1

First steps

1.1 Preliminaries

This book focuses on a class of random evolutions, in discrete time (by successive steps) on a discrete state space (finite or countable, with isolated elements), which satisfy a fundamental assumption, called the *Markov property*. This property can be described informally as follows: the evolution "forgets" its past and is "regenerated" at each step, retaining as sole past information for its future evolution its present state. The probabilistic description of such an evolution requires

- a law (probability measure) for drawing its initial state and
- a family of laws for drawing iteratively its state at the "next future instant" given its "present state," indexed by the state space.

Such a random evolution will be called a *Markov chain*.

Precise definitions can be found in the Appendix, Section A.3, but we give now the probabilistic framework. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ will be considered throughout. When $\mathcal{V}$ is discrete, usually its measurable structure is given by the collection of all subsets, and all functions with values in $\mathcal{V}$ are assumed to be measurable.

A random variable (r.v.) with values in a measurable state space $\mathcal{V}$ is a measurable function

$$X : \omega \in \Omega \mapsto X(\omega) \in \mathcal{V}.$$

Intuitively, the output $X(\omega)$ varies randomly with the input ω, which is drawn in Ω according to $\mathbb{P}$, and the measurability assumptions allow to assign a probability to events defined through X.

For the random evolutions under investigation, the natural random elements are sequences $(X_n)_{n\in\mathbb{N}}$ taking values in the same discrete state space $\mathcal{V}$, which are called

Markov Chains: Analytic and Monte Carlo Computations, First Edition. Carl Graham.

(random) *chains* or (discrete time) *processes*. Each X_n should be an r.v., and its law π_n on the discrete space $\mathcal{V}$ is then given by

$$\pi_n(x) = \pi_n(\{x\}) = \mathbb{P}(X_n = x) \ , \qquad x \in \mathcal{V} \ ,$$
$$\pi_n(A) = \mathbb{P}(X_n \in A) = \sum_{x \in A} \pi_n(x) \ , \qquad A \subset \mathcal{V} \ ,$$

and hence can be identified in a natural way with $(\pi_n(x))_{x \in \mathcal{V}}$.

Finite-dimensional marginals More generally, for any $k \geq 1$ and $n_1, \dots, n_k$ in $\mathbb{N}$, the random vector

$$(X_{n_1}, \dots, X_{n_k})$$

takes values in the discrete space $\mathcal{V}^k$, and its law $\pi_{n_1,\dots,n_k}$ can be identified with the collection of the

$$\pi_{n_1,\dots,n_k}(x_1, \dots, x_k) = \mathbb{P}(X_{n_1} = x_1, \dots, X_{n_k} = x_k) \ , \qquad x_1, \dots, x_k \in \mathcal{V}.$$

All these laws for $k \geq 1$ and $0 \leq n_1 < \cdots < n_k$ constitute the family of the *finite-dimensional marginals* of the chain $(X_n)_{n \geq 0}$ or of its law.

Law of the chain The r.v.

$$(X_n)_{n \geq 0} \ : \ \omega \mapsto (X_n(\omega))_{n \geq 0}$$

takes values in $\mathcal{V}^{\mathbb{N}}$, which is *uncountable* as soon as $\mathcal{V}$ contains at least two elements. Hence, its law *cannot*, in general, be defined by the values it takes on the elements of $\mathcal{V}^{\mathbb{N}}$. In the Appendix, Section A.3 contains some mathematical results defining the law of $(X_n)_{n \geq 0}$ from its finite-dimensional marginals.

Section A.1 contains some more elementary mathematical results used throughout the book, and Section A.2 a discussion on the total variation norm and on weak convergence of laws.

1.2 First properties of Markov chains

1.2.1 Markov chains, finite-dimensional marginals, and laws

1.2.1.1 First definitions

We now provide rigorous definitions.

Definition 1.2.1 *Let $\mathcal{V}$ be a discrete state space. A matrix $P = (P(x,y))_{x,y \in \mathcal{V}}$ is a transition matrix on $\mathcal{V}$, or also a Markovian or stochastic matrix, if*

$$P(x,y) \geq 0 \ , \qquad \sum_{y \in \mathcal{V}} P(x,y) = 1.$$

A sequence $(X_n)_{n\geq 0}$ of $\mathcal{V}$-valued random variables is a Markov chain on $\mathcal{V}$ with matrix P and initial law π_0 if, for every n in $\mathbb{N}$ and $x_0, \dots, x_n$ in $\mathcal{V}$,

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \pi_0(x_0)P(x_0, x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n).$$

Note that, by iteration, $(X_n)_{n\geq 0}$ is a Markov chain on $\mathcal{V}$ with matrix P if and only if for every n in $\mathbb{N}$ and $x_0, \dots, x_n, y$ in $\mathcal{V}$,

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = y) = \mathbb{P}(X_0 = x_0, \dots, X_n = x_n)P(x_n, y),$$

and that this is trivially true if $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = 0$.

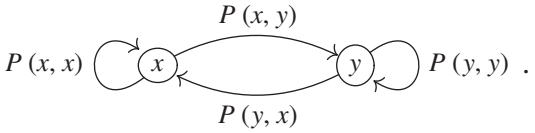
Markov chain evolution A family $P(x, \cdot)$ of laws on $\mathcal{V}$ indexed by $x \in \mathcal{V}$ is defined by

$$P(x, A) = \sum_{x \in A} P(x, y) \ , \qquad A \subset \mathcal{V}.$$

The evolution of $(X_n)_{n\geq 0}$ can be obtained by independent draws, first of X_0 according to π_0, and then iteratively of X_{n+1} according to $P(X_n, \cdot)$ for $n \geq 0$ without taking any further notice of the evolution before the present time n or of its actual value.

Inhomogeneous Markov chains A more general and complex evolution can be obtained by letting the law of the steps depend on the present instant of time, that is, using the analogous formulae with $P(n; x_n, y)$ instead of $P(x_n, y)$; this corresponds to a *time-inhomogeneous* Markov chain, but we will seldom consider this generalization.

Markov chain graph The graph of the transition matrix P, or of a Markov chain with matrix P, is the oriented marked graph with nodes given by the elements of $\mathcal{V}$ and directed links given by the ordered pairs (x, y) of elements of $\mathcal{V}$ such that $P(x, y) > 0$ marked by the value of $P(x, y)$. The restriction to the graph to $x \neq y$ in $\mathcal{V}$ is of the form [if $P(x, x)P(x, y)P(y, x)P(y, y) \neq 0$]



.

The graph and the matrix are equivalent descriptors for the random evolution. The links from x to x in the graph are redundant as they are marked by $P(x, x) = 1 - \sum_{y \neq x} P(x, y) > 0$, but illustrate graphically the possible transitions from x.

1.2.1.2 Conditional formulations

The last formula in Definition 1.2.1 can be written as

$$\mathbb{P}(X_{n+1} = y \mid X_0 = x_0, \dots, X_n = x_n) = P(x_n, y) \tag{1.2.1}$$

which is often used as the definition. Moreover, if f is nonnegative or bounded then

$$\mathbb{E}(f(X_{n+1}) \mid X_0 = x_0, \dots, X_n = x_n) = \mathbb{E}_{x_n}(f(X_1)) = \sum_{y \in \mathcal{V}} P(x_n, y) f(y).$$

For the sake of mathematical efficiency and simplicity, nonconditional expressions will be stressed, before possibly being translated into equivalent conditional formulations. As an example, Definition 1.2.1 immediately yields by summing over $x_0, \dots, x_{n-1}$ that

$$\mathbb{P}(X_{n+1} = y \mid X_n = x_n) = \frac{\mathbb{P}(X_n = x_n, X_{n+1} = y)}{\mathbb{P}(X_n = x_n)} = P(x_n, y) \ ,$$

which is not quite so obvious starting from (1.2.1).

1.2.1.3 Initial law, instantaneous laws

For a Markov chain $(X_n)_{n \geq 0}$, the law of π_n of X_n is called the instantaneous law at time n and π_0 the initial law. The notations $\mathbb{P}$ and $\mathbb{E}$ implicitly imply that π_0 is given and arbitrary, $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ for some law μ on $\mathcal{V}$ indicate that $\pi_0 = \mu$, and $\mathbb{P}_x$ and $\mathbb{E}_x$ indicate that $X_0 = x$. By linearity,

$$\mathbb{P}_\mu = \sum_{x \in \mathcal{V}} \mu(x) \mathbb{P}_x \ , \qquad \mathbb{E}_\mu = \sum_{x \in \mathcal{V}} \mu(x) \mathbb{E}_x .$$

A frequent abuse of notation is to write $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$, and so on.

Lemma 1.2.2 *Let $(X_n)_{n \geq 0}$ be a Markov chain with matrix P and initial law π_0. Then, $\mathbb{P}(X_{n+1} = y \mid X_n = x) = P(x, y)$ for n in $\mathbb{N}$ and x, y in $\mathcal{V}$, and the instantaneous laws $\pi_n = (\pi_n(x))_{x \in \mathcal{V}}$ are given by*

$$\pi_n(x) = \mathbb{P}(X_n = x) = \sum_{x_0, \dots, x_{n-1} \in \mathcal{V}} \pi_0(x_0) P(x_0, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x)$$

or in matrix notation

$$\pi_n = \pi_{n-1} P = \cdots = \pi_0 P^n .$$

Moreover, $(X_{nk})_{k \geq 0}$ is a Markov chain with matrix the nth matrix power of P

$$P^n = (P^n(x, y))_{x, y \in \mathcal{V}} .$$

Proof: This follows readily from Definition 1.2.1. ∎

1.2.1.4 Law on the canonical space of the chain

The notions in the Appendix, Section A.3.4, will now be used.

Definition 1.2.1 is actually a statement on the law of the Markov chain $(X_n)_{n \geq 0}$, which it characterizes by giving an explicit expression for its finite-dimensional marginals in terms of its initial law π_0 and transition matrix P.

Indeed, some rather simple results in measure theory show that there is uniqueness of a law on the canonical probability space $\mathcal{V}^{\mathbb{N}}$ with product σ-field having a given finite-dimensional marginal collection.

It is immediate to check that this collection is consistent [with respect to (w.r.t.) projections] and then the Kolmogorov extension theorem (Theorem A.3.10) implies that there is existence of a law $\mathbb{P}_{\pi_0}$ on the canonical probability space $\mathcal{V}^{\mathbb{N}}$ with the product σ-field such that the canonical (projection) process $(X_n)_{n\geq 0}$ has the given finite-dimensional marginal collection, which hence is a Markov chain with initial law π_0 and transition matrix P (see Corollary A.3.11).

The Kolmogorov extension theorem follows from a deep and general result in measure theory, the Caratheodory extension theorem.

1.2.2 Transition matrix action and matrix notation

1.2.2.1 Nonnegative and signed measures, total variation measure, andnorm

A (nonnegative) measure μ on $\mathcal{V}$ is defined by (and can be identified with) a collection $(\mu(x))_{x\in\mathcal{V}}$ of nonnegative real numbers and, in the sense of nonnegative series,

$$\mu : A \subset \mathcal{V} \mapsto \mu(A) := \sum_{x\in A} \mu(x) \in [0,\infty] := \mathbb{R}_+ \cup \{\infty\}.$$

A measure μ is *finite* if its total mass $\mu(\mathcal{V})$ is finite and then $\mu(A) < \infty$ for all $A \subset \mathcal{V}$. A measure is a *probability measure*, or a *law*, if $\mu(\mathcal{V}) = 1$.

For r in $\mathbb{R}$, let $r^+ = \max(r, 0)$ and $r^- = \max(-r, 0)$ denote the nonnegative and nonpositive parts of r, which satisfy $r = r^+ - r^-$ and $|r| = r^+ + r^-$.

For $\mu = (\mu(x))_{x\in\mathcal{V}}$ with $\mu(x) \in \mathbb{R}$, the measures μ^+, μ^-, and $|\mu|$ can be defined term wise. Then,

$$\mu = \mu^+ - \mu^-$$

is the minimal decomposition of μ as a difference of (nonnegative) measures, which have disjoint supports, and

$$|\mu| = \mu^+ + \mu^-$$

is called the *total variation measure* of μ.

If μ is such that $|\mu|$ is finite (equivalently, if both μ^+ and μ^- are finite), then we can extend it to a *signed measure* μ acting on subsets of $\mathcal{V}$ by setting, in the sense of absolutely converging series,

$$\mu : A \subset \mathcal{V} \mapsto \mu(A) := \mu^+(A) - \mu^-(A) = \sum_{x\in A} \mu(x) \in \mathbb{R} ,$$

and we can define its *total variation norm* by

$$\|\mu\|_{var} = |\mu|(\mathcal{V}) = \sum_{x\in\mathcal{V}} |\mu(x)| < \infty.$$

Note that $\mu(A) \leq |\mu|(A) \leq \|\mu\|_{var}$ for all $A \subset \mathcal{V}$.

The space $\mathcal{M} := \mathcal{M}(\mathcal{V})$ of all signed measures, furnished with the total variation norm, is a Banach space, which is isomorphic to the Banach space $\ell^1 := \ell^1(\mathcal{V})$ of summable sequences with its natural norm.

Probability measures or laws The space of probability measures

$$\mathcal{M}_+^1 := \mathcal{M}_+^1(\mathcal{V}) = \{\mu \in \mathcal{M} : \mu \geq 0,\ \|\mu\|_{var} = 1\}$$

is the intersection of the cone of nonnegative measures with the unit sphere. It is a closed subset of $\mathcal{M}$ and hence is complete for the induced metric.

Some properties of $\mathcal{M}$ and $\mathcal{M}_+^1$ are developed in the Appendix, Section A.2. Note that, according to the definition taken here, nonnegative measures with infinite mass are *not* signed measures.

Complex measures Spectral theory naturally involves complex extensions. For its purposes, complex measures can be readily defined, and the corresponding space $\mathcal{M}(\mathcal{V}, \mathbb{C})$, where the modulus in $\mathbb{C}$ is again denoted by $|\cdot|$, allows to define a total variation measure $\|\mu\|$ and total variation norm $\|\mu\|_{var} = \|\mu\|(\mathcal{V})$ for μ in $\mathcal{M}(\mathcal{V}, \mathbb{C})$. The Banach space is isomorphic to $\ell^1(\mathcal{V}, \mathbb{C})$. The real and imaginary parts of a complex measure are signed measures.

1.2.2.2 Line and column vectors, measure-function duality

In matrix notation, the functions f from $\mathcal{V}$ to $\mathbb{R}$ are considered as *column* vectors $(f(x))_{x\in\mathcal{V}}$, and nonnegative or signed measures μ on $\mathcal{V}$ as *line* vectors $(\mu(x))_{x\in\mathcal{V}}$, of infinite lengths if $\mathcal{V}$ is infinite. The integral of a function f by a measure μ is denoted by μf, in accordance with the matrix product

$$\mu f = \begin{pmatrix} \cdots & \mu(x) & \cdots \end{pmatrix} \begin{pmatrix} \vdots \\ f(x) \\ \vdots \end{pmatrix} = \sum_{x\in\mathcal{V}} \mu(x) f(x)\ ,$$

defined in $[0, \infty]$ in the sense of nonnegative series if $\mu \geq 0$ and $f \geq 0$ and in $\mathbb{R}$ in the sense of absolutely converging series if $\mu \in \mathcal{M}$ and $f \in L^\infty = L^\infty(\mathcal{V}) = L^\infty(\mathcal{V}, \mathbb{R})$, the Banach space of bounded functions on $\mathcal{V}$ with the uniform norm.

For A subset of $\mathcal{V}$, the indicator function $\mathbb{1}_A$ is defined by

$$\mathbb{1}_A(x) = 1\ ,\ \ x \in A\ , \qquad \mathbb{1}_A(x) = 0\ ,\ \ x \in \mathcal{V} - A.$$

For x in $\mathcal{V}$, the Dirac mass at x is the probability measure δ_x such that $\delta_x(A) = \mathbb{1}_A(x)$, that is,

$$\delta_x(A) = 1\ ,\ \ x \in A\ , \qquad \delta_x(A) = 0\ ,\ \ x \in \mathcal{V} - A.$$

For x and y in $\mathcal{V}$, it holds that

$$\delta_x(y) = \mathbb{1}_{\{x\}}(y) = 1\ ,\ \ x = y\ , \qquad \delta_x(y) = \mathbb{1}_{\{x\}}(y) = 0\ ,\ \ x \neq y\ ,$$

but δ_x will be represented by a line vector and $\mathbb{1}_{\{x\}}$ by a column vector.

If μ is a nonnegative or signed measure, then

$$\mu(A) = \mu \mathbb{1}_A.$$

Duality and total variation norm A natural duality bracket between the Banach spaces $\mathcal{M}$ and L^∞ is given by

$$(\mu, f) \in \mathcal{M} \times L^\infty \mapsto \mu f \in \mathbb{R} ,$$

and for μ in $\mathcal{M}$, it holds that

$$\mu f \leq \|\mu\|_{var} \|f\|_\infty , \quad \|\mu\|_{var} = \mu^+(\mathcal{V}) + \mu^-(\mathcal{V}) = \mu(\mathbb{1}_{\{x:\mu(x)>0\}} - \mathbb{1}_{\{x:\mu(x)<0\}}) ,$$

and hence that

$$\|\mu\|_{var} = \max_{f \in L^\infty,\ \|f\|_\infty \leq 1} \mu f. \tag{1.2.2}$$

Thus, $\mathcal{M}$ can be identified with a closed subspace of the dual of L^∞ with the strong dual norm, which is the norm as an operator on L^∞.

The space L^∞ can be identified with the space $\ell^\infty = \ell^\infty(\mathcal{V}) = \ell^\infty(\mathcal{V}, \mathbb{R})$ of bounded sequences, and this duality between $\mathcal{M}$ and L^∞ to the natural duality between ℓ^1 and ℓ^∞.

The operations between complex measures and complex functions are performed by separating the real and imaginary parts.

1.2.2.3 Transition matrices, actions on measures, and functions

A matrix P is a transition matrix if and only if each of its line vectors $P(x, \cdot)$ corresponds to a probability measure. Then, its column vector $P(\cdot, y)$ defines a nonnegative function, which is bounded by 1.

A transition matrix can be multiplied on its right by nonnegative functions and on its left by nonnegative measures, or on its right by bounded functions and on its left by signed measures. The order of these operations does not matter. The function Pf, the nonnegative or signed measure μP, and $\mu P f \in \mathbb{R} \cup \{\infty\}$ are given for $x \in \mathcal{V}$ by

$$\begin{aligned}
Pf(x) &= P(x, \cdot)f = \sum_{y \in \mathcal{V}} P(x, y) f(y) = \mathbb{E}_x(f(X_1)) , \\
\mu P(x) &= \mu P(\cdot, x) = \sum_{z \in \mathcal{V}} \mu(z) P(z, x) = \mathbb{P}_\mu(X_1 = x) , \\
\mu P f &= \sum_{x,y \in \mathcal{V}} \mu(x) P(x, y) f(y) = \mathbb{E}_\mu(f(X_1)) ,
\end{aligned}$$

in which the notations $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ are used *only* when μ is a *law*.

In matrix notation,

$$Pf = \begin{pmatrix} \ddots & \vdots & \ddots \\ \cdots & P(x,y) & \cdots \\ \ddots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \vdots \\ f(y) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ Pf(x) = \sum_{y\in\mathcal{V}} P(x,y)f(y) \\ \vdots \end{pmatrix},$$

$$\mu P = \begin{pmatrix} \cdots & \mu(z) & \cdots \end{pmatrix} \begin{pmatrix} \ddots & \vdots & \ddots \\ \cdots & P(z,x) & \cdots \\ \ddots & \vdots & \ddots \end{pmatrix}$$

$$= \begin{pmatrix} \cdots & \mu P(x) = \sum_{z\in\mathcal{V}} \mu(z)P(z,x) & \cdots \end{pmatrix},$$

$$\mu Pf = \begin{pmatrix} \cdots & \mu(x) & \cdots \end{pmatrix} \begin{pmatrix} \ddots & \vdots & \ddots \\ \cdots & P(x,y) & \cdots \\ \ddots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \vdots \\ f(y) \\ \vdots \end{pmatrix}$$

$$= \sum_{x,y\in\mathcal{V}} \mu(x)P(x,y)f(y).$$

Intrinsic notation The linear mapping

$$P : f \in L^\infty \mapsto Pf \in L^\infty$$

has matrix P in the canonical basis. Its dual, or adjoint, mapping on $\mathcal{M}$, w.r.t. the duality bracket $(\mu, f) \mapsto \mu f$, is given by

$$P^* : \mu \in \mathcal{M} \mapsto \mu P \in \mathcal{M},$$

and has the adjoint (or transpose) matrix, also denoted by P^*. In order to respect the vector space structure and identify linear mappings and their matrices in the canonical bases, we could write $P^*\mu$ instead of μP and $\langle \mu, f\rangle$ instead of μf and $\langle \mu, Pf\rangle = \langle P^*\mu, f\rangle$ instead of μPf.

1.2.2.4 Transition matrix products, many-step transition

If P and Q are both transition matrices on $\mathcal{V}$, then it is easy to check that the matrix product PQ with generic term

$$PQ(x,y) = \sum_{z\in\mathcal{V}} P(x,z)Q(z,y)$$

is a transition matrix on $\mathcal{V}$. Let

$$P^0 := I \text{ (identity matrix)}, \qquad P^n := P^{n-1}P = PP^{n-1}, \quad n \geq 1.$$

Then,

$$P^n(x,y) = \sum_{x_1,\ldots,x_{n-1}\in\mathcal{V}} P(x,x_1)P(x_1,x_2)\cdots P(x_{n-1},y)$$

is the probability for $(X_{nk})_{k\geq 0}$ to go in one step from x to y, and hence for $(X_k)_{k\geq 0}$ to do so in n steps. In particular,

$$\pi_0 = \delta_x \Rightarrow \pi_n(y) = P^n(x, y).$$

Chapman–Kolmogorov formula As $P^n = P^k P^{n-k}$ for $0 \leq k \leq n$, this yields the Chapman–Kolmogorov formula

$$P^n(x, y) = \sum_{z\in\mathcal{V}} P^k(x, z)P^{n-k}(z, y).$$

Probabilistic interpretation These algebraic formulae have simple probabilistic interpretations: the probability of going from x to y in n steps can be obtained as the sum of the probabilities of taking every n-step path allowing to do so, as well as the sum over all intermediate positions after k steps.

1.2.3 Random recursion and simulation

Many Markov chains are obtained in a natural way as a random recursion, or random iterative sequence, as follows.

Theorem 1.2.3 *Let $(F_k)_{k\geq 1}$ be a sequence of independent identically distributed (i.i.d.) random functions from $\mathcal{V}$ to $\mathcal{V}$. For instance, for some not necessarily discrete space $\mathcal{W}$ and $f : \mathcal{W} \times \mathcal{V} \to \mathcal{V}$ and i.i.d. r.v. $(\xi_k)_{k\geq 1}$ with values in $\mathcal{W}$,*

$$F_k : x \in \mathcal{V} \mapsto f(\xi_k, x) \in \mathcal{V}.$$

If X_0 is any $\mathcal{V}$-valued r.v. independent of $(F_k)_{k\geq 1}$, then $(X_n)_{n\geq 0}$ given for $n \geq 1$ by

$$X_n = F_n(X_{n-1}) = \cdots = F_n \circ \cdots \circ F_1(X_0) \ ,$$

and more precisely by

$$X_n(\omega) = F_n(\omega)(X_{n-1}(\omega)) = \cdots = F_n(\omega) \circ \cdots \circ F_1(\omega)(X_0(\omega)) \ , \quad \omega \in \Omega \ ,$$

is a Markov chain on $\mathcal{V}$ with matrix P given by

$$P(x, y) = \mathbb{P}(F_1(x) = y) \ , \qquad x, y \in \mathcal{V}.$$

Proof: For $x_0, x_1, \ldots, x_n$ in $\mathcal{V}$, it holds that

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, X_1 = x_1 \ldots, X_n = x_n) \\
&\quad = \mathbb{P}(X_0 = x_0, F_1(x_0) = x_1, \ldots, F_n(x_{n-1}) = x_n) \\
&\quad = \mathbb{P}(X_0 = x_0)\mathbb{P}(F_1(x_0) = x_1) \ldots \mathbb{P}(F_1(x_{n-1}) = x_n).
\end{aligned}$$

Thus, Definition 1.2.1 is satisfied for the above matrix P. ■

When a random sequence $(X_n)_{n\in\mathbb{N}}$ is defined in some particular way, there is often a natural interpretation in terms of a random recursion of the previous kind. This allows to prove that a $(X_n)_{n\in\mathbb{N}}$ is a Markov chain, without having to directly check the definition, or even having to explicit its matrix. Moreover, such a pathwise representation for the Markov chain may be used for its study or its simulation.

Any Markov chain, with arbitrary initial law π_0 and transition matrix P, can be thus represented, using an i.i.d. sequence $(\xi_k)_{k\geq 1}$ of uniform r.v. on $\mathcal{W} := [0,1]$. The state space is first enumerated as $\mathcal{V} = \{x_j : j = 0, 1, \dots\}$.

Let $\omega \in \Omega$. For the initial value, if j is determined by

$$\pi_0(x_0) + \cdots + \pi_0(x_{j-1}) < \xi_0(\omega) \leq \pi_0(x_0) + \cdots + \pi_0(x_{j-1}) + \pi_0(x_j),$$

then $X_0(\omega) = x_j$. For the transitions, for $n \geq 1$, if j is determined by

$$\begin{aligned} &P(X_{n-1}(\omega), x_0) + \cdots + P(X_{n-1}(\omega), x_{j-1}) \\ &\quad < \xi_n(\omega) \leq P(X_{n-1}(\omega), x_0) + \cdots + P(X_{n-1}(\omega), x_{j-1}) + P(X_{n-1}(\omega), x_j), \end{aligned}$$

then $X_n(\omega) = x_j$.

In theory, this allows for the simulation of the Markov chain, but in practice this is not necessarily the best way to do so.

Note that this representation yields a construction for an arbitrary Markov chain, starting from a rigorous construction of i.i.d. sequences of uniform r.v. on $[0,1]$, without having to use the more general Kolmogorov extension theorem (Theorem A.3.10).

Remark 1.2.4 *Very different random recursions can be associated with the same transition matrix, and result in diverse pathwise behavior, notably w.r.t. changes in the initial value.*

1.2.4 Recursion for the instantaneous laws, invariant laws

The instantaneous laws $(\pi_n)_{n\geq 0}$ satisfy the recursion

$$\pi_n = \pi_{n-1} P$$

with solution $\pi_n = \pi_0 P^n$. This is a linear recursion in dimension $\mathrm{Card}(\mathcal{V})$, and the affine constraint $\sum_{x\in\mathcal{V}} \pi_n(x) = 1$ allows to reduce it to an affine recursion in dimension $\mathrm{Card}(\mathcal{V}) - 1$. Note that $P^n(x,y) = \pi_n(y)$ for $\pi_0 = \delta_x$.

An elementary study of this recursion starts by searching for its fixed points. These are the only possible large n limits for the instantaneous laws, for any topology such that $\mu \mapsto \mu P$ is continuous.

By definition, a fixed point for this recursion is a law (probability measure) π such that $\pi = \pi P$, and is called an *invariant law* or a *stationary distribution* for (or of) the matrix P, or the Markov chain $(X_n)_{n\in\mathbb{N}}$.

Stationary chain, equilibrium If π is an invariant law and $\pi_0 = \pi$, then $\pi_n = \pi$ for all n in $\mathbb{N}$, and $(X_{n+k})_{k\in\mathbb{N}}$ is again a Markov chain with initial law π and transition matrix P. Then, $(X_n)_{n\in\mathbb{N}}$ is said to be stationary, or in equilibrium.

Invariant measures and laws In order to find an invariant law, one must:

1. Solve the linear equation $\mu = \mu P$.
2. Find which solutions $\mu \neq 0$ are nonnegative, that is, such that $\mu \geq 0$.
3. Normalize such μ by setting $\pi = \mu/\|\mu\|_{var}$, which is possible only if

$$\|\mu\|_{var} := \sum_{x\in\mathcal{V}} \mu(x) < \infty \qquad \text{(always true for finite } \mathcal{V}) \,.$$

A nonnegative measure $\mu \neq 0$ such that $\mu = \mu P$ is called an *invariant measure*. An invariant measure is said to be unique if it is unique up to a multiplicative factor, that is, if all invariant measures are proportional (to it).

Algebraic interpretation The invariant measures are the *left eigenvectors* for the eigenvalue 1 for the matrix P acting on nonnegative measures, that is, for the adjoint (or transposed) matrix P^* acting on nonnegative vectors. Note that the constant function 1 is a *right eigenvector* for the eigenvalue 1, as

$$\pi 1 = \sum_{x\in\mathcal{V}} \pi(x) = 1.$$

The possible convergence of $(\pi_n)_{n\geq 0}$ to an invariant law π is related to the moduli of the elements of the spectrum of the restriction of the action of P on the signed measures of null total mass.

More generally, the exact or approximate computation of P^n depends in a more or less explicit way on a spectral decomposition of P.

1.3 Natural duality: algebraic approach

An algebraic study based on the natural duality between the space of signed measures $\mathcal{M} \simeq \ell^1$ and the space of bounded functions $L^\infty \simeq \ell^\infty$ will provide some structural results. These results are quite complete for finite state spaces $\mathcal{V}$. The complete study for arbitrary discrete $\mathcal{V}$ will be done later using probabilistic techniques. A reader for which this is the main interest may go directly to Section 1.4.

1.3.1 Complex eigenvalues and spectrum

1.3.1.1 Some reminders

The eigenvalues of the operator

$$P : f \mapsto Pf \;\text{ on }\; L^\infty$$

are given by all $\lambda \in \mathbb{C}$ such that $\lambda I - P$ is not injective as an operator on $L^\infty(\mathcal{V}, \mathbb{C})$.

The eigenspace of λ is the kernel

$$\ker\ (\lambda I - P) \in L^\infty(\mathcal{V}, \mathbb{C})\ ,$$

and its nonzero elements are called eigenvectors. Hence, f in $L^\infty(\mathcal{V}, \mathbb{C}) - \{0\}$ is an eigenvector of λ if and only if

$$Pf = \lambda f\ ,$$

and then

$$P^n f = \lambda^n f\ , \qquad n \geq 1.$$

The generalized eigenspace of λ is given by

$$\bigcup_{k\geq 1} \ker\ ((\lambda I - P)^k) \in L^\infty(\mathcal{V}, \mathbb{C}).$$

If it contains strictly the eigenspace, then it contains some f and some eigenvector g such that

$$Pf = \lambda f + g\ ,$$

and then

$$P^n f = \lambda^n f + n\lambda^{n-1} g\ , \qquad n \geq 1.$$

An eigenvalue is said to be

- *semisimple* if the eigenspace and generalized eigenspace coincide,
- *simple* if these spaces have dimension 1.

Similar definitions involving $\mathcal{M}(\mathcal{V}, \mathbb{C})$ are given for the adjoint operator

$$P^*\ :\ \mu \in \mathcal{M} \mapsto \mu P \in \mathcal{M}.$$

Its eigenspaces are often said to be eigenspaces *on the left*, or *left eigenspaces*, of P. Those of

$$P\ :\ f \in L^\infty \mapsto Pf \in L^\infty$$

may accordingly be called eigenspaces on the right, or right eigenspaces, of P.

Hence, $\mu \in \mathcal{M}(\mathcal{V}, \mathbb{C})$ is a left eigenvector of λ if and only if

$$\mu P = \lambda \mu\ ,$$

and then

$$\mu P^n = \lambda^n \mu\ , \qquad n \geq 1.$$

If the generalized left eigenspace of λ contains strictly the eigenspace, then it contains some μ and some left eigenvector ν such that

$$\mu P = \lambda \mu + \nu\ ,$$

and then

$$\mu P^n = \lambda^n \mu + n\lambda^{n-1}\nu\ , \qquad n \geq 1.$$

The spectrum $\sigma(P)$ of P on L^∞ is given by

$$\sigma(P) = \{\lambda \in \mathbb{C} : \lambda I - P \text{ not invertible in } L^\infty(\mathcal{V}, \mathbb{C})\}.$$

It is a simple matter to check that, using $\mathcal{M}(\mathcal{V}, \mathbb{C})$ in the definition,

$$\sigma(P^*) = \sigma(P)$$

and mentions of "left" or "right" are useless.

The spectrum $\sigma(P)$ contains both the left and right eigenvectors. In finite dimensions, invertibility of an operator is the same as injectivity, and hence the spectrum, the left eigenspace, and the right eigenspace coincide, but it is not so in general.

If $\lambda \in \mathbb{C}$ is in the spectrum of an operator on a real vector space, such as P or P^*, then $\bar{\lambda}$ is also in the spectrum, and if moreover λ is an eigenvalue, then $\bar{\lambda}$ is an eigenvalue, and the corresponding (generalized) eigenspaces are conjugate.

If $\lambda \in \mathbb{R}$, then $L^\infty = L^\infty(\mathcal{V}, \mathbb{R})$ can be considered instead of $L^\infty(\mathcal{V}, \mathbb{C})$ in all definitions. Moreover, the real and complex (generalized) eigenspaces have same dimension.

1.3.1.2 Algebraic results for transition matrices

Theorem 1.3.1 *Let P be a transition matrix on $\mathcal{V}$.*

1. *The operator $P : f \mapsto Pf$ on $L^\infty(\mathcal{V}, \mathbb{C})$ and its dual operator $P^* : \mu \mapsto \mu P$ on $\mathcal{M}(\mathcal{V}, \mathbb{C})$ are bounded and have operator norm* 1.
2. *The spectrum $\sigma(P)$ is included in the complex unit disk, every left or right eigenvalue of modulus* 1 *is semisimple, and the constant function* 1 *is a right eigenvector of eigenvalue* 1 *for P.*
3. *If $\mu \in \mathcal{M}(\mathcal{V}, \mathbb{C})$ and $\mu P = \lambda\mu$ for $|\lambda| = 1$, then the total variation measure $|\mu| = (|\mu|(x))_{x\in\mathcal{V}}$ satisfies $|\mu|P = |\mu|$ and hence is an invariant measure.*

Proof: If $f \in L^\infty(\mathcal{V}, \mathbb{C})$, then clearly

$$\|Pf\|_\infty = \sup_{x\in\mathcal{V}} \left| \sum_{y\in\mathcal{V}} P(x,y) f(y) \right| \leq \|f\|_\infty.$$

If $f \equiv 1$, then $Pf \equiv 1$. If $|\lambda| > 1$, then the series

$$(\lambda I - P)^{-1} = \lambda^{-1}(I - \lambda^{-1}P)^{-1} = \lambda^{-1}(I + \lambda^{-1}P + (\lambda^{-1}P)^2 + \cdots)$$

converges in operator norm on L^∞, which is given by $\|Q\|_{\text{op}} = \sup_{\|f\|_\infty \leq 1} \|Qf\|_\infty$.

If an eigenvalue λ is not semisimple, then there exists f in the generalized eigenspace and g in the eigenspace such that

$$P^n f = \lambda^n f + n\lambda^{n-1} g \ , \qquad n \geq 1 \ ,$$

and $\|P^n f\|_\infty \leq \|f\|_\infty$ then implies that $|\lambda| < 1$.

The corresponding results for P^* can be obtained in a similar way, or by duality.

In particular, P^* has operator norm 1, and hence if $\mu \in \mathcal{M}(\mathcal{V}, \mathbb{C})$, then

$$\sum_{x\in\mathcal{V}} |\mu|P(x) := \| \ |\mu|P \ \|_{var} \leq \| \ |\mu| \ \|_{var} := \sum_{x\in\mathcal{V}} |\mu(x)| \ ,$$

and if moreover $\mu P = \lambda\mu$ for $|\lambda| = 1$, then

$$|\mu|P(x) = \sum_{y\in\mathcal{V}} |\mu(y)|P(y,x) \geq \left| \sum_{y\in\mathcal{V}} \mu(y)P(y,x) \right| = |\mu P(x)| = |\lambda\mu(x)| = |\mu(x)|$$

and necessarily $|\mu|P(x) = |\mu(x)| = |\mu|(x)$ for every x. ■

1.3.1.3 Uniqueness for invariant laws and irreducibility

A state x in $\mathcal{V}$ is *absorbing* if $P(x,x) = 1$ and then δ_x is an invariant law for P.

If $\mathcal{V}$ contains subsets $\mathcal{V}_i$ for $i \in I$ such that $P(x, \mathcal{V}_i) = 1$ for every $x \in \mathcal{V}_i$, these are said to be *absorbing* or *closed*. The restriction of P to each $\mathcal{V}_i$ is Markovian; if it has an invariant measure μ_i, then any convex combination $\sum_{i\in I} c_i\mu_i$ is an invariant measure on $\mathcal{V}$, and if the μ_i are laws then $\sum_i c_i\mu_i$ is an invariant law. (By abuse of notation, μ_i denotes the extension of the measure to $\mathcal{V}$ vanishing outside of $\mathcal{V}_i$.)

Hence, any uniqueness result for invariant measures or laws requires adequate assumptions excluding the above situation.

The standard hypothesis for this is that of *irreducibility*: a transition matrix P on $\mathcal{V}$ is *irreducible* if, for every x and y in $\mathcal{V}$, there exists $i := i(x,y) \geq 1$ such that $P^i(x,y) > 0$. Equivalently, there exists in the oriented graph of the matrix a path covering the whole graph (respecting orientation). This notion will be further developed in due time.

Lemma 1.3.2 *Let $\mathbb{P}$ be an irreducible transition matrix. If a measure μ satisfies $\mu P = \mu$, then either $\mu = 0$ or $\mu > 0$.*

Proof: Assume that there exists a state x such that $\mu(x) > 0$. For any state y, there exists $i \geq 1$ such that $P^i(x,y) > 0$. By iteration $\mu = \mu P = \cdots = \mu P^i$, and hence,

$$\mu(y) = \sum_{z\in\mathcal{V}} \mu(z)P^i(z,y) \geq \mu(x)P^i(x,y) > 0.$$

Hence, either $\mu > 0$ or $\mu = 0$. ■

Theorem 1.3.3 *Let $\mathbb{P}$ be an irreducible transition matrix. If $P^* : \mu \in \mathcal{M} \mapsto \mu P$ has 1 as an eigenvalue, then it is a simple eigenvalue, and its eigenspace is generated by an invariant law π, which is positive and unique.*

Proof: Let $\mu \neq 0$ be in $\mathcal{M}(\mathcal{V}, \mathbb{R})$ and satisfy $\mu P = \mu$. (This is enough, as $\lambda = 1$ is in $\mathbb{R}$.) Theorem 1.3.1 implies that $|\mu|$ is an invariant measure. As $\mu \neq 0$ and hence $|\mu| \neq 0$, Lemma 1.3.2 yields that $|\mu| > 0$, and an everywhere positive invariant law is given by

$$\pi = \frac{|\mu|}{|\mu|(\mathcal{V})}.$$

Moreover,

$$\mu^+ = \frac{1}{2}(|\mu| + \mu) , \qquad \mu^- = \frac{1}{2}(|\mu| - \mu) ,$$

are invariant measures or are zero and cannot be both zero. Lemma 1.3.2 yields that $\mu^+ > 0$ or $\mu^- > 0$, that is, that $\mu > 0$ or $\mu < 0$.

Hence, if π and π' are two invariant laws, then $\pi - \pi' = (\pi - \pi')P$ and hence either $\pi - \pi' = 0$ or $\pi - \pi' > 0$ or $\pi - \pi' < 0$. As $(\pi - \pi')(\mathcal{V}) = 0$, we conclude that $\pi - \pi' = 0$, hence the invariant law is unique.

The eigenvalue 1 is semisimple (see Theorem 1.3.1), hence it is simple. ■

This proof heavily uses techniques that are referred under the terminology "the maximum principle," which we will try to explain in Section 1.3.3.

1.3.2 Doeblin condition and strong irreducibility

A transition matrix P is *strongly irreducible* if there exists $i \geq 1$ such that $P^i > 0$.

Theorem 1.3.4 (Doeblin) *Let P be a transition matrix $\mathcal{V}$ satisfying the* Doeblin condition*: there exists $k \geq 1$ and $\varepsilon > 0$ and a law $\hat{\pi}$ on $\mathcal{V}$ such that*

$$P^k(x, y) \geq \varepsilon \hat{\pi}(y) , \qquad \forall x, y \in \mathcal{V}.$$

Then, there exists a unique invariant law π, which satisfies $\pi \geq \varepsilon \hat{\pi}$. Moreover, for any $\mu \in \mathcal{M}$ such that $\mu(\mathcal{V}) = 0$, it holds that

$$\|\mu P^n\|_{var} \leq (1 - \varepsilon)^{\lfloor n/k \rfloor} \|\mu\|_{var} \leq 2(1 - \varepsilon)^{\lfloor n/k \rfloor} \|\mu\|_{var} , \qquad n \geq 1 ,$$

which yields the exponential bounds, uniform on the initial law,

$$\sup_{x \in \mathcal{V}} \sum_{y \in \mathcal{V}} |P^n(x, y) - \pi(y)| \leq \sup_{\pi_0 \in \mathcal{M}_+^1} \|\pi_0 P^n - \pi\|_{var} \leq 2(1 - \varepsilon)^{\lfloor n/k \rfloor} , \quad n \geq 1.$$

The restriction of P to $\{\pi > 0\} := \{x \in \mathcal{V} : \pi(x) > 0\}$ is an irreducible transition matrix, which is strongly irreducible if $\{\pi > 0\}$ is finite.

Proof: Let us first assume the Doeblin condition to hold for $k = 1$. Let $\mu \in \mathcal{M}$ be such that $\mu(\mathcal{V}) := \sum_{x\in\mathcal{V}} \mu(x) = 0$. Then,

$$\begin{aligned}
\|\mu P\|_{var} &= \sum_{y\in\mathcal{V}} \left| \sum_{x\in\mathcal{V}} \mu(x)P(x,y) \right| && \text{(by definition)} \\
&= \sum_{y\in\mathcal{V}} \left| \sum_{x\in\mathcal{V}} \mu(x)(P(x,y) - \varepsilon\hat{\pi}(y)) \right| && \text{(as } \mu(\mathcal{V}) = 0\text{)} \\
&\leq \sum_{y\in\mathcal{V}} \sum_{x\in\mathcal{V}} |\mu(x)|(P(x,y) - \varepsilon\hat{\pi}(y)) && \text{(as } P(x,y) \geq \varepsilon\hat{\pi}(y)\text{)} \\
&\leq \|\mu\|_{var}(1-\varepsilon) && \text{(changing summation order).}
\end{aligned}$$

Moreover,

$$\mu P(\mathcal{V}) = \sum_{y\in\mathcal{V}} \sum_{x\in\mathcal{V}} \mu(x)P(x,y) = \sum_{x\in\mathcal{V}} \mu(x) \sum_{y\in\mathcal{V}} P(x,y) = 0$$

and iteration yields that

$$\|\mu P^n\|_{var} \leq \|\mu\|_{var}(1-\varepsilon)^n \,, \qquad n \geq 1.$$

If the Doeblin condition holds for an arbitrary $k \geq 1$, Theorem 1.3.1 and the result for $k = 1$ applied to P^k yield that

$$\|\mu P^n\|_{var} = \|\mu(P^k)^{\lfloor n/k \rfloor} P^{n-k\lfloor n/k \rfloor}\|_{var} \leq \|\mu(P^k)^{\lfloor n/k \rfloor}\|_{var} \leq \|\mu\|_{var}(1-\varepsilon)^{\lfloor n/k \rfloor}.$$

For any laws π_0 and π_0', it holds that $(\pi_0 - \pi_0')(\mathcal{V}) = 0$, and thus

$$\|\pi_0 P^n - \pi_0' P^n\|_{var} \leq (1-\varepsilon)^{\lfloor n/k \rfloor} \|\pi_0 - \pi_0'\|_{var} \leq 2(1-\varepsilon)^{\lfloor n/k \rfloor}$$

and this bound for arbitrary π_0 and $\pi_0' = \pi_0 P$ implies that $(\pi_0 P^n)_{n\geq 0}$ is a Cauchy sequence in the complete metric space $\mathcal{M}_+^1$ (by an exponential series bound).

Hence $(\pi_0 P^n)_{n\geq 0}$ converges to some law π, which by continuity must satisfy $\pi = \pi P$, and hence is an invariant law; this convergence also implies that the invariant law is unique.

Taking $\pi_0' = \pi$ and arbitrary π_0 or $\pi_0 = \delta_x$ for arbitrary x in $\mathcal{V}$ yield the bounds, which are uniform on the initial law. Moreover, for every y,

$$P^n(x,y) = \sum_{z\in\mathcal{V}} P^{n-1}(x,z)P(z,y) \geq \varepsilon\hat{\pi}(y) \sum_{z\in\mathcal{V}} P^{n-1}(x,z) = \varepsilon\hat{\pi}(y)$$

and taking the limit yields that $\pi(y) \geq \hat{\pi}(y)$.

If $\pi(x) > 0$ and $P(x,y) > 0$, then $\pi(y) \geq \pi(x)P(x,y) > 0$, and hence the restriction of P to $\{\pi > 0\}$ is Markovian.

If $\pi(y) > 0$, then, as $\lim_{n\to\infty} P^n(x,y) = \pi(y)$ for every x, there exists some $i(x,y) \geq 1$ such that $P^i(x,y) > 0$ for $i \geq i(x,y)$, hence this restriction is irreducible;

if moreover $\{\pi > 0\}$ is finite, then for $i \geq \max_{x,y\in\{\pi>0\}} i(x,y)$ it holds that $P^i > 0$ on $\{\pi > 0\}$, and hence the restriction of P is strongly irreducible. ■

Note that $\varepsilon \leq 1$ and that $\varepsilon = 1$ only in the trivial case in which $(X_n)_{n\geq 1}$ is a sequence of i.i.d. r.v. of law $\hat{\pi}$.

The Doeblin Condition (or strong irreducibility) is seldom satisfied when the state space $\mathcal{V}$ is infinite. For a finite state space, Section 4.2.1 will give verifiable conditions for strong irreducibility. The following result is interesting in these perspectives.

Corollary 1.3.5 *Let P be a strongly irreducible matrix on a finite state space $\mathcal{V}$. Then, P satisfies the Doeblin Condition (see Theorem 1.3.4) for*

$$k \geq 1 \text{ such that } P^k > 0\ ,\quad \varepsilon = \sum_{y\in\mathcal{V}} \min_{x\in\mathcal{V}} P^k(x,y) > 0\ ,\quad \hat{\pi}(y) = \frac{1}{\varepsilon} \min_{x\in\mathcal{V}} P^k(x,y)\ ,$$

and the conclusions of Theorem 1.3.4 hold with $\pi > 0$ on $\mathcal{V}$.

Proof: The proof is immediate. ■

1.3.3 Finite state space Markov chains

1.3.3.1 Perron–Frobenius theorem

If the state space $\mathcal{V}$ is finite, then the vector spaces $\mathcal{M}(\mathcal{V})$ and $L^\infty(\mathcal{V})$ have finite dimension $\mathrm{Card}(\mathcal{V})$.

Then, the eigenvalues and the dimensions of the eigenspaces and generalized eigenspaces of P, which are by definition those of P and P^*, are identical, and the spectrum is constituted of the eigenvalues.

A function f is *harmonic* if $Pf = f$. It is a right eigenvector for the eigenvalue 1.

Theorem 1.3.6 Perron–Frobenius *Let $\mathcal{V}$ be a finite state space and P a transition matrix on $\mathcal{V}$.*

1. *The spectrum $\sigma(P)$ of P is included in the complex unit disk, the eigenvalues with modulus 1 are semisimple, the constant functions are harmonic, and there exists an invariant law π.*
2. *If P is irreducible, then the invariant law π is unique and everywhere positive, the only harmonic functions are constant, and there exists an integer $d \geq 1$, called the period of P, such that the only eigenvalues with modulus 1 the dth complex roots of 1, and these eigenvalues are simple.*
3. *If P is strongly irreducible, then $d = 1$.*

Proof: The beginning of the proof is an application of Theorem 1.3.1. The fact that $P1 = 1$ implies that 1 is an eigenvalue and that the constant functions are harmonic. As the dimension is finite, it further implies that there exists a right eigenvector

$\mu \in \mathcal{M}$ for the eigenvalue 1, that is, satisfying $\mu P = \mu$, and Theorem 1.3.1 implies that the law $\pi = |\mu|/|\mu|(\mathcal{V})$ is invariant.

If P is irreducible, then Theorem 1.3.3 yields that the invariant law π is unique and satisfies $\pi > 0$, hence the eigenvalue 1 is simple, and as the dimension is finite, any harmonic function is a multiple of the constant function 1 and thus is a constant.

If P is strongly irreducible, then Corollary 1.3.5 holds. Let $\lambda \in \mathbb{C}$ satisfy $|\lambda| = 1$, and $\mu \in \mathcal{M}$ be such that $\mu P = \lambda\mu$. Then, for $n \geq 1$,

$$\mu P^n = \lambda^n \mu \ , \qquad \mu P^n = (\mu - |\mu|\pi)P^n + |\mu|\pi P^n = (\mu - |\mu|\pi)P^n + |\mu|\pi \ ,$$

and letting n go to infinity in the exponential bounds in Theorem 1.3.4 for $\mu - |\mu|\pi$, which satisfies $(\mu - |\mu|\pi)(\mathcal{V}) = 0$, shows that $\mu = |\mu|\pi$ and that $\lambda = 1$, *i.e.*, that the eigenvalue 1 is simple and that any other eigenvalue has modulus strictly <1.

If P is irreducible, then there exists $d \geq 1$ such that P^d is strongly irreducible on each class of a partition of $\mathcal{V}$, which allows to prove the result on the dth complex roots of 1 (see Section 4.2.1), in particular Definitions 4.2.1 and 4.2.6 and Theorems 4.2.4 and 4.2.7. ■

Theorem 4.2.7 will provide an extension of these results for infinite $\mathcal{V}$, by wholly different methods. (Saloff-Coste, L. (1997), Section 1.2) provides a detailed commentary on the Doeblin condition and the Perron–Frobenius theorem.

Maximum principle This terminology comes from the following short direct proof of the fact that if the state space $\mathcal{V}$ is finite and P is irreducible, then every harmonic function is constant. If f is harmonic on $\mathcal{V}$, then it attains its maximum in at least a state x. Moreover,

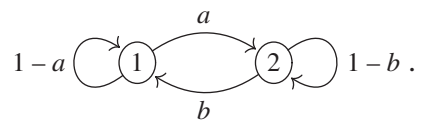
$$\max f = f(x) = P^i f(x) = \sum_{y \in \mathcal{V}} P^i(x, y) f(y) \ , \qquad i \geq 1 \ ,$$

and as $P^i(x, \cdot)$ is a probability measure, $f(y) = \max f$ for all y such that $P^i(x, y) > 0$. Thus, irreducibility yields that $f(y) = \max f$ for every $y \in \mathcal{V}$, and thus that f is constant.

1.3.3.2 Computation of the instantaneous and invariant laws

We are now going to solve the recursion for the instantaneous laws $(\pi_n)_{n \geq 0}$, and see how the situation deteriorates in practice very quickly as the size of the state space increases.

The chain with two states Let us denote the states by 1 and 2. There exists $0 \leq a, b \leq 1$ such that the transition matrix P and its graph are given by

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \qquad \text{  }$$

The recursion formula $\pi_n = \pi_{n-1}P$ then writes

$$\begin{aligned}(\pi_n(1), \pi_n(2)) &= (\pi_{n-1}(1), \pi_{n-1}(2)) \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \\ &= ((1-a)\pi_{n-1}(1) + b\pi_{n-1}(2),\ \ a\pi_{n-1}(1) + (1-b)\pi_{n-1}(2))\end{aligned}$$

and the affine constraint $\pi_n(1) + \pi_n(2) = 1$ allows to reduce this linear recursion in dimension 2 to the affine recursion, in dimension 1,

$$\pi_n(1) = (1-a-b)\pi_{n-1}(1) + b.$$

If $a = b = 0$, then $P = I$ and every law is invariant, and P is not irreducible as $P^n = I$. Else, the unique fixed point is $\frac{b}{a+b}$, the unique invariant law is $\pi = \left(\frac{b}{a+b}, \frac{a}{a+b}\right)$, and

$$\pi_n(1) = \left(\pi_0(1) - \frac{b}{a+b}\right)(1-a-b)^n + \frac{b}{a+b}$$

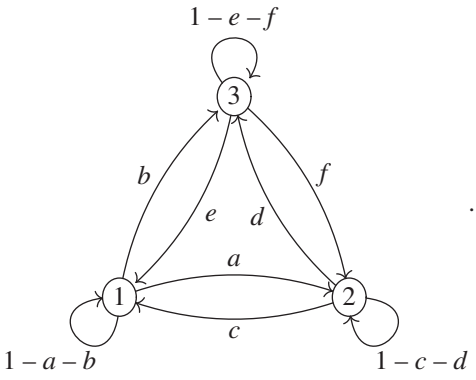
and the formula for $\pi_n(2)$ is obtained by symmetry or as $\pi_n(2) = 1 - \pi_n(1)$.

If $a = b = 1$, then $P = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ has eigenvalues 1 and -1, the latter with eigenvector $\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$, and the chain alternates between the states 1 and 2 and $\pi_n(1)$ is equal to $\pi_0(1)$ for even n and to $1 - \pi_0(1)$ for odd n.

If $(a, b) \notin \{(0, 0), (1, 1)\}$, then $\lim_{n\to\infty} \pi_n(1) = \frac{b}{a+b}$ and $\lim_{n\to\infty} \pi_n = \pi$ with geometric rate with reason $1 - a - b$.

The chain with three states Let us denote the states by 1, 2, and 3. There exists a, b, c, d, e, and f in $[0, 1]$, satisfying $a + b \leq 1$, $c + d \leq 1$, and $e + f \leq 1$, such that the transition matrix P and its graph are given by

$$\begin{pmatrix} 1-a-b & a & b \\ c & 1-c-d & d \\ e & f & 1-e-f \end{pmatrix},$$



.

As discussed above, we could reduce the linear recursion in dimension 3 to an affine recursion in dimension 2. Instead, we give the elements of a vectorial computation

in dimension 3, which can be generalized to all dimensions. This exploits the fact that 1 is an eigenvalue of P and hence a root of its characteristic polynomial $K(X) = \det(XI - P)$. Hence,

$$K(X) = \begin{vmatrix} X+a+b-1 & -a & -b \\ -c & X+c+d-1 & -d \\ -e & -f & X+e+f-1 \end{vmatrix}$$
$$= (X+a+b-1)(X+c+d-1)(X+e+f-1) - ade - bcf$$
$$- ac(X+e+f-1) - be(X+c+d-1) - df(X+a+b-1)$$

and by developing this polynomial and using the fact that 1 is a root, $K(X)$ factorizes into

$$(X-1)(X^2 + (a+b+c+d+e+f-2)X$$
$$+ ad + ae + af + bc + bd + bf + ce + cf + de - a - b - c - d - e - f + 1).$$

The polynomial of degree 2 is the characteristic polynomial of the affine recursion in dimension 2. It has two possible equal roots λ_1 and λ_2 in $\mathbb{C}$, and if $\lambda_1 \in \mathbb{C} - \mathbb{R}$, then $\lambda_2 = \bar{\lambda}_1$. Their exact theoretical expression is not very simple, as the discriminant of this polynomial does not simplify in general, but they can easily be computed on a case-by-case basis.

In order to compute P^n, we will use the Cayley–Hamilton theorem, according to which $K(P) = 0$ (nul matrix). The Euclidean division of X^n by $K(X)$ yields

$$X^n = Q(X)K(X) + a_n X^2 + b_n X + c_n \,, \qquad P^n = a_n P^2 + b_n P + c_n I \,,$$

and in order to effectively compute a_n, b_n, and c_n, we take the values of the polynomials for the roots of $K(X)$, which yields the linear system

$$\begin{cases} a_n + b_n + c_n & = 1 \\ \lambda_1^2 a_n + \lambda_1 b_n + c_n & = \lambda_1^n \\ \lambda_2^2 a_n + \lambda_2 b_n + c_n & = \lambda_2^n. \end{cases}$$

This system has rank 3 if the three roots are distinct.

If there is a double root λ (be it 1 or $\lambda_1 = \lambda_2$), then two of these equations are identical, but as the double root is also a root of $K'(X)$, a simple derivative yields a third equation $2\lambda a_n + b_n = n\lambda^{n-1}$, which is linearly independent of the two others.

If the three roots of $K(X)$ are equal, then they are equal to 1 and $P^n = P = I$.

If P is irreducible, then there exists a unique invariant law π, given by

$$\pi(1) = \frac{ce + cf + of}{ad + ae + af + bc + bd + bf + ce + cf + of},$$
$$\pi(2) = \frac{ae + af + bf}{ad + ae + af + bc + bd + bf + ce + cf + of},$$
$$\pi(3) = \frac{ad + bc + bd}{ad + ae + af + bc + bd + bf + ce + cf + of}.$$

The chain with a finite number of states Let $d = \mathrm{Card}(\mathcal{V})$. The above-mentioned method can be extended without any theoretical problem.

The Euclidean division $X^n = Q(X)K(X) + a_{n,d-1}X^{d-1} + \cdots + a_{n,1}X + a_{n,0}$ and $K(P) = 0$ yield that

$$P^n = a_{n,d-1}P^{d-1} + \cdots + a_{n,1}P + a_{n,0}I.$$

If $\lambda_1, \ldots, \lambda_r$ are the distinct roots of $K(X)$ and $m_1 \geq 1, \ldots, m_r \geq 1$ are their multiplicities, then

$$K(\lambda_i) = 0\ , \ldots, K^{(m_i-1)}(\lambda_i) = 0\ , \qquad 1 \leq i \leq r\ ,$$

is a system of $d = \sum_{i=1}^{r} m_i$ linearly independent equations for the d unknowns $a_{n,d-1}, \ldots, a_{n,0}$, which thus has a unique solution.

The enormous obstacle for the effective implementation of this method for computing P^n is that we must compute the roots of $K(X)$ first. The main information we have is that 1 is a root, and in general, computing the roots becomes a considerable problem as soon as $d \geq 4$. Once the roots are found, solving the linear system and finding the invariant laws is a problem only when d is much larger.

This general method is simpler than finding the reduced Jordan form J for P, which also necessitates to find the roots of the characteristic polynomial $K(X)$, and then solving a linear system to find the change-of-basis matrix M and its inverse M^{-1}. Then, $P^n = (MJM^{-1})^n = MJ^nM^{-1}$, where J^n can be made explicit.

1.4 Detailed examples

We are going to describe in informal manner some problems concerning random evolutions, for which the answers will obviously depend on some data or parameters. We then will model these problems using Markov chains.

These models will be studied in detail all along our study of Markov chains, which they will help to illustrate.

In these descriptions, random variables and draws will be supposed to be independent if not stated otherwise.

1.4.1 Random walk on a network

A particle evolves on a network $\mathcal{R}$, that is, on a discrete additive subgroup such as $\mathbb{Z}^d$. From x in $\mathcal{R}$, it chooses to go to $y = x + (y - x)$ in $\mathcal{R}$ with probability $p(y - x) \geq 0$, which satisfies $\sum_{z\in\mathcal{R}} p(z) = 1$. This can be, for instance, a model for the evolution of an electron in a network of crystals.

Some natural questions are the following:

- Does the particle escape to infinity?
- If yes, at what speed?

- With what probability does it reach a certain subset in finite time?
- What is the mean time for that?

1.4.1.1 Modeling

Let $(\xi_k)_{k\geq 1}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(\xi_1 = z) = p(z)$, and

$$X_n = X_{n-1} + \xi_n = \cdots = X_0 + \xi_1 + \cdots + \xi_n.$$

Theorem 1.2.3 shows that $(X_n)_{n\geq 0}$ is a Markov chain on $\mathcal{R}$, with a transition matrix, which is spatially homogeneous, or invariant by translation, given by

$$P(x, y) = \mathbb{P}(\xi_1 = y - x) = p(y - x)\ , \qquad x, y \in \mathcal{R}.$$

The matrix P restricted to the network generated by all z such that $p(z) > 0$ is irreducible. The constant measures are invariant, as

$$\sum_{x\in\mathcal{R}} P(x, y) = \sum_{x\in\mathcal{R}} p(y - x) = \sum_{z\in\mathcal{R}} p(z) = 1\ , \qquad \forall y \in \mathcal{R}.$$

If $\mathbb{E}(|\xi_1|) < \infty$, then the strong law of large numbers yields that

$$X_n = n\mathbb{E}(\xi_1) + o(n)\ , \quad \text{a.s.},$$

and for $\mathbb{E}(\xi_1) \neq 0$ the chain goes to infinity in the direction of $\mathbb{E}(\xi_1)$. The case $\mathbb{E}(\xi_1) = 0$ is problematic, and if $\mathbb{E}(|\xi_1|^2) < \infty$, then the central limit theorem shows that $X_n/\sqrt{n}$ converges in law to $\mathcal{N}(0, \mathrm{Cov}(\xi_1))$, which gives some hints to the long-time behavior of the chain.

Nearest-neighbor random walk For $\mathcal{R} = \mathbb{Z}^d$, this Markov chain is called a nearest-neighbor random walk when $P(x, y) = 0$ for $|x - y| > 1$, and the symmetric nearest-neighbor random walk when $P(x, y) = 1/2d$ for $|x - y| = 1$. These terminologies are used for other regular networks, such as the one in Figure 1.1.

1.4.2 Gambler's ruin

Two gamblers A and B play a game of head or tails. Gambler A starts with a fortune of $a \in \mathbb{N}$ units of money and Gambler B of $b \in \mathbb{N}$ units. At each toss, each gambler makes a bet of 1 unit, Gambler A wins with probability p and loses with probability $q = 1 - p$, and the total of the bets is given to the winner; a gambler thus either wins or loses 1 unit.

The game continues until one of the gamblers is *ruined*: he or she is left with a fortune of 0 units, the global winner with a fortune of $a + b = N$ units, and the game stops. This is illustrated in Figure 1.2.

When $p = q = 1/2$, the game is said to be *fair*, else to be *biased*.

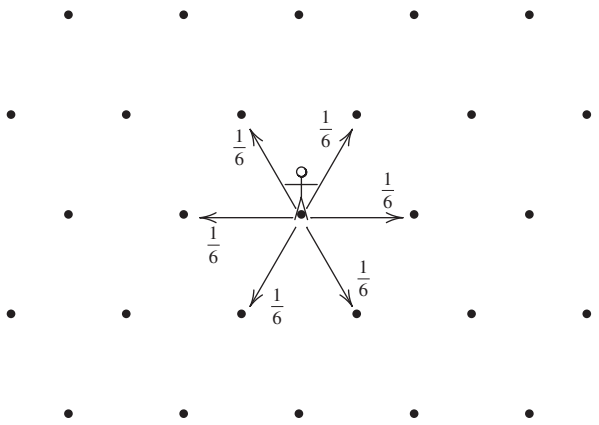


Figure 1.1 Symmetric nearest-neighbor random walk on regular planar triangular network.

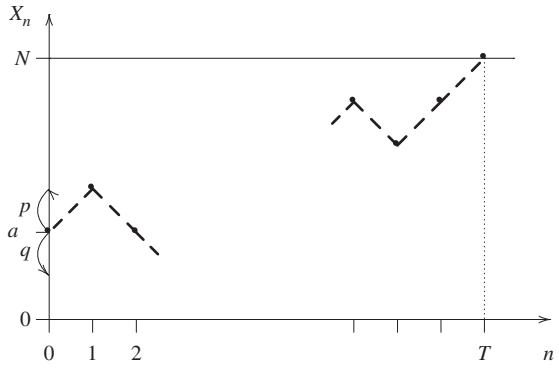


Figure 1.2 Gambler's ruin. Gambler A finishes the game at time T with a gain of $b = N - a$ units starting from a fortune of a units. The successive states of his fortune are represented by the • and joined by dashes. The arrows on the vertical axis give his probabilities of winning or losing at each toss.

As an example, Gambler A goes to a casino (Gambler B). He decides to gamble 1 unit at each draw of red or black at roulette and to stop either after having won a total of b units (what he or she would like to gain) or lost a total of a units (the maximal loss he or she allows himself).

Owing to the 0 and (most usually) the double 0 on the roulette, which are neither red nor black, the game is biased against him, and p is worth either $18/37 \simeq 0.4865$ if there is no double 0 or $18/38 \simeq 0.4737$ if there is one.

From a formal point of view, there is a symmetry in the game obtained by switching a and $b = N - a$ simultaneously with p and $q = 1 - p$. In practice, no casino allows a bias in favor of the gambler, nor even a fair game.

A unilateral case will also be considered, in which $a \in \mathbb{N}$ and $b = N = \infty$. In the casino example, this corresponds to a compulsive gambler, who will stop only when ruined. In all cases, the evolution of the gambler's fortune is given by a nearest-neighbor random walk on $\mathbb{Z}$, stopped when it hits a certain boundary.

Some natural questions are the following:

- What is the probability that Gambler A will be eventually ruined?
- Will the game eventually end ?
- If yes, what is the mean duration of the game?
- What is the law of the duration of the game (possibly infinite) ?

1.4.2.1 Stopped random walk

In all cases, the evolution of the gambler's fortune is given by a nearest-neighbor random walk on $\mathbb{Z}$ stopped when it hits a certain boundary.

1.4.2.2 Modeling

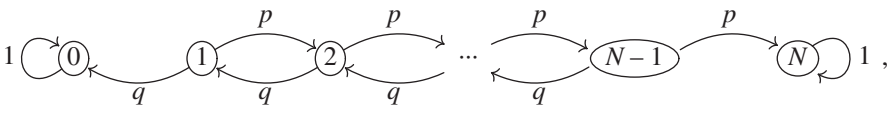
The evolution of the fortune of Gambler A can be represented using a sequence $(\xi_k)_{k\geq 1}$ of i.i.d. r.v. satisfying $\mathbb{P}(\xi_1 = 1) = p$ and $\mathbb{P}(\xi_1 = -1) = q = 1 - p$ by

$$X_n = X_{n-1} + \xi_n \mathbb{1}_{\{0<X_{n-1}<N\}} , \qquad n \geq 1 ,$$

where X_0 is its initial fortune a, or more generally a r.v. with values in $\{0, 1, \dots, N\}$ and independent of $(\xi_k)_{k\geq 1}$. Gambler B's fortune at time $n \geq 0$ is $N - X_n$.

Theorem 1.2.3 yields that $(X_n)_{n\geq 0}$ is a Markov chain on $\mathcal{V} = \{0, 1, \dots, N\}$ with matrix and graph given by

$$P(0, 0) = P(N, N) = 1; \quad P(x, x + 1) = p, \quad P(x, x - 1) = q, \quad 0 < x < N,$$



(1.4.3)

the other terms of P being 0.

The states 0 and $N = a + b$ are absorbing, hence P is not irreducible. The invariant measures μ are of the form $\mu(x) = 0$ if $0 < x < N$ with $\mu(0)$ and $\mu(N)$ arbitrary, and uniqueness does not hold (uniqueness being understood as "up to proportionality").

1.4.3 Branching process: evolution of a population

We study the successive generation of a population, constituted for instance of viruses in an organism, of infected people during an epidemic, or of neutrons during an atomic reaction.

The individuals of one generation disappear in the following, giving birth there each to k descendants with probability $p(k) \geq 0$, with $\sum_{k\in\mathbb{N}} p(k) = 1$. A classic subcase is that of a binary division: $p(2) = p > 0$ and $p(0) = 1 - p > 0$.

The result of this random evolution mechanism is called a branching process. It is also called a Galton–Watson process; the initial study of Galton and Watson, preceded by a similar study of Bienaymé, bore on family names in Great Britain.

Some natural questions are the following:

- What is the law of the number of individuals in the nth generation?
- Will the population become extinct, almost surely (a.s.), and else with what probability?
- What is the long-time population behavior when it does not become extinct?

1.4.3.1 Modeling

We shall construct a Markov chain $(X_n)_{n\geq 0}$ corresponding to the sizes (numbers of individuals) of the population along the generations.

Let $(\xi_{n,i})_{n,i\geq 1}$ be i.i.d. r.v. such that $\mathbb{P}(\xi_{1,1} = k) = p(k)$ for k in $\mathbb{N}$. We assume that the X_{n-1} individuals of generation $n-1$ are numbered $i \geq 1$ and that the ith one yields $\xi_{n,i}$ descendants in generation n, so that

$$X_n = \sum_{i=1}^{X_{n-1}} \xi_{n,i} \ , \qquad n \geq 1.$$

(An empty sum being null by convention.)

Figure 1.3 illustrates this using the genealogical tree of a population, which gives the relationships between individuals in addition to the sizes of the generations, and explains the term "branching."

The state space of $(X_n)_{n\geq 0}$ is $\mathbb{N}$, and Theorem 1.2.3 applied to $\xi_n = (\xi_{n,i})_{i\geq 1}$ yields that it is a Markov chain. The transition matrix is given by

$$P(0,0) = 1 \ , \qquad P(x,y) = \sum_{k_1+\cdots+k_x=y} p(k_1)\cdots p(k_x) \ , \quad x \geq 1 \ , \ y \geq 0 \ ,$$

and state 0 is absorbing. The matrix is not practical to use under this form.

It is much more practical to use generating functions. If

$$g(s) = \sum_{k\in\mathbb{N}} p(k)s^k = \mathbb{E}(s^{\xi_{1,1}})$$

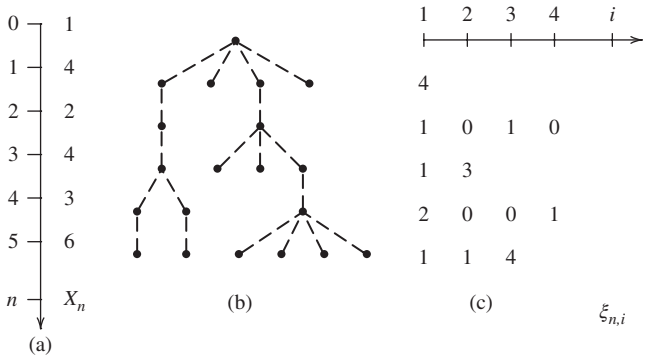


Figure 1.3 Branching process. (b) The genealogical tree for a population during six generations; • represent individuals and dashed lines their parental relations. (a) The vertical axis gives the numbers n of the generations, of which the sizes figure on its right. (c) The table underneath the horizontal axis gives the $\xi_{n,i}$ for $n \geq 1$ and $1 \leq i \leq X_{n-1}$, of which the sum over i yields X_n.

denotes the generating function of the reproduction law, the i.i.d. manner in which individuals reproduce yields that

$$\sum_{y \in \mathcal{Y}} P(x,y)s^y := \mathbb{E}_x(s^{X_1}) = \mathbb{E}_1(s^{X_1})^x = g(s)^x.$$

For n in $\mathbb{N}$, let

$$g_n(s) = \sum_{x \in \mathbb{N}} \mathbb{P}(X_n = x)s^x = \mathbb{E}(s^{X_n})$$

denote the generating function of the size of generation n. An elementary probabilistic computation yields that, for $n \geq 1$,

$$\begin{aligned} g_n(s) &= \mathbb{E}\left(s^{\sum_{i=1}^{X_{n-1}} \xi_{n,i}}\right) \\ &= \sum_{x \in \mathbb{N}} \mathbb{E}\left(s^{\sum_{i=1}^{x} \xi_{n,i}} \mathbb{1}_{\{X_{n-1}=x\}}\right) \\ &= \sum_{x \in \mathbb{N}} \mathbb{E}\left(s^{\sum_{i=1}^{x} \xi_{n,i}}\right) \mathbb{P}(X_{n-1} = x) \\ &= \sum_{x \in \mathbb{N}} g(s)^x \mathbb{P}(X_{n-1} = x) \end{aligned}$$

and hence that

$$g_n(s) = g_{n-1}(g(s)) = \cdots = g^{on}(g_0(s)) = g(g_{n-1}(s)).$$

We will later see how to obtain this result by Markov chain techniques and then how to exploit it.

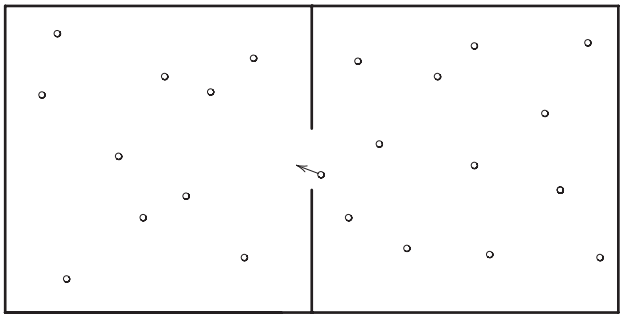


Figure 1.4 The Ehrenfest Urn. Shown at an instant when a particle transits from the right compartment to the left one. The choice of the particle that changes compartment at each step is uniform.

1.4.4 Ehrenfest's Urn

A container (urn, …) is constituted of two communicating compartments and contains a large number of particles (such as gas molecules). These are initially distributed in the two compartments according to some law, and move around and can switch compartment.

Tatiana and Paul Ehrenfest proposed a statistical mechanics model for this phenomenon. It is a discrete time model, in which at each step a particle is chosen uniformly among all particles and changes compartment. See Figure 1.4.

Some natural questions are the following:

- starting from an unbalanced distribution of particles between compartments, is the distribution of particles going to become more balanced in time?
- In what sense, with what uncertainty, at what rate?
- Is the distribution going to go through astonishing states, such as having all particles in a single compartment, and at what frequency?

1.4.4.1 Microscopic modeling

Let N be the number of molecules, and let the compartments be numbered by 0 and 1.

A microscopic description of the system at time $k \geq 0$ is given by

$$X_k = (X_k^i)_{1 \leq i \leq N} \text{ with values in } \{0,1\}^N ,$$

where the ith coordinate X_k^i is the number of the compartment in which the ith particle is located.

Starting from a sequence $(\xi_k)_{k \geq 1}$ of i.i.d. r.v. which are uniform on $\{1, \dots, N\}$, and an initial r.v. X_0 independent of this sequence, we define recursively X_k for $k \geq 1$ by changing the coordinate of rank ξ_k of X_{k-1}. This random recursion is a faithful rendering of the particle dynamics.

Theorem 1.2.3 implies that $(X_k)_{k\geq 0}$ is a Markov chain on $\{0,1\}^N$ with matrix given by

$$P(x,y) = \frac{1}{N} \text{ if } \sum_{i=1}^{N} |x^i - y^i| = 1 \ , \quad P(x,y) = 0 \text{ else.}$$

This is the symmetric nearest-neighbor random walk on the unit hypercube $\{0,1\}^N$. This chain is irreducible.

Invariant law This chain has for unique invariant law the uniform law π with density $\frac{1}{2^N}$.

As the typical magnitude of N is comparable to the Avogadro number 6.02×10^{23}, the number 2^N of configurations is enormously huge. Any computation, even for the invariant law, is of a combinatorial nature and will be most likely untractable.

1.4.4.2 Reduced macroscopic description

According to statistical mechanics, we should take advantage of the symmetries of the system, in order to stop following individual particles and consider collective behaviors instead.

A reduced macroscopic description of the system is the number of particles in compartment 1 at time $k \geq 0$, given in terms of the microscopic description by

$$S_k = \sum_{i=1}^{N} X_k^i \ , \quad \text{with values in } \{0,1,\dots,N\} \ ,$$

The information carried by S_k being less than the information carried by X_k, it is not clear that $(S_k)_{k\geq 0}$ is a Markov chain, but the symmetry of particle dynamics will allow to prove it.

For $x = (x_i)_{1\leq i\leq N} \in \{0,1\}^N$, let σ^x be the permutation of $\{1,\dots,N\}$ obtained by first placing in increasing order the i such that $x_i = 1$ and then by increasing order the i such that $x_i = 0$. Setting

$$\xi'_k = \sigma^{X_{k-1}}(\xi_k) \ ,$$

it holds that

$$S_k = S_{k-1} - \mathbb{1}_{\{S_{k-1}\leq \xi'_k\}} + \mathbb{1}_{\{S_{k-1} > \xi'_k\}} \ , \qquad k \geq 1.$$

For some deterministic f_k and g_k, using the random recursion for $(X_k)_{k\geq 0}$,

$$X_k = f_k(X_0, \xi_1, \dots, \xi_k) \ , \quad (X_0, \xi'_1, \dots, \xi'_k) = g_k(X_0, \xi_1, \dots, \xi_k) \ ,$$

and hence, for all $a \in \{0,1\}^N \times \{1,\dots,N\}^k$ and $z \in \{1,\dots,N\}$,

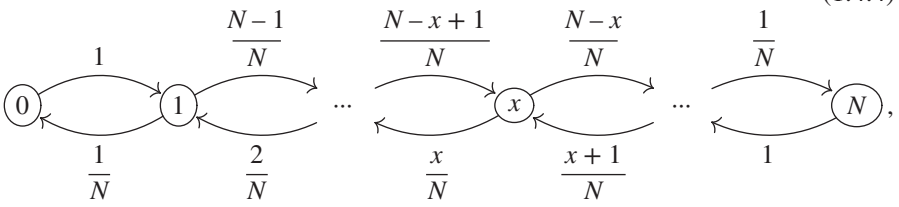
$$\begin{aligned} &\mathbb{P}((X_0, \xi'_1, \dots, \xi'_k) = a, \xi'_{k+1} = z) \\ &= \sum_{(x,z_1,\dots,z_k)\in g_k^{-1}(a)} \mathbb{P}((X_0, \xi_1, \dots, \xi_k) = (x, z_1, \dots, z_k), \end{aligned}$$

$$\sigma^{f_k(x,z_1,\dots,z_k)}(\xi_{k+1}) = z)$$
$$= \sum_{(x,z_1,\dots,z_k)\in g_k^{-1}(a)} \mathbb{P}((X_0,\xi_1,\dots,\xi_k) = (x,z_1,\dots,z_k))\frac{1}{N}$$
$$= \mathbb{P}((X_0,\xi'_1,\dots,\xi'_k) = a)\frac{1}{N}$$

as ξ_{k+1} is uniform and independent of $X_0,\xi_1,\dots,\xi_k$. Hence, ξ'_{k+1} is uniform on $\{1,\dots,N\}$ and independent of $(X_0,\xi'_1,\dots,\xi'_k)$. By a simple recursion, we conclude that the $(\xi'_k)_{k\geq 1}$ are i.i.d. r.v. which are uniform on $\{1,\dots,N\}$ and independent of X_0 and hence of S_0.

Thus, Theorem 1.2.3 yields that $(S_k)_{k\geq 0}$ is a Markov chain on $\{0,1,\dots,N\}$ with matrix Q and graph given by

$$Q(x,x+1) = \frac{N-x}{N}, \quad Q(x,x-1) = \frac{x}{N}, \quad 0 \leq x \leq N, \tag{1.4.4}$$



all other terms of Q being zero. As $(X_k)_{k\geq 0}$ is irreducible on $\{0,1\}^N$, it is clear that $(S_k)_{k\geq 0}$ is irreducible on $\{0,1,\dots,N\}$, and this can be readily checked.

Invariant law As the uniform law on $\{0,1\}^N$, with density $\frac{1}{2^N}$, is invariant for $(X_k)_{k\geq 0}$, a simple combinatorial computation yields that the invariant law for $(S_k)_{k\geq 0}$ is binomial $\mathcal{B}(N,1/2)$, given by

$$\beta = (\beta(x))_{x\in\{0,1,\dots,N\}}\,, \qquad \beta(x) = \frac{1}{2^N}\binom{N}{x}.$$

This law distributes the particles uniformly in both compartments, and this is preserved by the random evolution.

1.4.4.3 Some computations on particle distribution

At *equilibrium*, that is, under the invariant law, the X_k are uniform on $\{0,1\}^N$ and hence, the X_k^i for $i = 1,\dots,N$ are i.i.d. uniform on $\{0,1\}$. The strong law of large numbers and the central limit theorem then yield that

$$\frac{S_k}{N} \xrightarrow[N\to\infty]{\text{a.s.}} \frac{1}{2}\,, \qquad \sqrt{N}\left(\frac{S_k}{N} - \frac{1}{2}\right) \xrightarrow[N\to\infty]{\text{in law}} \mathcal{N}(0,(1/2)^2).$$

Hence, as N goes to infinity, the instantaneous proportion of molecules in each compartment converges to $1/2$ with fluctuations of order $1/\sqrt{N}$. For instance,

$$\mathbb{P}_\pi(|S_k - N/2| \geq a\sqrt{N}) \underset{N\to\infty}{\longrightarrow} 2\int_{2a}^{\infty} \mathrm{e}^{-\frac{x^2}{2}}\ \frac{dx}{\sqrt{2\pi}}\ ,$$

and as a numerical illustration, as $\int_{-\infty}^{4,5} \mathrm{e}^{-\frac{x^2}{2}}\ \frac{dx}{\sqrt{2\pi}} \simeq 0,999997$, the choice $2a = 4,5$ and $N = 6\times 10^{23}$ yields that $a\sqrt{N} \simeq 1,74\times 10^{12}$ and hence

$$\mathbb{P}_\pi(|S_k - 3\times 10^{23}| \geq 1,74\times 10^{12}) \simeq 6\times 10^{-6}.$$

For an arbitrary initial law, for $1 \leq i \leq N$ and $k \geq 1$,

$$\begin{aligned}\mathbb{P}(X_k^i = 1) &= \mathbb{P}(X_{k-1}^i = 1)\frac{N-1}{N} + \mathbb{P}(X_{k-1}^i = 0)\frac{1}{N}\\ &= \mathbb{P}(X_{k-1}^i = 1)\frac{N-2}{N} + \frac{1}{N}\end{aligned}$$

and the solution of this affine recursion, with fixed point $1/2$, is given by

$$\mathbb{P}(X_k^i = 1) = \frac{1}{2} + \left(\mathbb{P}(X_0^i = 1) - \frac{1}{2}\right)\left(\frac{N-2}{N}\right)^k.$$

Then, at geometric rate,

$$X_k^i \underset{k\to\infty}{\overset{\text{in law}}{\longrightarrow}} \frac{1}{2}(\delta_0 + \delta_1)\ , \qquad \mathbb{E}\left(\frac{1}{N}S_k\right) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{P}(X_k^i = 1) \underset{k\to\infty}{\longrightarrow} \frac{1}{2}\ ,$$

The rate $\frac{N-2}{N}$ seems poor, but the time unit should actually be of order $1/N$, and

$$\left(\frac{N-2}{N}\right)^{Nk} \underset{N\to\infty}{\longrightarrow} \mathrm{e}^{-2k}.$$

Explicit variance computations can also be done, which show that $\frac{1}{N}S_k$ converges in probability to $1/2$, but in order to go further some tools must be introduced.

1.4.4.4 Random walk, Fourier transform, and spectral decomposition

The Markov chain $(X_k)_{k\geq 0}$ on $\{0,1\}^N$ (microscopic representation) can be obtained by taking a sequence $(U_k)_{k\geq 1}$ of i.i.d. r.v. which are uniform on the vectors of the canonical basis, independent of X_0, and setting

$$X_k = X_{k-1} + U_k \pmod 2\ , \qquad k \geq 1.$$

This is a symmetric nearest-neighbor random walk on the additive group $\{0,1\}^N := (\frac{\mathbb{Z}}{2\mathbb{Z}})^N$, and we are going to exploit this structure, according to a technique adaptable to other random walks on groups.

For b and x in $\{0,1\}^N$ and for vectors $v = (v(x))_{x\in\{0,1\}^N}$ and $w = (w(x))_{x\in\{0,1\}^N}$, the canonical scalar products will be respectively denoted by

$$b \cdot x := \sum_{1\le i\le N} b_i x_i \ , \qquad \langle v, w\rangle := \sum_{x\in\{0,1\}^N} v(x)w(x).$$

Let us associate to each r.v. X on $\{0,1\}^N$ its characteristic function, which is the (discrete) Fourier transform of its law π_X given by, with the notation $\mathrm{e}_b = ((-1)^{b\cdot x})_{x\in\{0,1\}^N}$,

$$F_X : b \in \{0,1\}^N \mapsto \mathbb{E}((-1)^{b\cdot X}) = \sum_{x\in\{0,1\}^N} \mathbb{P}(X = x)(-1)^{b\cdot x} = \langle \pi_X, \mathrm{e}_b\rangle.$$

For $b \neq c$ in $\{0,1\}^N$ and any $1 \le i \le N$ such that $b_i \neq c_i$,

$$\langle \mathrm{e}_b, \mathrm{e}_b\rangle = \sum_{x\in\{0,1\}^N} (-1)^{2b\cdot x} = 2^N \ ,$$

$$\langle \mathrm{e}_b, \mathrm{e}_c\rangle = \sum_{x\in\{0,1\}^N \,:\, x_i=0} (-1)^{b\cdot x + c\cdot x}(1 + (-1)) = 0 \ ,$$

hence $(\mathrm{e}_b)_{b\in\{0,1\}^N}$ is an orthogonal basis of vectors, each with square product 2^N. This basis could easily be transformed into an orthonormal basis.

Fourier inversion formula From this follows the inversion formula

$$\pi_X = \frac{1}{2^N} \sum_{b\in\{0,1\}^N} F_X(b)\mathrm{e}_b \ ,$$

$$\pi_X(x) = \frac{1}{2^N} \sum_{b\in\{0,1\}^N} F_X(b)(-1)^{b\cdot x} = \frac{1}{2^N}\langle F_X, \mathrm{e}_x\rangle \ , \quad x \in \{0,1\}^N.$$

Fourier transform and eigenvalues For $b \in \{0,1\}^N$, setting

$$\lambda(b) = F_{U_1}(b) = \frac{1}{N}\sum_{i=1}^N (-1)^{b_i} = \frac{1}{N}\sum_{i=1}^N (1 - 2b_i) = \frac{N - 2\sum_{i=1}^N b_i}{N} \ ,$$

it holds that

$$F_{X_k}(b) = \lambda(b)F_{X_{k-1}}(b) = \cdots = \lambda(b)^k F_{X_0}(b)$$

and thus e_b is an eigenvector for the eigenvalue $\lambda(b)$ for the transition matrix. There are $N + 1$ distinct eigenvalues

$$\lambda_j = \frac{N - 2j}{N} \ , \qquad 0 \le j \le N \ ,$$

of which the eigenspace of dimension $\binom{N}{j}$ is generated by the e_b such that b has exactly j terms taking the value 1.

Spectral decomposition of the transition matrix This yields the spectral decomposition of P in an orthogonal basis and

$$\pi_{X_k} = \frac{1}{2^N} \sum_{b,y\in\{0,1\}^N} \lambda(b)^k \langle \pi_{X_0}, \mathrm{e}_b \rangle \mathrm{e}_b \ ,$$

$$\pi_{X_k}(x) = \frac{1}{2^N} \sum_{b,y\in\{0,1\}^N} \lambda(b)^k \pi_{X_0}(y)(-1)^{b\cdot(x+y)} \ , \qquad x \in \{0,1\}^N.$$

Long-time behavior This yields that $F_{X_k}(b)$ is $\mathcal{O}((\frac{N-2}{N})^k)$ for $b \notin \{0,1\}$ (constant vectors) and that

$$F_{X_k}(0) = 1 \ , \ F_{X_k}(1) = (-1)^k F_{X_0}(1) = (-1)^k (\mathbb{P}(S_0 \in 2\mathbb{N}) - \mathbb{P}(S_0 \in 2\mathbb{N}+1)).$$

For $\alpha = 2p-1 \in [-1,1] \iff p = \frac{1+\alpha}{2} \in [0,1]$, let F_α be such that $F_\alpha(0) = 1$, $F_\alpha(1) = \alpha$ and $F_\alpha(b) = 0$ for $b \notin \{0,1\}$, and π_α be the corresponding laws. Then, $F_\alpha = pF_1 + (1-p)F_{-1}$ and thus $\pi_\alpha = p\pi_1 + (1-p)\pi_{-1}$ and

$$\pi_1(x) = \frac{1}{2^N}(1 + (-1)^{1\cdot x}) \ , \quad \pi_{-1}(x) = \frac{1}{2^N}(1 - (-1)^{1\cdot x}) \ , \qquad x \in \{0,1\}^N \ ,$$

that is, π_1 and π_{-1} are the uniform laws respectively on

$$\left\{ x \in \{0,1\}^N \, : \, 1 \cdot x = \sum_{i=1}^N x_i \in 2\mathbb{N} \right\} \ , \quad \left\{ x \in \{0,1\}^N \, : \, \sum_{i=1}^N x_i \in 2\mathbb{N}+1 \right\}$$

Let

$$\alpha = \mathbb{P}(S_0 \in 2\mathbb{N}) - \mathbb{P}(S_0 \in 2\mathbb{N}+1) = 2\mathbb{P}(S_0 \in 2\mathbb{N}) - 1.$$

The law π_α is the mixture of π_1 andπ_{-1}, which respects the probability that S_0 be in $2\mathbb{N}$ and in $2\mathbb{N}+1$, and $\pi_{-\alpha}$ the mixture that interchanges these. If X_0 is of law π_α, then X_k is of law $\pi_{(-1)^k\alpha}$ and thus

$$F_{X_k}(b) - F_{(-1)^k\alpha}(b) = \lambda(b)^k (F_{X_0}(b) - F_\alpha(b))$$

and, as $F_{X_0}(0) = F_\alpha(0)$ and $F_{X_0}(1) = F_\alpha(1)$, for the Hilbert norm associated with $\langle \cdot, \cdot \rangle$ it holds that

$$\begin{aligned} \|\pi_{X_k} - \pi_{(-1)^k\alpha}\|^2 &= \frac{1}{2^N} \sum_{b\in\{0,1\}^N} \lambda(b)^{2k} (F_{X_0}(b) - F_\alpha(b))^2 \\ &\leq \left(\frac{N-2}{N}\right)^{2k} \|\pi_{X_0} - \pi_\alpha\|^2. \end{aligned}$$

In particular, the law of X_{2n} converges exponentially fast to π_α and the law of X_{2n+1} to $\pi_{-\alpha}$.

Periodic behavior The behavior we have witnessed is related to the notion of periodicity: S_{2n} is even or odd the same as S_0, and S_{2n+1} has opposite parity than S_0. This is obviously related to the fact that -1 is an eigenvalue.

1.4.5 Renewal process

A component of a system (electronic component, machine, etc.) lasts a random life span before failure. It is visited at regular intervals (at times n in $\mathbb{N}$) and replaced appropriately. The components that are used for this purpose behave in i.i.d. manner.

1.4.5.1 Modeling

At time 0, a first component is installed, and the ith component is assumed to have a random life span before replacement given by D_i, where the $(D_i)_{i\geq 1}$ are i.i.d. on $\{1, 2, \ldots\} \cup \{\infty\}$. Let D denote an r.v. with same law as D_i, representing a generic life span. It is often assumed that $\mathbb{P}(D = \infty) = 0$.

Let X_n denote the age of the component in function at time $n \geq 0$, with $X_n = 0$ if it is replaced at that time. Setting

$$T_0 = 0\ , \qquad T_k = T_{k-1} + D_k = D_1 + \cdots + D_k\ ,\ \ k \geq 1\ ,$$

it holds that $X_n = n - T_{k-1}$ sur $T_{k-1} \leq n < T_k$ (Figure 1.5).

The X_n are defined for all $n \geq 0$ as $\mathbb{P}(D \geq 1) = 1$. If $\mathbb{P}(D = \infty) = 0$, then all T_k are finite, a.s., else if $\mathbb{P}(D = \infty) > 0$, then there exists an a.s. finite r.v. K such that $D_K = \infty$ and $T_K = T_{K+1} = \cdots = \infty$.

This natural representation in terms of the life spans is *not* a random recursion of the kind discussed in Theorem 1.2.3. We will give a direct proof that $(X_n)_{n\geq 0}$ is a Markov chain and give its transition matrix.

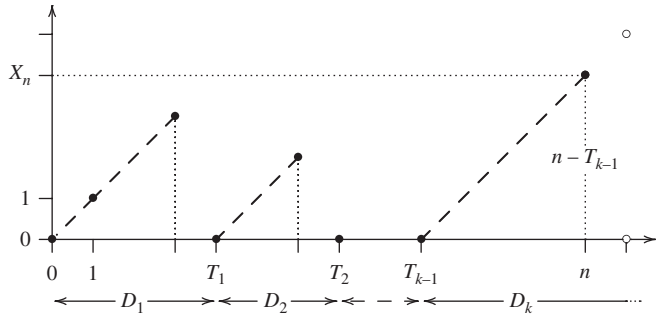


Figure 1.5 Renewal process. The • represent the ages at the discrete instants on the horizontal axis and are linearly interpolated by dashes in their increasing phases. Then, $X_n = n - T_{k-1}$ if $T_{k-1} \leq n < T_k = T_{k-1} + D_k$. The ∘ represent the two possible ages at time $n + 1$, which are $X_{n+1} = X_n + 1$ if $D_k > X_n + 1$ and $X_{n+1} = 0$ if $D_k = X_n$.

Note that $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = 0$ except if $x_0 = 0$ and x_k is in $\{0, x_{k-1} + 1\}$ for $1 \le k \le n$. These are the only cases to be considered and then

$$\{X_0 = x_0, \dots, X_n = x_n\} = \{D_1 = d_1, \dots, D_{k-1} = d_{k-1}, D_k > x_n\},$$

where k is the number of 0 in $(x_0, \dots, x_n)$ and 0, $t_1 = d_1$, $t_2 = d_1 + d_2$, $\dots$, $t_{k-1} = d_1 + \cdots + d_{k-1}$ are their ranks (Figure 1.5). This can be written as

$$k = \sum_{i=0}^{n} \mathbb{1}_{\{x_i = 0\}}\ , \qquad d_j = \inf\{i \ge 1 : x_{d_1 + \cdots + d_{j-1} + i} = 0\}\ , \quad 1 \le j \le k-1.$$

The independence of $(D_1, \dots, D_{k-1})$ and D_k yields that

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = x_n + 1) \\
&\quad = \mathbb{P}(D_1 = d_1, \dots, D_{k-1} = d_{k-1}, D_k > x_n, X_{n+1} = x_n + 1) \\
&\quad = \mathbb{P}(D_1 = d_1, \dots, D_{k-1} = d_{k-1}, D_k > x_n + 1) \\
&\quad = \mathbb{P}(D_1 = d_1, \dots, D_{k-1} = d_{k-1}) \mathbb{P}(D_k > x_n + 1) \\
&\quad = \mathbb{P}(D_1 = d_1, \dots, D_{k-1} = d_{k-1}, D_k > x_n)\ \frac{\mathbb{P}(D_k > x_n + 1)}{\mathbb{P}(D_k > x_n)} \\
&\quad = \mathbb{P}(X_0 = x_0, \dots, X_n = x_n)\ \frac{\mathbb{P}(D > x_n + 1)}{\mathbb{P}(D > x_n)}.
\end{aligned}$$

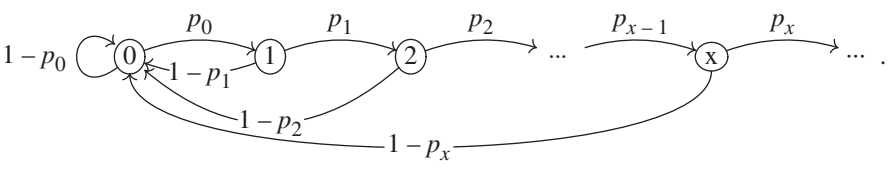
Moreover,

$$\begin{aligned}
&\mathbb{P}(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = 0) \\
&\qquad = \mathbb{P}(X_0 = x_0, \dots, X_n = x_n)\ \frac{\mathbb{P}(D = x_n + 1)}{\mathbb{P}(D > x_n)}
\end{aligned}$$

is obtained similarly, or by complement to 1.

Hence, the only thing that matters is the age of the component in function, and $(X_n)_{n \ge 0}$ is a Markov chain on $\mathbb{N}$ with matrix P and graph given by

$$P(x, x+1) = \frac{\mathbb{P}(D > x+1)}{\mathbb{P}(D > x)} = \mathbb{P}(D > x+1 | D > x) := p_x,$$
$$P(x, 0) = \mathbb{P}(D = x+1 | D > x) = 1 - p_x, \quad x \in \mathbb{N},$$



(1.4.5)

This Markov chain is irreducible if and only if $\mathbb{P}(D > k) > 0$ for every $k \in \mathbb{N}$ and $\mathbb{P}(D = \infty) < 1$.

A mathematically equivalent description Thus, from a mathematical perspective, we can start with a sequence $(p_x)_{x\in\mathbb{N}}$ with values in $[0, 1]$ and assume that a component with age x at an arbitrary time $n \in \mathbb{N}$ has probability p_x to pass the inspection at time $n + 1$, and else a probability $1 - p_x$ to be replaced then. In this setup, the law of D is determined by

$$\mathbb{P}(D > x) = p_0 \cdots p_{x-1} , \qquad x \geq 0.$$

This formulation is not as natural as the preceding one. It corresponds to a random recursion given by a sequence $(\xi_n)_{n\geq 1}$ of i.i.d. uniform r.v. on $[0, 1]$, independent of X_0, and

$$X_n = X_{n-1} + \sum_{x\geq 0} \mathbb{1}_{\{X_{n-1}=x\}}(\mathbb{1}_{\{\xi_n \leq p_x\}} - x\mathbb{1}_{\{\xi_n > p_x\}}) , \qquad n \geq 1.$$

The renewal process is often introduced in this manner, in order to avoid the previous computations. It is an interesting example, as we will discuss later.

Invariant measures and laws An invariant measure $\mu = (\mu(x))_{x\in\mathbb{N}}$ satisfies

$$\sum_{x\geq 0} \mu(x)(1 - p_x) = \mu(0) , \qquad \mu(x - 1)p_{x-1} = \mu(x) , \quad x \geq 1 ,$$

thus $\mu(x) = \mu(0)p_0 \cdots p_{x-1} = \mu(0)\mathbb{P}(D > x)$ that yields uniqueness, and existence holds if and only if

$$\mathbb{P}(D = \infty) := \lim_{x\to\infty} p_0 \cdots p_{x-1} = 0.$$

This unique invariant measure can be normalized, in order to yield an invariant law, if and only if it is finite, that is, if

$$\mathbb{E}(D) := \sum_{x\geq 0} \mathbb{P}(D > x) := \sum_{x\geq 0} p_0 \cdots p_{x-1} < \infty ,$$

and then

$$\pi(x) = \frac{\mathbb{P}(D > x)}{\mathbb{E}(D)} = \frac{p_0 \cdots p_{x-1}}{\sum_{y\geq 0} p_0 \cdots p_{y-1}} , \qquad x \geq 0.$$

Renewal process and Doeblin condition A class of renewal processes is one of the rare natural examples of infinite state space Markov chains satisfying the Doeblin condition.

Lemma 1.4.1 *Assume that there exists $m \geq 0$ such that $\inf_{x\geq m}(1 - p_x) > 0$. Then, the Markov chain satisfies the Doeblin condition for $k = m + 1$ and $\varepsilon = \inf_{x\geq m}(1 - p_x)$ and $\hat{\pi} = \delta_0$, and the conclusions of Theorem 1.3.4 hold.*

Proof: This can be checked easily. ∎

1.4.6 Word search in a character chain

A source emits an infinite i.i.d. sequence of "characters" of some "alphabet." We are interested in the successive appearances of a certain "word" in the sequence.

For instance, the characters could be 0 and 1 in a computer system, "red" or "black" in a roulette game, A, C, G, T in a DNA strand, or ASCII characters for a typewriting monkey. Corresponding words could be 01100010, red-red-red-black, GAG, and Abracadabra.

Some natural questions are the following:

- Is any word going to appear in the sequence?
- Is it going to appear infinitely often, and with what frequency?
- What is the law and expectation of the first appearance time?

1.4.6.1 Counting automaton

A general method will be described on a particular instance, the search for the occurrences of the word GAG.

Two different kinds of occurrences can be considered, *without* or *with* overlaps; for instance, GAGAG contains one single occurrence of GAG without overlaps but two with. The case without overlaps is more difficult, and more useful in applications; it will be considered here, but the method can be readily adapted to the other case.

We start by defining a counting automaton with four states $\emptyset$, G, GA, and GAG, which will be able to count the occurrences of GAG in any arbitrary finite character chain. The automaton starts in state $\emptyset$ and then examines the chain sequentially term by term, and:

- In state $\emptyset$: if the next state is G, then it takes state G, else it stays in state $\emptyset$,
- In state G: if the next state is A, then it takes state GA, if the next state is G, then it stays in state G, else it takes state $\emptyset$,
- In state GA: if the next state is G, then it takes state GAG, else it takes state $\emptyset$,
- In state GAG: if the next state is G, then it takes state G, else it takes state $\emptyset$.

Such an automaton can be represented by a graph that is similar to a Markov chain graph, with nodes given by its possible states and oriented edges between nodes marked by the logical condition for this transition (Figure 1.6).

This automation is now used on a sequence of characters given by an i.i.d. sequence $(\xi_n)_{n\geq 1}$ such that

$$\mathbb{P}(\xi_1 = \mathrm{A}) = p_{\mathrm{A}}\,, \qquad \mathbb{P}(\xi_1 = \mathrm{G}) = p_{\mathrm{G}}\,, \qquad \mathbb{P}(\xi_1 \notin \{\mathrm{A}, \mathrm{G}\}) = 1 - p_{\mathrm{A}} - p_{\mathrm{G}}\,,$$

satisfying $p_{\mathrm{A}} > 0$, $p_{\mathrm{G}} > 0$, and $p_{\mathrm{A}} + p_{\mathrm{G}} \leq 1$.

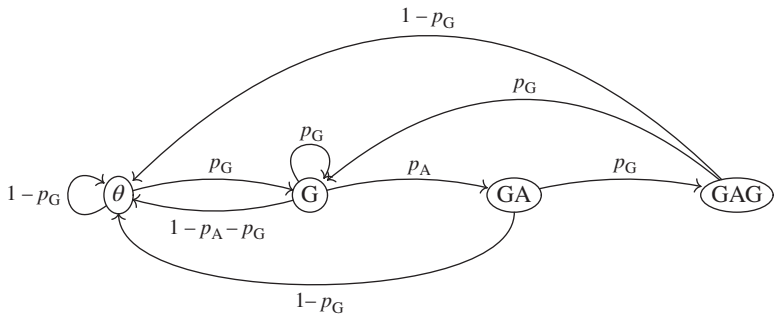


Figure 1.6 Search for the word GAG: Markov chain graph. The graph for the automaton is obtained by replacing p_A *by "if the next term is* A*,"* p_G *by "if the next term is* G*,"* $1-p_G$ *by "if the next term is not* G*," and* $1-p_A-p_G$ *by "if the next term is neither* A *nor* G*."*

Let $X_0 = \emptyset$, and X_n be the state of the automaton after having examined the nth character. Theorem 1.2.3 yields that $(X_n)_{n\geq 0}$ is a Markov chain with graph, given in Figure 1.6, obtained from the automaton graph by replacing the logical conditions by their probabilities of being satisfied.

Markovian description All relevant information can be written in terms of $(X_n)_{n\geq 0}$. For instance, if $T_0 = 0$ and T_i denotes the time of the ith occurrence (complete, without overlaps) of the word for $i \geq 1$, and N_k the number of such occurrences taking place before $k \geq 1$, then

$$T_i = \inf\{n > T_{i-1} : X_n = \mathrm{GAG}\}\,, \qquad N_k = \sum_{n=1}^{k} \mathbb{1}_{\{X_n=\mathrm{GAG}\}}.$$

The transition matrix $P = (P(x,y))_{x,y\in\{\emptyset,\mathrm{G},\mathrm{GA},\mathrm{GAG}\}}$ is irreducible and has for unique invariant law

$$\pi = \left(\frac{1-p_G-p_A p_G^2}{1+p_A p_G}\,, \frac{p_G}{1+p_A p_G}\,, \frac{p_A p_G}{1+p_A p_G}\,, \frac{p_A p_G^2}{1+p_A p_G} \right).$$

Occurrences with overlaps In order to search for the occurrences *with* overlaps, it would suffice to modify the automaton by considering the overlaps inside the word. For the word GAG, we need only modify the transitions from state GAG: if the next term is G, then the automaton should take state G, and if the next term is A, then it should take state GA, else it should take state $\emptyset$. For more general overlaps, this can become very involved.

1.4.6.2 Snake chain

We describe another method for the search for the occurrences *with* overlaps of a word $c_1 \cdots c_\ell$ of length $\ell \geq 1$ in an i.i.d. sequence $(\xi_n)_{n\geq 1}$ of characters from some alphabet $\mathcal{V}$.

Setting $Z_n = (\xi_n, \dots, \xi_{n+\ell-1})$, then

$$T_i = \inf\{n > T_{i-1} \,:\, Z_{n-\ell+1} = (c_1, \dots, c_\ell)\}$$

is the time of the ith occurrence of the word, $i \geq 1$ (with $T_0 = 0$), and

$$N_k = \sum_{n=\ell}^{k} \mathbb{1}_{\{Z_{n-\ell+1}=(c_1,\dots,c_\ell)\}}$$

is the number of such occurrences before time $k \geq \ell$. In general, $(Z_n)_{n\geq 1}$ is not i.i.d., but it will be seen to be a Markov chain.

More generally, let $(Y_n)_{n\geq 0}$ be a Markov chain on $\mathcal{V}$ of arbitrary matrix P and $Z_n = (Y_n, \dots, Y_{n+\ell-1})$ for $n \geq 0$. Then, $(Z_n)_{n\geq 0}$ is a Markov chain on $\mathcal{V}^\ell$ with matrix P_ℓ with only nonzero terms given by

$$P_\ell((x_1, \dots, x_\ell), (x_2, \dots, x_\ell, y)) = P(x_\ell, y)\ , \qquad x_1, \dots, x_\ell, y \in \mathcal{V}\ ,$$

called the snake chain of length ℓ for $(Y_n)_{n\geq 0}$. The proof is straightforward if the conditional formulation is avoided.

Irreducibility If P is irreducible, then P_ℓ is irreducible on its natural state space

$$\mathcal{V}_\ell = \{(x_1, \dots, x_\ell) \in \mathcal{V}^\ell \,:\, P(x_1, x_2) \cdots P(x_{\ell-1}, x_\ell) > 0\}.$$

Invariant Measures and Laws If μ is an invariant measure for $(Y_n)_{n\geq 0}$, then μ_ℓ given by

$$\mu_\ell(y_1, \dots, y_\ell) = \mu(y_1)P(y_1, y_2) \cdots P(y_{\ell-1}, y_\ell)$$

is immediately seen to be an invariant measure for $(Z_n)_{n\geq 0}$. If further μ is a law, then μ_ℓ is also a law.

In the i.i.d. case where $P(x, y) = p(y)$, the only invariant law for $(Y_n)_{n\geq 0}$ is given by $\pi(y) = p(y)$, and the only invariant law for $(Z_n)_{n\geq 0}$ by the product law $\pi_\ell(y_1, \dots, y_\ell) = p(y_1) \cdots p(y_\ell)$.

1.4.7 Product chain

Let P_1 and P_2 be two transition matrices on $\mathcal{V}_1$ and $\mathcal{V}_2$, and the matrices $P_1 \otimes P_2$ on $\mathcal{V}_1 \times \mathcal{V}_2$ have generic term

$$P_1 \otimes P_2((x^1, x^2), (y^1, y^2)) = P_1(x^1, y^1)P_2(x^2, y^2).$$

Then, $P_1 \otimes P_2$ is a transition matrix on $\mathcal{V}_1 \times \mathcal{V}_2$, as in the sense of product laws,

$$P_1 \otimes P_2((x^1, x^2), \cdot) = P_1(x^1, \cdot) \otimes P_2(x^2, \cdot).$$

see below for more details. See Figure 1.7.

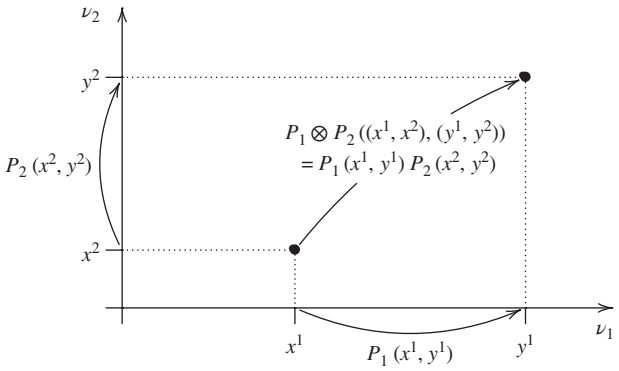


Figure 1.7 Product chain. The first and second coordinates are drawn independently according to P_1 and P_2.

The Markov chain $(X_n^1, X_n^2)_{n\geq 0}$ with matrix $P_1 \otimes P_2$ is called the product chain. Its transitions are obtained by independent transitions of each coordinate according, respectively, to P_1 and P_2. In particular, $(X_n^1)_{n\geq 0}$ and $(X_n^2)_{n\geq 0}$ are two Markov chains of matrices P_1 and P_2, which conditional on (X_0^1, X_0^2) are independent, and

$$(P_1 \otimes P_2)^n = P_1^n \otimes P_2^n\,, \qquad n \geq 0.$$

1.4.7.1 Invariant measures and laws

Immediate computations yield that if μ_1 is an invariant measure for P_1 and μ_2 for P_2, then the product measure $\mu_1 \otimes \mu_2$ given by

$$(\mu_1 \otimes \mu_2)(x_1, x_2) = \mu_1(x_1)\mu_2(x_2)$$

is invariant for $P_1 \otimes P_2$. Moreover,

$$\|\mu_1 \otimes \mu_2\|_{var} = \|\mu_1\|_{var} \times \|\mu_2\|_{var}$$

and thus if μ_1 and μ_2 are laws then $\mu_1 \otimes \mu_2$ is a law.

1.4.7.2 Irreducibility problem

The matrix $P = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ on $\{1, 2\}$ is irreducible and has unique invariant law the uniform law, whereas a Markov chain with matrix $P \otimes P$ alternates either between $(1, 1)$ and $(2, 2)$ or between $(1, 2)$ and $(2, 1)$, depending on the initial state and is not irreducible on $\{1, 2\}^2$. The laws

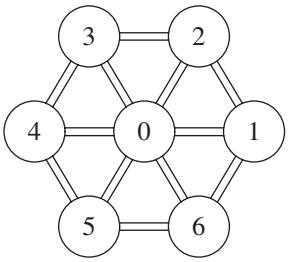
$$\frac{1}{2}(\delta_{(1,1)} + \delta_{(2,2)})\,, \qquad \frac{1}{2}(\delta_{(1,2)} + \delta_{(2,1)})\,,$$

are invariant for $P \otimes P$ and generate the space of invariant measures.

All this can be readily generalized to an arbitrary number of transition matrices.

Exercises

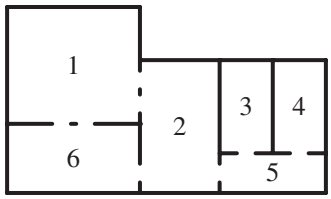
1.1 The space station, 1 An aimless astronaut wanders within a space station, schematically represented as follows:



The space station spins around its center in order to create artificial gravity in its periphery. When the astronaut is in one of the peripheral modules, the probability for him to go next in each of the two adjacent peripheral modules is twice the probability for him to go to the central module. When the astronaut is in the central module, the probability for him to go next in each of the six peripheral modules is the same.

Represent this evolution by a Markov chain and give its matrix and graph. Prove that this Markov chain is irreducible and give its invariant law.

1.2 The mouse, 1 A mouse evolves in an apartment, schematically represented as follows:



The mouse chooses uniformly an opening of the room where it is to go into a new room. It has a short memory and forgets immediately where it has come from.

Represent this evolution by a Markov chain and give its matrix and graph. Prove that this Markov chain is irreducible and give its invariant law.

1.3 Doubly stochastic matrices Let $P = (P(x, y))_{x,y\in\mathcal{V}}$ be a doubly stochastic matrix on a state space $\mathcal{V}$: by definition,

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

a) Find a simple invariant measure for P.

b) Prove that P^n is doubly stochastic for all $n \geq 1$.

c) Prove that the transition matrix for a random walk on a network is doubly stochastic.

1.4 The Labouchère system, 1 In a game where the possible gain is equal to the wager, the probability of gain p of the player at each draw typically satisfies $p \leq 1/2$ and even $p < 1/2$, but is usually close to $1/2$, as when betting on red or black at roulette. In this framework, the Labouchère system is a strategy meant to provide a means for earning in a secure way a sum $S \geq 1$ determined in advance.

The sum S is decomposed arbitrarily as a sum of $k \geq 1$ positive terms, which are put in a list. The strategy then transforms recursively this list, until it is empty.

At each draw, if $k \geq 2$ then the sum of the first and last terms of the list are wagered, and if $k \geq 1$ then the single term is wagered. If the gambler wins, he or she removes from the list the terms concerned by the wager. If the gambler loses, he or she retains these terms and adds at the end of the list a term worth the sum just wagered. The game stops when $k = 0$, and hence, the sum S has been won.

(Martingale theory proves that in realistic situations, for instance, if wagers are bounded or credit is limited, then with a probability close to 1 the sum S is indeed won, but with a small probability a huge loss occurs, large enough to prevent the gambler to continue the game and often to ever gamble in the future.)

a) Represent the list evolution by a Markov chain $(L_n)_{n\geq 0}$ on the set

$$\mathcal{V} = \bigcup_{k\geq 0} \mathbb{N}^k$$

of words of the form $n_1 \cdots n_k$. Describe its transition matrix Q and its graph. Prove that if L_n reaches $\emptyset$ (the empty word), then the gambler wins the sum S.

b) Let X_n be the length of the list (or word) L_n for $n \geq 0$. Prove that $(X_n)_{n\geq 0}$ is a Markov chain on $\mathbb{N}$ and give its matrix P and its graph.

1.5 Three-card Monte Three playing cards are lined face down on a cardboard box at time $n = 0$. At times $n \geq 1$, the middle card is exchanged with probability $p > 0$ with the card on the right and with probability $q = 1 - p > 0$ with the one on the left.

a) Represent the evolution of the three cards by a Markov chain $(Y_n)_{n\geq 0}$. Give its transition matrix Q and its graph. Prove that $(Y_n)_{n\geq 0}$ is irreducible. Find its invariant law ρ.

b) The cards are the ace of spades and two reds. Represent the evolution of the ace of spades by a Markov chain $(X_n)_{n\geq 0}$. Give its transition matrix P and its graph. Prove that it is irreducible. Find its invariant law π.

c) Compute P^n in terms of the initial law π_0 and p and $n \geq 1$. Prove that the law π_n of X_n converges to π as n goes to infinity, give an exponential convergence rate for this convergence, and find for which value of p the convergence is fastest.

1.6 Andy, 1 If Andy is drunk one evening, then he has one odd in ten to end up in jail, in which case will remain sober the following evening. If Andy is drunk one evening and does not end up in jail, then he has one odd in two to be drunk the following evening. If Andy stays sober one evening, then he has three odds out of four to remain sober the following evening.

It is assumed that $(X_n)_{n\geq 0}$ constitutes a Markov chain, where $X_n = 1$ if Andy on the n-th evening is drunk and ends up in jail, $X_n = 2$ if Andy then is drunk and does not end up in jail, and $X_n = 3$ if then he remains sober.

Give the transition matrix P and the graph for $(X_n)_{n\geq 0}$. Prove that P is irreducible and compute its invariant law. Compute P^n in terms of $n \geq 0$. What is the behavior of X_n when n goes to infinity?

1.7 Squash Let us recall the original scoring system for squash, known as English scoring. If the server wins a rally, then he or she scores a point and retains service. If the returner wins a rally, then he or she becomes the next server but no point is scored. In a game, the first player to score 9 points wins, except if the score reaches 8-8, in which case the returner must choose to continue in either 9 or 10 points, and the first player to reach that total wins.

A statistical study of the games between two players indicates that the rallies are won by Player A at service with probability $a > 0$ and by Player B at service with probability $b > 0$, each in i.i.d. manner.

The situation in which Player A has i points, Player B has j points, and Player L is at service is denoted by (i, j, L) in $\mathcal{V} = \{0, 1, \ldots, 10\}^2 \times \{A, B\}$.

a) Describe the game by a Markov chain on $\mathcal{V}$, assuming that if the score reaches 8-8 then they play on to 10 points (the play up to 9 can easily be deduced from this), in the two following cases: (i) all rallies are considered and (ii) only point scoring is considered.

b) Trace the graphs from arriving at 8-8 on the service of Player A to end of game.

c) A game gets to 8-8 on the service of Player A. Compute in terms of a and b the probability that Player B wins according to whether he or she elects to go to 9 or 10 points. Counsel Player B on this difficult choice.

1.8 Genetic models, 1 Among the individuals of a species, a certain gene can appear under $K \geq 2$ different forms called alleles.

In a **microscopic** (individual centered) model for a population of $N \geq 1$ individuals, these are arbitrarily numbered, and the fact that individual i carries allele a_i is coded by the state

$$(a_i)_{1 \leq i \leq N} \in \{1, \dots, K\}^N.$$

A **macroscopic** representation only retains the numbers of individuals carrying each allele, and the state space is

$$\{(n_1, \dots, n_K) \in \mathbb{N}^K \, : \, n_1 + \cdots + n_K = N\}.$$

We study two simplified models for the reproduction of the species, in which the population size is fixed, and where the selective advantage of every allele a w.r.t. the others is quantified by a real number $c(a) > 0$.

1. **Synchronous: Fisher–Wright model**: at each step, the whole population is replaced by its descendants, and in i.i.d. manner, each new individual carries allele a with a probability proportional both to $c(a)$ and to the number of old individuals carrying allele a.
2. **Asynchronous: Moran model**: at each step, an uniformly chosen individual is replaced by a new individual, which carries allele a with a probability proportional both to $c(a)$ and to the number of old individuals carrying allele a.
   a) Explain how to obtain the macroscopic representation from the microscopic representation
   b) Prove that each pair representation-model corresponds to a Markov chain. Give the transition matrices and the absorbing states.

1.9 Records Let $(X_i)_{i \geq 1}$ be i.i.d. r.v. such that $\mathbb{P}(X_i = 1) = p > 0$ and $\mathbb{P}(X_i = 0) = 1 - p > 0$, and R_n be the greatest number of consecutive 1 observed in $(X_1, \dots, X_n)$.

a) Show that $(R_n)_{n \geq 0}$ is not a Markov chain.

b) Let $X_0 := 0$, and

$$D_n = \inf\{k \geq 0 \, : \, X_{n-k} = 0\} \, , \qquad n \geq 0.$$

Prove that $(D_n)_{n \geq 0}$ is a Markov chain and give its transition matrix P. Prove that there exists a unique invariant law π and compute it.

c) Let $k \geq 0$,

$$S_k = \inf\{n \geq 0 \, : \, D_n = k\} \, , \quad Z_n = D_n \text{ if } n \leq S_k \, , \text{ else } Z_n = D_n.$$

Prove that $(Z_n)_{n \geq 0}$ is a Markov chain on $\{0, 1, \dots, k\}$ and give its transition matrix P_k.

d) Express $\mathbb{P}(R_n \geq k)$ in terms of Z_n, then of P_k. Deduce from this the law of R_n.

e) What is the probability of having at least 5 consecutive heads among 100 fair tosses of head-or-tails? One can use the fact that for $p = 1/2$,

$$P_5^{100} = \begin{pmatrix} 0,09659 & 0,04913 & 0,02499 & 0,01271 & 0,00647 & 0,81011 \\ 0,09330 & 0,04746 & 0,02414 & 0,01228 & 0,00625 & 0,81658 \\ 0,08683 & 0,04417 & 0,02247 & 0,01143 & 0,00581 & 0,82929 \\ 0,07412 & 0,03770 & 0,01918 & 0,00976 & 0,00496 & 0,85428 \\ 0,04913 & 0,02499 & 0,01271 & 0,00647 & 0,00329 & 0,90341 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

1.10 Incompressible mixture, 1 There are two urns, N white balls, and N black balls. Initially N balls are set in each urn. In i.i.d. manner, a ball is chosen uniformly in each urn and the two are interchanged. The white balls are numbered from 1 to N and the black balls from $N + 1$ to $2N$. We denote by A_n the r.v. with values in

$$\mathcal{V} = \{E \subset \{1, \dots, 2N\} : \mathrm{Card}(E) = N\}$$

given by the set of the numbers in the first urn just after time $n \geq 0$ and by

$$S_n = \sum_{i=1}^{N} \mathbb{1}_{\{i \in A_n\}}$$

the corresponding number of white balls.

a) Prove that $(A_n)_{n\in\mathbb{N}}$ is an irreducible Markov chain on $\mathcal{V}$ and give its matrix P. Prove that the invariant law π is unique and compute it.

b) Do the same for $(S_n)_{n\geq 0}$ on $\{0, 1, \dots, N\}$, with matrix Q and invariant law σ.

c) For $1 \leq i \leq 2N$ find a recursion for $\mathbb{P}(i \in A_n)$, and solve it in terms of n and $\mathbb{P}(i \in A_0)$. Do likewise for $\mathbb{E}(S_n)$. What happens for large n?

1.11 Branching with immigration The individuals of a generation disappear at the following, leaving there k descendants each with probability $p(k) \geq 0$, and in addition, $i \in \mathbb{N}$ immigrants appear with probability $q(i) \geq 0$, where $\sum_{k\geq 0} p(k) = \sum_{k\geq 0} q(k) = 1$. Let

$$g(s) = \sum_{k\in\mathbb{N}} p(k)s^k \,, \quad h(s) = \sum_{k\in\mathbb{N}} q(k)s^k \,, \qquad 0 \leq s \leq 1 \,,$$

be the generating functions for the reproduction and the immigration laws.

Similarly to Section 1.4.3, using X_0 with values in $\mathbb{N}$ and $\xi_{n,i}$ and ζ_n for $n \geq 1$ and $i \geq 1$ such that $\mathbb{P}(\xi_{n,i} = k) = p(k)$ and $\mathbb{P}(\zeta_n = k) = q(k)$ for k in $\mathbb{N}$,

all these r.v. being independent, let us represent the number of individuals in generation $n \in \mathbb{N}$ by

$$X_n = \zeta_n + \sum_{i=1}^{X_{n-1}} \xi_{n,i}.$$

Let G_n be the generating function of X_n.

a) Prove that $(X_n)_{n\in\mathbb{N}}$ is a Markov chain, without giving its transition matrix.

b) Compute G_n in terms of g, h, and G_{n-1}, then of h, g, n, and G_0.

c) If $x = \mathbb{E}(\xi_{n,i}) < \infty$ and $z = \mathbb{E}(\zeta_n) < \infty$, compute $\mathbb{E}(X_n)$ in terms of x, z, n, and $\mathbb{E}(X_0)$.

1.12 Single Server Queue Let $(A_n)_{n\geq 1}$ be i.i.d. r.v. with values in $\mathbb{N}$, with generating function $a(s) = \mathbb{E}(s^{A_1})$ and expectation $m = \mathbb{E}(A_1) < \infty$, and let X_0 be an independent r.v. with values in $\mathbb{N}$. Let

$$X_n = (X_{n-1} - 1)^+ + A_n \ , \qquad g_n(s) = \mathbb{E}(s^{X_n}).$$

a) Prove that $(X_n)_{n\geq 0}$ is a Markov chain with values in $\mathbb{N}$, which is irreducible if and only if $\mathbb{P}(A_1 = 0)\mathbb{P}(A_1 \geq 2) > 0$.

b) Compute g_n in terms of a and g_{n-1}.

c) It is now assumed that there exists an invariant law π for $(X_n)_{n\geq 0}$, with generating function denoted by g. Prove that

$$g(s)(s - a(s)) = \pi(0)(s - 1)a(s)$$

and that $\pi(0) = 1 - m$.

d) Prove that necessarily $m \leq 1$ and that $m = 1$ only in the trivial case where $\mathbb{P}(A_n = 1) = 1$.

e) Let $\mu = \sum_{x\in\mathbb{N}} \pi(x)x$. Prove that $\mu < \infty$ if and only if $\mathbb{E}(A_1^2) < \infty$, and then that for $\sigma^2 = Var(A_1)$, it holds that $\mu = \frac{1}{2}\left(m + \frac{\sigma^2}{1-m}\right)$.

1.13 Dobrushin mixing coefficient Let P be a transition matrix on $\mathcal{V}$, and

$$\rho_n = \frac{1}{2} \sup_{x,y\in\mathcal{V}} \|P^n(x,\cdot) - P^n(y,\cdot)\|_{var} \ , \qquad n \in \mathbb{N}.$$

a) Prove that $\rho_n \leq 1$ and that, for all laws μ and μ,

$$\|\mu P^n - \mu P^n\|_{var} \leq 2\rho_n.$$

b) Prove that

$$\rho_{n+m} \leq \rho_n \rho_m \ , \ \ m, n \geq 0 \ , \qquad \rho_n \leq \rho_k^{\lfloor n/k \rfloor} \ , \ \ k \geq 1.$$

One may use that $\inf_{c\in\mathbb{R}} \sup_{x\in M} |g(x) - c| \leq \frac{1}{2}\sup_{x,y\in M} |g(x) - g(y)|$.

c) Prove that if $k \geq 1$ is such that $\rho_k < 1$, then $(\mu P^n)_{n\in\mathbb{N}}$ is a Cauchy sequence, its limit is an invariant law π, and

$$\|\mu P^n - \pi\|_{var} \leq 2\rho_n \leq 2\rho_k^{\lfloor n/k \rfloor}.$$

d) Assume that P satisfies the Doeblin condition: there exists $k \geq 1$ and $\varepsilon > 0$ and a law $\hat{\pi}$ such that $P^k(x, \cdot) \geq \varepsilon\hat{\pi}$. Prove that $\rho_k \leq 1 - \varepsilon$. Compare with the result in Theorem 1.3.4.

e) Let $(F_i)_{i\geq 1}$ be a sequence of i.i.d. random functions from $\mathcal{V}$ to $\mathcal{V}$, and $X_0^x = x \in \mathcal{V}$ and $X_{n+1}^x = F_{n+1}(X_n^x)$ for $n \geq 0$, so that P is the transition matrix of the Markov chain induced by this random recursion, see Theorem 1.2.3 and what follows. Let

$$T_{x,y} = \inf\{n \geq 0 : \overset{x}{\underset{n}{\mathrm{X}}} = X_n^y\}.$$

Prove that

$$\rho_k \leq \sup_{x,y\in\mathcal{V}} \mathbb{P}(T_{x,y} > k).$$