Probability
and
Stochastic Processes

♣

Francis Comets[1], *Benoît Laslier*

University Paris Diderot

April 20, 2017

[1]Thank you for reporting typos to `comets@math.univ-paris-diderot.fr`

Contents

Preface

We study random sequences with particular dependence structure, mainly martingales and Markov chains. For martingales, the evolution in time is specified via the (conditional) mean only, though for Markov chains, via the complete one-step dynamics.

The course covers the necessary background to a 2nd year of master with orientation in pure or applied probability and/or mathematical statistics, as well as to "préparation à l'agrégation, option probabilités."

Part I

Convergence in law of real random sequences

Chapter 1

Convergence in law and tightness

1.1 Preliminaries

1.1.1 Probability spaces and random variables

Definition 1.1.1. *Let Ω be a set.*

1. *A collection $\mathcal{A}$ of subsets of Ω is a σ-field if*

$$(i)\ \emptyset, \Omega \in \mathcal{A}, \qquad (ii)\ A \in \mathcal{A} \implies A^c \in \mathcal{A},$$
$$(iii)\ (A_i \in \mathcal{A} \quad {}^\forall i \geq 1) \implies \cup_{i\geq 1} A_i \in \mathcal{A}.$$

2. *A measure μ on $\mathcal{A}$ is a mapping $\mu : \mathcal{A} \to [0, +\infty]$ such that*

$$(i)\ \mu(\emptyset) = 0,$$

and (ii) for all sequence $A_i \in \mathcal{A}$ such that $i \neq j \implies A_i \cap A_j = \emptyset$,

$$\mu\big(\cup_{i\geq 1} A_i\big) = \sum_{i\geq 1} \mu(A_i) \qquad (\sigma - \text{additivity}).$$

3. *A probability measure $\mathbb{P}$ is a measure with unit total mass: $\mathbb{P}(\Omega) = 1$.*

Such a triplet $(\Omega, \mathcal{A}, \mathbb{P})$ is called a probability space. For a probability measure, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. Probability measures are those functions on subsets of Ω with values in [0,1] which are non-decreasing and continuous:

$$A \subset B,\ A, B \in \mathcal{A} \implies \mathbb{P}(A) \leq \mathbb{P}(B)$$
$$A_n \subset A_{n+1},\ A_n \in \mathcal{A} \implies \lim_{n \nearrow \infty} (\nearrow) \mathbb{P}(A_n) = \mathbb{P}(\cup_n A_n)$$
$$A_n \supset A_{n+1},\ A_n \in \mathcal{A} \implies \lim_{n \nearrow \infty} (\searrow) \mathbb{P}(A_n) = \mathbb{P}(\cap_n A_n),$$

where the right-hand side of the last two formulas are the limit of the increasing or decreasing sequence A_n. For any measure we have the *union bound* $\mu(\cup_i A_i) \le \sum_i \mu(A_i)$, even if the union is not pairwise disjoint.

The σ-field generated by a class $\mathcal{C}$ of subsets of Ω is the smallest σ-field on Ω containing $\mathcal{C}$. It is given by

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{T}} \mathcal{T}\ , \qquad \text{where } \mathcal{T} \text{ ranges over all } \sigma - \text{fields with } \mathcal{T} \supset \mathcal{C}.$$

Example 1.1.2. *Borel sets. Let Ω be a topological space, and denote by $\mathcal{O}$ the class of open sets. The Borel σ-field $\mathcal{B}(\Omega)$ is the σ-field generated by the open sets.*

$$\mathcal{B}(\Omega) = \sigma(\mathcal{O}).$$

Except for trivial cases, σ-fields are not explicit. Further, two measures which coincide on $\mathcal{C}$ are not necessarily equal on $\sigma(\mathcal{C})$. The following is useful:

Proposition 1.1.3 (Identification of measures)**.** *Assume the class $\mathcal{C}$ is stable by finite intersection, i.e.*[(1)] $A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C}$. *Then, two measures μ_1, μ_2 which are equal on $\mathcal{C}$ and such that $\mu_1(\Omega) = \mu_2(\Omega) < \infty$, are equal on $\sigma(\mathcal{C})$.*

For instance, the class $\mathcal{C}$ of rectangles $\prod_{i=1}^d (a_i, b_i)$ with $a_i < b_i$ generates the Borel field $\mathcal{B}(\mathbb{R}^d)$, and is stable by finite intersection.

Theorem 1.1.4 (Lebesgue measure on $\mathbb{R}^d$, 1905)**.** *There exists a unique measure λ on the Borel field $\mathcal{B}(\mathbb{R}^d)$ which coincides with the volume on rectangles:*

$$\lambda\left(\prod_{i=1}^d (a_i, b_i)\right) = \prod_{i=1}^d (b_i - a_i).$$

Proposition 1.1.3 applies on every compact subset of $\mathbb{R}^d$. Since $\mathbb{R}^d$ is a countable union of such subsets, uniqueness follows. Existence is much more involved, it will not be discussed here.

Definition 1.1.5. *Let E another space equipped with a σ-field $\mathcal{E}$. A mapping $X : \Omega \to F$ is a* **random variable** *(abbrev. r.v.) if, for all $B \in \mathcal{F}$,*

$$\{X \in B\} \overset{\text{def.}}{=} X^{-1}(B) \quad \text{belongs to } \mathcal{A}.$$

Then, the formula

$$P_X(B) = \mathbb{P}\{X \in B\},$$

defines a probability measure P_X on $(F, \mathcal{F})$, that we call the **law** *of the random variable X.*

[(1)] id est

If necessary, we will the specify the σ-fields and write that $X : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ is measurable. If both Ω and E are topological spaces with $\mathcal{A}, \mathcal{B}$ the corresponding Borel fields, it is easy to see that any continuous $X : \Omega \to F$ is a random variable.

1.1.2 Distribution function

A real random variable (r.r.v.) is a r.v. taking values in $\mathbb{R}$ equipped with its Borel field. By Proposition 1.1.3, its law P_X is determined by its values on the intervals $(-\infty, t]$.

Definition 1.1.6. *The distribution function of X is $F_X : \mathbb{R} \to [0,1]$ given by $F_X(t) = \mathbb{P}(X \leq t)$.*

For instance, taking $\Omega = [0,1)$ with $\mathcal{A} = \mathcal{B}([0,1))$ and $\mathbb{P} = \lambda$ restricted to the unit interval, the continuous mapping $X(\omega) = \omega$ is a r.r.v. with distribution function

$$F_X(x) = x, x \in [0,1], \qquad F_X(x) = 0, x \leq 0, \quad F_X(x) = 1, x \geq 1.$$

X is known as a *random number between 0 and 1*, its law is the uniform law on $[0,1]$, denoted by $\mathrm{Unif}_{[0,1]}$.

Proposition 1.1.7 (Botanic of distribution functions)**.** .

(i) F_X is non-decreasing, cadlag[(2)], with

$$F_X(-\infty) \stackrel{\text{def.}}{=} \lim_{x \to -\infty} F_X(x) = 0, \qquad F_X(+\infty) \ = \ 1.$$

(ii) Conversely, any function $F : \mathbb{R} \to [0,1]$, which is non-decreasing, cadlag with limits $F(-\infty) = 0, F(+\infty) = 1$ is the distribution function of a r.r.v. X.

In particular, for any real x we have $\mathbb{P}(X = x) = F(x) - F(x^-)$, with the usual notation $F(x^-) = \lim_{y \nearrow x} F(x)$ for the left limit, and $\mathbb{P}(a < X \leq b) = F(b) - F(a)$. Thus discontinuity points of F_X are the atoms of the law of X.

It is plain to show (i) using the definition of a probability measure. We prove part (ii), as a device to generate r.r.v. with a specified distribution function using a random generator.

Proposition 1.1.8 (Simulation by inversion)**.** *Let $F : \mathbb{R} \to [0,1]$ non-decreasing, cadlag with $F(-\infty) = 0, F(+\infty) = 1$. Then, the* left-continuous inverse $F^{\leftarrow}$ *of F,*

$$F^{\leftarrow}(\omega) = \inf\{z \in \mathbb{R} : F(z) \geq \omega\},$$

[(2)]Form the french "Continue à Droite et pourvue de Limite à Gauche" *is such that*

$$F^{\leftarrow}(\omega) \le x \iff F(x) \ge \omega. \tag{1.1}$$

In particular, $X(\omega) = F^{\leftarrow}(\omega)$ *defines a r.r.v. on* $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1), \mathcal{B}([0,1)), \lambda)$ *with distribution function* F.

$\square$ In (1.1), the direction $\Leftarrow$ is clear. Conversely, if $F^{\leftarrow}(\omega) \le x$, for all $\varepsilon > 0$ we have $F(x+\varepsilon) \ge \omega$ by monotonicity of F, which implies that $F(x) \ge \omega$ by the cadlag property. Now, from (1.1) we obtain

$$\mathbb{P}(X \le x) = \lambda([0, F(x)]) = F(x),$$

for all $x \in \mathbb{R}$. $\blacksquare$

Remark 1.1.9. *Let* $F^{\rightarrow}(\omega) \stackrel{\text{def}}{=} \inf\{z \in \mathbb{R} : F(z) > \omega\} \ge F^{\leftarrow}(\omega)$. *Then the 2 r.v.'s*

$$F^{\rightarrow} = F^{\leftarrow} \quad \text{a.s.}$$

Indeed, the strict inequality holds iff ω *is a discontinuity of* $F^{\rightarrow}$, *which, as a non-decreasing function, has at most countably many discontinuities and then Lebesgue measure 0. So we can write*

$$\mathbb{P}(F^{\rightarrow}(\omega) \ne F^{\leftarrow}(\omega)) = \lambda_{|[0,1]}(\mathcal{C}_0(F^{\rightarrow})^{\complement}) = 0.$$

Lebesgue's decomposition of a measure on $\mathbb{R}$: Any Borel measure μ on the real line can be decomposed in a unique manner as a sum of 3 Borel measures,

$$\mu = \mu_{\text{cont}} + \mu_{\text{sing}} + \mu_{\text{pp}} \ ,$$

where

- $\mu_{\text{pp}} = \sum_i m_i \delta_{a_i}$ a purely atomic measure,
- $\mu_{\text{cont}} = p\lambda$ is the absolutely continuous part,
- μ_{sing} is the singular continuous part.

The three components are mutually singular, i.e., for any two of them μ_1, μ_2, there exists a Borel set $A \subset \mathbb{R}$ such that

$$\mu_1(A) = 0, \qquad \mu_2(A^c) = 0.$$

A singular continuous measure is singular to Lebesgue measure and has no atoms. We will see an example of singular continuous measure with the example 1.1.10 of the uniform law on the Cantor set.

Example 1.1.10 (Triadic Cantor set)**.** *Starting from $C_0 = [0,1)$, construct a sequence C_{n+1} of subsets of $[0,1)$ by discarding the middle interval when sharing every connected component of C_n into 3 equal length intervals. The construction yields*

$$C_0 = [0,1), \quad C_1 = \Big[0, \frac{1}{3}\Big) \cup \Big[\frac{2}{3}, 1\Big), \quad C_2 = \Big[0, \frac{1}{9}\Big) \cup \Big[\frac{2}{9}, \frac{1}{3}\Big) \cup \Big[\frac{2}{3}, \frac{7}{9}\Big) \cup \Big[\frac{8}{9}, 1\Big).$$

More generally, C_n is a union of 2^n intervals of length 3^{-n}. Consider the uniform law on C_n and its distribution function F_n: F_n is constant on the complement of these intervals, but increases linearly (with slope $(3/2)^n$) on every interval, and F_n is continuous. We claim that

$$F_n \longrightarrow F \qquad \text{uniformly}.$$

Indeed, it is easy to see that, for all x,

$$|F_{n+1}(x) - F_n(x)| \leq 2^{-(n+1)}, \qquad |F_{n+p}(x) - F_n(x)| \leq 2^{-(n+1)},$$

so $\sup_p \|F_{n+p} - F_n\|_\infty \to 0$ as $n \to \infty$. Thus F_n is a Cauchy sequence of continuous functions, it converges uniformly to a continuous limit, say F. Clearly, F increases, with $F(0) = 0, F(1) = 1$, which implies, with the continuity and using Proposition 1.1.7 (ii), that there exists a Borel probability measure μ on $[0,1)$ with distribution function F: μ is called the uniform law on the Cantor set C_∞, *as we justify now.*[(3)]

The sequence C_n is a decreasing sequence of compact sets, so it has a limit $C_\infty = \bigcap_n C_n$ which is compact and non-empty, with $\lambda(C_\infty) = 0$. In fact,

$$C_\infty = \{x = \sum_{i=1}^{\infty} x_i 3^{-i}; x_i = 0, 2 \ \forall i \geq 1\},$$

since the first digit equal to 1 in the decomposition of x in base 3 is the nth digit if and only if $x \in C_{n-1}, x \notin C_n$. Consider now the sequence of digits $(X_i)_i$ of $X \in C_\infty$: we claim that, if X is distributed according to μ (i.e., uniform on the Cantor set), the digits X_i are independent random variables, identically distributed taking value 0 *or* 2 *with probability 1/2. Indeed, for all n and all $(x_i)_{i \leq n} \in \{0,2\}^n$,*

$$\{X_i = x_i, i = 1, \dots n\} = \{X \in [y, y + 3^{-n})\} \quad \text{with } y = \sum_{i=1}^{n} x_i 3^{-i},$$

so that

$$\mathbb{P}(X_i = x_i, i = 1, \dots n) = \mathbb{P}(X \in [y, y+3^{-n})) = F(y+3^{-n})) - F(y) = \left(\frac{1}{2}\right)^n,$$

showing that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 2) = \frac{1}{2}$, and also independence of the X_i's.

[(3)] F_∞ is known as "devil staircase" and has many remarkable analytical properties. It is uniformly continuous (in fact, Holder continuous) and has zero derivative almost everywhere (e.g., on the complement on C_∞). As Vitali (1905) pointed out, the function is not the integral of its derivative even though the derivative exists almost everywhere.

1.2 Characteristic function

$$\Phi_X(t) = \mathbb{E}e^{itX} = \mathbb{E}\cos(tX) + i\,\mathbb{E}\sin(tX), \quad t \in \mathbb{R}.$$

When P_X has a density p_X, the characteristic function is the Fourier transform[(4)] of the density, $\Phi_X = \widehat{p_X}$ in L^1.

Proposition 1.2.1. *Formal properties of the function $\Phi_X : \mathbb{R} \to \mathbb{C}$.*

1. $|\Phi_X(t)| \leq \Phi_X(0) = 1$.
2. Φ_X *is uniformly continuous.*
3. $\Phi_{-X}(t) = \Phi_X(-t) = \overline{\Phi_X}(t)$.
4. Φ_X *is a positive definite function:* $\forall n \geq 1, t_1, \dots t_n \in \mathbb{R}, z_1, \dots z_n \in \mathbb{C}^n$,
$$\sum_{i,j=1}^{n} \Phi_X(t_i - t_j) z_i \bar{z}_j \geq 0.$$
Indeed, this term is equal to $\mathbb{E}|\sum_j z_j e^{it_j X}|^2 \geq 0$.
Conversely, if $\phi : \mathbb{R} \to \mathbb{C}$ *is continuous, positive definite with* $\phi(0) = 1$, *then* $\phi = \Phi_X$ *for some r.r.v. (Bochner theorem).*
5. *For X and Y independent,* $\Phi_{X+Y} = \Phi_X \cdot \Phi_Y$.
6. Φ_X *characterises the law P_X of X.*

For the last point, recall *Fourier's inversion formula* in L^1: if $p_X \in L^1$ and $\Phi_X \in L^1$, then

$$p_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_X(t) dt \qquad \lambda - a.e. \tag{1.2}$$

It has been extended as follows.

Theorem 1.2.2. *Let X be a r.r.v. For all $a < b$, we have* Paul Lévy's inversion formula*:*

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \Phi_X(t) dt = \mathbb{P}(X \in (a,b)) + \frac{1}{2}(\mathbb{P}(X = a) + \mathbb{P}(X = b)). \tag{1.3}$$

Thus, when $\Phi_X \in L^1$, the r.v. has a density w.r.t.[(5)] Lebesgue measure, which is given by (1.2) and therefore continuous.

---

(4) For $f \in L^1(\mathbb{R}, dx)$, the Fourier transform is $\hat{f}(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$

(5) Standard abbreviation for "with respect to"

Paul Lévy's inversion formula implies the last point of Prop. 1.2.1. Letting $a \to -\infty$, the right-hand side of (1.3) tends to $F_X(b^-) + \frac{1}{2}\mathbb{P}(X = b)$. So, knowing Φ_X, we know F_X at all continuity points of F_X, and then everywhere by right-continuity.

□ First express the integral

$$\begin{aligned}
Q_T(a,b) &= \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\Phi_X(t)dt \\
&= \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\int_{-\infty}^{\infty} e^{itx}P_X(dx)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} P_X(dx)\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}dt \qquad (1.4)\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} P_X(dx) q_T(a,b;x),
\end{aligned}$$

by Fubini theorem, which is valid for

$$\int_{\mathbb{R}} P_X(dx)\int_{-T}^{T}\left|\frac{e^{it(x-a)} - e^{it(x-b)}}{it}\right|dt \le \int_{\mathbb{R}} P_X(dx)\int_{-T}^{T}|b-a|dt = 2T(b-a) < \infty,$$

using that $y \mapsto e^{iy}$ is 1-Lipshitz continuous. But

$$\begin{aligned}
q_T(a,b;x) &= \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}dt \\
&= \int_{-T}^{T}\left([\text{odd}-\text{function}](t) + \frac{\sin[t(x-a)] - \sin[t(x-b)]}{t}\right)dt \\
&= 2\int_{0}^{T} \frac{\sin[t(x-a)] - \sin[t(x-b)]}{t}dt.
\end{aligned}$$

Now set

$$S(y) = \int_0^y \frac{\sin(t)}{t}dt.$$

Then,

$$\int_0^T \frac{\sin(\alpha t)}{t}dt = \left\{\begin{array}{cc} S(\alpha T), & \alpha > 0, \\ 0, & \alpha = 0, \\ -S(|\alpha|T), & \alpha < 0, \end{array}\right\} = \operatorname{sgn}(\alpha)S(|\alpha|T),$$

yielding

$$q_T(a,b;x) = 2\big[\operatorname{sgn}(x-a)S(|x-a|T) - \operatorname{sgn}(x-b)S(|x-b|T)\big].$$

We have $\int_0^\infty \frac{|\sin(t)|}{t}dt = \infty$ but

$$\lim_{t\to\infty} S(t) = \frac{\pi}{2}, \qquad (1.5)$$

and the function S is bounded. Therefore, $q_T(a,b;x)$ itself is bounded, and the convergence

$$\lim_{T\to\infty} q_T(a,b;x) = \begin{cases} -2(\pi/2)+2(\pi/2)=0, & x<a \text{ or } x>b, \\ \pi, & x=a \text{ or } x=b, \\ 2\pi, & x\in(a,b), \end{cases}$$

is dominated. By Lebesgue's theorem, we get from (1.4):

$$\lim_{T\to\infty} Q_T(a,b) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P_X(dx) \lim_{T\to\infty} q_T(a,b;x),$$

which is equal to the right-hand side of (1.3). This ends the proof of (1.3).

We now prove the other statement. When Φ_X is integrable, we get by dominated convergence

$$\lim_{T\to\infty} Q_T = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-ita}-e^{-itb}}{it}\Phi_X(t)dt$$

We already know that the L.H.S. is equal to $F_X(b)-F_X(a)$ for all $a,b\in \mathcal{C}^0(F_X)$ the set of continuity points of F_X. In turn, this implies that

$$|F_X(b)-F_X(a)| \le \frac{1}{2\pi}\|\Phi_X\|_1(b-a),$$

i.e., F_X is (Lipschitz-) continuous, so $\mathcal{C}^0(F_X)=\mathbb{R}$. Then the equality $F_X(b)=F_X(a)+\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}\Phi_X(t)dt$ holds everywhere, and by differentiating under the integral we see that F_X is a $\mathcal{C}^1$ function. ∎

Example 1.2.3. *With $X_n \sim \mathcal{U}_{C_n}$ and $X\sim\mathcal{U}_{C_\infty}$ uniform on the Cantor set, we have $X_n\to X$ in law.*

1.3 Convergence in law

Let E be topological and $X_n(n\ge 1), X$ some r.v. taking values in $(E,\mathcal{B}(E))$.

Definition 1.3.1. *The sequence X_n converges in law to X if*

$$\forall h\in\mathcal{C}_b(E;\mathbb{R}),\quad \lim_{n\to\infty}\mathbb{E}h(X_n)=\mathbb{E}h(X).$$

Then we write $X_n \xrightarrow{\text{law}} X \quad \text{as } n\to\infty.$

It means that the sequence of probability measures P_{X_n} on $(E,\mathcal{B}(E))$ converges weakly to P_X. We now consider the case $E=\mathbb{R}$. The different convergence modes relate to each other by

$$\boxed{\left.\begin{array}{c} X_n \to X \text{ in } L^p \;\Rightarrow \\ \text{or} \\ X_n \to X \text{ a.s.} \;\Rightarrow \end{array}\right\} \;\Rightarrow\; X_n \to X \text{ in } \mathbb{P}} \quad \Rightarrow \quad X_n \to X \text{ in law}$$

$$\boxed{\text{strong convergence}} \quad \Rightarrow \quad \text{weak convergence}$$

Strong convergence is a convergence of r.v.'s, though weak convergence is only convergence of the law of the r.v.'s. This implies that we can do operations with strong convergence, but not with weak convergence (without additional properties).

Definition 1.3.2. *The sequence X_n of real r.v.'s converges in law to X if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \forall x \in \mathcal{C}^0(F_X).$$

Before proving that the two definitions are equivalent, we give an example showing why it is necessary to discard the discontinuities of the limit. Consider (deterministic) r.v. $X_n = 1/n$ and $X = 0$. Of course $X_n \to X$ in law - as well as in any reasonable sense -, the respective distribution functions are

$$F_{X_n} = \mathbf{1}_{[1/n,+\infty)}, \quad F_X = \mathbf{1}_{[0,+\infty)},$$

and we see that F_{X_n} converge to F_X everywhere except at 0.

Proposition 1.3.3. *For $E = \mathbb{R}$, Definitions 1.3.1 and 1.3.2 agree.*

□ Step 1: Proof of $\Longrightarrow$.

Fix $x_0 \in \mathbb{R}$, and mollify the function $h = \mathbf{1}_{(-\infty,x_0]}$ into a piecewise linear approximation $h_\varepsilon(x) = 1 \wedge \big(1 - \varepsilon^{-1}(x - x_0)\big) \vee 0$. Using

$$(a)\ h_\varepsilon \in \mathcal{C}_b, \quad (b)\ h \le h_\varepsilon, \quad (c)\ h_\varepsilon(x) \le h(x - \varepsilon), \tag{1.6}$$

we obtain

$$F_{X_n}(x_0) \overset{(1.6.b)}{\le} \mathbb{E}h_\varepsilon(X_n) \underset{n\to+\infty}{\overset{(1.6.a)}{\longrightarrow}} \mathbb{E}h_\varepsilon(X) \overset{(1.6.c)}{\le} F_X(x_0 - \varepsilon)\ .$$

Letting $\varepsilon \searrow 0$ we conclude that $\limsup_{n\to\infty} F_{X_n}(x_0) \le F_X(x_0^-)$ for all x_0.

Consider now $h^\varepsilon(x) = h_\varepsilon(x + \varepsilon)$, which satisfies

$$(a)\ h^\varepsilon \in \mathcal{C}_b, \quad (b)\ h \ge h^\varepsilon, \quad (c)\ h^\varepsilon(x) \ge h(x + \varepsilon), \tag{1.7}$$

so that we get this time

$$F_{X_n}(x_0) \overset{(1.7.b)}{\ge} \mathbb{E}h^\varepsilon(X_n) \underset{n\to+\infty}{\overset{(1.7.a)}{\longrightarrow}} \mathbb{E}h^\varepsilon(X) \overset{(1.7.c)}{\ge} F_X(x_0 + \varepsilon)\ .$$

Finally,

$$\liminf_n F_{X_n}(x_0) \geq F_X(x_0 + \varepsilon), \quad \text{which } \searrow F_X(x_0) \quad \text{as } \varepsilon \searrow 0.$$

When $x_0 \in \mathcal{C}^0(F_X)$, we have $F_X(x_0^-) = F_X(x_0)$, and finally $\lim_n F_{X_n}(x_0) = F_X(x_0)$, completing the first step.

Step 2: : Proof of $\Longleftarrow$. We could give a pedestrian proof, but instead we prove a stronger result.

Theorem 1.3.4 (Skorokhod's representation theorem)**.** *Let F_n, F be distribution functions such that*

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in \mathcal{C}^0(F).$$

Then, there exist r.r.v. Y_n, Y defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $F_{Y_n} = F_n, F_Y = F$, and

$$Y_n \xrightarrow{\text{p.s.}} Y \ , \quad n \to \infty.$$

As it is well known that a.s. convergence implies convergence in law, the theorem completes the proof of Proposition 1.3.3. ∎

□ Proof of Theorem 1.3.4. Let $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1), \mathcal{B}([0,1)), \lambda)$,

$$Y^{\leftarrow}(\omega) = \inf\{x : F(x) \geq \omega\}, \quad Y^{\rightarrow}(\omega) = \inf\{x : F(x) > \omega\},$$

and the similar definition for $Y_n^{\leftarrow}, Y_n^{\rightarrow}$. We know that $Y^{\leftarrow} \leq Y^{\rightarrow}$ and also, from Remark 1.1.9, that the equality holds a.s. By Proposition 1.1.8, $Y_n \overset{\text{def}}{=} Y_n^{\leftarrow}$ and $Y \overset{\text{def}}{=} Y^{\leftarrow}$ have distribution functions F_n, F.

Fix $\omega \in [0,1]$, and $x > Y^{\rightarrow}(\omega)$ with $x \in \mathcal{C}^0(F)$. We then have $F(x) > \omega$ and $\lim_n F_n(x) = F(x)$, so $F_n(x) > \omega$ for large enough n and then $\limsup_n Y_n^{\rightarrow}(\omega) \leq x$. Since $\mathcal{C}^0(F)$ is at most countable, we can take a sequence $x_j \searrow Y^{\rightarrow}(\omega), x_j \in \mathcal{C}^0(F)$: by the previous argument for $x = x_j$,

$$\limsup_n Y_n^{\rightarrow}(\omega) \leq Y^{\rightarrow}(\omega). \tag{1.8}$$

Fix $x < Y^{\leftarrow}(\omega)$ with $x \in \mathcal{C}^0(F)$. We then have $F(x) < \omega$ and $\lim_n F_n(x) = F(x)$, so $F_n(x) < \omega$ for large enough n and then $\liminf_n Y_n^{\leftarrow}(\omega) \geq x$. By density as above,

$$\liminf_n Y_n^{\leftarrow}(\omega) \geq Y^{\leftarrow}(\omega). \tag{1.9}$$

Thus we have (1.8, 1.9) and $Y_n^{\leftarrow} \leq Y_n^{\rightarrow}$ for all ω, and also $Y^{\leftarrow} = Y^{\rightarrow}$ a.s. Hence all terms are equal on this set of full measure:

$$\lim_n Y_n^{\leftarrow} = \lim_n Y_n^{\rightarrow} = Y^{\leftarrow} = Y^{\rightarrow}$$

almost surely. ∎

1.4 Tightness

We address the sequential compactness of a sequence of probability measures on the Borel sets of the real line.

Lemma 1.4.1. *(Helly-Bray) Let $F_n, n \geq 1$, be a sequence of distribution functions on $\mathbb{R}$. Then, there exists a sequence $n_k \nearrow \infty$ as $k \nearrow \infty$ (an extractor) and a function $F : \mathbb{R} \to [0,1]$ such that*

$$\lim_{k\to\infty} F_{n_k}(x) = F(x), \qquad \forall x \in \mathcal{C}^0(F). \tag{1.10}$$

□ Enumerate the rational numbers, $\mathbb{Q} = \{q_1, q_2, \ldots\}$. The rough idea is to use compactness of [0,1] to extract a subsequence at each rational and then apply a diagonal procedure. Since $[0,1]$ is compact, we can

- Extract from $(F_n(q_1); n \geq 1)$ a converging sequence

$$F_{n(1,i)}(q_1) \overset{i\to\infty}{\longrightarrow} H(q_1).$$

- Since $(F_{n(1,i)}(q_2); i \geq 1)$ we can extract a converging sequence

$$F_{n(2,i)}(q_2) \overset{i\to\infty}{\longrightarrow} H(q_2).$$

- And so on for all q_k, we find and extractor $n(k,i)$ from $(n(k-1,m); m \geq 1)$ such that $F_{n(k,i)}(q_k)$ converges to some limit that we denote by $H(q_k)$.

Setting $n_i = n(i,i)$ for $i \geq 1$, we get

$$\lim_{i\to\infty} F_{n_i}(q) = H(q), \qquad \forall q \in \mathbb{Q}.$$

As a pointwise limit of non-decreasing functions, H is non-decreasing. We extend it into a cadlag function $F : \mathbb{R} \to [0,1]$ by

$$F(x) = \inf\{H(q); q \in \mathbb{Q}, q > x\}, \qquad x \in \mathbb{R}, \tag{1.11}$$

i.e., $F(x) = \lim_{q\to x, q>x, q\in\mathbb{Q}} H(q)$[(6)]. Indeed, F defined as above is non-decreasing, hence ladlag; it is also right-continuous at $x \in \mathbb{R}$ since the sets

$$\{q \in \mathbb{Q} : q > y\} \nearrow \{q \in \mathbb{Q} : q > x\} \qquad \text{as } y \searrow x.$$

Now we check (1.10). Let $x \in \mathcal{C}(F)$ and $q < x < q'$. First, observe by monotonicity that $F_{n_i}(q) \leq F_{n_i}(x) \leq F_{n_i}(q')$. Thus

$$H(q) \leq \liminf_i F_{n_i}(x) \leq \limsup_i F_{n_i}(x) \leq H(q').$$

[(6)]Observe that the value of $F(q)$ at a rational q may be strictly larger than $H(q)$

So for any $q_0 < x < q_0'$,

$$F(q_0) \le \liminf_i F_{n_i}(x) \le \limsup_i F_{n_i}(x) \le F(q_0').$$

For $x \in \mathcal{C}(F)$, both extreme terms converge to $F(x)$ as $q_0 \nearrow x, q_0' \searrow x$, yielding (1.10). ∎

Observe that, necessarily F is non-decreasing and cadlag with $F(+\infty) - F(-\infty) \le 1$. Then it is associated to a "sub-probability". However the strict inequality may occur.

Example 1.4.2. *$\mu_n = (1/2)(\delta_0 + \delta_n)$ does not converge weakly, because half of the mass goes to $+\infty$. Here, $F = (1/2)\mathbf{1}_{[0,+\infty)}$.*

The real line is not compact, and and to make sure the "sub-probability" is a probability, it suffices to localise the mass on compact subsets:

Definition 1.4.3. *A sequence $(\mu_n; n \ge 1)$ of probability measures on $E = \mathbb{R}$ is tight if for all $\varepsilon > 0$ there exists a compact K_ε such that*

$$\inf_n \mu_n(K_\varepsilon) \ge 1 - \varepsilon. \tag{1.12}$$

On the real line we will take $K_\varepsilon = [-A_\varepsilon, A_\varepsilon]$ without loss of generality (abbreviated WLOG).

Theorem 1.4.4. *(Prokhorov) If $(\mu_n; n \ge 1)$ is a tight sequence of probability measures on $\mathbb{R}$, we can extract a sub-sequence $(n_i)_i$ and a probability measure μ such that $\mu_{n_i} \to \mu$ weakly as $i \to \infty$.*
Equivalently, from any sequence ot r.r.v.'s X_n such that

$$\lim_{A \to +\infty} \sup_n \mathbb{P}(|X_n| \ge A) = 0,$$

we can extract a subsequence converging in law to a finite r.v.

□ Let $F_n(x) = \mu_n(]-\infty, x])$. By the Helly-Bray lemma, there exists n_i and F such that $F_{n_i} \to F$ on $\mathcal{C}^0(F)$ as $i \to \infty$. On the other hand, by tightness, for all $\varepsilon > 0$ there is some A_ε with

$$F_n(A_\varepsilon) - F_n(-A_\varepsilon) \ge 1 - \varepsilon, \qquad \forall n.$$

Increasing the value of A_ε if necessary, we can assume WLOG that both $A_\varepsilon, -A_\varepsilon$ are continuity points of F (recall that $\mathcal{C}^0(F)^c$ is at most countable). Taking $n = n_i$ and letting $i \to \infty$, we then get

$$F(A_\varepsilon) - F(-A_\varepsilon) \ge 1 - \varepsilon.$$

Holding for all $\varepsilon > 0$, this implies that $\lim_{A \to +\infty} F(A) - \lim_{a \to -\infty} F(a) = 1$, i.e. F is a distribution function. By Proposition 1.3.3, we have weak convergence $\mu_{n_i} \to \mu$ to the probability measure associated to F. ∎

Remark 1.4.5. *(Not In the Program) Caracterization of sequentially relatively compact sets of probability measures on Borel fields can be extended general topological spaces E.*

1. *A polish space E is a metric space which is separable (there is a countable dense subset) and complete (Cauchy sequences converge).*
2. *On a polish space, Prokhorov theorem holds: a sequence of probability measures is relatively compact if and only if (abbreviated iff) it is tight.*

1.5 Paul Lévy's continuity theorem

Recall that for r.r.v.'s X_n, X, pointwise convergence of characteristic functions is equivalent to converge in law:

$$X_n \xrightarrow{\text{law}} X \iff \forall t \in \mathbb{R}, \Phi_{X_n}(t) \to \Phi_X(t) \tag{1.13}$$

We start to note that for characteristic functions, as $n \to \infty$,

$$\Phi_{X_n} \to \Phi_X \text{ pointwise} \Leftrightarrow \Phi_{X_n} \to \Phi_X \text{ uniformly on compacts.}$$

□ Indeed, we now prove the surprising implication $\Rightarrow$: if $\Phi_{X_n} \to \Phi_X$ pointwise, we have $X_n \to X$, and by Skorohod representation theorem 1.3.4, $Y_n \to Y$ a.s. with new r.v.'s Y_n, Y having the same laws as X_n, X respectively. We now end by proving that: $X_n \xrightarrow{\mathbb{P}} X$ implies $\Phi_{X_n} \to \Phi_X$ uniformly on compacts. Using the standard notation

$$\mathbb{E}(Z; A) = \mathbb{E}(Z\mathbf{1}_A),$$

for any r.v. Z and any event A, we have

$$\begin{aligned}
\Phi_{X_n}(t) - \Phi_X(t) &= \Phi_{Y_n}(t) - \Phi_Y(t) \qquad \text{(same law)} \\
&= \mathbb{E}\big[e^{itY_n} - e^{itY}\big] \qquad \text{(same space)} \\
&= \mathbb{E}\big[e^{itY_n} - e^{itY}; |Y_n - Y| \le \varepsilon\big] + \mathbb{E}\big[e^{itY_n} - e^{itY}; |Y_n - Y| > \varepsilon\big] \\
|\Phi_{X_n}(t) - \Phi_X(t)| &\le \mathbb{E}\big[|e^{itY_n} - e^{itY}|; |Y_n - Y| \le \varepsilon\big] + \mathbb{E}\big[|e^{itY_n} - e^{itY}|; |Y_n - Y| > \varepsilon\big] \\
&\le \mathbb{E}\big[|t||Y_n - Y|; |Y_n - Y| \le \varepsilon\big] + \mathbb{E}\big[2; |Y_n - Y| > \varepsilon\big],
\end{aligned}$$

since $a \mapsto e^{ia}$ is Lipschitz continuous. Finally,

$$\sup_{|t| \le T} |\Phi_{X_n}(t) - \Phi_X(t)| \le T\varepsilon + 2\mathbb{P}(|Y_n - Y| > \varepsilon),$$

which tends to 0 as we let $n \to \infty$ and then $\varepsilon \to 0$. Therefore,

$$\lim_{n\to\infty} \sup_{|t| \le T} |\Phi_{X_n}(t) - \Phi_X(t)| = 0.$$

for all T. ∎

This indicates that there is a lot of structure in the requirement that Φ is a characteristic function. We now see that we don't need the requirement for the limit, but only its continuity at 0.

Theorem 1.5.1 (Lévy's continuity theorem)**.** *If a sequence of characteristic function Φ_{X_n} converges (pointwise on $\mathbb{R}$) to some limit ϕ, and if the limit is continuous at 0, then ϕ itself is a characteristic function and X_n converges in law.*

□ The proof argues with the laws, following the standard scheme: compactness and identification of the limit.

Step 1: Compactness. Fix $\varepsilon > 0$ and start by noting that

$$\Phi_{X_n}(t) + \Phi_{X_n}(-t) = 2\mathbb{E}\cos(tX_n)$$

is not larger than 2; the same holds with the limit ϕ. Note that $\phi(0) = 1$. Since ϕ is continuous at 0, we fix δ such that $|\phi(t) - 1| \leq \varepsilon$ for all $t \in [-\delta, \delta]$. Then

$$0 < \delta^{-1} \int_0^\delta (2 - \phi(t) - \phi(-t))dt < 2\varepsilon.$$

By dominated convergence, the integral is equal to the limit of the integral with Φ_{X_n} instead of ϕ, so, for n larger than some n_0,

$$\delta^{-1} \int_0^\delta (2 - \Phi_{X_n}(t) - \Phi_{X_n}(-t))dt < 2\varepsilon.$$

But the LHS writes

$$\begin{aligned}
\delta^{-1} \int_{-\delta}^{\delta} (1 - \mathbb{E}e^{itX_n})dt &= \delta^{-1} \int_{-\delta}^{\delta} \mathbb{E}(1 - e^{itX_n})dt \\
&\overset{\text{Fubini}}{=} \mathbb{E} \int_{-\delta}^{\delta} \frac{1 - e^{itX_n}}{\delta} dt \\
&= 2\mathbb{E}\left(1 - \frac{\sin(\delta X_n)}{\delta X_n}\right) \\
&\overset{\left|\frac{\sin u}{u}\right| \leq 1}{\geq} 2\mathbb{E}\left(1 - \frac{\sin(\delta X_n)}{\delta X_n}; |X_n| > 2/\delta\right) \\
&\overset{|\sin u| \leq 1}{\geq} 2\mathbb{E}\left(1 - \frac{1}{\delta |X_n|}; |X_n| > 2/\delta\right) \\
&\geq 2 \times \frac{1}{2} \times \mathbb{P}(|X_n| > 2/\delta).
\end{aligned}$$

Finally,

$$\mathbb{P}(|X_n| > 2/\delta) \leq 2\varepsilon,$$

which shows that $(X_n)_n$ is tight[(7)].

Step 2: Identification of the limit. First, ϕ is a characteristic function: indeed, by Prokhorov Proposition 1.4.4, we can extract from X_n a subsequence which converges in law to some X, and by (1.13), $\phi = \Phi_X$. Since Φ_{X_n} converges to $\phi = \Phi_X$, we have $X_n \to X$ in law by (1.13) again. ∎

1.6 Central limit theorem

This a version for independent but not identically distributed summands.

Theorem 1.6.1. *(Central limit theorem for sum of independent r.v.'s) Let $(X_k; k \geq 1)$ be independent r.r.v.'s in L^3, with $\mathbb{E}X_k = 0, \mathrm{Var}(X_k) = \sigma_k^2 > 0$. Define $s_n > 0$ by*

$$s_n^2 = \sum_{k=1}^n \sigma_k^2,$$

and assume

$$\lim_{n\to\infty} s_n^{-3} \sum_{k=1}^n \mathbb{E}(|X_k|^3) = 0. \tag{1.14}$$

Then,

$$\frac{S_n}{s_n} \xrightarrow{\text{law}} Z \sim \mathcal{N}(0,1), \qquad n \to \infty.$$

We start with some remarks: assumption (1.14) implies the following.

$$s_n \to \infty, \qquad \sup_{k\leq n} \frac{\sigma_k}{s_n} \to 0 \qquad \text{as } n \to \infty. \tag{1.15}$$

The first statement is clear, and the second follows from Jensen's inequality:

$$\left(\frac{\sigma_k}{s_n}\right)^3 \leq \frac{\mathbb{E}(|X_k|^3)}{s_n^3} \leq \frac{\sum_{k\leq n}\mathbb{E}(|X_k|^3)}{s_n^3} \to 0,$$

by (1.14). Then, the assumption means that all the summand are negligible in $\sum \sigma_k^2$. Roughly speaking, none of the terms is of the order of the full sum. Note that, when the X_k have different law, it is necessary to add an assumption making sure that "they are roughly of the same size"; otherwise, a single term could be significant in the total sum, and contribute with an arbitrary fluctuation. The message is the so-called *Gaussian error law*, which is well known all through science and engineering and relevant for a large class of phenomena: *many small independent errors add up to an almost Gaussian random result.*

[(7)] Controling finitely many terms ($n \leq n_0$) is not difficult

□ Proof of Theorem 1.6.1: By independence,

$$\Phi_n(t) := \mathbb{E}\exp\{itS_n/s_n\} = \prod_{k=1}^{n} \Phi_{X_k}(t/s_n).$$

By Taylor's expansion[8], since $\sigma_n \to \infty$, $\mathbb{E}X_k = 0$ and $\mathbb{E}X_k^2 = \sigma_k^2$, we get

$$\big|\Phi_{X_k}(t/s_n) - 1 + t^2\sigma_k^2/(2s_n^2)\big| \le |t|^3\mathbb{E}(|X_k|^3)/(6s_n^3). \tag{1.16}$$

But also

$$\big|\Phi_{X_k}(t/s_n) - 1\big| \le |t|^2\mathbb{E}(|X_k|^2)/(2s_n^2) \longrightarrow 0, \text{ uniformly in } k. \tag{1.17}$$

Therefore, with log the principal determination of the complex logarithm, we can write for all fixed t and n large enough,

$$\begin{aligned}
\Big|\sum_{k=1}^{n} \log \Phi_{X_k}(\frac{t}{s_n}) - [\Phi_{X_k}(\frac{t}{s_n}) - 1]\Big| &\le \sum_{k=1}^{n} \big|\Phi_{X_k}(\frac{t}{s_n}) - 1\big|^2 \\
&\le \frac{t^4}{s_n^4}\sum_{k\le n}\sigma_k^4 \qquad \text{(by (1.17))} \\
&\le t^4 \sup_{k\le n}\frac{\sigma_k^2}{s_n^2} \\
&\longrightarrow 0,
\end{aligned}$$

by (1.15). Finally, by (1.16),

$$\Big|\sum_{k=1}^{n} \log \Phi_{X_k}(\frac{t}{s_n}) - \frac{1}{2}\sum_{k\le n}\frac{t^2\sigma_k^2}{s_n^2}\Big| \le \frac{|t|^3\mathbb{E}(|X_k|^3)}{6s_n^3} + t^4\sup_{k\le n}\frac{\sigma_k^2}{s_n^2},$$

which tends to 0 as $n \to \infty$. The proof is complete. ∎

[8] Taylor formula with integral remainder:

$$\phi(x) = \sum_{k=0}^{n}\phi^{(k)}(0)\frac{x^k}{k!} + \int_0^1 \frac{(1-t)^n x^{n+1}}{n!}\phi^{(k+1)}(tx)dt$$

Part II

Discrete time martingales

Chapter 2

Conditional expectation

First recall a related notion.

2.1 A reminder on Independence

Definitions: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

- Two sub-sigma-fields $\mathcal{B}, \mathcal{C}$ of $\mathcal{A}$ are independent (for $\mathbb{P}$) iff

$$\mathbb{P}(B \bigcap C) = \mathbb{P}(B)\mathbb{P}(C), \qquad B \in \mathcal{B}, C \in \mathcal{C}.$$

- A family of sub-sigma-fields $\mathcal{B}_i, i \in I$ (with an arbitrary index set I) is independent iff for all choice of finite $J \subset I$ and $B_j \in \mathcal{B}_j, j \in J$,

$$\mathbb{P}(\bigcap_{i \in J} B_j) = \prod_{j \in J} \mathbb{P}(B_j).$$

- A family of r.v.'s $X_i, i \in I$ defined on Ω is independent iff the generated sigma-fields $\sigma(X_i)$ are independent. In this case, if we denote by $(E_i, \mathcal{E}_i)$ the space where X_i takes its values, the joint law of the family on the product space $(\prod_{i \in I} E_i, \otimes_{i \in I} \mathcal{E}_i)$ is the product law,

$$P_{(X_i)_{i \in I}} = \otimes_{i \in I} P_{X_i}$$

Intuitively, r.v.'s are independent if knowing some of them does not bring any information on the other ones.

To check independence, it is convenient to use the following consequence of monotone class theorem.

- Consider a family of sub-sigma-fields $\mathcal{B}_i, i \in I$, and for all i, a generating class $\mathcal{C}_i \subset \mathcal{B}_i, \sigma(\mathcal{C}_i) = \mathcal{B}_i$ which is stable by finite intersection. Assume that

$$\forall J \subset I \text{ finite}, C_j \in \mathcal{C}_j (j \in J), \qquad \mathbb{P}(\bigcap_{j \in J} C_j) = \prod_{j \in J} \mathbb{P}(C_j).$$

  Then, the sigma-fields $\mathcal{B}_i, i \in I$, are independent.

Stability. The fundamental property is that independence is preserved by disjoint grouping:

- Assume the sigma-fields $\mathcal{B}_i, i \in I$, are independent, and J is an arbitrary partition of I (J is a class of subsets of I, which are pairwise disjoint and which union is equal to I). For $j \in J$ define $\mathcal{C}_j = \sigma(\mathcal{B}_i; i \in j)$. Then, the sigma-fields $\mathcal{C}_j, j \in J$, are independent.

Let us be more specific. If a sequence $X_n, n \geq 1$, is independent, then it is also the case for:

- a subsequence $X_{n(k)}, k \geq 1$;
- a sequence obtained by non-overlaping groupings (i.e., $Y_k, k \geq 1$, defined from a increasing sequence $n_0 = 0 < n_1 < n_2 < n_3 < \ldots$ and $Y_k = (X_{n_{k-1}+1}, \ldots, X_{n_k})$);
- (measurable) functions of the terms (i.e., $Z_i, i \geq 1$, with $Z_i = f_i(X_i)$).

Combining these properties, we have many examples, e.g., for independent RRV's X_i,

$$X_1 + X_9, \quad X_3^4 + X_4 \cos(X_5), \quad \min\{(X_i)^+; i \geq 34\}$$

are independent.

2.2 Definition of Conditional Expectation

On some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ let X be an integrable r.r.v. and $\mathcal{B} \subset \mathcal{A}$ a σ-field. We view $\mathcal{B}$ as the available information, e.g. similar to the observable in physics when we don't have access to the full state of nature. Recall that the expectation $\mathbb{E}X$ represents the mean value of the r.v. X: Implicitely it means that we do not have any relevant information on the value X. An opposite situtation is when we observe X: then, the mean value of X given this information is X itself.

The correct notion of mean value of X observing the information in $\mathcal{B}$ is a $\mathcal{B}$-measurable r.v.

Definition 2.2.1. *For $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{B} \subset \mathcal{A}$ a σ-field, the* conditional expectation of X given *$\mathcal{B}$ (denoted by $E(X|\mathcal{B})$) is the unique random variable $Z \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ such that*

$$E(Z\mathbf{1}_B) = E(X\mathbf{1}_B) \qquad \forall B \in \mathcal{B}.$$

□ We prove existence and uniqueness, starting with the latter. If Z, Z' are two candidates, we have by linearity

$$E((Z - Z')\mathbf{1}_B) = E(X\mathbf{1}_B) - E(X\mathbf{1}_B) = 0, \quad B \in \mathcal{B}.$$

Taking $B = \{Z > Z'\} \in \mathcal{B}$, we get that $\mathbb{P}(Z > Z') = 0$. Hence, $Z = Z'$ a.s. So $E(X|\mathcal{B})$ is uniquely defined.

• Existence for $X \in L^2$: Recall that $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space for the scalar product $(X, Y) \mapsto \mathbb{E}(XY)$ (Riesz's theorem). Note that $L^2(\Omega, \mathcal{B}, \mathbb{P})$ can be imbedded in $L^2(\Omega, \mathcal{A}, \mathbb{P})$ as a linear space, as closed subset. Indeed, any sequence in $L^2(\Omega, \mathcal{B}, \mathbb{P})$ converging in $L^2(\Omega, \mathcal{A}, \mathbb{P})$, is a Cauchy sequence in the smaller space, so it converges there by Riesz's theorem. Then, the Hilbert projection theorem states that

$$\forall X \in L^2(\Omega, \mathcal{A}, \mathbb{P}), \exists! Z \in L^2(\Omega, \mathcal{B}, \mathbb{P}) : \|X - Z\| = \min_{U \in L^2(\mathcal{B})} \|X - U\|.$$

Then mapping $\pi^{\mathcal{B}} : L^2(\Omega, \mathcal{A}, \mathbb{P}) \to L^2(\Omega, \mathcal{B}, \mathbb{P})$, defined by $X \mapsto Z$, is linear and 1-Lipschitz continuous. Now, the crucial observation is that

$$X \in L^2(\Omega, \mathcal{A}, \mathbb{P}) \Longrightarrow \mathbb{E}(X|\mathcal{B}) = \pi^{\mathcal{B}} X \tag{2.1}$$

Indeed, $Z = \pi^{\mathcal{B}} X$ is characterized by

$$Z \in L^2(\mathcal{B}), \qquad X - Z \perp L^2(\mathcal{B}).$$

For all $U \in L^2(\mathcal{B})$, we then have $\mathbb{E}(XU) = \mathbb{E}(ZU)$. Taking $U = \mathbf{1}_B, B \in \mathcal{B}$, we recover the requirements of Definition 2.2.1: $Z \in L^2(\mathcal{B}), E(Z\mathbf{1}_B) = E(X\mathbf{1}_B)$.

• Extension to $X \in L^1$: For such X, let us decompose $X = X^+ - X^-$, where both $X^+ = \max\{X, 0\}$ and $X^- = \max\{-X, 0\}$ are integrable. By linearity, we see that it suffices to show the existence of $\mathbb{E}(X|\mathcal{B})$ for non-negative $X \in L^1$. Let $X_n = X \wedge n$, and $Y_n = \mathbb{E}(X_n|\mathcal{B})$ as defined for $X_n \in L^2$ as above. It is enough to show that

$$Y_n \geq 0, \qquad Y_n \leq Y_{n+1} \quad \text{a.s.}, \tag{2.2}$$

for then we have

$$Y = \lim_{n \nearrow \infty} (\nearrow) Y_n \in [0, \infty], \qquad Y \in \mathcal{B},$$

and by monotone convergence we have

$$\mathbb{E}(Y\mathbf{1}_B) = \lim_n \mathbb{E}(Y_n\mathbf{1}_B) = \lim_n \mathbb{E}(X_n\mathbf{1}_B) = \mathbb{E}(X\mathbf{1}_B)$$

for all $B \in \mathcal{B}$, proving that $Y = \mathbb{E}(X\mathcal{B})$. It remains to prove (2.2), and this is the object of the next proposition. ∎

Proposition 2.2.2.

$$X \in L^2, X \geq 0 \Longrightarrow \mathbb{E}(X|\mathcal{B}) \geq 0.$$

□ With $\mathbb{E}(X|\mathcal{B}) = \pi^{\mathcal{B}}(X)$ defined from linear considerations, it is not obvious how positivity comes. However, since $X \geq 0$, for any Z we have

$$|X - Z| \leq |X - Z^+|.$$

Integrating, we see that

$$\|X - Z\|_2^2 = \mathbb{E}[(X - Z)^2] \geq \mathbb{E}[(X - Z^+)^2] = \|X - Z^+\|_2^2,$$

i.e., $\|X - Z\|_2 \geq \|X - Z^+\|_2$. But Z^+ is $\mathcal{B}$-measurable if Z is $\mathcal{B}$-measurable, so by optimality of $Z = \pi^{\mathcal{B}}X$, we get

$$\pi^{\mathcal{B}}X = (\pi^{\mathcal{B}}X)^+ \quad a.s.,$$

so that $\mathbb{E}(X|\mathcal{B}) = \pi^{\mathcal{B}}X \geq 0$ a.s. ∎

Remark 2.2.3. *(i) Conditional expectation is easier in L^2.*

(ii) For X square-integrable, $\mathbb{E}(X|\mathcal{B})$ is the element of $L^2(\mathcal{B})$ closest to X in the sense of mean squares (best predictor).

(iii) This is similar to other non-conditional expectation: $\mathbb{E}(X)$ is the constant closest to X in this sense.

2.3 Properties

Proposition 2.3.1. *1. Tower property: For $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$ σ-fields,*

$$\mathbb{E}\Big(\mathbb{E}(X|\mathcal{B})|\mathcal{C}\Big) = \mathbb{E}\Big(X|\mathcal{C}\Big)$$

2. If X is $\mathcal{B}$-measurable, then $\mathbb{E}(X|\mathcal{B}) = X$.

3. Linearity: $\mathbb{E}(a_1X_1 + a_2X_2|\mathcal{B}) = a_1\mathbb{E}(X_1|\mathcal{B}) + a_2\mathbb{E}(X_2|\mathcal{B})$.

4. Positivity: $X \geq 0 \implies \mathbb{E}(X|\mathcal{B}) \geq 0$.

5. *Monotone convergence:* $0 \leq X_n \nearrow X \implies 0 \leq \mathbb{E}(X_n|\mathcal{B}) \nearrow \mathbb{E}(X|\mathcal{B})$.

6. *Fatou:* $X_n \geq 0 \implies \mathbb{E}(\liminf_n X_n|\mathcal{B}) \leq \liminf_n \mathbb{E}(X_n|\mathcal{B})$.

7. *Dominated convergence:*

$$(|X_n| \leq Y \in L^1, X_n \to X \ a.s) \implies \lim_n \mathbb{E}(X_n|\mathcal{B}) = \mathbb{E}(X|\mathcal{B}).$$

8. *Jensen: for* $\Phi : \mathbb{R} \to \mathbb{R}$ *convex with* $\Phi(X) \in L^1$,

$$\Phi(\mathbb{E}(X|\mathcal{B})) \leq \mathbb{E}(\Phi(X)|\mathcal{B})$$

9. *Weakly contracting:* $\|\mathbb{E}(X|\mathcal{B})\|_p \leq \|X\|_p, \qquad p \geq 1$.

10. *Take off what is measurable: If* $Y \in L^\infty(\mathcal{B})$, $\mathbb{E}(YX|\mathcal{B}) = Y\mathbb{E}(X|\mathcal{B})$. *(Extends to* $X \in L^p, Y \in L^q$ *with* $p^{-1} + q^{-1} = 1$.*)*

11. $X \perp\!\!\!\perp \mathcal{B} \implies E(X|\mathcal{B}) = E(X)$.

12. $\mathcal{C} \perp\!\!\!\perp \sigma(X, \mathcal{B}) \implies \mathbb{E}(X|\sigma(\mathcal{C}, \mathcal{B})) = \mathbb{E}(X|\mathcal{B})$.

2.4 Examples

How to compute conditional expectations.

2.4.1 Countable partition

On $(\Omega, \mathcal{A}, P)$, a measurable partition $\mathfrak{P} = \{A_1, \ldots, A_n, \ldots\}$ of Ω, with $\mathfrak{P}$ finite or countable,

$$\Omega = \cup_i A_i, \qquad A_i \bigcap A_j = \emptyset \quad (i \neq j).$$

Without loss of generality, we assume $P(A_i) > 0$. The σ-field $\mathcal{B}$ conveys the information of which of the A_i did occur.
If $X \in L^1$,

$$E(X|\mathcal{B}) = \frac{E(X\mathbf{1}_{A_n})}{P(A_n)} \qquad \text{if } \{\omega \in A_n\}, \tag{2.3}$$

i.e. $= E(X|A_n)$ on A_n.

□ We check the definition 2.2.1. Any $B \in \mathcal{B}$ is of the form

$$B = \bigcup_{n: A_n \subset B} A_n.$$

Then, denoting by Z the RHS of (2.3), we compute

$$\begin{aligned}
E(Z\mathbf{1}_B) &= E(Z\sum_{n:A_n\subset B}\mathbf{1}_{A_n}) && \text{(pairwise disjoint)}\\
&= \sum_{n:A_n\subset B} E(Z\mathbf{1}_{A_n}) && \text{(Fubini)}\\
&= \sum_{n:A_n\subset B} \frac{E(X\mathbf{1}_{A_n})}{P(A_n)}\times P(A_n)\\
&= E(X\mathbf{1}_B) && \text{(Fubini)},
\end{aligned}$$

which is the definition (2.2.1). ∎

This is the standard notion of expectation conditionally on a discrete r.v. Y. Indeed, labeling $\{y_1, y_2, \ldots\}$ the values of Y and setting $A_n = \{Y = y_n\}$, we recover the standard definition

$$E(X|Y) = E(X|Y = y_n) \quad \text{on the event } A_n = \{Y = y_n\}.$$

Example 2.4.1. *Let $(\Omega, \mathcal{A}, P)$ =([0,1), Borel field, Lebesgue measure). For $n \geq 1$, consider the finite partition $\Pi = \{[k/n, (k+1)/n); k = 0, \ldots, n-1\}$. Then,*

$$E(X|\mathcal{B}) = \sum_{k=0}^{n-1} \mathbf{1}_{[k/n,(k+1)/n)}(\omega)\times n\int_{k/n}^{(k+1)/n} X(\omega')d\omega'$$

the average value of the function X on the interval of the partition containing the current point ω.

2.4.2 Computation with independence

We want to generalize item 11 of Proposition 2.3.1.

Proposition 2.4.2. *Consider random variables $X \perp\!\!\!\perp Y$ for P, with values in $\mathcal{X}, \mathcal{Y}$. Let $h : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ measurable bounded. Denote*

$$\psi(x) = E(h(x, Y)), x \in \mathcal{X}.$$

Then

$$E(h(X, Y)|X) = \psi(X) \quad \text{a.s.}$$

We can replace the assumption on h by "$h(X, Y) \in L^1$".

□ We need to show

$$E[h(X, Y)\mathbf{1}_B(X)] = E[\psi(X)\mathbf{1}_B(X)]$$

for all measurable subset B of $\mathcal{X}$. We compute

$$\begin{aligned}
E[h(X,Y)\mathbf{1}_B(X)] &= \int_{\mathcal{X}\times\mathcal{Y}} h(x,y)\mathbf{1}_B(x)dP_{XY}(x,y) \\
&= \int_{\mathcal{X}\times\mathcal{Y}} h(x,y)\mathbf{1}_B(x)dP_X(x)dP_Y(y) \qquad \text{(indep.)} \\
&= \int_{\mathcal{X}} dP_X(x)\mathbf{1}_B(x)\int_{\mathcal{Y}} h(x,y)dP_Y(y) \qquad \text{(Fubini)} \\
&= E\mathbf{1}_B(X)\psi(X).
\end{aligned}$$

We get the desired result. ∎

Exercise 2.4.3. *Let X, Y be independent standard Gaussian. What is the characteristic function of the product XY ? What is the law of $Z = X_1X_2 + X_3X_4$ with independent standard Gaussian X_i's ?*

□ By independence,

$$E(e^{itXY}|X) = \psi(X), \qquad \psi(x) = E(e^{itxY}) = e^{-t^2x^2/2}. \tag{2.4}$$

Compute the characteristic function

$$\begin{aligned}
\Phi_X(t) &= E(e^{itXY}) \\
&= E(E(e^{itXY}|X)) \qquad \text{(Tower property)} \\
&= E(e^{-t^2x^2/2}) \qquad \text{(by (2.4))} \\
&= (1+t^2)^{-1/2},
\end{aligned}$$

by computing the Gaussian integral. Then, by independence,

$$\Phi_Z(t) = \Phi_{X_1X_2}(t)\Phi_{X_3X_4}(t) = (1+t^2)^{-1}.$$

Since Φ_Z is integrable, Z has a density p_Z, given by Fourier inversion,

$$p_Z(z) = (2\pi)^{-1}\int_{\mathbb{R}} e^{-itz}(1+t^2)^{-1}dt = (1/2)e^{-|z|}, \qquad z \in \mathbb{R},$$

that is, Z has the Laplace distribution. ∎

2.4.3 Conditional probabilities

Since conditional expectation has all the properties of the standard expectation, we look for a "conditional probability", i.e. a family of measure parametrized by the conditioning which expectation is equal to the conditional expectation. We run into difficulties.

Put

$$P(A|\mathcal{B}) = E(\mathbf{1}_A|\mathcal{B}).$$

This is a r.v., so we choose for all $A \in \mathcal{A}$ a representative $\omega \mapsto P(A|\mathcal{B})(\omega)$. For a given sequence $\alpha = (A_n)_n$ pairwise disjoint in $\mathcal{A}$, we have

$$P(\cup_n A_n|\mathcal{B}) = \sum_n P(A_n|\mathcal{B}), \tag{2.5}$$

for all ω on a set Ω_α with $P(\Omega_\alpha) = 1$.

Then, $P(\cdot|\mathcal{B})(\omega)$ is a probability measure if $\omega \in \bigcap_\alpha \Omega_\alpha$. The problem is that there are uncountably many such sequences. Nothing guarantees that $\bigcap_\alpha \Omega_\alpha$ has full probability, it may even be empty.

What we need is described in

Definition 2.4.4. *A regular version of the conditional probability (abbreviated r.v.c.p.) given $\mathcal{B}$ is a function $P_{\mathcal{B}}(\cdot, \cdot\cdot) : \Omega \times \mathcal{A} \to [0,1]$ such that*

(i) $\forall A \in \mathcal{A}, P_{\mathcal{B}}(\omega, A) = E(\mathbf{1}_A|\mathcal{B})$ *a.s.,*

(ii) for a.e. ω, $P_{\mathcal{B}}(\omega, \cdot)$ is a probability measure on $\mathcal{A}$.

The solution does not always exist. Conditional expectation always exists, but not conditional probability. When it exists, it has nice properties:

Proposition 2.4.5. *If there exists a r.v.c.p., we have*

$$E(X|\mathcal{B}) = \int_\Omega X(\omega')P_{\mathcal{B}}(\omega, d\omega') \qquad \text{a.s.} \tag{2.6}$$

for $X \in L^1$.

This explains why conditional expectation has all the properties of an expectation.

□ By Definition 2.4.4-(ii), the RHS is a.s. defined. (Note that $X \in L^1(P)$ implies that $X \in L^1(P_{\mathcal{B}}(\omega, \cdot))$ for a.e. ω.)

By Definition 2.4.4-(i), the equality (2.6) is true for $X = \mathbf{1}_A$. By linearity, it is true for X a simple r.v. (i.e., taking finitely many values). By linearity it suffices to consider the case of a nonnegative r.v. X. Let X_n be a sequence of simple functions with $0 \le X_n \nearrow X$ as $n \nearrow \infty$. By monotone convergence (in the conditional version of Proposition 2.3.1),

$$E(X_n|\mathcal{B}) \nearrow E(X|\mathcal{B}) \qquad \text{a.s.,}$$

though by standard monotone convergence for ω in the set under consideration at point (ii) of Definition 2.4.4,

$$\int_\Omega X_n(\omega')P_{\mathcal{B}}(\omega, d\omega') \nearrow \int_\Omega X(\omega')P_{\mathcal{B}}(\omega, d\omega').$$

$\blacksquare$

Conditional probability does not always exist, however there are results for existence: If Ω is a polish space with Borel field $\mathcal{A}$, then the conditional probability $P_{\mathcal{B}}$ exists. Here is a consequence, when $\Omega = \mathbb{R}$ or $\mathbb{R}^d$.

Proposition 2.4.6. *Let $X : \Omega \to \mathbb{R}^d$ a r.v. Then there exists a regular version of the conditional law (r.v.c.l.) of X given $\mathcal{B}$, i.e., a random measure*

$$\Omega \ni \omega \mapsto \mu(dx|\mathcal{B})(\omega) \in \mathcal{P}(\mathbb{R}^d)$$

on the Borel sigma-field of $\mathbb{R}^d$ such that

$$\mu(A|\mathcal{B}) = E(\mathbf{1}_{\{X \in A\}}|\mathcal{B}) \quad \text{a.s.}$$

Example 2.4.7. *Three important examples.*

1. *Discrete case: Let $\mathcal{B}$ generated by a partition Π. Recall that $B \mapsto P(B|A_i) = P(B \cap A_i)/P(A_i)$ defines a probability measure $P(d\omega|A_i)$ on $\mathcal{A}$. Then,*

   $$P_{\mathcal{B}}(\omega, d\omega') = P(d\omega'|A_i) \qquad \text{on } \{\omega \in A_i\}.$$

   Let I be the r.v. defined by $I = i \iff \omega \in A_i$. Then, $\mathcal{B} = \sigma(I)$, and we recover the natural definition of the conditional probability $P(\cdot|I)$.

2. *Densities: Assume $\Omega = \mathbb{R}^m \times \mathbb{R}^k, \omega = (x, y)$, $dP(x, y) = p(x, y)dxdy$ with a density p, $\mathcal{B} = \sigma(Y)$. Then, $p_{X|Y}(x, y) = p(x, y)/p_Y(y)$ with $p_Y(y) = \int_{\mathbb{R}^m} dxp(x, y)$ (the marginal) defines a density in x for all y (the conditional density), and*

   $$P_{\mathcal{B}}(x, y; dx', dy') = p_{X|Y}(x', y)dx'\delta_y(dy'),$$

   with δ_y the Dirac mass at y. Alternatively, if (X, Y) has a density, then a r.v.c.l. of X given Y is given by the conditional density.

3. *For X, Y independent, the conditional law $P_{X|Y}$ is equal to the marginal P_X.*

Exercise 2.4.8. *Let (X, Y) be uniform on the half-disk $\{x^2+y^2 \leq 1; y \geq 0\}$. Find the conditional law of Y given X and compute the value of $E(Y|X)$.*

Exercise 2.4.9. *Let (X, Y) a bivariate Gaussian $\mathcal{N}_2(m, \Sigma)$,*

$$m = (a, b), \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

with $\sigma_1, \sigma_2 \geq 0, |\rho| < 1$.
(i) Using conditional densities, prove that a r.v.c.l. of X *given* Y *is given by the Gaussian*

$$\mathcal{N}\Big(a + \rho\frac{\sigma_1}{\sigma_2}(Y-b); \sigma_1^2(1-\rho^2)\Big).$$

(Indication: start with the case $m = 0$*.)*
(ii) Prove directly that

$$X = a + \rho\frac{\sigma_1}{\sigma_2}(Y-b) + Z, \qquad \text{with } Z \perp\!\!\!\perp Y, \quad Z \sim \mathcal{N}(0, \sigma_1^2(1-\rho^2)).$$

Chapter 3

Martingales

A powerful tool.

3.1 Stochastic processes and filtrations

Two important definitions. Time is discrete, $n = 0, 1, \ldots$.

Definition 3.1.1. *A stochastic process $X = (X_n; n \geq 0)$ on $(\Omega, \mathcal{A}, P)$ is a sequence of r.v. with values in some measurable space $(E, \mathcal{E})$.*

As its name indicates, it represents a random object evolving in time. Its law P_X is a probability measure on $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$, it is determined by its finite-dimensional marginal according to Proposition 1.1.3.

Definition 3.1.2. *A filtration $\mathcal{F} = (\mathcal{F}_n; n \geq 0)$ on $(\Omega, \mathcal{A}, P)$ is an increasing sequence of sub-sigma fields of $\mathcal{A}$,*

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{A}, \qquad n \geq 0.$$

Not all events in $\mathcal{A}$ are observable at time n, only those in $\mathcal{F}_n$. Thus, $\mathcal{F}_n$ represents the information available at time n. It increases in time (as it should do!).

A process X is *adapted* to the filtration $\mathcal{F}$ if X_n is $\mathcal{F}_n$-measurable for all n.

Example 3.1.3. *Filtration generated by a stochastic process X:*

$$\mathcal{F}_n^X = \sigma(X_i, i \leq n\}.$$

By construction, any process X is $\mathcal{F}^X$-adapted.

3.2 Martingales

Definition 3.2.1. *A real-valued process X is a $\mathcal{F}$-martingale if, for all $n \geq 0$,*

$$X_n \in L^1(\mathcal{F}_n)$$
$$E(X_{n+1}|\mathcal{F}_n) = X_n. \tag{3.1}$$

Hence, a martingale is an adapted, integrable sequence which is constant in conditional expectation. A *supermartingale* [resp., *submartingale*] is by definition adapted, integrable, and such that

$$E(X_{n+1}|\mathcal{F}_n) \leq X_n \qquad [\text{resp.} E(X_{n+1}|\mathcal{F}_n) \geq X_n], \tag{3.2}$$

Note that, for a martingale,

$$EX_n = EX_0, \qquad E(X_{n+p}|\mathcal{F}_n) = X_n, \tag{3.3}$$

for all$n, p \geq 0$, and for a supermartingale

$$EX_n \leq EX_0, \qquad E(X_{n+p}|\mathcal{F}_n) \leq X_n.$$

Example 3.2.2 (Sum of independent r.r.v.)**.** *Let $(\xi_n)_n$ be an i.i.d. sequence of r.r.v., integrable with mean 0. Then,*

$$S_n = \sum_{1 \leq i \leq n} \xi_i$$

is a $\mathcal{F}^\xi = \mathcal{F}_X$-martingale. Indeed, it is adapted and integrable; By linearity and independence,

$$E(S_{n+1}|\mathcal{F}_n) = E(S_n + \xi_{n+1}|\mathcal{F}_n) = S_n + E(\xi_{n+1}) = S_n$$

a.s.

The simplest case is $\xi_i \sim$ Ber(p), with a parameter $p \in (0, 1)$, according to the result of the ith game: $P(\xi_i = \pm 1) = p$. The variable ξ_i describes the algebraic gain at time i placing a unit bet. The game is fair if $p = .5$, and then the total gain S_n is a martingale; for $p > .5$ it is a submartingale and the game is favourable.

Example 3.2.3 (Product of independent r.r.v.)**.** *Let $(\xi_n)_n$ be an i.i.d. sequence of r.r.v., integrable with mean 1. Then,*

$$M_n = \prod_{1 \leq i \leq n} \xi_i$$

is a $\mathcal{F}^{\mathbf{x}}$-martingale. Indeed,

$$E(M_{n+1}|\mathcal{F}_n) = E(M_n \times \xi_{n+1}|\mathcal{F}_n) = M_n \times E(\xi_{n+1}) = M_n$$

a.s.

Example 3.2.4 (Gradually discover a random variable). *Let $X \in L^1(\Omega, \mathcal{A}, P)$ and $\mathcal{F}$ a filtration on this space, with $\mathcal{F}_0 = \{\Omega, \emptyset\}$ the trivial σ-field. Then*

$$M_n = E(X|\mathcal{F}_n) \tag{3.4}$$

is a martingale by the tower property. Finer and finer details are progressively added to discover the full variable X. $M_n = E(X|\mathcal{F}_n)$ is a nice interpolation between $M_0 = E(X)$ and X.

3.3 Strategies, predictable processes

Definition 3.3.1. *A process $C = (C_n)_{n\geq 1}$ is $\mathcal{F}$-predictable if C_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$.*

An example to keep in mind is when you vary the bet amount you place in the example 3.2.2. Before the nth game you determine the bet C_n in view of the previous results. In this case, $C_n \geq 0$. The sequence $C = (C_n)_n$ is called a strategy. Your gain will then be Y with

$$Y_n = \sum_{1\leq k\leq n} C_k(S_k - S_{k-1}) =: (C \cdot X)_n \tag{3.5}$$

after the nth game. The above transform

$$S \mapsto Y = C \cdot X$$

given a strategy is called a [sub-]martingale transform by a predictable process C, or a stochastic integral in the continuous time case.

Theorem 3.3.2 (You cannot beat the system). .

1. *If C is predictable, non-negative and bounded, then*

$$S \text{ martingale [resp. submart., resp.supermart.]} \implies Y \text{ too}$$

2. *In the case of a martingale, there is no need of non-negativity assumption on C.*

3. *For $S_n \in L^2$ we can relax the assumption of boundedness for C into "$C_n \in L^2$ for all n".*

□ Observe that

$$E(Y_{n+1} - Y_n|\mathcal{F}_n) = E(C_{n+1}(S_{n+1} - S_n)|\mathcal{F}_n) = C_{n+1}E(S_{n+1} - S_n|\mathcal{F}_n)$$

has the sign of the conditional expectation $E(S_{n+1} - S_n|\mathcal{F}_n)$. This yields the first two claims. The third one is routine. ■

3.4 Stopping times

To do some surgery on random paths we need special random times.

Definition 3.4.1. *Let $(\Omega, \mathcal{A}, P, \mathcal{F})$ be a filtered probability space. A random time $T : \Omega \to [0, \infty]$ is a $\mathcal{F}$-stopping time if*

$$\{T \leq n\} \in \mathcal{F}_n, \qquad \forall n < \infty.$$

Note T may be ∞. $\mathcal{F}_\infty$ is understood as $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$ the limit σ-field. A constant $T = n$ is a stopping time, and if S, T are stopping times, then $S \vee T$ and $S \wedge T$ are stopping times.

Example 3.4.2 (Entrance time in a Borel set)**.** *If X is adapted and $A \in \mathcal{E}$, the entrance time of X in A is*

$$T = \inf\{n \geq 0 : X_n \in A\},$$

with the convention that $\inf \emptyset = \infty$*. We have*

$$\{T \leq n\} = \cup_{k \leq n} \{X_k \in A\} \in \mathcal{F}_n,$$

showing it is a stoppimg time.

Remark 3.4.3 (Discrete time is simple)**.** .
(i) T is a stopping time iff

$$\{T = n\} \in \mathcal{F}_n, \qquad n < \infty.$$

Indeed,

$$\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\},$$

and

$$\{T \leq n\} = \cup_{k=0}^n \{T = k\}.$$

(ii) For a stopping time T, we even have

$$\{T \geq n\} \in \mathcal{F}_{n-1}, \tag{3.6}$$

since this event is the complement of $\{T \leq n-1\}$.

A typical example in the framework of repeated games, see example 3.2.2, is the time T when the player decides to stop playing (in view of his results): $\{T = n\}$ is based on the view of $S_i, i \leq n$.
A counterexample is

$$T = \sup\{n \leq 100 : X_n \in A\},$$

with the convention $\sup \emptyset = 0$.

Definition 3.4.4. *The σ-field $\mathcal{F}_T$ of events determined prior to the stopping time T is*

$$\mathcal{F}_T = \left\{ A \in \mathcal{F}_\infty : A \bigcap \{T \le n\} \in \mathcal{F}_n, \forall n \right\}.$$

Stopped process: If X is $\mathcal{F}$-adapted and T is a stopping time, we define

$$X^T = (X_n^T, n \ge 0), \qquad X_n^T(\omega) = X_{T(\omega)\wedge n}(\omega). \tag{3.7}$$

In the previous typical example, S^T is the fortune process of the player who stops at time T. In full generality, X^T is $\mathcal{F}$-adapted, since

$$X_n^T = \sum_{i=0}^{n-1} X_i \mathbf{1}_{T=i} + X_n \mathbf{1}_{T \ge n},$$

where all the terms are $\mathcal{F}_n$-measurable.

3.4.1 Optional stopping

Theorem 3.4.5. .
(i) If X is a supermartingale and T a stopping time, then X^T is a supermartingale, and then $E(X_{T\wedge n}) \le EX_0$.
(ii) If X is a martingale and T a stopping time, then X^T is a martingale, and then $E(X_{T\wedge n}) = EX_0$.

□ To T we associate the strategy $C_n = \mathbf{1}_{\{T \ge n\}}$, i.e., we place a unit bet at the nth game if $T \ge n$. By (3.6), C is predictable. The transform is

$$\begin{aligned}
(C \cdot X)_n &= \sum_{k=1}^{n} C_k (X_k - X_{k-1}) \\
&= \sum_{k \ge 1} C_k (X_k - X_{k-1}) \mathbf{1}_{k \le n} \\
&= \sum_{k \ge 1} (X_k - X_{k-1}) \mathbf{1}_{k \le n \wedge T} \\
&= X_{n \wedge T} - X_0,
\end{aligned}$$

hence $X_{n\wedge T} = X_0 + (C \cdot X)_n$ and we apply Theorem 3.3.2 with a non-negative C. ∎

When T is a.s. finite, then $X_T : \omega \mapsto X_{T(\omega)}(\omega)$ is well-defined, and is $\mathcal{F}_T$-measurable. Note that $X_{T\wedge n}$ is necessarily integrable, $|X_{T\wedge n}| \le \sum_{k=1}^n |X_k - X_{k-1}|$. However, even if $T < \infty$ a.s., there is no reason for X_T to be integrable, and $EX_T \ne EX_0$ can occur!

Example 3.4.6 (attention!). *In the setup of example 3.2.2, with* $P(\xi = \pm 1) = 1/2$, *let*

$$T = \inf\{n \geq 0 : S_n = +1\},$$

the hitting time of 1 *for the simple symmetric random walk* S *starting from 0. It can be shown that* $T < \infty$ *a.s. When* $T < \infty$, *the value* S_T *is well-defined, and here it is equal to 1. Hence,* $S_T = 1$ *a.s., and*

$$ES_T = 1 \neq ES_0 = 0.$$

In order to have the equality some assumption is needed.

Theorem 3.4.7 (Doob's optional stopping theorem). *Let* X *be a supermartingale and* T *a stopping time. If* T *is a.s. bounded (i.e.,* $\exists C < \infty, T \leq C$ *a.s.), then*

$$EX_T \leq EX_0.$$

If X *is a martingale, the equality holds.*

□ We know already that $EX_{T\wedge n} \leq EX_0$ for all n, with equality in the case of a martingale. Take $n = [C]$. ∎

The assumptions can be refined, as shown in the next result.

Proposition 3.4.8 (Stopping theorem, continued). *With* X *a supermartingale [resp., martingale] and* T *an a.s. finite stopping time, the conclusions still hold in the two following cases:*

1. $\exists K < \infty : |X_{n\wedge T}| \leq K$ *a.s.*
2. $ET < \infty$ *and* $\exists K < \infty : |X_n - X_{n-1}| \leq K$.

□ We know already that $EX_{T\wedge n} \leq EX_0$ for all n. In case 1) we have $EX_{T\wedge n} \to EX_T$ by Lebesgue dominated convergence.
In case 2),

$$|X_{T\wedge n} - X_0| = |\sum_{k=1}^{T\wedge n}(X_k - X_{k-1})| \leq K(T \wedge n) \leq KT \in L^1.$$

Then, $X_{T\wedge n} - X_0 \to X_T - X_0$ in L^1 by dominated convergence, and also $EX_{T\wedge n} \to EX_T$. ∎

Proposition 3.4.9 (Stopping theorem, ended). *With* X *a non-negative supermartingale and* T *an a.s. finite stopping time. Then,* $0 \leq EX_T \leq EX_0$.

□ By Fatou's lemma,

$$\liminf_{n\to\infty} EX_{T\wedge n} \geq E \liminf_{n\to\infty} X_{T\wedge n} = X_T.$$

∎

Remark 3.4.10. *Doob's optional stopping time theorem essentially says that you can't increase your fortune on the average by buying and selling an asset whose price is a martingale. Precisely, if you buy the asset at some time and adopt any admissible strategy for deciding when to sell it, then the expected price at the time you sell is the price you originally paid. On financial markets, any asset prices are believed to behave approximately like martingales, at least in the short term. This is called the efficient market hypothesis: new information is instantly absorbed into the stock value, so expected value of the stock tomorrow should be the value today. If it were higher, statistical arbitrageurs would bid up today's price until this was not the case. In fact, one needs to incorporate other factors like interest, risk premium, etc. According to the fundamental axiom of asset pricing, the discounted price X_n/F_n , where F_n is a risk-free asset, is a martingale with respected to risk neutral probability.*

3.4.2 Example: Gambler's fortune process

Exercise 3.4.11 (Gambler's ruin)**.** *Let $\xi_i, i \geq 1$, be an i.i.d. sequence of Bernoulli random variables taking values ± 1 with probability $1/2$. Then,*

$$S_n = \xi_1 + \ldots + \xi_n$$

is a martingale. Let $a, b \geq 1$ be integers, and

$$T = \inf\{n \geq 0 : S_n = -a \text{ or } b\}.$$

Then, S is the algebraic gain of a player in a fair game, and T is the duration of the game, due to the ruin of the player ($S_T = -a$) or of its opponent ($S_T = b$). Prove that:

1. *$T < \infty$ a.s. if a, b are finite.*
2. *$P(S_T = -a) = b/(a+b), P(S_T = b) = a/(a+b)$.*
   [Hint: Use martingale S_n.]
3. *$ET = ab$.*
   [Hint: Check that $S_n^2 - n$ is a martingale, and use it.]

4. **From now on** $a = \infty$ **and** $b = 1$**.** *Then* T *is a.s. finite and*

$$Ez^T = z^{-1}[1 - (1 - z^2)^{1/2}] \quad \text{for } z \in (0, 1].$$

[Hint: Look for a relation $z(\theta)$ *between* z *and* θ *such that* $z^{-n} exp\{\theta S_n\}$ *is a martingale.]*

5. *Deduce that*

$$P(T = 2k - 1) = \frac{k \times (2(k-1))!}{(k!)^2 2^{2k-1}}, k = 1, 2 \dots.$$

6. *Check that for all* k*,* $\frac{d^k}{d\theta^k}\left(z(\theta)^{-n} exp\{\theta S_n\}\right)_{|\theta=0}$ *is a martingale. What do we find for* $k = 1, 2, 3$ *?*

Solution: 1. Observe that, with $\delta = (1/2)^{a+b}$,

$$P(T > (k+1)(a+b) | \mathcal{F}^S_{k(a+b)}) \leq 1 - \delta \quad \text{on } \{T > k(a+b)\}. \tag{3.8}$$

Indeed, if $T > k(a+b)$,

$$\begin{aligned} P(T \leq (k+1)(a+b) | \mathcal{F}^S_{k(a+b)}) &\geq P(\xi_n = +1, n = k(a+b)+1, \dots, (k+1)(a+b) | \mathcal{F}^S_{k(a+b)}) \\ &= P(\xi_n = +1, n = k(a+b)+1, \dots, (k+1)(a+b)) \\ &= \delta, \end{aligned}$$

which is the claim (3.8). Hence,

$$\begin{aligned} P(T > (k+1)(a+b)) &= EP(T > (k+1)(a+b) | \mathcal{F}^S_{k(a+b)}) \\ &= EE(\mathbf{1}_{\{T > (k+1)(a+b)\}} | \mathcal{F}^S_{k(a+b)}) \\ &= EE(\mathbf{1}_{\{T > k(a+b)\}} \mathbf{1}_{\{T > (k+1)(a+b)\}} | \mathcal{F}^S_{k(a+b)}) \\ &= E[\mathbf{1}_{\{T > k(a+b)\}} E(\mathbf{1}_{\{T > (k+1)(a+b)\}} | \mathcal{F}^S_{k(a+b)})] \\ &\leq (1 - \delta) E[\mathbf{1}_{\{T > k(a+b)\}}] \qquad \text{(by 3.8)} \\ &= (1 - \delta) P(T > k(a+b)) \\ &\leq (1 - \delta)^{k+1}, \end{aligned} \tag{3.9}$$

by iterating. Finally,

$$P(T = \infty) = \lim_{n \nearrow \infty} (\searrow) P(T > (k+1)(a+b) = 0,$$

which is the claim.

2. Write

$$0 = E(S_0) = ES_{n \wedge T} = E(S_n; n \leq T) - aP(S_T = -a, T < n) + bP(S_T = b, T < n),$$

and take the limit as $n \to \infty$.

3. With $Y_n = S_n^2 - n$, write

$$\begin{aligned} 0 &= E(Y_0) = EY_{n\wedge T} \\ &= E(S_n^2; n \le T) + a^2 P(S_T = -a, T < n) + b^2 P(S_T = b, T < n) - E(T \wedge n), \end{aligned}$$

and take the limit as $n \to \infty$.

4. First, we see from the second step that

$$P(T < \infty) = \lim_{a\to\infty} P(S \text{ hits } b \text{ before } -a) = \lim_{a\to\infty} \frac{b}{a+b} = 0. \qquad (3.10)$$

Compute now

$$Ee^{\theta\xi_n} = \frac{e^{\theta} + e^{-\theta}}{2} = \cosh(\theta),$$

and denote for short

$$z(\theta) = \cosh(\theta)^{-1} \in (0, 1]$$

Recalling the martingales associated to product of independent r.r.v., cf. Example 3.2.3, we deduce that the sequence

$$M_n(\theta) = z(\theta)^n exp\{\theta S_n\}, \qquad n \ge 0, \qquad (3.11)$$

is a martingale for all real θ. We have discovered infinitely many martingales in this problem ! By the optional stopping theorem,

$$EM_{n\wedge T}(\theta) = \mathbb{E}M_0 = 1,$$

for all finite n. We now take $\theta \ge 0$, in which case we can bound $M_{n\wedge T}(\theta) \le e^{\theta}$, since $z \in (0, 1]$. As $n \to \infty$, we have $\mathbb{E}M_{n\wedge T}(\theta) \to \mathbb{E}M_T(\theta)$ by dominated convergence. Since $S_T = 1$, we find

$$Ez(\theta)^T = e^{-\theta}, \qquad \theta \ge 0.$$

Now, we invert $z : [0, +\infty) \to (0, 1]$: since $e^{\theta} + e^{-\theta} = 2/z$, we solve $e^{2\theta} - (2/z)e^{\theta} + 1 = 0$, that is $e^{-\theta} = z^{-1}[1 - \sqrt{1 - z^2}]$, and the moment generating function G of T is

$$E(z^T) = \frac{1 - \sqrt{1 - z^2}}{z}, \qquad z \in (0, 1]. \qquad (3.12)$$

From Taylor series expansion, for all $n \in \mathbb{R}$,

$$(1 + u)^n = \sum_{k\ge 0} \binom{n}{k} u^k, \qquad |u| < 1,$$

with the binomial coefficient $\binom{n}{k} = \frac{n(n-1)...(n-k+1)}{k!}$, we have

$$\mathbb{E}(z^T) = \sum_{k\geq 1}(-1)^{k+1}\binom{1/2}{k}z^{2k-1}.$$

Therefore,

$$P(T = 2k-1) = (-1)^{k+1}\binom{1/2}{k} = \frac{k \times (2(k-1))!}{(k!)^2 2^{2k-1}}.$$

6. Since the sequence $M_n(\theta) = z(\theta)^n exp\{\theta S_n\}$ is a martingale for all θ, we want to differenciate the identity

$$E(M_{n+1}(\theta)|\mathcal{F}_n) = M_n(\theta).$$

Here, we easily justify the interchange of expectation and derivation for the integrand is bounded as well as its derivatives, so we get

$$E(\frac{d^k}{d\theta^k}M_{n+1}(\theta)|\mathcal{F}_n) = \frac{d^k}{d\theta^k}E(M_{n+1}(\theta)|\mathcal{F}_n) = \frac{d^k}{d\theta^k}M_n(\theta),$$

and we see that the $M_n^{(k)}(\theta)$ is a martingale for all θ. We compute

$$\begin{aligned}
M_n(\theta) &= e^{\theta S_n - n\ln\cosh\theta}, \\
M_n'(\theta) &= M_n(\theta)\Big[S_n - n\tanh\theta\Big], \\
M_n"\theta) &= M_n(\theta)\Big[(S_n - n\tanh\theta)^2 - n(1-\tanh^2\theta)\Big], \ldots
\end{aligned}$$

For $\theta = 0$,

$$M_n'(0) = S_n, \quad M"_n(0) = S_n^2 - n, \quad M_n^{(3)}(0) = S_n^3 - 3nS_n,$$

are martingales, and we recognize the first two ones. ∎

3.5 Doob's decomposition

Theorem 3.5.1 (Doob's decomposition)**.** *Let $(X_n)_n$ an adapted, integrable process. Then, there exists a unique decomposition*

$$X_n = X_0 + M_n + A_n, \qquad n \geq 0,$$

with M a martingale and A predictable integrable, with $M_0 = A_0$.

A and M are called the compensator and the martingale part of X.

□ *Analysis*: If such a decomposition exists, then necessarily

$$\begin{aligned} E(X_{n+1} - X_n | \mathcal{F}_n) &= E(M_{n+1} - M_n | \mathcal{F}_n) + E(A_{n+1} - A_n | \mathcal{F}_n) \\ &= A_{n+1} - A_n, \end{aligned}$$

for M a martingale and A predictable. By summing the previous equality and using $A_0 = 0$, we find that the

$$A_n = \sum_{k=1}^{n} E(X_k - X_{k-1} | \mathcal{F}_{k-1}), \tag{3.13}$$

and then

$$M_n = \sum_{k=1}^{n} \big(X_k - E(X_k | \mathcal{F}_{k-1})\big). \tag{3.14}$$

Synthesis: Formulae (3.13) and (3.14) define a martingale and a predictable integrable process such that $X_n = X_0 + M_n + A_n$. ■

Proposition 3.5.2. *Let X be $\mathcal{F}$-adapted, denote by $X_n = X_0 + M_n + A_n$ its Doob decomposition. Then,*

$$X \text{ is a submartingale} \iff A_{n-1} \le A_n \text{ a.s. for all } n \ge 1.$$

In this case, A is called an increasing process.

□ Proof: clear from the identity $E(X_n | \mathcal{F}_{n-1}) - X_{n-1} = A_n - A_{n-1}$. ■

Example 3.5.3 (Bellman's optimatility principle)**.** *For the gambler's fortune process in a favourable game, we look for the optimal strategy. Denote by ξ_n the issue of the nth game, a Bernoulli variable with parameter $p > 1/2$:*

$$P(\xi_n = +1) = p, \qquad P(\xi_n = -1) = q = 1 - p.$$

Assume the $\xi_n, n \ge 1$, are i.i.d. For a strategy $(C_n)_{n \ge 1}$, i.e. a sequence $C_n \ge 0$ of $\mathcal{F}_{n-1}$-measurable variables, the fortune at time n is

$$Z_n = Z_0 + \sum_{i=1}^{n} C_i \xi_i.$$

To keep it positive in all circumstances, we will assume $C_n \le Z_{n-1}$, and of course $Z_0 > 0$ (say, a constant). In fact, we will even assume

$$C_n \in [0, (1-\varepsilon) Z_{n-1}] \tag{3.15}$$

for some small constant ε.The problem is to determine the choice of C optimizing the **interest rate**

$$E \ln \frac{Z_n}{Z_0} \ .$$

To do so, we first write the Doob decomposition of $X_n = \ln(Z_n/Z_0)$ (integrability is guaranteed by (3.15)): by (3.13) we have

$$\begin{aligned}
A_n - A_{n-1} &= E\left(\ln \frac{Z_n}{Z_0} - \ln \frac{Z_{n-1}}{Z_0} | \mathcal{F}_{n-1}\right) \\
&= E\left(\ln \frac{Z_n}{Z_{n-1}} | \mathcal{F}_{n-1}\right) \\
&= E\left(\ln[1 + \frac{C_n \xi_n}{Z_{n-1}}] | \mathcal{F}_{n-1}\right) \\
&= \Psi\left(\frac{C_n}{Z_{n-1}}\right),
\end{aligned}$$

since $\frac{C_n}{Z_{n-1}}$ is $\mathcal{F}_{n-1}$-measurable and ξ_n is independent of $\mathcal{F}_{n-1}$. By proposition 2.4.2, we have

$$\Psi(r) = E(\ln(1 + r\xi)) = p \ln(1 + r) + q \ln(1 - r).$$

Then,

$$A_n = \sum_{k=1}^{n} \Psi\left(\frac{C_k}{Z_{k-1}}\right),$$

This expression is not simple, and the computation does not seem to be useful. However, it is natural to study the function Ψ,

$$\Psi'(r) = \frac{(p - q) - r}{1 - r^2},$$

so it has a unique maximum at $r = p - q$. Thus,

$$\forall r, \qquad \Psi(r) \leq \Psi(p - q) = p \ln(2p) + q \ln(2q) \stackrel{\text{def.}}{=} \alpha \in (0, \ln 2).$$

From this we derive that

$$Y_n = \ln \frac{Z_n}{Z_0} - n\alpha$$

is a supermartingale for all C, and

$$E(\ln \frac{Z_n}{Z_0}) \leq n\alpha \ . \tag{3.16}$$

On the other hand, by choosing

$$C_n = (p - q) Z_{n-1} \qquad \forall n \geq 1, \tag{3.17}$$

we achieve the equality in (3.16).

Therefore the optimal strategy is given by (3.17), which correspond for the player to bet at each step a proportion $p-q \in (0,1)$ of his current fortune. The optimal interest rate per unit time is α.

Exercise 3.5.4 (Likelihood ratio)**.** *Let μ and ν two probability measures on $(E, \mathcal{E})$. Assume $\mu \ll \nu$, and denote by h the Radon-Nikodym derivative $h = d\mu/d\nu$. Let $X = (X_i, i \leq n)$ an i.i.d. sequence with law ν. The likelihood ratio of μ with respect to ν observing the n-sample X, is defined in statistics as*

$$L_n = \prod_{i=1}^{n} h(X_i).$$

Prove that L_n is a martingale.

$\triangleright$ By definition of h,

$$Eh(X_i) = \int_E h(x) d\nu(x) = \int_E d\mu(x) = 1.$$

Then, L_n is a product of independent r.r.v. with mean 1, so it is a martingale by Example 3.2.3. $\triangleleft$

Chapter 4

Martingale convergence

4.1 Almost sure convergence

Fact F: If a monotone sequence of real numbers is bounded, then it has a finite limite.

Here, our fundamental result is:

Theorem 4.1.1 (Supermartingale bounded in L^1)**.** *If $X = (X_n)_n$ is a supermartingale such that $\sup_n E|X_n| < \infty$, then*

$X_\infty = \lim_{n\to\infty} X_n$ *exists a.s., and is a.s. finite.*

This implies the exact analogue of the above fact.

Corollary 4.1.2. *A non-negative supermartingale converges a.s. The limit is a non-negative, finite r.v.*

□ Indeed, $\sup_n E|X_n| = \sup_n EX_n \leq EX_0$. ■

Example 4.1.3 (Product of positive i.i.d.r.v's)**.** *Let $(\xi_i, i \geq 1)$ be positive i.i.d.r.v. with mean $E\xi = 1$, and*

$$M_n = \prod_{i=1}^{n} \xi_i$$

Then, M_n is a positive martingale. Hence,

$$M_n \to M_\infty \quad \text{a.s.}$$

If ξ is not constant, we have $M_\infty = 0$. Indeed, by Jensen inequality,

$$m = E \ln \xi_i < \ln E\xi_i = 0,$$

where the inequality is strict because ln *is strictly convex and* ξ_i *is not a.s. constant. Hence,*

$$M_n = \exp\{\sum_{i=1}^{n} \ln \xi_i\}, \quad \text{with} \quad \ln \xi_i \sim nm \to -\infty \quad \text{a.s.},$$

by the strong law of large numbers. Hence $M_n \to 0$.

Warning: the result does not implies convergence in L^1. See the previous example.

Lemma 4.1.4 (Doob's upcrossings lemma)**.** *For* $-\infty < a < b < +\infty$, *define the number of upcrossings of* X *from* a *to* b *in time* $[0, n]$ *by*

$$U_n[a,b] = \max\{k : \exists 0 \le s_1 < t_1 < s_2 < t_2 < \ldots < s_k < t_k \le n : X_{s_i} < a, X_{t_i} > b\}.$$

If $X = (X_n)_n$ *is a supermartingale, then, with* $x^- = \max\{-x, 0\}$,

$$EU_n[a,b] \le \frac{E[(X_n - a)^-]}{b-a}$$

□ Lemma $\implies$ Theorem: Consider

$$\begin{aligned} B &= \{\omega : (X_n(\omega))_n \text{ does not converge in } [-\infty, +\infty]\} \\ &= \left\{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right\} \\ &= \cup_{a,b\in\mathbb{Q}, a<b} \left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\} \\ &= \cup_{a,b\in\mathbb{Q}, a<b} \{U_\infty[a,b] = \infty\} \end{aligned}$$

By monotonicity,

$$\begin{aligned} EU_\infty[a,b] = \lim_{n\to\infty} (\nearrow) EU_n[a,b] &\le \frac{\sup_n E[(X_n - a)^-]}{b-a} \\ &\le \frac{\sup_n E[|X_n|] + a}{b-a} < \infty. \end{aligned}$$

Hence $U_\infty[a,b] < \infty$ a.s., and $P(B) = 0$. Hence, $X_n \to X_\infty \in \bar{\mathbb{R}}$, and it remains to show finiteness of the limit. Fatou's lemma implies

$$E|X_\infty| \le \liminf_n E|X_n| \le \sup_n E|X_n| < \infty,$$

hence not only X_∞ is finite but it is even integrable. ∎

□ Proof of the Lemma. Consider the transform $Y = C \cdot X$,

$$Y_n = \sum_{i=1}^{n} C_i (X_i - X_{i-1})$$

with $C_n \in [0,1]$ predictable with

$$\begin{cases} C_1 &= \mathbf{1}_{\{X_0<a\}} \\ C_n &= \mathbf{1}_{\{C_{n-1}=1, X_{n-1}\le b\}} + \mathbf{1}_{\{C_{n-1}=0, X_{n-1}<a\}} \end{cases}$$

For a path with upcrossings $0 \le s_1 < t_1 < s_2 < t_2 < \ldots < s_k < t_k \le n$ we have the identity

$$Y_n = \sum_{i=1}^{k} (X_{t_i} - X_{s_i}) + \mathbf{1}_{\{s_{k+1}<n\}}(X_n - X_{s_{k+1}}),$$

with $s_{k+1} = \inf\{i \ge t_k + 1 : X_i < a\}$. Since $X_{s_{k+1}} < a$ on $\{s_{k+1} < n\}$, we have

$$\begin{aligned} Y_n &\ge k(b-a) + \mathbf{1}_{\{s_{k+1}<n\}}(X_n - a) \\ &\ge k(b-a) - \mathbf{1}_{\{s_{k+1}<n\}}(X_n - a)^- \\ &\ge k(b-a) - (X_n - a)^- \end{aligned}$$

Yet, Y is a supermartingale by Theorem "you cannot beat the system". Hence,

$$EY_n \le EY_0 = 0,$$

and recalling that $U_n[a,b] = k$ above, this writes

$$(b-a)EU_n[a,b] \le E[(X_n - a)^-],$$

as wanted. ∎

Example 4.1.5 (Galton-Watson process)**.** *Let $\xi_i^n (i \ge 1, n \ge 0)$ a doubly-indexed sequence of i.i.d.r.v.'s taking integer values, and $Z_n \in \mathbb{N}$ the random sequence starting from $Z_0 = 1$, given by*

$$Z_{n+1} = \sum_{i \le Z_n} \xi_i^n$$

Assuming $m = E\xi_i^n$ finite, the sequence

$$W_n = m^{-n} Z_n$$

is a $\mathcal{F}$-martingale, with $\mathcal{F}_n = \sigma(\xi_i^k; i \ge 1, k = 0, \ldots, n-1)$. The positive martingale W_n converges a.s. to some non negative variable W_∞. By Fatou's lemma, $EW_\infty \in [0,1]$. The Laplace transform of a non negative r.v. W

$$\mathcal{L}_W(\lambda) = E\exp\{-\lambda W\},$$

is well defined from $\mathbb{R}^+$ to $(0,1]$, and it characterizes the law of W. By dominated convergence,

$$\mathcal{L}_{W_n}(\lambda) = E\exp\{-\lambda W_n\} \to \mathcal{L}_{W_\infty}(\lambda).$$

Let also $G(z) = Ez^\xi$ denote the generating function. From the recursive structure and the independence of the ξ_i^n, we can decompose

$$W_{n+1} = \frac{1}{m} \sum_{i \le \xi_1^0} W_n^i \ , \tag{4.1}$$

with $(W_n^i)_{i\ge1}$ independent with same distribution as W_n, and also independent of $(\xi_i^0, i \ge 1)$. Then,

$$\begin{aligned}
\mathcal{L}_{W_{n+1}}(\lambda) &= E\exp\{-\lambda W_{n+1}\} \\
&= E\Big[\exp\Big\{-\frac{\lambda}{m}\sum_{i\le\xi_1^0} W_n^i\Big\}\Big] \qquad \text{(by (4.1))} \\
&= EE\Big[\exp\Big\{-\frac{\lambda}{m}\sum_{i\le\xi_1^0} W_n^i\Big\}\Big|\xi_1^0\Big] \\
&= E\Big[\mathcal{L}_{W_n}(\lambda/m)^{\xi_1^0}\Big] \qquad \text{(indep.)} \\
&= G(\mathcal{L}_{W_n}(\lambda/m)). \qquad (4.2)
\end{aligned}$$

Taking the limit $n \to \infty$,

$$\mathcal{L}_{W_\infty}(\lambda) = G(\mathcal{L}_{W_\infty}(\lambda/m)).$$

By monotone convergence,

$$\mathcal{L}_{W_\infty}(\lambda) \searrow \sigma \stackrel{\text{def}}{=} P(W_\infty = 0), \qquad \text{as } \lambda \nearrow +\infty,$$

and we find that $\sigma = P(W_\infty = 0)$ is a fixed point in $[0,1]$ of the mapping G,

$$\sigma = G(\sigma) \ . \tag{4.3}$$

Example 4.1.6 (Polya urns, with 2 colors)**.** [(1)] *Initially the urn contains 1 red ball and 1 blue ball. At each time $n = 1, 2 \ldots$, a ball is randomly drawn in the urn and a new ball of the same color is then added. Let B_n, R_n, be the numbers of blue/red balls in the urn at time n, so that $B_n + R_n = n + 2$.*

[(1)] Polya urns are simplified model for emergence of standards, as we emphasize now. In economics and business, *network effects* are frequent: every user of a good or service has an effect on the value of that product to other people. When a network effect is present, the value of a product or service is dependent on the number of others using it.

A classic example is the telephone. The more people who own telephones, the more valuable the telephone is to each owner. This creates a positive externality because a user may purchase a telephone without intending to create value for other users, but does so in any case. Online social networks work in the same way, with sites like Twitter and Facebook becoming more attractive as more users join.

1. *The proportion $M_n = B_n/(B_n + R_n)$ is a martingale.*
   Let $\mathcal{F}_n = \sigma(B_i, i \le n) = \sigma(B_i, R_i, i \le n)$. Then, conditonnally on $\mathcal{F}_n$,

$$B_{n+1} = B_n + \begin{cases} 1 & \text{with probability} \quad M_n \\ 0 & \text{with probability} \quad 1 - M_n \end{cases}$$

   Hence,

$$E[B_{n+1}|\mathcal{F}_n] = B_n + M_n.$$

   Since $B_n + R_n = n + 2$ and $M_n = B_n/(B_n + R_n)$, , we get

$$E[M_{n+1}|\mathcal{F}_n] = \frac{1}{n+3}\Big(B_n + \frac{B_n}{n+2}\Big) = \frac{(n+3)B_n}{(n+3)(n+2)} = M_n.$$

2. *$M_n \to M_\infty$ a.s.*
   Since $M_n \in [0, 1]$, Theorem 4.1.1 applies.

3. *$P(B_n = k) = 1/(n + 1), 1 \le k \le n + 1$.*
   We prove it by recursion. Assume the relation holds for n and all k's.
   For $k = 2, \dots n + 1$, write

$$\begin{aligned} P(B_{n+1} = k) &= E[\mathbf{1}_{\{B_n=k-1,B_{n+1}-B_n=1\}} + \mathbf{1}_{\{B_n=k,B_{n+1}-B_n=0\}}] \\ &= E\Big[E[\mathbf{1}_{\{B_n=k-1,B_{n+1}-B_n=1\}} + \mathbf{1}_{\{B_n=k,B_{n+1}-B_n=0\}}|\mathcal{F}_n]\Big] \\ &= E\Big[\mathbf{1}_{\{B_n=k-1\}}P(B_{n+1} - B_n = 1|\mathcal{F}_n) \\ &\qquad\qquad + \mathbf{1}_{\{B_n=k\}}P(B_{n+1} - B_n = 0|\mathcal{F}_n)\Big] \\ &= E\Big[\mathbf{1}_{\{B_n=k-1\}}M_n + \mathbf{1}_{\{B_n=k\}}(1 - M_n)\Big] \\ &= \frac{k-1}{(n+2)(n+1)} + \frac{n+2-k}{(n+2)(n+1)} \\ &= \frac{1}{n+2}. \end{aligned}$$

   The last two cases are treated separately:

$$\begin{aligned} P(B_{n+1} = 1) &= E[\mathbf{1}_{\{B_n=1\}}P(B_{n+1}-B_n=0|\mathcal{F}_n)] \\ &= E[\mathbf{1}_{\{B_n=1\}}(1 - M_n)] \\ &= \frac{n+1}{(n+2)(n+1)}, \end{aligned}$$

   and

$$P(B_{n+1} = n+2) = E[\mathbf{1}_{\{B_n=n+1\}}P(B_{n+1}-B_n = 1|\mathcal{F}_n)] = \frac{n+1}{(n+2)(n+1)},$$

   yielding the desired result.

4. *M_∞ is uniform in $(0,1)$.*
   Indeed, we just showed that $M_n = B_n/(n+2)$ is uniform on the multiples of $(n+2)^{-1}$ in $(0,1)$.

5. *Generalization: Assume now that $B_0 = b, R_0 = r$ are fixed, arbitrary real numbers > 0, and that (B_n, R_n) evolves as above: it becomes (B_n+1, R_n) with probability $B_n/(B_n+R_n)$ at time $n+1$, or (B_n, R_n+1) with probability $R_n/(B_n+R_n)$. Show that, for all integer k,*

$$X_n^{(k)} = \frac{B_n(B_n+1)\dots(k-1+B_n)}{(b+r+n)(b+r+n+1)\dots(b+r+n+k-1)}$$

   is a martingale. (Note that for $k=1$ we simply have $X_n^{(1)} = M_n$.)
   This follows form computations similar to question 1.

6. *Deduce the moments of M_∞ in this general case, and then that M_∞ has a Beta(b,r) distribution,*

$$M_\infty \sim \frac{\Gamma(b+r)}{\Gamma(b)\Gamma(r)} u^{b-1}(1-u)^{r-1}\mathbf{1}_{(0,1)}(u).$$

   By the stopping theorem and boundedness of M,

$$X_\infty^{(k)} = \lim_n X_n^{(k)} \text{a.s.}, \quad E(X_\infty^{(k)}) = \lim_n E(X_n^{(k)}) = E(M_\infty^k).$$

   Since $E(X_n^{(k)}) = E(X_0^{(k)}) = \frac{b(b+1)\dots(b+k-1)}{(b+r+n)(b+r+n+1)\dots(b+r+n+k-1)}$, we can compute the moments of M_∞. But they characterize the law of the bounded variable M_∞. It suffices then to check that for Y a Beta(b,r) distribution, we have

$$E(Y^k) = \frac{b(b+1)\dots(b+k-1)}{(b+r+n)(b+r+n+1)\dots(b+r+n+k-1)}.$$

7. *Assume now that $b = B_0, r = R_0$ with $b, r > 1$. Then, M_n is still a martingale converging to some M_∞. Show that*

$$Z_n = \ln(B_n + R_n) - \ln(B_n - 1)$$

   is a supermartingale. Deduce that $M_\infty > 0$ a.s., and by symmetry, that $M_\infty < 1$.

Indeed, on the event $\{B_n = x\}$,

$$\begin{aligned}
E(Z_{n+1}-Z_n|\mathcal{F}_n) &= \ln\frac{n+3}{n+2} - E\Big(\ln\frac{B_{n+1}-1}{B_n-1}|\mathcal{F}_n\Big) \\
&= \ln\frac{n+3}{n+2} - E\Big(\mathbf{1}_{B_{n+1}=B_n+1}\ln\frac{B_{n+1}-1}{B_n-1}|\mathcal{F}_n\Big) \\
&= \ln\frac{n+3}{n+2} - \frac{x}{n+2}\ln\frac{x}{x-1} \\
&= \ln(1+\frac{1}{n+2}) + \frac{x}{n+2}\ln(1-\frac{1}{x}) \\
&\overset{m=n+2}{=} (\frac{1}{m} - \frac{1}{2m^2} + \dots) - \frac{x}{m}(\frac{1}{x} + \frac{1}{2x^2} + \dots) \\
&\leq 0.
\end{aligned}$$

Hence Z is a positive supermartingale. Thus, $Z_n \to Z_\infty$ finite. Finally, $M_\infty = \lim_n M_n = \lim_n \exp(-Z_n) = \exp(-Z_\infty)$ is strictly positive a.s.. Exchanging the role of X and Y, we get also that $M_\infty < 1$. (This was already known from the previous question. However our supermartingale approach is robust to small perturbations, since being a supermartingale requires only an inequality.)

8. *Polya's urn starting from $b = r = 1$ is a mixture of Bernoulli processes. Let Θ be uniform on [0,1], and $(\xi_i)_{i\geq 1}$ an i.i.d. Bernoulli sequence with parameter Θ and values in $\{0,1\}$. Denote by $S_n = \xi_1 + \ldots + \xi_n$. Then,*

$$(1+S_n)_{n\geq 0} \overset{\text{law}}{=} (B_n)_{n\geq 0},$$

with B_n from Polya's urn.
For $n \geq 1$, we use the notation

$$s_0^n = (s_i)_{0\leq i\leq n}$$

for a finite sequence. We assume that s_0^n is an admissible sequence for S, i.e., $s_0 = 0, s_i - s_{i-1} \in \{0,1\}$. Once Θ is chosen, then $(\xi_i)_{i\geq 1}$ is a heads-and-tails game,

$$P(S_0^n = s_0^n|\Theta) = \Theta^{s_n}(1-\Theta)^{n-s_n},$$

abbreviating $S_0^n = (S_i)_{0\leq i\leq n}$. With a uniform r.v. Θ, we have

$$P(S_0^n = s_0^n, \Theta \in d\theta) = \theta^{s_n}(1-\theta)^{n-s_n}\mathbf{1}_{[0,1]}(\theta)d\theta,$$

which is the joint law of S_0^n and Θ. The marginal law of S_0^n is

$$\begin{aligned}
P(S_0^n = s_0^n) &= \int_0^1 \theta^{s_n}(1-\theta)^{n-s_n}d\theta \\
&= \frac{\Gamma(s_n+1)\Gamma(n-s_n+1)}{\Gamma(n+2)} \\
&= \frac{s_n!(n-s_n)!}{(n+1)!}, \qquad (4.4)
\end{aligned}$$

with Γ the Euler function, satisfying $\Gamma(n) = (n-1)!$ at integer points. From this we compute the conditional law, setting $s_{n+1} = s_n + 1$,

$$
\begin{aligned}
P(S_{n+1} = s_n + 1 | S_0^n = s_0^n) &= \frac{P(S_0^{n+1} = s_0^{n+1})}{P(S_0^n = s_0^n)} \\
&= \frac{s_n + 1}{n+2} \qquad \text{(by(4.4))}
\end{aligned}
$$

Since $R_n + B_n = n+2$, we see that the conditional law of $S_{n+1}+1$ given $\mathcal{F}_n^S$ is the same as the of of B_{n+1} given $\mathcal{F}_n^B$. Since $B_0 = S_0 + 1 = 1$, we obtain by recursion that the law of B_0^n is the same as the one of $1 + S_0^n$.

Note that, by the standard law of large numbers, we have

$$
\lim_{n\to\infty} \frac{S_n}{n} = \Theta,
$$

$P(\cdot|\Theta)$-a.s. for all Θ, and then P-a.s. Since $M_n = B_n/(n+2)$ has same law as $(1+S_n)/(n+2)$ – and even the processes $(M_n)_n$ and $((1+S_n)/(n+2); n \geq 0)$ has same law – we recover that M_n converges a.s. to a limit which is uniformly distributed in [0,1].

4.2 Square integrable martingales

4.2.1 Generalities

The martingale M is square integrable if $M_n \in L^2(\Omega, \mathcal{A}, P)$ for all n. By Riesz's theorem, this space is a Hilbert space with scalar product $X, Y \mapsto \langle X, Y\rangle = E(XY)$.

• The increments $M_{n+1} - M_n, n \geq 0$, are orthogonal: $\forall Y \in L^2(\mathcal{F}_n)$,

$$
\begin{aligned}
\langle M_{n+1} - M_n, Y\rangle &= E[Y(M_{n+1} - M_n)] \\
&= EE[Y(M_{n+1} - M_n)|\mathcal{F}_n] \\
&= E(Y \times E[(M_{n+1} - M_n)|\mathcal{F}_n]) \\
&= 0,
\end{aligned}
$$

by the martingale property. So the increment $M_{n+1} - M_n \perp\!\!\!\perp \mathcal{F}_n$.

• Hence the sum $M_n = M_0 + \sum_{k=1}^n (M_k - M_{k-1})$ is orthogonal, and

$$
EM_n^2 = EM_0^2 + \sum_{k=1}^n E[(M_k - M_{k-1})^2]
$$

by Pythagoras' theorem.

• $X_n = M_n^2$ is a submartingale. This follows from a general fact, which applies with $\Phi(x) = x^2$.

Proposition 4.2.1. *If (M_n) is a martingale and $\Phi : \mathbb{R} \to \mathbb{R}$ is convex, the sequence $(\Phi(M_n); n \geq 0)$ is a submartingale provided that it is integrable.*

□ By Jensen inequality for conditional expectations,

$$E[\Phi(M_{n+1})|\mathcal{F}_n] \geq \Phi(E[M_{n+1}|\mathcal{F}_n]) = \Phi(M_n)$$

■

• Doob's decomposition then writes

$$M_n^2 = M_0^2 + A_n + N_n,$$

with a martingale N_n and a predictable (increasing) process

$$\begin{aligned} A_n &= \sum_{k=1}^{n} E[(M_k - M_{k-1})^2|\mathcal{F}_{k-1}] \\ &= \sum_{k=1}^{n} E[M_k^2 - M_{k-1}^2|\mathcal{F}_{k-1}]. \end{aligned} \tag{4.5}$$

Definition 4.2.2. *The bracket of the square integrable martingale M is the increasing process $\langle M \rangle_n = A_n$ given in (4.5). It is the unique predictable process starting at 0 such that*

$$M_n^2 - \langle M \rangle_n$$

is a martingale. In particular,

$$E(M_n^2) = E(M_0^2) + E\langle M \rangle_n$$

We denote by $\langle M \rangle_\infty$ its a.s. limit,

$$\langle M \rangle_\infty = \lim_{n \nearrow \infty} (\nearrow) \langle M \rangle_n \in \mathbb{R}^+ \cup \{\infty\}.$$

• If M, N are square integrable martingales, then $M + N$ is also a square integrable martingale, and we can write

$$M_n N_n = \frac{1}{2}\big[(M_n + N_n)^2 - M_n^2 - N_n^2\big].$$

We therefore define the *cross-bracket* of the two martingales as

$$\langle M, N \rangle_n = \frac{1}{2}\big[\langle M + N \rangle_n - \langle M \rangle_n - \langle N \rangle_n\big].$$

It is a (in fact, the unique) predictable process starting at 0 such that

$$M_n N_n - \langle M, N \rangle_n$$

is a martingale. Alternatively it can be written as

$$\langle M, N \rangle_n = \sum_{k=1}^{n} E[(M_k - M_{k-1})(N_k - N_{k-1})|\mathcal{F}_{k-1}].$$

When $M = N$ we recover the previous notion, i.e. $\langle M, M \rangle_n = \langle M \rangle_n$. The mapping $(M, N) \mapsto \langle M, N \rangle$ is bilinear and symmetric, with the positivity property that $\langle M, M \rangle_n = \langle M \rangle_n \geq 0$.

4.2.2 Bounded martingales in L^2

First, observe that, for a martingale M,

$$\sup_n EM_n^2 = E\langle M \rangle_\infty = \sum_{n \geq 1} E(M_n - M_{n-1})^2. \tag{4.6}$$

□ Indeed, $EM_n^2 = E\langle M \rangle_n = \sum_{k=1}^{n} E(M_k - M_{k-1})^2$ since $M_n^2 - \langle M \rangle_n$ is a martingale and by orthogonality of the increments. Now, let $n \to \infty$ and use monotone convergence theorem. ■

The martingale M is bounded in L^2 if $\sup_n EM_n^2 < \infty$. Then, being integrable, $\langle M \rangle_\infty < \infty$ a.s. The space $\mathcal{M}^2(\Omega, \mathcal{F}, P)$,

$$\mathcal{M}^2 = \{M : M\ \mathcal{F} - \text{martingale}, \sup_n EM_n^2 < \infty, M_0 = 0\}$$

is a $\mathbb{R}$-vector space. For $M, N \in \mathcal{M}^2$, we denote

$$\langle M, N \rangle_\infty = \frac{1}{2}[\langle M + N \rangle_\infty - \langle M \rangle_\infty - \langle N \rangle_\infty].$$

Then the mapping

$$(M, N) \mapsto (M; N) := E\langle M, N \rangle_\infty$$

is bilinear symmetric positive definite, it is a scalar product on $\mathcal{M}^2$.

Proposition 4.2.3. *The space $\mathcal{M}^2$ is a Hilbert space with this scalar product.*

□ Let M^k be a Cauchy sequence: $E\langle M^k - M^l \rangle_\infty \to 0$ as $k, l \to \infty$. Since the right-hand member is equal to the following supremum,

$$\sup_n E(M_n^k - M_n^l)^2 \to 0, \qquad k, l \to \infty, \tag{4.7}$$

it means that for all n, the sequence $(M_n^k)_k$ is a Cauchy sequence in the Hilbert space L^2, and then

$$M_n^k \xrightarrow{L^2} M_n^\infty \qquad \text{as } k \to \infty.$$

In the relation $M_n^k = E[M_{n+1}^k|\mathcal{F}_n]$, we can take the limit as $k \to \infty$, and we see that $M^\infty = (M_n^\infty)_n$ is a martingale. Moreover,

$$E(M_n^\infty)^2 = \lim_{k\to\infty} E(M_n^k)^2 = \lim_{k\to\infty} E\langle M^k\rangle_n \le \lim_{k\to\infty} E\langle M^k\rangle_\infty \le \sup_{k\to\infty} E\langle M^k\rangle_\infty < \infty,$$

showing that $M^\infty \in \mathcal{M}^2$. Finally, letting $l \to \infty$ in (4.7), we see that

$$\sup_n E(M_n^k - M_n^\infty)^2 = E\langle M^k, M^\infty\rangle_\infty \to 0, \qquad k \to \infty,$$

showing that $M^k \to M^\infty$ in $\mathcal{M}^2$. ∎

The previous result was for a sequence of martingale. The next one is for a single martingale.

Theorem 4.2.4 (Convergence of martingale bounded in L^2)**.** *A martingale which is bounded in L^2, converges a.s. and in L^2.*

□ By Theorem 4.1.1, there exists a r.v. M_∞ such that $M_n \to M_\infty$ a.s. On the other hand, by orthogonality of increments,

$$E(M_{n+k} - M_n)^2 = \sum_{\ell=1}^{k} E(M_{n+\ell} - M_{n+\ell-1})^2 \le \sum_{\ell=1}^{\infty} E(M_{n+\ell} - M_{n+\ell-1})^2.$$

Sending $k \to \infty$, we get by Fatou,

$$E(M_\infty - M_n)^2 \le \lim_{k\to\infty} E(M_{n+k} - M_n)^2 \le \sum_{\ell=1}^{\infty} E(M_{n+\ell} - M_{n+\ell-1})^2,$$

which tends to 0 as $n \to \infty$ by (4.6). Therefore, $M_n \to M_\infty$ in L^2. ∎

4.2.3 Applications

Proposition 4.2.5 (Random series)**.** *Let $(X_i)_i$ be independent, square-integrable centered random variables, $S_n = \sum_{i=1}^n X_i$ and let $\sigma_k^2 = \mathbb{V}ar X_k$. Then,*

1. *$\sum_k \sigma_k^2 < \infty \implies S_n$ converge a.s. and in L^2.*
2. *Assume $|X_n| \le K$ a.s.. Then, S_n converge a.s. $\implies \sum_k \sigma_k^2 < \infty$.*

Here are two comments:
(i) According to Kolmogorov 0-1 law, the event $\{S_n \text{ converge}\}$ has probability 0 or 1.
(ii) It may happen that S_n converge a.s. and $\sum_k \sigma_k^2 = \infty$. For instance, when $\sum_n \|X_n\|_1 < \infty$ the martingale, being bounded in L^1, converges a.s., and this is compatible with the divergence of the sum of variances. To be specific, for $n \geq 2$ take $X_n \in \{-n, 0, n\}$ with

$$P(X_n = \pm n) = n^{-5/2}, \quad P(X_n = 0) = 1 - 2n^{-5/2},$$

for $n \geq 2$.

$\Box$ of the proposition.

1. As a sum of independent, zero mean, square integrable r.v.'s, S_n is a L^2-martingale, which bracket is the deterministic sequence

$$\langle S \rangle_n = \sum_{k=1}^{n} \sigma_k^2.$$

Indeed,

$$\begin{aligned} \langle S \rangle_n - \langle S \rangle_{n-1} &= E[(S_n - S_{n-1})^2 | \mathcal{F}_{n-1}] \\ &= E[X_n^2 | \mathcal{F}_{n-1}] \\ &= E[X_n^2] \\ &= \sigma_n^2. \end{aligned}$$

So,

$\sum_k \sigma_k^2 < \infty \implies S_n$ bounded in $L^2 \implies S_n$ converge a.s. and in L^2,

by Theorem 4.2.4.

2. $N_n = S_n^2 - \sum_{k=1}^n \sigma_k^2$ is a martingale, and for each $C > 0$,

$$T = T(C) = \inf\{k : |S_k| > C\}$$

is a stopping time. By the stopping theorem, $EN_{n\wedge T} = EN_0 = 0$, which implies that

$$E \sum_{k=1}^{n\wedge T} \sigma_k^2 = ES_{n\wedge T}^2 \leq (C + K)^2,$$

since $|S_{n\wedge T-1}| \leq C, |X_{n\wedge T-1}| \leq K$. If S_n converges a.s., then it is bounded a.s.,

$$\begin{aligned} 1 &= P(\cup_{C\in\mathbb{N}}\{\sup_n |S_n| \leq C\}) \\ &= \lim_{C\nearrow\infty} P\{\sup_n |S_n| \leq C\} \\ &= \lim_{C\nearrow\infty} P(T(C) = \infty). \end{aligned}$$

Fix C such that $P(T = \infty) > 0$. Then,

$$(C + K)^2 \geq E[\sum_{k=1}^{n \wedge T} \sigma_k^2 \mathbf{1}_{\{T=\infty\}}] \geq (\sum_{k=1}^{n} \sigma_k^2) \times P(T = \infty)$$

Hence, $\sum_{k=1}^n \sigma_k^2$ is bounded independently of n. ∎

Proposition 4.2.6 (A.s. convergence of a martingale on the event where its bracket is bounded). *Let $M = (M_n)_n$ be a L^2 martingale. Then,*

$$\langle M \rangle_\infty < \infty \text{ a.s.} \implies M \text{ converges a. s.}$$

The result extends Theorem 4.2.4. We start with a lemma. Recall the notation X^T for the process $X = (X_n)_n$ stopped at T, defined by $X_n^T = X_{n \wedge T}$.

Lemma 4.2.7. *Let M be a L^2-martingale and T a stopping time. Then*

$$\langle M^T \rangle = \langle M \rangle^T$$

□ The martingale M^T is square integrable. We prove that its bracket is $\langle M \rangle_n^T = \langle M \rangle_{T \wedge n}$. The latter process is predictable, since for any Borel subset B of $\mathbb{R}$,

$$\{\langle M \rangle_{T\wedge n} \in B\} = \cup_{i=0}^{n-1}\{T = i, \langle M \rangle_i \in B\} \bigcup \Big[\{T \leq n-1\}^c \bigcap \{\langle M \rangle_n \in B\}\Big]$$

is in $\mathcal{F}_{n-1}$. Moreover,

$$(M_n^T)^2 - \langle M \rangle_n^T = \Big(M^2 - \langle M \rangle\Big)_n^T,$$

and the right-hand side is a (stopped) martingale. By uniqueness of Doob's decomposition, we have the desired conclusion. ∎

□ Proof of the proposition. For $k > 0$, consider the random time

$$S(k) = \inf\{n \geq 0 : \langle M \rangle_{n+1} > k\}.$$

Since $\langle M \rangle$ is predictable, $S(k)$ is a $\mathcal{F}$-stopping time. Consider the martingale $M^{S(k)}$. By the lemma, its bracket is $\langle M \rangle^{S(k)}$, which is uniformly bounded by k by definition. Then, Theorem 4.2.4 applies, showing that this new martingale converges a.s.. Hence, M_n converges a.s. on the event $\{S(k) = \infty\}$. Since $\{\langle M \rangle_\infty < \infty\} = \cup_k \{S(k) = \infty\}$ has probability 1, we see that M_n converges a.s. ∎

4.3 Complement: Maximal inequalities

The term of "maximal inequality" covers estimates for the maximum $\max\{X_k; k \leq n\}$ or $\max\{|X_k|; k \leq n\}$.

Theorem 4.3.1 (Doob's maximal inequality). *Let $(Z_n)_n$ a non-negative submartingale. Then, for all $\lambda > 0$,*

$$P(\max_{k\leq n} Z_k \geq \lambda) \leq \frac{EZ_n}{\lambda}$$

This is "super-Markov" inequality. The whole path is controled by the terminal time.

□ With $A_0 = \{Z_0 \geq \lambda\}$, $A_k = \{Z_i < \lambda, 0 \leq i \leq k-1, Z_k \geq \lambda\}$, we can write

$$A := \{\max_{k\leq n} Z_k \geq \lambda\} = A_0 \cup A_1 \cup \ldots \cup A_k$$

as a disjoint union. Moreover,

$$E(Z_n; A_k) \geq E(Z_k; A_k) \geq \lambda P(A_k),$$

since Z is a submartingale and $A_k \in \mathcal{F}_k$, and by Markov. By summing over k, we get, since $Z_n \geq 0$,

$$E(Z_n) \geq E(Z_n; A) = \sum_{k=0}^{n} E(Z_n; A_k) \geq \lambda \sum_{k=0}^{n} P(A_k),$$

implying the result. ■

Recall Bienaymé-Čebyšev inequality[2]: for a centered L^2-r.v. M, $P(|M| > \lambda) \leq EM^2/\lambda^2$.

Corollary 4.3.2 (Super-Bienaymé-Čebyšev). *If M is a square integrable martingale,*

$$\forall\lambda > 0, \qquad P(\max_{k\leq n} |M_k| > \lambda) \leq \frac{EM_n^2}{\lambda^2}$$

□ Apply the previous to $Z_n = M_n^2$. ■

Theorem 4.3.3 (Doob's L^P maximal inequality). *Let $(Z_n)_n$ a non-negative submartingale. Then, letting $Z_n^* = \max_{k\leq n} Z_k$, we have for all $p > 1$,*

$$\|Z_n^*\|_p \leq q\|Z_n\|_p, \qquad q = \frac{p}{p-1}.$$

□ To be written [??] ■

[2] Alternative spellings are Chebyshev, or Tchebicheff, or Чебышёв, ...

4.4 Uniform integrability

4.4.1 Uniform integrable random variables

Recall the standard notation, for an integrable r.v. X and an event A,

$$E[X; A] = E(X\mathbf{1}_A)$$

We start with an introductory lemma: the measure $dQ = |X|dP$ is absolutely continuous to P on $\mathcal{A}$.

Lemma 4.4.1 (Absolute continuity)**.** *Let $X \in L^1$. Then, $\forall \varepsilon > 0, \exists \delta > 0$ such that*

$$F \in \mathcal{A}, P(F) < \delta \implies E[|X|; F] < \varepsilon.$$

□ Argue by contradiction. Assume there is some $\varepsilon > 0$ and a sequence $(F_n)_n$ of events with

$$P(F_n) < 2^{-n}, \qquad E[|X|; F_n] > \varepsilon.$$

Let H be the event when the F_n's occur infinitely often,

$$H = \limsup_n F_n = \{\sum_n \mathbf{1}_{F_n} = \infty\}.$$

By Borel-Cantelli lemma, $P(H) = 0$. Apply Fatou's lemma to the sequence $|X|\mathbf{1}_{F_n^c}$:

$$\liminf_n E[|X|; F_n^c] \geq E[|X|; \liminf_n F_n^c] = E[|X|; H^c] = E|X|,$$

since H is negligible. Hence, we get $\limsup_n E[|X|; F_n] = 0$, a contradiction. ∎

Corollary 4.4.2. *Let $X \in L^1$ and $\varepsilon > 0$. Then $\exists K$ finite such that*

$$E[|X|; |X| > K] < \varepsilon$$

□ Recall Markov inequality, $P(|X| \geq K) \leq K^{-1}E|X|$. Choose $\delta(\varepsilon)$ as in Lemma 4.4.1 and $K > \delta^{-1}E|X|$. Then, the lemma yields the result. ∎

Exercise 4.4.3. *Find a direct proof of this statement using dominated convergence.*

We come to the point.

Definition 4.4.4. *A family $\mathcal{C}$ of integrable r.v.'s is uniformly integrable if*

$$\forall \varepsilon > 0, \exists K > \infty, \quad \sup_{X \in \mathcal{C}} E[|X|; |X| > K] < \varepsilon. \tag{4.8}$$

Note that:
(i) $\mathcal{C}$ uniformly integrable $\implies$ $\mathcal{C}'$ uniformly integrable for all $\mathcal{C}' \subset \mathcal{C}$
(ii) $\mathcal{C}$ uniformly integrable $\implies$ $\mathcal{C}$ bounded in L^1.

$\square$ of (ii): Choose $K = K(1)$ from the definition with $\varepsilon = 1$, and split for all $X \in \mathcal{C}$,

$$E|X| = E[|X|; |X| \leq K] + E[|X|; |X| > K] \leq K + 1,$$

which ends the proof. $\blacksquare$

Exercise 4.4.5. *Let $\mathcal{C}$ be uniformly integrable. For $X \in \mathcal{C}$ nonzero, define the probability measure μ_X on $\mathcal{B}(\mathbb{R})$ by*

$$d\mu_X(x) = (E|X|)^{-1}|x|dP_X(x).$$

Show that, for all $a > 0$,

$$\mathcal{C} \text{ uniformly integrable } \implies \{\mu_X; X \in \mathcal{C}, E|X| \geq a\} \text{ tight}$$

$\square$

$$\mu_X([-A, A]^c) = \frac{E[|X|; |X| > A]}{E|X|},$$

which vanishes (uniformly over $X \in \mathcal{C}$ with $E|X| \geq a$) as $A \to \infty$. $\blacksquare$

Counterexample 4.4.6. $(\Omega, \mathcal{A}, P)$ = *([0,1], Borel, Lebesgue), and* $\mathcal{C} = \{X_n, n \geq 1\}, X_n = n\mathbf{1}_{[0,1/n]}$. *Then,* $\|X_n\|_1 = 1, E[|X_n|; |X_n| > K] = EX_n = 1$ *when* $n \geq K$. *Hence,* $\mathcal{C}$ *is not uniformly integrable .*

The examples below are important:

Example 4.4.7. *(i) Bounded family in $L^p(p > 1)$: For $p > 1$,*

$$\sup_{X \in \mathcal{C}} \|X\|_p < \infty \ (p > 1) \implies \mathcal{C} \text{ uniformly integrable}$$

(ii) Dominated family: if $Y \in L^1_+$,

$$\mathcal{C} = \{X : |X| \leq Y\} \text{ is uniformly integrable}$$

(iii) Conditional expectations of an integrable r.v.: if $X \in L^1$,

$$\mathcal{C} = \{E(X|\mathcal{B}) : \mathcal{B} \subset \mathcal{A}\} \text{ is uniformly integrable}$$

where $\mathcal{B}$ ranges over the sub sigma-fields of $\mathcal{A}$.

The third example is fundamental.

□ (i) By Markov inequality, with $C = \sup_{X\in\mathcal{C}} \|X\|_p$,

$$E[|X|; |X| > K] \le \frac{E|X|^p}{K^{p-1}} \le \frac{C^p}{K^{p-1}},$$

which can be made less than ε taking K large.
(ii) Since $y \mapsto y\mathbf{1}_{y>K}$ is increasing and since $|X| \le Y$, bound

$$E[|X|; |X| > K] \le E[Y; Y > K],$$

and we conclude by applying corollary 4.4.2.
(iii) Note that $A_K = \{|E(X|\mathcal{B})| > K\} \in \mathcal{B}$, and write

$$\begin{aligned} E[|E(X|\mathcal{B})|; A_K] &\le E[E(|X||\mathcal{B}); A_K] \\ &= E[E(|X|\mathbf{1}_{A_K}|\mathcal{B})] && (A_K \in \mathcal{B}) \\ &= E[|X|; A_K] && \text{(tower prop.)} \end{aligned}$$

Let $\varepsilon > 0$. By lemma 4.4.1, we can find δ such that $E[|X|; A] < \varepsilon$ as soon as $P(A) < \delta$. We finish the proof by showing that, taking K large, we can make $P(A_K) < \delta$:

$$\begin{aligned} P(|E(X|\mathcal{B})| > K) &\le K^{-1}E|E(X|\mathcal{B})| && \text{(Markov ineq.)} \\ &\le K^{-1}EE(|X||\mathcal{B}) = K^{-1}E|X|, \end{aligned}$$

which indeed vanishes as $K \to \infty$ uniformly in $\mathcal{B}$. ■

4.4.2 Uniformly integrable sequences

We start with a warm up lemma: for dominated convergence, convergence in probability is enough.

Lemma 4.4.8 (Another Dominated Convergence theorem)**.**

$$\left.\begin{array}{c} X_n \xrightarrow{P} X \\ |X_n| \le Y \in L^1 \end{array}\right\} \implies X_n \xrightarrow{L^1} X$$

□ First, we prove that $X \in L^1$: there exists a subsequence $X_{n_j} \to X$ a.s., so $|X| \le Y$. For $\varepsilon > 0$, by Lemma 4.4.1 we choose $\delta > 0$ such that

$$P(F) < \delta \implies E[(Y + |X|); F] \le \varepsilon$$

Now, we split

$$\begin{aligned} E|X_n - X| &= E[|X_n - X|; |X_n - X| \le \varepsilon] + E[|X_n - X|; |X_n - X| > \varepsilon] \\ &\le \varepsilon + E[(Y + |X|); |X_n - X| > \varepsilon] \\ &\le 2\varepsilon \end{aligned}$$

as soon as n is large enough so that $P(|X_n - X| > \varepsilon) < \delta$. ■

Theorem 4.4.9 (Yet Another Dominated Convergence theorem)**.** *As $n \to \infty$,*

$$X_n \xrightarrow{L^1} X \iff \left\{ \begin{array}{c} X_n \xrightarrow{P} X \\ (X_n)_n \text{ uniformly integrable} \end{array} \right.$$

UI is a condition of domination !

□ Proof of $\Leftarrow$: Let f_K be a truncation function,

$$f_K(x) = x\mathbf{1}_{|x|\le K} + \operatorname{sign}(x)K\mathbf{1}_{|x|>K}.$$

This is a 1-Lipschitz function, andwe have for all n,

$$\begin{array}{rcl} E|f_K(X_n) - X_n| & = & E[|X_n| - K; |X_n| > K] \\ & \le & E[|X_n|; |X_n| > K] \\ & \le & \varepsilon, \end{array}$$

taking K larger than some K_1 by uniformly integrability. Similarly,

$$E|f_K(X) - X| \le \varepsilon$$

for $K \ge K_2$. Fix $K = K_1 \vee K_2$, note that, by continuity of f_K, as $n \to \infty$,

$$f_K(X_n) \xrightarrow{P} f_K(X), \qquad |f_K(X_n)| \le K,$$

which implies by Lemma 4.4.8

$$f_K(X_n) \xrightarrow{L^1} f_K(X).$$

Hence, $\|f_K(X_n) - f_K(X)\|_1 \le \varepsilon$ for all $n > n_0(K, \varepsilon)$. Then, we have

$$\begin{array}{rcl} E|X_n - X| & \le & E|X_n - f_K(X_n)| + E|f_K(X_n) - f_K(X)| + E|f_K(X) - X| \\ & \le & 3\varepsilon, \end{array}$$

for all such n.

Proof of $\Rightarrow$: Since convergence in L^1-norm implies convergence in probability, we just need to show uniformly integrability. Let $\varepsilon > 0$, and n_0 such that

$$n \ge n_0 \implies \|X_n - X\|_1 < \varepsilon/2 \tag{4.9}$$

By Lemma 4.4.1 we fix $\delta > 0$ such that

$$P(F) < \delta \implies \left\{ \begin{array}{l} E[|X|; F] \le \varepsilon \\ E[|X_n|; F] \le \varepsilon \ (n = 1, \dots, n_0) \end{array} \right. \tag{4.10}$$

For this δ, we fix K such that $K^{-1} \sup E|X_n| < \delta$. Then, by Markov inequality,

$$\forall n, \quad P(|X_n| > K) < \delta \tag{4.11}$$

On the one hand, for $n > n_0$, we have

$$\begin{aligned} E[|X_n|; |X_n| > K] &\leq E[|X|; |X_n| > K] + E|X_n - X| \\ &\leq 3\varepsilon/2 \end{aligned}$$

by (4.9) and (4.10), and on the other hand, for $n \leq n_0$, we have $E[|X_n|; |X_n| > K] \leq \varepsilon$ by (4.11) and (4.10). Then, the sequence is uniformly integrable. ∎

Corollary 4.4.10. .

$$\left. \begin{array}{l} X_n \stackrel{\text{law}}{\longrightarrow} X \\ (X_n)_n \text{ uniformly integrable} \end{array} \right\} \Longrightarrow \lim_{n\to\infty} EX_n = EX$$

□ By Skorohod embedding, we can find a probability space and a sequence $(\tilde{X}_n, n \geq 1)$ such that $\tilde{X}_n$ has the same law as X_n for all n, and $\tilde{X}_n \to \tilde{X}$ a.s. where the limit has the same law as X. Then, the previous theorem applies to new sequence $(\tilde{X}_n, n \geq 1)$ and its limit $\tilde{X}$, yielding

$$EX = E\tilde{X} = \lim_{n\to\infty} E\tilde{X}_n = \lim_{n\to\infty} EX_n$$

by equality in law. ∎

4.4.3 Uniformly integrable martingales

The martingale M is called uniformly integrable if the sequence $(M_n; n \geq 1)$ is uniformly integrable.

Since it is bounded in L^1, it converges a.s., and by Theorem 4.4.9, it converges also in L^1. The important point is

$$M_n = E[M_\infty|\mathcal{F}_n].$$

Indeed, we have $M_n = E[M_{n+p}|\mathcal{F}_n]$, and then,

$$\begin{aligned} M_n &= \lim_{p\to\infty} E[M_{n+p}|\mathcal{F}_n] \\ &= E[\lim_{p\to\infty} M_{n+p}|\mathcal{F}_n] \\ &= E[M_\infty|\mathcal{F}_n], \end{aligned}$$

by continuity of $Z \mapsto E[Z|\mathcal{F}_n]$ in L^1. We summarize:

Theorem 4.4.11. *If M is a uniformly integrable martingale, we have*

$$M_n \to M_\infty \text{ a.s. and in } L^1$$

to a limit such that

$$M_n = E[M_\infty|\mathcal{F}_n]. \tag{4.12}$$

Such a martingale is of the type (3.4), it is called "closed", because the sequence $(M_n; 0 \le n \le \infty)$ (including infinity; this is the natural closure) is a $(\mathcal{F}_n; n \le \infty)$-martingale.

For closed martingales, the optional stopping theorem takes its simplest form. First of all, M_T is well defined for all stopping times T, since it is equal to M_∞ when $T = \infty$. We have:

Proposition 4.4.12 (Stopping theorem for UI martingales)**.** *Let M be a uniformly integrable martingale, and $S \le T$ two stopping times. Then,*

$$M_S = E[M_T|\mathcal{F}_S].$$

□ Let M_∞ given by the previous theorem. We start to show that

$$E[M_\infty|\mathcal{F}_T] = M_T.$$

We simply write, on the event $\{T = n\}$ (where $n \le \infty$),

$$E[M_\infty|\mathcal{F}_T] = E[M_\infty|\mathcal{F}_n] = M_n,$$

which is exactly the claim.[(3)] This shows that $M_T \in L^1$. Then, we write

$$E[M_T|\mathcal{F}_S] = E\big[E[M_\infty|\mathcal{F}_T]|\mathcal{F}_S\big] = E[M_\infty|\mathcal{F}_S] = M_S$$

by the tower property. ■

With the same arguments as above, we get the following generalization:

Proposition 4.4.13. *An uniformly integrable submartingale X converges a.s. and in L^1-norm to a limit X_∞ satisfying*

$$X_n \le E[X_\infty|\mathcal{F}_n] \quad a.s.$$

Theorem 4.4.14 (Lévy's upward convergence theorem)**.** *Let $\xi \in L^1(\Omega, \mathcal{A}, P)$, $(\mathcal{F}_n)$ a filtration, and $M_n = E[\xi|\mathcal{F}_n]$. Then, M is uniformly integrable, and*

$$M_n \longrightarrow E[\xi|\mathcal{F}_\infty] \quad a.s. \text{ and in } L^1.$$

[(3)] In fact, the first equality is not completely obvious from the definition, and we leave the argument to the reader.

□ By Example 4.4.7, 3), we know that the martingale M is uniformly integrable, so Theorem 4.4.11 applies and yields a limit M_∞. What is left to prove is that $M_\infty = E[\xi|\mathcal{F}_\infty]$, which amounts, by Definition 2.2.1, to prove that

$$E[M_\infty - \xi; A] = 0, \qquad \forall A \in \mathcal{F}_\infty. \tag{4.13}$$

Since $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geq 1)$, the monotone class Theorem shows it is enough to prove (4.13) for all n and all $A \in \mathcal{F}_n$. In this case, we have

$$\begin{array}{rcll} E[M_\infty - \xi; A] & = & E[M_\infty; A] - E[\xi; A] & \\ & = & E\Big[E[M_\infty|\mathcal{F}_n]; A\Big] - E[\xi; A] & (A \in \mathcal{F}_n) \\ & = & E[M_n; A] - E[\xi; A] & (\text{by } (4.12)) \\ & = & E[M_n - \xi; A] & \\ & = & 0 & \end{array}$$

by definition of the conditional expectation. ∎

Example 4.4.15 (Gradually discover a function by its dyadic means)**.** *Let $(\Omega, \mathcal{A}, P)$ = ([0,1], Borel, Lebesgue), and $f \in L^1([0,1])$. With $\mathcal{F}_n = \sigma([k2^{-n}, (k+1)2^{-n}[; k = 0, \dots, 2^n - 1)$ the dyadic partition of nth generation, we consider*

$$\begin{array}{rcl} f_n = E[f|\mathcal{F}_n] : \ \Omega & \to & \mathbb{R} \\ \omega & \mapsto & 2^n \displaystyle\int_{k2^{-n}}^{(k+1)2^{-n}} f(s)ds \quad \text{with } k = \lfloor 2^n \omega \rfloor \end{array}$$

Theorem 4.4.14 shows that $f_n \to f_\infty$ a.s. and in L^1, with $f_\infty = E[f|\mathcal{F}_\infty] = f$ for the dyadic intervals generate the Borel σ-algebra. In particular, putting $I_n(s) = [k2^{-n}, (k+1)2^{-n}[, k = \lfloor 2^n \omega \rfloor$, the set

$$\left\{ s \in [0,1] : \lim_{n\to\infty} \frac{1}{|I_n(s)|} \int_{I_n(s)} f(u)du = f(s) \right\}$$

has full Lebesgue measure. This statement is similar to Lebesgue's differentiation theorem (1904), which states that for $s \mapsto \int_0^s f(u)du$ is almost everywhere differentiable, with $F' = f$. Actually, the multidimensional version (1910) is even closer: for $g \in L^1(\mathbb{R}^d), \lim_{r\to 0} |B(x,r)|^{-1} \int_{B(x,r)} g(y)dy = g(x)$ Lebesgue-almost everywhere. ($B(x,r)$ is the Euclidean ball.)

We end with a quick, deep and beautiful proof of the 0-1 law of Kolmogorov.

Corollary 4.4.16 (Kolmogorov's 0-1 law)**.** *Let $(\mathcal{A}_n, n \geq 1)$ be a sequence of independent sub-σ-fields of $\mathcal{A}$, $\mathcal{T}_n = \sigma(\mathcal{A}_i, i \geq n)$. Then, the tail field, $\mathcal{T} = \bigcap_{n\geq 1} \mathcal{T}_n$ is trivial:*

$$P(A) \in \{0,1\} \text{ for all } A \in \mathcal{T}.$$

$\square$ Let $A \in \mathcal{T}$ and $\xi = \mathbf{1}_A$. Set $\mathcal{F}_n = \sigma(\mathcal{A}_i; i \le n)$ for $n \le \infty$. Then, $\mathcal{T} \subset \mathcal{F}_\infty$. By Lévy's upward convergence,

$$E[\xi|\mathcal{F}_n] \to E[\xi|\mathcal{F}_\infty] = \xi \quad \text{a.s. and in } L^1,$$

with the last equality since ξ is measurable w.r.t. $\mathcal{T} \subset \mathcal{F}_\infty$. Since $\mathcal{T} \perp\!\!\!\perp \mathcal{F}_n$ for all $n < \infty$, and since $\cup_n \mathcal{F}_n$ is stable by finite intersection and generates $\mathcal{F}_\infty$, we have $\mathcal{T} \perp\!\!\!\perp \mathcal{F}_\infty$, and therefore $E[\xi|\mathcal{F}_\infty] = E\xi = P(A)$. So,

$$P(A) = \xi \in \{0,1\},$$

because ξ is an indicator function. We conclude that the number $P(A)$ is equal to 0 or 1, and also that the r.v. $\xi = \mathbf{1}_A$ is a.s. constant, which is totally equivalent a statement. ∎

4.5 Complement: Central limit theorem for martingales

In the following central limit theorem for square-integrable martingales, we extend the case of sum of independent r.v.'s.

Theorem 4.5.1. *Let $M_n = \sum_{i=1}^n X_i \in L^2(\Omega, \mathcal{A}, \mathbb{P}), \mathcal{F}_n = \sigma(X_i, i \le n)$ and $s_n^2 = \sum_{i=1}^n E(X_i^2)$. We assume that*

(i) $E(X_n|\mathcal{F}_{n-1}) = 0$,

(ii) $s_n \to \infty$, *and* $s_n^{-2} \sum_{i=1}^n E(X_i^2|\mathcal{F}_{i-1}) \xrightarrow{P} 1 \quad \text{as } n \to \infty$,

(iii) $\forall \varepsilon > 0, s_n^{-2} \sum_{i=1}^n E(X_i^2 \mathbf{1}_{\{|X_i| > \varepsilon s_n\}}|\mathcal{F}_{i-1}) \xrightarrow{P} 0 \quad \text{as } n \to \infty.$

Then

$$\frac{M_n}{s_n} \xrightarrow{\text{law}} Z$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0,1)$.

$\square$ The proof consists in controling the characteristic function of M_k/s_n step by step ($k = 1, \ldots, n$). To simplify, we give the proof in the case[4] of bounded increments,

$$|X_i| \le C < \infty. \tag{4.14}$$

[4]For a proof in the general case, see e.g. [corollaire 2.8.43 dans: D. Dacunha-Castelle, M. Duflo, "Probabilités et Statistiques", tome 2 de cours, Masson 1983].

Step 1: Fix n and let $\xi_k = X_k/s_n$. Then, for $u \in \mathbb{R}$, as $n \to \infty$,

$$Z_n := \prod_{k=1}^{n} E\left(e^{iu\xi_k}|\mathcal{F}_{k-1}\right) \longrightarrow e^{-u^2/2} \quad \text{in } L^1 - \text{norm}. \tag{4.15}$$

Because Z_n is bounded (by 1), it is enough, from Lemma 4.4.8, to prove convergence in probability. Define

$$\phi(u,x) = e^{iux} - 1 - iux + \frac{1}{2}u^2x^2 = \frac{-i}{2}\int_0^{ux}(ux-s)^2 e^{is}ds$$

where the last member is from Taylor integral formula, and note that

$$|\phi(u,x)| \leq \frac{|ux|^3}{6}. \tag{4.16}$$

Letting also $v_{k-1} = v_{k-1}^{(n)} = E(\xi_k^2|\mathcal{F}_{k-1})$, we have by assumption

$$\sum_{k=1}^{n} v_{k-1} \xrightarrow{P} 1 \tag{4.17}$$

as $n \to \infty$. Write

$$\begin{aligned} Z_n &= \prod_{k=1}^{n}\left[1 + 0 - \frac{1}{2}u^2 v_{k-1} + E(\phi(u,\xi_k)|\mathcal{F}_{k-1})\right] \\ &= \prod_{k=1}^{n}\left[1 - \frac{1}{2}u^2 v_{k-1} + O(v_{k-1}/s_n)\right] \end{aligned} \tag{4.18}$$

using (4.14), (4.16). The error term $O(v_{k-1}/s_n)$ is uniform, so all factors are close to 1 in the complex plane by choosing n large enough. Hence we can use the complex logarithm, for which $\log(1+v) = v + O(v^2)$, and derive

$$\begin{aligned} \log Z_n &= -\frac{u^2}{2}\left(\sum_{k\leq n} v_{k-1}\right) + O\left(\sum_{k\leq n}\frac{v_{k-1}}{s_n}\right) + O\left(\sum_{k\leq n} v_{k-1}^2\right) \\ &\xrightarrow{\mathbb{P}} -u^2/2, \end{aligned}$$

by (4.17), and since both

$$\sum_{k\leq n} v_{k-1}/s_n = s_n \times \sum_{k\leq n} v_{k-1}/s_n^2 \quad \text{and} \quad \sum_{k\leq n} v_{k-1}^2 \xrightarrow{P} 0.$$

The last limit follows from $\sum_{k\leq n} v_{k-1}^2 \leq C^2(\sum_{k\leq n} v_{k-1})/s_n^2$. This completes the first step.

Step 2: For fixed n and $k = 0, \dots, n$,

$$N_k = \left[\prod_{\ell=1}^{k} E\left(e^{iu\xi_\ell} | \mathcal{F}_{\ell-1} \right) \right]^{-1} e^{iuM_k/s_n}$$

is a martingale. By definition, a complex martingale has real and imaginary parts which are real martingales. Let $a \in (0, e^{-u^2/2})$, and

$$T = T_a^{(n)} = \inf \left\{ k : 1 \le k \le n, | \prod_{\ell=1}^{k} E\left(e^{iu\xi_\ell} | \mathcal{F}_{\ell-1} \right) | \le a \right\} - 1.$$

Then, T is a stopping time, even with the "-1" at the end of the definition (!), which makes the modulus of the denominator of N_k *larger* than a at times $k \le T$. The sequence $| \prod_{\ell=1}^{k} E\left(e^{iu\xi_\ell} | \mathcal{F}_{\ell-1} \right) |$ being non-increasing in k, we see that

$$\begin{aligned} P(T < n) &= P(|Z_n| \le a) \\ &\longrightarrow 0 \text{ as } n \to \infty, \end{aligned} \tag{4.19}$$

by step 1 and since $a < e^{-u^2/2}$. By the optional stopping theorem,

$$EN_{n \wedge T} = EN_0 = 1. \tag{4.20}$$

Observing that

$$|e^{iuM_n/s_n} - N_{n \wedge T} Z_n| \le 2 \times \mathbf{1}_{\{T < n\}},$$

we see from (4.19) that, in order to prove the convergence of the characteristic function $\Phi_{M_n/s_n}(u)$ to $\Phi_Z(u)$, it suffices to show that

$$E(N_{n \wedge T} Z_n) \to \exp\{-u^2/2\}$$

for all fixed u. Write

$$\begin{aligned} |E(N_{n \wedge T} Z_n) - \exp\{-u^2/2\}| &= |E(N_{n \wedge T}(Z_n - e^{-u^2/2}))| && \text{(by(4.20))} \\ &\le E(|N_{n \wedge T}| \times |Z_n - e^{-u^2/2}|) && \text{(def. of } T) \\ &\le a^{-1} E(|Z_n - e^{-u^2/2}|) \\ &\overset{n \to \infty}{\longrightarrow} 0, \end{aligned}$$

by Step 1. This ends the proof under the extra assumption (4.14). ∎

♠

Part III

Discrete Markov chains

Chapter 5

Markov Chains

We model a system evolving randomly in discrete time, it takes the states X_t at time $t = 0, 1, \ldots$. A natural and simple dependence is given by the following *Markov property*:

> **(M)** The future depends on the history only through the present.

Then $(X_t; t \in \mathbb{N})$ is called a Markov chain. For simplicity we focus on discrete state space.

5.1 Definition, properties

Let $(\Omega, \mathcal{A}, P)$ a probability space, E a finite or infinite countable space.

Definition 5.1.1. *A sequence $X = (X_t)_{t\geq 0}$ of r.v. on Ω with values in E is a Markov chain if*

$$P(X_{t+1} = y | X_0 = x_0, X_1 = x_1, \ldots X_t = x_t) = P(X_{t+1} = y | X_t = x_t) \quad (5.1)$$

for all $t \in \mathbb{N}$ and $x_0, x_1, \ldots x_t, y \in E$ with $P(X_i = x_i, 1 \leq i \leq t) > 0$.

This is property (M).

Proposition 5.1.2 (Random recursions)**.** *Let $(F, \mathcal{F})$ be a measurable space and $f : E \times F \times \mathbb{N} \to E$ measurable, $\xi = (\xi_t)_{t\geq 1}$ an i.i.d. sequence from Ω to F, and $X_0 \in E$ a r.v. on Ω, independent of the sequence ξ.Then, the r.v.'s $X_t : \Omega \to E$ defined recursively from X_0 by*

$$X_{t+1} = f(X_t, \xi_{t+1}, t), \qquad t \geq 0 \quad (5.2)$$

is a Markov chain, and the law of X_{t+1} given $X_t = x_t$ is the law of $f(x_t, \xi_{t+1}, t)$.

$\square$ Note that X_t is $\mathcal{F}_t$-measurable with $\mathcal{F}_t = \sigma(X_0, \xi_1, \dots \xi_t)$, so $\xi_{t+1} \perp\!\!\!\perp \mathcal{F}_t$ and then $\xi_{t+1} \perp\!\!\!\perp \mathcal{F}_t^X$ with $\mathcal{F}_t^X = \sigma(X_0, X_1, \dots X_t)$. Hence, on the event $X_t = x$, we have

$$\begin{aligned} P(X_{t+1} = y | \mathcal{F}_t^X) &= P(f(x, \xi_{t+1}, t) = y | \mathcal{F}_t^X) \\ &= P(f(x, \xi_{t+1}, t) = y) \end{aligned}$$

by independence. Then, X is Markov. Multiplying by $P(X_0 = x_0, \dots X_{t-1} = x_{t-1}, X_t = x)$ and summing over $x_0, \dots x_{t-1}$, we get

$$P(X_{t+1} = y | X_t = x) = P(f(x, \xi_{t+1}, t) = y)$$

All this proves both claims. ∎

Remark 5.1.3. *(i) Markov chains on infinite uncountable sets, like $E = \mathbb{R}$, are defined by the natural extension of definition 5.1.1. We stick to discrete state space to keep things simple and arguments transparent.*
(ii) It is not difficult to see that, if $P(X_{t+1} = y | X_0 = x_0, X_1 = x_1, \dots X_t = x_t)$ depends only on t, x_t, y - but not on $x_0, \dots, x_{t-1}$ - , then it is equal to $P(X_{t+1} = y | X_t = x_t)$ and the equality (5.1) holds. In this case X is a Markov chain.

5.2 Transition matrix

In most applications, the conditional laws $P(X_{t+1} = y | X_t = x)$ do not depend on t (time-homogeneous chain). For the recursive random sequence in Proposition 5.1.2 it means that f does not depend on t, and the ξ_t are i.d.

Definition 5.2.1. *(i) A stochastic matrix on E is a "matrix" $Q = (Q(x, y); x, y \in E)$ with*

$$Q(x, y) \geq 0 , \qquad \sum_{z \in E} Q(x, z) = 1 \qquad \forall x, y \in E,$$

(ii) A sequence $(X_t)_{t \geq 0}$ of r.v.'s $X_t : \Omega \to E$ is a Markov chain with transition Q if, for all $t \in \mathbb{N}$ and $x_0, \dots x_t, y \in E$,

$$P(X_{t+1} = y | X_0 = x_0, \dots X_t = x_t) = Q(x_t, y) \tag{5.3}$$

Denote by μ_0 the initial law of such a chain, i.e. the law of X_0, $\mu_0(x) = P(X_0 = x)$. By induction, the formula (5.3) implies that the joint law of $(X_i; 0 \leq i \leq t\}$ is

$$P(X_0 = x_0, X_1 = x_1, \dots X_t = x_t) = \mu_0(x_0) \prod_{i=1}^{t} Q(x_{i-1}, x_i) \tag{5.4}$$

Conversely, the relation (5.4) for all t and x_i implies (5.3) and characterizes the law of the chain with transition Q and initia law μ_0.

The RHS of (5.4) defines a probability measure on $E^{\mathbb{N}}$, by specifying the probability of cylinders $\{x_0\} \times \ldots \{x_t\} \times E \times E \times \ldots$. On $E^{\mathbb{N}}$, we denote by X_i the i-th coordinate mapping, given for $\mathbf{x} = (x_i; i \geq 0)$ by $X_i(\mathbf{x}) = \mathbf{x_i}$. Then, the sequence $X = (X_i)_i$ is, under this measure, a Markov chain with transition Q and initial distribution μ_0. It is called the *canonical chain*. Therefore, to each pair μ_0, Q we can associate a Markov chain with transition Q and initial law μ_0, and this chain is unique in law.

For $x \in E$, we denote by P_x this law with $\mu = \delta_x$.

Formula (5.4) has some important consequences.

Proposition 5.2.2. *(i) The n-th power Q^n of the matrix Q is a transition matrix, and we have*

$$P(X_{t+n} = y | X_t = x) = Q^n(x, y) \;, \qquad t \geq 0, \; x, y \in E.$$

Q^n is given by

$$Q^0 = \mathrm{Id}, \; Q^n(x, y) = \sum_{z \in E} Q^{n-1}(x, z) Q(z, y) \tag{5.5}$$

so that $Q^{m+n} = Q^m Q^n$ (product of matrices).

(ii) The law μ_t of X_t, i.e. $\mu_t(y) = P(X_t = y)$, is obtained from μ_0 by left multiplication:

$$\mu_t = \mu_0 Q^t \;, \qquad \text{i.e.} \quad \mu_t(y) = \sum_{x \in E} \mu_0(x) Q^t(x, y)$$

(Forward Kolmogorov equation)

(iii) For all function $f : E \to \mathbb{R}$, non negative or bounded,

$$E(f(X_t) | X_0 = x) = Q^t f(x) = \sum_{y \in E} Q^t(x, y) f(y)$$

(Backward Kolmogorov equation)

(iv) Moreover, for all fixed t, the sequence $(X_{t+n}, n \geq 0)$ is a Markov chain with transition Q and initial law μ_t.

5.3 Examples

Example 5.3.1 (Random walks)**.** *A particle moves on the lattice $\mathbb{Z}^d$ by jumping at each integer time t with a random jump ξ_t; Its position X_t is*

$$X_{t+1} = X_t + \xi_{t+1} \;, \qquad t \geq 0,$$

or, equivalently,

$$X_t = X_0 + \sum_{s=1}^{t} \xi_s$$

For a random walk, the sequence $(\xi_t)_t$ *is i.i.d. Letting* q *be the jump law* ($q(y) = P(\xi_1 = y)$)*, we see that the transition matrix* Q *est circulating,*

$$Q(x,y) = q(y-x) \ .$$

Two variants:

1. **RW with absorption:** *The particule moves as before in a subset* D *of* Z^d*, but gets trapped on the complement* D^c*. Recursion equations become nonlinear,*

$$X_{t+1} = \begin{cases} X_t + \xi_{t+1} & \text{if} \quad X_t \in D \\ X_t & \text{if} \quad X_t \notin D \end{cases}$$

   It is still a random recursion, with $f(x,\xi) = x + \xi \mathbf{1}_{x \in D}$*, and then a Markov chain.*

2. **Reflected RW:** *The particule moves freely in* D*, but if it tries to leave* D*, it is kept inside* D *by some reflection mechanism. This is the case of* $d = 1$ *and* $D = \mathbb{N}$ *in* $E = \mathbb{Z}$*, when* $X_{t+1} = |X_t + \xi_{t+1}|, t \geq 0$.

Example 5.3.2 (House of cards)**.** *When growing a house of cards, we have probability* p *of growing and extra floor, and complementary probability* $1-p$ *to completely ruin it, regardless of the current height. The height* $X_t \in \mathbb{N}$ *at time* t *is modeled as the random recursion*

$$X_{t+1} = \xi_{t+1}(X_t + 1)$$

with an i.i.d. Bernoulli sequence ξ *with parameter* p.

In this specific case we can explicitely compute the n*-step transtion:*

$$Q^n(x,y) = \begin{cases} (1-p)p^y & \text{if} & y < \min(x,n) \\ p^n & \text{if} & y = x+n \\ 0 & \text{otherwise} & \end{cases}$$

We see that, as $n \to \infty$*,we have for all* $x, y \in \mathbb{N}$,

$$Q^n(x,y) \to (1-p)p^y \ ,$$

i.e. the n*-step transition converges to a geometric distribution.*

Example 5.3.3 (Inventory)**.** *An electronic store sells a video game system. The stock is an integer r.v.* $X_t \geq 0$*. The demand between dates* t *and* $t+1$ *is a r.v.* ξ_{t+1}*, from an i.i.d. sequence with common law* $(q(y); y \in \mathbb{N})$*. The inventory control policy is, at each date* t *with* $X_t \leq a$*, to bring back the stock level to* b *(*$0 \leq a < b$*). Then,*

$$X_{t+1} = \begin{cases} (X_t - \xi_{t+1})^+ & \text{if} \quad X_t > a \\ (b - \xi_{t+1})^+ & \text{if} \quad X_t \leq a. \end{cases}$$

since demand is satisfied only within the limits of stock X_t *or* b*, hence the operator* $(\cdot)^+$ *above*[(1)]*.* $(X_t; t \geq 0)$ *is then Markov. What is the frequency of shortage (events* $\{X_t = 0\}$*) ? How to choose* a, b *to optimise the profit ?*

Example 5.3.4 (Queues)**.** *A server with* k *units (*$k \geq 1$*), can serve up to* k *clients between times* t *and* $t+1$*. All servers start and end their service at the same (integer) time, and the service time is equal to 1. The number of clients arriving between* t *and* $t+1$ *is Poisson distributed with parameter* θ*. The number* X_t *of clients present at time* t *is such that*

$$X_{t+1} = (X_t - k)^+ + \xi_{t+1},$$

it is a Markov chain. What are the asymptotics of X_t *in function of* k, θ *?*

Example 5.3.5 (Occurence of patterns)**.** *If* $(\xi_t; t \geq -1)$ *i.i.d. Bernoulli,*

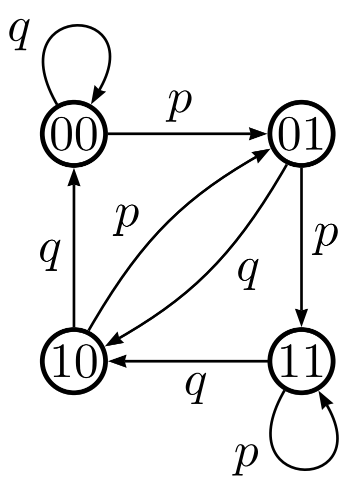
$$P(\xi_t = 1) = p, \qquad P(\xi_t = 0) = q = 1 - p,$$

consider

$$X_t = (\xi_{t-1}, \xi_t) \in \{00, 01, 10, 11\}.$$

Then, X is Markov, and the transitions can be depicted as:

(Recall ABRACADABRA)



[(1)] $x^+ = \max\{x, 0\}$, $x^- = \max\{-x, 0\}$, so that $x = x^+ - x^-$ and $|x| = x^+ + x^-$.

5.4 Construction, simulation

We show how to simulate a Markov chain with given transition Q and starting point $x_0 \in E$, using an i.i.d. sequence $(U_t; t \geq 1)$ uniform on $[0, 1)$.

Enumerate the elements of $E = \{y_1, y_2, \ldots\}$. For all $x \in E$, partition [0,1) into consecutive intervals $I(x, y_1), I(x, y_2), \ldots$ with length $Q(x, y_1), Q(x, y_2), \ldots$:

$$I(x, y_m) = [\sum_{i=1}^{m-1} Q(x, y_i), \sum_{i=1}^{m} Q(x, y_i)[$$

Then define $f : E \times [0, 1) \to E$ so that $f(x, u)$ is the point in E which corresponding interval contains u: precisely,

$$f(x, u) = y \quad \text{with} \quad u \in I(x, y)$$

The function f is measurable, and the sequence

$$X_{t+1} = f(X_t, U_{t+1}) \ , \qquad X_0 = x_0 \ ,$$

is a random recursion, with transition Q starting from x_0.

Remark 5.4.1. *Hence each Markov chain can be represented as a random recursion. In definition 5.1.2 we saw that a random recursion is Markov, hence the two notions are equivalent.*

Example 5.4.2. *For the simple random walk on $\mathbb{Z}$ with drift $2p - 1$ ($p \in [0, 1]$), the above construction reads:*

$$X_{t+1} = X_t + \big(2\mathbf{1}_{U_{t+1} < p} - 1\big)$$

Exercise 5.4.3. *Give a simulation for the case the initial condition X_0 is random with law μ.*

5.5 Strong Markov property

The hitting time[2] of a state $y \in E$ by the Markov chain X,

$$T_y = \inf\{t : t \geq 1, X_t = y\}$$

is a $\mathcal{F}^X$-stopping time. With the convention $\inf \emptyset = \infty$, we have

$$\{T_y < \infty\} = \bigcup_{t \geq 1} \{X_t = y\}$$

[2] It is important for our purpose below to exclude $t = 0$ from the infimum

On the event $\{T_y < \infty\}$, we consider the shifted sequence

$$(X_{T_y+t}; t \geq 0)$$

obtained by shifting the time by T_y; it represents the future of the chain after the first visit T_y to y.

The opposite notion, called the history before T_y, is defined as

$$(X_t; t \leq T_y)$$

Observe that this random variable is a sequence of *random length*, the length being finite on the event $\{T_y < \infty\}$. The possible values this variable can take on this event are the finite sequences $(y_0, y_1, \dots y_n)$ of states in E of length $n+1 \geq 2$, with $y_n = y$ but $y_t \neq y (1 \leq t \leq n-1)$. Such sequences belongs to the set

$$\Delta_y = \{(y_0, y_1, \dots y_n); n \geq 1, y_i \neq y \; \forall i = 1, \dots n-1, y_n = y\} . \tag{5.6}$$

Remark 5.5.1. *This history is in one-to-one correspondance with the stopped chain $(X_{t \wedge T_y}; t \geq 0)$. It generates the σ-field $\mathcal{F}^X_{T_y}$ of events prior to the stopping time T_y,*

$$\sigma(X_t; t \leq T_y) = \mathcal{F}^X_{T_y} , \quad \text{where} \quad \mathcal{F}^X_{T_y} = \Big\{ A \in \mathcal{F}^X_\infty ; A \bigcap \{T_y \leq n\} \in \mathcal{F}^X_n \; \forall n \Big\} .$$

The next result is fundamental. It consists in extending Markov property to some random times.

Proposition 5.5.2 (Strong Markov property)**.** *Let $(X_t; t \in \mathbb{N})$ a Markov chain with transition Q, and let $y \in E$. Conditionally on $T_y < \infty$, the shifted sequence $(X_{T_y+t}; t \geq 0)$ is a Markov chain with transition Q starting from y. It is independent of the history $(X_t; t \leq T_y)$.*

□ a) First, we note that, for all finite sequences $(y_0, \dots y_n)$ and $(x_1, \dots x_m)$ in E, we have

$$\begin{aligned} P\big[(X_t; t \leq T_y) = (y_0, \dots y_n); X_{T_y+t} = x_t, 1 \leq t \leq m)\big] \\ = P\big[(X_t; t \leq T_y) = (y_0, \dots y_n)\big] Q(y, x_1) Q(x_1, x_2) \dots Q(x_{m-1}, x_m) . \end{aligned} \tag{5.7}$$

Indeed, provided that $(y_0, \dots y_n) \in \Delta_y$, the LHS is equal to

$$\begin{aligned} P(X_0 = y_0, \dots X_{n-1} = y_{n-1}, X_n = y, X_{n+1} = x_1, \dots X_{n+m} = x_m) = \\ P(X_0 = y_0) Q(y_0, y_1) \dots Q(y_{n-1}, y) Q(y, x_1) \dots Q(x_{m-1}, x_m) , \end{aligned}$$

whereas the first factor of the RHS reduces to

$$P\big[(X_t; t \leq T_y) = (y_0, \dots y_n)\big] = P(X_0 = y_0) Q(y_0, y_1) \dots Q(y_{n-1}, y_n) .$$

Hence the equality follows in this case. If $(y_0, \dots y_n) \notin \Delta_y$, the two members are equal to zero.

b) Since the union of the events $\{(X_t; t \leq T_y) = (y_0, \dots y_n)\}$ for $(y_0, \dots y_n)$ ranging over Δ_y is $\{T_y < \infty\}$, the above formula implies by summing over Δ_y,

$$P[T_y < \infty; X_{T_y+t} = x_t, 1 \leq t \leq m)] = P[T_y < \infty]Q(y, x_1) \dots Q(x_{m-1}, x_m)$$

This is the first claim. The second one follows from (5.7). ∎

Hence, Markov chains restart afresh from their stopping times.

Chapter 6

Recurrent and ergodic Markov chains

6.1 Recurrence, transience

We consider the number of visits of the chain X to a state $x \in E$,

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

We denote by P_x the law of the chain starting from $X_0 = x$.

Theorem 6.1.1 (Recurrence vs transience)**.** *For every fixed $x_0 \in E$, only two cases can happen for the Markov chain X with transition matrix Q starting from x_0:*

1st case*: the chain a.s. comes back infinitely often to its initial state x_0,*

$$P_{x_0}(N_{x_0} = \infty) = 1$$

2nd case*: the number N_{x_0} of returns to x_0 is a.s. finite, and it is geometrically distributed on $\mathbb{N}$*

$$P_{x_0}(N_{x_0} = k) = (1-a)a^k \ , \quad k \geq 0, \tag{6.1}$$

with parameter equal to the return probability,

$$a = P_{x_0}(T_{x_0} < \infty) < 1.$$

This dychotomy is remarkable, it shows the strength of Markov property.

□ of the theorem: i) Consider the number N'_{x_0} of visits to x_0 of the shifted chain,

$$N'_{x_0} = \sum_{t \geq T_{x_0}+1} \mathbf{1}_{X_t = x_0} \ .$$

On the event $\{T_{x_0} < \infty\}$, the relation

$$N_{x_0} = 1 + N'_{x_0}$$

holds. The strong Markov property implies the second equality below:

$$P_{x_0}(N_{x_0} = k+1 | T_{x_0} < \infty) = P_{x_0}(N'_{x_0} = k | T_{x_0} < \infty) = P_{x_0}(N_{x_0} = k)$$

for all $k \in \mathbb{N} \cup \{\infty\}$. Therefore, by conditionning,

$$\begin{aligned} P_{x_0}(N_{x_0} = k+1) &= P_{x_0}(N_{x_0} = k+1 | T_{x_0} < \infty) \times P_{x_0}(T_{x_0} < \infty) \\ &= a \times P_{x_0}(N_{x_0} = k) \end{aligned}$$

for $k \in \mathbb{N} \cup \{\infty\}$. So $P_{x_0}(N_{x_0} = k) = Ca^k$ with $C = P_{x_0}(N_{x_0} = 0) = P_{x_0}(T_{x_0} = \infty)$, and two cases happen, according to $a = 1$ or not:

- If $a = 1$, the value of $P_{x_0}(N_{x_0} = k)$ doesn't depend on $k < \infty$, it is necessarily zero, as the general term of a convergent series, and thus $P_{x_0}(N_{x_0} = \infty) = 1$
- If $a < 1$, $P_{x_0}(N_{x_0} = k) = (1-a)a^k$ for $k \in \mathbb{N}$, and N_{x_0} has the claimed distribution.

∎

In view of the theorem, the following definition is most natural:

Definition 6.1.2. *The state x_0 is called* **recurrent** *in the first case,* **transient** *in the second case.*

Here is a criterion.

Proposition 6.1.3. *A necessary and sufficient condition for x_0 to be recurrent is*

$$E_{x_0}(N_{x_0}) = \sum_{n \geq 1} Q^n(x_0, x_0) = \infty$$

□ of Proposition 6.1.3. First note that, by Fubini,

$$E_{x_0}(N_{x_0}) = E_{x_0}(\sum_{t \geq 1} \mathbf{1}_{X_t = x_0}) = \sum_{t \geq 1} P_{x_0}(X_t = x_0) = \sum_{t \geq 1} Q^t(x_0, x_0)$$

If x_0 is recurrent $E_{x_0}(N_{x_0}) = \infty$, whereas if x_0 est transient,

$$E_{x_0}(N_{x_0}) = a/(1-a) < \infty$$

as it can be checked for a geometric distribution. ∎

Example 6.1.4. *The random walk on $\mathbb{Z}$, $X_n = \sum_{i=1}^n \xi_i$ with independent ξ_i Bernoulli distributed, $P(\xi_i = 1) = p, P(\xi_i = -1) = 1 - p$.*

(i) When $p \neq 1/2$, the law of large numbers

$$\lim_{n\to\infty} \frac{X_n}{n} = 2p - 1 \qquad a.s.$$

with a non-zero limit, implies that $X_n \to \infty$ a.s., and then $N_0 < \infty$ a.s.The origin is transient, as well as all points by translation invariance.

(ii) If $p = 1/2$, we study the divergence of the series in proposition 6.1.3.Terms with odd indices being zero, we estimate

$$\begin{aligned} \mathbb{P}_0(X_{2n} = 0) &= \binom{2n}{n} 2^{-2n} \\ &\sim (\pi n)^{-1/2} \end{aligned}$$

using Stirling's formula: $n! \sim n^n e^{-n}(2\pi n)^{1/2}$ as $n \to \infty$. This is a divergent Riemann's series, and the symmetric random walk is recurrent.

Many Markov chains have the property that any states can be reached from any other one. For such chain, all states are of the same nature.

Definition 6.1.5. *A transition matrix Q on E (and the associated Markov chains) are called* irreducible *if,*

$$\forall x, y \in E, \quad P_x(T_y < \infty) > 0.$$

This happens iff for every couple x, y, there exists a path $x, z_1, \dots z_k, y$ with $k \geq 0$ arbitrary such that $Q(x, z_1)Q(z_1, z_2)\dots Q(z_k, y) > 0$.

This means that all states communicate with positive probability.

□ Equivalence of the above conditions follows from

$$\{T_y < \infty\} = \{\exists n \geq 1 : X_n = y\}, \qquad \mathbb{P}_x(X_n = y) = Q^n(x, y),$$

and

$$Q^n(x, y) = \sum_{z_1, \dots z_{n-1}} Q(x, z_1)Q(z_1, z_2)\dots Q(z_{n-1}, y)$$

■

Example 6.1.6. *The random walk on $\mathbb{Z}$ with $p \in (0, 1)$ is irreducible, as well as the house of cards.*

On the contrary, if there exists an absorbing *state for the chain (i.e., some x with $Q(x, x) = 1$), the chain is not irreducible.*

Exercise 6.1.7. *The transitions*

$$Q_1 = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/6 & 1/2 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 3/4 & 1/4 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/6 & 1/2 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

are not irreducible on $\{1, 2, 3, 4, 5\}$. *But their restrictions to* $\{1, 2, 3\}$ *are irreducible, as well as that of* Q_2 *to* $\{4, 5\}$.

Definition and Proposition 6.1.8. *Let the transition* Q *be irreducible on* E. *Then, only two cases are possible for Markov chains with transition* Q.

- **1st case**: *All these chains a.s. visit infinitely often all states* $y \in E$,

$$P_x(N_y = \infty) = 1 \qquad \forall x, y$$

  In particular, all states are recurrent.

- **2nd case**: *All these chains visit finitely many times the state* $y \in E$,

$$P_x(N_y < \infty) = 1 \qquad \forall x, y$$

  In particular all states are transient.

The transition matrix Q *and the chains are called recurrent, resp. transient in the respective case.*

In the recurrent case, whatever the starting point is, the Markov chain visits infinitely often every site. This corresponds to an image of "complete mixing". In the transient case, the chain ends up but leaving every finite subset of the state space(1). This case can happen only if E is infinite(2), as we state now:

Corollary 6.1.9. *A transition* Q *irreducible on a finite state space is recurrent.*

The proof of the proposition requires a tool.

6.2 Excursions based at a recurrent point

Let x a recurrent state for the chain with transition Q. By definition, the random set $\{t : X_t = x\}$ is a.s. infinite for the chain starting from x. The times of successive visits to x, T_x^k defined recursively by

$$T_x^{k+1} = \min\{t : t > T_x^k, X_t = x\}$$

(1)Indeed the number of visits to F is $\sum_F N_y$, which is a.s. finite if F is finite.
(2)Since $\sum_E N_y = \infty$, at least one of the N_y is infinite if E is finite.

with $T_x^0 = 0$ by convention and $T_x^1 = T_x$ by definition, are a.s. finite. This allows us to consider the successive *excursions* based at x of the chain,

$$\mathcal{E}_k = (X_{T_x^k}, X_{T_x^k+1}, \dots X_{T_x^{k+1}}) \,, \qquad k \geq 0, \tag{6.2}$$

that the chain performs between its visits to x.

Each excursion is a random variable, a sequence of finite length ≥ 2, which elements are sites of E distinct of x except for the first and last ones which are equal to x. Denote by

$$u = (x, x_1, \dots x_n, x), \quad \text{with } n \geq 0, x_i \neq x, \ i = 1, \dots n,$$

such a sequence, and by U the set of such sequences. This set U is countable, and constitutes the space of possible values of an excursion. The excursions $\mathcal{E}_k$ are discrete r.v., and we denote by p the law of the first one,

$$p(u) = P_x(\mathcal{E}_0 = u), \qquad u \in U.$$

Lemma 6.2.1. *The excursions $\mathcal{E}_k, k \geq 0$, are independent and identically distributed. In other words, for any recurrent state x of the Markov chain, for all $k \geq 0$ and $u_i \in U, i = 1, \dots k$,*

$$P_x(\mathcal{E}_0 = u_0, \mathcal{E}_1 = u_1, \dots \mathcal{E}_k = u_k) = p(u_0)p(u_1)\dots p(u_k),$$

where $(\mathcal{E}_k, k \geq 0)$ denotes the infinite sequence of excursions based at x.

□ By induction on k with the strong Markov property. By the strong Markov property, $X_{T_x^1+\cdot}$ is independent from the excursion $\mathcal{E}_0$. (Since T_x is P_x-a.s. finite, the conditioning in Proposition 5.5.2 can be ignored.) Since the excursions $\mathcal{E}_1, \mathcal{E}_2, \dots$ depend only on the shifted chain, we see from Proposition 5.5.2 that

$$P_x(\mathcal{E}_0 = u_0, \mathcal{E}_1 = u_1, \dots \mathcal{E}_k = u_k) = p(u_0) P_x(\mathcal{E}_1 = u_1, \dots \mathcal{E}_k = u_k)$$

But, again by the strong Markov property, $X_{T_x^1+\cdot}$ has the law P_x of the original chain X and the excursions $\mathcal{E}_1, \mathcal{E}_2, \dots$ are the excursions of the shifted chain. Hence,

$$P_x(\mathcal{E}_1 = u_1, \dots \mathcal{E}_k = u_k) = P_x(\mathcal{E}_0 = u_1, \mathcal{E}_1 = u_2 \dots \mathcal{E}_{k-1} = u_k)$$

Then the induction is proved, and the equality in the lemma follows since for $k = 0$ is reduces to the definition of the law p. ∎

□ of proposition 6.1.8.

(i) Assume that there exists at least one recurrent state x. Then for $k \geq 0$, consider the chain starting from x and the events

$$A_k = \left\{ \exists t \in [T_x^k; T_x^{k+1}] : X_t = y \right\}$$

for another state $y \neq x$. Each A_k depends only on the excursion $\mathcal{E}_k$ and these events are defined in a similar manner from their respective excursion.Then the A_k are independent and have the same probability. If this probability was equal to 0, $P_x(\cup_k A_k) = 0$ contradicting the irreducibility assumption.

Hence $P_x(A_k) > 0$ and by the law of large numbers for the i.i.d. sequence $\mathbf{1}_{A_k}$, the A_k's occur infinitely often with probability one, and

$$P_x(N_y = \infty) = 1.$$

Then, $P_x(T_y < \infty) = 1$ and the shifted chain $X_{T_y+\cdot}$ visits y i.o. like the sequence X. This new chain is the Markov chain with transition Q starting from y, so that $P_y(N_y = \infty) = 1$: y itself is recurrent.

We showed that if there exists a recurrent state, then every state is recurrent. Moreover, if x is recurrent, $P_x(N_y = \infty) = 1$ for all $y \in E$. The proposition is then proved in the case there exists at least one recurrent state x.

(ii) In the opposite case, the formula (6.1), applied to y, combined with the strong Markov property for the chain P_x and the hitting time of y, yields

$$P_x(N_y = k) = \begin{cases} 1-b & \text{si} \quad k = 0, \\ b(1-a)a^{k-1} & \text{si} \quad k \geq 1 \end{cases}$$

with $b := P_x(T_y < \infty)$ and $a := P_y(T_y < \infty) < 1$. This result is clearly more precise than the one claimed in the proposition in the transient case. ∎

Example 6.2.2 (Ehrenfest's diffusion)**.** *A system of N particles is confined in a container divided in two parts separated by a porous wall. At time t we have X_t particle on the left side, and $N - X_t$ on the right. The time scale is chosen so that there is one and only one particle changing of side at a time. If this particle is chosen at random, we see that X_t is the Markov chain with transition*

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

and $Q(x, y) = 0$ otherwise. The chain is irreducible (for $x < y$ we have $Q(x, x+1)Q(x+1, x+2)\ldots Q(y-1, y) > 0$), hence recurrent by corollary 6.1.9. This may look weird, especially when N is very large ($N = 6\times 10^{23}$ is a value of interest) : starting from $X_0 = N/2$ (assume N is even), proposition 6.1.8 claims that the state N for which all particles are on the same side of the wall, will be reached a.s. ! Unexpected...

However this model is quite pertinent for diffusion of gaz molecules. The point is that the time needed to reach this strange state N is enormous, so enornous that we have no risk to see it happen during our life. The time scale is exponential. We will show later that the mean value of the hitting *time is* 2^N. *For* $N = 70$, *the mean time is* $2^N \simeq 10^{21}$ *roughly*[(3)], *and considering that particles experience* 100 *changes of side per seconds, we get that the mean time is* 10^{19} *seconds. This value is already larger than the age of universe, estimed to* 5×10^{17} *seconds. In statistical physics, the number of particles* N *is much larger –of the order of Avogadro number* 6.02×10^{23}–, *and the mean hitting time is absolutely enormous.*

6.3 Invariant distribution

For a recursive sequence $x_{t+1} = g(x_t)$ we study fixed points $x = g(x)$, they are potentially attractors of the sequence $(x_t)_t$. Here is the counterpart in the stochastic framework.

6.3.1 Definition, examples

Definition 6.3.1. *A probability* π *on* E *is invariant for the transition matrix* Q *if*

$$\pi Q = \pi \ , \qquad \text{i.e.,} \quad \sum_{x \in E} \pi(x) Q(x,y) = \pi(y) \quad \forall y \in E. \tag{6.3}$$

A positive measure on E *with (6.3) is an invariant measure.*

For the recursion $X_{t+1} = f(X_t, \xi_{t+1})$, this amounts to a law π such that

$$X_0 \sim \pi, \text{ independent of } \xi_1 \Rightarrow f(X_0, \xi_1) \sim \pi$$

In all generality, if X_0 has law π, X_1 also, and inductively all X_t too.

The definition writes

$$\sum_{x \neq y} \pi(x) Q(x,y) = \pi(y)[1 - Q(y,y)] \ , \qquad y \in E \ ,$$

i.e.: at equilibrium, at each site the mean arrival flow is equal to the mean departure flow.

By linearity of (6.3), the set of invariant probability measures [resp., positive measures] for a given transition Q is a convex set [resp., a convex cone], if it is not empty: indeed, $\lambda\pi_1 + (1-\lambda)\pi_2$ is invariant if π_1, π_2 are invariant, and $r\pi$ is an invariant measure if π is invariant and $r > 0$.

Exercise 6.3.2. *For the random walk on* $\mathbb{Z}/N\mathbb{Z}$ *with* $Q(x, x+1) = p, Q(x, x-1) = 1-p$ *(with* $p \in (0,1)$ *a parameter and addition modulo* N*), the uniform law is the unique invariant law.*

[(3)] Approximating $2^{10} = 1024$ by 10^3

Example 6.3.3. *Random walk on* $\mathbb{Z}$, $Q(x,x+1) = p, Q(x,x-1) = 1-p$ *(with some* $p \in (0,1)$*). We look for solutions* $\pi : \mathbb{Z} \to \mathbb{R}^+$ *of*

$$p\pi(x-1) + (1-p)\pi(x+1) = \pi(x) \ , \qquad x \in \mathbb{Z}, \tag{6.4}$$

which is a linear recursion of order 2. The characteristic equation

$$p + (1-p)r^2 = r \ ,$$

has roots $r \in \mathbb{R}$ *such that the measure* $\pi(x) = r^x$ *solves (6.4). The characteristic equation writes* $(r-1)\big((1-p)r - p\big) = 0$, *there are two cases.*

1st case, $p = 1/2$*: then* $r = 1$ *is double root, we look for solutions of the form* $\pi(x) = ax + b, x \in \mathbb{R}$*. Positivity of* π *requires* $a = 0, b \geq 0$*. The uniform measure*

$$\pi_1(x) = 1 \ , \qquad x \in \mathbb{Z},$$

is the unique (up to multiplicative factor) invariant measure, it has infinite mass.

2nd case, $p \neq 1/2$*: the roots* $r = 1, p/(1-p)$ *are distinct, solutions of (6.4) are of the form* $\pi(x) = a + b(p/(1-p))^x$*. Positivity implies* $a, b \geq 0$, *invariant measures are combinations of* π_1 *above and* π_2*, with*

$$\pi_2(x) = \left(\frac{p}{1-p}\right)^x, \qquad x \in \mathbb{Z}$$

which has also infinite mass. ∎

Exercise 6.3.4. *For* $\ell \in \mathbb{R}^d$*, consider the random walk on* $\mathbb{Z}^d$ *with*

$$Q(x, x+e) = \Big(\sum_{e':|e'|_1=1} \exp \ell \cdot e'\Big)^{-1} \exp \ell \cdot e \ , \qquad \text{for } e \in \mathbb{Z}^d, |e|_1 = 1 \ ,$$

and $Q(x,y) = 0$ *if* $|y-x|_1 \neq 1$*. (Steps to nearest neighbors, with a bias in the direction* ℓ*.) Check that the uniform measure* $\pi_1(x) = 1$ *is invariant, as well as*

$$\pi_2(x) = \exp 2\ell \cdot x$$

Example 6.3.5. *(Absence of non trivial invariant measure). Consider the totally assymetric random walk on the integers, given by* $Q(x,x+1) = p, Q(x,x) = 1-p$ *et* $Q(x,y) = 0$ *pour* $y \neq x, x+1$*. (It is a Bernoulli process,* $X_n = X_0 + \sum_{i=1}^n \xi_i$ *with* ξ *an i.i.d. Bernoulli sequence.) Viewed as a Markov chain on* $\mathbb{N}$*, it is not irreducible, and it does not have any invariant measures except zero. Indeed, invariance means*

$$\pi(0) = (1-p)\pi(0), \mu(x) = p\mu(x-1) + (1-p)\mu(x) \forall x \geq 1,$$

i.e. $\pi = 0$*. Now, viewed as a Markov on* $\mathbb{Z}$*, it leaves invariant the counting measure, but it is not irreducible.*

Exercise 6.3.6. *Find all the invariant measures for the following transition on $E = \{1, 2, 3\}$:*

$$Q = \begin{pmatrix} .4 & .6 & 0 \\ .2 & .8 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As shown in example 6.3.3, there may not exist any invariant law, and exercise 6.3.6 shows that there may be many.[(4)]

Thus, there are *existence and uniqueness* problems for the invariant law. We start with the finite case.

6.3.2 Finite state space

Elementary considerations of linear algebra lead to the following:

Proposition 6.3.7. *(Finite state space) When E is finite, we have:*
(i) Existence of an invariant law.
(ii) Uniqueness if the chain is irreducible.

□ (i) Indeed, the stochastic matrix Q satisfies

$$Q\mathbf{1} = \mathbf{1},$$

and so 1 is a left eigenvalue. Then 1 is also a right eigenvalue, and there exists an eigenvector $v \neq \vec{0}$, with $vQ = v$. We argue now, by positivity of the entries of Q, that we can choose v with non negative coordinates. Indeed, for all $y \in E$,

$$|v(y)| = |\sum_{x \in E} v(x)Q(x, y)| \leq \sum_{x \in E} |v(x)|Q(x, y),$$

where none of the inequalities can be strict, since the sums over $y \in E$ of the two members are equal. Hence $(|v(x)|; x \in E)$ is a right eigenvector[(5)]. Normalizing this eigenvector, we obtain an invariant law.

(ii) We start to show *Liouville property*: harmonic functions are constant. (A function $f : E \to \mathbb{R}$ is harmonic for Q if $Qf = f$.) Indeed, suppose there exists $y \in E$ such that $f(y) < \max_E f$. Since E is finite, f achieves its maximum at some $x \in E$. Since f is harmonic for Q, it is harmonic for Q^n for all n. Fix n s.t. $Q^n(x, y) > 0$, and write (*maximum principle* argument)

$$f(x) = Q^n f(x) = \sum_{z \neq y} Q^n(x, z)f(z) + Q^n(x, y)f(y) < f(x) \sum_z Q^n(x, z) = f(x),$$

[(4)]Examples 6.3.3 and 6.3.5 illustrate similar conclusions for invariant measures.
[(5)] Stochastic matrices are covered by Perron-Frobenius theory of positive matrices.

which is absurd. Hence f is constant, and Liouville property holds. Then the eigenspace of Q for eigenvalue 1 has dimension 1, and since its transposed has the same property, we derive that the set of row vectors v solutions de $vQ = v$ is a one-dimensional vector subspace. There is at most one invariant probability. ∎

6.3.3 Existence

We come back to the general case when E can be infinite. Here is an idea from dynamical systems to find invariant laws. Let μ_t the law of X_t, and consider the Cesáro mean

$$\nu_n = n^{-1} \sum_{t=0}^{n-1} \mu_t$$

It is a sequence of probability measures on E. We claim that all limit point (for the weak convergence) is invariant:

Proposition 6.3.8. *Assume for a subsequence $(n_k)_k$,*

$$\forall x \in E, \quad \lim_{k\to\infty} \nu_{n_k}(x) = \nu(x) \qquad \text{with} \quad \sum_{x\in E} \nu(x) = 1 \ .$$

Then, ν is an invariant probability for Q.

□ By Fatou's lemma,

$$\liminf_k \nu_{n_k} Q \geq \nu Q$$

(coordinatewise). But both $\nu_{n_k} Q$ and ν_{n_k} are sums of n_k terms, only 2 of which are different. We derive that

$$|\nu_{n_k} Q(y) - \nu_{n_k}(y)| \leq 2/n_k \to 0$$

as $k \to \infty$. Hence,

$$\nu \geq \nu Q \ .$$

This implies the equality, for if there exists y with $\nu(y) > \nu Q(y)$, we would get by summing[6] $\|\nu\| > \|\nu\|$ since $\|\nu\| < \infty$, which is a contradiction. ∎

A handy condition for invariance.

Definition and Proposition 6.3.9 (Reversible measure). *A measure π on E is reversible for the transition Q if*

$$\pi(x)Q(x,y) = \pi(y)Q(y,x) \ , \qquad x, y \in E \ . \tag{6.5}$$

A reversible measure is invariant.

[6]Denoting by $\|\nu\| = \sum_x \nu(x)$ the mass of ν

$\square$ Summing over x the equality (6.5) we get

$$\pi Q(y) = \sum_x \pi(y)Q(y,x) = \pi(y)$$

showing invariance. $\blacksquare$

Remark 6.3.10. *Invariant/reversible measure and flows. The quantity $\mu(x)Q(x,y)$ is the flow in one step of a mass $\mu(x)$ at x to site y by the transition Q. The property of invariance,*

$$\sum_{x\in E} \mu(x)Q(x,y) = \sum_{z\in E} \mu(y)Q(y,z) \ ,$$

means that the ingoing flow to y is equal to the outgoing flow from y. The property of reversiblity (6.5) *means that the flow on the edge (x,y) is equal to that on the reverse edge (y,x). It is a local condition, in contrast with invariance which is a global one.*

Exercise 6.3.11. *In example 6.3.3 of the random walk on $\mathbb{Z}$ with $p \neq 1/2$, check that π_2 is reversible, but π_1 is not.*

Condition (6.5) is called "detailed balance condition" because it details all pairs of states x, y instead of summing over x as in the definition of an invariant measure.

Exercise 6.3.12 (Ehrenfest's diffusion)**.** *Consider Ehrenfest model, with $E = \{0, 1, \ldots, N\}$ and*

$$Q(x,x-1) = \frac{x}{N} \ , \qquad Q(x,x+1) = \frac{N-x}{N} \ , \qquad 0 \le x \le N \ ,$$

$Q(x,y) = 0$ otherwise. The chain is irreducible on E finite has a unique invariant law π. Solving the reversibility equations, check that π is the binomial law $\mathcal{B}(N,1/2)$.

In the next exercise, we give another "detailed" condition, sufficient for invariance.

Exercise 6.3.13. *(i) Let Q a transition matrix on E, π a measure on E and $T : E \times E \to E$ such that $x \mapsto T_y(x)$ is a bijection for all y. Show that the conditions*

$$\pi(x)Q(x,y) = \pi(y)Q(y, T_y(x))) \ , \qquad x, y \in E \ ,$$

imply that π is invariant for Q.

(ii) Random walk on the discrete torus: Let $p \in (0,1)$, $E = \mathbb{Z}/N\mathbb{Z}, Q(x,x-1) = p, Q(x,x+1) = 1-p, Q(x,y) = 0$ if $y \neq x, x+1$ modulo N. Check that $T_y(x) = 2y - x \ [N]$ the symmetric of x w.r.t. y, satisfies the condition in (i) with the uniform probability.

Here is the connection with graph theory.

Example 6.3.14 (Weighted graph, resistor network, conductance network)**.** *A weighted graph is a pair* (E, c) *composed of a discrete "vertex set"* E *and weights* $c = (c_{x,y}; x, y \in E)$ *with* $c_{x,y} = c_{y,x} \geq 0$*. The other pair* (E, L) *with* $L = \{(x, y) : c_{x,y} > 0\}$ *is called an unoriented graph with vertices* E *and edges* L*. Assume that*

$$c_x \stackrel{\text{def}}{=} \sum_{y \in E} c_{x,y} \in (0, \infty) \qquad \forall x \in E.$$

Then,

$$Q(x, y) = \frac{c_{x,y}}{c_x}, \qquad x, y \in E,$$

defines a transition matrix on E*. The corresponding Markov chain is called the* random walk on the weighted graph (E, c)*. It is plain that*

$$\pi(x) = c_x, x \in E, \quad \text{is a reversible measure.}$$

Starting from x*, the walk jumps at time 1 along one of the edges* (x, y) *starting from* x *which is chosen with probability proportional to* $c_{x,y}$*. For that reason, the weight* $c_{x,y}$ *is also called* conductance *of the edge* (x, y)*, and the weighted graph is called conductance network.*

Now, one also defines the resistance *of an edge as the inverse of the conductance,*

$$r_{x,y} = \frac{1}{c_{x,y}} \in (0, \infty],$$

so infinite resistance means that it is impossible to jump directly from x *to* y *(or vice-versa). Then, the weighted graph is also called resistor network. For concreteness, consider the simple random walk in example 5.4.2, it is the resistor network*

$$E = \mathbb{Z}, \qquad c_{x,x+1} = (\frac{p}{1-p})^x, x \in \mathbb{Z} \quad \text{and} \quad c_{x,y} = 0 \quad \text{if} \quad y \neq x \pm 1:$$

When $p > 1/2$*, the conductance increases exponentially as we move to the right, which reflects that the walk drifts to* $+\infty$*.*

In full generality, the usual electrostatic theory (Ohm's and Kirchhoff's laws, electric potential, power, ...) has a full and deep counterpart here. We do not develop further here, but simply mention the online (small and wonderful) book of Doyle and Snell:
`https://math.dartmouth.edu/~doyle/docs/walks/walks.pdf`

6.3.4 Uniqueness

Proposition 6.3.15. *An irreducible Markov chain has at most one invariant law.*

□ (i) By irreducibility, any invariant law π is supported by the whole space E: Indeed if $\pi Q = \pi$ then $\pi Q^n = \pi$, and also

$$\pi(x) Q^n(x,y) \le \pi(y) \ , \qquad \forall x, y \in E, n \ge 1 \ .$$

If $\pi(y) = 0$ for some y, then $\pi = 0$, a contradiction.

(ii) If $\pi' \neq \pi$ is another invariant law, then

$$\pi''(x) := \min\{\pi(x), \pi'(x)\}$$

is a positive measure on E such that, componentwise,

$$\pi'' Q \le \pi''$$

(since $\pi'' Q(x) \le \pi Q(x) = \pi(x)$, and similarly with π'). Since these two measures have the same finite mass, $\pi'' \mathbf{1} = \pi'' Q \mathbf{1} \in [0,1]$, arguing by contradiction we see that

$$\pi'' = \pi'' Q \ .$$

Hence,

$$\mu = \frac{\pi - \pi''}{C} \ , \quad C = \sum_{x \in E} (\pi - \pi'')(x) > 0$$

it itself an invariant law for Q (note that $C > 0$ since $\pi \neq \pi'$). By (i), we have $\mu(x) > 0$ for all $x \in E$, so $\pi(x) > \pi'(x)$ for all x. This is a contradiction, and there is uniqueness. ∎

Now, a different idea coming from dynamical systems tells us that the invariant law is related to the mean return time.

Theorem 6.3.16 (Kac's lemma)**.** *If the irreducible recurrent transition Q has an invariant law π, then*

$$\pi(x) = \frac{1}{E_x T_x} \ , \qquad x \in E \ .$$

We will prove this result as a consequence of the ergodic theorem in the next section. Let us first state

Remark 6.3.17. *An irreducible recurrent Markov chain has a unique invariant measure, uniqueness being understood up to multiplicative factor. More precisely, for all x the measure*

$$\mu(y) = E_x \sum_{t=0}^{T_x - 1} \mathbf{1}_{X_t = y} = \sum_{t=0}^{\infty} P_x(X_t = y, t < T_x) \, , \tag{6.6}$$

with $T_x = \inf\{t \geq 1 : X_t = x\}$, is a non-trivial invariant measure. If it is normalizable, the corresponding probability measures coincide for different x's, and they coincide with Kac's formula. Cf. Pardoux, Th.2.19 for a proof.

6.4 Ergodic chains

An irreducible recurrent Markov chain visits infinitely often every site. The frequency of visits follows a law of large numbers.

Theorem 6.4.1 (Ergodic theorem)**.** *Let Q be an irreducible recurrent stochastic matrix on E. Then, for all $x, y \in E$, the following limit exists as $n \to \infty$,*

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{X_t=x} \longrightarrow \mu(x) = \frac{1}{E_x T_x}\ , \quad P_y - a.s. \tag{6.7}$$

Two cases can happen:

- *either μ is the invariant probability for the chain,*
- *or $\mu = 0$, and in this case, there is no invariant probability.*

The first case is called **ergodic** or **positive recurrent**, the second one **nul recurrent**. Ergodicity corresponds to short return times, i.e., integrable ones. On the contrary, an irreducible recurrent Markov chain does not have an invariant law if it takes "too long" to come back to the starting point.

Corollary 6.4.2 (Pointwise ergodic theorem)**.** *Consider $(X_n)_n$ an irreducible recurrent Markov chain on E, with invariant law π, and $f : E \to \mathbb{R}$ a bounded function . Then, for all $y \in E$, we have $P_y - a.s.$,*

$$\frac{1}{n}\sum_{t=1}^{n} f(X_t) \longrightarrow \int_E f d\pi = \sum_{x\in E} f(x)\pi(x) \tag{6.8}$$

The ergodic hypothesis formulated by Boltzmann is that the average of a experimental process over time and the average over the statistical ensemble are the same. The statistical ensemble is some probability distribution for the state of the system, which is preserved by the dynamics. For Markov dependence, under the above assumptions it is a theorem, and the "statistical ensemble" is the invariant law.

□ The proof of Theorem 6.4.1 will follow from independence of excursions.

1) We start to show (6.7) with $y = x$. Under P_x, the excursion lengths $(T_x^{k+1} - T_x^k)_k$ are i.i.d. non-negative r.v.'s according to Lemma 6.2.1. They

obey the strong law of large numbers[7], which writes

$$\frac{1}{k}T_x^k = \frac{1}{k}\sum_{l=1}^{k}(T_x^k - T_x^{k-1}) \longrightarrow E_xT_x \in [1,\infty] \ , \quad P_x - a.s. \text{ as } k\to\infty.$$

This implies

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{X_t=x} = \frac{1}{E_xT_x} \ , \quad P_x - a.s. \tag{6.9}$$

Indeed, for all $n \geq 0$ there exists a $k = k(n) \geq 0$ such that $T_x^k \leq n < T_x^{k+1}$, and then

$$\frac{k}{k+1}\frac{k+1}{T_x^{k+1}} \leq \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{X_t=x} = \frac{k}{n} \leq \frac{k}{T_x^k} \ .$$

Letting $n\to\infty$ we have $k = k(n) \to \infty$ and we get the desired result. From now on we set $\mu(x) = 1/E_xT_x$.

2) For $y \neq x$ another point in E, we have

$$|\sum_{t=1}^{n}\mathbf{1}_{X_{T_y+t}=x} - \sum_{t=1}^{n}\mathbf{1}_{X_t=x}| \leq T_y$$

which is finite P_x-a.s. by recurrence. Hence, from (6.9),

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{X_{T_y+t}=x} = \mu(x) \ , \quad P_x - p.s.$$

Recall the strong Markov property: the shifted chain $X_{T_y+\cdot}$ is a chain with transition Q starting from y, so the above statement is exactly (6.7).

3) The next step is to prove that $\mu \neq 0$ implies that μ is a probability. Consider $x \in E$ with $\mu(x) > 0$, i.e. $E_xT_x < \infty$. Consider also the excursions based at x: the number of visits to y during each excursion,

$$\sum_{t=T_x^k+1}^{T_x^{k+1}}\mathbf{1}_{X_t=y} \ , \quad k \geq 1 \ ,$$

is P_x a sequence of i.i.d.r.v.'s. By the law of large numbers,

$$\frac{1}{k}\sum_{t=1}^{T_x^k}\mathbf{1}_{X_t=y} \longrightarrow E_x[\sum_{t=1}^{T_x}\mathbf{1}_{X_t=y}] \ , \quad P_x - a.s. \text{ as } k\to\infty.$$

But the left-hand side is equal to

$$\frac{T_x^k}{k} \times \frac{1}{T_x^k}\sum_{t=1}^{T_x^k}\mathbf{1}_{X_t=y} \longrightarrow E_x[T_x] \times \mu(y) \ , \quad P_x - a.s. \text{ as } k\to\infty \ ,$$

[7] For an i.i.d. sequence non necessarily integrable but non-negative

using (6.7) with x and y interchanged and by dominated convergence. The above two limits are then equal,

$$E_x[\sum_{t=1}^{T_x} \mathbf{1}_{X_t=y}] = E_x[T_x] \times \mu(y) \ ,$$

which, by suming over $y \in E$, yields

$$E_x[T_x] = E_x[T_x] \times \sum_{x \in E} \mu(y) \ .$$

By assumption, $E_x[T_x] < \infty$, implying that μ is a probability measure as soon as there is some x with $\mu(x) > 0$.

4) We now check that μ is invariant if it is a probability. Indeed, combining (6.7) with dominated convergence theorem, we see that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} Q^t(y,x) = \mu(x) \tag{6.10}$$

Since μ is a probability, proposition 6.3.8 applies, and shows that μ is an invariant law.

5) It remains to show that when $\mu = 0$, there is no invariant probability. Otherwise, denote it by π, multiply by $\pi(y)$ the relation (6.10) and sum over $y \in E$: we obtain $\pi(x) = \mu(x) = 0$ for all x, a contradiction. ∎

□ Proof of Corollary 6.4.2. We first state a property about convergence of probability measures in a discrete set.

Proposition 6.4.3. *Let E be a discrete*[8] *set and $\nu, \nu_n (n \geq 1)$ be probability measures on E. As $n \to \infty$, the following statements are equivalent:*

(i) $\nu_n(x) \to \nu(x)$ *for all* x,

(ii) $\nu_n \to \nu$ *weakly,*

(iii) $\nu_n \to \nu$ *in* ℓ_1*-norm.*

Setting $\nu_n = \frac{1}{n}\sum_{t=1}^{n} \delta_{X_t}$ and $\nu = \pi$, we see from Theorem 6.4.1 that (i) holds true on an event of full P_y-measure. Then, we write

$$\begin{aligned} \left|\frac{1}{n}\sum_{t=1}^{n} f(X_t) - \sum_{x\in E} f(x)\pi(x)\right| &= \left|\sum_{x\in E} f(x)[\nu_n(x) - \pi(x)]\right| \\ &\leq \|f\|_\infty \|\nu_n - \pi\|_1, \end{aligned}$$

[8] This is the crucial assumption to have equivalence between weak and strong convergence as stated here.

which vanishes as $n \to \infty$ on this event. ∎

□ Proof of Proposition 6.4.3: We have clearly (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i). We end by proving (i) $\Longrightarrow$ (iii). Fix $\varepsilon > 0$. Since ν has mass 1, we can find A with $\nu(A) \geq 1 - \varepsilon$. By (i) and since A is finite, we can fix n_0 such that

$$\nu_n(A) = \sum_{x \in A} \nu_n(x) \geq 1 - 2\varepsilon, \qquad \forall n \geq n_0.$$

Thus,

$$\begin{aligned} \|\nu_n - \nu\|_1 &\leq \sum_{x \in A} |\nu_n(x) - \nu(x)| + \nu_n(A^{\complement}) + \nu(A^{\complement}), \\ \lim_{n \to \infty} \|\nu_n - \nu\|_1 &\leq 3\varepsilon, \end{aligned}$$

by (i) for A is finite, which yields (iii). ∎

Corollary 6.4.4. *Any irreducible Markov chain on a finite state space is ergodic.*

□ By corollary 6.1.9, we already know that Q is recurrent. Since E is finite, there is an invariant law by proposition 6.3.7. Hence, we are in the ergodic case. ∎

The theorem has another stronger corollary:

Remark 6.4.5. *An irreducible Markov chain which has an invariant probability is necessarily ergodic.*

□ The ergodic theorem remains trivially valid in the transient case: the sum in (6.7) is bounded and $\mu = 0$. Then, an irreducible Markov chain satisfies (6.7), and then (6.10). If π is invariant, multiply by $\pi(y)$ the relation (6.10) and sum over $y \in E$: we get $\mu = \pi$. ∎

Example 6.4.6 (Ehrenfest diffusion, the return). *Come back to the example 6.2.2. By exercise 6.3.12, the binomial $\pi = \mathcal{B}(N, 1/2)$ is a reversible probability. By the ergodic theorem, the asymptotic frequency of visits to x ($0 \leq x \leq N$) converges,*

$$\frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{X_t = x} \longrightarrow \binom{N}{x} 2^{-N}.$$

If N is large (and even, say), $\pi(N/2) \sim (2/\pi N)^{1/2}$ is much, much larger than $\pi(0) = 2^{-N}$. Precisely, for $c > 0$ and $I_c = [(N - cN^{1/2})/2, (N + cN^{1/2})/2]$, the Gaussian approximation of the binomial shows that

$$\pi(I_c) \sim \mathbb{P}(Z \in I_c)\,, \quad \text{with } Z \sim \mathcal{N}(0,1).$$

When c is of the order of several units, $\pi(I_c)$ is extremely close to 1 and Ehrenfest diffusion, starting from the fair configuration $X_0 = N/2$, has practically no chance to exit I_c within a reasonable time.

6.5 Aperiodic chains

Definition 6.5.1. *Let Q a transition matrix on E, and $x \in E$ a state for which $Q^n(x,x) > 0$ for some $n \geq 1$. The period of x is the greatest common divisor (GCD=PGCD in french) of the set $I(x) = \{n \geq 1 : Q^n(x,x) > 0\}$. The state x is aperiodic if it has period 1.*

Example 6.5.2. *The nearest neighbor random walk on $\mathbb{Z}^d$ has period 2. Its lazy version,*

$$Q(x,x) = q \in (0,1), \qquad Q(x,y) = (1-q)/(2d) \quad \text{if } |y-x|_1 = 1,$$

is aperiodic, because $Q(x,x) > 0$.

If Q is irreducible, all states have the same period. (Cf. [Pardoux].) The following is not difficult.

Exercise 6.5.3. *If Q is irreducible and has an aperiodic state, then all states are aperiodic.*

With $I(x)$ as in definition 6.5.1, consider the additive subgroup $J(x)$ of $\mathbb{Z}$ generated by $I(x)$. By definition of GCD, we have $J(x) = p\mathbb{Z}$ with p the period of x. In particular, x is aperiodic if $J(x) = \mathbb{Z}$.

Proposition 6.5.4. *The state x is aperiodic if and only if there exists an integer N such that $Q^n(x,x) > 0$ for all $n \geq N$.*

□ If $Q^n(x,x) > 0$ for $n \geq N$, the period of x divides N and $N+1$, then it is equal to 1.

Conversely, assume $J(x) = \mathbb{Z}$. By Bézout's theorem, there exist $a', b' \in I(x)$ and $c, d \in \mathbb{N}$ such that $ca' - db' = 1$. As $I(x)$ is stable by addition, there exist $a, b \in I(x) \cup \{0\}$ with $a = b+1$ (take $a = ca', b = db'$). If $b = 0$ or $b = 1$, we see that $N = 1$ fits. So we consider the case $b \geq 2$, and $b \in I(x), a = b+1 \in I(x)$. For $n \geq b(b-1)$, the Euclidean division of $x - b(b-1)$ by b writes

$$n - b(b-1) = qb + r,$$

with $q \geq 1$, $0 \leq r \leq b-1$. Observing that $r = r(a-b)$, we get

$$n = (b-1+q-r)b + ra,$$

with $b-1+q-r \geq 0$, and then $n \in I(x)$ since this set is stable. We get the desired result with $N = b(b-1)$. ∎

Proposition 6.5.5. *Let Q be irreducible on E. Then Q has period $p \geq 2$ iff there exists a finite partition $E = E_1 \cup E_2 \cup \ldots \cup E_p$ s.t. $Q(x, y) = 0$ except if $x \in E_i, y \in E_{i+1}$ (modulo p) for some $i \leq p$.*

Schematically, possible transitions for Q are such that

$$E_1 \to E_2 \to \ldots \to E_p \to E_1.$$

[Reference??]

Theorem 6.5.6. *Let Q irreducible on E, positive recurrent with invariant law π. If Q is aperiodic, then for all chain X with transition Q,*

$$\mathbb{P}(X_n = x) \longrightarrow \pi(x) , \qquad x \in E,$$

as $n \to \infty$.

□ We will introduce a *coupling* (called the "double chain"): a realization on the same probability space of our original chain together with another chain Y at equilibrium (starting from π) and independent of X. For $x_0 \in E$, let

$$T = \inf\{n \geq 0; X_n = Y_n = x_0\}$$

1- First we show that $T < \infty$ a.s. To do that, note that the couple (X_n, Y_n) is a Markov chain on $E \times E$ with transition

$$\tilde{Q}((x, y), (x', y')) = Q(x, x')Q(y, y') .$$

Since Q is irreducible aperiodic, $Q^n(x, x') > 0$ for all $n \geq N(x, x')$ by proposition 6.5.4, and $Q^n(x, x')Q^n(y, y') > 0$ for all $n \geq N(x, x') \vee N(y, y')$: thus, the product chain $\tilde{Q}$ is irreducible. But the product law $\pi \otimes \pi$ is invariant, then (X_n, Y_n) is positive recurrent, it will a.s. reach the state (x_0, x_0), and $T < \infty$.

2- We consider the sequence

$$Z_n = \begin{cases} X_n , & \text{si} \quad n \leq T , \\ Y_n , & \text{si} \quad n > T , \end{cases}$$

By the strong Markov property for the couple (X, Y), we see that Z is a (μ_0, Q)-Markov chain (with μ_0 the law of X_0), i.e. $X \overset{\text{law}}{=} Z$. (Indeed, $(X_n)_{n \leq T}$ is independent of $(Y_{T+n})_{n \geq 0}$, which is a Markov starting from x_0.) Then,

$$\begin{aligned} |\mathbb{P}(X_n = y) - \pi(y)| &= |\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)| \\ &= |\mathbb{P}(Z_n = y, n < T) - \mathbb{P}(Y_n = y, n < T) \\ &\qquad + \mathbb{P}(Z_n = y, n \geq T) - \mathbb{P}(Y_n = y, n \geq T)| \\ &= |\mathbb{P}(Z_n = y, n < T) - \mathbb{P}(Y_n = y, n < T)| \\ &\leq \mathbb{P}(n < T) \\ &\longrightarrow 0 \end{aligned}$$

as $n \to \infty$ by the first step. ∎

Example 6.5.7. *The random walk with parameter $p \in (0,1)$ on $\mathbb{Z}/m\mathbb{Z}$ is aperiodic m if m is odd, and periodic if m is even. Then, if m is odd, the law of X_n converges weakly to the uniform law.*

Example 6.5.8 (Metropolis-Hastings algorithm)**.**

Let π be a probability on a discrete space E with $\pi(x) > 0$ for all $x \in E$. Consider $L \subset E \times E \setminus \{(x,x); x \in E\}$, such that $(x,y) \in L \iff (y,x) \in L$. The pair is an unoriented graph, and, in addition, we assume it is connected: for all $x, y \in E$ there is a finite path $x_0 = x, x_1, \ldots, x_n = y$ with $(x_{i-1}, x_i) \in L$. We denote by $d(x) = \sharp\{y \neq x : (x,y) \in L\}$ the degree of the vertex x, that we assume to be finite for all x. Fix a function

$$h : \mathbb{R}_+^* \to (0,1] \qquad \text{with} \quad h(u) = u \times h(1/u), \tag{6.11}$$

which covers the two main examples:

$$h(u) = \min\{u, 1\}\ , \qquad \text{or} \quad h(u) = \frac{u}{1+u}. \tag{6.12}$$

Then, the formula

$$Q(x,y) = \begin{cases} \dfrac{1}{d(x)} \times h\left(\dfrac{\pi(y)/d(y)}{\pi(x)/d(x)}\right) & \text{if} \quad (x,y) \in L, \\ 1 - \sum_{z:(x,z)\in L} \frac{1}{d(x)} h\left(\frac{\pi(z)/d(z)}{\pi(x)/d(x)}\right) & \text{if} \quad y = x, \\ 0 & \text{if} \quad (x,y) \notin L, \end{cases}$$

defines a stochastic matrix: the weights $Q(x,y)$ add up to 1, and they are nonengative by definition of the degree and since $h \leq 1$. The corresponding Markov chain is irreducible, aperiodic, and π is invariant (reversible) measure: for $x \neq y$,

$$\begin{aligned} \pi(x)Q(x,y) &= \frac{\pi(x)}{d(x)}\, h\left(\frac{\pi(y)/d(y)}{\pi(x)/d(x)}\right) \\ &= \pi(y)\frac{\pi(x)/d(x)}{\pi(y)}\, h\left(\frac{\pi(y)/d(y)}{\pi(x)/d(x)}\right) \\ &= \frac{\pi(y)}{d(y)}\, h\left(\frac{\pi(y)/d(y)}{\pi(x)/d(x)}\right) \qquad \text{(by (6.11))} \\ &= \pi(y)Q(y,x) \end{aligned}$$

for $(x,y) \in L$, and $\pi(x)Q(x,y) = \pi(y)Q(y,x) = 0$ if $(x,y) \notin L$. Thus, the chain is ergodic, and by Theorem 6.5.6, the law of X_t weakly converges to π.

Simulation by MCMC: *The above procedure originates a popular method to simulate an arbitrary discrete probability π, called Monté Carlo Markov*

Chain (abbreviated MCMC or MC^2), which well suited when π is easily computed up to a constant factor (note that the algorithm depends only on ratios of π). One runs the chain X starting from an arbitrary point for a "large" time, and uses the value of X_t as a random variable with "approximate distribution" π. This is justified by Theorem 6.5.6, since the law of X_t at a large time t is close to the target distribution π.

Using the above chain was suggested first by Metropolis et al. in 1953 to simulate Boltzmann's distribution, for which the normalizing constant is akward. Many other procedures have been introduced since then, including heat-bath or Gibbs sampler when E is a product space.

Chapter 7

Martingales and Markov chains

Let X be Markov with transition Q on E discrete.

7.1 Harmonic functions

Let $f : E \to \mathbb{R}$ with $E_x|f(X_1)| < \infty$ for all $x \in E$. Observe that, on the event $\{X_{n-1} = x\}$,

$$\mathbb{E}(f(X_n)|\mathcal{F}_{n-1}) = Qf(x)$$

Doob's decomposition of the process $(f(X_n))_n$ writes

$$f(X_n) = f(X_0) + M_n + A_n \ , \quad n \geq 0$$

with $(A_n)_n$ predictable and $(M_n)_n$ a martingale. Recalling that

$$A_n - A_{n-1} = \mathbb{E}(f(X_n) - f(X_{n-1})|\mathcal{F}_{n-1}), M_n - M_{n-1} = f(X_n) - \mathbb{E}(f(X_n)|\mathcal{F}_{n-1}),$$

we have

$$\begin{aligned} A_n &= \sum_{k=1}^{n} \big(Qf(X_{k-1}) - f(X_{k-1})\big) \ , \\ M_n &= \sum_{k=1}^{n} \big(f(X_k) - Qf(X_{k-1})\big) \ . \end{aligned} \tag{7.1}$$

Definition 7.1.1. *A function $f : E \to \mathbb{R}$ such that $E_x|f(X_1)| < \infty$ for all $x \in E$ is called*

- *harmonic (for Q) if $Qf = f$,*
- *superharmonic if $Qf \leq f$ (i.e., $Qf(x) \leq f(x), \forall x \in E$),*

- *subharmonic if $Qf \geq f$.*

In the first case, $(f(X_n))_n$ is a martingale, a supermartingale, resp. a submartingale, in the other cases.

Example 7.1.2. *For a Markov chain X on $\mathbb{Z}$, we define the* drift *at x*

$$b(x) = \mathbb{E}(X_1 - X_0 | X_0 = x)$$

and the one-step variance

$$\sigma^2(x) = \mathbb{E}([X_1 - X_0 - b(X_0)]^2 | X_0 = x) \ .$$

If the second one is finite,

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

is a martingale, with bracket

$$\langle M \rangle_n = \sum_{k=0}^{n-1} \sigma^2(X_k)$$

In fact, definition 7.1.1 is too restrictive, because usually in practice, we cannot find functions which are harmonic on the full space (except constant ones).

Proposition 7.1.3. *Let $f : E \to \mathbb{R}$ with $E_x|f(X_1)| < \infty$ for all $x \in E$, and such that there exists $A \subset E$ such that*

$$Qf(x) = f(x) \qquad [\text{resp. } Qf(x) \leq f(x), \text{resp. } Qf(x) \geq f(x)] \qquad \forall x \in A \ .$$

Then, f is called harmonic [resp. superharmonic, resp. subharmonic] on A. With $T_{A^c} = \inf\{n \geq 0 : X_n \in A^c\}$ the exit time of A, then the sequence M_n,

$$M_n = f\left(X_{n \wedge T_{A^c}}\right)$$

is a martingale [resp. supermart., resp. submart.].

□ Doob's decomposition writes $f(X_n) = A_n + N_n$ with a martingale N, so

$$M_n = N_{n \wedge T_{A^c}},$$

which is a martingale by the stopping theorem. The predictible part A_m is given by

$$A_m = \sum_{k=1}^{m} \left(Qf(X_{k-1}) - f(X_{k-1})\right),$$

where all summands are 0 for $m \leq T_{A^c}$ (in the first case; the two other cases are similar). This ends the proof. ∎

Example 7.1.4 (Symmetric random walk). *This is the Markov chain on $\mathbb{Z}$ with transition $Q(x, x \pm 1) = 1/2$. Then,*

$$Qf(x) = \frac{1}{2}(f(x+1) + f(x-1))$$

is the mean of f on the neighbors of x. So, a superharmonic function is larger or equal to its mean on neighbouring points, which justifies the name. Supermartingales were introduced after superharmonic functions, then they received the similar name, though it may look surprising from their definition.

Similarly, we compute the increment of the compensator of $f(X_n)$,

$$Qf(x) - f(x) = \frac{1}{2}(f(x+1) - 2f(x) + f(x-1)) =: \frac{1}{2}\Delta f(x)$$

with Δ the (discrete) Laplacian in one space dimension. All these considerations extend to the d-dimensional case when the jumps are uniform on the $2d$ neighbors of the current state.

Example 7.1.5 (Birth and death process). *We describe a population of particles subject to creation and annihilation. At each integer moment, the population size X_t increases or decreases by one unit with respective probabilities $p_x, 1 - p_x$ if the current value is x. (At $x = 0$ only creation occurs, i.e., $p_0 = 1$.) Then, X is a Markov chain, with transition*

$$Q(x, y) = \begin{cases} p_x & \text{if} \quad y = x + 1 \\ 1 - p_x & \text{if} \quad y = x - 1 \\ 0 & \text{if} \quad |y - x| \neq 1. \end{cases}$$

If $p_x \in (0, 1)$ for all $x \geq 1$, the chain is irreducible, with period 2. We have

$$Qf(x) = p_x f(x+1) + (1 - p_x) f(x-1),$$

and for all $f : \mathbb{N} \to \mathbb{R}$,

$$M_n = f(X_n) - \sum_{t=0}^{n-1} \{p_{X_t}[f(X_t+1) - f(X_t)] + (1 - p_{X_t})[f(X_t-1) - f(X_t)]\}$$

is a L^2-martingale.

Example 7.1.6 (Polya's urn continued). *An urn contains balls with m different colors. At each integer time we select at random one ball in the urn, and we replace it with an extra ball of the same color. A model for emergence of standards in a decentralised economy: m standards are competing, each new client adopts the same choice as one of his friend, that we view as randomly picked in the full population.*

Let $\sum_{i=1}^m N_0^i = n_0$, and let $X_t^i = N_t^i/(n_0+t)$ denote the proportion of balls of colour i at time t. N_t^i evolves according to

$$N_{t+1}^i = N_t^i + \begin{cases} 0 & \text{avec probabilite} \quad 1 - X_t^i \\ 1 & \text{avec probabilite} \quad X_t^i \end{cases}$$

independently of everything else. The m-vector N_t is a Markov chain, which is time-inhomogeneous*: the transition at time t is, with $(e_i)_{i\le m}$ the canonical basis of $\mathbb{R}^m$,*

$$Q_t(x, x+e_i) = 1 - Q_t(x,x) = \frac{x_i}{n_0+t}, \qquad i = 1, \dots, m,$$

and $Q_t(x,y) = 0$ for $y \neq x, x+e_1, \dots, x+e_m$. For each $i = 1, \dots m$, $(N_t^i)_t$ is itself a Markov chain. $(X_t^i)_{t\ge 0}$ is a bounded martingale.

Set $N_t = N_t^1$ to simplify. Check that

$$Z_t^{(2)} = \frac{N_t^1(N_t^1+1)}{(n_0+t)(n_0+t+1)} \ ,$$

and more generally, for all k,

$$Z_t^{(k)} = \frac{N_t^1(N_t^1+1)\dots(N_t^1+k-1)}{(n_0+t)(n_0+t+1)\dots(n_0+t+k-1)} \ ,$$

are bounded martingales.

7.2 Potential theory

7.2.1 Potential and Green function in the transient case

The *potential matrix* or the *Green function* $G = (G(x,y))_{x,y\in E}$ is

$$G = I + Q + Q^2 + \dots \tag{7.2}$$

with coefficients in $[0,\infty]$. By Fubini,

$$G(x,y) = \sum_{n\ge 0} Q^n(x,y) = E_x N_y$$

with $N_y = \sum_{n\ge 0} \mathbf{1}_{X_n = y}$ the number of visits in y. As seen in the proof of proposition 6.1.8, we have

$$G(x,y) = P_x(T_y^- < \infty) G(y,y) \begin{cases} = \infty & \text{for } y \text{ recurrent} \\ < \infty & \text{for } y \text{ transient} \end{cases}$$

with $T_y^- = \inf\{t \geq 0 : X_t = y\}$, the first passage time in y.(1) We extend definition 7.1.1 in the case of non negative functions: we say that $h : E \to [0, \infty]$ is superharmonic if $Qh \leq h$.

Definition and Proposition 7.2.1 (Potential and charge)**.** *For all $g : E \to [0, \infty]$, the function Gg is superhamonic. It is called the potential of g,*

$$Gg(x) = E_x \sum_{n \geq 0} g(X_n)$$

and g is called the charge.

□ Write

$$QGg = Q \sum_{n \geq 0} Q^n g \overset{\text{Fubini}}{=} \sum_{n \geq 1} Q^n g \leq Gg \tag{7.3}$$

which proves the statement. ■

Formally, $(I - Q)G = I$, " so $G = (I - Q)^{-1}$ ". However, since $Q\mathbf{1} = \mathbf{1}$, $\lambda = 1$ is eigenvalue of Q, and so the last formula requires an appropriate definition of the inverse.

Proposition 7.2.2. *$\forall g \geq 0$, Gg is the smallest function $f : E \to [0, \infty]$ solution of* Poisson's equation*(2)* $g = f - Qf$*, which correct writing (because of possibly infinite values) est*

$$g + Qf = f \tag{7.4}$$

□ We saw that Gg solves Poisson's equation, cf. first and third members of (7.3). Conversely, if f solves (7.4), we get by iteration

$$f = g + Qg + Q^2 g + \ldots + Q^n g + Q^{n+1} f \ ,$$

and then $f \geq Qg$ by positivity of the last term. ■

In potential theory there is the famous decomposition of Frigyes Riesz(3).

---

(1)This definition is interesting only for a transient chain; Otherwise it holds $G(x, x) = \infty = G(x, y)$ making this function trivial. We can suppress this problem by making one part of the space absorbing -then we abandon irreducibility- or by killing it at rate $\lambda > 0$ (cf. section 7.2.2), or by compensating the divergence.

(2)Recall that in the case when X is the simple random walk on $\mathbb{Z}^d$, $Q - I$ is the discrete Laplacian, and then (7.4) is the discrete version of

$$-(2d)^{-1} \Delta f = g,$$

that is, the standard Poisson's equation from electrostatics. Cf. example 7.1.4.

(3)1880-1956, also named Frederic.

Proposition 7.2.3 (Riesz's decomposition)**.** *Every finite non-negative superharmonic function f has a unique decomposition*

$$f = h + Gg$$

with h non-negative harmonic and g non-negative. Moreover, for all $x \in E$,

$$\lim_{n\to\infty} f(X_n) = \lim_{n\to\infty} h(X_n)$$

exist P_x-a.s. and are equal, and

$$g = f - Qf \tag{7.5}$$

□ Since $0 \le Qf \le f < \infty$ for any finite non-negative superharmonic function f, formula (7.5) is meaningful and it defines a non-negative, finite g such that

$$\sum_{k=0}^{n-1} Q^k g = f - Q^n f \ .$$

As $n \to \infty$, the LHS converges to Gg, so the RHS has a limit,

$$h = \lim_{n\nearrow\infty} \searrow Q^n f \ , \qquad Gg = f - h \ . \tag{7.6}$$

The limit h is non-negative, finite, and harmonic since

$$h = \lim_n Q^n f = \lim_n Q^{n+1} f = \lim_n QQ^n f = Q \lim_n Q^n f = Qh$$

where the fourth equality comes from dominated convergence (dominated by f). For all x, $f(X_n), Gg(X_n), h(X_n)$ are non-negative P_x-supermartingales, they converge a.s. It remains to show $Gg(X_n) \to 0$ a.s. But

$$\begin{aligned} E_x Gg(X_n) = Q^n Gg(x) \quad &= \quad \sum_{m\ge n} Q^m g(x) \\ &\xrightarrow[n\to\infty]{} \quad 0 \end{aligned}$$

because the series $Gg(x) = \sum_{m\ge1} Q^m g(x)$ converges ($Gg(x) < \infty$). Thus $Gg(X_n) \to 0$ in $L^1(P_x)$, and its a.s.-limit is necessarily 0. ∎

Here is the fundamental example.

Example 7.2.4 (Hitting times and number of visits.)**.** *Let A be a subset of E, and*

$$T_A = \inf\{t \ge 0 : X_t \in A\} \ , \qquad N_A = \sum_{t\ge0} \mathbf{1}_{X_t\in A} \ .$$

Define

$$e_A(x) = P_x(T_A < \infty) \ , \qquad h_A(x) = P_x(N_A = \infty) \ . \tag{7.7}$$

Then, the function e_A is superharmonic with values in $[0,1]$, the function h_A is its harmonic component in Riesz's decomposition, $e_A = h_A + Gg_A$,

$$\lim_n h_A(X_n) = \lim_n e_A(X_n) = \mathbf{1}_{N_A=\infty} .$$

The function g_A is given by the "escape probability"

$$g_A(x) = \begin{cases} P_x(X_n \notin A, \forall n \geq 1) & \text{if } x \in A , \\ 0 & \text{if } x \notin A . \end{cases}$$

Before proving the claim, let's make the example more specific. Consider the case of the nearest neighbor random walk on $E = \mathbb{Z}^d$ for $d \geq 3$: $Q(x,y) = 1/(2d)$ if $|y-x|_1 = 1$, and $Q(x,y) = 0$ otherwise. Let A be a finite subset of E, say

$$A = [-m, m]^d$$

for concreteness. Then, by Polya's theorem, X is transient, so $h_A \equiv 0$ in (7.7), which implies that $e_A = Gg_A$ is a potential. It is the *electrostatic potential with prescribed value 1 on A* in a resistor network with unit conductance on nearest neighbor edges, as described in Example 6.3.14. The charge g_A creating this potential is concentrated on the "internal boundary" of A, that is on the $2d$ faces of the cube A,

$$x \in \text{support } g_A \iff |x|_\infty = m$$

In fact, the potential e_A is harmonic on A^c and in the "interior" of A, but superharmonic on the "internal boundary" of A.

$$\boxed{\begin{array}{rcl} \text{electrostatic potential (prescribed value 1 on } A) & = & P_x(T_A < \infty), \\ \text{charge on internal boundary of A} & = & P_x(\text{Escape from } A) \end{array}}$$

□ By Markov property,

$$\begin{aligned} Qe_A(x) &= \sum_y Q(x,y) e_A(y) \\ &= P_x(\exists t \geq 1, X_t \in A) \\ &\leq P_x(\exists t \geq 0, X_t \in A) \\ &= e_A(x) , \end{aligned}$$

so e_A is superharmonic. Also,

$$\begin{aligned} Q^n e_A(x) &= P_x(\exists t \geq n, X_t \in A) \\ &\searrow P_x(N_A = \infty) , \qquad \text{as } n \to \infty, \end{aligned}$$

which (cf. (7.6)) is the function h_A ! Now, if $N_A^{(n)}$ denotes the number of visits to A after time n, the event $\{N_A^{(n)} = \infty\}$ does not depend on n,

$$\begin{array}{rcll} h_A(X_n) & = & P_{X_n}(N_A = \infty) & \\ & = & P(N_A^{(n)} = \infty | \mathcal{F}_n) & \text{(Markov property)} \\ & = & P(N_A = \infty | \mathcal{F}_n) & \\ & \longrightarrow & P(N_A = \infty | \mathcal{F}_\infty) & (n \to \infty) \\ \lim_n h_A(X_n) & = & \mathbf{1}_{N_A = \infty} & \text{a.s.} \end{array}$$

by martingale convergence and noting that $\{N_A = \infty\} \in \mathcal{F}_\infty$. Finally,

$$\begin{array}{rcl} g_A(x) & = & e_A(x) - Qe_A(x) \\ & = & P_x(\exists t \geq 0, X_t \in A) - P_x(\exists t \geq 1, X_t \in A) \\ & = & P_x(X_0 \in A; \forall t \geq 1, X_t \notin A), \end{array}$$

ending the proof. ∎

Notion of corrector: given $f : E \to \mathbb{R}$, it is natural to look for a function $h : E \to \mathbb{R}$ such that

$$\begin{cases} h(X_n) \quad \text{is a martingale,} \\ f(X_n) - h(X_n) = o(1 + |f(X_n)|), \quad n \to \infty. \end{cases}$$

Its existence is not granted. When it exists, the difference $f - h$ is called the corrector f, for it corrects f into an harmonic function, without affecting its behavior at infinity. Riesz's decomposition provides such a function h, in the form $h(x) = \lim_{n\to\infty} E_x f(X_n)$; the next choice, depending on some origin $x_0 \in E$,

$$h(x) = \lim_{n\to\infty} [E_x f(X_n) - E_{x_0} f(X_n)]$$

is even better, allowing to suppress divergences. However, the existence of the limit is not clear in general.

7.2.2 Resolvant

What is potential theory in the recurrent case ? In short, the strategy is (i) to kill the chain at a positive rate λ, which forces it to be transient, (ii) to apply the above for the killed chain, (iii) to come back to our chain by letting $\lambda \searrow 0$.

For $\lambda > 0$ define the *resolvant matrix* $G_\lambda = (G_\lambda(x,y))_{x,y \in E}$ with coefficients

$$G_\lambda(x,y) = \sum_{n \geq 0} e^{-\lambda n} Q^n(x,y) \in [0, (1 - e^{-\lambda})^{-1}].$$

Since the series converge, we can write

$$G_\lambda = \sum_{n\geq 0} (e^{-\lambda}Q)^n = (I - e^{-\lambda}Q)^{-1}.$$

The sum of the series $G_\lambda(x,y)$ represents, for the chain killed at rate λ (i.e., with probability $1 - e^{-\lambda}$ at each step, independently of the jumps) the mean number of visits to y starting from x. Define also $\bar{G}_\lambda(x,y) = \sum_{n\geq 1} e^{-\lambda n} Q^n(x,y)$, so that $I + \bar{G}_\lambda = G_\lambda$. By Fubini,

$$\bar{G}_\lambda(x,y) = E_x \sum_{n\geq 1} e^{-\lambda n} \mathbf{1}_{X_n = y} = E_x \sum_{k\geq 1} e^{-\lambda T_y^k}$$

with $T_y^1 = \inf\{n \geq 1 : X_n = y\}, T_y^2 = \inf\{n \geq T_y^1 + 1 : X_n = y\}, , \ldots$ the successive passage times at y. By strong Markov property with the stopping time $T_y = T_y^1$, we have

$$\begin{aligned} \bar{G}_\lambda(x,y) &= E_x e^{-\lambda T_y} \big[1 + E_y \sum_{k\geq 1} e^{-\lambda T_y^k}\big] \\ &= E_x \big[e^{-\lambda T_y}\big](1 + \bar{G}_\lambda(y,y)) \end{aligned}$$

Thus, the Laplace transform of T_y is given by

$$E_x\big[e^{-\lambda T_y}\big] = \frac{\bar{G}_\lambda(x,y)}{1 + \bar{G}_\lambda(y,y)}$$

By monotone convergence, $\lim_{\lambda \searrow 0} G_\lambda(x,y) = G(x,y)$, whereas $\lim_{\lambda\searrow 0} E_x e^{-\lambda T_y} = P_x(T_y < \infty)$. In this way we recover that $P_x(T_x < \infty) = 1$ if and only if $G(x,x) = \infty$; And also that $G(y,y) < \infty$ implies $G(x,y) < \infty$, and then $P_x(T_y < \infty) = \bar{G}(x,y)/\bar{G}(y,y)$.

We stop here our discussion.

7.2.3 Green function in the recurrent case

In the recurrent case, the series (7.2) is infinite, so the Green function requires another definition.

Definition and Proposition 7.2.5. *Let Q be positive recurrent and irreducible, and assume also that the* Doeblin condition

$$\exists x_0 \in E : \quad \delta = \inf\{Q(x,x_0); x \in E\} > 0 \tag{7.8}$$

holds. Then, fixing an arbitrary $a \in E$, the series

$$G(x,y) = \sum_{n=0}^{\infty} \Big[Q^n(x,y) - Q^n(a,y)\Big] \tag{7.9}$$

converges in $\mathbb{R}$ *for all* $x, y \in E$*, and the sum* G *is called the* Green function*.* *If* $g : E \to \mathbb{R}$ *is bounded and centered for the invariant probability* π *of the chain*

$$\pi g \stackrel{\text{def.}}{=} \sum_x g(x)\pi(x) = 0,$$

the series $Gg(x) = \sum_y G(x,y)g(y)$ *is absolutely convergent, and the sum* Gg *is bounded and solves the* Poisson equation*:*

$$(I - Q)\, Gg = g\ .$$

Observe that Doeblin condition (7.8) implies aperiodicity, since the chain can stay put at x_0.

$\Box$ We follow the coupling argument of the proof of Theorem 6.5.6. The double chain (X_n, Y_n) is the Markov chain on $E \times E$ with transition

$$\tilde{Q}\big((x,y),(x',y')\big) = Q(x,x')Q(y,y')\ ,$$

it is irreducible, positive recurrent with invariant law $\pi \otimes \pi$. We start from $X_0 = x, Y_0 = a$. With x_0 from Doeblin condition (7.8), let

$$T = \inf\{n \geq 0; X_n = Y_n = x_0\}.$$

Then, T is a $\mathcal{F}_n = \mathcal{F}_n^{X,Y}$-stopping time, and

$$\begin{aligned}
\mathbb{P}(T > n+1) &= \mathbb{E}\Big(\mathbf{1}_{T>n}\mathbb{P}((X_{n+1}, Y_{n+1}) \neq (x_0,x_0)|\mathcal{F}_n)\Big) \\
&= \mathbb{E}\Big(\mathbf{1}_{T>n}(1 - Q(X_n,x_0)Q(Y_n,x_0))\Big) \qquad \text{(Markov property)} \\
&\leq \mathbb{E}\Big(\mathbf{1}_{T>n}(1-\delta^2)\Big) \qquad \text{(by (7.8)} \\
&\leq (1-\delta^2)^{n+1} \qquad (7.10)
\end{aligned}$$

by induction.

We consider now the coalescing coupling (Z_n, Y_n) with the sequence

$$Z_n = \begin{cases} X_n\ , & \text{if} \quad n \leq T\ , \\ Y_n\ , & \text{if} \quad n > T\ . \end{cases}$$

By the strong Markov property, we see that Z is a Markov chain starting at x with transition Q, i.e. $X \stackrel{\text{law}}{=} Z$. Then,

$$\begin{aligned}
|Q^n(x,y) - Q^n(a,y)| &= |\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)| \\
&= |\mathbb{P}(Z_n = y, n < T) - \mathbb{P}(Y_n = y, n < T) \\
&\qquad + \mathbb{P}(Z_n = y, n \geq T) - \mathbb{P}(Y_n = y, n \geq T)| \\
&= |\mathbb{P}(Z_n = y, n < T) - \mathbb{P}(Y_n = y, n < T)| \quad (7.11) \\
&\leq \mathbb{P}(n < T) \\
&\leq (1-\delta^2)^n
\end{aligned}$$

by (7.10), so that the series (7.9) is finite. From (7.11) we even have

$$\begin{aligned}\sum_{y\in E}|Q^n(x,y)-Q^n(a,y)| &\leq \sum_{y\in E}[\mathbb{P}(Z_n=y,n<T) \\ &\qquad +\mathbb{P}(Y_n=y,n<T)] \\ &\leq 2(1-\delta^2)^n. \end{aligned} \tag{7.12}$$

In particular,

$$|Gg(x)| \leq \sum_y\sum_n |Q^n(x,y)-Q^n(a,y)||g(y)| \leq 2\delta^{-2}\|g\|_\infty,$$

and Gg is bounded. Now we compute

$$\begin{aligned}(I-Q)\,Gg &= (I-Q)\lim_{N\to\infty}\sum_{n=0}^{N}\Big[Q^n g - Q^n g(a)\Big] \\ &= \lim_{N\to\infty}(I-Q)\sum_{n=0}^{N}\Big[Q^n g - Q^n g(a)\Big] \qquad \text{(Lebesgue)} \\ &= \lim_{N\to\infty}\sum_{n=0}^{N}(I-Q)\Big[Q^n g - Q^n g(a)\Big] \\ &= \lim_{N\to\infty}\sum_{n=0}^{N}(I-Q)Q^n g \qquad \text{(since } (I-Q)\mathbf{1}=0) \\ &= \lim_{N\to\infty}(g-Q^{N+1}g) \qquad \text{(telescopic sum)} \\ &= g-\pi g \qquad \text{(by Theorem 6.5.6)} \\ &= g\end{aligned}$$

since g is centered. ∎

Remark 7.2.6. *(i) Example 5.3.2 of the house of cards satisfies Doeblin condition with $x+0=0$ and $\delta=1-p$.*

(ii) We see in the above proof that, even though the Green function G obviously depends on the state $a\in E$ arbitrarily fixed in the definition (7.9), the product $(I-Q)G$ acting on centered functions does not.

7.3 Central limit theorem for ergodic Markov chains

For such a chain and $g:E\to\mathbb{R}$ bounded, the ergodic theorem tells us that

$$\frac{1}{n}\sum_{i=1}^{n}g(X_i)\longrightarrow \pi g, \qquad P_x-a.s.$$

The next question is to look at the fluctuations. By considering $g-\pi g$, we can and we will assume without loss of generality that g is centered for π. We start with an easy observation.

Proposition 7.3.1. *For all $f : E \to \mathbb{R}$ bounded and all $x \in E$, we have under P_x,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[f(X_i) - Qf(X_{i-1}) \Big] \xrightarrow{\text{law}} \mathcal{N}(0, \sigma^2)$$

with

$$\sigma^2 = \pi(f^2) - \pi((Qf)^2).$$

□ In order to use the central limit theorem for martingales, we consider

$$M_n = \sum_{i=1}^{n} Y_i \ , \qquad \text{with } Y_i = f(X_i) - Qf(X_{i-1}).$$

The increments Y_n are conditionally centered, and M_n is a $\mathcal{F}$-martingale, with $\mathcal{F}_n = \mathcal{F}_n^X$. The increments are bounded, $|Y_i| \leq 2\|f\|_\infty$. The conditional variance is

$$\mathbb{E}(Y_i^2 | \mathcal{F}_{i-1}) = v(X_{i-1}), \qquad v(x) = Q(f^2)(x) - \big(Qf\big)^2(x),$$

so

$$\langle M \rangle_n = \sum_{i=1}^{n} v(X_{i-1}). \tag{7.13}$$

By the ergodic theorem,

$$\frac{1}{n} \langle M \rangle_n \longrightarrow \pi v, \quad P_x - a.s.,$$

and also

$$\frac{s_n}{n} := \frac{1}{n} E_x(M_n^2) \longrightarrow \pi v.$$

We finally note that

$$\sigma^2 = \pi v,$$

since

$$\pi v = \pi Q(f^2) - \pi \Big[\big(Qf\big)^2 \Big] = \pi(f^2) - \pi \Big[\big(Qf\big)^2 \Big] = \sigma^2$$

by invariance of π. Two cases can happen:

(i) If $\sigma^2 > 0$, we can apply the central limit theorem for martingales, leading to

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} \Big[f(X_i) - Qf(X_{i-1}) \Big] \xrightarrow{\text{law}} \mathcal{N}(0, 1),$$

which is our claim;

(ii) If $\sigma^2 = 0$, then the claim trivially holds, since $M_n/\sqrt{n} \to 0$ in L^2. ∎

Remark 7.3.2. *We show that the second case can happen. Let $E = \{0, 1\}$, and*

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The chain is irreducible with invariant law $\pi = (1/2, 1/2)$ uniform. So it is ergodic, and Proposition 7.3.1 applies. We claim that $\sigma = 0$ for any choice of f. Indeed, the chain is "not very random", in the sense that X_n is determined by X_0: it is purely a dynamical system! In particular, $f(X_n) = Qf(X_{n-1})$, and so $M_n = 0$.

The previous result is not what we ask for in the introduction. The next one gives the answer to the question.

Theorem 7.3.3. *Let Q be positive recurrent and irreducible, with the Doeblin condition*

$$\exists x_0 \in E: \quad \delta = \inf\{Q(x, x_0); x \in E\} > 0.$$

Let $g : E \to \mathbb{R}$ bounded and centered for π. Then Under P_x,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) \xrightarrow{\text{law}} \mathcal{N}(0, \sigma_1^2)$$

with some $\sigma_1^2 \geq 0$.

□ By Proposition 7.2.5, we can solve the Poisson equation for g, and find an f such that

$$\Big(I - Q\Big) f = g.$$

In fact, $f = Gg$. Then,

$$\sum_{i=1}^{n} g(X_i) = \sum_{i=1}^{n} \Big[f(X_i) - Qf(X_{i-1})\Big] + Qf(X_n) - Qf(X_0),$$

where the last two terms are bounded in probability. By Proposition 7.3.1, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[f(X_i) - Qf(X_{i-1})\Big] + o_P(1) \xrightarrow{\text{law}} \mathcal{N}(0, \sigma_1^2),$$

with

$$\sigma_1^2 = \pi\big(Gg\big)^2 - \pi\big(QGg\big)^2,$$

ending the proof(4). ∎

(4) To keep notations compact, we use short notations, i.e.,

$$\pi\big(Gg\big)^2 = \sum_x \pi(x)\big(\sum_y G(x, y) g(y)\big)^2.$$

Remark 7.3.4. *Here is an example where $\sigma_1^2 = 0$. Let $E = \mathbb{Z}$, and Q a transition with*

$$\begin{cases} Q(0,x) = 0, & \text{for} \quad x \leq 0, \\ Q(0,x) > 0, & \text{for} \quad x > 0, \\ Q(x,-x) = 1 & \text{for} \quad x > 0, \\ Q(-x,0) = 1 & \text{for} \quad x < 0. \end{cases}$$

The chain is clearly irreducible, but also ergodic. Indeed, starting from 0 the chain is at 0 at all times 3, 6, 9,... multiple of 3. So $x = 0$ is recurrent – so the chain is recurrent – and $E_0 T_0 = 3 < \infty$ – so the chain is positive recurrent – . By Kac's lemma, $\pi(0) = 1/3$. Similarly, we easily see that

$$\forall x \neq 0, \qquad \pi(x) = \frac{1}{3} Q(0, |x|) \ .$$

The function $g(x) = x$ is odd and then centered for the invariant law. Under P_0, we have

$$\sum_{i=1}^{n} g(X_i) = 0 \quad \text{for } n = 0, 2 \text{ modulo } 3,$$

though

$$\left(\sum_{i=1}^{3n+1} g(X_i); n \geq 0 \right) \quad \text{is an i.i.d. sequence}$$

with law $Q(0, \cdot)$. So the sequence of partial sums $(\sum_{i=1}^{n} g(X_i), n \geq 1)$ is bounded in probability, and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) \xrightarrow{\mathbb{P}} 0,$$

which implies that the central limit theorem holds with $\sigma_1^2 = 0$.

Of course, in this example Doeblin condition is not satisfied, but at least we see that there is a problem of strict positivity of the variance in the central limit theorem even though there is some randomness in this example.

Example 7.3.5 (Asymptotic normality of estimates of transitions of a Markov chain)**.** *Consider X an irreducible ergodic Markov chain. The question we address is:*

```
How can we estimate
the invariant law $\pi$ and the transition probability $Q$
by observing the path during $n$ steps ?
```

Define

$$N_n(x,y) = \sum_{t=1}^{n} \mathbf{1}_{X_{t-1}=x, X_t=y} \ , \qquad N_n(x) = \sum_{t=1}^{n} \mathbf{1}_{X_{t-1}=x}$$

the number of jumps from x to y and the number of visits to x by time n. By ergodic theorem,

$$\widehat{\pi_n}(x) := \frac{N_n(x)}{n} \to \pi(x) \qquad a.s.,$$

and the ergodic theorem (for the chain $Z_n = (X_n, X_{n-1})$) yields

$$\frac{N_n(x,y)}{n} \to \pi(x)Q(x,y) \qquad a.s.$$

Therefore,

$$\widehat{Q_n}(x,y) := \frac{N_n(x,y)}{N_x} \to Q(x,y) \qquad a.s.$$

and the both estimators $\widehat{\pi_n}, \widehat{Q_n}$ approach the invariant law and the transition as the sample size increases. Furthermore, applying Proposition 7.3.1 we get

$$\frac{1}{\sqrt{n}}\Big[N_n(x,y) - N_n(x)Q(x,y)\Big] \xrightarrow{\text{law}} \mathcal{N}\Big(0, \pi(x)Q(x,y)[1-Q(x,y)]\Big).$$

Combining this with the above consequence of the ergodic theorem, we deduce

$$\sqrt{n\pi(x)}\Big[\widehat{Q_n}(x,y) - Q(x,y)\Big] \xrightarrow{\text{law}} \mathcal{N}\Big(0, Q(x,y)[1-Q(x,y)]\Big)\ .$$

From this, one derive a χ^2 test and other asymptotic statistical procedures.

Chapter 8

Bibliography

M. Benaïm, N. El Karoui: Promenade aléatoire. École Polytechnique. 2004

D. Williams: Probability with martingales. Cambridge University Press, 1991

And also:

R. Durrett: Probability: Theory and Examples. Duxbury 1995

G. Grimmett, D. Stirzaker: One thousand exercises in probability. Oxford University Press, 1992

G. Lawler, V. Limic: Random walk: a modern introduction. Cambridge University Press, 2010

J. Norris: Markov chains. Cambridge University Press, 1998

E. Pardoux: Processus de Markov et applications. Algorithmes, réseaux. Dunod, 2007