# Markov Chains on Metric Spaces. A short Course

Michel Benaïm & Tobias Hurth
Institut de Mathématiques,
Neuchâtel University

version 20/06/2022

In the memory of Marie Duflo

**For copyright reasons, this preview only includes the table of contents, the preface, and the first chapter of the book.**

# Contents

# Preface

This book is based on a series of lectures given over the recent years in Master's courses in probability. It provides a short, self-contained introduction to the ergodic theory of Markov chains in metric spaces.

Although primarily intended for graduate and postgraduate students, certain chapters (e.g. one and two) can be taught at the undergraduate level. Others (e.g. four and five) can be used as complements to courses in measure or ergodic theory. Basic knowledge in probability, measure theory, and calculus is recommended. A certain familiarity with discrete-time martingales is also useful, but the few results from martingale theory used in this book are all recalled in the appendix. Each chapter contains several exercises ranging from simple applications of the theory to more advanced developments and examples.

Whether in physics, engineering, biology, ecology, economics or elsewhere, Markov chains are frequently used to describe the random evolution of complex systems. The understanding and analysis of these systems requires, first of all, a good command of the mathematical techniques that allow to explain the long-term behavior of a general Markov chain living on a (reasonable) metric space. Presenting these techniques is, briefly put, our main objective. Questions that are central to this book and that will be recurrently visited are: under which conditions does such a chain have an invariant probability measure? If such a measure exists, is it unique? Does the empirical occupation measure of the chain converge? Does the law of the chain converge, and if so, in which sense and at which rate?

There are a variety of tools to address these questions. Some rely on purely measure-theoretic concepts that are natural generalizations of the ones developed for countable chains (i.e. chains living on countable state spaces). This includes notions of *irreducibility*, *recurrence* (in the sense of Harris), *petite and small sets*, etc. Other tools assume topological properties of the chain such as the *strong Feller* or *asymptotically strong Feller property* (in the sense of Hairer and Mattingly). However, when dealing with a specific model, measure-theoretic conditions - such as irreducibility - might be difficult to verify, and strong topological properties - such as the strong Feller condition - are seldom satisfied. A powerful approach is then to combine much weaker topological conditions - such as the (weak) Feller condition - with controllability properties of the system to prove that certain measure-theoretic conditions (e.g. irreducibility, existence of petite or small sets) are satisfied. This approach is largely developed here and is a key feature of this book.

The book is organized in eight chapters and a short appendix. Chapter 1 briefly defines Markov chains and kernels and gives their very first properties, the Markov and strong Markov properties. The end of the chapter gives a concise introduction to Markov chains in continuous time, also called Markov processes, as they appear in many examples throughout the book.

Chapter 2 is a self-contained mini course on countable Markov chains. Classical notions of *recurrence* (*positive* and *null*) and *transience* are introduced. These are powerful notions, but when students meet them for the first time and have to verify that a specific chain is either recurrent or transient, they are often disoriented. Thus, we have chosen to spend some time here to show how theses properties can be verified "in practice" with the help of suitable *Lyapunov functions*. We also explain how Lyapunov functions can be used to provide estimates on the moments (polynomial and exponential) of *hitting times* for a point or a finite set.

Certainly one of the most important results in the theory of countable chains is the *ergodic theorem*, which asserts that - for positive recurrent aperiodic chains - the law of the chain converges to a unique distribution. The final three sections of Chapter 2 are organized around this result. We first prove it quickly - by standard coupling - without any estimate on the rate of convergence. Then, the Lyapunov method is applied to investigate the behavior of *renewal processes* and provide short proofs of coupling theorems for these processes. Finally, relying on these coupling results, we revisit the ergodic theorem, this time with some convergence rates.

On uncountable state spaces, the simplest (and also the most natural) examples of Markov chains are given by *random dynamical systems* (also called *random iterative systems*). These are systems such that the state variable at time $n+1$ is a deterministic function of the state variable at time $n$ and a "random" input sampled from a sequence of i.i.d. random variables. Chapter 3 is devoted to this type of chains and explains how any given "abstract" Markov chain can be represented by a random dynamical system. Some interesting examples (Bernoulli convolutions, Propp-Wilson algorithm) are presented in exercises.

Chapter 4 starts with a detailed section on *weak convergence, tightness* and *Prohorov's theorem*. Then, *invariant probability measures* are defined and it is shown that, for a Feller chain, weak limit points for the family of *empirical occupation measures* are almost surely invariant probability measures. We discuss some practical tightness criteria (for the empirical occupation measures) based on Lyapunov functions. At this stage of the book, the reader understands that, under a reasonable control of the chain at infinity (obtained for instance by a Lyapunov function), uniqueness of the invariant probability measure equates stability: the empirical occupation measures converge almost surely to some (unique) distribution, regardless of the initial distribution. So we found it was a good place to discuss simple examples of *uniquely ergodic* chains (i.e. chains having a unique invariant probability measure). This is done in the third section of Chapter 4, where we analyze random dynamical systems obtained by random composition of contractions (or mappings that contract on average). The penultimate section of the chapter is devoted to ergodic theorems. We first prove several classical results (*Poincaré recurrence theorem, Birkhoff ergodic theorem*, and the *ergodic decomposition theorem*) and then show how they can be applied to Markov chains. Finally, we discuss invariant measures of continuous-time processes and explain how their properties (existence, ergodicity, uniqueness, ergodic decomposition, etc.) can be studied using discrete-time theory.

Chapter 5 is devoted to various notions of *irreducibility* which ensure unique ergodicity. We start with the measure-theoretic notion of irreducibility (also called $\psi$ irreducibility) and then move on to more topological conditions. The *accessible set* of a Feller chain is introduced and its relations with the support of invariant probability measures are investigated. We then consider strong Feller chains and prove that for such chains ergodic probability measures have disjoint support. We also prove the Hairer-Mattingly theorem, which says that the same property holds under the weaker assumption that the chain is *asymptotically strong Feller*. These results have the useful consequence that, on a connected set, if there is an invariant probability measure having full support, the chain is uniquely ergodic.

We then discuss in Chapter 6 the notions of *petite sets, small sets* and *(weak) Doeblin points* and show that the existence of an accessible weak Doeblin point implies irreducibility for (weak) Feller chains. This latter result is then applied to a variety of examples both in discrete time (random dynamical systems, random dynamical systems obtained by random switching between deterministic flows) and in continuous time (piecewise deterministic Markov processes, stochastic differential equations). This gives us the opportunity to show how the accessibility condition is naturally expressed as a control problem and how the Doeblin properties are naturally related to HÃűrmander type conditions (for random switching models, piecewise deterministic Markov processes and SDEs).

Chapter 7 introduces *Harris recurrence*. For uniquely ergodic chains, Harris recurrence equates to *positive recurrence*, meaning that for every bounded Borel (and not merely for every continuous) function, the Birkhoff averages of the function converge almost surely. We prove the important result that Harris recurrence (respectively positive recurrence) is implied by the existence of a *recurrent petite set* (respectively a petite set whose first return time is bounded in $L^1$). We also discuss simple useful criteria (relying on Lyapunov functions) ensuring that a set is recurrent and provide moment estimates on the return times.

Chapter 8 revolves around the celebrated Harris ergodic theorem. After revisiting the notions of *total variation distance* and *coupling* for two probability measures, we state a simple version of the Harris ergodic theorem where the entire state space is a petite set. Under this strong hypothesis, one has exponential convergence in total variation distance to the unique invariant probability measure. The same conclusion holds under the existence of a Lyapunov function that forces the Markov chain to enter a certain small set - a condition that is better adapted to noncompact state spaces, which are usually not petite. We give two different proofs for this latter version of Harris's ergodic theorem: first the recent proof by Hairer and Mattingly based on the ingenious construction of a semi-norm for which the Markov operator is a contraction. And second, a more classical proof using coupling arguments and ideas from renewal theory. More precisely, under uniform estimates on polynomial (respectively exponential) moments for the return times to an aperiodic and recurrent small set, we obtain polynomial (respectively exponential) convergence in total variation distance to the unique invariant probability measure. Finally, we present a condition, also due to Hairer and Mattingly, that yields exponential convergence to the unique invariant probability measure in a certain Wasserstein distance.

The appendix recalls the monotone class theorem and the few results from discrete time martingales that are used in the book.

More advanced textbooks include the excellent classical books by Meyn and Tweedie [49] and Duflo [22] and the more recent book by Douc, Moulines Priouret and Soulier [20]. The lecture notes by Hairer [31] contain some similar material and are also highly recommended.

**Acknowledgments** We first thank the Swiss National Science Foundation for its constant support over the years. This book has been prepared while we were working on the projects 200021-175728 and 200020-196999. We are grateful to Yuri Bakhtin, Jonathan Mattingly, and Tom Mountford for valuable remarks. This book has also indirectly benefited from the numerous discussions we have had during its preparation with our colleagues, students, and post-docs: Jean Baptiste Bardet, Antoine Bourquin, Charles-Edouard Brehier, Nicolas Champagnat, Bertrand Cloez, Carl-Eric Gauthier, Eva Locherbach, Florent Malrieu, Laurent Miclo, Pierre Monmarché, William Oçafrain, Fabien Panloup, Edouard Strickler, Sebastian Schreiber, Oliver Tough, Denis Villemonais, Pierre-André Zitt, Alexandre Zumbrunnen.

## Preliminaries

The general setting is the following. Throughout all this book, 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 $\sigma$-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

$$\|f\|_\infty := \sup_{x\in M} |f(x)|. \tag{1}$$

If $\mu$ 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 $\mu$. The rest of the notation is introduced in the main body of the text. Please also refer to the list of symbols at the end of the book.

# Chapter 1

# Markov Chains

This chapter introduces the basic objects of the book: Markov kernels and Markov chains. The Chapman Kolmogorov equation which characterizes the evolution of the law of a Markov chain, as well as the Markov and strong Markov properties are established. The last section briefly defines continuous time Markov processes.

## 1.1 Markov kernels

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:

$$Pg(x) := \int_M P(x, dy) g(y), \tag{1.1}$$

$$\mu P(G) := \int_M \mu(dx) P(x, G). \tag{1.2}$$

**Remark 1.1** For all $g \in B(M)$, we have $Pg \in B(M)$ and $\|Pg\|_\infty \le \|g\|_\infty$. 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 (1.1) 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 2.3.

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 $d(x, y) = \mathbf{1}_{x \ne y}$. The corresponding Borel $\sigma$-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.

## 1.2 Markov chains

In order to define Markov chains, we first need to introduce the (classical) notions of filtration and adapted processes. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. A *filtration* $\mathbb{F} = (\mathcal{F}_n)_{n \ge 0}$ is an increasing sequence of $\sigma$-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 $n \in \mathbb{N}$. If $X = (X_n)_{n\geq 0}$ is a family of random variables on $(\Omega, \mathcal{F}, \mathsf{P})$, the *canonical filtration* of $X$ is the filtration $\mathbb{F}^X = \{\mathcal{F}_n^X\}_{n\geq 0}$ where $\mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$ is the $\sigma$-field generated by $X_0, \dots, X_n$. With such a definition $X$ is always an *adapted* stochastic process on $(\Omega, \mathcal{F}, \mathbb{F}^X, \mathsf{P})$.

We can now define what a Markov chain is. Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ and a Markov kernel $P$ on $M$, a *Markov chain with kernel $P$ with respect to* $\mathbb{F}$ 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 $g \in B(M)$ (or 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.

**Proposition 1.4** *Let $(X_n)$ be a Markov chain with kernel $P$ with respect to $\mathbb{F}$. Then $(X_n)$ is always a Markov chain with kernel $P$ with respect to $\mathbb{F}^X$. This latter property is equivalent to*

$$\mathsf{E}(g(X_{n+1})h_0(X_0)...h_n(X_n)) = \mathsf{E}(Pg(X_n)h_0(X_0)...h_n(X_n))$$

*for all $n \in \mathbb{N}$, $h_0, \dots, h_n \in B(M)$, and $g \in B(M)$.*

**Proof** Suppose that $(X_n)$ is a Markov chain with kernel $P$ with respect to $\mathbb{F}$. Since $\mathcal{F}_n^X \subset \mathcal{F}_n$,

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

This proves the first statement. Multiplying the left hand side and right hand side of this equality by $h_0(X_0)...h_n(X_n)$ **QED**

**Remark 1.5** In view of Proposition 1.4, when we say that $(X_n)$ is a Markov chain with kernel $P$, we implicitly mean that it is a Markov chain with respect to $\mathbb{F}^X$.

**Proposition 1.6 (Chapman-Kolmogorov Equation)** *Let $(X_n)$ be a Markov chain with kernel $P$. Let $\mu_n$ 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 1.7 (countable space)** Let $(X_n)$ be a Markov chain on a countable state space $M$, with transition matrix $P$ and initial distribution $\mu_0$. The law $\mu_n$ 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 $n$th power of the matrix $P$. In matrix-vector notation, this identity can be written as

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

where $\mu_n$ and $\mu_0$ are row vectors. In particular, if $\mu_0$ 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 $\{x\}$, i.e.,

$$\mathsf{P}(X_n = x | X_0 = y) = P^n(y, x).$$

### Feller and strong Feller chains

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.

## 1.3 The canonical 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 $\sigma$-field $\mathcal{B}(M)^{\otimes\mathbb{N}}$ (see Exercise 1.9).

If $X_0$ has law $\nu$, we let $\mathbb{P}_\nu$ denote the law of $X$. That is the image measure of $\mathsf{P}$ by $X$. In particular, for all Borel sets $A_0, \ldots, A_k \subset M$,

$$\mathsf{P}(X_0 \in A_0, \ldots, X_k \in A_k) = \mathbb{P}_\nu\{\mathbf{x} \in M^{\mathbb{N}} : (x_0, \ldots, x_k) \in A_0 \times \ldots \times A_k\} \quad (1.3)$$

We let $\mathbb{E}_\nu$ denote the corresponding expectation. If $\nu$ is the Dirac measure at $x$, we use the standard notation $\mathbb{P}_x := \mathbb{P}_{\delta_x}$ and $\mathbb{E}_x := \mathbb{E}_{\delta_x}$.

**Proposition 1.8** **(i)** *Let $X = (X_n)_{n\geq 0}$ be a Markov chain with kernel $P$ and initial distribution $\nu$. Then for all Borel sets $A_0, \ldots, A_k \subset M$,*

$$\mathbb{P}_\nu\{\mathbf{x} \in M^{\mathbb{N}} : \, (x_0, \ldots, x_k) \in A_0 \times \ldots \times A_k\} = \int_{A_0} \nu(dx_0) \int_{A_1} P(x_0, dx_1) \ldots \int_{A_k} P(x_{k-1}, dx_k). \quad (1.4)$$

**(ii)** *Let $\Omega = M^{\mathbb{N}}$, and let $\mathcal{F} = \mathcal{B}(M)^{\otimes\mathbb{N}}$. Given a probability measure $\nu$ and a Markov kernel $P$ on $M$, there exists a unique probability measure $\mathbb{P}_\nu$ on $(\Omega, \mathcal{F})$ characterized by (1.4). On $(\Omega, \mathcal{F})$, the process $(X_n)_{n\geq 0}$ defined by $X_n(\mathbf{x}) = x_n$, is a Markov chain with kernel $P$ and initial law $\nu$, called the* canonical chain*.*

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

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

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

$$\begin{aligned} \mathsf{E}(h_0(X_0) \ldots h_k(X_k)) &= \mathbb{E}_\nu(h_0 \otimes \ldots \otimes h_k) \\ = \mathbb{E}_\nu(h_0 \otimes \ldots \otimes h_{k-1} P h_k) &= \nu[h_0 P[h_1 P[\ldots h_{k-1} P h_k] \ldots]]. \end{aligned} \quad (1.5)$$

The first equality is by definition of $\mathbb{E}_\nu$. The second equality follows from Proposition 1.4 and the last one follows from the second one by induction on $k$. This proves the first statement.

The existence of a unique probability measure $\mathbb{P}_\nu$ on $(\Omega, \mathcal{F})$ characterized by (1.4) is the celebrated Ionescu-Tulcea theorem (see, e.g., Theorem 2 in Chapter II.9 of [63]). Using the result from Exercise 1.9, it is not hard to check that the canonical process $(X_n)$ is a Markov chain on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}^X, \mathbb{P}_\nu)$, with initial distribution $\nu$ and kernel $P$. **QED**

**Exercise 1.9** Let $\mathcal{B}(M^n)$ (respectively $\mathcal{B}(M^{\mathbb{N}})$) denote the Borel $\sigma$-field over $M^n$ (respectively $M^{\mathbb{N}}$, endowed with the product topology). Let $\mathcal{B}(M)^{\otimes n}$ (respectively $\mathcal{B}(M)^{\otimes \mathbb{N}}$) denote the product $\sigma$-field over $M^n$ (respectively $M^{\mathbb{N}}$). Show that $\mathcal{B}(M)^{\otimes n} = \mathcal{B}(M^n)$ and $\mathcal{B}(M)^{\otimes \mathbb{N}} = \mathcal{B}(M^{\mathbb{N}})$.

*Hint:* For the inclusion $\subset$ one can use the fact that the projection $\pi_i : M^{\mathbb{N}} \to M, \mathbf{x} \mapsto x_i$ is continuous, hence measurable. Observe that this doesn't require the separability of $M$. For the converse implication, one can first show, using separability, that every open subset of $M^n$ is a countable union of product sets $O_1 \times \ldots \times O_n$ with $O_i$ open.

## 1.4 Markov and strong 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 1.10 (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 1.9, 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 (1.5), 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 *$\sigma$-field generated by $T$*, denoted $\mathcal{F}_T$, is the $\sigma$-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 1.11** **(i)** Show that $\mathcal{F}_T$ is indeed a $\sigma$-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 1.10.

**Proposition 1.12 (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**

## 1.5 Continuous time: Markov Processes

Although this book is about Markov chains in discrete time, it is useful to say a few words about Markov chains in continuous time, also called Markov processes, because they appear in many examples throughout the book. The definitions are modeled on discrete time.

A *Markov semigroup* on $M$ is a family $\{P_t\}_{t\geq 0}$ of Markov kernels on $M$ such that

**(i)** $P_0(x,\cdot) = \delta_x$;

**(ii)** For all $G \in \mathcal{B}(M)$, the mapping $(t,x) \to P_t(x,G)$ is measurable;

**(iii)** For all $t,s \geq 0$, $P_{t+s} = P_t \circ P_s$.

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and let $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ be a *continuous-time filtration*, i.e., a family of $\sigma$-fields such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for all $0 \leq s \leq t$. An $M$-valued *adapted* stochastic process on $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ is a family $(X_t)_{t\geq 0}$ of random variables defined on $(\Omega, \mathcal{F}, \mathsf{P})$, taking values in $M$ and such that $X_t$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.

A *Markov process* with semigroup $\{P_t\}_{t\geq 0}$ with respect to $\mathbb{F}$ is an adapted stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathsf{P})$ such that for all $g \in B(M)$ and $t,s \geq 0$,

$$\mathsf{E}(g(X_{t+s})|\mathcal{F}_t) = (P_s g)(X_t).$$

**Exercise 1.13** Suppose $M$ is countable. Let $(Y_n)$ be a Markov chain on $M$ with kernel $P$. Let $U_1, U_2, \ldots$ be a sequence of independent identically distributed random variables on $(0,\infty)$ having an exponential distribution of parameter $\lambda$, i.e., $\mathsf{P}(U_i > t) = e^{-\lambda t}$. Set $T_0 = 0$ and $T_n = U_1 + \ldots + U_n$ for $n \geq 1$. Let $(X_t)_{t\geq 0}$ be the continuous-time process defined by $X_t = Y_n$ for $T_n \leq t < T_{n+1}$. Show that $(X_t)$ is a Markov process with semigroup

$$P_t = e^{-\lambda t} e^{\lambda t P} := e^{-\lambda t} \sum_{k\geq 0} \frac{(\lambda t)^k P^k}{k!}.$$

### Feller processes

We use the following terminology. We say that the Markov semigroup $\{P_t\}_{t\geq 0}$ is *weak Feller* provided that

**(i)** $P_t(C_b(M)) \subset C_b(M)$ for all $t \geq 0$;

**(ii)** For all $f \in C_b(M)$ and $x \in M$, $\lim_{t \downarrow 0} P_t f(x) = f(x)$.

This definition implies that $P_t$ is Feller for all $t \geq 0$. Observe however that it is weaker than the usual definition of a *Feller semigroup* (see, e.g., [26], [59] or [45]), which assumes that

**(i)** $M$ is a locally compact metric space;

**(ii)** $\{P_t\}_{t \geq 0}$ is a strongly continuous semigroup on $C_0(M)$ (the set of continuous functions vanishing at infinity), meaning that

- **(a)** $P_t(C_0(M)) \subset C_0(M)$;
- **(b)** For all $f \in C_0(M)$, $\lim_{t \downarrow 0} \|P_t f - f\|_\infty = 0$.

**Remark 1.14** It is proved in [59, Proposition 2.4] that $[(a), (b)]$ above is equivalent to $[(a), (b)']$ where $(b)'$ is given by the (seemingly) weaker condition that

$$\lim_{t \downarrow 0} P_t f(x) = f(x)$$

for all $f \in C_0(M)$ and $x \in M$. As shown by the following exercise, this equivalence does not hold if $C_0(M)$ is replaced by $C_b(M)$.

**Exercise 1.15** Let $M = (0, \infty)$, and let $P_t$ be defined on $B(M)$ as

$$P_t f(x) = f\left(\frac{xe^t}{1 + x(e^t - 1)}\right).$$

Show that $\{P_t\}_{t \geq 0}$ is a weak Feller Markov semigroup which is not Feller.

# Bibliography


[1] D. Aldous and J. Fill, *Reversible Markov chains and random walks on graphs*.

[2] L. Arnold and W. Kliemann, *On unique ergodicity for degenerate diffusions*, Stochastics **21** (1987), no. 1, 41–61. MR 899954

[3] F. Aurzada, H. Döring, M. Ortgiese, and M. Scheutzow, *Moments of recurrence times for Markov chains*, Electron. Commun. Probab. **16** (2011), 296–303.

[4] Y. Bakhtin, *Some topics in ergodic theory*, https://cims.nyu.edu/%7Ebakhtin/2015-ergodic.pdf, 2015.

[5] Y. Bakhtin and T. Hurth, *Invariant densities for dynamical systems with random switching*, Nonlinearity **25** (2012), no. 10, 2937–2952. MR 2979976

[6] M Benaim, *Stochastic persistence*, arXiv preprint arXiv:1806.08450 (2018).

[7] M. Benaim and N. El Karoui, *Promenade aléatoire : Chaînes de Markov et Simulation*, Editions de l'Ecole Polytechnique, 2005.

[8] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt, *Qualitative properties of certain piecewise deterministic Markov processes*, Ann. Inst. Henri Poincaré Probab. Stat. **51** (2015), no. 3, 1040–1075.

[9] M. Benaïm and C. Lobry, *Lotka-Volterra with randomly fluctuating environments or "How switching between beneficial environments can make survival harder"*, Ann. Appl. Probab. **26** (2016), no. 6, 3754–3785.

[10] M BenaÃŤm, N Champagnat, W OÃĄafrain, and D Villemonais, *Degenerate processes killed at the boundary of a domain*, arXiv preprint arXiv:2103.08534 (2021).

[11] P. Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999, A Wiley-Interscience Publication.

[12] R. M. Blumenthal and H. H. Corson, *On continuous collections of measures*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, 1972, pp. 33–40.

[13] J. M. Bony, *Principe du maximum, inégalite de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés*, Ann. Inst. Fourier (Grenoble) **19** (1969), no. fasc. 1, 277–304 xii. MR 262881

[14] S. Cerrai, *A Hille–Yosida theorem for weakly continuous semigroups*, Semigroup Forum, vol. 49, Springer, 1994, pp. 349–367.

[15] W. L Chow, *Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann. **117** (1939), 98–105. MR 1880

[16] K. L. Chung, *Contributions to the theory of Markov chains. II*, Trans. Amer. Math. Soc. **76** (1954), 397–419.

[17] 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.

[18] C. Dellacherie and P.-A. 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.

[19] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev. **41** (1999), no. 1, 45–76.

[20] R. Douc, E. Moulines, P. Priouret, and P. Soulier, *Markov chains*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, 2018. MR 3889011

[21] R. M. Dudley, *Real analysis and probability*, Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, 2002, Revised reprint of the 1989 original.

[22] M. Duflo, *Random iterative models*, Applications of Mathematics (New York), vol. 34, Springer-Verlag, Berlin, 1997, Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

[23] J. Dugundji, *Topology*, Allyn and Bacon, Inc., Boston, Mass., 1966.

[24] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011.

[25] P. Erdös, *On a family of symmetric Bernoulli convolutions*, Amer. J. Math. **61** (1939), 974–976. MR 311

[26] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Series in Probability and Statistics, 1986.

[27] G. 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.

[28] A. Furman, *A brief introduction to ergodic theory*, Unpublished, http://homepages.math.uic.edu/ furman/papers.html, 2012.

[29] H. Furstenberg, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.

[30] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001, Reprint of the 1998 edition.

[31] M. Hairer, *Ergodic properties of markov processes*, Lecture given at the University of Warwick, 2006.

[32] ———, *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.

[33] ———, *Convergence of markov processes*, http://hairer.org/notes/Convergence.pdf, 2016.

[34] M. Hairer and J. C. Mattingly, *Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing*, Ann. of Math. (2) **164** (2006), no. 3, 993–1032.

[35] ———, *Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations*, The Annals of Probability **36** (2008), no. 6, 2050–2091.

[36] ———, *Yet another look at Harris' ergodic theorem for Markov chains*, Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab., vol. 63, Birkhäuser/Springer Basel AG, Basel, 2011, pp. 109–117.

[37] T. E. Harris, *The existence of stationary measures for certain Markov processes*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II, University of California Press, Berkeley and Los Angeles, 1956, pp. 113–124. MR 0084889

[38] Lars Hörmander, *Hypoelliptic second order differential equations*, Acta Math. **119** (1967), 147–171. MR 222474

[39] Kanji Ichihara and Hiroshi Kunita, *A classification of the second order degenerate elliptic operators and its probabilistic characterization*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **30** (1974), no. 3, 235–254.

[40] N Ikeda and S Watanabe, *Stochastic differential equations and diffusion processes*, North-Holland Mathematical Library, vol. 24, North-Holland Publishing Co., Amsterdam-New York; Kodansha, Ltd., Tokyo, 1981. MR 637061

[41] V. Jurdjevic, *Geometric control theory*, Cambridge Studies in Advanced Mathematics, vol. 52, Cambridge University Press, Cambridge, 1997. MR 1425878

[42] A. Katok and B. 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.

[43] Y. Kifer, *Ergodic theory of random transformations*, Progress in Probability and Statistics, vol. 10, Birkhäuser Boston, Inc., Boston, MA, 1986.

[44] H Kunita, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1990. MR 1070361

[45] Jean-François Le Gall et al., *Brownian motion, martingales, and stochastic calculus*, vol. 274, Springer, 2016.

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

[47] T. 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.

[48] R. Mañé, *Ergodic theory and differentiable dynamics*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 8, Springer-Verlag, Berlin, 1987, Translated from the Portuguese by Silvio Levy.

[49] S. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, second ed., Cambridge University Press, Cambridge, 2009, With a prologue by Peter W. Glynn.

[50] J. Moser, *On the volume elements on a manifold*, Trans. Amer. Math. Soc. **120** (1965), 286–294.

[51] D. Nualart, *The Malliavin calculus and related topics*, second ed., Probability and its Applications (New York), Springer-Verlag, Berlin, 2006. MR 2200233 (2006j:60004)

[52] E. Nummelin, *General irreducible Markov chains and nonnegative operators*, Cambridge Tracts in Mathematics, vol. 83, Cambridge University Press, Cambridge, 1984.

[53] Y. Peres, W. Schlag, and B. Solomyak, *Sixty years of Bernoulli convolutions*, Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), Progr. Probab., vol. 46, Birkhäuser, Basel, 2000, pp. 39–65. MR 1785620

[54] Y. Peres and B. Solomyak, *Absolute continuity of Bernoulli convolutions, a simple proof*, Math. Res. Lett. **3** (1996), no. 2, 231–239.

[55] J. W. Pitman, *Uniform rates of convergence for Markov chain transition probabilities*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **29** (1974), 193–227.

[56] G. Da Prato and J. Zabczyk, *Smoothing properties of transition semigroups in Hilbert spaces*, Stochastics and Stochastic Reports **35** (1991), no. 2, 63–77.

[57] J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995), vol. 9, 1996, pp. 223–252. MR 1611693

[58] A. N. Quas, *On representations of Markov chains by random smooth maps*, Bull. London Math. Soc. **23** (1991), no. 5, 487–492.

[59] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999. MR 1725357

[60] P. Robert, *Stochastic networks and queues*, french ed., Applications of Mathematics (New York), vol. 52, Springer-Verlag, Berlin, 2003, Stochastic Modelling and Applied Probability.

[61] W. Rudin, *Functional analysis*, second ed., International Series in Pure and Applied Mathematics, McGraw-Hill, Inc., New York, 1991. MR 1157815

[62] L. 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.

[63] 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.

[64] B. Solomyak, *On the random series* $\sum \pm\lambda^n$ *(an Erdös problem)*, Ann. of Math. (2) **142** (1995), no. 3, 611–625.

[65] D. W. Stroock and S. R. S. Varadhan, *On the support of diffusion processes with applications to the strong maximum principle*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory, 1972, pp. 333–359. MR 0400425

[66] H. J. Sussmann and V. Jurdjevic, *Controllability of nonlinear systems*, J. Differential Equations **12** (1972), 95–116. MR 0338882 (49 #3646)

[67] P. P. Varjú, *Recent progress on bernoulli convolutions*, 2016.

[68] C. Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new.

[69] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.

[70] W. Woess, *Random walks on infinite graphs and groups*, Cambridge Tracts in Mathematics, vol. 138, Cambridge University Press, Cambridge, 2000.