UNIVERSITÉ DE RENNES 1

Master STS mention mathématiques

Markov chains on measurable spaces

Lecture notes

Dimitri Petritis

Rennes
Preliminary draft of April 2012

UFR Mathématiques

Contents

1

Introduction

1.1 Motivating example

The simplest non-trivial example of a Markov chain is the following model. Let $\mathbb{X} = \{0,1\}$ (heads 1, tails 0) and consider two coins, one honest and one biased giving tails with probability $2/3$. Let

$$X_{n+1} = \begin{cases} \text{outcome of biased coin if} & X_n = 0 \\ \text{outcome of honest coin if} & X_n = 1. \end{cases}$$

For $y \in \mathbb{X}$ and $\pi_n(y) = \mathbb{P}(X_n = y)$, we have

$$\pi_{n+1}(y) = \mathbb{P}(X_{n+1} = y|\text{honest})\mathbb{P}(\text{honest}) + \mathbb{P}(X_{n+1} = y|\text{biased})\mathbb{P}(\text{biased}),$$

yielding

$$(\pi_{n+1}(0), \pi_{n+1}(1)) = (\pi_n(0), \pi_n(1)) \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix},$$

with $P_{00} = 2/3$, $P_{01} = 1/3$, and $P_{10} = P_{11} = 1/2$.

Iterating, we get

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 P^n$$

where $\boldsymbol{\pi}_0$ must be determined ad hoc (for instance by $\pi_0(x) = \delta_{0x}$ for $x \in \mathbb{X}$).

In this formulation, the probabilistic problem is translated into a purely algebraic problem of determining the asymptotic behaviour of $\boldsymbol{\pi}_n$ in terms of the asymptotic behaviour of P^n, where $P = (P_{xy})_{x,y \in \mathbb{X}}$ is a **stochastic matrix** (i.e. $P_{xy} \geq 0$ for all x and y and $\sum_{y \in \mathbb{X}} P_{xy} = 1$). Although the solution of this problem is an elementary exercise in linear algebra, it is instructive to give some details. Suppose we are in the non-degenerate case where $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ with $0 < a, b < 1$. Compute the spectrum of P and left and right eigenvectors:

$$\begin{aligned} \mathbf{u}^t P &= \lambda \mathbf{u}^t \\ P\mathbf{v} &= \lambda \mathbf{v}. \end{aligned}$$

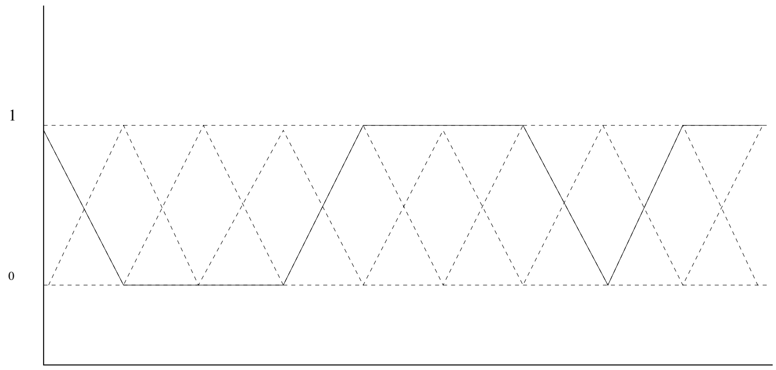

Figure 1.1: Dotted lines represent all possible and imaginable trajectories of heads and tails trials. Full line represents a particular realisation of such a sequence of trials. Note that only integer points have a significance, they are joined by line segments only for visualisation purposes.

Since the matrix P is not normal in general, the left (resp. right) eigenvectors corresponding to different eigenvalues are not orthogonal. However we can always normalise $\mathbf{u}_\lambda^t \mathbf{v}_{\lambda'} = \delta_{\lambda\lambda'}$. Under this normalisation, and denoting by $E_\lambda = \mathbf{v}_\lambda \otimes \mathbf{u}_\lambda^t$, we get

λ	$\mathbf{u}_\lambda$	$\mathbf{v}_\lambda$	E_λ
$\lambda_1 = 1$	$\begin{pmatrix} b \\ a \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\frac{1}{a+b}\begin{pmatrix} b & a \\ b & a \end{pmatrix}$
$\lambda_2 = 1-(a+b)$	$\begin{pmatrix} 1 \\ -1 \end{pmatrix}$	$\begin{pmatrix} a \\ -b \end{pmatrix}$	$\frac{1}{a+b}\begin{pmatrix} a & -a \\ -b & b \end{pmatrix}$

We observe that for λ and λ' being eigenvalues, $E_\lambda E_{\lambda'} = \delta_{\lambda\lambda'} E_\lambda$, i.e. E_λ are spectral projectors. The matrix P admits the **spectral decomposition**

$$P = \sum_{\lambda \in \operatorname{spec} P} \lambda E_\lambda,$$

where $\operatorname{spec} P = \{1, 1-(a+b)\} = \{\lambda \in \mathbb{C} : \lambda I - P \text{ is not invertible}\}$. Now E_λ being orthogonal projectors, we compute immediately, for all positive integers n,

$$P^n = \sum_{\lambda \in \operatorname{spec} P} \lambda^n E_\lambda = \lambda_1^n E_{\lambda_1} + \lambda_2^n E_{\lambda_2}$$

and since $\lambda_1 = 1$ while $|\lambda_2| < 1$, we get that $\lim_{n\to\infty} P^n = E_{\lambda_1}$.

Applying this formula for the numerical values of the example, we get that $\lim_{n\to\infty} \boldsymbol{\pi}_n = (0.6, 0.4)$.

1.2 Observations and questions

- In the previous example, a prominent role is played by the stochastic matrix P. In the general case, P is replaced by an operator, known as **stochastic kernel** that will be introduced and studied in chapter 2.
- An important question, that has been eluded in this elementary introduction, is whether exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ sufficiently large as to carry the whole sequence of random variables $(X_n)_{n\in\mathbb{N}}$. This question will be positively answered by the Ionescu Tulceă theorem and the construction of the **trajectory space** of the process in chapter 3.
- In the finite case, motivated in this chapter and fully studied in chapter 4, the asymptotic behaviour of the sequence (X_n) is fully determined by the spectrum of P, a purely algebraic object. In the general case (countable or uncountable), the spectrum is an algebraic and topological object defined as
$$\operatorname{spec} P = \overline{\{\lambda \in \mathbb{C} : \lambda I - P \text{ is not invertible}\}}.$$
  This approach exploits the topological structure of measurable functions on $\mathbb{X}$; a thorough study of the general case is performed, following the line of spectral theory of quasi-compact operators, in [14]. Although we don't follow this approach here, we refer the reader to the previously mentioned monograph for further study.
- Without using results from the theory of quasi-compact operators, we can study the formal power series $(I-P)^{-1} = \sum_{n=0}^{\infty} P^n$ (compare with the above definition of the spectrum). This series can be given both a probabilistic and an analytic significance realising a profound connection between harmonic analysis, martingale theory, and Markov chains. It is this latter approach that will be developed in chapter 5.
- Quantum Markov chains are objects defined on a quantum probability space. Now, quantum probability can be thought as a non-commutative extension of classical probability where real random variables are replaced by self-adjoint operators acting on a Hilbert space. Thus, once again the spectrum plays an important rôle in this context.

Exercise 1.2.1 ***(Part of the paper of December 2006)*** *Let (X_n) be the Markov chain defined in the example of section 1.1, with $\mathbb{X} = \{0,1\}$. Let $\eta_n^0(A) = \sum_{k=0}^{n-1} \mathbb{1}_A(X_k)$ for $A \subseteq \mathbb{X}$.*

1. *Compute $\mathbb{E}\exp(i\sum_{x\in\mathbb{X}} t_x \eta_n^0(\{x\}))$ as a product of matrices and vectors, for $t_x \in \mathbb{R}$.*
2. *What is the probabilistic significance of the previous expectation?*
3. *Establish a central limit theorem for $\eta_n^0(\{0\})$.*
4. *Let $f : \mathbb{X} \to \mathbb{R}$. Express $S_n(f) = \sum_{k=0}^{n-1} f(X_k)$ with the help of $\eta_n^0(A)$, with $A \subseteq \mathbb{X}$.*

2

Kernels

This chapter deals mainly with the general theory of kernels. The most complete reference for this chapter is the book [28]. Several useful results can be found in [24].

2.1 Notation

In the sequel, $(\mathbb{X},\mathscr{X})$ will be an arbitrary measurable space; most often the σ-algebra will be considered **separable** or **countably generated** (i.e. we assume that there exists a sequence $(F_n)_{n\in\mathbb{N}}$ of measurable sets $F_n \in \mathscr{X}$, for all n, such that $\sigma(F_n, n \in \mathbb{N}) = \mathscr{X}$). A closely related notion for a measure space $(\mathbb{X},\mathscr{X},\mu)$ is the μ-separability of the σ-algebra $\mathscr{X}$ meaning that there exists a separable σ-subalgebra $\mathscr{X}_0$ of $\mathscr{X}$ such that for every $A \in \mathscr{X}$ there exists a $A' \in \mathscr{X}_0$ verifying $\mu(A\triangle A') = 0$. We shall denote by $m\mathscr{X}$ the set of real (resp. complex) $(\mathscr{X},\mathscr{B}(\mathbb{R}))$- (resp. $(\mathscr{X},\mathscr{B}(\mathbb{C}))$)-measurable functions defined on $\mathbb{X}$. For $f \in m\mathscr{X}$, we introduce the norm $\|f\|_\infty = \sup_{x\in\mathbb{X}} |f(x)|$. Similarly, $b\mathscr{X}$ will denote the set of bounded measurable functions and $m\mathscr{X}_+$ the set of positive measurable functions.

Exercise 2.1.1 The set $(b\mathscr{X}, \|\cdot\|_\infty)$ is a Banach space.

Similar definitions apply by duality to the set of measures. Recall that a measure is a σ-additive function defined on $\mathscr{X}$ and taking values in $]-\infty,\infty]$. It is called positive if its set of values is $[0,\infty]$ and bounded if $\sup\{|\mu(F)|, F \in \mathscr{X}\} < \infty$. For every measure μ and every measurable set F, we can define positive and negative parts by

$$\mu^+(F) = \sup\{\mu(G), G \in \mathscr{X}, G \subseteq F\}$$

and

$$\mu^-(F) = \sup\{-\mu(G), G \in \mathscr{X}, G \subseteq F\}.$$

Note that $\mu = \mu^+ - \mu^-$. The **total variation** of μ is the measure $|\mu| = \mu^+ + \mu^-$. The total variation is bounded iff $|\mu|$ is bounded; we denote by $\|\mu\|_1 = |\mu|(\mathbb{X})$.

We denote by $\mathscr{M}(\mathscr{X})$ the set of σ-**finite** measures on $\mathscr{X}$ (i.e. measures μ for which there is an *increasing* sequence $(F_n)_{n\in\mathbb{N}}$ of measurable sets F_n such that $|\mu|(F_n) < \infty$ for all n and $\mathbb{X} = \cup^{\uparrow} F_n$.) Similarly, we introduce the set of bounded measures $b\mathscr{M}(\mathscr{X}) = \{\mu \in \mathscr{M}(\mathscr{X}) : \sup\{|\mu(A)|, A \in \mathscr{X}\} < \infty\}$. The sets $\mathscr{M}_+(\mathscr{X})$ and $b\mathscr{M}_+(\mathscr{X})$ are defined analogously for positive measures. The set $\mathscr{M}_1(\mathscr{X})$ denotes the set of probability measures on $\mathscr{X}$.

Exercise 2.1.2 The set $(b\mathscr{M}(\mathscr{X}), \|\cdot\|_1)$ is a Banach space.

Let $\mu \in \mathscr{M}(\mathscr{X})$ and $f \in \mathscr{L}^1(\mathbb{X}, \mathscr{X}, \mu)$. We denote in-distinctively, $\int_{\mathbb{X}} f d\mu = \int_{\mathbb{X}} f(x)\mu(dx) = \int_{\mathbb{X}} \mu(dx) f(x) = \mu(f) = \langle \mu, f \rangle$. As usual the space $L^1(\mathbb{X}, \mathscr{X}, \mu)$ is the quotient of $\mathscr{L}^1(\mathbb{X}, \mathscr{X}, \mu)$ by equality holding μ-almost everywhere. There exists a canonical isometry between $L^1(\mathbb{X}, \mathscr{X}, \mu)$ and $\mathscr{M}(\mathscr{X})$ defined by the two mappings:

$$
\begin{aligned}
f \mapsto \mu_f \quad &: \quad \mu_f(A) = \int_A f(x)\mu(dx), \text{ for } f \in L^1 \\
\nu \mapsto f_\nu \quad &: \quad f_\nu(x) = \frac{d\nu}{d\mu}(x), \text{ for } \nu \ll \mu.
\end{aligned}
$$

2.2 Transition kernels

Definition 2.2.1 Let $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{X}', \mathscr{X}')$ be two measurable spaces. A mapping $N : \mathbb{X} \times \mathscr{X}' \to]-\infty, +\infty]$ such that
- $\forall x \in \mathbb{X}$, $N(x, \cdot)$ is a measure on $\mathscr{X}'$ and
- $\forall A \in \mathscr{X}'$, $N(\cdot, A)$ is a $\mathscr{X}$-measurable function,

is called a **transition kernel** between $\mathbb{X}$ and $\mathbb{X}'$. We denote $(\mathbb{X}, \mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}', \mathscr{X}')$.

The kernel is termed
- **positive** if its image is $[0, +\infty]$,
- σ-**finite** if for all $x \in \mathbb{X}$, $N(x, \cdot) \in \mathscr{M}(\mathscr{X}')$,
- **proper** if there exists an increasing exhausting sequence $(A_n)_n$ of $\mathscr{X}'$-measurable sets such that, for all $n \in \mathbb{N}$, the function $N(\cdot, A_n)$ is bounded on $\mathbb{X}$, and
- **bounded** if its image is bounded.

Exercise 2.2.2 For a kernel, bounded implies proper implies σ-finite. A positive kernel is bounded if and only if the function $N(\cdot, \mathbb{X})$ is bounded.

Henceforth, we shall consider positive kernels (even when Markovian this qualifying adjective is omitted).

Definition 2.2.3 Let N be a positive transition kernel $(\mathbb{X},\mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}',\mathscr{X}')$. For $f \in m\mathscr{X}'_+$, we define a function on $m\mathscr{X}_+$, denoted by Nf, by the formula:

$$\forall x \in \mathbb{X},\ Nf(x) = \int_{\mathbb{X}'} N(x,dx')f(x') = \langle N(x,\cdot), f\rangle.$$

Remark: $f \in m\mathscr{X}_+$ is not necessarily integrable with respect to the measure $N(x,\cdot)$. The function Nf is defined with values in $[0,+\infty]$ by approximating by step functions. The definition can be extended to $f \in m\mathscr{X}'$ by defining $Nf = Nf^+ - Nf^-$ provided that the functions Nf^+ and Nf^- do not take simultaneously infinite values.

Remark: Note that the transition kernel $(\mathbb{X},\mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}',\mathscr{X}')$ transforms functions **contravariantly**. We have the expression $N(x,A) = N\mathbb{1}_A(x)$ for all $x \in \mathbb{X}$ and all $A \in \mathscr{X}'$.

Definition 2.2.4 Let $(\mathbb{X},\mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}',\mathscr{X}')$ be a positive kernel. For $\mu \in \mathscr{M}_+(\mathscr{X})$, we define a measure of $\mathscr{M}_+(\mathscr{X}')$, denoted by μN, by the formula:

$$\forall A \in \mathscr{X}',\ \mu N(A) = \int_{\mathbb{X}} \mu(dx)N(x,A) = \langle \mu, N(\cdot,A)\rangle.$$

Remark: The definition can be extended to $\mathscr{M}(\mathscr{X})$ by $\mu N = \mu^+ N - \mu^- N$.

Remark: Note that the transition kernel $(\mathbb{X},\mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}',\mathscr{X}')$ acts **covariantly** on measures. We have the expression $N(x,A) = \epsilon_x N(A)$ where ϵ_x is the Dirac mass on x.

2.3 Examples-exercises

2.3.1 Integral kernels

Let $\lambda \in \mathscr{M}_+(\mathscr{X}')$ and $n : \mathbb{X} \times \mathbb{X}' \to \mathbb{R}_+$ be a $\mathscr{X} \otimes \mathscr{X}'$-measurable function. Define for all $x \in \mathbb{X}$ and all $A \in \mathscr{X}'$

$$N(x,A) = \int_{\mathbb{X}} n(x,x')\mathbb{1}_A(x')\lambda(dx').$$

Then N is a positive transition kernel $(\mathbb{X},\mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}',\mathscr{X}')$, termed **integral kernel**. For $f \in m\mathscr{X}'_+$, $\mu \in \mathscr{M}_+(c = \mathscr{X})$, $A \in \mathscr{X}'$ and $x \in \mathbb{X}$ we have:

$$\begin{aligned} Nf(x) &= \int_{\mathbb{X}} n(x,x')f(x')\lambda(dx') \\ \mu N(A) &= \int_{\mathbb{X}}\int_{\mathbb{X}'} \mu(dx)n(x,x')(\mathbb{1}_A\lambda)(dx'), \end{aligned}$$

where $\mathbb{1}_A\lambda$ is a measure absolutely continuous with respect to λ having Radon-Nikodým density $\mathbb{1}_A$.

Some remarkable particular integral kernels are given below:

1. Let $\mathbb{X} = \mathbb{X}'$ be finite or countable sets, $\mathscr{X} = \mathscr{X}' = \mathscr{P}(\mathbb{X})$, and λ be the counting measure, defined by $\lambda(x) = 1$ for all $x \in \mathbb{X}$. Then the integral kernel N is defined, for $f \in m\mathscr{X}$, $\mu \in \mathscr{M}(\mathscr{X})$, by

$$\begin{aligned} Nf(x) &= \sum_{y\in\mathbb{X}} n(x,y)f(y), \text{ for all } x \in \mathbb{X} \\ \mu N(A) &= \sum_{x\in\mathbb{X};y\in A} \mu(x)n(x,y), \text{ for all } A \in \mathscr{X}. \end{aligned}$$

   In this discrete case, $N(x,y) \equiv N(x,y) \equiv n(x,y)$ are the elements of finite or infinite matrix whose columns are functions and rows are measures. If we impose further $N(x,\cdot)$ to be a probability, then $\sum_{y\in\mathbb{X}} n(x,y) = 1$ as was the case in the motivating example of chapter 3.1. In that case, the operator $N = (n(x,y))_{x,y\in\mathbb{X}}$ is a matrix whose rows sum up to one, termed **stochastic matrix** .

2. Let $\mathbb{X} = \mathbb{X}' = \mathbb{R}^d$ with $d \geq 3$, $\mathscr{X} = \mathscr{X}' = \mathscr{B}(\mathbb{R}^d)$, and λ be the Lebesgue measure in dimension d. The function given by the formula $n(x,x') = \|x-x'\|^{-d+2}$ for $x \neq x'$ defines the so-called Newtonian kernel. It allows expressing the electrostatic or gravitational potential U at x due to a charge density ρ via the integral formula:

$$U(x) = c\int_{\mathbb{R}^d} n(x,x')\rho(x')\lambda(dx').$$

   The function U is solution of the Poisson differenital equation $\frac{1}{2}\Delta U(x) = -\rho(x)$. The Newtonian kernel can be thought as the “inverse” operator of the Laplacian. We shall return to this “inverting” procedure in chapter 5.

3. Let $\mathbb{X} = \mathbb{X}'$, $\mathscr{X} = \mathscr{X}'$, and $\lambda \in \mathscr{M}(\mathscr{X})$ be arbitrary. Let $a, b \in b\mathscr{X}$ and define $n(x,x') = a(x)b(x')$. Then the integral kernel N is defined by

$$\begin{aligned} Nf(x) &= a(x)\langle b\lambda, f\rangle \\ \mu N(A) &= \langle \mu, a\rangle (b\lambda)(A). \end{aligned}$$

   In other words the kernel $N = a \otimes b\lambda$ is of the form function⊗measure (to be compared with the form [1] $E_\lambda = \mathbf{v}_\lambda \otimes \mathbf{u}^t_\lambda$ introduced in chapter 1).

1. Recall that functions are identified with vectors, measures with linear forms.

2.3.2 Convolution kernels

Let $\mathbb{G}$ be a topological semigroup (i.e. composition is associative and continuous) whose composition $\cdot$ is denoted multiplicatively. Assume that $\mathbb{G}$ is metrisable (i.e. there is a distance on $\mathbb{G}$ generating its topology) and locally compact (i.e. every point $x \in \mathbb{G}$ is contained in an open set whose closure is compact). Denote by $\mathscr{G}$ the Borel σ-algebra on $\mathbb{G}$. Define the convolution of two measures μ and ν by duality on continuous functions of compact support $f \in C_K(\mathbb{G})$ by $\langle \mu \star \nu, f\rangle = \int_{\mathbb{G}}\int_{\mathbb{G}} \mu(dx)\nu(dy) f(x\cdot y)$.

Let now $\mathbb{X} = \mathbb{X}' = \mathbb{G}$, $\mathscr{X} = \mathscr{X}' = \mathscr{G}$, and λ be a fixed positive Radon measure (i.e. $\lambda(K) < \infty$ for all compact measurable sets K) and define N by the formula $N(x, A) = (\lambda \star \epsilon_x)(A)$. Then N is a transtion kernel $(\mathbb{G}, \mathscr{G}) \overset{N}{\rightsquigarrow} (\mathbb{G}, \mathscr{G})$, called **convolution kernel**, whose action is given by

$$\begin{aligned} Nf(x) &= \langle \lambda \star \epsilon_x, f\rangle \\ &= \int_{\mathbb{G}}\int_{\mathbb{G}} \lambda(dy)\epsilon_x(dz) f(y\cdot z) \\ &= \int_{\mathbb{G}} \lambda(dy) f(y\cdot x). \end{aligned}$$

In particular, $N(x, A) = \langle \lambda \star \epsilon_x, \mathbb{1}_A\rangle = \int_{\mathbb{G}} \lambda(dy)\mathbb{1}_A(y\cdot x)$ yielding for the left action on measures $\mu N(A) = \langle \lambda \star \mu, \mathbb{1}_A\rangle$.

Suppose now that $\mathbb{G}$ instead of being merely a semigroup, is henceforth a group. Then the above formula becomes $N(x, A) = \lambda(Ax^{-1})$ and consequently

$$\langle \mu N, f\rangle = \langle \lambda \star \mu, f\rangle,$$

yielding[2] $\mu N = \lambda \star \mu$. This provides us with another reminding that N transforms differently arguments of functions and measures.

Definition 2.3.1 Let $\mathbb{V}$ be a topological space and $\mathscr{V}$ it Borel σ-algebra; the group $\mathbb{G}$ **operates** on $\mathbb{V}$ (we say that $\mathbb{V}$ is a $\mathbb{G}$-space), if there exists a continuous map $\alpha : \mathbb{G} \times \mathbb{V} \ni (x, v) \mapsto \alpha(x, v) \equiv x \circ v \in \mathbb{V}$, verifying $(x_1 \cdot x_2) \circ v = x_1 \circ (x_2 \circ v)$.

With any fixed measure $\lambda \in \mathscr{M}_+(\mathscr{G})$ we associate a kernel $(\mathbb{V}, \mathscr{V}) \overset{K}{\rightsquigarrow} (\mathbb{V}, \mathscr{V})$ by its action on functions $f \in m\mathscr{V}_+$ and measures $\mu \in \mathscr{M}(\mathscr{V})$:

$$\begin{aligned} Kf(v) &= \int_{\mathbb{G}} \lambda(dx) f(x\circ v), \text{ for any } v \in \mathbb{V} \\ \mu K(A) &= \int_{\mathbb{G}} \lambda(dx) \int_{\mathbb{V}} \mu(dv)\mathbb{1}_A(x\circ v), \text{ for any } A \in \mathscr{V}. \end{aligned}$$

2. Note however, that $Nf \neq \lambda \star f$ because the convolution $\lambda \star f(x) = \int_{\mathbb{G}} \lambda(dy) f(y^{-1}x)$. Hence on defining λ_* as the image of λ under the transformation $x \mapsto x^{-1}$, we have $Nf(x) = \int_{\mathbb{G}} \lambda(dy) f(y^{-1}x) = \lambda_* \star f(x)$.

When $\mathbb{V} = \mathbb{G}$ in the above formulae and identifying the operation $\circ$ with the group composition $\cdot$, we say that the group operates (from the left) on itself. In that case, we recover the convolution kernel of the beginning of this section as a particular case of the action on $\mathbb{G}$-spaces.

Some remarkable particular convolution kernels are given below:

1. Let $\mathbb{G}$ be the Abelian group $\mathbb{Z}$, whose composition is denoted additively. Let λ be the probability measure charging solely the singletons -1 and 1 with probability $1/2$. Then the convolution kernel N acts on measurable functions and measures by:

$$\begin{aligned} Nf(x) &= \frac{1}{2}\big(f(x+1)+f(x-1)\big) \\ \mu N(A) &= \frac{1}{2}\big(\mu(A-1)+\mu(A+1)\big). \end{aligned}$$

   This kernel corresponds to the transition matrix

$$n(x,y) = \begin{cases} 1/2 & \text{if } y-x=\pm 1 \\ 0 & \text{otherwise,} \end{cases}$$

   corresponding to the **simple symmetric random walk** on $\mathbb{Z}$.
   More generally, if $\mathbb{G} = \mathbb{Z}^d$ and λ is a probability measure whose support is a a generating set of the group, the transition kernel defines a random walk on $\mathbb{Z}^d$.

2. The same holds more generally for non Abelian groups. For example, let $G = \{a, b\}$ be a set of free generators and $\mathbb{G}$ the group generated by G via successive applications of the group operations. Then $\mathbb{G}$ is the so-called **free group** on two generators, denoted by $\mathbb{F}_2$. It is isomorphic with the homogeneous tree of degree 4. If λ is a probability supported by G and G^{-1}, then the convolution kernel corresponds to a random walk on $\mathbb{F}_2$.

3. Let $\mathbb{G}$ be the semigroup of $d \times d$ matrices with strictly positive elements and $\mathbb{V} = \mathbb{R}P^{d-1}$ be the real projective space[3]. If $\mathbb{R}P^{d-1}$ is thought as embedded in $\mathbb{R}^d$, the latter being equipped with a norm $\|\cdot\|$, define the action of a matrix $x \in \mathbb{G}$ on a vector $v \in \mathbb{R}P^{d-1}$ by $x \circ v = \frac{xv}{\|xv\|}$. Let λ be a probability measure on $\mathbb{G}$. Then the kernel K defined by $Kf(v) = \int_{\mathbb{G}} \lambda(dx) f(x \circ v)$ induces a random walk on the space $\mathbb{V}$.

3. The real projective space $\mathbb{R}P^n$ is defined as the space of all lines through the origin in $\mathbb{R}^{n+1}$. Each such line is determined as a nonzero vector of $\mathbb{R}^{n+1}$, unique up to scalar multiplication, and $\mathbb{R}P^n$ is topologised as the quotient space of $\mathbb{R}^{n+1} \setminus \{0\}$ under the equivalence relation $v \simeq \alpha v$ for scalars $\alpha \neq 0$. We can restrict ourselves to vectors of length 1, so that $\mathbb{R}P^n$ is also the quotient space $\mathbb{S}^n / v \simeq -v$ of the sphere with antipodal points identified.

2.3.3 Point transformation kernels

Let $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$ be two measurable spaces and $\theta : \mathbb{X} \to \mathbb{Y}$ a measurable function. Then, the function N defined on $\mathbb{X} \times \mathscr{Y}$ by $N(x, A) \equiv \mathbb{1}_A(\theta(x)) = \mathbb{1}_{\theta^{-1}(A)}(x) = \epsilon_{\theta(x)}(A)$ for $x \in \mathbb{X}$ and $A \in \mathscr{Y}$ defines a transition kernel $(\mathbb{X}, \mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{Y}, \mathscr{Y})$, termed **point transformation kernel** or **deterministic kernel**. Its action on functions and measures reads:

$$\begin{aligned}
Nf(x) &= \int_{\mathbb{Y}} N(x, dy) f(y) \\
&= \int_{\mathbb{Y}} \epsilon_{\theta(x)}(dy) f(y) \\
&= f(\theta(x)) = f \circ \theta(x) \\
\mu N(A) &= \int_{\mathbb{X}} \mu(dx) \mathbb{1}_{\theta^{-1}(A)}(x) \\
&= \mu(\theta^{-1}(A)) = \theta(\mu)(A).
\end{aligned}$$

Example 2.3.2 Let $\mathbb{X} = \mathbb{Y} = [0, 1]$, $\mathscr{X} = \mathscr{Y} = \mathscr{B}([0, 1])$ and θ the mapping given by the formula $\theta(x) = 4x(1 - x)$. The iterates of θ applied on a given point $x \in [0, 1]$ describe the trajectory of x under the **dynamical system** defined by the map. The corresponding kernel is deterministic.

Exercise 2.3.3 Determine the explicit form of the kernel in the previous example.

2.4 Markovian kernels

Definition 2.4.1 (Kernel composition) Let M and N be positive transition kernels $(\mathbb{X}, \mathscr{X}) \overset{M}{\rightsquigarrow} (\mathbb{X}', \mathscr{X}')$ and $(\mathbb{X}', \mathscr{X}') \overset{N}{\rightsquigarrow} (\mathbb{X}'', \mathscr{X}'')$. Then by MN we shall denote the positive kernel $(\mathbb{X}, \mathscr{X}) \overset{MN}{\rightsquigarrow} (\mathbb{X}'', \mathscr{X}'')$ defined, for all $x \in \mathbb{X}$ and all $A \in \mathscr{X}''$ by

$$MN(x, A) = \int_{\mathbb{X}'} M(x, dy) N(y, A).$$

Exercise 2.4.2 The composition of positive kernels is associative.

Remark: If N is a positive kernel $(\mathbb{X}, \mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X}, \mathscr{X})$ then N^n is defined for all $n \in \mathbb{N}$ by $N^0 = I$, where I is the identity kernel $(\mathbb{X}, \mathscr{X}) \overset{I}{\rightsquigarrow} (\mathbb{X}, \mathscr{X})$ defined by $I(x, A) = \epsilon_x(A)$ and recursively for $n > 0$ by $N^n = NN^{n-1}$.

Definition 2.4.3 (Kernel ordering) Let M and N be two positive kernels $(\mathbb{X},\mathscr{X}) \overset{M}{\rightsquigarrow} (\mathbb{Y},\mathscr{Y})$ and $(\mathbb{X},\mathscr{X}) \overset{M}{\rightsquigarrow} (\mathbb{Y},\mathscr{Y})$. We say $M \le N$ if $\forall f \in \mathscr{Y}_+$, the inequality $Mf \le Nf$ holds (point-wise).

Definition 2.4.4 (Markovian kernels) Let N be a positive kernel $(\mathbb{X},\mathscr{X}) \overset{N}{\rightsquigarrow} (\mathbb{X},\mathscr{X})$.

– The kernel N is termed **sub-Markovian** or **a transition probability** if $N(x,A) \le 1$ holds for all $x \in \mathbb{X}$ and all $A \in \mathscr{X}$.
– The kernel N is termed **Markovian** if $N(x,\mathbb{X}) = 1$ holds for all $x \in \mathbb{X}$.

Remark: It may sound awkward to term a sub-Markovian kernel, for which the strict inequality $N(x,\mathbb{X}) < 1$ may hold, transition probability. The reason for this terminology is that it is always possible to extend the space $\mathbb{X}$ to $\hat{\mathbb{X}} = \mathbb{X} \cup \{\partial\}$ by adjoining a particular point $\partial \notin \mathbb{X}$, called **cemetery**. The σ-algebra is extended analogously to $\hat{\mathscr{X}} = \sigma(\mathscr{X},\{\partial\})$. The kernel is subsequently extended to a Markovian kernel $\hat{N}$ defined by $\hat{N}(x,A) = N(x,A)$ for all $A \in \mathscr{X}$ and all $x \in \mathbb{X}$, assigning the missing mass to the cemetery by $\hat{N}(x,\{\partial\}) = 1 - N(x,\mathbb{X})$ for all $x \in \mathbb{X}$, and trivialising the additional point by imposing $\hat{N}(\partial,\{\partial\}) = 1$. It is trivially verified that $\hat{N}$ is now Markovian. Moreover, if N is already Markovian, obviously $\hat{N}(x,\{\partial\}) = 0$ for all $x \in \mathbb{X}$. Functions $f \in m\mathscr{X}$ are also extended to $\hat{f} \in m\hat{\mathscr{X}}$ coinciding with f on $\mathbb{X}$ and verifying $\hat{f}(\{\partial\}) = 0$.

We use henceforth the **reserved symbol** P to denote positive kernels that are transition probabilities. Without loss of generality, we can always assume that P are Markovian kernels, possibly at the expense of adjoining a cemetery point to the space $(\mathbb{X},\mathscr{X})$ according the previous remark, although the ˆ symbol will be omitted.

Exercise 2.4.5 Markovian kernels, as linear operators acting to the right on the Banach space $(b\mathscr{X}, \|\cdot\|_\infty)$ or to the left on $(b\mathscr{M}(\mathscr{X}), \|\cdot\|_1)$, are positive contractions.

2.5 Further exercises

1. Let $(\mathbb{X},\mathscr{X})$ be a measurable space, $g \in m\mathscr{X}_+$, $\nu \in \mathscr{M}_+(\mathscr{X})$, and $(\mathbb{X},\mathscr{X}) \overset{M}{\rightsquigarrow} (\mathbb{X},\mathscr{X})$ an arbitrary positive kernel. Define $N = g \otimes \nu$ and compute NM and MN.
2. Let P be a Markovian kernel and define $G^0 = \sum_{n=0}^{\infty} P^n$.
   – Show that G^0 is a positive kernel not necessarily finite on compact sets.
   – Show that $G^0 = I + PG^0 = I + G^0 P$.

3. Let P be a Markovian kernel and a, b arbitrary sequences of positive numbers such that $\sum_{n=0}^{\infty} a_n \leq 1$ and $\sum_{n=0}^{\infty} b_n \leq 1$. Define $G^0_{(a)} = \sum_{n=0}^{\infty} a_n P^n$ and compute $G^0_{(a)} G^0_{(b)}(x, A)$.
4. Let P be a Markovian kernel. For every $z \in \mathbb{C}$ define $G^0_z = \sum_{n=0}^{\infty} z^n P^n$.
   - Show that G^0_z is a finite kernel for all complex z with $|z| < 1$.
   - Is the linear operator $zI - P$ invertible for z in some subset of the complex plane?

3

Trajectory spaces

In this chapter a minimal construction of a general probability space is performed on which lives a general Markov chain.

3.1 Motivation

Let $(\mathbb{X}, \mathscr{X})$ be a measurable space. In the well known framework of the Kolmogorov axiomatisation of probability theory, to deal with a $\mathbb{X}$-valued random variable X we need an *ad hoc* probability space $(\Omega, \mathscr{F}, \mathbb{P})$ on which the the random variable is defined as a $(\mathscr{F}, \mathscr{X})$-measurable function. The law of X is the probability measure $\mathbb{P}_X$ on $(\mathbb{X}, \mathscr{X})$, image of $\mathbb{P}$ under X, i.e. $\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}$ for all $A \in \mathscr{X}$.

What is much less often stressed in the various accounts of probability theory is the profound significance of this framework. As put by Kolmogorov himself on page 1 of [21]: "…the field of probabilities is defined as a system of [sub]sets [of a universal set] which satisfy certain conditions. What the elements of this [universal] set represent is of no importance … ".

Example 3.1.1 (Heads and tails) Let $\mathbb{X} = \{0, 1\}$ and $p \in [0, 1]$. Modelling the outcomes X of a coin giving tails with probability p is equivalent to specifying $\mathbb{P}_X = p\epsilon_0 + (1 - p)\epsilon_1$.

Several remarks are necessary:

1. All information experimentally pertinent to the above example is encoded into the probability space $(\mathbb{X}, \mathscr{X}, \mathbb{P}_X)$.
2. The choice of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, on which the random variable X="the coin outcome" is defined, is not unique. Every possible choice

corresponds to a **possible physical realisation** used to model the random experiment. This idea clearly appears in the classical text [21].

3. According to Kolmogorov, in any random experiment, the only fundamental object is $(\mathbb{X}, \mathscr{X}, \mathbb{P}_X)$; with every such probability space, we can associate infinitely many pairs composed of an auxiliary probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a measurable mapping $X : \Omega \to \mathbb{X}$ such that $\mathbb{P}_X$ is the image of $\mathbb{P}$ under X.
4. This picture was further completed later by Loomis [22] and Sikorski [**?**]: in the auxiliary space $(\Omega, \mathscr{F}, \mathbb{P})$, the only important object is $\mathscr{F}$ since for every abstract σ-algebra $\mathscr{F}$, there exists a universal set Ω such that $\mathscr{F}$ can be realised as a σ-algebra of subsets of Ω (Loomis-Sikorski theorem).

Example 3.1.2 (Heads and tails revisited by the layman) When one tosses a coin on a table (approximated as an infinitely extending plane and very plastic), the space $\Omega = (\mathbb{R}_+ \times \mathbb{R}^2) \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{S}^2$ is used. This space encodes position $\mathbf{R}$ of the centre of the mass, velocity $\mathbf{V}$, angular momentum $\mathbf{M}$ and orientation $\mathbf{N}$ of the normal to the head face of the coin. The σ-algebra $\mathscr{F} = \mathscr{B}(\Omega)$ and $\mathbb{P}$ corresponds to some probability of compact support (initial conditions of the mechanical system). Newton equations govern the time evolution of the system and due to the large plasticity of the table, the coin does not bounce when it touches the table. Introduce the random time $T(\omega) = \inf\{t > 0 : R_3(t) = 0; \mathbf{V}(t) = 0, \mathbf{M}(t) = 0\}$ and the random variable

$$X(\omega) = \begin{cases} 0 & \text{if} \quad \mathbf{N}(T(\omega)) \cdot \mathbf{e}_3 = -1 \\ 1 & \text{if} \quad \mathbf{N}(T(\omega)) \cdot \mathbf{e}_3 = 1, \end{cases}$$

where $\mathbf{e}_3$ is the unit vector parallel to the vertical axis. A tremendously complicated modelling indeed!

Example 3.1.3 (Heads and tails revisited by the Monte Carlo simulator) Let $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}([0,1])$ and $\mathbb{P}$ be the Lebesgue measure on $[0,1]$. Then the outcome of a honest coin is modelled by the random variable

$$X(\omega) = \begin{cases} 0 & \text{if} \quad \omega < 1/2 \\ 1 & \text{if} \quad \omega > 1/2. \end{cases}$$

A simpler modelling but still needing a Borel σ-algebra on an uncountable set!

Example 3.1.4 (Heads and tails revisited by the mathematician) Let $\Omega = \{0,1\}$, $\mathscr{F} = \mathscr{P}(\Omega)$ and $\mathbb{P} = \frac{1}{2}\epsilon_0 + \frac{1}{2}\epsilon_1$. Then the outcome of a honest coin is modelled by the random variable $X = \text{id}$. Note further that $\mathbb{P}_X = \mathbb{P}$. Such a realisation is called **minimal**.

Exercise 3.1.5 (Your turn to toss coins!) Construct a minimal probability space carrying the outcomes for two honest coins.

Let $(\mathbb{X}, \mathscr{X})$ be an arbitrary measurable space and $X = (X_n)_{n\in\mathbb{N}}$ a countable family of $\mathbb{X}$-valued random variables. Two natural questions arise:

1. What is the significance of $\mathbb{P}_X$?
2. Does there exist a minimal probability space $(\Omega, \mathscr{F}, \mathbb{P})$ carrying the whole family of random variables?

3.2 Construction of the trajectory space

3.2.1 Notation

Let $(\mathbb{X}_k, \mathscr{X}_k)_{k\in\mathbb{N}}$ be a family of measurable spaces. For $n \in \mathbb{N} \cup \{\infty\}$, denote $\mathbb{X}^n = \times_{k=0}^n \mathbb{X}_k$ and $\mathscr{X}^n = \otimes_{k=0}^n \mathscr{X}_k$. **Beware** of the difference between subscripts and superscripts in the previous notation!

Definition 3.2.1 Let $(\mathbb{X}_k, \mathscr{X}_k)_{k\in\mathbb{N}}$ be a family of measurable spaces. The set

$$\mathbb{X}^\infty = \{x = (x_0, x_1, x_2, \ldots) : x_k \in \mathbb{X}_k, \forall k \in \mathbb{N}\}$$

is called the **trajectory space** on the family $(\mathbb{X}_k)_{k\in\mathbb{N}}$.

Remark: The trajectory space can be thought as the subset of all infinite sequences $\{x : \mathbb{N} \to \cup_{k\in\mathbb{N}} \mathbb{X}_k\}$ that are **admissible**, in the sense that necessarily for all k, $x_k \in \mathbb{X}_k$. When all spaces of the family are equal, i.e. $\mathbb{X}_k = \mathbb{X}$ for all k, then the space of admissible sequences trivially reduces to the set of sequences $\mathbb{X}^\mathbb{N}$.

Definition 3.2.2 Let $0 \le m \le n$. Denote by $\mathfrak{p}_m^n : \mathbb{X}^n \to \mathbb{X}^m$ the **projection** defined by the formula:

$$\mathfrak{p}_m^n(x_0, \ldots, x_m, \ldots, x_n) = (x_0, \ldots, x_m).$$

We simplify notation to $\mathfrak{p}_m \equiv \mathfrak{p}_m^\infty$ for the projection form $\mathbb{X}^\infty$ to $\mathbb{X}^m$. More generally, let $\emptyset \neq S \subseteq \mathbb{N}$ be an arbitrary subset of the indexing set. Denote by $\varpi_S : \mathbb{X}^\infty \to \times_{k\in S} \mathbb{X}_k$ the projection defined for any $x \in \mathbb{X}^\infty$ by $\varpi_S(x) = (x_k)_{k\in S}$. If $S = \{n\}$, we write ϖ_n instead of $\varpi_{\{n\}}$.

Remark: Obvsiously, $\mathfrak{p}_m = \varpi_{\{0,\ldots,m\}}$.

Definition 3.2.3 Let $(\mathbb{X}_k, \mathscr{X}_k)_{k\in\mathbb{N}}$ be a family of measurable spaces.

1. We call **family of rectangular sets** over the trajectory space the collection

$$\begin{aligned}\mathscr{R} &= \{\times_{k\in\mathbb{N}} A_k : A_k \in \mathscr{X}_k, \forall k \in \mathbb{N} \text{ and } \sharp\{k \in \mathbb{N} : A_k \neq \mathbb{X}_k\} < \infty \\ &= \{\times_{k\in\mathbb{N}} A_k : A_k \in \mathscr{X}_k, \forall k \in \mathbb{N} \text{ and } \exists N : k \geq N \Rightarrow A_k = \mathbb{X}_k\} \\ &= \cup_{N\in\mathbb{N}} \mathscr{R}_N,\end{aligned}$$

   where $\mathscr{R}_N = \times_{k=0}^{N} \mathscr{X}_k$.
2. We call **family of cylinder sets** over the trajectory space the collection

$$\mathscr{C} = \cup_{N\in\mathbb{N}} \{F \times (\times_{k>N} \mathbb{X}_k) : F \in \mathscr{X}^N\}.$$

The cylinder set $F \times (\times_{k>N} \mathbb{X}_k)$ will be occasionally denoted $[F]_N$.

Definition 3.2.4 The σ-algebra generated by the sequence of projections $(\varpi_k)_{k\in\mathbb{N}}$ is denoted by $\mathscr{X}^\infty$; i.e. $\mathscr{X}^\infty = \otimes_{k\in\mathbb{N}} \mathscr{X}_k = \sigma(\cup_{k\in\mathbb{N}} \varpi_k^{-1}(\mathscr{X}_k))$.

Exercise 3.2.5 $\mathscr{R}$ is a semi-algebra with $\sigma(\mathscr{R}) = \mathscr{X}^\infty$, while $\mathscr{C}$ is an algebra with $\sigma(\mathscr{C}) = \mathscr{X}^\infty$.

3.3 The Ionescu Tulceă theorem

The theorem of Ionescu Tulceă is a classical result [15] (see also[24, 2, 10] for more easily accessible references) of existence and unicity of a probability measure on the space $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ constructed out of finite dimensional marginals.

Theorem 3.3.1 (Ionescu Tulceă) *Let $(\mathbb{X}_n, \mathscr{X}_n)_{n\in\mathbb{N}}$ be a sequence of measurable spaces, $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ the corresponding trajectory space, $(N_{n+1})_{n\in\mathbb{N}}$ a sequence of transition probabilities*[1] $(\mathbb{X}^n, \mathscr{X}^n) \overset{N_{n+1}}{\rightsquigarrow} (\mathbb{X}_{n+1}, \mathscr{X}_{n+1})$, *and μ a probability on $(\mathbb{X}_0, \mathscr{X}_0)$. Define, for $n \geq 0$, a sequence of probabilities $\mathbb{P}_\mu^{(n)}$ on $(\mathbb{X}^n, \mathscr{X}^n)$ by initialising $\mathbb{P}_\mu^{(0)} = \mu$ and recursively, for $n \in \mathbb{N}$, by*

$$\mathbb{P}_\mu^{(n+1)}(F) = \int_{\mathbb{X}_0\times\cdots\times\mathbb{X}_{n+1}} \mathbb{P}_\mu^{(n)}(dx_0 \times \cdots \times dx_n) N_{n+1}((x_0, \ldots, x_n); dx_{n+1}) \mathbb{1}_F(x_0, \ldots, x_{n+1}),$$

for $F \in \mathscr{X}^{n+1}$. Then there exists a unique probability $\mathbb{P}_\mu$ on $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ such that for all n we have

$$\mathfrak{p}_n(\mathbb{P}_\mu) = \mathbb{P}_\mu^{(n)},$$

(i.e. $\mathfrak{p}_n(\mathbb{P}_\mu)(G) \equiv \mathbb{P}_\mu(\mathfrak{p}_n^{-1}(G)) = \mathbb{P}_\mu^{(n)}(G)$ for all $G \in \mathscr{X}^n$).

1. Beware of superscripts and subscripts

Before proving the theorem, note that for $F = A_0 \times \cdots \times A_n \in \mathscr{X}^n$, the term appearing in the last equality in the theorem 3.3.1 reads:

$$\mathbb{P}_\mu(\mathfrak{p}_n^{-1}(F)) = \mathbb{P}_\mu(A_0 \times \cdots \times A_n \times \mathbb{X}_{n+1} \times \mathbb{X}_{n+2} \times \dots)$$

i.e. $\mathbb{P}_\mu$ is the **unique probability** on the trajectory space $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ admitting $(\mathbb{P}_\mu^{(n)})_{n\in\mathbb{N}}$ as sequence of n-dimensional marginal probabilities.

A sequence $(\mathbb{P}_\mu^{(n)}, \mathfrak{p}_n)_{n\in\mathbb{N}}$ as in theorem 3.3.1 is called a **projective system**. The unique probability $\mathbb{P}_\mu$ on the trajectory space, is called the **projective limit** of the projective system, and is denoted

$$\mathbb{P}_\mu = \varprojlim_{n\to\infty} (\mathbb{P}_\mu^{(n)}, \mathfrak{p}_n).$$

The essential ingredient in the proof of existence of the projective limit is the **Kolmogorov compatibility condition** reading: $\mathbb{P}_\mu^{(m)} = \mathfrak{p}_m^n(\mathbb{P}_\mu^{(n)})$ for all m, n such that $0 \le m \le n$ or, in other words, $\mathfrak{p}_m = \mathfrak{p}_m^n \circ \mathfrak{p}_n$. The proof relies also heavily on standard results of the "monotone class type"; for the convenience of the reader, the main results of this type are reminded in the Appendix.

Proof of theorem 3.3.1: Denote by $\mathscr{C}_n = \mathfrak{p}_n^{-1}(\mathscr{X}^n)$. Obviously, $\mathscr{C} = \cup_{n\in\mathbb{N}}\mathscr{C}_n$. Since for $m \le n$ we have $\mathscr{C}_m \subseteq \mathscr{C}_n$, the collection $\mathscr{C}$ is an algebra on $\mathbb{X}^\infty$. In fact,

- Obviously $\mathbb{X}^\infty \in \mathscr{C}$.
- For all $C_1, C_2 \in \mathscr{C}$ there exist integers N_1 and N_2 such that $C_1 \in \mathscr{C}_{N_1}$ and $C_2 \in \mathscr{C}_{N_2}$; hence, choosing $N = \max(N_1, N_2)$, we see that the collection $\mathscr{C}$ is closed for finite intersections.
- Similarly, for every $C \in \mathscr{C}$ there exists an integer N such that $C \in \mathscr{C}_N$; hence $C^c \in \mathscr{C}_N \subseteq \mathscr{C}$, because $\mathscr{C}_N$ is a σ-algebra.

However, $\mathscr{C}$ is not a σ-algebra but it generates the full σ-algebra $\mathscr{X}^\infty$ (see exercise 3.2.5).

Next we show that if a probability $\mathbb{P}_\mu$ exists, then it is necessarily unique. In fact, for every $C \in \mathscr{C}$ there exists an integer N and a set $F \in \mathscr{X}^N$ such that $C = \mathfrak{p}_N^{-1}(F)$. Therefore, $\mathbb{P}_\mu(C) = \mathbb{P}_\mu(\mathfrak{p}_N^{-1}(F)) = \mathbb{P}_\mu^{(N)}(F)$. If another such measure $\mathbb{P}'_\mu$ exists (i.e. satisfying the same compatibility properties) we shall have similarly $\mathbb{P}'_\mu(C) = \mathbb{P}_\mu^{(N)}(F) = \mathbb{P}_\mu(C)$, hence $\mathbb{P}_\mu = \mathbb{P}'_\mu$ on $\mathscr{C}$. Define for a fixed $C \in \mathscr{C}$,

$$\Lambda_C = \{A \in \mathscr{X}^\infty : \mathbb{P}_\mu(C \cap A) = \mathbb{P}'_\mu(C \cap A)\} \subseteq \mathscr{X}^\infty.$$

Next we show that Λ_C is a λ-system. In fact

- Obviously $\mathbb{X}^\infty \in \Lambda_C$.
- If $B \in \Lambda_C$ then

$$\begin{aligned}\mathbb{P}_\mu(C \cap B^c) &= \mathbb{P}_\mu(C \setminus (C \cap B))\\ &= \mathbb{P}_\mu(C) - \mathbb{P}_\mu(C \cap B)\\ &= \mathbb{P}'_\mu(C) - \mathbb{P}'_\mu(C \cap B)\\ &= \mathbb{P}'_\mu(C \cap B^c).\end{aligned}$$

Hence $B^c \in \Lambda_C$.

– Suppose now that $(B_n)_{n\in\mathbb{N}}$ is a disjoint family of sets in Λ_C. Then σ-additivity establishes that

$$\mathbb{P}_\mu((\sqcup_{n\in\mathbb{N}} B_n)\cap C) = \sum_{n\in\mathbb{N}} \mathbb{P}_\mu(B_n \cap C) = \sum_{n\in\mathbb{N}} \mathbb{P}'_\mu(B_n\cap C) \le \mathbb{P}'_\mu(C) = \mathbb{P}_\mu(C).$$

Hence $\sqcup_{n\in\mathbb{N}} B_n \in \Lambda_C$ establishing thus that Λ_C is a λ-system.

Now stability of $\mathscr{C}$ under finite intersections implies that $\sigma(\mathscr{C}) = \lambda(\mathscr{C}) \subseteq \Lambda_C$ (see exercise A.1.6); consequently $\sigma(\mathscr{C}) \subseteq \Lambda_C \subseteq \mathscr{X}^\infty$. But $\sigma(\mathscr{C}) = \mathscr{X}^\infty$. Hence for all $C \in \mathscr{C}$, we have $\Lambda_C = \mathscr{X}^\infty$. Consequently, for all $A \in \mathscr{X}^\infty$ we have $\mathbb{P}_\mu(C\cap A) = \mathbb{P}'_\mu(C\cap A)$. Choose now an increasing sequence of cylinders $(C_n)_{n\in\mathbb{N}}$ such that $C_n \uparrow \mathbb{X}^\infty$. Monotone continuity of $\mathbb{P}_\mu$ yields then $\mathbb{P}_\mu = \mathbb{P}'_\mu$, proving unicity of $\mathbb{P}_\mu$.

To establish existence of $\mathbb{P}_\mu$ verifying the projective condition, it suffices to show that $\mathbb{P}_\mu$ is a premeasure on the algebra $\mathscr{C}$. For an arbitrary $C \in \mathscr{C}$, chose an integer N and a set $F \in \mathscr{X}^N$ such that $C = \mathfrak{p}_N^{-1}(F)$. On defining $\kappa(C) = \mathbb{P}_\mu^{(N)}(F)$, we note that κ is well defined on $\mathscr{C}$ and is a content. Further $\kappa(\mathbb{X}^\infty) = \pi(\mathfrak{p}_N^{-1}(\mathbb{X}^N)) = \mathbb{P}_\mu^{(N)}(\mathbb{X}^N) = 1$. To show that κ is a premeasure, it remains to show continuity at $\emptyset$ (see exercise A.1.8), i.e. for any sequence of cylinders (C_n) such that $C_n \downarrow \emptyset$, we must have $\kappa(C_n) \downarrow 0$, or, equivalently, if $\inf_n \kappa(C_n) > 0$, then $\cap_n C_n \neq \emptyset$. Let (C_n) be a decreasing sequence of cylinders such that $\inf \kappa(C_n) > 0$. Without loss of generality, we can represent these cylinders for all n by $C_n = F \times (\times_{k>n}\mathbb{X}_k)$ and $F \in \mathscr{X}^n$. For any $x_0 \in \mathbb{X}_0$, consider the sequence of functions

$$f_n^{(0)}(x_0) = \int_{\mathbb{X}_1} N_1(x_0, dx_1) \int_{\mathbb{X}_2} N_2(x_0x_1, dx_2) \ldots \int_{\mathbb{X}_n} N_n(x_0\cdots x_{n-1}, dx_n) \mathbb{1}_F(x_0\ldots x_n).$$

Obviously, for all n, $f_{n+1}^{(0)}(x_0) \le f_n^{(0)}(x_0)$ (why?). Monotone convergence yields then:

$$\int_{\mathbb{X}_0} \inf_n f_n^{(0)}(x_0)\mu(dx_0) = \inf_n \int_{\mathbb{X}_0} f_n^{(0)}(x_0)\mu(dx_0) = \inf_n \kappa(C_n) > 0.$$

Consequently, there exists a $\overline{x}_0 \in \mathbb{X}_0$ such that $\inf_n f_n^{(0)}(\overline{x}_0) > 0$. We can introduce similarly a decreasing sequence of two variables

$$f_n^{(1)}(\overline{x}_0, x_1) = \int_{\mathbb{X}_2} N_2(\overline{x}_0x_1, dx_2) \ldots \int_{\mathbb{X}_n} N_n(\overline{x}_0x_1\cdots x_{n-1}, dx_n) \mathbb{1}_F((\overline{x}_0\ldots x_n)$$

and show that

$$\int_{\mathbb{X}_1} \inf_n f_n^{(1)}(\overline{x}_0, x_1) N_1(\overline{x}_0, dx_1) = \inf_n \int_{\mathbb{X}_1} f_n^{(1)}(\overline{x}_0, x_1) N_1(\overline{x}_0, dx_1) > 0.$$

There exists then a $\overline{x}_1 \in \mathbb{X}_1$ such that $\inf_n f_n^{(1)}(\overline{x}_0, \overline{x}_1) > 0$. By recurrence, we show then that there exists $\overline{x} = (\overline{x}_0, \overline{x}_1, \ldots) \in \mathbb{X}^\infty$ such that for all k we have $\inf_n f_n^{(k)}(\overline{x}_0, \ldots, \overline{x}_k) > 0$. Therefore, $\overline{x} \in \cap_n C_n$ implying that $\cap_n C_n \neq \emptyset$. □

Remark: There are several other possibilities defining a probability on the trajectory space out of a sequence of finite dimensional data, for instance by fixing the sequence of conditional probabilities verifying the so called Dobrushin-Lanford-Ruelle (DLR) compatibility conditions instead of Komogorov compatibiltiy conditions for marginal probabilities (see [11] for instance). This construction naturally arises in Statistical Mechanics. The DLR condition is less rigid than the Kolmogorov condition: the projective system may fail to converge, can have a unique limit or have several convergent subsequences. Limiting probabilities are called **Gibbs measures** of a DLR projective system. If there are several different Gibbs measures, the system is said undergoing a **phase transition**.

Definition 3.3.2 A (discrete-time) **stochastic process** is the sequence of probabilities $(\mathbb{P}_\mu^{(n)})_{n\in\mathbb{N}}$ appearing in Ionescu-Tulceă existence theorem.

Definition 3.3.3 Suppose that for every n, the kernel $(\mathbb{X}^n, \mathscr{X}^n) \overset{N_{n+1}}{\rightsquigarrow} (\mathbb{X}_{n+1}, \mathscr{X}_{n+1})$ is a Markovian kernel depending merely on x_n (instead of the complete dependence on $(x_0,\ldots,x_n)$), i.e. for all $n \in \mathbb{N}$, there exists a Markovian kernel $(\mathbb{X}_n, \mathscr{X}_n) \overset{P_{n+1}}{\rightsquigarrow} (\mathbb{X}_{n+1}, \mathscr{X}_{n+1})$ such that $N_{n+1}((x_0,\ldots,x_n); A_{n+1}) = P_{n+1}(x_n; A_{n+1})$ for all $x_0,\ldots,x_n \in \mathbb{X}_n$ and all $A_{n+1} \in \mathscr{X}_{n+1}$. Then the stochastic process is termed a **Markov chain**.

Let $\Omega = \mathbb{X}^\infty$, $\mathscr{F} = \mathscr{X}^\infty$, and $(\mathbb{X}_n, \mathscr{X}_n) \overset{P_{n+1}}{\rightsquigarrow} (\mathbb{X}_{n+1}, \mathscr{X}_{n+1})$ be a sequence of Markovian kernels as in definition 3.3.3. Then theorem 3.3.1 guarantees, for every probability $\mu \in \mathscr{M}_1(\mathscr{X}_0)$, the existence of a unique probability $\mathbb{P}_\mu$ on $(\mathbb{X}^\infty, \mathscr{X}^\infty)$. Let $C = \mathfrak{p}_N^{-1}(A_0 \times \ldots \times A_N)$, for some integer $N > 0$, be a cylinder set. Then,

$$\begin{aligned}
\mathbb{P}_\mu(C) &= \mathbb{P}_\mu(A_0 \times \cdots \times A_N \times \mathbb{X}_{N+1} \times \mathbb{X}_{N+2} \times \cdots) \\
&= \mathbb{P}_\mu(\mathfrak{p}_N^{-1}(A_0 \times \cdots \times A_N)) \\
&= \mathfrak{p}_N(\mathbb{P}_\mu)(A_0 \times \cdots \times A_N) \\
&= \mathbb{P}_\mu^{(N)}(A_0 \times \cdots \times A_N) \\
&= \int_{A_0 \times \cdots \times A_{N-1}} \mathbb{P}_\mu^{(N-1)}(dx_0 \times \cdots \times dx_{N-1}) P_N(x_{N-1}; A_N) \\
&\vdots \\
&= \int_{A_0 \times \cdots \times A_N} \mu(dx_0) P_1(x_0, dx_1) \cdots P_N(x_{N-1}, dx_N).
\end{aligned}$$

On defining $X : \Omega \to \mathbb{X}^\infty$ by $X_n(\omega) = \omega_n$ for all $n \in \mathbb{N}$. we have on the other hand,

$$\begin{aligned}
\mathbb{P}_\mu(C) &= \mathbb{P}_\mu^{(N)}(A_0 \times \cdots \times A_N) \\
&= \mathbb{P}_\mu(\{\omega \in \Omega : X_0(\omega) \in A_0, \ldots X_N(\omega) \in A_N\}).
\end{aligned}$$

The **coordinate mappings** $X_n(\omega) = \omega_n$, for $n \in \mathbb{N}$, defined in the above framework provide the **canonical (minimal) realisation** of the sequence $X = (X_n)_{n\in\mathbb{N}}$.

Remark: If $\mu = \epsilon_x$ for some $x \in \mathbb{X}$, we write simply $\mathbb{P}_x$ or $\mathbb{P}_x^{(N)}$ instead of $\mathbb{P}_{\epsilon_x}$ or $\mathbb{P}_{\epsilon_x}^{(N)}$. Note further that for every $\mu \in \mathcal{M}_1(\mathcal{X}_0)$ we have $\mathbb{P}_\mu = \int_{\mathbb{X}_0} \mu(dx)\mathbb{P}_x$.

Remark: Giving

- the sequence of measurable spaces $(\mathbb{X}_n, \mathcal{X}_n)_{n\in\mathbb{N}}$,
- the sequence $(P_{n+1})_{n\in\mathbb{N}}$ of the Markovian kernels $(\mathbb{X}_n, \mathcal{X}_n) \overset{P_{n+1}}{\rightsquigarrow} (\mathbb{X}_{n+1}, \mathcal{X}_{n+1})$, and
- the initial probability $\mu \in \mathcal{M}_1(\mathcal{X}_0)$,

uniquely determines

- the trajectory space $(\mathbb{X}^\infty, \mathcal{X}^\infty)$,
- a unique probability $\mathbb{P}_\mu \in \mathcal{M}_1(\mathcal{X}^\infty)$, projective limit of the sequence of finite dimensional marginals, and
- upon identifying $(\Omega, \mathcal{F}) = (\mathbb{X}^\infty, \mathcal{X}^\infty)$, this construction also provides the canonical realisation of the sequence $X = (X_n)_{n\in\mathbb{N}}$ through the standard coordinate mappings.

Under these conditions, we say that X is an **(inhomogeneous) Markov chain**, and more precisely a $\mathsf{MC}((\mathbb{X}_n, \mathcal{X}_n)_{n\in\mathbb{N}}, (P_{n+1})_{n\in\mathbb{N}}, \mu))$. If for all $n \in \mathbb{N}$, we have $\mathbb{X}_n = \mathbb{X}$, $\mathcal{X}_n = \mathcal{X}$, and $P_n = P$, for some measurable space $(\mathbb{X}, \mathcal{X})$ and Markovian kernel $(\mathbb{X}, \mathcal{X}) \overset{P}{\rightsquigarrow} (\mathbb{X}, \mathcal{X})$, then we have a **(homogeneous) Markov chain** and more precisely a $\mathsf{MC}((\mathbb{X}, \mathcal{X}), P, \mu)$.

3.4 Weak Markov property

Proposition 3.4.1 (Weak Markov property) *Let X be a $\mathsf{MC}((\mathbb{X}, \mathcal{X}), P, \mu)$. For all $f \in b\mathcal{X}$ and all $n \in \mathbb{N}$,*

$$\mathbb{E}_\mu(f(X_{n+1})|\mathcal{X}^n) = Pf(X_n) \text{ a.s.}$$

Proof: By the definition of conditional expectation, we must show that for all $n \in \mathbb{N}$ and $F \in \mathcal{X}^n$, the measures α and β, defined by

$$\begin{aligned} \alpha(F) &= \int_F f(X_{n+1}(\omega))\mathbb{P}_\mu(d\omega) \\ \beta(F) &= \int_F Pf(X_n(\omega))\mathbb{P}_\mu(d\omega) \end{aligned}$$

coincide on $\mathcal{X}^n$. Let

$$\mathcal{A}_n \equiv \times_{k=0}^n \mathcal{X}_k = \{A_0 \times \cdots \times A_n, A_k \in \mathcal{X}_k, 0 \le k \le n\}.$$

Now, it is immediate that for all n, the family $\mathscr{A}_n$ is a π-system while $\sigma(\mathscr{A}_n) = \mathscr{X}_n$. It is therefore enough to check equality of the two measures on $\mathscr{A}_n$. On any $F \in \mathscr{A}_n$, of the form $F = A_0 \times \ldots \times A_n$, with $A_i \in \mathscr{X}$,

$$\alpha(F) = \int \mathbb{P}_\mu(\{\omega : X_0(\omega) \in A_0, \ldots, X_n(\omega) \in A_n, X_{n+1}(\omega) \in dx_{n+1}, X_{n+2}(\omega) \in \mathbb{X}, \ldots\}) f(x_{n+1}).$$

Hence

$$\begin{aligned} \alpha(F) &= \int \mathbb{P}_\mu(\mathfrak{p}_{n+1}^{-1}(A_0, \ldots, A_n, dx_{n+1})) f(x_{n+1}) \\ &= \int \mathbb{P}_\mu^{n+1}(F \times dx_{n+1}) f(x_{n+1}). \end{aligned}$$

Developing the right hand side of the above expression, we get

$$\begin{aligned} \alpha(F) &= \int_{A_0 \times \cdots \times A_n \times \mathbb{X}} \mu(dx_0) P(x_0, dx_1) \cdots P(x_n, dx_{n+1}) f(x_{n+1}) \\ &= \beta(F). \end{aligned}$$

□

Remark: The previous proof is an example of the use of some version of "monotone class theorems". An equivalent way to prove the weak Markov property should be to use the following conditional equality for all $f_0, \ldots, f_n, f \in b\mathscr{X}$:

$$\begin{aligned} \mathbb{E}_\mu(f_0(X_0) \ldots f_n(X_n) f(X_{n+1})) &= \mathbb{E}_\mu(f_0(X_0) \ldots f_n(X_n) \mathbb{E}_\mu(f(X_{n+1}) | \mathscr{X}^n)) \\ &= \mathbb{E}_\mu(f_0(X_0) \ldots f_n(X_n) P f(X_n)). \end{aligned}$$

Privileging measure formulation or integral formulation is purely a matter of taste.

The following statement can be used as an alternative definition of the Markov chain.

Definition 3.4.2 Let $(\Omega, \mathscr{F}, (\mathscr{F})_{n \in \mathbb{N}}, \mathbb{P})$ be a filtered probability space, $X = (X_n)_{n \in \mathbb{N}}$ a sequence of $(\mathbb{X}, \mathscr{X})$-valued random variables defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $(\mathbb{X}, \mathscr{X}) \overset{P}{\rightsquigarrow} (\mathbb{X}, \mathscr{X})$ a Markovian kernel. We say that X is a $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$ if

1. (X_n) is adapted to the filtration $(\mathscr{F}_n)$,
2. $\mathsf{Law}(X_0) = \mu$, and
3. for all $f \in b\mathscr{X}$, the equality $\mathbb{E}(f(X_{n+1}) | \mathscr{F}_n) = P f(X_n)$ holds almost surely.

We assume henceforth, unless stated differently, that $\Omega = \mathbb{X}^\infty$, $\mathscr{F} = \mathscr{X}^\infty$, $X_n(\omega) = \omega_n$ while $\mathscr{F}_n \simeq \mathscr{X}^n$. The last statement in the definition 3.4.2 implicitly implies that the conditional probability $\mathbb{P}(X_{n+1} \in A|\mathscr{F}_n)$ admits a regular version.

Let X be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$. Distinguish artificially the (identical) probability spaces $(\Omega,\mathscr{F},\mathbb{P}_\mu)$ and $(\mathbb{X}^\infty,\mathscr{X}^\infty,\mathbb{P}^X_\mu)$; equip further those spaces with filtrations $\mathscr{F}_n$ and $\mathscr{X}^n$ respectively. Assume that $X:\Omega\to\mathbb{X}^\infty$ is the canonical representation of X (i.e. $X=\mathbb{1}$). Let $G\in\mathscr{X}^\infty_+$ and $\Gamma = G\circ X:\Omega\to\mathbb{R}_+$ be a random variable (note that $\Gamma = G$!). Define the right shift $\theta:\Omega\to\Omega$ for $\omega=(\omega_0,\omega_1,\ldots)$ by $\theta(\omega)=(\omega_1,\omega_2,\ldots)$ and powers of θ by $\theta^0=\mathbb{1}$ and recursively for $n>0$ by $\theta^n=\theta\circ\theta^{n-1}$.

Theorem 3.4.3 *The weak Markov property is equivalent to the equality*

$$\mathbb{E}_\mu(\Gamma\circ\theta^n|\mathscr{F}_n)=\mathbb{E}_{X_n}(\Gamma),$$

holding for all $\mu\in\mathscr{M}_1(\mathscr{X})$, all $n\in\mathbb{N}$, and all $\Gamma\in m\mathscr{F}_+$ on the set $\{X_n\neq\partial\}$.

Remark: Define $\gamma(y)=\mathbb{E}_y(\Gamma)$ for all $y\in\mathbb{X}$. The meaning of the right hand side of the above equation is $\mathbb{E}_{X_n}(\Gamma)\equiv\gamma(X_n)$.

Exercise 3.4.4 Prove the theorem 3.4.3.

3.5 Strong Markov property

Definition 3.5.1 Let $(\Omega,\mathscr{F},(\mathscr{F}_n)_n)$ be a filtered space.

- A generalised random variable $T:\Omega\to\mathbb{N}\cup\{+\infty\}$ is called a **stopping time** (more precisely a $(\mathscr{F}_n)$-stopping time) if for all $n\in\mathbb{N}$ we have $\{\omega\in\Omega: T(\omega)=n\}\in\mathscr{F}_n$.
- Let T be a stopping time. The collection of events

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

  is the **trace σ-algebra**.

Definition 3.5.2 Let X be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ and $A\in\mathscr{X}$.

- The **hitting time** of A, is the stopping time

$$\tau^1_A=\inf\{n>0:X_n\in A\}.$$

- The **passage time** at A, is the stopping time

$$\tau^0_A=\inf\{n\geq 0:X_n\in A\}.$$

– The **death time** of X, is the stopping time

$$\zeta \equiv \tau^0_{\{\partial\}} = \inf\{n \geq 0 : X_n = \partial\}.$$

The symbols $\tau^\flat_A$, for $\flat \in \{0,1\}$ and ζ are reserved.

Remark: Note the difference between τ^0_A and τ^1_A. Suppose that the Markov chain starts with initial probability $\mu = \epsilon_x$ for some $x \in \mathscr{X}$. If $x \in A$, then $\tau^0_A = 0$, while τ^1_A may be arbitrary, even ∞. If $x \notin A$, then $\tau^0_A = \tau^1_A$. If the transition probability of the chain is sub-Markovian, the space is extended to include $\{\partial\}$ that can be visited with strictly positive probability. If the transition probability is already Markovian, we can always extend the state space $\mathbb{X}$ to contain $\{\partial\}$, the latter being a negligible event; the death time is then $\zeta = \infty$ a.s.

Remark: Another quantity, defined for all $F \in \mathscr{X}$ and $\flat \in \{0,1\}$ by

$$\eta^\flat(F) = \sum_{n \geq \flat} \mathbb{1}_F(X_n),$$

bears often the name "occupation time" in the literature. We definitely prefer the term **occupation measure** than occupation time because it is a (random) measure on $\mathscr{X}$. The symbol $\eta^\flat(\cdot)$ will be reserved in the sequel.

Exercise 3.5.3 Is the occupation measure a stopping time?

Let T be a stopping time. Then

$$\begin{aligned} X_T(\omega) &= \begin{cases} X_n(\omega) & \text{on} \quad \{T = n\} \\ \partial & \text{on} \quad \{T = \infty\} \end{cases} \\ &= \sum_{n \in \mathbb{N}} X_n \mathbb{1}_{\{T=n\}} + \partial \mathbb{1}_{\{T=\infty\}}, \end{aligned}$$

establishing the $(\mathscr{F}_T, \mathscr{X})$-measurability of X_T.

We already know that the right shift θ verifies $X_n \circ \theta^m(\omega) = X_n(\theta^m(\omega)) = X_{n+m}(\omega) = \omega_{n+m}$. On defining

$$\theta^T(\omega) = \begin{cases} \theta^n(\omega) & \text{on} \quad \{T = n\} \\ \omega_\partial & \text{on} \quad \{T = \infty\} \end{cases}$$

where $\omega_\partial = (\partial, \partial, \ldots)$, we see that $\theta^T \in \mathscr{F}_T$ and

$$\begin{aligned} X_{n+T}(\omega) &= X_n \circ \theta^T(\omega) \\ &= X_n(\theta^T(\omega)) \\ &= \begin{cases} X_{n+m}(\omega) & \text{on} \quad \{T = m\} \\ \partial & \text{on} \quad \{T = \infty\} \end{cases} \\ &= \sum_{m \in \mathbb{N}} X_{n+m} \mathbb{1}_{\{T=m\}} + \partial \mathbb{1}_{\{T=\infty\}} \\ &= X_{n+T} \mathbb{1}\{T < \infty\} + \partial \mathbb{1}_{\{T=\infty\}}. \end{aligned}$$

Theorem 3.5.4 (Strong Markov property) *Let $\mu \in \mathcal{M}_1(\mathcal{X})$ be an arbitrary probability and Γ a bounded random variable defined on $(\Omega, \mathcal{F}, \mathbb{P}_\mu)$. For every stopping time T we have:*

$$\mathbb{E}_\mu(\Gamma \circ \theta^T | \mathcal{F}_T) = \mathbb{E}_{X_T}(\Gamma),$$

with the two members of the above equality vanishing on $\{X_T = \partial\}$.

Proof: Define $\gamma(y) = \mathbb{E}_y(\Gamma)$ for $y \in \mathbb{X}$. We must show that for all $A \in \mathcal{F}_T$, we have on $\{X_T \neq \partial\}$:

$$\int_A \Gamma \circ \theta^T(\omega) \mathbb{P}_\mu(d\omega) = \int_A \gamma(X_T(\omega)) \mathbb{P}_\mu(d\omega).$$

Now,

$$A = [\sqcup_{n\in\mathbb{N}}(A \cup \{T = n\})] \sqcup [A \cup \{T = \infty\}].$$

On $\{T = \infty\}$, we have $\{X_T = \partial\}$. Hence on $\{X_T \neq \partial\}$, the above partition reduces to $A = \sqcup_{n\in\mathbb{N}}(A \cup \{T = n\})$. Hence, the sought left hand side reads:

$$\begin{aligned} \text{l.h.s.} &= \sum_{n\in\mathbb{N}} \int_{A\cap\{T=n\}} \Gamma \circ \theta^T(\omega) \mathbb{P}_\mu(d\omega) \\ &= \sum_{n\in\mathbb{N}} \int_{A\cap\{T=n\}} \Gamma \circ \theta^n(\omega) \mathbb{P}_\mu(d\omega) \\ &= \sum_{n\in\mathbb{N}} \int_{A\cap\{T=n\}} \gamma(X_n(\omega)) \mathbb{P}_\mu(d\omega), \end{aligned}$$

the last equality being a consequence of the weak Markov property because $A \cap \{T = n\} \in \mathcal{F}_n$. □

3.6 Examples-exercises

1. Let $\mathcal{T}_n = \sigma(X_n, X_{n+1}, \ldots)$ for all $n \in \mathbb{N}$ and recall that $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Prove that for all $A \in \mathcal{F}_n$ and all $B \in \mathcal{T}_n$, past and future become conditionally independent of present, i.e.

$$\mathbb{P}_\mu(A \cap B | \sigma(X_n)) = \mathbb{P}_\mu(A|\sigma(X_n)) \mathbb{P}_\mu(B|\sigma(X_n)).$$

2. Let $X = (X_n)_{n\in\mathbb{N}}$ be a sequence of independent $(\mathbb{X}, \mathcal{X})$-valued random variables identically distributed according to the law ν. Show that X is $\mathsf{MC}((\mathbb{X}, \mathcal{X}), P, \mu)$ with $P(x, A) = \nu(A)$ for all $x \in \mathbb{X}$ and all $A \in \mathcal{X}$.
3. Let $(\mathbb{X}, \mathcal{X})$ be a measurable space, $f \in b\mathcal{X}$, and x a point in $\mathbb{X}$. Define a sequence $X = (X_n)_{n\in\mathbb{N}}$ by $X_0 = x$ and recursively $X_{n+1} = f(X_n)$ for $n \in \mathbb{N}$; the trajectory of point x under the dynamical system f. Show that X is a $\mathsf{MC}((\mathbb{X}, \mathcal{X}), P, \mu)$ with $\mu = \epsilon_x$ and $P(y, A) = \epsilon_y(f^{-1}(A))$ for all $y \in \mathbb{Y}$ and all $A \in \mathcal{X}$.

4. Consider again the heads and tails example presented in chapter 1. Show that the sequence of outcomes is a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ with $\mathbb{X}=\{0,1\}$ and P the kernel defined by $P(x,\{y\})\equiv P(x,y)$ where $P=(P(x,y))_{x,y\in\mathbb{X}}=\begin{pmatrix}1-a & a\\ b & 1-b\end{pmatrix}$, with $0\le a,b\le 1$.
5. Let X be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ on a discrete (finite or countable) set $\mathbb{X}$.
   - Show that the kernel P can be uniquely represented as a **weighed directed graph** having $\mathbb{A}^0=\mathbb{X}$ as vertex set and $\mathbb{A}^1=\{(x,y)\in\mathbb{X}^2: P(x,y)>0\}$ as set of directed edges; any edge (x,y) is assigned a weight $P(x,y)$.
   - For every edge $a\in\mathbb{A}^1$ define two mappings $s,t:\mathbb{A}^1\to\mathbb{A}^0$, the source and terminal maps defined by

     $$\begin{array}{rcl}\mathbb{A}^1\ni a=(x,y) & \mapsto & s(a)=x\in\mathbb{A}^0,\\ \mathbb{A}^1\ni a=(x,y) & \mapsto & t(a)=y\in\mathbb{A}^0.\end{array}$$

     Let[2] $\mathbb{A}^2=\{ab\in\mathbb{A}^1\times\mathbb{A}^1: s(b)=t(a)\}$ be the set of composable edges (paths of length 2). Let $\alpha=\alpha_1\cdots\alpha_n$ be a sequence of n edges such that for all $i, 0\le i<n$, subsequent edges are composable, i.e. $\alpha_i\alpha_{i+1}\in\mathbb{A}^2$. Such an α is termed **combinatorial path** of length n; the set of all paths of length n is denoted by $\mathbb{A}^n$. Finally, define the space of **combinatorial paths** of indefinite length by $\mathbb{A}^*=\cup_{n=0}^{\infty}\mathbb{A}^n$. Show that $\mathbb{A}^*$ has a natural forest structure composed of rooted trees.
   - Let $\alpha\in\mathbb{A}^*$, denote by $|\alpha|$ the length of the skeleton α, and by $v(\alpha)$ the sequence of vertices appearing in α. Show that the cylinder set $[v(\alpha)]$ has probability

     $$\begin{array}{rcl}\mathbb{P}_\mu([v(\alpha)]) & = & \mu(s(\alpha_1))\times P(s(\alpha_1),t(\alpha_1))\\ & & \times\cdots\times P(s(\alpha_{|\alpha|}),t(\alpha_{|\alpha|})).\end{array}$$

6. Let $\xi=(\xi_n)_n$ be a sequence of $(\mathbb{Y},\mathscr{Y})$-valued i.i.d. random variables, such that $\mathsf{Law}(\xi_0)=\nu$. Let X_0 a $(\mathbb{X},\mathscr{X})$-valued random variable, independent of ξ such that $\mathsf{Law}(X_0)=\mu$, and $f:\mathbb{X}\times\mathbb{Y}\to\mathbb{X}$ be a bounded measurable function. Define for $n\ge 0$ $X_{n+1}=f(X_n,\xi_{n+1})$. Show that X is a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ with Markovian kernel given by

   $$P(x,A)=\nu(\{y\in\mathbb{Y}: f(x,y)\in A\}),$$

   for all $x\in\mathbb{X}$ and all $A\in\mathscr{X}$.
7. Let $\mathbb{X}$ be the Abelian group $\mathbb{Z}^d$; $\Gamma=\{\pm\mathbf{e}_1,\ldots,\pm\mathbf{e}_d\}$ a (minimal) generating set (i.e. $\mathbb{N}\Gamma=\mathbb{Z}^d$), and ν a probability on Γ. Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of independent Γ-valued random variables identically distributed according to ν. Define, for $x\in\mathbb{X}$ and $n\ge 0$, the sequence $X_n=x+\sum_{i=0}\xi_i$. Show

2. Note that $\mathbb{A}^2$ **is not** in general the Cartesian product $\mathbb{A}^1\times\mathbb{A}^1$. The graph is not in general a group; it has merely a semi-groupoid structure $(\mathbb{A}^0,\mathbb{A}^1,s,t)$.

that X is a Markov chain on $\mathbb{Z}^d$; determine its Markovian kernel and initial probability. (This Markov chain is termed a nearest-neighbour **random walk** on $\mathbb{Z}^d$, anchored at x.)

8. Let $\mathbb{X} = \mathbb{R}$ and $\mathscr{X} = \mathscr{B}(\mathbb{R}^d)$, $\nu \in \mathscr{M}_+(\mathscr{X})$, and $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of independent $\mathbb{X}$-valued random variables identically distributed according to ν. Define, for $x \in \mathbb{X}$ and $n \geq 0$, the sequence $X_n = x + \sum_{i=0} \xi_i$. Show that X is a Markov chain on $\mathbb{R}^d$; determine its Markovian kernel and initial probability. (This Markov chain is termed a random walk on $\mathbb{R}^d$, anchored at x.)
9. Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of independent $\mathbb{R}$-valued random variables identically distributed according to ν. Define, for $x \in \mathbb{R}_+$ and $n \geq 0$, the sequence $X = (X_n)$, by $X_0 = x$ and recursively $X_{n+1} = (X_n + \xi_{n+1})^+$. Show that X is a Markov chain on an appropriate space, determine its Markovian kernel and its initial probability.
10. Let $(\mathbb{G}, \mathscr{G})$ be a topological locally compact group with composition denoted multiplicatively, $\nu \in \mathscr{M}_+(\mathscr{G})$, and $(\xi_i)_{i\in\mathbb{N}}$ a sequence of independent $(\mathbb{G}, \mathscr{G})$-valued random variables identically distributed according to ν. Define, for $x \in \mathbb{R}_+$ and $n \geq 0$, the sequence $X = (X_n)$, by $X_0 = x$ and recursively $X_{n+1} = \xi_{n+1} X_n$. Show that X is a Markov chain on $(\mathbb{G}, \mathscr{G})$; determine its Markovian kernel and initial probability.
11. Let $\mathbb{X} = \mathbb{R}^d$, $\mathscr{X} = \mathscr{B}(\mathbb{R}^d)$, $\nu \in \mathscr{M}_+(\mathscr{X})$, and $(\xi_i)_{i\in\mathbb{N}}$ a sequence of independent $(\mathbb{X}, \mathscr{X})$-valued random variables identically distributed according to ν. Define $\Xi_0 = 0$ and, for $n \geq 1$, $\Xi_n = \sum_{i=1}^n \xi_i$. Let $x \in \mathbb{X}$ and define $X_0 = 0$ and, for $n \geq 1$, $X_n = x + \sum_{i=1}^n \Xi_i$.
   - Show that X is not a Markov chain.
   - Let $Y_n = \begin{pmatrix} X_{n+1} \\ X_n \end{pmatrix}$, for $n \in \mathbb{N}$. Show that Y is a Markov chain (on the appropriate space that will be determined); determine its Markovian kernel and initial probability.
   - When Y defined as above is proved to be a Markov chain, the initial process X is termed Markov chain of order 2. Give a plausible definition of a **Markov chain of order** k, for $k \geq 2$.
12. Let $Z = (Z_n)_n$, with $Z_n = (X_n, Y_n)$ for $n \in \mathbb{N}$, be a $\mathsf{CM}((\mathbb{X} \times \mathbb{Y}), (\mathscr{X} \otimes \mathscr{Y})), P, \mu)$. Show that neither $X = (X_n)_n$ nor $Y = (Y_n)_n$ are in general Markov chains. They are termed **hidden Markov chains**. Justify this terminology.
13. We have defined in the exercise section 2.5 the potential kernel G^0 associated with any Markov kernel P. Consistently with our notation convention, we define $G \equiv G^1 = \sum_{n=1}^\infty P^n$. Show that for all $x \in \mathbb{X}$ and all $F \in \mathscr{X}$,
$$G(x, F) = \mathbb{E}_x(\eta^1(F)).$$
14. We define for $\flat \in \{0, 1\}$
$$\begin{aligned} L_F^\flat(x) &= \mathbb{P}_x(\tau_F^\flat < \infty) \\ H_F^\flat(x) &= \mathbb{P}_x(\eta^\flat(F) = \infty). \end{aligned}$$

– How H_F^0 compares with H_F^1?
– Does the same comparison hold for L_F^0 and L_F^1?
– Are the quantities $H^\flat$ and $L^\flat$ kernels? If yes, are they associatively composable?
– Show that $L_F^\flat(x) = \mathbb{P}_x(\cup_{n\geq\flat}\{X_n \in F\})$.
– Show that $H_F^0(x) = \mathbb{P}_x(\cap_{m\geq 0} \cup_{n\geq m} \{X_n \in F\})$.

15. Let X be a $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$.
    – Suppose that $f \in b\mathscr{X}_+$ is a right eigenvector of P associated with the eigenvalue 1. Such a function is called **bounded harmonic function** for the kernel P. Show that the sequence $(f(X_n))_{n\in\mathbb{N}}$ is a $(\mathscr{F}_n)_{n\in\mathbb{N}}$-martingale.
    – Any $f \in m\mathscr{X}_+$ verifying point-wise $Pf \leq f$ is called **superharmonic**. Show that the sequence $(f(X_n))_{n\in\mathbb{N}}$ is a $(\mathscr{F}_n)_{n\in\mathbb{N}}$-supermartingale.

4

Markov chains on finite sets

Although the study of Markov chains on finite sets is totally elementary and can be reduced to the study of powers of finite-dimensional stochastic matrices, it is instructive to give the convergence theorem 4.3.17 thus obtained as a result of the spectral theorem — a purely algebraic result in finite dimension — of the Markovian kernel. This chapter greatly relies on [16]. Spectral methods can as a matter of fact be applied to more general contexts at the expense of more a sophisticated approach to spectral properties of the kernel. In the case the Markov kernel is a compact operator on a general state space, the formulation of the convergence theorem 4.3.17 remains valid almost *verbatim*. Another important reason for studying Markov chains on finite state spaces is their usefulness to the theory of stochastic simulations.

4.1 Basic construction

Let $\mathbb{X}$ be a discrete finite set of cardinality d; without loss of generality, we can always identify $\mathbb{X} = \{0, \ldots, d-1\}$. The set $\mathbb{X}$ will always be considered equipped with $\mathscr{X} = \mathscr{P}(\mathbb{X})$. A Markovian kernel $P = (P(x,y))_{x,y\in\mathbb{X}}$ on $\mathbb{X}$ will be a $d \times d$ matrix of positive elements — where we denote $P(x,y) \equiv P(x,\{y\})$ — that verifies, for all $x \in \mathbb{X}$, $\sum_{y\in\mathbb{X}} P(x,y) = 1$. Choosing ϵ_x, with some $x \in \mathbb{X}$, as starting measure, the theorem 3.3.1 guarantees that there is a unique probability measure $\mathbb{P}_x$ on the standard trajectory space $(\Omega, \mathscr{F})$ verifying

$$\begin{aligned} \mathbb{P}_x(\mathfrak{p}_n^{-1}(A_0 \times \cdots \times A_n)) &= \mathbb{P}_x(X_0 \in A_0 \ldots X_n \in A_n) \\ &= \sum_{\substack{x_0 \in A_0 \\ \vdots \\ x_n \in A_n}} P(x_0, x_1) \cdots P(x_{n-1}, x_n). \end{aligned}$$

In particular

$$\begin{aligned} \mathbb{P}_x(X_n = y) &= P^n(x,y) \\ \mathbb{P}_\mu(X_n = y) &= \sum_{x\in\mathbb{X}} \mu(x)P^n(x,y). \end{aligned}$$

As was the case in chapter 1, the asymptotic behaviour of the chain is totally determined by the spectral properties of the matrix P. However, we must now consider the general situation, not only the case of simple eigenvalues.

4.2 Some standard results from linear algebra

This section includes some elementary results from linear algebra as they can be found in [32, 33]. Denote by $\mathbb{M}_d(\mathbb{C})$ the set of $d \times d$ matrix with complex entries. With every matrix $A \in \mathbb{M}_d(\mathbb{C})$ and $\lambda \in \mathbb{C}$ associate the spaces

$$\begin{aligned} D_\lambda(A) &= \{\mathbf{v} \in \mathbb{C}^d : A\mathbf{v} = \lambda\mathbf{v}\} \\ &= \ker(A - \lambda I_d) \\ D^\lambda(A) &= \{\mathbf{v} \in \mathbb{C}^d : (A - \lambda I_d)^k \mathbf{v} = 0, \text{ for some } k \in \mathbb{N}\} \\ &= \cup_{k\in\mathbb{N}} \ker(A - \lambda I_d)^k, \end{aligned}$$

termed respectively right and **generalised right eigenspace** associated with the value λ. Obviously $D_\lambda(A) \subseteq D^\lambda(A)$ and if $D_\lambda(A) \neq \{0\}$ then λ is an **eigenvalue** of A.

Proposition 4.2.1 *For any $A \in \mathbb{M}_d(\mathbb{C})$, the following statements are equivalent:*

1. *λ is an eigenvalue of A,*
2. *$D_\lambda(A) \neq \{0\}$,*
3. *$D^\lambda(A) \neq \{0\}$,*
4. $\mathrm{rank}(A - \lambda I_d) < d$
5. *$\chi_A(\lambda) \equiv \det(A-\lambda I_d) = b(\lambda_1-\lambda)^{a_1} \cdots (\lambda_s-\lambda)^{a_s} = 0$ for some integer s, $0 < s \leq d$ and some positive integers $a_1,\ldots,a_s$, called the* ***algebraic multiplicities*** *of the eigenvalues $\lambda_1,\ldots,\lambda_s$.*

Proof: Equivalence of 1 and 2 is just the definition of the eigenvalue. The implication 2 ⇒ 3 is trivial since $D_\lambda(A) \subseteq D^\lambda(A)$. To prove equivalence, suppose conversely that a vector $v \neq 0$ lies in $D^\lambda(A)$; let $k \geq 1$ be the smallest integer such that $(A - \lambda I_d)^k v = 0$. Then the vector $w = (D - \lambda I_d)v$ is a non-zero element of $D_\lambda(A)$. To prove the equivalence of 4 and 5, note that by the equivalence of 1 and 2, λ is an eigenvalue if and only if $\ker(A-\lambda I_d)$ contains a non-zero element, i.e. $A - \lambda I_d$ is not injective which happens if and only if it is not invertible. The last statement is then equivalent to 4 and 5. □

We call **spectrum** of A the set $\mathrm{spec}(A) = \{\lambda_1, \ldots, \lambda_s\} = \{\lambda \in \mathbb{C} : A - \lambda - I_d$ is not invertible$\}$. For $\lambda \in \mathrm{spec}(A)$, we call **geometric multiplicity** of λ the dimension $g_\lambda = \dim D_\lambda(A)$.

Proposition 4.2.2 *Let $A \in \mathbb{M}_d(\mathbb{C})$.*

1. *There exists a unique polynomial $m_A \in \mathbb{C}[x]$, of minimal degree and leading numerical coefficient equal to 1 such that $m_A(A) = 0$. This polynomial is termed **minimal polynomial** of A.*
2. *m_A divides any polynomial $p \in \mathbb{C}[x]$ such that $p(A) = 0$. In particular, it divides the characteristic polynomial χ_A.*
3. *The polynomials χ_A and m_A have the same roots (possibly with different multiplicities).*
4. *If λ is a r-uple root of m_A then $D^\lambda(A) = \ker(A - \lambda I_d)^r$. In that case, $c_\lambda = r$ is the **generalised multiplicity** of λ, while the **algebraic multiplicity** is expressed as $a_\lambda = \dim D^\lambda(A) = \dim\ker(A - \lambda I_d)^{c_\lambda}$ and the minimal polynomial reads $m_A(\lambda) = (\lambda_1 - \lambda)^{c_{\lambda_1}} \cdots (\lambda_s - \lambda)^{c_{\lambda_s}}$.*

Proof: To prove 1 and 2, note that there exists a polynomial $p \in \mathbb{C}[x]$ such that $p(A) = 0$ (for instance consider the characteristic polynomial). Hence there exists also a polynomial of minimal degree m. Now, if $p(A) = 0$ perform division of p by m to write $p = qm + r$, where r is either the zero polynomial or has degree less than m. Now $r(A) = p(A) - q(A)m(A) = 0 - q(A)0 = 0$ and since m was chosen of minimal degree, then $r = 0$. Thus every polynomial vanisihing on A is divisible by m. In particular if m and m' are both minimal polynomials, they can differ only by a scalar factor. Fixing the leading coefficient of m to 1, uniquely determines $m \equiv m_A$.

To prove 3 note that if λ is a root of χ_A, i.e. is an eigenvalue of A, then there is a non-zero vector v with $Av = \lambda v$. Hence $0 = m_A(A)v = m(\lambda)v$, implying $m(\lambda) = 0$. Conversely, every root of m_A is a root of χ_A because m_A divides the characteristic polynomial.

To conclude, suppose that there is a vector $v \in D^\lambda(A)$ such that $w = (A - \lambda I_d)^r v \neq 0$. Write $m_A(x) = q(x)(x-\lambda)^r$. Then q and $(x-\lambda)^{n-r}$ are coprime; hence there are polynomials f and g with $f(x)q(x) + g(x)(x-\lambda)^{n-r} = 1$. Consequently,

$$\begin{aligned} w &= f(A)q(A)w + g(A)(A - \lambda I_d)^{n-r} w \\ &= f(A)m(A)v + g(A)(A - \lambda I_d)^n v \\ &= f(A)0 + g(A)0 \\ &= 0, \end{aligned}$$

which is a contradiction. □

Another possible characterisation of c_λ is $c_\lambda = \min\{n \in \mathbb{N} : \ker(A - \lambda I_d)^n = \ker(A-\lambda I_d)^{n+1}\}$. Moreover, the vector space $\mathbb{C}^d$ decomposes into the direct sum:

$$\mathbb{C}^d = \oplus_{\lambda \in \mathsf{spec}(A)} D^\lambda(A) = \oplus_{\lambda \in \mathsf{spec}(A)} \ker(A - \lambda I_d)^{c_\lambda}.$$

Definition 4.2.3 Let $A \in \mathbb{M}_d(\mathbb{C})$ and $\lambda \in \mathsf{spec}(A)$.

- If $c_\lambda = 1$ (i.e. $D^\lambda = D_\lambda$) then the eigenvalue λ is called **semisimple**.
- If $a_\lambda = 1 = \dim D^\lambda$ then the eigenvalue λ is called **simple**.

Example 4.2.4

- Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then $\mathsf{spec}(A) = \{0\}$ and $a_0 = c_0 = 2$ while $g_0 = 1$.
- Let $B = \begin{pmatrix} b & & \\ & \ddots & \\ & & b \end{pmatrix} \in \mathbb{M}_d(\mathbb{C})$. Then $\mathsf{spec}(B) = \{b\}$ and $a_b = d, c_b = 1$ while $g_b = d$.
- Let $C = \begin{pmatrix} b & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & b \end{pmatrix} \in \mathbb{M}_d(\mathbb{C})$. Then $\mathsf{spec}(C) = \{b\}$ and $a_b = c_b = d$ while $g_b = 1$.

Denote by E_λ the **spectral projector** on D^λ. Since the space admits the direct sum decomposition $\mathbb{C}^d = \oplus_{\lambda \in \mathsf{spec}(A)}$ it follows that spectral projections form a **resolution of the identity** $I_d = \sum_{\lambda \in \mathsf{spec}(A)} E_\lambda$.

Theorem 4.2.5 *Let $A \in \mathbb{M}_d(\mathbb{C})$ and U an open set in $\mathbb{C}$ (not necessarily connected), containing* $\mathsf{spec}(A)$. *Denote by $C^\omega(A)$ the set of analytic functions on U (i.e. indefinitely differentiable functions developable in Taylor series around any point of U) and let $f \in C^\omega_A$. Then,*

$$f(A) = \sum_{\lambda \in \mathsf{spec}(A)} \sum_{i=0}^{c_\lambda - 1} \frac{(A - \lambda I_d)^i}{i!} f^{(i)}(\lambda) E_\lambda.$$

Proof: See [7] pp. 555–559. □

Exercise 4.2.6 Choose any circuit B in U not passing through any point of the spectrum. Use analyticity of f, extended to matrices, to write by virtue of the residue theorem:

$$f(A) = \frac{1}{2\pi i}\oint_B f(\lambda)(\lambda I_d - A)^{-1}\, d\lambda.$$

The operator $R_A(\lambda) = (\lambda I_d - A)^{-1}$, defined for λ in an open domain of $\mathbb{C}$, is called **resolvent** of A.

Remark: Suppose that all the eigenvalues of a matrix A are semisimple ($c_\lambda = 1$ for all $\lambda \in \mathsf{spec}(A)$), then $f(A) = \sum_{\lambda\in\mathsf{spec}(A)} f(\lambda)E_\lambda$. In particular, $A^k = \sum_{\lambda\in\mathsf{spec}(A)} \lambda^k E_\lambda$.

Lemma 4.2.7 *Suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence of analytic functions in $C^\omega(A)$. The sequence of matrices $(f_n(A))_{n\in\mathbb{N}}$ converges if and only if, for all $\lambda \in \mathsf{spec}(A)$ and all integers k with $0 \le k \le c_\lambda - 1$, the numerical sequences $(f_n^{(k)}(\lambda))_{n\in\mathbb{N}}$ converge. If $f \in C^\omega(A)$, then $\lim_{n\to\infty} f_n(A) = f(A)$ if and only if, for all $\lambda \in \mathsf{spec}(A)$ and all integers k with $0 \le k \le c_\lambda - 1$, we have $\lim_{n\to\infty} f_n^{(k)}(\lambda) = f^{(k)}(\lambda)$.*

Proof: See [7] pp. 559–560. □

Proposition 4.2.8 *Let $A \in \mathbb{M}_d(\mathbb{C})$. The sequence of matrices $(\frac{1}{n}\sum_{p=1}^n A^p)_{n\in\mathbb{N}}$ converges if and only if $\lim_{n\to\infty} \frac{A^n}{n} = 0$.*

Proof: Define for all $n \ge 1$ and $z \in \mathbb{C}$, $f_n(z) = \frac{1}{n}\sum_{p=1}^n z^p$ and $g_n(z) = \frac{z^n}{n}$. We observe that the sequence $(f_n(\lambda))_{n\in\mathbb{N}}$ converges if and only if $|\lambda| \le 1$, the latter being equivalent to $\lim_{n\to\infty} g_n(\lambda) = 0$. For $k > 0$, the sequence $(f_n^{(k)}(\lambda))_{n\in\mathbb{N}}$ converges if and only if $|\lambda| < 1$, the latter being equivalent to $\lim_{n\to\infty} g_n^{(k)}(\lambda) = 0$. We conclude by theorem 4.2.5 and lemma 4.2.7. □

Definition 4.2.9 Let $A \in \mathbb{M}_d(\mathbb{C})$. We call **spectral radius** of A, the quantity denoted by $\mathsf{sr}(A)$ and defined by

$$\mathsf{sr}(A) = \max\{|\lambda| : \lambda \in \mathsf{spec}(A)\}.$$

Corollary 4.2.10 *For all $A \in \mathbb{M}_d(\mathbb{C})$, $\lim_{n\to\infty} \frac{A^n}{n} = 0$ if and only if the two following conditions are fulfilled:*

- $\mathsf{sr}(A) \le 1$, *and*
- *all peripheral eigenvalues of A, i.e. all $\lambda \in \mathsf{spec}(A)$ with $|\lambda| = 1$, are semisimple.*

4.3 Positive matrices

Definition 4.3.1 Let $x \in \mathbb{R}^d$ and $A \in \mathbb{M}_d(\mathbb{R})$. The vector x is called **positive**, and denoted $x \geq 0$, if for all $i = 1, \dots, d$, $x_i \geq 0$. It is called **strictly positive** if for all $i = 1, \dots, d$, $x_i > 0$. The matrix A is called positive (resp. strictly positive) if viewed as a vector of $\mathbb{R}^{d^2}$ is positive (resp. strictly positive). For arbitrary $x \in \mathbb{R}^d$ and $A \in \mathbb{M}_d(\mathbb{R})$, we denote by $|x|$ and $|A|$ the vector and matrix whose elements are given by $|x|_i = |x_i|$ and $|A|_{ij} = |A_{ij}|$.

Remark: The above positivity of elements must not be confused with positivity of quadratic forms associated with symmetric matrices. For example $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \geq 0$ but there exist $x \in \mathbb{R}^2$ such that $(x, Ax) \not\geq 0$.

Proposition 4.3.2

$$\begin{aligned} A \geq 0 \quad &\Leftrightarrow \quad [x \geq 0 \Rightarrow Ax \geq 0] \\ A > 0 \quad &\Leftrightarrow \quad [x \geq 0 \textit{ and } x \neq 0 \Rightarrow Ax \geq 0]. \end{aligned}$$

Exercise 4.3.3 Prove the previous proposition.

Definition 4.3.4 A matrix A is **reducible** if there exists a permutation matrix S such that $SAS^{-1} = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$. Otherwise, A is called **irreducible**.

Proposition 4.3.5 *For a positive matrix $A \in \mathbb{M}_d(\mathbb{R})$, the following statements are equivalent:*

- *A is irreducible,*
- *$(I_d + A)^{d-1} > 0$, and*
- *for all pairs of integers (i, j) with $1 \leq i, j \leq d$, there exists an integer $k = k(i, j)$ such that $(A^k)_{ij} > 0$.*

Exercise 4.3.6 Prove the previous proposition.

Theorem 4.3.7 (Brower's fixed point theorem) *Let $C \subseteq \mathbb{R}^d$ be a non-empty, closed, bounded,and convex set, and let $f : C \to C$ be continuous. Then, there exists a $x \in C$ such that $f(x) = x$.*

Proof: See [3], p. 176 for instance. □

Theorem 4.3.8 (Weak Perron-Frobenius theorem) *Let $A \in \mathbb{M}_d(\mathbb{R})$ be a positive matrix. Then $\mathsf{sr}(A)$ is an eigenvalue of A associated with a positive eigenvector v.*

Proof: Let λ be an eigenvalue of maximal modulus, i.e. with $|\lambda| = \mathsf{sr}(A)$. We can always normalise the eigenvector v of λ so that $\|v\|_1 = 1$. Then,

$$\mathsf{sr}(A)|v| = |\lambda v| = |Av| \leq A|v|.$$

If

$$C = \{x \in \mathbb{R}^d : x \geq 0, \sum_{i=1}^{d} x_i = 1, Ax \geq \mathsf{sr}(A)x\},$$

then C is closed, convex, non-empty (since it contains $|v|$), and bounded (since $0 \leq x_j \leq 1$ for all $i = 1, \ldots, d$). We distinguish two cases:

1. There exists a $x \in C$ such that $Ax = 0$. Then $\mathsf{sr}(A)x \leq Ax = 0$; consequently $\mathsf{sr}(A) = 0$ and the theorem is proved.
2. For all $x \in C$, $Ax \neq 0$. Define then $f : C \to \mathbb{R}^d$ by $f(x) = \frac{1}{\|Ax\|_1} Ax$. We observe that
   - For all $x \in C$, $f(x) \geq 0$, $\|f(x)\|_1 = 1$ and f continuous.
   - $Af(x) = \frac{1}{\|Ax\|_1} AAx \geq \frac{\mathsf{sr}(A)}{\|Ax\|_1} Ax = \mathsf{sr}(A) f(x)$. Therefore $f(C) \subseteq C$.
   - The theorem 4.3.7 ensures that there exists $y \in C : f(y) = y$.
   - $y \geq 0$ (since $y \in C$) and $f(y) = y$, therefore, y is an eigenvector associated with the eigenvalue $r = \|Ay\|_1$.
   - Hence $Ay = ry \geq \mathsf{sr}(A)y$, the last inequality holding since $y \in C$. Therefore $r \geq \mathsf{sr}(A)$.
   - Hence $r = \mathsf{sr}(A)$.

□

For $r \geq 0$ denote by

$$C_r = \{x \in \mathbb{R}^d : x \geq 0, \|x\|_1 = 1, Ax \geq rx\}.$$

Obviously every C_r is a convex and compact set.

Lemma 4.3.9 *Let $A \in \mathbb{M}_d(\mathbb{R})$ be a positive irreducible matrix; let $r \geq 0$ and $x \in \mathbb{R}^d$, with $x \geq 0$, be such that $Ax \geq rx$ and $Ax \neq rx$. Then there exists $r' > r$ such that $C_{r'} \neq \emptyset$.*

Proof: Let $y = (I_d + A)^{d-1}x$. Since the matrix A is irreducible and $x \geq 0$, thanks to the proposition 4.3.5, we get $y > 0$. For the same reason, $Ay - ry = (I_d + A)^{d-1}(Ax - rx) > 0$. On defining $r' = \min_j \frac{(Ay)_j}{y_j}$ we get $r' > r$. But then $Ay \geq r'y$ so that $C_{r'}$ contains the vector $y/\|y\|_1$. □

Lemma 4.3.10 *Let $A \in \mathbb{M}_d(\mathbb{R})$ be a positive irreducible matrix. If $x \in \mathbb{R}$ is an eigenvector of A and $x \geq 0$ then $x > 0$.*

Proof: Given such a vector x with $Ax = \lambda x$, we have that $\lambda \geq 0$. Then $x = \frac{1}{(1+\lambda)^{n-1}}(I_d + A)^{d-1}x > 0$, thanks to the proposition 4.3.5. □

Lemma 4.3.11 *Let $A, B \in \mathbb{M}_d(\mathbb{R})$ be matrices, with A irreducible and $|B| \leq A$. Then $\mathsf{sr}(B) \leq \mathsf{sr}(A)$. In the case of equality of the spectral radii, we have further:*

- *$|B| = A$, and*
- *for every eigenvector x of B associated with an eigenvalue of modulus $\mathsf{sr}(A)$, the vector $|x|$ is an eigenvector of A associated with $\mathsf{sr}(A)$.*

Proof: If λ is an eigenvalue of B of modulus $\mathsf{sr}(B)$ and x is the corresponding normalised eigenvector, then $\mathsf{sr}(B)|x| \leq |B||x| \leq A|x|$, so that $C_{\mathsf{sr}(B)} \neq \emptyset$. Hence, $\mathsf{sr}(B) \leq R = \mathsf{sr}(A)$.

In case of equality honding, then $|x| \in C_{\mathsf{sr}(A)}$ and $|x|$ is an eigenvector: $A|x| = \mathsf{sr}(A)|x| = \mathsf{sr}(B)|x| \leq |B||x|$ Hence, $(A - |B|)[x[\leq 0$, but since $|x| > 0$, from lemma 4.3.10, and $A - |B| \geq 0$, the equality $|B| = A$ follows. □

Theorem 4.3.12 (Strong Perron-Frobenius theorem) *Let $A \in \mathbb{M}_d(\mathbb{R})$ be a positive irreducible matrix. Then $\mathsf{sr}(A)$ is a simple eigenvalue of A associated with a strictly positive eigenvector v. Moreover $\mathsf{sr}(A) > 0$.*

Proof: Additionally, if λ is an eigenvalue associated with an eigenvectors v of unit norm, then $|v|$ is a vector of $C_{|\lambda|}$; in particular, the set $C_{\mathsf{sr}(A)}$ is non-empty. Conversely, if C_r is non-empty, then for $v \in C_r$:

$$r = r\|v\|_1 \leq \|Av\|_1 \leq \|A\|_1\|v\|_1 = \|A\|_1,$$

and therefore $r \leq \|A\|_1$. Further, the map $r \mapsto C_r$ is non-increasing with respect to inclusions and is "left continuous", in the sense $C_r = \cap_{s<r} C_s$. Define then $R = \sup\{r : C_r \neq \emptyset\}$; subsequently, $R \in [\mathsf{sr}(A), \|A\|_1]$. Decreasing with respect to inclusions implies that $r < R \Rightarrow C_r \neq \emptyset$.

If $x > 0$ of norm 1, then $Ax \geq 0$ and $Ax \neq 0$ since $A \geq 0$ and irreducible. From lemma 4.3.9 follws that $R > 0$; consequently, the set C_R being the intersection of a totally ordered family of non-empty compact sets is non-empty. For $x \in C_R$, the lemma 4.3.9 guarantees then that x is an eigenvector of A associated with the eigenvalue R. Observing that $R \geq \mathsf{sr}(A)$ implies then that $R = \mathsf{sr}(A)$, showing that $\mathsf{sr}(A)$ is the eigenvalue associated with the eigenvector x, and $\mathsf{sr}(A) > 0$. Lemma 4.3.10 guarantees then that $x > 0$.

It remains to show simplicity of the eigenvalue $\mathrm{sr}(A)$. The characteristic polynomial $\chi_A(\lambda)$ is seen as the composition of a d-linear form (the determinant) with polynomial vector valued functions (the columns of $\lambda I_d - A$). Now, if ϕ is a p-linear form and if $V_1(\lambda),\ldots,V_p(\lambda)$ are p polynomial vector-valued functions, then the polynomial $p(\lambda) = \phi(V_1(\lambda),\ldots,V_p(\lambda))$ has derivative:

$$p'(\lambda) = \phi(V_1'(\lambda),\ldots,V_p(\lambda)) + \ldots + \phi(V_1(\lambda),\ldots,V_p'(\lambda)).$$

Denoting $(\mathbf{e}^1,\ldots,\mathbf{e}^d)$ the canonical basis of $\mathbb{R}^d$ and writing A in terms of its column vectors $A = [a_1,\ldots,a_d]$, one obtains:

$$\begin{aligned}\chi_A'(\lambda) &= \det([\mathbf{e}^1, a_2,\ldots,a_d]) + \det([a_1,\mathbf{e}^2,\ldots,a_d]) + \ldots + \det([a_1,a_2,\ldots,\mathbf{e}^d]) \\ &= \sum_{j=1}^{d} \chi_{A_j}(\lambda),\end{aligned}$$

where $A_j \in \mathbb{M}_{d-1}(\mathbb{R})$ is the matrix obtained from A be deleting the j-th comumn and row. (This formula is obtained by developing the determinants with respect the j-th column.)

Denote by $B_j \in \mathbb{M}_d(\mathbb{R})$ the matrix obtained from A by replacing the j-th row and column by zeroes. This matrix is block-diagonal with two non-zero blocks and a block $0 \in \mathbb{M}_1(\mathbb{R})$. The two non-zero blocks can be put together by permutation of rows and columns to reconstruct A_j. Hence the eigenvalues of B_j are those of A_j and 0; therefore $\mathrm{sr}(B_j) = \mathrm{sr}(A_j)$. Further $|B_j| \geq A$ but $|B_j| \neq A$ because A is irreducible while B_j is block-diagonal, hence reducible. It follows by lemma 4.3.11 that $\mathrm{sr}(B_j) < \mathrm{sr}(A)$. Hence $\chi_{A_j}(\mathrm{sr}(A)) \neq 0$ with the same sign as $\lim_{t\to\infty} \chi_{A_j}(t) > 0$. Therefore, $\chi_A'(\mathrm{sr}(A)) > 0$ showing that $\mathrm{sr}(A)$ is a simple root.
□

Let $A \in \mathbb{M}_d(\mathbb{R})$ be a positive irreducible matrix. Denote by $\lambda_1, \cdots \lambda_s$ the eigenvalues of modulus $\mathrm{sr}(A)$. If $s = 1$, the matrix is called **primitive**, else **cyclic** of order s.

Exercise 4.3.13 Show that $A > 0$ implies that A is primitive.

Exercise 4.3.14 (Characterisation of the peripheral spectrum) $A \in \mathbb{M}_d(\mathbb{R})$ be a positive matrix. If $p = \sharp\{\lambda \in \mathrm{spec}(A) : |\lambda| = \mathrm{sr}(A)\}$, then $\{\lambda \in \mathrm{spec}(A) : |\lambda| = \mathrm{sr}(A)\} = \mathrm{sr}(A)\mathcal{U}_p$, where $\mathcal{U}_p$ is the group of p^{th} roots of unity. *Hint:* Associate the peripheral spectrum of A with its periodicity properties.

The spectrum of any matrix A decomposes into three disjoint components:

$$\mathrm{spec}(A) = \Sigma_{\bullet} \sqcup \Sigma_{\odot} \sqcup \Sigma_{\oslash},$$

where $\Sigma_\bullet = \{\mathsf{sr}(A)\}$ is the **maximal eigenvalue spectrum**, $\Sigma_\odot = \{\lambda \in \mathsf{spec}(A) : |\lambda| = \mathsf{sr}(A)\} \setminus \{\mathsf{sr}(A)\}$ is the **peripheral spectrum** and $\Sigma_\oslash = \{\lambda \in \mathsf{spec}(A) : |\lambda| < \mathsf{sr}(A)\}$ is the **contracting spectrum**. With the exception of $\Sigma_\bullet$ the other parts of the spectrum can be empty.

Lemma 4.3.15 *If P is a stochastic matrix, then* $\mathsf{sr}(P) = 1$.

Exercise 4.3.16 Prove the previous lemma.

Theorem 4.3.17 (Convergence theorem for Markov chains) *Let P be a stochastic matrix and E_1 the spectral projector to the eigenspace associated with the eigenvalue 1.*

1. *There exists a positive real constant K_1 such that for all $n \geq 1$,*

$$\|\frac{1}{n}\sum_{k=0}^{n-1} P^k - E_1\| \leq \frac{K_1}{n}$$

and, for every $\mu \in \mathcal{M}_1(\mathcal{X})$ (thought as a row vector),

$$\|\frac{1}{n}\sum_{k=0}^{n-1} \mu P^k - \mu E_1\| \leq \frac{K_1}{n}.$$

2. *If $\Sigma_\odot = \emptyset$, then there exist constants $K_2 < \infty$ and $0 < r < 1$ such that for all $n \in \mathbb{N}$,*

$$\|P^n - E_1\| \leq K_2 r^n.$$

3. *If the eigenvalue 1 is simple, then E_1 defines a unique invariant probability π (i.e. verifying $\pi P = \pi$) by $E_1 f = \pi(f)1$ for all $f \in b\mathcal{X}$.*

Proof: We only give the main idea of the proof, technical details are tedious but without any particular difficulty.

Since P is stochastic, by lemma 4.3.15 we have that $\mathsf{sr}(P) = 1$. Now $\|P\|_\infty = \max_{x\in\mathbb{X}} \sum_{y\in\mathbb{X}} P_{xy} = 1$ and for the same reason $\|P^n\|_\infty = 1$, for all n. Hence $\lim_{n\to\infty} \frac{P^n}{n} = 0$. Therefore, by corollary 4.2.10 the peripheral values are semi-simple.

Let us prove first assertion 2. Since $\Sigma_\odot = \emptyset$, it follows that $r' = \max\{|\lambda| : \lambda \in \mathsf{spec}(A), \lambda \neq 1\} < 1$. For every r in the spectral gap, i.e. $r' < r < 1$, we have from theorem 4.2.5, that

$$\|\sum_{\lambda\in\Sigma_\oslash}\sum_{i=0}^{c_\lambda - 1} \frac{(P-\lambda I_d)^i}{i!}\frac{d^i}{d\lambda^i}(\lambda^n)E_\lambda\| \leq K_1 r^n.$$

Hence $\|P^n - E_1\|_\infty \le K_1 r^n$.

Prove now assertion 1. For all $\lambda \in \Sigma_\odot$ there exists a $\theta = \theta(\lambda) \in]0, \pi[$ such that $\lambda = \exp(i\theta)$. Then $\epsilon = \min\{|\theta(\lambda)| : \lambda \in \Sigma_\odot\} > 0$. Therefore, $|\sum_{k=0}^{n-1} \lambda^k| \le \frac{1}{\sin(\epsilon/2)}$. Consequently,

$$\|\frac{1}{n}\sum_{k=0}^{n-1} P^k - E_1\|_\infty \le \frac{1}{n\sin(\epsilon/2)}\|\sum_{\lambda\in\Sigma_\otimes} E_\lambda\|_\infty + \frac{K_1'}{n}\sum_{k=0}^{n-1} r^k.$$

To prove 3, just remark that E_1 is a projector onto an one-dimensional space, hence it is a stable matrix (i.e. all its rows are equal) its row summing up to 1. □

4.4 Complements on spectral properties

The previous results establish that the spectrum of the stochastic matrix is contained in the unit disk of $\mathbb{C}$, the value 1 always being an eigenvalue. On the other hand, finer knowledge of the spectral values can be used to improve the speed of convergence that has been established in theorem 4.3.17. Besides, the study of the locus of the spectral values of an arbitrary stochastic matrix constitutes an interesting mathematical problem first posed by Kolmogorov. As an experimental fact stemming from numerical simulations, we know that the spectral values of stochastic matrices concentrate on a set strictly contained in the unit disk. It is therefore important to have better estimates of the localisation of the spectrum of the stochastic matrix within the unit disk.

4.4.1 Spectral constraints stemming from algebraic properties of the stochastic matrix

The next results [31] improve the localisation properties of the eigenvalues.

Proposition 4.4.1 (Gershgorin disks) *Let $A \in \mathcal{M}_d(\mathbb{R}_+)$ and define $a_x = \sum_{y\in\mathbb{X}:y\neq x} A_{xy}$ for $x \in \mathbb{X}$. Then, each eigenvalue of A is contained in at least one of the circular disks*

$$\mathbb{D}_x = \{\zeta \in \mathbb{C} : |\zeta - A_{xx}| \le a_x\},\ x \in \mathbb{X}.$$

Proof: Let $\lambda \in \mathsf{spec}(A)$. Then $\lambda v = Av$ for some vector $v \neq 0$ with $\|v\|_\infty = 1$. Let $x \in \mathbb{X}$ be such that $|v_x| = 1$. Since $|v_y| \le 1$ for all $y \in \mathbb{X}$, it follows that $|\lambda - A_{xx}| = |(\lambda - A_{xx})v_x| = |\sum_{y:y\neq x} A_{xy} v_y| \le \sum_{y:y\neq x} A_{xy}|v_y| \le a_x$. □

In the sequel we denote by SM_d the set of $d \times d$ stochastic matrices.

Corollary 4.4.2 *If $P \in \mathsf{SM}_d$, then each eigenvalue of P is contained in at least one of circular disks*

$$\mathbb{D}_i = \{\zeta \in \mathbb{C} : |\zeta - A_{xx}| \leq 1 - A_{xx}\} \ x \in \mathbb{X}.$$

Proposition 4.4.3 *Suppose the matrix $A \in \mathscr{M}_d(\mathbb{R}_+)$ has spectral radius $\mathsf{sr}(A) = r$; assume further that the eigenvalue r corresponds to the left eigenvector u (it will have necessarily strictly positive components: $u_x > 0$, for all $x \in \mathbb{X}$). Denote by e the vector with unit components, $S = \{w \in \mathbb{R}_+^d : \langle u \mid w \rangle = 1\}$, $T = \{v \in \mathbb{R}_+^d : \langle v \mid e \rangle = 1\}$, and $H = S \times T$. Let*

$$m = \inf_{(w,v) \in H} \langle v \mid Aw \rangle \ \text{and} \ M = \sup_{(w,v) \in H} \langle v \mid Aw \rangle.$$

Then, each eigenvalue $\lambda \neq r$ satisfies

$$|\lambda| \leq \min(M - r, r - m).$$

If further $m > 0$ (which occurs if and only if the matrix A has all its elements strictly positive) then the previous bound implies

$$|\lambda| \leq \frac{M - m}{M + m} r.$$

Proof: Define the matrices $B = A - me \otimes u^t$ and $C = Me \otimes u^t - A$. They are both positive since

$$\langle v \mid Bw \rangle = \langle v \mid Aw \rangle - m \langle v \mid e \otimes u^t \ w \rangle = \langle v \mid Aw \rangle - m \langle v \mid e \rangle \langle u \mid w \rangle \geq 0,$$

and similarly for C. Let $\lambda \neq r$ be an eigenvalue of A and z the corresponding right eigenvector, therefore $\langle u \mid w \rangle = 0$. Now

$$Bz = Az - me \otimes u^t \ w = Aw - m \langle u \mid w \rangle e = Aw = \lambda w,$$

and similarly $Cw = -\lambda w$. Hence, w is also an eigenvector of B and C corresponding respectively to the eigenvalues λ and $-\lambda$. On the other hand

$$u^t B = u^t A - m u^t e \otimes u^t = r u^t - m \langle u \mid e \rangle u^t = (r - m) u^t,$$

and similarly $u^t C = (M - r) u^t$. Since the eigenvector u has all its components strictly positive, it follows that $r - m = \mathsf{sr}(B)$ and $M - r = \mathsf{sr}(C)$. Hence $|\lambda| \leq \min(M - r, r - m)$. Now, if $m > 0$, two cases can occur:

– Either $M - r \geq r - m$, implying $M + m \geq 2r$ and

$$|\lambda| \leq r - m = \frac{2r - m - M}{2} + \frac{M - m}{2}.$$

Hence $|\lambda| \leq \frac{(M-m)/2}{1 + \frac{M+m-2r}{2|\lambda|}}$ and, since $|\lambda| < r$, finally

$$|\lambda| \leq \frac{(M - m)/2}{1 + \frac{M+m-2r}{2r}} = r \frac{M - m}{M + m}.$$

– Or $M-r<r-m$ hence $M+m<2r$, implying that

$$|\lambda| \le M-r < \frac{M-r}{2}+\frac{r-m}{2} = \frac{M-m}{2} = \frac{M-m}{2r}r < \frac{M-m}{M+m}r.$$

□

Note finally, the trivial observation that the spectrum of a stochastic matrix is symmetric around the real axis since the characteristic polynomial has only real coefficients.

4.4.2 Spectral constraints stemming from convexity properties of the stochastic matrix

Let us first recall some standard results concerning convex sets.

Definition 4.4.4 Let $A \subseteq \mathbb{V}$ be an arbitrary subset of a vector space $\mathbb{V}$. The **convex hull** of A, denoted by $\operatorname{co} A$, consists of all convex combinations from A, i.e.

$$\operatorname{co} A = \{x \in \mathbb{V} : \exists N \in \mathbb{N}^*, \exists v_i \in \mathbb{V}, \exists \alpha_i \in \mathbb{R}_+ (0 \le i \le N), \sum_{i=1}^{N} \alpha_i = 1, \text{ such that } x = \sum_{i=1}^{N} \alpha_i v_i\}.$$

Theorem 4.4.5 (Carathéodory's convexity theorem) *In an n-dimensional affine linear space, every vector in the convex hull of a nonempty set A can be written as a convex combination using no more than $n+1$ vectors from A.*

Exercise 4.4.6 Prove theorem 4.4.5. Hint: assume on the contrary that $k > n+1$ vectors are needed to express vectors in the convex hull and arrive in a contradiction.

Example 4.4.7 Let $\mathsf{PV}_d = \{p \in \mathbb{R}_+^d : \sum_{i=1}^d p_i = 1\}$; obviously any element of PV_d defines a probability measure on a finite set of cardinality d. The set PV_d is convex and although is defined as a subset of $\mathbb{R}_+^d$, the condition $\sum_{i=1}^d p_i = 1$ constrains this set to be a subset of an affine linear space of dimension $d-1$; therefore its dimension is $d-1$. The set of the d canonical unit vectors $\mathsf{extr}(\mathsf{PV}_d) = \{e_1, \ldots, e_d\}$ of $\mathbb{R}^d$ contains (the sole) elements of PV_d that cannot be written as a non-trivial convex combinations of others; they are the **extremal** elements of PV_d. All other elements of PV_d can be written as convex combinations $p = \sum_{i=1}^d p_i e_i$, therefore $\mathsf{PV}_d = \mathsf{co}(\mathsf{extr}(\mathsf{PV}_d))$. The latter means that the convex set PV_d is in fact a **simplex**, i.e. coincides with the convex hull of its extremal points.

Let $\mathbb{X}$ be a finite set with $\mathrm{card} X = d$; denote by $\mathsf{SM}_d = \{P \in \mathscr{M}_d(\mathbb{R}) : \sum_{y\in\mathbb{X}} P_{xy} = 1, \forall x \in \mathbb{X}\}$ the set of stochastic matrices. The stochastic matrices of $\mathsf{DM}_d = \{P \in \mathscr{M}_d(\{0,1\}) : \sum_{y\in\mathbb{X}} P_{xy} = 1, \forall x \in \mathbb{X}\} \subseteq \mathsf{SM}_d$ are called **deterministic transition matrices**. The reason for this terminology is obvious: deterministic transition matrices have precisely one 1 in every row, i.e. for every $D \in \mathsf{DM}_d$ and every $x \in \mathbb{X}$, there exists exactly one $y_{D,x} \in \mathbb{X}$ such that $D_{xy} = \delta y, y_{D,x}$. Therefore, there exists a bijection between DM and the set of functions $\{\beta : \mathbb{X} \to \mathbb{X}\}$, established by choosing for every D the map $\beta_D(x) = y_{D,x}$.

Proposition 4.4.8 *The set SM_d is convex. Its extremal points are the deterministic matrices $\mathrm{extr}(\mathsf{SM}_d) = \mathsf{DM}_d$. The set SM_d is not a simplex.*

Proof: Since $\mathsf{SM}_d = \mathsf{PV}_d^d$, convexity of SM_d follows from the convexity of every one of its rows. Moreover $\dim \mathsf{SM}_d = d(d-1)$.

To prove extremality of deterministic matrices, suppose on the contrary that for every $D \in \mathsf{DM}_d$ there exist $P, Q \in \mathsf{SM}_d$ and $\lambda \in]0,1[$ such that $D = \lambda P + (1-\lambda)Q$. Since D is deterministic,

$$\begin{aligned} 1 &= \lambda P_{x,\beta_D(x)} + (1-\lambda) Q_{x,\beta_D(x)}, \forall x \in \mathbb{X} \\ 0 &= \lambda P_{xy} + (1-\lambda) Q_{xy} \ \forall x \in \mathbb{X}, \forall y \in \mathbb{X} \setminus \{\beta_D(x)\}. \end{aligned}$$

Since $P_{xy}, Q_{xy} \in [0,1]$ for all x, y and $\lambda, (1-\lambda) \in]0,1[$, the above equations have a solution if and only if $P = Q = D$.

Finally since $\mathrm{card}\mathsf{DM}_d = d^d > d(d-1)+1$, it follows that $\mathsf{SM}_d \neq \mathrm{co}(\mathsf{DM}_d)$. □

A stochastic matrix is called **bi-stochastic** if both all its row- and column-sums equal 1; the set of $d \times d$ bistochastic matrices is denoted by BM_d.

It is worth noting that the map β_D is bijective if and only if D is a permutation matrix, i.e. has exactly one 1 in every row and every column. Therefore permutation matrices are deterministic bi-stochastic transition matrices. It is easy to show that the spectrum of deterministic transition matrices is contained in the set $\{0\} \cup \{\zeta \in \mathbb{C} : |\zeta| = 1\}$. (Hint: As a matter of fact, if D is a permutation matrix and the permutation has a cycle of lenght l, then the l roots of unity are in its spectrum. If D is a general deterministic transition matrix, the space $\mathbb{C}^d$ can be decomposed into a direct sum $F + F'$ so that $D\upharpoonright_F$ is a permutation, while $D\upharpoonright_{F'}$ is non-invertible.)

The extemal points of the set SM_d admit also a purely algebraic description. To describe it, we need a definition.

Definition 4.4.9 A linear map $A : \mathbb{C}^d \to \mathbb{C}^d$ is called a **lattice homomorphism** if

$|Av| = A|v|$ for all $v \in \mathbb{C}^d$, where $|v| = (|v_i|)_{i=1,\ldots,d}$. It is called a **lattice isomorphism** if A^{-1} exists and both A and A^{-1} are lattice homomorphisms.

Proposition 4.4.10 *A stochastic matrix $P \in \mathsf{SM}_d$ is a lattice homomorphism if and only if $P \in \mathsf{extr}(\mathsf{SM}_d) = \mathsf{DM}_d$.*

Proof: Exercise. □

Let $Z_k = \{\exp(\frac{il\pi}{k}), l = 0, \ldots, k-1\}$ denote the k-th roots of unity, for $k \in \mathbb{N}^*$. A first estimate of the spectrum of an arbitrary stochastic matrix is given by the following

Proposition 4.4.11 *For every $P \in \mathsf{SM}_d$, its spectrum verifies $\mathsf{spec}(P) \subseteq \mathsf{co}(\cup_{k=2}^{d} Z_k)$.*

Remark: Although the above result is not optimal, its proof is quite tricky in general. We give it below only in the simpler case where P is bistochastic since then the extremal points are the permutation matrices.

Partial proof: Every permutation matrix D is normal because $D^t D = I_d$. Now for any normal matrix N, it is known ([29] for instance) that the numerical range $\{\langle v | Nv \rangle, v \in \mathbb{C}^d, \|v\|_2 = 1\}$ equals $\mathsf{co}(\mathsf{spec}(N))$. On the other hand, the spectrum of d-dimensional permutation matrices is contained in $\cup_{k=2}^{d} Z_k$. Since permutation matrices are extremal points of the set of bistochastic matrices, any bistochastic matrix P admits a convex decomposition $P = \sum_{D \in \mathsf{BM}_d \cap \mathsf{DM}_d} \alpha_D D$ where $\alpha = (\alpha_D)$ is a probability on the set of permutation matrices [1]. Suppose now that $\lambda \in \mathsf{spec}(P)$ is an eigenvalue corresponding to the eigenvector u. Normalising v so that $\|v\|_2 = 1$, we have

$$\lambda = \lambda \langle v | v \rangle = \langle v | Pv \rangle = \sum_{D \in \mathsf{BM}_d \cap \mathsf{DM}_d} \alpha_D \langle v | Dv \rangle,$$

which implies that $\mathsf{spec}(P) \subseteq \mathsf{co}(\cup_{k=2}^{d} Z_k)$. □

Nevertheless, the numerical simulation, depicted in figure 4.1 below, demonstrates that the convex hull of the roots of unity provide us with an overestimate of the spectral set. An optimal estimate of the spectrum is given by a very tricky result obtained by Karpelevich [19], answering definitely the question asked by Kolmogorov. We present below the statement of the Karpelevich theorem in a shortened form due to Ito [17] and shown by this same author to be equivalent to the original formulation of [19].

1. Since the set of stochastic matrices is not a simplex, this decomposition is by no means unique.

Theorem 4.4.12 (Karpelevich [19], Ito [17]) *There exists a region $M_d \subseteq \mathbb{C}$ verifying:*

- *M_d is symmetric with respect to the real axis and contained in the unit disk.*
- *M_d intersects the unit circle at the points $\exp(2i\pi m/n)$ where m, n run over the relatively prime integers satisfying $0 \le m \le n \le d$.*
- *The boundary of M_d consists of these points and of curvilinear arcs connecting them in circular order.*
- *Let the endpoints of an arc be $\exp(2\pi i m_1/n_1)$ and $\exp(2\pi i m_2/n_2)$ (with $n_1 \le n_2$). Each of these arcs is given by the following parametric equation:*

$$\lambda^{n_2}(\lambda^{n_1} - s)^{[d/n_1]} = (1-s)^{[d/n_1]}\lambda^{n_1[d/n_2]},$$

where the real parameter s runs over the interval $[0,1]$.

Then for any $P \in \mathsf{SM}_d$, we have $\mathsf{spec}(P) \subseteq M_d$.

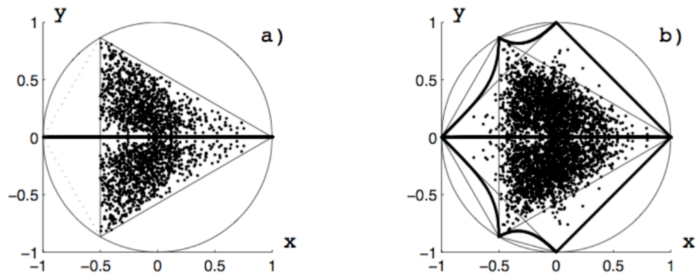

Figure 4.1: Distribution of spectral values for 3000 stochastic matrices of size d, randomly chosen in SM_d: a) Distribution for $d = 3$. The dotted lines depict the boundary of $\mathsf{co}(Z_2 \cup Z_3)$; obviously, this set overestimates the locus of spectral values. b) Same picture for $d = 4$; again we observe the overestimate. The optimal boundary of the locus, stemming from Karpelevich theorem [19], is curved and depicted as a thick curve in the figure; the locus of the spectral values is not convex any longer.

Exercise 4.4.13 The previous figure 4.1 is produced by "randomly choosing" 3000 matrices $P \in \mathsf{SM}_d$.

- What do you understand by the term "randomly choosen"?
- Propose an algorithm to "randomly chose" a $P \in \mathsf{SM}_d$. (Hint: can you construct a uniform probability measure on SM_d?)

Exercise 4.4.14 Since SM_d is convex and $\mathsf{extr}(\mathsf{SM}_d) = \mathsf{DM}_d$, every stochastic matrix P admits a (not necessarily) unique convex decomposition $P = \sum_{D \in \mathsf{DM}_d} \alpha_D D$ where $(\alpha_D)_D$ is a probability vector over DM_d.

- Using this decomposition and the deterministic mapping $\beta_D : \mathbb{X} \to \mathbb{X}$, uniquely defined from every D, express the matrix elements P^n_{xy}, for $n \in \mathbb{N}$ and $x, y \in \mathbb{X}$.

- Express the Markov evolution as an evolution of a random automaton (see [?] for details on automata and Turing machines).

5

Potential theory

In the case of finite state space, we have obtained precious information on the convergence properties of the sequence of kernels P^n by studying properties of the resolvent $R_P(\lambda) = (\lambda I - P)^{-1}$ through the spectral representation of P and theorem 4.2.5. Here, we shall pursue a different approach, starting from the development of the **resolvent**, $R_P(\lambda)$, into a formal power series

$$R_P(\lambda) = (\lambda I - P)^{-1} = \frac{1}{\lambda}(I + \sum_{n=1}^{\infty} \frac{P^n}{\lambda^n}).$$

5.1 Notation and motivation

We recollect here notation, largely already introduced in the previous chapters but scattered through the text.

Definition 5.1.1 Let P be a Markovian kernel $(\mathbb{X}, \mathscr{X}) \overset{P}{\rightsquigarrow} (\mathbb{X}, \mathscr{X})$.

1. For $\flat \in \{0,1\}$, $x \in \mathbb{X}$ and $A \in \mathscr{X}$, the **potential kernel** is the quantity
$$G^\flat(x, A) = \sum_{n \geq \flat} P^n(x, A).$$
2. For $z \in \mathbb{C}$, the **generating functional** of the sequence $(P^n)_n$ is given by the formal power series
$$G_z^\flat(x, A) = \sum_{n \geq \flat} z^n P^n(x, A).$$
3. Let $f \in m\mathscr{X}_+$. This function is called **harmonic** if $Pf = f$, **superharmonic** if $Pf \leq f$, **potential** if there exists another function $r \in m\mathscr{X}_+$ such that $f = Gr$; the function r is then called a **charge** for f.
4. Let $\phi \in \mathscr{M}_+(\mathscr{X})$. This measure is called **harmonic** if $\phi P = \phi$, **superharmonic** if $\phi P \leq \phi$, **potential** if there exists another measure $\rho \in m\mathscr{X}_+$ such that $\phi = \rho G$; the measure ρ is then called a **charge** for ϕ.

Remark: It is worth noting once again that the potential kernel G may be infinite even on compact sets.

Proposition 5.1.2 *Let $g \in m\mathscr{X}_+$ and $\mathscr{S}_g(Poisson)$ the set of solutions of the* ***Poisson equation*** *$g = (I - P)f$, i.e.*

$$\mathscr{S}_g(Poisson) = \{f \in m\mathscr{X}_+ : g = (I - P)f\}.$$

Then $G^0 g$ is the smallest element of $\mathscr{S}_g(Poisson)$.

Proof: Let $h = G^0 g$. Then $Ph + g = P\sum_{n=0}^{\infty} P^n g + g = \sum_{n=1}^{\infty} P^n g + g = G^0 g = h$. Then $(I - P)h = g$, hence $h \in \mathscr{S}_g$(Poisson).

Suppose now that $f \in \mathscr{S}_g$(Poisson). Observe that equalities

$$\begin{aligned} g + Pf &= f \\ Pg + P^2 f &= Pf \\ &\vdots \\ P^n g + P^{n+1} f &= P^n f \end{aligned}$$

imply $f = \sum_{k=0}^{n} P^k g + P^{n+1} f$, for all n. Due to the positivity of both f and P, we have that $\liminf_n P^{n+1} f \geq 0$. The latter implies $f \geq G^0 g = h$. □

A few comments are necessary to justify the terminology. Consider first a simple symmetric random walk on $\mathbb{Z}$. Then, if f is additionally bounded,

$$\begin{aligned} (I - P)f(x) &= f(x) - \frac{1}{2}[f(x+1) + f(x-1)] \\ &\simeq -\frac{1}{2}[f'(x+1/2) - f'(x-1/2)] \\ &\simeq -\frac{1}{2}\Delta f(x). \end{aligned}$$

This approximate formula holds also for the simple symmetric random on $\mathbb{Z}^d$ almost unchanged: $(I - P)f(x) \simeq -\frac{1}{2d}\Delta f(x)$.

Consider now the Maxwell equations for electric fields in dimension 3. The electric vector field E is expressed as the gradient of a potential scalar field f through: $E = -\nabla f$. Additionally its divergence is proportional to the electric charge density r through: $\nabla \cdot E = cr$ where c is a physical constant. Hence, $-\Delta f = cr$. It is a well known result of electromagnetic theory that the electric potential field is expressed in terms of the charge density by the Coulomb formula: $f(x) = c' \int_{\mathbb{R}^3} \frac{r(y)}{\|x-y\|} d^3 y$ (provided that the function $\frac{1}{\|y\|}$ is integrable with respect to the measure whose density with respect to the Lebesgue 3-dimensional measure is r. In this case, we have $f = Gr$ where G is the integral kernel $G(x,A) = \int_A g(x,y)d^3y$, with $g(x,y) = \frac{1}{\|x-y\|}$. Now comparing the two equations:

$$\begin{aligned} -\Delta f &= r \\ f &= Gr \end{aligned}$$

we conclude that the operator G "inverts" the operator $-\Delta$.

Come now back to proposition 5.1.2 applied to the random walk case: $g = (I-P)f \simeq -\Delta f$. Inverting formally, we get $f = G^0 g$. We can then "verify" that $-\Delta G^0 g = (I-P)\sum_{n=0}^{\infty} P^n g = g$.

Of course at this level of presentation all computations are purely formal. The purpose of this chapter is to show that all these formal results can be given a precise mathematical meaning and moreover provide us with a valuable tool for the study of Markov chains on general state spaces.

5.2 Martingale solution to Dirichlet's problem

Proposition 5.2.1 (The maximum principle) *Let $f, g \in m\mathscr{X}_+$ and $a \in \mathbb{R}_+$.*

1. *If $G^0 f \le G^0 g + a$ holds on $\{x \in \mathbb{X} : f(x) > 0\}$ then holds everywhere (i.e. on $\mathbb{X}$).*
2. *For all $x, y \in \mathbb{X}$, $G^0(x,y) = \mathbb{P}_x(\tau_y^0 < \infty) G^0(y,y)$.*
3. *For all $x \in \mathbb{X}$ and $B \in \mathscr{X}$, $G^0(x,B) \le \sup_{z\in B} G^0(z,B)$.*

Proof:

1. Let $A = \{f > 0\} \in \mathscr{X}$.

$$\begin{aligned}
G^0 f(x) &= \mathbb{E}_x(\sum_{n=0}^{\infty} f(X_n)) \\
&= \mathbb{E}_x(\sum_{n=0}^{\infty} f(X_n) \mathbb{1}_{\{\tau_A^0 < \infty\}}) + \mathbb{E}_x(\sum_{n=0}^{\infty} f(X_n) \mathbb{1}_{\{\tau_A^0 = \infty\}}) \\
&= \mathbb{E}_x(\sum_{n=0}^{\infty} f(X_n) \mathbb{1}_{\{\tau_A^0 < \infty\}}) \text{ (because } f = 0 \text{ on } A^c) \\
&= \mathbb{E}_x(\sum_{n=\tau_A^0}^{\infty} f(X_n) \mathbb{1}_{\{\tau_A^0 < \infty\}}) \\
&= \mathbb{E}_x(\sum_{n=0}^{\infty} f(X_n \circ \theta^{\tau_A^0}) \mathbb{1}_{\{\tau_A^0 < \infty\}}) \\
&= \mathbb{E}_x(\mathbb{E}(\sum_{n=0}^{\infty} f(X_n) \circ \theta^{\tau_A^0} \mathbb{1}_{\{\tau_A^0 < \infty\}} | \mathscr{F}_{\tau_A^0})) \\
&= \mathbb{E}_x(\mathbb{1}_{\{\tau_A^0 < \infty\}} \mathbb{E}_{X_{\tau_A^0}}(\sum_{n=0}^{\infty} f(X_n))) \\
&= \mathbb{E}_x(\mathbb{1}_{\{\tau_A^0 < \infty\}} G^0 f(X_{\tau_A^0}))) \\
&\leq \mathbb{E}_x(\mathbb{1}_{\{\tau_A^0 < \infty\}} G^0 g(X_{\tau_A^0})) + a\mathbb{P}_x(\tau_A^0 < \infty).
\end{aligned}$$

Now compute explicitly the expectations appearing in the previous formula to get

$$\begin{aligned}
G^0 f(x) &\leq \sum_{n=0}^{\infty} \mathbb{E}_x(\sum_{k\in\mathbb{N}} \mathbb{1}_{\{\tau_A^0 = k\}} \int_B P^k(x, dy) P^n(y, dz) g(z)) + a \\
&\leq \sum_{n\in\mathbb{N}} P^n(x, dz) g(z) + a \\
&= G^0 g(x) + a.
\end{aligned}$$

2. Apply the previous formula to $f = \mathbb{1}_{\{y\}}$.
3. Apply the previous formula to $f = \mathbb{1}_B$.

□

Proposition 5.2.2 *Let $B \in \mathscr{X}$ be such that* $\sup_{x\in B} G^0(x, B) < \infty$ *and $A = \{x \in \mathbb{X} : \mathbb{P}_x(\tau_B < \infty) > 0\}$. Then A is a countable union of sets $A_{n,k}$ for $n, k \in \mathbb{N}$ such that $G(x, A_{n,k})$ is bounded on $\mathbb{X}$ for all $n, k \in \mathbb{N}$.*

Proof: Define for $n \geq 0$ and $k \geq 1$:

$$A_{n,k} = \{x \in \mathbb{X} : P^n(x, B) > 1/k\}.$$

For all $x \in \mathbb{X}$ and $m, n \in \mathbb{N}$, we have:

$$\begin{aligned} P^{m+n}(x,B) &= \int_{\mathbb{X}} P^m(x,dy)P^n(y,B) \\ &\geq \int_{A_{n,k}} P^m(x,dy)P^n(y,B) \\ &\geq \frac{1}{k}P^n(x,A_{n,k}) \end{aligned}$$

establishing that $kG^0(x,B) \geq G^0(x,A_{n,k})$. Using now the maximum principle we show further for all $x \in \mathbb{X}$ that $G^0(x,A_{n,k}) \leq kG^0(x,B) \leq k\sup_{x\in B} G^0(x,B) < +\infty$. Moreover $A = \cup_{n,k} A_{n,k}$. □

Theorem 5.2.3 (Dirichlet's problem) *Let $B \in \mathscr{X}$ and $g \in m\mathscr{X}_+$. The set of solutions to the* ***Dirichlet's problem*** *is the set*

$$\mathscr{S}_g(\textit{Dirichlet}) = \{f \in m\mathscr{X}_+ : Pf = f \textit{ on } B^c \textit{ and } f = g \textit{ on } B\}.$$

The smallest element of $\mathscr{S}_g$(Dirichlet) is the function $h(x) = \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0<\infty\}}g(X_{\tau_B^0}))$.

Proof: If $x \in B$ then $\tau_B^0 = 0$ and consequently $h(x) = g(x)$, else $\tau_B^0 \geq 1$ and $h(x) = \mathbb{E}_x(\sum_{k=1}^\infty \mathbb{1}_{\{\tau_B^0=k\}}g(X_k))$.

$$\begin{aligned} Ph(x) &= \int_{\mathbb{X}} P(x,dy)h(y) \\ &= \int_B P(x,dy)g(y) + \int_{B^c} P(x,dy)h(y) \\ &= \int_B P(x,dy)g(y) + \sum_{k=1}^\infty \int_{B^c} P(x,dy)\mathbb{E}_y(\mathbb{1}_{\{\tau_B^0=k\}}g(X_k)) \\ &= \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0=1\}}g(X_1)) + \sum_{k=1}^\infty \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0=k+1\}}g(X_{k+1})) \\ &= \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0=1\}}g(X_1)) + \sum_{k=2}^\infty \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0=k\}}g(X_k)) \\ &= h(x). \end{aligned}$$

This establishes that $h \in \mathscr{S}_g$(Dirichlet).

Suppose now that f is another element of $\mathscr{S}_g$(Dirichlet). Define for all n: $Z_n = f(X_{\tau_B^0 \wedge n}) = \sum_{k=0}^{n-1} \mathbb{1}_{\{\tau_B^0 = k\}} f(X_k) + \mathbb{1}_{\{\tau_B^0 \geq n\}} f(X_n)$ and $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$. Then,

$$\begin{aligned}
\mathbb{E}(Z_{n+1}|\mathscr{F}_n) &= \sum_{k=0}^{n} \mathbb{1}_{\{\tau_B^0 = k\}} f(X_k) + \mathbb{1}_{\{\tau_B^0 > n\}} \mathbb{E}(f(X_{n+1})|\mathscr{F}_n) \\
&= \sum_{k=0}^{n} \mathbb{1}_{\{\tau_B^0 = k\}} f(X_k) + \mathbb{1}_{\{\tau_B^0 > n\}} Pf(X_n) \\
&= \sum_{k=0}^{n} \mathbb{1}_{\{\tau_B^0 = k\}} f(X_k) + \mathbb{1}_{\{\tau_B^0 > n\}} f(X_n) \\
&= Z_n,
\end{aligned}$$

the penultimate line of the above equality holding because on $\tau_B^0 > n \Rightarrow X_n \notin B \Rightarrow Pf(X_n) = f(X_n)$. Thus $(Z_n)_n$ is a positive $(\mathscr{F}_n)_n$-martingale. Therefore it converges almost surely to some random variable Z_∞. On $\{\tau_B^0 < \infty\}$, $Z_n \to f(X_{\tau_B^0}) = g(X_{\tau_B^0})$. On $\{\tau_B^0 = \infty\}$, $Z_n \to Z_\infty$ and consequently by Fatou inequality: $\mathbb{E}_x(Z_\infty) \leq \liminf_n \mathbb{E}_x Z_n$. We have finally:

$$\begin{aligned}
h(x) &= \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0 < \infty\}} g(X_{\tau_B^0})) \\
&= \mathbb{E}_x(\mathbb{1}_{\{\tau_B^0 < \infty\}} \lim_n Z_n) \\
&\leq \mathbb{E}_x(\lim_n Z_n) \\
&\leq \liminf_n \mathbb{E}_x(Z_n) \\
&= \mathbb{E}_x(Z_0) = f(x).
\end{aligned}$$

□

Theorem 5.2.4 (Riesz decomposition) *Let f be a finite superharmonic function. Then there exists a unique harmonic function h and a unique charge $r \in m\mathscr{X}_+$ such that $f = h + G^0 r$.*

Proof: For any $N \in \mathbb{N}$: $\sum_{k=0}^{N}(P^k f - P^{k+1} f) = f - P^{N+1} f$. Now $0 \leq Pf \leq f \Rightarrow f - Pf \geq 0$. It follows that $0 \leq \sum_{k=0}^{N} P^k(f - Pf) = f - P^{N+1} f$ is an increasing sequence of N implying that $P^N f$ is decreasing (and obviously minorised by 0), thus converging; let $h = \lim_n P^n f$. Now $Ph = P(\lim_n P^n f) = \lim_n P^{n+1} f = h$ proving that h is harmonic.

On the other hand, $0 \leq f - h = \lim_N \sum_{k=0}^{N} P^k(f - Pf) = G^0(f - Pf)$, showing that $f - h$ is the finite potential due to the positive charge $r = f - Pf$. Thus the decomposition $f = h + G^0 r$ is established.

To show unicity, suppose that h' and r' are another such pair satisfying

$$f = h + G^0 r = h' + G^0 r'.$$

Then

$$\begin{aligned} Pf &= P(h+G^0 r) = h + G^0 r - r = f - r \\ &= P(h' + G^0 r') = h' + G^0 r' - r' = f - r' \end{aligned}$$

Consequently $r = r'$ and therefore $h = h'$. □

Corollary 5.2.5 *A finite superharmonic function f is a potential if* $\lim_n P^n f = 0$. *If a finite superharmonic function is bounded from above by a finite potential, then its harmonic part is 0.*

Exercise 5.2.6 Let $\phi \in \mathscr{M}_+(\mathscr{X})$ be a superharmonic measure. Then there exists a unique harmonic measure α and a unique charge $\rho \in \mathscr{M}_+(\mathscr{X})$ such that $\phi = \alpha + \rho G^0$.

Lemma 5.2.7

1. *If s is P-superharmonic, then s is a potential if and only if* $\lim_n P^n s = 0$.
2. *If $0 \le s \le v$, s is P-superharmonic, and v a potential, then s is also a potential.*
3. *If s and s' are P-superharmonic, then $w = s \wedge s'$ is also P-superharmonic.*

Proof:

1. By theorem 5.2.4, s can be uniquely decomposed into $s = h + G^0 r$, where $h = \lim_n P^n s$ is the harmonic part of the Riesz decomposition. If this term vanishes, then s is a potential and conversely, if s is a potential then the harmonic part must vanish.
2. Since for every n, the inequality $0 \le P^n s \le P^n v$ holds and v is a potential, it follows from 1. that $\lim_n P^n v = 0$. Hence $\lim_n P^n s = 0$, and from 1. s is a potential.
3. We have $Pw \le Ps \le s$ and $Pw \le Ps' \le s'$; hence $Pw \le s \wedge s' = w$.

□

Exercise 5.2.8 First recall some definitions:

- A partially ordered set $(\mathbb{Y}, \preceq)$ is a **lattice** if any for any pair of points $x, y \in \mathbb{Y}$ there exists a supremum $s = x \vee y \in \mathbb{Y}$.
- An $\mathbb{R}$-vector space $\mathbb{Y}$ is **ordered** if is equipped with a partial order $\preceq$ compatible with the operations of the vector space, i.e. for all $x, y \in \mathbb{Y}$

$$\begin{aligned} x \preceq y &\Rightarrow \forall z \in \mathbb{Y}, x + z \preceq y + z \\ x \preceq y &\Rightarrow \forall \lambda \in \mathbb{R}_+, \lambda x \preceq \lambda y. \end{aligned}$$

- An ordered $\mathbb{R}$-vector space $\mathbb{Y}$, such that $(\mathbb{Y}, \preceq)$ is a lattice, is called a **Riesz space** or **vector lattice**.
- A subset C of a vector space is called a **cone** if for all $\lambda \in \mathbb{R}_+$, $\lambda C \subseteq C$.
- A subset C of a vector space is said **convex** if $C + C \subseteq C$.

After this recollection of definitions, show that

1. the set of of superharmonic functions of $m\mathscr{X}_+$ is a lattice and a convex cone,
2. the set of harmonic functions of $m\mathscr{X}_+$ is a Riesz space,
3. the set of bounded harmonic functions of $m\mathscr{X}_+$ is a subspace of the Banach space $(\mathbb{X}, \|\cdot\|_\infty)$.

Recall the notation previously defined in the exercise section 3.6.

Definition 5.2.9 For $B \in \mathscr{X}$ and $\flat \in \{0, 1\}$, we define

$$\begin{array}{rcl} L_B^\flat(x) & = & \mathbb{P}_x(\tau_B^\flat < \infty) \\ H_B^\flat(x) & = & \mathbb{P}_x(\eta^\flat(B) = \infty) \end{array}$$

When B is a singleton $B = \{x\}$, we write usually $L_x^\flat$ and $H_x^\flat$, instead of $L_{\{x\}}^\flat$ and $H_{\{x\}}^\flat$.

Proposition 5.2.10 *The function* L_B^0 *is finite superharmonic. Its harmonic part in the Riesz decomposition is* H_B^0. *Moreover,* $\lim_{n\to\infty} L_B^0(X_n) = \lim_{n\to\infty} H_B^0(X_n) = \mathbb{1}_{\{\eta^0(B)=\infty\}}$ *almost surely.*

Proof: Obviously $L_B^0(x) \leq 1$ for all x. To apply Riesz decomposition, we must establish its superharmonicity. Define $\tau_B^k = \inf\{n \geq k : X_n \in B\}$ and observe that $\mathbb{1}_{\{\tau_B^k < \infty\}} = \mathbb{1}_{\{\tau_B^0 < \infty\}} \circ \theta^k$. Strong Markov property yields then:

$$\begin{array}{rcl} P^k L_B^0(x) & = & \mathbb{E}_x(L_B^0(X_k)) \\ & = & \mathbb{E}_x(\mathbb{P}_{X_k}(\tau_B^0 < \infty)) \\ & = & \mathbb{E}_x(\mathbb{E}_x(\mathbb{1}_{\{\tau_B^0 < \infty\}} \circ \theta^k | \mathscr{F}_k)) \\ & = & \mathbb{P}_x(\tau_B^k < \infty) \\ & \leq & \mathbb{P}_x(\tau_B^0 < \infty) \\ & = & L_B^0(x). \end{array}$$

Since L_B^0 is shown to be finite superharmonic, repeating the reasoning appearing in the proof of theorem 5.2.4, we establish

$$\begin{array}{rcl} P^n L_B^0(x) & = & \mathbb{P}_x(\tau_B^n < \infty) \\ & \downarrow & \lim_{n\to\infty} {}^{\downarrow}\mathbb{P}_x(\tau_B^n < \infty) \\ & = & \mathbb{P}_x(\cap_{n\in\mathbb{N}}\{\tau_B^n < \infty\}) \\ & = & \mathbb{P}_x(\eta^0(B) = \infty) \quad (*) \\ & = & H_B^0(x) \le 1, \end{array}$$

the statement $(*)$ above holding because on defining $A = \cup_{n\in\mathbb{N}}\{\tau_B^n = \infty\}$ and $C = \{\eta^0(B) < \infty\}$, we have:

$$[\omega \in A] \Rightarrow [\exists N : \tau_B^N(\omega) = \infty] \Rightarrow [\eta^0(B)(\omega) \le N]$$

and

$$[\omega \in C] \Rightarrow [\exists N : \eta^0(B)(\omega) \le N][\Rightarrow [\exists M = M(N) : \tau_B^M(\omega) = \infty].$$

Thus it is established that H_B^0 is the harmonic part of L_B^0 (and is moreover finite).

Now, the previous conclusions imply that the sequences defined by $S_n = L_B^0(X_n)$ (resp. $M_n = H_B^0(X_n)$) are positive supermartingale (resp. martingale). As such, they converge almost surely to S_∞ (resp. M_∞). Additionally, both are bounded by 1, therefore they are uniformly integrable, establishing thus their convergence in L^1 and, for (M_n) we have the closure property:

$$M_n = \mathbb{E}(M_\infty|\mathscr{F}_n).$$

Note that for all $n \in \mathbb{N}$

$$\mathbb{1}\{\eta^0(B) = \infty\} = \mathbb{1}\{\eta^0(B) = \infty\} \circ \theta^n.$$

Consequently,

$$\begin{array}{rcl} M_n & = & H_B^0(X_n) \\ & = & \mathbb{E}_{X_n}(\mathbb{1}\{\eta^0(B) = \infty\}) \\ & = & \mathbb{E}_x(\mathbb{1}\{\eta^0(B) = \infty\} \circ \theta^n|\mathscr{F}_n) \\ & = & \mathbb{E}_x(\mathbb{1}\{\eta^0(B) = \infty\}|\mathscr{F}_n) \\ & \xrightarrow[n\to\infty]{} & \mathbb{1}\{\eta^0(B) = \infty\} \text{ by uniform integrability} \\ & = & M_\infty. \end{array}$$

Consider now the supermartingale $(S_n)_n$. Note that almost surely

$$S_\infty = \lim_{n\to\infty} L_B^0(X_n) \ge \lim_{n\to\infty} H_B^0(X_n) = M_\infty.$$

On the other hand

$$
\begin{aligned}
\mathbb{E}_x(S_\infty) &= \lim_{n\to\infty} \mathbb{E}_x(L_B^0(X_n)) \\
&= \lim_{n\to\infty} P^n L_B^0(x) \\
&= H_B^0(x) \\
&= \mathbb{E}_x(H_B^0(X_n)), \text{ for all } n \text{ (because } H_B^0(X_n) \text{ is a martingale)} \\
&= \mathbb{E}_x(\lim_{n\to\infty} H_B^0(X_n)) \\
&= \mathbb{E}_x(M_\infty) \text{ because of the convergence in } \mathscr{L}^1.
\end{aligned}
$$

Thus $S_\infty = M_\infty$ almost surely. □

5.3 Asymptotic and invariant σ-algebras

Definition 5.3.1 Let X be a $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$ and $\mathscr{T}_n = \sigma(X_m, m \geq n)$, for $n \in \mathbb{N}$.

- The σ-algebra $\mathscr{T}_\infty = \cap_{n\in\mathbb{N}} \mathscr{T}_n$ is called **tail σ-algebra** for the sequence X.
- A random variable Y on the path space $(\mathbb{X}^\infty, \mathscr{X}^\infty, \mathbb{P})$ is called **asymptotic** if there exists a sequence $(f_n)_n$ of measurable mappings $f_n : \mathbb{X}^\infty \to \mathbb{R}$ for $n \in \mathbb{N}$, such that $Y = f_n(X_n, X_{n+1}, \ldots)$ for all $n \in \mathbb{N}$.
- If $Y = 1\!\!1_A$ for some event $A \in \mathscr{X}^\infty$, then A is called **asymptotic event**.
- The class of all asymptotic events generates a σ-algebra, called the **asymptotic σ-algebra**, that is isomorphic to $\mathscr{T}_\infty$.

Exercise 5.3.2 Let X be a $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$, $g : \mathbb{X} \to \mathbb{R}_+$ a measurable function, and $(F_n)_n$ a fixed sequence of measurable sets in $\mathscr{X}$. Determine which of the following events (if any) are asymptotic:

1. $\{X_n \in F_n, \text{ infinitely often}\}$,
2. $\{X_n \in F_n, \text{ for all } n \geq k\}$, for some fixed k,
3. $\{\sum_{n\geq k} 1\!\!1_B(X_n) = \infty\}$, for fixed $B \in \mathscr{X}$ and $k \in \mathbb{N}$,
4. $\{\sum_{n\geq k} 1\!\!1_B(X_n) < \infty\}$, for fixed $B \in \mathscr{X}$ and $k \in \mathbb{N}$,
5. $\{\sum_{n\geq k} 1\!\!1_B(X_n) = l\}$, for fixed $B \in \mathscr{X}$ and $k, l \in \mathbb{N}$,
6. $\{\sum_{n\geq k} g(X_n) = \infty\}$, for fixed $k \in \mathbb{N}$,
7. $\{\sum_{n\geq k} g(X_n) < \infty\}$, for fixed $k \in \mathbb{N}$,
8. $\{\sum_{n\geq k} 1\!\!1_B(X_n) = c\}$, for fixed $k \in \mathbb{N}$ and $c \in \mathbb{R}_+$.

Definition 5.3.3 Let X be a $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$.

- A random variable Y on the path space $(\mathbb{X}^\infty, \mathscr{X}^\infty, \mathbb{P})$ is called **invariant** if there exists a measurable mapping $f : \mathbb{X}^\infty \to \mathbb{R}$, such that $Y = f(X_n, X_{n+1}, \ldots)$ for all $n \in \mathbb{N}$.

- If $Y = \mathbb{1}_A$ for some event $A \in \mathcal{X}^\infty$, then A is called **invariant event**.
- The class of all invariant events generates a σ-subalgebra of $\mathcal{T}_\infty$, called the **invariant σ-algebra**, denoted by $\mathcal{J}_\infty$.

Exercise 5.3.4 Which among the events defined in exercise 5.3.2 are elements of $\mathcal{J}_\infty$?

Remark: The probability $\mathbb{P}_\mu$ on the path space does not intervene in the definitions of either asymptotic or invariant events. On identifying random variables on $\mathbb{X}^\infty$ that coincide $\mathbb{P}$-almost surely, we determine a class of events that generate the so-called $\mathbb{P}$-asymptotic or $\mathbb{P}$-invariant σ-algebras, denoted respectively by $\mathcal{T}_\infty^{\mathbb{P}}$ or $\mathcal{J}_\infty^{\mathbb{P}}$.

Recall that a σ-algebra containing just two sets is called **trivial**. Criteria establishing triviality of the asymptotic σ-algebras are sought in the sequel since they will play an important role in establishing convergence results for Markov chains.

Definition 5.3.5 Let X be a $\mathsf{MC}((\mathbb{X},\mathcal{X}),P,\mu)$ and denote by $\mathcal{N} = \mathcal{P}(\mathbb{N})$. The process $(X_n,T_n)_n$ taking values in the measurable space $(\mathbb{X}\times\mathbb{N},\mathcal{X}\otimes\mathcal{N})$ with $T_{n+1} = T_n + 1$, is called the **spacetime process** associated with X.

Proposition 5.3.6 *Let $(X_n,T_n)_n$ be a spacetime process with $T_0 = k$ for some $k \in \mathbb{N}$. An event A is asymptotic for $(X_n)_n$ if and only if is invariant for $(X_n,T_n)_n$.*

Proof: Let $Y = \mathbb{1}_A$.

($\Rightarrow$) Since A is asymptotic for $(X_n)_n$, there exists a sequence $(f_n)_n$ in $m\mathcal{X}^\infty$ such that $Y = f_n(X_n,X_{n+1},\ldots)$ for all n. Let $(n_l)_{l\in\mathbb{N}}$ be a sequence of integers with $n_0 \geq k$ and $n_{l+1} = n_l + 1$ for all l. Define $\tilde{f} : (\mathbb{X}\times\mathbb{N})^\infty \to \mathbb{R}$ by $\tilde{f}((x_0,n_0),(x_1,n_1),\ldots) = f_{n_0-k}(x_0,x_1,\ldots)$. We obtain then that $Y = \tilde{f}((X_n,T_n),(X_{n+1},T_{n+1}),\ldots)$.

($\Leftarrow$) Since A is invariant for $(X_n,T_n)_n$, from the last equality, we obtain on defining $f_n(x_0,x_1,\ldots) = \tilde{f}(x_0,n+k),(x_1,n+k+1),\ldots)$ for every n that $Y = f_n(X_n,X_{n+1},\ldots)$.

□

5.4 Triviality of the asymptotic σ-algebra

Since with any X a $\mathsf{MC}((\mathbb{X},\mathcal{X}),P,\mu)$ is associated a spacetime chain, we use the tilde notational convention to denote quantities referring to the latter. Thus the kernel $P : \mathbb{X} \times \mathscr{X} \to [0,1]$ induces a kernel $\tilde{P}(\mathbb{X} \times \mathbb{N}) \times (\mathscr{X} \otimes \mathscr{N}) \to [0,1]$ by the formula

$$\tilde{P}((x,n), F \times \{n'\}) = \delta_{n+1,n'} P(x,F), \forall x \in \mathbb{X}, \forall F \in \mathscr{X}, \forall n \in \mathbb{N}.$$

The path space $\tilde{\mathbb{X}}^\infty = (\mathbb{X} \times \mathbb{N})^\infty$ will be identified with $\tilde{\Omega} = \{\tilde{\omega} : \mathbb{N} \to \mathbb{X} \times \mathbb{N}\}$. Similarly the spacetime process $(\tilde{X}_n)_n = (X_n, T_n)_n$ will be defined through its coordinate[1] representation: $\tilde{X}_n(\tilde{\omega}) = \tilde{\omega}_n = (X_n(\tilde{\omega}), T_n(\tilde{\omega}))$. Finally, we use the tilde convention to extend notions defined on functions f in $m\mathscr{X}$ to notions defined o functions $\tilde{f}$ in $m(\mathscr{X} \otimes \mathscr{N})$. For example, if h is an harmonic function for P, $\tilde{h}$ will be an harmonic function for $\tilde{P}$, verifying: $\tilde{h}(x,n) = \int_{\mathbb{X}} P(x,dy)\tilde{h}(y,n+1)$.

A random variable $\Xi : \Omega \to \mathbb{R}$, defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, is invariant if and only if $\Xi \circ \theta = \Xi$. With any $\Xi \in b\mathscr{F}$ is associated a $h \in b\mathscr{X}$ by $h(x) = \mathbb{E}_x(\Xi)$.

Theorem 5.4.1 – *Let $\Xi \in b\mathscr{F}_+$ be invariant. Then $h \in b\mathscr{X}$, defined by $h(x) = \mathbb{E}_x(\Xi)$, is harmonic. Additionally, $h \equiv 0 \Leftrightarrow \mathbb{P}_x(\Xi = 0) = 1$.*
– *Let $h \in b\mathscr{X}$ be harmonic. Then the random variable $\Xi = \liminf_n h(X_n)$ verifies $\mathbb{E}_x(\Xi) = h(x)$ for all x.*

Proof: For all n:

$$\begin{aligned} h(X_n) &= \mathbb{E}_{X_n}(\Xi) \\ &= \mathbb{E}_x(\Xi \circ \theta^n | \mathscr{F}_n) \text{ by Markov property} \\ &= \mathbb{E}_x(\Xi | \mathscr{F}_n) \text{ by invariance} \end{aligned}$$

For $n \geq 1$ the previous equality yields

$$\mathbb{E}_x(h(X_n)|\mathscr{F}_{n-1}) = \mathbb{E}_x(\mathbb{E}_x(\Xi|\mathscr{F}_n)|\mathscr{F}_{n-1}) = h(X_{n-1}),$$

establishing that $(h(X_n))_n$ is a martingale. Applying to $n = 1$:

$$\begin{aligned} \mathbb{E}_x(h(X_1)) &= h(x) \\ &= \mathbb{E}_x(\mathbb{E}_{X_1}(\Xi)) \\ &= \int_{\mathbb{X}} P(x,dy)\mathbb{E}_y(\Xi) \\ &= \int_{\mathbb{X}} P(x,dy)h(y), \end{aligned}$$

proving the harmonicity of h. Since $\Xi \in b\mathscr{F}$, this martingale is uniformly integrable, hence closed, meaning that $\lim_{n\to\infty} h(X_n) = \Xi$, $\mathbb{P}_x$-a.s. and for arbitrarily chosen x. The latter proves the equivalence $h \equiv 0 \Leftrightarrow \mathbb{P}_x(\Xi = 0) = 1$. The second statement of the theorem is just a consequence of the unicity of the limit. □

1. Note the trivial modification on the coordinate representation of the process $(X_n)_n$ that is necessary.

Corollary 5.4.2 *The following two conditions are equivalent:*

1. *All bounded harmonic functions are constant.*
2. *For every $\mu \in \mathcal{M}_1(\mathcal{X})$, the $\mathcal{J}_\infty^{\mathbb{P}_\mu}$ is trivial.*

Proof: The theorem 5.4.1 guarantees 1⇒2. To prove the converse, suppose that there exists a bounded harmonic function h and there exist $x, y \in \mathbb{X}$, with $x \neq y$, such that $h(x) \neq h(y)$. The theorem 5.4.1 guarantees the existence of an invariant random variable $\Xi \in b\mathcal{F}$, such that $h(z) = \mathbb{E}_z(\Xi)$. Let $\mu = \frac{1}{2}(\epsilon_x + \epsilon_y)$. Then $\mathbb{E}_\mu(\Xi|X_0 = x) = h(x) \neq h(y) = \mathbb{E}_\mu(\Xi|X_0 = y)$, thus Ξ is not almost surely constant.
□

Theorem 5.4.3 *A Markov chain X satisfies the condition*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{T}_n} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| = 0, \ \forall B \in \mathcal{F},$$

if and only if the σ-algebra $\mathcal{T}_\infty^{\mathbb{P}}$ is trivial.

Proof:

(⇒) Suppose the limit vanishes. For any $B \in \mathcal{T}_\infty \subseteq \mathcal{F}$, and $A = B$, we get $\mathbb{P}(B)^2 = \mathbb{P}(B)$. It follows that either $\mathbb{P}(B) = 0$ or $\mathbb{P}(B) = 1$. Therefore, up to $\mathbb{P}$-negligible events, the asymptotic σ-algebra $\mathcal{T}_\infty$ is trivial.

(⇐) For $B \in \mathcal{F}$, $A \in \mathcal{T}_n$:

$$\begin{aligned} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| &= |\int_A (\mathbb{P}(B|\mathcal{T}_n) - \mathbb{P}(B))d\mathbb{P}| \\ &\leq \int_{\mathbb{X}} |\int_{\mathbb{X}} (\mathbb{P}(B|\mathcal{T}_n) - \mathbb{P}(B))d\mathbb{P}|. \end{aligned}$$

Now, $(\mathbb{P}(B|\mathcal{T}_n))_n$ is a reverse martingale; as such it converges almost surely to $\mathbb{P}(B|\mathcal{T}_\infty)$. If the σ-algebra $\mathcal{T}_\infty^{\mathbb{P}}$ is trivial, $\mathbb{P}(B|\mathcal{T}_\infty) = \mathbb{P}(B)$.

□

Theorem 5.4.4 *Let X be a* MC$((\mathbb{X},\mathcal{X}), P, \mu)$ *and $\tilde{X}$ the corresponding spacetime chain* MC$((\mathbb{X}\times\mathbb{N}, \mathcal{X}\otimes\mathcal{N}), \tilde{P}, \tilde{\mu})$. *The following are equivalent:*

1. *For any $\tilde{\mu} \in \mathcal{M}_1(\mathcal{X}\otimes\mathcal{N})$ such that $\tilde{\mu}(\mathbb{X}\times\{k\}) = 1$ for some $k \in \mathbb{N}$, the σ-algebra $\tilde{\mathcal{J}}_\infty^{\tilde{\mathbb{P}}_{\tilde{\mu}}}$ is trivial.*
2. *For any $\mu \in \mathcal{M}_1(\mathcal{X})$, the σ-algebra $\mathcal{T}_\infty^{\mathbb{P}_\mu}$ is trivial.*
3. *For any $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$, we have $\lim_{n\to\infty} \|\mu P^n - \nu P^n\| = 0$.*
4. *The only bounded harmonic spacetime functions are the constants.*

Proof:

1 ⇒ 2 This follows from the equivalence established in proposition 5.3.6.

2 ⇒ 3 Let $x, y \in \mathbb{X}$, $x \neq y$, $\mu = \frac{1}{2}(\epsilon_x + \epsilon_y)$, and $A = \{X_0 = x\} \in \mathscr{F}$. For any $F \in \mathscr{X}$, we have $\{X_n \in F\} \in \mathscr{T}_n$ and by theorem 5.4.3,

$$\lim_{n\to\infty} |\mathbb{P}_\mu(A \cap \{X_n \in F\}) - \mathbb{P}_\mu(A)\mathbb{P}_\mu(\{X_n \in F\})| = 0.$$

Now,

$$\begin{aligned} \mathbb{P}_\mu(A \cap \{X_n \in F\}) &= \mathbb{P}_\mu(\{X_n \in F\}|A)\mathbb{P}_\mu(A) \\ &= P^n(x, F) \times \frac{1}{2} \end{aligned}$$

and

$$\begin{aligned} \mathbb{P}_\mu(A)\mathbb{P}_\mu(X_n \in F) &= \frac{1}{2}(\mathbb{P}_x(X_n \in F) + \mathbb{P}_y(X_n \in F)) \\ &= \frac{1}{4}\mathbb{P}_\mu(A)(P^n(x, F) + P^n(y, F)), \end{aligned}$$

yielding $|\mathbb{P}_\mu(A\cap\{X_n \in F\}) - \mathbb{P}_\mu(A)\mathbb{P}_\mu(X_n \in F)| = \frac{1}{4}|(P^n(x,F) - P^n(y,F)| \xrightarrow[n\to\infty]{} 0$, **uniformly** in $F \in \mathscr{X}$. Now for any two probabilities $\mu, \nu \in \mathscr{M}_1(\mathscr{X})$, the measure $\sigma_n = \mu P^n - \nu P^n$ is a signed measure in $\mathscr{M}(\mathscr{X})$. For any additive set function (hence for signed measures also) we have the Hahn decomposition (see [4], p. 441): For every n, there is a partition of $\mathbb{X}$ into disjoint sets $\mathbb{X} = H_n^- \sqcup H_n^+$ such that $\sigma_n(F) \leq 0$ for all $F \subseteq H_n^-$ and $\sigma_n(F) \geq 0$ for all $F \subseteq H_n^+$. Therefore, for all $F \in \mathscr{X}$, we define $\sigma_n^+(F) \equiv \sigma_n(F \cap H_n^+)$ and $\sigma_n^-(F) \equiv -\sigma_n(F \cap H_n^-)$, so that $\sigma_n(F) = \sigma_n^+(F) - \sigma_n^-(F)$ for two positive measures σ_n^+ and σ_n^- of disjoint support. Now, for all $G \in \mathscr{X}$, with $G \subseteq F$: $\sigma_n(G) \leq \sigma_n^+(G) \leq \sigma_n^+(F)$ with equality holding if $G = F \cap H_n^+$. Thus, $\sigma_n^+(F) = \sup_{G\in\mathscr{X};G\subseteq F} \sigma_n(G)$. Similarly, $\sigma_n^-(F) = -\inf_{G\in\mathscr{X};G\subseteq F} \sigma_n(G)$. Moreover, $\sigma_n^+(G) + \sigma_n^-(G) = |\sigma_n|(G)$ and $\|\sigma_n\|_1 = \sup_{F\in\mathscr{X}} |\sigma_n|(F) = |\sigma_n|(\mathbb{X})$. Use the explicit form of the measure σ_n to write:

$$\begin{aligned} \sigma_n(H_n^+) &= (\mu P^n - \nu P^n)(H_n^+) \\ &= \int_{\mathbb{X}}\int_{\mathbb{X}} (P^n(x, H_n^+) - P^n(y, H_n^+)\mu(dx)\nu(dy) \\ &\leq \int_{\mathbb{X}}\int_{\mathbb{X}} \sup_{F\in\mathscr{X}} |P^n(x,F) - P^n(y,F)|\mu(dx)\nu(dy). \end{aligned}$$

Now, $\sup_{F\in\mathscr{X}} |P^n(x,F) - P^n(y,F)| \longrightarrow 0$ and $|P^n(x,F) - P^n(y,F)| \leq 2$; dominated convergence theorem guarantees that $\sigma_n(H_n^+) \xrightarrow[n\to\infty]{} 0$. Similarly, we prove that $\sigma_n(H_n^-) \xrightarrow[n\to\infty]{} 0$. Thus,

$$\begin{aligned} \|\sigma_n\|_1 &= \|\mu P^n - \nu P^n\|_1 \\ &= \sigma_n^+(\mathbb{X}) + \sigma_n^-(\mathbb{X}) \\ &= \sigma_n^+(H_n^+) + |\sigma_n^-(H_n^-)| \\ &\xrightarrow[n\to\infty]{} 0. \end{aligned}$$

3 ⇒ 4 Let $\tilde{h}$ be an harmonic spacetime function that is bounded, i.e. there exists some constant $K \le \infty$, such that for all $x \in \mathbb{X}$ and all $m \in \mathbb{N}$: $|\tilde{h}(x,m)| \le K$. For any $x, y \in \mathbb{X}$ and any $m, n \in \mathbb{N}$:

$$\begin{aligned} |\tilde{h}(x,n) - \tilde{h}(y,n)| &= |\int_{\mathbb{X}} (P^m(x,dz) - P^m(y,dz))h(z,m+n)| \\ &\le K\|P^m(x,\cdot) - P^m(y,\cdot\|_1 \\ &\xrightarrow[n\to\infty]{} 0 \text{ by (3)}, \end{aligned}$$

implying that $\tilde{h}$ is, as a matter of fact, a function merely of the second argument, i.e. there exists a function $g : \mathbb{N} \to \mathbb{R}_+$ such that for all $x \in \mathbb{X}$ and all $n \in \mathbb{N}$, $g(n) = \tilde{h}(x,n)$. Using harmonicity of $\tilde{h}$, we get, for all $n \in \mathbb{N}$:

$$\begin{aligned} g(n) &= \tilde{h}(x,n), \ \forall x \in \mathbb{X} \\ &= \int_{\mathbb{X}} P(x,dy)\tilde{h}(y,n+1) \\ &= \int_{\mathbb{X}} P(x,dy)g(n+1) \\ &= g(n+1). \end{aligned}$$

Thus $\tilde{h}$ is in fact a constant.

4 ⇒ 1 Applying corollary 5.4.2 to the spacetime chain we conclude that $\tilde{\mathscr{I}}_\infty^{\tilde{\mathbb{P}}_{\tilde{\mu}}}$ is trivial.

□

Remark: The main interest of this result is that we only need to check harmonicity of merely **local** functions i.e. in $m\mathscr{X}$ instead of global properties of the whole sequence X.

6

Markov chains on denumerably infinite sets

Here the state space is a denumerably infinite set $\mathbb{X}$. Note however that most of the results stated in this chapter remain valid (as a matter of fact they are even easier) for denumerably finite sets. Since the approach developed here is different from the spectral methods developed in chapter 4, they provide another possible approach of the finite case. Conversely, most of the spectral methods can be adapted (with considerable effort however) to the denumerably infinite case. We can, without loss of generality, assume that $\mathscr{X} = \mathscr{P}(\mathbb{X})$. Recall also the notation introduced earlier for $\flat \in \mathbb{N}$.

$$
\begin{aligned}
L_B^\flat(x) &= \mathbb{P}_x(\tau_B^\flat < \infty) \\
H_B^\flat(x) &= \mathbb{P}_x(\eta^\flat(B) = \infty) = \mathbb{P}_x(\eta^0(B) = \infty).
\end{aligned}
$$

The standard reference for a first approach to Markov chains on denumerably infinite sets is [25]. This chapter concludes with some more advanced topics on the study of recurrence/transience with the help of Lyapunov functions, pertaining to the constructive theory of Markov chains [?].

6.1 Classification of states

We use the notation introduced in definition 5.2.9.

Definition 6.1.1 Let X be a $\mathsf{MC}((\mathbb{X},\mathscr{X}), P, \mu)$. The directed graph $(\mathbb{X},\mathbb{A})$ with vertex set $\mathbb{X}$ and oriented edge set $\mathbb{A} = \{(x, y) \in \mathbb{X} \times \mathbb{X} : P(x, y) > 0\}$ is called **graph** of the Markov chain.

Definition 6.1.2 Let $x, y \in \mathbb{X}$.

– We say y is **accessible** from x, and write $x \to y$, if $L^1_y(x) > 0$.
– We say x and y **communicate**, and write $x \leftrightarrow y$, if $x \to y$ and $y \to x$.

Remark: Accessibility of y from x means that there exists an integer $n = n(x, y)$ such that $P^n(x, y) > 0$, i.e. the graph of the Markov chain contains a finite path of directed edges leading from x to y.

Definition 6.1.3 A state $x \in \mathbb{X}$ is called **essential** if for each $y \in \mathbb{X}$ such that $x \to y$ follows that $y \to x$. We write $\mathbb{X}_e$ for the set of essential states.

Proposition 6.1.4 *Let P be a Markov transition matrix on $\mathbb{X}$.*

1. *For all $x \in \mathbb{X}$, there exists $y \in \mathbb{X}$ accessible from x.*
2. *Accessibility is transitive.*
3. *If $x \in \mathbb{X}_e$ then $x \leftrightarrow x$.*
4. *If $x \in \mathbb{X}_e$ and $x \to y$ then $y \in \mathbb{X}_e$ and $x \leftrightarrow y$.*
5. *Restricted on $\mathbb{X}_e$, communication is an equivalence relationship.*

Proof: The proof of all items is left as an exercise; we only prove item 4. Since $x \in \mathbb{X}_e$ and $x \to y$, then $y \to x$; therefore $x \leftrightarrow y$. Suppose that z is such that $y \to z$. We must show that $z \to y$. Now, since $x \to y$ and $y \to z$, transitivity implies that $x \to z$. But $x \in \mathbb{X}_e$, hence $z \to x$. Therefore, $z \to x$ and $x \to y$. Transitivity implies $z \to y$. □

Corollary 6.1.5 *We have the decomposition*

$$\mathbb{X} = (\sqcup_{[x]\in K}[x]) \sqcup \mathbb{X}_u,$$

*where $K = \mathbb{X}_e / \leftrightarrow$ is the set of **communicating classes** $K \ni [x] = \{y \in \mathbb{X}_e : x \leftrightarrow y\}$ and $\mathbb{X}_u$ is the set of **unessential states**.*

Definition 6.1.6 A subset $A \subseteq \mathbb{X}$ is called **absorbing** or **(stochastically) closed** if $A \neq \emptyset$ and for all $x \in A$, $\sum_{y\in A} P_{xy} = 1$. If for some $x \in \mathbb{X}$, we have $[x] = \mathbb{X}$, then the chain (or equivalently the transition kernel) is called **irreducible**.

Example 6.1.7

1. For the "ruin problem" with total fortune L, the set of essential states is $\mathbb{X}_e = \{0, L\} = [0] \sqcup [L]$ and the set of unessential ones is $\mathbb{X}_u = \{1, \ldots, L-1\}$.
2. For the simple symmetric random walk on $\mathbb{Z}$, the set of essential states is $\mathbb{X}_e = \mathbb{Z} = [0]$. The chain is irreducible.

Theorem 6.1.8 *Let $B \subseteq \mathbb{X}$ and $m_x(B) = \mathbb{E}_x(\tau_B^0) = \mathbb{R}_+ \cup \{+\infty\}$. Then the function $m_\cdot(B)$ is the minimal positive solution of the system*

$$\begin{aligned} m_x(B) &= 0 \text{ if } x \in B \\ m_x(B) &= 1 + \sum_{y \in B^c} P_{xy} m_y(B) \text{ if } x \notin B. \end{aligned}$$

Proof: Exercise. (Hint, use proposition 5.2.10). □

Definition 6.1.9 Let $x \in \mathbb{X}$ be such that $\mathbb{P}_x(\tau_x < \infty) > 0$. The **period** d_x of x is defined as

$$d_x = \gcd\{n \geq 1 : \mathbb{P}_x(X_n = x) > 0\}.$$

If $d_x = 1$, then the state x is called **aperiodic**.

Remark: We always have: $1 \leq d_x \leq \inf\{n \geq 1 : \mathbb{P}_x(X_n = x) > 0\}$.

Proposition 6.1.10 *Period is a class property, i.e.*

$$\forall x \in \mathbb{X}, \forall y \in [x] : d_x = d_y.$$

Proof: Since

$$\begin{aligned} [x \leftrightarrow y] &\Leftrightarrow [x \to y \text{ and } y \to x] \\ &\Leftrightarrow [\exists n_1 \geq 1 : \mathbb{P}_x(X_{n_1} = y) = \alpha > 0 \text{ and } \exists n_2 \geq 1 : \mathbb{P}_y(X_{n_2} = x) = \beta > 0], \end{aligned}$$

it follows that $\mathbb{P}_x(X_{n_1+n_2} = x) \geq \alpha\beta > 0$. Therefore, d_x is a divisor of $n_1 + n_2$. If $\mathbb{P}_y(X_n = y) > 0$, then $\mathbb{P}_x(X_{n_1+n+n_2} = x) \geq \alpha\beta\mathbb{P}_y(X_n = y) > 0$, hence d_x is a divisor of $n_1 + n + n_2$. Consequently, d_x is a divisor of n because it is already a divisor of $n_1 + n_2$. Therefore, d_x is a divisor of all $n \geq 1$ such that $\mathbb{P}_y(X_n = y) > 0$, or

$$d_x \leq d_y = \gcd\{n \geq 1 : \mathbb{P}_y(X_n = y) > 0\}.$$

Exchanging the roles of x and y, we get the reverse inequality, hence $d_x = d_y$. □

Lemma 6.1.11 *Let I be a non-empty set of strictly positive integers that is closed for addition (i.e. for all $x, y \in I$ we have $x + y \in I$) and $d = \gcd I$. Then, the set I contains all the sufficiently large multiples of d.*

Remark: In particular, if $d = 1$, then there exists an integer $n_0 = n_0(I)$ such that $n \geq n_0 \Rightarrow n \in I$.

Proof: Since $d = \gcd I$, there exists necessarily a finite set of integers $\{a_1, \ldots, a_k\} \subseteq I$ such that $d = \gcd\{a_1, \ldots, a_k\}$. If c is divisible by d then Bezout theorem (stating that for all a and b strictly positive integers, there exist $u, v \in \mathbb{Z}$ such that $\gcd(a, b) = au + bv$) implies that there exist $u_1, \ldots, u_k \in \mathbb{Z}$ such that $c = a_1 u_1 + \ldots + a_k u_k$. For all $i = 2, \ldots, k$ the Euclidean division by a_1 yields $u_i = v_i a_1 + r_i$ with $v_i \in \mathbb{Z}$ and $r_i \in \mathbb{N}$ with $0 \leq r_i < a_1$. Hence,

$$c = a_1(u_1 + a_2 v_2 + \ldots + a_k v_k) + (a_2 r_2 + \ldots + a_k r_k),$$

where the first parenthesis is in $\mathbb{Z}$ and the second in $\mathbb{N}$. Therefore, on dividing both members by a_1, we get

$$\begin{aligned} \mathbb{Z} \ni (u_1 + a_2 v_2 + \ldots + a_k v_k) &= \frac{c}{a_1} - (a_2 \frac{r_2}{a_1} + \ldots + a_k \frac{r_k}{a_1}) \\ &\geq \frac{c}{a_1} - (a_2 + \ldots + a_k). \end{aligned}$$

For c sufficiently large, the left hand side becomes strictly positive, hence $c \in I$.
□

Lemma 6.1.12 *Let P be irreducible and possessing an aperiodic state x (i.e. $d_x = 1$). Then for every pair of states $y, z \in \mathbb{X}$, there exists an integer $n_0 = n_0(y, z) \geq 0$ such that for all $n \geq n_0$, we have $P^n(y, z) > 0$. In particular, all states are aperiodic.*

Proof: Let $I_x = \{n \geq 1 : \mathbb{P}_x(X_n = x) > 0\}$. Then I_x is closed for addition because, if $m, n \in I_x$ then $\mathbb{P}_x(X_{n+m} = x) \geq \mathbb{P}_x(X_n = x)\mathbb{P}_x(X_m = x) > 0$. Hence, by previous lemma 6.1.11, the set I_x contains all integers larger than a certain n_0 (because $d_x = 1$). Irreducibility means that for arbitrary states y, z there exist integers m_1 and m_2 such that $P^{m_1}(y, x) > 0$ and $P^{m_2}(x, z) > 0$. Hence, for all $n \geq n_0$, we have

$$P^{m_1+n+m_2}(y, z) \geq P^{m_1}(z, y) P^n(x, x) P^{m_2}(x, z) > 0.$$

□

Exercise 6.1.13 Let $\mu \in \mathscr{P}(\mathbb{Z})$ and $(\xi_n)_n$ a sequence of independent random variables identically distributed with law μ. Denote by $r = \sum_{z=-\infty}^{-1} \mu(\{z\})$ and consider the Markov chain $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu) X$ with $\mathbb{X} = \mathbb{N}$ and $\mathscr{X} = \mathscr{P}(\mathbb{N})$ defined by $X_0 = x \in \mathbb{X}$ and $X_{n+1} = (X_n + \xi_{n+1})^+$ for all $n \geq 0$.

1. Determine P.
2. Using a simple probabilistic argument, show that, depending on the value of r, the state 0 is
   - either accessible,
   - or inaccessible.

3. Suppose that μ charges all even integers and only them. Consider the partition of the space $\mathbb{X}$ into $\mathbb{X}_0 = 2\mathbb{N}$ and $\mathbb{X}_1 = 2\mathbb{N}+1$.
   – Are the sets $\mathbb{X}_0$, $\mathbb{X}_1$ accessible from any point $x \in \mathbb{X}$? Are they (stochastically) closed?
   – Are the sets $\mathbb{X}_0$, $\mathbb{X}_1$ composed of essential states?

6.2 Recurrence and transience

Definition 6.2.1 Let $x \in \mathbb{X}$. We say the state x is
- **recurrent** if $\mathbb{P}_x(X_n = x \text{ i.o}) = 1$,
- **transient** if $\mathbb{P}_x(X_n = x \text{ i.o}) = 0$.

Exercise 6.2.2 Let X be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ and T a stopping time. Conditionally on $T<\infty$ and $X_T = y$ the sequence $\tilde{X} = (X_{T+n})_{n\in\mathbb{N}}$ is a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\epsilon_y)$.

For $x \in \mathbb{X}$, define the sequence $(\tau_x^{(n)})_{n\in\mathbb{N}}$ of random times [1] recursively by

$$\begin{aligned} \tau_x^{(1)} &= \tau_x^1 \\ \tau_x^{(n+1)} &= \inf\{k > \tau_x^{(n)} : X_k = x\} \text{ on the set } \{\tau_x^{(n)} < \infty\}, \end{aligned}$$

and

$$T_x^{n+1} = \begin{cases} \tau_x^{(n+1)} - \tau_x^{(n)} & \text{if } \tau_x^{(n)} < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 6.2.3 *For $n \in \mathbb{N}^*$, conditionally on $\tau_x^{(n)} < \infty$, the time T_x^{n+1} is independent of $\mathscr{F}_{\tau_x^{(n)}}$ and*

$$\mathbb{P}(T_x^{n+1} = k | \tau_x^{(n)} < \infty) = \mathbb{P}_x(\tau_x^1 = k),$$

for all k.

Proof: For x and n fixed, denote by $\sigma = \tau_x^{(n)}$. Then, σ is a stopping time such that $X_\sigma = x$ on $\{\sigma < \infty\}$. On that set, the sequence $(X_{\sigma+n})_{n\in\mathbb{N}}$ is a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\epsilon_x)$, independent of $X_0,\ldots,X_\sigma$. □

Recall the notation for the occupation measure $\eta^0(x) := \eta^0(\{x\}) := \sum_{n=0}^{\infty} \mathbb{1}_{\{x\}}(X_n)$. Consequently,

$$\mathbb{E}_x(\eta^0(x)) = \sum_{n\in\mathbb{N}} \mathbb{P}_x(X_n = x) = \sum_{n\in\mathbb{N}} P_{xx}^n = G_{xx}^0 = \sum_{n\in\mathbb{N}} \mathbb{P}_x(\eta^0(x) > n),$$

and $L_x^1(x) := L_{\{x\}}(x) = \mathbb{P}_x(\tau_x^1 < \infty)$.

---

1. Beware of the difference between the symbol $\tau_x^{(n)}$, defined here, and the symbol $\tau_x^n = \inf\{k \geq n : X_k = x\}$, defined in proposition 5.2.10.

Lemma 6.2.4 *For $n = 0, 1, 2, \ldots$, we have $\mathbb{P}_x(\eta^0(x) > n) = (L_x(x))^n$.*

Proof: If $X_0 = x$, then for all $n \in \mathbb{N}$, we have: $\{\eta^0(x) > n\} = \{\tau_x^{(n)} < \infty\}$. For $n = 1$ the claim $\mathbb{P}_x(\eta^0(x) > n) = L_x(x)^n$ is true. Suppose that it remains true for n. Then

$$\begin{aligned}
\mathbb{P}_x(\eta^0(x) > n+1) &= \mathbb{P}_x(\tau_x^{(n+1)} < \infty) \\
&= \mathbb{P}_x(\tau_x^{(n)} < \infty \text{ and } T_x^{n+1} < \infty) \\
&= \mathbb{P}_x(T_x^{n+1} < \infty | \tau_x^{(n)} < \infty)\mathbb{P}_x(\tau_x^{(n)} < \infty) \\
&= L_x(x)(L_x(x))^n.
\end{aligned}$$

□

Theorem 6.2.5 (Recurrence/transience dichotomy) *For a $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \epsilon_x)$ the following dichotomy holds:*

1. *Either $\mathbb{P}_x(\tau_x^1 < \infty) = 1$, implying that the state x is recurrent and $G^0(x,x) = +\infty$,*
2. *or $\mathbb{P}_x(\tau_x^1 < \infty) < 1$, implying that the state x is transient and $G^0(x,x) < +\infty$.*

In particular, every $x \in \mathbb{X}$ is either recurrent or transient.

Proof:

1. Obviously, $\mathbb{P}_x(\tau_x^1 < \infty) = L_x(x) = 1$. The previous lemma 6.2.4 yields

$$\mathbb{P}_x(\eta^0(x) = \infty) = \lim_{r\to\infty} \mathbb{P}_x(\eta^0(x) > r) = \lim_{r\to\infty} (L_x(x))^r = 1.$$

   Hence x is recurrent and

$$G^0(x,x) = \sum_{n\in\mathbb{N}} P^n(x,x) = \mathbb{E}_x(\eta^0(x)) = \infty.$$

2. If $L_x(x) = \mathbb{P}_x(\tau_x < \infty) < 1$, then

$$G^0(x,x) = \sum_{n\in\mathbb{N}} P^n(x,x) = \mathbb{E}_x(\eta^0(x)) = \sum_{r\geq 0} \mathbb{P}_x(\eta^0(x) > r) = \sum_{r\geq 0} (L_x(x))^r = \frac{1}{1 - L_x(x)} < \infty,$$

   implying that $\mathbb{P}_x(\eta^0(x) = \infty) = 0$ (since otherwise $\mathbb{E}_x(\eta^0(x))$ diverges) and transience of x.

□

Remark: Recurrence is equivalent to almost sure finiteness of return time. However, it may happen that $\mathbb{P}_x(\tau_x < \infty) = 1$ and $m_x(x) = \mathbb{E}_x(\tau_x) = \infty$; we say that x is **null recurrent**. If, on the contrary $m_x(x) = \mathbb{E}_x(\tau_x) < \infty$ occurs, we say that x is **positive recurrent**.

Exercise 6.2.6 Recurrence and transience are class properties.

In summarising, we have the following equivalences for $x \in \mathbb{X}$:

$$\begin{aligned} x \text{ recurrent} &\Leftrightarrow L_x(x) = 1 \Leftrightarrow \mathbb{P}_x(X_n = x \text{ i.o.}) = 1 \\ &\Leftrightarrow \mathbb{P}_x(\eta^0(x) = \infty) = 1 \Leftrightarrow H_x(x) = 1 \Leftrightarrow G^0(x,x) = \infty \\ x \text{ transient} &\Leftrightarrow L_x(x) < 1 \Leftrightarrow \mathbb{P}_x(X_n = x \text{ i.o.}) = 0 \\ &\Leftrightarrow \mathbb{P}_x(\eta^0(x) < \infty) = 1 \Leftrightarrow H_x(x) = 0 \Leftrightarrow G^0(x,x) < \infty. \end{aligned}$$

Definition 6.2.7 Let (X_n) be a Markov chain on a denumerably infinite set and $B \in \mathcal{X}$. The set B is called

- **transient** if $H_B(x) := \mathbb{P}_x(\liminf_n \{X_n \in B\}) = 0$ for all $x \in B$,
- **almost closed** if $0 < \mathbb{P}_x(\liminf_n \{X_n \in B\}) = \mathbb{P}_x(\limsup_n \{X_n \in B\})$,
- **recurrent** if B is a communicating class and $L_B(x) = 1$ for all $x \in B$.

The chain is called

- **recurrent** if for all $x \in \mathbb{X}$, $[x] = \mathbb{X}$ and x is recurrent,
- **transient** if all its recurrent classes are absorbing,
- **tending to infinity** if for every finite $F \in \mathcal{X}$, we have $\mathbb{P}_x(\tau_{F^c} < \infty) = 1$.

Exercise 6.2.8
1. Every recurrent class is stochastically closed.
2. All stochastically closed equivalence classes consisting of *finitely* many states are recurrent.

Exercise 6.2.9 If the chain is irreducible and transient, then every finite set $F \in \mathcal{X}$ is transient.

Exercise 6.2.10 If the chain is irreducible and recurrent then all bounded (or non-negative) harmonic functions on $\mathbb{X}$ are constant. (Therefore, if there exist a non-constant bounded (or non-negative) harmonic function, the chain is transient.)

6.3 Invariant measures

In this section, we consider a Markov chain X on $(\mathbb{X}, \mathcal{X})$ associated with the Markov operator P. We denote by $(\mathcal{F}_n)$ the natural filtration of the chain.

Definition 6.3.1 A measure $\mu \in \mathcal{M}_+(\mathcal{X})$ is said **invariant**, or **stationary**, or **equilibrium** , if $\mu P = \mu$.

Remark: Any P-harmonic measure μ is invariant and vice-versa. If μ is not a probability measure, we extend the notion of standard Markov chain $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ by introducing the measure $\mathbb{P}_\mu$ on the trajectory space defined by its desintegration $\mathbb{P}_\mu = \int_{\mathbb{X}} \mu(dx)\mathbb{P}_x$ into its probability components $\mathbb{P}_x$.

Theorem 6.3.2 *Let $A \in \mathscr{X}$; define the average occupation measure between excursions $\nu_x(A) = \mathbb{E}_x(\sum_{n=0}^{\tau_x-1} 1\!\!1_A(X_n))$. If P is irreducible and recurrent then*

1. *$\nu_x(x) = 1$, for all $x \in \mathbb{X}$,*
2. *$\nu_x P = \nu_x$,*
3. *$0 < \nu_x(y) < \infty$, for all $y \in \mathbb{X}$.*

Proof:

1. It is evident that between times 0 and $\tau_x - 1$, the chain X is at x only at instant 0.
2. For $n \geq 1$, $\{\tau_x \geq n\} = \{\tau_x < n\}^c = \{\tau_x \leq n-1\} \in \mathscr{F}_{n-1}$. Thus for y and z different from x:

$$\begin{aligned}\mathbb{P}_x(X_{n-1} = y, X_n = z, \tau_x \geq n) &= \mathbb{P}_x(X_{n-1} = y | X_n = z, \tau_x \geq n)\mathbb{P}_x(X_n = z, \tau_x \geq n)\\ &= \mathbb{P}_x(X_{n-1} = y, \tau_x \geq n)P(y,z).\end{aligned}$$

Since P is recurrent, we have $\mathbb{P}_x(\tau_x < \infty, X_{\tau_x} = x) = 1$. For $z \neq x$:

$$\begin{aligned}\nu_x(z) &= \mathbb{E}_x(\sum_{n=0}^{\tau_x-1} 1\!\!1_{\{z\}}(X_n)) = \mathbb{E}_x(\sum_{n=1}^{\tau_x} 1\!\!1_{\{z\}}(X_n))\\ &= \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = z, n \leq \tau_x)\\ &= \sum_{y\in\mathbb{X}} \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = z, X_{n-1} = y, n \leq \tau_x)\\ &= \sum_{y\in\mathbb{X}} \sum_{n=1}^{\infty} \mathbb{E}_x(1\!\!1_{X_{n-1}=y} 1\!\!1_{n\leq\tau_x} \mathbb{P}(X_n = z|\mathscr{F}_{n-1}))\\ &= \sum_{y\in\mathbb{X}} \sum_{n=1}^{\infty} \mathbb{E}_x(P(X_{n-1}, z) 1\!\!1_{X_{n-1}=y} 1\!\!1_{n\leq\tau_x})\\ &= \sum_{y\in\mathbb{X}} \sum_{n=0}^{\infty} \mathbb{P}_x(X_{n-1} = y, n \leq \tau_x)P(y,z)\\ &= \sum_{y\in\mathbb{X}} \mathbb{E}_x(\sum_{n=0}^{\tau_x-1} 1\!\!1_{\{y\}}(X_n = y))P(y,z)\\ &= \sum_{y\in\mathbb{X}} \nu_x(y)P(y,z).\end{aligned}$$

3. Irreducibility of P implies that for all $x, y \in \mathbb{X}$, there exist integers $n, m \geq 0$, depending on x, y such that $P^n(x, y) > 0$ and $P^m(y, x) > 0$. Hence

$$\nu_x(y) = \sum_{x' \in \mathbb{X}} \nu_x(x') P^n(x', y) \geq \nu_x(x) P^n(x, y) = P^n(x, y) > 0.$$

Similarly, $\nu_x(y) P^m(y, x) \leq \nu_x(x) = 1$, implying that $\nu_x(y) \leq \frac{1}{P^m(y,x)} < \infty$.

□

Remark: Thus, ν_x is a σ-finite measure that is harmonic for P.

Theorem 6.3.3 *Let P be irreducible and $\mu \in \mathcal{M}_+(\mathcal{X})$ such that*
- *μ is P-harmonic, and*
- *$\mu(x) = 1$ for some $x \in \mathbb{X}$.*

Then $\mu \geq \nu_x$. If P is further recurrent, then $\mu = \nu_x$.

Proof: Harmonicity of μ means that for all $z \in \mathbb{X}$, we have:

$$\begin{aligned}
\mu(z) &= \sum_{y_0 \in \mathbb{X}} \mu(y_0) P(y_0, z) \\
&= \mu(x) P(x, z) + \sum_{y_0 \neq x} \mu(y_0) P(y_0, z) \\
&= P(x, z) + \sum_{y_0 \neq x} (\sum_{y_1 \in \mathbb{X}} \mu(y_1) P(y_1, y_0)) P(y_0, z) \\
&= P(x, z) + \sum_{y_0 \neq x} P(x, y_0) P(y_0, z) + \sum_{y_0, y_1 \neq x} \mu(y_1) P(y_1, y_0) P(y_0, z) \\
&\vdots \\
&= P(x, z) + \sum_{y_0 \neq x} P(x, y_0) P(y_0, z) + \ldots + \sum_{y_0, \ldots, y_{n-1} \neq x} P(x, y_0) \cdots P(y_{n-1}, y_{n-1} P(y_{n-1}, z) \\
&\quad + \sum_{y_0, \ldots, y_{n-1} \neq x, y_n \in \mathbb{X}} \mu(y_n) P(y_n, y_{n-1}) \cdots P(y_0, z) \\
&\geq P(x, z) + \sum_{y_0 \neq x} P(x, y_0) P(y_0, z) + \ldots + \sum_{y_0, \ldots, y_{n-1} \neq x} P(x, y_0) \cdots P(y_{n-1}, y_{n-1} P(y_{n-1}, z) \\
&= \mathbb{P}_x(X_1 = z, \tau_x \geq 1) + \ldots + \mathbb{P}_x(X_n = z, \tau_x \geq n) \\
&\underset{n \to \infty}{\longrightarrow} \mathbb{E}_x(\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = z; \tau_x \geq n\}}) \\
&= \nu_x(z).
\end{aligned}$$

If P is additionally recurrent, then ν_x is harmonic by theorem 6.3.2 and $\kappa = \mu - \nu_x \geq 0$ by the previous result and further κ is harmonic as a difference of two harmonic measures. Irreducibility of P implies that for every $z \in \mathbb{X}$, there exists an integer n such that $P^n(z, x) > 0$. Now, harmonicity of κ reads:

$$\kappa(x) = \mu(x) - \nu_x(x) = 1 - 1 = 0 = \sum_{z' \in \mathbb{X}} \kappa(z') P^n(z', x) \geq \kappa(z) P^n(z, x),$$

implying that $0 \le \kappa(z) \le \frac{\kappa(x)}{P^n(z,x)} = 0$. □

Theorem 6.3.4 *Let P be irreducible. The following are equivalent:*

1. *All states are positive recurrent.*
2. *There exists a positive recurrent state.*
3. *Let $\pi \in \mathscr{M}_+(\mathscr{X})$ be defined by $\pi(x) = \frac{1}{m_x(x)}$, for all $x \in \mathbb{X}$. Then π is a P-harmonic probability measure.*

Proof:

$1 \Rightarrow 2$: Obvious.

$2 \Rightarrow 3$: Suppose that the state x is positive recurrent, hence recurrent. Irreducibility of P implies then recurrence of P. The theorem 6.3.2 guarantees then that ν_x is P-harmonic. Now,

$$\nu_x(\mathbb{X}) = \sum_{y \in \mathbb{X}} \nu_x(y) = \mathbb{E}_x\Big(\sum_{n=0}^{\tau_x - 1} \mathbb{1}_{\mathbb{X}}(X_n)\Big) = \mathbb{E}_x \tau_x = m_x(x) < \infty,$$

hence $\pi = \frac{\nu_x}{\nu_x(\mathbb{X})} = \frac{\nu_x}{m_x(x)}$ is a P-harmonic probability measure, verifying further $\pi(x) = \frac{\nu_x(x)}{m_x(x)} = \frac{1}{m_x(x)}$.

$3 \Rightarrow 1$: For every $x, y \in \mathbb{X}$, irreducibility of P implies the existence of an integer n such that $P^n(x,y) > 0$. Invariance of π reads: $\pi(y) = \sum_{z \in \mathbb{X}} \pi(z) P^n(z,y) \ge \pi(x) P^n(x,y) > 0$. Since π is a probability, $\pi(\mathbb{X}) = \sum_{y \in \mathbb{X}} \pi(y) = 1$, on defining $\lambda(y) = \frac{\pi(y)}{\pi(x)}$ yields $\lambda(x) = 1$ and harmonicity of λ. By virtue of theorem 6.3.3, $\lambda \ge \nu_x$. Consequently,

$$m_x(x) = \nu_x(\mathbb{X}) = \sum_{y \in \mathbb{X}} \le \frac{\sum_{y \in \mathbb{X}} \pi(y)}{\pi(x)} = \frac{\pi(\mathbb{X})}{\pi(x)} = \frac{1}{\pi(x)} < \infty,$$

implying positive recurrence of x, hence recurrence. Since P is irreducible and x is arbitrary P itself is recurrent and the previous inequality is as a matter of fact an equality. Therefore, each x is positive recurrent.

□

Exercise 6.3.5 Let X be a simple symmetric random walk on $\mathbb{Z}$. Show that

– the kernel P is irreducible and recurrent;
– the counting measure κ on $\mathbb{Z}$ is invariant;
– every other invariant measure for P is a scalar multiple of κ;
– the chain is null recurrent.

Exercise 6.3.6 Recurrence combined with irreducibility guarantees the existence of an invariant measure. Give an example showing that the converse is in general false if the invariant measure is not a *probability*.

6.4 Convergence to equilibrium

Although direct spectral methods, very much comparable to the finite case exist to prove convergence to equilibrium for chains on denumerably infinite spaces, we give here a purely probabilistic method, based on the idea of coupling.

Theorem 6.4.1 *Let X be a* $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ *and suppose that P is irreducible and aperiodic and possesses an invariant probability π. Then for every initial $\mu\in\mathscr{M}_1(\mathscr{X})$ we have*

$$\lim_{n\to\infty}\mathbb{P}_\mu(X_n=y)=\pi(y),\ \forall y\in\mathbb{X}.$$

In particular $\lim_{n\to\infty}P^n(x,y)=\pi(y)$, *for all $x,y\in\mathbb{X}$.*

Proof: Let Y be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\pi)$ independent of X. Considering simultaneously both chains, requires extending the trajectory space. Thus the sequence of pairs $(X_n,Y_n)_{n\in\mathbb{N}}$ is a $\mathsf{MC}((\mathbb{X}\times\mathbb{X},\mathscr{X}\otimes\mathscr{X}),\hat{P},\mu\otimes\pi)$, where $\hat{P}((x,y),(x',y'))=P(x,x')P(y,y')$. On the underlying abstract probability space $(\Omega,\mathscr{F},\mathbb{P})$ (extended if necessary to hold the sequence $(X_n,Y_n)_{n\in\mathbb{N}}$), we define, for an arbitrary fixed state $z_0\in\mathbb{X}$, the random time $T=\inf\{n\geq 1: X_n=Y_n=z_0\}$.

First step: establishing $\mathbb{P}_{\mu\otimes\pi}(T<\infty)=1$. Aperiodicity and irreducibility of P implies that there exists a sufficiently large integer n such that $\hat{P}^n((x,y),(x',y'))>0$, for all $x,x',y,y'\in\mathbb{X}$. Therefore, $\hat{P}$ is irreducible. Obviously $\hat{P}$ verifies the invariance condition $(\pi\otimes\pi)\hat{P}=\pi\otimes\pi$ with an invariant probability $\pi\otimes\pi\in\mathscr{M}_1(\mathscr{X}\otimes\mathscr{X})$; consequently, $\hat{P}$ is positive reurrent. As such, it verifies $\hat{\mathbb{P}}_{\hat{\mu}}(\hat{\tau}_{(z_0,z_0)}<\infty)=1$. Now, it is immediately seen that $T=\hat{\tau}_{(z_0,z_0)}$, establishing thus that the two independent chains X and Y will meet at point z_0 in an almost surely finite time.

Second step: coupling. Define new $(\mathbb{X},\mathscr{X})$-valued processes $(Z_n)_{n\in\mathbb{N}}$ and $(Z'_n)_{n\in\mathbb{N}}$ as follows:

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

$$Z'_n = \begin{cases} Y_n & \text{if } n<T \\ X_n & \text{if } n\geq T. \end{cases}$$

These processes allow introducing new $(\mathbb{X}\times\mathbb{X},\mathscr{X}\otimes\mathscr{X})$ processes W and W' by defining $W_n=(X_n,Y_n)$ and $W'_n=(Z_n,Z'_n)$, for all $n\in\mathbb{N}$. Now strong Markov property implies that the translated chain $(W_{T+n})_{n\in\mathbb{N}}$ is a $\mathsf{MC}((\mathbb{X}\times\mathbb{X},\mathscr{X}\otimes\mathscr{X}),\hat{P},\epsilon_{(z_0,z_0)})$, independent of $(X_0,Y_0),\ldots,(X_{T-1},Y_{T-1})$. Similarly, $(W'_n)_{n\in\mathbb{N}}$ is a $\mathsf{MC}((\mathbb{X}\times\mathbb{X},\mathscr{X}\otimes\mathscr{X}),\hat{P},\mu\otimes\pi)$, while $(Z_n)_{n\in\mathbb{N}}$ is a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$.

Third step: establishing convergence to equilibrium. By virtue of the definition of (Z_n), we have:

$$\hat{\mathbb{P}}_{\mu\otimes\pi}(Z_n = y) = \hat{\mathbb{P}}_{\mu\otimes\pi}(X_n = y; n < T) + \hat{\mathbb{P}}_{\mu\otimes\pi}(Y_n = y; n \geq T).$$

Now

$$\begin{aligned}
|\mathbb{P}_\mu(X_n = x) - \pi(x)| &= |\hat{\mathbb{P}}_{\mu\otimes\pi}(Z_n = x) - \pi(x)| \\
&= |\hat{\mathbb{P}}_{\mu\otimes\pi}(Z_n = x) - \mathbb{P}_\pi(Y_n = x)| \\
&= |\hat{\mathbb{P}}_{\mu\otimes\pi}(Z_n = x; n < T) - \hat{\mathbb{P}}_{\mu\otimes\pi}(Z'_n = x; n < T)| \\
&\leq \hat{\mathbb{P}}_{\mu\otimes\pi}(n < T) \\
&\underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}$$

the last line holding because $T < \infty$ almost surely. □

Proposition 6.4.2 *Let P be irreducible. Then, there exist an integer $d \geq 1$ and a partition $\mathbb{X} = \mathbb{X}_0 \sqcup \ldots \sqcup \mathbb{X}_{d-1}$ such that*

1. *if $x \in \mathbb{X}_i$ and $y \in \mathbb{X}_j$, then $P^n(x, y) = 0$ unless $n-(j-i)$ is divisible by d, and*
2. *for all $r \in \{0, \ldots, d-1\}$, for all $x, y \in \mathbb{X}_r$, there exists $n_0 = n_0(x, y)$ such that for all $n \geq n_0$ we have $P^{nd}(x, y) > 0$.*

Proof: Let $d = d_x$ for some $x \in \mathbb{X}$. Since the chain is irreducible, it is composed of a single communicating class, say $[x] = \mathbb{X}$. Hence, for all $y \in \mathbb{X}$, $d_y = d$. Let m_1 be such that $\mathbb{P}_x(X_{m_1} = y) > 0$ and m_2 such that $\mathbb{P}_y(X_{m_2} = x) > 0$. Then, $\mathbb{P}_y(X_{m_1+m_2} = y) \geq \mathbb{P}_y(X_{m_2} = x)\mathbb{P}_x(X_{m_1} = y) > 0$, therefore $m_1 + m_2$ is divisible by d. If $\mathbb{P}_x(X_n = y) > 0$, then $\mathbb{P}_y(X_{n+m_2} = y) > 0$; hence $n+m_2$ is divisible by d. Combined with the divisibility of $m_1 + m_2$ yields $n - m_1$ divisible by d; otherwise stated, there exists an integer $k \geq 0$ such that $n = kd + m_1$. But for the same reason, $m_1 = ld + r$ with some integer $l \geq 0$ and some $r \in \{0, \ldots, d-1\}$. Thus, $n = (k+l)d + r$, with $k+l \geq 0$, yielding $P^n(x, y) > 0$ only if $n = jd + r$. Chose an arbitrary $a \in \mathbb{X}$ and define

$$\mathbb{X}_r = \{y \in \mathbb{X} : \mathbb{P}_a(X_{md+r} = y) > 0, \text{ for an } m \geq 0\}.$$

If $x \in \mathbb{X}_i$, $y \in \mathbb{X}_j$, and $P^n(x, y) > 0$, then there exists $m_0 \geq 0$ such that $\mathbb{P}_a(X_{m_0 d+i} = x) > 0$, hence $\mathbb{P}_a(X_{m_0 d+i+n} = y) > 0$. Therefore, $n + m_0 d + i = kd + j$ for some $k \geq 0$, showing that $n - (j - i)$ is a multiple of d.

The second claim is evident. □

Remark: When $X_n \in \mathbb{X}_r$ implies $X_{n+1} \in \mathbb{X}_{r+1 \mod d}$, we say that $(\mathbb{X}_r)_{r=0,\ldots,d-1}$ are the **cyclic classes** of the chain. The corresponding partition is the **cycle decomposition** of the state space.

Proposition 6.4.3 *Let P be irreducible with period $d \geq 1$, $\mathbb{X}_0, \ldots, \mathbb{X}_{d-1}$ its cyclic classes, and $\mu \in \mathcal{M}_1(\mathcal{X})$ an arbitary probability such that $\mu(\mathbb{X}_0) = 1$. Then for all $r \in \{0, \ldots, d-1\}$ and all $y \in \mathbb{X}_r$,*

$$\lim_{n\to\infty} \mathbb{P}_\mu(X_{nd+r} = y) = \frac{d}{m_y(y)}.$$

In particular, for all $x \in \mathbb{X}_0$, we have $\lim_{n\to\infty} P^{nd+r}(x, y) = \frac{d}{m_y(y)}$.

Proof: On defining $\nu = \mu P^r$, we see that $\nu(\mathbb{X}_r) = 1$, by virtue of the previous theorem. Now the process $(\tilde{X}_n)_{n\in\mathbb{N}}$ defined by $\tilde{X}_n = X_{nd+r}$ is a $\mathsf{MC}(\mathbb{X}_r, \mathcal{X}_r), P^n, \nu)$ that is irreducible and aperiodic on $\mathbb{X}_r$. For $\tilde{\tau}_y = \inf\{n \geq 1 : \tilde{X}_n = y\}$, we have for all $y \in \mathbb{X}_r$: $\mathbb{E}_y(\tilde{\tau}_y) = \frac{m_y(y)}{d}$. Hence $\mathbb{P}_\mu(X_{nd+r} = y) = \tilde{\mathbb{P}}_\nu(\tilde{X}_n = y) \to \frac{m_y(y)}{d}$. □

Remark: If $m_y(y) = \infty$ (i.e. the state — hence the chain — is either transient or null recurrent), then $\lim_{n\to\infty} P^{nd+r}(x, y) = 0$, for all x. We say that the mass **escapes at infinity**.

6.5 Ergodic theorem

The archetype of ergodic[2] theorems in probability is the strong law of large numbers, stating that for a sequence of independent identically distributed random variables (ξ_n) with $\mathbb{E}\xi_1 = m \in [0, \infty]$, the "time" average $\frac{\xi_1 + \ldots + \xi_n}{n}$ converges almost surely to the "space" average m. This theorem has several extensions to variables that are neither independent nor identically distributed. For Markov chains it reads:

Theorem 6.5.1 (Ergodic theorem for Markov chains) *Let P be irreducible and $\mu \in \mathcal{M}_1(\mathcal{X})$. Then,*

1. *the empirical occupation probability*[3] *converges almost surely, more precisely*

$$\lim_{n\to\infty} \frac{\eta_n^0(x)}{n} = \frac{1}{m_x(x)}, \mathbb{P}_\mu\text{-a.s.};$$

2. The term "ergodic" has been coined by the Austrian physicist Ludwig Boltzmann. In its initial formulation it applied to macroscopic physical systems and gave, for the first time, an explanation why physical systems macroscopically evolve irreversibly towards equilibrium although microscopically they are governed by time-reversible physical laws. Boltzmann also proved, under the so called ergodic hypothesis, that if physical quantities are averaged over large times, statistical fluctuations disappear so that they attain a deterministic value. Nowadays, ergodic theory is an independent branch of mathematics, at the (common!) frontier of measure theory, geometry, group theory, operator algebras, stochastic processes, etc.
3. We recall that the occupation measure up to time n reads $\eta_n^0(x) = \sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k)$.

2. *if additionally, the chain is positive recurrent, then for all $f \in b\mathscr{X}$, the trajectory averages of f converge almost surely to the integral of f with respect to the invariant measure, more precisely*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(X_k) = \overline{f} := \int_{\mathbb{X}} \pi(dx) f(x) = \sum_{x\in\mathbb{X}} \pi(x) f(x),$$

where π is the unique invariant probability of the chain.

Proof:

1. If P is transient then every state is visited only finitely many times, so that $\eta^0(x) < \infty$ almost surely. Thus

$$\frac{\eta_n^0(x)}{n} \le \frac{\eta^0(x)}{n} \to 0 = \frac{1}{m_x(x)}$$

and the claim is proved.

Suppose therefore that P is recurrent and let $x \in \mathbb{X}$ be fixed. Recurrence of x entails $\tau_x < \infty$ almost surely. Thus the process $\tilde{X}$ defined, for all n, by $\tilde{X}_n = X_{\tau_x+n}$ is a $\mathsf{MC}(\mathbb{X}, \mathscr{X}), P, \epsilon_x)$ independent of $X_0, \ldots, X_{\tau_x}$. Thus, for $n > \tau_x$, we have: $\eta_n^0(x) = \eta_{\tau_x}^0(x) + \tilde{\eta}_{n-\tau_x}^0(x)$. Since $\frac{\eta_{\tau_x}^0(x)}{n} \le \frac{\tau_x}{n} \to 0$ (because $\mathbb{P}_x(\tau_x < \infty) = 1$), it is enough to consider $\mu = \epsilon_x$.

Recall that we have introduced the sequence of random times $T_x^{r+1} = \tau_x^{(r+1)} - \tau_x^{(r)}$ as the time span between two successive returns to x and shown in lemma 6.2.3 that $\mathbb{P}_x(T_x^{(r+1)} = n | \tau_x^{(r)} < \infty) = \mathbb{P}_x(\tau_x = n)$. Therefore, $\mathbb{E}_x(T_x^{(r)}) = m_x(x)$ for all r. Now, $T_x^{(1)} + \ldots + T_x^{(\eta_n^0(x)-1)} \le n-1$ while $T_x^{(1)} + \ldots + T_x^{(\eta_n^0(x))} \ge n$, so that

$$\frac{T_x^{(1)} + \ldots + T_x^{(\eta_n^0(x)-1)}}{\eta_n^0(x)} \le \frac{n}{\eta_n^0(x)} \le \frac{T_x^{(1)} + \ldots + T_x^{(\eta_n^0(x))}}{\eta_n^0(x)}.$$

Since the random variables $T_x^{(r)}$ are independent and identically distributed, the strong law of large numbers guarantees that $\lim_{n\to\infty} \frac{T_x^{(1)}+\ldots+T_x^{(n)}}{n} = m_x(x)$ almost surely. On the other hand, recurrence of P guarantees that $\lim_{n\to\infty} \eta_n^0(x) = \infty$. Combining these two results with the above inequalities implies that $\lim_{n\to\infty} \frac{n}{\eta_n^0(x)} = \frac{1}{m_x(x)}$.

2. If P possesses an invariant probability π, then

$$|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}| = |\sum_{x\in\mathbb{X}} (\frac{\eta_n^0(x)}{n} - \pi(x)) f(x)|.$$

Since $f \in b\mathscr{X}$, we can without loss of generality assume that $|f|_\infty \le 1$, so that for any subset $\mathbb{Y} \subseteq \mathbb{X}$ the right hand side of the above equation can be majorised by

$$\begin{aligned}|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k)-\overline{f}| &\leq |\sum_{x\in\mathbb{Y}}(\frac{\eta_n^0(x)}{n}-\pi(x))f(x)|+|\sum_{x\in\mathbb{Y}^c}(\frac{\eta_n^0(x)}{n}-\pi(x))f(x)| \\ &\leq |\sum_{x\in\mathbb{Y}}\frac{\eta_n^0(x)}{n}-\pi(x)|+\sum_{x\in\mathbb{Y}^c}(\frac{\eta_n^0(x)}{n}+\pi(x)) \\ &\leq 2\sum_{x\in\mathbb{Y}}|\frac{\eta_n^0(x)}{n}-\pi(x)|+2\sum_{y\in\mathbb{Y}^c}\pi(x).\end{aligned}$$

Now, $\pi(\mathbb{X})=1$, hence for all $\epsilon>0$, there exists a finite $\mathbb{Y}\subseteq\mathbb{X}$ such that $\pi(\mathbb{Y}^c)<\epsilon/4$. Since, on the other hand, $\mathbb{P}_\mu(\frac{\eta_n^0(x)}{n}\to\pi(x),$ for all $x\in\mathbb{X}\}=1$, for that ϵ we can chose $N=N(\omega,\epsilon)$ such that for $n\geq N$, $\sum_{y\in\mathbb{Y}}|\frac{\eta_n^0(x)}{n}-\pi(x)|<\epsilon/4$. Thus, finally, $|\frac{1}{n}\sum_{k=0}^{n-1}f(X_k)-\overline{f}|<\epsilon$.

□

6.6 Semi-martingale techniques for denumerable Markov chains

The general setting in this section will be that of a $X=(X_n)_{n\in\mathbb{N}}$. The chain will be of $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\epsilon_x)$; however, the underlying abstract probability space $(\Omega,\mathscr{F},\mathbb{P})$ will also be tacitly considered, either as identified with the trajectory space or larger than it. Measurable functions $f:\mathbb{X}\to\mathbb{R}_+$ will be also considered throughout the section verifying some conditions; applied on the chain, they transform it into a semi-martingale process $Z_n=f(X_n)$. The asymptotic properties of the chain are obtained as stability properties of the process obtained by transformation. The functions f allowing such a study are termed **Lyapunov functions**. They constitute an essential component of the constructive theory of Markov chains [8].

6.6.1 Foster's criteria

In [9], Foster established criteria for recurrence/transience of Markov chains. Here we give a slightly more general form of these criteria.

Theorem 6.6.1 (Foster's criterion for positive recurrence) *Let X be an irreducible $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$. The following are equivalent:*

1. *X is positive recurrent.*
2. *There exists a $f\in m\mathscr{X}_+$, an $\epsilon>0$, and a finite $F\in\mathscr{X}$ such that:*

(a) $\mathbb{E}(f(X_{n+1}) - f(X_n)|X_n = y) \le -\epsilon$ *if* $y \notin F$,

(b) $\mathbb{E}(f(X_{n+1})|X_n = y) \le \infty$ *if* $y \in F$.

Proof:

$1 \Rightarrow 2$: Let $x \in \mathbb{X}$ be fixed, $F = \{x\}$, and $\tau_F = \inf\{n \ge 1 : X_n \in F\}$. Positive recurrence entails that for every $\mathbb{E}_y(\tau_F) < \infty$ for all $y \in F$. Define subsequently:

$$f(y) = \begin{cases} 0 & \text{if } y \in F, \\ \mathbb{E}_y(\tau_F) & \text{if } y \notin F. \end{cases}$$

Now, for $y \notin F$:

$$\begin{aligned} \mathbb{E}(f(X_{n+1}) - f(X_n)|X_n = y) &= \sum_{z\in\mathbb{X}} P_{yz} f(z) - f(y) \\ &= \sum_{z\in F^c} P_{yz}\mathbb{E}_z(\tau_F) + \sum_{z\in F^c} P_{yz} \cdot 0 - \mathbb{E}_y(\tau_F) \\ &= \mathbb{E}_y(\tau_F - 1) - \mathbb{E}_y(\tau_F) = -1, \end{aligned}$$

the last line following from the observation, valid for all $y \in F$,

$$\begin{aligned} \sum_{z\in F^c} P_{yz}\mathbb{E}_z(\tau_F) &= \sum_{z\in F^c} P_{yz} \sum_{k=1}^{\infty} k\mathbb{P}_z(\tau_F = k) \\ &= \sum_{k=1}^{\infty} k\mathbb{P}_y(\tau_F = k+1) \\ &= \sum_{k=1}^{\infty} k\mathbb{P}_y(\tau_F - 1 = k) = \mathbb{E}_y(\tau_F - 1). \end{aligned}$$

This observation establishes claim a). Additionally, the last equation is equivalent to stating:

$$\infty > \mathbb{E}_y(\tau_F) - 1 = \sum_{z\in F^c} P_{yz}\mathbb{E}_z(\tau_F) = \sum_{z\in F^c} P_{yz} f(z) = \mathbb{E}(f(X_{n+1})|X_n = y),$$

establishing thus claim b).

$2 \Rightarrow 1$: For $n \ge 0$, define $\mathscr{F}_n = \sigma(X_0, \ldots, X_n)$, $W_n = f(X_n)$, and $\tau_F = \inf\{n \ge 1 : X_n \in F\}$. Then, claim a) entails

$$\mathbb{E}(W_{(n+1)\wedge\tau_F} - W_{n\wedge\tau_F}|X_n = y) \le -\epsilon \mathbb{1}_{\tau_F > n}.$$

Note that $\tau_F > n$ means, in particular, that $y \notin F$. For $y \notin F$, the last inequality and theorem C.1.1 imply $\mathbb{E}_y(\tau_F) \le \frac{\mathbb{E}_y(\tau_F)}{\epsilon} = \frac{f(y)}{\epsilon} < \infty$. Now, since for all y we have:

$$\begin{aligned} \mathbb{E}_y(\tau_F) &= 1 + \sum_{z\in F^c} P_{yz}\mathbb{E}_z(\tau_F) \le 1 + \frac{1}{\epsilon}\sum_{z\in F^c} P_{yz} f(z) \\ &\le 1 + \frac{1}{\epsilon}\sum_{z\in\mathbb{X}} P_{yz} f(z) = \frac{1}{\epsilon}\mathbb{E}(f(X_{n+1})|X_n = y) < \infty. \end{aligned}$$

In all situations, we have for all $z \in \mathbb{X}$, $\mathbb{E}_z(\tau_F) < \infty$, implying positive recurrence of the chain because for one z, hence for all (due to irreducibility), $\mathbb{E}_z(\tau_z) < \infty$.

□

Theorem 6.6.2 *Let X be an irreducible* $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$, *where* $\mathrm{card}\mathbb{X} = \aleph_0$. *The following are equivalent:*

1. *X is transient.*
2. *There exists a $f \in m\mathscr{X}_+$ and a set $A \in \mathscr{X}$ such that:*
   (a) $\mathbb{E}(f(X_{n+1}) - f(X_n)|X_n = y) \leq 0$,
   (b) *there exists $y \notin A$ such that* $f(y) < \inf_{z\in A} f(z)$.

Proof:

$2 \Rightarrow 1$: Define $W_n = f(X_{n\wedge\tau_A})$ for $n \geq 0$. Then, claim a) guarantees that (W_n) is a positive supermartingale. For arbitrary $x \in \mathbb{X}$, we have

$$\mathbb{E}_x W_n = \mathbb{E}_x W_0 + \sum_{k=0}^{n-1} \mathbb{E}_x(W_{k+1} - W_k).$$

On the event $\{\tau_A \leq n\}$ the summand on the right hand side of the above equality is 0 while on $\{\tau_A > n\}$ is negative. Therefore, in all siutations, $\mathbb{E}_x W_n \leq \mathbb{E}_x W_0 = f(x)$. Suppose that the chain is not transient. Then $\mathbb{P}_x(\tau_A < \infty) = 1$. Since $W_n \to W_\infty$ almost surely, by vitrue of the supermartingale convergence theorem, we have equivalently: $f(X_{n\wedge\tau_A}) \to f(X_{\tau_A})$, almost surely. Fatou's lemma yields:

$$\mathbb{E}_x(f(X_{\tau_A})) = \mathbb{E}_x(\liminf_n W_n) \leq \liminf_n \mathbb{E}_x W_n \leq \mathbb{E}_x W_0 = f(x),$$

which contradicts claim b) (because $X_{\tau_A} \in A$). We conclude *ad absurdum* that the chain must be transient.

$1 \Rightarrow 2$: Fix some $x \in \mathbb{X}$ and define $A = \{x\}$ and $f : \mathbb{X} \to \mathbb{R}_+$ by

$$f(y) = \begin{cases} 1 & \text{if } y \in A \\ \mathbb{P}_y(\tau_A < \infty) & \text{if } y \notin A. \end{cases}$$

Then, for $y \notin A$:

$$\begin{aligned} \mathbb{E}(f(X_{n+1}) - f(X_n)|X_n = y) &= \sum_{z\in\mathbb{X}} P_{yz} f(z) - f(y) \\ &= P_{yx} + \sum_{z\in A^c} P_{yz} f(z) - f(y) \\ &= \mathbb{P}_y(\tau_A = 1) + \sum_{z\in A^c} P_{yz} \sum_{N=1}^{\infty} \mathbb{P}_z(\tau_A = N) - \sum_{N=1}^{\infty} \mathbb{P}_y(\tau_A = N) \\ &= 0, \end{aligned}$$

guaranteeing thus a). If the chain is transient, then for all $y \neq x$:

$$f(y) = \mathbb{P}_y(\tau_A < \infty) < 1 = f(x) = \inf_{z \in A} f(z).$$

□

Since the state space $\mathbb{X}$ is not necessarily ordered, we need the following

Definition 6.6.3 On a denumerably infinite set $\mathbb{X}$, we say a function $f : \mathbb{X} \to \mathbb{R}_+$ **tends to infinity**, and write $f \to \infty$, if for every $n \in \mathbb{N}$, the set $\mathbb{X}_n = \{x \in \mathbb{X} : f(x) \leq n\}$ is finite.

Theorem 6.6.4 *Let X be an irreducible $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$, the set $\mathbb{X}$ being denumerably infinite. The following are equivalent:*

1. *X is recurrent.*
2. *There exists a positive measurable function $f : \mathbb{X} \to \mathbb{R}_+$ tending to infinity, and a finite set $F \in \mathscr{X}$ such that*

$$\mathbb{E}(f(X_{n+1}) - f(X_n) | X_n = y) \leq 0, \ \forall y \notin F.$$

Proof:

2 ⇒ 1: Let $W_n = f(X_{n \wedge \tau_F})$, for $n \geq 0$. From b) follows that (W_n) is a supermartingale; therefore, $W_\infty = \lim_n W_n$ exists almost surely. Fatou's lemma implies that

$$\mathbb{E}_x(W_\infty) \leq \liminf_n \mathbb{E}_x(W_n) \leq \mathbb{E}_x(W_0) = f(x).$$

Suppose now that X is transient. Irreducibility implies then that the finite set $F \in \mathscr{X}$ is transient, so that $\mathbb{P}_x(\tau_F = \infty) = s_x > 0$. Irreducibility implies that not only F but every finite set $F' \in \mathscr{X}$ is transient, i.e. $\eta^0(F') < \infty$ almost surely. Denote by $K_{F'} = \sup\{f(x), x \in F'\}$. It is then obvious (why?) that

$$\{\eta^0(F') < \infty\} \subseteq \{\liminf f(X_n) \geq K_{F'}\}.$$

For all $a \geq 0$, let $F_a = \{x \in \mathbb{X} : f(x) \leq a\}$. Then, for $c \geq K_F$, the set F_c is finite and contains F. Since F_c is finite, it follows that $\mathbb{P}_x(\eta^0(F_c) < \infty) = 1$. Hence

$$\begin{aligned} \mathbb{E}_x(W_\infty) &= \mathbb{E}_x(\liminf W_n) \geq \mathbb{E}_x(\liminf W_n \mathbb{1}_{\{\tau_F = \infty\}} \mathbb{1}_{\{\eta^0(F_c) < \infty\}}) \\ &\geq c\mathbb{P}_x(\eta^0(F_c) < \infty; \tau_F = \infty) \geq c s_x > 0. \end{aligned}$$

Since c is arbitrary, it follows that $\mathbb{E}_x(W_\infty) = \infty$, in contradiction with $\mathbb{E}_x(W_\infty) \leq f(x)$. We conclude *ad absurdum* that X is recurrent.

1 ⇒ 2: Since the set $\mathbb{X}$ is denumerably infinite, there exists a bijection $b : \mathbb{X} \to \mathbb{N}$. Introduce a new chain $\hat{X}$ on $\mathbb{X}$ evolving with the transition matrix $\hat{P}$ defined by

$$\hat{P}_{xy} = \begin{cases} 1 & \text{if } b(x) = b(y) = 0 \\ P_{xy} & \text{if } b(x) \geq 1. \end{cases}$$

Denote also by $\mathbb{X}_N = \{b \geq N\}$, $T(N) := \hat{\tau}_{\mathbb{X}_N}$, and define, for arbitrary $x \in \mathbb{X}$ and $N \in \mathbb{N}$,

$$\phi_x(N) = \begin{cases} 1 & \text{if } x \in \mathbb{X}_N, \\ \hat{\mathbb{P}}_x(T(N) < \infty) & \text{if } x \notin \mathbb{X}_N. \end{cases}$$

Since for all $a \in \mathbb{R}_+$, the set $F_a = \{f \leq a\}$ is finite, irreducibility of the chain implies that all F_a will be transient, i.e. $\eta^0(F_a) < \infty$ almost surely. Denote $K_{F_a} = \inf\{f(x), x \notin F_a\}$. Obviously $K_{F_a} \leq a$. Moreover, $\phi_{b^{-1}(0)}(N) = 0$ for all $N > 0$. It is also a simple observation (exercise[4]) that $\lim_{N\to\infty} \phi_x(N) = 0$ for all $x \in \mathbb{X}$. Therefore, for all $k \geq 1$, there exists an integer N_k (with (N_k) an increasing sequence), such that $\phi_x(N_k) < 1/2^k$ for all $x \in \{b \leq k\}$.
First step: Fix some N and start by defining $f(x) = \phi_x(N)$ and $F = \{b^{-1}(0)\}$. Then, for $x \notin F$:

$$\begin{aligned}
\mathbb{E}(f(X_{n+1}) - f(X_n) | X_n = x) &= \sum_{y\in\mathbb{X}} P_{xy} f(y) - f(x) \\
&= \sum_{y\in\{b\leq N-1\}} P_{xy} f(y) + \sum_{y\in\{b\geq N\}} P_{xy} f(y) - f(x) \\
&= \sum_{y\in\{b\leq N-1\}} P_{xy} \phi_y(N) + \sum_{y\in\{b\geq N\}} P_{xy} 1 - f(x) \\
&= \sum_{k=1}^{\infty} \sum_{y\in\{b\leq N-1\}} P_{xy} \mathbb{P}_y(T(N) = k) + \mathbb{P}_y(T(N) = 1) - f(x) \\
&= \sum_{k=1}^{\infty} \sum_{y\in\{b\leq N-1\}} P_{xy} \mathbb{P}_y(T(N) = k) + \mathbb{P}_x(T(N) = 1) - f(x) \\
&= \mathbb{P}_x(T(N) < \infty) - f(x) \\
&\leq 0.
\end{aligned}$$

Therefore, 2) is verified for this particular f.
Second step: Define now $f(x) = \sum_{k=1}^{\infty} \phi_x(N_k) < \infty$ (because $\phi_x(N_k) < 1/2^k$). Since every individual term of the series defining f verifies 2), it follows that f also does. It remains to show that f tends to infinity. Consider $f(b^{-1}(i)) = \sum_{k=1}^{\infty} \phi_{b^{-1}(i)}(N_k) < \infty$. On the other hand, $\lim_{i\to\infty} \phi_{b^{-1}(i)}(N_k) = 1$ for all k. Therefore

$$\infty = \sum_{k=1}^{\infty} 1 = \sum_{k=1}^{\infty} \liminf_i \phi_{b^{-1}(i)}(N_k) \leq \liminf_i \sum_{k=1}^{\infty} \phi_{b^{-1}(i)}(N_k) = \liminf_i f(b^{-1}(i)),$$

by Fatou's lemma. This remark concludes the proof.

4. The idea of the proof is the following: since X is recurrent, the chain starting at x returns to x in a finite time. Therefore the number of sites visited between two successive visits at x is finite. We conclude by choosing N sufficiently large, e.g. $N > \sup\{b(X_n) : n = 0, \ldots, \tau_x^0\}$.

□

Exercise 6.6.5 Let X be a simple random walk on $\mathbb{N}$ with transition kernel

$$P_{xy} = \begin{cases} 1 & \text{if } x=0, y=1; \\ p & \text{if } x>0, y=x-1; \\ 1-p & \text{if } x>0, y=x+1; \\ 0 & \text{otherwise;} \end{cases}$$

where $p \in]0,1[$. Let $\lambda = \ln \frac{p}{1-p}$. Show that

1. The chain is positive recurrent if, and only if, $\lambda > 0$.
2. The chain is null recurrent if, and only if, $\lambda = 0$.
3. The chain is transient if, and only if, $\lambda < 0$.

Exercise 6.6.6 Let $(p_x)_{x\in\mathbb{N}}$ be a sequence of independent and identically distributed $[0,1]$-valued random variables whose law has not atoms in 0 or 1. Let X be a simple random walk in the random environment (p_x) on $\mathbb{N}$ with transition kernel

$$P_{xy} = \begin{cases} 1 & \text{if } x=0, y=1; \\ p_x & \text{if } x>0, y=x-1; \\ 1-p_x & \text{if } x>0, y=x+1; \\ 0 & \text{otherwise.} \end{cases}$$

Let $\lambda = \mathbb{E}(\ln \frac{p_1}{1-p_1})$. Show that

1. The chain is positive recurrent if, and only if, $\lambda > 0$.
2. The chain is null recurrent if, and only if, $\lambda = 0$.
3. The chain is transient if, and only if, $\lambda < 0$.

6.6.2 Moments of passage times

7

Boundary theory for transient chains on denumerably infinite sets

7.1 Motivation

From the theory of harmonic functions in $\mathbb{R}^2$, we know (see [1] for instance) that every $h: \mathbb{R}^2 \to \mathbb{R}_+$ that is harmonic ($\Delta h = 0$) in the open unit disk $\mathbb{D} = B(0,1)$ is in one-to-one correspondence with a Borel measure μ^h on $\partial\mathbb{D} = \mathbb{S}^1$ through the Poisson kernel. In polar coordinates, this correspondence is defined through

$$h(r\exp(i\theta)) = \int_{\partial\mathbb{D}} K(r,\theta,\phi)\mu^h(d\phi),$$

where the Poisson kernel reads

$$K(r,\theta,\phi) = K((r,\theta);(1,\phi)) := \frac{1-r^2}{1-2r\cos(\theta-\phi)+r^2}.$$

The purpose of this section is to provide, in the case of a transient P, an analogue of the previous formula for P-harmonic, non-negative functions $h: \mathbb{X} \to \mathbb{R}_+$ through an integral Martin-Poisson representation

$$h(x) = \int_{\partial\mathbb{X}} K(x,\alpha)\mu^h(d\alpha).$$

Obviously, not only the boundary kernel K must be defined but also the meaning of the boundary of the state space $\partial\mathbb{X}$ and the measure μ^h. This representation will be useful in studying the asymptotic behaviour of the chain.

Several comments are necessary at this level. The first concerns the abstract space $\partial\mathbb{X}$ playing the role of the boundary. It will turn out that this space (more precisely a subset of this space) will be identified to the set of P-harmonic functions. Secondly, the mere notion of compactification implies the definition of a topology on the space that will defined in terms of the boundary kernel K; the corresponding boundary space is termed the Martin boundary. As for the convexity considerations, the aforementioned representation goes beyond the Krein-Milman theorem [1] to propose a Choquet decomposition [2] of the set of harmonic functions. Finally, a measure-theoretic notion of boundary can be introduced (the Poisson boundary), indexed by *bounded* harmonic functions, that is conceptually simpler than the topological one.

Boundary theory for transient Markov chains is far from being trivial; a certain number of important ramifications to other mathematical fields can be obtained as a by-product of this study. For this reason, only an elementary approach to the boundary theory of transient chains on denumerably infinite state spaces is given here. This chapter heavily relies on more complete original references as [6, 18, 35] or more pedagogical ones [20, 30].

7.2 Martin boundary

In the sequel, we assume that the chain is **transient**, i.e. for all $x, y \in \mathbb{X}$, we have $G^0(x, y) < \infty$; for most results **irreducibility** must also be assumed, although this latter condition can be slightly relaxed for some specific results where can be replaced by accessibility of all states from a reference point $o \in \mathbb{X}$.

Definition 7.2.1 (Boundary kernel) Suppose that there exists $o \in \mathbb{X}$ such that $G^0(o, y) > 0$ for all $y \in \mathbb{X}$. Define the **boundary kernel** K by

$$K(x, y) = \frac{G^0(x, y)}{G^0(o, y)}.$$

Lemma 7.2.2 *Assume that there exists a state $o \in \mathbb{X}$ from which all other states are accessible i.e. $\forall y \in \mathbb{X} : G^0(o, y) > 0$. Then there exist constants $C_x > 0$ (independent of y) such that*

$$K(x, y) \le C_x, \forall y \in \mathbb{X}.$$

1. It is recalled that the Krein-Milman theorem states: Let C be a non-empty compact convex subset of a topological vector space. Then $C = \overline{\text{co}}(\mathsf{Extr}(C))$.

2. It is recalled that the Choquet theorem states: Let C be a non-empty compact convex subset of a topological vector space $\mathbb{Y}$. Then for every $y \in C$ there exists a measure ν on $\mathscr{B}(\mathsf{Extr}(C))$ such that $y = \int_{\mathsf{Extr}(C)} \nu(\alpha)\alpha$.

Proof: We have

$$\begin{aligned}\sum_{z\in\mathbb{X}} P^m(x,z)G^0(z,y) &= \sum_{z\in\mathbb{X}}\sum_{n=0}^{\infty} P^m(x,z)P^n(z,y)\\ &= \sum_{n=m}^{\infty} P^n(x,y)\\ &= G^0(x,y) - \sum_{n=0}^{m-1} P^n(x,y).\end{aligned}$$

Since $\forall x\in\mathbb{X}$, the potential kernel verifies $G^0(o,x)>0$, it follows that there exists an integer $m\ge 0$ such that $P^m(o,x)>0$. Therefore $G^0(o,y)\ge\sum_{z\in\mathbb{X}} P^m(o,z)G^0(z,y)\ge P^m(o,x)G^0(x,y)$, hence $K(x,y)\le\frac{1}{P^m(o,x)}:=C_x<\infty$. □

Theorem 7.2.3 (Martin compactification) *For a Markov chain* $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ *which is irreducible and transient, there exist a compact metric space* $(\hat{\mathbb{X}}_M,\rho)$ *and a homeomorphic embedding* $\iota:\mathbb{X}\to\hat{\mathbb{X}}_M$.

Definition 7.2.4 (Martin boundary) For a Markov chain as in theorem 7.2.3 above, the set $\partial\mathbb{X}_M=\hat{\mathbb{X}}_M\setminus\mathbb{X}$ is called the **Martin boundary** of the chain.

Proof of theorem 7.2.3: For the denumerable family of constants $C=(C_x)_{x\in\mathbb{X}}$, as in lemma 7.2.2, and an arbitrary sequence $w=(w_x)_{x\in\mathbb{X}}\in\ell^1(\mathbb{X})$ of strictly positive real constants $w_x>0$, set

$$\rho(x,y):=\sum_{z\in\mathbb{X}} w_z\frac{|K(z,x)-K(z,y)|+|\delta_{zx}-\delta_{zy}|}{C_z+1},\quad\text{for}\quad x,y\in\mathbb{X}.$$

The condition $w\in\ell^1(\mathbb{X})$ guarantees that the series defining ρ converges uniformly in x,y. Additionally, $\rho(x,y)\ge 0$ for all x,y and $\rho(x,x)=0$ for all x; moreover, if $x\neq y$ then $\rho(x,y)>0$. Finally, for all $w,x,y\in\mathbb{X}$, the triangular inequality $\rho(w,y)\le\rho(w,x)+\rho(x,y)$ holds. Hence ρ is a distance on $\mathbb{X}$. Recall that for a sequence $(y_n)_{n\in\mathbb{N}}$ of points in $\mathbb{X}$, we write $\lim_{n\to\infty}y_n=\infty$ if the sequence eventually leaves any finite set $F\subset\mathbb{X}$ and never returns (converges within $\hat{\mathbb{X}}=\mathbb{X}\cup\{\infty\}$, the one-point compactification of $\mathbb{X}$). Consequently, a sequence $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in the metric space $(\mathbb{X},\rho)$ when

- either there exist $y\in\mathbb{X}$ and n_0 such that for all $n\ge n_0$: $y_n=y$ (so that (y_n) converges within $\mathbb{X}$);
- or else $\lim_{n\to\infty}y_n=\infty$ and $\lim_{n\to\infty}K(x,y_n)$ exists for every x.

Now the metric space $(\mathbb{X},\rho)$ is embedded as usual into a complete metric space $(\hat{\mathbb{X}}_M,\rho)$ by considering equivalence classes of Cauchy sequences. Consider now an arbitrary sequence $(y_n)_{n\in\mathbb{N}}$ of elements of $\mathbb{X}$. Either there exists a finite subset $F\subseteq\mathbb{X}$ visited infinitely often by the sequence (y_n), or the sequence leaves any finite set and (eventually) never returns. Consider these two cases separately:

– If $(y_n)_n$ returns infinitely often in F, then we can extract a subsequence $(y_{n_k})_k$ always remaining in F. Consequently, there exists a point $y \in F$ that is visited infinitely often by $(y_{n_k})_k$, i.e. there exists a subsequence $(y_{n_{k_l}})_{l\in\mathbb{N}}$ such that $y_{n_{k_l}} = y$ for all l hence, according to the first condition above, this subsequence is Cauchy.
– If (y_n) eventually leaves any finite set and never returns, the sequence of numbers $\kappa_n = K(z, y_n) \le \sup_{y\in\mathbb{X}} K(z, y) \le C_z < \infty$ is bounded. Since the interval $[0, C_z]$ is compact in $\mathbb{R}$, the sequence (κ_n) has a converging subsequence $\kappa_{n_k} = K(z, y_{n_k})$, for $k \in \mathbb{N}$, i.e. $\lim_{k\to\infty} K(z, y_{n_k})$ exists. Hence, according to the second condition above, the subsequence (y_{n_k}) is Cauchy.

Since the space $(\hat{\mathbb{X}}_M, \rho)$ is complete, we conclude that in both cases, from every sequence of $\mathbb{X}$ (hence of $\hat{\mathbb{X}}$) we can extract a converging subsequence. Hence the space $(\hat{\mathbb{X}}_M, \rho)$ is compact.

Since $\mathbb{X}$ is discrete, the boundary kernel $K(x, y)$ is a continuous function in y for each fixed $x \in \mathbb{X}$. By the definition of ρ it follows that

$$|K(x, y) - K(x, z)| \le \frac{C_x + 1}{w_x} \rho(y, z).$$

Thus, for each $x \in \mathbb{X}$, the function $K(x, \cdot)$ has a unique extension to $\hat{\mathbb{X}}_M$ as a continuous function of its second argument. Since $K(o, y) = 1$ for every $y \in \mathbb{X}$, it follows that $K(o, \alpha) = 1$ for all $\alpha \in \hat{\mathbb{X}}_M$. The complete compact metric space $(\hat{\mathbb{X}}_M, \rho)$ is the **Martin compactification** of $\mathbb{X}$. It obviously depends on P and possibly on the reference point o. □

Definition 7.2.5 For a transient Markov chain $\mathsf{MC}((\mathbb{X}, \mathscr{X}), P, \mu)$ and $\hat{\mathbb{X}}_M$ the Martin compactification defined in theorem 7.2.3, the function $K : \mathbb{X} \times \hat{\mathbb{X}}_M \to \mathbb{R}_+$ is called the **Martin kernel** corresponding to the transition kernel P and the reference point o.

Corollary 7.2.6 *Let K be the Martin kernel corresponding to the transient Markov kernel P and reference point o. Then*
– *for all $\alpha \in \partial\mathbb{X}_M$, $K(o, \alpha) = 1$,*
– *for all $x \in \mathbb{X}$, there exists $C_x > 0$ such that: $K(x, \alpha) \le C_x, \forall \alpha \in \partial\mathbb{X}_M$.*

Remark: The terms δ_{zx} and δ_{zy} in the definition of ρ in the previous proof are **not** necessary to make ρ a metric. If $\rho'(x, y) := \sum_{z\in\mathbb{X}} w_z \frac{|K(z,x)-K(z,y)|}{C_z+1}$, for $x, y \in \mathbb{X}$, then $\rho'(x, y) = 0$ still implies $x = y$ (why?). The *raison d'être* of these terms is more profound. Suppose that we had a sequence (y_n) of elements of $\mathbb{X}$ — eventually leaving any finite set (i.e. such that $y_n \to \infty$) — and a $y' \in \mathbb{X}$ such that for all $x \in \mathbb{X}$, $\lim_{n\to\infty} K(x, y_n) = K(x, y')$ implying that $\lim_{n\to\infty} \rho'(y_n, y') = 0$. Therefore what should be a non-convergent Cauchy sequence (y_n) converges actually to $y' \in \mathbb{X}$. The result should be that the set $\mathbb{X}$ would non be open in the completion $\hat{\mathbb{X}}_M$ and the boundary $\partial\mathbb{X}_M = \hat{\mathbb{X}}_M \setminus \mathbb{X}$ would not be closed in $\hat{\mathbb{X}}_M$. Using ρ instead prevents this phenomenon from occurring, since then $\rho(y_n, y') \geq \frac{w_z}{C_z+1} > 0$ whenever $y_n \neq y'$.

Lemma 7.2.7 *Let the boundary kernel K be defined as in definition 7.2.1. Then the function $K(\cdot,\alpha)$ is P-superharmonic for all $\alpha \in \partial\mathbb{X}_M$.*

Proof: Fix an arbitrary $y \in \mathbb{X}$ and define $r(y') = \frac{1\!\!1_{\{y\}}(y')}{G^0(o,y')}$. Then

$$G^0 r(x) = \sum_{n\in\mathbb{N}} \sum_{y'\in\mathbb{X}} P^n(x,y')r(y') = \sum_{n\in\mathbb{N}} \frac{P^n(x,y)}{G^0(o,y)} = K(x,y),$$

showing that $K(\cdot,y)$ is a potential — hence P-superharmonic — for every $y \in \mathbb{X}$. If $y_n \to \alpha \in \partial\mathbb{X}_M$, by Fatou's lemma

$$\sum_{z\in\mathbb{X}} P(x,z)K(z,\alpha) = \sum_{z\in\mathbb{X}} P(x,z) \lim_{n\to\infty} K(z,y_n) \leq \lim_{n\to\infty} \sum_{z\in\mathbb{X}} P(x,z)K(z,y_n) \leq \lim_{n\to\infty} K(x,y_n) = K(x,\alpha).$$

□

Since the boundary $\partial\mathbb{X}_M$ inherits the metric topology with which the set $\hat{\mathbb{X}}_M$ is endowed, it is meaningful to speak about the Borel σ-algebra $\mathscr{B}(\partial\mathbb{X}_M) = \mathscr{B}(\hat{\mathbb{X}}_M) \cap \partial\mathbb{X}_M$.

Exercise 7.2.8 Let $\nu \in \mathscr{M}_1(\mathscr{B}(\partial\mathbb{X}_M))$ and P a transient kernel. Then the function $s : \mathbb{X} \to \overline{\mathbb{R}}_+$, defined by $s(x) = \int_{\hat{\mathbb{X}}_M} K(x,\alpha)\nu(d\alpha)$, is finite P-superharmonic.

Theorem 7.2.9 (Poisson-Martin representation theorem) *Let P be a stochastic kernel on the denumerably infinite set $\mathbb{X}$, equipped with the exhaustive σ-algebra $\mathscr{X}$. Assume that*

1. *for all $x, y \in \mathbb{X}$, $G^0(x,y) < \infty$, and*
2. *there exists $o \in \mathbb{X}$ such that for all $y \in \mathbb{X}$, $G^0(o,y) > 0$.*

Define the boundary kernel K as in definition 7.2.1 and the Martin boundary $\partial\mathbb{X}_M$ as in definition 7.2.4. Then for any P-harmonic non-negative function h, there exists a measure μ^h on $\mathscr{B}(\partial\mathbb{X}_M)$ such that

$$h(x) = \int_{\partial\mathbb{X}_M} K(x,\alpha)\mu^h(d\alpha).$$

Proof: Since $\mathbb{X}$ is countably infinite, there exists a sequence $(\mathbb{X}_n)_{n\in\mathbb{N}}$ of finite sets such that $\mathbb{X}_n \uparrow \mathbb{X}$. Define $w_n = nG^0 \mathbb{1}_{\mathbb{X}_n}$ and $h_n(x) = h(x) \wedge w_n(x)$. Obviously w_n is a potential, therefore h_n is a potential by lemma 5.2.7 (claims 2 and 3). Thus, there exists a positive charge r_n such that

$$h_n(x) = \sum_{y\in\mathbb{X}} G^0(x,y)r_n(y) = \sum_{y\in\mathbb{X}} \frac{G^0(x,y)}{G^0(o,y)} G^0(o,y)r_n(y) = \int_{\mathbb{X}} K(x,y)\mu_n^h(dy),$$

where $\mu_n^h(\{y\}) = G^0(o,y)r_n(y)$. Recall that $K(o,y) = 1$ for all $y \in \mathbb{X}$, so that $\mu_n^h(\mathbb{X}) = h_n(o) \le h(o)$; subsequently, the sequence (μ_n^h) extends to a sequence of bounded measures on the compact metric space $\hat{\mathbb{X}}_M$, so that a subsequence $(\mu_{n_k}^h)_k$ converges weakly $\lim_{k\to\infty} \mu_{n_k}^h = \mu^h$ over continuous functions of $\hat{\mathbb{X}}_M$. Now, for every fixed $x \in \mathbb{X}$, the function $K(x,\cdot)$ is continuous on $\hat{\mathbb{X}}_M$, hence

$$\begin{aligned} h(x) &= \lim_n \int_{\hat{\mathbb{X}}_M} K(x,\alpha)\mu_n^h(d\alpha) \\ &= \int_{\hat{\mathbb{X}}_M} K(x,\alpha)\mu^h(d\alpha) \\ &= \sum_{y\in\mathbb{X}} G^0(x,y)\frac{\mu^h(\{y\})}{G^0(o,y)} + \int_{\partial\mathbb{X}_M} K(x,\alpha)\mu^h(d\alpha). \end{aligned}$$

P-harmonicity of h reads

$$0 = h(x) - Ph(x) = \frac{\mu^h(\{x\})}{G^0(o,x)} + \int_{\partial\hat{\mathbb{X}}_M} (K - PK)(x,\alpha)\mu^h(d\alpha), \forall x \in \mathbb{X}.$$

Since by the previous lemma 7.2.7, $K(\cdot,\alpha)$ is P-superharmonic for all $\alpha \in \partial\hat{\mathbb{X}}_M$, vanishing of the r.h.s. implies simultaneously that $\mu(\mathbb{X}) = 0$ — showing thus that μ^h is concentrated on $\partial\hat{\mathbb{X}}_M$ — and moreover, the measure μ^h is concentrated on the subset of $\alpha \in \partial\mathbb{X}_M$ on which $K(\cdot,\alpha)$ is P-harmonic (as a function of x). □

Remark: It is worth noting that the measure μ^h occurring in the previous representation is not unique. To ensure uniqueness, the integral must be further restricted on a smaller susbet of $\partial\mathbb{X}_M$ called the minimal boundary (see section 7.4).

Corollary 7.2.10 *Let K be the Martin kernel corresponding to a transient Markov kernel P and a reference point o. Let h be a non-negative P-harmonic function and μ^h a measure occurring in the Martin representation 7.2.9 of h. Then*

- *$\mu^h(\partial\mathbb{X}_M) = h(o) < \infty$, and*
- *the integral representing h converges whenever $\mu^h(\partial\mathbb{X}_M) < \infty$.*

Corollary 7.2.11 *Let s be P-superharmonic on $\mathbb{X}$ and $s(x) \ge 0$ for all $x \in \mathbb{X}$. Then under the same assumptions as for the theorem 7.2.9, there exists a non-negative* *measure μ on $\hat{\mathbb{X}}_M$ such that*

$$s(x) = \int_{\hat{\mathbb{X}}_M} K(x,\alpha)\mu(d\alpha), \forall x \in \mathbb{X}.$$

7.3 Convexity properties of sets of superharmonic functions

It is convenient to extend slightly the notion of superharmonicity. Recall that the space $\mathbb{X}$ contains a reference point o.

Definition 7.3.1 Let P be a Markov kernel and $t > 0$.

- A function $s : \mathbb{X} \to \mathbb{R}_+$ is called (non-negative) (P,t)**-superharmonic** if it verifies $Ps \le ts$. The set of (non-negative) (P,t)-superharmonic functions is denoted $\mathsf{SH}^+(P,t)$. When $t = 1$, we write $\mathsf{SH}^+(P)$.
- A function $h : \mathbb{X} \to \mathbb{R}_+$ is called (non-negative) (P,t)**-harmonic** if it verifies $Ps = ts$. The set of (non-negative) (P,t)-harmonic functions is denoted $\mathsf{H}^+(P,t)$. When $t = 1$, we write $\mathsf{H}^+(P)$.
- A function $f : \mathbb{X} \to \mathbb{R}_+$ is called a **normalised** if $f(o) = 1$. The set of normalised harmonic (resp. superharmonic) functions is denoted $\mathsf{H}^+_1(P,t)$ (resp. $\mathsf{SH}^+_1(P,t)$).
- A P-harmonic function (with $t = 1$) is called **minimal** if for every other P-harmonic function h_1, verifying $0 \le h_1 \le h$, we have $h_1 = ch$ for some constant c. We denote by $\mathsf{H}^+_m(P)$ the set of minimal and normalised P-harmonic functions.

Lemma 7.3.2 *Let P be an irreducible Markov kernel and $t > 0$. If $\mathsf{SH}^+_1(P,t) \neq \emptyset$, then $\mathsf{SH}^+_1(P,t)$ is compact in the topology of pointwise convergence and $\mathsf{SH}^+(P,t)$ is a convex cone with compact base.*

Proof: Let (s_n) be a sequence in $\mathsf{SH}^+_1(P,t)$ converging simply towards s. We have $Ps = P\lim_n s_n = P\liminf_n s_n \le \liminf_n Ps_n \le t\liminf_n s_n = ts$ by Fatou lemma. Additionally, $s(o) = 1$ by pointwise convergence. Hence $s \in \mathsf{SH}^+_1(P,t)$, showing that $\mathsf{SH}^+_1(P,t)$ is closed.

From the irreducibility of P follows that for every $x \in \mathbb{X}$, there exists $n_x \in \mathbb{N}$ such that $P^{n_x}(o,x) > 0$. Define thus $c_x = \frac{t^{n_x}}{P^{n_x}(o,x)} < \infty$. For $s \in \mathsf{SH}^+_1$ we have

$$P^{n_x}(o,x)s(x) \le P^{n_x}s(o) \le t^{n_x}s(o),$$

so that $s(x) \le c_x$. Since for every $x \in \mathbb{X}$, the set $[0,c_x]$ is compact in the usual topology of $\mathbb{R}$, the set $\times_{x\in\mathbb{X}}[0,c_x]$ is compact by Tychonoff's theorem. Further $\mathsf{SH}^+_1(P,t) \subseteq \times_{x\in\mathbb{X}}[0,c_x]$, hence compact as closed subset of a compact space.

Convexity of $\mathsf{SH}^+(P,t)$ is trivial. Any $s \in \mathsf{SH}^+(P,t)$ can be written $s = \beta b$ with $b \in \mathsf{SH}^+_1(P,t)$ by chosing $\beta = s(o)$. Hence $\mathsf{SH}^+(P,t)$ is a convex cone of base $\mathsf{SH}^+_1(P,t)$. □

Proposition 7.3.3 *Let P be a transient irreducible kernel and K the corresponding boundary kernel associated with the reference point o. Then*

$$\mathsf{Extr}(\mathsf{SH}^+_1(P)) = \{K(\cdot,y), y \in \mathbb{X}\} \cup \mathsf{H}^+_m(P).$$

Proof: Let $b \in \mathsf{SH}^+_1(P)$ and assume that b is extremal. Extremality of b means that its Riesz decomposition is either purely harmonic or purely potential.

Consider first the purely potential case, i.e. there exists a non-negative, non identically 0 charge r such that

$$\begin{aligned} b(x) &= \sum_{y\in\mathbb{X}} G^0(x,y)r(y) \\ &= \sum_{y\in\mathbb{X}} \frac{G^0(x,y)}{G^0(o,y)} G^0(o,y)r(y) \\ &= \sum_{y\in\operatorname{supp} r} K(x,y)c_y, \end{aligned}$$

where $c_y = G^0(o,y)r(y)$. Note also that $\sum_{y\in\mathbb{X}} c_y = \sum_{y\in\operatorname{supp} r} c_y = \sum_{y\in\mathbb{X}} G^0(o,y)r(y) = G^0 r(o) = b(o) = 1$. Hence, on $\operatorname{supp} r$, we have $c_y > 0$ and $\sum_{y\in\operatorname{supp} r} c_y = 1$. Since $PK(x,y) = K(x,y) - \frac{\delta_{x,y}}{G^0(o,y)}$, it follows that $K(\cdot,y)$ is harmonic on $\mathbb{X}\setminus\{y\}$ and strictly superharmonic on $\{y\}$. Now, the above expression $b(x) = \sum_{y\in\operatorname{supp} r} K(x,y)c_y$ provides a convex superposition of b in terms of superharmonic functions. Extremality of b implies that $\operatorname{card}\operatorname{supp} r = 1$, meaning that there exists some $y \in \mathbb{X}$ such that $b(\cdot) = K(\cdot,y)$.

Consider now the case of b being purely harmonic. Every convex combination of harmonic functions is still harmonic; extremality of b means then minimal harmonicity. We have thus established so far that $\mathsf{Extr}(\mathsf{SH}^+_1(P)) \subseteq \{K(\cdot,y), y \in \mathbb{X}\} \cup \mathsf{H}^+_m(P)$.

To establish the converse inclusion, consider first the case of a function $K(\cdot,y)$ for some $y \in \mathbb{X}$. We know that $K(\cdot,y) \in \mathsf{SH}^+(P)$ and since $K(o,y) = 1$ for all y, that K belongs to the base of the cone: $K(\cdot,y) \in \mathsf{SH}^+_1(P)$. Assume now that $K(\cdot,y) = \lambda b_1(\cdot) + (1-\lambda)b_2(\cdot)$, with $\lambda \in]0,1[$ and $b_1, b_2 \in \mathsf{SH}^+_1(P)$. Then $PK(x,y) = \lambda Pb_1(x) + (1-\lambda)Pb_2(x) \le K(x,y)$. Since further $b(\cdot) = \lambda b_1(\cdot) + (1-\lambda)b_2(\cdot)$ is bounded from above by the potential $K(\cdot,y)$ the harmonic part in its Riesz decomposition must be 0. Hence both b_1 and b_2 are potentials: $b_i = G^0 r_i$, $i = 1,2$. Since $K(\cdot,y)$ is harmonic in $\mathbb{X}\setminus\{y\}$, the same holds true for both b_i and the latter can occur only if $\operatorname{supp} r_i = \{y\}$. Now, degeneracy of the supports of r_i to the singleton $\{y\}$ means that b_1/b_2 is a constant and this constant can only be 1 since both b_i are normalised. Hence, we conclude that $b_1 = b_2 = b$ and consequently $K(\cdot, y) \in \mathsf{Extr}(\mathsf{SH}_1^+(P))$.

Similarly, if $h \in \mathsf{H}_m^+(P)$ we can always decompose into $h = \lambda b_1 + (1-\lambda) b_2$ with $b_i \in \mathsf{SH}_1^+(P)$. Now, harmonicity of h implies that $b_i \in \mathsf{H}_1^+(P)$ and minimality of h implies extremality $b_1 = b_2 = h$ and consequently the inclusion $\mathsf{H}_m^+(P) \subseteq \mathsf{Extr}(\mathsf{SH}_1^+(P))$. □

7.4 Minimal and Poisson boundaries

The theorem 7.2.9 guarantees that for every P-harmonic function h there exists a measure μ representing it through the Martin representation. This section deals with the questions that remain still unsettled:

- is the measure μ^h unique?
- is the Martin kernel $K(\cdot, \alpha)$, for fixed $\alpha \in \partial\mathbb{X}_M$, always harmonic so that for *every* probability μ on $\partial\mathbb{X}_M$, the integral $\int_{\partial\mathbb{X}_M} K(x, \alpha)\mu(d\alpha)$ defines a harmonic function on $\mathbb{X}$?

It will be shown in this section that the measure μ^h is unique (and defines a harmonic function whenever $\mu^h(\partial\mathbb{X}_M) < \infty$) provided that is restricted to a subset $\partial_m\mathbb{X}_M \subseteq \partial\mathbb{X}_M$.

Theorem 7.4.1 *Let P be an irreducible transient kernel. If $h \in \mathsf{H}_m^+(P)$ then the measure μ^h occurring in its Poisson-Martin representation is*

- *unique and*
- *a point mass.*

Proof: We have $\int_{\partial\mathbb{X}_M} \mu^h(d\alpha) = \int_{\partial\mathbb{X}_M} K(o, \alpha)\mu^h(d\alpha) = h(o) = 1$, hence μ^h is a probability. Assume there exists $B \in \mathscr{B}(\partial\mathbb{X}_M)$ such that $0 < \mu^h(B) < 1$; define then $h_B(x) = \frac{1}{\mu^h(B)} \int_B K(x, \alpha)\mu^h(d\alpha)$. We have then the non-trivial convex decomposition $h = \mu^h(B) h_B + (1 - \mu^h(B)) h_{B^c}$. Minimal harmonicity of h implies that $h = h_B = h_{B^c}$, hence, for all $x \in \mathbb{X}$ and all $B \in \mathscr{B}(\partial\mathbb{X}_M)$:

$$\int_B h(x)\mu^h(d\alpha) = h_B(x)\mu^h(B) = \int_B K(x, \alpha)\mu^h(d\alpha).$$

(Note that the above equalities hold trivially if $\mu^h(B) \in \{0, 1\}$.) Therefore $K(x, \cdot) = h(x)$ holds, for every $x \in \mathbb{X}$, μ^h-a.s. Let $A = \{\alpha \in \partial\mathbb{X}_M : h(x) = K(x, \alpha), \forall x \in \mathbb{X}\}$. Since $\mathbb{X}$ is countable, we have $\mu^h(A) = 1$, implying in particular that $A \neq \emptyset$. We conclude that there exists $\alpha \in \partial\mathbb{X}_M$ such that for all x, $h(x) = K(x, \alpha)$. Further, this α must be unique since by construction $K(\cdot,\alpha) \neq K(\cdot,\alpha')$ if $\alpha \neq \alpha'$. Consequently, μ^h is a Dirac mass on some $\alpha \in \partial\mathbb{X}_M$. □

Definition 7.4.2 (Minimal boundary) The set

$$\partial_m \mathbb{X}_M = \{\alpha \in \partial\mathbb{X}_M : K(\cdot,\alpha) \text{ is minimal } P\text{-harmonic}\}$$

is called the **minimal boundary**.

Corollary 7.4.3 *The set $\partial_m\mathbb{X}_M$ is a Borel subset of $\partial\mathbb{X}_M$. For every $h \in \mathsf{H}^+(P)$, there exists a unique μ^h on $\mathscr{B}(\partial\mathbb{X}_M)$ such that $\mu^h(\partial\mathbb{X}_M \setminus \partial_m\mathbb{X}_M) = 0$ and $h(x) = \int_{\partial\mathbb{X}_M} K(x,\alpha)\mu^h(d\alpha)$.*

Exercise 7.4.4 Prove the previous corollary 7.4.3.

Suppose now that $h, f \in \mathsf{H}^+(P)$ and $0 \leq h \leq f$ holds; then obvisously $g = f - h \in \mathsf{H}^+(P)$. Using the uniqueness of the representing measures supported by the minimal boundary $\partial_m\mathbb{X}_M$, we have $\mu^f = \mu^h + \mu^{f-h}$, implying that the inequality $\mu^h \leq \mu^f$ holds for the representing measures as well. Now suppose that for all x, $h(x) \leq M$, where $M > 0$ is a given constant, i.e. h is bounded P-harmonic. We denote $b\mathsf{H}^+(P)$ the set of bounded P-harmonic functions. The inequality $0 \leq h(x) \leq M$, holding for all $x \in \mathbb{X}$ implies that $\mu^h(B) \leq M\mu^1(B)$ holds for all $B \in \mathscr{B}(\partial_m\mathbb{X}_M)$ as well, where μ^1 is the representing measure of the constant harmonic function 1. Therefore the measure μ^h is absolutely continuous with respect to μ^1; the Radon-Nikodým derivative reads $p_h(\alpha) = \frac{d\mu^h}{d\mu^1}(\alpha)$.

Definition 7.4.5 (Poisson boundary) Let μ^1 be the unique measure representing the constant normalised P-harmonic function 1. We call **Poisson boundary** the set $\partial_P\mathbb{X}_M = \operatorname{supp}\mu^1$. For all $h \in b\mathsf{H}^+(P)$, we have the Poisson representation

$$h(x) = \int_{\partial_P\mathbb{X}_M} K(x,\alpha)p_h(\alpha)\mu^1(d\alpha),$$

where $p_h(\alpha) = \frac{d\mu^h}{d\mu^1}(\alpha)$.

Exercise 7.4.6 (Zero-one law) Let $h_1 \equiv 1$ (obviously in $\mathsf{H}_1^+(P)$). Show that

1. $h_1 \in \mathsf{H}_m^+(P) \Leftrightarrow [\forall h \in b\mathsf{H}^+(P) \Rightarrow h \equiv \text{const}]$, and
2. conclude that, if $h_1 \in \mathsf{H}_m^+(P)$, then for all $x \in \mathbb{X}$ and all $A \subseteq \mathbb{X}$, the function $H_A^0(x)$ can take only two values, 0 or 1.

7.5 Limit behaviour of the chain in terms of the boundary

Assume that (X_n) is transient and irreducible in $\mathbb{X}$. Transience means that (X_n) eventually leaves any bounded subset of $\mathbb{X}$. Now, even though the sequence (X_n) **does not** converge in $\mathbb{X}$, it does converge in the Martin compactification $\hat{\mathbb{X}}_M$ towards a random limit $X_\infty \in \partial_m \mathbb{X}_M$, i.e.

$$\mathbb{P}_x(\lim_{n\to\infty} X_n \text{ exists and equals } X_\infty \in \partial_m \mathbb{X}_M) = 1.$$

The purpose of this section (see theorem 7.5.6 below) is to determine the law of X_∞.

Lemma 7.5.1 *Let X be a transient MC$((\mathbb{X},\mathscr{X}),P,\mu)$ and $F \subseteq \mathbb{X}$ a finite set. Then there exists a charge $r_F \geq 0$, supported by the set F, such that*

$$L_F^0(x) := \mathbb{P}_x(\tau_F^0 < \infty) = G^0 r_F(x),$$

for all $x \in \mathbb{X}$.

Proof: From the very definition of L_F^0, we have

$$P^n L_F^0(x) = \mathbb{P}_x(X_k \in F, \text{ for some } k \geq n).$$

Define

$$r_F(x) := L_F^0(x) - PL_F^0(x) = \mathbb{P}_x(X_0 \in F; X_k \not\in F \text{ for } \ k \geq 1) \geq 0.$$

Hence L_F^0 is P-superharmonic [3]. Since F is finite and the chain transient, $\lim_n P^n L_F^0(x) = 0$, therefore from the Riesz decomposition theorem 5.2.4 follows that L_F^0 is a potential corresponding to the charge r_F. From the definition of r_F, namely $r_F(x) = \mathbb{P}_x(X_0 \in F; X_k \not\in F \text{ for } \ k \geq 1)$ follows immediately that $r_F(x) = r_F(x)\mathbb{1}_F(x)$ for all x, proving the claim that r_F is supported by F. □

From lemma 7.2.7, we know that $K(\cdot,\alpha)$ is P-superharmonic for all $\alpha \in \partial\mathbb{X}_M$. Denote $M(x,\alpha) = \lim_{n\to\infty} P^n K(x,\alpha)$ its harmonic part, stemming from the Riesz decomposition theorem 5.2.4. Recall also from proposition 5.2.10 that, for an arbitrary subset $A \subseteq \mathbb{X}$, the harmonic part in the Riesz decomposition of L_A^0 is H_A^0.

Lemma 7.5.2 *Let X be a transient MC$((\mathbb{X},\mathscr{X}),P,\mu)$. For an arbitrary subset $A \subseteq \mathbb{X}$, there exists a measure ν_A on $\mathscr{B}(\partial\mathbb{X}_M)$ such that*

$$H_A^0(x) = \int_{\overline{A}\cap\partial\mathbb{X}_M} M(x,\alpha)\nu_A(d\alpha).$$

3. A fact we already knew from proposition 5.2.10.

Proof: If A is finite, transience of the chain shows that the claim is correct with $\nu_A = 0$. Consider thus the infinite case and let $(A_l)_l$ be an increasing sequence of measurable finite sets, exhausting A. Then by lemma 7.5.1, $L^0_{A_l}(x) = \int_{A_l} K(x,y)\nu_{A_l}(dy)$, where

$$\nu_{A_l}(\{y\}) = G^0(o,y)r_{A_l}(y)$$

with $\text{supp}(\nu_{A_l}) \subseteq A_l \subseteq A$. Now

$$\nu_{A_l}(\hat{\mathbb{X}}_M) = \sum_{y\in A_l} G^0(o,y)r_{A_l}(y) = L^0_{A_l}(o) = \mathbb{P}_o(\tau^0_{A_l} < \infty) \le 1.$$

Hence, there exists a weakly converging subsequence $w - \lim \nu_{A_{l_k}} = \nu_A$ on $\hat{\mathbb{X}}_M$. Moreover, $\tau^0_{A_l} \downarrow \tau^0_A$. Hence,

$$L^0_A(x) = \mathbb{P}_x(\tau^0_A < \infty) = \int_{\overline{A}} K(x,\alpha)\nu_A(d\alpha),$$

generalising thus lemma 7.5.1 for arbitrary sets A. We have thus, for all $x \in \mathbb{X}$,

$$\begin{aligned} P^n L^0_A(x) &= \mathbb{P}_x(X_k \in A, \text{ for some } k \ge n) \\ &= \int_{\overline{A}} P^n K(x,\alpha)\nu_A(d\alpha) \\ &\to H^0_A(x). \end{aligned}$$

On the other hand, $P^n K(x,\alpha) \downarrow M(x,\alpha)$ for all $\alpha \in \hat{\mathbb{X}}_M$, while $P^n K(x,y) \downarrow 0$ for all $y \in \mathbb{X}$; the result follows immediately. □

We shall introduce a convenient trick, known as **relativisation**, to modify the asymptotic behaviour of the chain at infinity. Let $h \in \mathsf{H}^+_1(P)$ for an irreducible transient P. Since $h(o) = 1$, irreducibility of P implies that $h(x) > 0$ for all $x \in \mathbb{X}$. Define henceforth $P^{(h)}(x,y) = \frac{1}{h(x)}P(x,y)h(y)$ that is obviously non-negative for all $x, y \in \mathbb{X}$. Harmonicity of h implies that $\sum_{y\in\mathbb{X}} P^{(h)}(x,y) = \frac{1}{h(x)}h(x) = 1$; hence $P^{(h)}$ is a Markov kernel.

Notation 7.5.3 For any $h \in \mathsf{H}^+_1(P)$ denote by $P^{(h)}$ the **relativised Markov kernel** and by $\mathbb{P}^{(h)}_x$ the probability on the trajectory space induced by the kernel $P^{(h)}$ (with the same trajectories).

Exercise 7.5.4 (Relativised kernels) Let $h \in \mathsf{H}^+_1(P)$.

1. $f \in \mathsf{H}^+(P) \Leftrightarrow \frac{f}{h} \in \mathsf{H}^+(P^{(h)})$.
2. If $K^{(h)}$ denotes the boundary kernel of $P^{(h)}$, then $K^{(h)}(x,y) = \frac{1}{h(x)}K(x,y)$.
3. Let $h \in \mathsf{H}^+_1(P)$ and denote μ^h the unique measure on $\mathscr{B}(\partial_m\mathbb{X}_M)$ representing h, i.e. $h(x) = \int_{\partial_m\mathbb{X}_M} K(x,\alpha)\mu^h(d\alpha)$. Then μ^h represents the constant harmonic function 1 for the relativised kernel $K^{(h)}$.

The relativised chain follows the same trajectories as the original one; only probabilities are affected by the relativisation.

Proposition 7.5.5 *Let (X_n) be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\epsilon_x)$, with P irreducible and transient. For every $B \in \mathscr{B}(\mathscr{X}^\infty)$ we have*

$$\mathbb{P}_x(B) = \int_{\partial_m \mathbb{X}_M} \mathbb{P}_x^{(h_\alpha)}(B) K(x,\alpha)\mu^1(d\alpha),$$

where $h_\alpha(x) = K(x,\alpha)$ and $\mathbb{P}_x^{(h_\alpha)}$ denotes the probability on the path space due to the relativised kernel. More generally, for every $h \in \mathsf{H}_1^+(P)$,

$$\mathbb{P}_x^{(h)}(B) = \frac{1}{h(x)} \int_{\partial_m \mathbb{X}_M} \mathbb{P}_x^{(h_\alpha)}(B) K(x,\alpha)\mu^h(d\alpha).$$

Proof: It is enough to prove the formulae for the elementary cylindrical sets of the form $B = \{x_1\} \times \cdots \{x_n\} \times \mathbb{X} \times \mathbb{X} \times \cdots$. We have then

$$\begin{aligned} \mathbb{P}_x^{(h)}(B) &= P^{(h)}(x,x_1)\cdots P^{(h)}(x_{n-1},x_n) \\ &= \frac{1}{h(x)} P(x,x_1)\cdots P(x_{n-1},x_n) h(x_n) \\ &= \frac{1}{h(x)} \mathbb{E}_x(\mathbb{1}_{\{x_1,\ldots,x_n\}}(X_1,\cdots,X_n) h(X_n)). \end{aligned}$$

In particular

$$\mathbb{P}_x^{(h_\alpha)}(B) = \frac{1}{K(x,\alpha)} \mathbb{E}_x(\mathbb{1}_{\{x_1,\ldots,x_n\}}(X_1,\cdots,X_n) K(X_n,\alpha)).$$

Using the Poisson-Martin representation $h(x) = \int_{\partial_m \mathbb{X}_M} K(x,\alpha)\mu^h(d\alpha)$, we conclude that

$$\mathbb{P}_x^{(h)}(B) = \frac{1}{h(x)} \int_{\partial_m \mathbb{X}_M} \mathbb{P}_x^{(h_\alpha)}(B) K(x,\alpha)\mu^h(d\alpha).$$

The first statement is obtained immediately by applying the previous formula to the constant harmonic function 1. □

Theorem 7.5.6 *Let (X_n) be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\epsilon_x)$, where P is irreducible and transient. For all $A \in \mathscr{B}(\partial_m \mathbb{X}_M)$,*

$$\mathbb{P}_x(\lim_{n\to\infty} X_n = X_\infty \in A) = \int_A K(x,\alpha)\mu^1(d\alpha).$$

Proof: The event $\{\lim_{n\to\infty} X_n = X_\infty \in A\}$ is $\mathscr{X}^\infty$-measurable. From proposition 7.5.5, we conclude that

$$\mathbb{P}_x(\lim_{n\to\infty} X_n = X_\infty \in A) = \int_{\partial_m \mathbb{X}_M} \mathbb{P}_x^{(h_\alpha)}(\lim_{n\to\infty} X_n = X_\infty \in A) K(x,\alpha)\mu^1(d\alpha),$$

where $h_\alpha(x) = K(x,\alpha) \in \mathsf{H}^+_m(P)$.

It is enough to establish that $\mathbb{P}^{(h_\alpha)}_x(\lim_{n\to\infty} X_n = \alpha) = 1$ because then the r.h.s. of the previous formula reduces to $\int_A K(x,\alpha)\mu^1(d\alpha)$ which is preciseley the claim of the theorem.

Fix $\alpha \in \partial_m \mathbb{X}_M$ and $\epsilon > 0$ and let $A_\epsilon = \{x \in \mathbb{X} : \rho(x,\alpha) \geq \epsilon\}$. Obviously, $\alpha \not\in \overline{A_\epsilon}$. On denoting $\mathbb{P}^{(h_\alpha)}_x(X_n \in A_\epsilon \text{ i.o.})$, we observe that if $H^{(h_\alpha)}_{A_\epsilon}(x) = 0$ for all $x \in \mathbb{X}$ and all $\epsilon > 0$ then $\mathbb{P}^{(h_\alpha)}_x(X_n \in A^c_\epsilon \text{ eventually}) = 1$, and since $\epsilon > 0$ is arbitrary, the latter means $\mathbb{P}^{(h_\alpha)}_x(\lim_{n\to\infty} X_n = \alpha) = 1$. Hence it is enough to prove $H^{(h_\alpha)}_{A_\epsilon}(x) = 0$ for all $x \in \mathbb{X}$ and all $\epsilon > 0$.

Let $f = H^{(h_\alpha)}_{A_\epsilon}$; we know (see exercise 7.5.4 on relativised kernels) that $f \in \mathsf{H}^+(P^{h_\alpha}) \Leftrightarrow f(\cdot)K(\cdot,\alpha) \in \mathsf{H}^+(P)$. Since $K(\cdot,\alpha) \in \mathsf{H}^+_m(P)$, we conclude that any bounded $h \in \mathsf{H}^+(P^{(h_\alpha)})$ is constant. Hence (see exercise 7.4.6) $H^{h_\alpha}_{A_\epsilon}(x) \in \{0,1\}$ for all $x \in \mathbb{X}$ and all $\epsilon > 0$. It is therefore enough to exclude the possibility $H^{h_\alpha}_{A_\epsilon}(x) \equiv 1$.

By lemma 7.5.2, there exists a measure ν_{A_ϵ} on $\mathscr{B}(\partial\mathbb{X}_M)$ such that

$$\begin{aligned} 1 \geq H^{(h_\alpha)}_{A_\epsilon}(x) &= \int_{\overline{A_\epsilon}\cap\partial\mathbb{X}_M} M^{(h_\alpha)}(x,\beta)\nu_{A_\epsilon}(d\beta) \\ &= \int_{\overline{A_\epsilon}\cap\partial\mathbb{X}_M} \frac{M(x,\beta)}{K(x,\alpha)}\nu_{A_\epsilon}(d\beta), \end{aligned}$$

implying that $\int_{\overline{A_\epsilon}\cap\partial\mathbb{X}_M} M(x,\beta)\nu_{A_\epsilon}(d\beta) \leq K(x,\alpha)$. Thus, for all Borel subsets $A \subseteq \overline{A_\epsilon}\cap\partial\mathbb{X}_M$, we have $\int_A M(x,\beta)\nu_{A_\epsilon}(d\beta) \leq K(x,\alpha)$. Recall that $M(\cdot,\beta) \in \mathsf{H}^+(P)$ while $K(\cdot,\alpha) \in \mathsf{H}^+_m(P)$. From the very definition of minimality, we conclude that $\int_A M(x,\beta)\nu_{A_\epsilon}(d\beta) = C_A K(x,\alpha)$, and by applying this equality to $x = o$, we determine the constant $C_A = \int_A M(o,\beta)\nu_{A_\epsilon}(d\beta)$, so that for all $x \in \mathbb{X}$,

$$M(x,\beta) = M(o,\beta)K(x,\alpha), \text{ for } \nu_{A_\epsilon}\text{-a.e. } \beta \in \overline{A_\epsilon}\cap\partial\mathbb{X}_M.$$

If $H^{(h_\alpha)}_{A_\epsilon}(x) \neq 0$ then the zero-one law implies that $H^{(h_\alpha)}_{A_\epsilon}(x) = 1$, therefore

$$h(o) = \int_{\overline{A_\epsilon}\cap\partial_m\mathbb{X}_M} M(o,\beta)\nu_{A_\epsilon}(d\beta) = 1,$$

hence

$$L^0_{A_\epsilon}(x) = \mathbb{P}_o(\tau^0_{A_\epsilon} < \infty) = \int_{\overline{A_\epsilon}} K(o,\alpha)\nu_{A_\epsilon}(d\alpha) \leq 1$$

and $M(o,\beta) \leq K(o,\beta) \leq 1$. Consequently,

$$1 = h(o) = \int_{\overline{A_\epsilon}\cap\partial_m\mathbb{X}_M} M(o,\beta)\nu_{A_\epsilon}(d\beta) \leq 1,$$

implying that for all $x \in \mathbb{X}$ and $\alpha \in \partial_m \mathbb{X}_M$, $M(x,\beta) = K(x,\alpha)$, for ν_{A_ϵ}-a.e. $\beta \in \overline{A_\epsilon} \cap \partial\mathbb{X}_M$.

Now, by the Riesz decomposition of K, we have

$$K(x,\beta) = M(x,\beta) + G^0 r_\beta(x),$$

where r_β is a non-negative charge. But $K(o,\beta) = M(o,\beta) = 1$ hence $G^0 r_\beta(o) = 0$. Therefore, $G^0 r_\beta(x) = 0$ for all x, so that $M(x,\beta) = K(x,\alpha) = K(x,\beta)$ for $\nu_{A_\epsilon}-$ a.e.$\beta \in \overline{A_\epsilon} \cap \partial\mathbb{X}_M$. From the construction of the Martin boundary, if $K(x,\alpha) = K(x,\beta)$ for all $x \in \mathbb{X}$, then $\alpha = \beta$. Since $\alpha \not\in \overline{A_\epsilon}$, it follows that $\nu_{A_\epsilon}(\overline{A_\epsilon} \cap \partial\mathbb{X}_M) = 0$. Therefore $H^{h_\alpha}_{A_\epsilon} \equiv 0$ for all $\epsilon > 0$. □

Exercise 7.5.7 (Competition game [?]) Let $(\mathbf{X}_n)$ be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\epsilon_x)$ with $\mathbb{X} = \mathbb{Z}^2$ and Markov kernel P defined, for $x = (x_1,x_2)$ and $y = (y_1,y_2)$ elements of $\mathbb{Z}^2$, by

$$P(\mathbf{x},\mathbf{y}) = \begin{cases} \frac{1}{2} & y_1 = x_1, y_2 = x_2 + 1 \\ \frac{1}{2} & y_1 = x_1 + 1, y_2 = x_2 \\ 0 & \text{otherwise.} \end{cases}$$

1. The process is not irreducible.
2. Show that

$$G^0(\mathbf{x},\mathbf{y}) = \begin{cases} C^{y_1-x_1}_{y_1-x_1+y_2-x_2}(\frac{1}{2})^{y_1-x_1+y_2-x_2} & y_1 \le x_1 \wedge y_2 \le x_2, \\ 0 & y_1 < x_1 \vee y_2 < x_2. \end{cases}$$

3. Show that $K(\mathbf{x},\mathbf{y}) = 2^{x_1+x_2}(\frac{y_1}{y_1+y_2})^{x_1}(\frac{y_2}{y_1+y_2})^{x_2}(1 + \mathscr{O}(\frac{x_1^2}{y_1}) + \mathscr{O}(\frac{x_2^2}{y_2}))$.
4. Let $\mathbf{y}_k \to \infty$. Show that $\lim_k K(\mathbf{x},\mathbf{y}_k)$ exists if and only if $\frac{y_{k,1}}{y_{k,1}+y_{k,2}} \to \alpha$ with $\alpha \in [0,1]$. In that case, establish that

$$K(\mathbf{x},\alpha) = 2^{x_1+x_2}\alpha^{x_1}(1-\alpha)^{x_2}.$$

5. Conclude that $h \in \mathsf{SH}^+(P)$ on $\mathbb{X}$ if and only if there exists μ on $\mathscr{B}([0,1])$ such that

$$h(\mathbf{x}) = 2^{x_1+x_2}\int_0^1 \alpha^{x_1}(1-\alpha)^{x_2}\mu(d\alpha).$$

   In particular, the functions $h(\mathbf{x}) = K(\mathbf{x},\alpha)$ are harmonic.
6. Show that for all $\alpha \in [0,1]$, the function $K(\cdot,\alpha) \in \mathsf{H}^+_m(P)$.
7. Compute $K(\mathbf{x},\frac{1}{2})$ and conlcude that if $h \in b\mathsf{H}^+(P)$ then h is constant.
8. Establish $\partial\mathbb{X}_M = \partial_m\mathbb{X}_M = [0,1]$ while $\partial_P\mathbb{X}_M = \{\frac{1}{2}\}$.

7.6 Markov chains on discrete groups, amenability and triviality of the boundary

In this section $\mathbb{X}$ will be a countably infinite group. Random walks on $\mathbb{Z}^d$ provide a standard example of Markov chains on Abelian groups. However, the non Abelian case is also very interesting and will be provide us with non-trivial examples.

Definition 7.6.1 Let $\mathbb{X}$ be a countably infinite group (composition is denoted multiplicatively). The transition kernel is termed **group left invariant** if $P(x, y) = P(gx, gy)$ for all $g \in \mathbb{X}$. Right invariance is defined analogously.

Exercise 7.6.2 Let P be a transition kernel on the Abelian group $\mathbb{X} = \mathbb{Z}^d$ and μ an arbitrary probability on $\mathbb{X}$. Show that

- P defined by $P(x, y) = \mu(y - x)$ is group invariant,
- if h is P-harmonic, then so is the translated function $h_g(\cdot) := h(\cdot + g)$, for all $g \in \mathbb{X}$,
- any harmonic function h can be expressed as a convex combination of its translates.

Theorem 7.6.3 (Choquet-Deny) *Let P be a translation invariant and irreducible kernel on $\mathbb{X} = \mathbb{Z}^d$, and denote by $\Gamma_P = \{\beta \in \mathbb{R}^d : \sum_{y\in\mathbb{X}} P(0, y)\exp(\beta \cdot y) = 1\}$, where $\cdot$ stands for the scalar product on $\mathbb{R}^d$. Then*

- *$h \in \mathsf{H}^+_1(P)$ if, and only if, there exists a measure μ^h on $\mathscr{B}(\Gamma_P)$ such that $h(x) = \int_{\Gamma_P} \exp(\alpha \cdot x)\mu^h(d\alpha)$,*
- *$h \in \mathsf{H}^+_m(P)$ if, and only if, there exists an $\alpha \in \Gamma_P$ such that $h(x) = \exp(\alpha \cdot x)$ i.e. μ^h degenerates into a Dirac mass on α.*

Proof: If P is recurrent, then all $h \in \mathsf{H}^+(P)$ are constant (see exercise 6.2.10), hence the theorem holds with $\mu^h(d\alpha) = \epsilon_0(d\alpha)$. It is therefore enough to consider the case of transient P. Irreducibility of P implies that $\forall y \in \mathbb{X}$, $\exists n := n_y \geq 0$ such that $P^n(0, y) > 0$. Let $h \in \mathsf{H}^+_m(P)$.

$$\begin{aligned} h(x) &= \sum_{z\in\mathbb{X}} P^{n_y}(x, z)h(z) \\ &= \sum_{z\in\mathbb{X}} P^{n_y}(0, z - x)h(z) \\ &= \sum_{z\in\mathbb{X}} P^{n_y}(0, z)h(x + z) \\ &\geq P^{n_y}(0, y)h(x + y), \end{aligned}$$

hence $h(x+y) \le c_y h(x)$, where $c_y = \frac{1}{P^{n_y}(0,y)} < \infty$. Then for $h \in \mathsf{H}^+_m(P)$, we get $h(x+y) = d_y h(x)$ and applying for x equal to the reference point $o = 0$ we get $d_y = h(y)$. Hence $h(x+y) = h(x)h(y)$. Considering the case where x scans the set of basis elements of $\mathbb{Z}^d$, i.e. $x \in \{e_i, i = 1,\ldots,d\}$, we conclude that any multiplicative harmonic function must be of the form $h(x) = \exp(\alpha \cdot x)$. Additionally, if $\alpha \in \Gamma_P$ then such a function is necessarily normalised. If $h \in \mathsf{H}^+_1(P)$ then the representation $h(x) = \int_{\Gamma_P} \exp(\alpha \cdot x)\mu^h(d\alpha)$ holds immediately.

Conversely, assuming that $h(x) = \exp(\beta \cdot x)$ for some $\beta \in \Gamma_P$, implies that $h \in \mathsf{H}^+_1(P)$. If $f \in \mathsf{H}^+_1(P)$ is such that $0 \le f(x) \le Ch(x)$ for all x, then $0 \le \int_{\Gamma_P} \exp(\alpha \cdot x)\mu^f(d\alpha) \le C\exp(\beta \cdot x)$. Now, in $x = (x_1,\ldots,x_d)$ appearing in the previous inequalities, let $x_i \to \pm\infty$ for an $i \in \{1,\ldots,d\}$, and keep fixed all other x_j, for $j \ne i$. We conclude that necessarily $\operatorname{supp}\mu^f \subseteq \{\beta\}$ and consequently $f(x) = Ch(x)$ showing that $h \in \mathsf{H}^+_m(P)$. □

Corollary 7.6.4 *Let $X = (X_n)_n$ be an irreducible Markov chain on $\mathbb{X} = \mathbb{Z}^d$ with an invariant kernel P.*

- *If $\xi_{n+1} = X_{n+1} - X_n$ has 0 mean, there do not exist non-trivial functions in $\mathsf{H}^+(P)$.*
- *The constant function $h \equiv 1$ always belongs in $\mathsf{H}^+_m(P)$.*
- *If for all $n \ge 0$, $\sum_{y\in\mathbb{X}} P(0,y)\|y\|^n < \infty$ but $\sum_{y\in\mathbb{X}} P(0,y)\exp(\alpha\cdot x) = \infty$ for all $\alpha \ne 0$, then there do not exist non-trivial functions in $\mathsf{H}^+(P)$.*

Exercise 7.6.5 Prove the previous corollary 7.6.4.

Exercise 7.6.6 (Simple random walk on the homogeneous tree) Let $\mathbb{X} = T_d$, with $d \ge 3$, where T_d is the group generated by the d free involutions, i.e. the non-Abelian group with presentation $T_d = \langle a_1,\ldots,a_d \,|\, a_1^2 = \ldots = a_d^2 = e\rangle$ where e stands for the neutral element. Every $x \in T_d$ can be written in a unique way as a reduced word $x = a_{i_1}\cdots a_{i_l}$, for some integer l, with the constraint $i_k \ne i_{k+1}$ for all $k = 1,\ldots l-1$. The integer $l := l(x) := |x|$ is called the length of the word x. The inverse of x is immediately obtained as $x^{-1} = a_{i_l}\cdots a_{i_i}$. The Cayley graph of T_d is the graph $(\mathbb{V},\mathbb{A})$ where the vertex set $\mathbb{V} = T_d$ and the edge set $\mathbb{A}$ is the subset of $\mathbb{V}\times\mathbb{V}$ characterised by the condition $(x,y) \in \mathbb{A} \Leftrightarrow y = xa_i$, for some $i \in \{1,\ldots,d\}$. There exists an isomorphism between the Caley graph of T_d and the group T_d itself. Hence the group T_d can be metrised naturally by introducing the edge distance $\delta(x,y) = |x^{-1}y|$. The isotropic simple random walk on T_d is defined by the transition kernel

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

1. Write explicitly the condition for a function h to belong in $\mathsf{H}^+(P)$.

2. Introduce the sequence $Y_n = |X_n|$, for $n \in \mathbb{N}$. Show that $Y = (Y_n)_n$ is a Markov chain on $\mathbb{N}$ of transition kernel Q (determine Q).
3. Show that $\mathbb{P}_k(\lim \frac{Y_n}{n} = \frac{d_2}{d}) = 1$ for all $k \in \mathbb{N}$.
4. Conclude that $X = (X_n)_n$ is transient whenever $d \geq 3$.
5. Show that $L^0_y(x) = G^0(x, y)r(y)$ where $r(y) = \mathbb{P}_y(X_n \neq y, \forall n \geq 1)$. Argue that $\forall y \in \mathbb{X}$, $r(y) = r > 0$. Conclude that $G^0(x, y) = \frac{L^0_y(x)}{r}$.
6. Introducing the function $\phi(l) = \mathbb{P}_l(Y_n = 0, \text{for some } n \geq 0)$ and the reference point $o = e$, establish that $K(x, y) = \frac{\phi(|y^{-1}x|)}{\phi(|y^{-1}|)}$.
7. Show that ϕ verifies $\phi(0) = 1$ and for $l \geq 1$,

$$\phi(l) = \frac{1}{d}\phi(l-1) + \frac{d-1}{d}\phi(l+1).$$

   Show that a solution can be obtained in the form $\phi(l) = c_1\lambda_1^l + c_2\lambda_2^l$ where c_1, c_2 are constants and λ_1, λ_2 solutions of the quadratic equation $\lambda = \frac{1}{d} + \frac{d-1}{d}\lambda^2$. Conclude that $\phi(l) = (\frac{1}{d-1})^l$ for $l \geq 0$ and establish that $K(x, y) = (\frac{1}{d-1})^{|y^{-1}x| - |x|}$.
8. Let (y_n) be a sequence in $\mathbb{X}$ such that $\lim_n |y_n| = \infty$. Determine under which conditions the limit $\lim_n [|x^{-1}y_n| - |y_n|]$ exists for all $x \in \mathbb{X}$.
9. Let $A = \{\alpha = a_{i_1}a_{i_2}\cdots; i_k \in \{1, \ldots, d\} \wedge (i_k \neq i_{k+1})\forall k\}$ be the set of the leaves of the tree. For $x = a_{k_1}\cdots a_{k_{|x|}} \in \mathbb{X}$ and $\alpha = a_{i_1}a_{i_2}\cdots \in A$ written respectively as a finite and as an infinite reduced word, introduce $J : \mathbb{X} \times A \to \mathbb{N}$ by $J(x, \alpha) = |x| - 2\max\{j \leq |x| : k_n = i_n; 1 \leq n \leq j\}$. Show that $\lim K(x, y_n) = K(x, \alpha) = (\frac{1}{d-1})^{J(x,\alpha)}$.
10. $h \in \mathsf{H}^+(P) \Leftrightarrow h(x) = \int_A (\frac{1}{d-1})^{J(x,\alpha)}\mu^h(d\alpha)$, where μ^h is a measure on A.
11. Show that $\partial\mathbb{X}_M = \partial_m\mathbb{X}_M = A$.

Reste à rédiger: Amenability, amenable groups, Choquet-Deny theorem for non-Abelian groups, boundaries for groupoids and semi-groupoids.

8

Continuous time Markov chains on denumerable spaces

8.1 Embedded chain

8.2 Infinitesimal generator

8.3 Explosion

8.4 Semi-martingale techniques for explosion times

9

Irreducibility on general spaces

In this chapter, methods for dealing with Markov chains on general spaces are developed. The material presented in this and the following chapters heavily relies on the original work of Döblin [5]; a more pedagogical and concise account of this work is [27] or [26]. More complete treatement of the subject, far beyond the scope of this semestrial course, remain [23, **?**].

9.1 ϕ-irreducibility

It is well known that for Markov chains on discrete spaces, irreducibility means that for all $x, y \in \mathbb{X}$, there exists $n = n(x, y) \in \mathbb{N}$ such that $P^n(x, y) > 0$ or, equivalently, $L_y(x) = \mathbb{P}_x(\tau_y < \infty) > 0$. In the continuous case however, it may happen that for any pair $x, y \in \mathbb{X}$, $L_y(x) = \mathbb{P}_x(\tau_y < \infty) = 0$.

Example 9.1.1 Let $(\xi_i)_i$ be a sequence of independent random variables on $[-1, 1]$ identically distributed according to the uniform law on $[-1, 1]$. Then the sequence $(X_n)_n$ defined by $X_0 = x$ and $X_n = X_{n-1} + \xi_n$ for all $n \geq 1$ is a random walk on $\mathbb{R}$ verifying $L_y(x) = \mathbb{P}_x(\tau_y < \infty) = 0$ for all $x, y \in \mathbb{R}$.

Definition 9.1.2 A measurable set $F \in \mathscr{X}$ is called (stochastically) **closed** if $F \neq \emptyset$ and $P(x, F) = 1$ for all $x \in F$. A closed set that cannot be partitioned into two closed sets is called **indecomposable**.

Definition 9.1.3 Let $\phi \in \mathscr{M}_+(\mathscr{X})$.

– The measure ϕ is called an **irreducibility measure** for the Markov chain X if

$$A \in \mathscr{X}, \phi(A) > 0 \Rightarrow \forall x \in \mathbb{X}, L_A(x) > 0.$$

The chain is then called ϕ-**irreducible**.

– The chain is called ϕ-**recurrent** if

$$A \in \mathscr{X}, \phi(A) > 0 \Rightarrow \forall x \in \mathbb{X}, L_A(x) = 1.$$

Remark: If $\mathbb{X}$ is countable, the usual notions of irreducibility and recurrence are recovered on choosing for $\phi \in \mathscr{M}_+(\mathscr{X})$ the counting measure.

Given a probability $\nu \in \mathscr{M}_1(\mathscr{X})$ and a measure $\phi \in \mathscr{M}_+(\mathscr{X})$, we know that ν has a unique decomposition into $\nu = \nu_{ac} + \nu_s$, where $\nu_{ac} \ll \phi$ is the absolutely continuous part (i.e. $\forall A \in \mathscr{X}, \phi(A) = 0 \Rightarrow \nu_{ac}(A) = 0$), and $\nu_s \perp \phi$ is the singular part (i.e. $\exists A \in \mathscr{X}, \phi(A) = 0$ and $\nu_s(A^c) = 0$). Moreover, the Radon-Nikodým theorem states that there exists a unique $f \in L^1(\mathbb{X}, \mathscr{X}, \phi)$ such that, for all $A \in \mathscr{X}$, $\nu_{ac}(A) = \int_A f(x)\phi(dx)$, where $f(x) = \frac{d\mu}{d\phi}(x)$ is the Radon-Nikodým derivative.

Now, if $\phi \in \mathscr{M}_+(\mathscr{X})$ is an irreducibility measure, noting that for all $x \in \mathbb{X}$ and $n \in \mathbb{N}$, $P^n(x, \cdot)$ is a probability on $\mathscr{X}$, we can decompose $P^n(x, \cdot) = P^n_{ac}(x, \cdot) + P^n_s(x, \cdot)$. On denoting by $p_n(x, y) = \frac{dP^n_{ac}(x,\cdot)}{d\phi}(y)$, we have $P^n_{ac}(x, A) = \int_A p_n(x, y)\phi(dy)$. As far as the singular part is concerned, for all $x \in \mathbb{X}$ and $n \in \mathbb{N}$, $P^n(x, \cdot) \perp \phi$ means that there exists a set $\mathbb{X}_{x,n} \in \mathscr{X}$ such that $\phi(\mathbb{X}_{x,n}) = 0$ and $P^n(x, P^n(x, \cdot)^c) = 0$. In summarising,

$$P^n(x, A) = \int_{A \cap \mathbb{X}^c_{x,n}} p_n(x, y)\phi(dy) + P^n_s(x, A \cap \mathbb{X}_{x,n}).$$

Remark: One can easily show that for all $x \in \mathbb{X}$ and all $n \in \mathbb{N}$, the Radon-Nikodým density $p_n(x, \cdot)$ is measurable with respect to the second argument and similarly $p_n(\cdot, x)$ measurable with respect to the first argument. It is convenient to ensure that $p_n(\cdot, \cdot)$ is jointly measurable with respect to the two arguments.

Lemma 9.1.4 *Let $\phi \in \mathscr{M}_+(\mathscr{X})$ be an irreducibility measure for the chain. Then, any probability $\mu \in \mathscr{M}_1(\mathscr{X})$ that is equivalent to ϕ is an irreducibility measure.*

Proof: Since μ is equivalent to ϕ it follows that $\mu \ll \phi$ and $\phi \ll \mu$. Hence, for any $A \in \mathscr{X}$ either $\phi(A) = 0$ (and consequently $\mu(A) = 0$) or $\phi(A) > 0$ (and consequently $\mu(A) > 0$). Therefore, μ is also an irreducibility measure. □

Proposition 9.1.5 *Let $\phi \in \mathscr{M}_+(\mathscr{X})$ be an irreducibility measure for the chain. Then there exists an irreducibility measure $\psi \in \mathscr{M}_+(\mathscr{X})$ such any irreducibility measure $\phi' \in \mathscr{M}_+(\mathscr{X})$ is $\phi' \ll \psi$.*

Proof: See exercise 9.1.5 below. □

Definition 9.1.6 The measure ψ, whose existence is guaranteed by proposition 9.1.5, unique up to meausures thar are equivalent, is called the **maximal irreducibility measure**.

Exercise 9.1.7 (Part of the paper given on December 2006) Let $\phi \in \mathcal{M}_+(\mathcal{X})$ be an irreducibility measure for the chain.

1. Is it possible, without loss of generality, to assume that ϕ is a probability?
2. Let ψ be a measure, defined for every $A \in \mathcal{X}$ by $\psi(A) = \int \phi(dx) G_{1/2}(x, A)$. Denote by $\mathbb{X}_k = \{x \in \mathbb{X} : \sum_{n=1}^{k} P^n(x, A) > 1/k\}$. If $\psi(A) > 0$ is it true that $\phi(\mathbb{X}_k) > 0$ for some $k \geq 1$?
3. Does ϕ-irreducibility imply ψ-irreducibility?
4. Let $\phi' \in \mathcal{M}_+(\mathcal{X})$. Is it true that ϕ' is an irreducibility measure if and only if $\phi' \ll \psi$?
5. Can we conclude that two maximal irreducibility measures are equivalent? (The ordering implicitly assumed here is the absolute continuity $\ll$.)
6. Let ϕ' be an arbitrary finite irreducibility measure. Define $\psi'(A) = \int \phi'(dx) G_{1/2}(x, A)$ for all $A \in \mathcal{X}$. Is it true that ψ and ψ' are equivalent?
7. If for some $A \in \mathcal{X}$, we have $\psi(A) = 0$, show that $\psi(\{x \in \mathbb{X} : L_A(x) > 0\}) = 0$.

Exercise 9.1.8 Let (ξ_n) be a sequence of independent real-valued random variables, identically distributed according to the law ν. Suppose that the measure ν possesses an absolutely continuous part ν_{ac} with respect to the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with density f. Suppose further there exist constants $\beta > 0$ and $\delta > 0$ such that for all $x \in \mathbb{R}$ with $|x| \leq \beta$ we have $f(x) > \delta$. Construct a Markov chain (X_n) on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by the recurrence $X_0 = x$ and $X_{n+1} = X_n + \xi_{n+1}$, for $n \in \mathbb{N}$. Denote by P the Markovian kernel of this chain.

1. Let $C = \{x \in \mathbb{R} : |x| \leq \beta/2\}$. For an arbitrary Borel set $B \subseteq C$ and an arbitrary $x \in B$, minorise $P(x, B)$.
2. Let $\phi \in \mathcal{M}_+(\mathcal{B}(\mathbb{R}))$ be defined by

$$\phi(A) = \begin{cases} \lambda(A) & \text{if } A \in \mathcal{B}(\mathbb{R}), A \subseteq C, \\ 0 & \text{if } A \in \mathcal{B}(\mathbb{R}), A \subseteq C^c. \end{cases}$$

   Is the mesure ϕ an irreducibility measure for the chain (X_n)?

Exercise 9.1.9 Let $\psi \in \mathcal{M}_+(\mathcal{X})$ be a maximal irreducibility measure for the chain. A set $A \in \mathcal{X}$ is called **full** if $\psi(A^c) = 0$.

1. Show that every stochastically closed set F is full.
2. Let $A \in \mathcal{X}$ be full and define $F := \{x \in \mathbb{X} : G^0(x, A^c) = 0\}$.

- Show that $F \subseteq A$.
- Suppose that there exists $y \in F$ such that $P(y, A^c) > 0$. Show that this assumption contradicts the very definition of F.
- Conclude that any full set contains a closed set.

Theorem 9.1.10 *Let $\mathscr{X}$ be separable. Then the densities $p_n(\cdot,\cdot)$ are jointly measurable.*

Proof: By lemma 9.1.4, we can assume without loss of generality that ϕ is a probability. Since $P^n_{ac}(x,\cdot) \ll \phi(\cdot)$, it follows that for any $\epsilon > 0$, there exists a $\delta > 0$ such that for every $A \in \mathscr{X}$ with $\phi(A) < \delta$ it follows that $P^n_{ac}(x, A) < \epsilon$, for all $x \in \mathbb{X}$.

Since $\mathscr{X}$ is separable, we proceed as in theorem B.2.1; there exists a sequence $(B_l)_{l\in\mathbb{N}}$ of sets $B_n \subseteq \mathbb{X}$ such that $\mathscr{X} = \sigma(B_l, l \in \mathbb{N})$. Denote $\mathscr{G}_k = \sigma(B_1,\ldots,B_k)$ for $k \in \mathbb{N}$. Every such σ-algebra is composed by $2^{r(k)}$ possible unions of atoms $A_{k,1},\ldots,A_{k,r(k)}$ of $\mathscr{G}_k$. Every atom $A \in \mathscr{G}_k$ is of the form: $A = H_1 \cap \cdots \cap H_k$, each H_i being either B_i or B^c_i. For every $y \in \mathbb{X}$, there exists a unique atom $A \in \mathscr{G}_k$ such that $y \in A$; denote by $A_k(y)$ that particular atom. Define then

$$p_{n,k}(x, y) = \begin{cases} \frac{P^n(x,A_k(y))}{\phi(A_k(y))} & \text{if} \quad \phi(A_k(y)) > 0 \\ 0 & \text{if} \quad \phi(A_k(y)) = 0. \end{cases}$$

Each $p_{n,k}(\cdot,\cdot)$ being jointly measurable, so will be the quantity $q_n(\cdot,\cdot) = \liminf_{k\to\infty} p_{n,k}(\cdot,\cdot)$. Additionally, $p_{n,k}(x,\cdot) \in \mathscr{L}^1(\mathbb{X},\mathscr{G}_k,\phi)$ and for all $F \in \mathscr{G}_k$, we have $\int_F p_{n,k}(x, y)\phi(dy) = P^n_{ac}(x,F)$. This equation implies that $(p_{n,k}(x,\cdot))_k$ is a positive $(\mathscr{G}_k)$-martingale. Therefore, it converges almost surely to $q_n(x,\cdot)$ (already known to be jointly measurable).

Let $\epsilon > 0$, choose $\delta > 0$ (as at the beginning of the proof) and let $K \in]0,\infty[$ be such that $\frac{1}{K} = \frac{P^n_{ac}(x,\mathbb{X})}{K} < \delta$. Then, for all n and k,

$$\phi(\{x \in \mathbb{X} : p_{n,k}(x, y) > K\}) \leq \frac{1}{K}\int_{\mathbb{X}} p_{n,k}(x, y)\phi(dy) \leq \frac{P^n_{ac}(x,\mathbb{X})}{K} < \delta,$$

so that

$$\int_{\{z\in\mathbb{X}:p_{n,k}(x,z)>K\}} p_{n,k}(x, y)\phi(dy) = P^n_{ac}(x,\{z \in \mathbb{X} : p_{n,k}(x,z) > K\}) < \epsilon.$$

This shows that the martingale $(p_{n,k}(x,\cdot))_k$ is uniformly integrable, therefore the convergence takes place also in $\mathscr{L}^1$. Consequently, $q_n(x,\cdot)$ is a version of the Radon-Nikodým derivative $\frac{P^n_{ac}(x,\cdot)}{d\phi}(y)$. □

Remark: If $\mathscr{X}$ is not separable, the conclusion still holds (see [34], §14.13, pp. 147–149.)

Proposition 9.1.11 *There exist $\mathscr{X}\otimes\mathscr{X}$-measurable versions of the densities p_n of the Radon-Nikodým derivatives satisfying, for all $n \geq 1$, all k with $1 \leq k \leq n$ and all $x, y \in \mathbb{X}$:*

$$\begin{aligned} p_n(x,y) &\geq \int_{\mathbb{X}} P^{n-k}(x,dz)p_k(z,y) \\ &\geq \int_{\mathbb{X}} p_{n-k}(x,z)\phi(dz)p_k(z,y). \end{aligned}$$

Proof: For every k, let q_k be a $\mathscr{X}\otimes\mathscr{X}$-measurable version of the density of P^k_{ac}. Then we have for all $B \in \mathscr{X}$:

$$\begin{aligned} P^n(x,B) &= \int_{\mathbb{X}} P^{n-k}(x,dy)P^k(y,B) \\ &\geq \int_{\mathbb{X}} P^{n-k}(x,dy)\int_B q^k(y,z)\phi(dz) \\ &= \int_B (\int_{\mathbb{X}} P^{n-k}(x,dy)q^k(y,z))\phi(dz). \end{aligned}$$

Apply this inequality to $B = A \setminus \mathbb{X}_{x,n}$ for an arbitrary $A \in \mathscr{X}$. Then

$$\begin{aligned} P^n(x, A\setminus \mathbb{X}_{x,n}) &= \int_A q_n(x,z)\phi(dz) \text{ (because } \phi(\mathbb{X}_{x,n}) = 0) \\ &\geq \int_A (\int_{\mathbb{X}} P^{n-k}(x,dy)q^k(y,z))\phi(dz). \end{aligned}$$

because $\phi(\mathbb{X}_{x,n}) = 0$. Hence $\forall_\phi z \in \mathbb{X}$,

$$\begin{aligned} q_n(x,z) &\geq \int_{\mathbb{X}} (P^{n-k}(x,dy)q^k(y,z)) \\ &\geq \int_{\mathbb{X}} (q_{n-k}(x,dy)\phi(dy)q^k(y,z)). \end{aligned}$$

On defining recursively

$$\begin{aligned} p_1(x,z) &= q_1(x,z) \\ p_n(x,z) &= q_n(x,z) \vee \max_{1\leq k\leq n} \int \mathbb{X}(P^{n-k}(x,dy)q^k(y,z)), n \geq 2, \end{aligned}$$

we se that the sequence $(p_n)_n$ fulifills the requirements of the proposition. □

9.2 c-sets

When $\mathbb{X}$ is countable, or more generally when a general space $\mathbb{X}$ possesses a special point a such that $L_{\{a\}}(x) > 0$ for all x, then the situation is quite simple since segments of the Markov chain between successive visits to a are independent.

Example 9.2.1 Let $(\xi_n)_n$ be a sequence of independent random variables identically distributed with uniform probability on $[-1,1]$ and consider the Markov chain on $\mathbb{X} = \mathbb{R}_+$ defined by $X_{n+1} = (X_n + \xi_{n+1})^+$. Then $0 \in \mathbb{X}$ is a regenerating point.

In the general case, such points may not exist; they are subtituted by c-sets, "small sets" visible by the absolutely continuous part of the kernel.

Definition 9.2.2 Let $C \in \mathscr{X}$ and $\phi \in \mathscr{M}_+(\mathscr{X})$. The set C is called a c**-set** relatively to the measure ϕ if

- $\phi(C) > 0$ and
- $\exists n \in \mathbb{N} : r(n,C) \equiv \inf_{(x,y)\in C\times C} p_n(x,y) > 0$.

Theorem 9.2.3 *Suppose that $\mathscr{X}$ is separable and ϕ an irreducibility measure. Suppose that $A \in \mathscr{X}$ verifies*

1. $\phi(A) > 0$ *and*
2. $\forall B \in \mathscr{X}, B \subseteq A, \phi(B) > 0 \Rightarrow \forall x \in A, L_B(x) > 0$.

Then A contains a c-set.

The proof is long and tedious but without any conceptual difficulty, split for clarity into the two lemmata 9.2.4 an 9.2.5, proved below.

Denote by $\phi^2 = \phi \otimes \phi$ the measure on $\mathscr{X} \otimes \mathscr{X}$ induced by ϕ and for any $A \in \mathscr{X}$ by $A^2 = A \times A$ the corresponding rectangle. For any $U \subseteq \mathbb{X} \times \mathbb{X}$, we denote by

$$\begin{aligned} U_1(y) &= \{x \in \mathbb{X} : (x,y) \in U\}, y \in \mathbb{X} \\ U_2(x) &= \{y \in \mathbb{X} : (x,y) \in U\}, x \in \mathbb{X}, \end{aligned}$$

the horizontal section at ordinate y and the vertical section at abscissa x. Define now

$$R^{(m,n)} = \{(x,y) \in A^2 : p_m(x,y) \geq \frac{1}{n}\},$$

and

$$R = \cup_{m\geq 1} \cup_{n\geq 1} R^{(m,n)}.$$

Now

$$x \in A, y \in R_2(x) \Rightarrow \exists m \geq 1, \exists n \geq 1 : (x,y) \in R^{(m,n)}.$$

Thus, $R_2(x) = \cup_{m\geq 1} \cup_{n\geq 1} \{y \in A : p_m(x,y) \geq \frac{1}{n}\} \equiv B \subseteq A$.

Lemma 9.2.4 *Let ϕ be an irreducibility measure. For all $x \in \mathbb{X}$, $\phi(R_2(x)) > 0$.*

Proof: Instantiate hypothesis 2) of the theorem for $A = \mathbb{X}$.

$$\begin{aligned} \forall B \in \mathscr{X}, \phi(B) > 0 &\Rightarrow \forall x \in \mathbb{X} : L_B(x) = \mathbb{P}_x(\tau_B < \infty) > 0 \\ &\Rightarrow \forall x \in \mathbb{X} : \sum_{m\geq 1} \mathbb{P}_x(\tau_B = m) > 0 \\ &\Rightarrow \forall x \in \mathbb{X} : \sum_{m\geq 1} P^m(x, B) > 0 \ (*), \end{aligned}$$

because $\{X_m \in B\} = \{\tau_B = m\} \sqcup \{\tau_B < m, X_m \in B\}$ and consequently, $P^m(x,B) \geq \mathbb{P}_x(\tau_B = m)$. Now $P^m(x,B)$ differs from $\int_B p_m(x,y)\phi(dy)$ on $B \cap \mathbb{X}_{x,m}$ (recall that $\phi(\mathbb{X}_{x,m}) = 0$). On defining $\mathbb{X}_x = \cup_{m\in\mathbb{N}} \mathbb{X}_{x,m}$, we have that $\phi(\mathbb{X}_x) = 0$. Thus, without loss of generality, we can limit ourselves to measurable sets $B \in \mathscr{X}$, with $\phi(B) > 0$ and $B \cap \mathbb{X}_x = \emptyset$. These considerations imply that the inequality $(*)$ above, reads now $\int_B (\sum_{m\geq 1} p_m(x,y))\phi(dy) > 0$. Therefore,

$$\begin{aligned} \forall x \in \mathbb{X}, \forall_\phi y \in B &\Rightarrow \sum_{m\geq 1} p_m(x,y) > 0 \\ &\Rightarrow \exists m \geq 1 : p_m(x,y) > 0 \\ &\Rightarrow \exists m \geq 1, \exists n \geq 1 : p_m(x,y) \geq \frac{1}{n}. \end{aligned}$$

In summarising: $\forall x \in \mathbb{X}$: $\phi(\cup_{m\geq 1} \cup_{n\geq 1} \{(x,y) \in B^2 : p_m(x,y) \geq \frac{1}{n}\}) > 0$ hence $\phi(R_2(x)) > 0$. □

Lemma 9.2.5 *Let ϕ be an irreducibility measure.*

$$\phi(\{y \in \mathbb{X} : \phi(R_1(y)) > 0 \ \textit{and} \ \phi(R_2(y) > 0\}) > 0.$$

Proof: We know from lemma 9.2.4 that $\int \phi(R_2(x))\phi(dx) > 0$. Thus

$$\begin{aligned} 0 &< \int \phi(R_2(x))\phi(dx) \\ &= \int\int 1\!\!1_{R_2(x)}(y)\phi(dy) \\ &= \int (\int 1\!\!1_R(x,y)\phi(dx))\phi(dy) \\ &= \int \phi(R_1(y))\phi(dy), \end{aligned}$$

proving that we have both $\int \phi(R_1(y))\phi(dy) > 0$ and $\int \phi(R_2(y))\phi(dy) > 0$. To finish the proof, we must show $I > 0$, where:

$$\begin{aligned} I &\equiv \int \phi(R_1(y))\phi(R_2(y))\phi(dy) \\ &= \int (\int 1\!\!1_R(y,t)\phi(dt))(\int 1\!\!1_R(s,y)\phi(ds))\phi(dy) \\ &= \int\int (\int 1\!\!1_R(s,y) 1\!\!1_R(y,t)\phi(dy))\phi(ds))\phi(dt). \end{aligned}$$

Instantiating the previous lemma 9.2.4 for $A = \mathbb{X}$ yields $\phi(R_2(y)) > 0$, for all $y \in \mathbb{X}$. Combined with the conclusion $\int \phi(R_1(y))\phi(dy) > 0$ obtained above, we conclude that $I > 0$. □

Sketch of the proof of theorem 9.2.3: As was the case in the proof of lemma 9.2.5, ϕ-irreducibility guarantees that $\mathbb{1}_R(s,y)\mathbb{1}_R(y,t) > 0$ on a set of strictly positive measure: there exist m_1, n_1, m_2, n_2 such that $P^{m_1}(s,y) \geq \frac{1}{n_1}$ and $P^{m_2}(y,t) \geq \frac{1}{n_2}$. Define $F = R^{m_1,n_1}$ and $G = R^{m_2,n_2}$ and consider $(\Delta_n)_n$ be a sequence of finite $\mathscr{X}$-measurable partitions of A such that for all n, the partition Δ_{n+1} is refinement of Δ_n, i.e. every element of Δ_{n+1} is a finite disjoint union of elements of Δ_n. Additionally, every such partition of A induces a partition of A^2 because if $A_{n,\alpha} \in \Delta_n$ and $A_{n,\beta} \in \Delta_n$, for $\alpha, \beta = 1, \ldots, |\Delta_n|$, then the set $A_{n,\alpha} \times A_{n,\beta} \equiv A^2_{n,(\alpha,\beta)}$ will be an element of the induced partition. Now, for all $x \in A$, there exists a unique $\delta = \delta(n,x) \in \{1, \cdots, |\Delta_n|\}$ such that $x \in A_{n,\delta}$. Denote by $\mathscr{G}_n = \sigma(A^2_{n,(\alpha,\beta)}, \alpha, \beta \in \{1, \cdots, |\Delta_n|\})$ the σ-algebra on $\mathbb{X} \times \mathbb{X}$ generated by the partition of order n.

As was the case in the proof of theorem 9.1.11, it is enough to consider the case where ϕ is a probability. For every $n \geq 1$ and $x, y \in \mathbb{X}$, the set $D \equiv A^2_{n,(\delta(n,x),\delta(n,y))} \in \mathscr{G}_n$ and we have:

$$\int_D \frac{\phi^2(F \cap D)}{\phi^2(D)} \phi(dx)\phi(dy) = \int_D \mathbb{1}_F(x,y)\phi(dx)\phi(dy).$$

Since the sets of the form D are a π-system generating $\mathscr{G}_n$. Hence we shall have

$$\frac{\phi^2(F \cap D)}{\phi^2(D)} = \mathbb{E}(\mathbb{1}_F | \mathscr{G}_n), \text{ a.s.}$$

Now the sequence $(\mathbb{E}(\mathbb{1}_F|\mathscr{G}_n))_n$ is a uniformly integrable $(\mathscr{G}_n)_n$-martingale converging almost surely[1] to $\mathbb{1}_F$. Assuming that F is contained in the σ-algebra generated by the partitions, we have therefore,

$$\lim_{n\to\infty} \frac{\phi^2(F \cap A^2_{n,(\delta(n,x),\delta(n,y))})}{\phi^2(A^2_{n,(\delta(n,x),\delta(n,y))})} = \mathbb{1}_F(x,y), (x,y) \in A^2 \setminus N.$$

The same line of arguments leads to the conclusion that

$$\lim_{n\to\infty} \frac{\phi^2(G \cap A^2_{n,(\delta(n,x),\delta(n,y))})}{\phi^2(A^2_{n,(\delta(n,x),\delta(n,y))})} = \mathbb{1}_G(x,y), (x,y) \in A^2 \setminus N.$$

Since $\phi^2(N) = 0$, we have

$$\phi(\{y \in A : \phi(F_1(y)) > 0 \text{ and } \phi(G_2(y)) > 0\}) > 0,$$

1. That means that for all $(x,y) \in A^2$ but, may be for a negligible set N.

meaning that

$$\phi(\{y \in A : \phi(F_1(y) \setminus N_1(y)) > 0; \phi(G_2(y) \setminus N_2(y)) > 0\}) > 0.$$

There exist therefore $s_0, y_0, t_0 \in A$ such that $s_0 \in F_1(y_0) \setminus N_1(y_0)$ and $t_0 \in G_2(y_0) \setminus N_2(y_0)$. For n sufficiently large denote by $\alpha = \delta(n, s_0)$, $\beta = \delta(n, y_0)$, and $\gamma = \delta(n, t_0)$; since

$$\lim_{n\to\infty} \frac{\phi^2(F \cap A^2_{n,(\alpha,\beta)})}{\phi^2(A^2_{n,(\alpha,\beta)})} = \mathbb{1}_F,$$

when the right hand side does not vanish, there exists n_0 sufficiently large so that for $n \geq n_0$ we have

$$\phi^2(F \cap A^2_{n,(\alpha,\beta)}) \geq \frac{3}{4}\phi(A_{n,\alpha})\phi(A_{n,\beta})$$

and similarly

$$\phi^2(G \cap A^2_{n,(\alpha,\gamma)}) \geq \frac{3}{4}\phi(A_{n,\alpha})\phi(A_{n,\gamma}).$$

Left hand sides of the above equations read respectively $\int_{A_{n,\alpha}} \phi(F_2(s) \cap A_{n,\beta})\phi(ds)$ and $\int_{A_{n,\gamma}} \phi(G_1(t) \cap A_{n,\beta})\phi(dt)$. On defining

$$\begin{aligned} J &= \{s \in A_{n,\alpha} : \phi(F_2(s) \cap A_{n,\beta}) \geq \frac{3}{4}\phi(A_{n,\beta})\} \\ K &= \{t \in A_{n,\gamma} : \phi(G_1(t) \cap A_{n,\beta}) \geq \frac{3}{4}\phi(A_{n,\beta})\} \end{aligned}$$

we conclude that $\phi(J) > 0$ and $\phi(K) > 0$. We can further show that for all $x \in J$ and all $z \in K$, $\phi(F_2(x) \cap G_1(z)) \geq \frac{1}{2}\phi(A_{n,\beta})$. Now, $y \in F_2(x) \cap G_1(z)$ means that $(x, y) \in F$ and $(y, z) \in G$. Recalling the definitions of F and G given at the beginning of the proof, we have:

$$\begin{aligned} p_{m_1+m_2}(x,z) &\geq \int_{\mathbb{X}} p_{m_1}(x,y)p_{m_2}(y,z)\phi(dy) \\ &\geq \int_{F_2(x)\cap G_1(z)} p_{m_1}(x,y)p_{m_2}(y,z)\phi(dy) \\ &\geq \frac{\phi(A_{n,\beta})}{2n_1n_2} \\ &\equiv \lambda > 0. \end{aligned}$$

By hypothesis, for all $x \in K$ there exists an $m \geq 1$ and an $\epsilon > 0$ such that for $C = \{x \in K : P^m(x, J) > \epsilon\}$ we have $\phi(C) > 0$. For $x, y \in C$ then,

$$\begin{aligned} p_{m+m_1+m_2}(x,y) &\geq \int_J P^m(x,dz)p_{m+m_2}(z,y) \\ &\geq \phi(J)\lambda\epsilon > 0. \end{aligned}$$

Therefore C is a c-set. □

Corollary 9.2.6 *If $\mathscr{X}$ is separable and the chain is ϕ-irreducible, then all $A \in \mathscr{X}$ such that $\phi(A) > 0$ contain c-sets.*

Definition 9.2.7 Let $\phi \in \mathscr{M}_+(\mathscr{X})$ and C a c-set. Define

$$I(C) = \{n \geq 1 : r(n, C) \equiv \inf_{(x,y)\in C^2} p_n(x, y) > 0\}$$

and

$$d(C) = \gcd I(C).$$

Proposition 9.2.8 *Let C be a c-set with $d(C) = d$, for some $d \in \mathbb{N}$.*

1. *The set $I(C)$ contains all the sufficiently large multiples of d.*
2. *If the chain is irreducible then all c-sets $C' \subseteq C$ verify $d(C') = d$.*

Proof: Since $I(C)$ is closed for addition, (1) is an immediate corollary of lemma 6.1.11.

Recall the notation $r(n, C) = \inf_{(x,y)\in C^2} p_n(x, y)$. To prove (2), note that

$$\begin{aligned} n \in I(C) &\Leftrightarrow r(n, C) > 0, \\ n' \in I(C') &\Leftrightarrow r(n', C') > 0. \end{aligned}$$

Obviously $C' \subseteq C \Rightarrow I(C) \subseteq I(C') \Rightarrow d' = \gcd I(C') \leq d = \gcd I(C)$. Let $m_1, m_2 \in I(C)$, $n' \in I(C')$, $x, y \in C$. Hence

$$\begin{aligned} p_{m_1+n'+m_2}(x, y) &\geq \int_{C'\times C'} p_{m_1}(x, x')p_{n'}(x', y')p_{m_2}(y', y)\phi(dx')\phi(dy') \\ &\geq r(m_1, C)r(n', C')r(m_2, C)(\phi(C'))^2 \\ &> 0. \end{aligned}$$

Therefore, $m_1 + n' + m_2 \in I(C)$ and similarly we prove that $m_1 + 2n' + m_2 \in I(C)$. Therefore, d divides both $m_1 + 2n' + m_2$ and $m_1 + n' + m_2$, hence it divides their difference, i.e. n'. Hence d divides any integer in $I(C')$, therefore, $d \leq d'$. □

9.3 Cycle decomposition

Proposition 9.3.1 *Let $(\mathbb{Y}, \mathscr{Y}, (\mathscr{Y}_n)_{n\in\mathbb{N}}, \mathbb{P})$ be a filtered space, $\mathscr{Y}_\infty = \vee_{n\geq 1}\mathscr{Y}_n$, and $(A_i)_i$ a sequence of events such that $A_i \in \mathscr{Y}_\infty$ for all $i \geq 1$. Then*

1. $\lim_{n\to\infty} \mathbb{P}(\cup_{i=n}^\infty A_i | \mathscr{Y}_n) = \mathbb{1}_{\cap_{m=1}^\infty \cup_{i=m}^\infty A_1} = \mathbb{1}_{\limsup_i A_i}$*;*

2. $\lim_{n\to\infty} \mathbb{P}(\cap_{i=n}^{\infty} A_i|\mathscr{Y}_n) = \mathbb{1}_{\cup_{m=1}^{\infty}\cap_{i=m}^{\infty} A_1} = \mathbb{1}_{\liminf_i A_i}$.

Proof: The two assertions can be proved similarly; only proof of the first is given here. Note that for all $k \leq n$:

$$\begin{aligned} X_n &\equiv \mathbb{P}(\cup_{i=k}^{\infty} A_i|\mathscr{Y}_n) \\ &\geq \mathbb{P}(\cup_{i=n}^{\infty} A_i|\mathscr{Y}_n) \\ &\geq \mathbb{P}(\cap_{m=1}^{\infty}\cup_{i=m}^{\infty} A_i|\mathscr{Y}_n) \\ &= \mathbb{E}(\mathbb{1}_{\limsup_i A_i}|\mathscr{Y}_n) \\ &\equiv Y_n. \end{aligned}$$

Now, $(X_n)_n$ is a positive, uniformly integrable $(\mathscr{Y}_n)_n$-martingale. As such it converges almost surely and in $\mathscr{L}^1$ towards $\mathbb{1}_{\cup_{i=k}^{\infty} A_i}$. So is $(Y_n)_n$, therefore $\lim_n Y_n = \mathbb{1}_{\limsup_i A_i}$. Exploiting the convergence of the martingales at both ends of the above inequality, yields

$$\begin{aligned} \mathbb{1}_{\cup_{i=k}^{\infty} A_i} &\geq \limsup_n \mathbb{P}(\cup_{i=n}^{\infty} A_i|\mathscr{Y}_n) \\ &\geq \liminf_n \mathbb{P}(\cup_{i=n}^{\infty} A_i|\mathscr{Y}_n) \\ &\geq \mathbb{1}_{\limsup_i A_i}. \end{aligned}$$

Finally, taking the $k \to \infty$ limit of the leftmost term in the above inequality, we get $\lim_{k\to\infty} \mathbb{1}_{\cup_{i\geq k} A_i} = \mathbb{1}_{\cap_{k\geq 1}\cup_{i\geq k} A_i} = \mathbb{1}_{\limsup_i A_i}$. □

Recall (exercise 14 in section 3.6) that $L_B(x) = \mathbb{P}_x(\cup_{n\geq 1}\{X_n \in B\})$ and $H_B(x) = \mathbb{P}_x(\cap_{n=1}^{\infty}\cup_{m\geq n}\{X_m \in B\})$.

Definition 9.3.2 Let $A \in \mathscr{X}$. The set A is called **unessential** if for all $x \in \mathbb{X}$, we have $H_A(x) = 0$; otherwise, is called **essential**. If A is essential but can be written as a countable union of unessential sets, then is called **improperly essential**, otherwise **properly essential**.

Remark: The previous definition implicitly implies that a countable union of unessential sets is not necessarily unessential. To convince the reader that this is not a vain precaution, consider the following example: let $\mathbb{X} = \mathbb{N}$ and suppose that for all $x \in \mathbb{X}$, $P(x, y) = 0$ if $y < x$ while $P(x, x+1) > 0$ (for example $P(x, x+1) = 1$). Then all states are unessential. However $\mathbb{X} = \cup_{x\in\mathbb{X}}\{x\}$ is necessarily essential.

Proposition 9.3.3 *Let $F \in \mathscr{X}$ be a closed set and $F^\circ = \{x \in \mathbb{X} : L_F(x) = 0\}$. Then $\mathbb{X} \setminus (F \cup F^\circ)$ cannot be properly essential.*

Proof: Since F is closed, it follows that

$$F^\circ \equiv \{x \in \mathbb{X} : L_F(x) = 0\} = \{x \in F^c : L_F(x) = 0\}.$$

Thus, $(F^\circ)^c = \{x \in F^c : L_F(x) > 0\} \cup F$ and

$$\mathbb{X} \setminus (F \cup F^\circ) = \{x \in F^c : L_F(x) > 0\} = \cup_{m \geq 1} G_m,$$

where $G_m = \{x \in F^c : L_F(x) \geq \frac{1}{m}\}$. It is therefore enough to show that for all m the sets G_m are unessential.

From proposition 9.3.1 and strong Markov property, we conclude that

$$\begin{aligned} L_F(X_n) &= \mathbb{P}_x(\cup_{i \geq 0}\{X_{i+1} \circ \theta^n \in F\} | \mathscr{F}_n) \\ &= \mathbb{P}_x(\cup_{i \geq n}\{X_{i+1} \in F\} | \mathscr{F}_n) \\ &\to \mathbb{1}_{\limsup_n \{X_n \in F\}}. \end{aligned}$$

Thus $\lim_{n\to\infty} L_F(X_n) = \mathbb{1}_{\{X_n \in F \text{ i.o.}\}}$. On the set $\{X_n \in G_m \text{ i.o.}\}$ we have that for all $n \in \mathbb{N}$, there exists an integer $k_n \geq 1$ such that $X_{n+k_n} \in G_m$. Therefore, $L_F(X_{n+k_n}) \geq \inf_{y \in G_m} L_F(y) \geq 1/m$. Thus

$$\begin{aligned} \{0,1\} \ni \mathbb{1}_{\{X_n \in F \text{ i.o.}\}} &= \lim_n L_F(X_n) \\ &= \lim_n L_F(X_{n+k_n}) \\ &\geq 1/m, \end{aligned}$$

i.e. $\mathbb{1}_{\{X_n \in F \text{ i.o.}\}} = 1$. This shows that on $\{X_n \in G_m \text{ i.o.}\}$, we have $\{X_n \in F \text{ i.o.}\}$. However, F is closed, hence absorbing: if $X_n \in F$ infinitely often, then there exists an integer N such that for all $n \geq N$ we have $X_n \in F$, i.e. if $X_n \in F$ infinitely often, then F^c cannot be visited but at most a finite number of times. This is in contradiction with the assumption $X_n \in G_m$ infinitely often. □

Remark: It is worth quoting here a particularly vivid image used by Döblin [5] to explain the contradiction used in the above proof: “Si un promeneur a, en traversant une rue, une probabilité $p > 0$ d’être écrasé par une voiture, il ne saurait la traverser indéfiniment car il sera tué avant (nous admettons qu’il ne puisse pas mourir autrement).”

Definition 9.3.4 Let $(C_1, \ldots, C_d)$ be a collection of d ($d \geq 2$) disjoint measurable sets of $\mathscr{X}$. We say that $(C_1, \ldots, C_d)$ is **cycle** (more precisely a d-cycle) for the stochastic kernel P if for all $i = 1, \ldots, d-1$ and all $x \in C_i$ we have $P(x, C_i) = 1$ and for all $x \in C_d$, $P(x, C_1) = 1$.

Theorem 9.3.5 *Suppose that X is a ϕ-irreducible chain. Then there exists a d-cycle $\mathscr{C} = (C_1, \ldots, C_d)$ such that*

- *The set* $A = \mathbb{X} \setminus \cup_{i=1}^{d} C_i$ *is not properly essential and verifies* $\phi(A) = 0$.
- *If* $\mathscr{C}' = (C'_1, \ldots, C'_{d'})$ *is a* d'*-cycle, then* d' *divides* d *and for all* $i = 1, \ldots, d'$, *the set* C'_i *differs from the union of* d/d' *members of the cycle* $\mathscr{C}$ *by a* ϕ*-negligible set that is not properly essential.*

We shall prove this theorem only for the case of separable $\mathscr{X}$. The non-separable case is treated according the lines developed in [34] p. 147. The proof is split into two elementary lemmata, lemma 9.3.7 and lemma 9.3.8 below. We start by stating a

Definition 9.3.6 Let $\phi \in \mathscr{M}_+(\mathscr{X})$ be an irreducibility measure and d a fixed integer and C a c-set. For an integer k with $1 \leq k \leq d$, a measurable set $A \in \mathscr{X}$ is called k**-accessible** (relatively to ϕ and C) if there exists a measurable set $B \subseteq C$ with $\phi(B) > 0$ such that $\mathbb{P}_x(\cup_{n=0}^{\infty}\{X_{nd+k} \in A\}) > 0$, for all $x \in B$. If there exists some integer k so that A is k-accessible, then A is called **accessible**.

It is evident that $A \in \mathscr{X}$ is k-accessible if and only if

$$\exists B \in \mathscr{X}, B \subseteq C, \phi(B) > 0 : \forall x \in B, \mathbb{P}_x(\tau_A \in \mathbb{L}_k) > 0,$$

where $\mathbb{L}_k = d\mathbb{N} + k$.

Lemma 9.3.7 *Let* ϕ *be an irreducibility measure and* C *a* c*-set having periodicity* $d(C) = d$. *Then* C *is* d*-accessible.*

Proof: By lemma 6.1.11, it follows that there exists an integer $N_0 \geq 1$ such that , for $n \geq N_0$ the integer nd is contained in the set $I(C) = \{n \in \mathbb{N} : \inf_{(x,y)\in C^2} p_n(x,y)\}$. Therefore, for all $x \in C$,

$$\begin{aligned} \mathbb{P}_x(\cup_{l=1}^{\infty}\{X_{ld+d} \in C\}) &\geq \mathbb{P}_x(\cup_{l=N_0}^{\infty}\{X_{ld+d} \in C\}) \\ &\geq r((N_0+1)d, C)\phi(C) \\ &> 0. \end{aligned}$$

□

Lemma 9.3.8 *Let* ϕ *be an irreducibility measure and* C *a* c*-set having periodicity* $d(C) = d \geq 2$. *Then for any integer* k *with* $1 \leq k < d$, *the set* C *cannot be* k*-accessible.*

Proof: Suppose on the contrary that there exists a k with $1 \leq k < d$ such that C is k-accessible. Then, there exists a measurable $B \subseteq C$ with $\phi(B) > 0$ such that for all $x \in B$, there exists an integer $n_x = \inf\{n \geq 1 : P^{nd+k}(x,C) > 0\}$. Denote by $n_C = \sup\{n_x, x \in C\}$; since C is a c-set, $n_C < \infty$ and obviously $P^{n_C d+k}(x,C) > 0$. Without loss of generality, we can always assume that $n_C \geq N_0$, where N_0 is defined in the course of the proof of the previous lemma 9.3.7 (why?). Hence by lemma 6.1.11, we have $n_C d \in I(C)$. Let $z \in C$ and $m \in I(C)$. For all $x \in B$ (hence $x \in C$), we have thanks to proposition 9.1.11

$$p_{n_C d+k+m}(x,z) \geq \int_C P^{n_C d+k}(x,dy)p_m(y,z) \geq r(m,C)P^{n_C d+k}(x,C) > 0.$$

Therefore, $n_C d + k + m \in I(C)$; as such is divisible by d. But d already divides $n_C d + m$. Hence, d must also divide k, in contradiction with the definition of d as $\gcd I(C)$. □

Sketch of the proof of theorem 9.3.5: Since the chain is ϕ-irreducible, there exists a c-set C. Let $d = d(C) = \gcd\{n \geq 1 : \inf_{(x,y)\in C^2} p_n(x,y) > 0\}$. For $j = 1,\ldots,d$, define: $\hat{C}_j = \{x \in \mathbb{X} : P^{nd-j}(x,C) > 0 \text{ for some } n \geq 1\}$. Since C is a c-set, it follows that $\phi(C) > 0$ and since the chain is ϕ-irreducible it follows that for all $B \in \mathscr{X}$: $\phi(B) > 0 \Rightarrow \forall x \in \mathbb{X}, L_B(x) > 0$. Every $x \in \mathbb{X}$ belongs to at least a $\hat{C}_j$ for $j = 1,\ldots,d$. However, the sets $\hat{C}_j$ are not necessarily mutually disjoint.

Lemmata 9.3.7 and 9.3.8 guarantee that C is d-accessible (relatively to ϕ and C) while cannot be k-accessible for any k, with $1 \leq k < d$. We show similarly that for $1 \leq i \neq j \leq d$, $\hat{C}_i \cap \hat{C}_j$ is not accessible. (Exercise! Hint: follow the line of reasoning used within the proof of lemma 9.3.8.) Hence $A = \cup_{0<i<j\leq d}(\hat{C}_i \cap \hat{C}_j)$ is not accessible either. Consequently, $\mathbb{X} \setminus A$ is closed. Define $C_i = \hat{C}_i \setminus A$, for $i = 1,\ldots,d$. It is then elementary to show that $A = \mathbb{X} \setminus \cup_{i=1}^d C_i$. Since A is not accessible, ϕ-irreducibility implies that $\phi(A) = 0$. Applying proposition 9.3.3 to the closed set $\mathbb{X} \setminus A$ implies that $\mathbb{X} \setminus ((\mathbb{X} \setminus A) \cup (\mathbb{X} \setminus A)^\circ)$ is not properly essential. Now

$$(\mathbb{X} \setminus A)^\circ = \{x \in \mathbb{X} : L_{\mathbb{X}\setminus A}(x) = 0\}.$$

But $\phi(\mathbb{X} \setminus A) > 0$ so that by ϕ-irreducibility, for all $x \in \mathbb{X}$, we have that $L_{\mathbb{X}\setminus A}(x) > 0$; this implies $(\mathbb{X} \setminus A)^\circ = \emptyset$. In summarising,

$$\begin{aligned} \mathbb{X} \setminus ((\mathbb{X} \setminus A) \cup (\mathbb{X} \setminus A)^\circ) &= \mathbb{X} \setminus (\mathbb{X} \setminus A) \\ &= A, \end{aligned}$$

establishing thus that A is not properly essential.

The second assertion of the theorem is proved according to the same line of reasoning (exercise!). □

Proposition 9.3.9 *Let $A \in \mathscr{X}$ be a non empty set. Denote by $A^\infty = \{x \in \mathbb{X} : H_A(x) = 1\}$ and suppose that $\mathbb{X}$ is indecomposable and properly essential. Then the following are equivalent:*

1. *A is properly essential,*
2. $A^\circ = \emptyset$,
3. $A^\infty \neq \emptyset$.

Proof: An exercise (10.1.3), postponed to the next chapter 10. □

Exercise 9.3.10 Let $\mathbb{X}$ be indecomposable and F closed. Show that F^c cannot be properly essential. (Hint: Start by showing that $F \subseteq (F^c)^\circ$.)

10

Asymptotic behaviour for ϕ-recurrent chains

In this chapter we shall establish that the asymptotic σ-algebra, $\mathscr{T}_\infty^{\mathbb{P}_\mu}$, for a ϕ-recurrent chains is finite. If moreover the chain is aperiodic, the asymptotic σ-algebra is trivial.

10.1 Tail σ-algebras

Definition 10.1.1 For $A \in \mathscr{X}$, we denote by $\Lambda(A)$ the set

$$\Lambda(A) = \cap_{m=1}^\infty \cup_{i\geq m} \{X_i \in A\} = \{X_i \in A \text{ i.o.}\}.$$

Remark: Recall that $L_A(x) = \mathbb{P}_x(\tau_A < \infty) = \mathbb{P}_x(\cup_{n\geq 1}\{X_n \in A\})$ while $H_A(x) = \mathbb{P}_x(\Lambda(A))$.

Proposition 10.1.2 *Let* $A, B \in \mathscr{X}$.

1. *If* $i_B(A) := \inf_{x\in A} L_B(x) > 0$, *then, for all* $\mu \in \mathscr{M}_1(\mathscr{X})$, *we have* $\Lambda(A) \subseteq \Lambda(B)$, $\mathbb{P}_\mu$*-a.s. Consequently* $\mathbb{P}_\mu(\Lambda(A)) \leq \mathbb{P}_\mu(\Lambda(B))$.
2. *If* $s_B(A) := \sup_{x\in A} H_B(x) < 1$, *then, for all* $\mu \in \mathscr{M}_1(\mathscr{X})$, *we have* $\Lambda(A) \subseteq \Lambda^c(B)$, $\mathbb{P}_\mu$*-a.s. Consequently* $\mathbb{P}_\mu(\Lambda(A) \cap \Lambda(B)) = 0$.

Proof: Let $A_i = \{X_{i+1} \in A\}$ and $B_i = \{X_{i+1} \in B\}$.

To prove (1) note that $L_B(X_n) = \mathbb{P}_\mu(\cup_{i\geq n} B_i | \mathscr{F}_n)$ and by proposition 9.3.1, $\lim_{n\to\infty} L_B(X_n) = \mathbb{1}_{\Lambda(B)}$, $\mathbb{P}_\mu$-a.s. If $\omega \in \Lambda(A)$ then for all $m \geq 1$, there exists a $n \geq m$ such that $\omega \in \{X_n \in A\} \subseteq \{L_B(X_n) \geq i_B(A)\}$, the latter inclusion holding because $\inf_{x\in A} L_B(x) > 0$ (and $X_n \in A$). Hence, $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
\Lambda(A) &\subseteq \limsup_n \{L_B(X_n) \geq i_B(A)\} \\
&= \{\inf_m \sup_{n\geq m} L_B(X_n) \geq i_B(A)\} \\
&\subseteq \{\limsup_n L_B(X_n) > 0\} \\
&= \{\mathbb{1}_{\Lambda(B)} > 0\} \\
&= \{\mathbb{1}_{\Lambda(B)} = 1\} \\
&= \Lambda(B).
\end{aligned}$$

To prove (2) note that $H_B(X_n) = \mathbb{P}_\mu(\cap_{m\geq n} \cup_{k\geq m} B_k | \mathscr{F}_n)$ and by proposition 9.3.1, $\lim_{n\to\infty} H_B(X_n) = \mathbb{1}_{\Lambda(B)}$, $\mathbb{P}_\mu$-a.s. If $\omega \in \Lambda(A)$ then for all $m \geq 1$, there exists a $n \geq m$ such that $\omega \in \{X_n \in A\} \subseteq \{H_B(X_n) \leq s_B(A)\}$, the latter inclusion holding because $\sup_{x\in A} H_B(x) \geq s_B(A)$ (and $X_n \in A$). Hence, $\mathbb{P}_\mu$-a.s.

$$\begin{aligned}
\Lambda(A) &\subseteq \limsup_n \{H_B(X_n) \leq s_B(A)\} \\
&\subseteq \{\liminf_n H_B(X_n) < 1\} \\
&= \{\mathbb{1}_{H(B)} < 1\} \\
&= \{\mathbb{1}_{H(B)} = 0\} \\
&= \Lambda^c(B).
\end{aligned}$$

□

Exercise 10.1.3 (Proof of the proposition 9.3.9)

- For $n \geq 1$ define

$$A_n = \{x \in \mathbb{X} : H_A(x) < 1 - \frac{1}{n}; L_A(x) > \frac{1}{n}\}.$$

  Express $\mathbb{X}$ in terms of A°, A^∞ and the family $(A_n)_n$.
- Show that for all $n \geq 1$ the sets A_n cannot be properly essential. (Hint: use proposition 10.1.2 twice.)
- Determine $A^\circ \cap A^\infty$ and conclude that A^∞ and $(\mathbb{X} \setminus A) \cap A^\circ$ are disjoint.
- Show that A° and A^∞ cannot be simultaneously non-empty, establishing thus the equivalence of statements 2 and 3 of the proposition. (Hint: show they are closed and use indecomposability of $\mathbb{X}$.)
- Conlcude that either A° or A^∞ is properly essential.
- Assume statement 3 of the proposition holds. Write $A = \cup_n B_n$ with all B_n unessential and $D_n = \{x \in A^\infty : L_{\cup_{k=1}^n B_k}(x) > \frac{1}{n}\}$. Show that for some indices n the sets D_n must be properly essential.
- Show that the previous results contradicts the statement that all B_n are unessential.

– Conclude that 3 implies 1.
– Conversely suppose that 1 holds and assume that $A^\infty = \emptyset$. Show that one can then decompose A into a countable union of unessential sets.

Corollary 10.1.4 *If (X_n) is a ϕ-recurrent chain then*

1. *For all $B \in \mathscr{X}$, with $\phi(B) > 0$, we have $H_B(x) = 1$ for all $x \in \mathbb{X}$.*
2. *Every bounded harmonic function is constant.*

Proof: To show (1), recall that ϕ-recurrence means that for all $B \in \mathscr{X}$ with $\phi(B)$, we have $L_B(x) = 1$ for all $x \in \mathbb{X}$. Let $A = \mathbb{X}$ and $B \in \mathscr{X}$ as above; then $\inf_{x\in\mathbb{X}} L_B(x) > 0$. Hence (1) of proposition 10.1.2 is verified. For $\mu = \epsilon_x$, we get: $1 = \mathbb{P}_x(\Lambda(\mathbb{X})) = H_{\mathbb{X}}(x) \leq \mathbb{P}_x(\Lambda(B)) = H_B(x) \leq 1$. Summarising:

$$B \in \mathscr{X}, \phi(B) > 0 \Rightarrow \forall x \in \mathbb{X} : H_B(x) = 1.$$

To show (2), let $h \in b\mathscr{X}$ be harmonic. By theorem 5.4.1, there exists a bounded invariant function $\Xi \in b\mathscr{F}$ such that $\lim_n h(X_n) = \Xi$, $\mathbb{P}_x$-a.s. Define $\alpha_0 = \sup\{\alpha \in \mathbb{R} : \phi(\{x \in \mathbb{X} : h(x) > \alpha\}) > 0\}$. Since $h \in b\mathscr{X}$, it follows $\alpha_0 < \infty$. Therefore, for all $\alpha \in]0, \alpha_0[$, $\phi(\{x \in \mathbb{X} : h(x) \geq \alpha\}) > 0$. From (1) we get: $H_{\{x\in\mathbb{X}:h(x)\geq\alpha\}}(x) = 1$, for all $x \in \mathbb{X}$, meaning that $h(X_n) \geq \alpha$ for infinitely many indices. Consequently, $\Xi \geq \alpha$ $\mathbb{P}_x$-a.s. because $\Xi = \lim_n h(X_n)$. Similarly, for all $\alpha > \alpha_0$, we show that for $\alpha > \alpha_0$, w have $H_{\{x\in\mathbb{X}:h(x)\leq\alpha\}}(x) = 1$, for all $x \in \mathbb{X}$ and consequently $\Xi \leq \alpha_0$ $\mathbb{P}_x$-a.s. Thus, finally, $\mathbb{P}_x(\Xi = \alpha_0) = 1$ meaning that $h(x) = \mathbb{E}_x(\Xi) = \alpha_0$. □

10.2 Structure of the asymptotic σ-algebra

This section is devoted to the proof of the following

Theorem 10.2.1 *Let $\mu \in \mathscr{M}_1(\mathscr{X})$ and (X_n) a ϕ-recurrent Markov chain.*

1. *If the chain is aperiodic, then $\mathscr{T}_\infty^{\mathbb{P}_\mu}$ is trivial.*
2. *If the chain has period d and $(C_1, \ldots, C_d)$ is a d-cycle of (X_n), then $\mathscr{T}_\infty^{\mathbb{P}_\mu}$ is atomic and the atoms (modulo $\mathbb{P}_\mu$-negligible events) are those of the events*

$$E_i = \cup_{m\geq 1} \cap_{n\geq m} \{X_{nd+i} \in C_i\}, i = 1, \ldots, d$$

that verify the condition $\mathbb{P}_\mu(E_i) > 0$.

Proof: Only the periodic case will be proved. To establish triviality of the $\mathscr{T}_\infty^{\mathbb{P}_\mu}$ σ-algebra, it is enough to show that all bounded harmonic spatio-temporal functions are constant because then, from corollary 5.4.2, it follows that the invariant σ-algebra for the spatio-temporal chain is trivial and proposition 5.3.6 guarantees further that the asymptotic events for the chain are the invariant events for the spatio-temporal chain. Let $\tilde{h}$ be such a bounded harmonic spatio-temporal function. Define $\tilde{h}'(x,n) = \tilde{h}(x,n+1)$ for all $x \in \mathbb{X}$ and all $n \in \mathbb{N}$. Then $\tilde{h}'$ is also a bounded harmonic spatio-temporal function. We shall establish in the sequel that $\tilde{h}' = \tilde{h}$ showing that $\tilde{h}$ is independent of n, hence $\tilde{h}'(x,n) = h(x)$ for all x, where h will be a bounded harmonic, hence constant by (2) of previous corollary 10.1.4.

Suppose that there exists a $z_0 = (x_0, n_0)$ such that $\tilde{h}'(z_0) \neq \tilde{h}(z_0)$. Denote by $Z_n = (X_n, T_n)$ the spatio-temporal chain. The martingales $(\tilde{h}(Z_n))_n$ and $(\tilde{h}'(Z_n))_n$ converge respectively towards Ξ and Ξ', $\tilde{\mathbb{P}}_{z_0}$-a.s. Since we have supposed that $\tilde{h}'(z_0) \neq \tilde{h}(z_0)$, we have necessarily that $\tilde{\mathbb{P}}_{z_0}(\Xi \neq \Xi') > 0$. Assume that $\tilde{\mathbb{P}}_{z_0}(\Xi < \Xi') > 0$ (the other case is treated similarly). Then there exist $a, b \in \mathbb{R}$ with $a < b$ and $\delta > 0$ such that

$$\tilde{\mathbb{P}}_{z_0}(\Xi < a; \Xi' > b) = \delta > 0.$$

On denoting $A = \{z : \tilde{h}(z) < a\}$ and $B = \{z : \tilde{h}'(z) > b\}$, the previous relationship implies

$$\tilde{\mathbb{P}}_{z_0}(\cup_{m\geq 0} \cap_{n\geq m} \{Z_n \in A \cap B\}) \geq \delta.$$

Let $g(z_0) = \tilde{\mathbb{P}}_{z_0}(\cap_{n\geq m}\{Z_n \in A \cap B\})$. Since

$$\lim_n g(Z_n) = \mathbb{1}_{\cup_{m\geq 0}\cap_{n\geq m}\{Z_n \in A\cap B\}}, \tilde{\mathbb{P}}_{z_0}\text{-a.s.},$$

appropriately choosing z, the value $g(z)$ can be made arbitrarily close to 1 with strictly positive probability $\tilde{\mathbb{P}}_z$. Corollary 9.2.6 guarantees that any measurable set A with $\phi(A) >$ contains a C-set. Let C such a set; there exist $m > 0$ and $r(m,C) > 0$ such that for all $x, y \in C$, we have both $p_m(x,y) \geq r(m,C)$ and $p_{m+1}(x,y) \geq r(m,C)$ (due to the aperiodicity and lemma 6.1.11). Now,

$$\begin{aligned}
\phi(C) > 0 &\overset{10.1.4}{\Rightarrow} H_C(x) = 1, \forall x \in \mathbb{X} \\
&\Rightarrow \mathbb{P}_{x_0}(X_n \in C \text{ i.o.}) = 1 \\
&\Rightarrow \tilde{\mathbb{P}}_{z_0}(Z_n \in C \times \mathbb{N} \text{ i.o.}) = 1. \quad (*)
\end{aligned}$$

Observe also that

$$\begin{aligned}
\{\lim_n g(Z_n) = 1\} &= \cap_{p\geq 1} \cup_{m\geq 0} \cap_{n\geq m} \{|g(Z_n) - 1| < 1/p\} \\
&= \cap_{p\geq 1} A_p,
\end{aligned}$$

where $A_p = \cup_{m\geq 0}\cap_{n\geq m}\{|g(Z_n)-1| < 1/p\}$. Therefore, the condition $\tilde{\mathbb{P}}_{z_0}(\lim_n g(Z_n) = 1) > 0$ implies that for every $r \geq 1$, we have

$$\begin{array}{rcl} \tilde{\mathbb{P}}_{z_0}(\cup_{m\geq 0}\cap_{n\geq m}\{|g(Z_n)-1| < 1/r\} & = & \tilde{\mathbb{P}}_{z_0}(A_r) \\ & \geq & \tilde{\mathbb{P}}_{z_0}(\cap_{p\geq 1} A_p) \\ & > & 0. \ (**) \end{array}$$

In other words, there exists an increasing subsequence $(n_l)_l$, with $\lim_l n_l = \infty$ such that for every $\eta' > 0$, we have simultaneously for every l

$$\tilde{\mathbb{P}}_{z_0}(X_{n_l} \in C \text{ and } |g(Z_{n_l}) - 1| < \eta') > 0.$$

In fact, abbreviate $C_{n_l} = \{X_{n_l} \in C\}$ and $G_{n_l} = \{|g(Z_{n_l}) - 1| < \eta'\}$. Then the condition $\phi(C) > 0$ implies, by virtue of (*), that there exists a strictly increasing unbounded sequence of integers $(n_l)_l$ such that for all l, $\tilde{\mathbb{P}}_{z_0}(C_{n_l}) = 1$. Similarly, condition $\tilde{\mathbb{P}}_{z_0}(\lim_n g(Z_n) = 1) > 0$ implies, by virtue of (**) that there exists an integer $N > 0$ such that for all $n \geq N$, $\tilde{\mathbb{P}}_{z_0}(G_n) > 0$. Now

$$\begin{array}{rcl} 0 & < & \tilde{\mathbb{P}}_{z_0}(G_{n_l}) \\ & = & \tilde{\mathbb{P}}_{z_0}(C_{n_l} \cap G_{n_l}) + \tilde{\mathbb{P}}_{z_0}(C^c_{n_l} \cap G_{n_l}) \\ & = & \tilde{\mathbb{P}}_{z_0}(C_{n_l} \cap G_{n_l}). \end{array}$$

Therefore, there must exist a point $z_1 = (x_1, n_1)$ with $x_1 \in C$ and $1 - g(z_1) < r(m, C)\phi(C)/4$. Let $C' = C \cap \{x \in \mathbb{X} : (x, n_1 + m) \not\in A \cap B\}$. We have then:

$$\mathbb{P}_{x_1}(X_m \in C') \geq \int_{C'} p_m(x_1, y)\phi(dy) \geq r(m, C)\phi(C')$$

and

$$\mathbb{P}_{x_1}(X_m \in C') \leq \tilde{\mathbb{P}}_{z_1}(Z_m \not\in A \cap B) \leq 1 - g(z_1) < \eta\phi(C)/4.$$

Hence, $\phi(C') \leq \phi(C)/4$. We introduce similarly $C'' = C \cap \{x \in \mathbb{X} : (x, n_1 + m + 1) \not\in A \cap B\}$ and establish that $\phi(C'') \leq \phi(C)/4$ and conclude that $\phi(C' \cup C'') \leq \phi(C)/2$. The latter implies that C contains a point x that does not belong to $C' \cup C''$; for that point x, we have $(x, n_1 + m) \in A \cap B$ and $(x, n_1 + m + 1) \in A \cap B$. Now, for $n = n_1 + m$, $(x, n) \in A \cap B \Rightarrow (x, n) \in B \Rightarrow \tilde{h}(x, n) = \tilde{h}'(x, n+1) > b$, while $(x, n) \in A \cap B \Rightarrow (x, n) \in A \Rightarrow \tilde{h}(x, n) = \tilde{h}'(x, n + 1) < a$, a contradiction due to the hypothesis that $\tilde{h} \neq \tilde{h}'$.

The proof for the periodic case is based on theorem 9.3.5. □

Corollary 10.2.2 *Let (X_n) be a ϕ-recurrent chain and $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$.*

1. *If the chain is aperiodic, then* $\lim_n \|(\mu - \nu)P^n\| = 0$.

2. *If the chain has period d, then* $\| \frac{1}{d} \lim_n \sum_{k=0}^{d-1} (\mu - \nu) P^{n+k} \| = 0$.

Proof: If the chain is aperiodic, ϕ-recurrence guarantees the triviality of $\mathscr{T}_\infty^{\mathbb{P}_\mu}$ by theorem 10.2.1, for all $\mu \in \mathscr{M}_1(\mathscr{X})$. We conclude then by theorem 5.4.4. If the chain is periodic, then $(X_{nd+r})_n$ is ϕ-irreducible and aperiodic on the cycle C_r. □

Exercise 10.2.3 (Weakening of the condition of ϕ-recurrence) A chain is called **weakly ϕ-recurrent** if $A \in \mathscr{X}$ and $\phi(A) > 0$ imply that $\forall_\phi x \in \mathbb{X}$, $H_A(x) = 1$. Consider a weakly ϕ-recurrent chain and chose $\Xi = \mathbb{1}_A$ where A is an invariant set. If we define $h(x) = \mathbb{E}_x(\Xi)$ then for all $\mu \in \mathscr{M}_1(\mathscr{X})$, $\lim_n h(X_n) = \Xi$, $\mathbb{P}_\mu$-a.s.

1. If $0 < \alpha < 1$, show that the sets $B_1 = \{x \in \mathbb{X} : h(x) > \alpha\}$ and $B_2 = \{x \in \mathbb{X} : h(x) < \alpha\}$ cannot be simultaneously of strictly positive ϕ measure.
2. Conclude that $\forall_\phi y \in \mathbb{X}$, either $h(y) = 0$ or $h(y) = 1$.
3. Let $\mu \in \mathscr{M}_+$ and $\mathbb{P}_\mu(\cdot) = \int \mu(dx) \mathbb{P}_x(\cdot)$ (even when μ is not a probability). Show that if $\mu \ll \phi$, for all $A \in \mathscr{J}_\infty^{\mathbb{P}_\mu}$, either $\mathbb{P}_\mu(A) = 0$ or $\mathbb{P}_\mu(A) = 1$.
4. Conclude that if $\mu \ll \phi$, weak ϕ-recurrence extends theorem 10.2.1 by guaranteeing the triviality of the σ-algebra $\mathscr{J}_\infty^{\mathbb{P}_\mu}$.

11

Uniform ϕ-recurrence

11.1 Exponential convergence to equilibrium

We introduce the notion of **taboo probability kernel** by

$$_BP^m(x,A) = \mathbb{P}_x(X_m \in A; X_i \not\in B, \forall i = 1, \dots m-1).$$

Definition 11.1.1 A chain (X_n) is **uniformly ϕ-recurrent** if for all $A \in \mathscr{X}$ with $\phi(A) > 0$, the following $\lim_n \sum_{m=1}^n {}_AP^m(x,A) = 1$ holds *uniformly* in x.

Remark: Although elementary, it is worth stating explicitly for the sake of definiteness what is meant by uniformity in the previous definition:

$$\forall \epsilon > 0, \exists N : \forall x \in \mathbb{X}, \forall n \geq N \Rightarrow \sum_{m=1}^{n} {}_AP^m(x,A) \geq 1 - \epsilon.$$

Theorem 11.1.2 *Let $(X_n)_n$ be a uniformly ϕ-recurrent chain. Then there exist constants $C < \infty$ and $\rho < 1$ such that for all $\mu, \nu \in \mathscr{M}_1(\mathscr{X})$, and all n,*

1. *if the chain is aperiodic,*

$$\|(\mu - \nu)P^n\| \leq C\rho^n \|\mu - \nu\|,$$

and

2. *if the chain is d-periodic,*

$$\|\frac{1}{n}\sum_{k=1}^{n}(\mu-\nu)P^k\| \leq C\rho^n\|\mu-\nu\|,$$

Proof: Only the aperiodic case will be treated in detail. Without loss of generality, we can always assume that ϕ is a finite measure. Denote $\sigma_n = (\mu - \nu)P^n$.

We first prove that the sequence $(\|\sigma_n\|)_n$ is decreasing: recall that $|\sigma_n| = \sigma_n^+ + \sigma_n^-$. Then $|\sigma_{n+1}| = |\sigma_n P| \le \sigma_n^+ P + \sigma_n^- P$. It follows that

$$\begin{aligned} \|\sigma_{n+1}\| &= |\sigma_{n+1}|(\mathbb{X}) \\ &\le \sigma_n^+ P(\mathbb{X}) + \sigma_n^- P(\mathbb{X}) \\ &= \sigma_n^+(\mathbb{X}) + \sigma_n^-(\mathbb{X}) \\ &= \|\sigma_n\|. \end{aligned}$$

Suppose for the moment that there exist $n_1 \ge 1$ and $\rho_1 < 1$ such that

$$\|\sigma_{n_1}\| = \|(\mu - \nu)P^{n_1}\| \le \rho_1 \|\mu - \nu\| \quad (*)$$

holds. Iterating, we shall then have $\|\sigma_{n_1 k}\| = \|(\mu - \nu)P^{n_1 k}\| \le \rho_1^k \|\mu - \nu\|$ for all k. Therefore, the result holds for the subsequence $(\|\sigma_{n_1 k}\|)_k$ and since the initial sequence is decreasing, the result will hold for any ρ with $\sqrt[n_1]{\rho_1} < \rho < 1$. To conlude, it is henceforth enough to establish (*).

Note that for every pair $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$, there exist a pair of mutually singular finite measures $\mu', \nu' \in \mathcal{M}_+(\mathcal{X})$ such that $\mu - \nu = \mu' - \nu'$ with $\|\mu'\| = \|\nu'\| = \|\mu - \nu\|/2$. Then

$$\begin{aligned} \|(\mu - \nu)P^n\| &= \|(\mu' - \nu')P^n\| \\ &= \|\int \mu'(dx)\epsilon_x P^n - \int \nu'(dy)\epsilon_y P^n\| \\ &= \frac{1}{\|\mu'\|}\|\int\int \mu'(dx)\nu'(dx)(\epsilon_x - \epsilon_y)P^n\| \\ &\le \frac{1}{\|\mu'\|}\int\int \mu'(dx)\nu'(dx)\|(\epsilon_x - \epsilon_y)P^n\|. \end{aligned}$$

Therefore, it is enough to show the claim for $\mu = \epsilon_x$ and $\nu = \epsilon_y$ with $x \neq y$.

The chain is aperiodic and ϕ-recurrent, therefore, since for all $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$, the convergence $\lim_n \|(\mu - \nu)P^n\| = 0$ holds, for all $\delta > 0$, there exists a $n_0 \ge 1$ such that for $n \ge n_0$ we have $\|(\mu - \nu)P^n\| < \delta/4$. For fixed $x_0 \in \mathbb{X}$ and $\delta > 0$, we can find[1] a $B \in \mathcal{X}$ with $\phi(B) > 0$ and a $n_0 \ge 1$ such that for $z \in B$ and $n \ge n_0$ we have $\|(\epsilon_{x_0} - \epsilon_z)P^n\| < \delta/4$.

The chain (X_n) is uniformly ϕ-recurrent, therefore for $B \in \mathcal{X}$ with $\phi(B) > 0$, there exists a n_1 such that $\sum_{k=1}^{n_1} {}_B P^k(y, B) > 1 - \frac{\delta}{4}$. For $n \ge \max(n_0, n_1)$, for all

1. To establish the existence of such a set with strictly positive measure, on has to use the Egorov's theorem A.2.1.

$A \in \mathscr{X}$ and all $y \in \mathbb{X}$:

$$\begin{aligned} a_n &= |P^n(x_0, A) - P^n(y, A) - \\ &\quad \sum_{k=1}^{n_1} \int_B {}_B P^k(y, dz)(P^n(x_0, A) - P^{n-k}(z, A))| \\ &\leq P^n(x_0, A)(1 - \sum_{k=1}^{n_1} \int_B {}_B P^k(y, dz)) \\ &\quad + |P^n(y, A) - \sum_{k=1}^{n_1} \int_B {}_B P^k(y, dz) P^{n-k}(z, A)|. \end{aligned}$$

Now, for all $l \leq n$:

$$\begin{aligned} P^n(y, A) &= \int_{\mathbb{X}} P^l(y, dz) P^{n-l}(z, A) \\ &\geq \int_B P^l(y, dz) P^{n-l}(z, A) \\ &\geq \sum_{k=1}^{n} \int_B {}_B P^k(y, dz) P^{n-k}(z, A). \end{aligned}$$

Finally, a_n is majorised as

$$\begin{aligned} a_n &\leq (1 - (1 - \delta/4)) + |\sum_{k=n_1+1}^{n} \int_B P^k(y, dz) P^{n-k}(z, A)| \\ &\leq \delta/4 + \sum_{k=n_1+1}^{\infty} \int_B P^k(y, dz) \\ &\leq \delta/2. \end{aligned}$$

Consider now the remainder

$$\begin{aligned} b_n &= |\sum_{k=1}^{n} \int_B {}_B P^k(y, dz)(P^n(x_0, A) - P^{n-k}(z, A))| \\ &\leq \sum_{k=1}^{n} \int_B {}_B P^k(y, dz) |P^{n-k}(x_0, A) - P^{n-k}(z, A)| \quad (\equiv b_n^1) \\ &\quad + \sum_{k=1}^{n} \int_B {}_B P^k(y, dz) |P^n(x_0, A) - P^{n-k}(x_0, A)| \quad (\equiv b_n^2) \end{aligned}$$

From the choice of B as the support of the measure ν, it follows that $\lim_n |P^{n-k}(x_0, A) - P^{n-k}(z, A)| = 0$, uniformly in A; consequently $\lim_n b_n^1 = 0$, uniformly in A. From corollary 10.2.2, it follows that for every fixed k, we have $\| P^n(x_0, \cdot) - P^{n-k}(x_0, \cdot) \| = \| (\epsilon_{x_0} - \epsilon_{x_0} P^k) P^{n-k} \| \to 0$, hence $\lim_n b_n^2 = 0$, uniformly in A. Therefore, uniformly in A and y,

$$\lim_n |\sum_{k=1}^{n} \int_B {}_B P^k(y, dz)(P^n(x_0, A) - P^{n-k}(z, A)| = 0.$$

In conclusion,

$$|P^n(x_0,A) - P^n(y,A)| \le a_n + b_n \to 0,$$

uniformly in A and y. Since $|P^n(x,A)-P^n(y,A)| \le |P^n(x_0,A)-P^n(x,A)|+|P^n(x_0,A)-P^n(y,A)| \to 0$, there exists $\rho_1 < 1$ and an integer N, such that for all x, y and $\mu = \epsilon_x$ and $\nu = \epsilon_y$, we have $\|(\mu-\nu)P^N\| \le \rho_1 \|\mu-\nu\|$. □

11.2 Embedded Markov chains

If for some $A \in \mathscr{X}$ and all $x \in \mathbb{X}$ the condition $H_A(x) = 1$ holds, then $\mathbb{P}_x(\tau_A < \infty) = 1$. It is easy then to verify that the taboo potential kernel ${}_AG(x,B)$ is a Markovian kernel when restricted on $(A, \mathscr{X}_A)$, where $\mathscr{X}_A = \{B \in \mathscr{X} : B \subseteq A\}$:

$$\begin{aligned} {}_AG(x,A) &= \mathbb{E}_x(\sum_{k=1}^{\tau_A} \mathbb{1}_A(X_k)) \\ &= \sum_{n\ge 1} \mathbb{E}_x(\sum_{k=1}^{n} \mathbb{1}_A(X_k) | \tau_A = n)\mathbb{P}_x(\tau_A = n) \\ &= \mathbb{P}_x(\tau_A < \infty) \\ &= 1. \end{aligned}$$

Definition 11.2.1 Suppose that $(X_n)_n$ is a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$ with $\mu = \epsilon_x$ for some $x \in A$. Denote by $\tau_A^{(0)} = 0$ and recursively for $n \ge 1$,

$$\tau_A^{(n)} = \inf\{k > \tau_A^{(n-1)} : X_k \in A\}.$$

Then the process (Y_n) defined by $Y_n = X_{\tau_A^{(n)}}$ is a $\mathsf{MC}((A,\mathscr{X}_A),P_A,\epsilon_x)$, where $P_A(x,B) = {}_AG(x,B)$, for all $x \in A$ and $B \in \mathscr{X}_A$. The Markov chain (Y_n) is called **embedded in** A.

Definition 11.2.2 Let X be a $\mathsf{MC}((\mathbb{X},\mathscr{X}),P,\mu)$. A set $D \in \mathscr{X}$ is called a d**-set** if the chain embedded in D is uniformly ϕ-recurrent.

Proposition 11.2.3 *If the chain (X_n) on $(\mathbb{X},\mathscr{X})$ satisfies*

$$\forall A \in \mathscr{X}, \phi(A) > 0, \exists n > 0, \exists \epsilon > 0 : \forall x \in \mathbb{X} : \sum_{k=1}^{n} {}_AP^k(x,A) > \epsilon,$$

then the chain is uniformly ϕ-recurrent.

Proof: Observe that $\sum_{k=1}^{jn} {}_AP^k(x,A) = \sum_{k=1}^{jn} \mathbb{P}_x(X_k \in A; \tau_A > k-1)$. The hypothesis guarantees for all $x \in \mathbb{X}$:

$$\begin{aligned}\sum_{k=1}^{n} {}_AP^k(x,A) &= \sum_{k=1}^{n} \mathbb{P}_x(X_k \in A; \tau_A > k-1)\\ &= \sum_{k=1}^{n} \mathbb{P}_x(X_k \in A; \tau_A = k)\\ &= \sum_{k=1}^{n} \mathbb{P}_x(X_k \in A | \tau_A = k)\mathbb{P}_x(\tau_A = k)\\ &= \mathbb{P}_x(\tau_A \le n)\\ &\ge \epsilon.\end{aligned}$$

Hence $\sup_{x\in\mathbb{X}} \mathbb{P}_x(\tau_A > n) \le 1-\epsilon$. Therefore

$$\begin{aligned}\mathbb{P}_x(\tau_A > 2n) &= \mathbb{P}_x(\tau_A > n; X_{n+1} \notin A; \dots; X_{2n} \notin A)\\ &\le \mathbb{P}_x(\tau_A > n)\mathbb{P}_{X_n}(\tau_A > n)\\ &\le \mathbb{P}_x(\tau_A > n) \sup_{y\in\mathbb{X}} \mathbb{P}_y(\tau_A > n)\\ &\le (1-\epsilon)^2.\end{aligned}$$

We conclude by noting that $\mathbb{P}_x(\tau_A > jn) \le (1-\epsilon)^j$ and consequently that $\mathbb{P}_x(\tau_A \le jn) \ge 1-(1-\epsilon)^j \to 1$, uniformly in x. □

11.3 Embedding of a general ϕ-recurrent chain

Theorem 11.3.1 *Let (X_n) be a ϕ-recurrent chain. There exists an increasing sequence of measurable d-sets, $D_1 \subseteq D_2 \subseteq \cdots$, exhausting the space, i.e. $\cup_n D_n = \mathbb{X}$.*

Proof: Since (X_n) is ϕ-recurrent, is also ϕ-irreducible. By corollary 9.2.6, we know that there exists a c-set C, i.e. $\phi(C) > 0$ and there exists a $n_0 \ge 1$ and a $r(n_0, C) > 0$ such that $\inf_{(x,y)\in C^2} p_{n_0}(x,y) \ge r(n_0,C) > 0$. Let $s(x) = \inf\{m \ge 1 : \mathbb{P}_x(\cup_{i=1}^m \{X_i \in C\} \ge \frac{1}{m}\}$. Since C is a c-set, it follows that $s(x) < +\infty$. Obviously, s is a $\mathscr{X}$-measurable function. Define for all $n \in \mathbb{N}$, the measurable sets $D_n = \{x \in \mathbb{X} : s(x) \le n_0 + n\}$. The sequence (D_n) is increasing and $D_n \uparrow \mathbb{X}$. To conclude, it remains to show that every D_n is a d-set.

Let $A \subseteq D_n$ with $\phi(A) > 0$. Now ϕ-recurrence of (X_n) implies that for all z, we have $L_A(z) = 1$. Therefore, for any $\epsilon \in]0,1[$, there exist $n_1 \ge 1$, and $B \in \mathscr{X}$ with $B \subseteq C$ and $\phi(B) > 0$ such that for all $z \in B$, we have $\mathbb{P}_z(\cup_{k=1}^{n_1}\{X_k \in A\}) > \epsilon$. Then, for all $x \in D_n$,

$$\begin{aligned}
\mathbb{P}_x(\tau_A \le n + n_0 + n_1) &= \mathbb{P}_x(\cup_{k=1}^{n+n_0+n_1}\{X_k \in A\}) \\
&\ge \mathbb{P}_x(\cup_{k_1=1}^{n} \cup_{k_2=1}^{n_0+n_1} \{X_{k_1} \in C; X_{k_1+k_2} \in A\}) \\
&\ge \mathbb{P}_x(\{\tau_C \le n\} \cap \cup_{k=1}^{n_0+n_1}\{X_{k_1} \in C; X_{\tau_c+k} \in A\}) \\
&= \mathbb{P}_x(\{\tau_C \le n\})\mathbb{P}_x(\cup_{k=1}^{n_0+n_1}\{X_{\tau_c+k} \in A\}|\tau_c \le n) \\
&\ge \frac{1}{n_0+n} \inf_{y \in C} \mathbb{P}_y(\cup_{k=1}^{n_0+n_1}\{X_k \in A\}).
\end{aligned}$$

For all $y \in C$,

$$\begin{aligned}
\mathbb{P}_y(\cup_{k=1}^{n_0+n_1}\{X_k \in A\}) &= \mathbb{P}_y(\tau_A \le n_0 + n_1) \\
&\ge \mathbb{P}_y(\{\tau_B \le n_0\} \cap \cup_{k=1}^{n_1}\{X_{\tau_B+k} \in A\}) \\
&\ge \mathbb{P}_y(\{\tau_B \le n_0\}) \inf_{z \in B} \mathbb{P}_{z \in B}(\tau_A \le n_1) \\
&\ge \mathbb{P}_y(\{\tau_B \le n_0\})\epsilon \\
&\ge \mathbb{P}_y(X_{n_0} \in B)\epsilon \\
&\ge r(n_0, C)\phi(B)\epsilon.
\end{aligned}$$

In summarising, for all $x \in D_n$,

$$\begin{aligned}
\mathbb{P}_x(\cup_{k=1}^{n+n_0+n_1}\{X_k \in A\}) &= \mathbb{P}_x(\tau_A \le n + n_0 + n_1) \\
&= \sum_{k=1}^{n+n_0+n_1} {}_A P^k(x, A) \\
&\ge \frac{1}{n_0+n} r(n_0, C)\phi(B)\epsilon.
\end{aligned}$$

If $\hat{X}_k = X_{\tau_{D_n}^{(k)}}$ for $x \in D_n$, then $(\hat{X}_k)$ is obviously a chain embedded in D_n. Denoting by $\hat{}$ quantities associated with the embedded chain, the previous property guarantees that

$$\begin{aligned}
\hat{\mathbb{P}}_x(\hat{\tau}_A \le n + n_0 + n_1) &\ge \mathbb{P}_x(\cup_{k=1}^{n+n_0+n_1}\{X_k \in A\}) \\
&\ge \frac{1}{n_0+n} r(n_0, C)\phi(B)\epsilon.
\end{aligned}$$

The latter establishes that the chain embedded in D_n is uniformly ϕ-recurrent.
□

12

Invariant measures for ϕ-recurrent chains

12.1 Convergence to equilibrium

Definition 12.1.1 A measure $\pi \in \mathcal{M}_+(\mathcal{X})$ is called **invariant** for the kernel P, if $\pi = \pi P$. If additionally $\pi(\mathbb{X}) = 1$, then π is an **invariant probability**.

Theorem 12.1.2 *Let X be a ϕ-recurrent Markov chain and $\mu \in \mathcal{M}_1(\mathcal{X})$.*

1. *If the chain admits an invariant probability π and is aperiodic, then $\lim_{n\to\infty} \|\mu P^n - \pi\| = 0$; if the chain has period d, then $\lim_{n\to\infty} \|\frac{1}{d}\sum_{k=0}^{d} \mu P^n - \pi\| = 0$.*
2. *X admits at most one invariant probability.*
3. *If X is additionally uniformly ϕ-recurrent, then X has an invariant probability π and there exist constants $a > 0$ and $\rho < 1$ such that for all $n \in \mathbb{N}$*

$$\begin{aligned} \|(\mu - \pi)P^n\| &\leq a\rho^n \text{ for the aperiodic case} \\ \|\frac{1}{d}\sum_{k=0}^{d} \mu P^n - \pi\| &\leq a\rho^n \text{ for the } d\text{-periodic case.} \end{aligned}$$

Proof: The statement "X is ϕ-recurrent" is equivalent to

$$\forall A \in \mathcal{X}, \phi(A) > 0 \Rightarrow \forall x \in \mathbb{X}, H_A(x) = 1.$$

Thanks to the corollary 10.1.4, the latter means that every bounded harmonic function is constant. The theorem 5.4.4 guarantees then the triviality of the σ-algebra $\mathcal{T}_\infty^{\mathbb{P}_\mu}$ and equivalently $\lim_{n\to\infty} \|(\mu - \pi)P^n\| = 0$ for all probability measures π. If π is further an invariant probability, statement 1 follows immediately.

Statement 2 follows from 1. As a matter of fact, if μP^n has a limit, this limit is necessarily π.

To prove statement 3 for the aperiodic case, note that uniform ϕ-recurrence implies that

$$\lim_{n\to\infty} \|\mu P^n - \mu P^{n+m}\| = \lim_{n\to\infty} \|(\mu - \mu P^m)P^n\| = 0$$

uniformly in m, thanks to the theorem 11.1.2 stating exponential convergence to equilibrium. We conclude then that, uniformly in $A \in \mathscr{X}$, the sequence $\mu P^n(A)$ converges towards $\pi(A)$ and that π is a probability. Now,

$$\mu P^{n+1}(A) = \int \mu(dx) \int P^n(x, dy) P(y, A).$$

Taking limits in both sides and using the uniformity guaranteed by the theorem 11.1.2, we get $\pi(A) = \int \pi(dy)P(y, A)$ showing the invariance of π. □

12.2 Relationship between invariant and irreducibility measures

Theorem 12.2.1 *Let X be a ϕ-recurrent chain. Then there exists a measure $\pi \in \mathscr{M}_+(\mathscr{X})$ such that*

1. *π is invariant.*
2. *For any $\pi' \in \mathscr{M}_+(\mathscr{X})$ that is invariant, there is a constant c such that $\pi' = c\pi$.*
3. *$\phi \ll \pi$.*

Proof: Since X is ϕ-recurrent, theorem 11.3.1 guarantees the existence of d-sets. Let $A \in \mathscr{X}$, with $\phi(A) > 0$ such a d-set. By the very definition 11.3.1 of a d-set, the chain embedded in A will be uniformly ϕ-recurrent; theorem 12.1.2 then guarantees that it admits an invariant probability π_A. Define for all $B \in \mathscr{X}$ a measure (not necessarily a probability) by $\pi(B) = \int_A \pi_A(dx) P_A(x, B)$, where $P_A(x, B) = {}_AG(x, B)$. (Note that P_A is Markovian on A but not on $\mathbb{X}$.)

For all $B \subseteq A$, the definition $\pi(B) = \int_A \pi_A(dx) P_A(x, B)$ yields $\pi(B) = \pi_A(B)$, because π_A is an invariant measure of P_A on $(A, \mathscr{X}_A)$ where $\mathscr{X}_A = \{B \in \mathscr{X} : B \subseteq A\}$.

If the chain X is initiated with π_A, we have for all $B \in \mathscr{X}$ that

$$\begin{aligned}
\pi(B) &= \int_A \pi_A(dx) \sum_{k=1}^{\infty} {}_AP^k(x, B) \\
&= \sum_{k=1}^{\infty} \mathbb{P}_{\pi_A}(X_k \in B, X_i \not\in A, i = 1, \dots, k-1) \\
&= \int \pi_A(dx) \sum_{k=1}^{\infty} \mathbb{P}_x(\{X_k \in B; \tau_A \geq k\}).
\end{aligned}$$

We shall show that π is σ-finite. For fixed $A \in \mathscr{X}$ with $\phi(A) > 0$ and $m, n \geq 1$, define $\mathbb{X}^{(m,n)} = \{x \in \mathbb{X} : P^m(x, A) \geq 1/n\}$. Since $\phi(A) > 0$ and the chain is ϕ-recurrent, it follows that $\cup_{m,n}\mathbb{X}^{(m,n)} = \{x \in \mathbb{X} : L_A(x) > 0\} = \mathbb{X}$. Recall that

$$\pi(\mathbb{X}^{(m,n)}) = \int \pi_A(dx)\mathbb{E}_x(\sum_{k=1}^{\tau_A} \mathbb{1}_{\mathbb{X}^{(m,n)}}(X_k)).$$

We have now

$$\begin{aligned}
\mathbb{E}_x(\sum_{k=1}^{\tau_A} \mathbb{1}_{\mathbb{X}^{(m,n)}}(X_k)) &= \mathbb{E}_x(\sum_{k=\tau_{\mathbb{X}^{(m,n)}}}^{\tau_A} \mathbb{1}_{\mathbb{X}^{(m,n)}}(X_k))\mathbb{P}_x(\tau_{\mathbb{X}^{(m,n)}} \leq \tau_A) \\
&\leq \mathbb{E}_x(\tau_A - \tau_{\mathbb{X}^{(m,n)}}; \tau_{\mathbb{X}^{(m,n)}} \leq \tau_A) \\
&= \sum_{k=0}^{\infty} \mathbb{P}_x(\tau_A - \tau_{\mathbb{X}^{(m,n)}} > k; \tau_{\mathbb{X}^{(m,n)}} \leq \tau_A) \\
&\leq m\sum_{j=0}^{\infty} \mathbb{P}_x(\tau_A - \tau_{\mathbb{X}^{(m,n)}} > jm; \tau_{\mathbb{X}^{(m,n)}} \leq \tau_A) \\
&\leq m\sum_{j=1}^{\infty} (1 - \frac{1}{n})^j \\
&< +\infty.
\end{aligned}$$

Therefore, $\pi(\mathbb{X}^{(m,n)}) < +\infty$, establishing the σ-finiteness of π.

To prove invariance of π, write

$$\begin{aligned}
\int \pi(dx)P(x,B) &= \int_A \pi_A(dx)P(x,B) + \int_{A^c} \pi(dy)P(y,B) \\
&= \int_A \pi_A(dx)P(x,B) + \int_{A^c}\int_A \pi_A(dx)P_A(x,dy)P(y,B) \\
&= \int_A \pi_A(dx)[P(x,B) + \int_{A^c} P_A(x,dy)P(y,B)] \\
&= \int_A \pi_A(dx)P_A(x,B) \\
&\equiv \pi(B).
\end{aligned}$$

This remark concludes the proof of statement 1.

To prove 2, let $\pi \in \mathscr{M}_+(\mathscr{X})$ be an arbitrary invariant measure. Let $A \in \mathscr{X}$ be such that $\pi(A) = 1$ and denote by $\pi_A = \pi\restriction_{\mathscr{X}_A}$. Remark that $\pi_A(B) = \pi(B)$ for all $B \in \mathscr{X}_A$. We must show that the latter equals $\pi_A(dx)P_A(x,B)$ for all $B \in \mathscr{X}_A$ that will be sufficient in showing invariance of π_A for the chain embedded in A. We shall establish, by recurrence, that

$$\pi(B) = \sum_{k=1}^{n}\int_A \pi(dx)_A P^k(x,B) + \int_{A^c} \pi(dx)_A P^n(x,B). \quad (*)$$

Since for $n = 1$, we have ${}_AP^k(x,B) = P(x,B)$, the above equality is true for $n = 1$. Suppose that it remains true for n. The last term of $(*)$ reads:

$$\begin{aligned}\int_{A^c} \pi(dx)\,{}_AP^n(x,B) &= \int \pi(dy) \int_{A^c} P(y,dx)\,{}_AP^n(x,B) \\ &= \int_A \pi(dy) \int_{A^c} P(y,dx)\,{}_AP^n(x,B) + \int_{A^c} \pi(dy) \int_{A^c} P(y,dx)\,{}_AP^n(x,B) \\ &= \int_A \pi(dy)\,{}_AP^{n+1}(y,B) + \int_{A^c} \pi(dy)\,{}_AP^{n+1}(y,B).\end{aligned}$$

Using the recurrence hypothesis up to n, we get

$$\pi(B) = \sum_{k=1}^{n+1} \int_A \pi(dx)\,{}_AP^k(x,B) + \int_{A^c} \pi(dx)\,{}_AP^{n+1}(x,B).$$

Using thus $(*)$, that is proved true for all n, we get:

$$\pi(B) \geq \int_A \pi_A(dx) P_A(x,B),\ \forall B \in \mathscr{X}.$$

Let $F \in \mathscr{X}$ be a subset of A. The previous inequality is true for both $B = F$ and $B = A \setminus F$. Now

$$\begin{aligned} 1 &= \pi(A) \\ &= \pi(F) + \pi(A \setminus F) \\ &\geq \int_A \pi_A(dx) P_A(x,F) + \int_A \pi_A(dx) P_A(x, A\setminus F) \\ &= \int_A \pi_A(dx) P_A(x,A) \\ &= \int_A \pi_A(dx) \hat{\mathbb{P}}_x(\hat{\tau}_A = 1) \\ &= 1,\end{aligned}$$

where $\hat{}$ refers to the chain embedded in A. We conclude[1] that

$$\pi(F) = \int_A \pi_A(dx) P_A(x,F).$$

Consequently, if $\pi, \pi' \in \mathscr{M}_+(\mathscr{X})$ are non-trivial invariant measures and $A \in \mathscr{X}$ is such that $0 < \pi(A) < \infty$ and $0 < \pi'(A) < \infty$, then the measures $\frac{\pi}{\pi(A)}$ and $\frac{\pi'}{\pi'(A)}$ coincide on A with the unique invariant probability of the chain embedded in A (see exercise 12.2.3 below proving that it is as a matter of fact possible to choose an A such that $\pi(A) > 0$). Therefore, they must coincide everywhere thanks to the above arguments. This remarks proves statement 2.

To show statement 3, consider an invariant measure $\pi \in \mathscr{M}_+(\mathscr{X})$. For all $B \in \mathscr{X}$ and all $n \in \mathbb{N}$,

$$\pi(B) = \int \pi(dx) P(x,B) = \ldots = \int \pi(dx) P^n(x,B).$$

1. Use the trivial statement $[0 \leq a' \leq a \leq 1; 0 \leq b' \leq b \leq 1; 1 = a + b = a' + b'] \Rightarrow [a = a'; b = b']$.

Now, if $\pi(B) = 0$, then, for all $n \in \mathbb{N}$, the function $P^n(\cdot, B) = 0$ π-a.e. However, if $\phi(B) > 0$ then $L_B(x) > 0$, i.e. $\sum_n P^n(x, B) > 0$ for all x, by ϕ-recurrence. This leads to a contradiction. Therefore, $\phi \ll \pi$. □

Corollary 12.2.2 *Let X be a ϕ-recurrent chain and $\pi \in \mathcal{M}_+(\mathcal{X})$ a non-trivial invariant measure of X. Then X is π-recurrent. Additionally, for $B \in \mathcal{X}$, there is equivalence among*

1. $\pi(B) > 0$,
2. $\forall x \in \mathbb{X} : L_B(x) > 0$,
3. $\forall x \in \mathbb{X} : H_B(x) = 1$.

Proof: The existence of a non-trivial invariant measure π, is guaranteed by the previous theorem. We show first the implication $3 \Rightarrow 2$; for arbitrary $x \in \mathbb{X}$,

$$\begin{aligned} \mathbb{X} : H_B(x) = 1 &\Leftrightarrow \mathbb{P}_x(\{X_n \in B \text{ i.o. }\}) = 1 \\ &\Rightarrow \mathbb{P}_x(\cup_n \{X_n \in B\}) > 0 \\ &\Leftrightarrow L_B(x) > 0. \end{aligned}$$

Next, we show the implication $2 \Rightarrow 1$:

$$\begin{aligned} L_B(x) > 0 &\Leftrightarrow \mathbb{P}_x(\cup_n \{X_n \in B\}) > 0 \\ &\Leftrightarrow \sum_n P^n(x, B) > 0 \\ &\Leftrightarrow \sum_n t^n P^n(x, B) > 0, \forall t \in]0, 1[. \end{aligned}$$

Recalling that we denote by $G_t(x, B) = \sum_{n \geq 1} t^n P^n(x, B)$, $t \in]0, 1[$, we remark that $G_t(x, \mathbb{X}) = \frac{t}{1-t}$. Therefore $M_t(x, B) = \frac{1-t}{t} G_t(x, B)$ is a Markov kernel for all every $t \in]0, 1[$. Additionally, $L_B(x) > 0 \Leftrightarrow M_t(x, B) > 0$ for all $t \in]0, 1[$. Now, it π is an invariant measure for P, then, for all $A \in \mathcal{X}$, we have:

$$\int \pi(dx) M_t(x, A) = \frac{1-t}{t} \sum_{n \geq 1} t^n \int \pi(dx) P^n(x, A) = \pi(A).$$

Therefore, π is also invariant for M_t. Moreover, $L_B(x) > 0 \Leftrightarrow M_t(x, B) > 0$; subsequently, $\pi(B) = \int \pi(dx) M_t(x, B) > 0$, establishing thus 1.

Next we show the implication $1 \Rightarrow 3$. Let $F \in \mathcal{X}$ be an arbitrary closed set with $\phi(F) > 0$ and let $\mu \in \mathcal{M}_+(\mathcal{X})$ be a measure defined by

$$\mu = \begin{cases} \pi & \text{on} \quad \mathcal{X}_F = \{B \in \mathcal{X} : B \subseteq F\} \\ 0 & \text{on} \quad \mathcal{X}_{F^c} = \{B \in \mathcal{X} : B \subseteq F^c\}, \end{cases}$$

where π is an arbitrary invariant measure of the chain. First we observe that

$$\begin{aligned}\int \mu(dx)P(x,A) &= \int_F \mu(dx)P(x,A) + \int_{F^c} \mu(dx)P(x,A)\\ &= \int_F \pi(dx)P(x,A).\end{aligned}$$

Now,

- if $A \subseteq F^c$, since F is closed, it follows that $\int_F \pi(dx)P(x,A) \le \int_F \pi(dx)P(x,F^c) = 0 = \mu(A)$.
- if $A = F$, then $\int_F \pi(dx)P(x,F) = \pi(F) = \mu(F)$.
- if $A \subseteq F$, then $\int_F \pi(dx)P(x,A) \le \int_{\mathbb{X}} \pi(dx)P(x,A) = \pi(A) = \mu(A)$, so that $\int_{\mathbb{X}} \mu(dx)P(x,A) \le \mu(A)$. Now when $A \subseteq F$, it is also true that $F \setminus A \subseteq F$. Hence $\int_{\mathbb{X}} \mu(dx)P(x,F\setminus A) \le \mu(F\setminus A)$. Subsequently[2], since $\mu(F) = \int_{\mathbb{X}} \mu(dx)P(x,A) + \int_{\mathbb{X}} \mu(dx)P(x,F\setminus A) \le \mu(A) + \mu(F\setminus A) = \mu(F)$, it follows, in particular, that $\int_{\mathbb{X}} \mu(dx)P(x,A) = \mu(A)$.

The above observations establish thus the invariance of $\mu \in \mathcal{M}_+(\mathcal{X})$. Statement 2 of the theorem 12.2.1 guarantees that $\mu = c\pi$ for some constant c. Since, by definition $\mu(F^c) = 0$ it follows that $\pi(F^c) = 0$.

Suppose now that $B \in \mathcal{X}$ is such that $\pi(B) > 0$. Denote by $B^\circ = \{x \in \mathbb{X} : L_B(x) = 0\}$ and suppose that $B^\circ \neq \emptyset$. Define $F = B^\circ \cap B^c = \{x \in B^c : L_B(x) = 0\}$. For every $x \in F$, we have

$$\begin{aligned}L_B(x) = 0 &\Leftrightarrow \forall n \in \mathbb{N} : P^n(x,B) = 0\\ &\Leftrightarrow G(x,B) = 0\\ &\Leftrightarrow \forall n \in \mathbb{N} : P^n(x,B^c) = 1\\ &\Leftrightarrow G(x,B^c) > 0.\end{aligned}$$

Now $B^c = F \sqcup F'$ where $F' = \{x \in B^c : L_B(x) > 0\}$. Therefore, for all n, we have:

$$1 = P^n(x,B^c) = P^n(x,F) + P^n(x,F').$$

Suppose that there exists an integer n_1 such that $P^{n_1}(x,F') > 0$. Then

$$\begin{aligned}G(x,B) &\ge \sum_{n\in\mathbb{N}} P^{n+n_1}(x,B)\\ &= \int_{\mathbb{X}} P^{n_1}(x,dy)G(y,B)\\ &\ge \int_{F'} P^{n_1}(x,dy)G(y,B)\\ &> 0,\end{aligned}$$

because F' is equivalently $F' = \{y \in B^c : G(y,B) > 0\}$. Hence $L_B(x) > 0$, in contradiction with $x \in F$. It follows that for all $x \in F$ and all $n \in \mathbb{N}$, we have $P^n(x,F) = 1$, meaning that F is closed.

2. Use the elementary observations: $[0 \le a \le b$ and $0 \le a' \le b'$ and $b + b' = a + a' \le b + b'] \Rightarrow [b - a = a' - b'] \Rightarrow [a = b$ and $a' = b']$.

Since $F = B^\circ \cap B^c$ is closed, applying the previous result, we get that $\pi(F^c) = 0$. Now $F^c = \{x \in B^c : L_B(x) > 0\} \cup B$, therefore the sequence of inequalities $0 = \pi(F^c) \geq \pi(B) > 0$ leads to a contradiction, due to the assumption that $B^\circ \neq \emptyset$. Therefore, $B^\circ = \emptyset$, meaning that for all $x \in \mathbb{X}$, we have $L_B(x) > 0$. Subsequently, there exists $A \in \mathscr{X}$ with $\phi(A) > 0$ such that $\inf_{x\in A} L_B(x) > 0$. Applying proposition 10.1.2, we get $\Lambda(A) \subseteq \Lambda(B)$ and by ϕ-recurrence, that for all $x \in \mathbb{X}$, $1 = H_A(x) \leq H_B(x)$. The latter shows the implication

$$[\pi(B) > 0] \Rightarrow [\forall x \in \mathbb{X} : H_B(x) = 1],$$

implying π-recurrence and establishing the implication of statement 3 out of 1.
□

Exercise 12.2.3 Let $\phi \in \mathscr{M}_+(\mathscr{X})$ be a non-trivial measure and (X_n) a ϕ-recurrent chain with invariant measure $\pi \in \mathscr{M}_+(\mathscr{X})$. The purpose of this exercise is to show that it is possible to choose a measurable set A such that $0 < \pi(A) < \infty$

1. Argue why it is always possible to find a measurable set A such that $\phi(A) > 0$.
2. Choose a fixed measurable set A with $\phi(A) > 0$. Introduce the Markovian kernel M_t for $t \in]0,1[$ defined by
$$M_t = \frac{1-t}{t} \sum_{t=1}^{n} t^n P^n.$$
Show that for all $x \in \mathbb{X}$, $M_t(x, A) > 0$.
3. Conclude that $\pi(A) > 0$.
4. Introduce the family of sets $\mathbb{X}^{(m,n)} = \{x \in \mathbb{X} : P^m(x, A) \geq \frac{1}{n}\}$ for $m, n \geq 1$ and $\mathbb{Y}^{(M,N)} = \cup_{m=1}^{M} \cup_{n=1}^{N} \mathbb{X}^{(m,n)}$. Show that $\lim_{M,N} \mathbb{Y}^{(M,N)} = \mathbb{X}$ (see proof of theorem 12.2.1).
5. Conclude that there exist integers M_0 and N_0 such that $\phi(A \cap \mathbb{Y}^{(M_0,N_0)}) > \phi(A)/2$.
6. Conclude that $A' = A \cap \mathbb{Y}^{(M_0,N_0)}$ satisfies the sought conditions.

Theorem 12.2.4 *Let (X_n) be a* MC$((\mathbb{X}, \mathscr{X}), P, \mu)$ *with $\mathscr{X}$ separable and $\pi \in \mathscr{M}_+(\mathscr{X})$ an invariant measure for P. Suppose further that*

- *$\pi(\mathbb{X}) = \infty$, and*
- *(X_n) is π-recurrent.*

Then for all $\delta > 0$ and all $x \in \mathbb{X}$,

$$\lim_{n\to\infty} \frac{P^n(x, A)}{\pi(A) + \delta} = 0, \quad \text{uniformly in } A \in \mathscr{X}.$$

Proof: : Given in exercise 12.2.5. □

Exercise 12.2.5 (Proof of the theorem 12.2.4)

1. Show that if the conclusion were false one could find $\delta_0 > 0$, $x_0 \in \mathbb{X}$, $\theta_0 > 0$, a strictly increasing sequence of integers $(n_k)_{k\in\mathbb{N}}$, and a sequence of measurable sets $(A_k)_{k\in\mathbb{N}}$, such that
$$\frac{P^{n_k}(x_0, A_k)}{\pi(A_k) + \delta_0} \geq \theta_0, \forall k \in \mathbb{N}.$$
2. Use corollary 10.2.2 and the extension of Egorov's theorem A.2.2, to establish that it is possible to choose $B \in \mathscr{X}$ with $\pi(B) > \frac{1}{\theta_0}$ such that for all $y \in B$,
$$|P^{n_k}(x_0, A_k) - P^{n_k}(y, A_k)| \leq \delta_0\theta_0/2.$$
3. Use previous results to establish that
$$\pi(A_k) \geq \theta_0(\pi(A_k) + \frac{\delta_0}{2})\pi(B).$$
4. Conclude *ad absurdum*.

12.3 Ergodic properties

13

Quantum Markov chains

A

Complements in measure theory

A.1 Monotone class theorems

On general sets, σ-algebras are technically complicated objects. However, they are fundamental in the definition of a measure. To prove for instance that two measures coïncide, we must show that they charge equally every set of the σ-algebra. To palliate this difficulty, it is often possible to show equality of measures on some easier object than the σ-algebra itself, typically a monotone class.

A.1.1 Set systems

Definition A.1.1 Let Ω be an arbitrary non-empty universal set and $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ a given class of subsets of Ω. Then $\mathscr{A}$ is **a π-system, a semi-algebra, an algebra, or a σ-algebra** if the following conditions hold:

π-system	**semi-algebra**	**algebra**	**σ-algebra**
	$\Omega \in \mathscr{A}$	$\Omega \in \mathscr{A}$	$\Omega \in \mathscr{A}$
$\mathscr{A}$ stable in finite intersections	$\mathscr{A}$ stable in finite intersections	$\mathscr{A}$ stable in finite intersections	$\mathscr{A}$ stable in countable intersections
	For $A \in \mathscr{A}$ there exists a finite family of mutually disjoint sets $(A_1, \ldots, A_n)$ such that $A^c = \sqcup_{i=1}^{n} A_i$	$A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$	$A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$

Exercise A.1.2 1. Show that σ-algebra $\Rightarrow$ algebra $\Rightarrow$ semi-algebra $\Rightarrow$ π-system.

2. Provide some explicit examples for each class of sets in the previous definition.
3. Let $\mathscr{A}_1 = \{A \in \mathscr{P}(\Omega) : A \text{ countable or } A^c \text{ countable}\}$ and $\mathscr{A}_2 = \{A \in \mathscr{P}(\Omega) : A \text{ finite or } A^c \text{ finite}\}$. Which class do they belong to?

Definition A.1.3 Let Ω be an arbitrary non-empty universal set and $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ a given class of subsets of Ω and $(A_n)_n$ a sequence of sets of $\mathscr{A}$. Then $\mathscr{A}$ is **a monotone classe, a Dynkin system, or a λ-system** if the following conditions hold:

monotone class	**Dynkin system**	**λ-system**
	$\Omega \in \mathscr{A}$	$\Omega \in \mathscr{A}$
$A_n \uparrow A \Rightarrow A \in \mathscr{A}$	$A_1 \subseteq A_2 \Rightarrow A_2 \setminus A_1 \in \mathscr{A}$	$A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$
$A_n \downarrow A \Rightarrow A \in \mathscr{A}$	$\mathscr{A}$ is a monotone class	(A_n) mutually disjoint $\Rightarrow \cup_n A_n \in \mathscr{A}$

Exercise A.1.4 Let Ω be a universal set and $\mathscr{J}$ an arbitrary non-empty indexing set of a collection of set systems $(\mathscr{A}_j)_{j\in\mathscr{J}}$, with $\mathscr{A}_j \in \mathscr{P}(\Omega)$ for every $j \in \mathscr{J}$. If for all $j \in \mathscr{J}$, the system $\mathscr{A}_j$ is of one of the types: algebra, σ-algebra, monotone class, Dynkin system, π-system, so is the system $\cap_{j\in\mathscr{J}}\mathscr{A}_j$.

Definition A.1.5 Let $\mathscr{E} \subseteq \mathscr{P}(\Omega)$ be an arbitrary non-empty collection of subsets of a universal set Ω. We call **closure** of type γ, for $\gamma \in \{\alpha, \sigma, m, \mathscr{D}, \lambda\}$, and denote by $\gamma(\mathscr{E})$, the sets

- $\alpha(\mathscr{E}) = \cap\{\mathscr{A} : \mathscr{E} \subseteq \mathscr{A} \subseteq \mathscr{P}(\Omega); \mathscr{A} \text{ is an algebra}\}$,
- $\sigma(\mathscr{E}) = \cap\{\mathscr{A} : \mathscr{E} \subseteq \mathscr{A} \subseteq \mathscr{P}(\Omega); \mathscr{A} \text{ is a } \sigma\text{-algebra}\}$,
- $m(\mathscr{E}) = \cap\{\mathscr{A} : \mathscr{E} \subseteq \mathscr{A} \subseteq \mathscr{P}(\Omega); \mathscr{A} \text{ is a monotone class}\}$,
- $\mathscr{D}(\mathscr{E}) = \cap\{\mathscr{A} : \mathscr{E} \subseteq \mathscr{A} \subseteq \mathscr{P}(\Omega); \mathscr{A} \text{ is a Dynkin system}\}$,
- $\lambda(\mathscr{E}) = \cap\{\mathscr{A} : \mathscr{E} \subseteq \mathscr{A} \subseteq \mathscr{P}(\Omega); \mathscr{A} \text{ is a } \lambda\text{-system}\}$.

Exercise A.1.6 Let γ be the closure operator defined above and $\mathscr{E} \subseteq \mathscr{P}(\Omega)$ for some universal set Ω. Then

1. γ is isotone,
2. if $\mathscr{E}$ is an algebra, then $\sigma(\mathscr{E}) = m(\mathscr{E})$,
3. if $\mathscr{E}$ is a π-system, then $\sigma(\mathscr{E}) = \mathscr{D}(\mathscr{E})$,
4. if $\mathscr{E}$ is a π-system, then $\sigma(\mathscr{E}) = \lambda(\mathscr{E})$,
5. if $\mathscr{E}$ is a π-system and a λ-system then $\mathscr{E}$ is a σ-algebra.

A.1.2 Set functions and their extensions

Definition A.1.7 Let Ω be an arbitrary non-empty universal set and $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ a given class of subsets of Ω. A set function $\mu : \mathscr{A} \to \mathbb{R}_+ \cup \{+\infty\}$ is **a content, a pre-measure, a measure, or a probability** if the following conditions hold:

Set function type	properties of $\mathscr{A}$	properties of μ
Content	semi-algebra	finitely additive
Pre-measure	semi-algebra	σ-additive
Measure	σ-algebra	σ-additive
Probability	σ-algebra	σ-additive and $\mu(\Omega) = 1$

Exercise A.1.8 Let $\mu : \mathscr{A} \to \mathbb{R}_+ \cup \{+\infty\}$ be a content. For sequences $(A_n)_{n\in\mathbb{N}}$ of sets in $\mathscr{A}$, consider the following statements:

1. μ is a pre-measure.
2. μ is sequentially continuous for increasing sequences, i.e. if $A_n \uparrow A$ then $\mu(A_n) \uparrow \mu(A)$.
3. μ is sequentially continuous for decreasing sequences of sets of finite content, i.e. if $A_n \downarrow A$ and $\mu(A_1) < +\infty$, then $\mu(A_n) \downarrow \mu(A)$.
4. μ is sequentially continuous for sequences of sets decreasing to $\emptyset$ of finite content , i.e. if $A_n \downarrow \emptyset$ and $\mu(A_1) < +\infty$, then $\mu(A_n) \downarrow 0$.

Then, $1 \Leftrightarrow 2 \Rightarrow 3 \Rightarrow 4$. If, additionally, $\mu(\Omega) < +\infty$ then all statements are equivalent.

Exercise A.1.9 Let $\mu : \mathscr{A} \to \mathbb{R}_+ \cup \{+\infty\}$ be a set function defined on some class $\mathscr{A}$ of sets.

1. If $\mathscr{A}$ is a semi-algebra and μ is finitely additive, then there exists a content $\hat{\mu}$ on $\alpha(\mathscr{A})$ such that $\hat{\mu}(A) = \mu(A)$ for all $A \in \mathscr{A}$. Moreover, $\hat{\mu}$ is a pre-measure on $\alpha(\mathscr{A})$.
2. If $\mathscr{A}$ is an algebra and μ is a pre-measure, then there exists a measure $\hat{\mu}$ on $\sigma(\mathscr{A})$ such that $\hat{\mu}(A) = \mu(A)$ for all $A \in \mathscr{A}$.

Exercise A.1.10 Let $\mathscr{A} \subseteq \mathscr{P}(\Omega)$ be a π-system generating a σ-algebra $\mathscr{F}$, i.e. $\sigma(\mathscr{A}) = \mathscr{F}$. Suppose that $\mathbb{P}$ and $\mathbb{P}'$ are two probabilities on $\mathscr{F}$ coinciding on $\mathscr{A}$. Then $\mathbb{P} = \mathbb{P}'$, i.e. they coincide on $\mathscr{F}$.

A.2 Uniformity induced by Egorov's theorem and its extension

Recall first the elementary version of Egorov's theorem

Theorem A.2.1 (Egorov's theorem) *Let* $(\mathbb{X}, \mathscr{X})$ *be a measurable space and* $\phi \in \mathscr{M}_+(\mathscr{X})$ *a finite measure. Suppose* (f_n) *be a sequence of functions such that, for all* n*, the function* f_n *is* $\mathscr{X}$*-measurable and* ϕ*-a.e. finite and that* f *is a finite* $\mathscr{X}$*-measurable function. If* $\lim_n f_n = f$ ϕ*-a.e. Then, for every* $\epsilon > 0$ *there exist a set* $B \in \mathscr{X}$*, with* $\phi(B) < \epsilon$*, such that* $\lim_n f_n = f$ *uniformly on* $\mathbb{X} \setminus B$*.*

The conclusion is in general false if ϕ is not finite. Therefore, the following generalisation, valid for the case of infinite total mass, is non trivial.

Theorem A.2.2 (Extension of Egorov's theorem) *Let* $(\mathbb{X}, \mathscr{X})$ *be a measurable space and* $\phi \in \mathscr{M}_+(\mathscr{X})$ *(not necessarily of finite mass). Suppose* (f_n) *be a sequence of functions such that, for all* n*, the function* f_n *is* $\mathscr{X}$*-measurable and* ϕ*-a.e. finite and that* f *is a finite* $\mathscr{X}$*-measurable function. If* $\lim_n f_n = f$ ϕ*-a.e., then there exists a sequence of measurable sets* (B_k) *such that* $\phi(\mathbb{X} \setminus \cup_{k=0}^{\infty} B_k) = 0$ *and* $\lim_n f_n = f$*, uniformly on every* B_k*.*

Proof: See Halmos [13] p. 90. □

B

Some classical martingale results

B.1 Uniformly integrable martingales

It is an elementary result that for every $\xi \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$ and every $\epsilon > 0$ we can chose a $\delta = \delta(\xi, \epsilon) > 0$ such that for every $F \in \mathscr{F}$, $\mathbb{P}(F) \le \delta \Rightarrow \int_F |\xi(\omega)| \mathbb{P}(d\omega) < \epsilon$. Otherwise stated, for every $\epsilon > 0$, there exists $K = K(\xi, \epsilon) \in [0, \infty[$ such that $\int_{\{|\xi| > K\}} |\xi(\omega)| \mathbb{P}(d\omega) < \epsilon$. For the sake of clarity, we recall the following

Definition B.1.1 A class $\mathscr{E} \subseteq \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$ is called **uniformly integrable** if for all $\epsilon > 0$, there exists a constant $K = K(\epsilon) \ge 0$ such that for all $\xi \in \mathscr{E}$ we have $\int_{\{|\xi| > K\}} |\xi(\omega)| \mathbb{P}(d\omega) < \epsilon$.

Remark: A uniformly integrable class is bounded in L^1. The converse is in general false. Therefore the following lemma provides us with two useful sufficient conditions of uniform integrability.

Lemma B.1.2 *Let* $\mathscr{E} \subseteq \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$.

- *If there exists* $p > 1$ *such that* $\mathbb{E}(|\xi|^p) < \infty$ *for all* $\xi \in \mathscr{E}$*, then* $\mathscr{E}$ *is uniformly integrable.*
- *If there exists* $\zeta \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}_+)$ *such that* $|\xi| \le \zeta$ *for all* $\xi \in \mathscr{E}$*, then* $\mathscr{E}$ *is uniformly integrable.*

Lemma B.1.3 *Let* $\xi \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$. *Then the class of random variables* $\mathscr{E} = \{\mathbb{E}(\xi|\mathscr{G}), \mathscr{G}$ *sub-*σ*-algebra of* $\mathscr{F}\}$ *is uniformly integrable.*

The main result on uniform martingales needed in this course is the following

Theorem B.1.4 *Let $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathbb{P})$ be a filtered space and (M_n) a $(\mathscr{F}_n)$-martingale such that the class $\mathscr{E} = \{M_n, n \in \mathbb{N}\}$ is uniformly integrable. Then, $M_\infty = \lim M_n$ exists almost surely and in L^1. Additionally, the martingale is **closed**, i.e. $M_n = \mathbb{E}(M_\infty|\mathscr{F}_n)$ for all n (so that $M_\infty = \lim \mathbb{E}(M_\infty|\mathscr{F}_n)$).*

Proof: See [34]. □

B.2 Martingale proof of Radon-Nikodým theorem

Theorem B.2.1 *Let $(\Omega, \mathscr{G}, \phi)$ be a probability space and suppose that $\mathscr{G}$ is separable. Let ψ be a finite measure on $(\Omega, \mathscr{G})$ that is absolutely continuous with respect to ϕ. Then there exist a random variable $D \in \mathscr{L}^1(\Omega, \mathscr{G}, \phi)$ such that $\psi = D\phi$. The random variable D is called (a version of) the **Radon-Nikodým derivative** of ψ with respect to ϕ, denoted $D = \frac{d\psi}{d\phi}$.*

Proof: For all $\epsilon > 0$, there exists $\delta > 0$ such that

$$G \in \mathscr{G}; \phi(G) < \delta \Rightarrow \psi(G) < \epsilon.$$

Since $\mathscr{G}$ is separable, there exists a sequence (G_n) generating $\mathscr{G}$. Define the filtration $\mathscr{G}_n = \sigma(G_1, \ldots, G_n)$ for $n \geq 1$. Obviously, $\mathscr{G}_n$ is composed of $2^{r(n)}$ possible unions of atoms $A_{n,j}$, $j = 1, \ldots, r(n)$ of $\mathscr{G}_n$. Note that every atom is of the form $H_1 \cap \cdots \cap H_n$ where $H_i = G_i$ or $H_i = G_i^c$. Since for every $\omega \in \Omega$ there exists precisely one atom containing it, we can define for every $n \geq 1$ the function $D_n : \Omega \to \mathbb{R}_+$ just by its restriction on atoms, namely:

$$A_{n,k} \ni \omega \mapsto D_n(\omega) = \begin{cases} \frac{\psi(A_{n,k})}{\phi(A_{n,k})} & \text{if } \phi(A_{n,k}) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Decomposing an arbitrary $G \in \mathscr{G}_n$ in its constituting atoms: $G = A_{n,i_1} \sqcup \ldots \sqcup A_{n,i_l}$, we compute

$$\int_G D_n d\phi = \sum_{j=1}^{l} \int_{A_{n,i_j}} D_n d\phi = \sum_{j=1}^{l} \psi(A_{n,i_j}) = \psi(G).$$

Thus, D_n is a version of the Radon-Nikodým derivative on $\mathscr{G}_n$. The previous equalities guarantee similarly that (D_n) is positive $(\mathscr{G}_n)$-martingale; therefore $D_\infty = \lim_n D_n$ exists almost surely. The measure ψ being of positive mass, there always exists constant K such that $\psi(\Omega)/K < \delta$, for the δ determined in the beginning of the proof. Now Markov inequality reads:

$$\phi(D_n > K) \leq \frac{1}{K} \int_\Omega D_n d\phi = \frac{1}{K} \psi(\Omega) < \delta.$$

The set $\{D_n > K\}$ being $\mathscr{G}$ measurable, the sequence (D_n) becomes uniformly integrable since

$$\int_{\{D_n>K\}} D_n d\phi = \psi(\{D_n > K\}) < \epsilon.$$

Consequently, the convergence $D_n \to D_\infty$ takes places also in $\mathscr{L}^1$. Therefore, the set functions $\mathscr{G} \to \mathbb{R}_+$ defined by

$$\begin{array}{rcl} G & \mapsto & \psi(G) \\ G & \mapsto & \int_G D_\infty d\phi = \lim_n \int_G D_n d\phi, \end{array}$$

coincide on the π-system $\cup\mathscr{G}_n$ and consequently on the generated σ-algebra $\mathscr{G}$, by virtue of exercise A.1.10. □

B.3 Reverse martingales

We only need elementary results concerning reverse martingales, as they can be found in chapter 1 of [12].

Definition B.3.1 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space $(Z_n)_{n\in\mathbb{N}}$ a sequence of real-valued random variables defined it and $(\mathscr{T}_n)_{n\in\mathbb{N}}$ a sequence of decreasing sub-σ-algebras of $\mathscr{F}$. The sequence (Z_n) is called a $(\mathscr{T}_n)$-**reverse martingale** or **backwards martingale** if for all $n \in \mathbb{N}$,

- $Z_n \in m\mathscr{T}_n$,
- $\mathbb{E}|Z_n| < \infty$, and
- $\mathbb{E}(Z_n|\mathscr{T}_m) = Z_m$, for all $m \geq n$.

Theorem B.3.2 *If (Z_n) is a $(\mathscr{T}_n)$-reverse martingale, then Z_n converges almost surely to finite limit Z_∞.*

Proof: See [12], p. 18. □

Remark: (Z_n) is a $(\mathscr{T}_n)$-reverse martingale if, and only if, (Z_{n-i+1}) is a $(\mathscr{T}_{n-i+1})$-martingale for $1 \leq i \leq n$.

C

Semi-martingale techniques

We regroup here some results, effectively used [8] in the constructive theory of Markov chains, to obtain sharp criteria for recurrence/transience or precise estimate of the existing moments of the recurrence time.

C.1 Integrability of the passage time

Theorem C.1.1 *Let* $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\in\mathbb{N}}, \mathbb{P})$ *be a filtered probability space and* $(Z_n)_{n\in\mathbb{N}}$ *a* $\mathbb{R}_+$ *valued* $(\mathscr{F}_n)_{n\in\mathbb{N}}$*-adapted sequence of random variables defined on it. For* $c > 0$ *we denote by* $\sigma_c = \inf\{n \geq 1 : Z_n \leq c\}$*. We assume that* $\mathbb{P}(Z_0 = z) = 1$ *for some* $z > 0$*. If* $z > c$ *and there exists* $\epsilon > 0$ *such that for all* $n \geq 0$*:*

$$\mathbb{E}(Z_{(n+1)\wedge\sigma_c} - Z_{n\wedge\sigma_c} | \mathscr{F}_n) \leq -\epsilon \mathbb{1}_{\{\sigma_c > n\}},$$

then $\mathbb{E}(\sigma_c) < \infty$*.*

Proof: Taking expectations on both sides of the strong supermartingale condition yields:

$$\mathbb{E}(Z_{(n+1)\wedge\sigma_c} - Z_{n\wedge\sigma_c}) \leq -\epsilon \mathbb{P}(\sigma_c > n).$$

Using telescopic cancellation to perform the summation over n and taking into account the positivity of the random variables we get:

$$0 \leq \mathbb{E}(Z_{(n+1)\wedge\sigma_c}) \leq -\epsilon \sum_{k=0}^{n} \mathbb{P}(\sigma_c > k) + z.$$

The latter inequalities imply

$$\mathbb{E}(\sigma_c) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbb{P}(\sigma_c > k) < \frac{z}{\epsilon} < \infty.$$

□

Bibliography

[1] Sheldon Axler, Paul Bourdon, and Wade Ramey. *Harmonic function theory*, volume 137 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2001. 85

[2] Heinz Bauer. *Maß- und Integrationstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, second edition, 1992. 18

[3] Claude Berge. *Topological spaces*. Dover Publications Inc., Mineola, NY, 1997. Including a treatment of multi-valued functions, vector spaces and convexity, Translated from the French original by E. M. Patterson, Reprint of the 1963 translation. 36

[4] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication. 62

[5] W. Doblin. Éléments d'une théorie générale des chaînes simples constantes de Markoff. *Ann. École Norm. (3)*, 57:61–111, 1940. 105, 116

[6] Joseph L. Doob. *Classical potential theory and its probabilistic counterpart*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1984 edition. 86

[7] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. 34, 35

[8] G. Fayolle, V. A. Malyshev, and M. V. Men′shikov. *Topics in the constructive theory of countable Markov chains*. Cambridge University Press, Cambridge, 1995. 79, 151

[9] F. G. Foster. On the stochastic matrices associated with certain queuing processes. *Ann. Math. Statistics*, 24:355–360, 1953. 79

[10] Peter Gänssler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 1977. 18

[11] Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988. 21

[12] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics. 149

[13] Paul R. Halmos. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y., 1950. 146

[14] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001. 3

[15] C. T. Ionescu Tulceă. Mesures dans les espaces produits. *Atti Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat. (8)*, 7:208–211 (1950), 1949. 18

[16] Marius Iosifescu. *Finite Markov processes and their applications*. John Wiley & Sons Ltd., Chichester, 1980. Wiley Series in Probability and Mathematical Statistics. 31

[17] Hisashi Ito. A new statement about the theorem determining the region of eigenvalues of stochastic matrices. *Linear Algebra Appl.*, 267:241–246, 1997. 45, 46

[18] Vadim A. Kaimanovich. Measure-theoretic boundaries of Markov chains, 0-2 laws and entropy. In *Harmonic analysis and discrete potential theory (Frascati, 1991)*, pages 145–180. Plenum, New York, 1992. 86

[19] F.I. Karpelevich. On the characteristic roots of matrices with nonnegative elements. *Izv. Akad. Nauk SSSR Ser. Mat.*, 15:361–383, 1951. 45, 46

[20] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp. *Denumerable Markov chains*. Springer-Verlag, New York, second edition, 1976. With a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40. 86

[21] A. Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag, Berlin, 1977. Reprint of the 1933 original. 15, 16

[22] L. H. Loomis. The lattice theoretic background of the dimension theory of operator algebras. *Mem. Amer. Math. Soc.*, 1955(18):36, 1955. 16

[23] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993. 105

[24] Jacques Neveu. *Bases mathématiques du calcul des probabilités*. Préface de R. Fortet. Deuxième édition, revue et corrigée. Masson et Cie, Éditeurs, Paris, 1970. 5, 18

[25] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original. 65

[26] Esa Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1984. 105

[27] Steven Orey. *Lecture notes on limit theorems for Markov chain transition probabilities*. Van Nostrand Reinhold Co., London, 1971. Van Nostrand Reinhold Mathematical Studies, No. 34. 105

[28] D. Revuz. *Markov chains*, volume 11 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1984. 5

[29] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987. 45

[30] Stanley A. Sawyer. Martin boundaries and random walks. In *Harmonic functions on trees and buildings (New York, 1995)*, volume 206 of *Contemp. Math.*, pages 17–44. Amer. Math. Soc., Providence, RI, 1997. 86

[31] Helmut H. Schaefer. *Banach lattices and positive operators*. Springer-Verlag, New York, 1974. Die Grundlehren der mathematischen Wissenschaften, Band 215. 41

[32] Denis Serre. *Matrices*, volume 216 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. Theory and applications, Translated from the 2001 French original. 32

[33] Karlheinz Spindler. *Abstract algebra with applications. Vol. I*. Marcel Dekker Inc., New York, 1994. Vector spaces and groups. 32

[34] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991. 108, 117, 148

[35] Wolfgang Woess. *Random walks on infinite graphs and groups*, volume 138 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2000. 86

Index