D-MATH, D-PHYS

Probability Theory

Fall Semester 2017

Prof. Dr. Alain-Sol Sznitman

(Lecture notes of P. Nolin based on previous lecture notes in German of A.S. Sznitman)

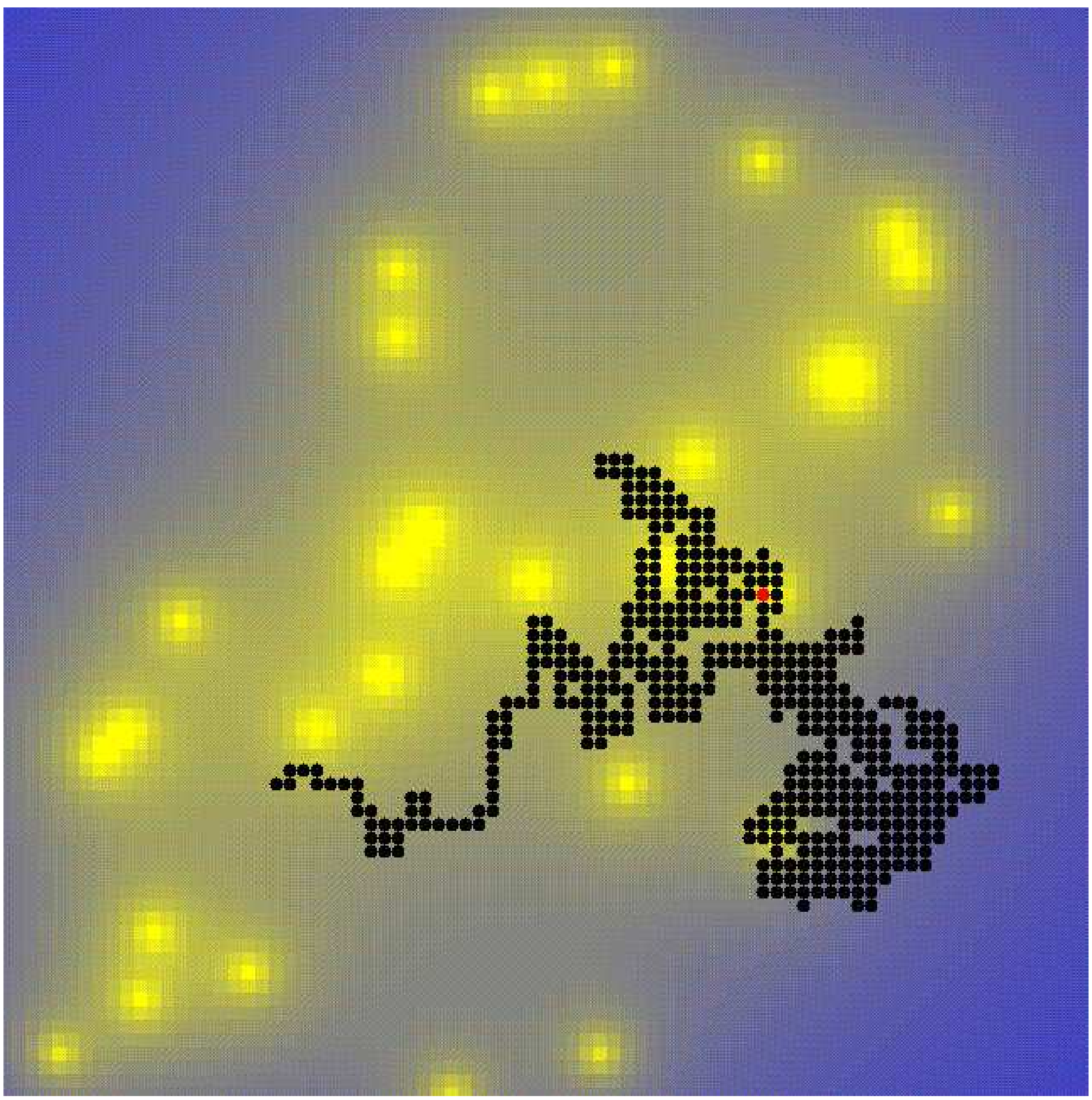


Table of Contents

0 Introduction

The goal of this course is mostly to study **stochastic processes in discrete time**, i.e. infinite sequences $X_0,\ X_1,\ X_2, \ldots,\ X_n, \ldots$ of random variables. Usually, $0, 1, 2, \ldots$ play the role of time.

In particular, we would like to introduce the **notions** and **tools** allowing to describe such stochastic processes. Very often, we will be interested in the **asymptotic behavior** of the sequence $X_0, X_1, \ldots, X_n, \ldots$

Example 0.1. Series with stochastic coefficients: One can verify that

$$\overline{X}_n = \sum_{k=1}^{n} \frac{(-1)^k}{k} \underset{n\to\infty}{\longrightarrow} -\log 2, \qquad \text{while}$$
$$\widetilde{X}_n = \sum_{k=1}^{n} \frac{1}{k} \underset{n\to\infty}{\longrightarrow} \infty \quad \text{(no absolute convergence)}\,.$$

What happens if we do not choose the preceding sign to be $(-1)^k$, but random instead?

$$X_n = \sum_{k=1}^{n} \frac{Z_k}{k}\,, \quad \begin{array}{l}\text{where } Z_1, Z_2, \ldots \text{ are independent, with} \\ P[Z_i = 1] = P[Z_i = -1] = \frac{1}{2}\,.\end{array}$$

How does one determine, in general, whether such stochastic series converge or not? ◯

Probability theory is a relatively new mathematical subject (Kolmogorov's axioms – 1933), even if questions related to it were considered quite early (Bernoulli, Fermat, Pascal – 17th century).

However, probability theory has numerous connections with other fields of pure mathematics, as well as with applications.

Example 0.2.

1) *Connection with Partial Differential Equations:*

We consider the Simple Random Walk (SRW) on the two-dimensional square grid $\mathbb{Z}^2$, with starting point $x \in \mathbb{Z}^2$, $X_0 = x, X_1, X_2, \ldots, X_n, \ldots$

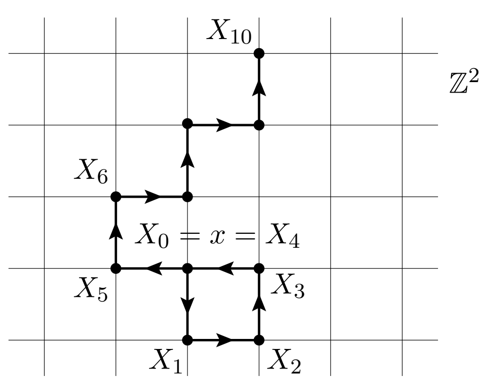


Fig. 0.1: A possible trajectory of the random walk

(the random walk makes successive "independent" jumps, each time to one of its four neighbors). Let $g(\cdot)$ be a continuous function on $\mathbb{R}^2$, and D a disk with center 0 and radius R.

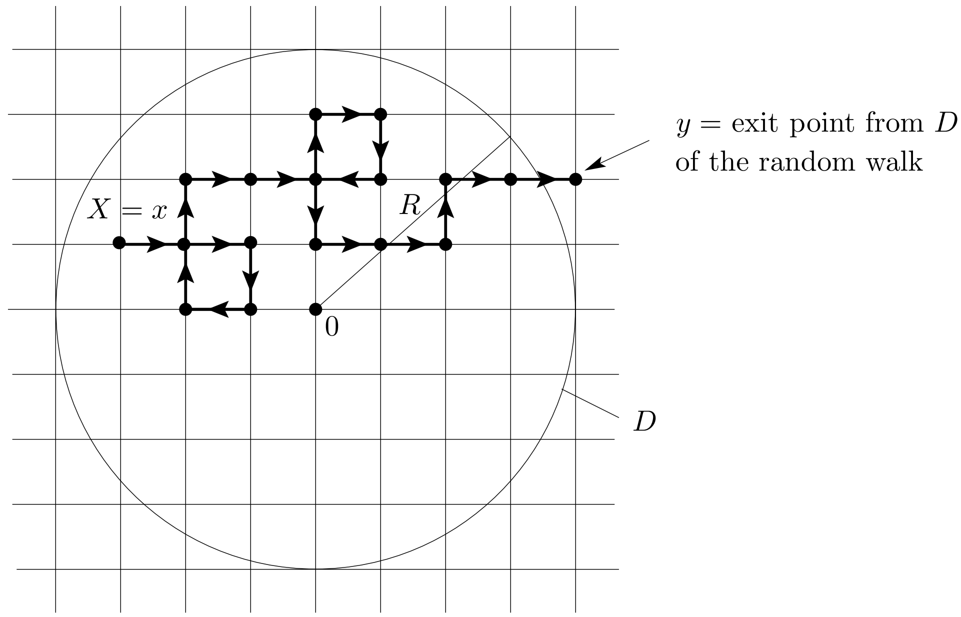


Fig. 0.2

One can consider the exit point of the random walk from D (i.e. the first point outside D that is visited by the random walk), and then define the function

$$u(x) = E[\underset{\substack{\uparrow \\ \text{exit point from } D \text{ of the random walk with starting point } x\,.}}{g(Y)}], \quad x \in \mathbb{Z}^2.$$

It can be proved that u is a **solution of the discrete Dirichlet problem**:

$$\begin{cases} \Delta_{\text{disc}}\, u(x) \;=\; 0\,, & x \in \mathbb{Z}^2 \cap D\,, \\ \qquad u(x) \;=\; g(x)\,, & x \in \mathbb{Z}^2 \cap D^c\,, \end{cases}$$

where

$$\Delta_{\text{disc}}\, u(y) = \frac{1}{4}\big(u(y+e_1) + u(y+e_2) + u(y-e_1) + u(y-e_2)\big) - u(y)$$

is the so-called discrete Laplacian, and

$$e_1 = (1,0), \quad e_2 = (0,1)\,.$$

2) *Connection with financial mathematics:*

The exchange rate of a currency (e.g. US dollar / Swiss Franc) can often be seen as a random process, and thus be modeled accurately with stochastic processes.

The computation of "call options" is for instance a successful application of methods from probability theory, in particular the theory of martingales (Chapter 4).

For example, someone can obtain, through the purchase of a "call option", the right to buy in two months 100 US dollars at a price of SFr. 95.

If the exchange rate of the US dollar is lower than SFr. 0.95 in two months, then he will of course not exercise his right. On the contrary, if the rate happens to be higher than SFr. 0.95, then his contract allows him to buy 100 US dollars at a price of SFr. 95.

The question is now: what is the fair price for such a contract? Here, methods from the theory of martingales give an answer to such questions in certain cases. $\bigcirc$

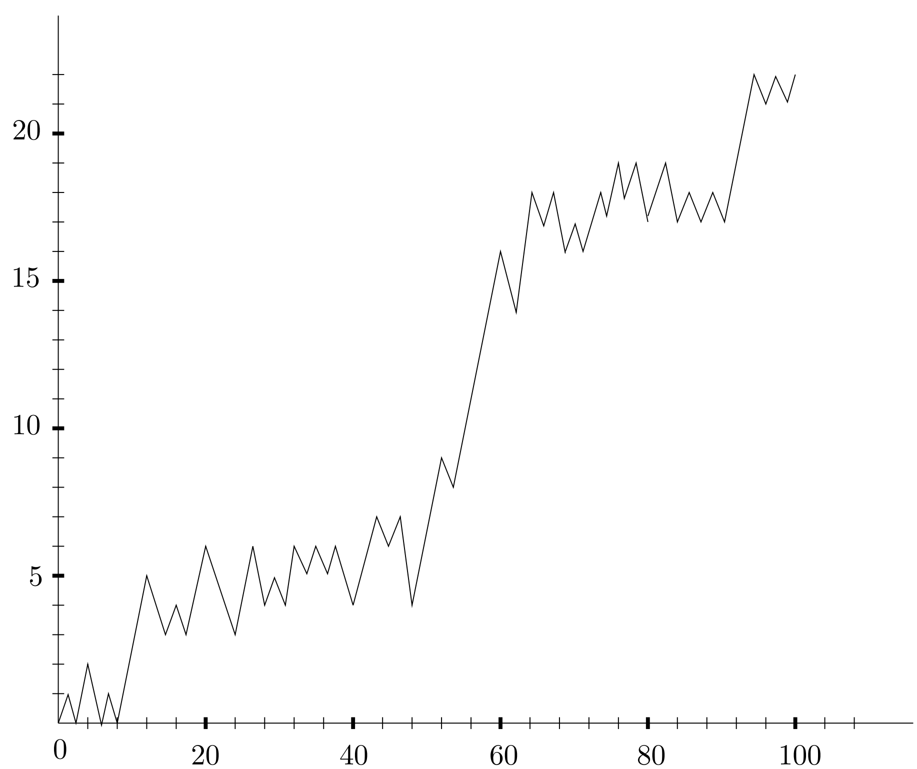


Fig. 0.3: Simple Random Walk

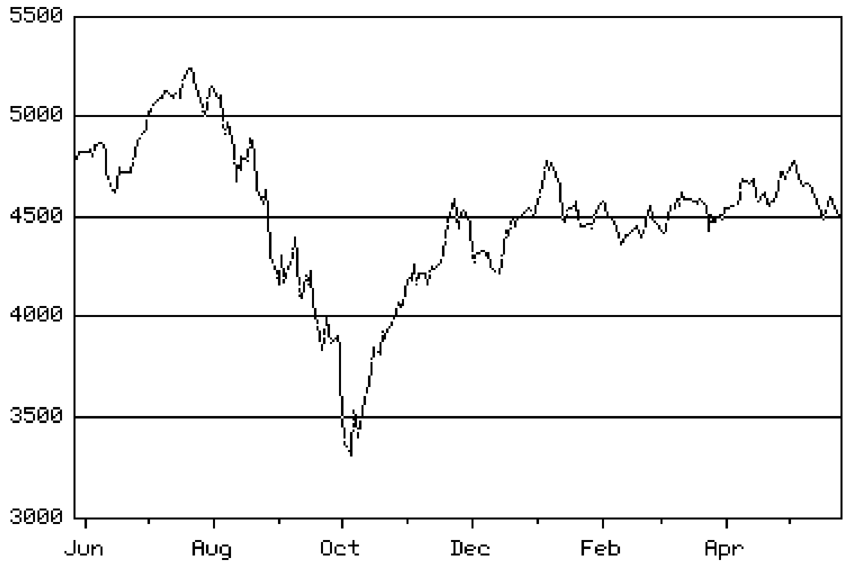


Fig 0.4: Swiss Performance Index

There are of course many further examples of connections and applications of probability theory (e.g. with physics). We will discuss some of them in the course of the lecture.

Plan of the course:

Chap. 1: Basic notions, Law of Large Numbers

Chap. 2: Central Limit Theorem, characteristic functions

Chap. 3: Martingales

Chap. 4: Random walks, Markov chains

References:

Probability: Theory and Examples, R. Durrett, Duxbury Press (1996).

Probability with Martingales, D. Williams, Cambridge University Press (1991).

Wahrscheinlichkeitstheorie, H. Bauer, de Gruyter Lehrbuch, 4. Auflage (1991).

Wahrscheinlichkeitstheorie, A. Klenke, Springer (2006).

1 Basic notions, Law of Large Numbers

1.1 Probability spaces, random variables

We start with the axioms of Kolmogorov. A random experiment will be modeled by a **probability space** $(\Omega, \mathcal{A}, P)$, where:

(1.1.1)

- Ω "the sample space" is a non-empty set,
- $\mathcal{A}$ *is* a σ**-algebra** on Ω (i.e. $\mathcal{A}$ is a family of subsets of Ω, with $\Omega \in \mathcal{A}$; $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$; and for each sequence A_i, $i \geq 1$, $A_i \in \mathcal{A}$, one has $\bigcup_{i\geq 1} A_i \in \mathcal{A}$),
- P is a probability measure on $(\Omega, \mathcal{A})$ (i.e. P is a map: $\mathcal{A} \xrightarrow{P} [0,1]$, with $P(\Omega) = 1$, and for each sequence A_i, $i \geq 1$, of pairwise disjoint elements of $\mathcal{A}$ (i.e. $A_i \in \mathcal{A}$, $i \geq 1$, and $A_i \cap A_j = \phi$, $i \neq j$), it holds that $P(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1} P(A_i)$),

in other words, P is a **normalized measure** on $(\Omega, \mathcal{A})$.

- $\omega \in \Omega$ is called an "elementary event".
- $A \in \mathcal{A}$ is called an "event".
- Intuitively speaking, $A \in \mathcal{A}$ models a possible question related to the random experiment: by running the experiment, one obtains an elementary event ω, and asks the question: "does ω lie in A?"
- $P(A)$ (with $A \in \mathcal{A}$) describes the relative likelihood of a positive answer to the previous question ("does ω lie in A?"), when we conduct the random experiment many times.

Example 1.1.

1) $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ "the Borel σ-algebra on $\mathbb{R}$", which means the smallest σ-algebra[1] on $\mathbb{R}$ that contains all open subsets of $\mathbb{R}$, and for $A \in \mathcal{B}(\mathbb{R})$:

$$P(A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{(x-m)^2}{2\sigma^2}\Big\}\, dx \,. \qquad (1.1.2)$$

Here, we have the **normal distribution** with parameters $m \in \mathbb{R}$, $\sigma > 0$ (notation: $\mathcal{N}(m, \sigma^2)$).

[1] We remind the reader that in general, for any $\mathcal{E} \subset \mathcal{P}(\Omega)$, there exists a smallest σ-algebra $\sigma(\mathcal{E})$ satisfying $\mathcal{E} \subset \sigma(\mathcal{E})$. This σ-algebra is called **σ-algebra generated by** $\mathcal{E}$, and it can be defined as $\sigma(\mathcal{E}) = \cap \mathcal{A}'$, where this intersection runs over all σ-algebras $\mathcal{A}' \subset \mathcal{P}(\Omega)$ with $\mathcal{E} \subset \mathcal{A}'$. We leave as an exercise to check that this definition indeed gives rise to a σ-algebra with the desired property, and that in particular, $\sigma(\sigma(\mathcal{E})) = \sigma(\mathcal{E})$, and $\mathcal{E} \subset \mathcal{E}' \Longrightarrow \sigma(\mathcal{E}) \subset \sigma(\mathcal{E}')$.

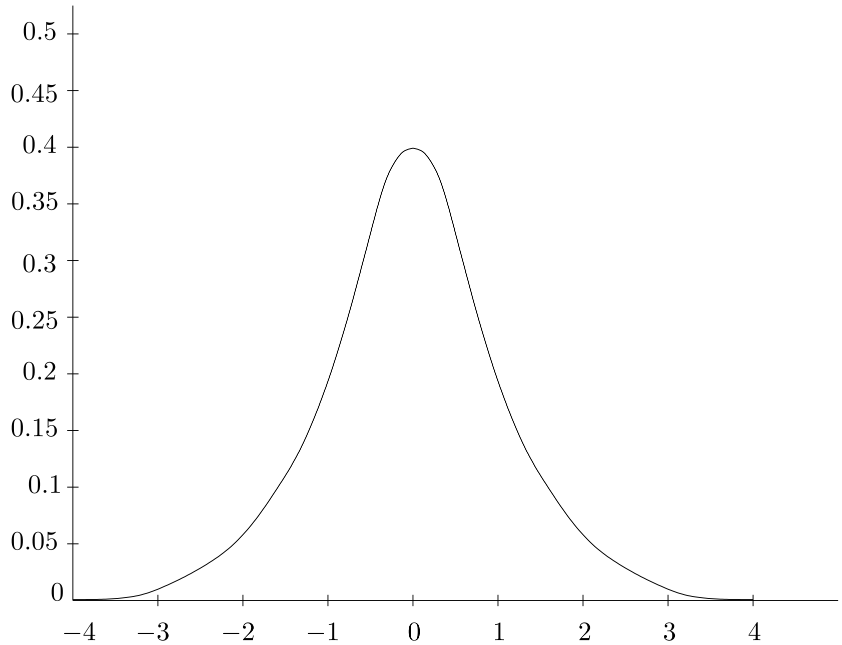


Fig. 1.1: Density of the standard normal distribution: $f(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$

2) $\Omega = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ the power set of $\mathbb{N}$, i.e. the family of all subsets of $\mathbb{N}$, and for $A \subseteq \mathbb{N}$:

$$P(A) = \sum_{n \in A} e^{-\lambda}\, \frac{\lambda^n}{n!}\,, \tag{1.1.3}$$

the **Poisson distribution** with parameter $\lambda > 0$.

3) *Product spaces:* $(\Omega_1, \mathcal{A}_1, P_1)$, $(\Omega_2, \mathcal{A}_2, P_2)$ two probability spaces.

One can construct a new probability space $(\Omega, \mathcal{A}, P)$, defined by

(1.1.4)

- $\Omega = \Omega_1 \times \Omega_2$,
- $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$, i.e. $\mathcal{A}$ is the smallest σ-algebra on Ω that contains all sets of the form $A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$,
- P is the unique probability measure on $(\Omega, \mathcal{A})$ with $P(A_1 \times A_2) = P_1(A_1) \cdot P_2(A_2)$ for all $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$.

$P = P_1 \times P_2$ means **product** of P_1 and P_2 (see the course on measure theory, or Durrett, p. 423). ○

One can also consider, of course, the product of n probability spaces $(\Omega_1, \mathcal{A}_1, P_1), \ldots,$ $(\Omega_n, \mathcal{A}_n, P_n)$.

For instance, if all $(\Omega_i, \mathcal{A}_i, P_i)$, $1 \le i \le n$, are chosen as in Example 1.1 1) with $m = 0$ and $\sigma = 1$ (standard normal distribution), one obtains the n-dimensional standard normal distribution:

$$\Omega = \mathbb{R}^n, \;\; \mathcal{A} = \mathcal{B}(\mathbb{R}^n) \quad \text{(Borel } \sigma\text{-algebra on } \mathbb{R}^n\text{)},$$
$$P(A) = \int_A \frac{1}{(2\pi)^{n/2}} \exp\Big\{ -\frac{x_1^2 + \cdots + x_n^2}{2} \Big\}\, dx_1 \dots dx_n \quad \text{for } A \in \mathcal{B}(\mathbb{R}^n)\,. \tag{1.1.5}$$

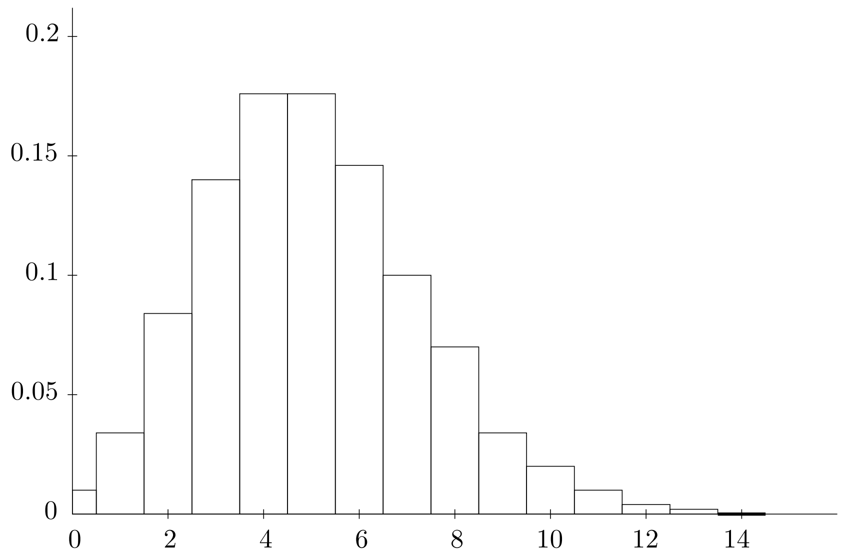


Fig. 1.2: Poisson distribution: $p(n) = e^{-\lambda}\, \frac{\lambda^n}{n!}$, with parameter $\lambda = 5$

Probability spaces often contain too much information, and we thus introduce the notion of random variables.

Definition 1.2. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A map $X \colon \Omega \to \mathbb{R}$ is called a* ***random variable*** *if*

$$X^{-1}(B) \overset{\text{def.}}{=} \{\omega \in \Omega : \; X(\omega) \in B\} \in \mathcal{A} \;\; \textit{for each } B \in \mathcal{B}(\mathbb{R}), \qquad (X^{-1}(B) \overset{\text{notation}}{=} \{X \in B\}) \tag{1.1.6}$$

(In other words: X is a ***measurable map*** *from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).*

Remark 1.3. The family of Borel sets $\mathcal{B}(\mathbb{R})$ is rather abstract, and the property "X is a random variable" actually is equivalent to the concrete condition:

$$X^{-1}((-\infty, a]) \; \big(\overset{\text{notation}}{=} \{X \le a\} \big) \in \mathcal{A} \;\; \text{for all } a \in \mathbb{R}\,. \tag{1.1.7}$$

Proof.

- (1.1.6) $\Longrightarrow$ (1.1.7): clear.
- (1.1.7) $\Longrightarrow$ (1.1.6): the family of $B \subseteq \mathbb{R}$ such that $X^{-1}(B) \in \mathcal{A}$ is a σ-algebra: indeed,

$$X^{-1}(\mathbb{R}) = \Omega,\ X^{-1}(B^c) = \big(X^{-1}(B)\big)^c,\ X^{-1}\big(\bigcup_{i\geq 1} B_i\big) = \bigcup_{i\geq 1} X^{-1}(B_i)\,.$$

Because of (1.1.7), this σ-algebra contains all sets $(-\infty, a]$, $a \in \mathbb{R}$, so that it contains $\mathcal{B}(\mathbb{R})$, and (1.1.6) is thus satisfied. $\square$

With a random variable X on $(\Omega, \mathcal{A}, P)$, one can associate a probability measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the so-called **distribution of** X. One has

$$\text{for all } B \in \mathcal{B}(\mathbb{R}),\ \mu_X(B) = P(\{X \in B\}). \tag{1.1.8}$$

μ_X is a probability measure, since: $\mu_X(\mathbb{R}) = P(\Omega) = 1$, and for $B_i \in \mathcal{B}(\mathbb{R})$, $i \geq 1$, pairwise-disjoint, it holds that

$$\Big\{X \in \bigcup_{i\geq 1} B_i\Big\} = \bigcup_{i\geq 1} \{X \in B_i\} \quad \leftarrow \text{pairwise-disjoint events} \in \mathcal{A}\,,$$

and thus

$$\mu_X\Big(\bigcup_{i\geq 1} B_i\Big) = P\Big(\bigcup_{i\geq 1} \{X \in B_i\}\Big) = \sum_{i\geq 1} P(\{X \in B_i\}) = \sum_{i\geq 1} \mu_X(B_i)\,.$$

Example 1.4. In the setting of (1.1.5), $X : \mathbb{R}^n \to \mathbb{R}$, $(x_1, \dots, x_n) \stackrel{X}{\longmapsto} x_1$ is a random variable, and for $B \in \mathcal{B}(\mathbb{R})$, one has:

$$\begin{aligned}
\mu_X(B) &= P(\{X \in B\}) \\
&= \int_{B\times\mathbb{R}^{n-1}} \frac{1}{(2\pi)^{n/2}} \ \exp\Big\{-\frac{x_1^2 + \dots + x_n^2}{2}\Big\}\, dx_1 \dots dx_n \\
&= \int_B \frac{1}{(2\pi)^{1/2}} \ \exp\Big\{-\frac{x^2}{2}\Big\}\, dx\,,
\end{aligned}$$

so μ_X is the standard normal distribution on $\mathbb{R}$. $\bigcirc$

The **distribution function** $F(\cdot)$ of a random variable X on $(\Omega, \mathcal{A}, P)$ is defined as follows: $F : \mathbb{R} \to [0, 1]$, with

$$F(x) = P[\{X \leq x\}] \stackrel{(1.1.8)}{=} \mu_X((-\infty, x]) \ \text{ for all } x \in \mathbb{R}\,. \tag{1.1.9}$$

This function has the following three **properties:**

$$\begin{aligned}
&\text{i)} \quad F(\cdot) \text{ is non-decreasing}\,,\\
&\text{ii)} \quad \lim_{x\to\infty} F(x) = 1,\ \lim_{x\to-\infty} F(x) = 0\,,\\
&\text{iii)} \quad F \text{ is right-continuous}\,.
\end{aligned} \tag{1.1.10}$$

Proof. i): clear from (1.1.9).

iii): if $x \in \mathbb{R}$, then for each sequence $x_n \downarrow x$, one has $\cap_n(-\infty, x_n] = (-\infty, x]$ and $(-\infty, x_n]$ is decreasing. From this, it follows that $F(x) = \mu_X((-\infty, x]) = \lim_n \mu_X((-\infty, x_n]) = \lim_n F(x_n)$, which gives iii).

ii) is proved similarly. $\square$

With the help of distribution functions, one can in fact derive a complete description of all possible probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

Proposition 1.5. *(Lebesgue-Stieltjes)*

For each function $F : \mathbb{R} \to [0,1]$ *satisfying* (1.1.10), *there exists a unique probability measure* μ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with:*

(1.1.11) $$F(x) = \mu((-\infty, x]) \text{ for all } x \in \mathbb{R}\,.$$

Proof. *Existence:*

Consider $\Omega = (0,1)$, $\mathcal{A} = \mathcal{B}(0,1)$, $P =$ Lebesgue measure on $(0,1)$, and define for $\omega \in (0,1)$

(1.1.12) $$X(\omega) = \sup\{y \in \mathbb{R};\ F(y) < \omega\}\,,$$

X plays the role of the inverse function of F.

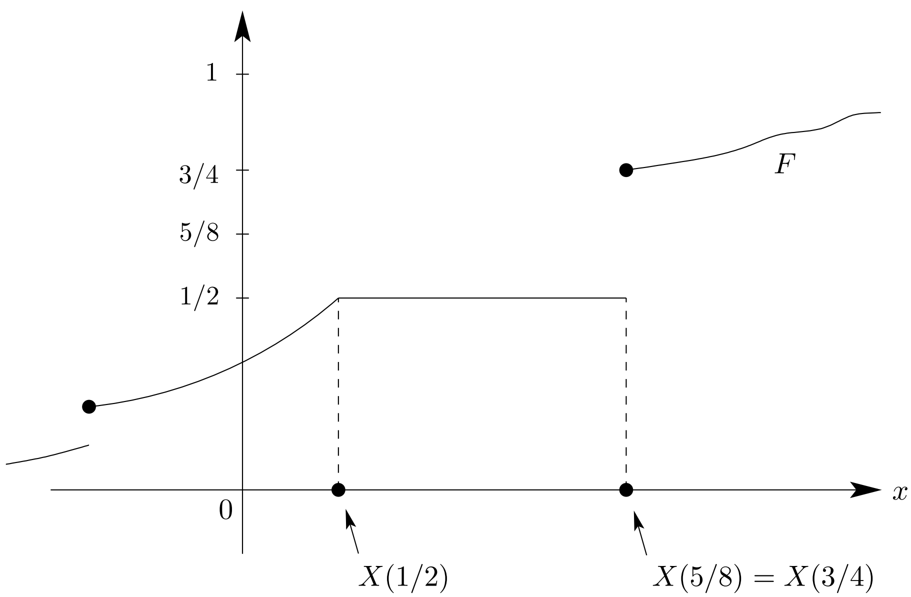


Fig. 1.3

We claim that:

$$\{\omega: X(\omega) \le x\} = \{\omega: \omega \le F(x)\} \quad \forall x \in \mathbb{R}\,. \tag{1.1.13}$$

The existence of μ in (1.1.11) follows from this: we deduce from (1.1.13) that X is a random variable, and that the distribution function of X is equal to F. As for proving (1.1.13):

"$\supseteq$": Let $\omega \in (0,1)$ with $\omega \le F(x)$. Then $x \notin \{y : F(y) < \omega\}$, which implies $x \ge X(\omega)$.

"$\subseteq$": Let $\omega \in (0,1)$ with $F(x) < \omega$. From the right-continuity of F, there exists an $\epsilon > 0$ with $F(x+\epsilon) < \omega$, and consequently $X(\omega) \ge x + \epsilon > x$. This means $F(x) < \omega \Longrightarrow X(\omega) > x$.

Uniqueness:

We can see from (1.1.11) that $\mu((a,b]) = F(b) - F(a)$, $a < b$, is uniquely determined, and thus $\mu((a,b)) = \lim_n \mu((a, b - \frac{1}{n}])$, $a < b$, as well. Also, by σ-additivity, $\mu(O)$ is uniquely determined for each open set O in $\mathbb{R}$ (since it can be written as a countable union of pairwise disjoint open intervals). The uniqueness of μ follows (see the course on measure theory, or Dynkin's lemma below). □

Remark 1.6. The proof of existence is constructive. One can use (1.1.12) to simulate a general distribution, if one is already able to simulate the uniform distribution on $(0,1)$. □

We will now study a few further properties of random variables:

Proposition 1.7. *Let $X_1, \dots, X_n$ be random variables on $(\Omega, \mathcal{A}, P)$, and $f : \mathbb{R}^n \to \mathbb{R}$ a measurable map, then*

$$f(X_1, X_2, \dots, X_n) \text{ is a random variable}\,. \tag{1.1.14}$$

Proof. One has
$$f(X_1, X_2, \dots, X_n) = f \circ X\,,$$
with $X : \omega \in \Omega \to (X_1(\omega), \dots, X_n(\omega)) \in \mathbb{R}^n$. If we show that X from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is measurable, then it will follow that $f \circ X$ from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable as well since f is measurable, in other words that $f(X_1, \dots, X_n)$ is a random variable.

But one has, for $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\{\omega: X(\omega) \in A_1 \times \cdots \times A_n\} = \bigcap_{i=1}^{n} \{X_i \in A_i\},$$

and $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra on $\mathbb{R}^n$ that contains all such $A_1 \times \cdots \times A_n$. The measurability of X now follows as in (1.1.7). □

A classical application of (1.1.14) is for example:

$$\text{Let } X_1, \dots, X_n \text{ be random variables,} \\ \text{then } X_1 + X_2 + \cdots + X_n \text{ is a random variable too}. \tag{1.1.15}$$

As a conclusion of this section, we will discuss a very useful property of the notion of random variable. Namely, the class of random variables is closed under countable inf-, liminf-, sup-, and limsup- operations.

In order to present the full power of this result, a small generalization will however be needed: we will consider **random variables with values in** $[-\infty, \infty]$, i.e. $X^{-1}((a, +\infty]) \in \mathcal{A}$ for all $a \in \mathbb{R}$. Such random variables are also sometimes called **numerical random variables**.

Proposition 1.8. *Let $X_1, X_2, \dots, X_n, \dots$ be random variables with values in $[-\infty, +\infty]$, then $\inf_n X_n$, $\sup_n X_n$, $\liminf_n X_n$, and $\limsup_n X_n$ are numerical random variables as well.*

Indeed, for all $a \in \mathbb{R}$,

$$\{\inf_n X_n < a\} = \bigcup_n \{X_n < a\}, \quad \text{and}$$
$$\{\sup_n X_n > a\} = \bigcup_n \{X_n > a\}\ .$$

Hence, $\inf_n X_n$ and $\sup_n X_n$ are measurable, and thus (numerical) random variables.

From this, it also follows that

$$\limsup_n X_n = \inf_m \big(\sup_{n \geq m} X_n\big) \text{ and}$$
$$\liminf_n X_n = \sup_m \big(\inf_{n \geq m} X_n\big)$$

are (numerical) random variables too.

Using the previous proposition, one can easily see that the **set of convergence** of the sequence X_n,

$$\Omega_0 = \{\limsup_n X_n = \liminf_n X_n\} \subset \Omega,$$

lies in $\mathcal{A}$. When $P(\Omega_0) = 1$, we say that the sequence X_n P**-almost surely** (abbreviated as P-a.s.) **converges**.

1.2 Expectation

In this section, we will recall a few classical and useful results from the course on measure theory.

Intuitively speaking, the expectation (or expected value) of a random variable X on $(\Omega, \mathcal{A}, P)$ corresponds to the average value taken by X, when one repeats the random experiment modeled by $(\Omega, \mathcal{A}, P)$.

The expectation of a random variable X on $(\Omega, \mathcal{A}, P)$ with

$$\int_\Omega |X| \, dP < \infty \tag{1.2.1}$$

is defined mathematically as:

$$E[X] = \int_\Omega X \, dP \, . \tag{1.2.2}$$

If X and Y are random variables satisfying (1.2.1), one has (see Durrett, Appendix)

$$E[aX + bY] = a\,E[X] + b\,E[Y] \qquad \text{for all } \ a, b \in \mathbb{R}\,, \tag{1.2.3}$$

$$E[X] \geq E[Y] \qquad \text{if } \ X \geq Y \ , \tag{1.2.4}$$

and as a special case of (1.2.4),

$$E[\,|X|\,] \geq |E[X]| \ . \tag{1.2.5}$$

Jensen's Inequality:

For a random variable X satisfying property (1.2.1), and a convex function $\varphi : \mathbb{R} \to \mathbb{R}$ (i.e. $\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y)$, $x, y \in \mathbb{R}$, $\lambda \in [0,1]$) with $E[|\varphi(X)|] < \infty$, one has

$$\varphi(E[X]) \leq E[\varphi(X)] \ . \tag{1.2.6}$$

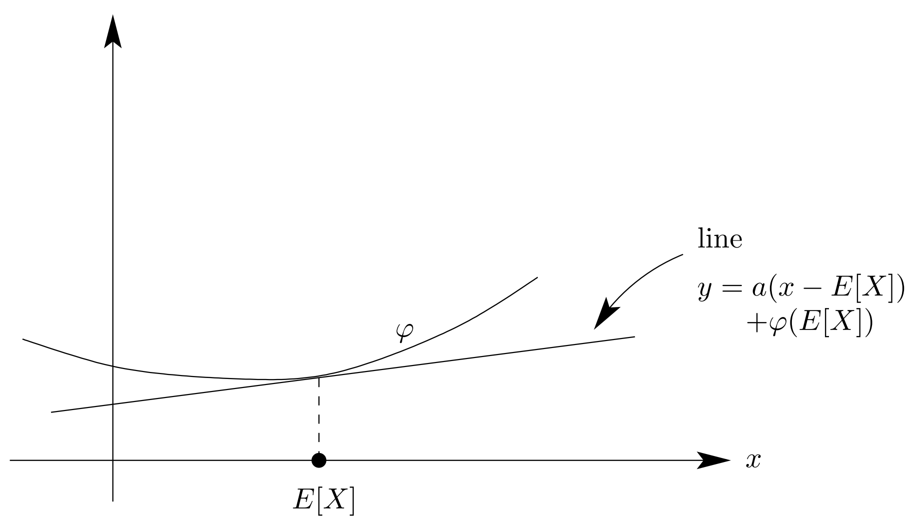


Fig. 1.4

Special case:

$$E[X]^2 \leq E[X^2] \ . \tag{1.2.7}$$

The difference $\mathrm{Var}(X) \stackrel{\text{def.}}{=} E[X^2] - E[X]^2 \geq 0$ is called **variance of** X.

Hölder's Inequality:

Let $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, one has:

$$
\begin{aligned}
&E[\,|XY|\,] \le \|X\|_p \, \|Y\|_q, \ \text{ where} \\
&\qquad\qquad \|X\|_r = E[\,|X|^r]^{1/r}, \ r \in [1,\infty), \ \text{ and} \\
&\qquad\qquad \|X\|_\infty = \inf\{M : \ P[\,|X| > M] = 0\} \ .
\end{aligned} \tag{1.2.8}
$$

For $p = q = 2$, one obtains in particular the **Cauchy-Schwarz Inequality:**

$$
E[\,|XY|\,] \le \|X\|_2 \, \|Y\|_2 \ . \tag{1.2.9}
$$

Lemma 1.9. *(Fatou)*

Let $X_n \ge 0$ be a sequence of random variables with values in $[0, \infty]$, then:

$$
E\Big[\liminf_n \ X_n\Big] \le \liminf_n \ E[X_n] \ . \tag{1.2.10}
$$

Theorem 1.10. *(Monotone convergence (Beppo Levi))*

Let X_n be a sequence of random variables with $X_n \ge 0$ and $X_n \uparrow X$, then

$$
E[X_n] \uparrow E[X] \ . \tag{1.2.11}
$$

Theorem 1.11. *(Dominated convergence (Lebesgue))*

Let X, Y, and X_n be random variables with $X_n \overset{P-\text{a.s.}}{\longrightarrow} X$, $|X_n| \le Y$ for all n, and $E[Y] < \infty$, then:

$$
\lim_{n\to\infty} \ E[X_n] = E[X] \ . \tag{1.2.12}
$$

Chebyshev's Inequality:

For $\varphi : \mathbb{R} \to [0, \infty)$ a measurable function, $A \in \mathcal{B}(\mathbb{R})$, and X a random variable, one has

$$
\begin{aligned}
&\inf\{\varphi(x), \, x \in A\} \ P[X \in A] \\
&\le \int_{X\in A} \varphi(X)\, dP \big(\overset{\text{notation}}{=} E[\varphi(X); \, X \in A]\big) \\
&\le E[\varphi(X)] \ .
\end{aligned} \tag{1.2.13}
$$

Proof. We can write

$$
\inf\{\varphi(x), \, x \in A\} \, 1_{\{X\in A\}} \le \varphi(X) \, 1_{\{X\in A\}} \le \varphi(X)
$$

(we adopt the usual convention from measure theory that $0 \cdot \infty = 0$), and (1.2.13) follows by integration. $\square$

Special case:

$$a^2\, P[\,|X| \geq a] \leq E[X^2]\,. \tag{1.2.14}$$

Image of a probability measure through a measurable map:
(change of variable formula)

We consider a probability space $(\Omega, \mathcal{A}, P)$, a measurable space $(S, \mathcal{S})$, and a measurable map $h : (\Omega, \mathcal{A}) \longrightarrow (S, \mathcal{S})$.

One can define **the image of** P **through** h (denoted by $h \circ P$ or $P \circ h^{-1}$) as the following probability measure:

$$(h \circ P)(B) = P[h^{-1}(B)] \qquad \forall B \in \mathcal{S}\,. \tag{1.2.15}$$

The fact that $h \circ P$ is a probability measure on $(S, \mathcal{S})$ follows from analogous reasons as for the distribution of a random variable (see (1.1.8)).

Example 1.12. Consider $P(dx) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}\, x^2\}dx$ (standard normal distribution) on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The image of P through the exponential function $x \to \exp(x)$ is the distribution:

$$Q(dy) = (\exp \circ P)(dy) = \frac{1}{\sqrt{2\pi}}\ \exp\Big\{-\frac{1}{2}\,(\log y)^2\Big\}\ 1(y > 0)\ \frac{dy}{y}\,.$$

(log-normal distribution)

$\bigcirc$

Proposition 1.13. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space,* $(\Omega, \mathcal{A}) \xrightarrow{h} (S, \mathcal{S})$ *a measurable map, and* Y *a random variable on* $(S, \mathcal{S})$*. One has*

$$\int_S |Y|\; d(h \circ P) < \infty \Longleftrightarrow \int_\Omega |Y \circ h|\, dP < \infty\,, \tag{1.2.16}$$

and if (1.2.16) *is satisfied, then*

$$E^{(h \circ P)}[Y] = E^P[Y \circ h]\,. \tag{1.2.17}$$

Proof. We approximate Y in four steps (this is sometimes called "measure-theoretic induction"):

a) $Y = 1_B,\ B \in \mathcal{S}$. In this case, (1.2.16) is clearly satisfied, and

$$(h \circ P)(B) \overset{\text{def.}}{=} P[h^{-1}(B)] = E^P[1_B \circ h]\,,$$

which is (1.2.17).

b) $Y = \sum_{m=1}^n\ c_m\ 1_{B_m},\ \ c_m \in \mathbb{R},\ B_m \in \mathcal{S}$. Again, (1.2.16) is satisfied, and (1.2.17) follows from a) and the linearity of expected value.

c) $Y \geq 0$. Let us introduce

$$Y_n = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \, 1\Big\{\frac{k}{2^n} \leq Y < \frac{k+1}{2^n}\Big\} + n \, 1\{n \leq Y\}$$

(in other words: $Y_n(s)$ is, for $s \in S$, the minimum of n and the largest dyadic number with complexity n that is smaller than $Y(s)$). One has

$$\begin{cases} Y_n \uparrow Y, \text{ and } Y_n \text{ is as in b)}, \\ Y_n \circ h \uparrow Y \circ h. \end{cases}$$

Using the monotone convergence theorem, we obtain

$$\begin{aligned} & E^{(h \circ P)}[Y] \overset{\text{(monotone convergence)}}{=} \lim_{n \to \infty} \uparrow E^{h \circ P}[Y_n] \\ & \overset{(b)}{=} \lim_{n \to \infty} \uparrow E^P[Y_n \circ h] \overset{\text{(monotone convergence)}}{=} E^P[Y \circ h] \, . \end{aligned}$$

Hence, (1.2.16) and (1.2.17) follow.

d) Y a general random variable on $(S, \mathcal{S})$. Consider

$$\begin{aligned} Y_+(s) &= \max(Y(s), 0), \; s \in S, \\ Y_-(s) &= \max(-Y(s), 0), \; s \in S \, . \end{aligned}$$

The property $E^{(h \circ P)}[\,|Y|\,] < \infty$ is then equivalent to

$$E^{(h \circ P)}[Y_+] < \infty \;\text{ and }\; E^{h \circ P}[Y_-] < \infty \, .$$

A similar statement holds for $Y \circ h$, $Y_+ \circ h$ $(= (Y \circ h)_+)$, and $Y_- \circ h (= (Y \circ h)_-)$, with respect to P.

The equivalence (1.2.16) then follows from these observations and c), and (1.2.17) comes from writing

$$\begin{aligned} E^{h \circ P}[Y] &= E^{h \circ P}[Y_+] - E^{h \circ P}[Y_-], \\ E^P[Y \circ h] &= E^P[Y_+ \circ h] - E^P[Y_- \circ h], \end{aligned}$$

and using c). □

1.3 Independence

We first present the elementary definition of independence for two events, and then a series of generalizations of this elementary definition.

Let $(\Omega, \mathcal{A}, P)$ be a probability space, $B, C \in \mathcal{A}$ **are said to be independent** if

(1.3.1) $$P[B \cap C] = P[B] \, P[C] \, .$$

If for instance $P(B) > 0$, then (1.3.1) is equivalent to

$$\frac{P[C \cap B]}{P[B]} = P[C|B] = P[C] \text{ (conditional probability)} .$$

This allows one to interpret (1.3.1) as follows:

"The occurrence of B has no influence on the occurrence of C" (and similarly for C, if $P[C] > 0$).

Two **σ-algebras $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$ are said to be independent** if

$$P[B \cap C] = P[B]\, P[C] \qquad \forall B \in \mathcal{B}, \quad \forall C \in \mathcal{C} . \tag{1.3.2}$$

(1.3.3) Two **random variables** X, Y on $(\Omega, \mathcal{A}, P)$ **are said to be independent** if the σ-algebras $\sigma(X) = \{X^{-1}(A);\ A \in \mathcal{B}(\mathbb{R})\}$ and $\sigma(Y) = \{Y^{-1}(A);\ A \in \mathcal{B}(\mathbb{R})\}$ that they generate are independent (in the sense of (1.3.2))

(we leave as an exercise to check that $\sigma(X)$ and $\sigma(Y)$ are indeed σ-algebras). These definitions can be generalized to the case of a larger number (more than two) of σ-algebras or random variables.

From such a generalized definition, we expect for example that from the property

(1.3.4) "X_1, X_2, X_3, X_4, X_5 are independent random variables", it follows that "$\exp\{X_1 + X_2\}$, $\frac{1}{1+X_3^2+X_4^2+X_5^2}$ are independent random variables".

But we immediately encounter a problem here, as the following example shows:

Example 1.14.

$$\Omega = \{a, b, c, d\}, \quad \mathcal{A} = \mathcal{P}(\Omega) ,$$

$$P[\{a\}] = P[\{b\}] = P[\{c\}] = P[\{d\}] = \frac{1}{4} ,$$

then

$$A = \{a, b\}, \;\; B = \{b, c\}, \;\; C = \{c, a\}$$

are pairwise independent, but A and $B \cap C = \{c\}$ are not independent at all, since $A \cap (B \cap C) = \emptyset$! ○

As a consequence, in order to define the desired generalization in the case of a larger number of σ-algebras (or random variables), one needs more than simply pairwise independence for each pair of σ-algebras (or random variables).

Definition 1.15. *Consider $(\Omega, \mathcal{A}, P)$ a probability space. The sub-σ-algebras $\mathcal{B}_1, \dots, \mathcal{B}_n$ of $\mathcal{A}$ are said to be independent if:*

$$P[B_1 \cap \dots \cap B_n] = P[B_1]\, P[B_2] \dots P[B_n] \qquad \forall B_1 \in \mathcal{B}_1, \dots, \forall B_n \in \mathcal{B}_n . \tag{1.3.5}$$

Remark 1.16. Each subsequence of $\mathcal{B}_1, \dots, \mathcal{B}_n$ (for instance $\mathcal{B}_1, \mathcal{B}_5, \mathcal{B}_8$, if $n \geq 8$) is also independent.

(1.3.6) Independence of random variables $X_1, \dots, X_n$ can be defined in an analogue way as the independence of the σ-algebras $\sigma(X_1), \dots, \sigma(X_n)$ that they generate.

□

It is now more than time that we investigate whether these definitions make sense. For this, we will devise a very useful tool: Dynkin's lemma.

Definition 1.17. *A family $\mathcal{D}$ of subsets of Ω is called a* ***Dynkin system*** *(or a λ-system) if it satisfies*

(1.3.7)

i) $\Omega \in \mathcal{D}$,

ii) $A \in \mathcal{D} \Longrightarrow A^c \in \mathcal{D}$,

iii) *for each sequence $A_i, i \geq 1$, of* ***pairwise disjoint*** *elements from $\mathcal{D}$ (i.e. $A_i \in \mathcal{D}$, $i \geq 1$, and $A_i \cap A_j = \phi$, $i \neq j$), one has $(\bigcup_{i \geq 1} A_i) \in \mathcal{D}$*

(the difference with the notion of σ-algebra lies in iii)*).*

(1.3.8) *A family $\mathcal{C}$ of subsets of Ω is called a π*-***system*** *if $C \cap C' \in \mathcal{C}$ for $C, C' \in \mathcal{C}$*

(in other words: $\mathcal{C}$ is closed under $\cap$).

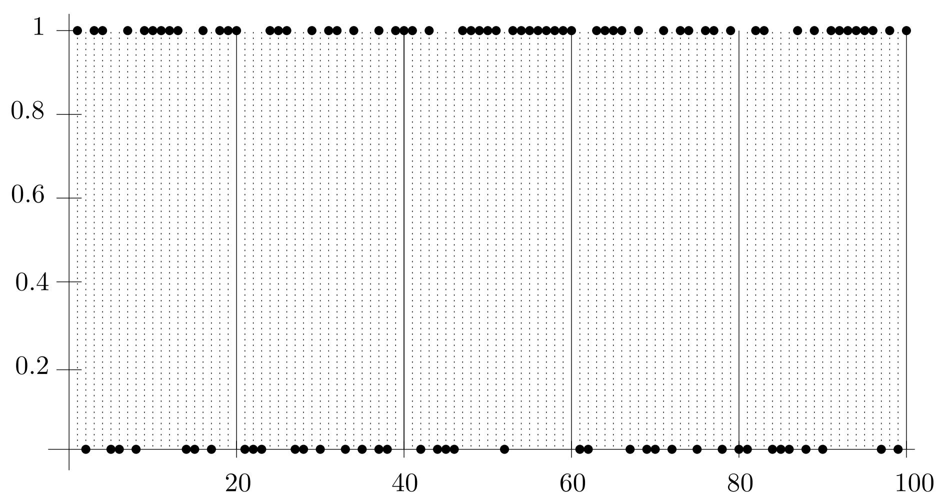


Fig. 1.4: Independent Bernoulli(1/2) random variables

Lemma 1.18. *(Dynkin)*

Let $\mathcal{D}$ be a Dynkin system, and $\mathcal{C}$ a π-system on Ω with $\mathcal{C} \subseteq \mathcal{D}$. Then one has

(1.3.9) $$\mathcal{D} \supset \sigma(\mathcal{C}) \text{ } (\leftarrow \textit{the } \sigma\textit{-algebra generated by } \mathcal{C}).$$

The proof of Lemma 1.18 is rather abstract. As a motivation, we will first examine a few consequences and applications of this lemma.

For many applications, the following **general principle** is of interest:

(1.3.10) Let P, Q be two probability measures on $(\Omega, \mathcal{A})$, then the family $\mathcal{D} = \{A \in \mathcal{A} : \; P(A) = Q(A)\}$ is a Dynkin system.

Proof. i) and ii) are clear, and iii) follows by σ-additivity.

As a consequence of Dynkin's lemma and (1.3.10), we obtain

(1.3.11) Let P, Q be two probability measures on $(\Omega, \mathcal{A})$, $\mathcal{C} \subseteq \mathcal{A}$ a π-system such that $\forall C \in \mathcal{C},\; P(C) = Q(C)$, then one has:

$$\forall B \in \sigma(\mathcal{C}), \;\; P(B) = Q(B)\,.$$

□

As a concrete application of (1.3.11), one has for instance:

Proposition 1.19. *Let X, Y be independent random variables on $(\Omega, \mathcal{A}, P)$, with distributions μ and ν respectively. Then, the image of P on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ through the map $\omega \overset{\phi}{\longmapsto} (X(\omega), Y(\omega))$ is exactly $\mu \otimes \nu$.*

For a measurable function h: $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, one has:

$$E[\,|h(X,Y)|\,] < \infty \iff h \text{ is } \mu \otimes \nu \text{ integrable.} \tag{1.3.12}$$

In this case, one has furthermore

$$E[h(X,Y)] = \int_{\mathbb{R}^2} h(x,y)\, d\mu(x) d\,\nu(y)\;. \tag{1.3.13}$$

In particular:

$$E[XY] = E[X]\,E[Y] \qquad \text{if } E[\,|X|\,],\; E[\,|Y|\,] < \infty\,. \tag{1.3.14}$$

Proof. Thanks to (1.2.16) and (1.2.17), we only need to show that

$$\mu \otimes \nu = \phi \circ P\;. \tag{1.3.15}$$

Define $\mathcal{C} = \{A_1 \times A_2 : \; A_1, A_2 \in \mathcal{B}(\mathbb{R})\}$. $\mathcal{C}$ is a π-system, and for $A = A_1 \times A_2 \in \mathcal{C}$, one has

$$\begin{aligned} &\mu \otimes \nu(A) = \mu(A_1)\,\nu(A_2), \;\; \text{and} \\ &(\phi \circ P)(A) = P[X \in A_1,\, Y \in A_2] \\ &\overset{\text{(independence)}}{=} P[X \in A_1]\,P[Y \in A_2] = \mu(A_1)\,\nu(A_2)\,. \end{aligned} \tag{1.3.16}$$

Since one has also $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^2)$, the claim follows from (1.3.11) and (1.3.16). □

Remark 1.20. The uniqueness part in the proposition of Lebesgue-Stieltjes (see (1.1.11)) can be proved analogously. In this case, one chooses as a π-system $\mathcal{C}$ the family of intervals $(a, b]$, with $a \le b$ in $\mathbb{R}$. $\square$

A further application of Dynkin's lemma is the following:

Theorem 1.21. *Let $\mathcal{C}_1, \dots, \mathcal{C}_n \subseteq \mathcal{A}$ be π-systems with $\Omega \in \mathcal{C}_i$, and*

$$\forall C_1 \in \mathcal{C}_1, \dots, \forall C_n \in \mathcal{C}_n, \qquad P[C_1 \cap \dots \cap C_n] = P[C_1] \dots P[C_n]\,. \tag{1.3.17}$$

(1.3.18) *Then the σ-algebras $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n)$ are independent.*

Proof. For fixed $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$, we consider the family $\mathcal{D}_1$ of all $D \in \mathcal{A}$ with

$$P[D \cap C_2 \cap \dots \cap C_n] = P[D]\, P[C_2] \dots P[C_n]\ . \tag{1.3.19}$$

Then one has

$$\mathcal{D}_1 \text{ contains } \mathcal{C}_1 \text{ (using (1.3.17))}\,. \tag{1.3.20}$$

(1.3.21)

$\mathcal{D}_1$ is a Dynkin system:

i), ii) are clear.

iii) Consider $D_\ell, \ell \ge 1$, in $\mathcal{D}_1$ pairwise disjoint, and $D = \bigcup_\ell D_\ell$, then one has:

$$P[D \cap C_2 \cap \dots \cap C_n] \overset{(\sigma-\text{addit.})}{=} \sum_{\ell \ge 1} P[D_\ell \cap C_2 \cap \dots \cap C_n]$$
$$\overset{(D_\ell \in \mathcal{D}_1)}{=} \sum_{\ell \ge 1} P[D_\ell] P[C_2] \dots P[C_n] \overset{(\sigma-\text{addit.})}{=} P[D] P[C_2] \dots P[C_n].$$

Thanks to (1.3.20), (1.3.21) and Lemma 1.18, one has (1.3.19) for $D \in \sigma(\mathcal{C}_1)$.

One can now define the family $\mathcal{D}_2$ of all sets $D \in \mathcal{A}$ with:

$$P[A \cap D \cap C_3 \cap \dots \cap C_n] = P[A]\, P[D]\, P[C_3] \dots P[C_n]\,, \tag{1.3.22}$$

for fixed $C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$ and arbitrary $A \in \sigma(\mathcal{C}_1)$. In a similar way, one can see that $\mathcal{D}_2 \supset \mathcal{C}_2$ and $\mathcal{D}_2$ is a Dynkin system.

We conclude from this, using Dynkin's lemma, that (1.3.22) holds true for $A \in \sigma(\mathcal{C}_1)$, $D \in \sigma(\mathcal{C}_2)$, $C_3 \in \mathcal{C}_3, \dots, C_n \in \mathcal{C}_n$, and so on. $\square$

Corollary 1.22.

- *Let $\mathcal{F}_{i,j}$, $1 \le i \le n$, $1 \le j \le m(i)$, be independent σ-algebras.*

(1.3.23) *Then the σ-algebras $\mathcal{G}_i = \sigma\Big(\bigcup_{j=1}^{m(i)} \mathcal{F}_{i,j}\Big)$, $1 \le i \le n$, are also independent.*

- *Let $X_{i,j}$, $1 \le i \le n$, $1 \le j \le m(i)$, be independent random variables, and f_i: $\mathbb{R}^{m(i)} \to \mathbb{R}$ be measurable functions,*

$$\text{then the random variables } Y_i = f_i(X_{i,1}, \dots, X_{i,m(i)}),\ 1 \le i \le n, \text{ are independent too.} \tag{1.3.24}$$

In particular, (1.3.4) *holds, as desired.*

Proof.

- (1.3.23) $\Longrightarrow$ (1.3.24): Since $\sigma(Y_i) \subset \mathcal{G}_i \stackrel{\text{def.}}{=} \sigma(X_{i,1}, \dots, X_{i,m(i)})$, and using (1.3.23), $\mathcal{G}_1, \dots, \mathcal{G}_n$ are independent.

- (1.3.23): Define $\mathcal{C}_i$ as the family of subsets of Ω of the form $\bigcap_{j=1}^{m(i)} A_{i,j}$, where $A_{i,j} \in \mathcal{F}_{i,j}$. Then (1.3.17) holds, and (1.3.23) follows from (1.3.18).

$\square$

Finally, we arrive to

Proof of Lemma 1.18.

$$\begin{array}{c} \mathcal{D} \supset \mathcal{C} \leftarrow \pi\text{−system} . \\ \uparrow \\ \text{Dynkin system} \end{array}$$

(1.3.9): We must show $\mathcal{D} \supset \sigma(\mathcal{C})$. We define the Dynkin system generated by $\mathcal{C}$:

$$\mathcal{D}(\mathcal{C}) = \bigcap_{\mathcal{D}' \supset \mathcal{C}, \mathcal{D}' \text{ Dynkin system}} \mathcal{D}', \tag{1.3.25}$$

and we will show that

$$\mathcal{D}(\mathcal{C}) = \sigma(\mathcal{C}) \ \text{(and so (1.3.9))} .$$

$\mathcal{D}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$ is clear, since each σ-algebra is a Dynkin system. Hence, we only need to prove that

$$\sigma(\mathcal{C}) \subseteq \mathcal{D}(\mathcal{C}) .$$

This statement follows from

$$\mathcal{D}(\mathcal{C}) \text{ is a } \sigma\text{-algebra} , \tag{1.3.26}$$

which we now prove. It follows from (1.3.25) that (see below)

$$\mathcal{D}(\mathcal{C}) \text{ is } \bigcap\text{-closed} . \tag{1.3.27}$$

With (1.3.27), we obtain

$$\begin{array}{l} \mathcal{D}(\mathcal{C}) \text{ is } \bigcup\text{-closed (using } A \cup B = (A^c \cap B^c)^c, \\ \text{(1.3.7) ii), and (1.3.27)) .} \end{array}$$

Hence, $\mathcal{D}(\mathcal{C})$ is closed under countable unions, since for A_n, $n \ge 1$, $A_n \in \mathcal{D}(\mathcal{C})$, one has:

$$\bigcup_{n \ge 1} A_n = \bigcup_{n \ge 1} B_n \backslash B_{n-1}, \text{ with } B_0 = \emptyset,\ B_n = \bigcup_{i=1}^{n} A_i \in \mathcal{D}(\mathcal{C}),$$

and

$$(B_n \backslash B_{n-1}) \in \mathcal{D}(\mathcal{C}), \quad \text{since} \quad (B_n \backslash B_{n-1})^c = \underset{\text{pairwise disjoint in } \mathcal{D}(\mathcal{C})}{\underset{\nwarrow \ \nearrow}{B_n^c \cup B_{n-1}}} \in \mathcal{D}(\mathcal{C})\,.$$

Consequently, $\bigcup_{n\geq 1} A_n \in \mathcal{D}(\mathcal{C})$ follows with (1.3.7) iii). As $\Omega \in \mathcal{D}(\mathcal{C})$, and $\mathcal{D}(\mathcal{C})$ is closed under taking complement, we obtain (1.3.26).

It is thus enough to prove (1.3.27):

First step:

(1.3.27): $A \in \mathcal{D}(\mathcal{C})$, $B \in \mathcal{C} \Longrightarrow A \cap B \in \mathcal{D}(\mathcal{C})$.

We define, for fixed $B \in \mathcal{C}$:

$$\mathcal{D}_B = \{A \subseteq \Omega :\ A \cap B \in \mathcal{D}(\mathcal{C})\}\,.$$

- $\mathcal{D}_B$ is a Dynkin system, since:

  i) $\Omega \in \mathcal{D}_B$: clear.

  ii) $A \in \mathcal{D}_B \Longrightarrow A^c \cap B = \underbrace{B}_{\in \mathcal{C}} \setminus \underbrace{A \cap B}_{\in \mathcal{D}(\mathcal{C})} \in \mathcal{D}(\mathcal{C})$

  $$\text{(since } (B \backslash (A \cap B))^c = \underset{\text{pairwise disjoint in } \mathcal{D}(\mathcal{C})}{\underset{\nwarrow \qquad \nearrow}{B^c \cup (A \cap B))}}$$

  $\Longrightarrow A^c \in \mathcal{D}_B\,.$

  iii) $A_1, \ldots, A_n, \cdots \in \mathcal{D}_B$ pairwise disjoint $\Longrightarrow$

  $$\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B = \bigcup_{i=1}^{\infty} \underbrace{(A_i \cap B)}_{\in \mathcal{D}(\mathcal{C}),\ \text{pairwise disjoint}} \in \mathcal{D}(\mathcal{C})\,.$$

  Hence, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}_B$.

- $\mathcal{D}_B \supset \mathcal{C}$. Consequently $\mathcal{D}_B \supset \mathcal{D}(\mathcal{C})$, and the statement follows.

Second step to prove* (1.3.27)*:

We define, for fixed $A \in \mathcal{D}(\mathcal{C})$,

$$\mathcal{D}_A = \{B \subseteq \Omega :\ A \cap B \in \mathcal{D}(\mathcal{C})\}\,.$$

- $\mathcal{D}_A$ is a Dynkin system: the proof is analogous to i), ii), iii) above.

- $\mathcal{D}_A \supset \mathcal{C}$ thanks to the *first step*.

It follows that $\mathcal{D}_A \supset \mathcal{D}(\mathcal{C})$, and $A \in \mathcal{D}(\mathcal{C})$, $B \in \mathcal{D}(\mathcal{C}) \Longrightarrow A \cap B \in \mathcal{D}(\mathcal{C})$: we deduce (1.3.27). $\square$

Up to now, we have only defined the independence of finite sequences of σ-algebras or random variables. An **infinite sequence** of σ-algebras $\mathcal{B}_i, i \geq 1$ (resp. random variables $X_i, i \geq 1$), **is said to be independent** if each finite subsequence of $\mathcal{B}_i$ (resp. of X_i) is independent. A sequence of events A_i, $i \geq 1$, is called independent if the random variables 1_{A_i}, $i \geq 1$, are independent.

Lemma 1.23. *(Borel-Cantelli)*

For a sequence A_i, $i \geq 1$, of events, we define

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \bigcap_{n\geq 1} \Big(\bigcup_{m \geq n} A_m \Big) \\ &= \{\omega \in \Omega : \ \omega \textit{ lies in infinitely many } A_n\}, \end{aligned} \tag{1.3.28}$$

$$\begin{aligned} \liminf_{n\to\infty} A_n &= \bigcup_{n\geq 1} \Big(\bigcap_{m \geq n} A_m \Big) \\ &= \{\omega \in \Omega : \textit{only finitely many } A_n \textit{ do not contain } \omega \}. \end{aligned} \tag{1.3.29}$$

The notations $\limsup\limits_n A_n$, $\liminf\limits_n A_n$ come from the identities

$$1_{\limsup\limits_n A_n} = \limsup_n 1_{A_n}, \ 1_{\liminf\limits_n A_n} = \liminf_n 1_{A_n} . \tag{1.3.30}$$

Lemma 1.24. *(First lemma of Borel Cantelli)*

Let $A_n, n \geq 1$, be a sequence of events on $(\Omega, \mathcal{A}, P)$, then:

$$\sum_n P(A_n) < \infty \Longrightarrow P[\limsup A_n] = 0 . \tag{1.3.31}$$

Proof. Using the monotone convergence theorem (1.2.11), we have

$$E\Big[\sum_{n=1}^{\infty} 1_{A_n}\Big] = \sum_{n=1}^{\infty} P(A_n) < \infty ,$$

so that

$$\sum_{n=1}^{\infty} 1_{A_n} < \infty \ \ P\text{-a.s} \ \ \Longrightarrow P[\limsup A_n] = 0 ,$$

which completes the proof. $\square$

The converse is false without further assumptions, as the following example shows.

Example 1.25. $\Omega = (0,1)$, $\mathcal{A} = \mathcal{B}(0,1)$, $P =$ Lebesgue measure on $(0,1)$, $A_n = \big(0, \frac{1}{n}\big)$, then $\limsup A_n = \emptyset$, but $\sum P(A_n) = \infty$. $\bigcirc$

Lemma 1.26. *(Second lemma of Borel Cantelli)*

Let A_n, $n \geq 1$, be a sequence of ***independent*** *events on $(\Omega, \mathcal{A}, P)$. Then:*

$$\sum_n P(A_n) = \infty \Longrightarrow P[\limsup A_n] = 1 . \tag{1.3.32}$$

Proof. We show $P[(\limsup A_n)^c] = P[\liminf A_n^c] = 0$. From the inequality $1-x \le e^{-x} (x \in \mathbb{R})$, we see that for $m < M$,

$$P\Big[\bigcap_{k=m}^{M} A_k^c\Big] \overset{\text{independent}}{=} \prod_{k=m}^{M} P[A_k^c] = \prod_{k=m}^{M} (1 - P(A_k))$$
$$\le \exp\Big\{ -\sum_{k=m}^{M} P(A_k)\Big\} \underset{M\to\infty}{\longrightarrow} 0 \;\; (\text{using (1.3.32)}) \; .$$

It follows that $P[\bigcap_{k\ge m} A_k^c] = 0$, $\forall m \ge 1 \Longrightarrow P[\liminf A_n^c] = 0$. $\square$

Examples

1) Let $X_n, n \ge 1$, be a sequence of independent $N(0,\sigma^2)$-distributed random variables, with $\sigma > 0$. From the second lemma of Borel-Cantelli, it follows that

$$P\text{-a.s.}, \qquad \limsup_n X_n = \infty.$$

Proposition 1.27.

$$(1.3.33) \qquad P\text{-a.s.}, \qquad \limsup_n \frac{X_n}{\sigma\sqrt{2\log n}} = 1 \, .$$

The proof consists essentially of two steps, establishing first an upper bound, and then a lower bound. For that, we will make use of the following lemma.

Lemma 1.28. *For $x > 0$, one has*

$$(1.3.34) \qquad \Big(x + \frac{1}{x}\Big)^{-1} e^{-x^2/2} \overset{i)}{\le} \int_x^\infty e^{-y^2/2} dy \overset{ii)}{\le} \frac{1}{x}\, e^{-x^2/2} .$$

Proof.

i) $\frac{1}{x} e^{-x^2/2} = \int_x^\infty (1 + \frac{1}{y^2}) e^{-y^2/2} dy \le (1 + \frac{1}{x^2}) \int_x^\infty e^{-y^2/2} dy$. Moreover, one has also $x(1 + \frac{1}{x^2}) = x + \frac{1}{x}$, which establishes the first inequality.

ii) $\int_x^\infty e^{-y^2/2} dy \le \frac{1}{x} \int_x^\infty y e^{-y^2/2} dy = \frac{1}{x} e^{-x^2/2}$.

$\square$

Proof of Proposition 1.27.

First step: (upper bound)

$$(1.3.35) \qquad P\text{-a.s.}, \qquad \limsup_n \frac{X_n}{\sigma\sqrt{2\log n}} \le 1.$$

For the proof of this upper bound, we will use Lemma 1.24. Let us choose an arbitrary $\varepsilon > 0$, and set

$$A_n = \Big\{ X_n > (1+\varepsilon)\sigma\sqrt{2\log n}\Big\}, \;\; n \ge 1 \, .$$

Then, one has

$$
\begin{aligned}
P[A_n] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{(1+\varepsilon)\sigma\sqrt{2\log n}}^{\infty} e^{\frac{-y^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{(1+\varepsilon)\sqrt{2\log n}}^{\infty} e^{-y^2/2} dy \\
&\overset{(1.3.34)}{\leq} \frac{1}{\sqrt{2\pi}} \frac{1}{(1+\varepsilon)\sqrt{2\log n}} e^{-(1+\varepsilon)^2 \log n} = \frac{1}{\sqrt{2\pi}} \frac{1}{(1+\varepsilon)\sqrt{2\log n}} \frac{1}{n^{(1+\varepsilon)^2}},
\end{aligned}
$$

and so $\sum_n P[A_n] < \infty$. It follows from Lemma 1.24 that $P[\limsup_n A_n] = 0$, and thus

$$
P\text{-a.s.,} \qquad \text{for large } n, \ X_n \leq (1+\varepsilon)\sigma\sqrt{2\log n}.
$$

Hence,

$$
P\text{-a.s.,} \qquad \limsup_n \frac{X_n}{\sigma\sqrt{2\log n}} \leq (1+\varepsilon).
$$

(1.3.35) then follows with $\varepsilon \searrow 0$.

Second step: (lower bound)

$$
P\text{-a.s.,} \qquad \limsup_n \frac{X_n}{\sigma\sqrt{2\log n}} \geq 1. \tag{1.3.36}
$$

For the proof of this lower bound, we will use the second lemma of Borel-Cantelli. In a similar way as in the first step, we choose some $0 < \varepsilon < 1$, and set

$$
B_n = \left\{ X_n > (1-\varepsilon)\sigma\sqrt{2\log n} \right\}, \ n \geq 1.
$$

The events $B_n, n \geq 1$, are then independent, and

$$
\begin{aligned}
P[B_n] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{(1-\varepsilon)\sigma\sqrt{2\log n}}^{\infty} e^{\frac{-y^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{(1-\varepsilon)\sqrt{2\log n}}^{\infty} e^{-y^2/2} dy \\
&\overset{(1.3.34)}{\geq} \frac{1}{\sqrt{2\pi}} \left((1-\varepsilon)\sigma\sqrt{2\log n} + \frac{1}{(1-\varepsilon)\sigma\sqrt{2\log n}} \right)^{-1} e^{-(1-\varepsilon)^2 \log n} \\
&\geq \frac{1}{n^a}, \ \text{for } n \geq n_0(a,\varepsilon) \text{ and } (1-\varepsilon)^2 < a < 1\,.
\end{aligned}
$$

It follows that $\sum_n P[B_n] = \infty$: the second lemma of Borel-Cantelli implies that $P[\limsup_n B_n] = 1$, and thus

$$
P\text{-a.s.,} \qquad \limsup_n \frac{X_n}{\sigma\sqrt{2\log n}} \geq (1-\varepsilon)\,.
$$

By letting $\varepsilon \searrow 0$, we obtain (1.3.36). The claim (1.3.33) follows. □

2) (Length of the longest gap in a sequence of independent $0-1$ Bernoulli random variables with parameter $p = 1/2$).

We consider a sequence $X_i, i \geq 1$, of independent Bernoulli($\frac{1}{2}$)-distributed random variables on a probability space $(\Omega, \mathcal{A}, P)$.

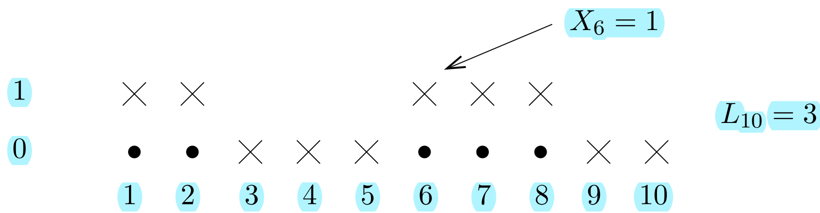


Fig. 1.5

The length $L_n(\omega)$ of the longest gap in the sequence $X_i(\omega)$, $1 \le i \le n$, is

$$
\begin{aligned}
L_n(\omega) \;=\; \max\big\{&m \in \{0,\dots,n\},\ \exists k \in \{1,\dots,n\}: \\
&k+m-1 \le n,\quad X_k(\omega) = X_{k+1}(\omega) = \dots = X_{k+m-1}(\omega) = 0\big\}\,.
\end{aligned}
\tag{1.3.37}
$$

Our goal is then to study the asymptotic behavior of L_n. The following proposition tells us at which speed $L_n(\omega)$ grows to infinity – its order of magnitude – for a typical ω:

Proposition 1.29.

$$
P\text{-}a.s.,\qquad \frac{L_n}{\log_2(n)} \underset{n\to\infty}{\longrightarrow} 1\,.
\tag{1.3.38}
$$

Proof.

First step: For all $\epsilon > 0$, one has

$$
P\text{-a.s.},\qquad \overline{\lim_n}\, L_n/\log_2(n) \le 1+\epsilon\,.
\tag{1.3.39}
$$

For $2^m \le n < 2^{m+1}$, $m \ge 1$, one has

$$
L_n > (1+\epsilon)\log_2(n) \Longrightarrow L_{2^{m+1}} > (1+\epsilon)\log_2(2^m) = (1+\epsilon)m\,.
$$

Hence,

$$
\begin{aligned}
&P\big[\,\overline{\lim_n}\, L_n/\log_2(n) > 1+\epsilon\big] \le P[\limsup_m A_m], \quad \text{where} \\
&A_m = \{L_{2^{m+1}} > (1+\epsilon)m\}\,.
\end{aligned}
\tag{1.3.40}
$$

Now, we have

$$
\begin{aligned}
P[A_m] &\le P\Big[\bigcup_{k=0}^{2^{m+1}-1} \{X_{k+1} = \dots = X_{k+[(1+\epsilon)m]} = 0\}\Big] \\
&\le 2^{m+1}\cdot P[X_1 = \dots = X_{[(1+\epsilon)m]} = 0] \overset{\text{independence}}{=} 2^{m+1}\Big(\frac{1}{2}\Big)^{[(1+\epsilon)m]} \\
&\le 2^{m+2}\Big(\frac{1}{2}\Big)^{(1+\epsilon)m} = 2^{2-\epsilon m}\,.
\end{aligned}
$$

Consequently, $\sum_m P(A_m) < \infty$, so that (1.3.39) follows from (1.3.40), using the first lemma of Borel-Cantelli.

Second step: Let us choose $\epsilon \in (0,1)$, then one has

$$\text{(1.3.41)} \qquad P\text{-a.s.}, \qquad \underline{\lim_n}\, L_n / \log_2(n) \geq (1-\epsilon)\,.$$

(1.3.41) indeed follows from

$$\text{(1.3.42)} \qquad \begin{aligned} &P[\liminf_m B_m] = 1, \text{ where } B_m \overset{\text{def.}}{=} \{\omega\text{: there exists “a gap} \\ &\text{of length} \geq [(1-\tfrac{\epsilon}{2})\,m] \text{ in the block } \{2^m+1,\dots,2^{m+1}\}\text{”}\} \\ &\overset{\text{def.}}{=} \bigcup_{2^m \leq k \leq 2^{m+1} - [(1-\frac{\epsilon}{2})m]} \{0 = X_{k+1} = X_{k+2} = \dots = X_{k+[(1-\frac{\epsilon}{2})m]}\}\,. \end{aligned}$$

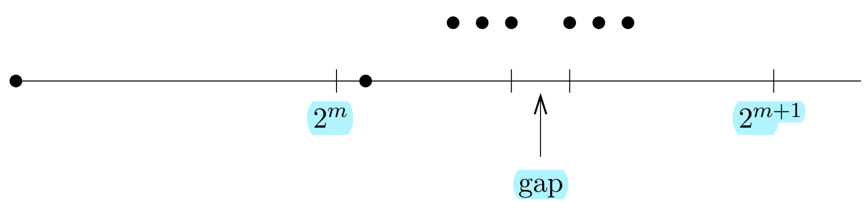


Fig. 1.6

Furthermore, we see that

$$\begin{aligned} B_m^c \subset \Big\{\omega:\ &\text{in each block } \Big\{2^m + \ell\big[\big(1-\frac{\epsilon}{2}\big)m\big] + 1, \\ &2^m + (\ell+1)\Big[\Big(1-\frac{\epsilon}{2}\Big)m\Big]\Big\},\ 0 \leq \ell < \Big[2^m/\Big[\Big(1-\frac{\epsilon}{2}\Big)m\Big]\Big], \\ &\text{one has } X_i = 1 \text{ for at least one } i\Big\}\,. \end{aligned}$$

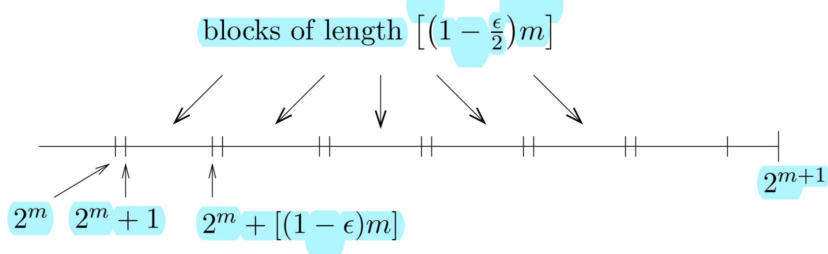


Fig. 1.7

These blocks are pairwise disjoint, which implies, thanks to (1.3.22):
for $a_m = \big[2^m/[(1-\frac{\epsilon}{2})m]\big]$ = number of blocks,

$$\begin{aligned} P[B_m^c] &\leq \Big(P\Big[X_i = 1, \text{ for one } i \in \Big\{1,\dots,\Big[\Big(1-\frac{\epsilon}{2}\Big)m\Big]\Big\}\Big]\Big)^{a_m} \\ &= \Big(1 - P\Big[X_i = 0, \text{ for all } i \in \Big\{1,\dots,\Big[\Big(1-\frac{\epsilon}{2}\Big)m\Big]\Big\}\Big]\Big)^{a_m} \\ &\leq \Big(1 - \Big(\frac{1}{2}\Big)^{(1-\frac{\epsilon}{2})m}\Big)^{a_m} = \exp\Big\{a_m \log\Big(1 - \Big(\frac{1}{2}\Big)^{(1-\frac{\epsilon}{2})m}\Big)\Big\}\,. \end{aligned}$$

Since

$$a_m \log\left(1 - \left(\frac{1}{2}\right)^{(1-\frac{\epsilon}{2})m}\right) \underset{m\to\infty}{\sim} -a_m \cdot \left(\frac{1}{2}\right)^{(1-\frac{\epsilon}{2})m} \underset{m\to\infty}{\sim} \frac{-2^{\frac{\epsilon}{2}\,m}}{[(1-\frac{\epsilon}{2})m]}\,,$$

we have $\sum P[B_m^c] < \infty \Longrightarrow P[\limsup B_m^c] = 0$ using the first lemma of Borel Cantelli, and (1.3.42) follows readily. $\square$

Kolmogorov's 0-1 Law

Let $X_i, i \geq 1$, be a sequence of random variables on $(\Omega, \mathcal{A}, P)$. For $n \geq 1$, let us define the σ-algebra $\mathcal{F}_n$, corresponding to the "**future of the sequence $(X_i)_{i\geq 1}$ after time n**", as:

(1.3.43)
$$\begin{aligned} \mathcal{F}_n &= \sigma(X_n, X_{n+1}, X_{n+2} \dots) \\ &\overset{\text{def.}}{=} \text{the smallest } \sigma\text{-algebra that contains all } \sigma(X_n), \\ &\quad\ \sigma(X_{n+1}),\ \sigma(X_{n+2}) \dots \end{aligned}$$

$\mathcal{F}_n = \sigma(\bigcup_{i\geq n} \sigma(X_i))$

(in other words, $\sigma(\bigcup_{i\geq n} \sigma(X_i))$). For $p > n$, one has

$$\sigma(X_n, X_{n+1}, \dots, X_p) \subseteq \sigma(X_n, X_{n+1}, X_{n+2} \dots)\,,$$

and for $n \geq 1$, $\sigma(X_n, X_{n+1}, X_{n+2} \dots)$ is the smallest σ-algebra that contains all $\sigma(X_n, \dots, X_p)$, $p > n$.

$= \sigma(\bigcup_{p\geq n} \sigma(X_n, \dots, X_p))$

For instance, all partial sums $\sum_{k=10}^{p} X_k$, with $p > 10$, are

$$\sigma(X_{10}, X_{11} \dots) - \text{measurable}\,.$$

One can then define the σ-algebra of the "**distant future of the sequence** $(X_i)_{i\geq 1}$", denoted by $\mathcal{F}_\infty$, as:

(1.3.44)
$$\mathcal{F}_\infty = \bigcap_{n\geq 1} \mathcal{F}_n \quad (\text{also called "}\textbf{asymptotic } \boldsymbol{\sigma}\textbf{-algebra}\text{"})\,.$$

For example, it contains the set of convergence (in $[-\infty, +\infty]$) of the series $\sum X_k$:

$$\Omega_1 = \left\{\omega \in \Omega,\ \underline{\lim}_p \sum_{k=1}^{p} X_k(\omega) = \overline{\lim}_p \sum_{k=1}^{p} X_k(\omega)\right\} \in \mathcal{F}_\infty\,,$$

since for each $n \geq 1$,

(1.3.45)
$$\begin{aligned} &\Omega_1 = \left\{\omega \in \Omega,\ \underline{\lim}_p \sum_{k=n}^{p} X_k(\omega) = \overline{\lim}_p \sum_{k=n}^{p} X_k(\omega)\right\} \in \mathcal{F}_n \\ \Longrightarrow\ &\Omega_1 \in \bigcap_{n\geq 1} \mathcal{F}_n = \mathcal{F}_\infty\,. \end{aligned}$$

Analogously, it also contains the set of convergence in $\mathbb{R}$ of the series $\sum X_k$:

$$\begin{aligned} \Omega_2 = \Omega_1 \cap & \Big\{\underline{\lim}_p \sum_{k=1}^p X_k > -\infty\Big\} \\ & \cap \Big\{\overline{\lim}_p \sum_{k=1}^p X_k < \infty\Big\} \in \mathcal{F}_\infty \,. \end{aligned} \tag{1.3.46}$$

In the case of **independent random variables** $X_i, i \geq 1$, we have a very particular property of the σ-algebra $\mathcal{F}_\infty$:

Theorem 1.30. *(Kolmogorov's 0-1 law)*

*If the $X_i, i \geq 1$, are independent, then the asymptotic σ-algebra $\mathcal{F}_\infty$ is **trivial**, i.e.*

$$A \in \mathcal{F}_\infty \Longrightarrow P(A) = 0 \ \text{or}\ 1\,. \tag{1.3.47}$$

For example, $P(\Omega_1) = 0$ or 1 for the set Ω_1 from (1.3.45).

Proof.

First step: Let $n > 1$ be fixed. We prove:

$$\begin{aligned} & \forall A \in \sigma(X_1, \dots, X_{n-1}),\ \forall B \in \mathcal{F}_n = \sigma(X_n, X_{n+1} \dots), \\ & \qquad P[A \cap B] = P[A]\, P[B]\,. \end{aligned} \tag{1.3.48}$$

Indeed, (1.3.48) holds for all $A \in \sigma(X_1, \dots, X_{n-1})$, $B \in \sigma(X_n, \dots, X_p)$ $(p \geq n)$, thanks to (1.3.23). However, $\sigma(X_1, \dots, X_{n-1})$ and $\bigcup_{p \geq n} \sigma(X_n, \dots, X_p)$ are π-systems, and $\sigma(\bigcup_{p\geq n} \sigma(X_n, \dots, X_p)) = \mathcal{F}_n$. Claim (1.3.48) follows using (1.3.17).

Second step:

$$\begin{aligned} & \forall A \in \sigma(X_1, \dots, X_n, \dots) = \mathcal{F}_1,\ \forall B \in \mathcal{F}_\infty, \ \text{ one has} \\ & \qquad P[A \cap B] = P[A]\, P[B]\,. \end{aligned} \tag{1.3.49}$$

Thanks to (1.3.48) and $\mathcal{F}_\infty \subseteq \mathcal{F}_n$, this equality holds for

$$\begin{aligned} & A \in \bigcup_{n>1} \sigma(X_1, \dots, X_{n-1}) \leftarrow \pi\text{-system}, \quad \text{and} \\ & B \in \mathcal{F}_\infty (\sigma\text{-algebra} \Longrightarrow \pi\text{-system})\,. \end{aligned}$$

Similarly, (1.3.49) follows from (1.3.17), and the fact that $\mathcal{F}_1$ is the smallest σ-algebra containing $\bigcup_{n>1} \sigma(X_1, \dots, X_{n-1})$.

We finally see that for $A = B \in \mathcal{F}_\infty \subset \mathcal{F}_1$, (1.3.49) implies that

$$P(A) = P(A \cap A) = P(A)^2 \Longrightarrow P(A) = 0 \ \text{ or } \ 1\,.$$

$\square$

Up to now, we have hardly discussed the question of existence for sequences of independent random variables. We conclude this section by mentioning the following result.

Theorem 1.31.

$$\text{(1.3.50)} \qquad \begin{array}{l} \textit{Let } \mu \textit{ be a probability measure on } (\mathbb{R}, \mathcal{B}(\mathbb{R})), \textit{ then there exists} \\ \textit{a probability space } (\Omega, \mathcal{A}, P) \textit{ with a sequence } X_i, i \geq 1, \\ \textit{of independent } \mu\textit{-distributed random variables on } (\Omega, \mathcal{A}, P)\,. \end{array}$$

An "abstract" proof of this statement, as a consequence of Kolmogorov's extension theorem and the construction of products of infinitely many measures, can be found in the Appendix of Durrett. We also refer the reader to Durrett, p. 26. A "concrete proof", with $\Omega = [0,1)$, $\mathcal{A} = \mathcal{B}([0,1))$, $P =$ Lebesgue measure on $[0,1)$, can also be given, as will be explained in an exercise.

1.4 Convergence of stochastic series

We consider a sequence $X_i, i \geq 1$, of **independent random variables** on $(\Omega, \mathcal{A}, P)$.

In this section, we would like to develop concrete criteria to determine whether the series $\sum_k X_k$ converges.

As in (1.3.46), we define the set of convergence in $\mathbb{R}$ of the series $\sum X_k$ as

$$\text{(1.4.1)} \qquad \begin{array}{l} \Omega_2 = \Big\{\omega \in \Omega,\ \underline{\lim}_p \displaystyle\sum_{k=1}^p X_k = \overline{\lim}_p \displaystyle\sum_{k=1}^p X_k \in \mathbb{R}\Big\} \in \mathcal{F}_\infty, \ \text{ where} \\ \mathcal{F}_\infty = \text{asymptotic } \sigma\text{-algebra of the } X_i, i \geq 1\,. \end{array}$$

Thanks to Kolmogorov's 0–1 law (1.3.47), we know that

$$\text{(1.4.2)} \qquad P(\Omega_2) = 0 \ \text{ or } \ 1\,.$$

Under which conditions on the $X_i, i \geq 1$, does one have $P(\Omega_2) = 1$?

Notation:

$$\text{(1.4.3)} \qquad S_0 = 0, \ \ S_n = \sum_{k=1}^n X_k, \ \ n \geq 1\,.$$

Kolmogorov's Inequality:

Let $X_1, \dots, X_n$ be independent random variables with $E[X_i^2] < \infty$ and $E[X_i] = 0$, then

$$\text{(1.4.4)} \qquad \forall u > 0,\ P\Big[\max_{1 \leq k \leq n} |S_k| \geq u\Big] \leq \frac{1}{u^2}\,\mathrm{Var}(S_n) = \frac{1}{u^2}\,\sum_{i=1}^n \mathrm{Var}(X_i)\,.$$

Remark 1.32. (1.4.4) is an example of a **maximal inequality**, i.e. the variance of the final term S_n of the sequence $S_0, S_1, \dots, S_n$ controls the behavior of $\max_{1 \leq k \leq n} |S_k|$.

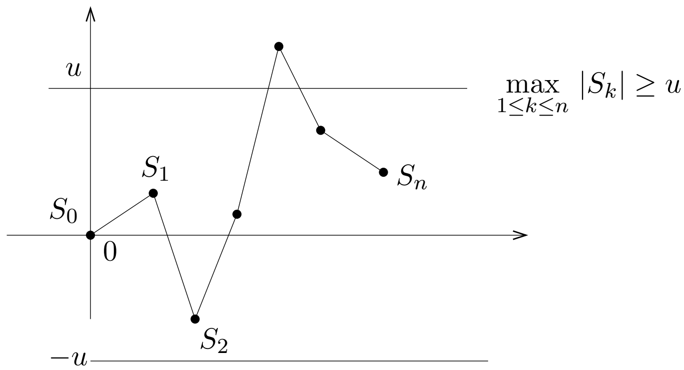


Fig. 1.8

$\square$

Proof. We first decompose $\{\max_{1 \le k \le n} |S_k| \ge u\}$ as

$$\{\max\ |S_k| \ge u\} = \bigcup_{k=1}^{n} \underbrace{A_k}_{\textbf{pairwise disjoint}}, \quad \text{where} \tag{1.4.5}$$

$$\begin{aligned} A_k &= \{\, |S_k| \ge u, \text{ and } |S_j| < u \text{ for all } j < k\} \\ &= \{\omega\colon \text{ the first time } j \text{ at which } |S_j| \ge u \text{ is exactly } k\}\ . \end{aligned} \tag{1.4.6}$$

One can easily see that $A_k \in \sigma(X_1, X_2, \dots, X_k)$ (later, we will link decompositions such as those in (1.4.5) - (1.4.6) with the notion of **stopping times**). Now, we have

$$\begin{aligned} E[S_n^2] &\ge \sum_{k=1}^{n} \int_{A_k} S_n^2\, dP = \sum_{k=1}^{n} \int_{A_k} (S_k + S_n - S_k)^2\, dP \\ &= \sum_{k=1}^{n} \int_{A_k} S_k^2 + 2S_k \cdot (S_n - S_k) + (S_n - S_k)^2\, dP \\ &\ge \sum_{k=1}^{n} \int_{A_k} S_k^2\, dP + 2 \sum_{k=1}^{n} \int_{A_k} S_k \cdot (S_n - S_k)\, dP\ . \end{aligned} \tag{1.4.7}$$

Let us note that

$$\begin{array}{ll} 1_{A_k} \cdot S_k & \text{is } \sigma(X_1, \dots, X_k)\text{-measurable, and} \\ (S_n - S_k) & \text{is } \sigma(X_{k+1}, \dots, X_n)\text{-measurable}. \end{array}$$

These two random variables are thus independent (see (1.3.23)), and

$$\int_{A_k} S_k(S_n - S_k)\, dP \overset{(1.3.14)}{=} E[1_{A_k}\, S_k]\ \underbrace{E[S_n - S_k]}_{\overset{\|}{0}} = 0\ .$$

We conclude from (1.4.7) that

$$
\begin{aligned}
E[S_n^2] &\geq \sum_{k=1}^n \int_{A_k} S_k^2 \, dP \\
&\geq \sum_{k=1}^n u^2 \, P[A_k] \ \ (\text{since } 1_{A_k} S_k^2 \geq u^2 \, 1_{A_k}, \text{ see (1.4.6)}) \\
&= u^2 \, P\Big[\bigcup_{k=1}^n A_k\Big] \overset{(1.4.5)}{=} u^2 \, P[\max_{1 \leq k \leq n} |S_k| \geq u] \, .
\end{aligned} \tag{1.4.8}
$$

Finally: $E[S_n] = 0 \Longrightarrow E[S_n^2] = \mathrm{Var}(S_n)$, and

$$
E[S_n^2] = E\Big[\Big(\sum_{k=1}^n X_k\Big)^2\Big] = E\Big[\sum_{k=1}^n X_k^2 + 2 \sum_{1 \leq k < k' \leq n} X_k \, X_{k'}\Big] \, , \tag{1.4.9}
$$

where for $k < k'$, $E[X_k \, X_{k'}] \overset{(1.3.14)}{=} E[X_k] \, E[X_{k'}] = 0$. Hence, we obtain that

$$
\mathrm{Var}(S_n) = E[S_n^2] = \sum_{k=1}^n E[X_k^2] = \sum_{k=1}^n \mathrm{Var}(X_k) \, .
$$

$\square$

Remark 1.33. A similar calculation leads to:

$$
\begin{aligned}
&\text{Let } Y_i \text{ be } \textbf{independent with } \boldsymbol{E[Y_i^2]} < \infty, \ 1 \leq i \leq n, \\
&\text{then } \mathbf{Var}\ \Big(\sum_{i=1}^n \boldsymbol{Y_i}\Big) = \sum_{i=1}^n \mathbf{Var}(\boldsymbol{Y_i}) \, .
\end{aligned} \tag{1.4.10}
$$

$\square$

As an application of Kolmogorov's inequality, we obtain the following theorem:

Theorem 1.34.

$$
\begin{aligned}
&\textit{Let } X_k, k \geq 1, \textit{ be independent with } \textstyle\sum \ \mathit{Var}(X_k) < \infty, \\
&\textit{and } E[X_k] = 0 \textit{ for all } k \geq 1. \textit{ Then } \textstyle\sum X_k \textit{ converges } P\textit{-a.s.}
\end{aligned} \tag{1.4.11}
$$

(i.e. $P[\Omega_2] = 1$, *see* (1.4.1)*).*

Remark 1.35. The $X_k, k \geq 1$, are pairwise orthogonal in $L^2(\Omega, \mathcal{A}, P)$, and since $\sum \|X_k\|_2^2 = \sum E(X_k^2) < \infty$, the series $\sum X_k$ also converges in $L^2(\Omega, \mathcal{A}, P)$. $\square$

Example 1.36. In the introduction (Chapter 0), we considered $\sum_{k \geq 1} \frac{Z_k}{k}$, where $Z_i, i \geq 1$, are independent with $P[Z_i = -1] = P[Z_i = 1] = \frac{1}{2}$. We have

$$
\sum_{k \geq 1} \mathrm{Var}\Big(\frac{Z_k}{k}\Big) = \sum_{k \geq 1} \frac{1}{k^2} < \infty \, ,
$$

and the series $\sum_k \frac{Z_k}{k}$ converges P-a.s. $\bigcirc$

Proof. We will show that $S_n, n \ge 0$, is a Cauchy sequence P-a.s. We define, for $M \ge 1$,

$$W_M = \sup_{m,n \ge M} |S_m - S_n| \,. \tag{1.4.12}$$

It holds that $W_M \downarrow W_\infty$ as $M \to \infty$, and (1.4.11) would follow from

$$P[W_\infty = 0] = 1 \;\; (\text{i.e. } W_\infty = 0 \;\; P\text{-a.s.}) \,. \tag{1.4.13}$$

Let us prove (1.4.13): for $\epsilon > 0$ and $M \ge 1$, one has

$$\sup_{m \ge M} \{ |S_m - S_M| \} \le \epsilon \Longrightarrow \underbrace{\sup_{m,n \ge M} \{ |S_m - S_n| \}}_{\substack{\| \\ W_M}} \le 2\epsilon \,.$$

Hence,

$$\begin{aligned} P[W_M > 2\epsilon] \;&\le\; P\big[\sup_{m \ge M} |S_m - S_M| > \epsilon\big] \\ &=\; \lim_{N \to \infty} \uparrow P\big[\sup_{M \le m \le N} |S_m - S_M| > \epsilon\big] \,, \end{aligned} \tag{1.4.14}$$

since

$$\big\{ \sup_{m \ge M} |S_m - S_M| > \epsilon \big\} = \bigcup_{N \ge M} \big\{ \sup_{M \le m \le N} |S_m - S_M| > \epsilon \big\} \,,$$

where the latter sets form an increasing sequence[2] in N.

From the equality $S_m - S_M = \sum_{k=1}^{m-M} X_{M+k}$, $m > M$, and Kolmogorov's inequality, we find:

$$\begin{aligned} P\big[\sup_{M \le m \le N} |S_m - S_M| > \epsilon\big] \;&\le\; \frac{1}{\epsilon^2} \sum_{k=M+1}^{N} \mathrm{Var}(X_k) \\ &\le\; \frac{1}{\epsilon^2} \sum_{k>M}^{\infty} \mathrm{Var}(X_k) \,. \end{aligned} \tag{1.4.15}$$

Thanks to (1.4.14), we obtain for $\epsilon > 0$, $M \ge 1$,

$$P[W_\infty > 2\epsilon] \le P[W_M > 2\epsilon] \le \frac{1}{\epsilon^2} \sum_{k>M} \mathrm{Var}(X_k) \underset{M \to \infty}{\longrightarrow} 0,$$

and (1.4.13) follows. $\square$

[2]We remind the reader that the notation "lim ↑", resp. "lim ↓", is often used to stress that one is taking the limit of a non-decreasing, resp. non-increasing, sequence.

Theorem 1.37. *(Kolmogorov's Three-Series Theorem)*

Let $X_k, k \geq 1$, be independent random variables, $A > 0$, and $Y_k = X_k\ 1\{\,|X_k| \leq A\}$. Statements (1.4.16) *and* (1.4.17) *are then equivalent:*

$$\sum_k X_k \text{ converges } P\text{-a.s.} \tag{1.4.16}$$

$$\begin{aligned} &\text{i)} \quad \sum_{k=1}^{\infty} P(\,|X_k| > A) < \infty\,, \\ &\text{ii)} \quad \sum_k E[Y_k] \text{ converges}\,, \\ &\text{iii)} \quad \sum_{k=1}^{\infty} \mathrm{Var}(Y_k) < \infty\,. \end{aligned} \tag{1.4.17}$$

Example 1.38. Consider

$$X_k = \frac{Z_k}{k^\alpha},\ k \geq 1, \text{ with } \alpha > 0 \text{ and } Z_k,\ k \geq 1, \text{ independent with}$$
$$P[Z_i = -1] = P[Z_i = 1] = \tfrac{1}{2}\,.$$

Let us apply the theorem above with $A = 1$. Then

$$Y_k = X_k,\ P[\,|X_k| > 1] = 0,\ E[Y_k] = 0, \text{ and } \mathrm{Var}(Y_k) = \frac{1}{k^{2\alpha}}\,.$$

Hence, i), ii), iii) satisfied $\Longleftrightarrow \alpha > \frac{1}{2}$, and

$$\begin{aligned} \sum X_k \quad &\text{converges } P\text{-a.s. for } \alpha > \frac{1}{2}\,, \\ &\text{diverges } \quad P\text{-a.s. for } \alpha \leq \frac{1}{2}\,. \end{aligned}$$

○

Proof. We only prove (1.4.17) $\Longrightarrow$ (1.4.16) (the converse (1.4.16) $\Longrightarrow$ (1.4.17) is a bit more complicated, it can be proved using martingales, defined in Chapter 3).

Let us define $\widetilde{Y}_k = Y_k - E[Y_k]$. Then the $\widetilde{Y}_k$ are independent, $E[\widetilde{Y}_k] = 0$, and $\sum \mathrm{Var}(\widetilde{Y}_k) = \sum \mathrm{Var}(Y_k) < \infty$ (using iii)). Thanks to (1.4.11), one has $\sum \widetilde{Y}_k$ converges P-a.s., from which it follows that (using also ii))

$$\sum Y_k = \sum \widetilde{Y}_k + \sum E[Y_k] \text{ converges } P\text{-a.s.} \tag{1.4.18}$$

Thanks to the first lemma of Borel Cantelli and i), it now follows that

$$P\big[\liminf\ |X_k| \leq A\big] = 1 - P\big[\limsup\ |X_k| > A\big] = 1\,. \tag{1.4.19}$$

On the set $\liminf \{|X_k| \leq A\}$, it holds that

$$\sum X_k(\omega) \text{ converges} \Longleftrightarrow \sum Y_k(\omega) \text{ converges}\,.$$

Thanks to (1.4.18) and (1.4.19), $P\big[\sum X_k(\omega) \text{ converges}\big] = 1$. □

1.5 Law of Large Numbers

In this section, we consider sequences $X_k, k \geq 1$, of random variables that all possess the same distribution. We want to investigate the asymptotic behavior of

$$\frac{S_n}{n}\,, \tag{1.5.1}$$

where

$$S_0 = 0,\;\; S_n = \sum_{k=1}^{n} X_k,\;\; n \geq 1\,. \tag{1.5.2}$$

Terminology:

Let $Y_n, n \geq 1$, and Y be random variables on $(\Omega, \mathcal{A}, P)$. The sequence Y_n is said to **converge in probability** to Y (**Notation:** $Y_n \xrightarrow{P} Y$) if:

$$\forall \epsilon > 0,\;\; \lim_{n\to\infty} P[\,|Y_n - Y| \geq \epsilon] = 0\,. \tag{1.5.3}$$

We will discuss a **weak** and a **strong** Law of Large Numbers.

For the weak law, we show that under certain hypotheses, $\frac{S_n}{n}$ converges in probability. For the strong law, it is then proved that $\frac{S_n}{n}$ converges P-a.s.

The terminology comes from the observation that

$$Y_n \to Y \;\; P\text{-a.s.} \;\; \Longrightarrow Y_n \xrightarrow{P} Y \tag{1.5.4}$$

since:

$$\lim_n\; P[\,|Y_n - Y| \geq \epsilon] = \lim_n \int 1\{\,|Y_n - Y| \geq \epsilon\}\, dP = 0$$

using the dominated convergence theorem (1.2.12).

Using Chebyshev's inequality, one has

$$\begin{aligned} &P[\,|Y_n - Y| \geq \epsilon] \leq \epsilon^{-p}\, E[\,|Y_n - Y|^p],\;\; \epsilon > 0,\; p \in [1,\infty), \\ &\text{so that}\quad Y_n \xrightarrow{L^p} Y \Longrightarrow Y_n \xrightarrow{P} Y\,. \end{aligned} \tag{1.5.5}$$

Weak Law of Large Numbers:

Let $X_k, k \geq 1$, be identically distributed, uncorrelated (i.e. $E[X_k^2] < \infty$, and $\mathrm{Cov}(X_k, X_{k'}) \overset{\text{def.}}{=} E\big[(X_k - E[X_k])(X_{k'} - E[X_{k'}])\big] = 0$ for $k \neq k'$) random variables. Then:

$$\frac{S_n}{n} \;\; \begin{array}{l}\text{converges in } L^2 \text{ (and thus also in probability)} \\ \text{to } \mu = E[X_k], k \geq 1\,.\end{array} \tag{1.5.6}$$

Proof. Let us write $\frac{S_n}{n} - \mu = \frac{1}{n} \sum_{k=1}^n \widetilde{X}_k$, where $\widetilde{X}_k = X_k - \mu$. Then,

$$\begin{aligned}
E\Big[\Big(\frac{S_n}{n} - \mu\Big)^2\Big] &= \frac{1}{n^2}\, E\Big[\Big(\sum_{k=1}^n \widetilde{X}_k\Big)^2\Big] \\
&= \frac{1}{n^2}\,\Big(E\Big[\sum_{k=1}^n \widetilde{X}_k^2\Big] + 2\underbrace{E\Big[\sum_{1\le k<k'\le n} \widetilde{X}_k\, \widetilde{X}_{k'}\Big]}_{\overset{\|}{0} \text{ since the } X_k \text{ are uncorrelated}}\Big) \\
&= \frac{n}{n^2}\, E[\widetilde{X}_1^2] = \frac{\mathrm{Var}(X_1)}{n} \to 0\,.
\end{aligned}$$

□

Example 1.39. (Shannon's theorem)

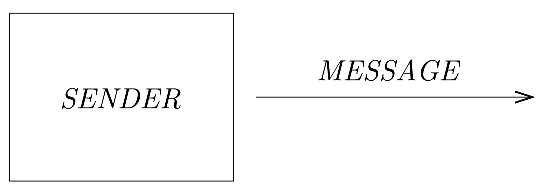


Fig. 1.9

A person sends a message, modeled by a sequence $X_1, X_2, \dots$ of independent random variables with values in $\{1, \dots, r\}$ ("finite alphabet"), and a common distribution $0 < p(k) = P[X = k]$, $k = 1, \dots, r$.

For $n \ge 1$, we consider

$$\begin{aligned}
\pi_n(\omega) &= p(X_1(\omega)) \cdot p(X_2(\omega)) \cdots p(X_n(\omega)) \\
&= \text{probability to observe the exact sequence} \\
&\quad\; (X_1(\omega), \dots, X_n(\omega)) \text{ for the first } n \text{ "letters".}
\end{aligned}$$

Thanks to (1.5.6),

$$\begin{aligned}
-\frac{1}{n}\, \log \pi_n(\omega) &= -\frac{1}{n} \sum_{k=1}^n \log p(X_k(\omega)) \xrightarrow[n\to\infty]{P} E[-\log p(X_1)] \\
&= -\sum p(k) \log p(k) \overset{\text{def.}}{=} H\,.
\end{aligned} \tag{1.5.7}$$

H is called the sender's **entropy per character**. It is a **measure of the quantity of information** contained in the message.

Thanks to (1.5.7), for $\epsilon > 0$ fixed and large n,

$$\pi_n(\omega) \in [\exp\{-n(H+\epsilon)\},\ \exp\{-n(H-\epsilon)\}]$$

with probability $\ge 1 - \epsilon$.

Strong Law of Large Numbers (Etemadi (1981)):

Let $X_k, k \geq 1$, be **pairwise independent**, identically distributed random variables, with $E[\,|X_k|\,] < \infty$. Then

$$\frac{S_n}{n} \longrightarrow E[X_1] \qquad P\text{-a.s.} \tag{1.5.8}$$

○

Remark 1.40. The **first proof** of the Strong Law of Large Numbers (**Kolmogorov**) is less general, it assumes that the random variables X_k are independent[3] and identically distributed, with $E[\,|X_k|\,] < \infty$. The original proof is based on the three-series theorem, and the connection between stochastic series and the Law of Large Numbers comes from Kronecker's lemma (see Durrett, pp. 51-53). □

Lemma 1.41. *Let $a_n, n \geq 1$, and $x_n, n \geq 1$, be two sequences of real numbers with $a_n \uparrow \infty$. If $\sum_{n\geq 1} \frac{x_n}{a_n}$ converges, then $\frac{1}{a_n}(\sum_{k=1}^n x_k)$ converges to 0.*

Proof of (1.5.8).

First reduction:

$X_k^+ \overset{\text{def.}}{=} \max(X_k, 0)$, $k \geq 1$, and $X_k^- \overset{\text{def.}}{=} \max(-X_k, 0)$ are two sequences of pairwise independent, identically distributed, integrable random variables. If we prove (1.5.8) for X_k^+ and X_k^-, instead of X_k, then (1.5.8) will hold for X_k, $k \geq 1$, as well (by writing $X_k = X_k^+ - X_k^-$). Therefore:

(1.5.9) From now on, we will assume (without loss of generality) that $X_k \geq 0$.

Second reduction: (this reduction is not needed if $E[X_1^2] < \infty$).

Define $Y_k = X_k \, 1\{X_k \leq k\}$ ("truncated" variable), and $A = \liminf\{Y_k = X_k\}$. Then

$$P(A) = 1 - P[\limsup\{Y_k \neq X_k\}], \text{ and } \sum_k P[Y_k \neq X_k] \leq \sum_k P[X_k \geq k]$$

$$\overset{\substack{\text{identically} \\ \text{distributed}}}{=} \sum_k E[1\{X_1 \geq k\}] \overset{\substack{\text{monotone} \\ \text{convergence}}}{=} E\Big[\underbrace{\sum_{k\geq 1} 1\{k \leq X_1\}}_{=[X_1]\leq X_1}\Big] \leq E[X_1] < \infty\,.$$

Using the first lemma of Borel-Cantelli, one has $P(A^c) = 0$, and so $P(A) = 1$. For $\omega \in A$, one has

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow E[X_1] \Longleftrightarrow \frac{Y_1 + \cdots + Y_n}{n} \longrightarrow E[X_1]\,.$$

Our claim (1.5.8) would follow, under assumption (1.5.9), from

$$\frac{T_n}{n} \longrightarrow E[X_1] \;\; P\text{-a.s., where } \; T_0 = 0,\; T_n = Y_1 + \cdots + Y_n\,. \tag{1.5.10}$$

[3] Note that this is a much stronger hypothesis than simply pairwise independence.

Third reduction: (1.5.10) follows from

$$\text{(1.5.11)} \qquad \text{for all fixed } \alpha > 1, \quad \frac{T_{\overset{\text{integer part}}{[\alpha^n]}}}{[\alpha^n]} \underset{n}{\longrightarrow} E[X_1] \qquad P\text{-a.s.}$$

Indeed, consider $\alpha_M = 1 + \frac{1}{M}$, and define

$$\widetilde{\Omega} = \bigcap_{M \geq 1} \Omega_M, \quad \text{where } \Omega_M = \Big\{\omega \in \Omega, \lim_n \frac{T_{[\alpha_M^n]}}{[\alpha_M^n]} = E[X_1]\Big\}$$

$$P[\Omega_M] = 1 \ \forall M \geq 1 \Longrightarrow P[\widetilde{\Omega}] = 1 \ .$$

Consider now $\omega \in \widetilde{\Omega}$ and $\alpha_M = 1 + \frac{1}{M}$ fixed.

Notation: $k(n) \overset{\text{def.}}{=} [\alpha_M^n]$, $n \geq 1$. Then, since $Y_i \geq 0$, there holds

$$\text{(1.5.12)} \qquad \frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)} \quad \text{for } k(n) \leq m < k(n+1) \ .$$

Clearly, one also has $\lim_n \frac{k(n+1)}{k(n)} = \alpha_M$, since

$$\frac{\alpha_M^{n+1} - 1}{\alpha_M^n} \leq \frac{k(n+1)}{k(n)} \leq \frac{\alpha_M^{n+1}}{\alpha_M^n - 1} \ .$$

From (1.5.11) - (1.5.12), we obtain, for $\omega \in \widetilde{\Omega}$ and $M \geq 1$:

$$\frac{1}{\alpha_M} E[X_1] \leq \varliminf_n \frac{T_n}{n} \leq \varlimsup_n \frac{T_n}{n} \leq \alpha_M E[X_1] \ ,$$

and it follows, for $M \to \infty$, that for $\omega \in \widetilde{\Omega}$, $\lim_n \frac{T_n}{n} = E[X_1]$.

Hence, there only remains to prove (1.5.11):

Thanks to Chebyshev's inequality and the pairwise independence hypothesis, one has, for $\epsilon > 0$,

$$\begin{aligned}
&\sum_{n=1}^{\infty} P[\,|T_{k(n)} - E[T_{k(n)}]\,| > \epsilon k(n)] \leq \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{\mathrm{Var}(T_{k(n)})}{k(n)^2} \\
\text{(1.5.13)} \qquad &= \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{k(n)^2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) \\
&\overset{\text{Fubini}}{=} \frac{1}{\epsilon^2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \cdot \sum_{k(n) \geq m} \frac{1}{k(n)^2} \ .
\end{aligned}$$

For $n \geq 1$, one has $k(n) = [\alpha^n] \geq \frac{\alpha^n}{2}$ (if $k(n) \geq 2$: $k(n) \geq \alpha^n - 1$; if $k(n) = 1$: clear). Consequently,

$$\sum_{\underbrace{k(n)\geq m}_{\substack{\Updownarrow \\ \alpha^n \geq m}}} \frac{1}{k(n)^2} \leq 4 \sum_{\alpha^n \geq m} \alpha^{-2n} = \frac{4\alpha^{-2n_0(m)}}{(1-\alpha^{-2})} \leq \frac{4}{(1-\alpha^{-2})m^2}\,, \tag{1.5.14}$$

where $n_0(m) =$ the smallest n with $\alpha^n \geq m$.

From (1.5.13) and (1.5.14), we get:

$$\sum_{n=1}^{\infty} P[\,|T_{k(n)} - E[T_{k(n)}]\,| > \epsilon k(n)] \leq \frac{4}{\epsilon^2(1-\alpha^{-2})} \sum_{m=1}^{\infty} \frac{E[Y_m^2]}{m^2}\,. \tag{1.5.15}$$

Remark 1.42. If $E[X_1^2] < \infty$, one has

$$\sum_{m=1}^{\infty} \frac{E[Y_m^2]}{m^2} \leq \sum_{m=1}^{\infty} \frac{E[X_1^2]}{m^2} < \infty\,.$$

One can conclude, using (1.5.15) and the Borel-Cantelli lemma, that

$$\overline{\lim_n} \;\Big|\, \frac{T_{k(n)}}{k(n)} - \frac{E[T_{k(n)}]}{k(n)} \,\Big| \leq \epsilon \qquad P\text{-a.s.} \tag{1.5.16}$$

Actually, in the case $E[X_1^2] < \infty$, one does not need to introduce the variables Y_k at all; one can directly work with the variables X_k. One can write simply

$$P\text{-a.s.} \qquad \overline{\lim_n} \;\Big|\, \frac{S_{k(n)}}{k(n)} - \frac{E[S_{k(n)}]}{k(n)} \,\Big| = \overline{\lim_n} \;\Big|\, \frac{S_{k(n)}}{k(n)} - E[X_1] \,\Big| \leq \epsilon,$$

and obtain the analogue of (1.5.11) in this way. □

In the general case, one has:

Lemma 1.43.

$$\sum_{m=1}^{\infty} \frac{E[Y_m^2]}{m^2} < \infty\,, \tag{1.5.17}$$

$$\frac{E[T_n]}{n} \underset{n\to\infty}{\longrightarrow} E[X_1]\,. \tag{1.5.18}$$

We first complete the proof of (1.5.11): using (1.5.15) and (1.5.17), we deduce (1.5.16). Thanks to (1.5.18), we obtain, for $\epsilon > 0$,

$$P\Big[\overline{\lim_n} \;\Big|\, \frac{T_{k(n)}}{k(n)} - E[X_1] \,\Big| \leq 2\epsilon\Big] = 1 \Longrightarrow P\Big[\bigcap_{M\geq 1} \Big\{\overline{\lim_n} \;\Big|\, \frac{T_{k(n)}}{k(n)} - E[X_1] \,\Big| \leq \frac{2}{M}\Big\}\Big] = 1,$$

and (1.5.11) follows. □

Proof of Lemma 1.43.

(1.5.18): $\frac{E[T_n]}{n} = \frac{E[Y_1]+\cdots+E[Y_n]}{n}$, and

$$0 \le E[X_1] - E[Y_k] = E[X_1\, 1\{X_1 > k\}] \underset{k\to\infty}{\overset{\substack{\text{dominated}\\ \text{convergence}}}{\longrightarrow}} 0\,.$$

We deduce $E[Y_k] \underset{k\to\infty}{\longrightarrow} E[X_1]$ and $\frac{E[T_n]}{n} \underset{\text{(Cesaro convergence)}}{\longrightarrow} E[X_1]$.

$$\text{(1.5.17):} \quad E[Y_m^2] = E\Big[2 \int_0^\infty 1(y \le Y_m) y\, dy\Big] \overset{\text{Fubini}}{=} 2 \int_0^\infty y\, P[Y_m \ge y] dy$$
$$= 2 \int_0^m y\, P[Y_m \ge y] dy \quad \text{(since } Y_m \le m\text{)}.$$

Hence, $E[Y_m^2] \le 2 \int_0^m y\, P[X_1 \ge y] dy$ (using $Y_m = X_m\, 1\{X_m \le m\} \le X_m$).

From this, we deduce

$$\text{(1.5.19)} \quad \begin{aligned} \sum_{m=1}^\infty \frac{E[Y_m^2]}{m^2} &\le 2 \sum_{m=1}^\infty \frac{1}{m^2} \int_0^m y\, P[X_1 \ge y] dy \\ &\overset{\substack{\text{monotone}\\ \text{convergence}}}{=} 2 \int_0^\infty \sum_{m \ge 1} \frac{1}{m^2}\, 1(y \le m) \cdot y\, P[X_1 \ge y] dy\,. \end{aligned}$$

For $y \ge 2$, we obtain an upper bound for $\sum_{m \ge y} \frac{1}{m^2}$:

$$\sum_{m \ge y} \frac{1}{m^2} \le \sum_{m \ge y} \int_{m-1}^m \frac{dx}{x^2} \le \int_{y-1}^\infty \frac{dx}{x^2} = \frac{1}{(y-1)}\,.$$

Furthermore, $\sum_{m=1}^\infty \frac{1}{m^2} = 1 + \sum_{m=2}^\infty \frac{1}{m^2} \le 1 + 1 = 2$. Using (1.5.19):

$$\begin{aligned} \sum_{m=1}^\infty \frac{E[Y_m^2]}{m^2} &\le 2 \int_0^2 2 \underbrace{y}_{\le 2} P[X_1 \ge y] dy + 2 \int_2^\infty \Big(\frac{y}{y-1}\Big) P[X_1 \ge y]\, dy \\ &\le 8 \int_0^\infty P[X_1 \ge y]\, dy \\ &\overset{\text{Fubini}}{=} 8E\Big[\int_0^{X_1} dy\Big] = 8E[X_1]\,. \end{aligned}$$

Claim (1.5.17) thus follows. □

Remark 1.44.

- It was sufficient to consider geometric subsequences (1.5.11).
- Even if X_1 does not necessarily possess a finite second moment, one can still work with the truncated variables Y_k, and although $E[Y_k^2]$ diverges, it does not do so too fast (1.5.17).

$\square$

We next discuss an application of the strong Law of Large Numbers:

Example 1.45. (renewal process)

We consider random variables $X_1, X_2, \dots$ which are independent, identically distributed, positive (i.e. $X_i > 0$), with $E[X_i] < \infty$. The variables

$$T_n = X_1 + \cdots + X_n,\ n \geq 1\,, \tag{1.5.20}$$

model, for instance, the successive arrival times of customers in a queue, or occurrence times of failures in an electrical system. Set

$$N_t(\omega) = \sum_{n \geq 1} 1\{T_n(\omega) \leq t\} = \sup\{n \geq 0,\ T_n(\omega) \leq t\} \qquad \text{(with the convention : } T_0 = 0) \tag{1.5.21}$$

to be the number of customers (or failures) that arrived (occurred) up to time t. $\bigcirc$

Proposition 1.46.

$$P\text{-}a.s. \qquad \frac{N_t(\omega)}{t} \underset{t\to\infty}{\longrightarrow} \frac{1}{E[X_1]}\,. \tag{1.5.22}$$

Proof. Thanks to the strong Law of Large Numbers,

$$\frac{T_n(\omega)}{n} \longrightarrow E[X_1] \qquad P\text{-a.s.} \tag{1.5.23}$$

Consider some fixed ω for which (1.5.23) is satisfied. Then, for $t > 0$,

$$T_{N_t(\omega)} \leq t < T_{N_t(\omega)+1}$$

using (1.5.21), and we deduce

$$\frac{T_{N_t(\omega)}}{N_t(\omega)} \leq \frac{t}{N_t(\omega)} \leq \frac{T_{N_t(\omega)+1}}{N_t(\omega)+1}\ \frac{N_t(\omega)+1}{N_t(\omega)}\,. \tag{1.5.24}$$

Since $N_t(\omega) \to \infty$ for $t \to \infty$, we obtain, by (1.5.23) and (1.5.24),

$$\lim_{t\to\infty} \frac{t}{N_t(\omega)} = E[X_1] \qquad P\text{-a.s.}$$

$\square$

2 Central Limit Theorem, characteristic functions

2.1 Motivation, goals

We know, from the strong Law of Large Numbers, that for a sequence of independent, identically distributed, integrable random variables X_k, $k \geq 1$, with $E[X_k] = 0$,

$$\frac{S_n}{n} \longrightarrow 0 \;\; P\text{-a.s.}\,,$$

where

$$S_0 = 0, \;\; S_n = X_1 + \dots + X_n \,.$$

The Law of Large Numbers says that S_n grows sublinearly in n. A natural question that we can then ask ourselves is: at which speed (with which order of magnitude) does S_n increase?

We examine **a concrete example** first, with

$$P[X_k = 1] = P[X_k = -1] = \tfrac{1}{2} \,. \tag{2.1.1}$$

We define

$$Z_n = \frac{S_n}{\sqrt{n}}, \;\; n \geq 1 \,. \tag{2.1.2}$$

One has

$$E[Z_n] = 0 \quad \text{and} \quad \mathrm{Var}(Z_n) = \frac{1}{n}\,\mathrm{Var}(S_n) = \frac{n}{n}\,\mathrm{Var}(X_1) = 1 \,. \tag{2.1.3}$$

It is thus plausible that the right order of magnitude for Z_n neither decreases nor increases in n, i.e. that it stays ~ 1. In order to study this more closely, let us introduce a sequence $k(n) \in \mathbb{Z}$ with

$$\frac{2k(n)}{\sqrt{2n}} \underset{n\to\infty}{\longrightarrow} x \in \mathbb{R}, \;\; |k(n)| \leq n, \tag{2.1.4}$$

and consider

$$P\Big[Z_{2n} = \underbrace{\frac{2k(n)}{\sqrt{2n}}}_{\text{close to } x}\Big] = P[S_{2n} = 2k(n)] =$$

$$P\,[\underbrace{X_1 + \dots + X_{2n} = 2k(n)}_{\substack{\text{number of } 1:\;\; n+k(n) \\ \text{number of } -1:\;\; n-k(n)}}] = \binom{2n}{n+k(n)}\, 2^{-2n} \,.$$

It comes from Stirling's formula that

$$m! \overset{\text{def.}}{=} 1 \cdot 2 \cdot 3 \dots (m-1)m \underset{m\to\infty}{\sim} m^m\, e^{-m}\, \sqrt{2\pi m} \,, \tag{2.1.5}$$

where $a_m \sim b_m$ means $\frac{a_m}{b_m} \underset{m\to\infty}{\longrightarrow} 1$. Hence,

$$\binom{2n}{n+k(n)} = \frac{(2n)!}{(n+k(n))!(n-k(n))!} \sim \text{ (thanks to } \tfrac{k(n)}{n} \to 0 \text{ and (2.1.5))}$$

$$\frac{(2n)^{2n}}{(n+k(n))^{n+k(n)}(n-k(n))^{n-k(n)}} \ \frac{\sqrt{2\pi(2n)}}{\sqrt{2\pi(n+k(n))}\,\sqrt{2\pi(n-k(n))}} \sim$$

$$2^{2n}\left(1+\frac{k(n)}{n}\right)^{-(n+k(n))}\left(1-\frac{k(n)}{n}\right)^{-(n-k(n))}(\pi n)^{-1/2}\,.$$

It follows that

$$P\Big[Z_{2n} = \frac{2k(n)}{\sqrt{2n}}\Big] \ \overset{(2.1.4)}{\underset{n\to\infty}{\sim}} \ \frac{1}{\sqrt{\pi n}} \cdot \left(1+\frac{k(n)}{n}\right)^{-(n+k(n))}\left(1-\frac{k(n)}{n}\right)^{-(n-k(n))}. \tag{2.1.6}$$

We note that

$$\left(1+\frac{k(n)}{n}\right)^{n+k(n)}\left(1-\frac{k(n)}{n}\right)^{n-k(n)} = \left(1-\frac{k^2(n)}{n^2}\right)^{n}\left(1+\frac{k(n)}{n}\right)^{k(n)}\left(1-\frac{k(n)}{n}\right)^{-k(n)},$$

and $$\left(1-\frac{k^2(n)}{n^2}\right)^{n} = \exp\Big\{\underbrace{n\log\left(1-\frac{k^2(n)}{n^2}\right)}\Big\}\,.$$

$$\underset{n\to\infty}{\sim} -n\,\frac{k^2(n)}{n^2} = -\frac{k^2(n)}{n} \overset{(2.1.4)}{\longrightarrow} -\frac{x^2}{2}\text{ , using } \log(1+x) = x+o(x) \text{ as } x\to 0$$

Hence, $(1-\frac{k^2(n)}{n^2})^n \underset{n\to\infty}{\longrightarrow} \exp\{-\frac{x^2}{2}\}$. In an analogue way,

$$\left(1+\frac{k(n)}{n}\right)^{k(n)} \longrightarrow \exp\Big\{\frac{x^2}{2}\Big\} \quad \text{since } \ k(n)\cdot\frac{k(n)}{n} \longrightarrow \frac{x^2}{2}\,,$$

$$\text{and} \quad \left(1-\frac{k(n)}{n}\right)^{-k(n)} \longrightarrow \exp\Big\{\frac{x^2}{2}\Big\} \quad \text{since } \ k(n)\cdot\frac{k(n)}{n} \longrightarrow \frac{x^2}{2}.$$

From this, we obtain

$$P\Big[Z_{2n} = \frac{2k(n)}{\sqrt{2n}}\Big] \sim \frac{1}{(\pi n)^{1/2}}\ \exp\Big\{-\frac{x^2}{2}\Big\}\,, \tag{2.1.7}$$

or, in other words,

$$\lim_{n\to\infty}\ \sqrt{\frac{n}{2}}\ P\Big[Z_{2n} = \frac{2k(n)}{\sqrt{2n}}\Big] = \frac{1}{\sqrt{2\pi}}\ \exp\Big\{-\frac{x^2}{2}\Big\}\,. \tag{2.1.8}$$

It is therefore **plausible** (here, in order to argue rigorously, we would need to prove some uniform convergence in (2.1.8), for values of x in a bounded interval) that for $a < b$,

$$P[a \le Z_{2n} \le b] = \sum_{2k \in [a\sqrt{2n}, b\sqrt{2n}]} \sqrt{\frac{2}{n}}\ \sqrt{\frac{n}{2}}\ P\Big[Z_{2n} = \frac{2k}{\sqrt{2n}}\Big] \overset{(2.1.8)}{\approx}$$

$$\sum_{2k \in [a\sqrt{2n}, b\sqrt{2n}]} \sqrt{\frac{2}{n}}\ \frac{1}{\sqrt{2\pi}}\ \exp\Big\{-\frac{1}{2}\Big(\underbrace{\frac{2k}{\sqrt{2n}}}_{\overset{\parallel}{\sqrt{\frac{2}{n}} \cdot k}}\Big)^2\Big\} \xrightarrow{\text{Riemann sum}} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx\,.$$

In this way, we obtain an almost rigorous proof of the **de Moivre Laplace theorem** (de Moivre: 1667-1754, Laplace 1749-1827):

For $a < b$, one has

$$P[a \le Z_n \le b] \underset{n \to \infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx\,. \tag{2.1.9}$$

It is clear that the combinatorial argument above is not general. It is the goal of this chapter to develop general methods that allow one to prove statements like (2.1.9).

Let us now describe quickly the **strategy** that we will apply.

We will first introduce the **characteristic function** of Z_n:

$$\varphi_{Z_n}(t) \overset{\text{def.}}{=} E[\exp\{i\, t\, Z_n\}]\,. \tag{2.1.10}$$

One has

$$\begin{aligned}
\varphi_{Z_n}(t) &= E\Big[\exp\Big\{i\,\frac{t}{\sqrt{n}}\,(X_1 + \dots + X_n)\Big\}\Big] \\
&= E\Big[\prod_{k=1}^n \exp\Big\{i \frac{t}{\sqrt{n}}\, X_k\Big\}\Big] \underset{(1.3.14)}{\overset{\text{independence}}{=}} \prod_{k=1}^n E\Big[\exp\Big\{i\,\frac{t}{\sqrt{n}}\, X_k\Big\}\Big] \\
&= E\Big[\exp\Big\{i\,\frac{t}{\sqrt{n}}\, X_1\Big\}\Big]^n \overset{(2.1.1)}{=} \Big(\cos\,\frac{t}{\sqrt{n}}\Big)^n,
\end{aligned}$$

and for $t \in \mathbb{R}$:

$$\cos\frac{t}{\sqrt{n}} \overset{\substack{\text{Taylor} \\ \text{expansion}}}{=} 1 - \frac{t^2}{2n} + \frac{\epsilon(n)}{n}\,, \quad \epsilon(n) \overset{n \to \infty}{\longrightarrow} 0\,.$$

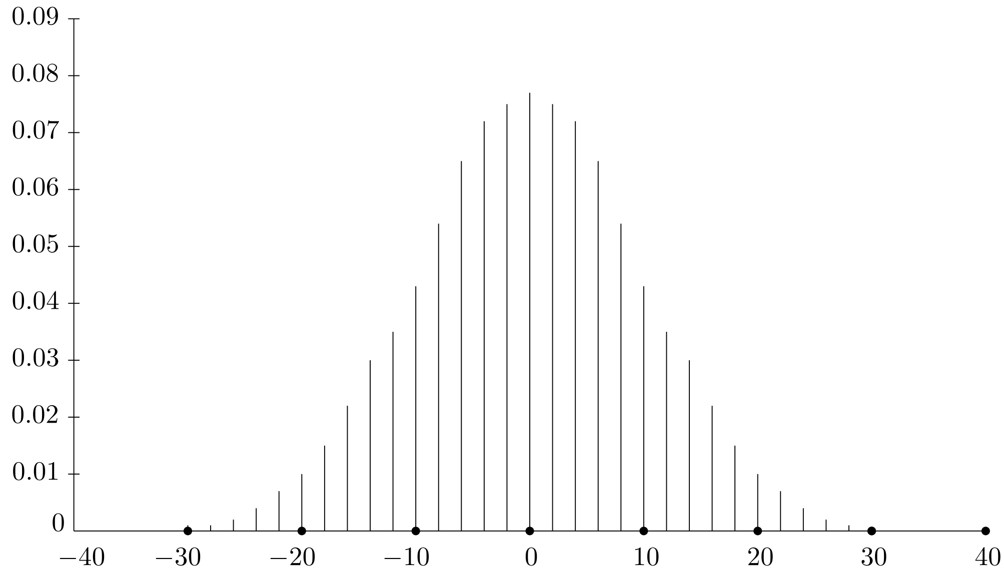


Fig. 2.1: Distribution of S_{100}

Hence,

$$\begin{aligned}\varphi_{Z_n}(t) &= \Big(\cos\frac{t}{\sqrt{n}}\Big)^n = \Big(1-\frac{t^2}{2n}+\frac{\epsilon(n)}{n}\Big)^n \\ &= \exp\Big\{\underbrace{n\log\Big(1-\frac{t^2}{2n}+\frac{\epsilon(n)}{n}\Big)}_{\sim n\left(-\frac{t^2}{2n}+\frac{\epsilon(n)}{n}\right)\to -\frac{t^2}{2}}\Big\} \xrightarrow{n\to\infty} \exp\Big\{-\frac{t^2}{2}\Big\}. \end{aligned} \tag{2.1.11}$$

We will see that

$$\exp\Big\{-\frac{t^2}{2}\Big\} = \int e^{itx}\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\,dx \tag{2.1.12}$$

(i.e. $\exp\{-\frac{x^2}{2}\}$ is the characteristic function of the standard normal distribution). The method that we develop in this chapter will provide (2.1.9) as a consequence of (2.1.11) and (2.1.12).

The first step is to introduce a notion of convergence, for which statements like (2.1.9) make sense.

2.2 Weak convergence

Definition 2.1. *A* ***sequence of distributions*** μ_n *on* $\mathbb{R}$ *(i.e. probability measures on* $(\mathbb{R},\mathcal{B}(\mathbb{R}))$*)* ***converges weakly*** *to the distribution* μ *on* $\mathbb{R}$ *(Notation:* $\mu_n \xrightarrow{w} \mu$*) if*

(2.2.1) *$F_n(y) \overset{n\to\infty}{\longrightarrow} F(y)$ for all points of continuity y of F, where $F_n(\cdot) = \mu_n((-\infty,\cdot])$ and $F(\cdot) = \mu((-\infty,\cdot])$ are the corresponding distribution functions.*

A ***sequence of random variables*** *X_n (on possibly different probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$)* ***converges in distribution*** *to the random variable X on $(\Omega, \mathcal{A}, P)$ if the distributions μ_n of the X_n converge weakly to the distribution μ of X.*

Remark 2.2. We will introduce later a condition equivalent to (2.2.1) which can easily be generalized to other spaces (e.g. $\mathbb{R}^d$). The drawback of this equivalent definition is that it is a little less intuitive than (2.2.1). □

Example 2.3.

1) *De Moivre Laplace:* (2.1.9) is equivalent to

$$Z_n \text{ converges in distribution to an } \mathcal{N}(0,1) \text{ random variable .} \tag{2.2.2}$$

(2.1.9) $\Longrightarrow$ (2.2.2):

Consider $y \in \mathbb{R}$, $F_n(\cdot)$ the distribution functions of Z_n, $\epsilon > 0$, and $M > |y|$ such that

$$\frac{1}{\sqrt{2\pi}} \int_{(-\infty,-M]\cup[M,+\infty)} e^{-\frac{x^2}{2}}\, dx \le \frac{\epsilon}{2} .$$

Using (2.1.9), one has, for $n \ge n_0$,

$$P[-M \le Z_n \le M] \ge 1 - \epsilon . \tag{2.2.3}$$

Hence,

$$\begin{aligned} F_n(y) = P[Z_n \le y] \quad &= \quad P[-M \le Z_n \le y] + \underbrace{P[Z_n < -M]}_{\le \epsilon \text{ using } (2.2.3)} \\ & \qquad n \to \infty \ \Big\downarrow \ (2.1.9) \\ & \qquad \int_{-M}^{y} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{x^2}{2}\Big\} dx \end{aligned}$$

and

$$\begin{aligned} \overline{\lim_n} \Big| F_n(y) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} dx \Big| &\le \overline{\lim_n}\, P[Z_n \le -M] + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-M} e^{-\frac{x^2}{2}} dx \\ &\le \epsilon + \frac{\epsilon}{2} = \frac{3\epsilon}{2} . \end{aligned}$$

(2.2.2) follows for $\epsilon \to 0$.

(2.2.2) $\Longrightarrow$ (2.1.9): One has, for $a < b$,

$$P[a \le Z_n \le b] = P[Z_n \le b] - P[Z_n < a] ,$$

so

$$\underline{\lim_n}\, P[a \le Z_n \le b] \ge \lim_n\ F_n(b) - \lim_n\ F_n(a) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\Big\{-\frac{x^2}{2}\Big\} dx,$$

and for $a' < a$,

$$\begin{aligned} \overline{\lim_n}\, P[a \le Z_n \le b] \quad &\le \quad \overline{\lim_n}\, P[a' < Z_n \le b] \\ &= \quad \overline{\lim_n}\, \big(F_n(b) - F_n(a')\big) = \frac{1}{\sqrt{2\pi}} \int_{a'}^{b} \exp\Big\{-\frac{x^2}{2}\Big\}\, dx . \end{aligned}$$

If we let $a' \uparrow a$, we find

$$\lim_n \; P[a \le Z_n \le b] = \frac{1}{\sqrt{2\pi}} \int_a^b \; \exp\Big\{ -\frac{x^2}{2} \Big\} \, dx \, .$$

2) Consider $\mu_n(dx) = \delta_{\frac{1}{n}}(dx)$. Then $\mu_n \xrightarrow{w} \mu = \delta_0$, since the corresponding distribution functions are equal to

$$F_n(y) = 1\Big\{ y \ge \frac{1}{n} \Big\} \;\text{ and }\; F(y) = 1\{y \ge 0\} \, ,$$

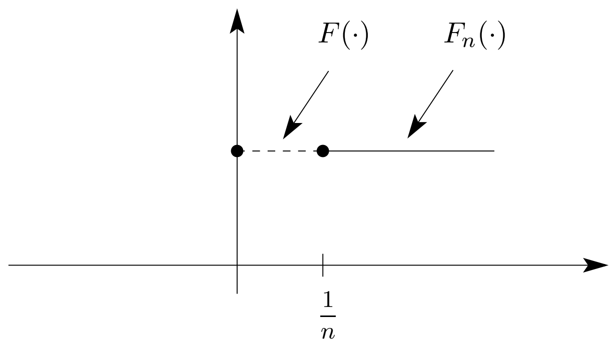


Fig. 2.2

so that we have $F_n(y) \longrightarrow F(y)$ for $y \ne 0$, and (2.2.1) holds (note that $F_n(0) = 0$ does not converge to $F(0) = 1$).

3) X_p a geometrically distributed random variable with success probability p:

$$P[X_p = k] = p(1-p)^{k-1} \qquad \forall k \ge 1.$$

(**Interpretation**: X_p is the **location of the first success** in an infinite sequence of independent Bernoulli variables with parameter p: $(Y_1, \, Y_2, \dots, \, Y_n, \dots)$). If we let the success parameter p tend to 0, we obtain

(2.2.4) $\quad$ $p\,X_p$ converges in distribution to an exponentially distributed random variable with parameter 1.

Indeed: for $y \le 0$,

$$P[p\,X_p \le y] \le P[X_p \le 0] \underset{p \to 0}{\longrightarrow} \; 0,$$

and for $y > 0$,

$$P[p\,X_p \le y] = 1 - P\Big[X_p > \frac{y}{p} \Big] = 1 - \sum_{k > \frac{y}{p}} p(1-p)^{k-1} =$$

$$1 - \frac{p}{(1-(1-p))} \cdot (1-p)^{[\frac{y}{p}]} = 1 - (1-p)^{[\frac{y}{p}]} \, .$$

One has

$$p \cdot \left[\frac{y}{p}\right] \overset{p \to 0}{\longrightarrow} y \ \left(\text{since } p\,\frac{y}{p} \ge p\left[\frac{y}{p}\right] \ge p\Big(\frac{y}{p} - 1\Big)\right)$$

and consequently, $(1-p)^{[\frac{y}{p}]} \underset{p \to 0}{\longrightarrow} e^{-y}$. We thus have, for $y > 0$,

$$P[p\,X_p \le y] \underset{p \to 0}{\longrightarrow} 1 - e^{-y} ,$$

and (2.2.4) follows.

4) *Poisson approximation*

Let X_n be binomial(n, p_n) distributed random variables, with $n \cdot p_n \longrightarrow \lambda > 0$. Then

(2.2.5) X_n converges in distribution to a Poisson distributed random variable X with parameter λ .

For fixed $k \ge 0$,

$$\begin{aligned} P[X_n = k] &= \frac{n!}{k!(n-k)!}\ p_n^k\,(1-p_n)^{n-k} \\ &= \underbrace{\frac{n(n-1)\cdot \ldots \cdot (n-k+1)}{n \ \cdot \ n \ \ \cdot \ \ldots \ \cdot \ \ n}}_{\downarrow\, n \to \infty \atop 1} \cdot \frac{1}{k!}\ \underbrace{(p_n \cdot n)^k}_{\downarrow\, n \to \infty \atop \lambda^k} \ \Big(1 - \frac{p_n \cdot n}{n}\Big)^{n-k}, \end{aligned}$$

and since $\frac{p_n \cdot n}{n} \cdot (n-k) \longrightarrow \lambda$, one has $(1 - \frac{p_n \cdot n}{n})^{n-k} \underset{n \to \infty}{\longrightarrow} e^{-\lambda}$. Hence,

$$P[X_n = k] \longrightarrow \frac{e^{-\lambda}}{k!}\ \lambda^k = P[X = k] ,$$

and for $y \in \mathbb{R}$,

$$P[X_n \le y] = \sum_{k \le y} P[X_n = k] \underset{n \to \infty}{\longrightarrow} \sum_{k \le y} P[X = k] = P[X \le y] .$$

Our claim (2.2.5) thus follows.

5) *Order statistics*

Let $X_1, \ldots, X_{2n+1}$ be independent, uniformly distributed random variables on $(0,1)$. Define:

$$\begin{aligned} V_{n+1} &= (n+1)\text{th smallest value of } \{X_1, X_2, \ldots, X_{2n+1}\} \\ &= \min\{\max\{X_k, k \in J\},\ |J| = n+1\} \\ &= \max\{\min\{X_k, k \in J\},\ |J| = n+1\} \\ &= \text{median value of the numbers } \{X_1, X_2, \ldots, X_{2n+1}\} . \end{aligned}$$

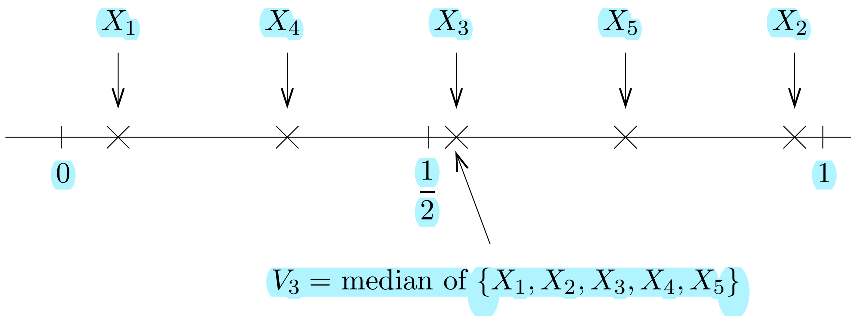


Fig. 2.3

For $x \in (0,1)$,

$$
\begin{aligned}
P[V_{n+1} \le x] &= P[\underbrace{\#\{k,\ X_k \le x\}}_{\text{binomial}(n,x)\text{ distributed}} \ge n+1] \\
&= \sum_{k=n+1}^{2n+1} \binom{2n+1}{k} x^k (1-x)^{2n+1-k} \,.
\end{aligned}
\tag{2.2.6}
$$

By differentiating (2.2.6), we obtain the density $f_{n+1}(\cdot)$ of V_{n+1}:

$$
\begin{aligned}
&f_{n+1}(x) = \\
&\sum_{k=n+1}^{2n+1} \frac{(2n+1)!}{k!(2n+1-k)!}\, 1\{0<x<1\}[k\, x^{k-1}(1-x)^{2n+1-k} \\
&- (2n+1-k)\, x^k (1-x)^{2n-k}] = \\
&(2n+1) \sum_{k=n+1}^{2n} 1\{0<x<1\} \Big[\binom{2n}{k-1} x^{k-1}(1-x)^{2n-(k-1)} \\
&- \binom{2n}{k} x^k (1-x)^{2n-k}\Big] + (2n+1)\, 1\{0<x<1\} \binom{2n}{2n} x^{2n} \\
&\overset{\text{simplifications}}{=} (2n+1) \binom{2n}{n} x^n (1-x)^n \, 1\{0<x<1\} \,.
\end{aligned}
\tag{2.2.7}
$$

$\bigcirc$

Remark 2.4. It can be showed analogously that the kth smallest value V_k of $\{X_1, \dots, X_{2n+1}\}$ has density

$$
f_k(x) = (2n+1) \binom{2n}{k-1} x^{k-1} (1-x)^{2n+1-k}\, 1\{0<x<1\},
\tag{2.2.8}
$$

$1 \le k \le 2n+1$. $\square$

Define

$$
Y_n = 2\Big(V_{n+1} - \tfrac{1}{2}\Big) \sqrt{2n} \,.
\tag{2.2.9}
$$

Proposition 2.5.

(2.2.10) $\quad Y_n$ *converges in distribution to an* $\mathcal{N}(0,1)$ *random variable* .

Proof. One has

$$P[Y_n \le y] = P\Big[V_{n+1} \le \frac{1}{2} + \frac{y}{2\sqrt{2n}}\Big],$$

and Y_n has density

$$g_n(y) = f_{n+1}\Big(\frac{1}{2} + \frac{y}{2\sqrt{2n}}\Big) \times \frac{1}{2\sqrt{2n}} .$$

For fixed $y \in \mathbb{R}$ and n large enough,

$$\begin{aligned} g_n(y) &= (2n+1)\binom{2n}{n}\Big(\frac{1}{2} + \frac{y}{2\sqrt{2n}}\Big)^n \Big(\frac{1}{2} - \frac{y}{2\sqrt{2n}}\Big)^n \times \frac{1}{2\sqrt{2n}} \\ &= \underbrace{\binom{2n}{n} \times \Big(\frac{1}{2}\Big)^{2n}}_{\substack{\sim \frac{1}{\sqrt{\pi n}} \text{ using (2.1.7)} \\ \text{with } k(n)=0}} \times \Big(\underbrace{1 - \frac{y^2}{2n}}_{\downarrow}\Big)^n \times \frac{2n+1}{2\sqrt{2n}}, \\ &\qquad\qquad\qquad\qquad\qquad\quad \exp\Big\{-\frac{y^2}{2}\Big\} \end{aligned}$$

and since $\frac{1}{\sqrt{\pi n}} \; \frac{2n+1}{2\sqrt{2n}} \longrightarrow \frac{1}{\sqrt{2\pi}}$, we obtain:

(2.2.11) $$g_n(y) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{y^2}{2}\Big\} \quad \text{for all } \; y \in \mathbb{R} .$$

□

Lemma 2.6. *(Scheffe)*

Let $h_n(\cdot)$, $h(\cdot)$ *be density functions on* $\mathbb{R}$ *(i.e.* $h_n, h \ge 0$ *are measurable functions, with* $\int h_n = \int h = 1$*).*

(2.2.12) *From* $h_n(x) \overset{n\to\infty}{\longrightarrow} h(x)$ *for* $x \in \mathbb{R}$, *it follows that* $\|h_n - h\|_{L^1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$.

Proof.

$$\int |h_n(x) - h(x)| dx = 2 \int (h(x) - h_n(x))^+ dx$$

since $\int (h(x) - h_n(x))^+ dx - \int (h(x) - h_n(x))^- dx = \int h(x)dx - \int h_n(x) dx = 0$. Hence,

$$\int \underbrace{(h(x) - h_n(x))^+}_{\le h(x)} dx \xrightarrow{\substack{\text{dominated} \\ \text{convergence}}} 0 .$$

□

Thanks to (2.2.11) - (2.2.12), one has $\|g_n - g\|_{L^1(\mathbb{R})} \to 0$, where $g(y) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{y^2}{2}\}$. It follows, for $y \in \mathbb{R}$, that:

$$\lim_{n\to\infty} P[Y_n \le y] = \lim_{n\to\infty} \int_{-\infty}^{y} g_n(y)dy = \int_{-\infty}^{y} g(y)dy .$$

6) $\mu_n(dy) = \frac{1}{2}\,\delta_0(dy) + \frac{1}{2}\,\delta_n(dy)$. One has:

$$\begin{aligned} F_n(y) &= \frac{1}{2}\,1\{y \ge 0\} + \frac{1}{2}\,1\{y \ge n\} \\ n \to \infty &\downarrow \\ F(y) &= \frac{1}{2}\,1\{y \ge 0\}\;. \end{aligned}$$

This is not a distribution function, and so the sequence μ_n does not converge weakly.

Fig 2.4

A fraction of the mass "disappears to $+\infty$".

The next proposition gives two further conditions that are equivalent to weak convergence.

Proposition 2.7. *The following conditions are equivalent:*

(2.2.13)
$$\mu_n \xrightarrow{w} \mu,$$

(2.2.14) *there exist random variables Y_n, $n \ge 1$, and Y on a common probability space $(\Omega, \mathcal{A}, P)$ such that $\mu_n =$ distribution of Y_n, $\mu =$ distribution of Y, and $Y_n \to Y$ P-a.s.,*

(2.2.15)
$$\int_{\mathbb{R}} f\, d\mu_n \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f\, d\mu \qquad \text{for all } f \in C_b(\mathbb{R}),$$

where $C_b(\mathbb{R}) \overset{\text{def.}}{=} \{$*bounded continuous functions on* $\mathbb{R}\}$.

Proof. We show (2.2.13) $\Longrightarrow$ (2.2.14) $\Longrightarrow$ (2.2.15) $\Longrightarrow$ (2.2.13).

(2.2.13) $\Longrightarrow$ (2.2.14):

We choose $\Omega = (0,1)$, $\mathcal{A} = \mathcal{B}((0,1))$, $P =$ Lebesgue measure on $(0,1)$, and we define, for $\omega \in (0,1)$,

(2.2.16)
$$\begin{aligned} Y_n(\omega) &= \sup\{y \in \mathbb{R},\, F_n(y) < \omega\} \\ &\qquad \text{(where } F_n \text{ is the distribution function of } \mu_n\text{)}, \\ Y(\omega) &= \sup\{y \in \mathbb{R},\, F(y) < \omega\} \\ &\qquad \text{(where } F \text{ is the distribution function of } \mu\text{)}\;. \end{aligned}$$

Thanks to the proof of (1.1.13) in Chapter 1, we know that $\mu_n =$ distribution of Y_n, $\mu =$ distribution of Y.

Y and Y_n are non-decreasing on $(0,1)$. We define

$$\begin{aligned} &\Omega_0 = \{\omega \in (0,1) \,:\, Y(\omega) = \widetilde{Y}(\omega)\}\,, \\ &\text{where } \widetilde{Y}(\omega) = \inf\{y \in \mathbb{R} \,:\, F(y) > \omega\}\ . \end{aligned} \tag{2.2.17}$$

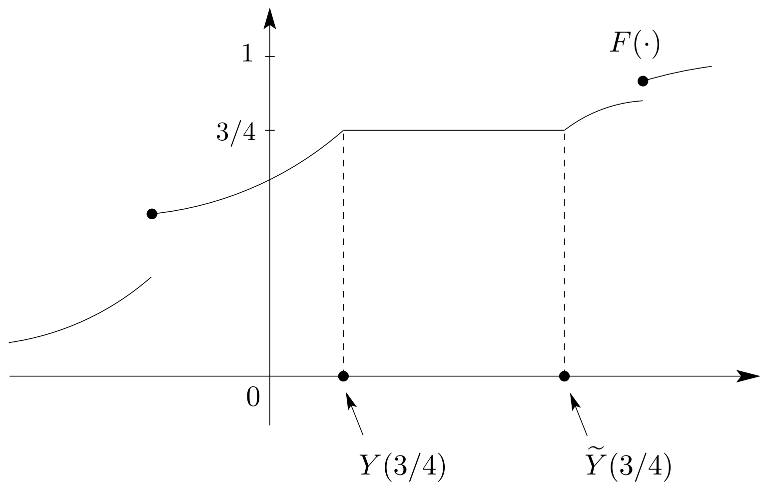


Fig. 2.5

Note that for $\omega < \omega'$,

$$Y(\omega) \le \widetilde{Y}(\omega) \le Y(\omega') \le \widetilde{Y}(\omega')\ .$$

Hence, $Y(\cdot)$ has a discontinuity in each $\omega \in \Omega \backslash \Omega_0$, and since $Y(\cdot)$ is non-decreasing, we have

$$\Omega \backslash \Omega_0 \text{ is at most countable}\,. \tag{2.2.18}$$

We have thus $P(\Omega_0) = 1$, and it suffices to show that

$$\text{for all } \omega \in \Omega_0,\ \ \lim_n\ Y_n(\omega) = Y(\omega)\ . \tag{2.2.19}$$

Choose $\omega \in \Omega_0$. Let $y < Y(\omega)$ be a point of continuity of $F(\cdot)$. Then $F(y) < \omega$ (from the definition of $Y(\cdot)$), and for n large enough, $F_n(y) < \omega \Longrightarrow y \le Y_n(\omega)$. Hence,

$$y \le \varliminf_n Y_n(\omega) \tag{2.2.20}$$

for y as above. If we let $y \uparrow Y(\omega)$ (using that $F(\cdot)$ has only at most countably many points of discontinuity), we obtain $Y(\omega) \le \varliminf_n Y_n(\omega)$.

Let $y > Y(\omega)$ be a point of continuity of $F(\cdot)$. Then

$$F(y) > \omega \ \ (\text{since } \omega \in \Omega_0 \text{ and } y > Y(\omega) = \widetilde{Y}(\omega) = \inf\{z : F(z) > \omega\})\,.$$

We thus have, for n large enough, $F_n(y) > \omega \Longrightarrow Y_n(\omega) \le y \Longrightarrow \overline{\lim}_n\, Y_n(\omega) \le Y(\omega)$, and (2.2.19) follows readily.

(2.2.14) $\Longrightarrow$ (2.2.15):

If we take $f \in C_b(\mathbb{R})$, then

$$\int f\, d\mu_n = E[f(Y_n)] \xrightarrow[]{\substack{\text{dominated}\\ \text{convergence}}} E[f(Y)] = \int f\, d\mu\,.$$

(2.2.15) $\Longrightarrow$ (2.2.13):

Let $y \in \mathbb{R}$ be a point of continuity of $F(\cdot)$, and the function $g_\epsilon(\cdot)$: $\mathbb{R} \to [0,1]$ defined by

$$g_\epsilon(x) = \begin{cases} 1 & \text{for } x \le y\,, \\ 0 & \text{for } x \ge y+\epsilon\,, \\ \text{linear} & \text{for } x \in [y, y+\epsilon] \end{cases}$$

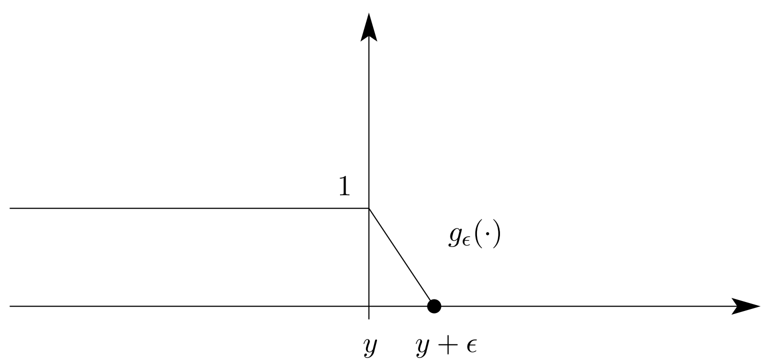


Fig. 2.6

(note that $g_\epsilon(\cdot) \in C_b(\mathbb{R})$). Then:

$$\begin{array}{rcl} & \overset{\displaystyle g_\epsilon(\cdot) \le 1_{(-\infty, y+\epsilon]}}{\downarrow} & \\ F(y+\epsilon) = \mu((-\infty, y+\epsilon]) & \ge & \displaystyle\int g_\epsilon(x)\,\mu(dx) = \lim_n \int g_\epsilon(x)\mu_n(dx) \\ & \overset{\displaystyle (g_\epsilon \ge 1 \text{ on } (-\infty, y])}{\downarrow} & \\ & \ge & \displaystyle\overline{\lim_n}\, \mu_n((-\infty, y]) = \overline{\lim_n}\, F_n(y)\,. \end{array}$$

For $\epsilon \to 0$, we obtain $\overline{\lim_n}\ F_n(y) \leq F(y)$. In an analogue way, we consider the function $h_\epsilon(\cdot)$: $\mathbb{R} \to [0,1]$ defined by

$$h_\epsilon(x) = \begin{cases} 1 & \text{for } x \leq y - \epsilon\ , \\ 0 & \text{for } x \geq y\ , \\ \text{linear} & \text{for } x \in [y-\epsilon, y]\ . \end{cases}$$

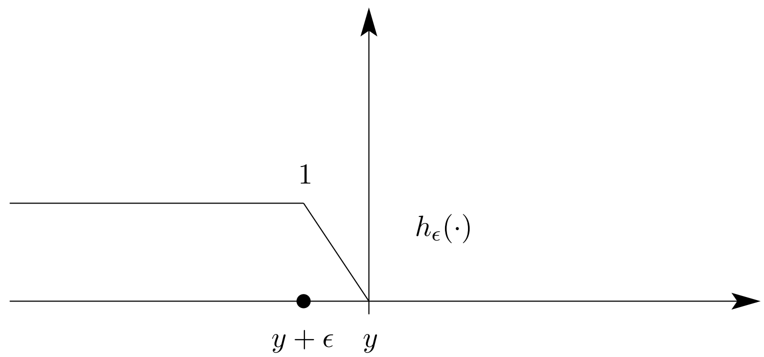


Fig. 2.7

Then

$$\begin{aligned} F(y-\epsilon) = \mu((-\infty, y-\epsilon]) &\leq \int h_\epsilon(x)\mu(dx) = \lim_n \int h_\epsilon(x)\mu_n(dx) \\ &\leq \underline{\lim}_n\ \mu_n((-\infty, y]) = \underline{\lim}_n\ F_n(y)\ . \end{aligned}$$

For $\epsilon \to 0$, we get

$$\underset{\substack{\uparrow \\ \text{points of continuity } y \text{ of } F}}{F(y)} \quad \leq \quad \underline{\lim}_n\ F_n(y)\ .$$

We deduce that $F(y) = \lim_n\ F_n(y)$ for all points of continuity y of F, i.e. $\mu_n \xrightarrow{w} \mu$. □

Remark 2.8. (2.2.15) can easily be generalized to other spaces (for example $\mathbb{R}^d$), and it is often used directly as a definition for weak convergence. □

Theorem 2.9. *(Helly)*

Let $F_n(\cdot)$, $n \geq 1$, be a sequence of distribution functions on $\mathbb{R}$, then there exists a subsequence $F_{n(k)}$ and a right-continuous non-decreasing function $F(\cdot)$: $\mathbb{R} \to [0,1]$ such that

(2.2.21) $$F(y) = \lim_{k\to\infty}\ F_{n(k)}(y) \qquad \textit{for all points of continuity } y \textit{ of } F\ .$$

Remark 2.10. 1) $F(\cdot)$ is not necessarily a distribution function, as Example 6 shows: for all $y \in \mathbb{R}$,

$$F_n(y) = \frac{1}{2}\, 1\{y \geq 0\} + \frac{1}{2}\, 1\{y \geq n\} \underset{n\to\infty}{\longrightarrow} F(y) = \frac{1}{2}\, 1\{y \geq 0\}\,. \tag{2.2.22}$$

We will later give a sufficient condition on the sequence F_n from which it follows that $F(\cdot)$ in (2.2.21) is a distribution function.

2) Helly's theorem can be interpreted as a compactness result. For that, one defines the so-called "vague convergence" of a sequence of sub-probability measures ν_n on $\mathbb{R}$ (i.e. $\nu_n \in \mathcal{M}_+(\mathbb{R})$ with $\nu_n(\mathbb{R}) \leq 1$) to a sub-probability measure ν as:

$$\int f\, d\nu_n \longrightarrow \int f\, d\nu \qquad \forall f \in \underset{\substack{\| \\ \text{continuous with compact support}}}{C_{\text{comp.}}(\mathbb{R})}.$$

This allows one to reformulate Helly's theorem as follows: "any sequence of sub-probability measures ν_n on $\mathbb{R}$, $n \geq 1$, possesses a subsequence that converges vaguely to a sub-probability measure ν". □

Proof. We write $\mathbb{Q} = \{q_1, q_2, q_3, \dots\}$. Note that $F_n(q_1) \in [0,1]$ for all $n \geq 1$

$\Longrightarrow$ there exists an infinite set $\mathbb{N}_1 \subseteq \mathbb{N}\backslash\{0\}$ such that

$$\lim_{n\in\mathbb{N}_1} F_n(q_1) = G(q_1)\,.$$

In an analogous way, one has $F_n(q_2) \in [0,1]$ for all $n \in \mathbb{N}_1$

$\Longrightarrow$ there exists an infinite set $\mathbb{N}_2 \subseteq \mathbb{N}_1$ such that

$$\lim_{n\in\mathbb{N}_2} F_n(q_2) = G(q_2), \quad \lim_{n\in\mathbb{N}_2} F_n(q_1) = G(q_1)\,.$$

By induction, we obtain a decreasing sequence $\mathbb{N}_1 \supseteq \mathbb{N}_2 \supseteq \cdots \supseteq \mathbb{N}_k \supseteq \dots$ of infinite subsets of $\mathbb{N}\backslash\{0\}$ such that for all $k \geq 1$,

$$\lim_{n\in\mathbb{N}_k} F_n(q_\ell) = G(q_\ell) \qquad (1 \leq \ell \leq k)\,. \tag{2.2.23}$$

Using Cantor's diagonal argument, we deduce the existence of an infinite set $\mathbb{N}_*$ such that $\mathbb{N}_*\backslash\mathbb{N}_k$ is finite for each k, and so

$$\forall \ell \geq 1, \qquad \lim_{n\in\mathbb{N}_*} F_n(q_\ell) = G(q_\ell)\,.$$

In other words,

$$\text{for all } r \in \mathbb{Q}, \text{ one has } \lim_{n\in\mathbb{N}_*} F_n(r) = G(r)\,. \tag{2.2.24}$$

$G(\cdot)$ is a non-decreasing function $\mathbb{Q} \to [0,1]$, and

$$F(x) \overset{\text{def.}}{=} \inf\{G(q) :\ q \in \mathbb{Q}\,,\, q > x\}$$

is non-decreasing as well. Furthermore, $F(\cdot)$ is **right-continuous**: indeed, for $x \in \mathbb{R}$ and $\epsilon > 0$, one can choose $q > x$ rational such that $F(x) \le G(q) < F(x) + \epsilon$, which implies

$$F(x) \le F(y) \le F(x) + \epsilon \qquad \text{for } \; x \le y < q \; .$$

We now show (2.2.21). Let $y \in \mathbb{R}$ be a point of continuity of $F(\cdot)$, and $\epsilon > 0$. Let us consider rational numbers

$$r_1 < r_2 < y < s$$

such that

$$F(y) - \epsilon < F(r_1) \le F(r_2) \le F(y) \le F(s) < F(y) + \epsilon \; .$$

Then

$$\begin{aligned} \varliminf_{n \in \mathbb{N}_*} \; F_n(r_2) &= G(r_2) \ge F(r_1), \\ \varlimsup_{n \in \mathbb{N}_*} \; F_n(s) &= G(s) \le F(s) \; . \end{aligned}$$

We thus have, for $n \in \mathbb{N}_*$ large enough,

$$F(y) - \epsilon \le F_n(r_2) \le F_n(y) \le F_n(s) \le F(y) + \epsilon \; ,$$

and so $\lim_{n \in \mathbb{N}_*} \; F_n(y) = F(y)$. Hence, (2.2.21) follows. □

Proposition 2.11. *If a sequence of distributions μ_n on $\mathbb{R}$ is **tight**, i.e.*

(2.2.25) $$\forall \epsilon > 0, \; \exists M > 0, \; \sup_{n \ge 1} \; \mu_n([-M, M]^c) \le \epsilon \; ,$$

then any subsequential limit (in the sense of (2.2.21)*) of $F_n(\cdot)$ (the distribution functions of μ_n) is a distribution function too.*

Proof. Let $F(\cdot)$ be non-decreasing, right-continuous with

$$F(y) = \lim_{k \to \infty} \; F_{n_k}(y) \qquad \text{for all points of continuity } y \text{ of } F \; .$$

Choose $\epsilon > 0$, and then $M > 0$ such that

$$\sup_n \; \mu_n([-M, M]^c) \le \epsilon \; .$$

Consider some points of continuity $y_1 > M$ and $y_2 < -M$ of $F(\cdot)$, then

$$\begin{aligned} F(y_2) &= \lim_k \; F_{n_k}(y_2) \le \varlimsup_k \mu_{n_k}((-\infty, -M)) \le \epsilon, \\ \text{and} \quad F(y_1) &= \lim_k \; F_{n_k}(y_1) \ge \varliminf_k F_{n_k}(M) \ge 1 - \epsilon \; . \end{aligned}$$

It follows that $\lim_{y \to -\infty} \; F(y) = 0$ and $\lim_{y \to \infty} \; F(y) = 1$. $F(\cdot)$ is thus a distribution function on $\mathbb{R}$. □

An important consequence of the last proposition and Helly's theorem is the following:

Corollary 2.12.

(2.2.26) *Any tight sequence of distributions $\mu_n, n \ge 1$, on $\mathbb{R}$ possesses a weakly convergent sub-sequence $\mu_{n_k} \xrightarrow{w} \mu$.*

2.3 Characteristic functions

Definition 2.13. *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The* ***characteristic function φ of μ*** *is the function from $\mathbb{R}$ to $\mathbb{C}$ defined by*

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} \mu(dx) \qquad \forall t \in \mathbb{R} \,. \tag{2.3.1}$$

The ***characteristic function of a random variable*** *X is the characteristic function of the distribution of X, i.e.*

$$\varphi(t) = E[e^{itX}] = \int e^{itx} \mu_X(dx) \,. \tag{2.3.2}$$

Remark 2.14. The characteristic function $\varphi(\cdot)$ of a distribution μ is nothing else but the Fourier transform of μ.

First properties: Let μ be a distribution on $\mathbb{R}$, one has

$$\varphi(0) = 1 \,, \tag{2.3.3}$$

$$\begin{aligned} &\forall t \in \mathbb{R}, \quad |\varphi(t)| \le 1 \ \text{ since} \\ &|\varphi(t)| \le \int |e^{itx}| \, \mu(dx) = \int \mu(dx) \,, \end{aligned} \tag{2.3.4}$$

$$\begin{aligned} &\forall t \in \mathbb{R}, \quad \varphi(-t) = \overline{\varphi(t)} \ \text{ since} \\ &\varphi(-t) = \int e^{-itx} \, \mu(dx) = \overline{\int e^{itx} \, \mu(dx)} = \overline{\varphi(t)} \,, \end{aligned} \tag{2.3.5}$$

$$\begin{aligned} &\varphi(\cdot) \text{ is } \textbf{uniformly continuous} \text{ since} \\ &|\varphi(t+h) - \varphi(t)| = \Big| \int (e^{i(t+h)\cdot x} - e^{itx}) \mu(dx) \Big| \\ &\qquad\qquad \le \int |e^{ihx} - 1| \, | \underbrace{e^{itx}}_{=1} | \, \mu(dx), \\ &\text{and using Lebesgue's theorem,} \\ &\qquad \int |e^{ihx} - 1| \, \mu(dx) \underset{h \to 0}{\longrightarrow} 0 \,. \end{aligned} \tag{2.3.6}$$

$$\begin{aligned} &\text{Let } X, Y \text{ be independent random variables on } (\Omega, \mathcal{A}, P), \\ &\text{then } \varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) \text{ for all } t \in \mathbb{R}, \text{ since} \\ &\varphi_{X+Y}(t) = E[e^{it(X+Y)}] \overset{\text{independence}}{=} E[e^{itX}] \, E[e^{itY}] \\ &\qquad = \varphi_X(t) \, \varphi_Y(t) \,. \end{aligned} \tag{2.3.7}$$

□

We start with a few examples of characteristic functions.

Example 2.15.

1) For some $a < b$, let us consider $\mu =$ uniform on $[a, b]$, i.e.

$$\mu(dx) = \frac{1}{b-a}\, 1_{[a,b]}(x)dx,$$

then:

$$\varphi(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \begin{cases} 1 & \text{for } t = 0, \\ \dfrac{e^{itb} - e^{ita}}{it(b-a)} & \text{for } t \neq 0. \end{cases} \tag{2.3.8}$$

2) $\mu =$ standard normal distribution. We have, for all $t \in \mathbb{R}$,

$$\varphi(t) = \int_{\mathbb{R}} e^{itx - \frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \overset{\text{(symmetry)}}{=} \int_{\mathbb{R}} \cos(tx)\, e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}\,.$$

Note that for $h \neq 0$,

$$\left|\frac{1}{h}\big(\cos((t+h)x) - \cos(tx)\big)\right| = \left|\,\frac{1}{h} \int_{tx}^{(t+h)x} \underbrace{\sin u}_{\in[-1,1]}\, du\,\right| \leq |x|\,,$$

and

$$\frac{1}{h}\,(\varphi(t+h) - \varphi(t)) = \int_{\mathbb{R}} \frac{\cos((t+h)x) - \cos(tx)}{h}\, e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}\,.$$

The dominating function $|x|\, \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ is Lebesgue integrable. It follows from Lebesgue's theorem that

$$\begin{aligned} \varphi'(t) &= \lim_{h \to 0} \frac{\varphi(t+h) - \varphi(t)}{h} \\ &= \int_{\mathbb{R}} -x \sin(tx)\, e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}\,. \end{aligned} \tag{2.3.9}$$

With the help of one integration by parts, one obtains

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \underbrace{\sin(tx)}_{=u}\, \underbrace{(-x\, e^{-\frac{x^2}{2}})}_{=v'}\, dx &= 0 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t \cos(tx)\, e^{-\frac{x^2}{2}} dx \\ &= -t\, \varphi(t)\,. \end{aligned}$$

Hence,

$$\begin{aligned} &\varphi'(t) = -t\varphi(t)\,, \\ \text{and} \quad &\Big(\exp\Big\{\frac{t^2}{2}\Big\}\, \varphi(t)\Big)' = t\, \exp\Big\{\frac{t^2}{2}\Big\}\, \varphi(t) - \exp\Big\{\frac{t^2}{2}\Big\}\, t\varphi(t) \\ &\qquad = 0\,. \end{aligned} \tag{2.3.10}$$

It follows that $\exp\{\frac{t^2}{2}\}\,\varphi(t) = \exp\{\frac{0^2}{2}\}\varphi(0) = 1$, and so we obtain that for all $t \in \mathbb{R}$,

$$\varphi(t) = \exp\Big\{-\frac{t^2}{2}\Big\}. \tag{2.3.11}$$

3) $Y : \mathcal{N}(m, \sigma^2)$ distributed. Consider X with standard normal distribution, then

$$Y \overset{\text{distribution}}{=} \sigma\, X + m.$$

Hence,

$$\begin{aligned}
\varphi_Y(t) &= E[\exp\{itY\}] \\
&= E[\exp\{it(m + \sigma X)\}] \\
&= \exp\{itm\}\,\varphi_X(t\sigma) \\
&= \exp\{itm - \frac{\sigma^2}{2}\,t^2\}\,.
\end{aligned}$$

In other words,

$$\forall t \in \mathbb{R}, \qquad \varphi_{\mathcal{N}(m,\sigma^2)}(t) = \exp\Big\{itm - \frac{\sigma^2}{2}\,t^2\Big\}\,. \tag{2.3.12}$$

Uniqueness Property:

(2.3.13) If two distributions μ, ν on $\mathbb{R}$ have the same characteristic function, then $\mu = \nu$.

Proof. Consider $h \in L^1(\mathbb{R}, dx)$, and set

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\,h(t)dt\,. \tag{2.3.14}$$

Then:

$$\begin{aligned}
\int f(x)d\mu(x) &= \int\Big(\int \frac{1}{2\pi}\,e^{-ixt}h(t)dt\Big)d\mu(x) \\
&\overset{\text{Fubini}}{=} \frac{1}{2\pi}\int h(t)\Big(\int e^{-ixt}d\mu(x)\Big)dt \\
&= \frac{1}{2\pi}\int \varphi_\mu(-t)h(t)dt\,.
\end{aligned}$$

Analogously,

$$\int f(x)\,d\nu(x) = \frac{1}{2\pi}\int \varphi_\nu(-t)h(t)dt\,,$$

and so

$$\int f(x)d\mu(x) = \int f(x)d\nu(x)\,. \tag{2.3.15}$$

If we choose $h(t) = \exp\{imt - \frac{\sigma^2}{2}\, t^2\}$, then

$$
\begin{aligned}
f(x) &= \frac{1}{2\pi} \int \exp\Big\{i(m-x)t - \frac{\sigma^2}{2}\, t^2\Big\}\, dt \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int \exp\Big\{i(m-x)t - \frac{\sigma^2}{2}\, t^2\Big\}\, \frac{\sigma dt}{\sqrt{2\pi}} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\, \varphi_{N(0,\frac{1}{\sigma^2})}(m-x) \overset{(2.3.12)}{=} \frac{1}{\sqrt{2\pi}\sigma}\, \exp\Big\{-\frac{(x-m)^2}{2\sigma^2}\Big\}\,.
\end{aligned}
$$

Take $\sigma = \frac{1}{n^2}$. It follows from (2.3.15) that:

$$
\begin{aligned}
&\int \frac{n^2}{\sqrt{2\pi}}\, \exp\Big\{-\frac{(x-m)^2}{2}\, n^4\Big\}\, d\mu(x) = \\
&\int \frac{n^2}{\sqrt{2\pi}}\, \exp\Big\{-\frac{(x-m)^2}{2}\, n^4\Big\}\, d\nu(x)\,.
\end{aligned}
\tag{2.3.16}
$$

One has

$$
\begin{aligned}
&\int_{a-\frac{1}{n}}^{b+\frac{1}{n}} \frac{n^2}{\sqrt{2\pi}}\, \exp\Big\{-\frac{(x-m)^2}{2}\, n^4\Big\} dm \\
&= P\Big[a - \frac{1}{n} \le x + \frac{X}{n^2} \le b + \frac{1}{n}\Big], \quad \text{where } X \text{ is } \mathcal{N}(0,1) \text{ distributed}, \\
&= P[(a-x)n^2 - n \le X \le (b-x)n^2 + n] \underset{n\to\infty}{\longrightarrow} 1_{[a,b]}(x) = \begin{cases} 1 & \text{if } x \in [a,b]\,, \\ 0 & \text{if } x \notin [a,b]\,. \end{cases}
\end{aligned}
\tag{2.3.17}
$$

If we integrate Equation (2.3.16) over $m \in [a-\frac{1}{n},\ b+\frac{1}{n}]$, we obtain, using Fubini's theorem and (2.3.17) (as well as Lebesgue's theorem), for $n \to \infty$:

$$
\begin{array}{c}
\underbrace{\int_{\mathbb{R}} \int_{a-\frac{1}{n}}^{b+\frac{1}{n}} \frac{n^2}{\sqrt{2\pi}}\, \exp\Big\{-\frac{(x-m)^2}{2}\, n^4\Big\}\, dm\, d\mu(x)} \\
\downarrow \\
\int_{\mathbb{R}} 1_{[a,b]}(x) d\mu(x) = \mu([a,b]) = \nu([a,b]) = \int_{\mathbb{R}} 1_{[a,b]}(x) d\nu(x) \\
\uparrow \\
= \overbrace{\int_{\mathbb{R}} \int_{a-\frac{1}{n}}^{b+\frac{1}{n}} \frac{n^2}{\sqrt{2\pi}}\, \exp\Big\{-\frac{(x-m)^2}{2}\, n^4\Big\}\, dm\, d\nu(x)}
\end{array}
$$

and so $\mu = \nu$ by Dynkin's lemma (more precisely, one of its consequences), which completes the proof. $\square$

As a consequence of this uniqueness property, some results can be read directly from the characteristic function, and then transferred to the distribution itself, as we now explain.

Example 2.16.

1) $\forall t \in \mathbb{R}$, $\varphi_X(t) \in \mathbb{R} \overset{(2.3.5)+\text{uniqueness property}}{\Longrightarrow} X$ and $-X$ have the same distribution.

2) $\mu =$ Poisson (λ), $\lambda > 0$, then for all $t \in \mathbb{R}$,

$$\varphi_\lambda(t) = \sum_{n \geq 0} e^{-\lambda}\, e^{itn}\, \frac{\lambda^n}{n!} = \exp\{\lambda(e^{it} - 1)\}, \tag{2.3.18}$$

and thus: $\varphi_\lambda(t)\, \varphi_{\lambda'}(t) = \varphi_{\lambda+\lambda'}(t)$. Using (2.3.7), we obtain:

(2.3.19) If X, Y are two independent random variables, with distribution Poisson (λ), resp. Poisson (λ'), then $X + Y$ has distribution Poisson $(\lambda + \lambda')$.

3) $\mu =$ Cauchy with parameter $a > 0$, i.e.

$$\mu(dx) = \frac{1}{\pi}\, \frac{a}{x^2 + a^2}\, dx,$$

then for all $t \in \mathbb{R}$,

$$\varphi_a(t) = \int_{-\infty}^{+\infty} \frac{a e^{itx}}{a^2 + x^2}\, \frac{dx}{\pi} = \exp\{-a\,|t|\} \tag{2.3.20}$$

(see exercises). Hence, a property similar to (2.3.19) follows from the identity $\varphi_a(t)\varphi_{a'}(t) = \varphi_{a+a'}(t)$.

4) Let $N, X_1, X_2, \dots, X_n, \dots$ be independent random variables, where N is Poisson (λ) distributed, and the X_k are μ distributed. Define

$$Y = X_1(\omega) + X_2(\omega) + \cdots + X_{N(\omega)}(\omega) = S_{N(\omega)}(\omega)\ .$$

The distribution of Y is called the compound Poisson distribution, and

$$\begin{aligned}
\varphi_Y(t) = {} & E[\exp\{itY\}] = E[\exp\{itS_N\}] \\
= {} & \sum_{n=0}^{\infty} E[\exp\{itS_n\}\ 1_{\{N=n\}}] \\
\overset{\text{independence}}{=} {} & \sum_{n=0}^{\infty} E[\exp\{itS_n\}]\, e^{-\lambda}\, \frac{\lambda^n}{n!} \\
= {} & \sum_{n=0}^{\infty} \varphi_\mu(t)^n\ e^{-\lambda}\, \frac{\lambda^n}{n!} = \exp\{\lambda(\varphi_\mu(t) - 1))\}\ .
\end{aligned}$$

Hence, we have

$$\varphi_Y(t) = \exp\{-\lambda(1 - \varphi_\mu(t))\} . \tag{2.3.21}$$

In particular,

$$\varphi_{Y_{\lambda+\lambda'}}(t) = \varphi_{Y_\lambda}(t)\, \varphi_{Y_{\lambda'}}(t),$$

so a property similar to (2.3.19) holds. ○

As we now explain, there is an important connection between the differentiability of the characteristic function $\varphi_\mu(\cdot)$ in a neighborhood of 0, and the "decay" of μ close to $\pm\infty$.

Proposition 2.17. *Let μ be a distribution on $\mathbb{R}$, then for all $u > 0$,*

$$\mu\Big(\Big\{x : |x| \ge \frac{2}{u}\Big\}\Big) \le \frac{1}{u} \int_{-u}^{u} (1 - \varphi(t))dt . \tag{2.3.22}$$

If for some $k \ge 1$, μ has a finite kth moment (i.e. $\int |x|^k d\mu < \infty$), then $\varphi_\mu(\cdot)$ is k times continuously differentiable, and one has

$$\varphi_\mu^{(\ell)}(t) = \int (ix)^\ell e^{itx} d\mu \qquad (0 \le \ell \le k) . \tag{2.3.23}$$

Proof. (2.3.22): For $x \ne 0$,

$$\begin{aligned}\int_{-u}^{u} (1 - e^{itx})dt &= 2u - \frac{e^{iux} - e^{-iux}}{ix} \\ &= 2u - 2\,\frac{\sin(ux)}{x} = 2u\Big(1 - \frac{\sin ux}{ux}\Big) \ge 0 .\end{aligned}$$

Hence,

$$\begin{aligned}&\frac{1}{u} \int_{-u}^{u} (1 - \varphi(t))dt \overset{\text{Fubini}}{=} \int_{\mathbb{R}} \frac{1}{u} \int_{-u}^{u} (1 - e^{itx})dt\, d\mu(x) \\ &= 2 \int_{\mathbb{R}} \underbrace{1 - \frac{\sin ux}{ux}}_{\ge 0}\, d\mu(x) \ge 2 \int_{|x| \ge \frac{2}{u}} \underbrace{\Big(1 - \frac{1}{|ux|}\Big)}_{\ge \frac{1}{2}} d\mu(x) \\ &\ge \mu\Big(\Big\{x : |x| \ge \frac{2}{u}\Big\}\Big),\end{aligned}$$

and (2.3.22) follows.

(2.3.23): We consider $k \ge 1$ fixed, and we argue by induction. For $\ell = 0$, (2.3.23) holds. If (2.3.23) holds for $\ell < k$, then it follows that:

$$\frac{1}{h}\Big(\varphi_\mu^{(\ell)}(t+h) - \varphi_\mu^{(\ell)}(t)\Big) = \int_{\mathbb{R}} (ix)^\ell\, \frac{e^{i(t+h)x} - e^{itx}}{h}\, \mu(dx) .$$

Note that

$$\frac{|e^{i(t+h)x} - e^{itx}|}{|h|} = \frac{|e^{ihx} - 1|}{|h|} = \frac{|\int_0^{hx} ie^{iu}\, du|}{|h|} \le |x| \ ,$$

and $|(ix)^\ell| \times |x| = |x|^{\ell+1} \in L^1(d\mu)$, by assumption. Using Lebesgue's theorem, we obtain

$$\lim_{h\to 0} \frac{1}{h} \left(\varphi_\mu^{(\ell)}(t+h) - \varphi_\mu^{(\ell)}(t)\right) = \int_{\mathbb{R}} (ix)^{\ell+1}\, e^{itx}\, \mu(dx) \ .$$

In other words, (2.3.23) holds for $\ell + 1$. Our property then follows by induction. □

Theorem 2.18. *(Continuity)*

Let $\mu_n, n \ge 1$, be a sequence of distributions on $\mathbb{R}$.

(2.3.24) *If $\mu_n \xrightarrow{w} \mu$, then $\forall t \in \mathbb{R}$, $\varphi_{\mu_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_\mu(t)$.*

(2.3.25) *If for all $t \in \mathbb{R}$, $\varphi_{\mu_n}(t)$ converges, and $\varphi_\infty(t) = \lim_n \varphi_{\mu_n}(t)$ is continuous in 0, then there exists a distribution μ_∞ on $\mathbb{R}$ with: $\forall t \in \mathbb{R}$, $\varphi_{\mu_\infty}(t) = \varphi_\infty(t)$, and $\mu_n \xrightarrow{w} \mu_\infty$.*

Proof.

(2.3.24): $e^{itx} \in C_b(\mathbb{R})$, so (2.3.24) follows from (2.2.15).

(2.3.25): The sequence $\mu_n, n \ge 1$, is tight. Indeed, let us consider $\epsilon > 0$: since φ_∞ is continuous in 0 (and $\varphi_\infty(0) = 1$), we can choose $u > 0$ small enough so that

$$\frac{\epsilon}{2} \ge \frac{1}{u} \int_{-u}^{u} (1 - \varphi_\infty(t))dt \overset{\text{Lebesgue}}{=} \lim_n \frac{1}{u} \int_{-u}^{u} (1 - \varphi_n(t))dt \ .$$

Hence, for $n \ge n_0$,

$$\epsilon \ge \frac{1}{u} \int_{-u}^{u} (1 - \varphi_n(t))dt \overset{(2.3.22)}{\ge} \mu_n\Big(\Big[-\frac{2}{u}, \frac{2}{u}\Big]^c\Big),$$

which shows that μ_n is tight.

Thanks to (2.2.26), each subsequence $\mu_{n(k)}$ possesses a weakly convergent subsequence $\mu_{n(k(\ell))} \underset{\ell\to\infty}{\xrightarrow{w}} \nu$.

Using (2.3.24), $\varphi_\nu(\cdot) = \varphi_\infty(\cdot) \Longrightarrow \nu$ is uniquely determined. We write $\mu_\infty \overset{\text{def.}}{=} \nu$.

Then $\mu_n \xrightarrow{w} \mu_\infty$, since otherwise, there would exist a point of continuity y of $F_\infty(\cdot) \overset{\text{def.}}{=} \mu_\infty((-\infty, \cdot])$, $\epsilon > 0$, and a subsequence $n'(k)$ with

$$\forall k \ge 1, \qquad |F_{n'(k)}(y) - F_\infty(y)| \ge \epsilon \ .$$

This contradicts the existence of a subsequence $\mu_{n'(k(\ell))}$ that converges to μ_∞. □

An application: the symmetric stable distributions

We show the existence of distributions on $\mathbb{R}$ with characteristic functions:

$$\varphi_\alpha(t) = \exp\{-c|t|^\alpha\},\ t \in \mathbb{R},\ \text{ where } \alpha \in (0,2],\ c > 0\,. \tag{2.3.26}$$

Obviously, the case $\alpha = 2$ corresponds to the $\mathcal{N}(0, \sigma^2 = 2c)$ distribution. We discuss the case $0 < \alpha < 2$.

Consider first a random variable X with density

$$f(x) = \frac{\alpha}{2}\ \frac{1}{|x|^{\alpha+1}}\ 1_{\{|x|\ge 1\}}$$

(i.e. X and $-X$ have the same distribution, and $P[|X| > x] = \frac{1}{x^\alpha},\ x \ge 1$).

Let $\Psi(\cdot)$ be the characteristic function of X. We prove:

Lemma 2.19.

$$1 - \Psi(t) \underset{t\to 0}{\sim} K\,|t|^\alpha, \qquad \textit{for } K > 0 \textit{ a constant.} \tag{2.3.27}$$

Proof. Let us assume (without loss of generality) that $t > 0$. One has

$$\begin{aligned} 1 - \Psi(t) &\overset{\text{(symmetry)}}{=} \int_1^\infty \frac{(2 - e^{itx} - e^{-itx})}{2}\ \alpha\ \frac{dx}{x^{\alpha+1}} \\ &= \int_1^\infty (1 - \cos tx)\ \alpha\ \frac{dx}{x^{\alpha+1}} = t^\alpha \int_t^\infty (1 - \cos u)\alpha\, \frac{du}{u^{\alpha+1}} \end{aligned}$$

(where we set $x = \frac{u}{t}$). Since $1 - \cos u \underset{u\to 0}{\sim} \frac{u^2}{2}$, one has

$$\int_0^\infty (1 - \cos u) \frac{du}{u^{\alpha+1}} < \infty \ \text{ for } \ \alpha \in (0,2)\,.$$

(2.3.27) thus follows with $K = \alpha \int_0^\infty (1 - \cos u)\ \frac{du}{u^{\alpha+1}}$. $\square$

We now consider, for $n \ge 1$, a random variable with compound Poisson distribution like in Example 4 above:

$$Y_n(\omega) = \frac{X_1(\omega)}{n} + \frac{X_2(\omega)}{n} + \cdots + \frac{X_{N(\omega)}(\omega)}{n}\,, \tag{2.3.28}$$

where N, and X_k, $k \ge 1$, are independent, with distribution Poisson (n^α), resp. the same distribution as X. Using (2.3.21), we have

$$\varphi_{Y_n}(t) = \exp\Big\{ - n^\alpha \Big(1 - \Psi\Big(\frac{t}{n}\Big)\Big)\Big\}\,. \tag{2.3.29}$$

One has $n^\alpha(1 - \Psi(\frac{t}{n})) \overset{(2.3.27)}{\sim} n^\alpha\, K\, |\frac{t}{n}|^\alpha = K|t|^\alpha$, and so

$$\lim_{n\to\infty} \varphi_{Y_n}(t) = \exp\{-K\,|t|^\alpha\}\,. \tag{2.3.30}$$

Thanks to the continuity theorem, we deduce the existence of a random variable Y with $\varphi_Y(t) = \exp\{-K|t|^\alpha\}$. From this, it follows that $(\frac{c}{K})^{\frac{1}{\alpha}} Y = Z$ has characteristic function $\exp\{-c|t|^\alpha\}$.

The distributions with a characteristic function of the form (2.3.26) are called "symmetric stable distributions". Explicit formulas for the densities are known only in the two cases

$$\begin{aligned} \alpha = 2 &: \quad \text{normal distribution,} \\ \alpha = 1 &: \quad \text{Cauchy distribution (Example 3 above)}\,. \end{aligned}$$

2.4 Central Limit Theorem

We can now apply the whole theory that was developed in Sections 2.2 and 2.3. As we will see, the de Moivre Laplace theorem corresponds to a general phenomenon. The Central Limit Theorem shows that the **normal distribution** plays a **particularly important role**, since it describes the **behavior of the fluctuations** of a sum $S_n = X_1 + \cdots + X_n$ of many independent, identically distributed random variables, in the case when the X_i possess a finite second moment.

Theorem 2.20. *(Central limit)*

Let $X_1, X_2, \ldots, X_n, \ldots$ be independent, identically distributed random variables with $E[X_1^2] < \infty$, and denote

$$\sigma^2 \stackrel{\text{def.}}{=} \mathrm{Var}(X_1), \quad \mu \stackrel{\text{def.}}{=} E[X_1]\,. \tag{2.4.1}$$

Let us furthermore assume that $\sigma^2 > 0$ (i.e. that X_1 is not a.s. constant). Then, as $n \to \infty$,

$$Z_n \stackrel{\text{def.}}{=} \frac{S_n - n\mu}{\sigma\sqrt{n}}, \quad \text{where } S_n = X_1 + \cdots + X_n,\ n \geq 1\,, \tag{2.4.2}$$

converges in distribution to a random variable with standard normal distribution.

Proof. We can write, in an analogue way to (2.1.10):

$$\varphi_{Z_n}(t) = E\Big[\exp\Big\{\frac{it}{\sigma\sqrt{n}}\,\widetilde{S}_n\Big\}\Big], \quad \text{where}$$

$$\widetilde{S}_n = \widetilde{X}_1 + \cdots + \widetilde{X}_n, \quad \text{and } \widetilde{X}_i = X_i - E[X_i]\,.$$

Hence,

$$\begin{aligned} \varphi_{Z_n}(t) &\stackrel{\text{independence}}{=} \prod_{i=1}^n E\Big[\exp\Big\{\frac{it}{\sigma\sqrt{n}}\,\widetilde{X}_i\Big\}\Big] \\ &\stackrel{\text{identical distribution}}{=} E\Big[\exp\Big\{\frac{it}{\sigma\sqrt{n}}\,\widetilde{X}_1\Big\}\Big]^n = \varphi_{\widetilde{X}_1}\Big(\frac{t}{\sigma\sqrt{n}}\Big)^n\,. \end{aligned} \tag{2.4.3}$$

Thanks to (2.3.23), we also know that $\varphi_{\widetilde{X}_1}(\cdot)$ is twice continuously differentiable, so that:

$$\varphi'_{\widetilde{X}_1}(u) = E[i\ \widetilde{X}_1\ e^{iu\widetilde{X}_1}] \implies \varphi'_{\widetilde{X}_1}(0) = 0\ ,$$
$$\varphi''_{\widetilde{X}_1}(u) = -E[\widetilde{X}_1^2\ e^{iu\widetilde{X}_1}] \implies \varphi''_{\widetilde{X}_1}(0) = -\sigma^2\ .$$

Using a Taylor expansion, we find:

$$\begin{aligned} \varphi_{\widetilde{X}_1}(u) &= \varphi_{\widetilde{X}_1}(0) + u\,\varphi'_{\widetilde{X}_1}(0) + \frac{u^2}{2}\ \big(\varphi''_{\widetilde{X}_1}(0) - \epsilon(u)\big) \quad (\text{where } \epsilon(u) \to 0 \text{ for } u \to 0)\ , \\ &= 1 - \frac{u^2}{2}\ (\sigma^2 + \epsilon(u))\ . \end{aligned}$$

Using (2.4.3), we obtain

$$\begin{aligned} \varphi_{Z_n}(t) &= \Big(1 - \frac{t^2}{2\sigma^2 n}\ \Big(\sigma^2 + \epsilon\Big(\frac{t}{\sigma\sqrt{n}}\Big)\Big)\Big)^n \\ &= \exp\Big\{\underbrace{n\log\Big(1 - \frac{t^2}{2\sigma^2 n}\ \Big(\sigma^2 + \epsilon\Big(\frac{t}{\sigma\sqrt{n}}\Big)\Big)\Big)}_{\underset{n\to\infty}{\sim}\ n \times -\frac{t^2}{2\sigma^2 n}(\sigma^2 + \epsilon(\frac{t}{\sigma\sqrt{n}})) \longrightarrow -\frac{t^2}{2}}\Big\}\ . \end{aligned}$$

Consequently, $\lim_n \varphi_{Z_n}(t) = \exp\{-\frac{t^2}{2}\}$, and (2.4.2) follows from (2.3.11) and (2.3.25). $\square$

If $E[|X_1|^3] < \infty$, one can actually estimate quantitatively the difference

$$\sup_{x\in\mathbb{R}} |F_n(x) - \mathcal{N}(x)|\ ,$$

where

$$F_n(x) = P[Z_n \le x] \quad (\text{distribution function of } Z_n),$$
$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\ du \quad (\text{distribution function of the standard normal distribution}).$$

One has indeed (see Durrett, p.108):

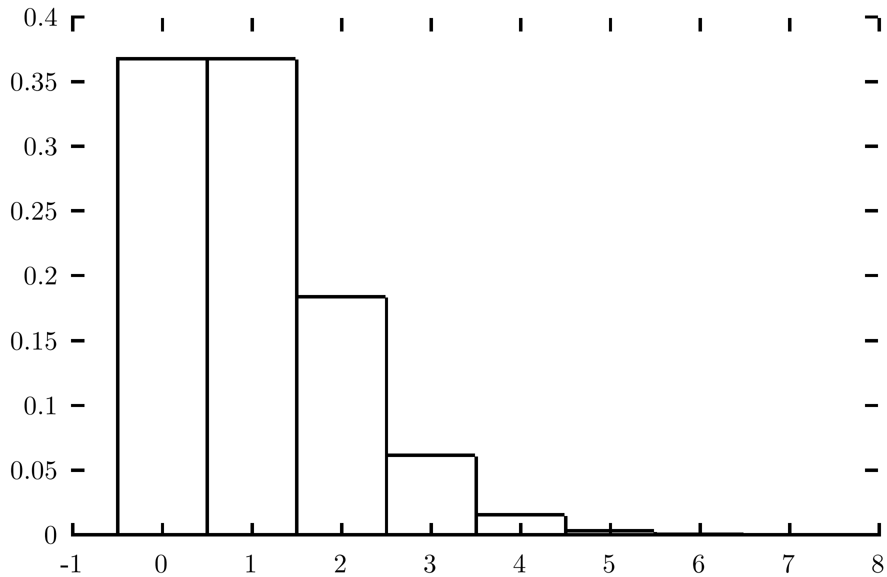


Fig. 2.8: Poisson distribution with expectation 1

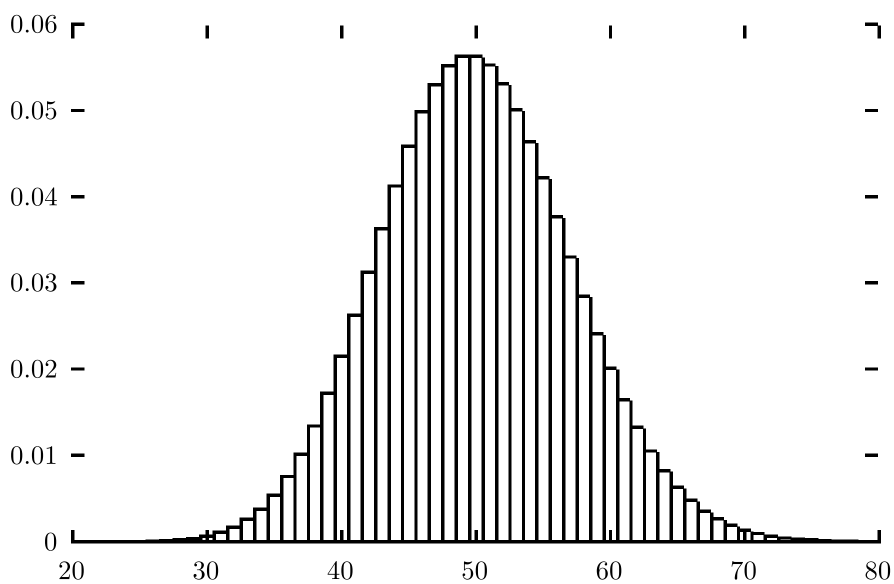


Fig. 2.9: Its 50th convolution power

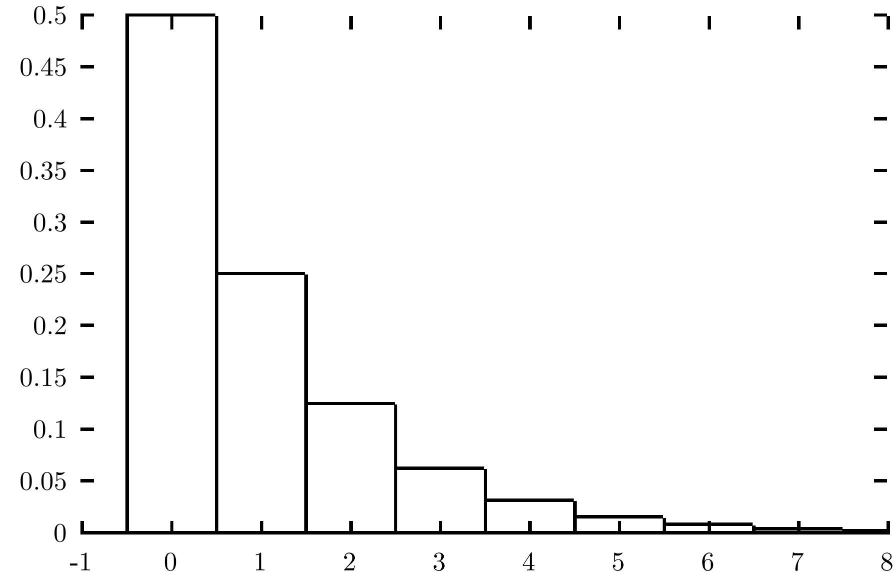


Fig. 2.10: Geometric distribution with expectation 1

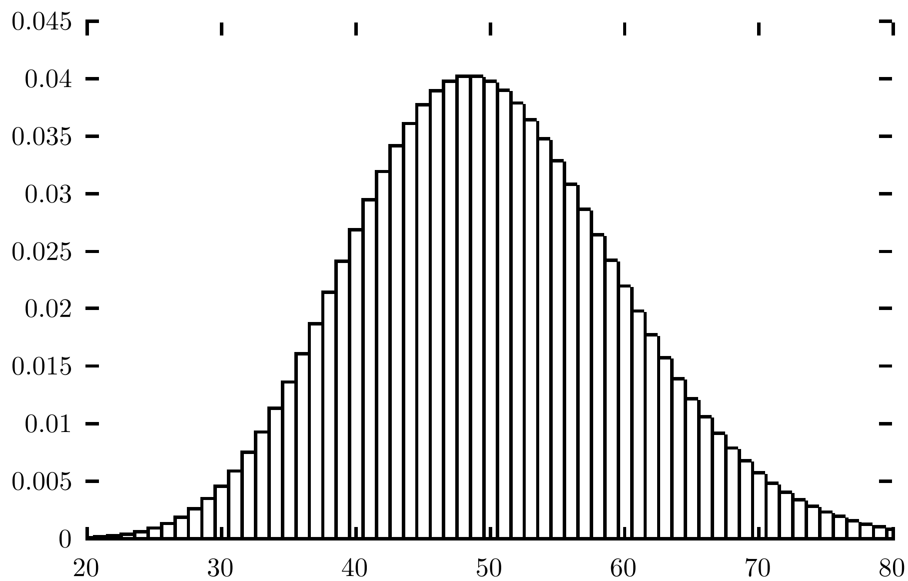


Fig. 2.11: Its 50th convolution power

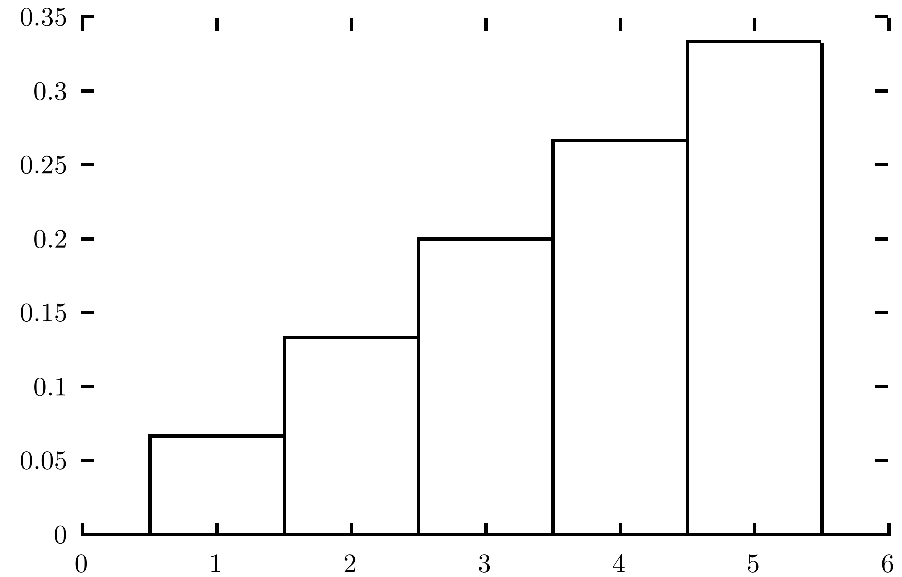


Fig. 2.12: A triangular distribution

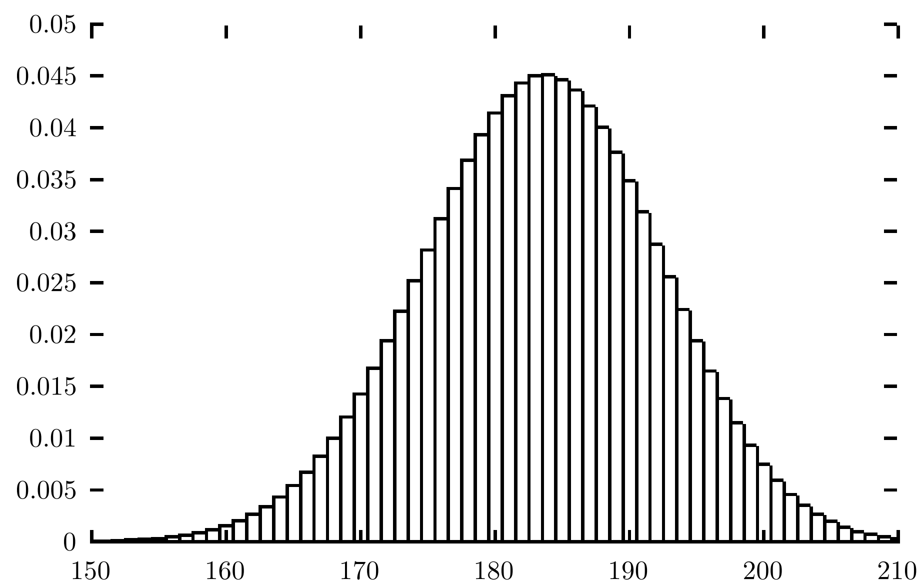


Fig. 2.13: Its 50th convolution power

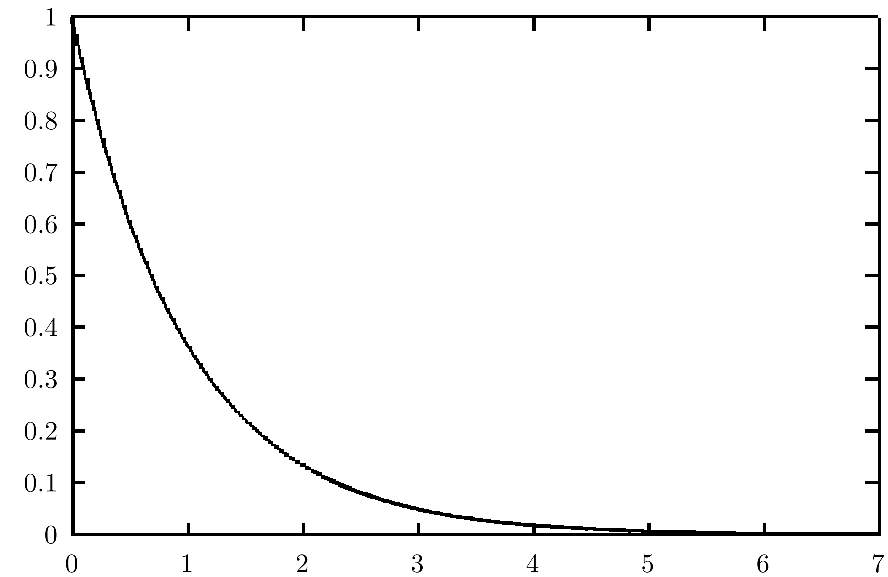


Fig. 2.14: Density of the exponential distribution (with expectation 1)

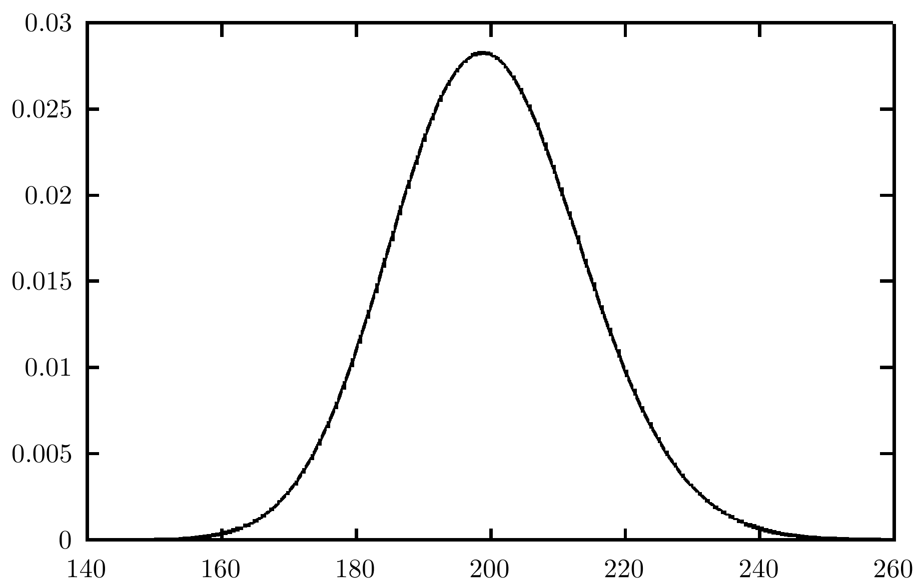


Fig. 2.15: Its 200th convolution power

Theorem 2.21. *(Berry-Esseen)*

Let $X_1, \dots, X_n$ be independent and identically distributed random variables, with

$$E[|X_i|^3] = \rho < \infty,\ E[X_i] = 0,\ E[X_i^2] = \sigma^2,$$

then one has

$$\sup_{x \in \mathbb{R}} |F_n(x) - \mathcal{N}(x)| \le \frac{3\rho}{\sigma^3 \sqrt{n}} \,. \tag{2.4.4}$$

Remark 2.22. The order of magnitude cst$/\sqrt{n}$ in (2.4.4) is correct in general. If, for instance, $P[X_i = 1] = P[X_i = -1] = \frac{1}{2}$, then

$$P[S_{2n} = 0] \overset{(2.1.7)}{\sim} \frac{1}{\sqrt{\pi n}}, \quad n \to \infty \,.$$

For symmetry reasons, one has also

$$\begin{aligned} P[S_{2n} \le 0] &= P[S_{2n} < 0] + P[S_{2n} = 0] \\ &= \tfrac{1}{2}\left(P[S_{2n} < 0] + P[S_{2n} > 0] + 2P[S_{2n} = 0]\right) \\ &= \tfrac{1}{2}\big(\underbrace{P[S_{2n} \in \mathbb{R}]}_{=1} + P[S_{2n} = 0]\big) \\ &= \tfrac{1}{2} + \tfrac{1}{2}\, P[S_{2n} = 0] \,. \end{aligned}$$

Hence, we obtain:

$$\begin{aligned} P[S_{2n} \le 0] - \mathcal{N}(0) &= P[S_{2n} \le 0] - \tfrac{1}{2} \\ &= \tfrac{1}{2}\, P[S_{2n} = 0] \sim \tfrac{1}{2\sqrt{\pi n}} \,. \end{aligned}$$

□

Complement: the Lindeberg-Feller theorem

We now present a generalization of the Central Limit Theorem in the case of "triangular arrays":

(2.4.5) $X_{n,m}, 1 \le m \le n$, are, for each $n \ge 1$, independent and integrable random variables with

$$E[X_{n,m}] = 0 \quad \text{for } 1 \le m \le n \,. \tag{2.4.6}$$

Example 2.23. If $X_i, i \ge 1$, are i.i.d. integrable random variables with $E[X_i] = \mu$, $i \ge 1$, then

$$X_{n,m} = \frac{1}{\sqrt{n}}\,(X_m - \mu), \quad \text{for } 1 \le m \le n,$$

constitutes an example of a triangular array, with

$$\sum_{m=1}^{n} X_{n,m} = \frac{S_n - n\mu}{\sqrt{n}}, \quad \text{where } S_n = X_1 + \dots + X_n, \text{ for } n \ge 1 \,.$$

○

Theorem 2.24. *(Lindeberg-Feller)*

Let $X_{n,m}$, $1 \le m \le n$, be a triangular array of random variables, in the sense of (2.4.5), (2.4.6), *with finite second moment. If*

$$\sum_{m=1}^{n} E[X_{n,m}^2] \underset{n\to\infty}{\longrightarrow} \sigma^2 > 0\,, \tag{2.4.7}$$

and

$$\text{for any } \epsilon > 0, \quad \lim_{n\to\infty} \sum_{m=1}^{n} E[X_{n,m}^2;\, |X_{n,m}| > \epsilon] = 0\,, \tag{2.4.8}$$

then

(2.4.9) *$X_{n,1} + \cdots + X_{n,n}$ converges in distribution, as n tends to ∞, to an $\mathcal{N}(0,\sigma^2)$-distributed random variable.*

Remark 2.25. The Central Limit Theorem, see (2.4.1), (2.4.2), is a direct consequence of the Lindeberg-Feller theorem. Consider a sequence $X_i, i \ge 1$, of i.i.d. random variables so that (2.4.1) is satisfied. Define $X_{n,m}$, $1 \le m \le n$, as in the example above. Then:

$$\sum_{m=1}^{n} E[X_{n,m}^2] = \frac{1}{n} \sum_{m=1}^{n} E[(X_m - \mu)^2] = \sigma^2 \overset{\text{def.}}{=} \mathrm{Var}(X_1) > 0\,,$$

and for any $\epsilon > 0$,

$$\begin{aligned}\sum_{m=1}^{n} E\big[X_{n,m}^2;\, |X_{n,m}| > \epsilon\big] &= \frac{1}{n} \sum_{m=1}^{n} E\Big[(X_m - \mu)^2;\, \Big|\frac{X_m - \mu}{\sqrt{n}}\Big| > \epsilon\Big] \\ &= E\big[(X_1 - \mu)^2;\, |X_1 - \mu| > \sqrt{n}\,\epsilon\big] \underset{n\to\infty}{\longrightarrow} 0 \quad \text{(dominated convergence)}\,.\end{aligned}$$

This implies that (2.4.7) and (2.4.8) are satisfied, and using (2.4.9), $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ converges in distribution to a random variable with standard normal distribution, as in the Central Limit Theorem (see just below (2.4.2)). $\square$

Proof of Theorem 2.24. We start with

Lemma 2.26.

$$\Big|e^{iu} - 1 - iu + \frac{1}{2}\,u^2\Big| \le \min\Big(\frac{|u|^3}{6},\, u^2\Big) \qquad \textit{for all } u \in \mathbb{R}\,, \tag{2.4.10}$$

$$\big|\log(1+z) - z\big| \le 2\,|z|^2 \qquad \textit{for all } z \textit{ in } \mathbb{C} \textit{ with } |z| < \frac{1}{2}\,, \tag{2.4.11}$$

where $\log w = \log \rho + i\theta$, *for* $w = \rho e^{i\theta}$, *with* $\rho > 0$ *and* $\theta \in (-\pi, \pi)$.

Proof. For a function f analytic in $D = \{z \in \mathbb{C} : |z| < r\}$ and $z \in D$, Taylor's formula implies that for $n \geq 1$,

$$\begin{aligned} f(z) - f(0) - f'(0)\,z - \cdots - f^{(n-1)}(0)\,\frac{z^{n-1}}{(n-1)!} &\overset{\text{def}}{=} R_n(f,z) \\ &= \int_0^1 f^{(n)}(tz)(1-t)^{n-1} dt\,\frac{z^n}{(n-1)!}\,. \end{aligned}$$

Furthermore, one has

$$\big|R_n(f,z)\big| \leq \sup_{0 \leq t \leq 1} |f^{(n)}(tz)|\,\frac{|z|^n}{n!}\,.$$

In particular for $f(z) = e^{iz}$, we obtain, with $n = 3$ and 2:

$$\Big|e^{iu} - 1 - iu + \frac{1}{2}\,u^2\Big| \leq \frac{|u|^3}{6} \ \text{ for } u \in \mathbb{R}\,,$$

$$|e^{iu} - 1 - iu| \leq \frac{|u|^2}{2} \ \text{ for } u \in \mathbb{R}, \quad \text{and thus } \Big|e^{iu} - 1 - iu + \frac{1}{2}\,u^2\Big| \leq u^2\,.$$

This shows (2.4.10). In the case of $f(z) = \log(1+z)$, $|z| < \frac{1}{2}$, we obtain

$$|\log(1+z) - z| \leq \sup_{0 \leq t \leq 1} \frac{1}{|1+tz|^2}\,\frac{|z|^2}{2} \leq 2\,|z|^2,$$

as claimed in (2.4.11). □

We now return to the proof of Theorem 2.24. We show that for all $t \in \mathbb{R}$,

(2.4.12) $$\lim_{n\to\infty} E\big[e^{it(X_{n,1}+\cdots+X_{n,n})}\big] = e^{-\frac{t^2}{2}\,\sigma^2}\,.$$

Then, (2.4.9) will follow from (2.3.12) and (2.3.25). We use the following notations:

$$\varphi_{n,m}(t) = E[e^{it\,X_{n,m}}], \quad \sigma^2_{n,m} = \mathrm{Var}(X_{n,m}),$$

$$w_{n,m}(t) = \varphi_{n,m}(t) - 1 + \frac{t^2}{2}\,\sigma^2_{n,m}\,.$$

Using independence (see (2.4.5)), one has

(2.4.13) $$E\Big[e^{it(X_{n,1}+\cdots+X_{n,n})}\Big] = \prod_{m=1}^n \varphi_{n,m}(t) \qquad \text{for all } t \in \mathbb{R}\,.$$

We now prove the following:

(2.4.14) $$\lim_{n\to\infty} \sum_{m=1}^n |w_{n,m}(t)| = 0 \qquad \text{for all } t \in \mathbb{R}\,.$$

Indeed, for $\epsilon > 0$ and $t \in \mathbb{R}$, one has, using (2.4.6),

$$\begin{aligned}|w_{n,m}(t)| &= \Big|E\Big[e^{it\,X_{n,m}} - 1 - it\,X_{n,m} + \frac{t^2}{2}\,X_{n,m}^2\Big]\Big| \\ &\overset{(2.4.10)}{\leq} E\Big[\min\Big(\frac{|tX_{n,m}|^3}{6},\ |tX_{n,m}|^2\Big)\Big] \\ &\leq E\Big[\frac{|tX_{n,m}|^3}{6};\ |X_{n,m}| \leq \epsilon\Big] + E\big[|tX_{n,m}|^2;\ |X_{n,m}| > \epsilon\big]\,,\end{aligned}$$

so that

$$\sum_{m=1}^n |w_{n,m}(t)| \leq \frac{\epsilon}{6}\,|t|^3 \sum_{m=1}^n E[X_{n,m}^2] + t^2 \sum_{m=1}^n E\big[X_{n,m}^2;\ |X_{n,m}| > \epsilon\big]\,.$$

Now, if we first send n to infinity, and then ϵ to 0, we obtain (2.4.14) as a consequence of (2.4.7) and (2.4.8). Note that

(2.4.15) $$\lim_{n\to\infty} \sup_{1\leq m\leq n} \sigma_{n,m}^2 = 0\,,$$

since for all $\epsilon > 0$, one has

$$\sup_{1\leq m\leq n} \sigma_{n,m}^2 \leq \sup_{1\leq m\leq n} \Big(E[X_{n,m}^2;\ |X_{n,m}| > \epsilon] + \epsilon^2\Big),$$

and thus, using (2.4.8),

$$\overline{\lim_{n\to\infty}} \sup_{1\leq m\leq n} \sigma_{n,m}^2 \leq \epsilon^2\,.$$

If we let ϵ tend to 0, then (2.4.15) follows. Let us then set, for $t \in \mathbb{R}$,

(2.4.16) $$z_{n,m}(t) = \varphi_{n,m}(t) - 1 = -\frac{t^2}{2}\,\sigma_{n,m}^2 + w_{n,m}(t)\,.$$

For large n, and $1 \leq m \leq n$, one has, thanks to (2.4.14) and (2.4.15),

$$|z_{n,m}(t)| \leq \frac{t^2}{2}\,\sigma_{n,m}^2 + |w_{n,m}(t)| < \frac{1}{2}\,.$$

Using (2.4.11), we obtain

(2.4.17) $$\Big|\log \varphi_{n,m}(t) - z_{n,m}(t)\Big| \leq 2\,|z_{n,m}(t)|^2\,,$$

and so

(2.4.18) $$\begin{aligned}\Big|\sum_{m=1}^n (\log \varphi_{n,m} - z_{n,m})\Big| &\leq 2\sum_{m=1}^n \Big(-\frac{t^2}{2}\,\sigma_{n,m}^2 + w_{n,m}(t)\Big)^2 \\ &\leq \sum_{m=1}^n \big(t^4\,\sigma_{n,m}^4 + 4w_{n,m}(t)^2\big) \qquad (\text{since } (a+b)^2 \leq 2a^2 + 2b^2) \\ &\leq t^4 \sup_{1\leq m\leq n} \sigma_{n,m}^2 \sum_{m=1}^n \sigma_{n,m}^2 + 4\Big(\sum_{m=1}^n |w_{n,m}(t)|\Big)^2 \underset{n\to\infty}{\longrightarrow} 0\,,\end{aligned}$$

using (2.4.7), (2.4.15), and (2.4.14) in the last step. Thanks to (2.4.18), (2.4.16), and (2.4.14), we obtain

$$\lim_{n\to\infty}\left|\sum_{m=1}^{n}\log\varphi_{n,m}(t)+\frac{t^2}{2}\sum_{m=1}^{n}\sigma_{n,m}^2\right|=0 \qquad \text{for all } t\in\mathbb{R},$$

and using (2.4.7),

$$\lim_{n\to\infty}\sum_{m=1}^{n}\log\varphi_{n,m}(t)=-\frac{t^2}{2}\,\sigma^2 \qquad \text{for all } t\in\mathbb{R}. \tag{2.4.19}$$

Hence, for all $t\in\mathbb{R}$,

$$\lim_{n\to\infty}\prod_{m=1}^{n}\varphi_{n,m}(t)=e^{-\frac{t^2}{2}\sigma^2},$$

which proves (2.4.12), using (2.4.13). □

Example 2.27. Consider $Y_m, m\ge 1$, some independent Bernoulli distributed random variables, with respective success probabilities

$$P[Y_m=1]=\frac{1}{m}. \tag{2.4.20}$$

Define $S_n=Y_1+\cdots+Y_n$. Then, one has $E[Y_m]=\frac{1}{m}$, $\mathrm{Var}(Y_m)=\frac{1}{m}-\frac{1}{m^2}$, and thus

$$\begin{aligned} &E[S_n]=1+\frac{1}{2}+\cdots+\frac{1}{n}\sim\log n,\\ \text{and}\quad &\mathrm{Var}(S_n)=\mathrm{Var}(Y_1)+\cdots+\mathrm{Var}(Y_n)\sim\log n, \end{aligned} \tag{2.4.21}$$

as $n\to\infty$. We set

$$X_{n,m}=\frac{(Y_m-\frac{1}{m})}{\sqrt{\log n}}, \qquad \text{so that } E[X_{n,m}]=0, \text{ and } \sum_{m=1}^{n}E[X_{n,m}^2]\underset{n\to\infty}{\longrightarrow}1.$$

Since $|X_{n,m}|\le(\log n)^{-1/2}$, one also has, for $\epsilon>0$,

$$\sum_{m=1}^{n}E\big[X_{n,m}^2;|X_{n,m}|>\epsilon\big]=0 \qquad \text{if } \sqrt{\log n}\,\epsilon>1.$$

In other words, hypotheses (2.4.7) and (2.4.8) of the Lindeberg-Feller theorem are satisfied, and consequently

$$\frac{\big(S_n-\sum_{m=1}^{n}\frac{1}{m}\big)}{\sqrt{\log n}} \quad \text{converges in distribution to an } N(0,1)\text{-distributed random variable}. \tag{2.4.22}$$

Note also that for all $n \geq 2$,

$$\sum_{m=1}^{n-1} \frac{1}{m} \geq \int_1^n \frac{dx}{x} = \log n \geq \sum_{m=2}^{n} \frac{1}{m},$$

$$\text{and so} \quad \left| \sum_{m=1}^{n} \frac{1}{m} - \log n \right| \leq 1.$$

Thanks to the equivalence of (2.2.13) and (2.2.14), it follows that

$$\frac{S_n - \log n}{\sqrt{\log n}} \quad \begin{array}{l}\text{converges in distribution, as } n \to \infty, \\ \text{to an } N(0,1)\text{-distributed random variable}.\end{array} \tag{2.4.23}$$

$\square$

3 Martingales

3.1 Conditional expectation

We start this chapter by introducing the notion of conditional expectation. This is a quite abstract notion, therefore we first discuss two concrete examples.

Example 3.1.

1) We consider two urns A and B. The urn A (resp. B) contains $X(\omega)$ (resp. $Y(\omega)$) balls. The random variables X and Y are independent, Poisson (λ_A) and Poisson (λ_B) distributed, with $\lambda_A, \lambda_B > 0$, and they are defined on a common probability space $(\Omega, \mathcal{A}, P)$.

Our **first goal** is to **find the distribution of the number of balls in A, given that the total number of balls $T = X + Y$ is equal to n**. In other words, we compute

$$P\big[X = k | T = n\big], \;\; 0 \le k \le n \;, \tag{3.1.1}$$

where we use the notation for "conditional probability"

$$P\big[C|D\big] \stackrel{\text{def.}}{=} \frac{P[C \cap D]}{P[D]}$$

("probability of C given D"), for events C, D with $P[D] > 0$.

One has:

$$\begin{aligned} P\big[X = k|T = n\big] &= \frac{P[X = k, T = n]}{P[T = n]} = \frac{P[X = k, Y = n-k]}{P[T = n]} \\ &\stackrel{\text{independence}}{=} \underset{\substack{\uparrow \\ T \text{ is Poisson } (\lambda_A + \lambda_B) \text{ distributed}}}{\frac{P[X = k]\, P[Y = n-k]}{P[T = n]}} \\ &= e^{-\lambda_A}\, \frac{\lambda_A^k}{k!} \cdot e^{-\lambda_B}\, \frac{\lambda_B^{n-k}}{(n-k)!} \cdot \frac{1}{e^{-(\lambda_A+\lambda_B)} \frac{(\lambda_A+\lambda_B)^n}{n!}} \\ &= \frac{n!}{k!(n-k)!} \Big(\frac{\lambda_A}{\lambda_A+\lambda_B}\Big)^k \Big(\frac{\lambda_B}{\lambda_A+\lambda_B}\Big)^{n-k} \\ &= \tbinom{n}{k} p_A^k (1-p_A)^{n-k}, \end{aligned} \tag{3.1.2}$$

where $p_A = \frac{\lambda_A}{\lambda_A+\lambda_B}$ and $1 - p_A = p_B = \frac{\lambda_B}{\lambda_A+\lambda_B}$.

In other words: if the total number T of balls is equal to n, then the number X of balls in A has a binomial distribution with parameter (n, p_A).

Hence, the "expected number of balls in A given that $T = n$" is:

$$E[X|T = n] \stackrel{\text{def.}}{=} \sum_{k=0}^{n} k\, P[X = k|T = n] = n \cdot p_A \;. \tag{3.1.3}$$

Define the random variable

$$Z = p_A \cdot T \quad \text{“the conditional expectation of } X \text{ given } T\text{”}, \tag{3.1.4}$$

Z is a linear function of T, so that

$$Z \text{ is } \sigma(T) \longrightarrow \mathcal{B}(\mathbb{R}) \text{ measurable} . \tag{3.1.5}$$

Let $C \in \sigma(T)$, then C is of the form $\{T \in I\}$ for a certain $I \subseteq \mathbb{N}$. **With the help of $Z = T \cdot p_A$, we do not need to know the joint distribution of X and T any more to calculate $E[1_C \cdot X]$**, since one has:

$$\begin{aligned} E[1_C \cdot X] &= E\big[1_{\{T\in I\}} \cdot X\big] = \sum_{n\in I} E\big[1_{\{T=n\}} \cdot X\big] \\ &= \sum_{n\in I} E\big[X|T=n\big] \cdot P[T=n] \overset{(3.1.3)}{=} \sum_{n\in I} np_A\, P[T=n] \\ &= E[1_C \cdot Z] \ \text{ for } C \in \sigma(T) . \end{aligned} \tag{3.1.6}$$

The usual notation for Z is: $Z = E[X|T]$, or $E[X|\sigma(T)]$.

Z has a further interpretation: consider

$$\begin{aligned} H &= \Big\{ f(T) \ : \ \sum_{n=0}^{\infty} f^2(n)\, \frac{(\lambda_A+\lambda_B)^n}{n!}\, e^{-(\lambda_A+\lambda_B)} < \infty \Big\} \\ &= \big\{ f(T) \ : \ E[f^2(T)] < \infty \big\} . \end{aligned} \tag{3.1.7}$$

In fact, H is the set $L^2(\Omega, \sigma(T), P)$ of square-integrable $\sigma(T)$-measurable functions (so a Hilbert space!). One has $Z \in H$, and for $Z' = f(T) \in H$,

$$\begin{aligned} E\big[(X-Z')^2\big] &= E\big[(X-Z+Z-Z')^2\big] \\ &= E\big[(X-Z)^2\big] + 2E\big[(X-Z)(Z-Z')\big] + E\big[(Z-Z')^2\big] . \end{aligned} \tag{3.1.8}$$

Z and Z' are functions of T, and thanks to (3.1.6) one has

$$\begin{aligned} E\big[Z\cdot(Z-Z')\big] &= \sum_{n=0}^{\infty} E\big[Z\cdot(\underbrace{Z-Z'}_{\text{constant on } \{T=n\}})\, 1\{T=n\}\big] \\ &\overset{(3.1.6)}{=} \sum_{n=0}^{\infty} E\big[X\cdot(Z-Z')\, 1\{T=n\}\big] \\ &\overset{\text{Lebesgue}}{=} E\big[X\cdot(Z-Z')\big] \\ &\quad (\text{since } \ X, Z, Z' \in L^2 \Longrightarrow X(Z-Z') \in L^1) . \end{aligned}$$

We obtain

$$E\big[(X-Z')^2\big] \geq E\big[(X-Z)^2\big] \text{ for } Z' \in H\,. \tag{3.1.9}$$

That is to say, Z is the orthogonal projection of $X(\in L^2(\Omega,\mathcal{A},P))$ onto $H = L^2(\Omega,\sigma(T),P)$, i.e. Z minimizes $E\big[(X-Z')^2\big]$ for $Z' \in H$, "Z is the best forecast for X among all L^2-functions $f(T)$".

2) Consider the random variables X and T on $\mathbb{R}^2$ with joint density $f(x,t) > 0$, $x,t \in \mathbb{R}^2$, and $E[|X|] < \infty$.

The conditional density of X given T is defined as

$$f(x|t) = \frac{f(x,t)}{\displaystyle\int_{\mathbb{R}} f(u,t)du} \quad \text{"density in } x \text{ for fixed } t\text{"}\,. \tag{3.1.10}$$

One can define the "conditional expectation of X given T" as

$$Z = \varphi(T), \text{ where } \varphi(t) = \int_{\mathbb{R}} x\, f(x|t)dx\,. \tag{3.1.11}$$

We can note that Z satisfies equations (3.1.5) - (3.1.6). Indeed, (3.1.5) is clear, and if we consider

$$C = \{T \in \underset{\substack{\uparrow \\ I \in \mathcal{B}(\mathbb{R})}}{I}\} \in \sigma(T),$$

then

$$\begin{aligned} E[X\,1_C] =&\ E[X \cdot 1_I \circ T] = \iint x \cdot 1_I(t)\, f(x,t)dxdt \\ =&\ \int \varphi(t) \cdot 1_I(t) \times \underbrace{\int_{\mathbb{R}} f(u,t)du}_{\text{density of } T}\, dt = E[\varphi(T)\, 1_I \circ T] = E[Z \cdot 1_C]\,. \end{aligned}$$

○

We have just seen two examples in which the conditional expectation of a random variable X given the information of a certain sub-σ-algebra ($= \sigma(T)$) of $\mathcal{A}$ was defined. Our goal now is to give a general construction, and then to study its properties.

Theorem 3.2. *Let X be an integrable random variable on $(\Omega,\mathcal{A},P)$, and $\mathcal{F}$ a sub-σ-algebra of $\mathcal{A}$. Then there exists a random variable Z such that*

$$Z \text{ is } \mathcal{F}\text{-measurable and integrable,} \tag{3.1.12}$$

$$E[X \cdot 1_F] = E[Z \cdot 1_F] \text{ for all } F \in \mathcal{F}\,. \tag{3.1.13}$$

Z is uniquely determined, up to sets with P-measure zero, by (3.1.12) - (3.1.13)*. Moreover,*

$$\text{if } X \geq 0, \text{ then also } Z \geq 0 \ P\text{-a.s.} \tag{3.1.14}$$

Notation: $Z = E[X|\mathcal{F}]$ "the conditional expectation of X given $\mathcal{F}$".

Proof.

a) ***Existence:***

- Consider first $X \geq 0$. The property is an application of the Radon-Nikodym theorem: let Q be a measure on $(\Omega, \mathcal{A})$ defined by

$$Q(A) = \int_A X\,dP \;\Big(\text{i.e. } \frac{dQ}{dP} = X\Big)\,. \tag{3.1.15}$$

  Since $\mathcal{F} \subseteq \mathcal{A}$, we can introduce the restrictions $\widetilde{Q}$ and $\widetilde{P}$ of, respectively, Q and P to $(\Omega, \mathcal{F})$. For $F \in \mathcal{F}$ with $\widetilde{P}(F) = 0$, one has

$$\widetilde{P}(F) = P(F) = 0 \implies \widetilde{Q}(F) = Q(F) = \int_F X\,dP = 0\,,$$

  i.e. $\widetilde{Q} \ll \widetilde{P}$ ("$\widetilde{Q}$ is absolutely continuous with respect to $\widetilde{P}$").

$$\overset{\substack{\text{Radon-Nikodym}\\ \text{theorem}}}{\Longrightarrow} \quad \text{there exists a } Z \geq 0 \text{ in } L^1(\Omega, \mathcal{F}, \widetilde{P}) \text{ such that}$$

$$\underset{E[X\,1_F]}{\underset{\parallel}{\widetilde{Q}}}\;(F) = \int_{\underbrace{F}_{\mathcal{F}\text{-measurable}}} \underbrace{Z}_{\swarrow}\,d\widetilde{P} \overset{\substack{(P \text{ and } \widetilde{P}\\ \text{coincide on } \mathcal{F})}}{=} \int_F Z\,dP\,,$$

  i.e. Z satisfies (3.1.12) - (3.1.13). Of course, one also has (3.1.14).

- *General case:*

  We write $X = X^+ - X^-$, where $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$. Then, the construction above produces Z^+, Z^- satisfying (3.1.12) - (3.1.13) with respect to X^+, X^-. Hence, $Z = Z^+ - Z^-$ satisfies (3.1.12) - (3.1.13) with respect to $X = X^+ - X^-$.

b) ***Uniqueness:***

Let Z^1, Z^2 be given that both satisfy (3.1.12) - (3.1.13). Consider $D = Z^1 - Z^2$. Then D satisfies (3.1.12) and $E[D \cdot 1_F] = 0$, $\forall F \in \mathcal{F} \implies 0 = E[D \cdot Y] = E[|D|]$ for $Y = 1\{D > 0\} - 1\{D < 0\}$. This implies that $D = 0$, P-a.s. □

Example 3.3. 1) Let $A_i \in \mathcal{A}$, $1 \leq i \leq N \leq \infty$, be pairwise disjoint events with $P(A_i) > 0$ and $\bigcup A_i = \Omega$. Set $\mathcal{F} = \sigma(A_i, 1 \leq i \leq N)$ ($\mathcal{F}$ is the family of sets $\bigcup_{i \in I} A_i$, where $I \subseteq \{1, \ldots, N\}$). Consider $X \in L^1(\Omega, \mathcal{A}, P)$ and $Z = E[X|\mathcal{F}]$.

Thanks to (3.1.12), Z is $\mathcal{F}$-measurable $\implies Z = a_i$ on $A_i, \forall i$, and using (3.1.13):

$$a_i\,P[A_i] = E[Z\,1_{A_i}] \overset{(3.1.13)}{=} E[X\,1_{A_i}] \implies a_i = E[X\,1_{A_i}]/P[A_i] = E[X|A_i]\,.$$

Hence,

$$E[X|\mathcal{F}] = \sum_{i=1}^{N} E[X|A_i]\, 1_{A_i} \ . \tag{3.1.16}$$

2) *Special case:*

For $\mathcal{F} = \{\phi, \Omega\}$ (i.e. $A_1 = \Omega, N = 1$),

$$E[X|\mathcal{F}] = E[X] \tag{3.1.17}$$

"the best prediction of X, when one has no information, is $E[X]$".

3) X and $\mathcal{F}$ are independent. Then for $F \in \mathcal{F}$, $E[X\, 1_F] = E[X] \cdot P[F]$. Hence,

$$E[X|\mathcal{F}] = E[X] \ \ P\text{-a.s.} \tag{3.1.18}$$

We now investigate a few properties of the conditional expectation.

Expectation:

For $F = \Omega \in \mathcal{F}$, (3.1.13) implies that

$$E\big[E[X|\mathcal{F}]\big] = E\big[E[X|\mathcal{F}] \cdot 1_\Omega\big] = E[X \cdot 1_\Omega] = E[X] \ , \tag{3.1.19}$$

in other words, $E[X|\mathcal{F}]$ has the same expectation as X.

Linearity:

For all $a, b \in \mathbb{R}$, $X, Y \in L^1(\Omega, \mathcal{A}, P)$, one has

$$E\big[aX + bY|\mathcal{F}\big] = aE\big[X|\mathcal{F}\big] + bE\big[Y|\mathcal{F}\big] \ \ P\text{-a.s.} \tag{3.1.20}$$

(this is clear from the definition).

Jensen's inequality:

Let φ: $\mathbb{R} \to \mathbb{R}$ be a convex function, X a random variable on $(\Omega, \mathcal{A}, P)$ with $E[|X|]$ and $E[|\varphi(X)|] < \infty$, and $\mathcal{F}$ a sub-σ-algebra of $\mathcal{A}$. Then

$$\varphi\big(E[X|\mathcal{F}]\big) \le E\big[\varphi(X)|\mathcal{F}\big] \ \ P\text{-a.s.} \tag{3.1.21}$$

$\bigcirc$

Proof. **a)** Let us consider $\varphi(x) = ax + b$. Then it is clear that

$$\begin{aligned} E\big[aX + b|\mathcal{F}\big] &= aE\big[X|\mathcal{F}\big] + bE\big[1|\mathcal{F}\big] \\ &= aE\big[X|\mathcal{F}\big] + b = \varphi\big(E\big[X|\mathcal{F}\big]\big) \ \ P\text{-a.s.} \end{aligned} \tag{3.1.22}$$

b) *General case:*

One can write $\varphi(x) = \sup_{n \ge 1} \ \varphi_n(x)$, with φ_n of the form $\varphi_n(x) = a_n x + b_n$ for some $a_n, b_n \in \mathbb{R}$ (see Figure): for example, when φ is not an affine function, introduce

$$H_\varphi = \{(a,b) \in \mathbb{R}^2 \ : \ \forall x \in \mathbb{R}, \ \varphi(x) \ge ax + b\} \, ,$$

and check (using the fact that φ is continuous, and lies above its "support lines") that

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \sup_{(a,b)\in H_\varphi} (ax+b) = \sup_{(a,b)\in H_\varphi \cap \mathbb{Q}^2} (ax+b)\ .$$

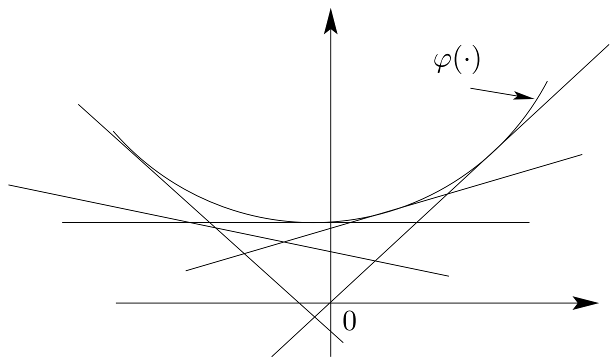


Fig. 3.1

Then, for all $n \geq 1$,

$$E\big[\varphi(X)|\mathcal{F}\big] \underset{P\text{-a.s.}}{\overset{(3.1.14)}{\geq}} E\big[\varphi_n(X)|\mathcal{F}\big] \underset{P\text{-a.s.}}{\overset{(3.1.22)}{\geq}} \varphi_n\big(E\big[X|\mathcal{F}\big]\big)\ .$$

It follows that P-a.s. (here, we use the fact that the supremum is over a countable family),

$$E\big[\varphi(X)|\mathcal{F}\big] \geq \sup_n \varphi_n\big(E\big[X|\mathcal{F}\big]\big) = \varphi\big(E\big[X|\mathcal{F}\big]\big)\ .$$

□

Corollary 3.4. *Consider $X \in L^p(\Omega, \mathcal{A}, P)$, $1 \leq p \leq \infty$, and $\mathcal{F}$ a sub-σ-algebra of $\mathcal{A}$, then*

(3.1.23) $$E[X|\mathcal{F}] \in L^p(\Omega,\mathcal{A},P), \text{ and } \big\|E[X|\mathcal{F}]\big\|_p \leq \big\|X\big\|_p\ .$$

Proof. For $p \in [1,\infty)$, the claim follows from (3.1.21). For $p = \infty$, it comes from

$$P\text{-a.s.} \ -M \leq X \leq M, \ \text{ where } \ M = \|X\|_\infty, \ \text{ and (3.1.14)}\ .$$

□

Theorem 3.5. *Consider $X \in L^1(\Omega,\mathcal{A},P)$, and Y an $\mathcal{F}$-measurable random variable with $E[|XY|] < \infty$. Then*

(3.1.24) $$E\big[XY|\mathcal{F}\big] \overset{P\text{-a.s.}}{=} E\big[X|\mathcal{F}\big]\,Y\ .$$

Proof. **a)** *Special case $Y = 1_B$:*

In this case, $E[X|\mathcal{F}]\cdot Y$ is $\mathcal{F}$-measurable and integrable, so that for $C \in \mathcal{F}$,

$$E\big[XY\cdot 1_C\big] = E\big[X1_{\underbrace{B\cap C}_{\in\mathcal{F}}}\big] \overset{(3.1.13)}{=} E\big[E[X|\mathcal{F}]\,1_{B\cap C}\big] = E\big[E[X|\mathcal{F}]\,Y\cdot 1_C\big]\ .$$

Hence, $E[X|\mathcal{F}]\,Y$ satisfies (3.1.12), (3.1.13), and (3.1.24) holds.

b) $X \geq 0,\ Y \geq 0$.

We consider an increasing sequence Y_n of $\mathcal{F}$-measurable step functions with $Y_n \uparrow Y$. Thanks to a), one has, for $C \in \mathcal{F}$,

$$\begin{array}{rcl} E\big[X \cdot Y_n \cdot 1_C\big] & = & E\big[E[X|\mathcal{F}]\, Y_n \cdot 1_C\big] \\ \text{monotone convergence} \Big\downarrow n\to\infty & & \Big\downarrow n\to\infty \ \text{monotone convergence, using (3.1.14) to prove that the sequence is non-decreasing} \\ E\big[X \cdot Y \cdot 1_C\big] & = & E\big[E[X|\mathcal{F}]\, Y \cdot 1_C\big]\,. \end{array}$$

For $C = \Omega \in \mathcal{F}$, we obtain $E[X|\mathcal{F}]\, Y \in L^1(\Omega, \mathcal{F}, P)$, and for a general $C \in \mathcal{F}$, we see that (3.1.13) is satisfied. Hence, (3.1.24) holds true.

c) *General case:*

We set $X = X^+ - X^-$, $Y = Y^+ - Y^-$, where $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$, and likewise for Y: (3.1.24) then follows by using b). □

A simple application of the last theorem is the following

Special case:

Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{F} \subseteq \mathcal{A}$ a sub-σ-algebra, and X an $\mathcal{F}$-measurable integrable random variable. Then

$$E[X|\mathcal{F}] \overset{P\text{-a.s.}}{=} X\,. \tag{3.1.25}$$

A further application is the

Theorem 3.6. *Consider $X \in L^2(\Omega, \mathcal{A}, P)$ and $\mathcal{F}$ a sub-σ-algebra of $\mathcal{A}$, then*

$$\begin{array}{l} E[X|\mathcal{F}] \text{ is the orthogonal projection of } X \text{ onto} \\ \text{the sub-Hilbert space } L^2(\Omega, \mathcal{F}, P) \text{ of } L^2(\Omega, \mathcal{A}, P)\,. \end{array} \tag{3.1.26}$$

Proof. Set $Z = E[X|\mathcal{F}] \in L^2(\Omega, \mathcal{F}, P)$, using (3.1.23). It suffices to show that

$$E\big[(X - Z)^2\big] \leq E\big[(X - Z')^2\big] \ \text{ for } \ Z' \in L^2(\Omega, \mathcal{F}, P)\,. \tag{3.1.27}$$

Analogously to (3.1.8), one has

$$E\big[(X - Z')^2\big] = E\big[(X - Z)^2\big] + 2E\big[(X - Z)(Z - Z')\big] + E\big[(Z - Z')^2\big]\,,$$

and

$$\begin{aligned} E\big[X \cdot (Z - Z')\big] &\overset{(3.1.19)}{=} E\big[E[X \cdot (\underbrace{Z - Z'}_{\mathcal{F}\text{-measurable and } |X \cdot (Z - Z')| \in L^1})|\mathcal{F}]\big] \\ &\overset{(3.1.24)}{=} E\big[E[X|\mathcal{F}](Z - Z')\big] = E\big[Z \cdot (Z - Z')\big]\,. \end{aligned}$$

It follows that $E\big[(X - Z)(Z - Z')\big] = 0$, and

$$E\big[(X - Z')^2\big] = E\big[(X - Z)^2\big] + E\big[(Z - Z')^2\big] \geq E\big[(X - Z)^2\big]\,. \qquad \square$$

Remark 3.7. (3.1.26) can be interpreted as "$\boldsymbol{E[X|\mathcal{F}]}$ **is the best prediction of** $\boldsymbol{X}$ **among the** $\boldsymbol{\mathcal{F}}$**-measurable functions in** $\boldsymbol{L^2}$". □

As a last property, we consider the so-called tower property of conditional expectations

Proposition 3.8. *Let $\mathcal{F}_1 \subset \mathcal{F}_2$ be sub-σ-algebras of $\mathcal{A}$, and $X \in L^1(\Omega, \mathcal{A}, P)$, then*

$$E\big[E[X|\mathcal{F}_2]|\mathcal{F}_1\big] = E[X|\mathcal{F}_1] \;\; P\text{-a.s. (tower property)} \tag{3.1.28}$$

$$E\big[E[X|\mathcal{F}_1]|\mathcal{F}_2\big] = E[X|\mathcal{F}_1] \;\; P\text{-a.s.} \tag{3.1.29}$$

Proof.

- (3.1.28): Consider $F \in \mathcal{F}_1 \subseteq \mathcal{F}_2$, then

$$E\big[E[X|\mathcal{F}_2] \cdot 1_F\big] \overset{(3.1.13)}{=} E\big[X \cdot 1_F\big] \overset{(3.1.13)}{=} E\big[E[X|\mathcal{F}_1] \cdot 1_F\big] \,.$$

  Hence, $E[X|\mathcal{F}_1]$ satisfies (3.1.12) and (3.1.13) with respect to $E[X|\mathcal{F}_2]$, and (3.1.28) follows.

- (3.1.29): using (3.1.25) and $\mathcal{F}_1 \subseteq \mathcal{F}_2$, one has P-a.s.

$$E\big[E[X|\mathcal{F}_1]|\mathcal{F}_2\big] \overset{(3.1.25)}{=} E\big[X|\mathcal{F}_1\big] \,.$$ □

3.2 Martingales

Terminology:

Let $(\Omega, \mathcal{A}, P)$ be a probability space. An increasing sequence $\mathcal{F}_n, n \geq 0$, of sub-σ-algebras of $\mathcal{A}$ (i.e. $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots \subseteq \mathcal{A}$) is called a **filtration**.

Example 3.9. Consider a sequence $X_n, n \geq 0$, of random variables on $(\Omega, \mathcal{A}, P)$. We define the sub-σ-algebras $\mathcal{F}_n, n \geq 0$, of $\mathcal{A}$ by

$$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n), \;\; n \geq 0 \,. \tag{3.2.1}$$

$\mathcal{F}_n$ corresponds to the information that is contained in the random variables $X_0, X_1, \dots, X_n$, in other words, to the information that is available at time n: the filtration so-obtained is called the filtration generated by the random variables $X_n, n \geq 0$.

A sequence $X_n, n \geq 0$, of random variables is said to be $\mathcal{F}_n$-**adapted** if for all $n \geq 0$, X_n is $\mathcal{F}_n$-measurable (this is automatically satisfied if we take $\mathcal{F}_n$ to be the filtration generated by the X_n). ○

Definition 3.10. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $\mathcal{F}_n$, $n \geq 0$, a filtration. An $\mathcal{F}_n$-adapted sequence $X_n, n \geq 0$, of integrable random variables is called a martingale (resp. supermartingale, resp. submartingale) if*

$$\forall n \geq 0, \;\; E\big[X_{n+1}|\mathcal{F}_n\big] = X_n \quad P\text{-a.s. (martingale)}\,, \tag{3.2.2}$$

$$\forall n \geq 0, \;\; E\big[X_{n+1}|\mathcal{F}_n\big] \leq X_n \quad P\text{-a.s. (supermartingale)}\,, \tag{3.2.3}$$

$$\forall n \geq 0, \;\; E\big[X_{n+1}|\mathcal{F}_n\big] \geq X_n \quad P\text{-a.s. (submartingale)}\,. \tag{3.2.4}$$

Example 3.11. 1) Consider $X_i, i \geq 1$, independent and identically distributed, with $E[|X_i|] < \infty$ and $E[X_i] = 0$. Let $S_n, n \geq 0$, be the **random walk**

$$S_0 = 0, \;\; S_n = X_1 + \cdots + X_n, \; n \geq 1\,,$$
$$\mathcal{F}_0 = \{\phi, \Omega\}, \;\; \mathcal{F}_n = \sigma(X_1, \dots, X_n), \; n \geq 1\,.$$

Then $S_n, n \geq 0$, is $\mathcal{F}_n$-adapted, and for $n \geq 0$ one has

$$\begin{aligned} E\big[S_{n+1}|\mathcal{F}_n\big] \; &= \; E\big[S_n + X_{n+1}|\mathcal{F}_n\big] \\ = \; E\big[\underset{\substack{\uparrow \\ \mathcal{F}_n-\text{measurable}}}{S_n}|\mathcal{F}_n\big] + E\big[\underset{\substack{\uparrow \\ \text{independent of } \mathcal{F}_n}}{X_{n+1}}|\mathcal{F}_n\big] &\overset{(3.1.18)-(3.1.24)}{=} S_n + E[X_{n+1}] = S_n\,. \end{aligned} \tag{3.2.5}$$

Hence,

$$S_n \text{ is an } \mathcal{F}_n\text{-martingale}\,. \tag{3.2.6}$$

2) Same setting as in 1), but now with $E[X_i^2] = \sigma^2 < \infty$. Set

$$M_n = S_n^2 - n\sigma^2, \; n \geq 0\,. \tag{3.2.7}$$

M_n, $n \geq 0$, is $\mathcal{F}_n$-adapted, integrable, and

$$\begin{aligned} E\big[M_{n+1} - M_n|\mathcal{F}_n\big] \; &= \; E\big[S_{n+1}^2 - S_n^2 - \sigma^2|\mathcal{F}_n\big] \\ = \; E\big[(S_n + X_{n+1})^2 - S_n^2 - \sigma^2|\mathcal{F}_n\big] \; &= \; E\big[2\underbrace{S_n}_{\mathcal{F}_n-\text{measurable}} \cdot \underset{\substack{\uparrow \\ \text{independent of } \mathcal{F}_n}}{X_{n+1} + X_{n+1}^2} - \sigma^2|\mathcal{F}_n\big] \\ = \; 2S_n\, E\big[X_{n+1}|\mathcal{F}_n\big] + E\big[X_{n+1}^2|\mathcal{F}_n\big] - \sigma^2 & \\ = \; 2S_n\, E[\underset{\substack{\| \\ 0}}{X_{n+1}}] + \sigma^2 - \sigma^2 = 0\,. & \end{aligned}$$

Hence, $E\big[M_{n+1}|\mathcal{F}_n\big] = E\big[M_n|\mathcal{F}_n\big] = M_n, \; n \geq 0$.

$$\begin{aligned} &M_n \text{ is an } \mathcal{F}_n\text{-martingale, and} \\ &S_n^2 \text{ is an } \mathcal{F}_n\text{-submartingale}\,. \end{aligned} \tag{3.2.8}$$

3) We consider now the **asymmetric simple random walk** on $\mathbb{Z}$: $X_i, i \geq 1$, are independent and identically distributed, with $P[X_i = 1] = p, \; P[X_i = -1] = 1 - p$, for some $p \neq \frac{1}{2}$.

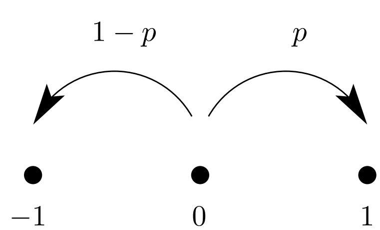


Fig. 3.2

S_n and $\mathcal{F}_n$, $n \ge 0$, are defined as in 1). Set

(3.2.9) $$M_n = \Big(\frac{1-p}{p}\Big)^{S_n} .$$

M_n is $\mathcal{F}_n$-adapted, and integrable (since $|S_n| \le n$). One also has

$$E[M_{n+1}|\mathcal{F}_n] = E\Big[\underset{\substack{\uparrow \\ \mathcal{F}_n\text{-measurable}}}{M_n} \cdot \underset{\substack{\uparrow \\ \text{independent of } \mathcal{F}_n}}{\underbrace{\Big(\frac{1-p}{p}\Big)^{X_{n+1}}}} \Big|\mathcal{F}_n\Big] = M_n\, E\Big[\Big(\frac{1-p}{p}\Big)^{X_1}\Big] ,$$

where
$$E\Big[\Big(\frac{1-p}{p}\Big)^{X_1}\Big] = p\Big(\frac{1-p}{p}\Big) + (1-p)\,\frac{p}{1-p} = 1 .$$

From this, we conclude that

(3.2.10) $$M_n \text{ is an } \mathcal{F}_n\text{-martingale} .$$

Note also that for $p > \frac{1}{2}$, $(\frac{1-p}{p}) < 1$ and $S_n \to +\infty$ P-a.s. In the same way, for $p < \frac{1}{2}$, one has $\frac{1-p}{p} > 1$ and $S_n \to -\infty$ P-a.s. Hence, we see that

(3.2.11) $$M_n \longrightarrow 0 \;\; P\text{-a.s.} ,$$

even though $E[M_n] = E[M_{n-1}] = \cdots = E[M_0] = 1$, thanks to (3.2.10).

4) *Radon-Nikodym derivatives:*

Let $(\Omega, \mathcal{F})$ be a measurable space, and $\mathcal{F}_n$, $n \ge 0$, a filtration with $\sigma(\bigcup_{n\ge 0} \mathcal{F}_n) = \mathcal{F}$. Also, let μ and ν be two probability measures on $(\Omega, \mathcal{F})$ with restrictions μ_n, resp. ν_n, on $(\Omega, \mathcal{F}_n)$. We assume that

(3.2.12) $$\mu_n \ll \nu_n ,$$

and define

(3.2.13) $$M_n = \frac{d\mu_n}{d\nu_n} , \;\; n \ge 0 .$$

Then $M_n, n \ge 0$, is an $\mathcal{F}_n$-adapted sequence of integrable random variables on $(\Omega, \mathcal{F}, \nu)$. Moreover, one has for $n \ge 0$ and $A \in \mathcal{F}_n$,

$$\begin{aligned}\int_{A\in\mathcal{F}_n} \underbrace{M_{n+1}}_{\mathcal{F}_{n+1}\text{-measurable}} d\nu &= \int_A M_{n+1}\, d\nu_{n+1} = \int_A \frac{d\mu_{n+1}}{d\nu_{n+1}}\, d\nu_{n+1} = \mu_{n+1}(A) \\ &= \mu_n(A) = \int_A \frac{d\mu_n}{d\nu_n}\, d\nu_n = \int_A M_n\, d\nu_n = \int_A M_n\, d\nu .\end{aligned}$$

This means that

(3.2.14) $$M_n \text{ is an } \mathcal{F}_n\text{-martingale (for } \nu) .$$

As a concrete example, consider: $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}([0,1))$,

$$\mathcal{F}_n = \sigma\Big(\Big[0, \frac{1}{2^n}\Big),\ \Big[\frac{1}{2^n}, \frac{2}{2^n}\Big), \dots, \Big[\frac{2^n-1}{2^n}, 1\Big)\Big), \ \ \nu = \text{Lebesgue-measure}\,,$$

and μ any probability measure on $(\Omega, \mathcal{F})$ (the condition $\mu_n \ll \nu_n$ is then automatically satisfied). ○

We now give a few direct consequences of Definitions (3.2.2) - (3.2.4):

Proposition 3.12. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $(\mathcal{F}_n)_{n\geq 0}$ a filtration. Also, let M_n, $n \geq 0$, be an $\mathcal{F}_n$-supermartingale. One has P-a.s.*

$$E\big[M_n|\mathcal{F}_m\big] \leq M_m, \ \ \textit{for}\ \ 0 \leq m \leq n\,. \tag{3.2.15}$$

Analogously, for an $\mathcal{F}_n$-submartingale M_n, one has P-a.s.

$$E\big[M_n|\mathcal{F}_m\big] \geq M_m, \ \ \textit{for}\ \ 0 \leq m \leq n\,, \tag{3.2.16}$$

and for an $\mathcal{F}_n$-martingale M_n, P-a.s.

$$E\big[M_n|\mathcal{F}_m\big] = M_m, \ \ 0 \leq m \leq n\,. \tag{3.2.17}$$

Proof. Let us first make the following simple remarks:

$$M_n \text{ is a submartingale} \Longleftrightarrow -M_n \text{ is a supermartingale}\,, \tag{3.2.18}$$

$$M_n \text{ is a martingale} \Longleftrightarrow M_n \text{ is a sub- and a supermartingale}\,. \tag{3.2.19}$$

□

Hence, we can obtain (3.2.16) and (3.2.17) as consequences of (3.2.15).

Proof of (3.2.15): We proceed by induction. For $n = m$, (3.2.15) holds. If we now assume that (3.2.15) holds for $n = m + k$, then one also has

$$\begin{aligned} P\text{-a.s.} \qquad E\big[M_{m+k+1}|\mathcal{F}_m\big] &\overset{(3.1.28)}{=} E\big[E[M_{m+k+1}|\mathcal{F}_{m+k}]|\mathcal{F}_m\big] \\ \leq \ E\big[M_{m+k}|\mathcal{F}_m\big] &\overset{\substack{\text{induction}\\ \text{hypothesis}}}{\leq} M_m\,, \end{aligned}$$

and our claim (3.2.15) follows. □

Proposition 3.13. *Let $M_n, n \geq 0$, be an $\mathcal{F}_n$-martingale, and φ a convex function with $E\big[|\varphi(M_n)|\big] < \infty$ for all $n \geq 0$. Then*

$$\varphi(M_n) \ \textit{is an}\ \mathcal{F}_n\textit{-submartingale}\,. \tag{3.2.20}$$

Proof. One has, for all $n \geq 0$,

$$\begin{aligned} P\text{-a.s.} \ \ E\big[\varphi(M_{n+1})\,|\,\mathcal{F}_n\big] &\overset{\substack{\text{Jensen's}\\ \text{inequality}}}{\geq} \varphi\big(E[M_{n+1}|\mathcal{F}_n]\big) \\ &= \ \varphi(M_n)\,. \end{aligned} \tag{3.2.21}$$

□

Special case:

$$\text{(3.2.22)} \qquad \begin{array}{l}\text{Consider } 1 \le p < \infty, \text{ and } M_n, n \ge 0, \text{ a martingale in } L^p, \\ \text{then } |M_n|^p, n \ge 0, \text{ is a submartingale}.\end{array}$$

A similar result holds if M_n is only a submartingale, with the further assumption that φ is non-decreasing.

Proposition 3.14. *Let $M_n, n \ge 0$, be a submartingale, and φ a convex non-decreasing function with $E\big[|\varphi(M_n)|\big] < \infty$, for all $n \ge 0$. Then*

$$\text{(3.2.23)} \qquad \varphi(M_n) \textit{ is an } \mathcal{F}_n\textit{-submartingale} .$$

Proof. same as in (3.2.21). □

Special case:

$$\text{(3.2.24)} \qquad \begin{array}{l}\text{Let } M_n \text{ be a submartingale}, \\ \text{then } M_n \vee b \overset{\text{def.}}{=} \max(M_n, b) \text{ is a submartingale},\end{array}$$

$$\text{(3.2.25)} \qquad \begin{array}{l}\text{Let } M_n \text{ be a supermartingale}, \\ \text{then } M_n \wedge a \overset{\text{def.}}{=} \min(M_n, a) \text{ is a supermartingale}.\end{array}$$

Remark 3.15. The assumption "φ non-decreasing" in the previous proposition is important, as the following example shows. Consider

$$\text{(3.2.26)} \qquad M_n = -\frac{1}{n},\ n \ge 1,\ M_0 = -1 \ \text{(deterministic variables!)}.$$

This is certainly a submartingale, but

$$\text{(3.2.27)} \qquad M_n^2 = \frac{1}{n^2},\ n \ge 1, \quad M_0^2 = 1 ,$$

is a supermartingale! □

3.3 Stopping times

We now discuss the notion of stopping times, which plays an important role when studying martingales and Markov chains. We have already seen an example of a stopping time in the proof of Kolmogorov's inequality (Section 1.4). In what follows, we consider a probability space $(\Omega, \mathcal{A}, P)$, equipped with a filtration $\mathcal{F}_n, n \ge 0$.

Definition 3.16. *A random variable $N : \Omega \longrightarrow \mathbb{N} \cup \{\infty\}$ is called an $\mathcal{F}_n$-stopping time if*

$$\text{(3.3.1)} \qquad \textit{for all } n < \infty, \quad \{N = n\} \in \mathcal{F}_n .$$

"The decision to stop at time n depends only on the information up to time n".

Example 3.17. **1)** A fixed time $N(\omega) \equiv n_0$ is a stopping time, since for each n, one has $\{N = n\} = \emptyset$ or Ω (and these two sets belong to $\mathcal{F}_n$, since it is a σ-algebra).

For the next examples, we consider a sequence X_n, $n \geq 0$, of $\mathcal{F}_n$-adapted random variables.

2) Let us consider the "first time at which X_n becomes strictly positive" (i.e. first visit to $(0, +\infty)$),

$$N(\omega) \overset{\text{def.}}{=} \inf\{n \geq 0 :\ X_n(\omega) > 0\} \tag{3.3.2}$$

(defined as $+\infty$ if: $\forall n \geq 0$, $X_n(\omega) \leq 0$). This is a stopping time, since

$$\{N = 0\} = \{X_0 > 0\} \in \mathcal{F}_0, \quad \text{and for } n \geq 1,\ n < +\infty,$$

$$\{N = n\} = \{X_0 \leq 0,\ X_1 \leq 0,\ X_2 \leq 0, \dots, X_{n-1} \leq 0,\ X_n > 0\} \in \mathcal{F}_n\ .$$

Generalization: For $A \in \mathcal{B}(\mathbb{R})$, one can define the first visit time of X_n to A as

$$T_A(\omega) \overset{\text{def.}}{=} \inf\{n \geq 0 : X_n(\omega) \in A\} \tag{3.3.3}$$

(again, $= +\infty$ if: $\forall n \geq 0$, $X_n(\omega) \notin A$). This is a stopping time, for analogous reasons.

3) The first visit time of X_n to $B(\in \mathcal{B}(\mathbb{R}))$ strictly after visiting $A(\in \mathcal{B}(\mathbb{R}))$,

$$T(\omega) \overset{\text{def.}}{=} \inf\{n > T_A(\omega) : X_n(\omega) \in B\}\,, \tag{3.3.4}$$

defined as $+\infty$ if $T_A(\omega) = +\infty$, or if $T_A(\omega) < +\infty$ and $X_n(\omega) \notin B$ for all $n > T_A(\omega)$, is a stopping time since

$$\{T = 0\} = \emptyset \in \mathcal{F}_0, \quad \text{and for all } n \geq 1,$$

$$\{T = n\} = \bigcup_{m=0}^{n-1} \Big(\underbrace{\{T_A = m\}}_{\in \mathcal{F}_m \subseteq \mathcal{F}_n} \cap \underbrace{\{X_{m+1} \notin B, \dots, X_{n-1} \notin B, X_n \in B\}}_{\in \mathcal{F}_n} \Big) \in \mathcal{F}_n\ .$$

4) Consider $\Omega = \mathbb{R}^{\mathbb{N}_*}$, $\mathcal{A} = \bigotimes_{i \geq 1} \mathcal{B}(\mathbb{R})$, P arbitrary, $X_i(\omega) = \omega_i$, and the "last time at which X_n is non-negative" (i.e. last visit of X_n to $[0, +\infty)$),

$$S(\omega) \overset{\text{def.}}{=} \sup\{n \geq 1 : X_n(\omega) \geq 0\}\,, \tag{3.3.5}$$

defined as 0 if $X_n(\omega) < 0$ for all n. This is not a stopping time: indeed,

$$\{S = 1\} = \{X_1 \geq 0\} \cap \bigcap_{i \geq 2} \{X_i < 0\} \notin \mathcal{F}_1 = \sigma(X_1)\ .$$

In this setting, one can define the so-called σ-algebra of the stopping time N, denoted by $\mathcal{F}_N$, that contains all information known up to time $N(\omega)$ (which is random):

$$\mathcal{F}_N \overset{\text{def.}}{=} \{A \in \mathcal{A}\ :\ \forall n \in \mathbb{N},\ A \cap \{N = n\} \in \mathcal{F}_n\}\ . \tag{3.3.6}$$

On $\{N = n\}$, $A \in \mathcal{F}_N$ depends only on $X_1, \dots, X_n$, i.e. it depends only on the information available at time n. $\bigcirc$

Remark 3.18. For a fixed stopping time $N \equiv n_0 \in \mathbb{N}$, one has $\mathcal{F}_N = \mathcal{F}_{n_0}$ (clear from (3.3.6)). $\square$

3.4 Convergence theorem

We start with a characterization of the submartingale property.

Proposition 3.19. *(Doob's Decomposition)*

$X_n, n \geq 0$, is an $\mathcal{F}_n$-submartingale $\Longleftrightarrow$ $X_n, n \geq 0$, is of the form:

$$X_n = M_n + A_n,\ n \geq 0, \tag{3.4.1}$$

$$\text{where } M_n \text{ is an } \mathcal{F}_n\text{-martingale,} \tag{3.4.2}$$

$$\begin{aligned} &A_n \text{ is } \mathcal{F}_{n-1}\text{-measurable and integrable for all } n \geq 1, \\ &\text{and } 0 = A_0 \leq A_1 \leq A_2 \leq \dots\ P\text{-a.s.} \end{aligned} \tag{3.4.3}$$

Moreover, M_n and A_n are uniquely determined by (3.4.1), (3.4.2), and (3.4.3) *(up to sets of P-measure zero).*

Proof.

a) $\Longrightarrow$ *Uniqueness:*
For $n \geq 0$, one has necessarily (P-a.s.):

$$\begin{array}{ccccc} E\big[X_{n+1} - X_n|\mathcal{F}_n\big] & = & E\big[M_{n+1} - M_n|\mathcal{F}_n\big] & + & E\big[A_{n+1} - A_n|\mathcal{F}_n\big] \\ \| & & \| & & \| \ (3.2.8) \\ E\big[X_{n+1}|\mathcal{F}_n\big] - X_n & & 0 & & A_{n+1} - A_n\ . \end{array}$$

Hence,

$$A_0 = 0 \ \text{ and } \ A_{n+1} - A_n = E[X_{n+1}|\mathcal{F}_n] - X_n \ \text{ for } \ n \geq 0\ , \tag{3.4.4}$$

and $M_n = X_n - A_n \Longrightarrow M_n$ and A_n are uniquely determined.

Existence:

With the help of (3.4.4), we define A_n, $n \ge 0$, by $A_0 = 0$ and for $n \ge 0$, $A_{n+1} - A_n = E[X_{n+1}|\mathcal{F}_n] - X_n \ge 0$, P-a.s. (since X_n is a submartingale), and $M_n = X_n - A_n$. Then (3.4.1) and (3.4.3) are satisfied. Furthermore, one also has, for $n \ge 0$,

$$\begin{aligned} E\big[M_{n+1} - M_n|\mathcal{F}_n\big] &= E\big[X_{n+1} - X_n - (A_{n+1} - A_n)|\mathcal{F}_n\big] \\ &\overset{(3.4.3)}{=} E\big[X_{n+1}|\mathcal{F}_n\big] - X_n - (A_{n+1} - A_n) = 0 \end{aligned}$$

so that M_n, $n \ge 0$, is an $\mathcal{F}_n$-martingale.

b) $\Longleftarrow$ Each $X_n = M_n + A_n$ for which (3.4.2), (3.4.3) are satisfied is, of course, an $\mathcal{F}_n$-submartingale. □

In Doob's decomposition, the non-decreasing sequence $(A_n)_{n \ge 0}$ satisfies the condition

$$A_n \text{ is } \mathcal{F}_{n-1}\text{-measurable (and } A_0 = 0) \,.$$

This brings us to the following definition.

Definition 3.20. *Let $(\Omega, \mathcal{F})$ be a measurable space, and $\mathcal{F}_n, n \ge 0$, a filtration. A sequence $H_n, n \ge 1$, of random variables is said to be* ***predictable*** *if*

$$\forall n \ge 1,\ H_n \text{ is } \mathcal{F}_{n-1}\text{-measurable} \,. \tag{3.4.5}$$

Example 3.21. Let T: $\Omega \to \mathbb{N} \cup \{\infty\}$ be an $\mathcal{F}_n$-stopping time.

$$\begin{aligned} &H_n \overset{\text{def.}}{=} 1\{T \ge n\},\ n \ge 1, \ \text{ is predictable, since:} \\ &\{T \ge n\}^c = \{T \le n-1\} \in \mathcal{F}_{n-1} \,. \end{aligned} \tag{3.4.6}$$

○

Let $M_n, n \ge 0$, be an $\mathcal{F}_n$-sub- (resp. super-) martingale, and H_n, $n \ge 1$, a predictable sequence. Let us introduce the new sequence

$$\begin{aligned} (H \cdot M)_n &= \sum_{m=1}^{n} H_m(M_m - M_{m-1}) \qquad (n \ge 1)\,, \\ &= 0 \qquad (n = 0)\,. \end{aligned} \tag{3.4.7}$$

$(H \cdot M)_n$, $n \ge 0$, is a discrete version of what is called a stochastic integral "$\int_0^t H_s\, dM_s$".

First interpretation of (3.4.7) as "gambling winnings"

We consider a coin flipping game:

$$\begin{aligned} X_n &= 1 \quad \text{if the coin comes up Tails on the } n\text{th throw } (n \ge 1)\,, \\ &= -1 \quad \text{if the coin comes up Heads on the } n\text{th throw } (n \ge 1)\,. \end{aligned}$$

In the nth period of time, the gambler plays with a stake of H_n SFr. If the coin shows Tails, then he wins H_n SFr., and if the coin shows Heads, then he loses H_n SFr.

Let us consider the simple random walk with increments X_i, $i \geq 1$:

$$M_n = X_1 + \cdots + X_n \qquad (n \geq 1)\,, \\ = 0 \qquad (n = 0)\,.$$

The information that is available to the player at time n (≥ 1) (right before playing) is described through the σ-algebra

$$\mathcal{F}_{n-1} = \sigma(X_1, \dots, X_{n-1})\,.$$

The hypothesis that H_n, $n \geq 1$, is a predictable sequence (H_n is a "strategy") is then a natural assumption: at each time, the player can only use the information that is available to him in order to decide his next move.

The winnings of the player at time n are given by

$$\begin{aligned} W_n &= H_1\,X_1 + H_2\,X_2 + \cdots + H_n\,X_n = \sum_{m=1}^{n} H_m(M_m - M_{m-1}) \\ &= (H \cdot M)_n\,. \end{aligned} \tag{3.4.8}$$

Second interpretation, in terms of financial mathematics

Assume M_n is the value of a US dollar in Swiss francs on day n (we assume that $M_n \geq 0$ holds). On day n, an investor possesses F_n Swiss francs and H_n dollars. On day $n = 0$, his entire fortune is invested in Swiss francs, i.e. $H_0 = 0$.

His fortune on day n is then (in Swiss francs)

$$V_n = F_n + H_n\,M_n\,. \tag{3.4.9}$$

At the end of each day, the investor chooses a new distribution for his fortune V_n (in SFr. and dollars): he decides, at the end of day n, to have H_{n+1} as a new account balance in dollars, having thus F_{n+1}, H_{n+1} as new balances in SFr and USD at the end of day n. One has:

$$V_n = F_n + H_n \cdot M_n = F_{n+1} + H_{n+1} \cdot M_n \tag{3.4.10}$$

(here, $F_n \leq 0$ or $H_n \leq 0$ are possible, and correspond to a loan on the associated account).

The choice of H_{n+1} as a new balance in dollars is made with the help of the information that is available to the investor at the end of day n. Hence, we assume that

$$H_n,\ n \geq 1, \text{ is predictable}\,. \tag{3.4.11}$$

Note also that (3.4.9), (3.4.10) imply, for $n \geq 0$,

$$\begin{aligned} V_{n+1} - V_n &= H_{n+1}\,M_{n+1} + F_{n+1} - H_{n+1}\,M_n - F_{n+1} \\ &= H_{n+1}(M_{n+1} - M_n)\,. \end{aligned}$$

Hence,

$$
\begin{aligned}
V_n &= (V_n - V_{n-1}) + (V_{n-1} - V_{n-2}) + \cdots + (V_1 - V_0) + V_0 \\
&= V_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1}) \\
&= V_0 + (H \cdot M)_n \, .
\end{aligned}
\tag{3.4.12}
$$

In other words, $V_0 + (H \cdot M)_n$ is the **fortune of the investor on day** n (in SFr). $H_n, n \geq 1$, is then his investment strategy (in dollars).

A usual restriction on valid investment strategies is the following: only predictable $H_n, n \geq 1$, for which

$$
V_0 + (H \cdot M)_n \geq 0 \qquad \text{for all } n \geq 0 \tag{3.4.13}
$$

are allowed.

Theorem 3.22. *Let $X_n, n \geq 0$, be a super- (resp. sub-) martingale, and $H_n \geq 0$, $n \geq 1$, a predictable sequence of random variables with H_n bounded for each $n \geq 1$. Then*

$$
(H \cdot X)_n \text{ is a super- (resp. sub-) martingale} \, . \tag{3.4.14}
$$

Proof. We assume that X_n is a supermartingale (the case of a submartingale is handled in an analogous way). For $n \geq 0$,

$$
\begin{aligned}
E\big[(H \cdot X)_{n+1} - (H \cdot X)_n \,|\, \mathcal{F}_n\big] \;&=\; E\big[\underbrace{H_{n+1}}_{\mathcal{F}_n\text{-measurable}} (X_{n+1} - X_n) \,|\, \mathcal{F}_n\big] \\
=\; \underset{\geq 0}{H_{n+1}} \;&\underbrace{E\big[X_{n+1} - X_n |\, \mathcal{F}_n\big]}_{\leq 0 \;\; (X_n \text{ is a supermartingale})} \;\leq\; 0 \, .
\end{aligned}
$$

Consequently, $(H \cdot X)_n$, $n \geq 0$, is a supermartingale. □

Corollary 3.23. *Let $X_n, n \geq 0$, be a martingale, and $H_n, n \geq 1$, predictable, with H_n bounded for all n. Then $(H \cdot X)_n$ is a martingale.*

Corollary 3.24. *(Optional Stopping Theorem (first version))*

Let $X_n, n \geq 0$, be an $\mathcal{F}_n$- (resp. super-, resp. sub-) martingale, and N an $\mathcal{F}_n$-stopping time. Then

$$
X_{N \wedge n} \text{ is a (resp. super-, resp. sub-) martingale} \, . \tag{3.4.15}
$$

Proof. We consider only the case when X_n is a supermartingale. Set $H_n = 1\{N \geq n\}$, $n \geq 1$. Then, for $n \geq 1$,

$$
\begin{aligned}
&(H \cdot X)_n = \sum_{m=1}^{n} 1\{N \geq m\}(X_m - X_{m-1}) = X_{N \wedge n} - X_0 \\
&M_n \equiv X_0, \; \forall n \geq 0, \text{ is, of course, a martingale and} \\
&X_{N \wedge n} = (H \cdot X)_n + M_n \text{ is a supermartingale, thanks to (3.4.14)} \, .
\end{aligned}
\tag{3.4.16}
$$

□

The next result gives a useful estimate for the possible fluctuations of a super- (resp. sub-) martingale during a time interval $[0, u]$. We first need some notation.

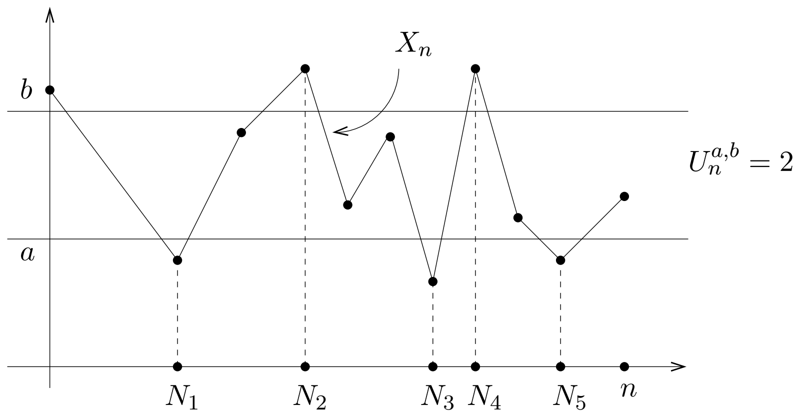


Fig. 3.3

Consider $a < b$ in $\mathbb{R}$, and X_n, $n \geq 0$, an $\mathcal{F}_n$-submartingale. We define an increasing sequence of $\mathcal{F}_n$-stopping times $N_j = N_j^{X,[a,b]}$ by

(3.4.17)
$$
\begin{aligned}
N_1 &= \inf\{m \geq 0,\ X_m \leq a\} \\
&\quad (N_1(\omega) = \infty \quad \text{if } \{\dots\} = \emptyset) \\[1em]
N_2 &= \inf\{m > N_1,\ X_m \geq b\} \\
&\quad (N_2 = \infty \ \text{ if } N_1 = \infty \ \text{ or } \ \{m > N_1(\omega), X_m(\omega) \geq b\} = \emptyset) \\[1em]
\underset{(k \geq 2)}{N_{2k-1}} &= \inf\{m > N_{2k-2},\ X_m \leq a\} \\
&\quad (\text{where } N_{2k-1}(\omega) = \infty \text{ if } N_{2k-2}(\omega) = \infty \text{ or } \{\dots\} = \emptyset) \\[1em]
N_{2k} &= \inf\{m > N_{2k-1},\ X_m \geq b\} \\
&\quad (\text{where } N_{2k}(\omega) = \infty \text{ if } N_{2k-1} = \infty \text{ or } \{\dots\} = \emptyset)
\end{aligned}
$$

The $N_j, j \geq 1$, are $\mathcal{F}_n$-stopping times (see Example 3 above, (3.3.4)).

We define the number of upward crossings ("upcrossings") of $[a, b]$ during the time interval $[0, n]$ by

(3.4.18)
$$U_n^{a,b}\,(\omega) = \sup\{k \geq 1,\ N_{2k}(\omega) \leq n\} \ \ (= 0 \ \text{ if } \ \{\dots\} = \emptyset)\,.$$

Proposition 3.25. *(Upcrossing Inequality)*

Consider a submartingale $X_n, n \geq 0$. Then for all $a < b$ in $\mathbb{R}$,

(3.4.19)
$$(b - a)\ E\big[U_n^{a,b}\big] \leq E\big[(X_n - a)^+\big] - E\big[(X_0 - a)^+\big]\,.$$

Proof. We set $Y_n = (X_n - a)^+$. Thanks to (3.2.24), Y_n stays a submartingale. One has of course

$$\begin{aligned} & N_j^{X,[a,b]} = N_j^{Y,[0,b-a]}\,, \quad j \ge 1\,, \\ \text{and} \qquad & U_n^{a,b,X} = U_n^{0,b-a,Y}\,. \end{aligned}$$

In other words, we can consider the upward crossings of $[0, b-a]$ by the random variables $Y_n, n \ge 0$.

We consider the following sequence of random variables

$$H_m = \begin{cases} 1 & \text{on } \bigcup_{k \ge 1} \{N_{2k-1} < m \le N_{2k}\}\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{3.4.20}$$

Note that

$$\{N_{2k-1} < m \le N_{2k}\} \;=\; \{\overset{\text{stopping time}}{N_{2k-1}} \le m-1\} \;\cap\; \{\underset{\text{stopping time}}{N_{2k}} \le m-1\}^c \in \mathcal{F}_{m-1}.$$

Hence, H_m is a predictable sequence (the interpretation of H as an investment strategy is the following: one buys 1 USD when the value of a dollar falls below a, and one sells this dollar when its value goes above b). One has

$$(b-a)\, U_n \;\le\; \overbrace{(H \cdot Y)_n}^{\text{profit at time } n} \Big(= \sum_{m=1}^n H_m(Y_m - Y_{m-1})\Big)\,, \tag{3.4.21}$$

since with each upward crossing of $[0, b-a]$, one makes a profit $\ge (b-a)$, and there may also be an incomplete final upcrossing, which provides a positive contribution (since $Y_n \ge 0$). One has

$$\begin{aligned} Y_n \quad &= \quad Y_0 + ((1-H)\cdot Y)_n + (H \cdot Y)_n \\ &\overset{(3.4.21)}{\ge} \quad Y_0 + ((1-H)\cdot Y)_n + (b-a)\, U_n\,. \end{aligned}$$

Using (3.4.14), we obtain

$$E\big[((1-H)\cdot Y)_n\big] \ge 0\,. \tag{3.4.22}$$

Hence,

$$(b-a)\, E[U_n] \le E[Y_n] - E[Y_0]\,.$$

This proves (3.4.19). □

Remark 3.26. It is remarkable that for the opposite investment strategy $K_n = 1 - H_n$, one also has $E[(K \cdot Y)_n]$ and $E[(K \cdot X)_n] \ge 0$. In this case, one keeps 1 USD until the price falls below level a, one then sells this dollar, waits until it becomes more expensive than b, and buys the dollar back, and so on. On average, one still makes a profit, if "the price of a USD is a submartingale". □

Application of the upcrossing inequality:

Theorem 3.27. *(Martingale Convergence Theorem)*

Let $X_n, n \ge 0$, be a submartingale with $\sup_{n\ge 0} E[X_n^+] < \infty$, then

$$\begin{aligned} &\textit{the sequence } X_n(\omega) \textit{ converges } P\textit{-a.s.,} \\ &\textit{to some } X(\omega) \textit{ with } E[|X|] < \infty \ . \end{aligned} \tag{3.4.23}$$

Proof. Consider $a < b$ in $\mathbb{R}$. Due to (3.4.19), one has, for $n \ge 0$,

$$E[U_n^{a,b}] \le \frac{E[(X_n - a)_+]}{b-a} \le \frac{E[X_n^+] + |a|}{b-a} \le \text{cst } < \infty \ . \tag{3.4.24}$$

If we define, analogously to (3.4.18), the total number of upward crossings of $[a,b]$ by the random variables X_n as

$$U_\infty^{a,b}(\omega) = \sup\{k \ge 1, \ N_{2k}(\omega) < \infty\}, \tag{3.4.25}$$

then one has $U_\infty^{a,b}(\omega) = \lim\limits_n \uparrow \ U_n^{a,b}(\omega)$. Using (3.4.24), we obtain

$$E[U_\infty^{a,b}] \overset{\substack{\text{monotone} \\ \text{convergence}}}{=} \lim_n \uparrow \ E[U_n^{a,b}] \le \text{cst}\,.$$

Hence, $P\big[\{U_\infty^{a,b} < \infty\}\big] = 1$, and

$$P\Big[\bigcap_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{U_\infty^{a,b} < \infty\}\Big] = 1 \ . \tag{3.4.26}$$

Note that

$$\big\{\underline{\lim}\, X_n < a < b < \overline{\lim}\, X_n\big\} \subseteq \{U_\infty^{a,b} = \infty\},$$

and that

$$\big\{\underline{\lim}\, X_n < \overline{\lim}\, X_n\big\} = \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \big\{\underline{\lim}\, X_n < a < b < \overline{\lim}\, X_n\big\} \subseteq \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{U_\infty^{a,b} = \infty\}$$

has probability 0, thanks to (3.4.26). Hence, $X_n(\omega)$ converges P-a.s. to some $X(\omega)$. Furthermore,

$$E[X_0] \le E[X_n] = E[X_n^+] - E[X_n^-]\,,$$

and thus

$$E[X_n^-] \le E[X_n^+] - E[X_0] \le \text{cst} \Longrightarrow \sup_n \ E\big[|X_n|\big] < \infty \ .$$

Fatou's lemma then implies

$$E\big[|X|\big] \le \underline{\lim}\, E\big[|X_n|\big] < \infty \ .$$

The convergence theorem thus follows. $\square$

Corollary 3.28. *Let $X_n \geq 0$ be a supermartingale, then*

(3.4.27) $\quad X_n$ *converges P-a.s. to some $X \geq 0$, with $E[X] \leq E[X_0]$.*

Proof. $Y_n = -X_n$ is a submartingale with $Y_n^+ = 0$. Using the convergence theorem, $X_n \to X$ P-a.s., and furthermore,

$$E[X] \underset{\text{Fatou}}{\leq} \underline{\lim}\, E[X_n] \underset{\text{supermartingale}}{\leq} E[X_0]\ .$$

$\square$

Remark 3.29. In general, the hypotheses of the convergence theorem are not enough to show that X_n converges to X in L^1. For instance, we have seen (Example 3 above) that $M_n = (\frac{1-p}{p})^{S_n}$ is a martingale when S_n is an asymmetric random walk on $\mathbb{Z}$ ($P[X_i = 1] = p$, $P[X_i = -1] = 1 - p$, $p \neq \frac{1}{2}$).

We have also seen that $M_n \to 0$ P-a.s. But one has $\|M_n\|_1 = E[M_n] = E[M_0] = 1$, so M_n does not converge to 0 in L^1. $\square$

3.5 Some examples and applications

In this section, we want to discuss a few examples and applications of the optional stopping theorem, and the convergence theorem.

A) Branching process: the Galton-Watson chain

The Galton-Watson chain is a model that describes the evolution of a population.

At time 0, there is one particle. This particle has then a certain number of descendants, the first generation. The distribution of this number of descendants is called ν (ν is a probability measure on $\mathbb{N}$). Each particle of the first generation has then, independently of the other ones, a number of descendants with distribution ν. These descendants form the second generation, and so on. If a generation happens to have no descendants, then the population dies out (the numbers of particles in the later generations are all equal to 0).

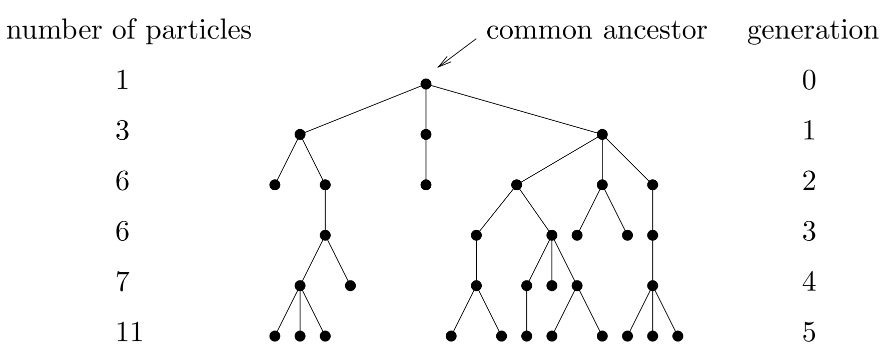


Fig. 3.4

A natural question is then: how does the number of particles in the nth generation behave asymptotically, for large n?

Mathematical construction of the model:

We consider independent ν-distributed random variables ξ_i^n, $i, n \ge 1$. We define the number of particles in the nth generation $Z_n, n \ge 0$, as:

$$\begin{aligned} & Z_0 = 1\,, \\ \text{and} \quad & Z_{n+1} = \xi_1^{n+1} + \xi_2^{n+1} + \dots + \xi_{Z_n}^{n+1} \quad \text{if} \;\; Z_n > 0\,, \\ & \phantom{Z_{n+1}} = 0 \quad \text{if} \;\; Z_n = 0\,. \end{aligned} \tag{3.5.1}$$

- $Z_n, n \ge 0$, is the so-called Galton-Watson chain (we will see later that $Z_n, n \ge 0$, is a Markov chain).

- ν is the distribution of the number of descendants for a single individual (sometimes called offspring distribution).

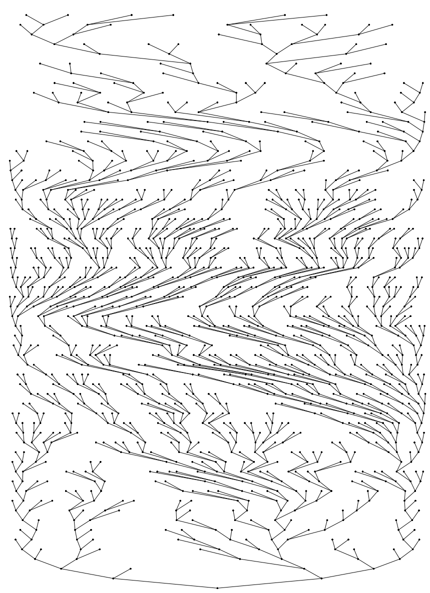


Fig. 3.5: An (atypical) simulation of a branching process (offspring distribution $\frac{1}{2}\,\delta_0 + \frac{1}{2}\,\delta_2$)

We assume that

$$\nu(0) \neq 1 \text{ and } \nu(1) \neq 1 \tag{3.5.2}$$

(otherwise, the model is trivial), and that the mean number of descendants of a particle satisfies

$$m = \sum_{k=0}^{\infty} k\,\nu(k) < \infty\,. \tag{3.5.3}$$

We define the filtration $\mathcal{F}_n, n \ge 0$, as

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \;\; \mathcal{F}_n = \sigma\left(\xi_i^m, \; 1 \le m \le n, \; i \ge 1\right). \tag{3.5.4}$$

Proposition 3.30.

$$M_n \overset{\text{def.}}{=} \frac{Z_n}{m^n} \;\; \textit{is an } \mathcal{F}_n\textit{-martingale}\,. \tag{3.5.5}$$

Proof. Thanks to (3.5.1), (3.5.3), Z_n is $\mathcal{F}_n$-measurable and integrable. One has

$$\begin{aligned} E\big[Z_{n+1}|\mathcal{F}_n\big] &= E\Big[\sum_{k=0}^{\infty} Z_{n+1}\, 1\{Z_n = k\}|\mathcal{F}_n\Big] \\ &\overset{\substack{\text{monotone}\\ \text{convergence}}}{=} \sum_{k=0}^{\infty} E\big[Z_{n+1}\, 1\{Z_n = k\}|\mathcal{F}_n\big] \end{aligned}$$

(here, we are actually using an analogue of the usual monotone convergence theorem, for conditional expectations: the proof is left as an exercise).
Using (3.5.1), we find

$$\begin{aligned} E\big[Z_{n+1}|\mathcal{F}_n\big] &= \sum_{k=1}^{\infty} E\big[(\xi_1^{n+1} + \dots + \xi_k^{n+1})\, 1\underbrace{\{Z_n = k\}}_{\mathcal{F}_n\text{-measurable}}|\mathcal{F}_n\big] \\ &= \sum_{k=1}^{\infty} 1\{Z_n = k\}\; E\big[\underbrace{(\xi_1^{n+1} + \dots + \xi_k^{n+1})}_{\text{independent of } \mathcal{F}_n}|\mathcal{F}_n\big] \\ &= \sum_{k=1}^{\infty} mk \cdot 1\{Z_n = k\} = mZ_n\,. \end{aligned} \tag{3.5.6}$$

Hence, M_n, $n \ge 0$, is a non-negative martingale. Thanks to (3.4.27), M_n then converges P-a.s. to $M_\infty \ge 0$, with $E[M_\infty] \le 1 = E[M_0]$.

Subcritical case: $0 < m < 1$

$$M_n = \frac{Z_n}{m^n} \overset{P\text{-a.s.}}{\longrightarrow} M_\infty \in [0,\infty)\,.$$

Note that $Z_n > 0$ implies that $M_n \ge \frac{1}{m^n}$, and $\frac{1}{m^n} \uparrow +\infty$ as $n \to \infty$, so

$$\text{in the case } 0 < m < 1,\; Z_n = 0 \text{ for } n \text{ large enough, } P\text{-a.s.} \tag{3.5.7}$$

Critical case: $m = 1$

$$\mathbb{N} \ni Z_n(\omega) = M_n(\omega) \overset{P\text{-a.s.}}{\longrightarrow} M_\infty \in [0,\infty)\,.$$

In other words,

$$Z_n(\omega) = M_\infty(\omega) \in \mathbb{N} \text{ for } n \text{ large enough, } P\text{-a.s.} \tag{3.5.8}$$

One also has, for $k \geq 1$, $n_0 \geq 0$,

$$
\begin{aligned}
P\Big[\bigcap_{n \geq n_0} \{Z_n = k\}\Big] &= P\big[\{Z_{n_0} = k\} \cap \bigcap_{n > n_0} \{\xi_1^n + \cdots + \xi_k^n = k\}\big] \\
&\overset{\text{independence}}{=} P[Z_{n_0} = k] \cdot \prod_{n > n_0} \underbrace{P[\xi_1^n + \cdots + \xi_k^n = k]}_{<\,1 \text{ thanks to } (3.5.2)} = 0
\end{aligned} \tag{3.5.9}
$$

(because of (3.5.2) and $m = 1$, the distribution ν cannot be concentrated on a single value, and $P[\xi_1^n + \cdots + \xi_k^n = k] = 1$ is impossible). Thanks to (3.5.8), (3.5.9), one has $P[M_\infty \geq 1] = 0$, and thus

(3.5.10) in the case $m = 1$, $Z_n = 0$ for n large enough, P-a.s.

Supercritical case: $m > 1$

Consider the "generating function"

$$
\varphi(s) \overset{\text{def.}}{=} \sum_{k=0}^{\infty} s^k \, \nu(k) \; \Big(= \int_{\mathbb{N}} s^x \, d\nu(x) \Big), \quad s \in [0,1] \,. \tag{3.5.11}
$$

The function φ is continuous, non-decreasing on $[0,1]$, and one has, for $s \in [0,1)$,

$$
\varphi'(s) = \sum_{k=1}^{\infty} k s^{k-1} \nu(k), \quad \varphi''(s) = \sum_{k=2}^{\infty} k(k-1) \, s^{k-2} \nu(k) \geq 0 \,.
$$

Hence, $\lim_{s \to 1} \varphi'(s) = m > 1$ (and $\varphi'(0) = \nu(1)$).

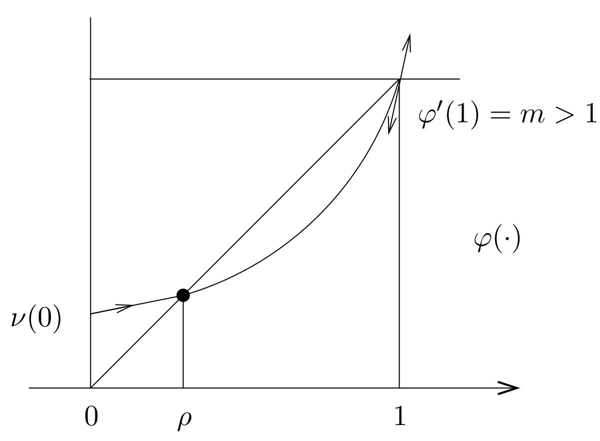


Fig. 3.6

Furthermore, $\varphi'(1) = m > 1$ and $\varphi(1) = 1 \Longrightarrow \varphi(1-\epsilon) < 1 - \epsilon$ for $\epsilon > 0$ small. Moreover, $\varphi(0) = \nu(0) \geq 0$, so there exists a $\rho \in [0,1)$ with $\rho = \varphi(\rho)$, and ρ is unique, since φ is strictly convex ($\varphi''(s) > 0$ for $s \in (0,1)$).

Let us set $\theta_n = P[Z_n = 0]$, $n \geq 0$. Obviously, $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$, and θ_n is non-decreasing in n. Intuitively speaking, it is clear that **conditionally on $\boldsymbol{Z_1 = k \geq 1}$, $\boldsymbol{Z_{n+1}}$ is distributed as the sum of $\boldsymbol{k}$ independent copies of $\boldsymbol{Z_n}$**. This claim follows by induction, using (3.5.1). Consequently,

$$\begin{aligned} \theta_{n+1} &= E\big[E\big[1\{Z_{n+1} = 0\}|Z_1\big]\big] = \nu(0) + \sum_{k=1}^{\infty} \nu(k)\, \theta_n^k \\ &= \varphi(\theta_n),\ n \geq 0\ . \end{aligned} \tag{3.5.12}$$

Hence, $\theta_0 = 0 \leq \theta_1 = \nu(0) \leq \rho \overset{\text{induction}}{\Longrightarrow} \theta_n \uparrow$ and $\theta_n \leq \rho$.

We deduce convergence: $\theta_m \to \theta_\infty = P\big[\bigcup_{n \geq 1} \{Z_n = 0\}\big] \leq \rho$, where θ_∞ satisfies $\theta_\infty = \varphi(\theta_\infty)$. It follows that

$$P\Big[\bigcup_{n \geq 1} \{Z_n = 0\}\Big] = \rho \in [0,1)\ . \tag{3.5.13}$$

To summarize, with a positive probability $1-\rho$, the population of particles never dies out, and we also know that $\frac{Z_n}{m^n} \to M_\infty$ P-a.s.

Let us finally mention that Kesten and Stigum have shown that

$$\int_{\mathbb{N}} x(\log x)_+\, \nu(dx) < \infty \Longrightarrow P[M_\infty > 0] = 1 - \rho > 0\ .$$

B) Asymmetric random walk

We keep the notations of Example 3 in Section 3.2: we denote the random walk by $S_n, n \geq 0$ (with $S_0 = 0$), and we suppose that

$$p = P[X_i = 1] = 1 - P[X_i = -1] \in \Big(\frac{1}{2}, 1\Big)\ .$$

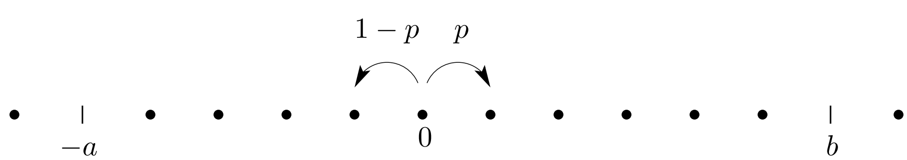


Fig. 3.7

Consider $a, b \geq 1$, and

$$T = \inf\{n \geq 0,\ S_n = -a \ \text{ or } \ b\} \tag{3.5.14}$$

the exit time of S_n from $(-a, b)$.

We know from (3.2.9) - (3.2.10) that $h(S_n)$ is a martingale, where

$$h(x) = \Big(\frac{1-p}{p}\Big)^x,\ \ x \in \mathbb{Z}\,. \tag{3.5.15}$$

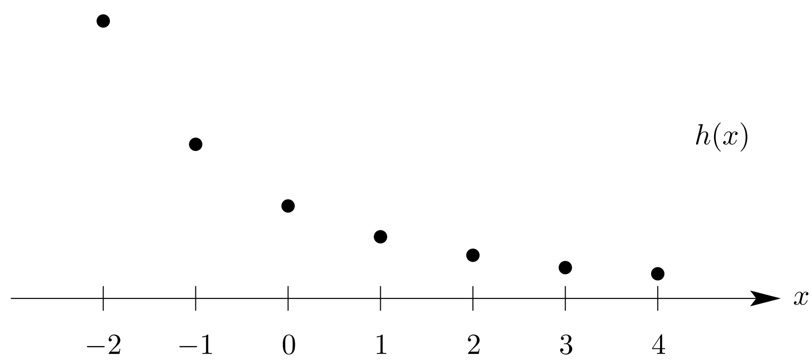


Fig. 3.8

The optional stopping theorem implies that:

$$1 = E\big[h(S_0)\big] = E\big[h(S_{T\wedge n})\big] \ \text{ and } \ 0 \le h(S_{T\wedge n}) \le h(-a)$$

$$\Big\downarrow \ n \to \infty \quad \text{(dominated convergence)}$$

$$E\big[h(S_T)\big] = h(-a)\,P[S_T = -a] + h(b)\,P[S_T = b]\ .$$

It follows that $1 = h(-a)\,P[S_T = -a] + h(b)(1 - P[S_T = -a])$, and thus

(3.5.16)
$$P[S_T = -a] = \frac{1 - h(b)}{h(-a) - h(b)} = \frac{1 - r^b}{r^{-a} - r^b}\ ,$$
$$P[S_T = b] = \frac{h(-a) - 1}{h(-a) - h(b)} = \frac{r^{-a} - 1}{r^{-a} - r^b}\ ,$$

with $r = \frac{1-p}{p} \in (0,1)$.

If we choose $a = 1$, and let $b \to \infty$, we obtain

(3.5.17)
$$P[T_{-1} < \infty] = \lim_{b\to\infty} \uparrow P[S_T = -1] = r = \frac{1-p}{p} < 1\ ,$$

where $T_{-1} = \inf\{n \ge 0,\ S_n = -1\}$.

C) Call options: the Cox Ross Rubinstein model

We keep the same notations as in Section 3.2: $M_n, n \ge 0$, describes the price of a risky asset on day n (for example, 1 USD in SFr). We assume that there exist $0 < a < b$ such that

(3.5.18)
$$\frac{M_{n+1}}{M_n} \overset{\text{def.}}{=} D_{n+1} \in \{a, b\},\ \ 0 \le n,\ \ n+1 \le N\,,$$

and $M_0 = 1$. Our space Ω is simply $\Omega = \{a,b\}^N$ $(N \ge 1)$.

A call option is a contract between A and B, that gives A the possibility to buy 1 USD from B at a stipulated price v on day N. Of course, in the case when $M_N < v$, A does not exercise his right to buy. In the case when $M_N > v$, A buys 1 USD from B at price v, and then he sells this USD at price M_N, making in this way a profit of $M_N - v$. In other words, the contract for A corresponds to a profit potential of $(M_N - v)_+$.

Question: what is the "fair price" for such a contract? How much should A pay B on day 0, so that the contract is fair? Before we handle this question, we will make a few preliminary remarks.

For an initial fortune $V_0 \in \mathbb{R}$, and a valid (or admissible) investment strategy H_n, $1 \le n \le N$, i.e. a predictable sequence with

$$V_0 + (H \cdot M)_n \ge 0, \quad 0 \le n \le N \ , \tag{3.5.19}$$

the fortune obtained on day N is equal to, thanks to (3.4.12):

$$V_N = V_0 + (H \cdot M)_N \ . \tag{3.5.20}$$

If $1 < a < b$, then an initial fortune $V_0 = 0$ allows one to obtain "without any risk" $V_N > 0$. Indeed, by (3.5.18), if we choose $H_n = 1$, $1 \le n \le N$, then $V_N = \sum_{n=1}^N \underbrace{(M_n - M_{n-1})}_{>0} > 0$ (and H_n is valid).

Analogously, if $a < b < 1$, then $H_n = -1$ is a valid investment strategy, and

$$V_n = \sum_{m=1}^{n} \underbrace{(M_{m-1} - M_m)}_{>0} > 0, \quad 1 \le n \le N \ .$$

In both cases, one can obtain safely a positive fortune on day N, without investing any capital. One speaks here of **arbitrage**.

On the contrary, if $a < 1 < b$, a real risk subsists (theoretically). We can define a (purely artificial) probability measure Q on Ω, such that under Q, M_n, $0 \le n \le N$, is a martingale: we simply choose D_n, $1 \le n \le N$, independent with distribution ν, where

$$\nu = \frac{b-1}{b-a}\,\delta_a + \frac{1-a}{b-a}\,\delta_b \qquad \Big(\text{it satisfies the condition} \int x\, d\nu(x) = 1\Big) \ . \tag{3.5.21}$$

In principle, this probability measure Q has nothing to do with the description of the real statistical properties of M_n. It is a "purely mathematical construction".

In parallel, we have a description of M_n, $0 \le n \le N$: we write P for the corresponding probability measure (one should be careful not to confuse P and Q).

Due to (3.4.14), $(H \cdot M)_n$ is a Q-martingale for each investment strategy H_n, $1 \le n \le N$. It follows that

$$V_0 = E^Q\big[V_0 + (H \cdot M)_N\big] = E^Q[V_N] \ .$$

Hence, if we find (V_0, H) such that $V_N = (M_N - v)_+$, then $V_0 = E^Q[(M_N - v)_+]$ as well. In other words: in order to obtain $(M_N - v)_+$ as a capital on day N through (V_0, H), the initial fortune must be equal to $E^Q[(M_N - v)_+]$.

Proposition 3.31. *There exists a unique valid investment strategy H with initial fortune $V_0 = E^Q[(M_N - v)_+]$ such that*

$$E^Q\big[(M_N - v)_+\big] + (H \cdot M)_N = (M_N - v)_+ \ . \tag{3.5.22}$$

Proof.

Uniqueness: Define the martingale

$$Z_n = E^Q\big[(M_N - v)_+ | \mathcal{F}_n\big], \ \ 0 \le n \le N \ . \tag{3.5.23}$$

Then (using $D_n = M_n/M_{n-1}$),

$$Z_n = E^Q\left[\left(M_n \cdot \prod_{k=n+1}^{N} D_k - v\right)_+ | \mathcal{F}_n\right] .$$

Now, $\prod_{k=n+1}^{N} D_k$ is independent of $\mathcal{F}_n$, and one has

$$Z_n = c(n, M_n) \ , \tag{3.5.24}$$

where

$$\begin{aligned} c(n,x) &= E^Q\left[\left(x \prod_{k=n+1}^{N} D_k - v\right)_+\right] \\ &= \sum_{j=0}^{N-n} \binom{N-n}{j} \nu(a)^j \, \nu(b)^{N-n-j} (x a^j \, b^{N-n-j} - v)_+ \ . \end{aligned} \tag{3.5.25}$$

Using (3.5.24), we obtain:

$$\begin{aligned} H_n(M_n - M_{n-1}) &= Z_n - Z_{n-1} = c(n, M_n) - c(n-1, M_{n-1}) \\ \Longrightarrow H_n \, M_{n-1}(D_n - 1) &= c(n, M_{n-1} \, D_n) - c(n-1, M_{n-1}) \ . \end{aligned}$$

Since H_n and M_{n-1} are $\mathcal{F}_{n-1}$-measurable, and do not depend on D_n,

$$\begin{cases} H_n \, M_{n-1}(b-1) = c(n, M_{n-1} \, b) - c(n-1, M_{n-1}) \\ H_n \, M_{n-1}(a-1) = c(n, M_{n-1} \, a) - c(n-1, M_{n-1}) \end{cases} \tag{3.5.26}$$

and we obtain, by subtraction,

$$H_n = \frac{c(n, M_{n-1} \, b) - c(n, M_{n-1} \, a)}{(b-a) M_{n-1}}, \ \ 1 \le n \le N \ . \tag{3.5.27}$$

This proves uniqueness.

Existence:

We define the predictable sequence H_n, $1 \le n \le N$, via (3.5.27). Then, for $n \in \{1, \dots, N\}$,

$$
\begin{aligned}
(b-1)H_n\, M_{n-1} &= \frac{(b-1)}{(b-a)}\,\big(c(n, M_{n-1}\,b) - c(n, M_{n-1}\,a)\big) \\
&= \nu(a)\big(c(n, M_{n-1}\,b) - c(n, M_{n-1}\,a)\big) \\
&= c(n, M_{n-1}\,b) - \nu(b)\,c(n, M_{n-1}\,b) - \nu(a)\,c(n, M_{n-1}\,a) \\
&= c(n, M_{n-1}\,b) - E^Q\big[\underbrace{c(n, M_n)}_{\overset{\|}{Z_n}}\,|\mathcal{F}_{n-1}\big] \\
&= c(n, M_{n-1}\,b) - Z_{n-1} \\
&= c(n, M_{n-1}\,b) - c(n-1, M_{n-1})\,.
\end{aligned}
$$

Analogously,

$$
\begin{aligned}
(a-1)H_n\, M_{n-1} &= \frac{(a-1)}{(b-a)}\,\big(c(n, M_{n-1}\,b) - c(n, M_{n-1}\,a)\big) \\
&= c(n, M_{n-1}\,a) - \nu(b)\,c(n, M_{n-1}\,b) - \nu(a)\,c(n, M_{n-1}\,a) \\
&= c(n, M_{n-1}\,a) - c(n-1, M_{n-1})\,.
\end{aligned}
$$

In other words, we obtain that (3.5.26) holds, so

$$
\begin{aligned}
&H_n(M_n - M_{n-1}) = Z_n - Z_{n-1}\,, \\
\text{and}\quad &V_0 + (H\cdot M)_n = Z_n \ge 0,\ 0 \le n \le N, \ \text{ since }\ V_0 = Z_0\,.
\end{aligned}
$$

Hence, H is valid, and (3.5.22) holds. □

Before we come back to the question of a "fair price" for the call option, it is of interest to describe explicitly the investment strategy determined in (3.5.22) - (3.5.27) in the special case $N = 1$.

Special case $N = 1$:

We have $c(1, x) = (x - v)_+$, and thanks to (3.5.27) and $M_0 = 1$,

$$
\begin{aligned}
H_1 &= \big((b-v)_+ - (a-v)_+\big)/(b-a)\,, \\
V_0 &= \big((1-a)(b-v)_+ + (b-1)(a-v)_+\big)/(b-a)\,.
\end{aligned}
\tag{3.5.28}
$$

- In the case $v > b$, the profit potential of the call option is $(M_1 - v)_+ = 0$ (the right to buy is never exercised!), and $H_1 = V_0 = 0$.

- In the case $a > v$, the profit potential of the call option is $(M_1 - v)_+ = M_1 - v$ (the right to buy is always exercised), $V_0 = 1 - v$, $H_1 = 1$ (one buys 1 USD "at the end of day 0").

- In the case $a < v < b$, the right to buy is exercised only part of the time. One has

$$V_0 = \frac{1-a}{b-a} \cdot (b-v), \quad H_1 = \frac{b-v}{b-a} \ (< 1)\ ,$$

and (3.5.22) is simply

$$(M_1 - v)_+ = (1-a)\,\frac{b-v}{b-a} + \frac{b-v}{b-a}\,(M_1 - \underset{\substack{\parallel \\ M_0}}{1}\,)\,.$$

□

Interpretation of the proposition:

With an initial fortune of $V_0 = E^Q[(M_N - v)_+]$, one can always achieve the profit potential $(M_N - v)_+$ of the call option on day N, by using the investment strategy (3.5.27) ("exact replication strategy"). It follows that $V_0 = E^Q[(M_N - v)_+]$ can be interpreted as the **fair price of the call option**.

In a certain sense, V_0 is the price of "total safety", the statistical description P of the model does not play any role here!

3.6 Doob's inequality, convergence in L^p

In this section, we would like to discuss the convergence properties of martingales in L^p spaces. We start with a generalization of Kolmogorov's inequality (see (1.4.4)): let X_i be independent random variables with $E[X_i^2] < \infty$, $E[X_i] = 0$, then for all $u > 0$,

$$P\big[\max_{1 \le k \le n} |S_k| \ge u\big] \le \frac{1}{u^2}\,\mathrm{Var}(S_n).$$

Doob's inequality: *Let $X_m, m \ge 0$, be a submartingale, and $\lambda > 0$. Then*

$$\lambda P(A) \le E[X_n\, 1_A] \le E[X_n^+]\ , \tag{3.6.1}$$

where $A = \{\max_{0 \le m \le n} X_m \ge \lambda\}$.

Proof. Define $T = \inf\{m \ge 0, X_m \ge \lambda\}$. Then $H_n = 1\{T < n\}$, $n \ge 1$, is a predictable sequence with values in $[0,1]$. Thanks to (3.4.14), $(H \cdot X)_n$, $n \ge 0$, is a submartingale. We have

$$(H \cdot X)_n = \sum_{m=1}^{n} 1\{T < m\}(X_m - X_{m-1}) = X_n - X_{T \wedge n},\ n \ge 0\,,$$

and we obtain in this way, for $n \ge 0$,

$$E\big[(H \cdot X)_n\big] = E[X_n] - E[X_{T \wedge n}] \ge 0\,. \tag{3.6.2}$$

Note that on A, $X_{T\wedge n} \geq \lambda$, and on A^c, $T \wedge n = n$. Using (3.6.2), we obtain

$$E[X_n 1_A] + E[X_n 1_{A^c}] \geq \underbrace{E[X_{T\wedge n} 1_A]}_{\geq \lambda P[A]} + E[X_n 1_{A^c}] \,.$$

It follows that

$$\lambda P[A] \leq E[X_n 1_A] \leq E[X_n^+] \,. \qquad \square$$

Remark 3.32.

- (3.6.1) is a maximal inequality: it allows one to estimate $\max_{0\leq m\leq n} X_m$ with the help of the final value X_n.
- Kolmogorov's inequality is a consequence of (3.6.1): simply choose $X_n = S_n^2$, which is a submartingale, and $\lambda = u^2$ in (3.6.1). $\square$

Doob's inequality allows us to investigate the **convergence properties of martingales in $L^p, p > 1$**.

Proposition 3.33. *Let X_n be a submartingale, and $p \in (1, \infty)$. Set $\overline{X}_n = \max_{0\leq m\leq n} X_m^+$, then*

$$\|\overline{X}_n\|_p \leq \left(\frac{p}{p-1}\right) \|X_n\|_p \tag{3.6.3}$$

(i.e. $X_n \in L^p \Longrightarrow \overline{X}_n \in L^p$ and (3.6.3)).

Corollary 3.34. $(p > 1)$. *Let X_n be a martingale with $\sup_n E[|X_n|^p] < \infty$. Then $X_n \to X_\infty$ P-a.s. and in L^p. Furthermore, $\big\| \sup_n |X_n| \big\|_p \leq \frac{p}{p-1} \sup_n \|X_n\|_p$.*

Proof of Corollary 3.34. Thanks to the convergence theorem (3.4.23), $X_n \to X_\infty$ P-a.s. Furthermore, $|X_n|$ is a submartingale, and using (3.6.3), one has (monotone convergence)

$$E\Big[\Big(\sup_n |X_n|\Big)^p\Big] = \lim_n \uparrow E\Big[\Big(\sup_{0\leq m\leq n} |X_m|\Big)^p\Big] \leq \sup_n \left(\frac{p}{p-1}\right)^p E\big[|X_n|^p\big] < \infty \,.$$

It follows from Lebesgue's theorem that

$$\lim_n E\big[\underbrace{|X_\infty - X_n|^p}_{\leq 2^p \sup_{m\geq 0} |X_m|^p}\big] = 0 \,. \qquad \square$$

Proof of Proposition 3.33. Without loss of generality, we assume that $X_n \in L^p$, and, thanks to (3.2.24), that $X_n = X_n^+$ and (3.2.23) that $X_m \in L^p$ for $0 \leq m \leq n$. Then one has

$$\begin{aligned} E\big[(\overline{X}_n)^p\big] &= E\Big[\int_0^{\overline{X}_n} p\lambda^{p-1} d\lambda\Big] \\ &= E\Big[\int_0^{\infty} p\lambda^{p-1} 1\{\overline{X}_n \geq \lambda\} d\lambda\Big] \\ &\overset{\text{Fubini}}{=} \int_0^{\infty} p\lambda^{p-1} P\big[\overline{X}_n \geq \lambda\big]\, d\lambda \,. \end{aligned}$$

We know that for $\lambda > 0$, $\lambda\, P[\overline{X}_n \ge \lambda] \le E[X_n\, 1\{\overline{X}_n \ge \lambda\}]$ holds (using (3.6.1)). Hence,

$$\begin{aligned} E\big[(\overline{X}_n)^p\big] \;&\le\; \int_0^\infty p\lambda^{p-2} E[X_n\, 1\{\overline{X}_n \ge \lambda\}]\, d\lambda \\ &\overset{\text{Fubini}}{=} \Big(\frac{p}{p-1}\Big)\, E\Big[X_n \int_0^{\overline{X}_n} (p-1)\lambda^{p-2} d\lambda\Big] \\ &= \Big(\frac{p}{p-1}\Big)\, E\big[X_n (\overline{X}_n)^{p-1}\big] \\ &\overset{\text{Hölder}}{\le} \Big(\frac{p}{p-1}\Big) \cdot \|X_n\|_p \cdot E\big[(\overline{X}_n)^p\big]^{(1-\frac{1}{p})}\,. \end{aligned}$$

Consequently, either we have $E[(\overline{X}_n)^p] = 0$, or we can divide the previous inequality by $\|\overline{X}_n\|_p^{p-1}$. In both cases, it follows that

$$\|\overline{X}_n\|_p \le \Big(\frac{p}{p-1}\Big)\, \|X_n\|_p\,.$$

Our claim (3.6.3) follows. □

The following theorem and the subsequent proposition describe the convergence properties of martingales in L^p, $p > 1$.

Theorem 3.35. $(1 < p < \infty)$

Let $X_n, n \ge 0$, be an $(\mathcal{F}_n)$-martingale. The following properties are equivalent:

(3.6.4) $\quad \sup_{n \ge 0} \|X_n\|_p < \infty\,,$

(3.6.5) $\quad E\big[\sup_{n \ge 0} |X_n|^p\big] < \infty\,,$

(3.6.6) $\quad X_n$ *converges in* L^p,

(3.6.7) $\quad$ *there exists* $X \in L^p$ *such that for all* $n \ge 0$, $X_n = E\big[X \,|\, \mathcal{F}_n\big]$.

Proof. (3.6.4) $\Longrightarrow$ (3.6.5) and (3.6.5) $\Longrightarrow$ (3.6.6): see corollary below (3.6.3).

(3.6.6) $\Longrightarrow$ (3.6.7): choose $0 \le n \le m$, and $A \in \mathcal{F}_n$, then:

$$E[X_m\, 1_A] = E[X_n\, 1_A]\,.$$

For $m \to \infty$, X_m converges in L^p to X_∞. Hence,

(3.6.8) $$E[X_\infty\, 1_A] = E[X_n\, 1_A], \text{ for all } A \in \mathcal{F}_n,$$

and (P-a.s.)

$$E[X_\infty \,|\, \mathcal{F}_n] = X_n,\ n \ge 0\,.$$

This proves (3.6.7).

(3.6.7) $\Longrightarrow$ (3.6.4): see (3.1.23). □

The connection between X in (3.6.7) and X_∞ is explained in the next proposition.

Proposition 3.36. $(1 < p < \infty)$

In the case when $X_n = E[X \,|\, \mathcal{F}_n]$, $n \ge 0$, with $X \in L^p$, then

$$\text{(3.6.9)} \qquad \begin{aligned} & X_n \text{ converges } P\text{-a.s. and in } L^p \text{ to } X_\infty, \\ \text{where} \quad & X_\infty = E\big[X \,|\, \mathcal{F}_\infty\big] \; P\text{-a.s., with } \mathcal{F}_\infty = \sigma\Big(\bigcup_{n\ge 0} \mathcal{F}_n\Big). \end{aligned}$$

Proof. We already know that X_n converges P-a.s. and in L^p. Using (3.6.8), one also has, for $n \ge 0$ and $A \in \mathcal{F}_n$,

$$E[X_\infty 1_A] = E[X_n 1_A] = E\big[E[X \,|\, \mathcal{F}_n]\, 1_A\big] = E[X\, 1_A].$$

Using Dynkin's lemma (1.3.9), it follows that

$$E[X_\infty 1_A] = E[X\, 1_A], \text{ for all } A \in \sigma\Big(\bigcup_{n\ge 0} \mathcal{F}_n\Big) = \mathcal{F}_\infty,$$

and (3.6.9) is thus proved. □

Example 3.37. Let us consider the Galton-Watson chain. We keep the same notations as in Section 3.5, and we also assume that the distribution ν of the number of descendants has a finite second moment, i.e. $\sum_0^\infty k^2\, \nu(k) < \infty$, $m = \sum_{k=0}^\infty k\nu(k)$.

We consider the martingale $M_n = Z_n / m^n$. The following computations show (by induction) that M_n actually lies in L^2:

$$\begin{aligned}
& E\big[(M_n - M_{n-1})^2 | \mathcal{F}_{n-1}\big] = E\Big[\Big(\frac{Z_n}{m^n} - \frac{Z_{n-1}}{m^{n-1}}\Big)^2 | \mathcal{F}_{n-1}\Big] \\
= \; & m^{-2n}\, E\big[(Z_n - m\, Z_{n-1})^2 | \mathcal{F}_{n-1}\big] \\
\overset{\substack{\text{monotone} \\ \text{convergence}}}{=} \; & m^{-2n} \sum_{k=0}^\infty E\big[(Z_n - m\, Z_{n-1})^2 1\{Z_{n-1} = k\} | \mathcal{F}_{n-1}\big] \\
\text{(3.6.10)} \qquad = \; & m^{-2n} \sum_{k=0}^\infty E\big[(\underbrace{\xi_1^n + \dots + \xi_k^n}_{\text{independent of } \mathcal{F}_{n-1}} - mk)^2\, 1\underbrace{\{Z_{n-1} = k\}}_{\mathcal{F}_{n-1}\text{-measurable}} | \mathcal{F}_{n-1}\big] \\
= \; & m^{-2n} \sum_{k=0}^\infty 1\{Z_{n-1} = k\}\, E\big[\underbrace{(\xi_1^n + \dots + \xi_k^n - mk)^2}_{=((\xi_1^n - m) + \dots + (\xi_k^n - m))^2}\big] \\
= \; & m^{-2n} \sum_{k=0}^\infty k\sigma^2\, 1\{Z_{n-1} = k\} = m^{-2n}\, \sigma^2\, Z_{n-1} = m^{-(n+1)} \sigma^2\, M_{n-1},
\end{aligned}$$

where $\sigma^2 = \mathrm{Var}(\xi)$, and furthermore, one has (this computation holds for a general martingale in L^2)

$$E\big[M_n^2|\mathcal{F}_{n-1}\big] = E\big[M_{n-1}^2 + 2(M_n - M_{n-1})\,M_{n-1} + (M_n - M_{n-1})^2|\mathcal{F}_{n-1}\big]$$

$$(3.6.11)\qquad = M_{n-1}^2 + 2M_{n-1}\,\underbrace{E\big[(M_n - M_{n-1})|\mathcal{F}_{n-1}\big]}_{\substack{\| \\ 0 \text{ (martingale)}}} + E\big[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}\big]$$

$$= M_{n-1}^2 + E\big[(M_n - M_{n-1})^2|\mathcal{F}_{n-1}\big]\,.$$

Using (3.6.10), we obtain

$$E\big[M_n^2|\mathcal{F}_{n-1}\big] = M_{n-1}^2 + \frac{\sigma^2}{m^{n+1}}\,M_{n-1},\ n \ge 1\,.$$

Consequently,

$$(3.6.12)\qquad E[M_n^2] = E[M_{n-1}^2] + \frac{\sigma^2}{m^{n+1}} \overset{\text{induction}}{=} E[M_0^2] + \sigma^2 \sum_{k=1}^{n} \frac{1}{m^{k+1}} = 1 + \sigma^2 \sum_{k=2}^{n+1} m^{-k}\,.$$

In the supercritical case $m > 1$, we obtain from (3.6.12)

$$\sup E[M_n^2] < \infty \overset{\text{corollary}}{\Longrightarrow} M_n \underset{L^2}{\overset{P\text{-a.s.}}{\longrightarrow}} M_\infty \ge 0\,,\ \text{ with } E[M_\infty] = 1 = E[M_0]\,.$$

In particular, M_∞ is not P-a.s. $= 0$, and one has

$$(3.6.13)\qquad \frac{Z_n}{m^n}\,(= M_n) \overset{P\text{-a.s.}}{\longrightarrow} M_\infty\,.$$

Of course, $\bigcup_{n\ge1}\{Z_n = 0\} \subseteq \{M_\infty = 0\}$. One can show (see Durrett, ex. 5.3.12) that

$$\bigcup_{n\ge1}\{Z_n = 0\} \overset{P\text{-a.s.}}{=} \{M_\infty = 0\}$$

(i.e. the symmetric difference between the two sides has P-measure zero). In other words: in the supercritical case, either the population dies out, or it grows in a geometric way, like m^n, P-a.s. $\bigcirc$

We would like now to examine the **convergence properties of martingales in L^1**. Contrary to the case $p \in (1,\infty)$, there exist martingales $M_n, n \ge 0$, that converge in L^1 without having $\sup_n |M_n| \in L^1$ (see exercises). The key concept here is a property called uniform integrability.

Definition 3.38. *A family of random variables X_i, $i \in I$, on $(\Omega, \mathcal{A}, P)$ is said to be* uniformly integrable *if*

$$\lim_{M\to\infty} \sup_{i\in I} E\big[|X_i|\, 1_{\{|X_i|>M\}}\big] = 0 . \tag{3.6.14}$$

In particular, a finite family of L^1 random variables is uniformly integrable (the proof is left as an exercise).

Example 3.39. Let $X_i, i \in I$, be a family of random variables on $(\Omega, \mathcal{A}, P)$ with $\sup_I E\big[\varphi(|X_i|)\big] = A < \infty$, where $\varphi(\cdot)$: $\mathbb{R}_+ \to \mathbb{R}_+$ is measurable, and $\lim_{u\to\infty} \varphi(u)/u = +\infty$. Then $X_i, i \in I$, is uniformly integrable. Indeed, choose $\epsilon > 0$, then there exists $M > 0$ such that

$$\inf_{u\ge M} \frac{\varphi(u)}{u} \ge \frac{A}{\epsilon} . \tag{3.6.15}$$

Hence, for $i \in I$,

$$\begin{aligned} E\big[|X_i|\, 1\{|X_i| > M\}\big] &\le \frac{\epsilon}{A}\, E\Big[\frac{\varphi(|X_i|)}{|X_i|} \cdot |X_i|\, 1\{|X_i| > M\}\Big] \\ &\le \frac{\epsilon}{A}\, E\big[\varphi(|X_i|)\big] \le \epsilon . \end{aligned}$$

(3.6.14) thus follows. ○

A further example of uniformly integrable random variables is provided by the following result.

Proposition 3.40. *Consider $X \in L^1(\Omega, \mathcal{A}, P)$. The family*

$$\big\{E[X|\mathcal{F}];\ \mathcal{F} \text{ sub-}\sigma\text{-algebra of } \mathcal{A}\big\} \tag{3.6.16}$$

is uniformly integrable.

Proof. We first note that

$$\lim_{\eta\to 0} \sup_{A\in\mathcal{A}, P[A]\le\eta} E\big[|X|\, 1_A\big] = 0 . \tag{3.6.17}$$

Indeed, for $M > 0$,

$$E\big[|X|\, 1_A\big] \le E\big[|X|\, 1\{|X| > M\}\big] + \underbrace{M\, P[A]}_{\le M\eta} ,$$

and Lebesgue's theorem implies $\lim\limits_{M\to\infty} E\big[|X|\, 1\{|X| > M\}\big] = 0$, hence (3.6.17) follows by choosing $M = \eta^{-1/2}$.

Moreover, one has, for $\mathcal{F}$ a sub-σ-algebra of $\mathcal{A}$, and $M > 0$,

$$\begin{aligned} E\Big[\big|E[X|\mathcal{F}]\big|\,1\big\{|E[X|\mathcal{F}]| > M\big\}\Big] &\overset{\text{Jensen}}{\leq} E\Big[E\big[|X|\,|\mathcal{F}\big]\,1\underbrace{\big\{E[|X|\,|\mathcal{F}] > M\big\}}_{\in\mathcal{F}}\Big] \\ &= E\Big[|X|\,1\big\{E[|X|\,|\mathcal{F}] > M\big\}\Big] \end{aligned}$$

and

$$P\big[E[|X|\,|\mathcal{F}] > M\big] \overset{\text{Chebyshev}}{\leq} \frac{1}{M}\,E\big[E[|X|\,|\mathcal{F}]\big] = \frac{E[|X|]}{M}\,.$$

If we choose $\eta = \frac{E[|X|]}{M}$, our claim (3.6.16) follows thanks to (3.6.17). □

We will need the following proposition:

Proposition 3.41. *Let $X_n, n \geq 0$, and X be random variables in L^1 with $X_n \to X$ P-a.s. The following three properties are then equivalent:*

(3.6.18) $$\{X_n, n \geq 0\} \text{ is uniformly integrable}\,,$$

(3.6.19) $$X_n \xrightarrow{L^1} X\,,$$

(3.6.20) $$(\infty >)\; E\big[|X_n|\big] \underset{n\to\infty}{\longrightarrow} E\big[|X|\big] < \infty\,.$$

Proof. (3.6.18) $\Longrightarrow$ (3.6.19): for $M > 0$, one has

(3.6.21) $$\begin{aligned} E\big[|X_n - X|\big] \;\leq\; & E\big[|X_n - X|\,1\{|X_n| \leq M, |X| \leq M\}\big] \\ & + 3E\big[|X_n|\,1\{|X_n| > M\}\big] + 3E\big[|X|\,1\{|X| > M\}\big]\,. \end{aligned}$$

Consider $\epsilon \in (0,1)$. Thanks to (3.6.18), we can choose M_0 such that

(3.6.22) $$M \geq M_0 \Longrightarrow \sup_n\; E\big[|X_n|\,1\{|X_n| > M\}\big] \leq \frac{\epsilon}{6}\,.$$

Fatou's lemma implies

$$E\big[|X|\big] \leq \underline{\lim}\; E\big[|X_n|\big] \leq \frac{\epsilon}{2} + M_0 \leq M_0 + 1\,.$$

We can thus choose $M > 0$ such that the sum of the last two terms in (3.6.21) is smaller than ϵ uniformly in n. It follows that

$$\overline{\lim}\; E\big[|X_n - X|\big] \leq \overline{\lim_n}\; E\big[|X_n - X|\,1\{|X_n| \leq M, |X| \leq M\}\big] + \epsilon$$

$$\big\downarrow \text{ Lebesgue}$$

$$0$$

$$= \epsilon, \quad \epsilon > 0 \;\text{ arbitrary}\,.$$

(3.6.19) $\Longrightarrow$ (3.6.20): this follows from

$$\big|E\big[|X_n|\big] - E\big[|X|\big]\big| \le E\big[|X_n - X|\big] \ .$$

(3.6.20) $\Longrightarrow$ (3.6.18): take $\epsilon > 0$, and define, for $M > 1$, the continuous function $\psi_M(.)$ satisfying

$$\psi_M(x) = \begin{cases} x & \text{for } x \in [0, M-1], \\ 0 & \text{for } x \ge M, \\ \text{linear} & \text{for } x \in [M-1, M]. \end{cases}$$

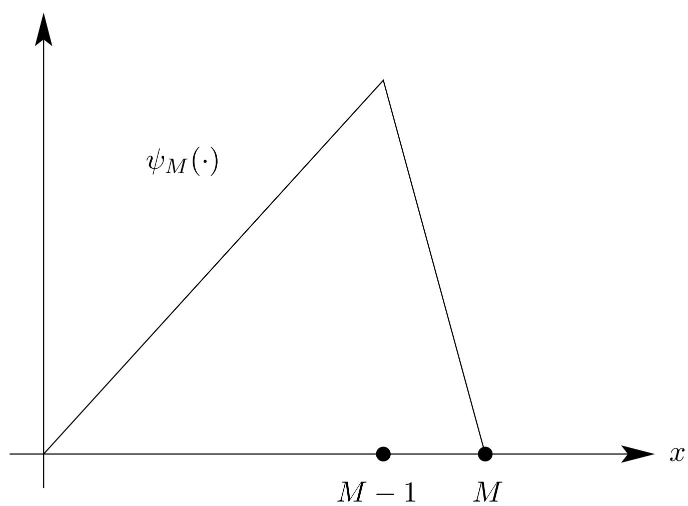


Fig. 3.9

Choose M_0 such that $0 \le E\big[|X|\big] - E\big[\psi_{M_0}(|X|)\big] \le \frac{\epsilon}{2}$. Lebesgue's theorem implies that

$$\lim_n \ E\big[\psi_{M_0}(|X_n|)\big] = E\big[\psi_{M_0}(|X|)\big] \ . \tag{3.6.23}$$

Then,

$$\begin{aligned} \overline{\lim_n} \, E\big[|X_n| \, 1\{|X_n| > M_0\}\big] \ &\le \ \overline{\lim_n} \, \big(E\big[|X_n|\big] - E\big[\psi_{M_0}(|X_n|)\big]\big) \\ &= \ \lim_n \ E\big[|X_n|\big] - \lim_n \ E\big[\psi_{M_0}(|X_n|)\big] \\ &\overset{(3.6.20)-(3.6.23)}{=} \ E\big[|X|\big] - E\big[\psi_{M_0}(|X|)\big] \le \frac{\epsilon}{2} \ . \end{aligned} \tag{3.6.24}$$

$X_n \in L^1$ for all n, and thanks to (3.6.24) and Lebesgue's theorem, we can choose $M \ge M_0$ such that $\sup_n \ E\big[|X_n| \, 1\{|X_n| > M\}\big] \le \epsilon$. $\square$

As an application of the notion of uniform integrability, we have the following theorem for martingales:

Theorem 3.42. *Let $M_n, n \geq 0$, be a martingale. The following three properties are equivalent:*

$$\{M_n, n \geq 0\} \text{ is uniformly integrable}, \tag{3.6.25}$$

$$M_n \text{ converges in } L^1, \tag{3.6.26}$$

$$\text{there exists } X \in L^1(\Omega, \mathcal{A}, P) \text{ with } M_n = E[X|\mathcal{F}_n] \text{ for all } n \geq 0. \tag{3.6.27}$$

Proof. (3.6.25) $\Longrightarrow$ (3.6.26): Thanks to (3.6.25), one has $\sup_n \; E[|M_n|] < \infty$ (see for instance (3.6.22)). Hence, $M_n \to M_\infty$ P-a.s. (convergence theorem), and (3.6.26) follows, using (3.6.19).

(3.6.26) $\Longrightarrow$ (3.6.27): Consider $n < m$, and $A \in \mathcal{F}_n$. One has:

$$E[M_n \, 1_A] = E[M_m \, 1_A] \overset{(3.6.26)}{\longrightarrow} E[M_\infty \, 1_A],$$

from which it follows that

$$M_n = E[M_\infty | \mathcal{F}_n], \; n \geq 0 \,. \tag{3.6.28}$$

(3.6.27) $\Longrightarrow$ (3.6.25): this follows from (3.6.16). □

In a very similar way to (3.6.9), one has the following proposition:

Proposition 3.43. *Assume that $M_n = E[X \mid \mathcal{F}_n]$, $n \geq 0$, with $X \in L^1$. One has*

$$\begin{aligned} &M_n \text{ converges } P\text{-a.s. and in } L^1 \text{ to } M_\infty, \\ \text{and} \quad &M_\infty = E[X \mid \mathcal{F}_\infty], \text{ where } \mathcal{F}_\infty = \sigma\Big(\bigcup_{n \geq 0} \mathcal{F}_n\Big). \end{aligned} \tag{3.6.29}$$

Proof. same as for (3.6.9). □

Example 3.44. (Radon-Nikodym derivative)

We keep the same notations as in Section 4.2, Example 4). We assume that

$$\mu \ll \nu \; \text{ on } \; (\Omega, \mathcal{F}), \tag{3.6.30}$$

and we define

$$X = \frac{d\mu}{d\nu} \,,$$

then condition (3.2.12) is automatically satisfied ($\mu_n \ll \nu_n$ on $(\Omega, \mathcal{F}_n)$). Furthermore (see the proof of existence for the conditional expectation),

$$M_n \overset{\text{def.}}{=} \frac{d\mu_n}{d\nu_n} = E^\nu[X|\mathcal{F}_n], \;\; n \geq 0 \,. \tag{3.6.31}$$

Using (3.6.29), one has

$$M_n \underset{L^1}{\overset{\nu\text{-a.s.}}{\longrightarrow}} X \quad \Big(\text{since } \mathcal{F} = \sigma\Big(\bigcup_{m\geq 0} \mathcal{F}_m\Big)\Big) .$$

In the case when (3.2.12) holds, but (3.6.30) is not satisfied, one can show that $M_n \overset{\nu\text{-a.s.}}{\longrightarrow} X \in L^1(\Omega, \mathcal{F}, \nu)$, and

$$\text{(3.6.32)} \qquad \forall A \in \mathcal{F}, \;\; \mu(A) = \int_A X \, d\nu + \mu\Big(A \cap \underbrace{\{\overline{\lim_n} \, M_n = \infty\}}_{\text{set with } \nu\text{-measure equal to } 0}\Big) ,$$

see Durrett, p.242 (it uses the Radon-Nikodym decomposition of μ with respect to ν). Of course, M_n does not converge in L^1 to X when (3.6.30) is not satisfied (otherwise, we would have $\mu = X \cdot \nu \Longrightarrow \mu \ll \nu\,!$). ○

4 Random walks, Markov chains

4.1 Random walks

In this section, we study random walks, which are paths

(4.1.1) $$S_0, S_1, S_2, \dots, S_n, \dots,$$

where $S_0 = 0$, $S_n = X_1 + \dots + X_n$, $n \geq 1$, and the X_i, $i \geq 1$, are independent, identically distributed random variables (possibly with values in $\mathbb{R}^d$).

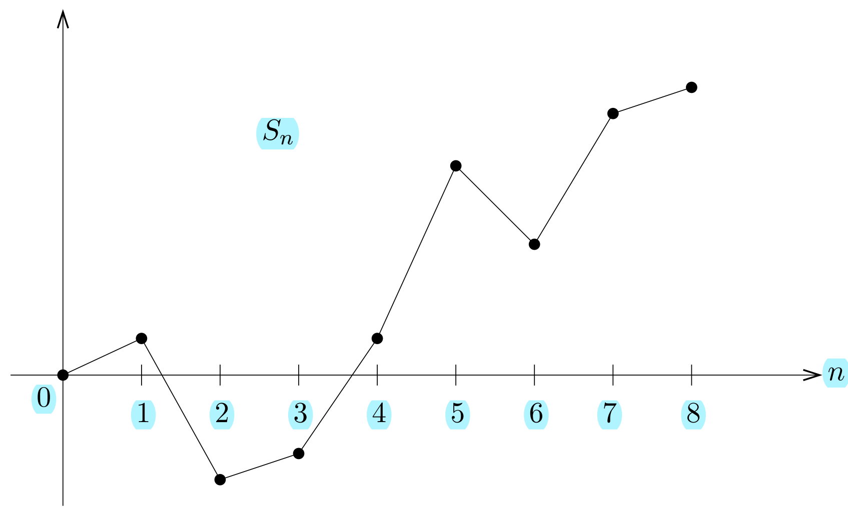


Fig 4.1

Unlike in previous chapters, we are no longer interested in the behavior of the distribution of S_n for large n, but rather in analyzing the **path** $\big(S_n(\omega)\big)_{n \in \mathbb{N}}$, for a "typical" ω in Ω.

Here are some examples of questions we are interested in:

- Case $d = 1$: does $S_n(\omega)$ converge, for a typical ω, to $+\infty$ or $-\infty$? Could the path display a different behavior, and how?

- Case $d \geq 1$: does $S_n(\omega)$ come back infinitely often near 0, for a typical ω (this is called the **recurrence** property)?

The Law of Large Numbers already provides a partial answer to such questions, at least in certain cases.

Example 4.1. We assume that the X_i, $i \geq 1$, have distribution

(4.1.2) $$P[X_i = 1] = p, \quad P[X_i = -1] = 1 - p\,,$$

for some fixed parameter $p \in (0,1)$. In the case $p = \frac{1}{2}$, $\big(S_n(\omega)\big)_{n \in \mathbb{N}}$ is called the **(symmetric) simple random walk on $\mathbb{Z}$**.

In the case $p \neq \frac{1}{2}$, one has $E[X_1] = 1 \cdot p - 1 \cdot (1-p) = 2p - 1 \neq 0$, and the Law of Large Numbers readily implies that

$$\frac{S_n(\omega)}{n} \underset{n\to\infty}{\longrightarrow} 2p-1 \qquad P\text{-a.s.}\,, \tag{4.1.3}$$

so in particular

$$\begin{aligned} & S_n(\omega) \underset{n\to\infty}{\longrightarrow} +\infty \quad P\text{-a.s.} \quad \text{if } p > \frac{1}{2}, \\ \text{and} \quad & S_n(\omega) \underset{n\to\infty}{\longrightarrow} -\infty \quad P\text{-a.s.} \quad \text{if } p < \frac{1}{2}\,. \end{aligned}$$

The case $p = \frac{1}{2}$ (simple random walk) is somewhat less clear. In this case, (4.1.3) does not give any direct answer to the question "does S_n converge to $+\infty$ or $-\infty$?".

We will see that for $p = \frac{1}{2}$, one has:

$$\overline{\lim_n}\, S_n(\omega) = +\infty\,, \quad \underline{\lim_n}\, S_n(\omega) = -\infty\,,$$

and $S_n(\omega)$ visits 0 infinitely often, P-a.s. (i.e. for a typical ω). $\bigcirc$

Random walks play a particularly important role in probability theory, they constitute fundamental examples of Markov chains and (in the case when $E[|X_i|] < \infty$, and $E[X_i] = 0$) martingales.

4.1.1 Hewitt-Savage 0-1 Law

We now prove a 0-1 law for random walks, the Hewitt-Savage 0-1 law. In order to present it, we will construct the **independent random variables** $X_1, \dots, X_n, \dots$ **on a specific probability space**. We introduce

$$\Omega = \mathbb{R}^{\mathbb{N}_*} = \{(\omega_1, \omega_2, \dots),\, \omega_i \in \mathbb{R}\}\,, \tag{4.1.4}$$

and we equip Ω with the **product σ-algebra**

$$\begin{aligned} \mathcal{A} = \; & \text{the smallest } \sigma\text{-algebra containing all sets of the form} \\ & B_1 \times \cdots \times B_m \times \mathbb{R} \times \cdots \times \mathbb{R} \times \cdots = \{\omega : \omega_i \in B_i,\ 1 \leq i \leq m\}\,, \\ & \text{with } m \text{ arbitrary}\,,\ B_i \in \mathcal{B}(\mathbb{R})\ \ (1 \leq i \leq m)\,. \end{aligned} \tag{4.1.5}$$

We consider on $(\Omega, \mathcal{A})$ the **canonical coordinate functions** $X_i, i \geq 1$:

$$X_i(\omega) = \omega_i\,. \tag{4.1.6}$$

The X_i are then **random variables on $(\Omega, \mathcal{A})$**, since:

$$\{X_i \in B\} = \mathbb{R} \times \mathbb{R} \times \cdots \times B_i \times \mathbb{R} \times \cdots \times \mathbb{R} \times \cdots\,.$$

Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and let us consider, on any probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$, some independent random variables $\widetilde{X}_i$, $i \ge 1$, with distribution μ. We can then define the map Φ from $\widetilde{\Omega}$ to Ω:

$$\begin{aligned} &\widetilde{\omega} \in \widetilde{\Omega} \longmapsto \Phi(\widetilde{\omega}) = (\widetilde{X}_1(\widetilde{\omega}), \widetilde{X}_2(\widetilde{\omega}), \dots) \in \Omega = \mathbb{R}^{\mathbb{N}_*} \\ &\qquad\qquad\qquad\qquad \text{“the sequence of values } \widetilde{X}_i(\widetilde{\omega})\text{”} \,. \end{aligned} \tag{4.1.7}$$

Φ is measurable from $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ to $(\Omega, \mathcal{A})$ since

$$\Phi^{-1}(B_1 \times \dots \times B_m \times \mathbb{R} \times \dots \times \mathbb{R} \times \cdots) = \bigcap_{i=1}^{m} \{\widetilde{X}_i \in B_i\} \in \widetilde{\mathcal{A}},$$

and consequently, using the same argument as in Chapter 1, (1.1.7), one has: $\Phi^{-1}(A) \in \widetilde{\mathcal{A}}$ for all $A \in \mathcal{A}$.

As a probability measure on $(\Omega, \mathcal{A})$, we consider the image measure of $\widetilde{P}$ under Φ:

$$P = \Phi \circ \widetilde{P}\,. \tag{4.1.8}$$

Then one has, for all $m \ge 1$, $B_1, \dots, B_m \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} &P[B_1 \times \dots \times B_m \times \mathbb{R} \times \dots \times \mathbb{R} \times \cdots] \\ &\overset{(4.1.6)}{=} P[X_1 \in B_1, \dots, X_m \in B_m] \\ &= \widetilde{P}\big[\Phi^{-1}(B_1 \times \dots \times B_m \times \mathbb{R} \times \dots \times \mathbb{R} \times \cdots)\big] \\ &= \widetilde{P}\Big[\bigcap_{i=1}^{m} \{\widetilde{X}_i \in B_i\}\Big] = \prod_{i=1}^{m} \mu(B_i)\,. \end{aligned} \tag{4.1.9}$$

Note that the family of all sets of the form $B_1 \times \dots \times B_m \times \mathbb{R} \times \dots \times \mathbb{R} \times \cdots$ with $m \ge 1$, $B_i \in \mathcal{B}(\mathbb{R})$, $1 \le i \le m$, forms a π-system, and this π-system generates $\mathcal{A}$ by definition.

As a consequence of (4.1.9), we see that **P is uniquely determined by μ**, and that the **$X_i, i \ge 1$**, on $(\Omega, \mathcal{A}, P)$ are **independent random variables** with **distribution μ**.

Notation: The standard notation for $\mathcal{A}$ in (4.1.5) and P in (4.1.8) is

$$\mathcal{A} = \bigotimes_{i \ge 1} \mathcal{B}(\mathbb{R}) \quad \text{and} \quad P = \underset{\substack{\uparrow \\ \text{“infinite product probability measure”}}}{\bigotimes_{i \ge 1}} \mu\,. \tag{4.1.10}$$

$(\Omega, \mathcal{A}, P)$ is called “infinite product probability space”.

Remark 4.2.

1) In the construction above, one can of course replace the space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, when the distribution μ is d-dimensional.

2) If we consider functions of the path, such as $\overline{\lim}\, \widetilde{S}_n$ or $\underline{\lim}\, \frac{\widetilde{S}_n}{n}$, their distributions do not depend on the probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, P)$ on which the independent random variables $\widetilde{X}_i$ with distribution μ are considered. Indeed,

$$\widetilde{X}_i(\widetilde{\omega}) = X_i \circ \Phi(\widetilde{\omega}),\ i \ge 1 \Longrightarrow \widetilde{S}_n(\widetilde{\omega}) = S_n \circ \Phi(\widetilde{\omega}), \tag{4.1.11}$$

and thus, for example, $\overline{\lim}\,\widetilde{S}_n = \overline{\lim}\,S_n \circ \Phi$. Using $P = \Phi \circ \widetilde{P}$, it follows that $\overline{\lim}\,\widetilde{S}_n$ and $\overline{\lim}\,S_n$ have the same distribution.

Let μ be a distribution on $\mathbb{R}$ (or $\mathbb{R}^d$), and $(\Omega, \mathcal{A}, P)$ be the infinite product probability space, as above. We now introduce the **symmetric events of $\boldsymbol{\mathcal{A}}$**. $\square$

Definition 4.3. $A \in \mathcal{A}$ *is said to be symmetric if for all* $m \geq 1$ *and* $\sigma \in \mathcal{S}_m$ ($\stackrel{\text{def.}}{=}$ *set of all permutations of* $\{1, \dots, m\}$), *one has:*

$$
\begin{aligned}
&\omega = (\omega_1, \omega_2, \dots, \omega_m, \dots) \in A \Longleftrightarrow \\
&\sigma.\omega \stackrel{\text{def.}}{=} (\omega_{\sigma(1)}, \dots, \omega_{\sigma(m)}, \omega_{m+1}, \omega_{m+2}, \dots) \in A
\end{aligned}
\tag{4.1.12}
$$

(in other words: $A =$ inverse of A under σ.

$$\| $$

(Notation) $(\sigma.)^{-1}(A))$.

It is clear from the definition that the **family** of all **symmetric events $\boldsymbol{\mathcal{P}}$** forms a **sub-$\boldsymbol{\sigma}$-algebra of $\boldsymbol{\mathcal{A}}$**, since one has:

$$
(\sigma.)^{-1}(\Omega) = \Omega, \qquad (\sigma.)^{-1}(A^c) = \big((\sigma.)^{-1}(A)\big)^c \quad \text{for all } A \in \mathcal{A},
$$

$$
\text{and} \qquad (\sigma.)^{-1}\Big(\bigcup_{i \geq 1} A_i\Big) = \bigcup_{i \geq 1} (\sigma.)^{-1}(A_i) \quad \text{for all } A_i \in \mathcal{A},\ i \geq 1\,.
$$

We now present a few examples of symmetric events.

Example 4.4.

1) Let $S_n = X_1 + \dots + X_n$, $n \geq 1$ (see (4.1.6)), in the case $d = 1$, and $B \in \mathcal{B}(\mathbb{R})$. Then

$$
\limsup_{n \geq 1} \{S_n \in B\} = \{\omega \in \Omega : \ S_n(\omega) \in B \ \text{ infinitely often}\}
\tag{4.1.13}
$$

is symmetric, since for $\sigma \in \mathcal{S}_m$, $\omega = (\omega_1, \omega_2, \dots)$, and $n \geq m$,

$$
\begin{aligned}
S_n(\sigma.\,\omega) &= \omega_{\sigma(1)} + \dots + \omega_{\sigma(m)} + \omega_{m+1} + \dots + \omega_n \\
&= S_n(\omega)\,.
\end{aligned}
\tag{4.1.14}
$$

2) Thanks to (4.1.14), one has, for $\omega \in \Omega$ and $\sigma \in \mathcal{S}_m$,

$$
\limsup S_n(\omega) = \limsup S_n(\sigma.\omega),
\tag{4.1.15}
$$

and for $B \in \mathcal{B}([-\infty, +\infty])$,

$$
\{\limsup S_n \in B\} \in \mathcal{P}
\tag{4.1.16}
$$

(in other words, $\limsup S_n$ is measurable from $(\Omega, \mathcal{P})$ to $([-\infty, +\infty], \mathcal{B}([-\infty, +\infty]))$).

3) Let $\mathcal{F}_\infty = \cap_{n\geq 1} \sigma(X_n, X_{n+1}, \dots)$ be the asymptotic σ-algebra associated with the random variables X_i. Then

$$\mathcal{F}_\infty \subseteq \mathcal{P}\,, \tag{4.1.17}$$

since for $\sigma \in \mathcal{S}_m$, one has $\mathcal{F}_\infty \subseteq \mathcal{F}_{m+1} = \sigma(X_{m+1}, X_{m+2}, \dots)$, and for $B \in \mathcal{F}_{m+1}$,

$$\omega \in B \Longleftrightarrow \sigma.\,\omega = (\omega_{\sigma(1)}, \dots, \omega_{\sigma(m)}, \omega_{m+1}, \omega_{m+2}, \dots) \in B$$

"B does not depend on the first m components".

Note also that $\mathcal{F}_\infty \subsetneq \mathcal{P}$, since

$$A \overset{\text{def.}}{=} \liminf\{S_n = 0\} = \{\omega : S_n(\omega) = 0 \;\text{ for } n \text{ large enough}\} \in \mathcal{P}\,,$$

thanks to (4.1.15), but $A \notin \mathcal{F}_2 = \sigma(X_2, X_3, \dots)$, since

$$\omega = (0, 0, 0, \dots) \in A \quad \text{but} \quad (1, 0, 0, \dots) \notin A$$

"A depends on the first component ω_1 of ω". ◯

In the case of an infinite product probability space $(\Omega, \mathcal{A}, P)$, the Hewitt-Savage 0-1 law generalizes Kolmogorov's 0-1 law:

Theorem 4.5. (*Hewitt-Savage 0-1 law*) *Let* $d \geq 1$,

$$\textit{for } A \in \mathcal{P}, \textit{ one has either } P(A) = 0 \textit{ or } 1\,. \tag{4.1.18}$$

We first discuss an application of the Hewitt-Savage 0-1 law, to analyze the asymptotic behavior of the 1-dimensional random walk.

Theorem 4.6. *Let* $S_n, n \geq 0$, *be a random walk on* $\mathbb{R}$. *Then, one of the following four possibilities holds:*

$$\begin{aligned}
&\text{a) } \forall n \geq 0,\ \ S_n = 0 && P\text{-a.s.,}\\
&\text{b) } \lim_{n\to\infty} S_n = +\infty && P\text{-a.s.,}\\
&\text{c) } \lim_{n\to\infty} S_n = -\infty && P\text{-a.s.,}\\
&\text{d) } \varliminf_{n\to\infty} S_n = -\infty \ \ \textit{and} \ \ \varlimsup_{n\to\infty} S_n = +\infty && P\text{-a.s.}
\end{aligned} \tag{4.1.19}$$

Proof. It follows from (4.1.16) that $\varlimsup_n S_n$ is $\mathcal{P}$-measurable, and thanks to (4.1.18), there exists one $c \in [-\infty, +\infty]$ for which

$$\varlimsup_n S_n = c \qquad P\text{-a.s.} \tag{4.1.20}$$

(c is simply the same as $\inf\{a \in [-\infty, +\infty];\ P[\overline{\lim}\ S_n \le a] = 1\}$). Note that $S_{n+1} = X_1 + S'_n$, where $S'_n = X_2 + \cdots + X_{n+1}$, $n \ge 1$, is also a random walk, with the same distribution as S_n. Consequently (see Remark 2 above),

$$c \overset{P\text{-a.s.}}{=} \overline{\lim_n}\ S_{n+1} = X_1 + \overline{\lim}\ S'_n \overset{P\text{-a.s.}}{=} X_1 + c\,. \tag{4.1.21}$$

If we have $c \in \mathbb{R}$, then

$$\begin{aligned} c \overset{P\text{-a.s.}}{=} X_1 + c \quad &\Longrightarrow \quad X_1 \overset{P\text{-a.s.}}{=} 0 \\ \Longrightarrow \quad \forall n \ge 0,\ S_n = 0 \quad &\Longrightarrow \quad \text{a) holds}\,. \end{aligned} \tag{4.1.22}$$

In the case when $c = \pm\infty$, then in particular X_1 is not P-a.s. $= 0$. We can argue similarly with $\underline{\lim}_n\ S_n$, and find that $\underline{\lim}_n\ S_n \overset{P\text{-a.s.}}{=} c' \in [-\infty, +\infty]$. From the condition that X_1 is not P-a.s. $= 0$, it follows, as in (4.1.22), that $c' \notin \mathbb{R}$, i.e. $c' = \pm\infty$.

In other words: if $c \in \{+\infty, -\infty\}$, then one also has $c' \in \{+\infty, -\infty\}$. It follows that we are in one of cases b), c), or d). $\square$

Example 4.7.

1) X_1 and $-X_1$ have the same distribution, and X_1 is not P-a.s. $= 0$. Then $-S_n, n \ge 1$, and $S_n, n \ge 1$, are two random walks with the same distribution. We have

$$\overline{\lim_n}\ S_n \overset{P\text{-a.s.}}{=} c \in \{+\infty, -\infty\}, \quad \text{and} \quad \overline{\lim_n}(-S_n) \overset{P\text{-a.s.}}{=} -\underline{\lim_n}\ S_n.$$

Hence, $\underline{\lim}_n\ S_n \overset{P\text{-a.s.}}{=} -\overline{\lim}_n\ S_n = -c \in \{+\infty, -\infty\} \Longrightarrow c = +\infty$, and **(4.1.19) d) holds**.

In particular, this holds for the **simple random walk on** $\mathbb{Z}$.

2) $X_1 = 1 + X'_1$, where **X'_1 is Cauchy distributed**, that is, $\mu_{X'_1}(dx) = \frac{1}{\pi}\ \frac{1}{1+x^2}\,dx$. Note that $E[|X_1|] = \infty$, so we are not allowed to use the Law of Large Numbers!

Define $S'_n = X'_1 + \cdots + X'_n$, then $\frac{S'_n}{n}$ is Cauchy distributed as well, since

$$\underbrace{\varphi_{\frac{S'_n}{n}}(t)}_{\text{characteristic function}} \overset{\text{independence}}{=} \varphi_{X'_1}\Big(\frac{t}{n}\Big)^n = \exp\Big\{-n\,\Big|\frac{t}{n}\Big|\Big\} = \varphi_{X'_1}(t)\,.$$

Hence, one has

$$\begin{aligned} &P[S_n \ge 0] \ge P[S'_n \ge 0] = P\Big[\frac{S'_n}{n} \ge 0\Big] = \frac{1}{2}\,, \\ \text{and} \qquad &P[S_n \le 0] = P\Big[\frac{S'_n}{n} \le -1\Big] = P[X'_1 \le -1] > 0\,. \end{aligned}$$

(4.1.19) b) and c) are not possible: indeed, in case b),

$$S_n \overset{P\text{-a.s.}}{\longrightarrow} +\infty \Longrightarrow P[S_n \le 0] \underset{n\to\infty}{\overset{\text{Lebesgue}}{\longrightarrow}} 0\,,$$

which is a contradiction, and similarly, in case c),

$$S_n \overset{P\text{-a.s.}}{\longrightarrow} -\infty \Longrightarrow P[S_n \geq 0] \underset{n\to\infty}{\longrightarrow} 0 \,.$$

Hence, (4.1.19) d) holds:

$$S_n \text{ is of the same type } d) \text{ as } S_n'! \tag{4.1.23}$$

○

Let us now turn to the proof of the Hewitt-Savage 0-1 law.

Proof. Consider a symmetric event $A \in \mathcal{P}$, it suffices to show that

$$P(A)^2 = P(A) \,, \tag{4.1.24}$$

since then it would follow that $P(A) = 0$ or 1. We need the following approximation result.

Lemma 4.8. *Let $B \in \mathcal{A}$, then one has:*

$$\text{for any } \epsilon > 0, \text{ there exists } C \in \bigcup_{n\geq 1} \sigma(X_1,\dots,X_n) \text{ with } E[|1_B - 1_C|] \leq \epsilon \,. \tag{4.1.25}$$

Proof. Let $\mathcal{D}$ be the class of all $B \in \mathcal{A}$ for which (4.1.25) is satisfied. Then $\mathcal{D}$ is a Dynkin system:

- $\Omega \in \mathcal{D}$ is clear.
- For $B \in \mathcal{D}$ and $\epsilon > 0$, let $C \in \bigcup_{n\geq 1} \sigma(X_1,\dots,X_n)$ be as in (4.1.25). Then one has $E[|1_{B^c} - 1_{C^c}|] = E[|1_B - 1_C|] \leq \epsilon$. Hence, $B^c \in \mathcal{D}$.
- Let $B_i \in \mathcal{D}$ be pairwise disjoint events, $\epsilon > 0$, and $C_i \in \bigcup_{n\geq 1} \sigma(X_1,\dots,X_n)$ with $E[|1_{B_i} - 1_{C_i}|] \leq \frac{\epsilon}{2^{i+1}}$.

Note that for $m \geq 1$, one has

$$\Big|1_{\bigcup_{i=1}^m B_i} - 1_{\bigcup_{i=1}^m C_i}\Big| \leq \sum_{i=1}^m |1_{B_i} - 1_{C_i}|, \tag{4.1.26}$$

since $x \in \bigcup_{i=1}^m B_i$ and $x \notin \bigcup_{i=1}^m C_i \Longrightarrow \exists j \in \{1,\dots,m\}$ with $x \in B_j \backslash C_j$, and analogously, $x \in \bigcup_{i=1}^m C_i$ and $x \notin \bigcup_{i=1}^m B_i \Longrightarrow \exists k \in \{1,\dots,m\}$ with $x \in C_k \backslash B_k$.

Let us choose m such that $\sum_{i=m+1}^\infty P(B_i) \leq \frac{\epsilon}{2}$. Then one has, thanks to (4.1.26),

$$\begin{aligned} E\Big[\Big|1_{\bigcup_{i=1}^\infty B_i} - 1_{\bigcup_{i=1}^m C_i}\Big|\Big] &\leq \sum_{i=m+1}^\infty P[B_i] + \sum_{i=1}^m E\big[|1_{B_i} - 1_{C_i}|\big] \\ &\leq \frac{\epsilon}{2} + \epsilon \sum_{i=1}^m \frac{1}{2^{i+1}} \leq \epsilon \Longrightarrow \bigcup_{i=1}^\infty B_i \in \mathcal{D} \,. \end{aligned}$$

Clearly, $\mathcal{D}$ contains the π-system $\bigcup_{n\geq 1}\sigma(X_1,\dots,X_n)$. Dynkin's lemma (1.3.9) thus implies that $\mathcal{D}=\mathcal{A}(=\sigma(\bigcup_{n\geq 1}\sigma(X_1,\dots,X_n)))$. This establishes (4.1.25). $\square$

Let $\epsilon>0$. Using (4.1.25), there exist $n\geq 1$ and $\widetilde{A}\in\sigma(X_1,\dots,X_n)$ such that

$$E[|1_A-1_{\widetilde{A}}|]\leq\epsilon.$$

Let σ be the permutation on $\{1,\dots,2n\}$ defined by

$$\begin{aligned} j\in\{1,\dots,n\} &\mapsto \sigma(j)=j+n,\\ j\in\{n+1,\dots,2n\} &\mapsto \sigma(j)=j-n. \end{aligned}$$

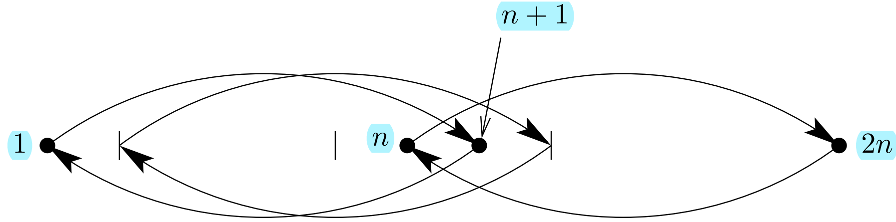


Fig. 4.2

Note that the image of $P=\bigotimes_{i\geq 1}\mu$ under $\sigma_{\bullet}$ is exactly P, since

$$\begin{aligned} &P[(\sigma_{\bullet})^{-1}(B_1\times\dots\times B_{2n}\times\dots\times B_m\times\mathbb{R}\times\dots\mathbb{R})]\\ &= P[X_{\sigma(1)}\in B_1,\dots,X_{\sigma(2n)}\in B_{2n},\ X_{2n+1}\in B_{2n+1},\dots,X_m\in B_m]\\ &\overset{\text{independence}}{=} \prod_{i=1}^m \mu(B_i)\ . \end{aligned}$$

Then,

$$\begin{aligned} \epsilon\geq E\big[|1_A-1_{\widetilde{A}}|\big] &= E\big[|1_A-1_{\widetilde{A}}|\circ\sigma_{\bullet}\big]\\ = \underbrace{E\big[|1_A\circ\sigma_{\bullet}-1_{\widetilde{A}}\circ\sigma_{\bullet}|\big] = E\big[|1_A-1_{(\sigma_{\bullet})^{-1}(\widetilde{A})}|\big]}_{A \text{ is symmetric}}\ . \end{aligned} \tag{4.1.27}$$

Now, it follows from the construction of σ, and $\widetilde{A}\in\sigma(X_1,\dots,X_n)$, that $\widetilde{B}\overset{\text{def.}}{=}(\sigma_{\bullet}^{-1})(\widetilde{A})\in\sigma(X_{n+1},\dots,X_{2n})$. Hence, $\widetilde{A}$ and $\widetilde{B}$ are independent. The triangle inequality thus implies that

$$\begin{aligned} &|P(A)-P(A)^2-(P(\widetilde{A})-P(\widetilde{A})\,P(\widetilde{B}))|\\ &\leq 2|P(A)-P(\widetilde{A})|+|P(A)-P(\widetilde{B})|\leq 3\epsilon\ , \end{aligned} \tag{4.1.28}$$

and

$$P(\widetilde{A})-P(\widetilde{A})\,P(\widetilde{B})\overset{\text{independence}}{=}P(\widetilde{A})-P(\widetilde{A}\cap\widetilde{B})\ .$$

However,

$$\begin{aligned}
\Big|1_{\widetilde{A}\cap\widetilde{B}} - \overbrace{1_A \cdot 1_A}^{1_A}\Big| &\le |1_{\widetilde{B}}(1_{\widetilde{A}} - 1_A)| + |1_A(1_{\widetilde{B}} - 1_A)| \\
&\le |1_{\widetilde{A}} - 1_A| + |1_{\widetilde{B}} - 1_A| \,.
\end{aligned}$$

It follows from (4.1.27) that

$$|P(\widetilde{A}) - P(\widetilde{A} \cap \widetilde{B})| \le 3\epsilon$$

(triangle inequality). Using (4.1.28), we finally obtain

$$|P(A) - P(A)^2| \le 3\epsilon + 3\epsilon = 6\epsilon,$$

where $\epsilon > 0$ is arbitrary. This completes the proof of (4.1.24). □

4.1.2 Strong Markov property, Wald's identity

Stopping times play a particularly important role when studying random walks. As we now explain, if N is a finite stopping time, then the "random walk after time N", i.e. $(S_{N+n} - S_N)_{n\ge 0}$, has the same distribution as the original random walk $(S_n)_{n\ge 0}$, and it is independent of $\mathcal{F}_N$. This property satisfied by the random walk $(S_n)_{n\ge 0}$ is known as the strong Markov property.

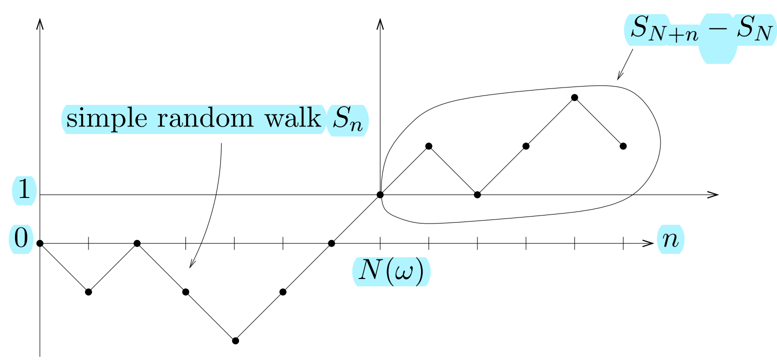


Fig. 4.3

Let us consider $X_i, i \ge 1$, some independent random variables on $(\Omega, \mathcal{A}, P)$ with distribution μ. In what follows, we will always consider the natural filtration $(\mathcal{F}_n)$, defined by

$$\begin{aligned}
&\mathcal{F}_0 = \{\phi, \Omega\} \quad \text{"trivial } \sigma\text{-algebra"}\,, \\
&\mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n),\ n \ge 1\,.
\end{aligned} \tag{4.1.29}$$

Theorem 4.9. *Let N be an $(\mathcal{F}_n)$-stopping time such that $P(N < \infty) > 0$, then*

$$\begin{aligned}
&\textit{on } \{N < \infty\}, \textit{ the random variables } X_{N+n},\ n \ge 1\,, \\
&\textit{are independent, } \mu\textit{-distributed, and independent of } \mathcal{F}_N\,.
\end{aligned} \tag{4.1.30}$$

In other words, the random variables X_{N+n}, $n \ge 1$, *on the probability space*

$$\Big(\overbrace{\{N<\infty\}}^{\text{“new space”}},\ \mathcal{A}\cap\{N<\infty\},\ Q(\cdot) \overset{\text{def.}}{=} \frac{1}{P[N<\infty]}\cdot P[\cdot\cap\{N<\infty\}]\Big)$$

are independent, μ*-distributed random variables, which are independent of* $\mathcal{F}_N \cap \{N<\infty\}$*. Hence,*

$$\text{(4.1.31)}\qquad \begin{array}{l} \textit{on } \{N<\infty\},\ (S_{N+n}-S_N)_{n\ge 0} \textit{ is independent of } \mathcal{F}_N, \textit{ and} \\ \textit{it possesses the same distribution as the random walk } (S_n)_{n\ge 0}\,. \end{array}$$

Proof. (4.1.31) follows directly from (4.1.30): indeed, on $\{N<\infty\}$, one has $S_{N+n}-S_N=0$ for $n=0$, and $S_{N+n}-S_N = X_{N+1}+\dots+X_{N+n}$ for $n\ge 1$. □

Proof of (4.1.30). It is enough to show that for $A\in\mathcal{F}_N$ and for $B_1,\dots,B_k\in\mathcal{B}(\mathbb{R})$, $k\ge 1$, one has

$$\text{(4.1.32)}\qquad Q\Big(A\cap\{N<\infty\}\cap\bigcap_{j=1}^k\{X_{N+j}\in B_j\}\Big) = Q(A\cap\{N<\infty\})\prod_{j=1}^k \mu(B_j)\,.$$

For $n\in\mathbb{N}$, one has

$$\text{(4.1.33)}\qquad \begin{aligned} & P\big[\underbrace{A\cap\{N=n\}}_{\in\mathcal{F}_n}\overset{\text{Notation for “}\bigcap\text{”}}{\overset{\downarrow}{,}}\ X_{n+j}\in B_j,\ 1\le j\le k\big] \\ \overset{\text{independence}}{=}\ & P\big[A\cap\{N=n\}\big]\,P\big[X_{n+j}\in B_j,\ 1\le j\le k\big] \\ =\ & P\big[A\cap\{N=n\}\big]\prod_{j=1}^k \mu(B_j)\,. \end{aligned}$$

By summing over n, we obtain

$$\begin{aligned} & P\Big[A\cap\Big(\bigcup_{n\ge 0}\{N=n\}\Big),\ X_{N+j}\in B_j,\ 1\le j\le k\Big] \\ =\ & \sum_{n\ge 0} P\big[A\cap\{N=n\},\ X_{N+j}\in B_j,\ 1\le j\le k\big] \\ =\ & \sum_{n\ge 0} P\big[A\cap\{N=n\},\ X_{n+j}\in B_j,\ 1\le j\le k\big] \\ \overset{(4.1.33)}{=}\ & \sum_{n\ge 0} P\big[A\cap\{N=n\}\big]\prod_{j=1}^k\mu(B_j) = P\big[A\cap\{N<\infty\}\big]\prod_{j=1}^k\mu(B_j)\,. \end{aligned}$$

This proves (4.1.32). □

We now prove a useful identity, which is specific to random walks.

Theorem 4.10. *(Wald's identity).*

If $E[|X_1|] < \infty$, and N is an $(\mathcal{F}_n)$-stopping time with $E[N] < \infty$, then

$$E\big[|S_N|\big] < \infty\,, \quad \text{and} \quad E[S_N] = E[X_1]\,E[N]\,. \tag{4.1.34}$$

Proof. This follows from the optional stopping theorem, since $(S_n - nE[X_1])$, $n \geq 0$, is an $(\mathcal{F}_n)$-martingale. This uses the fact that $(\mathcal{F}_n)$ is the filtration generated by the independent random variables X_i, $i \geq 1$: X_{n+1} is thus independent of $\mathcal{F}_n$, which implies $E[X_{n+1}|\mathcal{F}_n] = E[X_{n+1}] = E[X_1]$. $\square$

Note that in the previous proof, it is important that N is a stopping time for the natural filtration $(\mathcal{F}_n)$ (generated by the X_i). Wald's identity does not necessarily hold true if N is a stopping time with respect to a larger filtration (the key property that X_{n+1} is independent of $\mathcal{F}_n$ would be sufficient).

Example 4.11.

1) Let S_n be the simple random walk, and define $N = \inf\{n \geq 0, S_n = 1\}$.

We know that $N < \infty$ P-a.s., but one also has

$$E[N] = \infty\,! \tag{4.1.35}$$

Otherwise, we could write

$$1 = E[S_N] = E[X_1] \cdot E[N] = 0 \cdot E[N] = 0\,,$$

which is obviously a contradiction.

2) Let S_n be the simple random walk. Let us also consider $a, b \in \mathbb{Z}$ with $a < 0 < b$, and

$$N = \inf\{n \geq 0, S_n \notin (a,b)\} = \inf\{n \geq 0, S_n = a \text{ or } b\}\,.$$

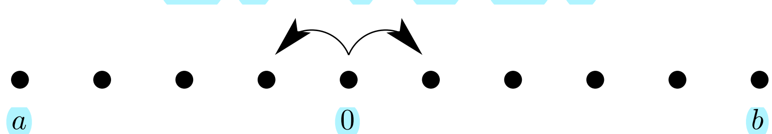


Fig. 4.4

Then $E[N] < \infty$, since for $k \geq 0$, one has

$$\{X_{k+1} = 1, \dots, X_{k+b-a} = 1\} \subseteq \{N \leq k + (b-a)\}$$
"$b - a$ successive increments $= 1 \Longrightarrow$ the random walk exits from (a,b)" .

Hence,

$$\big\{N > n(b-a)\big\} \subseteq \bigcap_{\ell=0}^{n-1} \big\{X_{\ell(b-a)+1} = 1, \dots, X_{(\ell+1)(b-a)} = 1\big\}^c,$$

and

$$\begin{aligned}
& P[N > n(b-a)] \\
(4.1.36) \qquad & \overset{\text{independence}}{\le} \prod_{\ell=0}^{n-1} \Big(1 - P\big[X_{\ell(b-a)+1} = 1, \dots, X_{(\ell+1)(b-a)} = 1\big]\Big) \\
& = \big(1 - 2^{-(b-a)}\big)^n
\end{aligned}$$

(note that since the trials are independent, the geometric distribution arises). This implies

$$E[N] = \sum_{k=1}^{\infty} P[N \ge k] < \infty .$$

Thanks to (4.1.34), one has

$$E[S_N] = E[X_1]\, E[N] = 0 \,, \tag{4.1.37}$$

and one can also write

$$E[S_N] = aP[S_N = a] + bP[S_N = b] = aP[S_N = a] + b(1 - P[S_N = a]) \,.$$

Hence,

$$P[S_N = a] = \frac{b}{b-a}, \quad P[S_N = b] = \frac{-a}{b-a} \tag{4.1.38}$$

(which can also be obtained directly via martingale techniques).

3) Renewal process (see Section 1.5)

We consider $X_i > 0$ with $E[X_i] < \infty$, $T_n = X_1 + \dots + X_n$ "arrival time of the nth customer" (with the convention $T_0 = 0$), and $N_t = \sup\{n \ge 0,\ T_n \le t\}$ "number of customers up to time t (≥ 0)".

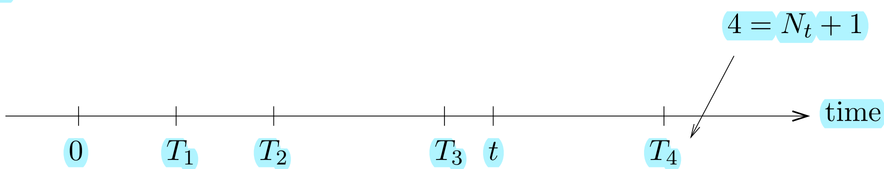


Fig. 4.5

$S \overset{\text{def.}}{=} \inf\{n \ge 0,\ T_n > t\} = N_t + 1$ is a stopping time since

$$\{S = 0\} = \emptyset, \quad \text{and} \quad \{S = n\} = \{T_1 \le t, \dots, T_{n-1} \le t,\ T_n > t\} \in \mathcal{F}_n \,.$$

Analogously to (4.1.36), one can show that $E[S] < \infty$, and one has the identity

$$E[T_{N_t+1}] = E[X_1]\, \big(E[N_t] + 1\big) \,. \tag{4.1.39}$$

$\bigcirc$

4.1.3 Recurrence *vs* transience

We consider a random walk $S_n, n \geq 0$, on $\mathbb{R}^d$, $d \geq 1$:

$$S_0 = 0\,, \quad S_n = X_1 + \cdots + X_n\,,$$

where the $X_i, i \geq 1$, are independent with distribution μ on $\mathbb{R}^d$.

Thanks to the Hewitt-Savage 0-1 law, we know that for $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\limsup_{n \geq 1} \{S_n \in B\} = \big\{\omega \in \Omega : \; S_n(\omega) \in B \;\text{ infinitely often}\big\} \tag{4.1.40}$$

has either probability 0 or 1.

If $P\big[\limsup_{n\geq 1}\{S_n \in B\}\big] = 1$, then "the random walk visits the set B infinitely often" P-a.s., which brings us to the following definition. We use the notation: for $x \in \mathbb{R}^d$, $\|x\| = \sup_{1 \leq i \leq d} |x^i|$.

Definition 4.12.

• $x \in \mathbb{R}^d$ *is called a recurrence value of the random walk* $S_n, n \geq 0$, *if*

$$\forall \epsilon > 0, \quad P\Big[\limsup_{n \geq 1} \big\{\|S_n - x\| < \epsilon\big\}\Big] = 1\,. \tag{4.1.41}$$

Notation: $\mathcal{R} =$ *set of all recurrence values.*

• $x \in \mathbb{R}^d$ *is called a possible value of the random walk* $S_n, n \geq 0$, *if*

$$\forall \epsilon > 0, \; \exists n \geq 0 \;\; \textit{with} \;\; P\big[\|S_n - x\| < \epsilon\big] > 0\,. \tag{4.1.42}$$

Notation: $\mathcal{M} =$ *set of all possible values.*

Example 4.13. Let us consider the simple random walk on $\mathbb{Z}^d$, obtained for

$$\mu(dy) = \frac{1}{2d} \sum_{\substack{e \in \mathbb{Z}^d \\ |e|=1}} \delta_e(dy).$$

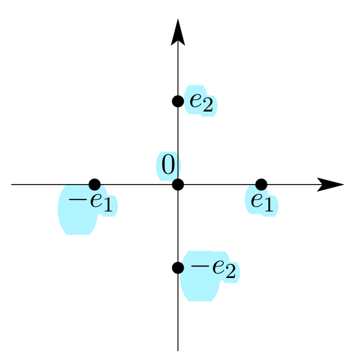
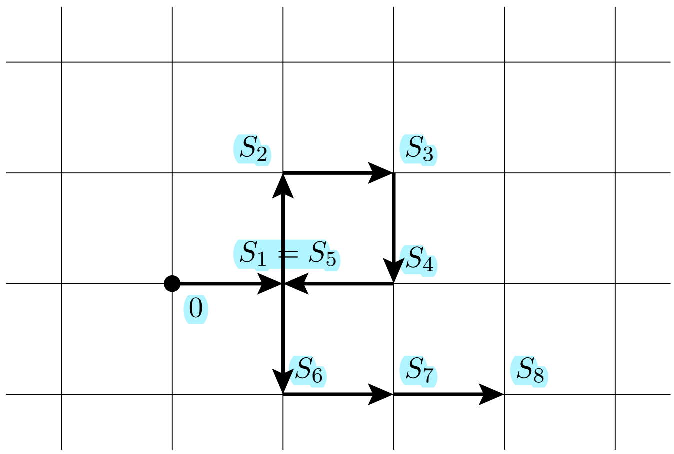


Fig. 4.6

If at time n, the random walk is at position S_n, then it jumps at time $n+1$ to one of the $2d$ neighboring sites.



a possible path of $S_n, 0 \le n \le 8$

Fig. 4.7

Note that each $x \in \mathbb{Z}^d$ is a possible value, and that $S_n \in \mathbb{Z}^d$ P-a.s., for all $n \ge 0$. Hence, $\mathcal{M} = \mathbb{Z}^d$. $\bigcirc$

The set $\mathcal{R}$ of recurrence values is somewhat more complicated to determine. At this point, we know that in the case of dimension $d = 1$, one has $\mathcal{R} = \mathbb{Z}$ since

$$P\Big[\underline{\lim_{n\to\infty}}\, S_n = -\infty,\ \overline{\lim_{n\to\infty}}\, S_n = +\infty\Big] = 1$$

(see Example 1) after the Hewitt-Savage 0-1 law, Section 3.2).

We start with a theorem that describes the structure of the set $\mathcal{R}$ of recurrence values. Note that $\mathcal{R} \subseteq \mathcal{M}$ (the set of possible values).

Theorem 4.14. *One has either*

(4.1.43)
$$\begin{aligned} &\text{i)} \quad \mathcal{R} = \emptyset, \\ \textit{or} \quad &\text{ii)} \quad \mathcal{R} \textit{ is a closed subgroup of } \mathbb{R}^d . \end{aligned}$$

Moreover, in case ii), *one has* $\mathcal{R} = \mathcal{M}$ *(set of possible values).*

Proof. $\mathcal{R}$ is closed, since $x \notin \mathcal{R} \Longrightarrow \exists \epsilon_0 > 0$ such that

$$P\big[\limsup \{\|S_n - x\| < \epsilon_0\}\big] = 0 ,$$

and so $B(x, \epsilon_0) \subseteq \mathcal{R}^c$.

We now assume that $\mathcal{R} \neq \emptyset$, and we show

(4.1.44)
$$x \in \mathcal{R},\ \ y \in \mathcal{M} \Longrightarrow x - y \in \mathcal{R} .$$

Our result (4.1.43) ii) will then follow, since (i) for $x = y \in \mathcal{R}$, one has $x - y = 0 \in \mathcal{R}$, (ii) $0 - y = -y \in \mathcal{R}$ (i.e. the inverse stays in $\mathcal{R}$), and (iii) for $x, y \in \mathcal{R}$, one also has $\underbrace{-x}_{\in \mathcal{R}} - y = -(x+y) \in \mathcal{R} \Longrightarrow x + y \in \mathcal{R}$.

Consequently, $\mathcal{R}$ is a closed subgroup of $\mathbb{R}^d$. In addition, for $x = 0 \in \mathcal{R}$, $y \in \mathcal{M}$, one obtains $-y \in \mathcal{R}$, and so $y \in \mathcal{R}$. This implies $\mathcal{M} = \mathcal{R}$, which completes the proof. □

Proof of (4.1.44). Consider $\epsilon > 0$, $x \in \mathcal{R}$, $y \in \mathcal{M}$. Then one has, for an $n_0 \geq 0$,

$$P\big[\|S_{n_0} - y\| < \epsilon\big] > 0\,.$$

Since $x \in \mathcal{R}$, one also has, for $m \geq 0$,

$$\begin{aligned} 0 &= P\left[\|S_n - x\| > \epsilon,\ \forall n \geq n_0 + m\right] \\ &\geq P\left[\{\|S_{n_0} - y\| < \epsilon\} \cap \{\|S_{n_0+k} - S_{n_0} - (x-y)\| \geq 2\epsilon,\ \forall k \geq m\}\right] \\ &\overset{(4.1.31)}{=} P\left[\|S_{n_0} - y\| < \epsilon\right] \cdot P\big[\|S_k - (x-y)\| \geq 2\epsilon,\ \forall k \geq m\big]\,. \end{aligned}$$

Hence, for $m \geq 0$,

$$P\left[\|S_k - (x-y)\| \geq 2\epsilon,\ \forall k \geq m\right] = 0\,,$$

and so

$$P\left[\limsup_k \big\{\|S_k - (x-y)\| < 2\epsilon\big\}\right] = 1\,,$$

where $\epsilon > 0$ is arbitrary. Our claim (4.1.44) follows. □

Remark 4.15.

- A closed subgroup of $\mathbb{R}$ is either equal to $\mathbb{R}$, or to $a\,\mathbb{Z} = \{a \cdot k; k \in \mathbb{Z}\}$ for some $a \geq 0$.
- A closed subgroup of $\mathbb{R}^d$ is "of the form $\mathbb{R}^p \times \mathbb{Z}^q$ with $p + q \leq d$", i.e.

$$\Big\{\sum_{i=1}^d t_i\, e_i;\ \text{ where } t_i \in \mathbb{R},\ 1 \leq i \leq p,\ t_i \in \mathbb{Z},\ p+1 \leq i \leq p+q\Big\},$$

  for some basis $(e_i)_{1 \leq i \leq d}$ of $\mathbb{R}^d$ (see Bourbaki, Topologie Générale VII, 2).

Terminology:

- In the case (4.1.43) i) (i.e. $\mathcal{R} = \emptyset$), the random walk is called **transient**.
- In the case (4.1.43) ii), the random walk is called **recurrent**. □

Now, we would like to present a concrete criterion to decide whether a random walk is recurrent or transient.

Theorem 4.16. *Let $S_n, n \geq 0$, be a random walk on $\mathbb{R}^d$. One has either*

$$\text{(4.1.45)} \qquad \begin{aligned} & \text{i)} \quad \sum_{n\geq 0} P\big[\|S_n\| < \epsilon\big] < \infty, \quad \forall \epsilon > 0\,, \\ \textit{or} \quad & \text{ii)} \quad \sum_{n\geq 0} P\big[\|S_n\| < \epsilon\big] = \infty, \quad \forall \epsilon > 0 \end{aligned}$$

(with $\|\cdot\|$ the sup-norm).

(4.1.46) *The random walk is transient in case* i)*, and it is recurrent in case* ii)*.*

Proof. We first show (4.1.45).

It is enough to prove that for any $\epsilon_0 > 0$,

$$\text{(4.1.47)} \qquad \sum_{n\geq 0} P\big[\|S_n\| < \epsilon_0\big] < \infty \implies \forall M \geq 1,\ \sum_{n\geq 0} P\big[\|S_n\| < M\epsilon_0\big] < \infty\,.$$

Indeed, since the function ϕ: $\epsilon > 0 \mapsto \sum_{n\geq 0} P\big[\|S_n\| < \epsilon\big] \in [0,\infty]$ is non-decreasing, it follows from (4.1.47) that either $\phi(\epsilon) < \infty$, $\forall \epsilon > 0$, or $\phi(\epsilon) = \infty$, $\forall \epsilon > 0$. □

Our claim (4.1.47) is a consequence of the following lemma.

Lemma 4.17. *Consider $\epsilon > 0$ and $M \geq 2$, then*

$$\text{(4.1.48)} \qquad \sum_{n=0}^{\infty} P\big[\|S_n\| < M\epsilon\big] \leq (2M)^d \sum_{n=0}^{\infty} P\big[\|S_n\| < \epsilon\big]\,.$$

Proof.

$$\text{(4.1.49)} \qquad \sum_{n=0}^{\infty} P\big[\|S_n\| < M\epsilon\big] \leq \sum_{n=0}^{\infty} \sum_{k\in\{-M,\dots,M-1\}^d} P\big[S_n \in k\epsilon + [0,\epsilon)^d\big]\,.$$

Let us fix some $k \in \{-M,\dots,M-1\}^d$. Defining the stopping time $T_k = \inf\big\{m \geq 0,\ S_m \in k\epsilon + [0,\epsilon)^d\big\}$, we obtain

$$\text{(4.1.50)} \qquad \begin{aligned} \sum_{n=0}^{\infty} P\big[S_n \in k\epsilon + [0,\epsilon)^d\big] &= \sum_{n=0}^{\infty}\sum_{m=0}^{n} P\big[S_n \in k\epsilon + [0,\epsilon)^d,\ T_k = m\big] \\ &\overset{\text{Fubini}}{=} \sum_{m=0}^{\infty}\sum_{n\geq m} P\big[\underbrace{S_n \in k\epsilon + [0,\epsilon)^d}_{\subseteq \{\|S_n - S_m\| < \epsilon\}},\ T_k = m\big] \quad \text{(on the event } \{T_k = m\}\text{)} \\ &\leq \sum_{m=0}^{\infty}\sum_{n\geq m} P\Big[\|S_n - S_m\| < \epsilon,\ \underbrace{T_k = m}_{\in\mathcal{F}_m}\Big] \\ &\overset{\text{independence}}{=} \sum_{m=0}^{\infty}\sum_{n\geq m} P\big[\|S_n - S_m\| < \epsilon\big]\cdot P[T_k = m] \\ &= \sum_{n=0}^{\infty} P\big[\|S_n\| < \epsilon\big]\cdot \underbrace{\sum_{m=0}^{\infty} P[T_k = m]}_{\leq 1} \leq \sum_{n=0}^{\infty} P\big[\|S_n\| < \epsilon\big]\,. \end{aligned}$$

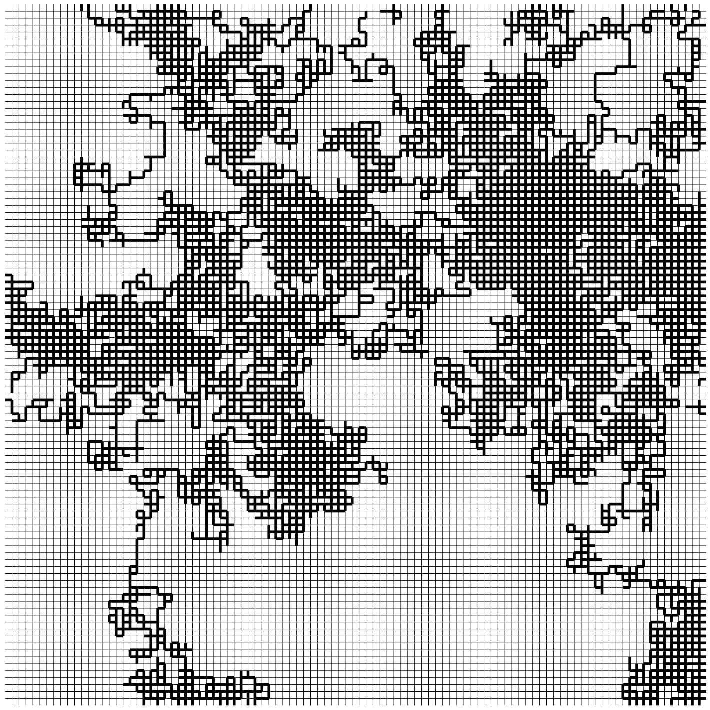


Fig. 4.8: Path of a symmetric random walk
(starting point in the middle of the picture, 10^7 steps)

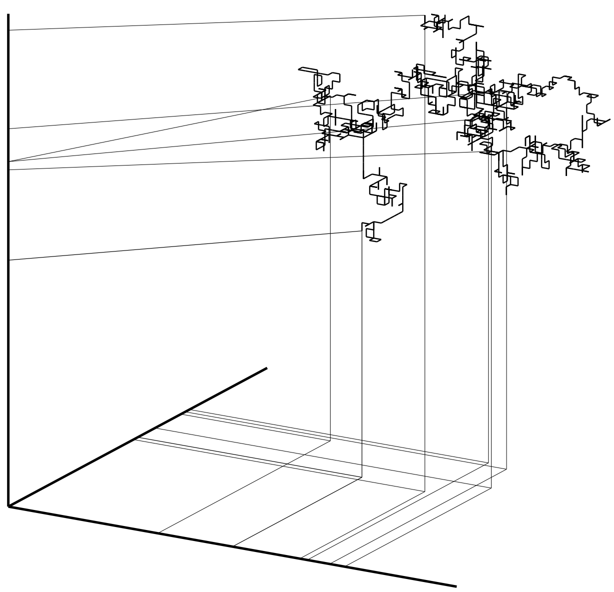


Fig. 4.9: Path of a three-dimensional symmetric random walk

It follows from (4.1.49) - (4.1.50) that

$$\sum_{n=0}^{\infty} P\big[\|S_n\| < M\epsilon\big] \leq \Big|\underbrace{\{-M,\dots,M-1\}^d}_{=(2M)^d}\Big| \sum_{n=0}^{\infty} P\big[\|S_n\| < \epsilon\big]\,.$$

□

We now show (4.1.46).

In case (4.1.45) i), one has $P\big[\limsup\{\|S_n\| < \epsilon\}\big] = 0$ for all $\epsilon > 0$ (using the first Borel-Cantelli lemma), which implies $0 \notin \mathcal{R}$, and thus, using (4.1.43), that $\mathcal{R} = \emptyset$: the random walk is transient.

There remains to prove that in case (4.1.45) ii), the random walk is recurrent, which will come from the following lemma.

Lemma 4.18. *Let* $\epsilon > 0$,

$$\sum_{n\geq 0} P\big[\|S_n\| < \epsilon\big] = \infty \implies P\big[\limsup\big\{\|S_n\| < 2\epsilon\big\}\big] = 1\,. \tag{4.1.51}$$

Proof. One has

$$\liminf\big\{\|S_n\| \geq \epsilon\big\} = \bigcup_{m=0}^{\infty} \underbrace{\big\{\|S_m\| < \epsilon,\ \|S_n\| \geq \epsilon,\ \forall n \geq m+1\big\}}_{\text{pairwise disjoint events}}$$

"m is the last time at which $\|S_n\| < \epsilon$ holds". Hence,

$$\begin{aligned}
1 &\geq \sum_{m=0}^{\infty} P\big[\|S_m\| < \epsilon,\ \|S_n\| \geq \epsilon,\ \ \forall n \geq m+1\big] \\
&\geq \sum_{m=0}^{\infty} P\big[\underbrace{\{\|S_m\| < \epsilon\}}_{\text{independent}} \cap \big\{\underbrace{\|S_n - S_m\| \geq 2\epsilon,\ \forall n \geq m+1}_{\text{independent}}\big\}\big] \\
&= \underbrace{\sum_{m=0}^{\infty} P\big[\|S_m\| < \epsilon\big]}_{=\,\infty \text{ by assumption}} \cdot P\big[\|S_\ell\| \geq 2\epsilon,\ \ \forall \ell \geq 1\big]\,.
\end{aligned} \tag{4.1.52}$$

Hence, $P[\|S_\ell\| \geq 2\epsilon,\ \forall \ell \geq 1] = 0$. In the same way, define for $k \geq 2$, $m \geq 0$,

$$A_m = \big\{\|S_m\| < \epsilon,\ \|S_n\| \geq \epsilon,\ \ \forall n \geq m+k\big\}\,.$$

For $\omega \in A_m$, the last time $m_0(\omega)$ at which $\|S_m\| < \epsilon$ is finite, and

$$\begin{aligned}
m \leq m_0(\omega) < m+k &\implies m_0(\omega) - k < m \leq m_0(\omega) \\
&\implies k \geq \sum_{m=0}^{\infty} 1_{A_m}(\omega),\ \ \forall \omega \in \Omega\,.
\end{aligned}$$

We thus obtain, analogously to (4.1.52),

$$k \geq \sum_{m=0}^{\infty} P[A_m] \geq \underbrace{\sum_{m=0}^{\infty} P\big[\|S_m\| < \epsilon\big]}_{=\infty} \cdot P\big[\|S_\ell\| \geq 2\epsilon,\ \forall \ell \geq k\big]\,. \tag{4.1.53}$$

It follows that $P\big[\|S_\ell\| \geq 2\epsilon,\ \forall \ell \geq k\big] = 0$. Hence, $P\big[\liminf_\ell \|S_\ell\| \geq 2\epsilon\big] = 0$, from which (4.1.51) follows. $\square$

Special case:

(4.1.54) the distribution μ of the $X_i, i \geq 1$, is concentrated on $\mathbb{Z}^d$, i.e. $\mu(\mathbb{Z}^d) = 1$.

In this case, $\mathcal{M} \subseteq \mathbb{Z}^d$. One has either

$$\infty > \sum_{n\geq 0} P\big[\|S_n\| < 1\big] \overset{(4.1.54)}{=} \sum_{n\geq 0} P[S_n = 0]$$

and the random walk is transient, or

$$\infty = \sum_{n\geq 0} P\big[\|S_n\| < 1\big] = \sum_{n\geq 0} P[S_n = 0]$$

and the random walk is recurrent.

In other words: from the convergence (resp. divergence) of the series $\sum_{n\geq 0} P[S_n = 0]$, one can deduce the transience (resp. the recurrence) of S_n.

The following lemma provides a convenient way to estimate $P[S_n = 0]$.

Lemma 4.19. *Let μ be given satisfying* (4.1.54)*, and let φ be the characteristic function of μ:*

$$\varphi(t) = \int_{\mathbb{Z}^d} \exp\{it\cdot x\}\, \mu(dx), \quad t \in \mathbb{R}^d\,. \tag{4.1.55}$$

Then,

$$P[S_n = 0] = \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \varphi(t)^n dt, \quad n \geq 1\,. \tag{4.1.56}$$

Proof. Note that the characteristic function of S_n is equal to $\varphi(t)^n$. It is thus enough to show (4.1.56) for $n = 1$. Now,

$$\begin{aligned}
\frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \varphi(t) dt &= \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \int_{\mathbb{Z}^d} e^{it\cdot x}\, \mu(dx)\, dt \\
&\overset{\text{Fubini}}{=} \int_{\mathbb{Z}^d} \Big(\frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} e^{it\cdot x}\, dt\Big)\, \mu(dx)\,.
\end{aligned}$$

Moreover, for $k \in \mathbb{Z}$,

$$\frac{1}{2\pi} \int_{(-\pi,\pi)} e^{it\cdot k}\, dt = \begin{cases} 1 & \text{for } k = 0\,, \\ \dfrac{(e^{i\pi k} - e^{-i\pi k})}{ik} = 0 & \text{for } k \neq 0\,. \end{cases}$$

Hence,

$$\frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} e^{it\cdot x}\, dt = \begin{cases} 1 \;\text{ for } x = 0 \\ 0 \;\text{ for } x \in \mathbb{Z}^d \backslash \{0\}\,, \end{cases}$$

which implies

$$\frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \varphi(t)\, dt = \mu(\{0\})\,.$$

□

Example 4.20. For the simple random walk,

$$\mu(dx) = \frac{1}{2d} \sum_{\substack{e\in\mathbb{Z}^d \\ |e|=1}} \delta_e(dx)\,,$$

and

(4.1.57) $$\varphi(t) = \frac{1}{2d} \sum_{\substack{e\in\mathbb{Z}^d \\ |e|=1}} e^{it\cdot e} = \frac{1}{d} \sum_{i=1}^d \cos t_i\,, \quad \text{where } t = (t_1, \dots, t_d)\,.$$

Using (4.1.56), we obtain

$$P[S_n = 0] = \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \varphi(t)^n\, dt\,.$$

For $r \in (0,1)$, we can then consider the (summable) series

(4.1.58) $$\begin{aligned} \sum_{n\geq 0} r^n P[S_n = 0] &= \sum_{n\geq 0} \frac{1}{(2\pi)^d} r^n \int_{(-\pi,\pi)^d} \varphi(t)^n\, dt \\ &\overset{\text{Lebesgue}}{=} \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \sum_{n\geq 0} r^n \varphi(t)^n\, dt = \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \frac{1}{1 - r\varphi(t)}\, dt \\ &= \frac{1}{(2\pi)^d} \int_{(-\pi,\pi)^d} \frac{1}{1 - \frac{r}{d} \sum\limits_{i=1}^d \cos t_i}\, dt\,. \end{aligned}$$

We now let $r \uparrow 1$ in this equality. On the one hand,

$$\lim_{r\uparrow 1} \sum_{n\geq 0} r^n P[S_n = 0] \overset{\substack{\text{monotone} \\ \text{convergence}}}{=} \underbrace{\sum_{n\geq 0} P[S_n = 0]}_{\uparrow \text{ possibly } +\infty}\,.$$

On the other hand, let us divide the integral on the right-hand side into two parts: one has

$$\lim_{r\uparrow 1}\int_{(-\pi,\pi)^d\setminus(-\frac{\pi}{2},\frac{\pi}{2})^d}\underbrace{\frac{1}{1-\frac{r}{d}\sum_{i=1}^d \cos t_i}}_{\substack{\uparrow\\ \cos t_i\,\le\, 0 \text{ for at least one } i}}\,dt$$

$$\overset{\substack{\text{dominated}\\ \text{convergence}}}{=}\int_{(-\pi,\pi)^d\setminus(-\frac{\pi}{2},\frac{\pi}{2})^d}\frac{1}{1-\frac{1}{d}\sum_{i=1}^d \cos t_i}\,dt$$

(the integrand can be dominated by $\frac{1}{1-\frac{d-1}{d}} = d$), and the monotone convergence theorem implies that

$$\lim_{r\uparrow 1}\int_{(-\frac{\pi}{2},\frac{\pi}{2})^d}\frac{1}{1-\frac{r}{d}\sum_{i=1}^d \cos t_i}\,dt = \underbrace{\int_{(-\frac{\pi}{2},\frac{\pi}{2})^d}\frac{1}{1-\frac{1}{d}\sum_{i=1}^d \cos t_i}\,dt}_{\substack{\uparrow\\ \text{possibly } +\infty}}$$

(the integrand is non-decreasing in r since all $\cos t_i \ge 0$). We thus obtain the identity

$$\sum_{n\ge 0} P[S_n = 0] = \frac{1}{(2\pi)^d}\int_{(-\pi,\pi)^d}\frac{1}{1-\frac{1}{d}\sum_{i=1}^d \cos t_i}\,dt\,, \tag{4.1.59}$$

where both sides can be equal to $+\infty$. Now, for $t \to 0$,

$$1-\frac{1}{d}\sum_{i=1}^d \cos t_i \sim \frac{1}{2d}\,|t|^2 = \frac{1}{2d}\,(t_1^2+\cdots+t_d^2)\,.$$

As a consequence, the integral in (4.1.59) converges for $d \ge 3$, and it diverges for $d \le 2$. In other words:

(4.1.60) the **simple random walk** is **recurrent** for $d = 1, 2$,
transient for $d \ge 3$.

○

Further applications of (4.1.45) - (4.1.46) can be found in the exercises.

Remark 4.21. As a conclusion, let us mention a final (somewhat surprising) example. We know that in dimension one, for the symmetric stable distribution μ with parameter $\alpha \in (0,2)$ (i.e. $\varphi(t) = \exp\{-c\,|t|^\alpha\}$, $c > 0$ fixed), one has, for symmetry reasons (see (4.1.19)): P-a.s., $\overline{\lim}\, S_n = +\infty$ and $\underline{\lim}\, S_n = -\infty$.

Nevertheless, one can show that **for $\boldsymbol{\alpha < 1}$, the random walk is transient on $\mathbb{R}$!** "The random walk passes over 0 by making large jumps". □

4.2 Markov chains: an introduction

Markov chains are stochastic processes $X_0, \dots, X_n, \dots$ taking values in a state space S (equipped with a σ-algebra $\mathcal{S}$) such that: for each n, the best forecast of the future after time n, given the past up to this time n, depends only on the information contained in X_n. More precisely, one has:

$$E\big[f(X_n, \dots, X_{n+k}) \,|\, \mathcal{F}_n\big] = E\big[f(X_n, \dots, X_{n+k}) \,|\, \sigma(X_n)\big] \quad P\text{-a.s.},$$

for all $n, k \geq 0$, and all bounded measurable functions $f : S^{k+1} \to \mathbb{R}$, where $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ contains the information on the past up to time n.

4.2.1 Stochastic kernels

Stochastic kernels play an important role in the construction of canonical Markov chains, and also of general canonical stochastic processes.

Definition 4.22. *Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be two measurable spaces. A stochastic kernel K from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ is a map from $\Omega_1 \times \mathcal{A}_2$ to $[0,1]$ such that*

(4.2.1) *for all $A \in \mathcal{A}_2$, $\omega_1 \in \Omega_1 \mapsto K(\omega_1, A)$ is $\mathcal{A}_1$-measurable,*

(4.2.2) *for all $\omega_1 \in \Omega_1$, the map $A \in \mathcal{A}_2 \mapsto K(\omega_1, A) \in [0,1]$ is a probability measure on $(\Omega_2, \mathcal{A}_2)$.*

Notation: $K(\omega_1, d\omega_2)$ *(due to* (4.2.2)*).*

Example 4.23.

1) $\Omega_1 = \Omega_2 = \mathbb{Z}$, $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{P}(\mathbb{Z})$ (the power set of $\mathbb{Z}$). Then

$$K(x, dy) = \frac{1}{2}\big(\delta_{x+1}(dy) + \delta_{x-1}(dy)\big), \text{ for } x \in \mathbb{Z}\,, \tag{4.2.3}$$

is a stochastic kernel from $\mathbb{Z}$ to $\mathbb{Z}$. This kernel is directly related to the simple random walk on $\mathbb{Z}$.

2) $\Omega_1 = \Omega_2 = \mathbb{N}$ $(= \{0, 1, \dots\})$, $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{P}(\mathbb{N})$, ν a distribution on $\mathbb{N}$. Then

$$\begin{aligned} K(x, dy) &= \delta_0(dy) \quad \text{if } x = 0, \\ &= P[\xi_1 + \dots + \xi_x \in dy] \quad \text{if } x \geq 1\,, \end{aligned} \tag{4.2.4}$$

where $\xi_i, i \geq 1$, are i.i.d. random variables with distribution ν, is a stochastic kernel from $\mathbb{N}$ to $\mathbb{N}$. This kernel is directly related to the Galton-Watson chain, see (3.5.1).

3) $(\Omega_i, \mathcal{A}_i)$ measurable spaces $(i = 1, 2)$, ν a probability measure on $(\Omega_2, \mathcal{A}_2)$, and $f(\omega_1, \omega_2) \geq 0$ a measurable function on $\Omega_1 \times \Omega_2$, with $\int f(\omega_1, \omega_2)\, d\nu(\omega_2) = 1$ for all $\omega_1 \in \Omega_1$. Then

$$K(\omega_1, d\omega_2) = f(\omega_1, \omega_2)\, \nu(d\omega_2), \quad \omega_1 \in \Omega_1 \tag{4.2.5}$$

(i.e. $K(\omega_1, A) = \int_A f(\omega_1, \omega_2)\, d\nu(\omega_2)$ for $A \in \mathcal{A}_2$) is a stochastic kernel from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$. ○

Proposition 4.24. *Let P_1 be a probability measure on $(\Omega_1, \mathcal{A}_1)$, and K a stochastic kernel from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$. Then, there exists a unique probability measure P on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ such that*

$$P[A_1 \times A_2] = \int_{A_1} K(\omega_1, A_2)\, P_1(d\omega_1) \quad \text{for all } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\,. \tag{4.2.6}$$

Furthermore, a measurable function $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is P-integrable if and only if

$$\int_{\Omega_1} \Big(\int_{\Omega_2} |f(\omega_1, \omega_2)|\, K(\omega_1, d\omega_2) \Big)\, P_1(d\omega_1) < \infty\,, \tag{4.2.7}$$

and in this case,

$$\int_{\Omega_1 \times \Omega_2} f\, dP = \int_{\Omega_1} \Big(\int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \Big)\, P_1(d\omega_1)\,. \tag{4.2.8}$$

Notation: $P = P_1 \times K$, P is called semi-product of P_1 and K.

Proof.

- The uniqueness of P is clear, as the collection $\{A_1 \times A_2; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ is a π-system that generates $\mathcal{A}_1 \times \mathcal{A}_2$ (see Dynkin's lemma, (1.3.9)).

• *Existence:* Using Dynkin's lemma, one sees that

$$\text{for all } A \in \mathcal{A}_1 \otimes \mathcal{A}_2, \quad \omega_1 \mapsto \int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \text{ is } \mathcal{A}_1\text{-measurable}\,. \tag{4.2.9}$$

If we set

$$P[A] = \int_{\Omega_1} \Big(\int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \Big)\, P_1(d\omega_1) \text{ for } A \in \mathcal{A}_1 \times \mathcal{A}_2\,, \tag{4.2.10}$$

then

$$P[\Omega_1 \times \Omega_2] = 1\,,$$

and for a sequence of pairwise disjoint $A_i, i \geq 1$, with $A_i \in \mathcal{A}_1 \otimes \mathcal{A}_2$ for all $i \geq 1$, we obtain

$$\begin{aligned}
P\Big(\bigcup_{i \geq 1} A_i \Big) &= \int_{\Omega_1} \Big(\int_{\Omega_2} \sum_{i \geq 1} 1_{A_i}(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \Big)\, P_1(d\omega_1) \\
&\overset{\sigma\text{-additivity}}{=} \int_{\Omega_1} \sum_{i \geq 1} \Big(\int_{\Omega_2} 1_{A_i}(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \Big)\, P_1(d\omega_1) \\
&\overset{\substack{\text{monotone} \\ \text{convergence}}}{=} \sum_{i \geq 1} \int_{\Omega_1} \Big(\int_{\Omega_2} 1_{A_i}(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \Big)\, P_1(d\omega_1) \\
&= \sum_{i \geq 1} P(A_i)\,.
\end{aligned}$$

In other words, P is a probability measure on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$. Using an approximation by a non-decreasing sequence of step functions, we can then obtain the following generalization of (4.2.10):

$$\begin{aligned}
&\int_{\Omega_1 \times \Omega_2} f\, dP = \int_{\Omega_1} \Big(\int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \Big)\, dP_1(\omega_1) \\
&\qquad\qquad \text{for all } f \geq 0 \text{ measurable on } \Omega_1 \times \Omega_2\,.
\end{aligned} \tag{4.2.11}$$

The remaining claims of the proposition then follow as in the proof of (1.2.16), (1.2.17). □

Remark 4.25. In the case when $K(\omega_1, d\omega_2) = P_2(d\omega_2)$ (i.e. the kernel K does not depend on ω_1), the measure P in the previous proposition (see (4.2.6)) is simply the product measure $P_1 \times P_2$.

Stochastic kernels are also useful to compute conditional expectations, as we now explain. Let Z_1, Z_2 be measurable maps on $(\Omega, \mathcal{A}, P)$:

$$Z_1 : (\Omega, \mathcal{A}) \to (S_1, \mathcal{S}_1), \quad Z_2 : (\Omega, \mathcal{A}) \to (S_2, \mathcal{S}_2)\,. \tag{4.2.12}$$

We denote by P_1 the distribution of Z_1 on $(S_1, \mathcal{S}_1)$ under P, and we assume that the distribution Q of $Z = (Z_1, Z_2)$ on $(S = S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$ is of the form

$$Q = P_1 \times K\,, \tag{4.2.13}$$

where K is a stochastic kernel from $(S_1, \mathcal{S}_1)$ to $(S_2, \mathcal{S}_2)$.

Proposition 4.26. *(under* (4.2.12), (4.2.13)*) Let $f : (S, \mathcal{S}_1 \otimes \mathcal{S}_2) \to \mathbb{R}$ be a measurable function such that $f(Z_1, Z_2) \in L^1(\Omega, \mathcal{A}, P)$. Then,*

$$E^P\big[f(Z_1, Z_2) \,|\, \sigma(Z_1)\big] \overset{P\text{-a.s.}}{=} \int_{S_2} f(Z_1(\omega), s_2)\, K(Z_1(\omega), ds_2)\,. \tag{4.2.14}$$

Proof. Thanks to (1.2.16) and (4.2.7), one has $f \in L^1(S, Q)$, and

$$\int |f|\, dQ = \int_{S_1} \Big(\int_{S_2} |f(s_1, s_2)\,|\, K(s_1, ds_2) \Big)\, P_1(ds_1) < \infty\,.$$

Hence, $\phi : S_1 \to \mathbb{R}$ defined by

$$\begin{aligned}
s_1 \mapsto \phi(s_1) &= \int_{S_2} f(s_1, s_2)\, K(s_1, ds_2) \quad \text{if } \int_{S_2} |f(s_1, s_2)\,|\, K(s_1, ds_2) < \infty\,, \\
&= 0 \quad \text{if } \int_{S_2} |f(s_1, s_2)\,|\, K(s_1, ds_2) = \infty\,,
\end{aligned} \tag{4.2.15}$$

is an $\mathcal{S}_1$-measurable function, and $\int_{S_2} |f(s_1, s_2)| K(s_1, ds_2) = \infty$ holds only on a set with P_1-measure zero. Moreover, ϕ is P_1-integrable. Hence, $\phi(Z_1)$ is $\sigma(Z_1)$-measurable, and P-integrable. We also know that $\sigma(Z_1) = \{Z_1^{-1}(B);\, B \in \mathcal{S}_1\}$, and for $B \in \mathcal{S}_1$, we obtain

$$\begin{aligned}
&E^P\big[f(Z_1, Z_2)\, 1_B \circ Z_1\big] \overset{(1.2.17)}{=} \int_S f(s_1, s_2)\, 1_B(s_1)\, dQ(s_1, s_2) \\
&\underset{(4.2.8)}{\overset{(4.2.13)}{=}} \int_{S_1} 1_B(s_1) \Big(\int_{S_2} f(s_1, s_2)\, K(s_1, ds_2) \Big)\, P_1(ds_1) = \int_{S_1} 1_B\, \phi\, dP_1 \overset{(1.2.17)}{=} E^P\big[\phi(Z_1)\, 1_B \circ Z_1\big]\,.
\end{aligned}$$

This equality shows that $E^P[f(Z_1, Z_2)|\sigma(Z_1)] \overset{P\text{-a.s.}}{=} \phi(Z_1)$, and (4.2.14) follows. □

Example 4.27. Let X_1, X_2 be independent random variables with distribution μ, resp. ν. Consider a $\mu \otimes \nu$-integrable function f on $\mathbb{R}^2$, then

$$E\big[f(X_1, X_2) \,|\, \sigma(X_1)\big] = \int_{\mathbb{R}} f(X_1, y)\, d\nu(y) \quad P\text{-a.s.} \tag{4.2.16}$$

○

A further connection with conditional expectations leads to the following definition.

Definition 4.28. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $\mathcal{F}$ a sub-σ-algebra of $\mathcal{A}$. A stochastic kernel K from $(\Omega, \mathcal{F})$ to $(\Omega, \mathcal{A})$ is called a regular conditional probability of P given $\mathcal{F}$ if*

$$K(\omega, A) = E\big[1_A \,|\, \mathcal{F}\big] \quad P\text{-a.s.} \quad \text{for all } A \in \mathcal{A}\,, \tag{4.2.17}$$

and there exists $N \in \mathcal{F}$ with $P(N) = 0$ such that

$$K(\omega, F) = 1_F(\omega) \text{ for all } F \in \mathcal{F} \text{ and } \omega \in \Omega \backslash N\,. \tag{4.2.18}$$

Example 4.29. Let μ be a distribution on $\mathbb{R}$, and K a stochastic kernel from $\mathbb{R}$ to $\mathbb{R}$. We define a probability measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ by $P = \mu \times K$, and we denote by X_1, X_2 the canonical coordinates on $\mathbb{R}^2$. Define $\mathcal{F} = \sigma(X_1)$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^2)$, and let R be the stochastic kernel from $(\mathbb{R}^2, \mathcal{F})$ to $(\mathbb{R}^2, \mathcal{A})$ defined by

$$R(\omega, A) = \int 1_A(X_1(\omega), y)\; K(X_1(\omega), dy) \quad \text{for all } A \in B(\mathbb{R}^2)\,. \tag{4.2.19}$$

Using (4.2.14) (in the case $X_1 = Z_1$, $X_2 = Z_2$, $P = Q, 1_A = f$ on $\mathbb{R}^2$), R satisfies (4.2.17). Furthermore, for all $F \in \sigma(X_1)$, one has $1_F = 1_C \circ X_1$ for some $C \in B(\mathbb{R})$, and so

$$R(\omega, F) = \int 1_C \circ X_1(\omega)\, K(X_1(\omega), dy) = 1_C \circ X_1(\omega) = 1_F(\omega)$$

for $\omega \in \Omega$. Hence, (4.2.18) holds, and

$$R \text{ is a regular conditional probability of } P \text{ given } \sigma(X_1)\,. \tag{4.2.20}$$

○

Remark 4.30. In the case when Ω is a Polish space (i.e. $(\Omega, \mathcal{C})$ is a topological space, and the topology $\mathcal{C}$ is generated by a metric d such that (Ω, d) is complete and separable), $\mathcal{A}$ is the Borel σ-algebra on Ω, and $\mathcal{F}$ is generated by a countable collection, then there exists a regular conditional probability of P (on $(\Omega, \mathcal{A})$) given $\mathcal{F}$ (see Stroock: "Probability Theory: an Analytic View", p.256). □

4.2.2 Ionescu-Tulcea theorem, construction of Markov chains

We start with a general result (the Ionescu-Tulcea theorem) that can be used to construct stochastic processes $X_0, \dots, X_n, \dots$. The idea of the construction is the following.

If a kernel K_n is associated with each time $n \geq 1$, describing the distribution of X_n given $X_0, \dots, X_{n-1}$, and if an initial distribution is also chosen for X_0, then the theorem constructs a probability measure corresponding to the distribution of the $X_n, n \geq 0$.

More precisely, we consider a sequence of measurable spaces

$$(S_i, \mathcal{S}_i)_{i \ge 0}\,, \tag{4.2.21}$$

and we define the sequence of successive product spaces

$$\begin{aligned} &\Omega_0 = S_0, \;\; \Omega_1 = S_0 \times S_1, \;\; \dots \;\; \Omega_n = S_0 \times S_1 \times \dots \times S_n, \;\; \dots \\ &\mathcal{A}_0 = \mathcal{S}_0, \;\; \mathcal{A}_1 = \mathcal{S}_0 \otimes \mathcal{S}_1, \;\; \dots \;\; \mathcal{A}_n = \underbrace{\mathcal{S}_0 \otimes \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n}_{\text{product } \sigma\text{-algebra}}, \;\; \dots \end{aligned} \tag{4.2.22}$$

We are also given

- an initial distribution on S_0:

$$P_0 \text{ probability measure on } (S_0, \mathcal{S}_0)\,, \tag{4.2.23}$$

- a sequence of kernels:

$$K_n \text{ stochastic kernel from } (\Omega_{n-1}, \mathcal{A}_{n-1}) \text{ to } (S_n, \mathcal{S}_n),\; n \ge 1\,. \tag{4.2.24}$$

Furthermore, we define by induction, by means of (4.2.6),

$$\begin{aligned} &Q_0 = P_0 \text{ on } (\Omega_0 = S_0,\, \mathcal{A}_0 = \mathcal{S}_0), \quad \text{and for } n \ge 0, \\ &Q_{n+1} = Q_n \times K_{n+1} \text{ on } (\Omega_{n+1}, \mathcal{A}_{n+1})\,. \end{aligned} \tag{4.2.25}$$

We obtain in this way a sequence Q_n, $n \ge 0$, of probability measures on $(\Omega_n, \mathcal{A}_n)$, and we consider the infinite product space

$$\Omega = \prod_{i=0}^{\infty} S_i = \{\omega = (x_0, x_1, x_2, \dots): \; x_i \in S_i, \; \forall i \ge 0\}\,, \tag{4.2.26}$$

the canonical coordinates

$$X_i(\omega) = x_i \in S_i \quad \text{for } \omega = (x_0, x_1, x_2, \dots) \in \Omega,\; i \ge 0\,, \tag{4.2.27}$$

the product σ-algebra

$$\begin{aligned} \mathcal{A} &= \sigma(X_n, n \ge 0) \\ &= \sigma(A_0 \times \dots \times A_k \times S_{k+1} \times S_{k+2} \times \dots;\; k \ge 0, A_i \in \mathcal{S}_i, 0 \le i \le k) \\ &= \sigma(A \times S_{k+1} \times S_{k+2} \times \dots;\; k \ge 0, A \in \mathcal{A}_k)\,, \end{aligned} \tag{4.2.28}$$

and the projections

$$\pi_n : \Omega \to \Omega_n, \;\; \text{with } \pi_n(\omega) = (x_0, \dots, x_n) \;\text{ for } \omega = (x_0, x_1, x_2, \dots) \in \Omega,\; n \ge 0\,. \tag{4.2.29}$$

Theorem 4.31. *(Ionescu-Tulcea)*

There exists a unique probability measure Q on $(\Omega, \mathcal{A})$ such that for all $n \geq 0$,

$$\pi_n \circ Q = Q_n \quad \textit{(see (4.2.25))}. \tag{4.2.30}$$

In particular, for any bounded random variable f on $(\Omega_n, \mathcal{A}_n)$, one has

$$\begin{aligned} &E^Q[f(X_0, X_1, \dots, X_n)] \\ &= \int_{S_0} P_0(dx_0) \int_{S_1} K_1(x_0, dx_1) \dots \int_{S_n} K_n(x_0, x_1, \dots, x_{n-1}, dx_n)\, f(x_0, \dots, x_n). \end{aligned} \tag{4.2.31}$$

Proof. (4.2.31) is a direct consequence of (4.2.30), using (1.2.17), (4.2.8), and (4.2.25). Let us show (4.2.30).

Uniqueness of Q: Thanks to (4.2.30), Q is uniquely determined on

$$\mathcal{B} = \{A \times S_{k+1} \times S_{k+2} \times \dots \;:\; k \geq 0, A \in \mathcal{A}_k\}. \tag{4.2.32}$$

Since $\mathcal{B}$ is a π-system and $\sigma(\mathcal{B}) = \mathcal{A}$, Q is uniquely determined on $\mathcal{A}$ thanks to Dynkin's lemma, cf. (1.3.9).

Existence: We define Q on $\mathcal{B}$, cf. (4.2.32), by

$$Q(A \times S_{k+1} \times S_{k+2} \times \dots) = Q_k(A) \quad \text{for all } k \geq 0, A \in \mathcal{A}_k. \tag{4.2.33}$$

(4.2.34) Then Q is well-defined on $\mathcal{B}$ through (4.2.33).

Indeed, consider $0 \leq \ell \leq n$, and $A \in \mathcal{A}_\ell$, $B \in \mathcal{A}_n$, with

$$A \times S_{\ell+1} \times \dots = B \times S_{n+1} \times \dots$$

In the case when $n = \ell$, one has $A = B$ and $Q_\ell(A) = Q_n(B)$. If $n > \ell$, then

$$\begin{aligned} Q_{\ell+1}(A \times S_{\ell+1}) &\overset{(4.2.25)}{=} Q_\ell \times K_{\ell+1}(A \times S_{\ell+1}) \\ &\overset{(4.2.8)}{=} \int_A Q_\ell(dx_0 \dots dx_\ell)\, \underbrace{K_{\ell+1}(x_0, \dots, x_\ell, S_{\ell+1})}_{\overset{\shortparallel}{1}} \\ &= Q_\ell(A), \end{aligned}$$

$Q_n(A \times S_{\ell+1} \times \dots \times S_n) = Q_\ell(A)$ holds by induction, and since $A \times S_{\ell+1} \times \dots \times S_n = B$, (4.2.34) follows.

Moreover, (4.2.32) implies that

$$\mathcal{B} \text{ is an algebra on } \Omega \tag{4.2.35}$$

(i.e. $\Omega \in \mathcal{B}$, $B \in \mathcal{B} \Longrightarrow B^c \in \mathcal{B}$, and it follows from $B_1, \dots, B_n \in \mathcal{B}$ that $\bigcup_{i=1}^n B_i \in \mathcal{B}$), and one has, thanks to (4.2.33),

$$\begin{aligned} &Q(\Omega) = 1, \\ \text{and} \quad &Q(B_1 \cup B_2) = Q(B_1) + Q(B_2) \text{ when } B_1, B_2 \in \mathcal{B} \text{ with } B_1 \cap B_2 = \emptyset, \end{aligned} \tag{4.2.36}$$

since we can find $k \geq 0$, $A_1 \in \mathcal{A}_k$, $A_2 \in \mathcal{A}_k$, such that $B_i = A_i \times S_{k+1} \times S_{k+2} \times \dots$ for $i = 1, 2$, and so

$$Q(B_1 \cup B_2) \overset{(4.2.33)}{=} Q_k(A_1 \cup A_2) = Q_k(A_1) + Q_k(A_2) \overset{(4.2.33)}{=} Q(B_1) + Q(B_2)\,.$$

Carathéodory's extension theorem (see lecture notes on measure theory, or Durrett p. 400) implies the existence of a probability measure Q on $\sigma(\mathcal{B}) \overset{(4.2.28)}{=} \mathcal{A}$ as an extension of Q in (4.2.33), as soon as we show the σ-additivity of Q on the algebra, i.e.

$$\begin{aligned}&\text{If } B_i, i \geq 1, \text{ and } B \text{ are in } \mathcal{B}, \text{ with } B_i, i \geq 1, \text{ pairwise disjoint,}\\ &\text{and } B = \bigcup_{i \geq 1} B_i, \text{ then } Q(B) = \sum_{i \geq 1} Q(B_i)\,.\end{aligned} \tag{4.2.37}$$

For B, B_i, $i \geq 1$, as above, we can define $\widetilde{B}_n = B \backslash (\bigcup_{i=1}^n B_i) \in \mathcal{B}$. Then, $\widetilde{B}_n \downarrow \emptyset$, and using (4.2.36) ("additivity of Q"),

$$Q(B) = Q(\widetilde{B}_n) + \sum_{i=1}^n Q(B_i)\,.$$

Hence, (4.2.37) follows from

$$\begin{aligned}&\text{For any decreasing sequence } \widetilde{B}_n, n \geq 1, \text{ with } \widetilde{B}_n \in \mathcal{B},\\ &\text{and } \bigcap_{n \geq 1} \widetilde{B}_n = \emptyset, \text{ one has } \lim_n Q(\widetilde{B}_n) = 0\,.\end{aligned} \tag{4.2.38}$$

Consider $\widetilde{B}_n$, $n \geq 1$, as in (4.2.38), then we can construct a sequence

$$\overline{B}_k = A_k \times S_{k+1} \times \dots,\ k \geq 0, \text{ with } A_k \in \mathcal{A}_k, \text{ and } \overline{B}_k \downarrow \emptyset\,, \tag{4.2.39}$$

such that $\widetilde{B}_n = \overline{B}_{k(n)}$ is a sub-sequence of $\overline{B}_k, k \geq 0$. The claim (4.2.38) follows if we show that

$$\lim_k Q(\overline{B}_k) = 0\,. \tag{4.2.40}$$

We show (4.2.40) by contradiction, assuming that

$$\inf_{k \geq 0} Q(\overline{B}_k) > \varepsilon > 0\,. \tag{4.2.41}$$

Because of (4.2.39), one has

$$\begin{aligned}&A_{k+1} \subseteq A_k \times S_{k+1}, \text{ for } k \geq 0, \text{ and } Q(\overline{B}_k) \overset{(4.2.33)}{=} Q_k(A_k) \overset{(4.2.25)}{=}\\ &\int_{S_0} P_0(dx_0) \underbrace{\int_{S_1} K_1(x_0, dx_1) \dots \int_{S_k} K_k\big((x_0, \dots, x_{k-1}), dx_k\big) 1_{A_k}(x_0, \dots, x_k)}_{\overset{\shortparallel}{f_{0,k}(x_0)}}\,,\end{aligned} \tag{4.2.42}$$

and

$$(4.2.43) \qquad Q(\overline{B}_{k+1}) = Q_{k+1}(A_{k+1}) = \int_{S_0} P_0(dx_0)$$

$$\underbrace{\int_{S_1} K_1(x_0, dx_1) \dots \int_{S_k} K_k\big((x_0, \dots, x_{k-1}), dx_k\big) \int_{S_{k+1}} K_{k+1}\big((x_0, \dots, x_k), dx_{k+1}\big) 1_{A_{k+1}}(x_0, \dots, x_{k+1})}_{\substack{\| \\ f_{0,k+1}(x_0)}}.$$

Using (4.2.42), one has

$$1_{A_{k+1}}(x_0, \dots, x_{k+1}) \le 1_{A_k}(x_0, \dots, x_k)\, 1_{S_{k+1}}(x_{k+1}),$$

and thus

$$(4.2.44) \qquad f_{0,k}(x_0) \ge f_{0,k+1}(x_0) \ \text{ for all } x_0 \in S_0, k \ge 1\,.$$

Using (4.2.41), (4.2.43), and monotone convergence, there exists $\overline{x}_0 \in S_0$ such that

$$(4.2.45) \qquad \begin{aligned} &\inf_{k \ge 1} \underbrace{f_{0,k}(\overline{x}_0)}_{\substack{\| }} > 0 \\ &\int_{S_1} K_1(\overline{x}_0, dx_1) \int_{S_2} K_2\big((\overline{x}_0, x_1), dx_2\big) \ \dots \\ &\int_{S_k} K_k\big((\overline{x}_0, x_1, \dots, x_{k-1}), dx_k\big)\, 1_{A_k}(\overline{x}_0, x_1, \dots, x_k)\,. \end{aligned}$$

In the same way, it follows from (4.2.45) that there exists $\overline{x}_1 \in S_1$ such that

$$(4.2.46) \qquad \begin{aligned} &\inf_{k \ge 2} f_{1,k}(\overline{x}_0, \overline{x}_1) > 0\,, \ \text{where } f_{1,k}(x_0, x_1) = \\ &\int_{S_2} K_2\big((x_0, x_1), dx_2\big) \dots \int_{S_k} K_k\big((x_0, x_1, \dots x_{k-1}), dx_k\big)\, 1_{A_k}(x_0, x_1, \dots, x_k)\,. \end{aligned}$$

By induction, we construct a sequence $\overline{x}_\ell, \ell \ge 0$, with $\overline{x}_\ell \in S_\ell$ for $\ell \ge 0$, and

$$(4.2.47) \qquad \begin{aligned} &\inf_{k \ge \ell+1} \int_{S_{\ell+1}} K_{\ell+1}\big((\overline{x}_0, \dots, \overline{x}_\ell), dx_{\ell+1}\big) \dots \\ &\int_{S_k} K_k\big((\overline{x}_0, \dots, \overline{x}_\ell), x_{\ell+1}, \dots, x_{k-1}, dx_k\big)\, 1_{A_k}(\overline{x}_0, \dots, \overline{x}_\ell, x_{\ell+1}, \dots, x_k) > 0\,. \end{aligned}$$

In particular, for $k = \ell + 1$ and $\ell \ge 0$, we obtain

$$0 < \int_{S_{\ell+1}} K_{\ell+1}\big((\overline{x}_0, \dots, \overline{x}_\ell), dx_{\ell+1}\big) \ \underbrace{1_{A_{\ell+1}}(\overline{x}_0, \dots, \overline{x}_\ell, x_{\ell+1})}_{\le 1_{A_\ell}(\overline{x}_0, \dots, \overline{x}_\ell)} \le 1_{A_\ell}(\overline{x}_0, \dots, \overline{x}_\ell)$$

and so

$$(\overline{x}_0, \dots, \overline{x}_\ell) \in A_\ell, \text{ for all } \ell \geq 0\,. \tag{4.2.48}$$

Thanks to (4.2.39), one has for all $\ell \geq 0$,

$$\overline{\omega} \overset{\text{def.}}{=} (\overline{x}_0, \overline{x}_1, \dots) \in \overline{B}_\ell\,,$$

which is a contradiction, since $\bigcap\limits_{\ell \geq 0} \overline{B}_\ell \overset{(4.2.39)}{=} \emptyset$.

We have thus proved (4.2.40), and the existence of a probability measure Q on $\mathcal{A} = \sigma(\mathcal{B})$, as an extension of Q from (4.2.33), follows. □

Remark 4.32. Kolmogorov's extension theorem (see Neveu "Probability Theory", Chap. 3 §3) enables the construction of probability measures on arbitrarily large product spaces $\prod_{i \in I} X_i$ (I not necessarily countable). However, the $(X_i, \mathcal{B}_i)$, $i \in I$, must then satisfy a certain regularity condition (e.g. that the X_i are Polish spaces, and for all $i \in I$, $\mathcal{B}_i$ is the Borel σ-algebra on X_i). □

We now present some consequences of the Ionescu-Tulcea theorem.

1) Construction of product probability measures on $\prod_{i \geq 0} S_i$:

We consider $(S_i, \mathcal{S}_i)$ as in (4.2.21), and for all $i \geq 0$, μ_i is a probability measure on $(S_i, \mathcal{S}_i)$. We define K_n, cf. (4.2.24), by

$$K_n\big((x_0, \dots, x_{n-1}), dx_n\big) = \mu_n(dx_n),\ n \geq 1\,. \tag{4.2.49}$$

Hence, it follows from (4.2.25) that

$$Q_n = \mu_0 \otimes \dots \otimes \mu_n \quad \text{for all } n \geq 0\,. \tag{4.2.50}$$

The Ionescu-Tulcea theorem produces a unique probability measure Q on $(\Omega = \prod_{i=0}^\infty S_i, \mathcal{A})$, cf. (4.2.28), such that

$$\pi_n \circ Q = \mu_0 \otimes \dots \otimes \mu_n \quad \text{for all } n \geq 0\,. \tag{4.2.51}$$

Notation: $Q = \bigotimes\limits_{i=0}^{\infty} \mu_i$ "infinite product measure".

2) Construction of canonical Markov chains

Now, $(S_i, \mathcal{S}_i) = (S, \mathcal{S})$ for all $i \geq 0$, and we consider a probability measure μ on $(S, \mathcal{S})$ (the "initial distribution") and a stochastic kernel K from $(S, \mathcal{S})$ to $(S, \mathcal{S})$ (the "transition kernel"). We define, for $n \geq 1$,

$$K_n\big((x_0, x_1, \dots, x_{n-1}), dx_n\big) = K(x_{n-1}, dx_n)\,. \tag{4.2.52}$$

The Ionescu-Tulcea theorem produces a unique probability measure P_μ on $(\Omega = S^{\mathbb{N}}, \mathcal{A})$ such that for any bounded random variable f on $\prod_{i=0}^n S$, $n \geq 0$, one has

$$\begin{aligned} &E^{P_\mu}\big[f(X_0, \dots, X_n)\big] \\ &= \int_S \mu(dx_0) \int_S K(x_0, dx_1) \dots \int_S K(x_{n-1}, dx_n)\, f(x_0, \dots, x_n)\,. \end{aligned} \tag{4.2.53}$$

Notation: $P_x \overset{\text{def.}}{=} P_{\delta_x}$, for $x \in S$.

$(\Omega, \mathcal{A}, (P_x)_{x\in S})$ is called the canonical (time-homogeneous) Markov chain with state space S and transition kernel K. We denote by $(\theta_n)_{n\geq 0}$ the shift operators on Ω: for $\omega = (x_0, x_1, x_2, \dots)$,

$$\theta_n\, \omega = (x_n, x_{n+1}, x_{n+2}, \dots) \in \Omega\,,$$

and by $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$, $n \geq 0$, the canonical filtration on $(\Omega, \mathcal{A})$.

Proposition 4.33. *The map*

$$(x, B) \in S \times \mathcal{A} \mapsto P_x[B] \text{ is a stochastic kernel from } (S, \mathcal{S}) \text{ to } (\Omega, \mathcal{A})\,. \tag{4.2.54}$$

For all $n \geq 0$, $B \in \mathcal{A}$, one has

$$E^{P_\mu}\big[1_B \circ \theta_n \,|\, \mathcal{F}_n\big] \overset{P_\mu\text{-a.s.}}{=} \underbrace{P_{X_n(\omega)}}_{\text{function of } X_n(\omega)}[B]\,. \tag{4.2.55}$$

*In particular, the **Markov property** holds:*

$$E^{P_\mu}\big[1_C \,|\, \mathcal{F}_n\big] = E^{P_\mu}\big[1_C \,|\, \sigma(X_n)\big] \quad P_\mu\text{-a.s.} \tag{4.2.56}$$

for all $C \in \sigma(X_n, X_{n+1}, \dots)$.

Proof.

- (4.2.54): for all $x \in S$, $B \mapsto P_x[B]$ is a probability measure on $(\Omega, \mathcal{A})$. For all $B \in \sigma(X_0, \dots, X_n)$, $x \in S \mapsto P_x[B]$ is $\mathcal{S}$-measurable, thanks to (4.2.53).

  Since $\bigcup_{n\geq 0} \sigma(X_0, \dots, X_n) \overset{\text{def}}{=} \mathcal{B}$ is a π-system with $\sigma(\mathcal{B}) = \mathcal{A}$, the claim follows by using Dynkin's lemma, (1.3.9).

- (4.2.55): Consider $A \in \mathcal{F}_n$, then for any bounded random variable $f\colon \prod_{i=0}^k S \to \mathbb{R}$,

  $$\begin{aligned} &E^{P_\mu}\big[1_A\, f(X_n, X_{n+1}, \dots, X_{n+k})\big] \\ &\overset{(4.2.53)}{=} E^{P_\mu}\big[1_A\, E^{P_{X_n}}[f(X_0, X_1, \dots, X_k)]\big]\,. \end{aligned}$$

  In particular, for all $B \in \mathcal{B}$,

  $$E^{P_\mu}[1_A\, 1_B \circ \theta_n] = E^{P_\mu}\big[1_A\, P_{X_n}[B]\big]\,. \tag{4.2.57}$$

  The collection of all $B \in \mathcal{A}$ satisfying (4.2.57) is a Dynkin system. Thanks to Dynkin's lemma, one has (4.2.57) for all $B \in \mathcal{A}$. Since $A \in \mathcal{F}_n$ is arbitrary, we obtain (4.2.55).

- (4.2.56): any $C \in \sigma(X_n, X_{n+1}, \dots)$ is of the form $1_C = 1_B \circ \theta_n$, with $B \in \mathcal{A}$. Due to (4.2.55), $E^{P_\mu}[1_B \circ \theta_n \,|\, \mathcal{F}_n]$ is equal to a $\sigma(X_n)$-measurable function up to P_μ-equivalence. We deduce (4.2.56). $\square$

Intuitively speaking, the Markov property (4.2.56) means the following: the best forecast for the future after time n of the sequence $X_i, i \geq 0$, given the past up to time n, depends only on the information that is contained in X_n.

Markov chains and martingales

We consider the canonical Markov chain with state space S, transition kernel K, and initial distribution μ (see (4.2.53)). In this setting, we can construct a variety of $(\mathcal{F}_n)$-martingales.

Proposition 4.34. *For f a bounded and measurable function on $(S, \mathcal{S})$,*

$$\begin{aligned} M_n &= f(X_n) - \sum_{k=0}^{n-1} (Kf - f)(X_k) \quad n \geq 1\,, \\ &= f(X_0) \quad n = 0\,, \end{aligned} \tag{4.2.58}$$

where $Kf(x) = \int_S f(x')K(x, dx')$ for $x \in S$, is an $(\mathcal{F}_n)$-martingale under P_μ.

Furthermore, if f is uniformly positive (i.e. $f \geq \alpha > 0$), then

$$\begin{aligned} I_n &= f(X_n) \prod_{k=0}^{n-1} \Big(\frac{f}{Kf}\Big)(X_k) \quad n \geq 1\,, \\ &= f(X_0) \quad n = 0\,, \end{aligned} \tag{4.2.59}$$

is an $(\mathcal{F}_n)$-martingale under P_μ.

Proof. (4.2.58): M_n is bounded and $(\mathcal{F}_n)$-adapted. Moreover, for $n \geq 0$,

$$\begin{aligned} E^{P_\mu}\big[M_{n+1} - M_n \,|\, \mathcal{F}_n\big] &= E^{P_\mu}\big[f(X_{n+1}) - f(X_n) - (Kf - f)(X_n) \,|\, \mathcal{F}_n\big] \\ &= E^{P_\mu}\big[f(X_{n+1}) \,|\, \mathcal{F}_n\big] - Kf(X_n) \\ &\overset{(4.2.55)}{=} E^{P_{X_n}}[f(X_1)] - Kf(X_n) = 0\,. \end{aligned}$$

(4.2.59): I_n is bounded and $(\mathcal{F}_n)$-adapted. Furthermore, for $n \geq 0$,

$$\begin{aligned} E^{P_\mu}\big[I_{n+1} \,|\, \mathcal{F}_n\big] &= E^{P_\mu}\Big[\frac{f(X_{n+1})}{Kf(X_n)}\, I_n \,|\, \mathcal{F}_n\Big] \\ &= \frac{I_n}{Kf(X_n)}\, E^{P_\mu}\big[f(X_{n+1}) \,|\, \mathcal{F}_n\big] \overset{(4.2.55)}{=} I_n\,. \end{aligned}$$

□

Example 4.35. In the case of the simple random walk on $\mathbb{Z}$, one has $Kf(x) = \frac{f(x+1)+f(x-1)}{2}$ for x in $\mathbb{Z}$. Hence,

$$M_n = f(X_n) - \sum_{k=0}^{n-1} \frac{1}{2}\,\big[f(X_k + 1) + f(X_k - 1) - 2f(X_k)\big] \quad n \geq 0\,, \tag{4.2.60}$$

resp.

$$I_n = f(X_n) \, \exp\Big\{ - \sum_{k=0}^{n-1} \log \, \Big(\frac{f(X_k+1) + f(X_k-1)}{2f(X_k)}\Big)\Big\} \quad n \ge 0\,, \tag{4.2.61}$$

is an $(\mathcal{F}_n)$-martingale under each P_x, $x \in \mathbb{Z}$, if f is bounded, resp. bounded and uniformly positive.

Furthermore, one has P_x-a.s., $|X_n - x| \le n$ for all $n \ge 0$. From this, one can easily see that the martingale property of $(M_n)_{n\ge0}$, resp. $(I_n)_{n\ge0}$, holds under all P_x, for all, resp. all uniformly positive, functions f. For instance, if $f(x) = e^{\alpha x}$, one sees from (4.2.61) that

$$e^{\alpha X_n - n \log\cosh\alpha} \quad n \ge 0\,, \tag{4.2.62}$$

is an $(\mathcal{F}_n)$-martingale under each P_x. ○

Index