Markov Chains on Metric Spaces. A short Course

Michel Benaïm & Tobias Hurth
Institut de Mathématiques,
Neuchâtel University

version 10/12/2020

In the memory of Marie Duflo

These notes, written for graduate students, are an introduction to Markov chains on metric spaces and their ergodic properties. More advanced textbooks include the excellent classical books by Meyn and Tweedy [31] and Duflo [14]. The lecture notes by Hairer[22] contain some similar material and are also highly recommended. Basic knowledge in probability, measure theory, and analysis as well as some familiarity with elementary Markov chain theory is recommended.

Keywords: Markov Chains, Invariant Measures, Ergodicity, Irreducibility, Ergodic Behavior, Small and Petite sets, Harris Theorem

Contents

Chapter 1

Markov Chains

The general setting is the following. Throughout we let M denote a separable (there exists a countable dense subset) metric space with metric d (e.g. $\mathbb{R}$, $\mathbb{R}^n$) equipped with its Borel σ-field $\mathcal{B}(M)$. We let $B(M)$ (respectively $C_b(M)$) denote the set of real-valued bounded measurable (respectively bounded continuous) functions on M equipped with the norm

$$\|f\|_\infty := \sup_{x \in M} |f(x)|. \tag{1.1}$$

If μ is a (non-negative) measure on M and $f \in L^1(\mu)$ (or $f \geq 0$ measurable), we let

$$\mu f := \int_M f(x)\mu(dx)$$

denote the integral of f with respect to μ.

1.1 Markov kernels and chains

1.1.1 Markov kernel

A Markov *kernel* on M is a family of measures

$$P = \{P(x, \cdot)\}_{x \in M}$$

such that

(i) For all $x \in M$, $P(x, \cdot) : \mathcal{B}(M) \to [0, 1]$ is a probability measure;

(ii) For all $G \in \mathcal{B}(M)$, the mapping $x \in M \mapsto P(x, G) \in \mathbb{R}$ is measurable.

The Markov kernel P acts on functions $g \in B(M)$ and measures (respectively probability measures) according to the formulae:

$$Pg(x) := \int_M P(x,dy)g(y), \tag{1.2}$$

$$\mu P(G) := \int_M \mu(dx)P(x,G). \tag{1.3}$$

Remark 1.1 For all $g \in B(M)$, we have $Pg \in B(M)$ and $\|Pg\|_\infty \leq \|g\|_\infty$. Boundedness is immediate and measurability easily follows from the condition **(ii)** defining a Markov kernel (use for example the monotone class theorem from the appendix).

Remark 1.2 The term $Pg(x)$ can also be defined by (1.2) for measurable functions $g : M \to \mathbb{R}$ that are nonnegative, but not necessarily bounded. For such g, $Pg(x)$ is an element of $[0,\infty]$. This will play a role in the study of Lyapunov functions starting in Section **??**.

We let P^n denote the operator recursively defined by $P^0 g := g$ and $P^{n+1}g := P(P^n g)$ for $n \in \mathbb{N}$. Or, equivalently,

$$P^0(x,\cdot) := \delta_x \text{ and } P^{n+1}(x,G) := \int_M P^n(x,dy)P(y,G)$$

for all $n \in \mathbb{N}$ and for all $G \in \mathcal{B}(M)$. Here and throughout these notes, $\mathbb{N}$ is the set of nonnegative integers (including 0). The set of positive integers (excluding 0) will be denoted by $\mathbb{N}^*$.

Example 1.3 (countable space) Suppose M is countable. We can turn M into a separable (and complete) metric space by endowing it with the discrete metric $d(x,y) = \mathbf{1}_{x \neq y}$. The corresponding Borel σ-field is the collection of all subsets of M. A *Markov transition matrix* on M is a map $P : M \times M \to [0,1]$ such that

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

for all $x \in M$. This gives rise to a Markov kernel Q defined by

$$Q(x,G) := \sum_{y \in G} P(x,y)$$

for all $G \subset M$. Since there is a one-to-one correspondence between transition matrices and kernels on M, we shall identify P with Q and refer to it at times as a transition matrix and at times as a kernel.

1.1.2 Markov chain

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. A *filtration* $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ is an increasing sequence of σ-fields: $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all $n \in \mathbb{N}$. The data $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ is called a *filtered* probability space. An $M-$valued *adapted* stochastic process on $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ is a family $(X_n)_{n\geq 0}$ of random variables defined on $(\Omega, \mathcal{F}, \mathsf{P})$, taking values in M and such that X_n is $\mathcal{F}_n-$measurable for all $n \in \mathbb{N}$.

Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ and a Markov kernel P on M, a *Markov chain with kernel P on* $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ is an $M-$valued adapted stochastic process (X_n) on $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ such that

$$\mathsf{P}(X_{n+1} \in G | \mathcal{F}_n) = P(X_n, G)$$

for all $n \in \mathbb{N}$ and for all $G \in \mathcal{B}(M)$. Equivalently,

$$\mathsf{E}(g(X_{n+1}) | \mathcal{F}_n) = Pg(X_n)$$

for all $n \in \mathbb{N}$ and for all $g \in B(M)$ and all functions $g : M \to \mathbb{R}$ that are measurable and nonnegative. Here, $\mathsf{E}(\cdot|\mathcal{F}_n)$ denotes conditional expectation with respect to $\mathcal{F}_n$, and $\mathsf{P}(X_{n+1} \in G|\mathcal{F}_n) := \mathsf{E}(\mathbf{1}_{X_{n+1}\in G}|\mathcal{F}_n)$. In the appendix, we recall the definition of conditional expectation and list some of its basic properties, which will be used without further comment throughout the text.

If $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ is unambiguous, we simply say that (X_n) is a Markov chain with kernel P. If only one of the data $\mathbb{F}, \mathsf{P}$ or $(\mathbb{F}, \mathsf{P})$ is ambiguous, we may say that (X_n) is a *Markov chain with respect to* $\mathbb{F}$ (P, $(\mathbb{F}, \mathsf{P})$).

Given a Markov kernel P and a probability measure ν on M, there always exists a Markov chain (X_n) with kernel P and such that X_0 has law ν. As outlined in Remark 1.6, this follows from the Ionescu-Tulcea theorem.

Proposition 1.4 (Chapman-Kolmogorov Equation) *Let (X_n) be a Markov chain with kernel P. Let μ_n denote the law of X_n. Then, for every $n \in \mathbb{N}$,*

$$\mu_{n+1} = \mu_n P = \mu_0 P^{n+1}.$$

Proof For every $g \in B(M)$,

$$\mu_{n+1} g = \mathsf{E}(g(X_{n+1})) = \mathsf{E}(\mathsf{E}(g(X_{n+1})|\mathcal{F}_n)) = \mathsf{E}(Pg(X_n)) = \mu_n Pg.$$

QED

Example 1.5 (countable space) Let (X_n) be a Markov chain on a countable state space M, with transition matrix P and initial distribution μ_0. The law μ_n of the random variable X_n then satisfies

$$\mu_n(\{x\}) = \sum_{y \in M} \mu_0(\{y\}) P^n(y,x), \quad \forall x \in M,$$

where P^n is the nth power of the matrix P. In matrix-vector notation, this identity can be written as

$$\mu_n = \mu_0 P^n,$$

where μ_n and μ_0 are row vectors. In particular, if μ_0 is the Dirac measure at a point $y \in M$, then the law of X_n assigns mass $P^n(y,x)$ to every singleton $\{x\}$.

The Markov kernel P (or the associated Markov chain (X_n)) is said to be *Feller* if it takes bounded continuous functions into bounded continuous functions. It is said to be *strong Feller* if it takes bounded Borel functions into bounded continuous functions. If M is countable and equipped with the discrete metric, then every function on M is continuous. In particular, every Markov kernel on a countable set is strong Feller.

1.2 Markov and strong Markov properties

1.2.1 The law of a Markov chain

Let $X = (X_n)_{n \geq 0}$ be a Markov chain with kernel P. Then X can be seen as a random variable on $(\Omega, \mathcal{F}, \mathsf{P})$ taking values in the *space of trajectories*

$$M^{\mathbb{N}} := \{\mathbf{x} = (x_i)_{i \in \mathbb{N}} \ : x_i \in M\}$$

equipped with the product σ-field $\mathcal{B}(M)^{\otimes \mathbb{N}}$ (see Exercise 1.7).

If X_0 has law ν, we let $\mathbb{P}_\nu$ denote the law of X (i.e. the image measure of P by X) and $\mathbb{E}_\nu$ the corresponding expectation. If ν is the Dirac measure at x, we use the standard notation $\mathbb{P}_x := \mathbb{P}_{\delta_x}$ and $\mathbb{E}_x := \mathbb{E}_{\delta_x}$.

Given $k \in \mathbb{N}$ and $h_0, \dots, h_k \in B(M)$, we let $h_0 \otimes \dots \otimes h_k$ denote the map on $M^{\mathbb{N}}$ defined as

$$h_0 \otimes \dots \otimes h_k(\mathbf{x}) := h_0(x_0) \dots h_k(x_k).$$

For further reference such a map will be called a *product map* of length $k+1$. Then

$$\begin{aligned} \mathsf{E}(h_0(X_0)\dots h_k(X_k)) &= \mathbb{E}_\nu(h_0 \otimes \dots \otimes h_k) \\ = \mathbb{E}_\nu(h_0 \otimes \dots \otimes h_{k-1} P h_k) &= \nu[h_0 P[h_1 P[\dots h_{k-1} P h_k]\dots]]. \end{aligned} \tag{1.4}$$

The first equality is by definition of $\mathbb{E}_\nu$. The last one follows from the second one by induction on k. For the second equality, write

$$\begin{aligned} \mathsf{E}(h_0(X_0)\dots h_k(X_k)) &= \mathsf{E}(\mathsf{E}(h_0(X_0)\dots h_k(X_k)|\mathcal{F}_{k-1})) \\ &= \mathsf{E}(h_0(X_0)\dots h_{k-1}(X_{k-1}) P h_k(X_{k-1})). \end{aligned}$$

In particular, for all Borel sets $A_0, \dots, A_k \subset M$,

$$\begin{aligned} \mathsf{P}(X_0 \in A_0, \dots, X_k \in A_k) &= \mathbb{P}_\nu\{\mathbf{x} \in M^{\mathbb{N}} :\ (x_0, \dots, x_k) \in A_0 \times \dots \times A_k\} \\ &= \int_{A_0} \nu(dx_0) \int_{A_1} P(x_0, dx_1) \dots \int_{A_k} P(x_{k-1}, dx_k). \end{aligned}$$

Remark 1.6 (The canonical chain) The formula above can be used to show that for every Markov kernel P and for every probability measure ν on M, there exists a Markov chain (X_n) with kernel P and X_0 distributed according to ν.

Indeed, let $\Omega = M^{\mathbb{N}}$, and let $\mathcal{F} = \mathcal{B}(M)^{\otimes \mathbb{N}}$. For $n \in \mathbb{N}$, set $X_n(\omega) := \omega_n$ and let $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$, i.e. $\mathcal{F}_n$ is the smallest σ-field over Ω with respect to which $X_0, \dots, X_n$ are measurable. The pair $(\Omega, \mathcal{F})$ is called the *canonical space,* (X_n) the *canonical process* and $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ the *natural filtration* with respect to (X_n).

Now, ν is a probability measure on $(M, \mathcal{B}(M))$, and for every $n \in \mathbb{N}$ and $(\omega_0, \dots, \omega_n) \in M^{n+1}$,

$$\mathcal{P}(\omega_0, \dots, \omega_n; \cdot) := P(\omega_n, \cdot)$$

defines a probability measure on $(M, \mathcal{B}(M))$. Moreover,

$$(\omega_0, \dots, \omega_n) \mapsto \mathcal{P}(\omega_0, \dots, \omega_n; A)$$

is $\mathcal{B}(M^{n+1})-$measurable for every $A \in \mathcal{B}(M)$. By the Ionescu-Tulcea theorem (see, e.g., Theorem 2 in Chapter II.9 of [38]), there exists a unique probability measure $\mathbb{P}_\nu$ on $(\Omega, \mathcal{F})$ such that for every $n \in \mathbb{N}$ and $A_0, \dots, A_n \in \mathcal{B}(M)$,

$$\begin{aligned} & \mathbb{P}_\nu(\omega_0 \in A_0, \dots, \omega_n \in A_n) \\ = & \int_{A_0} \nu(d\omega_0) \int_{A_1} \mathcal{P}(\omega_0; d\omega_1) \dots \int_{A_n} \mathcal{P}(\omega_0, \dots, \omega_{n-1}; d\omega_n) \\ = & \int_{A_0} \nu(d\omega_0) \int_{A_1} P(\omega_0, d\omega_1) \dots \int_{A_n} P(\omega_{n-1}, d\omega_n). \end{aligned} \tag{1.5}$$

Using the result from Exercise 1.7, it is not hard to check that the canonical process (X_n) is a Markov chain on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}_\nu)$, with initial distribution ν and kernel P. The chain (X_n) is called the *canonical chain* with initial distribution ν and kernel P. A probability measure of the form in (1.5) is called a *Markov measure*.

Exercise 1.7 Let $\mathcal{B}(M^n)$ (respectively $\mathcal{B}(M^{\mathbb{N}})$) denote the Borel σ-field over M^n (respectively $M^{\mathbb{N}}$, endowed with the product topology). Let $\mathcal{F}_n$ be the σ-field over $M^{\mathbb{N}}$ generated by the *canonical projection* $\pi_n : M^{\mathbb{N}} \to M^{n+1}$ defined as $\pi_n(\mathbf{x}) := (x_i)_{i=0,\dots,n}$, and let $\mathcal{B}(M)^{\otimes\mathbb{N}}$ be the σ-field generated by the union of $\mathcal{F}_n$, $n \geq 0$. Show that $\mathcal{F}_n = \pi_n^{-1}(\mathcal{B}(M^{n+1}))$ for all $n \in \mathbb{N}$, and $\mathcal{B}(M)^{\otimes\mathbb{N}} = \mathcal{B}(M^{\mathbb{N}})$.

1.2.2 The Markov properties

For $n \in \mathbb{N}$, we let $\Theta^n : M^{\mathbb{N}} \to M^{\mathbb{N}}$ denote the *shift operator* defined by

$$\Theta^n(\mathbf{x}) := (x_{n+k})_{k\geq 0}.$$

The following proposition known as the *Markov property* easily follows from the definitions.

Proposition 1.8 (Markov Property) *Let $H : M^{\mathbb{N}} \to \mathbb{R}$ be a nonnegative or bounded measurable function and X a Markov chain with kernel P. Then*

$$\mathsf{E}(H(\Theta^n \circ X)|\mathcal{F}_n) = \mathbb{E}_{X_n}(H).$$

Proof Assume without loss of generality that H is bounded. Indeed, if H is non-negative and unbounded, there is an increasing sequence of bounded non-negative functions that converges pointwise to H, and one can apply the monotone convergence theorem. The set of bounded H satisfying the required property is a vector space, containing the constant functions and closed under bounded monotone convergence. Therefore, by the monotone class theorem (given in the appendix) and by Exercise 1.7, it suffices to check the property when $H = h_0 \otimes \ldots \otimes h_k$ is a product map. We proceed by induction on k. If $k = 0$, this is immediate. If the property holds for all product maps of length $k+1$, then

$$\begin{aligned} &\mathsf{E}(h_0(X_n)\ldots h_k(X_{n+k})h_{k+1}(X_{n+k+1})|\mathcal{F}_n) \\ &= \mathsf{E}(h_0(X_n)\ldots h_k(X_{n+k})\mathsf{E}(h_{k+1}(X_{n+k+1})|\mathcal{F}_{n+k})|\mathcal{F}_n) \\ &= \mathsf{E}(h_0(X_n)\ldots h_k(X_{n+k})Ph_{k+1}(X_{n+k})|\mathcal{F}_n) = \mathbb{E}_{X_n}(h_0 \otimes \ldots \otimes h_k Ph_{k+1}). \end{aligned}$$

By (1.4), this last term equals $\mathbb{E}_{X_n}(h_0 \otimes \ldots \otimes h_{k+1})$. **QED**

A *stopping time* on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ is a random variable $T : \Omega \to \mathbb{N} \cup \{\infty\}$ such that for all $n \in \mathbb{N}$, the event $\{T = n\} = T^{-1}(\{n\})$ lies in $\mathcal{F}_n$. The *σ-field generated by* T, denoted $\mathcal{F}_T$, is the σ-field consisting of all events $A \in \mathcal{F}$ such that

$$A \cap \{T = n\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}.$$

Exercise 1.9 **(i)** Show that if $\mathcal{F}_T$ is indeed a σ-field.

(ii) Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of stopping times on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ such that $T_n \leq T_{n+1}$ for every $n \in \mathbb{N}$. Show that $\mathcal{A}_n := \mathcal{F}_{T_n}$, $n \in \mathbb{N}$, defines a filtration on $(\Omega, \mathcal{F}, \mathsf{P})$.

The following proposition generalizes Proposition 1.8.

Proposition 1.10 (Strong Markov Property) *Let $H : M^{\mathbb{N}} \to \mathbb{R}$ be a nonnegative or bounded measurable function, X a Markov chain, and T a stopping time living on the same filtered probability space as X. Then*

$$\mathsf{E}(H(\Theta^T \circ X)|\mathcal{F}_T)\mathbf{1}_{T<\infty} = \mathbb{E}_{X_T}(H)\mathbf{1}_{T<\infty}.$$

Proof It suffices to show that for all $n \in \mathbb{N}$,

$$\mathsf{E}(H(\Theta^n \circ X)\mathbf{1}_{T=n}|\mathcal{F}_T) = \mathbb{E}_{X_n}(H)\mathbf{1}_{T=n}.$$

The right-hand side is $\mathcal{F}_T-$measurable, and for all $A \in \mathcal{F}_T$,

$$\mathsf{E}(H(\Theta^n \circ X)\mathbf{1}_{T=n}\mathbf{1}_A) = \mathsf{E}(\mathbb{E}_{X_n}(H)\mathbf{1}_{T=n}\mathbf{1}_A)$$

by the Markov property (because $\mathbf{1}_{T=n}\mathbf{1}_A$ is $\mathcal{F}_n-$measurable). This proves the result. **QED**

Chapter 2

Countable Markov Chains

This chapter presents the basic theory of countable Markov chains. The assumption that M is countable makes the proofs easier and permits to introduce, in a simple setting, some of the key notions (such as *invariant probability measures, irreducibility, positive recurrence, etc.*) that will be revisited in the subsequent chapters. Furthermore, some of the results given here, in particular in Section 2.3, will be used later to prove the main results in Chapter 6. We assume here that M is a countable set equipped with the σ-field $\mathcal{S}$ of all subsets of M, and (X_n) is a Markov chain on M with Markov kernel (or matrix) $P = P(x, y)_{x,y \in M}$. In most of this chapter, we assume without loss of generality that $\Omega = M^{\mathbb{N}}, \mathcal{F} = \mathcal{S}^{\otimes \mathbb{N}}, X_n(\omega) = \omega_n$, and $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, i.e. (X_n) is the canonical chain introduced in Remark 1.6.

2.1 Recurrence and transience

For $x \in M$, we let

$$\tau_x := \inf\{k \geq 1 : \; X_k = x\}$$

denote the first time ≥ 1 at which the chain hits x,

$$\tau_x^{(n)} := \inf\{k > \tau_x^{(n-1)} : \; X_k = x\},$$

the n^{th} time of hitting x (with $\tau_x^{(0)} := 0$), and

$$N_x := \sum_{k \geq 1} \mathbf{1}_{\{X_k = x\}} \in \mathbb{N} \cup \{\infty\}$$

the number of visits of x at or after time 1. We adopt the convention that $\inf \emptyset = +\infty$. A point x is said to be *recurrent* if

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

and *transient* otherwise.

Given $x, y \in M$ and $k \in \mathbb{N}^*$, we say that x *leads to* y *in* k *steps*, written $x \rightsquigarrow^k y$, if $P^k(x,y) > 0$. We say that x *leads to* y, written $x \rightsquigarrow y$, if $x \rightsquigarrow^k y$ for some $k \in \mathbb{N}^*$. The chain is called *irreducible* if $x \rightsquigarrow y$ for all $x, y \in M$. To any Markov chain on a countable set M with transition matrix P, one can associate a weighted directed graph as follows: Let M be the set of vertices. For any $x, y \in M$, not necessarily distinct, there is a directed edge of weight $P(x,y)$ going from x to y if and only if $P(x,y) > 0$. The chain is then irreducible if and only if the associated directed graph is connected, i.e. for any $x, y \in M$ there is a path from vertex x to vertex y that moves along directed edges. Note that a general notion of irreducibility will be defined in Chapter 5 and that every countable irreducible chain (as defined here) satisfies this general definition.

Exercise 2.1 Let $(X_n)_{n\geq 0}$ be a Markov chain on $\mathbb{Z} \setminus \{0\}$ whose transition matrix P is given by

$$P(i, i+1) = P(i, -i) = 1/2,\ i \geqslant 1$$
$$P(-1, 1) = P(i, i+1) = 1,\ i \leqslant -2.$$

Draw the weighted directed graph associated with (X_n) and determine whether the chain is irreducible.

Proposition 2.2 **(i)** *If x is transient, then $N_x < \infty$ a.s. and for all $k \geq 0$,*

$$\mathbb{P}_x(N_x = k) = a^k(1-a),$$

where $a = \mathbb{P}_x(\tau_x < \infty)$. In particular,

$$\mathbb{E}_x(N_x) = \sum_{k\geq 1} P^k(x,x) = \frac{a}{1-a} < \infty.$$

(ii) *If x is recurrent, then $\mathbb{P}_x(N_x = \infty) = 1$,*

$$\mathbb{E}_x(N_x) = \sum_{k\geq 1} P^k(x,x) = \infty,$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k=x\}} = \frac{1}{\mathbb{E}_x(\tau_x)}$$

$\mathbb{P}_x$*−a.s.*

(iii) *If the chain is irreducible, then either all points are recurrent or all points are transient. In the recurrent case, for all* $x, y \in M$,

$$\mathbb{P}_x(\tau_y < \infty) = 1 \text{ and } \mathbb{E}_x(N_y) = \infty.$$

In the transient case, for all $x, y \in M$,

$$\mathbb{E}_x(N_y) < \infty.$$

Proof (**i**). Using the strong Markov property,

$$\mathbb{P}_x(N_x = k) = \mathbb{P}_x(\tau_x^{(k)} < \infty; \tau_x^{(k+1)} = \infty) = (1-a)\mathbb{P}_x(\tau_x^{(k)} < \infty)$$

and

$$\mathbb{P}_x(\tau_x^{(k)} < \infty) = a\mathbb{P}_x(\tau_x^{(k-1)} < \infty) = \ldots = a^k.$$

(**ii**). If x is recurrent, then, using again the strong Markov property,

$$\mathbb{P}_x(\tau_x^{(n)} < \infty) = \mathbb{P}_x(\tau_x^{(n-1)} < \infty) = \ldots = 1.$$

Hence $\mathbb{P}_x(N_x = \infty) = 1$ and thus $\mathbb{E}_x(N_x) = \infty$.

For all $n \geq 1$, there exists $k(n) \geq 0$ such that $\tau_x^{(k(n))} \leq n < \tau_x^{(k(n)+1)}$. Furthermore, the random variables $(\tau_x^{(n+1)} - \tau_x^{(n)})_{n\geq 0}$ are, under $\mathbb{P}_x$, i.i.d. Thus, by the strong law of large numbers for nonnegative i.i.d. random variables,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k=x\}} = \lim_{n\to\infty} \frac{k(n)}{\tau_x^{(k(n))}} = \frac{1}{\mathbb{E}_x(\tau_x)}.$$

(**iii**). If the chain is irreducible, for all $x, y \in M$ there exist $i, j \geq 1$ and $\varepsilon > 0$ such that $P^i(x,y) \geq \varepsilon, P^j(y,x) \geq \varepsilon$. Thus $P^{k+i+j}(x,x) \geq \varepsilon^2 P^k(y,y)$ for all $k \geq 1$. Therefore, we have the implication

$$\sum_{k\geq 1} P^k(y,y) = \infty \quad \Rightarrow \quad \sum_{k\geq 1} P^k(x,x) = \infty,$$

proving that x is recurrent whenever y is recurrent and y is transient whenever x is transient.

Suppose the chain is recurrent. Fix $x, y \in M$ such that $x \neq y$ (for $x = y$ the statement holds trivially true). By irreducibility, recurrence, and the strong Markov property,

$$\varepsilon := \mathbb{P}_x(\exists k < \tau_x :\ X_k = y) > 0.$$

Thus, again using the strong Markov property,

$$\begin{aligned}\mathbb{P}_x(\tau_y > \tau_x^{(n+1)}) =&\mathbb{E}_x(\mathbb{P}_x(\tau_y > \tau_x^{(n+1)} | \mathcal{F}_{\tau_x^{(n)}}))\\ =&\mathbb{E}_x((1 - \mathbb{P}_x(\exists k < \tau_x :\ X_k = y))\mathbf{1}_{\tau_y > \tau_x^{(n)}})\\ =&(1-\varepsilon)\mathbb{P}_x(\tau_y > \tau_x^{(n)}) = \ldots = (1-\varepsilon)^{n+1}.\end{aligned}$$

Thus $\mathbb{P}_x(\tau_y > \tau_x^{(n+1)}) \to 0$ as $n \to \infty$, showing that $\mathbb{P}_x(\tau_y < \infty) = 1$. The two statements about $\mathbb{E}_x(N_y)$ follow from the identity

$$\mathbb{E}_x(N_y) = \mathbb{P}_x(\tau_y < \infty)(1 + \mathbb{E}_y(N_y)),$$

which itself follows from the strong Markov property, and is valid for both recurrent and transient chains. **QED**

Remark 2.3 Transience doesn't imply that $\mathbb{P}_x(\tau_y < \infty) < 1$ for all x, y. Consider the chain on $\mathbb{N}$ whose transition matrix is given by

$P(x, x+1) = p \in (\frac{1}{2}, 1), P(x+1, x) = 1-p$ for all $x \in \mathbb{N}$ and $P(0,0) = 1-p$.

By the strong law of large numbers, $\mathbb{P}_x(\tau_y < \infty) = 1$ for all $x < y$ and the chain is transient.

Example 2.4 (Pólya walks) The Pólya walk on $\mathbb{Z}^d$ is the Markov chain with transition matrix

$$P(x,y) = \frac{1}{2d}\mathbf{1}_{\{x \sim y\}},$$

where $x \sim y \Leftrightarrow \sum_{i=1}^d |x_i - y_i| = 1$. In 1921, Pólya proved that the associated chain is recurrent for $d \leq 2$ and transient for $d \geq 3$.

The proof for $d = 1$ goes as follows. Clearly

$$P^{2k+1}(0,0) = 0 \text{ and } P^{2k}(0,0) = \frac{1}{2^{2k}}\binom{2k}{k}.$$

Stirling's formula ($\ln(n!) = n(\ln(n) - 1) + \frac{1}{2}(\ln(n) + \ln(2\pi)) + O(\frac{1}{n})$) then yields

$$P^{2k}(0,0) \sim \frac{1}{\sqrt{2\pi k}}.$$

This proves that $\sum_k P^k(0,0) = \infty$, hence the recurrence.

For $d = 2$, recurrence can be deduced from Exercise 2.5 below. The proof of transience for $d \geq 3$ is slightly more involved and can be found in classical textbooks (see e.g [5] or Woess's book [41] for a more advanced textbook on Markov chains on graphs and groups).

Exercise 2.5 Let $X_n = (X_n^1, \ldots, X_n^d)$, where the $(X_n^i), i = 1, \ldots, d$ are independent Pólya walks on $\mathbb{Z}$. Show that (X_n) is recurrent if and only if $d \leq 2$. Deduce from this result the recurrence of the Pólya walk on $\mathbb{Z}^2$.

2.1.1 Positive recurrence

A recurrent point x is called *positive recurrent* if $\mathbb{E}_x(\tau_x) < \infty$ and *null recurrent* otherwise.

A probability measure π on M is called *invariant* for a transition matrix P if $\pi P = P$, or equivalently,

$$\pi(x) = \sum_{y \in M} \pi(y) P(y, x)$$

for all $x \in M$. Here, we write $\pi(x)$ instead of $\pi(\{x\})$ to highlight the link with matrix-vector notation. Precisely, if $M = \{1, \ldots, N\}$ or $M = \mathbb{N}^*$, and if $x \in M$, then $\pi(x)$ is the xth entry of the row vector $\pi = (\pi(\{1\}), \pi(\{2\}), \ldots, \pi(\{N\}))$ or $\pi = (\pi(\{1\}), \pi(\{2\}), \ldots)$. If π is invariant for P and if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \geq 1$ (see Example 1.5).

The next result shows that for an irreducible recurrent kernel, either all points are positive recurrent or all points are null recurrent. Moreover, positive recurrence equates the existence of an invariant probability measure.

Theorem 2.6 *Suppose P is irreducible. Then the following assertions are equivalent:*

(a) *There exists an invariant probability measure π for P;*

(b) *There exists a positive recurrent point.*

Under these equivalent conditions:

(i) *All the points are positive recurrent;*

(ii) *For every initial probability distribution ν on M, and $x \in M$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k=x\}} = \pi(x) = \frac{1}{\mathbb{E}_x(\tau_x)}$$

$\mathbb{P}_\nu-$a.s. (in particular π is unique);

(iii) *For all $x \in M$ and $f : M \to \mathbb{R}$ bounded or $f : M \to [0,\infty]$,*

$$\pi f = \frac{\mathbb{E}_x(\sum_{k=0}^{\tau_x-1} f(X_k))}{\mathbb{E}_x(\tau_x)};$$

(iv) *For all $x, y \in M$, $\mathbb{E}_y(\tau_x) < \infty$.*

Proof For all $x \in M$, $\sum_{k=1}^{n} \mathbf{1}_{\{X_k=x\}} = \mathbf{1}_{\{\tau_x<\infty\}} \sum_{k=\tau_x}^{n} \mathbf{1}_{\{X_k=x\}}$. Then, using irreducibility and Proposition 2.2, one has for every probability measure ν on M

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \mathbf{1}_{\{X_k=x\}}}{n} = \frac{\mathbf{1}_{\{\tau_x<\infty\}}}{\mathbb{E}_x(\tau_x)} \tag{2.1}$$

$\mathbb{P}_\nu-$a.s., with the convention that the right-hand term is zero if x is transient. Suppose now that π is an invariant probability measure. By irreducibility and the relation $\pi(x) = \sum_y \pi(y)P(y,x)$, one sees that $\pi(x) > 0$ for all $x \in M$. Taking $\mathbb{E}_\pi$-expectation on both sides of (2.1) and using dominated convergence gives

$$0 < \pi(x) = \frac{\mathbb{P}_\pi(\tau_x < \infty)}{\mathbb{E}_x(\tau_x)}.$$

This implies $\mathbb{E}_x(\tau_x) < \infty$ so that x is positive recurrent. By Proposition 2.2 (**iii**), recurrence implies $\mathbb{P}_\pi(\tau_x < \infty) = 1$. Thus $\pi(x) = \frac{1}{\mathbb{E}_x(\tau_x)}$. Suppose now that there exists a positive recurrent point x. Let π be the probability measure defined as in Assertion (**iii**) of Theorem 2.6. We claim that π is an invariant probability measure (compare with Exercise 4.24). For all $f \in B(M)$,

$$\mathbb{E}_x(\tau_x)\ \pi f = \mathbb{E}_x(\sum_{k\geq 0} \mathbf{1}_{\{k<\tau_x\}} f(X_k)) = \mathbb{E}_x(\sum_{k\geq 0} \mathbf{1}_{\{k<\tau_x\}} f(X_{k+1}))$$

because $f(X_{\tau_x}) = f(x)$. Thus, using the Markov property and Fubini's theorem,

$$\begin{aligned}\mathbb{E}_x(\tau_x)\ \pi f =& \sum_{k\geq 0} \mathbb{E}_x(\mathsf{E}(f(X_{k+1})\mathbf{1}_{\{k<\tau_x\}}|\mathcal{F}_k)) \\ =& \mathbb{E}_x(\sum_{k\geq 0} \mathbf{1}_{\{k<\tau_x\}} Pf(X_k)) = \mathbb{E}_x(\tau_x)\pi(Pf).\end{aligned}$$

This shows that $\pi P f = \pi f$, hence $\pi P = \pi$.

It remains to prove Assertion (**iv**). Let $x \neq y \in M$. By irreducibility one can choose $k \geq 1$ such that $P^k(x,y) > 0$. Let $\tau_{k,x} := \inf\{n \geq k\ : X_n = x\}$. Then $\tau_{k,x} \leq \tau_x^{(k)}$ and, consequently,

$$\mathbb{E}_x(\tau_{k,x}) \leq \mathbb{E}_x(\tau_x^{(k)}) = \frac{k}{\pi(x)}.$$

Here the last equality follows from Assertion (**ii**) and the strong Markov property. By the Markov property,

$$\mathbb{E}_x(\tau_{k,x}) = k + \mathbb{E}_x(\mathbb{E}_{X_k}(\tau_x \mathbf{1}_{\{X_k \neq x\}})) \geq k + P^k(x,y)\mathbb{E}_y(\tau_x).$$

This shows that

$$\mathbb{E}_y(\tau_x) \leq \frac{k(1-\pi(x))}{\pi(x)P^k(x,y)} < \infty.$$

QED

An irreducible kernel (or chain) satisfying one of the equivalent conditions (**a**) or (**b**) of Theorem 2.6 is called a positive recurrent kernel (chain).

Corollary 2.7 *If M is finite and P is irreducible, then P is positive recurrent.*

Proof The set $\mathcal{P}(M)$ of probability measures on M is nothing but the unit simplex in $\mathbb{R}^d$ with d the cardinality of M. By Brouwer's fixed point theorem (see, e.g., Corollary XVI.2.2 in [15]), the map $\mathcal{P}(M) \ni \pi \mapsto \pi P \in \mathcal{P}(M)$ has a fixed point, which is then an invariant probability measure for P. **QED**

Remark 2.8 The proof of Corollary 2.7 shows that every Markov chain on a finite set, possibly non-irreducible, always admits (at least) one invariant probability measure.

Exercise 2.9 Show that the Pólya walks on $\mathbb{Z}$ and $\mathbb{Z}^2$ are null recurrent. (*Hint:* Show that they don't have any invariant probability measure.)

Exercise 2.10 [Reversibility] Let π be a probability measure on M. A Markov kernel P is said to be *reversible* with respect to π if $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x, y \in M$.

(i) Show that if P is reversible with respect to π, then π is invariant for P.

(ii) Show that if P is reversible with respect to π and if $\pi(x) > 0$ for all $x \in M$, then $Pf(x) := \sum_{y\in M} P(x,y)f(y)$ defines a self-adjoint operator on the Hilbert space $l^2(\pi) := \{f : M \to \mathbb{R} : \sum_{x\in M} \pi(x)|f(x)|^2 < \infty\}$ with inner product $\langle f, g\rangle := \sum_{x\in M} \pi(x)f(x)g(x)$, i.e. $\langle Pf, g\rangle = \langle f, Pg\rangle$ for all $f, g \in l^2(\pi)$.

(iii) Give an example of a Markov kernel P and a probability measure π such that π is invariant for P, but P is not reversible with respect to π.

An interesting consequence of Theorem 2.6 (**iii**) is the next proposition, which relates moments of the first return time to x to π-mean moments of the hitting time of x.

Proposition 2.11 *Suppose P is positive recurrent with invariant probability measure π. Then for every nonnegative function $\psi : \mathbb{N} \to \mathbb{R}_+$ and every $x \in M$,*

$$\mathbb{E}_\pi(\psi(\tau_x)) = \pi(x)\mathbb{E}_x(\sum_{k=1}^{\tau_x} \psi(k)).$$

In particular, for every $\lambda > 0$,

$$\mathbb{E}_\pi(e^{\lambda\tau_x}) = \pi(x)\frac{e^\lambda}{e^\lambda - 1}[\mathbb{E}_x(e^{\lambda\tau_x}) - 1];$$

And for every $p \geq 0$,

$$\mathbb{E}_\pi(\tau_x^p) \leq \pi(x)\frac{\mathbb{E}_x[(\tau_x + 1)^{p+1}] - 1}{p+1}.$$

Proof Fix $\psi : \mathbb{N} \to \mathbb{R}_+$ and $x \in M$. By Theorem 2.6 (**iii**) applied to $f(y) := \mathbb{E}_y(\psi(\tau_x))$, one has

$$\mathbb{E}_\pi(\psi(\tau_x)) = \pi(x)\mathbb{E}_x(\sum_{k\geq 0} \mathbf{1}_{\tau_x>k}\mathbb{E}_{X_k}(\psi(\tau_x))) = \pi(x)\sum_{k\geq 0} \mathbb{E}_x(\mathbf{1}_{\tau_x>k}\mathbb{E}_{X_k}(\psi(\tau_x))).$$

But, by the Markov property,

$$\mathbb{E}_x(\mathbf{1}_{\tau_x>k}\mathbb{E}_{X_k}(\psi(\tau_x))) = \mathbb{E}_x(\mathbb{E}_x(\psi(\tau_x-k)\mathbf{1}_{\tau_x>k}|\mathcal{F}_k)) = \mathbb{E}_x(\psi(\tau_x-k)\mathbf{1}_{\tau_x>k}).$$

This proves the result. **QED**

2.1.2 Null recurrence

Although an irreducible null recurrent chain has no invariant probability measure (for otherwise it would be positive recurrent) it always has an unbounded invariant measure.

Theorem 2.12 *Suppose P is irreducible and null recurrent. Given $x \in M$, let π be the measure on M defined by*

$$\pi f = \mathbb{E}_x(\sum_{k=0}^{\tau_x-1} f(X_k))$$

for $f : M \to \mathbb{R}$ nonnegative. Then π is σ-finite ($\pi(y) < \infty$ for all $y \in M$), positive ($\pi(y) > 0$ for all $y \in M$), unbounded ($\pi(M) = \infty$), and invariant under P ($\pi = \pi P$). Every other σ-finite invariant measure is proportional to π.

Proof ***FIX*** **QED**

2.1.3 Recurrence and Lyapunov functions

By Proposition 2.2, the divergence (respectively convergence) of the series $\sum_{k\geq 1} P^k(x,x)$ is a criterion for the recurrence (transience) of the point x, but such a criterion may be difficult to verify in practice. We discuss here other criteria based on Lyapounov functions, a tool that will play a key role in the next chapters.

Given $C \subset M$, we let

$$\tau_C = \tau_C^{(1)} := \inf\{n \geq 1 \, : X_n \in C\},$$

and

$$\tau_C^{(k+1)} := \inf\{n > \tau_C^{(k)} \ : X_n \in C\}$$

for all $k \geq 1$. We also set $\tau_C^{(0)} := 0$. The next proposition shows that, whenever P is irreducible, recurrence (respectively positive recurrence) of the chain equates recurrence (positive recurrence) of any finite subset.

Proposition 2.13 *Suppose P is irreducible and let $C \subset M$ be a nonempty finite set such that for all $x \in C$, $\mathbb{P}_x(\tau_C < \infty) = 1$ (respectively $\mathbb{E}_x(\tau_C) < \infty$). Then P is recurrent (respectively positive recurrent).*

Proof Let $x \in C$. Then, since $\mathbb{P}_y(\tau_C < \infty) = 1$ for all $y \in C$, the strong Markov property implies that (X_n) visits C infinitely often $\mathbb{P}_x$-almost surely. Since C is finite, it follows that $\mathbb{P}_x$-almost surely, there is $y \in C$ such that $N_y = \infty$. If P was transient, we would have by Proposition 2.2 that $\mathbb{P}_x(\bigcup_{y \in C}\{N_y = \infty\}) = 0$, a contradiction. Hence P is recurrent.

Suppose now that $K := \max_{x \in C} \mathbb{E}_x(\tau_C) < \infty$. Let Q be the Markov kernel on C defined by $Q(x,y) := \mathbb{P}_x(X_{\tau_C} = y)$ for $x, y \in C$. Since C is finite, Q admits an invariant probability measure π (see Remark 2.8). Thus, if X_0 has law π, then X_{τ_C} has also law π. It follows (by a proof similar to the proof of Theorem 2.6 (**iii**), or by Exercise 4.24) that the measure μ defined by

$$\mu f = \frac{\mathbb{E}_\pi(\sum_{k=0}^{\tau_C - 1} f(X_k))}{\mathbb{E}_\pi(\tau_C)}$$

is invariant for P. Note here that $\mathbb{E}_\pi(\tau_C) \leq K < \infty$. This proves positive recurrence. **QED**

Exercise 2.14 Suppose P is irreducible, $C \subset M$ is finite and for all $x \in M \setminus C$, $\mathbb{P}_x(\tau_C < \infty) = 1$. Show that P is recurrent. *Hint:* If $M \setminus C \neq \emptyset$, prove that for all $x \in C$, $\mathbb{P}_x(\tau_{M \setminus C} < \infty) = 1$ and then use Proposition 2.13.

The next result extends and generalizes Proposition 2.13.

Proposition 2.15 *Suppose P is irreducible and let $C \subset M$ be a finite set.*

(i) *Assume that for some $\lambda_0 > 0$ and all $x \in C$, $\mathbb{E}_x(e^{\lambda_0 \tau_C}) < \infty$. Then there exists $\lambda \in (0, \lambda_0]$ such that for all $x, y \in C$,*

$$\mathbb{E}_x(e^{\lambda \tau_y}) < \infty.$$

(ii) *Let $p \geq 1$ and suppose that for all $x \in C$, $\mathbb{E}_x(\tau_C^p) < \infty$. Then for all $x, y \in C$,*

$$\mathbb{E}_x(\tau_y^p) < \infty.$$

Proof (**i**). First assume that $M = C$. In this case there exists, by irreducibility, some $\varepsilon > 0$ such that for all $x, y \in M$ and $k := \mathsf{card}(M)$, $\mathbb{P}_x(\tau_y > k) \leq 1 - \varepsilon$. Therefore, by the Markov property and induction on $n \geq 1$,

$$\mathbb{P}_x(\tau_y > nk) = \mathbb{E}_x(\mathbf{1}_{\tau_y > (n-1)k})\mathbb{P}_{X_{(n-1)k}}(\tau_y > k)) \leq (1-\varepsilon)^n.$$

Thus, for all $n \geq 0$,

$$\mathbb{P}_x(\tau_y > n) \leq \mathbb{P}_x(\tau_y > k[\frac{n}{k}]) \leq (1-\varepsilon)^{\frac{n}{k}-1},$$

where $[\frac{n}{k}]$ is the largest integer less than or equal to $\frac{k}{n}$. Hence, for $\alpha > 0$ so small that $e^{k\alpha}(1-\varepsilon) < 1$,

$$\mathbb{E}_x(e^{\alpha\tau_y}) \leq \sum_{n=1}^{\infty} e^{\alpha n}\mathbb{P}_x(\tau_y \geq n) < \infty.$$

We now turn to the proof of the first statement in full generality. Let

$$Y_n = X_{\tau_C^{(n)}}.$$

Such a definition makes sense because, by recurrence, $\tau_C^{(n)} < \infty$ almost surely. For all $y \in C$, set $\sigma_y := \inf\{n \geq 1 : \; Y_n = y\}$. For $x \in C, (Y_n)$ is a C-valued Markov chain on the probability space $(M^{\mathbb{N}}, \mathcal{B}(M^{\mathbb{N}}), \mathbb{P}_x)$, with respect to the filtration $\{\mathcal{F}_{\tau_C^{(n)}}\}_n$, and with Markov kernel $Q(a,b) := \mathbb{P}_a(X_{\tau_C} = b)$ introduced in the proof of Proposition 2.13. Thus, by what precedes,

$$\max_{x,y \in C} \mathbb{E}_x(e^{\alpha\sigma_y}) < \infty \tag{2.2}$$

for some $\alpha > 0$.

By assumption, $\max_{x \in C} \mathbb{E}_x(e^{\lambda_0 \tau_C}) \leq e^{\alpha_0}$ for some $\alpha_0 \geq 0$. By Jensen's inequality, for all $t \in [0,1], \mathbb{E}_x(e^{t\lambda_0\tau_C}) \leq \mathbb{E}_x(e^{\lambda_0\tau_C})^t \leq e^{t\alpha_0}$. Choose $\lambda \in (0, \frac{\lambda_0}{2}]$ so small that $2\lambda\alpha_0 \leq \lambda_0\alpha$. Then

$$\max_{x \in C} \mathbb{E}_x(e^{2\lambda\tau_C}) \leq e^{\alpha}.$$

Set $M_n := e^{(2\lambda\tau_C^{(n)} - n\alpha)}$. The previous inequality combined with the strong Markov property shows that (M_n) is a supermartingale under $\mathbb{P}_x$ with respect to the filtration $\{\mathcal{F}_{\tau_C^{(n)}}\}_n$. Therefore, using Theorem A.4 on optional stopping, $(M_{n\wedge\sigma_y})$ is again a supermartingale, and in particular $\mathbb{E}_x(M_{n\wedge\sigma_y}) \leq \mathbb{E}_x(M_0) = 1$. Together with Hölder's inequality, this yields for all $x, y \in C$

$$\mathbb{E}_x(e^{\lambda\tau_C^{(n\wedge\sigma_y)}}) \leq \mathbb{E}_x(M_{n\wedge\sigma_y})^{1/2}\mathbb{E}_x(e^{\alpha(n\wedge\sigma_y)})^{1/2} \leq \mathbb{E}_x(e^{\alpha\sigma_y})^{1/2} < \infty.$$

Thus,

$$\mathbb{E}_x(e^{\lambda\tau_y}) = \mathbb{E}_x(e^{\lambda\tau_C^{(\sigma_y)}}) < \infty$$

for all $x, y \in C$.

(**ii**). By the assumption and the strong Markov property, there exists $K \geq 0$ such that for every $n \geq 0$,

$$\mathbb{E}_x(|\tau_C^{(n+1)} - \tau_C^{(n)}|^p|\mathcal{F}_{\tau_C^{(n)}}) = \mathbb{E}_{Y_n}(\tau_C^p) \leq K^p.$$

Therefore, with $\|\cdot\|_p = \mathbb{E}_x(|\cdot|^p)^{1/p}$,

$$\|\tau_y\|_p = \|\tau_C^{(\sigma_y)}\|_p = \|\sum_{i\geq 0}(\tau_C^{(i+1)} - \tau_C^{(i)})\mathbf{1}_{i<\sigma_y}\|_p \leq \sum_{i\geq 0}\|(\tau_C^{(i+1)} - \tau_C^{(i)})\mathbf{1}_{i<\sigma_y}\|_p.$$

Now

$$\mathbb{E}_x(|\tau_C^{(i+1)} - \tau_C^{(i)}|^p\mathbf{1}_{i<\sigma_y}) = \mathbb{E}_x(\mathbb{E}_x(|\tau_C^{(i+1)} - \tau_C^{(i)}|^p|\mathcal{F}_{\tau_C^{(i)}})\mathbf{1}_{i<\sigma_y}) \leq K^p\mathbb{P}_x(\sigma_y > i).$$

Thus

$$\|\tau_y\|_p \leq K\sum_{i\geq 0}\mathbb{P}_x(\sigma_y > i)^{1/p} < \infty,$$

because, as seen in the beginning of the proof, the law of σ_y has a geometric tail. **QED**

A consequence of Proposition 2.15 is a short proof of the following classical result due to Chung [9].

Corollary 2.16 (Chung's theorem on moments of return times) *Suppose P is irreducible and for some $u \in M$ and $p \geq 1$, $\mathbb{E}_u(\tau_u^p) < \infty$. Then for all $x, y \in M$, $\mathbb{E}_x(\tau_y^p) < \infty$.*

Proof Slightly modifying the proof of Theorem 2.6 (**iv**), one has for all $x, y \in M$ the implication $\mathbb{E}_x(\tau_x^p) < \infty \Rightarrow \mathbb{E}_y(\tau_x^p) < \infty$. Fix $y \in M$ and choose $C = \{u, y\}$. Then $\mathbb{E}_y(\tau_u^p) < \infty$, and by Proposition 2.15 (**ii**), $\mathbb{E}_u(\tau_y^p) < \infty$. Since $\tau_y \leq \tau_u + \tau_y \circ \theta_{\tau_u}$, we obtain by the strong Markov property $\mathbb{E}_y(\tau_y^p) \leq 2^{p-1}(\mathbb{E}_y(\tau_u^p) + \mathbb{E}_u(\tau_y^p)) < \infty$. Thus, for any $x \in M$, $\mathbb{E}_x(\tau_y^p) < \infty$. **QED**

Lyapunov functions

In brief, a *Lyapunov function* is a map $V : M \to [1, \infty)$ such that $PV - V \leq 0$ outside a certain subset $C \subset M$. Lyapunov functions are practical tools to ensure that the assumptions of Propositions 2.13 and 2.15 are satisfied.

A map $V : M \to \mathbb{R}_+$ is called *proper* if for every $R > 0$, the set $\{x \in M : V(x) \leq R\}$ is finite. If M is finite, every map $V : M \to \mathbb{R}_+$ is proper. If M is countably infinite and $(x_n)_{n\geq 1}$ is any enumeration of the elements of M, $V : M \to \mathbb{R}_+$ is proper if and only if $\lim_{n\to\infty} V(x_n) = \infty$.

Apart from the first assertion, the following result is a consequence of a more general result (Proposition 6.11) that will be proved later.

Theorem 2.17 *Let P be a Markov kernel, let $V : M \to [1, \infty)$ be a map, and let $C \subset M$ be nonempty. Consider the following conditions:*

(a) *P is irreducible, $PV - V \leq 0$ on $M \setminus C$ and V is proper;*

(b) *$PV - V \leq -1$ on $M \setminus C$ and $PV < \infty$ on C;*

(b') *Condition* (**b**) *and in addition*

$$\sup_{x\in M} \mathbb{E}_x(|V(X_1) - V(x)|^p) < \infty$$

for some $p \geq 1$;

(c) *$PV - V \leq -\lambda V$ on $M \setminus C$ for some $\lambda \in (0, 1)$ and $PV < \infty$ on C.*

Then, for all $x \in M$,

(i) *Under Condition* (**a**),

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

(ii) *Under Condition* (**b**),

$$\mathbb{E}_x(\tau_C) \leq PV(x) + 1;$$

(iii) *Under Condition* ($\mathbf{b}'$),

$$\mathbb{E}_x(\tau_C^p) \le c(1 + V(x)^p)$$

for some constant $c > 0$ that depends on p but doesn't depend on x;

(iv) *Under Condition* ($\mathbf{c}$),

$$\mathbb{E}_x(e^{\lambda \tau_C}) \le \mathbb{E}_x(e^{-\log(1-\lambda)\tau_C}) \le \frac{1}{1-\lambda} PV(x).$$

In particular, if P is irreducible and if C is finite, Conditions ($\mathbf{a}$), ($\mathbf{b}$), ($\mathbf{b}'$), ($\mathbf{c}$) *respectively ensure recurrence of P, positive recurrence of P, p-th moments for the hitting times τ_y under $\mathbb{P}_x$ for every $y \in C$ and $x \in M$, and exponential moments for τ_y under $\mathbb{P}_x$ for every $x, y \in C$.*

Proof We only prove the first assertion. The other three follow from Proposition 6.11 to be proved later. When P is irreducible and when C is finite, recurrence, positive recurrence, p-th moments, and exponential moments of hitting times are direct consequences of Propositions 2.13 and 2.15. (To obtain $\mathbb{E}_x(\tau_y^p) < \infty$ for $x \in M \setminus C$, consider the finite set $\tilde{C} := C \cup \{x\}$ and note that $\tau_{\tilde{C}} \le \tau_C$.)

By irreducibility, the chain is either recurrent or transient. If it is recurrent, $\mathbb{P}_x(\tau_C < \infty) = 1$ for every $x \in M$ by Proposition 2.2. Suppose the chain is transient. For $x \in M \setminus C$, the sequence $V_n := V(X_{n \wedge \tau_C})$ is under $\mathbb{P}_x$ a supermartingale because $\mathbb{E}_x(V_{n+1} - V_n | \mathcal{F}_n) = (PV(X_n) - V(X_n))\mathbf{1}_{\tau_c > n} \le 0$. Thus, being nonnegative, (V_n) converges $\mathbb{P}_x$-almost surely to some random variable V_∞ taking values in $[0, \infty)$ (apply Theorem A.6 to the submartingale $(-V_n)$). This shows that $V(X_n)$ converges $\mathbb{P}_x$-almost surely on $\{\tau_C = \infty\}$. On the other hand, by transience (Proposition 2.2 (**iii**)) and by the assumption that V is proper, $\limsup_{n\to\infty} V(X_n) = \infty$ $\mathbb{P}_x$-almost surely, and therefore $\mathbb{P}_x(\tau_C < \infty) = 1$. And for $x \in C$, we have by the Markov property

$$\mathbb{P}_x(\tau_C < \infty) = \mathbb{P}_x(X_1 \in C) + \mathbb{E}_x(\mathbf{1}_{X_1 \in M \setminus C} \mathbb{P}_{X_1}(\tau_C < \infty)) = 1.$$

QED

Exercise 2.18 Suppose $V : M \to [1, \infty)$ is a proper map. Show that Condition (**c**) in Theorem 2.17 for a nonempty finite set C is equivalent to the existence of constants $0 \leq \rho < 1$ and $\kappa \geq 0$ such that

$$PV \leq \rho V + \kappa.$$

Show that under such a condition, every invariant probability measure π satisfies

$$\pi V \leq \frac{\kappa}{1-\rho} < \infty.$$

See Corollary 4.23 for a proof of the second assertion.

2.2 Convergence in distribution

2.2.1 Aperiodicity

We start with a general definition of aperiodicity. Let $R \subset \mathbb{N}^*$ be a (nonempty) set closed under addition. That is

$$i, j \in R \Rightarrow i + j \in R.$$

The *period* of R is defined as its greatest common divisor. If this period is 1, R is said to be *aperiodic*. Aperiodic sets enjoy the following useful property, that will be used repeatedly throughout the book.

Proposition 2.19 *Let R be aperiodic. Then there exists $n_0 \in \mathbb{N}$ such that $n_0 + \mathbb{N} = \{n \in \mathbb{N} : n \geq n_0\} \subset R$.*

Proof There exist, by aperiodicity, $a_1, \dots, a_l \in R$ whose greatest common divisor is 1. (To see this, take any element of R and call it a_1; then a_1 has a finite number of divisors strictly greater than 1, which we denote by $d_2, \dots, d_l$; for $2 \leq i \leq l$, pick a_i from $R(y)$ such that d_i does not divide a_i; such a_i exists because the greatest common divisor of R is 1). By Bézout's identity, there exist $q_1, \dots q_l \in \mathbb{Z}$ such that $\sum_i q_i a_i = -1$. Set $a := \sum_{i : q_i > 0} q_i a_i$. The set $R(y)$ being closed under addition, both a and $a + 1 = \sum_{i : q_i < 0} -q_i a_i$ lie in $R(y)$. Every $n \geq a^2$ can be written as $n = ka + r = (k - r)a + r(a + 1)$ for some $r \in \{0, \dots, a - 1\}$ and $k \geq a$. Thus, every $n \geq a^2$ is an element of R. **QED**

We now turn to the definition of aperiodicity for a countable Markov chain. Given a kernel P on M and $x \in M$, let $R(x) := \{k \geq 1 : x \leadsto^k x\}$ be the set of possible return times to x. The *period* of x, $per(x)$, is defined as the period of $R(x)$ and x is called *aperiodic* whenever $R(x)$ is. The kernel (or the chain) is said to be aperiodic if all points $x \in M$ are aperiodic.

Proposition 2.20 *Suppose P is irreducible. Then*

(i) *All points $x \in M$ have the same period;*

(ii) *P is aperiodic if and only if for all $x, y \in M$ there exists $n(x, y) \in \mathbb{N}$ such that $x \leadsto^n y$ for all $n \geq n(x, y)$.*

Proof (**i**). Let $x, y \in M$. By irreducibility, there exist $i, j \in \mathbb{N}^*$ such that $x \leadsto^i y$ and $y \leadsto^j x$. Thus $i + j \in R(x)$ and for all $k \in R(y), i + j + k \in R(x)$. Therefore, $per(x)$ divides $i + j$ and $i + j + k$, hence k, for all $k \in R(y)$. Thus $per(x) \leq per(y)$ and by symmetry $per(x) = per(y)$.

(**ii**). The "if" part is obvious. We prove the "only if" part. Given $y \in M$, there exists, by Proposition 2.19, $n_0 \in \mathbb{N}$ such that $n \in R(y)$ for all $n \geq n_0$. If now x is another point in M, $x \leadsto^i y$ for some i by irreducibility, hence $x \leadsto^n y$ for all $n \geq n_0 + i$. **QED**

An immediate useful consequence of Proposition 2.20 is the next result. Given two Markov kernels P and $\tilde{P}$ respectively defined on the countable state space M and $\tilde{M}$, we let $P \otimes \tilde{P}$ denote the Markov kernel on $M \times \tilde{M}$ corresponding to two independent chains with kernels $P, \tilde{P}$. That is

$$(P \otimes \tilde{P})((x, x'); (y, y')) := P(x, y)\tilde{P}(x', y').$$

Corollary 2.21 *If P and $\tilde{P}$ are both irreducible and aperiodic, so is $P \otimes \tilde{P}$. If in addition P and $\tilde{P}$ are positive recurrent, so is $P \otimes \tilde{P}$.*

Proof Note that $(P \otimes \tilde{P})^n = P^n \otimes \tilde{P}^n$ for every $n \in \mathbb{N}^*$. Thus, irreducibility (and aperiodicity) of $P \otimes \tilde{P}$ follows from Proposition 2.20 (**ii**), applied to P and $\tilde{P}$c. Also, if π and $\tilde{\pi}$ are invariant probability measures for P and $\tilde{P}$, so is $\pi \otimes \tilde{\pi}$ (defined as $((\pi \otimes \tilde{\pi})(x, x') := \pi(x)\tilde{\pi}(x'))$ for $P \otimes \tilde{P}$. By Theorem 2.6, this proves positive recurrence. **QED**

Exercise 2.22 Give an example of an irreducible and positive recurrent kernel P such that $P \otimes P$ is not irreducible, and an example of an irreducible recurrent kernel P such that $P \otimes P$ is irreducible and transient.

Exercise 2.23 Show that if $P \otimes \tilde{P}$ is irreducible, then both P and $\tilde{P}$ are irreducible. Also show that if $P \otimes \tilde{P}$ is irreducible and recurrent, then both P and $\tilde{P}$ are recurrent.

Exercise 2.24 Consider the Markov chain $(X_n)_{n\geq 0}$ from Exercise 2.1.

(i) Find the period of the chain.

(ii) Show that the chain is positive recurrent and find its unique invariant probability measure.

2.2.2 The convergence theorem

We now state and prove the main result of this section, the convergence theorem for irreducible aperiodic Markov chains.

Theorem 2.25 *Suppose P is irreducible and aperiodic. Let μ be a probability measure on M.*

(i) *If P is positive recurrent with invariant probability measure π, then*

$$\lim_{n\to\infty} \sup_{z\in M} |\mu P^n(z) - \pi(z)| = 0.$$

(ii) *If P is not positive recurrent, then for all $z \in M$,*

$$\lim_{n\to\infty} \mu P^n(z) = 0.$$

Proof Let $(X_n, Y_n)_{n\in\mathbb{N}}$ be the canonical chain on $(M\times M)^{\mathbb{N}}$ (i.e. $(X_n, Y_n)(\omega, \tilde{\omega}) := (\omega_n, \tilde{\omega}_n)$), and let

$$\tau_\Delta := \inf\{n \geq 1 \ : (X_n, Y_n) \in \Delta\},$$

where $\Delta := \{(x, x) : \ x \in M\}$ is the diagonal of M. Throughout the proof, we write $\mathbb{P}_\alpha$ for the Markov measure on $(M \times M)^{\mathbb{N}}$ with kernel $P \otimes P$ and initial distribution α. By Corollary 2.21, $P \otimes P$ is irreducible, hence either recurrent or transient.

Case 1: $P \otimes P$ is recurrent. For all $x, y, z \in M$,

$$\begin{aligned}
\mathbb{P}_{x,y}(X_n = z) &= \mathbb{P}_{x,y}(X_n = z; \tau_\Delta > n) + \mathbb{P}_{x,y}(X_n = z; \tau_\Delta \leq n) \\
&= \mathbb{P}_{x,y}(X_n = z; \tau_\Delta > n) + \mathbb{P}_{x,y}(Y_n = z; \tau_\Delta \leq n) \\
&\leq \mathbb{P}_{x,y}(\tau_\Delta > n) + \mathbb{P}_{x,y}(Y_n = z),
\end{aligned}$$

where the second equality follows from the strong Markov property and the fact that $X_{\tau_\Delta} = Y_{\tau_\Delta}$. Interchanging the roles of X_n and Y_n, one also has

$$\mathbb{P}_{x,y}(Y_n = z) \le \mathbb{P}_{x,y}(\tau_\Delta > n) + \mathbb{P}_{x,y}(X_n = z).$$

Hence

$$|P^n(x,z) - P^n(y,z)| = |\mathbb{P}_{x,y}(X_n = z) - \mathbb{P}_{x,y}(Y_n = z)| \le \mathbb{P}_{x,y}(\tau_\Delta > n),$$

and by integration

$$|\mu P^n(z) - \nu P^n(z)| \le \mathbb{P}_{\mu\otimes\nu}(\tau_\Delta > n) \tag{2.3}$$

for every probability measure ν on M and every $z \in M$. By recurrence of $P \otimes P$ (and Proposition 2.2 (**iii**)), one has for every $x, y \in M$ that $\mathbb{P}_{x,y}(\tau_\Delta > n) \to 0$ as $n \to \infty$. Thus

$$\lim_{n\to\infty} \sup_{z\in M} |\mu P^n(z) - \nu P^n(z)| = 0 \tag{2.4}$$

by dominated convergence. In light of Exercise 2.23, there are two subcases: P is either positive recurrent or null recurrent. If P is positive recurrent, (2.4) applied to $\nu = \pi$, the invariant probability measure of P, proves part (**i**) of the theorem. If P is null recurrent, let π be an unbounded invariant measure of P (see Theorem 2.12). For any nonempty finite set $A \subset M$, set $\pi_A(x) := \frac{\pi(x)\mathbf{1}_A(x)}{\pi(A)}$. Then, $\pi_A \le \frac{\pi}{\pi(A)}$, whence

$$\pi_A P^n(z) \le \frac{\pi P^n(z)}{\pi(A)} = \frac{\pi(z)}{\pi(A)}.$$

Therefore, by (2.4) applied to $\nu = \pi_A$,

$$\limsup_{n\to\infty} \mu P^n(z) \le \lim_{n\to\infty} |\mu P^n(z) - \pi_A P^n(z)| + \frac{\pi(z)}{\pi(A)} = \frac{\pi(z)}{\pi(A)}.$$

Letting $A \uparrow M$ proves (**ii**) in this case because $\pi(M) = \infty$.

Case 2: $P \otimes P$ is transient. By Proposition 2.2 (**i**),

$$[P^n(z,z)]^2 = (P\otimes P)^n((z,z);(z,z)) \to 0$$

as $n \to \infty$, for all $z \in M$. By irreducibility of P, this implies that $P^n(x,z) \to 0$ for all $x, z \in M$. Thus $\mu P^n(z) \to 0$ by dominated convergence. This proves (**ii**) in Case 2. **QED**

As shown below, the convergence in Theorem 2.25 is exponential if there exists a proper map that satisfies Condition (**c**) of Theorem 2.17 for a nonempty finite set C (see also Exercise 2.18).

Theorem 2.26 *Suppose P is irreducible and aperiodic, and that there exists a proper map $V : M \to [1, \infty)$ and constant $0 \le \rho < 1, \kappa \ge 0$ such that*

$$PV \le \rho V + \kappa.$$

Then P is positive recurrent and, denoting by π its invariant probability measure,

(i) *One has $\pi V \le \frac{\kappa}{1-\rho} < \infty$;*

(ii) *There exist constants $0 \le \gamma < 1$ and $c \ge 0$ such that for every probability measure μ on M,*

$$\sup_{z \in M} |\mu P^n(z) - \pi(z)| \le c\gamma^n(\mu V + 1), \quad \forall n \in \mathbb{N}.$$

Corollary 2.27 *Suppose M is finite and P irreducible and aperiodic, with invariant probability measure π. Then there exist constants $0 \le \gamma < 1$ and $c \ge 0$ such that for every probability measure μ on M,*

$$\sup_{z \in M} |\mu P^n(z) - \pi(z)| \le c\gamma^n, \quad \forall n \in \mathbb{N}.$$

Proof Take $V \equiv 1$ in Theorem 2.26. **QED**

Proof [of Theorem 2.26]. We use the same notation, $P \otimes P, (X_n, Y_n), \Delta$, etc., as in the proof of Theorem 2.25.

Positive recurrence follows from Exercise 2.18 and Theorem 2.17. Assertion (**i**) follows from Exercise 2.18. By Inequality (2.3) from the proof of Theorem 2.25, it suffices to derive an exponential upper bound on $\mathbb{P}_{\mu\otimes\pi}(\tau_\Delta > n)$ in order to prove Assertion (**ii**). Pick $x^* \in M$ and choose $\varepsilon > 0$ small enough so that $V(x^*) \le \frac{\kappa}{\varepsilon}$ and $\rho + \varepsilon < 1$. Set $W(x, y) := V(x) + V(y)$, $x, y \in M$. Then

$$(P \otimes P)W(x, y) = PV(x) + PV(y) \le \rho W(x, y) + 2\kappa,$$

so that $(P \otimes P)W \le (\rho + \varepsilon)W$ on the complement of the set

$$C := \{(x, y) : \; W(x, y) \le \frac{2\kappa}{\varepsilon}\}.$$

By Theorem 2.17 (**iv**) and Assertion (**i**), we then obtain, for some positive constant c depending on κ, ρ and ε,

$$\mathbb{E}_{\mu\otimes\pi}(e^{(1-\rho-\varepsilon)\tau_C}) \le \frac{(\mu \otimes \pi)(P \otimes P)W}{\rho + \varepsilon} \le \frac{\rho(\mu V + \pi V) + 2\kappa}{\rho + \varepsilon} \le c(1 + \mu V).$$

Since V is proper, the set C is finite, and Proposition 2.15 (**i**) together with $(x^*, x^*) \in C$ yield the existence of $\lambda > 0$ such that

$$\max_{(x,y)\in C} \mathbb{E}_{(x,y)}(e^{\lambda\tau_{(x^*,x^*)}}) < \infty.$$

Thus

$$\begin{aligned}\mathbb{P}_{\mu\otimes\pi}(\tau_\Delta > n) \leq & \mathbb{P}_{\mu\otimes\pi}(\tau_{(x^*,x^*)} > n)\\ \leq & \mathbb{P}_{\mu\otimes\pi}(\tau_C > n/2) + \mathbb{E}_{\mu\otimes\pi}(\mathbb{P}_{(X_{\tau_C},Y_{\tau_C})}(\tau_{(x^*,x^*)} > n/2))\\ \leq & ce^{-\lambda n/2}(1+\mu V)\end{aligned}$$

for some other constant c. Inequality (2.3) concludes the proof. **QED**

2.3 Application to renewal theory

Let $(\Delta_i)_{i\geq 1}$ be a sequence of i.i.d. random variables living on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and taking values in $\mathbb{N}$. Let Δ_0 be another $\mathbb{N}$-valued random variable on $(\Omega, \mathcal{F}, \mathsf{P})$, independent of $(\Delta_i)_{i\geq 1}$ but having a possibly different distribution. Set

$$T_n := \Delta_0 + \Delta_1 + \ldots + \Delta_n.$$

The sequence $T := (T_n)_{n\in\mathbb{N}}$ is called a *renewal process*; $T_0 = \Delta_0$ is the *delay* of the process, and $\{T_n, n \geq 0\}$ is the set of *renewal times*. Observe that T is a Markov chain with respect to the filtration $\mathcal{F}_n := \sigma(\Delta_0, \ldots, \Delta_n)$, whose transition matrix has entries $A(i,j) := \mathsf{P}(\Delta_1 = j - i)$.

Let

$$p_k := \mathsf{P}(\Delta_1 = k)$$

for $k \in \mathbb{N}$. We say that T is *aperiodic* if $p_0 \neq 1$ and $\{k \geq 1 : p_k > 0\}$ is an aperiodic set as defined in Section 2.2.1. We say that T is L^p if Δ_1 is in L^p, i.e. $\sum_{k\in\mathbb{N}} k^p p_k < \infty$.

To fix ideas, one can imagine that a certain device breaks down and is replaced by a generic device at times $T_0, T_1, \ldots$. The lifespan of the initial device is distributed as Δ_0 and the lifespan of the replacement devices are distributed as Δ_1.

From now on we shall assume that T is aperiodic. For all $n \in \mathbb{N}$, let

$$\varsigma_n := \min\{k \geq 0 \ : T_k \geq n\}.$$

Then $\varsigma_n < \infty$ P-almost surely so that

$$X_n := T_{\varsigma_n} - n$$

is well-defined. A key observation is the following:

The set of renewal times for T equals the zero set of (X_n).

That is

$$\{T_n : n \in \mathbb{N}\} = \{n \in \mathbb{N} : X_n = 0\}.$$

It is easily checked that with respect to the filtration $\{\mathcal{F}_{\varsigma_n}\}$, (X_n) is a Markov chain on $\mathbb{N}$ whose transition matrix is given by

$$P(k, k-1) = 1 \text{ for } k \geq 1,$$

$$P(0, k) = \frac{p_{k+1}}{1 - p_0} \text{ for } k \in \mathbb{N},$$

and

$$P(k, l) = 0 \text{ for } k \geq 1, l \neq k - 1.$$

Let $K := \sup\{k \geq 1 : p_k > 0\} \in \mathbb{N}^* \cup \{\infty\}$ and $M := \{0, \ldots, K-1\}$ (with the convention that $M = \mathbb{N}$ if $K = \infty$). Then $X_n \in M$ for n large enough (precisely $n \geq (X_0 - K + 1)^+$). On M, the chain (X_n) is irreducible, recurrent, and aperiodic (by aperiodicity of T).

Exercise 2.28 Verify the claims made about (X_n). In particular, show that (X_n) is a Markov chain with the transition matrix given above, and that (X_n) restricted to M is irreducible, recurrent, and aperiodic.

On $\mathbb{N}^{\mathbb{N}}$, let $\mathbb{P}_0$ be the Markov measure with initial distribution δ_0 and kernel P. The corresponding expectation is denoted by $\mathbb{E}_0$. On $\mathbb{N}^{\mathbb{N}}$, we define

$$\tau_0(\mathbf{x}) := \inf\{n \geq 1 : x_n = 0\}.$$

Then

$$\mathbb{E}_0(\tau_0) = \sum_{k \geq 0} (1 + k) P(0, k) = \frac{\mathsf{E}(\Delta_1)}{1 - p_0} = \mathsf{E}(\Delta_1 | \Delta_1 > 0) \in (0, \infty],$$

where the expectation of a random variable X conditional on an event A of positive probability is defined as $\mathsf{E}(X|A) := \mathsf{E}(X \mathbf{1}_A)/\mathsf{P}(A)$. The equation $\mathbb{E}_0(\tau_0) = \mathsf{E}(\Delta_1)/(1 - p_0)$ implies that (X_n) is positive recurrent if and only if T is L^1.

Exercise 2.29 Assume that (X_n), restricted to M, is positive recurrent. Express the unique invariant probability measure for the transition matrix P in terms of the p_k's.

As a consequence of Theorem 2.25, we obtain the following classical renewal theorem.

Theorem 2.30 *Assume that T is aperiodic. Then*

$$\lim_{k\to\infty} \sum_{n=0}^{\infty} \mathsf{P}(T_n = k) = \frac{1}{\mathsf{E}(\Delta_1)},$$

with the convention that the right-hand side is zero if $\mathsf{E}(\Delta_1) = \infty$.

Proof Let $N_k := \sum_{n\geq 0} \mathbf{1}_{\{T_n=k\}}$. Then

$$N_k = \mathbf{1}_{\{X_k=0\}}(1 + \sum_{i\geq 1} \mathbf{1}_{\{T_i'=0\}}),$$

where

$$T_i' := \Delta_{\varsigma_k+1} + \ldots + \Delta_{\varsigma_k+i}.$$

Thus $\mathsf{E}(N_k) = \mathsf{E}(\mathsf{E}(N_k|\mathcal{F}_{\varsigma_k})) = \mathsf{P}(X_k = 0)\frac{1}{1-p_0}$, and by Theorems 2.25 and 2.6,

$$\lim_{k\to\infty} \mathsf{P}(X_k = 0) = \frac{1}{\mathbb{E}_0(\tau_0)}.$$

This proves the result. **QED**

2.3.1 Coupling of renewal processes

Suppose that T is L^1, and let $\tilde{T}$ be another aperiodic L^1 renewal process independent of T with

$$\tilde{T}_n = \tilde{\Delta}_0 + \tilde{\Delta}_1 + \ldots + \tilde{\Delta}_n.$$

The distribution of $(\tilde{\Delta}_i)_{i\geq 0}$ may be different from the one of $(\Delta_i)_{i\geq 0}$. We are interested in the first time $\tau > 0$ that is a renewal time for both T and $\tilde{T}$. Equivalently, with $\tilde{X}_n$ defined in analogy to X_n,

$$\tau := \inf\{n \geq 1 :\ X_n = \tilde{X}_n = 0\}.$$

We know that (X_n) is absorbed by M in finite time and that it is aperiodic and positive recurrent on M. Hence, $(X_n, \tilde{X}_n)$ is absorbed by $M \times \tilde{M}$ in finite time ($\tilde{M}$ defined in analogy to M) and, by Corollary 2.21, it is positive recurrent on $M \times \tilde{M}$. In particular,

$$\mathsf{P}(\tau < \infty) = \mathbb{P}_{\alpha\otimes\tilde{\alpha}}(\tau_{0,0} < \infty) = 1, \tag{2.5}$$

where α (respectively $\tilde{\alpha}$) denotes the law of Δ_0 (respectively $\tilde{\Delta}_0$), and where

$$\tau_{0,0}(\mathbf{x}, \mathbf{y}) := \inf\{n \geq 1 : \ x_n = y_n = 0\}, \quad \mathbf{x}, \mathbf{y} \in \mathbb{N}^{\mathbb{N}}.$$

In turns out that whenever $\Delta_0, \tilde{\Delta}_0$ and $\Delta_1, \tilde{\Delta}_1$ are in L^p for some $p \geq 1$, the same is true for τ. A proof of this fact can be found for instance in Lindvall's book [29] and goes back to Pitman's seminal paper [34]. We provide here a short proof (different from Lindvall's) based on Proposition 2.15 and Theorem 2.17.

Theorem 2.31 *Suppose T and $\tilde{T}$ are aperiodic and in L^p for some $p \geq 1$. Then there exists a constant $c > 0$, independent of the distributions of Δ_0 and $\tilde{\Delta}_0$, such that $\mathsf{E}(\tau^p) \leq c(1 + \mathsf{E}(\Delta_0^p) + \mathsf{E}(\tilde{\Delta}_0^p))$.*

Proof Let $Q := P \otimes \tilde{P}$ denote the kernel of $(X_n, \tilde{X}_n)$. Let V be the function defined on $\mathbb{N} \times \mathbb{N}$ by $V(i,j) = 2\max(i,j) + 1$. One has

$$QV(i,j) - V(i,j) = -2 \text{ for } i \neq 0, j \neq 0,$$

and (by integrability of Δ_1 and dominated convergence)

$$\lim_{j\to\infty} QV(0,j) - V(0,j) = \lim_{j\to\infty} \mathsf{E}(\max(2\Delta_1 - 2j - 2, -2) | \Delta_1 > 0) = -2.$$

Similarly, $\lim_{i\to\infty} QV(i,0) - V(i,0) = -2$. Condition (**b**) of Theorem 2.17 is then satisfied for the Markov process $(X_n, \tilde{X}_n)$ on $\mathbb{N} \times \mathbb{N}$, with $C = \{(i,j) \in \mathbb{N} \times \mathbb{N} : \ V \leq R\}$ and R large enough. Condition (**b**$'$) is easily seen to be satisfied as well because Δ_1 and $\tilde{\Delta}_1$ are in L^p. Therefore, there is $c > 0$ such that for all $(i,j) \in \mathbb{N} \times \mathbb{N}$,

$$\mathbb{E}_{i,j}(\tau_{0,0}^p) \leq 2^{p-1}(\mathbb{E}_{i,j}(\tau_C^p) + \max_{(i,j)\in C} \mathbb{E}_{i,j}(\tau_{0,0}^p)) \leq c(1 + \max(i,j)^p). \tag{2.6}$$

Here, the first inequality follows from the strong Markov property and inequality $\tau_{0,0} \leq \tau_C + \tau_{0,0} \circ \Theta_{\tau_C}$. The last inequality follows from Theorem 2.17

(**iii**) and from the proof of Proposition 2.15. Note that while $(X, \tilde{X})$ is not necessarily irreducible on $\mathbb{N} \times \mathbb{N}$ and thus a key assumption of Proposition 2.15 is not satisfied, the proof still goes through because any point $(i, j) \in \mathbb{N} \times \mathbb{N}$ leads to $(0, 0)$. Integrating the inequality in (2.6) with respect to $\alpha \otimes \tilde{\alpha}$, the law of $(\Delta_0, \tilde{\Delta}_0) = (X_0, \tilde{X}_0)$, gives the result. **QED**

Theorem 2.32 *Suppose T and $\tilde{T}$ are aperiodic and*

$$\mathsf{E}(e^{\lambda_0 \Delta_1}) + \mathsf{E}(e^{\lambda_0 \tilde{\Delta}_1}) < \infty$$

for some $\lambda_0 > 0$. Then there exist $0 < \lambda \leq \lambda_0$ and $c > 0$ such that

$$\mathsf{E}(e^{\lambda \tau}) \leq c(1 + \mathsf{E}(e^{\lambda_0 \Delta_0}) + \mathsf{E}(e^{\lambda_0 \tilde{\Delta}_0})).$$

Proof The proof is similar to the proof of Theorem 2.31. Set $V(i, j) := e^{\lambda_0 i} + e^{\lambda_0 j}$. Then $QV(i, j) \leq e^{-\lambda_0} V(i, j) + \kappa$ with

$$\kappa := \mathsf{E}(e^{\lambda_0 \Delta_1} | \Delta_1 > 0) + \mathsf{E}(e^{\lambda_0 \tilde{\Delta}_1} | \tilde{\Delta}_1 > 0).$$

Condition (**c**) of Theorem 2.17 is then satisfied for any $0 < \lambda < 1 - e^{-\lambda_0}$ and $C = \{(i, j) \in \mathbb{N} \times \mathbb{N} : \ V(i, j) \leq R\}$ with R sufficiently large given the choice of λ (see also Exercise 2.18). Then, relying on $\tau_{0,0} \leq \tau_C + \tau_{0,0} \circ \Theta_{\tau_C}$, the strong Markov property, Theorem 2.17 (**iv**), and the proof of Proposition 2.15, we obtain

$$\mathbb{E}_{i,j}(e^{\lambda \tau_{0,0}}) \leq c(1 + V(i, j)), \quad \forall (i, j) \in \mathbb{N} \times \mathbb{N}$$

for some $c > 0$ and some $\lambda \in (0, 1 - e^{-\lambda_0})$. Integrating this inequality with respect to the law of $(\Delta_0, \tilde{\Delta}_0)$ gives the desired result. **QED**

2.4 Convergence rates for positive recurrent chains

We revisit here the ergodic theorems from Subsection 2.2, Theorems 2.25 and 2.26, with the help of Theorems 2.31 and 2.32.

Let M be countable and let $(X_n, Y_n)_{n \in \mathbb{N}}$ be the canonical chain on $(M \times M)^{\mathbb{N}}$. Let P be an irreducible, aperiodic, and positive recurrent kernel on M. If π denotes the invariant probability measure of P, we have seen in the proofs of Theorems 2.25 and 2.26 that for every probability measure μ on M and every $x^* \in M$,

$$\sup_{x\in M} |\mu P^n(x) - \pi(x)| \le \mathbb{P}_{\mu\otimes\pi}(\tau_{(x^*,x^*)} > n),$$

where $\mathbb{P}_{\mu\otimes\pi}$ is the Markov measure with kernel $P \otimes P$ and initial distribution $\mu \otimes \pi$, and where $\tau_{(x^*,x^*)} = \inf\{n \ge 1 : \ X_n = Y_n = x^*\}$.

Let $(\tau_{x^*}^{(n)})$ (respectively $(\tilde\tau_{x^*}^{(n)})$) denote the successive hitting times of x^* by (X_n) (respectively (Y_n)). Then, for any probability measures α, β on M, the processes $T := (\tau_{x^*}^{(n+1)})_{n\ge 0}$ and $\tilde T := (\tilde\tau_{x^*}^{(n+1)})_{n\ge 0}$ living on the probability space $((M\times M)^{\mathbb{N}}, \mathcal{B}((M\times M)^{\mathbb{N}}), \mathbb{P}_{\alpha\otimes\beta})$ are two independent renewal processes and $\tau_{(x^*,x^*)}$ is nothing but the first common renewal time for T and $\tilde T$.

The Markov inequality, Theorems 2.31, 2.32, and Proposition 2.11 lead to the following result.

Theorem 2.33 *Let P be irreducible, aperiodic, and positive recurrent, with invariant probability measure π. Let $x^* \in M$.*

(i) *If $\mathbb{E}_{x^*}(\tau_{x^*}^p) < \infty$ for some $p \ge 2$, then there exists $c \ge 0$ such that for every probability measure μ on M and for every $n \in \mathbb{N}^*$,*

$$\sup_{x\in M} |\mu P^n(x) - \pi(x)| \le \frac{1}{n^{p-1}} c(1 + E_\mu(\tau_{x^*}^{p-1})).$$

(ii) *If $\mathbb{E}_{x^*}(e^{\lambda_0 \tau_{x^*}}) < \infty$ for some $\lambda_0 > 0$, then there exist $0 < \lambda < \lambda_0$ and $c \ge 0$ such that for every probability measure μ on M and for every $n \in \mathbb{N}$,*

$$\sup_{x\in M} |\mu P^n(x) - \pi(x)| \le e^{-\lambda n} c(1 + \mathbb{E}_\mu(e^{\lambda_0 \tau_{x^*}})).$$

Combined with Theorem 2.17, Proposition 2.15, and the strong Markov property, we recover and extend Theorem 2.26.

Corollary 2.34 *Let P be irreducible, aperiodic, and positive recurrent, with invariant probability measure π. Let $V : M \to [1, \infty)$ and let $C \subset M$ be as in Theorem 2.17 ((**b**$'$) or (**c**)) with C finite. Then*

(i) *Under Condition (**b**$'$) of Theorem 2.17 for $p \ge 2$, there is $c \ge 0$ such that for every probability measure μ on M and for every $n \in \mathbb{N}^*$,*

$$\sup_{x\in M} |\mu P^n(x) - \pi(x)| \le \frac{1}{n^{p-1}} c(1 + \mu V^p).$$

(ii) *Under Condition* (**c**) *of Theorem 2.17, there are* $c, \lambda > 0$ *such that for every probability measure* μ *on* M *and for every* $n \in \mathbb{N}$,

$$\sup_{x \in M} |\mu P^n(x) - \pi(x)| \leq e^{-\lambda n} c(1 + \mu V).$$

Notes

The book by Aldous and Fill [1] contains numerous interesting identities for the mean hitting times $(\mathbb{E}_x(\tau_y))$, the occupation times $(\mathbb{E}_x(N_y))$ and their relation to the rate of convergence. Convergence rates for finite Markov chains, in term of the geometry of the chain, are thoroughly investigated in the monograph by Saloff-Coste [37] and the book by Levin, Peres and Wilmer [28]. A nice extension of Chung's theorem (Proposition 2.16) can be found in the recent paper [2]. , *** Pitman, Lindvall etc ****

Chapter 3

Random Dynamical Systems

Numerous examples of Markov chains in the applied literature are given by *random dynamical systems* (also called *random iterative systems*). These are defined as follows.

Let $(\Theta, \mathcal{A}, m)$ be a probability space,

$$F : \Theta \times M \to M$$

$$(\theta, x) \mapsto F_\theta(x),$$

a measurable map, and $(\theta_n)_{n\geq 1}$ a sequence of independent identically distributed (i.i.d.) $\Theta-$valued random variables having law m. Consider an $M-$valued process recursively defined by

$$X_{n+1} := F_{\theta_{n+1}}(X_n) \tag{3.1}$$

for some given random variable X_0.

Proposition 3.1 *Assume that X_0 is a random variable independent of (θ_n). Then (X_n) is a Markov chain on M whose Markov kernel is given by*

$$P(x, G) = m(\theta \in \Theta : F_\theta(x) \in G). \tag{3.2}$$

If furthermore F_θ is continuous for m-almost every θ, then P is Feller.

Proof The proof follows (almost) directly from the definitions. Measurability of $x \mapsto P(x, G)$ is a by-product of Fubini's theorem since $P(x, G) = \int_\Theta \mathbf{1}_G \circ F_\theta(x) m(d\theta)$. The Feller property follows by continuity under the integral. **QED**

Exercise 3.2 [additive noise] Suppose $M = \mathbb{R}^n$ (or an abelian locally compact group), $\Theta = M$, $F : M \to M$,

$$F_\theta(x) = F(x) + \theta$$

and $m(d\theta) = h(\theta)d\theta$ with $h \in L^1(d\theta)$. Here $d\theta$ stands for the Lebesgue measure (or the Haar measure) on M. Let P denote the corresponding Markov kernel given by (3.2).

Given $x \in M$, let $U_x : L^1(dx) \to L^1(dx)$ be the translation operator defined as $U_x(g)(y) := g(y-x)$. Show that for all $f \in B(M)$,

$$|Pf(x) - Pf(y)| \leq \|f\|_\infty \|U_{F(x)}(h) - U_{F(y)}(h)\|_1.$$

Deduce that P is strong Feller whenever F is continuous. One can use (or better, prove) that for all $g \in L^1(dx), x \in M \mapsto U_x(g) \in L^1(dx)$ is continuous.

The kernel P defined by (3.2) is called the *Markov kernel induced by* (F, m). The sequence of random maps (F^n) defined by

$$F^n := F_{\theta_n} \circ F_{\theta_{n-1}} \circ \ldots \circ F_{\theta_1}$$

is called the *random dynamical system induced by* (F, m).

Note that, by Chapman-Kolmogorov, the law of $F^n(x)$ is determined by P ($F^n(x)$ has law $P^n(x, \cdot)$) but, as shown by the next example, P is not sufficient to characterize the law of F^n.

Example 3.3 This example is due to Kifer [27]. Let $M = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle, $\Theta = [0,1]$, and $m(d\theta) = d\theta$ the uniform Lebesgue measure. Let $f : S^1 \to S^1$ be any, say continuous, map and $F_\theta(z) = e^{2i\pi\theta} f(z)$. Then $P(z, \cdot)$ is the uniform measure on S^1 for every $z \in S^1$, but the random dynamical system (RDS) induced by (F, m) clearly depends on the choice of f. For instance, if $f(z) = z$, F^n preserves the distance between points, while for $f(z) = z^2$, F^n locally increases the distance exponentially.

Example 3.4 This example is due to Diaconis and Freedman [12]. Let $M = [0,1]$ be the closed unit interval, and

$$P(x, dy) = \frac{1}{2x} \mathbf{1}_{[0,x]}(y)dy + \frac{1}{2(1-x)} \mathbf{1}_{[x,1]}(y)dy.$$

Here we adopt the convenient convention that $\frac{\mathbf{1}_{[0,x]}(y)}{x} dy = \delta_0(dy)$ for $x = 0$ and $\frac{\mathbf{1}_{[x,1]}(y)}{(1-x)} dy = \delta_1(dy)$ for $x = 1$. In words, if the chain is at x it moves to a point y randomly chosen in the right interval $[x, 1]$ (respectively left interval $[0, x]$) with probability $1/2$.

Let $F : (0, 1) \times [0, 1] \to [0, 1]$ be defined by

$$F_\theta(x) := 2\theta x \mathbf{1}_{\theta<1/2} + [x + (2\theta - 1)(1 - x)]\mathbf{1}_{\theta\geq 1/2}.$$

Then P is induced by (F, dx).

3.1 Representation of Markov chains by RDS

Proposition 3.1 shows that every RDS defines a Markov chain. Here we briefly discuss the converse problem and consider the question of representing a Markov chain by a suitable RDS.

A *transformation space* is a set of maps $f : M \to M$ closed under composition. Let $\mathbf{T}$ be a transformation space and P a Markov kernel on M.

We say that P can be represented by $\mathbf{T}$ if there exists a probability space $(\Theta, \mathcal{A}, m)$ and a measurable map $F : \Theta \times M \to M$ such that

(i) $F_\theta \in \mathbf{T}$ for all $\theta \in \Theta$;

(ii) P is induced by (F, m).

Recall that a separable metric space M is called *Polish* if it is complete. The following result is folklore.

Theorem 3.5 *If M is a Borel subset of a Polish space, then any Markov kernel on M can be represented by a space $\mathbf{T}$ of measurable maps with $(\Theta, \mathcal{A}, m) = ((0, 1), \mathcal{B}((0, 1)), \lambda)$ and λ the Lebesgue measure on $(0, 1)$.*

Proof When M is a Borel subset of $\mathbb{R}$, the proof is constructive and makes F explicit. Indeed, let G_x be the cumulative distribution function of $P(x, .)$, i.e.

$$G_x(t) = P(x, (-\infty, t]).$$

For all $\theta \in (0, 1)$ and $x \in M$, set

$$F_\theta(x) := G_x^{-1}(\theta),$$

where $G_x^{-1} : (0, 1) \to \mathbb{R}$, the generalized inverse of G_x, is defined as

$$G_x^{-1}(u) := \inf\{t \in \mathbb{R} : \; G_x(t) \geq u\}.$$

Then

$$\lambda(\theta \in (0,1) : \; F_\theta(x) \leq t) = \lambda(\theta \in (0,1) : \; \theta \leq G_x(t)) = G_x(t).$$

The proof in the general case follows from the following abstract result of measure theory: Every Borel subset M of a Polish space is *isomorphic* to a Borel subset of $[0,1]$. That is, there exists a Borel set $\tilde{M} \subset [0,1]$ and a bi-measurable bijection $\Psi : M \to \tilde{M}$ (meaning that both Ψ and its inverse are Borel measurable). Chapter 13 of Dudley's book [13] contains a detailed proof of this result. Exercise 4.9 treats the particular case where M is compact or locally compact.

Given such a Ψ and a Markov kernel P on M, let $\tilde{P}$ be the Markov kernel on $\tilde{M}$ defined as $\tilde{P}(x, A) := P(\Psi^{-1}(x), \Psi^{-1}(A))$. Then $\tilde{P}$ is induced by $(\tilde{F}, \lambda)$ for some measurable $\tilde{F} : (0,1) \times \tilde{M} \to \tilde{M}$ so that P is induced by (F, λ) with $F_\theta(x) = \Psi^{-1}(\tilde{F}_\theta(\Psi(x)))$. **QED**

Blumenthal and Corson [8] prove the following result (see also Kifer [27], Theorem 1.2).

Theorem 3.6 *(Blumenthal and Corson, 1972). Let M be a connected and locally connected compact metric space. Let P be a Feller Markov kernel such that $P(x, \cdot)$ has full support for all $x \in M$, i.e. for all $x \in M$ and for every closed set F strictly contained in M, we have $P(x, F) < 1$. Then P may be represented by $\mathbf{T} = C^0(M, M)$ (the space of continuous maps $f : M \to M$).*

The question of representation by smooth maps has been considered by Quas [35]. Before stating Quas' theorem, we state a result due to Jürgen Moser from which it will be deduced.

Let M be a smooth (C^∞) compact orientable Riemannian manifold without boundary, with normalized Riemannian probability measure λ. If $\rho : M \to \mathbb{R}_+$ is a C^1 density on M and $\Phi : M \to M$ a C^1 diffeomorphism, we let $\Phi^*\rho$ denote the image of ρ by Φ. That is

$$(\Phi^*\rho)(\Phi(x)) = \frac{\rho(x)}{|J\Phi(x)|},$$

where $J\Phi(x)$ is the Jacobian of Φ, i.e. the determinant of the derivative $D\Phi(x) : T_xM \to T_{\Phi(x)}M$. In other words, if X is a random variable with density ρ, then $\Phi(X)$ is a random variable with density $\Phi^*\rho$.

In 1965, Moser [32], using the "homotopy trick" argument, proved part **(i)** of the following result in the C^∞ case. For every positive integer k and

$0 \leq \alpha < 1$, we let $C^{k+\alpha}(M)$ denote the space of $C^{k+\alpha}$ (C^k with α-Hölder kth derivatives if $\alpha > 0$) functions $h : M \to \mathbb{R}$ endowed with the $C^{k+\alpha}$ topology,

$$E_t^{k+\alpha} := \{h \in C^{k+\alpha}(M) : \int_M h(x)\lambda(dx) = t\},$$

and $D^{k+\alpha} := \{\rho \in E_1^{k+\alpha} : \rho(x) > 0 \; \forall x \in M\}$ the space of positive $C^{k+\alpha}$ densities. Plainly, $E_1^{k+\alpha}$ is a closed subset of $C^{k+\alpha}(M)$ which can be identified with the Banach space $E_0^{k+\alpha}$, and $D^{k+\alpha}$ is an open subset of $E_1^{k+\alpha}$.

Theorem 3.7 *(Moser, 1965). Let ρ_0 be a positive C^k density for some $k \geq 1$. Then*

(i) *For any positive C^k density ρ, there exists a C^k diffeomorphism Φ_ρ on M with the property that*

$$\Phi_\rho^* \rho_0 = \rho;$$

(ii) *The C^k diffeomorphisms Φ_ρ from part* **(i)** *can be chosen in such a way that the mapping*

$$D^k \times M \to M,$$
$$(\rho, x) \mapsto \Phi_\rho(x)$$

is C^k.

Proof Let $\rho_t = \rho_0 + t(\rho - \rho_0)$ for $0 \leq t \leq 1$. We look for a family of diffeomorphisms $(\Phi_t)_{t\in[0,1]}$ such that $\Phi_t^*\rho_0 = \rho_t$ for all $t \in [0, 1]$. That is

$$j(t, x)\rho_t(\Phi_t(x)) = \rho_0(x), \tag{3.3}$$

where $j(t, x)$ is the Jacobian of Φ_t, evaluated at x. More precisely, we look for a family of vector fields $\{X_t\}_{t\in[0,1]}$ on M such that $\Phi_t(x)$ is the solution to the non-autonomous Cauchy problem

$$\frac{dy}{dt} = X_t(y)$$

with initial condition $y(0) = x$. Using Jacobi's formula for the derivative of the determinant of a matrix-valued function, one obtains that $j(t, x)$ solves

$$\frac{dj}{dt} = div(X_t)[\Phi_t(x)]j(t)$$

with initial condition $j(0,\cdot) \equiv 1$. Thus, taking the time derivative of (3.3) and setting $y := \Phi_t(x)$, $\eta := \rho_0 - \rho$ gives

$$div(X_t)(y)\rho_t(y) - \eta(y) + < \nabla\rho_t(y), X_t(y) >_y = 0.$$

That is

$$div(\rho_t X_t)(y) = \eta(y).$$

If one sets $X_t = \nabla U/\rho_t$ the problem reduces to finding a function $U : M \to \mathbb{R}$ such that

$$\Delta U = div(\nabla U) = \eta, \tag{3.4}$$

where one should recall that $\eta = \rho_0 - \rho$.

Since

$$\int_M \eta(x)\lambda(dx) = 0,$$

(3.4) admits a solution, and we may define $\Delta^{-1}\eta$ as the particular solution

$$x \mapsto 2\int_0^\infty Q_t\eta(x)dt,$$

where $Q_t\eta(x) := E(\eta(W_t)|W_0 = x)$ and W_t a Brownian motion on M. Furthermore, by Schauder's estimates (see e.g. Chapter 6 in [20]) Δ^{-1} maps continuously $E_0^{k-1+\alpha}(M)$ into $C^{k+1+\alpha}(M)$ for every positive integer k and $0 < \alpha < 1$. This makes the vector field

$$X_t^\rho := \nabla U/\rho_t$$

a C^k vector field. It also implies that the continuous mapping

$$[0,1] \times D^k \times M \to TM,$$

$$(t,\rho,x) \mapsto X_t^\rho(x)$$

is C^k.

Let $t \mapsto \Phi_t(\rho,x)$ denote the solution to the Cauchy problem $\frac{dy}{dt} = X_t^\rho(y)$ with initial condition $\Phi_0(\rho,x) = x$. It then follows, from standard results on differential equations that $x \mapsto \Phi_t(\rho,x)$ is a C^k diffeormorphism for all $(t,\rho) \in [0,1] \times D^k$, and that $(x,\rho) \mapsto \Phi_t(\rho,x)$ is C^k for all $t \in [0,1]$. To conclude the proof, set $\Phi_\rho(x) := \Phi_1(\rho,x)$. **QED**

From Moser's theorem we deduce the following result proved by Quas [35] in the C^∞ case.

Corollary 3.8 *(Quas, 1991). Let P be a Markov kernel on M, a smooth compact orientable connected Riemannian manifold without boundary. Assume that for each $x \in M$, P_x has a $C^k, k \geq 1$, positive density ρ_x with respect to the Riemannian measure, and that $x \in M \mapsto \rho_x \in D^k$ is $C^r, r \geq 0$. Then P may be represented by $\mathbf{T} = C^r(M, M)$.*

Proof Let $\rho_0 = \rho_{x_0}$ for some $x_0 \in M$ and let $\Psi_x = \Phi_{\rho_x}$ denote the C^k diffeormorphism produced by Moser's Theorem (Theorem 3.7). Then

$$P(x, G) = P(x_0, \Psi_x^{-1}(G)).$$

Let $\mathbf{T} = C^r(M, M)$ and $f_y \in \mathbf{T}$ be defined by $f_y(x) := \Psi_x(y)$. Then

$$P(x, G) = m(f \in \mathbf{T} : f(x) \in G),$$

where m is the image of P_{x_0} by the mapping $y \in M \mapsto f_y \in \mathbf{T}$. **QED**

While $x \mapsto \rho_x$ is C^r, the map $\rho \mapsto \Phi_\rho(y)$ is C^k, and k may be less than r — so why is f_y in C^r?

Chapter 4

Invariant and Ergodic Probability Measures

4.1 Weak convergence of probability measures

Let $\mathcal{P}(M)$ denote the set of probability measures on $(M, \mathcal{B}(M))$. A sequence $\{\mu_n\} \subset \mathcal{P}(M)$ is said to *converge weakly* to $\mu \in \mathcal{P}(M)$, written

$$\mu_n \Rightarrow \mu,$$

provided

$$\lim_{n\to\infty} \mu_n f = \mu f$$

for all $f \in C_b(M)$. The following theorem, known as Portmanteau theorem, gives equivalent conditions to weak convergence. Note that this theorem is true in any metric space (without assumption of separability or completeness).

Let $U_b(M) \subset C_b(M)$ (resp. $L_b(M) \subset U_b(M)$) denote the set of bounded and uniformly continuous (resp. bounded and Lipschitz) mappings $f : M \to \mathbb{R}$.

Theorem 4.1 (Portmanteau theorem) *Let $\{\mu_n\} \subset \mathcal{P}(M)$ and $\mu \in \mathcal{P}(M)$. The following conditions are equivalent:*

(a) $\mu_n \Rightarrow \mu$;

(b) $\mu_n f \to \mu f$ *for all* $f \in U_b(M)$;

(c) $\mu_n f \to \mu f$ *for all* $f \in L_b(M)$;

(d) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for all closed sets* $F \subset M$;

(e) $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$ *for all open sets* $O \subset M$;

(f) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all* $A \in \mathcal{B}(M)$ *such that* $\mu(\partial A) = 0$, *where* $\partial A := \overline{A} \setminus int(A)$ *denotes the boundary of* A.

Proof $(a) \Rightarrow (b) \Rightarrow (c)$ is clear and $(d) \Leftrightarrow (e)$ holds by set complementation. Assume (c). Let F be a closed set, $\varepsilon > 0$, and $f_\varepsilon(x) := (1 - \frac{d(x,F)}{\varepsilon})^+$, where $d(x,F) := \inf_{y\in F} d(x,y)$. Then $1 \geq f_\varepsilon \geq \mathbf{1}_F$ and $f_\varepsilon \in L_b(M)$. Thus, $\limsup \mu_n(F) \leq \limsup \mu_n f_\varepsilon = \mu f_\varepsilon$ and, by dominated convergence, $\mu f_\varepsilon \to \mu(F)$ as $\varepsilon \to 0$. This proves that $(c) \Rightarrow (d)$. Assume (d). Let $A \in \mathcal{B}(M)$ be such that $\mu(\partial A) = 0$. Let F be the closure of A and O its interior. Then $\mu(F) = \mu(O)$ and, by (d) and (e), $\liminf \mu_n(A) \geq \liminf \mu_n(O) \geq \mu(O)$ and $\limsup \mu_n(A) \leq \limsup \mu_n(F) \leq \mu(F)$. This proves that $(d),(e) \Rightarrow (f)$. It remains to show that $(f) \Rightarrow (a)$. Assume (f) and let $f \in C_b(M)$. Replacing f by $f + c$ for some $c > 0$ if necessary, we can assume that $f \geq 0$. For all $a \geq 0$, the set $\{f > a\}$ is open and its boundary is contained in $\{f = a\}$. Furthermore, the set of $a \geq 0$ such that $\mu(\{f = a\}) > 0$ is at most countable (as the set of discontinuity points of the cumulative distribution function $a \mapsto \mu(\{f \leq a\})$). Thus, by Fubini's theorem, (f), and dominated convergence, $\mu_n f = \int_0^{\|f\|_\infty} \mu_n(f > a)da \to \int_0^{\|f\|_\infty} \mu(f > a)da = \mu f$.
QED

The following corollary is often useful.

Corollary 4.2 *Let* $f \in B(M)$ *and let* D_f *denote the set of discontinuities of* f. *If* $\mu_n \Rightarrow \mu$ *and* $\mu(D_f) = 0$, *then* $\mu_n f \to \mu f$.

Proof Let $\mu_n^f := \mu_n(f^{-1}(\cdot))$ be the image measure of μ_n by f. It suffices to show that $\mu_n^f \Rightarrow \mu^f$. Indeed, let $g(t) := t$ for $|t| \leq \|f\|_\infty$, and $g(t) := \text{sign}(t)\|f\|_\infty$ for $|t| > \|f\|_\infty$. Then $\mu_n^f g = \mu_n f$ and $\mu^f g = \mu f$. To prove that $\mu_n^f \Rightarrow \mu^f$, we rely on assertion (d) of the Portmanteau Theorem. Let F be a closed subset of $\mathbb{R}$. Then $\limsup \mu_n^f(F) \leq \limsup \mu_n(\overline{f^{-1}(F)}) \leq \mu(\overline{f^{-1}(F)})$ because $\mu_n \Rightarrow \mu$. Now, $\overline{f^{-1}(F)} \subset D_f \cup f^{-1}(F)$ so that $\mu(\overline{f^{-1}(F)}) = \mu(f^{-1}(F)) = \mu^f(F)$. **QED**

Exercise 4.3 For $\varepsilon, \delta > 0$ let $A_{\varepsilon,\delta}$ be the set of $x \in M$ such that $|f(y) - f(z)| \geq \varepsilon$ for some $y, z \in B(x,\delta)$. Show that $D_f = \cup_{n\in\mathbb{N}^*} \cap_{m\in\mathbb{N}^*} A_{1/n,1/m}$ and that D_f is measurable (even if f is not).

Exercise 4.4 Let P be a Markov kernel on a metric space M. Show that P is Feller if and only if the map $\varphi : M \to \mathcal{P}(M),\ x \mapsto P(x, \cdot)$ is continuous (where $\mathcal{P}(M)$ is equipped with the topology of weak convergence).

The space $\mathcal{P}(M)$ equipped with the topology of weak convergence is actually a metric space, as shown by the next proposition.

Proposition 4.5 *There exists a countable family $\{f_n\}_{n\geq 0} \subset C_b(M)$ such that*

$$D(\mu, \nu) := \sum_{n\geq 0} \frac{1}{2^n} \min(|\mu f_n - \nu f_n|, 1)$$

is a distance on $\mathcal{P}(M)$ whose induced topology is the topology of weak convergence. That is

$$\mu_n \Rightarrow \mu \Leftrightarrow D(\mu_n, \mu) \to 0.$$

Remark 4.6 Unless when M is compact, the family $\{f_n\}_{n\geq 0}$ is not dense in $C_b(M)$ (see Exercise 4.8).

Proof If M is compact, $C_b(M)$ is separable (see Exercise 4.7) and it suffices to choose a dense sequence $\{f_n\} \subset C_b(M)$. If M is not compact, $C_b(M)$ is no longer separable (see Exercise 4.8), but we shall prove that there exists a metric $\tilde{d}$ on M, topologically equivalent to d, making M homeomorphic to a subset of a compact metric space. It will then follow that $U_b(M, \tilde{d})$, the space of bounded uniformly continuous functions on $(M, \tilde{d})$, is separable. (Here one should recall that two topologically equivalent metrics may yield distinct sets of uniformly continuous functions.)

Replacing d by $\frac{d}{1+d}$ (which remains a distance on M inducing the same topology as d), we can assume that $d \leq 1$. Let $\{a_n\}_{n\geq 0} \subset M$ be countable and dense, and let $H : M \to [0,1]^{\mathbb{N}}$ be the map defined by

$$H(x) := (d(x, a_n))_{n\geq 0}.$$

By Tychonoff's Theorem (see, e.g., Theorem 2.2.8 in [13]), $[0,1]^{\mathbb{N}}$ is a compact metric space. A metric for $[0,1]^{\mathbb{N}}$ is given by

$$e(\mathbf{x}, \mathbf{y}) = \sum_{k\geq 0} \frac{|x_k - y_k|}{2^k},$$

where $\mathbf{x} = (x_k)_{k\geq 0}, \mathbf{y} = (y_k)_{k\geq 0}$. Set

$$\tilde{d}(x, y) := e(H(x), H(y)).$$

It is not hard to check that $\tilde{d}$ is a metric on M inducing the same topology as d. The spaces $(M, \tilde{d})$ and $(H(M), e)$ are thus isometric. Let $K := \overline{H(M)}$. Then K is compact (as a closed subset of a compact space) and thus, there exists a countable and dense family $\{g_n\} \subset C_b(K)$. Let $f \in U_b(M, \tilde{d})$. Since H is an isometry, the map $f \circ H^{-1} : H(M) \to \mathbb{R}$ is uniformly continuous. It then extends to a continuous map $\hat{f} : \overline{H(M)} \to \mathbb{R}$. By density of $\{g_n\}$, there exists, for all $\varepsilon > 0$, some n such that

$$\|f - g_n \circ H\|_\infty = \sup_{\mathbf{x} \in H(M)} |f \circ H^{-1}(\mathbf{x}) - g_n(\mathbf{x})| \le \sup_{\mathbf{x} \in K} |\hat{f}(\mathbf{x}) - g_n(\mathbf{x})| \le \varepsilon.$$

This proves that the sequence $\{f_n\}$, with $f_n := g_n \circ H$, is dense in $U_b(M, \tilde{d})$. Now, by Theorem 4.1 (b) and density of $\{f_k\}$, $\mu_n \Rightarrow \mu$ if and only if $\mu_n f_k \to \mu f_k$ for all $k \in \mathbb{N}$. This is equivalent to $D(\mu_n, \mu) \to 0$. **QED**

Exercise 4.7 Let K be a compact metric space (and thus also a Polish space). Using the proof of Proposition 4.5, show that K is homeomorphic to a compact subset of $[0,1]^{\mathbb{N}}$, equipped with the metric e. We now identify K with a subset of $[0,1]^{\mathbb{N}}$. Let P be the set of real-valued functions on $[0,1]^{\mathbb{N}}$ of the form $p(\mathbf{x}) = q(x_0, \dots, x_n)$, where $q : [0,1]^{n+1} \to \mathbb{R}$ is a polynomial in $(n+1)$ variables with rational coefficients. Use the Stone-Weierstrass theorem to show that $P|_K = \{p|_K : p \in P\}$ is dense in $C(K)$. *This shows that $C(K)$ is separable. Since $C_b(K)$ is a subset of the separable metric space $C(K)$, it is itself separable.*

Exercise 4.8 Let $\mathcal{X}$ be a topological space. Suppose that there exists an uncountable family $\{O_\alpha\}$ of open sets such that $O_\alpha \cap O_\beta = \emptyset$ for $\alpha \neq \beta$. Show that $\mathcal{X}$ is not separable. Show that $C_b(\mathbb{R})$, the set of continuous bounded functions on $\mathbb{R}$, is not separable. *Hint:* Let $f \in C_b(\mathbb{R})$ be such that $f(n) = 0$ and $f = 1$ on $[n+1/(n+1), n+1-1/(n+1)]$ for all $n \in \mathbb{N}^*$. Set $f_x(t) := f(x+t)$ and consider the family $\{O_x\}_{x \in (0,1)}$, where $O_x := \{g \in C_b(\mathbb{R}) : \|f_x - g\|_\infty < 1/2\}$.

Exercise 4.9 [Borel Isomorphism] We say that two measurable spaces X and Y are isomorphic if there exists a bi-measurable bijection $\Psi : X \to Y$, meaning that both Ψ and Ψ^{-1} are measurable. It turns out that every Borel subset M of a Polish space is isomorphic to a Borel subset of $[0,1]$ (see Remark 4.10). The purpose of this exercise is to prove this result when M is compact or locally compact and separable.

(i) Let $\{0,1\}^{\mathbb{N}^*}$ be equipped with the product topology and Borel σ-field. Show that $\{0,1\}^{\mathbb{N}^*}$ is a metric space with the metric d defined as

$$d(\omega, \alpha) := \sum_{i \geq 1} \frac{|\omega_i - \alpha_i|}{2^i}.$$

(ii) Show that the map

$$\Psi : \{0,1\}^{\mathbb{N}^*} \to [0,1],$$
$$\omega \mapsto \sum_{i \geq 1} \frac{\omega_i}{2^i}$$

is 1-Lipschitz continuous.

(iii) Let $\tilde{I} \subset \{0,1\}^{\mathbb{N}^*}$ be the set of ω such that $\omega_i = 0$ for infinitely many i and $\omega_j = 1$ for infinitely many j. Show that $\tilde{I}$ is a Borel subset of $\{0,1\}^{\mathbb{N}^*}$ and that $\Psi|_{\tilde{I}}$ (Ψ restricted to $\tilde{I}$) is a homeomorphism onto $\Psi(\tilde{I})$, i.e. a continuous bijection with continuous inverse.

(iv) Show that $[0,1]$ and $\{0,1\}^{\mathbb{N}^*}$ are isomorphic. *Hint:* Use (**iii**) and the fact that the complement of $\tilde{I}$ in $\{0,1\}^{\mathbb{N}^*}$ is countably infinite.

(v) Show that there is a homeomorphism between $\{0,1\}^{\mathbb{N}^*}$ and $\{0,1\}^{\mathbb{N}^* \times \mathbb{N}^*}$, equipped with the metric

$$e(A, B) := \sum_{j \geq 1} \frac{d((A_{i,j})_{i \geq 1}, (B_{i,j})_{i \geq 1})}{2^j}.$$

Then show that $[0,1]$ and $[0,1]^{\mathbb{N}^*}$ are isomorphic. Relying on the proof of Proposition 4.5, deduce that every compact (or locally compact separable) metric space is isomorphic to a Borel subset of $[0,1]$. *Hint:* Any locally compact separable metric space can be written as a countable union of compact sets, see, e.g., Theorem XI.6.3 in [15].

Remark 4.10 Theorem 13.1.1 in [13] implies the following: If M is a Borel subset of a Polish space, and if B is a Borel subset of $[0,1]$ whose cardinality equals the cardinality of M, then M and B are isomorphic. Since the cardinality of a Borel subset of a Polish space is either finite, countably infinite, or the cardinality of the continuum, every such set is in fact isomorphic to a large class of Borel subsets of $[0,1]$.

One of the main advantages of the distance defined in Proposition 4.5 is that it allows to verify weak convergence by testing the condition $\mu_n f \to \mu f$ over a countable set of functions.

Two other classical distances over $\mathcal{P}(M)$ are the following:

Prohorov metric For any $A \subset M$ and $\varepsilon > 0$, let

$$A^\varepsilon := \{y \in M : \ d(y, A) < \varepsilon\}.$$

For all $\mu, \nu \in \mathcal{P}(M)$ the *Prohorov* distance (also called the Lévy-Prohorov distance) between μ and ν is defined as

$$\pi(\mu, \nu) := \inf \{\varepsilon > 0 \ : \ \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(M)\} . \tag{4.1}$$

Fortet-Mourier metric Let $L_b(M) \subset C_b(M)$ be the space of bounded Lipschitz maps equipped with the norm

$$\|f\|_{bl} = \|f\|_\infty + Lip(f),$$

where

$$Lip(f) := \sup\{\frac{|f(x) - f(y)|}{d(x, y)} \ : \ (x, y) \in M^2, x \neq y\}.$$

For all $\mu, \nu \in \mathcal{P}(M)$ the *Fortet-Mourier* distance between μ and ν is defined as

$$\rho(\mu, \nu) := \sup\{|\mu f - \nu f| \ : f \in L_b(M), \|f\|_{bl} \leq 1\}. \tag{4.2}$$

A proof of the following result can be found in Dudley [13].

Theorem 4.11 *The maps π and ρ are distances on $\mathcal{P}(M)$. Let $\{\mu_n\} \subset \mathcal{P}(M)$ and $\mu \in \mathcal{P}(M)$. The following conditions are equivalent:*

(a) $\mu_n \Rightarrow \mu$;

(b) $\rho(\mu_n, \mu) \to 0$;

(c) $\pi(\mu_n, \mu) \to 0$.

Proof We only prove that (**a**) $\Leftrightarrow$ (**b**). For more details and the proof of (**b**) $\Leftrightarrow$ (**c**), see Dudley [13]. The implication (**b**) $\Rightarrow$ (**a**) follows from assertion

(c) of Theorem 4.1. Conversely assume (**a**). We first assume that M is complete. Fix $\varepsilon > 0$. By Ulam's theorem (or Prohorov's theorem, Theorem 4.13 below), one can choose $K \subset M$ compact such that

$$\mu(K) > 1 - \varepsilon \tag{4.3}$$

Let $K_\varepsilon = \{x \in M \ : d(x, K) < \varepsilon\}$. By assertion (e) of Theorem 4.1,

$$\mu_n(K_\varepsilon) > 1 - \varepsilon \tag{4.4}$$

for n sufficiently large. By the Ascoli theorem, the unit ball $L_{b,1} := \{f \in L_b : \ \|f\|_{bl} \leq 1\}$ restricted to K is a compact subset of $C_b(K)$. There exists then a finite set $\{f_1, \ldots, f_N\} \subset L_{b,1}$ such that for all $f \in L_{b,1}$ there is some $i \in \{1, \ldots, N\}$ such that $|f(x) - f_i(x)| \leq \varepsilon$ for all $x \in K$. Since f and f_i have a Lipschitz constant ≤ 1, we also get that

$$|f(x) - f_i(x)| \leq 3\varepsilon \tag{4.5}$$

for all $x \in K_\varepsilon$. Now

$$|\mu_n f - \mu f| \leq |(\mu_n - \mu) f_i| + |(\mu_n - \mu)((f - f_i)\mathbf{1}_{K_\varepsilon})| + |(\mu_n - \mu)((f - f_i)\mathbf{1}_{M\setminus K_\varepsilon})|$$

Thus, using inequalities (4.3), (4.4) and (4.5), we get

$$\rho(\mu_n, \mu) \leq \max_{1\leq i\leq N} |(\mu_n - \mu) f_i| + 8\varepsilon.$$

This proves (**b**). If M is not complete, we can replace it by its completion $\tilde{M}$. Any map $f \in L_b$ extends to a bounded Lipschitz map on $\tilde{M}$ and the measures (μ_n) and μ can be seen as measures on $\tilde{M}$ so that the proof goes through. **QED**

Remark 4.12 Theorem 4.1 is true in any (not necessarily separable) metric space. The equivalences in Theorem 4.11 require separability (but not completeness).

4.1.1 Tightness and Prohorov Theorem

A set $\mathcal{P} \subset \mathcal{P}(M)$ is said to be *tight* (sometimes called uniformly tight) if for every $\epsilon > 0$ there exists a compact set $K \subset M$ such that

$$\mu(K) \geq 1 - \varepsilon$$

for all $\mu \in \mathcal{P}$. Observe in particular that if M is compact, every subset of $\mathcal{P}(M)$ is tight. A set $\mathcal{P} \subset \mathcal{P}(M)$ is said to be *relatively compact* if it has compact closure in $\mathcal{P}(M)$ (equipped with one of the distances π, ρ or any other distance characterizing weak convergence). Finally, it is said to be *totally bounded* if for every $\varepsilon > 0$, there is a finite set $A \subset \mathcal{P}$ such that for every $\mu \in \mathcal{P}$ there is $\nu \in A$ with $d(\mu, \nu) < \varepsilon$. Here, d can be the Prohorov metric, the Fortet-Mourier metric, or any other metric on $\mathcal{P}(M)$ characterizing weak convergence.

The following theorem usually referred to as Prohorov's theorem asserts that tightness and relative compactness are equivalent in a Polish space (complete and separable). Here the assumption that M is a Polish space is crucial. For otherwise the implication (**b**) $\Rightarrow$ (**a**) might be false (see e.g. Billingsley [7] or Dudley [13, Chapter 11.5] for a proof).

Theorem 4.13 (Prohorov theorem) *Assume M is a Polish space (i.e a complete separable metric space). Then the following assertions are equivalent:*

(a) *$\mathcal{P}$ is tight;*

(b) *$\mathcal{P}$ is relatively compact;*

(c) *Every sequence $\{\mu_n\} \subset \mathcal{P}$ has a convergent subsequence $\mu_{n_k} \Rightarrow \mu \in \mathcal{P}(M)$;*

(d) *$\mathcal{P}$ is totally bounded for π or ρ.*

Remark 4.14 The latter property shows that $\mathcal{P}(M)$ is complete for ρ or π since every Cauchy sequence is totally bounded.

Tightness Criteria

We conclude this subsection with a simple practical Lyapunov-type condition ensuring tightness of a sequence of probabilities.

A measurable map $V : M \to \mathbb{R}$ is called *proper* if for all $R \in \mathbb{R}$ the set

$$\{V \leq R\} = \{x \in M : V(x) \leq R\}$$

has compact closure.

Proposition 4.15 *Let $V : M \to \mathbb{R}^+$ be a proper map and let $\{\mu_n\}$ be a sequence in $\mathcal{P}(M)$ such that*

$$\limsup_{n\to\infty} \mu_n V \le K < \infty.$$

Then $\{\mu_n\}$ is tight. Assume furthermore that V is continuous. Then

(i) *For every limit point μ of $\{\mu_n\}$, $\mu V \le K$;*

(ii) *Let $H : M \to \mathbb{R}$ be a continuous function such that $G = \frac{V}{1+|H|}$ is proper. If $\mu_n \Rightarrow \mu$ then $\mu_n H \to \mu H$.*

Proof Fix $\varepsilon > 0$ and let $R > 0$ be large enough so that $\limsup_{n\to\infty} \mu_n V \le \varepsilon R$. By the Markov inequality, $\limsup_{n\to\infty} \mu_n\{V > R\} \le \limsup_{n\to\infty} \frac{\mu_n V}{R} \le \varepsilon$. Let now $\mu = \lim \mu_{n_k}$ be a limit point of $\{\mu_n\}$. Then for all $R > 0$ $\mu(V \wedge A) = \lim_{k\to\infty} \mu_{n_k}(V \wedge A) \le K$. Thus $\mu V \le K$ by monotone convergence. We pass to the proof of the last statement. Let $G = \frac{V}{1+|H|}$. For all $R \in \mathbb{R} \setminus D_G$ with D_G at most countable, $\mu\{G = R\} = 0$ and, therefore,

$$\lim_{n\to\infty} \mu_n(H\mathbf{1}_{G\le R}) = \mu(H\mathbf{1}_{G\le R}).$$

On the other hand $\mu_n(|H|\mathbf{1}_{G>R}) \le \mu_n(\frac{V}{G}\mathbf{1}_{G>R}) \le \frac{1}{R}\mu_n(V)$. Thus

$$\lim_{R\to\infty} \limsup_{n\to\infty} \mu_n(|H|\mathbf{1}_{G>R}) = 0$$

and, similarly,

$$\lim_{R\to\infty} \mu(|H|\mathbf{1}_{G>R}) = 0.$$

This proves the result. **QED**

4.2 Invariant probability measures

Given a Markov kernel P, a measure (respectively a probability measure) μ is called P-*invariant* or simply *invariant* if

$$\mu P f = \mu f \tag{4.6}$$

for all $f \in B(M)$, where Pf is defined by (1.2). Equivalently,

$$\mu P = \mu,$$

where μP is defined by (1.3).

Exercise 4.16 Let $(\theta_n)_{n\geq 1}$ be an i.i.d. sequence of random variables with distribution m, where m is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For every $\theta \in \mathbb{R}$, let

$$F_\theta : \mathbb{R} \to \mathbb{R}, x \mapsto x + \theta.$$

Show that the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is an invariant measure for the Markov kernel induced by (F, m).

Remark 4.17 Let $\mathcal{C}$ denote a space of bounded, measurable mappings $f : M \to \mathbb{R}$, closed under multiplication and such that $\mathcal{B}(M) = \sigma(\mathcal{C})$ (the smallest σ-field making elements of $\mathcal{C}$ measurable). By a monotone class argument (see the appendix, Theorem A.1), it suffices to check (4.6) on $\mathcal{C}$ to prove $P-$invariance of $\mu \in \mathcal{P}(M)$.

For instance, one can choose $\mathcal{C}$ to be the set of bounded, continuous functions; or, if $C_b(M)$ is separable, a countable set of continuous functions dense in the set of bounded continuous functions and closed under multiplication.

We let $\mathsf{Inv}(P)$ denote the set of P-invariant probability measures. The set $\mathsf{Inv}(P)$ might be empty as shown by the following two examples.

Example 4.18 Let $M = [0, 1]$ and $f : M \to M$ be the map defined by $f(x) = x/2$ for $x \neq 0$ and $f(0) = 1$. Then, the (deterministic) chain $X_{n+1} = f(X_n)$ has no invariant probability measure. For otherwise the Poincaré recurrence theorem (see Theorem 4.41 below) would imply that such a measure is δ_0, but $f(0) = 1$.

Example 4.19 Consider the pair (F, m) introduced in Exercise 4.16. Let us assume in addition that $\int_{\mathbb{R}} |\theta| \; m(d\theta) < \infty$ and set $\alpha := \int_{\mathbb{R}} \theta \; m(d\theta)$. While the corresponding Markov kernel P has the Lebesgue measure as an invariant measure, P does not admit any invariant probability measures if $\alpha \neq 0$.

To see this, let μ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then there is $K > 0$ such that $\mu([-K, K]) > 0$. If μ was invariant for P, the Markov chain $(X_n)_{n\in\mathbb{N}}$ induced by (F, m) and with $X_0 \sim \mu$ would satisfy

$$0 < \mu([-K, K]) = \mu P^n([-K, K]) = \mathsf{P}(|X_n| \leq K), \quad \forall n \in \mathbb{N}^*.$$

But if $\alpha > 0$ ($\alpha < 0$), one has $\lim_{n\to\infty} X_n = \infty$ ($\lim_{n\to\infty} X_n = -\infty$) P-almost surely by the law of large numbers. Hence $\lim_{n\to\infty} \mathsf{P}(|X_n| \leq K) = 0$, a contradiction.

Given a Markov chain (X_n) on M, the associated family of *empirical occupation measures* is defined as

$$\nu_n := \frac{1}{n}\sum_{i=0}^{n-1}\delta_{X_i}, \quad n \in \mathbb{N}^*. \tag{4.7}$$

Notice that each ν_n is a random element of $\mathcal{P}(M)$.

A sufficient condition ensuring existence of invariant probability measures is given by the following classical theorem (see e.g. [14]).

Theorem 4.20 *Let (X_n) denote a Markov chain (defined on $(\Omega, \mathcal{A}, \mathbb{F}, \mathsf{P})$) on M with kernel P that is Feller. Then the following statements hold.*

(i) P*-almost surely, every limit point of the family of empirical occupation measures $(\nu_n)_{n\geq 1}$ is P-invariant;*

(ii) *If $(\nu_n)_{n\geq 1}$ is tight with positive P-probability, then $\mathsf{Inv}(P)$ is nonempty.*

Proof (**i**). Let $f \in B(M)$. Set $U_{n+1} := f(X_{n+1}) - Pf(X_n), M_0 := 0$ and $M_{n+1} := M_n + U_{n+1}$, for $n \geq 0$. Then (M_n) is an L^2 martingale, whose quadratic variation (see the section on martingale theory in the appendix) verifies

$$\langle M\rangle_{n+1} - \langle M\rangle_n = \mathsf{E}(U_{n+1}^2|\mathcal{F}_n) = Pf^2(X_n) - (Pf)^2(X_n) \leq 2||f||_\infty^2.$$

Hence by the strong law of large numbers for martingales (see Theorem A.8 in the appendix),

$$0 = \lim_{n\to\infty}\frac{M_n}{n} = \lim_{n\to\infty}\nu_n f - \nu_n(Pf) \tag{4.8}$$

almost surely. Let $\{f_k\} \subset C_b(M)$ be as in Proposition 4.5. Then, by the Feller property, Pf_k is in $C_b(M)$ for all k and, consequently, with probability one

$$\nu f_k - \nu(Pf_k) = 0$$

for every limit point ν of $\{\nu_n\}$ and every $k \in \mathbb{N}$. Thus, $\nu = \nu P$.

(**ii**). Let $\omega \in \Omega$ such that $(\nu_n(\omega))_{n\geq 1}$ is tight and all of its limit points are P-invariant. By Prohorov's theorem, $(\nu_n(\omega))_{n\geq 1}$ admits at least one limit point, so $\mathsf{Inv}(P)$ is nonempty. **QED**

Corollary 4.21 *If M is compact and P is Feller, $\mathsf{Inv}(P)$ is a nonempty compact convex subset of $\mathcal{P}(M)$. Convexity of $\mathsf{Inv}(P)$ holds for arbitrary metric spaces and Markov kernels.*

Tightness Criteria for Empirical Occupation Measures

When M is noncompact, the tightness of the empirical occupation measures (ν_n) can be ensured by the existence of a convenient Lyapunov function. This is a proper map $V : M \to \mathbb{R}_+$ such that $PV - V$ is "sufficiently" negative.

Corollary 4.22 *Let $V : M \to \mathbb{R}_+$ be a proper map. Assume that $PV \leq V$ and that $\mathsf{E}(V(X_0)) < \infty$. Then the family of empirical occupation measures (ν_n) is almost surely tight.*

Proof The sequence $\{V_n = V(X_n)\}$ being a nonnegative supermartingale with $\mathsf{E}(V_0) < \infty$, it converges almost surely to some finite random variable V_∞ (see Theorem A.6 in the appendix). This implies that $\nu_n V \to V_\infty$ almost surely and the result follows from Proposition 4.15. **QED**

Another result, in the same spirit, is

Corollary 4.23 *Let $V : M \to \mathbb{R}_+$ be a proper map. Assume that*

$$PV \leq \rho V + \kappa,$$

with $\kappa \geq 0, 0 \leq \rho < 1$, and $\mathsf{E}(V(X_0)) < \infty$. Then

$$\limsup_{n\to\infty} \nu_n \sqrt{V} \leq \frac{\sqrt{\kappa}}{1-\sqrt{\rho}}$$

almost surely. In particular, (ν_n) is tight. The set $\mathsf{Inv}(P)$ is a nonempty compact convex subset of $\mathcal{P}(M)$ and for all $\mu \in \mathsf{Inv}(P)$, $\mu V \leq \frac{\kappa}{1-\rho}$.

Proof Set $W = \sqrt{V}$. Then by Jensen's inequality, $PW \leq \sqrt{PV} \leq \sqrt{\rho V + \kappa} \leq \sqrt{\rho} W + \sqrt{\kappa}$. Set $LW(x) = PW(x) - W(x), M_0 = 0$ and

$$M_n = W(X_n) - W(X_0) - \sum_{k=0}^{n-1} LW(X_k)$$

for all $n \geq 1$. Then (M_n) is an L^2 martingale whose quadratic variation process is given as $\langle M \rangle_0 = 0$ and

$$\langle M \rangle_{n+1} - \langle M \rangle_n = \mathsf{E}((M_{n+1} - M_n)^2 | \mathcal{F}_n)) = PV(X_n) - (PW)^2(X_n) \leq PV(X_n)$$

for $n \geq 0$. Thus $\mathsf{E}(\langle M \rangle_n) \leq \sum_{i=0}^n \mathsf{E}(P^{i+1}V(X_0)) \leq n\frac{\kappa}{1-\rho} + \frac{\rho}{1-\rho}\mathsf{E}(V(X_0))$ where the last inequality easily follows from the the assumption on V. Then by the second strong law of large number for L^2 martingale (Theorem A.8 (iv) in the appendix) $\frac{M_n}{n} \to 0$ almost surely. Now, because $-LW \geq (1-\sqrt{\rho})W - \sqrt{\kappa}$,

$$(1-\sqrt{\rho})\nu_n W \leq \sqrt{\kappa} + \frac{M_n}{n} + \frac{W(X_0)}{n}.$$

This, combined with Proposition 4.15 proves the first statement.

By Theorem 4.20, $\mathsf{Inv}(P)$ is nonempty. Let $\mu \in \mathsf{Inv}(P)$. For all $n \in \mathbb{N}^*$

$$P^n V \leq \rho^n V + \kappa \frac{1-\rho^n}{1-\rho} \leq \rho^n V + \kappa \frac{1}{1-\rho}.$$

Thus, by invariance and Jensen's inequality,

$$\mu(V \wedge M) = \mu P^n (V \wedge M) \leq \mu(P^n V \wedge M) \leq \mu((\rho^n V + \frac{\kappa}{1-\rho}) \wedge M).$$

Letting $n \to \infty$ in the right-hand term and using dominated convergence shows that $\mu(V \wedge M) \leq \frac{\kappa}{1-\rho}$. Then, $\mu V \leq \frac{\kappa}{1-\rho}$ by monotone convergence. Compactness follows from Proposition 4.15 and Prohorov's theorem. **QED**

Exercise 4.24 [Invariant measures and mean-occupation] Let (X_k) be a Markov chain, T a finite stopping time (i.e. $T < \infty$ a.s) and let ν be the "mean occupation measure up to time T" defined for all $f \in B(M), f \geq 0$, as

$$\nu f := \mathsf{E}(\sum_{k=0}^{T-1} f(X_k)).$$

(i) Show that $\nu(Pf) - \nu f = \mathsf{E}(f(X_T)) - E(f(X_0))$.

(ii) Show that if X_0 and X_T have the same distribution, and $E(T) < \infty$, then $\frac{\nu}{\nu(1)}$ is an invariant probability measure for the chain.

4.2.1 Excessive measures

A measure μ is called *excessive* provided

$$\mu P \leq \mu.$$

Lemma 4.25 *Every finite excessive measure is invariant.*

Proof If μ is a finite excessive measure, then $\mu P(A) \leq \mu(A)$ and $\mu(M) - \mu P(A) = \mu P(A^c) \leq \mu(A^c) = \mu(M) - \mu(A)$. **QED**

Given two measures α and β on M, one calls α *absolutely continuous* with respect to β and writes $\alpha \ll \beta$ if for every $A \in \mathcal{B}(M)$, $\beta(A) = 0$ implies that $\alpha(A) = 0$. One says that α and β are *mutually singular* and writes $\alpha \perp \beta$ if there is $A \in \mathcal{B}(M)$ such that $\alpha(A) = \beta(A^c) = 0$. Let μ and ν be measures on M. By Lebesgue's decomposition theorem (see, e.g., Theorem 3.8 in [17]), $\nu = \nu_{ac} + \nu_s$, where $\nu_{ac} \ll \mu$ and $\nu_s \perp \mu$. Equivalently,

$$\nu(dx) = h(x)\mu(dx) + \mathbf{1}_A(x)\nu(dx),$$

where $h \in L^1(\mu)$ and $\mu(A) = 0$.

Lemma 4.26 *Let $\mu, \nu \in \mathsf{Inv}(P)$. Then the absolutely continuous and the singular parts of ν with respect to μ are invariant measures.*

Proof Write $\nu(dx) = h(x)\mu(dx) + \mathbf{1}_A(x)\nu(dx)$ with $h \in L^1(\mu)$ and $\mu(A) = 0$. By invariance, $\mu(A) = \int P(x, A)\mu(dx) = 0$, so that $P(x, A) = 0$ for $\mu-$almost every $x \in M$. Thus, for any Borel set B,

$$\int P(x, B)h(x)\mu(dx) = \int P(x, B \cap A^c)h(x)\mu(dx) \leq \nu(B \cap A^c) = (h\mu)(B).$$

This proves that $h(x)\mu(dx)$ is finite and excessive, hence invariant. Since $\mathbf{1}_A(x)\nu(dx) = \nu(dx) - h(x)\mu(dx)$ and since $\nu \in \mathsf{Inv}(P)$, the measure $\mathbf{1}_A(x)\nu(dx)$ is invariant as well. **QED**

4.2.2 Ergodic probability measures

Let $\mu \in \mathsf{Inv}(P)$. A bounded, measurable function g is called $(P, \mu)-$invariant provided $Pg = g$ $\mu-$almost surely. A set $A \in \mathcal{B}(M)$ is called $(P, \mu)-$invariant if $\mathbf{1}_A$ is $(P, \mu)-$invariant.

An invariant probability measure μ is called *ergodic* (for P) if every (P, μ)-invariant function is μ-almost surely constant. (A function $f : M \to \mathbb{R}$ is called μ-almost surely constant if there is $c \in \mathbb{R}$ such that $f(x) = c$ for μ-almost every $x \in M$.)

Lemma 4.27 *A probability measure $\mu \in \mathsf{Inv}(P)$ is ergodic if and only if every (P,μ)-invariant set has μ-measure 0 or 1.*

Proof Suppose first that $\mu \in \mathsf{Inv}(P)$ is not ergodic. Then there exists a bounded, measurable function h such that $Ph = h$ μ-almost surely and for every $c \in \mathbb{R}$,

$$\mu(\{x \in M : \ h(x) = c\}) < 1.$$

It follows that for some $c \in \mathbb{R}$, $A := \{x \in M : \ h(x) > c\}$ has μ-measure different from 0 and 1.

Claim: A is (P,μ)-invariant.

Proof of the claim: By Jensen's inequality, $|Ph| \le P|h|$. Since $\mu(P|h| - |h|) = 0$ by P-invariance of μ, and since $Ph = h$ μ-almost surely, this proves that $|h|$ is (P,μ)-invariant as well. Hence, $\max(0,h) = \frac{1}{2}(h + |h|)$ is $(P,\mu)-$invariant. Similarly,

$$h_n := \min(n\max(0,h-c),1)$$

is (P,μ)-invariant for every $n \ge 1$. Since $\lim_{n\to\infty} h_n = \mathbf{1}_A$, $\mathbf{1}_A$ is (P,μ)-invariant as the pointwise limit of a uniformly bounded sequence of (P,μ)-invariant functions. This proves the claim and one direction of the lemma.

For the converse direction, let μ be ergodic and let A be a (P,μ)-invariant set. Then $\mathbf{1}_A$ is a (P,μ)-invariant function, and ergodicity of μ implies that there is $c \in \mathbb{R}$ such that $\mathbf{1}_A$ is μ-almost surely equal to c. Necessarily, $c \in \{0,1\}$, whence it follows that $\mu(A) \in \{0,1\}$. **QED**

Remark 4.28 One usually defines a *harmonic* map as a measurable map (bounded or nonnegative) such that $Pf = f$. Note that a harmonic map is $(P,\mu)-$invariant for every $\mu \in \mathsf{Inv}(P)$.

A probability measure $\mu \in \mathsf{Inv}(P)$ is called *extremal* if it cannot be written as $\mu = (1-t)\mu_0 + t\mu_1$ with $\mu_0, \mu_1 \in \mathsf{Inv}(P), 0 < t < 1$ and $\mu_0 \neq \mu_1$. Notice that an extremal invariant probability measure cannot be written as the sum of two nontrivial invariant measures that are mutually singular. This fact will be used below in the proof of Proposition 4.29, part (**ii**).

Proposition 4.29 **(i)** *An invariant probability measure μ is ergodic if and only if it is extremal.*

(ii) *Two distinct ergodic probability measures are mutually singular.*

Proof (**i**). Suppose that μ is nonergodic. By Lemma 4.27, there exists a $(P,\mu)-$invariant set A such that $0 < \mu(A) < 1$. Let $\mu_A(.) := \mu(A \cap .)/\mu(A)$. We claim that for every $g \in B(M)$,

$$P(g\mathbf{1}_A) = (Pg)\mathbf{1}_A$$

$\mu-$almost surely. Indeed, by the Cauchy-Schwarz inequality,

$$|P(g\mathbf{1}_A)|^2 \leq P(g^2)P(\mathbf{1}_A) = P(g^2)\mathbf{1}_A$$

$\mu-$almost surely. Thus $P(g\mathbf{1}_A)\mathbf{1}_{A^c} = 0$ $\mu-$almost surely and, interchanging the roles of A and A^c, $P(g\mathbf{1}_{A^c})\mathbf{1}_A = 0$ $\mu-$almost surely. On the other hand,

$$\begin{aligned} P(g\mathbf{1}_A) - (Pg)\mathbf{1}_A &= [P(g\mathbf{1}_A) - Pg]\mathbf{1}_A + P(g\mathbf{1}_A)\mathbf{1}_{A^c} \\ &= -P(g\mathbf{1}_{A^c})\mathbf{1}_A + P(g\mathbf{1}_A)\mathbf{1}_{A^c} = 0. \end{aligned}$$

This proves the claim. Therefore,

$$\begin{aligned} \mu_A(Pg) &= \frac{1}{\mu(A)}\mu((Pg)\mathbf{1}_A) = \frac{1}{\mu(A)}\mu(P(g\mathbf{1}_A)) \\ &= \frac{1}{\mu(A)}\mu(g\mathbf{1}_A) = \mu_A(g). \end{aligned}$$

This proves that μ_A is an invariant probability measure. Similarly, μ_{A^c} is an invariant probability measure, and since $\mu = \mu(A)\mu_A + (1 - \mu(A))\mu_{A^c}$, the probability measure μ is nonextremal.

Suppose now that μ is ergodic, and that $\mu = (1-t)\mu_0 + t\mu_1$ with $\mu_0, \mu_1 \in \mathsf{Inv}(P)$ and $t \in [0,1]$. If $t \neq 0$, $\mu_1 \ll \mu$. Hence, there exists $h \in L^1(\mu)$ such that $\mu_1(dx) = h(x)\mu(dx)$. Then, by Jensen's inequality,

$$\begin{aligned} 0 \leq \mu((Ph-h)^2) = \mu((Ph)^2 - 2hPh + h^2) &\leq \mu(Ph^2 - 2hPh + h^2) \\ = 2\mu h^2 - 2\mu_1 Ph &= 2\mu h^2 - 2\mu_1 h = 0, \end{aligned}$$

from which it follows that $Ph = h$ $\mu-$almost surely, and -by ergodicity and the fact that μ and μ_1 are probability measures- $h = 1$. As a result, $t = 1$ and μ is extremal.

(**ii**). Let μ and ν be ergodic. Write $\nu(dx) = h(x)\mu(dx) + \mu_s(dx)$ with μ_s singular with respect to μ and $h \in L^1(\mu)$. By Lemma 4.26, $h(x)\mu(dx)$ and μ_s are invariant, and by extremality either $h = 0$ or $\mu_s = 0$. If $\mu_s = 0$, the proof just above shows that h must be 1. **QED**

4.3 Unique ergodicity

We say that (X_n) or P is *uniquely ergodic* if the set of P-invariant probability measures has cardinality one. An immediate consequence of the preceding section is

Proposition 4.30 *If P is uniquely ergodic, then its invariant probability measure is ergodic.*

While a deterministic dynamical system is rarely uniquely ergodic (see Section 4.4 for a definition of ergodic probability measures for deterministic dynamical systems), this property is much more often satisfied by random dynamical systems and Markov chains. We start with a simple situation, which can be seen as a random version of the Banach fixed point theorem.

4.3.1 Unique ergodicity of random contractions

Throughout this subsection, let M be a complete, separable metric space. Recall that a map $f : M \to M$ is a *contraction* if its Lipschitz constant

$$Lip(f) := \sup \left\{ \frac{d(f(x), f(y))}{d(x,y)} \ : x \neq y \right\}$$

is < 1. By the *Banach fixed point theorem*, a contraction f has a unique fixed point x^*, and for all $x \in M$, $f^n(x) \to x^*$ at an exponential rate. Here, using the notation of Section **??**, we shall consider a Markov chain recursively defined by

$$X_{n+1} = F_{\theta_{n+1}}(X_n)$$

under the assumption that the maps F_θ are contracting on average.

More precisely, we assume that for each $\theta \in \Theta$, the map F_θ is Lipschitz continuous, and we let $l_\theta := Lip(F_\theta)$. Note that, by separability, the supremum in the definition of the Lipschitz constant can be chosen over a countable set, so that l_θ is measurable in θ.

We say that the family $\{F_\theta\}$ is *contracting on average* if $\int \log(l_\theta)^+ m(d\theta) < \infty$ and

$$\int \log(l_\theta) m(d\theta) =: -\alpha < 0.$$

Here, we allow for α to be $+\infty$. The next result is classical and has been proved in several places. Here we follow the approach of Diaconis and Freedman [12].

Theorem 4.31 *Assume that* $\{F_\theta\}$ *is contracting on average and that*

$$\int \log(d(F_\theta(x_0), x_0))^+ m(d\theta) < \infty \tag{4.9}$$

for some $x_0 \in M$. *Then the induced Markov chain has a unique invariant probability measure* μ^*, *and* X_n *converges in distribution to* μ^*. *In other words, for every probability measure* μ *on* M,

$$\mu P^n \Rightarrow \mu^*.$$

If we furthermore assume that $\alpha < \infty$,

$$A := \sup_\theta \mid \log(l_\theta) + \alpha \mid < \infty$$

and

$$B := \int d(F_\theta(x_0), x_0) m(d\theta) < \infty,$$

then, for every $x \in M$ *there is* $C(x) > 0$ *such that*

$$\rho(\delta_x P^n, \mu^*) \leq C(x) e^{-n\beta}, \quad \forall n \in \mathbb{N},$$

where ρ *stands for the Fortet-Mourier distance (see* (4.2)*), and* $\beta := \min\{\alpha/4, \alpha^2/(32A^2)\}$.

Proof For all $x \in M$, set $X_n^x := F_{\theta_n} \circ \ldots \circ F_{\theta_1}(x)$ and $Y_n^x := F_{\theta_1} \circ \ldots \circ F_{\theta_n}(x)$. The idea of the proof is to show that (Y_n^x) converges almost surely (and thus in law) to some random variable Y_∞ independent of x. Since X_n^x and Y_n^x have the same law, this implies that (X_n^x) converges in law to Y_∞.

To shorten notation, set $l_n := l_{\theta_n}, L_n := \prod_{i=1}^n l_i$, and $Y_n := Y_n^{x_0}$ for x_0 as in (4.9). By the strong law of large numbers, $\mathsf{P}-$almost surely,

$$\lim_{n\to\infty} \frac{\log(L_n)}{n} = -\alpha \in [-\infty, 0). \tag{4.10}$$

Thus, P-almost surely,

$$\limsup_{n\to\infty} \frac{\log(d(Y_n^x, Y_n))}{n} \leq -\alpha$$

because $d(Y_n^x, Y_n^y) \leq L_n d(x, y)$. We shall now show that (Y_n) is almost surely Cauchy, by completeness of M hence convergent.

For all $n, p \in \mathbb{N}$,

$$d(Y_{n+p}, Y_n) \leq \sum_{i=0}^{p-1} d(Y_{n+i+1}, Y_{n+i}) \leq \sum_{i\geq 0} L_{n+i} d(F_{\theta_{n+i+1}}(x_0), x_0). \tag{4.11}$$

Let $0 < \varepsilon < \alpha/2$. Then

$$\sum_{n\geq 1} \mathsf{P}(\log d(F_{\theta_n}(x_0), x_0) \geq \varepsilon n) \leq \sum_{n\geq 1} \mathsf{P}(\log(d(F_{\theta_1}(x_0), x_0))^+ \geq \varepsilon n)$$

$$\leq \frac{1}{\varepsilon} \mathsf{E}(\log(d(F_{\theta_1}(x_0), x_0))^+) < \infty.$$

(Here, we used that $\sum_{n\geq 1} \mathsf{P}(\xi \geq n) \leq \mathsf{E}(\xi)$ for every nonnegative random variable ξ, as well as the integrability condition in (4.9).) Thus, by Borel-Cantelli, $\limsup \frac{\log d(F_{\theta_n}(x_0), x_0)}{n} \leq \varepsilon$ almost surely. Combined with (4.11) and (4.10), it follows that, almost surely, for n large enough,

$$d(Y_{n+p}, Y_n) \leq \sum_{i\geq 0} e^{-(n+i)(\alpha - 2\varepsilon)}.$$

This concludes the proof of the first statement, with μ^* the law of the limiting random variable Y_∞ (see also Exercise 4.32).

We now pass to the second statement. For every bounded Lipschitz function f with $\|f\|_{bl} \leq 1$ and for every $\delta > 0$,

$$|\delta_x P^n f - \mu^* f| = |\mathsf{E}(f(Y_n^x) - f(Y_\infty))| \leq \delta + 2\mathsf{P}(d(Y_n^x, Y_\infty) \geq \delta). \tag{4.12}$$

First observe that by (4.11),

$$d(Y_n^x, Y_\infty) \leq d(Y_n^x, Y_n) + d(Y_n, Y_\infty) \leq L_n d(x, x_0) + \sum_{i\geq 0} L_{n+i} d(x_0, F_{\theta_{n+i+1}}(x_0)).$$

By Markov's inequality,

$$\mathsf{P}(d(x_0, F_{\theta_n}(x_0)) \geq e^{\varepsilon n}) \leq B e^{-\varepsilon n}$$

and by a standard Chernoff inequality (see Exercise 4.33 below),

$$\mathsf{P}(L_n \geq e^{(-\alpha+\varepsilon)n}) \leq e^{-n(\varepsilon^2/2A^2)}.$$

Thus

$$\mathsf{P}(d(Y_n^x, Y_\infty) \geq e^{(-\alpha+\varepsilon)n} d(x, x_0) + \sum_{i \geq 0} e^{(-\alpha+\varepsilon)(n+i)} e^{\varepsilon(n+i)})$$

$$\leq e^{-n(\varepsilon^2/2A^2)} + \sum_{i \geq 0} \left(e^{-(n+i)(\varepsilon^2/2A^2)} + B e^{-\varepsilon(n+i)} \right).$$

Choose $\varepsilon = \alpha/4$. Then

$$\mathsf{P}(d(Y_n^x, Y_\infty) \geq e^{-n\alpha/2}(d(x, x_0) + \frac{1}{1 - e^{-\alpha/2}}))$$

$$\leq e^{-n\alpha^2/(32A^2)}(1 + \frac{1}{1 - e^{-\alpha^2/(32A^2)}}) + B e^{-n\alpha/4} \frac{1}{1 - e^{-\alpha/4}},$$

and we obtain the desired estimate with the help of (4.12). **QED**

Exercise 4.32 Let P be a Markov kernel on a separable metric space M, and let μ^* be a Borel probability measure on M such that for every $x \in M$, $\delta_x P^n$ converges weakly to μ^* as $n \to \infty$. Show that if P is Feller, then μ^* is the unique invariant probability measure for P.

Exercise 4.33 [Chernoff bounds] Let X be an L^1 random variable with zero mean. Assume that $E(e^{\lambda_0 X}) < \infty$ for some $\lambda_0 > 0$. Let $g(\lambda) := \ln(E(e^{\lambda X}))$.

(i) Show that for all $\varepsilon > 0$ and $0 \leq \lambda \leq \lambda_0$,

$$\mathsf{P}(X \geq \varepsilon) \leq e^{-\lambda\varepsilon + g(\lambda)}$$

and

$$\mathsf{P}(X \geq \varepsilon) \leq e^{-g^*(\varepsilon)},$$

where

$$g^*(\varepsilon) := \sup_{0 \leq \lambda \leq \lambda_0} (\lambda\varepsilon - g(\lambda)).$$

(ii) Assume $|X| \leq A < \infty$. Show that $g(\lambda) \leq \frac{A^2\lambda^2}{2}$ and $g^*(\varepsilon) \geq \frac{\varepsilon^2}{2A^2}$. *Hint:* For the first inequality, it may help to use convexity of g.

(iii) Let (X_n) be a sequence of i.i.d. random variables with the same distribution as X. Show that

$$\mathsf{P}(X_1 + \ldots + X_n \geq n\varepsilon) \leq e^{-ng^*(\varepsilon)}.$$

4.4 Ergodic theorems

4.4.1 Classical results from ergodic theory

We first recall some basic definitions from ergodic theory. There are numerous textbooks on the subject including Cornfeld, Fomin, Sinaï [10], Mañe [30], Katok and Haselblatt [26].

Let $(X, \mathcal{F})$ be a measurable space and $T : X \to X$ a measurable mapping. A probability measure $\mathbb{P}$ over X is said to be $T-$*invariant* (or simply invariant) if

$$\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$$

for all $A \in \mathcal{F}$. Given such a $\mathbb{P}$, a measurable function $g : X \to \mathbb{R}$ is said to be $(T, \mathbb{P})-$*invariant* if $g \circ T = g$ $\mathbb{P}-$almost surely, and a measurable set $A \in \mathcal{F}$ is said to be $(T, \mathbb{P})-$*invariant* if $\mathbf{1}_A$ is $(T, \mathbb{P})-$invariant. One also defines a $T-$*invariant set* (or simply invariant set) as a set $A \in \mathcal{F}$ such that $T^{-1}(A) = A$. Note that this definition of invariance makes no reference to the measure $\mathbb{P}$ and that a $T-$invariant set is clearly $(T, \mathbb{P})-$invariant.

A $T-$invariant probability measure $\mathbb{P}$ is said to be $T-$*ergodic* (or simply ergodic) provided that every $(T, \mathbb{P})-$invariant function is $\mathbb{P}-$almost surely constant.

Example 4.34 A *periodic point* of period $d \geq 1$ for T, is a point $x \in X$ such that $T^d(x) = x$ and $T^i(x) \neq x$ for $i = 1, \ldots, d-1$. Given such a point, the measure

$$\frac{1}{d}(\delta_x + \delta_{T(x)} + \ldots + \delta_{T^{d-1}(x)})$$

is $T-$ergodic.

Remark 4.35 One sometimes says that T is *ergodic with respect to* $\mathbb{P}$ to mean that $\mathbb{P}$ is $T-$ergodic.

Proposition 4.36 *The following assertions are equivalent:*

(a) *The probability* $\mathbb{P}$ *is* $T-$ergodic*;*

(b) *Every* $(T, \mathbb{P})-$*invariant set has* $\mathbb{P}-$*measure* 0 *or* 1*;*

(c) *Every* $T-$*invariant set has* $\mathbb{P}-$*measure* 0 *or* 1.

Proof (**a**) $\Rightarrow$ (**b**) $\Rightarrow$ (**c**) are obvious. To show that (**c**) $\Rightarrow$ (**b**), let A be a $(T,\mathbb{P})-$invariant set. The set

$$\tilde{A} := \{x \in X : T^k(x) \in A \text{ for infinitely many } k \in \mathbb{N}\}$$

is invariant. Hence, by (**c**), $\mathbb{P}(\tilde{A}) \in \{0,1\}$. If $x \in A \setminus \tilde{A}$ there exists $k \geq 1$ such that $x \in A \setminus T^{-k}(A)$, and if $x \in \tilde{A} \setminus A$ there exists $k \geq 1$ such that $x \in T^{-k}(A) \setminus A$. It then follows that

$$A\Delta\tilde{A} \subset \bigcup_{k\geq 1} A\Delta T^{-k}(A).$$

Thus

$$\mathbb{P}(A\Delta\tilde{A}) \leq \sum_{k\geq 1} \mathbb{P}(A\Delta T^{-k}(A)).$$

Now

$$\mathbb{P}(A\Delta T^{-k}(A)) \leq \sum_{i=0}^{k-1} \mathbb{P}(T^{-i}(A)\Delta T^{-(i+1)}(A)) = k\mathbb{P}(A\Delta T^{-1}(A)) = 0.$$

It remains to prove that (**b**) $\Rightarrow$ (**a**). Let h be $(T,\mathbb{P})-$invariant. Then, for each $c \in \mathbb{R}$, the set $\{x \in X : h(x) > c\}$ is $(T,\mathbb{P})-$invariant and the result follows. **QED**

Exercise 4.37 [Rotations] Let $S^1 = \mathbb{R}/\mathbb{Z}, \alpha \in S^1$, and $T_\alpha : S^1 \mapsto S^1$ be the rotation $x \mapsto x + \alpha$. Describe the invariant and ergodic probabilities of T_α. Show that when α is *irrational* (i.e $\alpha = \xi + \mathbb{Z}$ with $\xi \in (0,1) \setminus \mathbb{Q}$) T_α is uniquely ergodic, and more precisely, that the normalized Lebesgue measure λ on S^1 is the unique invariant probability for T_α.

Exercise 4.38 Let $k \geq 2$ be an integer and $\mathrm{Z}^k : S^1 \mapsto S^1, x \to kx$. Show that λ is ergodic for Z^k. Show that Z^k has infinitely many periodic points, hence infinitely many ergodic measures.

Exercise 4.39 [Shift] Let $M = \{0,1\}^{\mathbb{N}^*}$ and Θ the shift map on M defined as $\Theta(\omega)_i = \omega_{i+1}$. Show that:

(a) For all $n \geq 1$, Θ has 2^n periodic orbits of period n, and that the set of periodic points is dense in M;

(b) There is point $x \in M$ whose orbit is dense in M;

(c) The probability $(\frac{1}{2}(\delta_0 + \delta_1))^{\otimes \mathbb{N}^*}$ is ergodic for Θ;

(d) There exists a continuous surjective map $\Psi : M \mapsto S^1$ such that

$$\Psi \circ \Theta = \mathrm{Z}^2 \circ \Psi$$

where Z^2 is like in Exercise 4.38. Hint: one can use exercise 4.9.

Using (d), prove that Z^2 posseses a dense orbit and give an alternative proof of the results of Exercise 4.38 when $k = 2$.

Exercise 4.40 Let $T : S^1 \times S^1 \mapsto S^1 \times S^1, (x, y) \to (x + \alpha, y + x)$ with α irrational. Show that $\lambda \otimes \lambda$ is ergodic. Hint : one can use the fact that every $f \in L^2(\lambda \otimes \lambda)$ writes as a Fourier series $f(x, y) = \sum_{k,l \in \mathbb{Z}^2} c_{kl} e_k(x) e_l(y)$ where $e_k(x) = e^{2i\pi kx}$ and $\sum_{k,l} |c_{kl}|^2 < \infty$.

The first important result from ergodic theory is the Poincaré recurrence theorem. Notice that there is no assumption here that $\mathbb{P}$ is ergodic.

Theorem 4.41 (Poincaré recurrence theorem) *Let $\mathbb{P}$ be a $T-$invariant probability measure. For every measurable set $A \subset X$,*

$$\mathbb{P}(A) = \mathbb{P}(\{x \in A : \ T^n(x) \in A \ \textit{for infinitely many } n\}).$$

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

$$B_N := \{x \in A : \ \{T^n(x)\}_{n \geq N} \subset X \setminus A\}.$$

Then $T^{-n}(B_1) \cap B_1 = \emptyset$ for all $n \geq 1$. Hence $T^{-n}(B_1) \cap T^{-m}(B_1) = \emptyset$ for all $m, n \in \mathbb{N}$ and $n \neq m$. Thus

$$1 \geq \sum_{n \in \mathbb{N}} \mathbb{P}(T^{-n}(B_1)) = \sum_{n \in \mathbb{N}} \mathbb{P}(B_1)$$

and $\mathbb{P}(B_1) = 0$. Replacing T with T^N proves that $\mathbb{P}(B_N) = 0$. **QED**

Let $\mathcal{I}$ denote the set of all invariant sets. Then $\mathcal{I}$ is a σ-field. The next result is the celebrated pointwise Birkhoff ergodic theorem. The proof given here follows [26] and goes back to Neveu.

Theorem 4.42 (Birkhoff ergodic theorem) *Let $\mathbb{P}$ be a $T-$invariant probability measure. Let $f \in L^1(\mathbb{P})$. Then $\hat{f} := \mathbb{E}(f|\mathcal{I})$ is $(T, \mathbb{P})-$invariant and*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i = \hat{f}$$

$\mathbb{P}-$almost surely. In particular, if $\mathbb{P}$ is $T-$ergodic, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i = \mathbb{E}(f)$$

$\mathbb{P}-$almost surely.

Proof For $f \in L^1(\mathbb{P})$, set $S_n(f)(x) := \sum_{i=0}^{n-1} f \circ T^i(x)$ and $\hat{f} := \mathbb{E}(f|\mathcal{I})$. We claim that

$$\hat{f} < 0, \quad \mathbb{P} - \text{almost surely} \implies \limsup_{n\to\infty} \frac{S_n(f)}{n} \leq 0, \quad \mathbb{P} - \text{almost surely.}$$

Let us first derive the theorem from the claim. For $\varepsilon > 0$, set $f_\varepsilon := f - \hat{f} - \varepsilon$. Then $\hat{f}_\varepsilon = -\varepsilon < 0$, and since $\hat{f}$ is $(T, \mathbb{P})-$invariant (the proof is easy and left to the reader),

$$\limsup_{n\to\infty} \frac{S_n(f)}{n} - \hat{f} - \varepsilon = \limsup_{n\to\infty} \frac{S_n(f_\varepsilon)}{n} \leq 0.$$

Thus, ε being arbitrary, $\limsup_{n\to\infty} \frac{S_n(f)}{n} \leq \hat{f}$. Similarly, $\liminf_{n\to\infty} \frac{S_n(f)}{n} \geq \hat{f}$.

We now move on to the proof of the claim. For $n \in \mathbb{N}^*$ and $x \in X$, let

$$F_n(x) := \max\{S_k(f)(x) \ : k = 1, \ldots, n\},$$

$F_\infty(x) := \lim_{n\to\infty} F_n(x) \in \mathbb{R} \cup \{\infty\}$, and $A := \{F_\infty = \infty\}$. Clearly

$$\limsup_{n\to\infty} \frac{S_n(f)}{n} \leq 0$$

on $X \setminus A$ and it suffices to prove that $\mathbb{P}(A) = 0$. Now observe that $F_{n+1} - F_n \circ T = f - \min(0, F_n \circ T)$. Consequently, $A \in \mathcal{I}$ and $(F_{n+1} - F_n \circ T)$ decreases to $f - \min(0, F_\infty \circ T)$. In particular, by monotone convergence, $\lim_{n\to\infty} \mathbb{E}((F_{n+1} - F_n \circ T)\mathbf{1}_A) = \mathbb{E}(f\mathbf{1}_A) = \mathbb{E}(\hat{f}\mathbf{1}_A)$. By $T-$invariance of $\mathbb{P}$, the left-hand side is nonnegative. Hence, if $\hat{f} < 0$ $\mathbb{P}-$almost surely, then necessarily $\mathbb{P}(A) = 0$. **QED**

The next theorem, known as the ergodic decomposition theorem, shows that every invariant measure on a Borel subset of a Polish space equipped with the Borel $\sigma-$field can be written as a "sum" of ergodic measures.

Theorem 4.43 (Ergodic decomposition theorem) *Let M be a Borel subset of a Polish space, with Borel σ-field $\mathcal{B}(M)$. Let $T : M \to M$ be a measurable transformation. Every T-invariant probability measure $\mathbb{P}$ can be decomposed as*

$$\mathbb{P}(\cdot) = \int_M P(x, \cdot)\ \mathbb{P}(dx),$$

where P is a Markov kernel on $(M, \mathcal{B}(M))$ such that $P(x, \cdot)$ is ergodic for $\mathbb{P}$-almost every x.

Before proving the ergodic decomposition theorem, we state without proof a lemma that can be deduced from Theorem 10.2.2 in [13] and the monotone class theorem in the appendix.

Lemma 4.44 *Let M be a Borel subset of a Polish space, with Borel σ-field $\mathcal{B}(M)$. Let $\mathbb{P}$ be a probability measure on $(M, \mathcal{B}(M))$, and let $\mathcal{A}$ be a sub-σ-field of $\mathcal{B}(M)$. Then there exists a Markov kernel P on $(M, \mathcal{B}(M))$ such that for every $f \in B(M)$, Pf is a representative of $\mathbb{E}(f|\mathcal{A})$, i.e. Pf is $\mathcal{A}$-measurable and $\mathbb{E}(\mathbf{1}_A Pf) = \mathbb{E}(\mathbf{1}_A f)$ for every $A \in \mathcal{A}$.*

Proof [Theorem 4.43] Recall that $\mathcal{I}$ denotes the σ-field of T-invariant sets in $\mathcal{B}(M)$. By Lemma 4.44, there is a Markov kernel P on $(M, \mathcal{B}(M))$ such that for every $f \in B(M)$, Pf is a representative of $\mathbb{E}(f|\mathcal{I})$. This yields

$$\mathbb{P}(A) = \mathbb{E}(\mathbb{E}(\mathbf{1}_A|\mathcal{I})) = \int_M P(x, A)\ \mathbb{P}(dx), \quad \forall A \in \mathcal{B}(M).$$

As a subset of a separable metric space, M is separable (see part (**i**) of Exercise 4.45). Proposition 4.5 implies the existence of a countable family $\{f_n\}_{n\in\mathbb{N}} \subset C_b(M)$ such that for every $\mu, \nu \in \mathcal{P}(M)$, $\mu = \nu$ if and only if $\mu f_n = \nu f_n$ for all $n \in \mathbb{N}$. For every $n \in \mathbb{N}$, Pf_n is a representative of $\mathbb{E}(f_n|\mathcal{I})$, and $x \mapsto P(x, T^{-1}(\cdot))f_n = P(f_n \circ T)(x)$ is a representative of $\mathbb{E}(f_n \circ T|\mathcal{I})$. Since $\mathbb{P}$ is T-invariant, we have $\mathbb{E}(f_n|\mathcal{I}) = \mathbb{E}(f_n \circ T|\mathcal{I})$ for every $n \in \mathbb{N}$, hence

$P(x, \cdot)$ is T-invariant for $\mathbb{P}$-almost every x. To show that $P(x, \cdot)$ is ergodic for $\mathbb{P}-$almost every x, we follow the proof of Theorem 6.2 in [16]. Since M is a separable metric space, the σ-field $\mathcal{B}(M)$ is countably generated, i.e. there is a countable family of sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\mathcal{B}(M) = \sigma(A_n : n \in \mathbb{N})$ (see part (**ii**) of Exercise 4.45). As a result, $L^1(M, \mathcal{B}(M), \mathbb{P})$ is separable (see parts (**i**) and (**ii**) of Exercise 4.46). Since the set $\{\mathbf{1}_A : \ A \in \mathcal{I}\}$ is contained in $L^1(M, \mathcal{B}(M), \mathbb{P})$, it is also separable in the L^1-topology, so there is a countable family $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{I}$ such that for every $A \in \mathcal{I}$ and for every $\epsilon > 0$, there is $n \in \mathbb{N}$ with $\mathbb{P}(A \triangle A_n) < \epsilon$. Let $\mathcal{I}_0 := \sigma(A_n : \ n \in \mathbb{N})$. By definition, $\mathcal{I}_0$ is a countably generated sub-σ-field of $\mathcal{I}$. Moreover, $\mathcal{I}_0$ and $\mathcal{I}$ are $\mathbb{P}$-equivalent, i.e. for every $A \in \mathcal{I}$, there is $B \in \mathcal{I}_0$ such that $\mathbb{P}(A \triangle B) = 0$ (see part (**iii**) of Exercise 4.46). As $\mathcal{I}$ need not be countably generated (see Exercise 4.48), we will work with $\mathcal{I}_0$ in the remainder of the proof. Applying Lemma 4.44 to $\mathcal{I}_0$, we obtain a Markov kernel Q on $(M, \mathcal{B}(M))$ such that for every $f \in B(M)$, Qf is a representative of $\mathbb{E}(f|\mathcal{I}_0)$. Let $\{f_n\}_{n\in\mathbb{N}} \subset C_b(M)$ be as above. For $n \in \mathbb{N}$, consider the function $h_n := Qf_n$. Since $\mathcal{I}$ and $\mathcal{I}_0$ are $\mathbb{P}$-equivalent, $\mathbb{E}(f_n|\mathcal{I}) = \mathbb{E}(f_n|\mathcal{I}_0)$. As a result, there is $M^1 \in \mathcal{B}(M)$ such that $\mathbb{P}(M^1) = 1$ and for every $x \in M^1$, $P(x, \cdot)$ is $T-$invariant and

$$h_n(x) = P(x, \cdot) f_n, \quad \forall n \in \mathbb{N}.$$

Hence, $Q(x, \cdot) = P(x, \cdot)$ is T-invariant for every $x \in M^1$. By Birkhoff's ergodic theorem (Theorem 4.42), there is $M^2 \subset M^1$ such that $\mathbb{P}(M^2) = 1$ and for every $x \in M^2$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f_n(T^k(x)) = h_n(x), \quad \forall n \in \mathbb{N}.$$

And as both $Q(\cdot, A_n)$ and $\mathbf{1}_{A_n}$ are representatives of $\mathbb{E}(\mathbf{1}_{A_n}|\mathcal{I}_0)$, there is $M^3 \subset M^2$ such that $\mathbb{P}(M^3) = 1$ and $Q(x, A_n) = \mathbf{1}_{A_n}(x)$ for every $x \in M^3$ and $n \in \mathbb{N}$. Finally, as $Q(\cdot, M^3)$ is a representative of $\mathbb{E}(\mathbf{1}_{M^3}|\mathcal{I}_0)$ and as $\mathbb{P}(M^3) = 1$, there is $M^4 \subset M^3$ such that $\mathbb{P}(M^4) = 1$ and $Q(x, M^3) = 1$ for every $x \in M^4$. Let us show that $Q(x, \cdot)$ is ergodic for every $x \in M^4$, which will complete the proof of the ergodic decomposition theorem. Fix $x \in M^4$ and $A \in \mathcal{I}$. In light of Proposition 4.36, it is enough to show that $Q(x, A) \in \{0, 1\}$. If $Q(x, A) = 0$, we are done. If $Q(x, A) > 0$, consider the probability measure

$$\nu(B) := \frac{Q(x, A \cap B)}{Q(x, A)}, \quad B \in \mathcal{B}(M).$$

Since $\nu(A) = 1$, it suffices to show that $Q(x, \cdot) = \nu$, which will follow from

$$h_n(x) = \nu f_n, \quad \forall n \in \mathbb{N}. \tag{4.13}$$

Set

$$[x] := \bigcap_{A \in \mathcal{I}_0 : x \in A} A.$$

By part (**i**) of Exercise 4.47, one has

$$[x] = \bigcap_{n : x \in A_n} A_n \cap \bigcap_{n : x \notin A_n} A_n^c \tag{4.14}$$

and $[x] \in \mathcal{I}_0$. Fix $n \in \mathbb{N}$. Since h_n is $\mathcal{I}_0-$measurable, it is constant on the set $[x]$ by part (**ii**) of Exercise 4.47. Therefore, we have for every $y \in [x] \cap M^3$

$$h_n(x) = h_n(y) = \lim_{N \to \infty} \sum_{k=0}^{N-1} f_n(T^k(y)).$$

Since $x \in M^3$, the representation of $[x]$ in (4.14) implies $Q(x, [x]) = 1$ and thus $Q(x, [x] \cap M^3) = 1$. Since $Q(x, \cdot)$ is T-invariant, another application of Birkhoff's ergodic theorem then yields that the constant $h_n(x)$ is a representative of $\mathsf{E}_{Q(x,\cdot)}(f_n|\mathcal{I})$, where $\mathsf{E}_{Q(x,\cdot)}$ denotes expectation with respect to $Q(x, \cdot)$. Consequently,

$$h_n(x)Q(x, A) = \mathsf{E}_{Q(x,\cdot)}(\mathbf{1}_A h_n(x)) = \mathsf{E}_{Q(x,\cdot)}(\mathbf{1}_A f_n) = \int_M \mathbf{1}_A(z) f_n(z) \; Q(x, dz).$$

Dividing both sides by $Q(x, A)$ gives (4.13). **QED**

Exercise 4.45 [Properties of separable metric spaces] Let (M, d) be a separable metric space.

(i) Let A be any subset of M. Show that A with the metric induced from M is itself a separable metric space.

(ii) Show that the Borel σ-field $\mathcal{B}(M)$ is countably generated.

Exercise 4.46 For an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, prove the following statements:

(i) If $\mathcal{F}$ is countably generated, then $(\Omega, \mathcal{F}, \mathbb{P})$ is separable, i.e. there is a countable family $\mathcal{D} \subset \mathcal{F}$ such that for every $A \in \mathcal{F}$ and $\epsilon > 0$ there is $B \in \mathcal{D}$ with $\mathbb{P}(A \triangle B) < \epsilon$.

(ii) If $(\Omega, \mathcal{F}, \mathbb{P})$ is separable, then $L^1(\Omega, \mathcal{F}, \mathbb{P})$ is a separable metric space.

(iii) If $(\Omega, \mathcal{F}, \mathbb{P})$ is separable, then for every $A \in \mathcal{F}$ there is $B \in \sigma(\mathcal{D})$ such that $\mathbb{P}(A \triangle B) = 0$.

Exercise 4.47 Let $(\Omega, \mathcal{F})$ be a measurable space, let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ be a countable family of sets, and let $\mathcal{A} := \sigma(A_n : n \in \mathbb{N})$. For $x \in \Omega$, set

$$[x]_{\mathcal{A}} := \bigcap_{A\in\mathcal{A}: x\in A} A.$$

(i) Show that for every $x \in \Omega$,

$$[x]_{\mathcal{A}} = \bigcap_{n: x\in A_n} A_n \cap \bigcap_{n: x\notin A_n} A_n^c,$$

and deduce that $[x]_{\mathcal{A}} \in \mathcal{A}$.

(ii) Let $f : \Omega \to \mathbb{R}$ be $\mathcal{A}$-measurable and let $x \in \Omega$. Show that f is constant on $[x]_{\mathcal{A}}$.

The next exercise shows that $\mathcal{I}$, the σ-field of T-invariant sets, need not be countably generated.

Exercise 4.48 Consider the irrational rotation T_α of exercise 4.37 with α irrational. Let $\mathcal{I}$ be the σ-field of T_α-invariant sets. Use the formula from part (**i**) of Exercise 4.47 to show that $\mathcal{I}$ is not countably generated, even though $\mathcal{B}(S^1)$ is.

4.4.2 Application to Markov chains

Consider now the canonical chain introduced in Remark 1.6 in Section 1.2.1. Let $\Theta : M^{\mathbb{N}} \to M^{\mathbb{N}}$ be the shift operator defined by $\Theta(\omega)_n := \omega_{n+1}$ and let $\mathbb{P}_\nu$ be the law of the canonical chain with initial distribution ν and kernel P. Recall that $\mathbb{P}_\nu$ is a probability measure over $M^{\mathbb{N}}$ characterized by (1.5).

Proposition 4.49 **(i)** *$\mathbb{P}_\nu$ is $\Theta-$invariant if and only if $\nu \in \mathsf{Inv}(P)$.*

(ii) *Let $\nu \in \mathsf{Inv}(P)$ and let $h \in L^1(\mathbb{P}_\nu)$ be $(\Theta, \mathbb{P}_\nu)-$invariant. For $x \in M$ such that $h \in L^1(\mathbb{P}_x)$, let*

$$\overline{h}(x) := \mathbb{E}_x(h) = \int h d\mathbb{P}_x.$$

Then

(a) $h(\omega) = \overline{h}(\omega_0)\ \mathbb{P}_\nu -$ *almost surely;*

(b) $\overline{h}$ *is $(P, \nu)-$invariant.*

(iii) $\mathbb{P}_\nu$ *is $\Theta-$ergodic if and only if ν is $P-$ergodic.*

Proof (**i**) follows easily from the definitions.

(**ii**). Let $h \in L^1(\mathbb{P}_\nu)$ be $(\Theta, \mathbb{P}_\nu)-$invariant. For $n \in \mathbb{N}$, set $h_n := \mathbb{E}_\nu(h|\mathcal{F}_n)$. By Doob's martingale convergence theorem (Theorem A.7 in the appendix), h_n converges $\mathbb{P}_\nu-$almost surely, hence in probability, to h. In particular, for all $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}_\nu(|h_{n+1} - h_n| > \varepsilon) = 0$. By $(\Theta, \mathbb{P}_\nu)-$invariance of h and by the Markov property from Proposition 1.8,

$$h_n = \mathbb{E}_\nu(h \circ \Theta^n|\mathcal{F}_n) = \mathbb{E}_{\omega_n}(h) = \overline{h}(\omega_n).$$

Thus,

$$\mathbb{P}_\nu(|h_{n+1} - h_n| > \varepsilon) = \mathbb{P}_\nu(|\overline{h}(\omega_{n+1}) - \overline{h}(\omega_n)| > \varepsilon). \tag{4.15}$$

Since $\nu \in \mathsf{Inv}(P)$, (**i**) implies that $\mathbb{P}_\nu$ is $\Theta-$invariant. The expression on the right-hand side of (4.15) thus equals $\mathbb{P}_\nu(|\overline{h}(\omega_1) - \overline{h}(\omega_0)| > \varepsilon)$, which proves that $h_n = h_0 = h$. Also, by the Markov property, $P\overline{h}(x) = \mathbb{E}_x(\mathbb{E}_{X_1}(h)) = \mathbb{E}_x(\mathbb{E}_x(h \circ \Theta|\mathcal{F}_1)) = \mathbb{E}_x(h \circ \Theta)$. And as h is $(\Theta, \mathbb{P}_\nu)-$invariant, we have for $\nu-$almost every $x \in M$ that $\mathbb{E}_x(h \circ \Theta) = \overline{h}(x)$.

(**iii**). Let ν be $P-$ergodic. We will show that every $(\Theta, \mathbb{P}_\nu)-$invariant function $h \in L^1(\mathbb{P}_\nu)$ is $\mathbb{P}_\nu-$almost surely constant. In particular, every $(\Theta, \mathbb{P}_\nu)-$invariant set has $\mathbb{P}_\nu-$measure 0 or 1, so $\mathbb{P}_\nu$ is $\Theta-$ergodic by Proposition 4.36. If $h \in L^1(\mathbb{P}_\nu)$ is $(\Theta, \mathbb{P}_\nu)-$invariant, then $\overline{h}$ is $\nu-$almost surely constant by (**ii**) and $P-$ergodicity of ν. By (**ii**), this proves that h is $\mathbb{P}_\nu-$almost surely constant. Conversely, assume that $\mathbb{P}_\nu$ is $\Theta-$ergodic. Let A be a $(P, \nu)-$invariant set. Set $\tilde{A} := \{\omega \in M^{\mathbb{N}} : \omega_0 \in A\}$. Then $\mathbb{P}_\nu(\tilde{A} \cap \Theta^{-1}(\tilde{A})) = \int_A \nu(dx)P(x, A) = \nu(A) = \mathbb{P}_\nu(\tilde{A})$. This shows that $\tilde{A}$ is $(\Theta, \mathbb{P}_\nu)-$invariant. Hence $\nu(A) = \mathbb{P}_\nu(\tilde{A}) \in \{0, 1\}$. **QED**

Theorem 4.50 *Let P be a Markov kernel, $\mu \in \mathsf{Inv}(P)$, and $h \in L^1(\mathbb{P}_\mu)$. Then there exist a set $N \in \mathcal{B}(M)$ and a function $\overline{h} \in L^1(\mu)$ such that $\mu(N) = 1$ and, for all $x \in N$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} h \circ \Theta^k(\omega) = \overline{h}(x)$$

$\mathbb{P}_x$−almost surely. If μ is ergodic, then $\overline{h}(x) = \mathbb{E}_\mu(h)$.

Proof By Birkhoff's ergodic theorem, $\frac{1}{n}\sum_{k=0}^{n-1} h\circ\Theta^k(\omega)$ converges $\mathbb{P}_\mu$−almost surely to a $(\Theta, \mathbb{P}_\mu)$−invariant function $\hat{h} \in L^1(\mathbb{P}_\mu)$. According to (**ii**) of Proposition 4.49, $\hat{h}(\omega) = \overline{h}(\omega_0)$ $\mathbb{P}_\mu$−almost surely, where $\overline{h}(\omega_0) := \mathbb{E}_{\omega_0}(h)$. To conclude the proof, we use the fact that $\mathbb{P}_\mu(\cdot) = \int_M \mathbb{P}_x(\cdot)\mu(dx)$. **QED**

Exercise 4.51 [Skew product chains] Let M, N be two metric spaces and

$$T : M \times N \mapsto N,$$

$$(x, y) \to T_x(y)$$

a measurable map. Let (X_n) be a M-valued Markov chain defined on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ and $Y_0 \in N$ be a $\mathcal{F}_0$ measurable random variable. Consider the stochastic process (Y_n) defined by

$$Y_{n+1} = T_{X_n}(Y_n).$$

(i) Show that (X_n, Y_n) is a Markov chain on $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$.

(ii) Suppose $\mu \in \mathcal{P}(M)$ is an invariant probability for (X_n) and $\nu \in \mathcal{P}(N)$ is T_x invariant for all $x \in M$. Show that $\mu \otimes \nu$ is invariant for (X_n, Y_n).

(iii) (inspired by Lemma 2.1 in [19]) We suppose in addition that μ is the **unique** invariant probability of (X_n) and that for all $x \in M$ T_x is 1-Lipchitz. That is

$$d(T_x(y), T_x(z)) \le d(y, z)$$

for all $x \in M, y, z \in N$. Show that:

(a) For all $f \in L_b(M \times N)$, μ almost all $x \in M$ and **all** $y \in \mathsf{supp}(\nu)$

$$\mathbb{P}_{x,y}(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k, Y_k) = (\mu \otimes \nu)(f)) = 1.$$

(b) If $\mathsf{supp}(\nu) = N$, then (X_n, Y_n) is uniquely ergodic.

(iv) Using *(iii)* show that the map defined in Exercise 4.40 is uniquely ergodic. Deduce that for all β irrational, the sequence $(n^2\beta)$ is equidistributed on S^1 (choose $\beta = 2\alpha$. See [18], corollaries 1.12 and 1.13).

Exercise 4.52 [Markov rotations] With the notation of the preceding exercise 4.51, we assume here that $M = \{1, \ldots, n\}, N = S^1$, (X_n) is a Markov chain on M whose transition matrix K is irreducible, and that for all $i \in M, T_i(y) = y + \alpha_i$ for some $\alpha_i \in S^1$.

A *circuit* for K is a sequence $(i_1, \ldots, i_d)$ of $d \geq 1$ distinct points such that $K(i_k, i_{k+1}) > 0$ for $k = 1, \ldots, d$ and $i_{d+1} = i_1$. The purpose of this exercise is to show that the chain (X_n, Y_n) is uniquely ergodic if and only if there exists a circuit $(i_1, \ldots, i_d)$ such that $\alpha_{i_1} + \ldots + \alpha_{i_d}$ is irrational.

(i) (preliminary) Let D be a diagonal matrix whose entries $\theta_1, \ldots, \theta_n$ are complex numbers having modulus 1. Consider the linear equation

$$Ku = Du \tag{4.16}$$

with $u \in \mathbb{C}^n$. Assume that $u \in \mathbb{C}^n$ is a nonzero solution to (4.16). Show that:

(a) $|u_i| = |u_1|$ for $i = 1, \ldots, n$;

(b) $K_{ij} > 0 \Rightarrow u_j = \theta_i u_i$;

(c) For every circuit $(i_1, \ldots, i_d)$ $\theta_{i_1} \ldots \theta_{i_d} = 1$.

Prove that there exists a nonzero solution to (4.16) if and only if for every circuit $(i_1, \ldots, i_d)$ $\theta_{i_1} \ldots \theta_{i_d} = 1$.

(ii) Let μ be the invariant measure of (X_n) and $f = (f_1, \ldots, f_n) \in L^2(\mu \otimes \lambda)$ Set $f_j(x) = \sum_{k \in \mathbb{Z}} u_j(k) e^{2i\pi kx}$ with $\sum_{k \in \mathbb{Z}} |u_j(k)|^2 < \infty$. Show that $Pf = f$ if and only if $Ku(k) = D^k u(k)$ for all $k \in \mathbb{Z}$; where D is the diagonal matrix with entries $e^{2i\pi\alpha_1}, \ldots, e^{2i\pi\alpha_n}$ and $u(k) = (u_j(k))_{j=1,\ldots,n}$. Here P stands for the kernel of (X_n, Y_n).

(iii) Prove the desired result.

The next theorem is the ergodic decomposition theorem for a Markov kernel.

Theorem 4.53 *Let M be a Borel subset of a Polish space and let P be a Markov kernel on $(M, \mathcal{B}(M))$. Every P-invariant probability measure μ can be decomposed as*

$$\mu(\cdot) = \int_M Q(x, \cdot)\ \mu(dx), \tag{4.17}$$

where Q is a Markov kernel on $(M, \mathcal{B}(M))$ such that $Q(x, \cdot)$ is P-ergodic for μ-almost every x.

Proof The proof we give here is taken from unpublished lecture notes by Yuri Bakhtin. Let $\mathcal{I}(P, \mu)$ be the collection of (P, μ)-invariant sets in $\mathcal{B}(M)$. In Exercise 4.54, it is shown that $\mathcal{I}(P, \mu)$ is a σ-field. By Lemma 4.44, there is a Markov kernel Q on $(M, \mathcal{B}(M))$ such that for every $f \in B(M)$, Qf is a representative of $\mathsf{E}_\mu(f|\mathcal{I}(P, \mu))$, where E_μ denotes expectation with respect to μ. In complete analogy to the proof of Theorem 4.43, this yields the representation in (4.17).

It remains to show that $Q(x, \cdot)$ is P-ergodic for μ-almost every $x \in M$. Let $(\tilde{M}, d)$ be a Polish space such that M is a Borel subset of $\tilde{M}$. The space $\tilde{M}^{\mathbb{N}}$ equipped with the metric

$$e(\omega, \alpha) := \sum_{i \in \mathbb{N}} 2^{-i} \frac{d(\omega_i, \alpha_i)}{1 + d(\omega_i, \alpha_i)}$$

is Polish as well; the corresponding Borel σ-field equals the product σ-field $\mathcal{B}(\tilde{M})^{\otimes \mathbb{N}}$. Thus, $M^{\mathbb{N}}$ is a Borel subset of the Polish space $\tilde{M}^{\mathbb{N}}$. By part (**i**) of Proposition 4.49, the Markov measure $\mathbb{P}_\mu$ on $(M^{\mathbb{N}}, \mathcal{B}(M)^{\otimes \mathbb{N}})$ is Θ-invariant. Hence, by the ergodic decomposition theorem (Theorem 4.43), there is a Markov kernel $\mathcal{P}$ on $(M^{\mathbb{N}}, \mathcal{B}(M)^{\otimes \mathbb{N}})$ such that

$$\mathbb{P}_\mu(\cdot) = \int_{M^{\mathbb{N}}} \mathcal{P}(\omega, \cdot)\ \mathbb{P}_\mu(d\omega),$$

and $\mathcal{P}(\omega, \cdot)$ is Θ-ergodic for $\mathbb{P}_\mu$-almost every $\omega \in M^{\mathbb{N}}$. Moreover, as seen in the proof of Theorem 4.43, $\mathcal{P}f$ is a representative of $\mathbb{E}_\mu(f|\mathcal{I})$ for every $f \in B(M^{\mathbb{N}})$, where $\mathcal{I}$ is the σ-field of Θ-invariant sets in $\mathcal{B}(M)^{\otimes \mathbb{N}}$. We will now relate the Markov kernels Q and $\mathcal{P}$ by showing that $\mathbb{P}_\mu-$almost surely,

$$\mathcal{P}(\omega, \cdot) = \mathbb{P}_{Q(\omega_0, \cdot)}(\cdot).$$

Let $\{F_n\}_{n \in \mathbb{N}} \subset C_b(M^{\mathbb{N}})$ such that for every $\mathsf{P}, \mathsf{Q} \in \mathcal{P}(M^{\mathbb{N}})$, $\mathsf{P} = \mathsf{Q}$ if and only if $\mathsf{P}F_n = \mathsf{Q}F_n$ for all $n \in \mathbb{N}$. In Exercise 1.7, we introduced the canonical projections $\pi_n : \ M^{\mathbb{N}} \to M^{n+1}, \ \omega \mapsto (\omega_i)_{i=0,\ldots,n}$. We use π_0 to define the σ-field

$$\mathcal{J} := \{\pi_0^{-1}(A) : \ A \in \mathcal{I}(P,\mu)\}.$$

Claim: The σ-fields $\mathcal{J}$ and $\mathcal{I}$ are $\mathbb{P}_\mu$-equivalent.

Proof of the claim: Let $S \in \mathcal{I}$ and define $\varphi(x) := \mathbb{P}_x(S)$ and

$$A := \{x \in M : \ \varphi(x) = 1\}.$$

By Proposition 4.49, part (**ii**)(**b**), φ is (P,μ)-invariant. Since every $x \in A$ such that $\varphi(x) = P\varphi(x)$ satisfies $P(x,A) = 1$, it follows with part (**i**) of Exercise 4.54 that $A \in \mathcal{I}(P,\mu)$, and hence $\pi_0^{-1}(A) \in \mathcal{J}$. By part (**ii**)(**a**) of Proposition 4.49,

$$\mathbf{1}_S(\omega) = \varphi(\omega_0)$$

for $\mathbb{P}_\mu$-almost every ω. In particular, $\varphi(\omega_0) \in \{0,1\}$ $\mathbb{P}_\mu$-almost surely, so

$$\mathbf{1}_S = \mathbf{1}_{\pi_0^{-1}(A)}, \quad \mathbb{P}_\mu - a.s.$$

Hence,

$$\mathbb{P}_\mu(S \triangle \pi_0^{-1}(A)) = 0.$$

Let us now fix a set $S \in \mathcal{J}$. Then there is $A \in \mathcal{I}(P,\mu)$ such that $S = \pi_0^{-1}(A)$. Set $\tilde{S} := A^{\mathbb{N}} = \bigcap_{n\in\mathbb{N}} \pi_n^{-1}(A^{n+1})$. A simple induction argument using $A \in \mathcal{I}(P,\mu)$ implies that $\mathbb{P}_\mu(\pi_n^{-1}(A^{n+1})) = \mu(A)$ for all $n \in \mathbb{N}$. Continuity of $\mathbb{P}_\mu$ from above yields

$$\mathbb{P}_\mu(\tilde{S}) = \mu(A) = \mathbb{P}_\mu(S).$$

Since $\tilde{S} \subset S$, it follows that $\mathbb{P}_\mu(S \triangle \tilde{S}) = 0$. And in the proof of Proposition 4.36, it was shown that for any $\tilde{S} \in \mathcal{I}(P,\mu)$ there is $\hat{S} \in \mathcal{I}$ such that $\mathbb{P}_\mu(\hat{S} \triangle \tilde{S}) = 0$. This completes the proof of the claim.

We now complete the proof of Theorem 4.53. Since $\mathcal{I}$ and $\mathcal{J}$ are $\mathbb{P}_\mu$-equivalent, we have for every $n \in \mathbb{N}$ that the representatives of $\mathbb{E}_\mu(F_n|\mathcal{I})$ and the representatives of $\mathbb{E}_\mu(F_n|\mathcal{J})$ are representatives of $\mathbb{E}_\mu(F_n|\sigma(\mathcal{I},\mathcal{J}))$. The function $\mathcal{P}F_n$ is a representative of $\mathbb{E}_\mu(F_n|\mathcal{I})$ and thus also of $\mathbb{E}_\mu(F_n|\sigma(\mathcal{I},\mathcal{J}))$. Let

$$\mathcal{F}_0 := \{\pi_0^{-1}(A) : A \in \mathcal{B}(M)\}.$$

For $n \in \mathbb{N}$, consider the functions

$$\overline{F_n} : M \to \mathbb{R}, \ x \mapsto \mathbb{E}_x(F_n)$$

and

$$G_n : M^{\mathbb{N}} \to \mathbb{R},\ \omega \mapsto \overline{F_n}(\omega_0).$$

By the Markov property from Proposition 1.8, G_n is a representative of $\mathbb{E}_\mu(F_n|\mathcal{F}_0)$. As a result,

$$\mathbb{E}_\mu(F_n|\mathcal{J}) = \mathbb{E}_\mu(\mathbb{E}_\mu(F_n|\mathcal{F}_0)|\mathcal{J}) = \mathbb{E}_\mu(G_n|\mathcal{J}).$$

Next, observe that $\omega \mapsto Q\overline{F_n}(\omega_0)$ is a representative of $\mathbb{E}_\mu(G_n|\mathcal{J})$, and thus also of $\mathbb{E}_\mu(F_n|\mathcal{J})$ and $\mathbb{E}_\mu(F_n|\sigma(\mathcal{I},\mathcal{J}))$. This shows that

$$\begin{aligned} 1 =& \mathbb{P}_\mu\left(\left\{\omega \in M^{\mathbb{N}} : \mathcal{P}F_n(\omega) = Q\overline{F_n}(\omega_0)\right\}\right) \\ =& \mathbb{P}_\mu\left(\left\{\omega \in M^{\mathbb{N}} : \mathcal{P}(\omega,\cdot)F_n = \mathbb{P}_{Q(\omega_0,\cdot)}F_n\right\}\right), \end{aligned}$$

and hence

$$\mathbb{P}_\mu\left(\left\{\omega \in M^{\mathbb{N}} : \mathcal{P}(\omega,\cdot) = \mathbb{P}_{Q(\omega_0,\cdot)}\right\}\right) = 1.$$

Let $S \in \mathcal{B}(M)^{\otimes\mathbb{N}}$ such that $\mathbb{P}_\mu(S) = 1$ and for every $\omega \in S$, $\mathcal{P}(\omega,\cdot) = \mathbb{P}_{Q(\omega_0,\cdot)}$ and $\mathcal{P}(\omega,\cdot)$ is Θ-ergodic. By part (**iii**) of Proposition 4.49, $Q(\omega_0,\cdot)$ is P-ergodic for every $\omega \in S$. Since $S \in \mathcal{B}(M)^{\otimes\mathbb{N}}$ and since π_0 is continuous, the set $\pi_0(S)$ is analytic (see Theorem 13.2.1 in [13]). Theorem 13.2.6 in [13] implies that there are $A, N \in \mathcal{B}(M)$ and $B \subset N$ such that $\mu(N) = 0$ and $\pi_0(S) = A \cup B$. It follows that

$$1 = \mathbb{P}_\mu(S) \le \mathbb{P}_\mu(\pi_0^{-1}(A \cup N)) = \mu(A \cup N) \le \mu(A) + \mu(N) = \mu(A),$$

which completes the proof. **QED**

Exercise 4.54 Let

$$\mathcal{I}(P,\mu) := \{A \in \mathcal{B}(M) :\ \mathbf{1}_A = P(\cdot, A)\ \mu - a.s.\}$$

be the collection of (P,μ)-invariant sets in $\mathcal{B}(M)$.

(i) Show that

$$\mathcal{I}(P,\mu) = \{A \in \mathcal{B}(M) :\ \mu(\{x \in A :\ P(x, A^c) > 0\}) = 0\}.$$

(ii) With the help of the representation in part (**i**), show that $\mathcal{I}(P,\mu)$ is a σ-field.

Chapter 5

Irreducibility

5.1 Resolvent and ξ-irreducibility

Given a (nonzero) measure ξ on M, P is called ξ-*irreducible* if for every Borel set $A \subset M$ and every $x \in M$

$$\xi(A) > 0 \Rightarrow \exists k \geq 0, P^k(x, A) > 0.$$

Equivalently,

$$\xi(A) > 0 \Rightarrow R_a(x, A) > 0$$

where $R_a(.,.)$ is the *Resolvent Kernel* defined as

$$R_a(x, A) = (1-a)\sum_{k\geq 0} a^k P^k(x, A)$$

for some $0 < a < 1$.

Remarks 5.1

(i) Let (X_n) be a Markov chain with kernel P and (Δ_n) a sequence of i.i.d random variables independent from (X_n) having a geometric distribution with parameter a, that is

$$\mathsf{P}(\Delta_i = k) = a^k(1-a),\ k \in \mathbb{N};$$

Then R_a is the Kernel of the sampled chain $Y_n = X_{T_n}$ with

$$T_n = \sum_{i=1}^{n} \Delta_i;$$

(ii) P and R_a have the same invariant probabilities;

(iii) If P is ξ-*irreducible*, then **for all** $n \in \mathbb{N}, x \in M$ and $A \in \mathcal{B}(M)$ such that $\xi(A) > 0$ there exists $k \geq n$ such that $P^k(x, A) > 0$.

Exercise 5.2 (i) Check the assertions of the preceding remark

(ii) Using the notation of Remark 5.1, show that for all $m \in \mathbb{N}^*$ T_m has a *negative bimomial* distribution with parameters (a, m). That is

$$\mathsf{P}(T_m = k) = \left(\begin{array}{c} k+m-1 \\ m-1 \end{array} \right) a^k (1-a)^m$$

for all $k \in \mathbb{N}$. Let $Y_n^m = X_{T_{nm}}$. Show that $(Y_n^m)_n$ is a Markov chain with kernel R_a^m.

Example 5.3 (Doeblin condition) Suppose that $R_a(x, A) \geq \xi(A)$ for some non zero measure ξ. Then P is $\xi-$ irreducible.

Example 5.4 (Countable chains) If M is countable and P is irreducible in the usual sense (see Chapter 2) then it is ξ-irreducible for $\xi = \sum_x \delta_x$.

Theorem 5.5 *Suppose that P is ξ-irreducible. Then P admits at most one invariant probability.*

Proof The assumption implies that ξ is absolutely continuous with respect to every invariant probability, but since distinct ergodic probabilities are mutually singular (Proposition 4.29), there is at most one such probability. **QED**

5.2 The accessible set

With the exception of a few particular cases (like Examples 5.3 and 5.4) it is in general not an easy task to verify that a Markov chain is $\xi-$ irreducible. A purely topological notion of irreducibility is defined below. Combined with the existence of certain points satisfying a local Doeblin condition (see section 5.4), this will ensure $\xi-$ irreducibility.

Recall that the (topological) *support* of a measure μ is the closed set $\mathsf{supp}(\mu)$ defined as the intersection of all closed sets $F \subset M$ such that $\mu(M \setminus F) = 0$. It enjoys the following properties:

(a) $\mu(M \setminus \mathsf{supp}(\mu)) = 0$;

(b) $x \in \mathsf{supp}(\mu)$ if and only if $\mu(O) > 0$ for every open set O containing x.

Exercise 5.6 Prove that assertions $a), b)$ above hold in any separable metric space. Use (or prove) the fact that such a space has a countable basis of open sets.

We define the set of points that are *accessible* from $x \in M$ as

$$\Gamma_x = \mathsf{supp}(R_a(x, \cdot)).$$

Equivalently, y is accessible from x if for every neighborhood U of y there exists $k \geq 0$ such that $P^k(x, U) > 0$.

For $C \subset M$, we let $\Gamma_C = \cap_{x \in C} \Gamma_x$ denote the set of points that are accessible from C and $\Gamma := \Gamma_M$ the set of *accessible points*. Note that Γ_C is a closed (but possibly empty) set. We say that P is (topologically) *indecomposable* if $\Gamma \neq \emptyset$.

Remark 5.7 If P is ξ-irreducible, then it is indecomposable and

$$supp(\xi) \subset \Gamma.$$

The converse implication is false in general (see Theorem 5.5 and Remark 5.10) but true for strong Feller chains (see Proposition 5.17).

Proposition 5.8 *Assume P is Feller and topologically indecomposable. Then*

(i) $P(x, \Gamma) = 1$ *for all* $x \in \Gamma$;

(ii) $\Gamma \subset supp(\mu)$ *for all* $\mu \in \mathsf{Inv}(P)$;

(iii) *If* Γ *has nonempty interior,* $supp(\mu) = \Gamma$ *for all* $\mu \in \mathsf{Inv}(P)$;

(iv) *If* Γ *is compact, there exists* $\mu \in \mathsf{Inv}(P)$ *such that* $supp(\mu) = \Gamma$;

(v) *If* Γ *is compact and* $g : \Gamma \mapsto \mathbb{R}$ *is a continuous and harmonic on* Γ *(i.e* $Pg(x) = g(x)$ *for all* $x \in \Gamma$*), then* g *is constant.*

Proof (i) amounts to proving that $\mathsf{supp}(P(x,.)) \subset \Gamma$. Let $x^* \in \mathsf{supp}(P(x,.))$ and O an open set containing x^*. Then $\delta = P(x,O) > 0$. By Feller continuity and the Portmanteau theorem 4.1 $V = \{y \in M \ : \ P(y,O) > \delta/2\}$ is an open set containing x. Let $z \in M$ and $k \in \mathbb{N}$ be such that $P^k(z,V) > 0$ (recall that $x \in \Gamma$). Then

$$P^{k+1}(z,O) \geq \int_V P^k(z,dy)P(y,O) \geq \frac{\delta}{2} P^k(z,V) > 0.$$

This proves that $x^* \in \Gamma$.

(ii) Let $x \in \Gamma, U$ a neighborhood of x and μ an invariant probability. Then $\mu(U) = \int \mu(dy)R(y,U) > 0$.

(iii) By invariance $\mu(\Gamma) = \int_\Gamma \mu(dx)R(x,\Gamma) + \int_{\Gamma^c} \mu(dx)R(x,\Gamma)$ and since - by (i)- $R(x,\Gamma) = 1$ for all $x \in \Gamma$ it follows that $\int_{\Gamma^c} \mu(dx)R(x,\Gamma) = 0$. If furthermore Γ has nonempty interior, then $R(x,\Gamma) > 0$ for all x, so that $\mu(\Gamma^c) = 0$. This proves that $\mathsf{supp}(\mu) \subset \Gamma$.

(iv) By (i), Feller continuity and Theorem 4.20, there exists an invariant probability μ with $\mu(\Gamma) = 1$. Hence the result.

(v) By (i) we can assume without loss of generality that $\Gamma = M$. By compactness, accessibility and Feller continuity, for every open set $O \subset M$ there exists a finite covering of M by open sets $U_1, \ldots, U_k$, integers $n_1, \ldots, n_k$ and $\delta > 0$ such that $P^{n_i}(x,O) \geq \delta$ for all $x \in U_i$. Thus $\mathbb{P}_x(\tau_O > n) \leq (1-\delta)$ for $n = \max(n_1, \ldots, n_k)$, hence $\mathbb{P}_x(\tau_O > kn) \leq (1-\delta)^k$ by the Markov property. Thus $\mathbb{P}_x(\tau_O < \infty) = 1$. The assumption that g is harmonic makes $(g(X_n))$ a bounded martingale. It then converges $\mathbb{P}_x$ almost surely. If g is nonconstant there exist $a < b$ such that $\{g < a\}$ and $\{g > b\}$ are nonempty open sets, and, by what precedes, (X_n) visits infinitely often these sets $\mathbb{P}_x$ almost surely. A Contradiction. **QED**

Remark 5.9 The inclusion $\Gamma \subset \mathsf{supp}(\mu)$ doesn't require Feller continuity.

Remark 5.10 The inclusion $\Gamma \subset \mathsf{supp}(\mu)$ may be strict when Γ has empty interior as shown by following exercise 5.11.

Exercise 5.11 Let $F : \{0,1\} \times [0,1] \mapsto [0,1]$ be the map defined by

$$F(0,x) = ax, \ F(1,x) = bx(1-x),$$

where $0 < a < 1$ and $1 < b < 4$. Let (X_n) be the Markov chain on $[0,1]$ defined by $X_{n+1} = F(\theta_{n+1}, X_n)$, $X_0 = x > 0$ where (θ_n) is an i.i.d Bernoulli sequence with distribution $(1-p)\delta_0 + p\delta_1$ for some $0 < p < 1$. Show that $\Gamma = \{0\}$ and that when $(1-p)\log a + p \log b > 0$ there exists an invariant probability μ such that $\mu(\{0\}) = 0$, hence $\mathsf{supp}(\mu) \not= \subset \Gamma$.

Other examples where the inclusion $\Gamma \subset \mathsf{supp}(\mu)$ can be found in [3] and [4].

In case P is uniquely ergodic on a compact set, it is topologically indecomposable.

Proposition 5.12 *Suppose M is compact, P Feller and uniquely ergodic. Then P is indecomposable and $\Gamma = \mathsf{supp}(\mu)$.*

Proof By Proposition 5.8 it suffices to prove that Γ is nonempty. By Theorem 4.20, $\frac{1}{n}\sum_{k=1}^n P^k(x,\cdot) \Rightarrow \mu$ for all $x \in M$. Hence for any open set O such that $\mu(O) > 0$ $\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P^k(x,O) > 0$. Thus $R(x,O) > 0$. **QED**

A partial converse to Proposition 5.12 is the following result.

Recall that $L_b(M)$ is the set of bounded Lipschitz real valued functions on M.

Proposition 5.13 *Assume that M is compact, P is Feller, Γ has nonempty interior and that for all $f \in L_b(M)$ the sequence $(P^n f)_{n\geq 1}$ is equicontinuous. Then P is uniquely ergodic.*

Proof By equicontinuity of $(P^n f)_{n\geq 1}$, the sequence $(\overline{f}_n)_{n\geq 1}$ defined by

$$\overline{f}_n = \frac{\sum_{k=1}^n P^k f}{n}$$

is also equicontinuous, hence relatively compact in $C_b(M)$ by Arzelà-Ascoli theorem. Let g be a limit point of $(\overline{f}_n)_{n\geq 1}$. Then g is continuous and $Pg = g$. By Proposition 5.8 (v), $g|_\Gamma$ is a constant C_f. Let now μ and ν be two invariant probabilities. Then $\mu P f = \mu f$ implies that $\mu(\overline{f}_n) = \mu(f)$. Therefore $\mu(f) = \mu(g) = \mu(g|_\Gamma) = C_f$ Similarly $\nu(f) = C_f$ This proves that $\mu = \nu$. **QED**

Exercise 5.14 Deduce from Proposition 5.13 that the irrational rotation T_α (see Exercise 4.37) is uniquely ergodic.

Exercise 5.15 Let M be a compact space. Using the notation of Chapter 3 and Section 4.3.1, consider the the Markov chain on M recursively defined by

$$X_{n+1} = F_{\theta_{n+1}}(X_n).$$

Assume that Θ is a metric space, $(\theta, x) \to F_\theta(x)$ is continuous and that for each $\theta \in \Theta$ F_θ is Lipschitz with Lipchitz constant l_θ. Assume furthermore that

(i) $\int l_\theta m(d\theta) \leq 1$ (compare with the condition of Theorem 4.31);

(ii) For every $x \in M$ and every open set $O \subset M$, there exists a sequence $\theta_1, \ldots, \theta_n$ with $\theta_i \in \mathsf{supp}(m)$ such that $f_{\theta_n} \circ \ldots f_{\theta_1}(x) \in O$.

Show that (X_n) is uniquely ergodic.

Remark 5.16 It is important to emphasize here that the condition that Γ has non empty interior is not sufficient to ensure uniqueness of the invariant probability. For instance, Furstenberg in a remarkable work [19] (see also [30]) has shown that for a convenient choice of $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and β a smooth map on $S^1 := \mathbb{R}/\mathbb{Z}$, the diffeomorphism

$$T : S^1 \times S^1 \mapsto S^1 \times S^1,$$

$$(x, y) \mapsto x + \alpha, y + \beta(x)$$

is minimal (all the orbits are dense) but non uniquely ergodic.

Another example is given by the Ising Model on $\mathbb{Z}^2$. This is a Feller Markov process on a compact set $M = \{-1, 1\}^{\mathbb{Z}^2}$ which all points are accessible (i.e $\Gamma = M$) and which admits (at low temperature) several invariant probabilities (See [24] Example 2.3 for a discussion and further references).

Proposition 5.17 *Suppose that P is topologically* indecomposable *and that for some $x^* \in \Gamma$ and all $A \in \mathcal{B}(M)$ $x \mapsto P(x, A)$ is lower semi-continuous at x^*. Then P is ξ irreducible for $\xi = P(x^*, .)$, In particular P admits at most one invariant probability.*

Proof Let A be such that $P(x^*, A) > 0$. Then for all $x \in M$ there exists O neighborhood of x^* and $n \geq 0$ such that $P^n(x, O) > 0$ and $P(y, A) > 0$ for all $y \in O$ (by lower semi continuity of $x \mapsto P(x, A)$ at x^*.) Thus $P^{n+1} 1_A(x) \geq \int_O P^n(x, dy) P(y, A) > 0$. **QED**

Note that the assumption that $x \mapsto P(x, A)$ is lower semi-continuous at x^* is automatically satisfied if P is strong Feller. Hence Proposition 5.17 gives a practical tool to ensure that a strong Feller chain is uniquely ergodic. Another result is the following

Proposition 5.18 *Suppose that P is strong Feller. Then*

(i) *Two distinct ergodic probabilities have disjoint support;*

(ii) *The support of an invariant non-ergodic probability is disconnected;*

(iii) *If M is connected and P has an invariant probability having full support, P is uniquely ergodic.*

Proof (i) Let μ, ν be two distinct ergodic probabilities. By Proposition 4.29 they are mutually singular. Hence there exists a Borel set $A \subset M$ such that $\mu(A) = 1$ and $\nu(A) = 0$. The set $\{x \in M : \ P(x, A) = 1\}$ is closed (strong Feller property) and has μ measure 1 because $1 = \mu(A) = \int \mu(dx)P(x, A)$. Thus $\mathsf{supp}(\mu) \subset \{x \in M : \ P(x, A) = 1\}$. Similarly $\mathsf{supp}(\nu) \subset \{x \in M : P(x, M \setminus A) = 1\}$.

(ii) Let μ be invariant and let A be such that $P\mathbf{1}_A = \mathbf{1}_A$ μ-almost surely and $0 < \mu(A) < 1$. Set $f = P\mathbf{1}_A$. Then $f(x) \in \{0, 1\}$ for μ almost every x and (by the strong Feller property) f is continuous. Thus f restricted to $\mathsf{supp}(\mu)$ takes values in $\{0, 1\}$ If now $\mathsf{supp}(\mu)$ is connected then f restricted to $\mathsf{supp}(\mu)$ is constant and $\mu(A) \in \{0, 1\}$. (iii) follows from (ii). **QED**

5.3 The asymptotic strong Feller property

The asymptotic strong Feller property was introduced in [25] by Hairer and Mattingly to prove uniqueness for the invariant probability measure of the Navier–Stokes equation on the two-dimensional torus, subject to degenerate stochastic forcing. Before we define this property, we introduce some notation.

Let (M, d^*) be a separable metric space, with $\mathcal{P}(M)$ the space of probability measures on $(M, \mathcal{B}(M))$. One important idea in this section is to consider a whole family of metrics on M, but throughout, d^* will be the metric that gives rise to the topology on M, and in particular induces the σ-field $\mathcal{B}(M)$.

For any bounded metric d on M, we let $\mathrm{Lip}_1(d)$ denote the set of $\mathcal{B}(M)$-measurable functions $\phi : M \to \mathbb{R}$ such that

$$|\phi(x) - \phi(y)| \leq d(x,y), \quad \forall x, y \in M.$$

Notice that $\mathrm{Lip}_1(d)$ contains all constant functions. If the metric d is continuous with respect to the topology induced by d^* and if $\mathcal{B}_d(M)$ denotes the Borel σ-field with respect to d, then $\mathrm{Lip}_1(d)$ is equal to the set of $\mathcal{B}_d(M)$-measurable functions $\phi : M \to \mathbb{R}$ such that $|\phi(x) - \phi(y)| \leq d(x,y)$ for all $x, y \in M$. For $\mu, \nu \in \mathcal{P}(M)$, we define

$$\|\mu - \nu\|_d := \sup_{\phi \in \mathrm{Lip}_1(d)} (\mu\phi - \nu\phi).$$

Boundedness of d guarantees that every function in $\mathrm{Lip}_1(d)$ is bounded and thus integrable with respect to any Borel probability measure on M.

Exercise 5.19 Let d^* be bounded. Show that $(\mu, \nu) \mapsto \|\mu - \nu\|_{d^*}$ defines a bounded metric on $\mathcal{P}(M)$.

Remark 5.20 If $\delta(x,y) := \mathbf{1}_{x \neq y}$ is the discrete metric, then

$$\|\mu - \nu\|_\delta = \tfrac{1}{2}|\mu - \nu| := \tfrac{1}{2}\sup\{|\mu f - \nu f| : f \in B(M),\ \|f\|_\infty \leq 1\},$$

where $|\mu - \nu|$ is the so-called total variation distance between μ and ν. The latter will play a key role in Chapter 7.

We call a metric d on M *continuous* if it is continuous as a function from $M \times M$ to $[0, \infty)$, where $M \times M$ has the topology induced by the product metric $(d^* \star d^*)((x,y),(x',y')) := d^*(x,x') + d^*(y,y')$. Notice in particular that d^* itself is continuous. A sequence of metrics $(d_n)_{n\geq 1}$ on M is called *nondecreasing* if for every $n \in \mathbb{N}^*$,

$$d_{n+1}(x,y) \geq d_n(x,y), \quad \forall x, y \in M.$$

Recall that $\delta(x,y) := \mathbf{1}_{x \neq y}$ and that δ_x is the Dirac measure that assigns mass 1 to $\{x\}$.

Definition 5.21 (Hairer, Mattingly) *We say that a Markov kernel P on M is asymptotic strong Feller at $x \in M$ if there exist a nondecreasing sequence $(n_k)_{k\geq 1}$ of positive integers and a nondecreasing sequence $(d_k)_{k\geq 1}$ of continuous metrics on M such that*

$$\lim_{k\to\infty} d_k(y,z) = \delta(y,z), \quad \forall y, z \in M,$$

and

$$\inf_{\substack{x\in U\subset M,\\ U open}} \limsup_{k\to\infty} \sup_{y\in U} \|\delta_x P^{n_k} - \delta_y P^{n_k}\|_{d_k} = 0.$$

We call P asymptotic strong Feller if it is asymptotic strong Feller at every $x \in M$.

Since $(d_k)_{k\geq 1}$ is nondecreasing and converges to a bounded metric, each metric d_k is, of course, bounded.

5.3.1 Strong Feller implies asymptotic strong Feller

In this subsection, we show that every strong Feller Markov kernel also has the asymptotic strong Feller property. The proof of this statement makes use of the ultra Feller property, which we now define. A Markov kernel P on M is called *ultra Feller* if for every $x \in M$,

$$\inf_{\substack{x\in U\subset M,\\ U open}} \sup_{y\in U} \|\delta_x P - \delta_y P\|_{\delta} = 0,$$

where δ denotes the discrete metric. The following statement corresponds to Theorem 1.6.6 in [23]. It is due to Dellacherie and Meyer, see [11].

Proposition 5.22 *Let P and Q be strong Feller Markov kernels on M. Then the Markov kernel PQ is ultra Feller.*

The proof of Proposition 5.22 we present here is taken from [23]. It is an adaptation of an argument due to Seidler. We begin by stating two lemmas.

Lemma 5.23 *Let P be a strong Feller Markov kernel on M. Then there exists $\pi \in \mathcal{P}(M)$ such that $P(x,\cdot) \ll \pi$ for every $x \in M$.*

Proof Since M is separable, there is a dense sequence $(x_n)_{n\geq 1}$ of elements of M. We define the probability measure

$$\pi(A) := \sum_{n=1}^{\infty} 2^{-n} P(x_n, A), \quad A \in \mathcal{B}(M).$$

To obtain a contradiction, assume there is $x \in M$ such that $P(x,\cdot)$ is not absolutely continuous with respect to π. Then there is $A \in \mathcal{B}(M)$ such that

$\pi(A) = 0$ and $P(x, A) > 0$. Let $f := \mathbf{1}_A \in B(M)$. Since P is strong Feller, Pf is continuous. We have $Pf(x) = P(x, A) > 0$. Since $\pi(A) = 0$, we have $0 = P(x_n, A) = Pf(x_n)$ for every $n \in \mathbb{N}^*$. But then continuity of Pf and the fact that (x_n) is dense in M imply that $Pf \equiv 0$, a contradiction. **QED**

The following real-analysis lemma corresponds to Corollay 1.6.3 in [23], where a proof can be found. Recall from the proof of Lemma 4.44 in Section 4.4 that a σ-field $\mathcal{F}$ is called *countably generated* if there exists a countable family of sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\mathcal{F} = \sigma(A_n : n \in \mathbb{N})$.

Lemma 5.24 *Let $(\Omega, \mathcal{F}, \pi)$ be a measure space such that $\mathcal{F}$ is countably generated. Let (ϕ_n) be a bounded sequence in $L^\infty(\Omega, \mathcal{F}, \pi)$. Then there exist a subsequence $(\phi_{n_k})_{k\geq 1}$ and $\phi \in L^\infty(\Omega, \mathcal{F}, \pi)$ such that*

$$\lim_{k\to\infty} \int_\Omega \phi_{n_k}(x) f(x)\ \pi(dx) = \int_\Omega \phi(x) f(x)\ \pi(dx), \quad \forall f \in L^1(\Omega, \mathcal{F}, \pi).$$

We proceed to the proof of Proposition 5.22.

Proof [of Proposition 5.22] Since Q is strong Feller, Lemma 5.23 yields existence of a probability measure π on $(M, \mathcal{B}(M))$ such that $Q(x, \cdot) \ll \pi$ for every $x \in M$. To obtain a contradiction, suppose that the kernel PQ is not ultra Feller. Then there are $x \in M$ and $\varepsilon > 0$ such that for every open neighborhood U of x,

$$\sup_{y\in U} \|\delta_x PQ - \delta_y PQ\|_\delta > \varepsilon.$$

For $r > 0$ and $y \in M$, let $B_r(y) := \{z \in M : d^*(y, z) < r\}$ be the open d^*-ball of radius r centered at y. Then for every $n \in \mathbb{N}^*$, there is $y_n \in B_{1/n}(x)$ such that

$$\|\delta_x PQ - \delta_{y_n} PQ\|_\delta > \varepsilon.$$

According to Remark 5.20,

$$\sup_{\phi\in B(M):\|\phi\|_\infty\leq 1} (PQ\phi(x) - PQ\phi(y_n)) > 2\varepsilon, \quad \forall n \in \mathbb{N}^*,$$

where the expression on the left denotes the total variation distance between $\delta_x PQ$ and $\delta_{y_n} PQ$. As a result, there is a sequence $(\phi_n)_{n\geq 1}$ in $B(M)$ such that $\|\phi_n\|_\infty \leq 1$ and

$$PQ\phi_n(x) - PQ\phi_n(y_n) > 2\varepsilon, \quad \forall n \in \mathbb{N}^*. \tag{5.1}$$

Since M is a separable metric space, part (**ii**) of Exercise 4.45 implies that the σ-field $\mathcal{B}(M)$ is countably generated. And since (ϕ_n) is a bounded sequence in $L^\infty(M, \mathcal{B}(M), \pi)$, Lemma 5.24 implies that there exist a subsequence $(\phi_{n_k})_{k\geq 1}$ and a function $\phi \in L^\infty(M, \mathcal{B}(M), \pi)$ such that

$$\lim_{k\to\infty} \int_M \phi_{n_k}(x) f(x)\ \pi(dx) = \int_M \phi(x) f(x)\ \pi(dx), \quad \forall f \in L^1(M, \mathcal{B}(M), \pi).$$

Since $Q(x, \cdot) \ll \pi$ for every $x \in M$, we have that for every $x \in M$ there is $h_x \in L^1(M, \mathcal{B}(M), \pi)$ with $Q(x, dy) = h_x(y)\ \pi(dy)$. Then, for every $x \in M$,

$$\lim_{k\to\infty} Q\phi_{n_k}(x) = Q\phi(x).$$

To keep notation short, set $\psi_k := Q\phi_{n_k}$ for every $k \in \mathbb{N}^*$, and set $\psi := Q\phi$. We also introduce the functions $(\rho_j)_{j\geq 1}$ defined by

$$\rho_j(x) := \sup_{k\geq j} |\psi_k(x) - \psi(x)|, \quad x \in M,$$

and note that $\lim_{j\to\infty} \rho_j(x) = 0$ for every $x \in M$. For every $k \geq 1$,

$$\|\psi_k\|_\infty \leq \|\phi_{n_k}\|_\infty \leq 1 \quad \text{and} \quad \|\rho_k\|_\infty \leq \|\psi\|_\infty + \sup_{l\geq 1} \|\psi_l\|_\infty \leq \|\phi\|_\infty + 1,$$

so bounded convergence implies that

$$\lim_{k\to\infty} P\psi_k(x) = P\psi(x) \tag{5.2}$$

and

$$\lim_{j\to\infty} P\rho_j(x) = 0$$

for every $x \in M$. For every $m \in \mathbb{N}^*$,

$$\limsup_{j\to\infty} P\rho_j(y_{n_j}) \leq \limsup_{j\to\infty} P\rho_m(y_{n_j}) = P\rho_m(x)$$

because (ρ_j) is a nonincreasing sequence of nonnegative functions in $B(M)$, $\lim_{j\to\infty} y_{n_j} = x$, and P is strong Feller. Since the estimate above holds for every $m \in \mathbb{N}^*$ and since $\lim_{m\to\infty} P\rho_m(x) = 0$, it follows that

$$\lim_{j\to\infty} P\rho_j(y_{n_j}) = 0. \tag{5.3}$$

Consequently,

$$\begin{aligned}
&\limsup_{k\to\infty}\left(PQ\phi_{n_k}(x) - PQ\phi_{n_k}(y_{n_k})\right)\\
\leq &\limsup_{k\to\infty}|P\psi_k(x) - P\psi(x)|\\
&+\limsup_{k\to\infty}|P\psi(x) - P\psi(y_{n_k})| + \limsup_{k\to\infty}|P\psi(y_{n_k}) - P\psi_k(y_{n_k})|\\
\leq &\limsup_{k\to\infty} P\rho_k(y_{n_k}) = 0,
\end{aligned}$$

where we used (5.2), the assumption that P is strong Feller, and (5.3). This contradicts (5.1). **QED**

We are now ready to state and prove the main result of this subsection.

Proposition 5.25 *Let P be a Markov kernel on a separable metric space (M, d^*). If P is strong Feller, then it is also asymptotic strong Feller.*

Proof Consider the sequence of continuous metrics

$$d_k(x, y) := 1 \wedge (kd^*(x, y)), \quad k \in \mathbb{N}^*,$$

where $a \wedge b$ denotes the minimum of a and b. The sequence is clearly nondecreasing, and

$$\lim_{k\to\infty} d_k(x, y) = \delta(x, y), \quad \forall x, y \in M.$$

If P is strong Feller, then Proposition 5.22 implies that P^2 is ultra Feller. Therefore

$$0 = \inf_{\substack{x\in U\subset M,\\ U open}} \sup_{y\in U} \|\delta_x P^2 - \delta_y P^2\|_\delta. \tag{5.4}$$

Since $(d_k)_{k\geq 1}$ is nondecreasing and converges pointwise to δ, the sequence of functions $f_k(y) := \|\delta_x P^2 - \delta_y P^2\|_{d_k}$ is nondecreasing and dominated by $f(y) := \|\delta_x P^2 - \delta_y P^2\|_\delta$. Thus, for every open neighborhood U of x,

$$\limsup_{k\to\infty}\sup_{y\in U} f_k(y) \leq \sup_{y\in U}\lim_{k\to\infty} f_k(y) \leq \sup_{y\in U} f(y).$$

Together with (5.4) and $n_k := 2$ for all $k \geq 1$, this yields

$$\inf_{\substack{x\in U\subset M,\\ U open}} \limsup_{k\to\infty}\sup_{y\in U} \|\delta_x P^{n_k} - \delta_y P^{n_k}\|_{d_k} = 0.$$

QED

Remark 5.26 If P is a Markov kernel on a separable metric space such that P^n is strong Feller for some $n \in \mathbb{N}^*$, then P is asymptotic strong Feller. This follows if one replaces P^2 in the proof of Proposition 5.25 with P^{2n}.

The following exercise shows that the converse of Proposition 5.25 does not hold, i.e. there are Markov kernels which are asymptotic strong Feller but not strong Feller.

Exercise 5.27 Consider the mapping

$$F : \mathbb{R}^2 \to \mathbb{R}^2, \ (x_1, x_2) \mapsto (x_2, x_1).$$

For $(x, \theta) \in \mathbb{R}^2 \times \mathbb{R}$, set $F_\theta(x) := F(x) + \theta e_1$, where $e_1 := (1, 0)^\top$ (cf. part (**ii**) of Exercise 5.40 in Subsection 5.4.1). Let m be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that is absolutely continuous with respect to Lebesgue measure.

(i) Show that the Markov kernel P corresponding to the random dynamical system (F, m) is not strong Feller. *Hint:* Consider for instance the function $f(x_1, x_2) := \mathbf{1}_{x_2 \geq 0}$.

(ii) Use the result from Exercise **??** to show that P^2 is strong Feller, and conclude that P is asymptotic strong Feller.

5.3.2 Uniqueness of the invariant probability measure

The following theorem, first shown in [25], provides an important justification for introducing the asymptotic strong Feller property. It can be seen as a strengthening of Proposition 5.18, part (**i**), for Polish spaces.

Theorem 5.28 (Hairer, Mattingly) *Let (M, d^*) be a Polish space, i.e. a complete and separable metric space, and let P be a Markov kernel on $(M, \mathcal{B}(M))$. Let μ, ν be ergodic probability measures with respect to P. If P is asymptotic strong Feller at a point $x \in \mathsf{supp}(\mu) \cap \mathsf{supp}(\nu)$, then $\mu = \nu$. In particular, if P is asymptotic strong Feller, then two distinct ergodic probability measures have disjoint support.*

The proof of Theorem 5.28 requires several tools we yet need to introduce. We therefore postpone it to the end of this subsection. First, we define the important notions of coupling and lower semicontinuity. Let (X, e) be an arbitrary metric space and let $\mu, \nu \in \mathcal{P}(X)$. A *coupling* of μ and ν is a probability measure Γ on $(X^2, \mathcal{B}(X) \otimes \mathcal{B}(X))$ such that

$$\Gamma(A \times X) = \mu(A), \quad \Gamma(X \times A) = \nu(A), \quad \forall A \in \mathcal{B}(X).$$

We denote by $\mathcal{C}(\mu, \nu)$ the set of couplings of μ and ν.

Exercise 5.29 Assume in addition that X is separable and let $\mathcal{P}(X^2)$ be the set of Borel probability measures on X^2, endowed with the topology of weak convergence. Show that for every $\mu, \nu \in \mathcal{P}(X)$, $\mathcal{C}(\mu, \nu)$ is a closed subset of $\mathcal{P}(X^2)$.

Let $f : X \to \mathbb{R}$ be a function. We say that f is *lower semicontinuous* (respectively, *upper semicontinuous*) at a point $x_0 \in X$ if

$$f(x_0) \leq \liminf_{x \to x_0} f(x), \quad \text{resp. } f(x_0) \geq \limsup_{x \to x_0} f(x).$$

Clearly, f is continuous at a point $x_0 \in X$ if and only if f is both upper and lower semicontinuous at x_0. We call f lower semicontinuous (respectively, upper semicontinuous) if f is lower (upper) semicontinuous at every point $x_0 \in X$.

Exercise 5.30 Let $f : X \to [0, \infty)$ be a function.

(i) Show that
$$\tilde{f}(x) := \inf\{f(y) + e(x, y) : y \in X\}$$
defines a continuous function on X.

(ii) Show that f is lower semicontinuous if and only if there exists a non-decreasing sequence $(f_n)_{n \geq 1}$ of continuous functions from X to $[0, \infty)$ that converges pointwise to f. *Hint:* Consider the functions $f_n(x) := \inf\{f(y) + ne(x, y) : y \in X\}, n \in \mathbb{N}^*$, and use part (**i**).

The following statement, cited without proof here, can be found in [39] (see Particular Case 5.16 of Theorem 5.10 for the formula and Theorem 4.1 for existence of a minimizing coupling). It is an instance of the famous Kantorovich–Rubinstein duality theorem. The term *duality* refers to the asserted equivalence of a maximization and a minimization problem.

Theorem 5.31 *Let (M, d^*) be a Polish space and let d be a bounded metric on M that is lower semicontinuous as a function from the product metric space $(M \times M, d^* \star d^*)$ to $[0, \infty)$. Then, for every $\mu, \nu \in \mathcal{P}(M)$, we have*

$$\|\mu - \nu\|_d = \inf_{\Gamma \in \mathcal{C}(\mu,\nu)} \int_{M^2} d(x, y)\ \Gamma(dx, dy)$$

and the infimum on the right is attained.

Lemma 5.32 *Let (M, d^*) be a Polish space, let $(d_n)_{n\geq 1}$ be a nondecreasing sequence of continuous metrics on M, and let d be a bounded metric on M such that*

$$\lim_{n\to\infty} d_n(x, y) = d(x, y), \quad \forall x, y \in M.$$

Then, for every $\mu, \nu \in \mathcal{P}(M)$, we have

$$\lim_{n\to\infty} \|\mu - \nu\|_{d_n} = \|\mu - \nu\|_d.$$

Proof Let $\mu, \nu \in \mathcal{P}(M)$. Since $(d_n)_{n\geq 1}$ is nondecreasing and since d is bounded, we have

$$\|\mu - \nu\|_{d_n} \leq \|\mu - \nu\|_{d_{n+1}} \leq \|\mu - \nu\|_d < \infty, \quad \forall n \in \mathbb{N}^*.$$

Therefore,

$$l := \lim_{n\to\infty} \|\mu - \nu\|_{d_n}$$

exists and is less than or equal to $\|\mu - \nu\|_d$. By Theorem 5.31, there are couplings $(\Gamma_n)_{n\geq 1}$ of μ and ν such that

$$\|\mu - \nu\|_{d_n} = \int_{M^2} d_n(x, y)\ \Gamma_n(dx, dy), \quad \forall n \in \mathbb{N}^*.$$

Since μ and ν are Borel probability measures on a Polish space, they are tight, i.e. for every $\varepsilon > 0$ there is a compact set $K \subset M$ such that $\mu(K), \nu(K) > 1-\varepsilon$ (see, e.g., Theorem 7.1.4 in [13]). Hence, by Exercise 5.33 below, the family of couplings $(\Gamma_n)_{n\geq 1}$ is tight as well. By Prohorov's theorem, $(\Gamma_n)_{n\geq 1}$ admits a subsequence that converges weakly to a probability measure $\Gamma_\infty \in \mathcal{P}(M^2)$. And by Exercise 5.29, $\Gamma_\infty \in \mathcal{C}(\mu, \nu)$. For simplicity, we denote the convergent subsequence again by $(\Gamma_n)_{n\geq 1}$. For $n \leq m$, we have

$$\int_{M^2} d_n(x, y)\ \Gamma_m(dx, dy) \leq \int_{M^2} d_m(x, y)\ \Gamma_m(dx, dy) = \|\mu - \nu\|_{d_m} \leq l.$$

Since each d_n is continuous and bounded, and since Γ_m converges weakly to Γ_∞, we have

$$\lim_{m\to\infty}\int_{M^2} d_n(x,y)\ \Gamma_m(dx,dy) = \int_{M^2} d_n(x,y)\ \Gamma_\infty(dx,dy).$$

Thus,

$$\int_{M^2} d_n(x,y)\ \Gamma_\infty(dx,dy) \le l.$$

By monotone convergence,

$$l \ge \int_{M^2} d(x,y)\ \Gamma_\infty(dx,dy) \ge \inf_{\Gamma\in\mathcal{C}(\mu,\nu)}\int_{M^2} d(x,y)\ \Gamma(dx,dy). \tag{5.5}$$

Since d is the pointwise limit of a nondecreasing sequence of continuous functions, Exercise 5.30 implies that d is lower semicontinuous. Hence, by virtue of Theorem 5.31, the expression on the right side of (5.5) equals $\|\mu-\nu\|_d$. We have thus shown that $l \ge \|\mu-\nu\|_d$, and together with $l \le \|\mu-\nu\|_d$ one obtains $\lim_{n\to\infty}\|\mu-\nu\|_{d_n} = \|\mu-\nu\|_d$. **QED**

Exercise 5.33 Let (X,e) be a metric space and let $\mu,\nu\in\mathcal{P}(X)$ be tight. Show that $\mathcal{C}(\mu,\nu)\subset\mathcal{P}(X^2)$ is a tight family of probability measures.

Lemma 5.34 *Let (M,d^*) be a separable metric space, let P be a Markov kernel on $(M,\mathcal{B}(M))$, and let d be a metric on M that is bounded by 1. Assume further that there are $\epsilon>0$ and $U\in\mathcal{B}(M)$ such that*

$$\sup_{x,y\in U}\|\delta_x P - \delta_y P\|_d \le \epsilon.$$

Let $\mu,\nu\in\mathcal{P}(M)$ and set $\alpha := \mu(U)\wedge\nu(U)$. Then

$$\|\mu P - \nu P\|_d \le 1-\alpha(1-\epsilon).$$

Proof Since d is bounded by 1, we have $\|\mu P-\nu P\|_d\le 1$, so the assertion holds if $\alpha=0$. If $\alpha>0$, define for $A\in\mathcal{B}(M)$ the Borel probability measures

$$\mu^U(A) := \frac{\mu(A\cap U)}{\mu(U)}, \qquad \nu^U(A) := \frac{\nu(A\cap U)}{\nu(U)},$$
$$\bar\mu(A) := \frac{\mu(A)-\alpha\mu^U(A)}{1-\alpha}, \qquad \bar\nu(A) := \frac{\nu(A)-\alpha\nu^U(A)}{1-\alpha},$$

and observe that

$$\begin{aligned}\mu =&(1-\alpha)\bar{\mu} + \alpha\mu^U,\\ \nu =&(1-\alpha)\bar{\nu} + \alpha\nu^U.\end{aligned}$$

Exercise 5.35 below and the fact that $\mu^U(U^c) = \nu^U(U^c) = 0$ yield

$$\|\mu^U P - \nu^U P\|_d \le \int_{U^2} \|\delta_x P - \delta_y P\|_d \; \mu^U(dx)\; \nu^U(dy).$$

By assumption, the expression on the right is bounded from above by ϵ. The triangle inequality for $\|\cdot\|_d$ then implies

$$\|\mu P - \nu P\|_d \le (1-\alpha)\|\bar{\mu}P - \bar{\nu}P\|_d + \alpha\|\mu^U P - \nu^U P\|_d \le 1 - \alpha + \alpha\epsilon.$$

QED

Exercise 5.35 Let (M, d^*) be a separable metric space, let P be a Markov kernel on $(M, \mathcal{B}(M))$, and let d be a bounded metric on M. Show that

$$\|\mu P - \nu P\|_d \le \int_{M^2} \|\delta_x P - \delta_y P\|_d \; \mu(dx)\nu(dy), \quad \forall \mu, \nu \in \mathcal{P}(M).$$

We are now ready to prove Theorem 5.28.

Proof [of Theorem 5.28] Let $x \in \mathsf{supp}(\mu) \cap \mathsf{supp}(\nu)$ such that P is asymptotic strong Feller at x. Then there exist a nondecreasing sequence $(n_k)_{k\ge1}$ of positive integers as well as a nondecreasing sequence $(d_k)_{k\ge1}$ of continuous metrics on M such that $\lim_{k\to\infty} d_k(y,z) = \delta(y,z)$, $y, z \in M$, and

$$\inf_{\substack{x\in U\subset M,\\ U open}} \limsup_{k\to\infty} \sup_{y\in U} \|\delta_x P^{n_k} - \delta_y P^{n_k}\|_{d_k} = 0.$$

Let U be an open neighborhood of x and let $K \in \mathbb{N}$ such that

$$\sup_{y\in U} \|\delta_x P^{n_k} - \delta_y P^{n_k}\|_{d_k} < \frac{1}{4}, \quad \forall k \ge K.$$

Since $\|\cdot\|_d$ satisfies the triangle inequality for every metric d on (M, d^*), we have

$$\sup_{y,z\in U} \|\delta_y P^{n_k} - \delta_z P^{n_k}\|_{d_k} < \frac{1}{2}, \quad \forall k \ge K.$$

Set $\alpha := \mu(U) \wedge \nu(U)$. Lemma 5.34 implies

$$\|\mu P^{n_k} - \nu P^{n_k}\|_{d_k} \leq 1 - \frac{\alpha}{2}, \quad \forall k \geq K.$$

Since μ and ν are invariant probability measures,

$$\|\mu - \nu\|_{d_k} \leq 1 - \frac{\alpha}{2}, \quad \forall k \geq K.$$

As

$$\lim_{k\to\infty} \|\mu - \nu\|_{d_k} = \|\mu - \nu\|_\delta$$

by Lemma 5.32, it follows that

$$\|\mu - \nu\|_\delta \leq 1 - \frac{\alpha}{2}.$$

Since $x \in \mathsf{supp}(\mu) \cap \mathsf{supp}(\nu)$, we have $\alpha > 0$, so $\|\mu - \nu\|_\delta < 1$. In particular, for every $A \in \mathcal{B}(M)$,

$$2 > |\mu(\mathbf{1}_A - \mathbf{1}_{A^c}) - \nu(\mathbf{1}_A - \mathbf{1}_{A^c})| = 2|\mu(A) - \nu(A)|$$

in view of Remark 5.20. This implies that μ and ν are not mutually singular. Since μ and ν are ergodic, it follows from part (**ii**) of Proposition 4.29 that $\mu = \nu$. **QED**

5.4 Petite sets, small sets and Doeblin points

We call a measurable set C a *petite set* if there exist $a \in (0,1)$ and some nonzero Borel measure ξ on M such that

$$R_a(x, A) \geq \xi(A)$$

for all $x \in C$ and $A \in \mathcal{B}(M)$. We call the set C a *small set* if there is a nonzero Borel measure ξ on M such that

$$P(x, A) \geq \xi(A)$$

for all $x \in C$ and $A \in \mathcal{B}(M)$. Clearly, every small set is petite.

Remark 5.36 In the terminology of Meyn and Tweedie [31] (Chapter 5), a ν_α*-petite set* for a probability measure α on $\mathbb{N}$ is a set $C \in \mathcal{B}(M)$ such that

$$\sum_{n=0}^{\infty} \alpha(n) P^n(x, A) \geq \nu_\alpha(A), \quad \forall x \in C,\ A \in \mathcal{B}(M),$$

where ν_α is some nonzero Borel measure on M. A ν_m*-small set* for $m \in \mathbb{N}^*$ is a set $C \in \mathcal{B}(M)$ such that

$$P^m(x, A) \geq \nu_m(A), \quad \forall x \in C,\ A \in \mathcal{B}(M),$$

where ν_m is a nonzero Borel measure on M. With these definitions, the class of petite sets defined above is equal to the class of sets that are ν_{Δ_a}-petite for some $a \in (0,1)$, where

$$\Delta_a(k) := a^k(1-a),\ k \in \mathbb{N}.$$

Our notion of a small set corresponds to the notion of a ν_1-small set.

We call a point $x^* \in M$ a *weak Doeblin point* (respectively a *Doeblin point*) if x^* has a neighborhood that is a petite set (respectively a small set).

The importance and usefulness of these notions will be highlighted in Chapters 6 and 7. The following proposition extends Example 5.3.

Proposition 5.37 *Assume that there exists an accessible weak Doeblin point for P. Then P is ξ-irreducible.*

Proof By assumption, there exists an open set C and a non trivial measure ξ such that $C \cap \Gamma \neq \emptyset$ and $R_a(x, .) \geq \xi(.)$ for all $x \in C$. Let $p_k = \sum_{i=0}^{k}(1-a)^2 a^i a^{k-i} = (k+1)(1-a)^2 a^k$. Then, for all A measurable and $x \in M$

$$\sum_{k \geq 0} p_k P^k(x, A) = R_a^2(x, A) = \int R_a(x, dy) R_a(y, A) \geq R_a(x, C)\xi(A).$$

By accessibility $R_a(x, C) > 0$. **QED**

5.4.1 Doeblin points for random dynamical systems

Let Θ be a nonempty open subset of $\mathbb{R}^d$ for some $d \in \mathbb{N}^*$, let $\mathcal{A}$ be the Borel σ-field on Θ, and let m be a probability measure on $(\Theta, \mathcal{A})$. For $n \in \mathbb{N}^*$, the n-fold product measure $m \otimes \ldots \otimes m$ will be denoted by m^n. Let $k \in \mathbb{N}^*$ and let M be a nonempty open subset of $\mathbb{R}^k$, with Borel σ-field $\mathcal{B}(M)$. Let $F : \Theta \times M \to M$ be a C^1-mapping. Recall from Chapter **??** that the pair (F, m) induces a random dynamical system with associated Markov kernel

$$P(x, G) = m(\{\theta \in \Theta : \ F_\theta(x) \in G\}), \quad (x, G) \in M \times \mathcal{B}(M).$$

For $n \in \mathbb{N}^*$ and $x \in M$, let

$$\varphi_{n,x} : \Theta^n \to M, \ (\theta_1, \ldots, \theta_n) \mapsto (F_{\theta_n} \circ \ldots \circ F_{\theta_1})(x).$$

The following proposition is essentially Lemma 6.3 in [6].

Proposition 5.38 *Let $x^* \in M$, $n \in \mathbb{N}^*$, and $\theta^* = (\theta_1^*, \ldots, \theta_n^*) \in \Theta^n$ such that the following conditions hold.*

(a) *The Jacobian matrix $D\varphi_{n,x^*}(\theta)|_{\theta=\theta^*}$ has rank k;*

(b) *There is a neighborhood $V \subset \Theta^n$ of θ^* such that $m^n(\cdot \cap V)$ is absolutely continuous with respect to $\lambda^{nd}(\cdot \cap V)$, where λ^{nd} is the Lebesgue measure on $\mathbb{R}^{nd}$. The corresponding probability density function ρ has a representative $\hat{\rho}$ such that*

$$c := \inf_{\theta \in V} \hat{\rho}(\theta) > 0.$$

Under these conditions, x^* is a Doeblin point with respect to the Markov kernel P^n, and in particular a weak Doeblin point with respect to P.

Proof Since $D\varphi_{n,x^*}(\theta)|_{\theta=\theta^*}$ is a $(k \times nd)$-matrix of rank k, we have either $k = nd$ or $k < nd$. To avoid repeating ourselves, we will only prove the slightly more complicated case $k < nd$. The case $k = nd$ can be easily derived by making small modifications to the proof for $k < nd$. Assume without loss of generality that the first k columns of $D\varphi_{n,x^*}(\theta)|_{\theta=\theta^*}$ are linearly independent. We will often write points $\theta \in \Theta^n$ as $\theta = (\theta^{(k)}, \theta^{(nd-k)})$, where $\theta^{(k)} \in \mathbb{R}^k$ is the vector consisting of the first k components of θ, and where $\theta^{(nd-k)}$ is the vector of the remaining $(nd-k)$ components. For $x \in M$, consider the C^1 mapping

$$G_x : \Theta^n \to M \times \mathbb{R}^{nd-k},\ \theta = (\theta^{(k)}, \theta^{(nd-k)}) \mapsto (\varphi_{n,x}(\theta), \theta^{(nd-k)}).$$

We also define the C^1 mapping

$$H : \Theta^n \times M \to M \times \mathbb{R}^{nd-k} \times M,\ (\theta, x) \mapsto (G_x(\theta), x).$$

Since

$$\begin{aligned} \det DH(\theta, x)|_{\theta=\theta^*, x=x^*} &= \det DG_{x^*}(\theta)|_{\theta=\theta^*} \\ &= \det D_{\theta^{(k)}} \varphi_{n,x^*}(\theta^{(k)}, (\theta^*)^{(nd-k)})|_{\theta^{(k)}=(\theta^*)^{(k)}} \neq 0, \end{aligned}$$

the inverse function theorem implies that there is an open neighborhood W of (θ^*, x^*) such that the restriction of H to W, denoted by H_W, is a C^1 diffeomorphism. By intersecting W with an open subset of $V \times M$ that contains (θ^*, x^*) and calling the resulting set W again, we may assume without loss of generality that $\theta \in V$ for every $(\theta, x) \in W$. The set $H(W)$ is a neighborhood of $H(\theta^*, x^*) = (\varphi_{n,x^*}(\theta^*), (\theta^*)^{(nd-k)}, x^*)$, so there are open neighborhoods Z_0 of $\varphi_{n,x^*}(\theta^*)$, T_0 of $(\theta^*)^{(nd-k)}$, and U_0 of x^* such that $Z_0 \times T_0 \times U_0 \subset H(W)$. Let $W_0 := H_W^{-1}(Z_0 \times T_0 \times U_0)$. For any $x \in U_0$, set

$$V_x := \{\theta \in \Theta^n :\ (\theta, x) \in W_0\}.$$

It is straightforward to check that for every $x \in U_0$, the restriction of G_x to V_x is a C^1 diffeomorphism that satisfies $G_x(V_x) = Z_0 \times T_0$.

Let $x \in U_0$ and $A \in \mathcal{B}(M)$. We have

$$P^n(x, A) \geq P^n(x, A \cap Z_0) = \int_{\varphi_{n,x}^{-1}(A \cap Z_0)} m^n(d\theta).$$

Since $G_x^{-1}((A \cap Z_0) \times T_0) \subset \varphi_{n,x}^{-1}(A \cap Z_0)$, the expression on the right is bounded from below by

$$\int_{G_x^{-1}((A \cap Z_0) \times T_0)} m^n(d\theta) \geq \int_{V_x \cap G_x^{-1}((A \cap Z_0) \times T_0)} m^n(d\theta).$$

As $V_x \subset V$, the integral on the right equals

$$\int_{V_x \cap G_x^{-1}((A \cap Z_0) \times T_0)} \hat{\rho}(\theta)\ \lambda^{nd}(d\theta) \geq c \int_{V_x \cap G_x^{-1}((A \cap Z_0) \times T_0)} \lambda^{nd}(d\theta).$$

There is no loss of generality in assuming that V and U_0 are each contained in a compact set. Since the mapping $(\theta, x) \mapsto \det DG_x(\theta)$ is continuous, we have

$$\hat{c} := \sup_{\theta \in V, x \in U_0} |\det DG_x(\theta)| < \infty.$$

Hence,

$$P^n(x, A) \geq \frac{c}{\hat{c}} \int_{V_x \cap G_x^{-1}((A \cap Z_0) \times T_0)} |\det DG_x(\theta)| \ \lambda^{nd}(d\theta).$$

Since the restriction of G_x to V_x is a diffeomorphism, the change of variables formula (see for instance Theorem 2.47 in [17]) implies that the expression on the right equals

$$\frac{c}{\hat{c}} \lambda^{nd-k}(T_0) \lambda^k(A \cap Z_0).$$

The measure $\xi(A) := \frac{c}{\hat{c}} \lambda^{nd-k}(T_0) \lambda^k(A \cap Z_0)$ on $(M, \mathcal{B}(M))$ is nontrivial and does not depend on $x \in U_0$, so U_0 is a small set with respect to the kernel P^n. As U_0 is a neighborhood of x^*, the point x^* is a Doeblin point with respect to P^n. **QED**

Example 5.39 (additive noise) *Recall the setting of Exercise* **??**: *We have* $M = \Theta = \mathbb{R}^k$, $F : M \to M$, $F_\theta(x) := F(x) + \theta$ *for* $(\theta, x) \in \Theta \times M$, *and* $m(d\theta) = h(\theta)\, d\theta$, *where* $h \in L^1(d\theta)$. *Assume in addition that* F *is* C^1, *which implies that* $(\theta, x) \mapsto F_\theta(x)$ *is* C^1 *as well. Finally, suppose that there are a nonempty open set* $V \subset \Theta$ *and a representative* $\hat{h}$ *of* h *such that*

$$\inf_{\theta \in V} \hat{h}(\theta) > 0.$$

For every $x^* \in M$ *and* $\theta^* \in \Theta$,

$$D\varphi_{1,x^*}(\theta)|_{\theta=\theta^*} = \mathbf{1}_{k \times k},$$

where $\mathbf{1}_{k \times k}$ *is the identity matrix of dimensions* $(k \times k)$. *Since* $\mathbf{1}_{k \times k}$ *has rank* k, *every pair* $(x^*, \theta^*) \in M \times V$ *satisfies the conditions of Proposition 5.38. Hence, every point* $x^* \in M$ *is a Doeblin point with respect to the Markov kernel* $P(x, G) = m(\{\theta \in \Theta : \ F_\theta(x) \in G\})$.

Exercise 5.40 [degenerate additive noise] Let m be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, with probability density function h. Assume further that there are a nonempty open interval $I \subset \mathbb{R}$ and a representative $\hat{h}$ of h such that

$$\inf_{\theta \in I} \hat{h}(\theta) > 0.$$

Show the following statements.

(i) Let $F : \mathbb{R}^2 \to \mathbb{R}^2, F = (F_1, F_2)^\top$ be a C^1 map and let $(x^*, \theta_1^*) \in \mathbb{R}^2 \times I$ such that

$$\partial_{x_1} F_2(F(x^*) + \theta_1^* e_1) \neq 0,$$

where $e_1 = (1,0)^\top$. Set

$$F_\theta(x) := F(x) + \theta e_1, \quad (x, \theta) \in \mathbb{R}^2 \times \mathbb{R}.$$

Then x^* is a weak Doeblin point for the Markov kernel associated with (F, m).

(ii) Let $k \geq 2$, and let $F : \mathbb{R}^k \to \mathbb{R}^k$ be defined by $F(x_1, \ldots, x_k) := (x_k, x_1, x_2, \ldots, x_{k-1})^\top$. Set

$$F_\theta(x) := F(x) + \theta e_1, \quad (x, \theta) \in \mathbb{R}^k \times \mathbb{R},$$

where $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^k$. Then any point $x^* \in \mathbb{R}^k$ is a weak Doeblin point for the Markov kernel associated with (F, m).

5.4.2 Piecewise deterministic Markov processes

Let $N \in \mathbb{N}^*$ and let μ be a probability measure on the finite set $E := \{1, \ldots, N\}$. Let $\nu_1, \ldots, \nu_N$ be Borel probability measures on $\mathbb{R}_+ := (0, \infty)$. On the set $\Theta := \mathbb{R}_+ \times E$, we define the probability measure

$$m(T \times \{i\}) := \mu(i)\nu_i(T), \quad T \in \mathcal{B}(\mathbb{R}_+),\ i \in E.$$

Let M be a nonempty open subset of $\mathbb{R}^k$ for some $k \in \mathbb{N}^*$, equipped with the Euclidean metric and the corresponding Borel σ-field. On M, we consider a family of vector fields $(G_i)_{i \in E}$ that are C^∞ smooth and forward complete, i.e. for every $i \in E$, $G_i : M \to \mathbb{R}^k$ is C^∞, and for every $x_0 \in M$, the initial-value problem

$$\begin{aligned} \frac{dx}{dt} =& G_i(x(t)), \quad t > 0, \\ x(0) =& x_0 \end{aligned}$$

has a unique solution, denoted by $t \mapsto \Phi_i(x_0, t)$, that is defined for all $t \geq 0$ and satisfies $\Phi_i(x_0, t) \in M$ for all $t \geq 0$. For every $\theta = (t, i) \in \Theta$, define

$$F_\theta : M \to M, \ x \mapsto \Phi_i(x, t).$$

Smoothness of the vector fields $(G_i)_{i \in E}$ implies that the map $(t, x) \mapsto F_{(t,i)}(x)$ is C^∞ smooth for every $i \in E$, see for instance Theorem 17.9 in [**?**].

Exercise 5.41 Show that the pair (F, m) gives rise to a random dynamical system, i.e. verify that m is indeed a probability measure and that $(\theta, x) \mapsto F_\theta(x)$ is measurable.

In words, the Markov chain $(X_n)_{n \in \mathbb{N}}$ on M induced by the random dynamical system (F, m) can be described as follows: Pick an index $i \in E$ and a time $t \in \mathbb{R}_+$ according to the distribution m. Starting at $X_0 \in M$, flow along the vector field G_i for time t. The point $X_1 \in M$ is the point reached at time t. Then, pick $j \in E$ and $s \in \mathbb{R}_+$, according to m and independent of i and t, and flow along G_j for time s, leading to the point $X_2 \in M$, etc. The following exercise gives a concrete example of such a Markov chain.

Exercise 5.42 Let $\mu_1, \mu_2, \mu_3 \geq 0$ such that $\mu_1 + \mu_2 + \mu_3 = 1$, and let $\lambda_1, \lambda_2, \lambda_3, \alpha_1, \alpha_2, \alpha_3 > 0$. On $\Theta := \mathbb{R}_+ \times \{1, 2, 3\}$, define the probability measure m by

$$m(T \times \{i\}) := \mu_i \lambda_i \int_T e^{-\lambda_i t} \, dt, \quad T \in \mathcal{B}(\mathbb{R}_+), \ i \in \{1, 2, 3\}.$$

For $\theta = (t, i) \in \Theta$, let

$$F_\theta : \mathbb{R} \to \mathbb{R}, \ x \mapsto \begin{cases} x + \alpha_1 t, & i = 1, \\ x - \alpha_2 t, & i = 2, \\ x e^{-\alpha_3 t}, & i = 3. \end{cases}$$

Prove the following statements.

(i) If $\mu_3 > 0$, then the Markov kernel P associated with (F, m) admits a unique invariant probability measure. *Hint*: Use Theorem 4.31 on random contractions.

Consider the function

$$f(t, i) := (-1)^{i+1} \alpha_i t, \quad (t, i) \in \Theta$$

and the Borel measure

$$a(A) := m\left(\{\theta \in \Theta : f(\theta) \in A\}\right), \quad A \in \mathcal{B}(\mathbb{R}).$$

(ii) Show that if $\mu_3 = 0$, then the Markov kernel P satisfies

$$P(x, G) = a(\{\xi \in \mathbb{R} : x + \xi \in G\}), \quad x \in \mathbb{R}, G \in \mathcal{B}(\mathbb{R}).$$

(iii) Deduce from part (**ii**) that if $\mu_3 = 0$ and $\mu_1\alpha_1/\lambda_1 \neq \mu_2\alpha_2/\lambda_2$, then P does not admit any invariant probability measures. *Hint*: See Example 4.19.

The main result in this section is a sufficient condition for the existence of a weak Doeblin point with respect to the Markov kernel P induced by (F, m). This condition will be formulated in terms of the Lie algebra generated by $(G_i)_{i\in E}$. The *Lie bracket* of two C^1 vector fields G and H on a nonempty open subset M of $\mathbb{R}^k$ is itself a vector field on M, defined as

$$[G, H](x) := DH(x)G(x) - DG(x)H(x), \quad x \in M.$$

Here, $DG(x)$ and $DH(x)$ denote the Jacobian matrices of G and H, respectively, evaluated at the point x. The products $DH(x)G(x)$ and $DG(x)H(x)$ are to be understood as matrix-vector products.

Exercise 5.43 [Properties of Lie brackets]

(i) Show that the Lie bracket $[\cdot, \cdot]$ is bilinear and antisymmetric, i.e. for any C^1 vector fields A, B, C and for any $\lambda \in \mathbb{R}$, one has

$$[\lambda A, C] = \lambda[A, C], \quad [A + B, C] = [A, C] + [B, C], \quad [A, B] = -[B, A].$$

Why is this enough to deduce linearity for the second argument?

(ii) To a vector field A on M, one can associate the operator on $C^\infty(M, \mathbb{R})$ that maps $f \in C^\infty(M, \mathbb{R})$ to $x \mapsto \langle A(x), \nabla f(x)\rangle$. Here, $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{R}^k$ and ∇f denotes the gradient of f. This operator is usually identified with A, so one writes Af for the image of f under the operator. Let A and B be C^2 vector fields on M. Show that

$$[A, B] = AB - BA,$$

where AB and BA should be interpreted as compositions of the operators A and B.

(iii) Use the result from (**ii**) to prove the *Jacobi identity*: For C^3 vector fields A, B, C, one has

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0.$$

We inductively define a sequence of families of vector fields by $\mathrm{G}_0 := \{G_i\}_{i\in E}$ and $\mathrm{G}_{n+1} := \mathrm{G}_n \cup \{[G_i, V] : i \in E, V \in \mathrm{G}_n\}$ for $n \in \mathbb{N}$. Recall that the *linear span* of a set S contained in some vector space is the set of all (finite) linear combinations of elements in S. We say that the *weak bracket condition* holds at a point $x \in M$ if the linear span of $\{V(x) : V \in \cup_{n\in\mathbb{N}} \mathrm{G}_n\}$ is equal to the full space $\mathbb{R}^k$. As alluded to earlier, this condition admits an alternative formulation in terms of the Lie algebra generated by $(G_i)_{i\in E}$. The latter is defined as the smallest linear subspace $\mathcal{L}$ of the vector space of C^∞ vector fields on M that is closed under Lie brackets ($[G,H] \in \mathcal{L}$ for all $G, H \in \mathcal{L}$) and contains $(G_i)_{i\in E}$.

Exercise 5.44 Let $\mathcal{L}$ denote the Lie algebra generated by $(G_i)_{i\in E}$.

(i) Show that $\mathrm{G}_n \subset \mathcal{L}$ for all $n \in \mathbb{N}$.

(ii) Deduce from (**i**) that the weak bracket condition at a point x implies that $\{V(x) : V \in \mathcal{L}\} = \mathbb{R}^k$.

(iii) Show that $\mathcal{G}$, the linear span of $\bigcup_{n\in\mathbb{N}} \mathrm{G}_n$, is closed under Lie brackets. *Hint*: This will follow once it is shown that for every $n \in \mathbb{N}$, $A \in \mathrm{G}_n$, and $B \in \mathcal{G}$, one has $[A,B] \in \mathcal{G}$. The Jacobi identity from Exercise 5.43 may be helpful.

(iv) Conclude that the weak bracket condition holds at a point $x \in M$ if and only if $\{V(x) : V \in \mathcal{L}\} = \mathbb{R}^k$.

Theorem 5.45 *If the weak bracket condition holds at a point $x^* \in M$, then there is $n \in \mathbb{N}$ such that x^* is a Doeblin point with respect to P^n. In particular, x^* is then a weak Doeblin point with respect to P.*

The proof of Theorem 5.45 relies on a slight generalization of Proposition 5.38. To state this generalization, let T be a nonempty open subset of $\mathbb{R}^d$, $d \in \mathbb{N}^*$, and let E be a finite set. Let m be a probability measure on $\Theta := T \times E$, equipped with the product σ-field of $\mathcal{B}(T)$ and the power set of E. As in Section 5.4.1, the n-fold product measure $m \otimes \ldots \otimes m$ will be denoted by m^n. Let M be a nonempty open subset of $\mathbb{R}^k$, $k \in \mathbb{N}^*$, with Borel σ-field $\mathcal{B}(M)$. Let $F : \Theta \times M \to M$ be a map such that for every $i \in E$, $(t,x) \mapsto F_{(t,i)}(x)$ is C^1. For $n \in \mathbb{N}^*$, $\mathbf{i} = (i_1, \ldots, i_n) \in E^n$, and $x \in M$, let

$$\varphi^{\mathbf{i}}_{n,x} : T^n \to M,\ (t_1, \ldots, t_n) \mapsto (F_{(t_n,i_n)} \circ \times \circ F_{(t_1,i_1)})(x).$$

Proposition 5.46 *Let $x^* \in M$, $n \in \mathbb{N}^*$, and $\mathbf{t}^* = (t_1^*, \ldots, t_n^*) \in T^n$ such that the following conditions hold.*

1. *There is $\mathbf{i} \in E^n$ such that the Jacobian matrix $D\varphi^{\mathbf{i}}_{n,x^*}(\mathbf{t})|_{\mathbf{t}=\mathbf{t}^*}$ has rank k;*
2. *There is a neighborhood $V \subset T^n$ of $\mathbf{t}^*$ such that $m^n((\cdot \cap V) \times \{\mathbf{i}\})$ is absolutely continuous with respect to $\lambda^{nd}(\cdot \cap V)$. The corresponding probability density function ρ has a representative $\hat{\rho}$ such that*

$$c := \inf_{\mathbf{t} \in V} \hat{\rho}(\mathbf{t}) > 0.$$

Under these conditions, x^* is a Doeblin point with respect to P^n, and in particular a weak Doeblin point with respect to P.

Exercise 5.47 Prove Proposition 5.46. *Hint:* The proof of Proposition 5.38 can almost be repeated verbatim.

With Proposition 5.46 in hand, the proof of Theorem 5.45 reduces to checking conditions (**i**) and (**ii**) of the proposition. While condition (**ii**) follows almost immediately from the definition of m, establishing condition (**i**) requires a link between the weak bracket condition and the full-rank condition on the Jacobian matrix of $\varphi^{\mathbf{i}}_{n,x}$. This link is provided by a result from geometric control theory that is due to Sussmann and Jurdjevic.

Proof [of Theorem 5.45] Let $x^* \in M$ be a point where the weak bracket condition holds. The setting of piecewise deterministic Markov processes introduced at the beginning of this section is clearly covered by the more general setting of Proposition 5.46, with $T = \mathbb{R}_+$. By part (1) of Theorem 5 in BH, there are $n > k$, $\mathbf{i} \in E^n$, and $\mathbf{t}^* = (t_1^*, \ldots, t_n^*) \in \mathbb{R}^n_+$ such that $D\varphi^{\mathbf{i}}_{n,x^*}(\mathbf{t})|_{\mathbf{t}=\mathbf{t}^*}$ has rank k.

Let $V := (0,R)^n$ for some $R > 0$ so large that $\mathbf{t}^* \in V$. For Borel sets $A_1, \ldots, A_n \subset (0,R)$ and $A := A_1 \times \ldots \times A_n$, we have

$$m^n(A \times \{\mathbf{i}\}) = \prod_{l=1}^n m(A_l \times \{i_l\}) = \prod_{l=1}^n \mu(i_l)\lambda_{i_l} \int_{A_l} e^{-\lambda_{i_l} t}\, dt = \int_A \rho(\mathbf{t})\, d\mathbf{t},$$

where

$$\rho(\mathbf{t}) := \prod_{l=1}^{n} \mu(i_l)\lambda_{i_l} e^{-\lambda_{i_l} t_l}$$

and thus $\inf_{\mathbf{t}\in V} \rho(\mathbf{t}) > 0$. The theorem then follows with the help of Proposition 5.46. **QED**

Chapter 6

Harris and Positive Recurrence

6.1 Stability and positive recurrence

Let (X_n) denote a Markov chain (defined on $(\Omega, \mathcal{A}, \mathbb{F}, \mathsf{P})$) on M with kernel P. Recall that we let

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_k}$$

denote its *empirical occupation measure.*

If there exists $\pi \in \mathcal{P}(M)$ such that, **for every initial distribution** μ (i.e law of X_0), (ν_n) converge weakly to π (i.e $\mathbb{P}_\mu(\nu_n \Rightarrow \pi) = 1$) the kernel P (respectively the chain (X_n)) is called *stable* and π is called the *stationary distribution* of P (respectively (X_n)). If furthermore $\mathbb{P}_\mu(\nu_n f \to \pi f) = 1$ for **every bounded measurable function** $f : M \mapsto \mathbb{R}$ then P (respectively (X_n)) is called *positively recurrent.*

Recall that P is said *uniquely ergodic* if $\mathsf{Inv}(P)$ has cardinal one. The following proposition easily follows from the definitions and Proposition 4.5.

Proposition 6.1 *If P is positively recurrent with stationary distribution π, then it is stable, uniquely ergodic and $\mathsf{Inv}(P) = \{\pi\}$. If P is Feller and stable with stationary distribution π, then it is uniquely ergodic and $\mathsf{Inv}(P) = \{\pi\}$.*

Remark 6.2 A Feller stable Markov chain is not necessarily positively recurrent. For instance, let $X_n \in \mathbb{R}$ be recursively defined as

$$X_{n+1} = \frac{1}{2} X_n + \xi_{n+1}$$

where (ξ_n) are independent uniformly random variables taking values in $\{-1,1\}$. Then it is not hard to show that (X_n) is Feller stable and that π, its stationary distribution, is the uniform distribution over $[-2,2]$. On the other hand, for $X_0 = 0$, $X_n \in D = \{\sum_{k=0}^m 2^{-k}\theta_k \ : \theta_k \in \{-1,1\}, m \in \mathbb{N}\}$ so that $\nu_n(D) = 1$ while $\pi(D) = 0$.

Another example (borrowed from [14]) is the following. Let P be the Kernel on $[0,\infty[$ defined by $P(0,0) = 1$ and for $x > 0$, $P(x,0) = 1 - P(x,x/2) = 2^{-x}$. This kernel is δ_0-irreducible, Feller and admits δ_0 as (unique) invariant probability. It is stable (since $X_n \leq \frac{X_0}{2^n}$) but is not positively recurrent, because the probability that X_n never touches 0 is positive.

.

Exercise 6.3 Let (X_n) be the deterministic system on $S^1 = \mathbb{R}/\mathbb{Z}$ defined by $X_{n+1} = (X_n + \alpha) \mod 1$ where $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Show that (X_n) is stable but not positively recurrent.

Proposition 6.4 **(i)** *Suppose that P is Feller, uniquely ergodic and that (for every initial distribution) $\{\nu_n\}$ is almost surely tight. Then P is stable.*

(ii) *Suppose that P is strongly Feller and stable. Then P is positively recurrent.*

Proof The first part follows from Theorem 4.20. If P is strongly Feller, then for every bounded measurable f, Pf is continuous so that, $\nu_n(Pf) - \pi(Pf) \to 0$. By invariance of π, $\pi(Pf) = \pi f$ and, as shown in the proof of Theorem 4.20 $\nu_n(Pf) - \nu_n f \to 0$. **QED**

Remark 6.5 A Feller (even strongly Feller) uniquely ergodic Kernel on a non compact space is not necessarily stable. For instance, let P be the Kernel on $\mathbb{N}$ defined as $P(0,0) = 1$ and for $n \geq 1$ $P(n, n-1) = 1 - p, P(n, n+1) = p$ with $1 > p > 1/2$. Then δ_0 is the unique invariant probability of this Markov chain but the chain is not stable since $\mathbb{P}_x(X_n \to \infty) > 0$ for all $x > 0$. Another (similar) example on $\mathbb{R}^n$ is given by the deterministic linear dynamical system, $X_{n+1} = aX_n$ with $a > 1$.

6.2 Harris recurrence

The chain (X_n) is called *Harris recurrent* if there exists a non zero measure ξ such that for every Borel set $A \subset M$ and every $x \in M$

$$\xi(A) > 0 \Rightarrow \mathbb{P}_x(\limsup_{n\to\infty} \mathbf{1}_A(X_n) = 1\} = 1.$$

Note that an Harris recurrent chain is ξ-irreducible. Recall that an *harmonic* function is a measurable function $h : M \mapsto \mathbb{R}$ such that

$$Ph = h.$$

Theorem 6.6 *Suppose that (X_n) is Harris recurrent. Then every bounded harmonic function is constant.*

Proof Let h be bounded and harmonic. Let (X_n^x) denote the chain having P as transition kernel and initial condition $X_0^x = x$. Then $Y_n = h(X_n^x)$ is a bounded (in particular uniformly integrable) martingale. Hence, by Doob's convergence theorem (Theorem A.7 in the appendix), $\lim_{n\to\infty} Y_n = Y_\infty$ exists almost surely and $\mathsf{E}(Y_\infty|\mathcal{F}_n) = Y_n$. Given $a \in \mathbb{R}$ let $\{h \geq a\}$ (respectively $\{h \leq a\}, \{h = a\}$ be the set of $u \in M$ such that $h(u) \geq a$ (respectively $\leq, =$). If $\xi(\{h \geq a\}) > 0$ then (X_n^x) enters $\{h \geq a\}$ infinitely often. Thus $Y_\infty \geq a$ so that $Y_n = \mathsf{E}(Y_\infty|\mathcal{F}_n) \geq a$. In particular, $h(x) = Y_0 \geq a$. Similarly if $\xi(\{h \leq a\}) > 0$ then $h(x) \leq a$. Let now a be such that $\{h = a\} \neq \emptyset$. Then $\xi(\{h \neq a\}) = \xi(\cup_{n\in\mathbb{N}}\{a - (n+1)^{-1} \leq h \leq a + (n+1)^{-1}\}^c) = 0$. This proves that $h = a$. **QED**

Positive recurrence and Harris recurrence are intimately linked as shown by the next important theorem.

Theorem 6.7 *The following assertion are equivalent:*

(a) *P is Harris recurrent and $\mathsf{Inv}(P) \neq \emptyset$;*

(b) *P is positively recurrent;*

(c) *There exists $\pi \in \mathsf{Inv}(P)$ such that for all $f \in L^1(\pi)$ and every initial distribution μ,*

$$\mathbb{P}_\mu(\lim_{n\to\infty} \nu_n(f) = \pi(f)))1.$$

Proof $c \Rightarrow b \Rightarrow a$ is immediate. Conversely, if P is Harris recurrent with an invariant probability π then P is uniquely ergodic. Let $f \in L^1(\pi), \mathcal{A} = \{\omega \in M^{\mathbb{N}} : \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f \circ \theta^k(\omega) = \pi f\}$ and $g(x) = \mathbb{P}_x(\mathcal{A})$. By the ergodic theorem $g(x) = 1$ π almost surely. We now claim that g is harmonic, which with Theorem 6.6 proves the result. To prove the claim we use the invariance of $\mathcal{A}$ under θ and the Markov property:

$$g(x) = \mathbb{E}_x(\mathbf{1}_{\mathcal{A}}) = \mathbb{E}_x(\mathbf{1}_{\mathcal{A}} \circ \theta) = \mathbb{E}_x(\mathbb{E}_x(\mathbf{1}_{\mathcal{A}} \circ \theta)|\mathcal{F}_1)) = \mathbb{E}_x(g(X_1)) = Pg(x).$$

QED

Theorem 6.8 *Suppose P is strong Feller, uniquely ergodic with an invariant probability π having full support. Then the equivalent conditions of Theorem 6.7 hold true.*

Proof Let $f \in L^1(\pi)$ and let g be defined like in the proof of Theorem 6.7. We have seen that g is harmonic. Since P is strong Feller g is continuous, and by the ergodic theorem $g(x) = 1$ for π almost all x. The set $\{x \in M : g(x) = 1\}$ is then a closed set containing the support of π. This proves that P is positively recurrent. **QED**

Corollary 6.9 *Suppose P is strong Feller with an invariant probability π having full support. If M is connected, then the equivalent conditions of Theorem 6.7 hold true.*

Proof follows from Theorem 6.8 and Proposition 5.18 **QED**

6.2.1 Petite sets and Harris recurrence

A convenient and practical way to ensure that a chain is Harris recurrent is to exhibit a *recurrent petite set.*

Given a Borel set $C \subset M$ we say that $x \in M$ *leads almost surely to C* if $\mathbb{P}_x(\tau_C < \infty) = 1$ where

$$\tau_C = \min\{k \geq 1 : X_k \in C\}.$$

We say that C is *recurrent* if every $x \in M$ leads almost surely to C.

For further reference, we define the successive return times in C recursively by

$$\tau_C^{(n+1)} = \min\{k > \tau_C^{(n)} : \ X_k \in C\}$$

with $\tau_C^{(0)} = 0$.

Proposition 6.10 *Let $C \subset M$ be a recurrent petite set. Then (X_n) is Harris recurrent.*

Proof It easily follows from the definition of a petite set (see Section 5.4), that for all $x \in C$ and A Borel, $\mathbb{P}_x(\tau_A < \infty) \geq \xi(A)$. Thus, using the strong Markov property, for all $x \in M$,

$$\mathbb{P}_x(\tau_A < \infty) \geq \mathbb{P}_x(\exists k \geq \tau_C \ : X_k \in A) = \mathbb{E}_x(\mathbb{P}_{X_{\tau_C}}(\tau_A < \infty)) \geq \xi(A).$$

Therefore, by the Markov property, for all $n \in \mathbb{N}$

$$\mathsf{P}(\tau_A < \infty | \mathcal{F}_n) = \mathbb{P}_{X_n}(\tau_A < \infty) \geq \xi(A).$$

The first term of this inequality converges to $\mathbf{1}_{\tau_A < \infty}$ (see Theorem A.7 in the appendix). Thus $\mathbb{P}_x(\tau_A < \infty) = 1$ for all x, whenever $\xi(A) > 0$. By the strong Markov property, this implies that $X_n \in A$ infinitely often. **QED**

6.3 Recurrence criteria and Lyapunov functions

We discuss here simple useful criteria, based on Lyapounov functions, ensuring that a set is recurrent. It also provide moments estimates of the return times. Conditions (a) and (b) of the next results are folklore (see the notes at the end of the chapter). We learned condition (a') from Philippe Robert (see [36], Proposition 8 in Chapter 8).

Proposition 6.11 *Let $V : M \mapsto [1, \infty[$ be a measurable map and $C \subset M$ a Borel set. Assume that for all $x \in C$ $PV(x) < \infty$ and that one of the three following conditions hold:*

(a) *$PV - V \leq -1$ on $M \setminus C$;*

(a') *Condition (a) and $\sup_{x \in M} \mathbb{E}_x(|V(X_1) - V(x)|^p) < \infty$ for some $p > 1$;*

(b) *$PV - V \leq -\lambda V$ on $M \setminus C$ for some $1 > \lambda > 0$.*

Then for all $x \in M$

(i) $\mathbb{E}_x(\tau_C) \leq PV(x) + 1$ *under condition (a);*

(ii) $\mathbb{E}_x(\tau_C^p) \leq c(1 + V^p(x))$ *for some constant $c > 0$, under condition (a');*

(iii) $\mathbb{E}_x(e^{\lambda \tau_C}) \leq \mathbb{E}_x(e^{-\log(1-\lambda)\tau_C}) \leq \frac{1}{1-\lambda} PV(x)$ *under condition (b).*

In particular, C is a recurrent set.

Proof Let $V_n = V(X_{n \wedge \tau_C}) + (n \wedge \tau_C)$. Then $(V_n)_{n \geq 1}$ is a supermartingale. Indeed, for all $n \geq 1$

$$\mathsf{E}(V_{n+1} - V_n | \mathcal{F}_n) = \mathsf{E}(V_{n+1} - V_n | \mathcal{F}_n) \mathbf{1}_{\tau_C > n} = (PV(X_n) - V(X_n)) \mathbf{1}_{\tau_C > n} \leq 0.$$

Thus $\mathbb{E}_x(n \wedge \tau_C) \leq \mathbb{E}_x(V_n) \leq \mathbb{E}_x(V_1) = PV(x) + 1$. This proves the first assertion. The proof of assertion (iii) is similar. Set $V_n = \frac{V(X_{n \wedge \tau_C})}{(1-\lambda)^{n \wedge \tau_C}}$. Then $(V_n)_{n \geq 1}$ is a supermartingale. Thus

$$\mathbb{E}_x(e^{-\log(1-\lambda) n \wedge \tau_C}) \leq \mathbb{E}_x(V_n) \leq \mathbb{E}_x(V_1) = \frac{PV(x)}{1-\lambda}.$$

We now prove assertion (ii), following Robert ([36], Proposition 8, Chapter 8).

We claim that for all $x > -1$

$$(1 + x)^p \leq 1 + px + C_p r(x) \tag{6.1}$$

where

$$r(x) = x^2 (1 + |x|)^{p-2} \text{ and } C_p = \frac{p(p-1)}{4}$$

for $p \geq 2$; And

$$r(x) = |x|^p \text{ and } C_p = 1$$

for $1 < p < 2$. Indeed, by Taylor-Lagrange formula, for all $x > -1$,

$$(1 + x)^p = 1 + px + \frac{p(p-1)}{2} x^2 R(x)$$

with $R(x) = \int_0^1 (1 - s)(1 + sx)^{p-2} ds$. Thus $|R(x)| \leq \frac{1}{2}(1 + |x|)^{p-2})$ for $p \geq 2$. For $1 < p < 2$, and $x > 0, |R(x)| \leq \int_0^1 (1 - s) s^{p-2} x^{p-2} = \frac{1}{p(p-1)} x^{p-2}$ while for

$1 < p < 2$ and $-1 < x \leq 0$ $|R(x)| \leq 1$ (because $s \in [0,1] \mapsto (1-s)(1+sx)^{p-2}$ is decreasing, hence bounded above by 1). This proves the claim.

Now set

$$Z_n = 1 + \varepsilon(V(X_n) + \frac{n}{2})$$

where $\varepsilon > 0$, and

$$\Delta_{n+1} = V(X_{n+1}) - V(X_n) + \frac{1}{2}.$$

Then

$$Z_{n+1}^p = Z_n^p(1 + \frac{\varepsilon\Delta_{n+1}}{Z_n})^p,$$

so that by (6.1) and condition (a),

$$\mathbb{E}_x(Z_{n+1}^p|\mathcal{F}_n) \leq Z_n^p[1 - \frac{p\varepsilon}{2Z_n} + C_p\mathbb{E}_x(r(\frac{\varepsilon\Delta_{n+1}}{Z_n})|\mathcal{F}_n)]$$

on the event $\tau_C > n$. Now, it is easy to check that $r(\frac{\varepsilon\Delta_{n+1}}{Z_n}) \leq \frac{\varepsilon^2}{Z_n}(1+|\Delta_{n+1}|)^p)$ for $p \geq 2$, and $r(\frac{\varepsilon\Delta_{n+1}}{Z_n}) \leq \varepsilon^p\frac{|\Delta_{n+1}|^p}{Z_n}$ for $1 < p < 2$. Thus, for $\varepsilon > 0$ small enough, condition (a) and (a') make $(Z_{n\wedge\tau_C}^p)$ a supermartingale. The end of the proof is like the proof of (i). **QED**

Remark 6.12 if V is a Lyapounov function in the sense that $PV \leq \rho V + \kappa$ with $0 \leq \rho < 1$ and $\kappa \geq 0$; the assumptions of the Proposition 6.11 (b) hold true with $0 < \lambda < 1 - \rho$ and $C = \{x \in M \ : V(x) \leq \frac{\kappa+1}{1-\rho-\lambda}\}$. Compare to Proposition 4.23.

The next proposition extends assertion (iii) of Proposition 6.11 and gives an alternative condition (to conditions $(a), (a')$) to control the moments of τ_C. The proof is based on a beautiful argument used in section 4.1 of Hairer's notes [**?**].

Proposition 6.13 *Let $V : M \mapsto [1,\infty[$ be a measurable map and $C \subset M$ a Borel set. Let $\varphi : [0,\infty[\mapsto \mathbb{R}_+^*$ be a C^1 concave function and let $h : [1,\infty[\mapsto [0,\infty[$ be the map defined by*

$$h(x) = \int_1^x \frac{ds}{\varphi(s)}.$$

Assume that for all $x \in C$ $PV(x) < \infty$ and that for all $x \in M \setminus C$

$$PV(x) - V(x) \leq -\varphi(V(x)).$$

Then, for all $x \in M \setminus C$

$$\mathbb{E}_x(h^{-1}(\tau_C)) \leq V(x)$$

and, for all $x \in C$

$$\mathbb{E}_x(h^{-1}(\tau_C)) \leq h^{-1}(h(PV(x)) + 1).$$

Proof First observe that $\varphi' \geq 0$ (for otherwise by concavity φ could not be > 0). For $x \geq 1$ and $t \geq 0$ set $H(t,x) = h^{-1}(h(x) + t)$. It is readily seen that that

$$\frac{\partial H}{\partial t}(t,x) = \varphi(H(t,x)) = \varphi(x)\frac{\partial H}{\partial x}(t,x). \tag{6.2}$$

Thus

$$\frac{\partial^2 H}{\partial x^2}(t,x) = \frac{(\varphi'(H(t,x)) - \varphi'(x))\varphi(H(t,x))}{\varphi(x)^2} \leq 0. \tag{6.3}$$

In particular, H is convex in t and concave in x.

It follows that for all $n \geq 0$

$$H(n+1, V(X_{n+1})) - H(n, V(X_n)) =$$

$$H(n+1, V(X_{n+1})) - H(n+1, V(X_n)) + H(n+1, V(X_n)) - H(n, V(X_n))$$
$$\leq \frac{\partial H}{\partial x}(n+1, V(X_n))(V(X_{n+1}) - V(X_n)) + \frac{\partial H}{\partial t}(n+1, V(X_n)).$$

Therefore, on the event $\{X_n \notin C\}$,

$$\mathbb{E}(H(n+1, V(X_{n+1})) - H(n, V(X_n))|\mathcal{F}_n)$$
$$\leq -\varphi(V(X_n))\frac{\partial H}{\partial x}(n+1, V(X_n)) + \frac{\partial H}{\partial t}(n+1, V(X_n)) \leq 0.$$

Here the first inequality follows from the hypotheses on V and the second from equation (6.2). This makes the process $(H(n \wedge \tau_C, V(X_{n\wedge\tau_C}))_{n\geq 1}$ a supermartingale. Thus

$$\mathbb{E}_x(h^{-1}(n\wedge\tau_C)) \leq \mathbb{E}_x(H(n\wedge\tau_C, V(X_{n\wedge\tau_C}))) \leq \mathbb{E}_x(H(1, V(X_1)) \leq H(1, PV(x))$$

where the last inequality follows from concavity of H in x and Jensen inequality. In case $x \in M \setminus C$, by monotony and concavity of h

$$h(PV(x)) \leq h(V(x) - \varphi(V(x))) \leq h(V(x)) - h'(V(x))\varphi(V(x)) = h(V(x)) - 1.$$

Thus $H(1, PV(x)) \leq V(x)$. This proves the result. **QED**

6.3.1 Subsets of recurrent sets

Let $C \subset M$ be a recurrent set for the chain (X_n) (for instance the sublevel set $\{V \leq R\}$ of a Lyapounov function) and $U \subset C$ a measurable smaller subset (for instance the neighborhood of a Doeblin point). It is often desirable to deduce recurrence properties of U from recurrence properties of C. This short section discusses two such results.

The *induced chain* on C is the process $(Y_n)_{n\geq 1}$ defined as

$$Y_n = X_{\tau_C^{(n)}}.$$

Exercise 6.14 Verify that $(Y_n)_{n\geq 1}$ is a Markov chain on C.

Proposition 6.15 *Let $C \subset M$ be a nonempty recurrent set and $U \subset C$ a measurable subset. Suppose that there exists $k \geq 1$ and $0 < \varepsilon \leq 1$ such that for all $x \in C$*

$$\mathbb{P}_x(\exists i \in \{1, \ldots, k\}\, Y_i \in U) \geq \varepsilon$$

where (Y_n) stands for the induced chain on C. Then

(i) *U is recurrent;*

(ii) *If $\sup_{x\in C} \mathbb{E}_x(\tau_C^p) < \infty$ for some $p \geq 1$, then*

$$\sup_{x\in C} \mathbb{E}_x(\tau_U^p) < \infty$$

(iii) *If $\sup_{x\in C} \mathbb{E}_x(e^{\lambda_0 \tau_C}) < \infty$ for some $\lambda_0 > 0$, then*

$$\sup_{x\in C} \mathbb{E}_x(e^{\lambda \tau_U}) < \infty$$

for some $0 < \lambda \leq \lambda_0$.

Proof For all $x \in M, \mathbb{P}_x$ almost surely

$$\mathbf{1}_{\tau_U < \infty} = \lim_{n\to\infty} \mathbb{P}_x(\tau_U < \infty | \mathcal{F}_{\tau_C^{(n)}}) = \lim_{n\to\infty} \mathbb{P}_{Y_n}(\tau_U < \infty) \geq \varepsilon.$$

Here the first equality follows from the Martingale convergence theorem A.7 and the second from the strong Markov property. This proves that U is recurrent.

Let $\sigma_U = \min\{n \geq 1 \ : Y_n \in U\}$. The proofs of assertions (ii) and (iii) now follow from the identity $\tau_U = \tau_C^{(\sigma_U)}$, exactly as in the proof of Proposition 2.15 $(i),(ii)$. The verification is an easy exercise left to the reader. **QED**

When P is Feller, the existence of a compact recurrent set C makes every accessible open set U recurrent. More precisely,

Proposition 6.16 *Suppose that P is Feller. Let $C \subset M$ be a nonempty compact set, $x^* \in M$ an accessible point from C (i.e $x^* \in \Gamma_C$) and U a neighborhood of x^*.*

(i) *If C is recurrent, so is U;*

(ii) *If $U \subset C$ and $\sup_{x\in C} \mathbb{E}_x(\tau_C^p) < \infty$ for some $p \geq 1$, then*

$$\sup_{x\in C} \mathbb{E}_x(\tau_U^p) < \infty$$

(iii) *If $U \subset C$ and $\sup_{x\in C} \mathbb{E}_x(e^{\lambda_0 \tau_C}) < \infty$ for some $\lambda_0 > 0$, then*

$$\sup_{x\in C} \mathbb{E}_x(e^{\lambda \tau_U}) < \infty$$

for some $0 < \lambda \leq \lambda_0$.

Proof For $\varepsilon > 0$ and $i \in \mathbb{N}^*$ let $O(\varepsilon, i) = \{x \in M \ : \ P^i(x,U) > \varepsilon\}$. By Feller continuity and Portemanteau's theorem 4.1, $O(\varepsilon,i)$ is an open set. By accessibility of x^* the family $\{O(\varepsilon,i), \varepsilon > 0, i \in \mathbb{N}^*\}$ covers C. Thus, by compactness, there exist $\varepsilon > 0$ and a finite set $I \subset \mathbb{N}$ such that $C \subset \cup_{i\in I} O(\varepsilon,i)$. This shows that, for all $x \in C$,

$$\mathbb{P}_x(\tau_U \leq k) \geq \varepsilon \tag{6.4}$$

with $k = \max I$. Assertions (ii) and (iii) then follow from Proposition 6.15 because, for all $x \in C$,

$$\mathbb{P}_x(\exists i \in \{1,\ldots,k\}\ Y_i \in U) \geq \mathbb{P}_x(\tau_U \leq k) \geq \varepsilon.$$

The proof of first assertion is similar to the proof of the first assertion in Proposition 6.15. Namely, for all $x \in M, \mathbb{P}_x$ almost surely,

$$\mathbf{1}_{\tau_U < \infty} = \lim_{n\to\infty} \mathbb{P}_x(\tau_U < \infty | \mathcal{F}_{\tau_C^{(n)}}) = \lim_{n\to\infty} \mathbb{P}_{Y_n}(\tau_U < \infty) \geq \varepsilon.$$

Thus $\mathbb{P}_x(\tau_U < \infty) = 1$. **QED**

6.4 Petite sets and positive recurrence

We have seen (Proposition 6.10) that the existence of a recurrent petite set for a Markov chain makes it Harris recurrent. If, in addition, the return times to the set are bounded in L^1, then it is positively recurrent.

Theorem 6.17 *Let $C \subset M$ be a recurrent petite set such that*

$$\sup_{x \in C} \mathbb{E}_x(\tau_C) < \infty.$$

Then the equivalent conditions of Theorem 6.7 hold true.

Before proving this theorem, we start with a proposition relating the recurrence properties of the chain (X_n) and the sampled chain $Y_n := X_{T_n}$, where

$$T_n := \Delta_1 + \ldots + \Delta_n$$

for $n \geq 1$, $T_0 := 0$, and $(\Delta_i)_{i \geq 1}$ is a sequence of i.i.d. random variables taking on values in $\mathbb{N}$.

Recall that in the particular case where Δ_i has a geometric distribution with parameter a, (i.e. $\mathsf{P}(\Delta_i = n) = a^n(1-a)$ for all $n \in \mathbb{N}$) then (Y_n) has kernel R_a.

The *hazard rate* of Δ_i is the sequence

$$\lambda(n) = \mathsf{P}(\Delta_i = n | \Delta_i \geq n) = \frac{\mathsf{P}(\Delta_i = n)}{\mathsf{P}(\Delta_i \geq n)}, \quad n \in \mathbb{N}.$$

For a geometric distribution with parameter a, the hazard rate is constant and equals $1 - a$.

Exercise 6.18 Suppose Δ_i has a negative binomial distribution with parameters (a, m) (see Exercice 5.2 (ii)). Prove that $\lambda(n)$ is nondecreasing and converges to $1 - a$. In particular,

$$\inf_{n \in \mathbb{N}} \lambda(n) = \lambda(0) = (1-a)^m.$$

The next result is an easy consequence of the memoryless property when Δ_i has a geometric distribution (prove it as an exercise) and this is exactly what we'll need for the proof of Theorem 6.17. It is however interesting to point out that it remains valid under the weaker assumption that the hazard rate of Δ_i is bounded below. Tom Mountford helped us with the proof of this proposition and suggested the minorization condition on the hazard rate.

Proposition 6.19 *Let* $(\Delta_n), (T_n)$ *be as above, i.e.* (Δ_n) *is an i.i.d. sequence of* $\mathbb{N}$*-valued random variables and* $T_n := \Delta_1 + \ldots + \Delta_n$. *Assume that there is* $\alpha \in (0,1)$ *such that*

$$\inf_{n\in\mathbb{N}} \lambda(n) \geq 1-\alpha > 0.$$

Let $\mathcal{N} = \{n_1 < n_2 < \ldots < n_k < \ldots\} \subset \mathbb{N}$ *be an infinite set of integers and*

$$\tau_{\mathcal{N}} := \min\{n \geq 1 :\ T_n \in \mathcal{N}\}.$$

Then

(i) $\mathsf{P}(\tau_{\mathcal{N}} < \infty) = 1$;

(ii) $\mathsf{P}(T_{\tau_{\mathcal{N}}} > n_i) \leq \alpha^i$ *for all* $i \geq 1$;

(iii) $\mathsf{E}(\Delta_1)\mathsf{E}(\tau_{\mathcal{N}}) \leq n_1 + \sum_{i\geq 1}(n_{i+1}-n_i)\alpha^i$;

(iv) *If* $\lambda(n) = 1-\alpha$ *for all* $n \in \mathbb{N}$ *(meaning that* Δ_i *has a geometric distribution with parameter* α*), inequalities* (**ii**) *and* (**iii**) *are equalities.*

Proof (i). For $n \geq 1$, let $\mathcal{F}_n := \sigma(\Delta_1, \ldots, \Delta_n)$ and $v(n) := \mathsf{P}(\exists i \geq 0 :\ T_i = n)$. We claim that $v(n) \geq 1-\alpha$ for all $n \geq 1$. One has

$$\begin{aligned} v(n) =& \mathsf{E}(\mathsf{P}(\exists i \geq 0 :\ T_i = n|\mathcal{F}_1)) \\ =& v(n)\mathsf{P}(\Delta_1 = 0) + \mathsf{E}(v(n-\Delta_1)\mathbf{1}_{0<\Delta_1<n}) + \mathsf{P}(\Delta_1 = n). \end{aligned}$$

Thus, $v(1) = \lambda(1) \geq 1-\alpha$. Suppose now that $v(i) \geq 1-\alpha$ for $i = 1, \ldots, n-1$. Then

$$v(n)\mathsf{P}(\Delta_1 > 0) \geq (1-\alpha)\mathsf{P}(0 < \Delta_1 < n) + \mathsf{P}(\Delta_1 = n) \geq (1-\alpha)\mathsf{P}(\Delta_1 > 0).$$

This proves the claim by induction. It follows from what precedes that $\mathsf{P}(\tau_{\mathcal{N}} < \infty|\mathcal{F}_n) \geq 1-(1-\alpha)^n$, so that P-almost surely

$$\mathbf{1}_{\tau_{\mathcal{N}}<\infty} = \lim_{n\to\infty} \mathsf{P}(\tau_{\mathcal{N}} < \infty|\mathcal{F}_n) = 1.$$

(ii). For $k \geq 1$, let $S_k := \min\{i \geq 0 :\ n_k \leq T_i < n_{k+1}\} \in \mathbb{N} \cup \{\infty\}$. Then

$$\mathsf{P}(T_{\tau_{\mathcal{N}}} > n_{k+1}) = \mathsf{P}(T_{\tau_{\mathcal{N}}} > n_{k+1}; S_k < \infty) + \mathsf{P}(T_{\tau_{\mathcal{N}}} > n_{k+1}; S_k = \infty).$$

Using the strong Markov property,

$$\mathsf{P}(T_{\tau_{\mathcal{N}}} > n_{k+1}; S_k < \infty) = \mathsf{E}(\mathsf{P}(T_{\tau_{\mathcal{N}}} > n_{k+1}|\mathcal{F}_{S_k})\mathbf{1}_{\{S_k<\infty\}})$$

$$= \mathsf{E}((1 - v(n_{k+1} - T_{S_k}))\mathbf{1}_{\{T_{\tau_\mathcal{N}} > n_k\}}\mathbf{1}_{\{S_k < \infty\}}) \leq \alpha \mathsf{P}(T_{\tau_\mathcal{N}} > n_k; S_k < \infty).$$

On the other hand,

$$\begin{aligned} &\mathsf{P}(T_{\tau_\mathcal{N}} > n_{k+1}; S_k = \infty) \\ =&\sum_{i \geq 0} \mathsf{P}(\{T_0, T_1, \ldots, T_i\} \cap \{n_1, \ldots, n_k\} = \emptyset; T_i < n_k; T_{i+1} > n_{k+1}), \end{aligned}$$

and

$$\begin{aligned} &\mathsf{P}(\{T_0, T_1, \ldots, T_i\} \cap \{n_1, \ldots, n_k\} = \emptyset; T_i < n_k; T_{i+1} > n_{k+1} | \mathcal{F}_i) \\ &\quad = \mathbf{1}_{\{T_0, T_1, \ldots, T_i\} \cap \{n_1, \ldots, n_k\}} \mathbf{1}_{T_i < n_k} \mathsf{P}(\Delta_{i+1} > n_{k+1} - T_i | \mathcal{F}_i) \\ &\quad \leq \alpha \mathbf{1}_{\{T_0, T_1, \ldots, T_i\} \cap \{n_1, \ldots, n_k\}} \mathbf{1}_{T_i < n_k} \mathsf{P}(\Delta_{i+1} \geq n_{k+1} - T_i | \mathcal{F}_i) \end{aligned}$$

by the assumption on the hazard rate of (Δ_i). Therefore,

$$\begin{aligned} &\mathsf{P}(T_{\tau_\mathcal{N}} > n_{k+1}; S_k = \infty) \\ \leq&\alpha \sum_{i \geq 0} \mathsf{E}(\mathbf{1}_{\{T_0, T_1, \ldots, T_i\} \cap \{n_1, \ldots, n_k\}} \mathbf{1}_{T_i < n_k} \mathsf{P}(\Delta_{i+1} \geq n_{k+1} - T_i | \mathcal{F}_i)) \\ =&\alpha \mathsf{P}(T_{\tau_\mathcal{N}} > n_k; S_k = \infty). \end{aligned}$$

Finally we have shown that

$$\mathsf{P}(T_{\tau_\mathcal{N}} > n_{k+1}) \leq \alpha \mathsf{P}(T_{\tau_\mathcal{N}} > n_k).$$

(iii). Let $M_n := T_n - \mathsf{E}(T_n) = T_n - nm$, where $m := \mathsf{E}(\Delta_i)$. Then (M_n) is an $(\mathcal{F}_n)$-martingale with zero mean. Thus, by part (**ii**) of Theorem A.4,

$$\mathsf{E}(M_{n \wedge \tau_\mathcal{N}}) = 0 = \mathsf{E}(T_{\tau_\mathcal{N} \wedge n}) - m\mathsf{E}(\tau_\mathcal{N} \wedge n),$$

and, by monotone convergence,

$$m\mathsf{E}(\tau_\mathcal{N}) = \mathsf{E}(T_{\tau_\mathcal{N}}) = \sum_{k \geq 1} n_k \mathsf{P}(T_{\tau_\mathcal{N}} = n_k) = \sum_{k \geq 0} (n_{k+1} - n_k) \mathsf{P}(T_{\tau_\mathcal{N}} > n_k)$$

with the convention $n_0 := 0$.

(iv). This follows immediately from the proofs of (**ii**) and (**iii**). **QED**

Proof of Theorem 6.17 In view of Theorem 6.7 and Proposition 6.10 it suffices to show that there exists an invariant probability for (X_n).

First observe that we can always assume that $\xi(C) > 0$ where ξ is the minorizing measure of R_a. Indeed, let $\xi_k(\cdot) = a^k \int \xi(dy)P^k(y,\cdot)$. Then for all $x \in C$

$$R_a(x,\cdot) \geq a^k R_a P^k(x,\cdot) \geq \xi_k(\cdot)$$

so that ξ_k is another minorizing measure. Now, there exists k such that $\xi_k(C) > 0$, for otherwise we would have $P^k(y,C) = 0$ for all k and ξ almost all y, in contradiction with the assumption that C is recurrent. Replacing ξ by such a ξ_k proves our claim.

Let $\tau_C < \tau_C^{(2)} < \tau_C^{(3)} < \ldots$ be the successive times at which (X_n) enters C. That is $\tau_C^{(k+1)} = \min\{n > \tau_C^{(k)} : X_n \in C\}$. By assumption (iii) (of the theorem to be proved) and the strong Markov property

$$\mathbb{E}_x(\tau_C^{(k)}) \leq kM$$

for all $x \in C$. Let (Y_n) be the chain with kernel R_a, $\tau_C^Y = \min\{n \geq 1 : Y_n \in C\}$ and $Q(x,\cdot)$ the kernel on C defined by $Q(x,A) = \mathbb{P}_x(Y_{\tau_C^Y} \in A)$ for all Borel set $A \subset C$. By Proposition 6.19 (i), $\tau_C^Y < \infty$ a.s so that Q is a Markov kernel (i.e $Q(x,C) = 1$). Furthermore $Q(x,A) \geq R_a(x,A) \geq \varepsilon\psi(A)$ with $\varepsilon = \xi(C)$ and $\psi(A) = \frac{\xi(A)}{\varepsilon}$. In other words, Q is a Markov kernel whose full state space (here C) is a small set. Then, by a theorem that will be proved later (Theorem 7.7 in Chapter 7), Q has a (unique) invariant probability π. If Y_0 is distributed according to π so is $Y_{\tau_C^Y}$ and by Proposition 6.19 (iii)

$$\mathbb{E}_\pi(\tau_C^Y) \leq \frac{M}{a}.$$

By Exercise 4.24 this implies that (Y_n) (or equivalently R_a), hence (X_n), admits an invariant probability. **QED**

6.4.1 Positive recurrence for Feller chains

The next results give some (much more tractable) conditions ensuring that a Feller chain is positively recurrent.

Theorem 6.20 *Let P be Feller. Assume that there exist a compact recurrent set C such that $\sup_{x\in C} \mathbb{E}_x(\tau_C) < \infty$, and an accessible weak Doeblin point*

$x^* \in \mathsf{Int}(C)$ (the interior of C). Then the equivalent conditions of Theorem 6.7 hold true.

Proof By assumption there exists $U \subset C$ a neighborhood of x^*, and a non trivial measure ξ such that $R_a(x, \cdot) \geq \xi(\cdot)$ for all $x \in U$. By Proposition 6.16 U is recurrent and $\sup_{x \in C} \mathbb{E}_x(\tau_U) < \infty$. We can then apply Theorem 6.17, with U in place of C. This proves the result **QED**

Corollary 6.21 *Let P be Feller. Assume that there exists a weak Doeblin accessible point, a proper map $V : M \mapsto \mathbb{R}_+$ and nonnegative constants R, M such that $PV \leq V - 1$ on $\{V > R\}$, and $PV \leq M$ on $V \leq R$. Then the equivalent conditions of Theorem 6.7 hold true.*

Proof Let x^* be the weak Doeblin accessible point. Choose R large enough so that $V(x^*) < R$. Set $C = \{V \leq R\}$ and apply Proposition 6.11 (a) and Theorem 6.20. **QED**

Theorem 6.22 *Let P be Feller. Assume that there exists a weak Doeblin accessible point and that for all $x \in M$ the empirical occupation measure (ν_n) is $\mathbb{P}_x$ almost surely tight (this is true for instance under the assumptions of corollary). Then the equivalent conditions of Theorem 6.7 hold true.*

Proof By assumption there exists an open accessible petite set C. By Proposition 5.37 and 4.20, there exists a unique invariant probability π for P and $\nu_n \Rightarrow \pi$ $\mathbb{P}_x$ almost surely for all $x \in M$. Since C is open and accessible $\pi(C) > 0$ (see Proposition 5.8 (ii)) and, by Portmanteau theorem, $\liminf \nu_n(C) \geq \pi(C)$. This proves that every point x leads almost surely to C. The result then follows from Proposition 6.10 and Theorem 6.7 **QED**

Chapter 7

Harris and Orey Ergodic Theorems

7.1 Total variation distance

Recall that $B(M)$ is the set of real valued bounded measurable maps on M. For $f \in B(M), \|f\|_\infty$ is defined by (1.1). Given two probability measures α and β on M the *total variation distance* between α and β is defined by

$$|\alpha - \beta| = \sup\{|\alpha(f) - \beta(f)| \ : \ f \in B(M), \ \|f\|_\infty \leq 1\}. \tag{7.1}$$

It is easy to verify that this defines a distance on $\mathcal{P}(M)$.

Note that if K is a Markov kernel on M

$$|\alpha K - \beta K| \leq |\alpha - \beta| \tag{7.2}$$

because K maps $\{f \in B(M), \ \|f\|_\infty \leq 1\}$ into itself.

Proposition 7.1 *Let $\alpha, \beta \in \mathcal{P}(M)$.*

(i)

$$|\alpha - \beta| = 2 \sup_{A \in \mathcal{B}(M)} \alpha(A) - \beta(A).$$

(ii) *Assume α and β are absolutely continuous with respect to $\gamma \in \mathcal{P}(M)$ with respective densities p and q. Then*

$$|\alpha - \beta| = \int |p - q| d\gamma.$$

(iii) *The space $\mathcal{P}(M)$ equipped with the total variation distance is complete.*

Proof We begin by assertion (ii). For all $f \in B(M)$ with $\|f\| \leq 1$, $|\alpha(f) - \beta(f)| \leq \int |p-q| d\gamma$ so that $|\alpha-\beta| \leq \int |p-q| d\gamma$. Conversely, set $f = 1_{p>q} - 1_{p<q}$. Then $\alpha(f) - \beta(f) = \int |p-q| d\gamma$.

We now pass to the proof of (i). We can always assume that for some $\gamma \in \mathcal{P}(M)$ α and β are absolutely continuous with respect to γ. It suffices for instance to choose $\gamma = \frac{\alpha+\beta}{2}$. Then,

$$|\alpha - \beta| = \int_G (p-q) d\gamma + \int_{M \setminus G} (q-p) d\gamma = 2(\alpha(G) - \beta(G))$$

with $G = \{p > q\}$. Also for all $A \in \mathcal{B}(M)$ $\alpha(A) - \beta(A) \leq \alpha(A \cap G) - \beta(A \cap G) \leq (\alpha(G) - \beta(G))$. Our last job is to prove the completeness. Let (μ_n) be a Cauchy sequence for the total variation. Then, in view of (i), for every Borel set A $(\mu_n(A))$ is a Cauchy sequence in $\mathbb{R}$, hence converge to some number $\mu(A)$. By the Cauchy property, the convergence is uniform in A. That is $\sup_{A \in \mathcal{B}(M)} |\mu_n(A) - \mu(A)| \to 0$. From this it is easy to verify that μ is a probability over M. **QED**

Exercise 7.2 For $f : M \mapsto \mathbb{R}$, let $\Delta(f) = \sup\{\frac{|f(x)-f(y)|}{2} : x, y \in M\}$. Show that

$$|\alpha - \beta| = \sup\{|\alpha(f) - \beta(f)| : f \text{ measurable}, \Delta(f) \leq 1\}.$$

Remark 7.3 Although the total variation distance (7.1) and the Fortet-Mourier distance (4.2) look very similar, they induce quite different topologies on $\mathcal{P}(M)$. Clearly,

$$\rho(\alpha, \beta) \leq |\alpha - \beta|$$

so that convergence in total variation implies weak convergence; but the converse is clearly false. Let, for example, X be a random variable on $\mathbb{R}$ whose law P_X is absolutely continuous with respect to the Lebesgue measure dx (e.g a Gaussian random variable) and $X_n = \frac{X}{n}$. Then $X_n \to 0$ almost surely, hence, $\mathsf{P}_{X_n} \Rightarrow \delta_0$ while $|\mathsf{P}_{X_n} - \delta_0| = 2$ by Proposition 7.1, (i).

Remark 7.4 (Total variation of signed measures) A *finite signed measure* on M is a map $\mu : \mathcal{B}(M) \mapsto \mathbb{R}$ such that $\mu(\emptyset) = 0$ and which is σ-additive. That is

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$$

for any family $\{A_n\} A_n \in \mathcal{B}(M)$ having disjoint elements. The Hahn-Jordan decomposition theorem (see [13], Theorem 5.6.1) asserts that such a measure can be written as

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

where μ^+ and μ^- are nonnegative measure that are mutually singular: There exists $D \in B(M)$ such that for all $A \in \mathcal{B}(M)$ $\mu^+(A) = \mu(A \cap D)$ and $\mu^-(A) = -\mu(A \cap D^c)$. The total variation measure of μ is the non negative measure $\mu^+ + \mu^-$ and the its total variation norm is

$$|\mu| = \mu^+(M) + \mu^-(M) = \sup\{|\mu(f)| \ : \ f \in B(M), \ \|f\|_\infty \leq 1\}.$$

When M is a metric compact space, the topological dual $C^*(M)$ of $C(M)$ can be identified with the space of bounded signed measures equipped with the total variation norm, so that convergence in total variation coincides with (strong) convergence in $C^*(M)$. We refer the reader to [13], chapter 7 for more details and a proof of this latter point.

Exercise 7.5 Use the Hahn-Jordan decomposition to show assertion (i) of Proposition 7.1.

7.1.1 Coupling

Given $\alpha, \beta \in \mathcal{P}(M)$, a *coupling* of α and β is a random vector (X, Y) defined on some probability space $(\Omega, \mathcal{A}, \mathsf{P})$ taking values in $M \times M$ such that X has distribution α and Y has distribution β.

Proposition 7.6 *Let $\alpha, \beta \in \mathcal{P}(M)$. Then*

(i) *(Coupling Inequality) For every coupling (X, Y) of (α, β)*

$$|\alpha - \beta| \leq 2P(X \neq Y);$$

(ii) *(Maximal coupling) There exists a coupling (X, Y) of (α, β) such that*

$$|\alpha - \beta| = 2P(X \neq Y).$$

Proof (i). For all $A \in \mathcal{B}(M)$

$$\mathsf{P}(X \in A) - \mathsf{P}(Y \in A) = \mathsf{P}(X \in A; X \neq Y) - \mathsf{P}(Y \in A; X \neq Y) \leq \mathsf{P}(X \neq Y).$$

This inequality, combined with Proposition 7.1 (i), proves (i).

(ii). Assume (without loss of generality) that $d\alpha = pd\gamma$ and $d\beta = qd\gamma$ for some $\gamma \in \mathcal{P}(M)$ (e.g $\gamma = (\alpha + \beta)/2$). Then, by Proposition 7.1 (ii),

$$|\alpha - \beta| = \int |p - q| d\gamma = 2(1 - \varepsilon)$$

where $\varepsilon = \int (p \wedge q) d\gamma$. If $\varepsilon = 0$ α and β are mutually singular and any coupling satisfies the equality $|\alpha - \beta| = 2P(X \neq Y) = 2$. If $\varepsilon \neq 0$, let $U \in M, V \in M, W \in M, \Theta \in \{0, 1\}$ be independent random variables having distributions $\frac{1}{\varepsilon} p \wedge q d\gamma$, $\frac{1}{1-\varepsilon}(p - p \wedge q)d\gamma$, $\frac{1}{1-\varepsilon}(q - p \wedge q)d\gamma$ and $(1-\varepsilon)\delta_0 + \varepsilon\delta_1$. Set $X = \Theta U + (1 - \Theta)V$ and $Y = \Theta U + (1 - \Theta)W$. Then $\mathsf{P}(X \neq Y) = \mathsf{P}(\Theta = 0) = (1 - \varepsilon)$ and (X, Y) is a coupling of (α, β). **QED**

7.2 Harris convergence theorems

Throughout all this section P is a Markov kernel on M . Recall (see Section 5.4) that a set $C \subset \mathcal{B}(M)$ is called a *small set* for P if there exists a nontrivial measure ξ on M (called the minorizing measure of C) such that

$$P(x, \cdot) \geq \xi(\cdot) \tag{7.3}$$

for all $x \in C$. Recall also that a point is called a Doeblin point if it has a neighborhood which is a small set.

7.2.1 Geometric convergence

The importance of small sets is emphasized by the following simple version of Harris's theorem.

Theorem 7.7 *Let $m \in \mathbb{N}, m \geq 1$. Suppose M is a small set for P^m. Then for all $\alpha, \beta \in \mathcal{P}(M)$*

$$|\alpha P^n - \beta P^n| \leq (1 - \varepsilon)^{[n/m]} |\alpha - \beta|$$

where $0 < \varepsilon = \xi(M) \leq 1$ and ξ is like in (7.3). Furthermore P has a unique invariant probability π and

$$|\alpha P^n - \pi| \leq (1 - \varepsilon)^{[n/m]} |\alpha - \pi|.$$

Proof First suppose $m = 1$. Set $\psi = \frac{\xi}{\xi(M)}, \varepsilon = \xi(M)$ and

$$K(x, \cdot) = \frac{P(x, \cdot) - \varepsilon\psi(\cdot)}{1 - \varepsilon}.$$

Then, K is a Markov kernel and $\alpha P = \varepsilon\psi + (1-\varepsilon)\alpha K$ so that

$$|\alpha P - \beta P| = (1-\varepsilon)|\alpha K - \beta K| \leq (1-\varepsilon)|\alpha - \beta|$$

where the last inequality follows from (7.2). Hence, $\alpha \mapsto \alpha P$ is a contraction, for the total variation distance. Then

$$|\alpha P^n - \beta P^n| \leq (1-\varepsilon)^n |\alpha - \beta|$$

and $\alpha \mapsto \alpha P$ has a unique fixed point, by application of the Banach fixed point theorem, because the space of probability measures endowed with the total variation distance is complete.

If now $m \geq 1$ set $Q = P^m$. Write $n = km + r$ for $r \in \{0, \ldots, m-1\}$ and

$$|\alpha P^n - \beta P^n| = |\alpha P^r Q^k - \beta P^r Q^k| \leq (1-\varepsilon)^k |\alpha P^r - \beta P^r| \leq (1-\varepsilon)^k |\alpha - \beta|.$$

To conclude, notice that if π is invariant for P^m then $\frac{1}{m}\sum \pi P^k$ is invariant for P. **QED**

Remark 7.8 Theorem 7.7 doesn't require that P is Feller (and not even that M is a metric space).

Aperiodic small sets

A measurable set $C \subset M$ is said to be *aperiodic* if the set

$$R(C) = \{k \geq 1 \ : \inf_{x \in C} P^k(x, C) > 0\}$$

is nonempty and aperiodic as defined in Section 2.2.1.

Exercise 7.9 **(a)** Let P be Feller and let $U \subset M$ be an open, accessible (i.e $R_a(x, U) > 0$ for all $x \in M$) small set. Show that $R(U)$ is nonempty.

(b) Construct a Feller Markov chain having an open recurrent set for which $R(U) = \emptyset$.

Let $x^* \in M$ be an accessible Doeblin point for P Feller. We say that x^* is *aperiodic* if it has a neighboring small set U which is aperiodic. Observe that if $\xi(U) > 0$ (where ξ stands for the minorizing measure of U) then x^* is aperiodic.

Proposition 7.10 *Assume P is Feller. Let $x^* \in M$ be an accessible and aperiodic Doeblin point and let $C \subset M$ be a compact set. Then there exists $m \geq 1$ such that C is a small set for P^m.*

Proof Let U be a small set, neighborhood of x^* with $R(U)$ aperiodic. Then, by aperiodicity, there exists $n_0 \in \mathbb{N}$ such that $k \in R(U)$ for all $k \geq n_0$ (see Proposition 2.19).

For $\delta > 0$ and $k \in \mathbb{N}^*$ let $O(\delta, k) = \{x \in M : P^k(x, U) > \delta\}$. By Feller continuity and Portemanteau's theorem 4.1, $O(\delta, k)$ is an open set. By assumption (x^* accessible) the family $\{O(\delta, k), \delta > 0, k \in \mathbb{N}^*\}$ covers M. Thus, by compactness, there exist $\delta > 0$ and integers $k_1, \ldots, k_n$ such that $C \subset \cup_{i=1}^n O(\delta, k_i)$. For $x \in O(\delta, k_i)$ and $k > n_0$

$$P^{k_i+k}(x, .) \geq \int_U P^{k_i}(x, dy) P^k(y, .)$$

$$\geq \int_U P^{k_i}(x, dy) P^{k-1}(y, U)\xi(.) \geq \delta \inf_{y \in U} P^{k-1}(y, U)\xi(.).$$

Here ξ stands for the minorizing measure of U. Thus, for $m = \max\{k_1, \ldots, k_n\} + n_0 + 1$ and some $\delta' > 0$

$$P^m(x, .) \geq \delta' \xi(.).$$

QED

Theorem 7.7 and this latter proposition imply the following useful result for Feller chain on compact sets.

Corollary 7.11 *Assume P is Feller on M compact and that there exists an accessible and aperiodic Doeblin point. Then the conclusion of Theorem 7.7 holds.*

When M is not compact, the assumption (made in Theorem 7.7 or used in Corollary 7.11) that all the space is a small set is usually not satisfied. A sufficient condition ensuring geometric convergence is given by the existence of a small set and a Lyapunov function forcing the system to enter this small set. A classical proof relying on coupling and renewal properties will be given in the next section. Hairer and Mattingly in [21] gave an alternative beautiful proof based on the construction of a suitable semi-norm making P a contraction. This proof is given below.

Theorem 7.12 (Harris, Hairer & Mattingly) *Assume that there exist:*

(a) *A measurable map* $V : M \mapsto \mathbb{R}_+, 0 < \rho < 1$ *and* $\kappa \geq 0$ *such that*

$$PV \leq \rho V + \kappa;$$

(b) *A probability measure* ψ *on* M *and* $0 < \varepsilon \leq 1$ *such that*

$$P(x, \cdot) \geq \varepsilon \psi(\cdot)$$

for all $x \in V_R := \{x \in M\ : V(x) \leq R\}$ *and* $R \geq 2\kappa/(1-\rho)$.

Then, there exist a (unique) invariant probability π *for* P *and constants* $0 \leq \gamma < 1, C > 0$ *such that for all* $f : M \mapsto \mathbb{R}$ *measurable with* $\|f\|_V := \sup \frac{|f(x)|}{1+V(x)} < \infty$,

$$|P^n f(x) - \pi(f)| \leq C\gamma^n (1 + V(x)) \|f\|_V$$

for all $x \in M$.

Proof For $\beta > 0$ and $f : M \mapsto \mathbb{R}$ measurable, possibly unbounded, let

$$\|f\|_\beta = \sup\{\frac{|f(x) - f(y)|}{2 + \beta(V(x) + V(y))}\ :\ x, y \in M\}.$$

We claim that for some $1 \geq \beta > 0$ and $0 \leq \gamma < 1$,

$$\|f\|_\beta \leq 1 \Rightarrow \|Pf\|_\beta \leq \gamma. \tag{7.4}$$

Assume the claim is proved. Observe that $\|f\|_1 \leq \|f\|_\beta \leq \frac{1}{\beta}\|f\|_1 \leq \frac{1}{\beta}\|f\|_V$. Then

$$\|P^n f\|_1 \leq \|P^n f\|_\beta \leq \gamma^n \|f\|_\beta \leq \gamma^n \beta^{-1} \|f\|_V.$$

Equivalently

$$|P^n f(x) - P^n f(y)| \leq \gamma^n \beta^{-1} \|f\|_V (2 + V(x) + V(y)).$$

Thus,

$$|P^n f(x) - \pi f| \leq \int |P^n f(x) - P^n f(y)| \pi(dy) \leq \gamma^n \beta^{-1} \|f\|_V (2 + V(x) + \pi V).$$

where π is some (hence unique) invariant probability (see Exercise 7.13). This proves the result.

We now pass to the proof of the claim. Let f be such that $\|f\|_\beta \leq 1$.

If $V(x) + V(y) \geq R$. Then

$$|Pf(x) - Pf(y)| = |\int (f(u) - f(v)) \delta_x P(du) \delta_y P(dv)|$$

$$\leq \int |f(u) - f(v)| \delta_x P(du) \delta_y P(dv) \leq 2 + \beta PV(x) + \beta PV(y)$$

$$\leq 2 + 2\beta\kappa + \rho\beta(V(x) + V(y)) \leq \gamma_1(2 + \beta(V(x) + V(y)))$$

where

$$\gamma_1 = \frac{\beta(2\kappa + \rho R) + 2}{\beta R + 2} \in]0, 1[.$$

The last inequality follows from the fact that for all $\rho, r > 0$ and $a \geq 2\rho$,

$$t \geq r \Rightarrow a + \rho t \leq \gamma_1(2 + t)$$

where γ_1 is the solution to $a + \rho r = \gamma_1(2 + r)$. It suffices to set $a = 2 + 2\beta\kappa$ and $r = \beta R$.

Suppose now that $V(x) + V(y) \leq R$. In particular, $x, y \in V_R$. Like in the proof of Theorem 7.7, write $Pf = (1 - \varepsilon)Kf + \varepsilon\psi(f)$ where for all $x \in V_R$, $K(x, \cdot)$ is a Markov operator. Thus

$$|Pf(x) - Pf(y)| = (1 - \varepsilon)|Kf(x) - Kf(y)| \leq (1 - \varepsilon)(1 + \beta(KV(x) + KV(y)).$$

Also $(1 - \varepsilon)KV(x) = PV(x) - \varepsilon\psi V \leq \rho V(x) + \kappa$. Thus

$$|Pf(x) - Pf(y)| \leq (1 - \varepsilon) + 2\beta\kappa + \rho\beta(V(x) + V(y)) \leq \gamma_2(2 + \beta(V(x) + V(y))$$

with $\gamma_2 = max(\rho, 1 - \varepsilon + \beta\kappa)$. Finally it suffices to choose $\beta\kappa < \varepsilon$ and to set $\gamma = \max(\gamma_1, \gamma_2)$. **QED**

Exercise 7.13 **(i)** Suppose that is M is Polish space, P is Feller and that there exists a proper and continuous map $V : M \mapsto \mathbb{R}_+$ satisfying assumption (a) of Theorem 7.12. Show that the set of $\mathsf{Inv}(P)$ is nonempty. Hint: Use Corollary 4.23.

(ii) Suppose only that M is measurable space. Show that $\mathcal{P}_V(M) = \{\mu \in \mathcal{P}(M) : V \in L^1(\mu)\}$ is complete for the distance $d_\beta(\mu, \nu) = |\mu - \nu|_\beta$ defined in the proof of Theorem 7.12. Deduce that, under the assumptions of Theorem 7.12, there exists a unique invariant probability for P. Hint: Use inequality (7.4) to show that

$$|\mu P - \nu P|_\beta \leq \gamma |\mu - \nu|_\beta \tag{7.5}$$

for some $0 \leq \gamma < 1$ and $\beta > 0$.

Corollary 7.14 *Suppose P is Feller and that and that there exists a proper map $V : M \mapsto \mathbb{R}_+$ satisfying assumption (a) of Theorem 7.12. Suppose furthermore that there exists an accessible aperiodic Doeblin point. Then the conclusion of Theorem 7.12 hold true.*

Proof Choose $R > \frac{2\kappa}{(1-\rho)^2}$. The set $C = \{V \leq R\}$ is a compact set (because V is proper) and petite for some P^m by Proposition 7.10. Since $P^m V \leq \rho^m V + \frac{\kappa}{1-\rho}$, Theorem 7.12 applies to P^m and the result follows. **QED**

7.2.2 Coupling, splitting and polynomial convergence

This section is the natural counterpart of Section 2.4 on countable chains. It revisits the convergence theorems of the previous section and relate the rate of convergence to the moments of the return time to a recurrent small set.

Theorem 7.15 *Let $C \subset M$ be an aperiodic, recurrent small set for P.*

(i) *If $\sup_{x \in C} \mathbb{E}_x(\tau_C) < \infty$, then P is positively recurrent and, letting π denote its invariant probability,*

$$|\mu P^n - \pi| \to 0$$

for every probability $\mu \in \mathcal{P}(\mathcal{M})$.

(ii) *If* $\sup_{x\in C}\mathbb{E}_x(\tau_C^p)<\infty$ *for some* $p\geq 2$, *then there exists* $c\geq 0$ *such that for every probability* $\mu\in\mathcal{P}(\mathcal{M})$ *and for every* $n\in\mathbb{N}^*$,

$$|\mu P^n-\pi|\leq\frac{1}{n^{p-1}}c(1+E_\mu(\tau_C^{p-1})).$$

(iii) *If* $\sup_{x\in C}\mathbb{E}_x(e^{\lambda_0\tau_C})<\infty$ *for some* $\lambda_0>0$, *then there exist* $0<\lambda<\lambda_0$ *and* $c\geq 0$ *such that for every probability measure* μ *on* M *and for every* $n\in\mathbb{N}$,

$$|\mu P^n-\pi|\leq e^{-\lambda n}c(1+\mathbb{E}_\mu(e^{\lambda_0\tau_C})).$$

Proof Positive recurrence follows from Theorem 6.17. The rest of the proof relies on a coupling argument that goes back to Harris ***FIX*** and Nummelin [33] . We proceed in two steps.

Step 1. We first assume that C is an *atom*, meaning that there exists a probability ξ on M such that for all $x\in C, P(x,\cdot)=\xi(\cdot)$. In this situation the proof is very much like the proof given for a countable Markov chain (Theorem 2.33). Let (X_n) and (Y_n) be two independent chains (induced by P), $\mathbb{P}_{\mu\otimes\nu}$ the law of $((X_n,Y_n))_{n\geq 0}$ when (X_0,Y_0) has law $\mu\otimes\nu$, and let

$$\tau_{C\times C}=\min\{n\geq 1:\ X_n\in C,Y_n\in C\}.$$

Because C is an atom, for all $\mu,\nu\in\mathcal{P}(\mathcal{M})$

$$\mathbb{P}_{\mu\otimes\nu}(X_n\in\cdot;\tau_{C\times C}<n)=\mathbb{P}_{\mu\otimes\nu}(Y_n\in\cdot;\tau_{C\times C}<n).$$

Hence

$$|\mu P^n-\pi|=|\mu P^n-\pi P^n|\leq\mathbb{P}_{\mu\otimes\pi}(\tau_{C\times C}\geq n).$$

Let now $(\tau_C^{(n)})$ (respectively $(\tilde\tau_C^{(n)})$) denote the successive hitting times of C by (X_n) (respectively (Y_n)). The assumption that C is an aperiodic atom make the processes $T:=(\tau_C^{(n+1)})_{n\geq 0}$ and $\tilde T:=(\tilde\tau_C^{(n+1)})_{n\geq 0}$ two aperiodic independent renewal processes (see Section 2.3) and $\tau_{C\times C}$ is their first common renewal time. The additional assumption that $\sup_{x\in C}\mathbb{E}_x(\tau_C)<\infty$ make these processes L^1 (as defined in Section 2.3) so that $\tau_{C\times C}<\infty$ almost surely (see Equation (2.5) and the discussion preceding it). This proves the first assertion by letting $\nu=\pi$, the invariant probability of P. To prove the second assertion, by Markov inequality and Theorem 2.31 we get that for all $0<q\leq p$,

$$|\mu P^n-\pi|\leq\frac{1}{n^q}\mathbb{E}_{\mu\otimes\pi}(\tau^q_{C\times C})\leq\frac{1}{n^q}(c(1+\mathbb{E}_\mu(\tau_C^q)+\mathbb{E}_\pi(\tau_C^q)).$$

The problem then reduces to estimate $\mathbb{E}_\pi(\tau_C^q)$. Here again, the assumption that C is an atom will prove to be very useful. Like for countable Markov chains, π can be explicitly written as

$$\pi(f) = \frac{\mathbb{E}_x(f(X_1) + \ldots + f(X_{\tau_C}))}{\mathbb{E}_x(\tau_C)} = \pi(C)\mathbb{E}_x(f(X_1) + \ldots + f(X_{\tau_C}))$$

for any $x \in C$ and all $f \geq 0$ measurable. The proof is similar to the proof of assertion (iii) in Theorem 2.6 (compare to Exercise 4.24) and left to the reader. Applying this formulae to the map $y \mapsto \mathbb{E}_y(\psi(\tau_C))$ for some nonnegative function ψ leads to

$$\mathbb{E}_\pi(\psi(\tau_C)) = \pi(C)\mathbb{E}_x(\sum_{k=0}^{\tau_C - 1} \psi(k)),$$

for all $x \in C$, exactly as in Proposition 2.11. In particular

$$\mathbb{E}_\pi(\tau_C^q) \leq \pi(C)\mathbb{E}_x(\tau_C^{q+1})$$

for all $x \in C$. This concludes the proof of the second assertion.

The proof of the third assertion is similar. By Markov inequality and Theorem 2.32 there exists $0 < \lambda \leq \lambda_0$ such that

$$|\mu P^n - \pi| \leq e^{-\lambda n}\mathbb{E}_{\mu\otimes\pi}(e^{\lambda\tau_{C\times C}}) \leq e^{-\lambda n}c(1 + \mathbb{E}_\mu(e^{\lambda_0\tau_C}) + \mathbb{E}_\pi(e^{\lambda_0\tau_C})).$$

On the other hand, for all $x \in C$

$$\mathbb{E}_\pi(e^{\lambda_0\tau_C}) = \Pi(C)\mathbb{E}_x(\frac{e^{\lambda_0\tau_C} - 1}{e^{\lambda_0} - 1}).$$

<u>Step 2</u>. We suppose now that C is a small set with minorizing measure ξ. Let $\varepsilon = \xi(M), \psi(\cdot) = \frac{\xi(\cdot)}{\varepsilon}$ and let K be the kernel on C defined by

$$K(x,\cdot) = \frac{P(x,\cdot) - \varepsilon\psi(\cdot)}{1-\varepsilon}.$$

The idea of the splitting method consists to constructs (X_n) with the help of an auxiliary sequence $(I_n), I_n \in \{0,1\}$. If $X_n \notin C$, then I_n is set to 0. If $X_n \in C$, I_n is randomly chosen according to a Bernoulli distribution with parameter ε. At the next step, X_{n+1} is distributed according to

$$P(X_n,\cdot)\mathbf{1}_{\{X_n\in M\setminus C\}} + [(1-I_n)K(X_n,\cdot) + I_n\psi(\cdot)]\mathbf{1}_{\{X_n\in C\}}.$$

More formally, consider the Markov kernel Q defined on

$$\mathcal{M} = \{(x,i) \in M \times \{0,1\} \,: x \notin C \Rightarrow i = 0\},$$

by: For all $x \in M \setminus C$

$$\begin{aligned} Q(x,0,dy \times \{0\}) &= P(x,dy)(1 - \varepsilon \mathbf{1}_C(y)), \\ Q(x,0,dy \times \{1\}) &= P(x,dy)\varepsilon \mathbf{1}_C(y), \end{aligned}$$

and for all $x \in C$

$$\begin{aligned} Q(x,0,dy \times \{0\}) &= K(x,dy)(1 - \varepsilon \mathbf{1}_C(y)), \\ Q(x,0,dy \times \{1\}) &= K(x,dy)\varepsilon \mathbf{1}_C(y), \\ Q(x,1,dy \times \{0\}) &= \psi(dy)(1 - \varepsilon \mathbf{1}_C(y)), \\ Q(x,1,dy \times \{1\}) &= \psi(dy)\varepsilon \mathbf{1}_C(y). \end{aligned}$$

We let (X_n, I_n) denote the canonical process on $(\Omega, \mathcal{F}) = (\mathcal{M}^{\mathbb{N}}, \mathcal{B}(\mathcal{M})^{\otimes \mathbb{N}})$, $\mathcal{F}_n = \sigma((X_i, I_i)_{i \leq n})$, and for each $\nu \in \mathcal{P}(\mathcal{M}), \mathbb{P}_\nu$ the Markov measure on Ω making (X_n, I_n) a Markov chain with kernel Q (with respect to $(\mathcal{F}_n)$) and initial law ν. As usual we write $\mathbb{P}_{x,i}$ for $\mathbb{P}_{\delta_{(x,i)}}$. We shall also use the following convenient notation:

$$\mathbb{P}_x := \mathbb{P}_{(x,0)} \text{ if } x \in M \setminus C,$$

$$\mathbb{P}_x := (1-\varepsilon)\mathbb{P}_{(x,0)} + \varepsilon \mathbb{P}_{(x,1)} \text{ if } x \in C.$$

Let $\mathcal{G}_n = \sigma((X_i)_{i \leq n})$. It is not hard to verify (but still a good recommended exercise) that

$$\mathbb{P}_\nu(X_{n+1} \in \cdot | \mathcal{G}_n) = P(X_n, \cdot)$$

for all $n \geq 1$ and $\nu \in \mathcal{P}(\mathcal{M})$; and that

$$\mathbb{P}_x(X_1 \in \cdot) = P(x, \cdot).$$

This shows that $(X_n)_{n \geq 0}$ is a Markov chain with kernel P and initial value $X_0 = x$ on $(\Omega, \mathcal{F}, (\mathcal{G}_n), \mathbb{P}_x)$.

We claim that:

(a) $C \times \{1\}$ is a recurrent aperiodic atom for Q;

(b) If for some $p \geq 1$, $\sup_{x \in C} \mathbb{E}_x(\tau_C^p) < \infty$, then there exist $a, b \geq 0$ such that for all $(x,i) \in \mathcal{M}$

$$\mathbb{E}_{x,i}(\tau_{C \times 1}^p) \leq a \mathbb{E}_x(\tau_C^p) + b;$$

(c) If for some $\lambda_0 > 0$, $\sup_{x \in C} \mathbb{E}_x(e^{\lambda_0 \tau_C}) < \infty$, then there exist $a \geq 0$ and $0 < \lambda \leq \lambda_0$ such that for all $(x,i) \in \mathcal{M}$

$$\mathbb{E}_{x,i}(e^{\lambda \tau_{C\times 1}}) \leq a \mathbb{E}_x(e^{\lambda \tau_C}).$$

Assume the claims are proved. Then, by step 1 (X_n, I_n) is positively recurrent, so is (X_n) and $P^n(x,\cdot) = \mathbb{P}_x(X_n \in \cdot)$ converges in total variation, as $n \to \infty$ toward π, the invariant probability of P. If $\sup_{x\in C} \mathbb{E}_x(\tau_C^p) < \infty$ then, by (b) in the claim, $\sup_{x \in C} \mathbb{E}_{x,1}(\tau^p_{C\times 1}) < \infty$. Thus, by step 1,

$$|P^n(x,A) - \pi(A)| = |\mathbb{P}_x(X_n \in A) - \pi(A)| \leq \frac{1}{n^{p-1}} c(1 + \mathbb{E}_x(\tau^p_{C\times 1})) \leq \frac{1}{n^{p-1}} c(1 + a\mathbb{E}_x(\tau^p_C) + b).$$

This proves the second assertion. The third one is similar.

We now prove the claims. Clearly $C \times 1$ is an atom for Q. Identify C with the subset of $\mathcal{M}$ consisting of points (x,i) such that $x \in C$. Then (under this identification) $C \times 1 \subset C$ and we rely on Proposition 6.15 to prove the claim. By the assumption that C is recurrent for P, for all $x \in M$

$$1 = \mathbb{P}_x(\tau_C < \infty) = (1-\varepsilon)\mathbb{P}_{(x,0)}(\tau_C < \infty) + \varepsilon \mathbb{P}_{(x,1)}(\tau_C < \infty).$$

Thus for all $(x,i) \in \mathcal{M}$ $\mathbb{P}_{(x,i)}(\tau_C < \infty)$ showing that C is recurrent for Q. Also,

$$\mathbb{P}_{(x,i)}((X_{\tau_C}, I_{\tau_C}) \in C \times 1) = \varepsilon$$

because, $\mathbb{P}_{(x,i)}((X_{\tau_C}, I_{\tau_C}) \in C \times 1 | \mathcal{G}_{\tau_C}) = \varepsilon$. Thus, by Proposition 6.15 (i), $C \times 1$ is recurrent for Q. We now prove that it is aperiodic. For $x \in C, j, k \geq 1$

$$\begin{aligned}\mathbb{P}_{(x,1)}(X_{j+k} \in C, I_{j+k} = 1) = \varepsilon \mathbb{P}_{(x,1)}(X_{j+k} \in C) &\geq \varepsilon \mathbb{E}_{(x,1)}(\mathbf{1}_{\tau_C = j} P^k(X_{\tau_C}, C)) \\ &\geq \varepsilon \mathbb{P}_{(x,1)}(\tau_C = j) \inf_{x \in C} P^k(x,C).\end{aligned}$$

Since $C \times 1$ is an atom, $\mathbb{P}_{(x,1)}(\tau_C = j)$ doesn't depend on $x \in C$ and is > 0 for some $j = j_0 \geq 1$. By aperiodicity of C for P, there exists $n_0 \in \mathbb{N}$ such that for all $k \geq n_0$

$$\inf_{x \in C} P^k(x,C) > 0.$$

Therefore $\inf_{x\in C} \mathbb{P}_{(x,1)}(X_k \in C, I_k = 1) > 0$ for all $k \geq n_0 + j_0$. This proves aperiodicity and concludes the proof of claim (a).

If $\sup_{x \in C} \mathbb{E}_x(\tau_C^p) < \infty$, then $\sup_{x\in C} \mathbb{E}_{(x,i)}(\tau^p_C) < \infty$ and by Proposition 6.15 (ii) $\sup_{x \in C} \mathbb{E}_{(x,i)}(\tau^p_{C\times 1}) < \infty$. Now

$$\tau_{C\times 1} \leq \tau_C + \tau_{C\times 1} \circ \theta_{\tau_C},$$

so that,

$$\mathbb{E}_{(x,i)}(\tau^p_{C\times 1}) \le 2^{p-1}(\mathbb{E}_{(x,i)}(\tau^p_C) + \sup_{x\in C, i=0,1} \mathbb{E}_{(x,i)}(\tau^p_{C\times 1})) = a\mathbb{E}_{(x,i)}(\tau^p_C) + b.$$

Claim (c) is proved similarly. **QED**

Remark 7.16 It is interesting to compare Theorems 7.12 and 7.15 (iii). Under the assumptions of Theorem 7.12, the set $C = \{V \le R\}$ with $R \ge \frac{2\kappa}{1-\rho}$, satisfies condition (iii) of Theorem 7.15 (with $\lambda_0 = \frac{1-\rho}{2}$). This follows from Proposition 6.11 (iii) or Proposition 6.13 (choose $\phi(s) = \lambda_0 s$). Then, by Theorem 7.15, $|P^n f(x) - \pi(f)| \le e^{-\lambda n} c(1 + V(x))\|f\|_\infty$ for all $f \in B(M)$. Observe however that the conclusion of Theorems 7.12 is stronger, in the sense that it allows to deal with functions that are unbounded but only bounded by $1 + V$.

Appendices

Appendix A

Monotone class and Martingales

A.1 Monotone class theorem

A set $H \subset B(M)$ is said to be *stable by bounded monotone convergence* if $f_n \in H$ and $0 \leq f_n \leq f_{n+1} \leq 1$ implies that $f = \lim_n f_n \in H$.

Theorem A.1 (Monotone class theorem) *Let $H \subset B(M)$ be a vector space of bounded functions containing the constant functions and stable by bounded monotone convergence. Let $C \subset B(M)$ be a set stable by multiplication and let $\sigma(C)$ denote the sigma algebra generated by C (i-e the smallest sigma algebra making the elements of C measurables). If $C \subset H$, then H contains every bounded $\sigma(C)$-measurable function.*

A.2 Conditional expectation

We recall here the definition of conditional expectation and give some of its basic properties. More details and proofs can be found in standard textbooks such as [5].

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, and let $\mathcal{B}$ be a σ-field contained in $\mathcal{F}$. Let X be a real-valued random variable such that $\mathsf{E}(|X|) < \infty$. Then there exists a real-valued random variable Z with $\mathsf{E}(|Z|) < \infty$ such that

(i) Z is $\mathcal{B}-$measurable;

(ii) For all $A \in \mathcal{B}$, we have

$$\mathsf{E}(Z\mathbf{1}_A) = \mathsf{E}(X\mathbf{1}_A).$$

The random variable Z is unique in the following sense: If Z' is any other random variable satisfying $\mathsf{E}(|Z'|) < \infty$ and the conditions in **(i)** and **(ii)**, then $\mathsf{P}(Z' = Z) = 1$. In other words, the space of equivalence classes $L^1(\Omega, \mathcal{B}, \mathsf{P})$ has a unique element Z satisfying the condition in **(ii)**. This element of $L^1(\Omega, \mathcal{B}, \mathsf{P})$ is called the *conditional expectation* of X given $\mathcal{B}$, and is denoted by $\mathsf{E}(X|\mathcal{B})$. If we write $Y = \mathsf{E}(X|\mathcal{B})$ for some $\mathcal{B}$-measurable random variable Y, we mean that Y is a representative of the equivalence class $\mathsf{E}(X|\mathcal{B})$.

One can also define conditional expectation for nonnegative random variables: Let $X : \Omega \to [0, \infty]$ be measurable, i.e. $\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$ for every set $A \subset [0, \infty]$ such that $A \setminus \{\infty\}$ is a Borel subset of $[0, \infty)$. For every $n \in \mathbb{N}$, let $X_n := X \wedge n$ and let Z_n be a $\mathcal{B}$-measurable random variable such that $\mathsf{E}(|Z_n|) < \infty$ and $\mathsf{E}(Z_n \mathbf{1}_A) = \mathsf{E}(X_n \mathbf{1}_A)$ for every $A \in \mathcal{B}$. By changing the values of (Z_n) on a set of measure 0 if necessary, one can assume that $(Z_n(\omega))_{n\in\mathbb{N}}$ is nondecreasing for every $\omega \in \Omega$. The function

$$Z(\omega) := \lim_{n\to\infty} Z_n(\omega)$$

then maps from Ω to $[0, \infty]$ and satisfies the conditions in **(i)** and **(ii)**. If $Z' : \Omega \to [0, \infty]$ is any other random variable satisfying **(i)** and **(ii)**, then $\mathsf{P}(Z = Z') = 1$. On the set of $\mathcal{B}$-measurable functions from Ω to $[0, \infty]$, consider the equivalence relation given by equality P-almost surely. The conditional expectation of X given $\mathcal{B}$, denoted by $\mathsf{E}(X|\mathcal{B})$, is defined as the unique equivalence class that satisfies **(ii)**.

Theorem A.2 (Properties of conditional expectation) *Let X be a random variable, with $\mathsf{E}(|X|) < \infty$ or $X \in [0, \infty]$, and let $\mathcal{B}$ be a σ-field contained in $\mathcal{F}$. Then,*

(i) $\mathsf{E}(\mathsf{E}(X|\mathcal{B})) = \mathsf{E}(X)$*;*

(ii) *If $\mathsf{E}(|X|) < \infty$ (resp. $X \in [0, \infty]$), we have for every $\mathcal{B}$-measurable random variable Y with $\mathsf{E}(|XY|) < \infty$ (resp. $Y \in [0, \infty]$)*

$$\mathsf{E}(XY|\mathcal{B}) = Y\mathsf{E}(X|\mathcal{B}),$$

with the convention that $0 \cdot \infty = 0$;

(iii) *For every σ-field $\mathcal{A}$ contained in $\mathcal{B}$, we have*

$$\mathsf{E}(\mathsf{E}(X|\mathcal{B})|\mathcal{A}) = \mathsf{E}(X|\mathcal{A}).$$

This is often called tower property*.*

A.3 Martingales

Here, we recall the few results from martingale theory that are used in this course. As for conditional expectation, there are many introductory texts on probability theory that provide more details and proofs, e.g. [40] or [5].

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ be a filtered probability space. We let $\mathcal{F}_\infty$ denote the σ field generated by $\cup_{n\geq 0}\mathcal{F}_n$. A sequence (M_n) of adapted (i.e M_n is $\mathcal{F}_n$ measurable) and L^1 real valued random variables is called a *martingale* (respectively, a *submartingale*, respectively a *supermartingale*) if

$$\mathsf{E}(M_{n+1}|\mathcal{F}_n) = M_n \text{ resp. } \geq, \text{ resp. } \leq$$

for all $n \geq 0$.

A simple, but useful consequence of Jensen inequality is the following.

Proposition A.3 *Let (M_n) be a martingale (resp. a submartingale) and ϕ a convex function (resp. a convex non decreasing function) such that $\phi(M_n) \in L^1$; then $(\phi(M_n))$ is a submartingale .*

It is often useful to extends the martingale (sub, super) property to stopping times. Doob's optional stopping theorem shows that this is the case for bounded stopping times.

Theorem A.4 (Optional stopping) *Let $M = (M_n)$ be a martingale (resp. submartingale, supermartingale).*

(i) *If T is a stopping time, then $(M_{n\wedge T})_{n\geq 0}$ is a martingale (resp. submartingale, supermartingale);*

(ii) *If $S \leq T$ are stopping times bounded by some constant N, then*

$$E(M_T|\mathcal{F}_S) = M_S \textit{ resp. } \geq, \textit{ resp. } \leq .$$

Proof (i) For all $n \in \mathbb{N}$

$$M_{n+1\wedge T} - M_{n\wedge T} = (M_{n+1} - M_n)\mathbf{1}_{\{T>n\}}.$$

Taking the conditional expectation with respect to $\mathcal{F}_n$ proves the result.

(ii) Assume (M_n) is a martingale. Proving that $E(M_T|\mathcal{F}_S) = M_S$ amounts to prove that for all $A \in \mathcal{F}_S$ and $0 \leq k \leq N$, $\mathsf{E}(M_T\mathbf{1}_{A\cap\{S=k\}}) = \mathsf{E}(M_k\mathbf{1}_{A\cap\{S=k\}}))$.

$$\mathsf{E}(M_T\mathbf{1}_{A\cap\{S=k\}}) = \sum_{i=k}^{N}\mathsf{E}(M_i\mathbf{1}_{\{T=i\}}\mathbf{1}_{A\cap\{S=k\}}) = \sum_{i=k}^{N}\mathsf{E}(\mathsf{E}(M_N|\mathcal{F}_i)\mathbf{1}_{\{T=i\}}\mathbf{1}_{A\cap\{S=k\}}))$$

$$= \sum_{i=k}^{N} \mathsf{E}(\mathsf{E}(M_N \mathbf{1}_{\{T=i\}} \mathbf{1}_{A\cap\{S=k\}} | \mathcal{F}_i)) = \mathsf{E}(M_N \mathbf{1}_{A\cap\{S=k\}})$$
$$= \mathsf{E}(\mathsf{E}(M_N \mathbf{1}_{A\cap\{S=k\}} | \mathcal{F}_k)) = \mathsf{E}(M_k \mathbf{1}_{A\cap\{S=k\}}).$$

The proof for sub and supermartingales is similar. **QED**

Corollary A.5 (Doob's inequality) *Let (X_n) be a non negative sub martingale. Then for all $\alpha > 0$*

$$\mathsf{P}(\sup_{0\le i\le N} X_i \ge \alpha) \le \frac{\mathsf{E}(X_N)}{\alpha}.$$

Proof Let $T = \min\{i \ge 0 \ : X_i \ge \alpha\}$. Then $T \wedge N$ is a stopping time bounded by N, so that by the optional stopping theorem

$$\mathsf{E}(X_N) \ge \mathsf{E}(X_{N\wedge T}) = \mathsf{E}(X_N \mathbf{1}_{T>N}) + \mathsf{E}(X_T \mathbf{1}_{T\le N}) \ge \alpha \mathsf{P}(T \le N).$$

QED

The two following theorems are classical convergence results due to Doob.

Theorem A.6 *Let (M_n) be a submartingale. Assume that $\sup_n \mathsf{E}(M_n^+) < \infty$. Then there exists $M_\infty \in L^1$ such that $M_n \to M_\infty$ almost surely.*

Theorem A.7 *Let (M_n) be a martingale. Then the following assertions are equivalent:*

(a) *(M_n) is uniformly integrable;*

(b) *(M_n) converges almost surely and in L^1 to some random variable M_∞;*

(c) *$M_n = \mathsf{E}(M|\mathcal{F}_n)$ for some $M \in L^1$.*

Furthermore, in case (c) $\lim_{n\to\infty} M_n = M_\infty = \mathsf{E}(M|\mathcal{F}_\infty)$.

Let (M_n) be an L^2 martingale (i.e $M_n \in L^2$), the *quadratic variation* of (M_n) is the process $(\langle M\rangle_n)$ recursively defined as

$$\langle M\rangle_0 = 0, \langle M\rangle_{n+1} - \langle M\rangle_n = \mathsf{E}((M_{n+1} - M_n)^2 | \mathcal{F}_n) = \mathsf{E}(M_{n+1}^2 | \mathcal{F}_n) - M_n^2.$$

Note that $(\langle M\rangle_n)$ is nondecreasing, predictable (i.e M_n is $\mathcal{F}_{n-1}$ measurable) and that $(M_n^2 - \langle M\rangle_n)_n$ is a zero mean martingale. We let $\langle M\rangle_\infty = \lim_{n\to\infty} \langle M\rangle_n$.

Theorem A.8 (Strong law of large numbers) *Let (M_n) be a L^2 martingale. Then,*

(i) *If $\mathsf{E}(\langle M\rangle_\infty) = \sum_{k\geq 0} \mathsf{E}((M_{k+1} - M_k)^2) < \infty$, then (M_n) converges almost surely and in L^2 to some random variable M_∞;*

(ii) *On $\langle M_\infty\rangle < \infty$ (M_n) converges almost surely to some finite random variable M_∞;*

(iii) *On $\langle M\rangle_\infty = \infty$ $\lim_{n\to\infty} \frac{M_n}{\langle M\rangle_n} = 0$ a.s.*

(iv) *If $\sup_n \mathsf{E}(\frac{\langle M\rangle_n}{n}) < \infty$, then $\lim_{n\to\infty} \frac{M_n}{n} = 0$ a.s.*

Proof We only prove the last statement, which is sufficient in this book and whose proof is short.By Doob's inequality, for all $n \in \mathbb{N}$

$$\mathsf{P}(\sup_{2^n\leq k\leq 2^{n+1}} \frac{|M_k|}{k} \geq \varepsilon) \leq \mathsf{P}(\sup_{k\leq 2^{n+1}} |M_k|^2 \geq \varepsilon^2 2^{2n}) \leq \frac{1}{\varepsilon^2 2^{2n}} \mathsf{E}(\langle M\rangle_{2^n}) \leq C\frac{1}{\varepsilon^2 2^n}$$

and the result follows from Borel-Cantelli lemma **QED**

Bibliography

[1] David Aldous and Jim Fill, *Reversible markov chains and random walks on graphs.*

[2] Frank Aurzada, Hanna Döring, Marcel Ortgiese, and Michael Scheutzow, *Moments of recurrence times for Markov chains*, Electron. Commun. Probab. **16** (2011), 296–303. MR 2811181

[3] M. Benaïm, S. Le Borgne, F. Malrieu, and P-A. Zitt, *Qualitative properties of certain piecewise deterministic markov processes*, Annales de l'IHP **51** (2015), no. 3, 1040 – 1075.

[4] M. Benaïm and C. Lobry, *Lotka Volterra in fluctuating environment or "how switching between beneficial environments can make survival harder"*, Annals of Applied Probability (2016), no. 6, 3754–3785.

[5] Michel Benaim and Nicole El Karoui, *Promenade alÃl'atoire : ChaÃőnes de markov et simulation*, Editions de l'Ãl'cole Polytechnique, 2005.

[6] Michel Benaïm, Stéphane Le Borgne, Florent Malrieu, and Pierre-André Zitt, *Qualitative properties of certain piecewise deterministic Markov processes*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015), no. 3, 1040–1075. MR 3365972

[7] P. Billingsley, *Convergence of probability measures. Second edition*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc.,New York, 1999.

[8] H. H. Blumenthal, R. M.; Corson, *On continuous collections of measures.*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability theory (1972), 33–34.

[9] K. L. Chung, *Contributions to the theory of Markov chains. II*, Trans. Amer. Math. Soc. **76** (1954), 397–419. MR 63603

[10] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ, *Ergodic theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 245, Springer-Verlag, New York, 1982, Translated from the Russian by A. B. Sosinskiĭ. MR 832433

[11] Claude Dellacherie and Paul-André Meyer, *Probabilités et potentiel. Chapitres IX à XI*, revised ed., Publications de l'Institut de Mathématiques de l'Université de Strasbourg [Publications of the Mathematical Institute of the University of Strasbourg], XVIII, Hermann, Paris, 1983, Théorie discrète du potential. [Discrete potential theory], Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics], 1410. MR 727641

[12] Persi Diaconis and David. Freedman, *Iterated random functions*, SIAM Review **41** (1999), 45–76.

[13] R. M. Dudley, *Real Analysis and Probability* , Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, 2002.

[14] M. Duflo, *Random Iterative Models*, Springer, 2000.

[15] James Dugundji, *Topology*, Allyn and Bacon, Inc., Boston, Mass., 1966. MR 0193606

[16] Manfred Einsiedler and Thomas Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011. MR 2723325

[17] Gerald B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication. MR 1681462

[18] Alex Furman, *A brief introduction to ergodic theory*, Unpublished, http://homepages.math.uic.edu/ furman/papers.html, 2012.

[19] H. Furstenberg, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.

[20] David Gilbarg and Neil S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001, Reprint of the 1998 edition. MR 1814364

[21] M. Hairer and J. Mattingly, *Yet another look at harris ergodic theorem for markov chains*, Seminar on Stochastic Analysis, Random Fields and Applications VI (Dalang, ed.), Progress in Probability, 2011, pp. 109–117.

[22] Martin Hairer, *Ergodic properties of markov processes*, Lecture given at The University of Warwick, 2006.

[23] ———, *Ergodic properties of a class of non-Markovian processes*, Trends in stochastic analysis, London Math. Soc. Lecture Note Ser., vol. 353, Cambridge Univ. Press, Cambridge, 2009, pp. 65–98. MR 2562151

[24] ———, *Convergence of markov processes*, Unpublished, http://hairer.org/notes/Convergence.pdf), 2016.

[25] Martin Hairer and Jonathan C. Mattingly, *Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing*, Ann. of Math. (2) **164** (2006), no. 3, 993–1032. MR 2259251

[26] Anatole Katok and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995, With a supplementary chapter by Katok and Leonardo Mendoza. MR 1326374

[27] Yuri Kifer, *Ergodic theory of random transformations*, Progress in Probability and Statistics, vol. 10, Birkhäuser Boston, Inc., Boston, MA, 1986. MR 884892

[28] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson. MR 2466937

[29] Torgny Lindvall, *Lectures on the coupling method*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1992, A Wiley-Interscience Publication. MR 1180522

[30] R. Mané, *Ergodic theory and differentiable dynamics*, Results in Mathematics and Related Areas, vol. 8, Springer-Verlag, Berlin, 1987, Translated from the Portuguese by Silvio Levy.

[31] S. P. Meyn and R. L. Tweedy, *Markov Chains and Stochastic Stability*, Springer, 1993, Revised in 2005, version availaible on the web.

[32] J. Moser, *On the volume elements on a manifold.*, Trans. Amer. Math. Soc. (1965), 286–294.

[33] Esa Nummelin, *General irreducible Markov chains and nonnegative operators*, Cambridge Tracts in Mathematics, vol. 83, Cambridge University Press, Cambridge, 1984. MR 776608

[34] J. W. Pitman, *Uniform rates of convergence for Markov chain transition probabilities*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **29** (1974), 193–227. MR 373012

[35] Anthony N. Quas, *On representations of Markov chains by random smooth maps*, Bull. London Math. Soc. **23** (1991), no. 5, 487–492. MR 1141021

[36] Philippe Robert, *Stochastic networks and queues*, french ed., Applications of Mathematics (New York), vol. 52, Springer-Verlag, Berlin, 2003, Stochastic Modelling and Applied Probability. MR 1996883

[37] Laurent Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. MR 1490046

[38] A. N. Shiryaev, *Probability*, second ed., Graduate Texts in Mathematics, vol. 95, Springer-Verlag, New York, 1996, Translated from the first (1980) Russian edition by R. P. Boas. MR 1368405

[39] Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR 2459454

[40] David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991. MR 1155402

[41] Wolfgang Woess, *Random walks on infinite graphs and groups*, Cambridge Tracts in Mathematics, vol. 138, Cambridge University Press, Cambridge, 2000. MR 1743100

Index