}}\Big] p(z,y) && \text{(Fubini)}\\
&= \sum_{z\in E} \mu_x(z) p(z,y) && (7.5)
\end{aligned}$$

with the convention $0 \times \infty = 0$. Thus, μ_x satisfies the equation of invariant measures. Yet, we still need to check that

$$\text{(a)} \quad \mu_x(y) > 0 \qquad \text{and} \qquad \text{(b)} \quad \mu_x(y) < \infty.$$

whenever y is in the same recurrence class as x. By iterating (7.5), we find that, for any $n \in \mathbb{N}$,

$$\mu_x(y) = \sum_{z\in E} \mu_x(z) p^n(z,y) \quad \text{for all } y \in E. \tag{7.6}$$

If y is in the same recurrence class as x, there exists n_0 such that $p^{n_0}(x,y) > 0$. Noticing that, by definition, $\mu_x(x) = 1$, we find that $\mu_x(y) \geq \mu_x(x)\, p^{n_0}(x,y) > 0$ which proves (a). On the other hand, there also exists n_1 such that $p^{n_1}(y,x) > 0$. Using again (7.6), we find that

$$1 = \mu_x(x) = \sum_{z\in E} \mu_x(z)p^{n_1}(z,x) \geq \mu_x(y)p^{n_1}(y,y)$$

which proves (b). □

Remark 7.9

If E is finite, the kernel p is a stochastic matrix. In particular, it has non-negative entries and the Perron-Frobenius Theorem tells us that there exists an eigenvector for the eigenvalue 1 (the spectral radius). Thus, we recover the existence of an invariant measure in that case.

Theorem 7.10

Assume that the chain is irreducible and recurrent. Then the invariant measure is unique, up to a constant multiplication.

Proof. Fix $x \in E$ and let μ_x be the invariant measure defined by (7.4). Let ν denote another invariant measure. We first prove that

$$\nu(y) \geq \nu(x)\mu_x(y) \quad \text{for all } y \in E. \tag{7.7}$$

In view of the monotone convergence theorem, we just need to show that, for any n,

$$\nu(y) \geq \nu(x)\,\mathbf{E}_x\Big[\sum_{i=0}^{n\wedge(T_x-1)} \mathbf{1}_{\{X_i=y\}}\Big] \quad \text{for all } y \in E. \tag{7.8}$$

Notice that, if $x = y$, then (7.8) holds trivially. Thus, let us assume $y \neq x$. We prove the result by induction. For $n = 0$, the identity is true because the right hand side is 0. Suppose that the inequality holds for n. Then

$$\begin{aligned}
\nu(y) &= \sum_{z\in E} \nu(z)\, p(z,\, y) \qquad \text{(stationarity)}\\
&\geq \nu(x) \sum_{z\in E} \mathbf{E}_x\Big[\sum_{i=0}^{n\wedge(T_x-1)} \mathbf{1}_{\{\mathbf{X}_i = z\}}\Big] p(z, y) \qquad \text{(induction assumption)}\\
&= \nu(x) \sum_{z\in E} \sum_{i=0}^{n} \mathbf{E}_x\Big[\mathbf{1}_{\{\mathbf{X}_i=z,\, i\leq T_x-1\}}\Big] p(z, y) \qquad \text{(Fubini)}\\
&= \nu(x) \sum_{z\in E} \sum_{i=0}^{n} \mathbf{E}_x\Big[\mathbf{1}_{\{\mathbf{X}_i=z,\, i\leq T_x-1\}}\, \mathbf{1}_{\{\mathbf{X}_{i+1}=y\}}\Big] \qquad \text{(Markov property using } \{i\leq T_x-1\}\in\mathscr{F}_i)\\
&= \nu(x) \sum_{i=0}^{n} \mathbf{E}_x\Big[\mathbf{1}_{\{i\leq T_x-1\}}\, \mathbf{1}_{\{\mathbf{X}_{i+1}=y\}}\Big]. \qquad \text{(Fubini)}\\
&= \nu(x)\, \mathbf{E}_x\Big[\sum_{i=0}^{n\wedge(T_x-1)} \mathbf{1}_{\{\mathbf{X}_{i+1}=y\}}\Big]\\
&= \nu(x)\, \mathbf{E}_x\Big[\sum_{i=1}^{(n+1)\wedge T_x} \mathbf{1}_{\{\mathbf{X}_{i}=y\}}\Big]\\
&= \nu(x)\, \mathbf{E}_x\Big[\sum_{i=0}^{(n+1)\wedge(T_x-1)} \mathbf{1}_{\{\mathbf{X}_{i}=y\}}\Big], \qquad (y\neq x)
\end{aligned}$$

which prove the equality for $n+1$. Thus, (7.7) holds. Using the invariance of ν, we can write, for any $n\in\mathbb{N}$,

$$\nu(x) \;=\; \sum_{z\in E} \nu(z)\, p^n(z,\, x) \;\geq\; \sum_{z\in E} \nu(x)\, \mu_x(z)\, p^n(z,\, x) \;=\; \nu(x)\, \mu_x(x) = \nu(x),$$

where we used that $\mu_x(x) = 1$ for the last equality. In particular, this means that the inequality in the previous equation is, in fact, an equality and therefore $\nu(z) = \nu(x)\, \mu_x(z)$ holds for all z such that $p^n(z,\, x) > 0$. Irreducibility ensures that for any $z\in E$, there exists $n\in\mathbb{N}$ such that $p^n(z, x) > 0$, which allows us to conclude that $\nu = c\,\mu_x$, with $c := \nu(x) < \infty$. Finally $c\neq 0$ because otherwise ν would be the null measure. This completes the proof of the theorem. □

The theorem above tells us that all the invariant measures of a irreducible Markov chain are proportional. Thus, either they all have infinite mass or they all have finite mass in which case there exists a unique one with unit mass (*i.e.* a probability distribution). This dichotomy separates irreducible recurrent Markov chains into two disctinct sub-classes: **null recurrent** chains and **positive recurrent** chains.

Corollary 7.11: definition of positive/null recurrence

Assume that the chain is irreducible and recurrent. Then

- either there exists an invariant probability measure π on E. In this case, we say that the chain is **positive recurrent**. Furthermore, we have

$$\mathbf{E}_x[T_x] = \frac{1}{\pi(x)} \qquad \text{for all } x \in E.$$

- or all the invariant measures have infinite mass. In this case, we say that the chain is **null recurrent**. Furthermore, we have

$$\mathbf{E}_x[T_x] = \infty, \qquad \text{for all } x \in E.$$

When E is a finite, then only the first situation occurs so every recurrent chain is positive recurrent.

Proof. Let $x \in E$. Let μ_x denote the invariant measure given by (7.4). Its total mass is

$$\mu_x(E) = \sum_{y \in E} \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1} \mathbf{1}_{\{\mathbf{X}_i = y\}}\Big] = \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1}\sum_{y\in E} \mathbf{1}_{\{\mathbf{X}_i = y\}}\Big] = \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1} 1\Big] = \mathbf{E}_x[T_x].$$

This shows that in the null recurrent case the expectation of the return time to any point is infinite. Suppose now that $\mathbf{X}$ is positive recurrent which means $\mu_x(E) = \mathbf{E}_x[T_x] \in]0, \infty[$. Let π denote the (unique) invariant probability measure. By proportionality, we have

$$\pi(y) = \frac{\mu_x(y)}{\mu_x(E)} = \frac{\mu_x(y)}{\mathbf{E}_x[T_x]} \quad \text{for all } y \in E.$$

Choosing $y = x$, we conclude that

$$\pi(x) = \frac{1}{\mathbf{E}_x[T_x]}.$$

□

Proposition 7.12

Assume that X is an irreducible Markov chain. We have the equivalence

$$X \text{ is positive recurrent} \quad \Longleftrightarrow \quad \text{there exists an invariant probability measure.}$$

Proof. We have $\Rightarrow$ by definiton of positive recurrence. So, we just need to prove that, if an invariant probability exists, then the chain is recurrent (hence positive recurrent). By irreductibility, it suffices to show that there is at least one recurrent state. Let π denote the invariant probability. Since it is not identically 0, we can fix $y \in E$ such that $\pi(y) > 0$. Recall that G denote the Green function of the chain. According to 3. of Proposition 6.4, we have

$$\sum_{n=0}^{\infty} p^n(x,y) = G(x,y) \le G(y,\, y),$$

Multiplying on both sides by $\pi(x)$ and then summing over $x \in E$, we obtain

$$\sum_{n=0}^{\infty} (\pi.p^n)(y) \le \pi(E)\, G(y,y) = G(y,y).$$

Since π is invariant, we have $\pi p^n = \pi$, $\forall n \in \mathbb{N}$, so that

$$\sum_{n=0}^{\infty} \pi(y) \leq G(y, y).$$

This shows that $G(y, y) = \infty$ since $\pi(y) > 0$. Thus, y is a recurrent state. $\square$

Remark 7.13

1. The proposition above shows that **an irreducible transient Markov Chain does not have any invariant probability distribution**. However, it can still admit invariant distribution with infinite mass. In fact, it may admit more than one. Take for instance the biased random walk on $\mathbb{Z}$ with transitions $\mathbf{P}(\mathbf{X}_{n+1} = \mathbf{X}_n+1) = 1-\mathbf{P}(\mathbf{X}_{n+1} = \mathbf{X}_n-1) = p$ with $p \neq 1/2$. We have already observed that both the counting measure and the measure $\mu(x) = \left(\frac{p}{1-p}\right)^x$ are invariant and are not proportional to each other.

2. There exist irreducible Markov chains that do not admit any invariant measure (but they are necessarily transient). Consider for instance the Markov chain on $\mathbb{N}$ with transition kernel

$$p(i, j) = \begin{cases} \alpha_i & \text{if } j = i+1, \\ 1-\alpha_i & \text{if } j = 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $(\alpha_i)_{i\geq 0}$ is a sequence of number in $]0, 1[$ such that $\sum(1-\alpha_i) < \infty$. If ν is is a invariant measure, then it must satisfy the balance equations:

$$\nu(0) = \sum_{i=0}^{\infty}(1-\alpha_i)\nu(i) \quad \text{and} \quad \nu(i+1) = \nu(i)\alpha_i \text{ for all } i \geq 0.$$

Thus, we deduce that $\nu(i) = \nu(0)\prod_{j=0}^{i-1}\alpha_j$. Assuming that $\nu(0) \neq 0$ (otherwise $\nu = 0$) and replacing $\nu(i)$ by its expression inside the balance equation for $\nu(0)$ and then simplifying by $\nu(0)$ on both sides, we find that

$$1 = \sum_{i=0}^{\infty}(1-\alpha_i)\prod_{j=0}^{i-1}\alpha_j = 1 - \prod_{j=0}^{\infty}\alpha_j < 1$$

which leads to a contradiction. This shows that there is no invariant measure.

Exercice 7.14

Show that, for any any $n \in \mathbb{N} \cup \{+\infty\}$, there exists an irreducible transient Markov chain on some space E such that the vector space generated by the invariant measures has dimension exactly n. [Hint: case $n = 0$ is treated in the remark above. For $1 \leq n < \infty$, consider a biased random walk on n copy of $\mathbb{N}$ joined at 0. For $n = \infty$, consider a biased random walk on a binary tree.]

Corollary 7.15

Let X be a random walk on an electrical network $(E, \mathcal{E}, c)$. Suppose that the chain is irreducible (equivalently that the graph is connected). Then

$$X \text{ is positive recurrent} \quad \Longleftrightarrow \quad \sum_{x,y\in E} c(x,y) < \infty.$$

Proof. We already noticed that π defined by (7.1) is an invariant measure. Its mass is $\sum_{x\in E}\pi(x) = \sum_{x,y\in E} c(x,y)$. Thus, if this quantity is finite, then the walk is positive recurrent thank to Proposition 7.12. Otherwise, it is either transient or null recurrent but cannot be positive recurrence by unicity of the invariant mesure up to a multiplicative factor. $\square$

Remark 7.16

The simple random walk on a graph $(E, \mathcal{E})$ is a particular case of random walk on an electrical network where the conductances are constant: $c(x,y) = 1$ for all pair of neighbors (x,y). Thus, if the graph is infinite and connected, the walk is either transient or null recurrent but it cannot be positive recurrent. In particular, **the simple random walk on $\mathbb{Z}^d$ is null recurrent for $\mathbf{d = 1, 2}$ and transient for $\mathbf{d \geq 3}$**.

Ehrenfest Urns. We conclude this section with a striking numerical application of the theory to the Ehrenfest urn model. Recall that we have a box with two compartments that contains a total number of N molecules and that, at each step, one molecule chosen uniformly switches its compartment. Since the chain is irreducible and the state space is finite, it is necessarily positive recurrent. We have already seen that its invariant probability distribution is the Binomial(N,1/2) distribution (which is reversible for the chain).

Suppose now that $N = 10^{23}$ (about the Alvogadro number) and that 10^{10} molecules switch places every second. Using Corollary 7.11, we can explicitly compute the expected time between two observations when both compartments have exactly the same number of molecules:

$$\mathbf{E}_{N/2}[T_{N/2}] = \frac{1}{\mathbf{P}(\text{Binom}(N,1/2) = N/2)} = \frac{2^N}{\binom{N}{N/2}} \simeq \sqrt{\frac{\pi}{2}N} \simeq 40 \text{ seconds}.$$

Similarly, we can compute the expected time between two observations when the first compartment is empty:

$$\mathbf{E}_0[T_0] = \frac{1}{\mathbf{P}(\text{Binom}(N,1/2) = 0)} = \frac{2^N}{\binom{N}{N}} = 2^{10^{23}} \gg \text{lifetime of the universe !}$$

These computation provide a convincing argument to explain why some microscopic reversible recurrent processes in thermodynamics/statistical mechanics seem irreversible: the recurrence time is so huge that returns to low probability states are never observed in practice!

Transience	Recurrence	
	null recurrence	positive recurrence
each site is visited only finitely many times a.s. $\mathbf{P}_x(T_x < \infty) < 1$	each site is visited infinitely often a.s. $\mathbf{P}_x(T_x < \infty) = 1$	
There can be any number of invariant measures. (possibly 0).	An invariant measure exist and is unique up to a multiplicative constant	
Any invariant measure μ must have infinite mass *i.e.* $\mu(E) = \infty$.	The invariant measures have infinite mass *i.e.* $\mu(E) = \infty$.	The invariant measures have finite mass *i.e.* $\mu(E) < \infty$. $\Rightarrow$ There exists a unique invariant probability distribution π.
$\mathbf{E}_x[T_x] = \infty$	$\mathbf{E}_x[T_x] = \infty$	$\mathbf{E}_x[T_x] = \frac{1}{\pi(x)} < \infty$

Figure 1.1: Summary of the properties of an irreducible Markov chain depending on its type.

8. Limit theorems

In the previous section, we have seen that, for an irreducible positive recurrent Markov chain, its invariant probability measure quantify how often each state is visited. We will now complete this picture by studying the asymptotic distribution of X_n for large n. We will see that it converges, under reasonable assumptions, toward the invariant measure.

The first result states a convergence for the Cesàro sums.

Theorem 8.1

Suppose that the Markov chain X is irreducible and recurrent. Let ν be an arbitrary initial distribution. Let μ be an invariant measure for X. For any functions $f : E \to \mathbb{R}_+$ and $g : E \to \mathbb{R}_+$ such that $0 < \sum_E g(x)\mu(x) < \infty$, we have

$$\frac{\sum_{i=0}^n f(X_i)}{\sum_{i=0}^n g(X_i)} \underset{n\to\infty}{\longrightarrow} \frac{\sum_E f(x)\mu(x)}{\sum_E g(x)\mu(x)}, \qquad \mathbf{P}_\nu\text{-a.s.}$$

If the chain is positive recurrent, then the invariant measures have finite mass and so can choose $g = 1$ which proves that:

Corollary 8.2

Assume that the chain is irreducible and positive recurrent. Let π be the unique invariant probability. For any $f : E \to \mathbb{R}_+$, we have

$$\frac{1}{n}\sum_{i=0}^n f(X_i) \underset{n\to\infty}{\longrightarrow} \sum_E f(x)\pi(x), \qquad \mathbf{P}_\nu\text{-a.s.}$$

Proof of Theorem 8.1. Fix $x \in E$. In view of equation (4.2), we just need to prove the result when the chain starts from $\nu = \delta_x$. Define by induction the sequence of stopping times

$$\begin{cases} T^{(0)} & := & 0, \\ T^{(n+1)} & := & \inf\{i > T^{(n)} : X_i = x\}. \end{cases}$$

With this definition, we have $T^{(1)} = T_x$. Notice that all these times are finite $\mathbf{P}_x$-a.s. since the chain is assumed to be recurrent. Define also

$$\xi_n := \sum_{i=T^{(n)}}^{T^{(n+1)}-1} f(X_i), \qquad \text{for } n \in \mathbb{N}.$$

Using the strong Markov property. It is easily checked by induction that $(\xi_n, n \geq 0)$ is a sequence of i.i.d. random variables. Let μ_x be the measure in Theorem 7.7:

$$\mu_x(y) := \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1} \mathbf{1}_{\{X_i=y\}}\Big], \qquad \text{for } y \in E.$$

Since the chain is irreducible and recurrent, we know that $\mu_x = c\mu$ for some $c > 0$ (Theorem 7.10). Thus, we get

$$\mathbf{E}_x[\xi_0] = \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1} f(X_i)\Big] = \mathbf{E}_x\Big[\sum_{i=0}^{T_x-1}\sum_{y\in E} f(y)\,\mathbf{1}_{\{X_i=y\}}\Big]$$

$$= \sum_{y\in E} f(y)\mathbf{E}_x\Big[\sum_{i=0}^{T_x-1}\mathbf{1}_{\{X_i=y\}}\Big] = \sum_{y\in E} f(y)\,\mu_x(y) = c\sum_{y\in E} f(y)\,\mu(y)$$

Applying the strong law of large numbers, we find that

$$\frac{1}{n}\sum_{i=0}^{n-1}\xi_i \underset{n\to\infty}{\longrightarrow} c\sum_{y\in E} f(y)\,\mu(y), \qquad \mathbf{P}_x\text{-a.s.}$$

Define now $N_n(x) := \sum_{i=1}^{n}\mathbf{1}_{\{X_i=x\}}$ for any $n \geq 0$. This quantity represents the number of hits at state x by the chain up to time n; as such, we have $T^{(N_n(x))} \leq n < T^{(N_n(x)+1)}$ and therefore

$$\sum_{i=0}^{n} f(X_i) \leq \sum_{i=0}^{T^{(N_n(x)+1)}-1} f(X_i) = \sum_{j=0}^{N_n(x)} \xi_j,$$

which implies (because $N_n(x)\to\infty$ as $n\to\infty$) that

$$\limsup_{n\to\infty}\frac{\sum_{i=0}^{n} f(X_i)}{N_n(x)} \leq \mathbf{E}[\xi_0] = c\sum_{y\in E} f(y)\,\mu(y) \qquad \mathbf{P}_x\text{-a.s.}$$

Similarly, we have

$$\sum_{i=0}^{n} f(X_i) \geq \sum_{i=0}^{T^{(N_n(x))}} f(X_i) \geq \sum_{i=0}^{T^{(N_n(x))}-1} f(X_i) = \sum_{j=0}^{N_n(x)-1} \xi_j,$$

and therefore

$$\liminf_{n\to\infty}\frac{\sum_{i=0}^{n} f(X_i)}{N_n(x)} \geq \mathbf{E}[\xi_0] = c\sum_{y\in E} f(y)\,\mu(y) \qquad \mathbf{P}_x\text{-a.s.}$$

We conclude that

$$\frac{\sum_{i=0}^{n} f(X_i)}{N_n(x)} \underset{n\to\infty}{\longrightarrow} c\sum_{y\in E} f(y)\,\mu(y) \qquad \mathbf{P}_x\text{-a.s.}$$

The same result holds with g in place of f. Taking the ratio, the constant c and the term $N_n(x)$ cancel so we obtain the convergence stated in the theorem. $\square$

Remark 8.3

Let $F, G \subset E$ such that $\mu(G) < \infty$. Applying Theorem 8.1 with $f(z) = \mathbf{1}_{z\in F}$ and $g(z) = \mathbf{1}_{z\in G}$, we observe that

$$\frac{\text{number of visits of the set } F \text{ before time } n}{\text{number of visits of the set } G \text{ before time } n} \underset{n\to\infty}{\longrightarrow} \frac{\mu(F)}{\mu(G)} \qquad \mathbf{P}_\nu\text{-a.s.} \tag{8.1}$$

Choose $F = \{x\}$ for $x \in E$. If the chain is positive recurrent, then $\mu(E) < \infty$ so we can take

$G = E$ which yields

$$\frac{\text{number of visits of } x \text{ before time } n}{n} \underset{n\to\infty}{\longrightarrow} \pi(x) > 0 \qquad \mathbf{P}_\nu\text{-a.s.} \tag{8.2}$$

where π is the unique invariant probability measure. On the other hand, if X is null recurrent, then $\mu(E) = \infty$. By taking larger and larger (finite) subset G, we conclude that

$$\frac{\text{number of visits of } x \text{ before time } n}{n} \underset{n\to\infty}{\longrightarrow} 0 \qquad \mathbf{P}_\nu\text{-a.s.} \tag{8.3}$$

(8.2) and (8.3) explain the terminology "positive recurrent" and "null recurrent": in the first case, the walk spends a positive fraction of its time at each site of E whereas it spends a null fraction in the second case.

Corollary 8.4

Suppose that the chain X is irreducible.

- If X positive recurrent, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} p^k(x,y) = \pi(y) > 0 \qquad \text{for all } x, y \in E,$$

  where π is the unique invariant probability measure.

- If X is null recurrent or transient

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} p^k(x,y) = 0 \qquad \text{for all } x, y \in E.$$

Proof. We observe that $\frac{\text{number of visits of } y \text{ before time } n}{n} \le 1$ for all n. Suppose that X is recurrent. In view of (8.2) and (8.3) and making use of the dominated convergence theorem we deduce that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} p^k(x,y) &= \lim_{n\to\infty} \mathbf{E}_x\Big[\frac{\text{number of visits of } y \text{ before time } n}{n}\Big] \\ &= \begin{cases} \pi(y) & \text{if } X \text{ is positive recurrent,} \\ 0 & \text{if } X \text{ is null recurrent.} \end{cases} \end{aligned}$$

Finally, if the chain is transient, then the Green function $G(x,y)$ is finite so the Cesàro sum also converges to 0. □

The convergence of the Cesàro sum $\frac{1}{n}\sum_{k=0}^{n} p^k(x,y)$ does not guaranty the convergence of $p^n(x,\, y)$ itself. Consider for instance the Markov chain on $E = \{0,\, 1\}$ with transition kernel $p(0,1) = p(1,0) = 1$ and $p(0,0) = p(1,1) = 0$ *i.e.* the chain jumps back and forth (deterministically) between 0 to 1. Clearly, we have

$$p^n(0,1) = \begin{cases} 0 & \text{if } n \text{ is even,} \\ 1 & \text{if } n \text{ is odd.} \end{cases}$$

Hence it does not converge (but the Cesàro sum converges to $1/2$). We will now see that this "cyclic" phenomenon is the only thing that can prevent $p^n(x,y)$ from converging. To do so, we introduce the notion of periodicity.

Definition 8.5

Let $x \in E$. We define the set

$$\mathcal{D}_x := \{n \geq 0 : \, p^n(x,x) > 0\}.$$

The greatest common divisor of $\mathcal{D}_x$, denoted by d_x, is called the **period** of x.

Proposition 8.6

Assume that the chain is irreducible.

1. All states have the same period d.
2. Suppose that $d = 1$. For each $x \in E$, there exists $n_0 \in \mathbb{N}$ such that,

$$p^n(x,x) > 0 \qquad \text{for all } n \geq n_0$$

Proof. 1. Fix $x, y \in E$ and denote by d_x, d_y their period. By irreducibility, there exist k, l such that $p^k(x,y) > 0$ and $p^l(y,x) > 0$. For any $m \in \mathcal{D}_y$, we have $p^m(y,y) > 0$ so that

$$p^{m+k+l}(x,x) \geq p^k(x,y)p^m(y,y)p^l(y,x) > 0.$$

This means that $m+k+l \in \mathcal{D}_x$ *i.e.* d_x divides $m+k+l$. In particular, $0 \in \mathcal{D}_y$ thus d_x divides $k+l$. But then, d_x divides any element $m \in \mathcal{D}_y$ so it divides its gcd d_y. By symmetry, $d_x = d_y$.

2. Fix $x \in E$. Let us first remark that, if $u, v \in \mathcal{D}_x$, then

$$p^{u+v}(x,x) \geq p^u(x,x)p^v(x,x)$$

which means that $u+v \in \mathcal{D}_x$. Thus, $\mathcal{D}_x$ is stable by addition (it is a semigroup). Suppose now that $d = 1$. By Bezout identity, there exist $u_1, \dots u_k \in \mathcal{D}_x$ and $a_1, \dots a_k \in \mathbb{Z}$ such that

$$\sum_{i=1}^{k} a_i u_i = 1.$$

(here, we apply Bezout identity to the infinite set but $\mathcal{D}_x$ is infinite but this is not a problem because $\gcd(\mathcal{D}_x \cap [0,N]) = 1$ for all N large enough since it is an integer-valued decreasing function converging to 1). Let

$$q = \sum_{i | a_i \geq 0} a_i u_i \qquad \text{and} \qquad q' = -\sum_{i | a_i < 0} a_i u_i$$

so that $q - q' = 1$. Moreover, $q, q' \in \mathcal{D}_x$ thanks to the semigroup property of $\mathcal{D}_x$. We choose $n_0 = q'^2$. Let $n \geq n_0$. The euclidian division of n by q' gives $n = bq' + r$ for some $b \geq q'$ and $r < q'$. In particular, $b > r$. But then, using that $1 = q - q'$, we can write

$$n = bq' + r = bq' + r(q - q') = (b - r)q' + rq.$$

Since $(b - r)$ and r are both positive and q and q' belong to $\mathcal{D}_x$, we deduce from the semigroup property that $n \in \mathcal{D}_x$ which means $p^n(x, x) > 0$ as requested. $\square$

Definition 8.7

The common period d of all states of an irreducible Markov chain is called the **period of the chain**. In the case $d = 1$. We say that the chain is **aperiodic**.

Remark 8.8

1. The simple random on $\mathbb{Z}^d$, $d \geq 1$ has period 2.
2. Any irreducible chain for which there exists a state x such that $p(x, x) > 0$ is aperiodic.

Exercice 8.9

Let X be a Markov chain on some electrical network $(E, \mathcal{E}, c)$. The graph $(E, \mathcal{E})$ is said to be **bipartite** if its vertices can be colored with two colors in such way that no two adjacent vertices share the same color. Show that, if $(E, \mathcal{E})$ is bipartite, then the chain has period 2 and otherwise it is aperiodic. In particular, since any irreducible reversible Markov chain admits an electrical network representation, its period is either 1 or 2.

Theorem 8.10

Assume that the chain is irreducible, positive recurrent, and aperiodic. Then for all $x \in E$,

$$\sum_{y \in E} |p^n(x, y) - \pi(y)| \to 0, \qquad n \to \infty,$$

where π is the unique invariant probability.

Remark 8.11

The convergence stated in the theorem above is called **convergence in total variation norm**. This is a strong convergence which implies in particular the convergence in law of X_n to π.

Proof of Theorem 8.10. The formula

$$\hat{\mathbf{p}}(\mathbf{x},\,\mathbf{y}) := p(x_1,\,y_1)\,p(x_2,\,y_2), \qquad \mathbf{x} := (x_1,\,x_2) \in E^2,\ \mathbf{y} := (y_1,\,y_2) \in E^2,$$

defines a transition probability on $E \times E$. Let $((X_n^1,\,X_n^2)_{n\in\mathbb{N}},\ (\widehat{\mathbf{P}}_{\mathbf{x}})_{\mathbf{x}\in E^2})$ be the associated canonical Markov chain. We remark that $\hat{\mathbf{p}}$ is irreducible. Indeed, if $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ are two states of $E \times E$, by the previous proposition, there exist n_1 and n_2 such that $p^n(x_1,\,y_1) > 0$, for all $n \geq n_1$ and $p^n(x_2,\,y_2) > 0$, for all $n \geq n_2$. Setting $m = \max\{n_1,\,n_2\}$, we conclude that $\hat{\mathbf{p}}^m(\mathbf{x},\,\mathbf{y}) > 0$ which shows the irreducibility. We also note that $\pi \otimes \pi$ is an invariant probability measure for $\hat{\mathbf{p}}$:

$$\begin{aligned}
\sum_{\mathbf{x}\in E^2} (\pi\otimes\pi)(\mathbf{x})\,\hat{\mathbf{p}}(\mathbf{x},\,\mathbf{y}) &= \sum_{\mathbf{x}\in E^2} \pi(x_1)\pi(x_2)\,p(x_1,\,y_1)p(x_2,\,y_2) \\
&= \sum_{x_1\in E} \pi(x_1)\,p(x_1,\,y_1) \sum_{x_2\in E} \pi(x_2)\,p(x_2,\,y_2) \\
&= \pi(y_1)\pi(y_2) = (\pi\otimes\pi)(\mathbf{y}).
\end{aligned}$$

The chain being irreducible with an invariant probability, it follows from Proposition 7.12 that it is positive recurrent. In particular, the stopping time

$$N := \inf\{n \in \mathbb{N} :\ X_n^1 = X_n^2\}.$$

is a.s. finite under an initial distribution (because N is smaller than the return time to any diagonal state (z, z) which is finite). The reason for which we have introduced this bi-dimensional chain is that we can write

$$p^n(x,\,y) - \pi(y) = \widehat{\mathbf{P}}_{\pi\otimes\delta_x}\big(X_n^2 = y\big) - \widehat{\mathbf{P}}_{\pi\otimes\delta_x}\big(X_n^1 = y\big) = \widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}}\big].$$

Splitting this equality depending on whether $N \leq n$ or $N \geq n$, we get that

$$\begin{aligned}
p^n(x,\,y) - \pi(y) &= \widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})\big] \\
&\quad + \sum_{k=0}^{n}\sum_{z\in E} \widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N=k,\,X_k^1=X_k^2=z\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})\big]. \qquad (8.4)
\end{aligned}$$

Using the Markov property at time $n-k$, we find that, for $k \in \{0,\,1,\,\cdots,\,n\}$ and $z \in E$,

$$\begin{aligned}
\widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N=k,\,X_k^1=X_k^2=z\}}\mathbf{1}_{\{X_n^2=y\}}\big] &= \widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N=k,\,X_k^1=X_k^2=z\}}\big]\,p^{n-k}(z,\,y) \\
&= \widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N=k,\,X_k^1=X_k^2=z\}}\mathbf{1}_{\{X_n^1=y\}}\big],
\end{aligned}$$

which means that the double sum on the right-hand side of (8.4) vanishes. Accordingly,

$$\begin{aligned}
\sum_{y\in E} |p^n(x,\,y) - \pi(y)| &= \sum_{y\in E} \big|\widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N>n\}}(\mathbf{1}_{\{X_n^2=y\}} - \mathbf{1}_{\{X_n^1=y\}})\big]\big| \\
&\leq \sum_{y\in E} \widehat{\mathbf{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{N>n\}}(\mathbf{1}_{\{X_n^2=y\}} + \mathbf{1}_{\{X_n^1=y\}})\big] \\
&= 2\,\widehat{\mathbf{P}}_{\pi\otimes\delta_x}(N > n),
\end{aligned}$$

which converges to 0 when $n \to \infty$, since N is finite $\hat{\mathbf{P}}_{\pi\otimes\delta_x}$-a.s. $\square$