

LIMIT THEOREMS & APPLICATIONS

M2 MATHÉMATIQUES DE L'ALÉATOIRE
UNIVERSITÉ PARIS SACLAY & INSTITUT POLYTECHNIQUE DE PARIS

CYRIL MARZOUK¹

21st September 2026

¹CMAP, École polytechnique.

ABOUT THIS DOCUMENT. This is a preliminary version of lecture notes for this master course and it will certainly considerably evolve. Please use it with care! Any remark, typo, error that you may spot will be appreciated, in order to improve these notes.

SUGGESTION OF REFERENCES. The content of this course is quite classical, although it can be split into several different courses, and it may be found in numerous books, with different approaches. We may suggest the following books with some personal comments. The point is to find one or two books that you enjoy reading and find complementary to the lectures.

The first two parts of these notes were mostly inspired by Billingsley and the third one by Sato.

- For about the whole course:
 - Fristedt & Gray - A Modern Approach to Probability Theory. Covers all the course (and much more) with a similar approach.
 - Kallenberg - Foundations of modern probability. An encyclopaedia! Many results, and very condensed.
 - Stroock - Probability theory an analytic view. Clearly analysis-oriented, proofs are typically very different from this course.
- For generalities on weak convergence (Part I) and applications to stochastic processes (Part II):
 - Billingsley - Convergence of Probability Measures. An excellent reference with the theory (not too deep) and many discussions.
 - Ethier & Kurtz - Markov processes characterization and convergence. Stochastic (Markov) process oriented, Chapter 3 presents the general theory of weak convergence before specialising to processes; the approach is quite different from ours.
 - Klenke - Probability Theory. A Comprehensive Course. Starts from the very beginning, very pedagogical.
 - Parthasarathy - Probability measures on metric spaces. If you want to see more topology and higher generalities.
- For infinitely divisible laws, stable laws, and Lévy processes (Part III):
 - Feller - An Introduction to Probability Theory and Its Applications, Volume 2. Excellent book in many respect! The treatment of infinitely divisible laws is particularly enjoyable.
 - Kyprianou - Fluctuations of Lévy Processes with Applications. We will only cover the very beginning, this shows you just about how much you can with Lévy processes.
 - Sato - Lévy Processes and Infinitely Divisible Distributions. Excellent introduction to the topic (and more advanced result!); theory oriented.

CONTENTS

I	Weak convergence in metric spaces	5
1	Introduction to weak convergence	6
1.1	Topological primer	6
1.2	Definition and first properties	9
1.3	The Portmanteau theorem	10
1.4	Test functions	12
1.5	Joint convergence, independence	14
1.6	Proving weak convergence	15
1.7	Exercises	17
2	Convergence of random vectors	19
2.1	Distribution functions on the line	19
2.2	Distribution functions in higher dimension	21
2.3	Tightness	24
2.4	Characteristic functions	26
2.5	Characteristic functions and weak convergence	29
2.6	The moment method	31
2.7	Exercises	33
3	<i>Intermezzo: product spaces</i>	36
3.1	Product σ -algebra	36
3.2	Extension theorems	39
3.3	Product topology	42
3.4	Weak convergence of sequences	43
3.5	Exercises	44
4	The topology of weak convergence	45
4.1	Prohorov's theorem	45
4.2	Metric on probability measures	48
4.3	Topology of the Lévy–Prohorov distance	50
4.4	Exercises	51
5	Other modes of convergence & coupling	53
5.1	Convergence of random variables	53
5.2	The Skorokhod representation theorem	55
5.3	Topology of convergence in probability	56
5.4	Coupling	57
5.5	Total variation distance	58
5.6	Exercises	62

II	Convergence of stochastic processes	63
6	Continuous stochastic processes	64
6.1	Stochastic processes vs. random functions	64
6.2	Separability and completeness	65
6.3	Marginal laws	66
6.4	The compact sets	68
6.5	Tightness criteria	70
6.6	Exercises	75
7	Random walks & Brownian motion	77
7.1	Donsker's theorem & Brownian motion	77
7.2	Finite-dimensional marginals	78
7.3	Tightness	79
7.4	Some applications	80
7.5	Exercises	82
8	Discontinuous stochastic processes	86
8.1	Cadlag functions	86
8.2	The Skorokhod distance	88
8.3	A complete distance	90
8.4	Marginal laws	90
8.5	Compact sets and tightness	93
8.6	Exercises	94
III	Infinitely divisible laws & Lévy processes	95
9	Infinitely divisible distributions	96
9.1	Poisson convergence	96
9.2	Infinitely divisible distributions	98
9.3	Limits of compound Poisson distributions	101
9.4	The Lévy–Khintchine formula	104
9.5	A convergence criterion	106
9.6	Exercises	109
10	Lévy processes	111
10.1	Lévy processes	111
10.2	Poisson processes	112
10.3	Compound Poisson processes	115
10.4	Lévy–Itô decomposition	118
10.5	Construction of the martingale	119
10.6	Poisson random measure of jumps	121
10.7	Exercises	123
11	Some properties of Lévy processes	125
11.1	Structure of the jumps	125
11.2	Finite and infinite variation	126
11.3	Increasing Lévy processes	128
11.4	Moments of a Lévy process	130

11.5	Exercises	134
12	Stable laws & processes	135
12.1	Stable laws	135
12.2	Stable Lévy processes	138
12.3	Domain of attraction	139
12.4	The Lévy–Khintchine formula	141
12.5	Moments and variation	145
12.6	An explicit formula for the characteristic function	146
12.7	Exercises	147
13	Domain of attraction of stable laws	148
13.1	Slow variation	148
13.2	Limits of normalised random walks	150
13.3	Proof of the sufficiency	152
13.4	Proof of the necessity	155
13.5	Exercises	155
IV	Appendix	157
A	Poisson random measures	158
A.1	Random measures	159
A.2	Poisson random measures	161
A.3	Stability properties	163
A.4	Integration against a Poisson measure	165
A.5	Compensated integrals	167
A.6	Space-time measures	169
A.7	Exercises	173

PART I

WEAK CONVERGENCE IN METRIC SPACES

CHAPTER 1

INTRODUCTION TO WEAK CONVERGENCE

In this first chapter, we define the general topological framework of this course, then we define weak convergence and review some basic properties with which the reader may already be familiar (perhaps just in $\mathbb{R}$ or Euclidean spaces). This chapter mainly follows the book of Billingsley.

CONTENTS

1.1	Topological primer	6
1.2	Definition and first properties	9
1.3	The Portmanteau theorem	10
1.4	Test functions	12
1.5	Joint convergence, independence	14
1.6	Proving weak convergence	15
1.7	Exercises	17

We shall only consider probability measures on a metric space (E, d) , which we shall always equip with the Borel σ -algebra $\mathcal{B}(E)$ generated by the open sets.

1.1 TOPOLOGICAL PRIMER

The general framework in which the theory of weak convergence / convergence in distribution demonstrates its power is that of Polish spaces.

DEFINITION 1.1. A *Polish space* is a separable topological space (it contains a countable dense subset) whose topology can be induced by a distance that makes it complete (every Cauchy sequence converges).

A topological space can have many different metrics that induce its topology. Some properties of E are topological properties, that are shared by all these metrics and are preserved by homeomorphisms, such as separability, but some others are metric properties, in that they depend on the precise distance that we use, such as completeness. As another example, continuity of a function $f : E \rightarrow \mathbb{R}$ is a topological property, but Lipschitz-continuity is a metric property.

EXAMPLE 1.2. The open interval $(-1, 1)$ naturally comes with the Euclidean distance. It is separable but clearly not complete. However, one may define an equivalent distance by $d'(x, y) = |\tan(\pi x/2) - \tan(\pi y/2)|$ for example, which inherits completeness from $\mathbb{R}$. Hence, the interval $(-1, 1)$ with the Euclidean topology is a Polish space!

This example might seem artificial and another solution here would be to work with the completion $[-1, 1]$ instead, but Chapter 8 will display an important case where we really want to replace the original, incomplete, distance by an equivalent, complete, one.

This being said...

In practice, we will assume that our Polish space E is equipped with a distance d that is complete and we will use the adjective ‘Polish’ for such a complete separable metric space.

EXAMPLES

Consider the following metric spaces; some of them will be further discussed in later chapters.

- (i) Euclidean spaces $\mathbb{R}^d$ with $d \geq 1$ are separable (just take $\mathbb{Q}^d$) and complete.
- (ii) The space of continuous functions from $[0, 1]$ to $\mathbb{R}$ equipped with the supremum norm is separable (take polynomials with rational coefficients) and complete.
- (iii) The space of bounded continuous functions from $[0, \infty)$ to $\mathbb{R}$ equipped with the supremum norm is *not* separable; it is for the uniform norm on compact sets (and complete for both).
- (iv) Similarly the space $\{0, 1\}^{\mathbb{N}}$ of binary sequences is *not* separable for the ℓ^∞ norm, but it is for the product topology (and complete for both).
- (v) If G is a countable, connected and locally finite graph with a distinguished root vertex v , let $B(v, n)$ denotes the subgraph made only of those vertices that can be reached from the root v by traversing at most n edges (i.e. the ball of radius n centred at the root). Then define *local distance* between such rooted graphs as:

$$d_{\text{loc}}((G_1, v_1), (G_2, v_2)) = \frac{1}{1 + \sup\{n \geq 0 : B_1(v_1, n) = B_2(v_2, n)\}}.$$

This space is separable and complete; here the countable set of finite graphs is dense.

- (vi) The space of probability measures on a metric space (E, d) can be endowed with the *Lévy–Prohorov distance*:

$$d_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(C) \leq \nu(C^\varepsilon) + \varepsilon \text{ and } \nu(C) \leq \mu(C^\varepsilon) + \varepsilon \text{ for all closed sets } C\},$$

where C^ε is the set of points that lie at distance less than ε from C . We shall study in quite depth this distance, showing that if E is separable, then the space of probability measures is separable (those that are supported by finitely many elements in the countable dense set of E and with rational point masses are dense), and complete if E is. Also convergence for d_{LP} is equivalent to weak convergence.

- (vii) The space of all compact subsets of a separable metric space (E, d) can be endowed with the *Hausdorff distance*:

$$d_H(A, B) = \inf\{r > 0 : A \subset B^r \text{ and } B \subset A^r\}.$$

It is then separable: take all the finite collections of elements in the countable dense set of E . It is complete when (E, d) is.

- (viii) The space of compact metric spaces can be endowed with the *Gromov–Hausdorff pseudo-distance*:

$$d_{GH}((E_1, d_1), (E_2, d_2)) = \inf d_H(\phi(E_1), \psi(E_2)),$$

where the infimum is taken over all metric spaces (E, d) and all isometric embeddings $\phi : E_1 \rightarrow E$ and $\psi : E_2 \rightarrow E$, and where d_H is the Hausdorff distance on the compact sets of E . This is only a pseudo-distance, in which two metric spaces lie at distance 0 if and only if they are isometric to each other. The associated quotient space is a genuine metric space, which is separable and complete.

The focus of this course, in the second and third part, will be on stochastic processes, a.k.a. random functions, but convergence in distribution of random graphs, random measures, compact subsets, or compact metric spaces, using the above topologies, are actively studied in contemporary research.

COMPACTNESS

Let us recall some standard definitions.

DEFINITION 1.3. A metric space is:

- *compact* when every cover by open sets admits a finite subcover,
- *pre-compact* (or *totally bounded*) when for every $\varepsilon > 0$, it can be covered by finitely many balls with radius ε .

A subset of a metric space is *relatively compact* when its closure is compact.

Note that a pre-compact space is separable. Compactness is a very enjoyable setting when considering convergence of sequences, however many interesting spaces are not compact (such as the above examples). Completeness then allows to reduce compactness to pre-compactness only as shown by the following important result.

PROPOSITION 1.4. A metric space is compact if and only if it is both pre-compact and complete.

Proof. Clearly compactness implies pre-compactness: just extract from the cover made of all balls with radius ε a finite subcover. Also compactness is equivalent in a metric space to the fact that every sequence has a convergent subsequence. Since a Cauchy sequence with a convergent subsequence must be convergent, then compactness implies completeness.

Conversely, suppose the space pre-compact and complete and let $(x_n)_{n \geq 1}$ be any sequence. For any $k \geq 1$, pre-compactness implies that we may cover the space by finitely many balls with radius $1/k$, so one of these balls must contain infinitely many elements of $(x_n)_{n \geq 1}$. By a diagonal extraction, we may construct a subsequence of $(x_n)_{n \geq 1}$ that satisfies $d(x_{n_i}, x_{n_j}) < 2/n_i$ for every $n_j > n_i$. This subsequence is Cauchy and therefore convergent. $\square$

SEPARABILITY

Separability is crucial in measure theory because we are a priori only allowed to use countable set operations (unions and intersections). Here are a few consequences of separability that may not hold in full generality.

REMARK 1.5. In a separable metric space, an open set is the union of countably many open balls. Thus the Borel σ -algebra is generated by the open balls and a random variable $X : (\Omega, \mathcal{A}) \rightarrow (E, d)$ is measurable as soon as $\{X \in B\} \in \mathcal{A}$ for every open ball B . In a non-separable metric space, the inclusion

$$\sigma(\{\text{open balls}\}) \subset \mathcal{B}(E) = \sigma(\{\text{open sets}\})$$

may be strict.

Another reason to assume separability is when we consider product spaces. We will review some general facts about product spaces in Chapter 3, but let us here recall a few basics. The product set $E^2 = E \times E$ can be equipped with a metric, e.g. $d_2((x, x'), (y, y')) = d(x, x') \vee d'(x', y')$, and further a Borel σ -algebra $\mathcal{B}(E^2)$. This product set can also be equipped with the product of the Borel σ -algebras

$$\mathcal{B}(E)^{\otimes 2} = \mathcal{B}(E) \otimes \mathcal{B}(E) = \sigma(\{A \times A' : A, A' \in \mathcal{B}(E)\}).$$

While it is always true that $\mathcal{B}(E)^{\otimes 2} \subset \mathcal{B}(E^2)$, equality may not hold, but it does so for separable spaces.

Recall that the product σ -algebra is made up to satisfy:

$$(X, Y) \text{ is measurable for } \mathcal{B}(E)^{\otimes 2} \iff X \text{ and } Y \text{ are measurable for } \mathcal{B}(E).$$

In other words, this is the σ -algebra that makes a pair of random variables a random variable in the space of pairs. For a separable metric space, we infer that for any two random variables X and Y , the pair is measurable for the Borel σ -algebra on E^2 and hence, by composition,

$$d(X, Y) \text{ is measurable.}$$

This ensures for example that convergence in probability is well-defined:

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X \quad \text{when} \quad \forall \varepsilon > 0, \quad \mathbb{P}(d(X_n, X) > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

Similarly for almost sure convergence

$$X_n \xrightarrow[n \rightarrow \infty]{a.s.} X \quad \text{when} \quad \mathbb{P}(\{X_n \rightarrow X\}) = 1,$$

after writing $\{X_n \rightarrow X\} = \bigcap_{k \geq 1} \bigcup_{N \geq 1} \bigcap_{n \geq N} \{d(X_n, X) \leq 1/k\}$, which is a measurable set provided that $d(X_n, X)$ is measurable.

1.2 DEFINITION AND FIRST PROPERTIES

Here is the notion of interest in this course. For a probability measure μ on (E, d) and a measurable function $f : E \rightarrow \overline{\mathbb{R}}$, we shall write the integral as:

$$\mu(f) = \int f \, d\mu$$

whenever it is well-defined, that is, when either $f \in [0, \infty]$, or either $f^+ = \max(f, 0)$ has $\mu(f^+) < \infty$ or $f^- = \max(-f, 0)$ has $\mu(f^-) < \infty$.

DEFINITION 1.6. Let μ and μ_n for $n \geq 1$ be probability measures on (E, d) . We say that μ_n *converges weakly* to μ when

$$\mu_n(f) \xrightarrow[n \rightarrow \infty]{} \mu(f)$$

for every continuous and bounded function $f : E \rightarrow \mathbb{R}$. We denote it by $\mu_n \Rightarrow \mu$.

We shall for the moment keep the notation of probability measures and weak convergence, and the reader is left to translate everything in terms of random variables.

DEFINITION 1.7. If X is a random variable defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in (E, d) , then we let $\mathcal{L}(X)$ denote its law, which is the probability μ on E given by $\mu(B) = \mathbb{P}(X \in B)$ for every $B \in \mathcal{B}(E)$. We say that random variables X_n *converge in distribution* (or *in law*) to a random variable X when $\mathcal{L}(X_n)$ converges weakly to $\mathcal{L}(X)$, namely when

$$\mathbb{E}[f(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X)]$$

for every f continuous and bounded.

Observe that a probability μ on (E, d) is uniquely characterised by its values on all open sets or, equivalently, on all closed sets. This holds thanks to the $\pi - \lambda$ lemma, also called Dynkin's lemma or monotone class lemma, because these collections of sets are closed under finite intersections (also called π -systems). We can alternatively be more explicit, see Exercise 1.

DEFINITION 1.8. Let $A \subset E$ and $x \in E$, then the distance from x to A is

$$d(x, A) = \inf_{y \in A} d(x, y).$$

For every $r > 0$, the r -enlargement of A is the (open) set

$$A^r = \{x \in E : d(x, A) < r\}.$$

PROPOSITION 1.9. If μ and ν are two probability measures on E such that $\mu(f) = \nu(f)$ for every function f that is Lipschitz-continuous and bounded, then $\mu = \nu$. This implies that the weak limit of probability measures, if it exists, is unique since two limits would coincide on every (Lipschitz-)continuous and bounded functions.

Proof. Since two probability measures that agree on every closed set are identical, then the idea is then to approximate the indicator of any closed set C by Lipschitz and bounded functions. Precisely for every $n \geq 1$, we easily check that

$$\mathbb{1}_C \leq \max(0, 1 - nd(x, C)) \leq \mathbb{1}_{C^{1/n}}.$$

The function f_n in the middle is n -Lipschitz, so

$$\mu(C) \leq \mu(f_n) = \nu(f_n) \leq \nu(C^{1/n}).$$

As $n \rightarrow \infty$, then right-hand side converges to $\nu(C)$, so we conclude that $\mu(C) \leq \nu(C)$, and symmetrically that $\nu(C) \leq \mu(C)$ for every closed set C , which implies that $\mu = \nu$. $\square$

1.3 THE PORTMANTEAU THEOREM

Viewing a probability measure on (E, d) as a function on $\mathcal{B}(E)$, a natural notion of convergence would be a pointwise convergence, i.e. $\mu_n \rightarrow \mu$ when $\mu_n(A) \rightarrow \mu(A)$ for all $A \in \mathcal{B}(E)$. In terms of random variables, this would mean $\mathbb{P}(X_n \in A) \rightarrow \mathbb{P}(X \in A)$ for all $A \in \mathcal{B}(E)$. This is actually too strong because of boundary effects; for example, if U_n has the uniform distribution on the set $\{k/n : 1 \leq k \leq n\}$ and U the uniform distribution on $[0, 1]$, then for f continuous and bounded, it holds

$$\mathbb{E}[f(U_n)] = \frac{1}{n} \sum_{k=1}^n f(k/n) \xrightarrow{n \rightarrow \infty} \int_0^1 f(x) dx = \mathbb{E}[f(U)]$$

so $U_n \rightarrow U$ in distribution as we naturally want. However, since $\mathbb{Q}$ is countable, then

$$\mathbb{P}(U_n \in \mathbb{Q}) = 1 \quad \text{whereas} \quad \mathbb{P}(U \in \mathbb{Q}) = 0.$$

Even more striking, for a deterministic sequence $(x_n)_n$ that converges in E to some x , but such that $x_n \neq x$ for every n , the random variables $X_n = x_n$ almost surely converge in distribution to $X = x$ almost surely, but $\mathbb{P}(X_n = x) = 0$ for every n .

The next result compares this notion with the weak convergence. Recall that the boundary of a set $A \subset E$ is the set ∂A given by its closure minus its interior. It is a closed set, hence measurable for the Borel σ -algebra $\mathcal{B}(E)$.

THEOREM 1.10 (Portmanteau). Let μ and $(\mu_n)_{n \geq 1}$ be probability measures on (E, d) . The following are equivalent:

- (i) $\mu_n \rightarrow \mu$ weakly,

- (ii) For every closed set $C \subset \mathbb{R}^d$ it holds $\limsup_n \mu_n(C) \leq \mu(C)$,
- (iii) For every open set $O \subset \mathbb{R}^d$ it holds $\liminf_n \mu_n(O) \geq \mu(O)$,
- (iv) For every Borel set $B \subset \mathbb{R}^d$ such that $\mu(\partial B) = 0$ it holds $\lim \mu_n(B) = \mu(B)$.

Proof. Suppose that $\mu_n \rightarrow \mu$ weakly. Fix C a closed set and recall the sequence of Lipschitz functions

$$f_k(x) = \max(0, 1 - kd(x, C)),$$

which lies between $\mathbb{1}_C$ and $\mathbb{1}_{C^{1/k}}$. Then

$$\limsup_{n \rightarrow \infty} \mu_n(C) \leq \limsup_{n \rightarrow \infty} \mu_n(f_k) = \mu(f_k) \leq \mu(C^{1/k}).$$

Letting $k \rightarrow \infty$, the right-hand side converges to $\mu(C)$.

Clearly (ii) $\iff$ (iii) by complementation; let us prove that, combined, they both imply (iv). Fix $A \in \mathcal{B}(E)$, let $O \subset A$ denote its interior and $C \supset A$ its closure, so $\partial A = C \setminus O$. Then we know that

$$\liminf_{n \rightarrow \infty} \mu_n(A) \geq \liminf_{n \rightarrow \infty} \mu_n(O) \geq \mu(O),$$

and

$$\limsup_{n \rightarrow \infty} \mu_n(A) \leq \limsup_{n \rightarrow \infty} \mu_n(C) \leq \mu(C).$$

If $\mu(\partial A) = 0$, then $\mu(O) = \mu(A) = \mu(C)$ so the previous inequalities are all equalities: the $\liminf$ and $\limsup$ coincide and thus

$$\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A).$$

The only tricky part is to prove that (iv) $\implies$ (i). Let f be continuous and bounded; replacing f by $f - \inf f$ if necessary, let us assume that $f \in [0, K]$ for some $K > 0$. Observe that for any $x \in E$ we have

$$f(x) = \int_0^K \mathbb{1}_{t \leq f(x)} dt$$

so by Fubini's theorem, if we let $A_t^f = \{x \in E : f(x) \geq t\}$, then

$$\mu(f) = \int_0^K \mu(A_t^f) dt.$$

Since f is continuous, then $\partial A_t^f = f^{-1}(t) = \{x \in \mathbb{R}^d : f(x) = t\}$. These sets are disjoint for different values of t , so for any $k \geq 1$, by σ -additivity, there can only be at most k different values of t for which $\mu(f^{-1}(t)) \geq 1/k$. The set $D = \{t \in [0, K] : \mu(\partial A_t^f) > 0\}$ is therefore countable and in particular has zero Lebesgue measure. Now for any $t \in D^c$ we have $\mu(\partial A_t^f) = 0$ so $\mu(A_t^f) = \lim_n \mu_n(A_t^f)$. We conclude that

$$\mu_n(f) = \int_0^K \mu_n(A_t^f) dt \xrightarrow{n \rightarrow \infty} \int_0^K \mu(A_t^f) dt = \mu(f),$$

by the dominated convergence theorem applied to the Lebesgue integral on $[0, K]$. $\square$

Recall that a discrete space is a topological space E where every subset is both open and closed. The topology can be induced by the so-called discrete metric $d(x, y) = \mathbb{1}_{x \neq y}$ that puts every pair $x \neq y$ at distance 1 of each other. The associated Borel σ -algebra is then made of all the subsets $\mathcal{B}(E) = \mathcal{P}(E)$. Typical examples are countable sets where each element is isolated, such as $\mathbb{Z}^d$ or $\{n^{-1}, n \in \mathbb{N}\}$, but $\{n^{-1}, n \in \mathbb{N}\} \cup \{0\}$ is not discrete when equipped with the Euclidean distance.

COROLLARY 1.11 (Discrete case). Let μ_n and μ be probability measures on a countable discrete space E . Then $\mu_n \Rightarrow \mu$ if and only if $\mu_n(\{x\}) \rightarrow \mu(\{x\})$ for every $x \in E$.

Proof. Since every subset is both open and closed, then the boundary is empty, so by the previous theorem, weak convergence is equivalent to $\mu_n(B) \rightarrow \mu(B)$ for every subset B , in particular for singletons.

Conversely, suppose we have the convergence for all singletons, then it extends to any finite set B by writing:

$$\mu_n(B) = \sum_{x \in B} \mu_n(x) \xrightarrow{n \rightarrow \infty} \sum_{x \in B} \mu(x) = \mu(B).$$

Since E is countable, then there exists an enumeration $(x_n)_{n \geq 1}$ of its elements. Letting $E_K = \{x_1, \dots, x_K\}$, we have $\mu(E_K^c) = \sum_{k \geq K+1} \mu(x_k) \rightarrow 0$ as $K \rightarrow \infty$. Fix $\varepsilon > 0$, then there exists K such that $\mu(E_K^c) \leq \varepsilon$. Since E_K is finite, then

$$\mu_n(E_K^c) = 1 - \mu_n(E_K) \xrightarrow{n \rightarrow \infty} 1 - \mu(E_K) \leq \varepsilon.$$

Now if $B \subset E$ is any subset, then write

$$|\mu_n(B) - \mu(B)| \leq |\mu_n(B \cap E_K) - \mu(B \cap E_K)| + \mu_n(B \cap E_K^c) + \mu(B \cap E_K^c).$$

On the right, the last two term are bounded above by ε while the first one tends to 0 as $n \rightarrow \infty$ since $B \cap E_K$ is finite. We conclude since ε is arbitrary. $\square$

EXERCISE 1.12. Still relying on Theorem 1.10, provide a more ‘down to Earth’ proof of Corollary 1.11 in the case $E = \mathbb{Z}$ by noting that for every $j \in \mathbb{Z}$, we have $X = j \iff X \in [j - 1/2, j + 1/2] \iff X \in (j - 1/2, j + 1/2)$.

1.4 TEST FUNCTIONS

In practice, it may be useful to check the convergence only for more restrictive functions than all continuous and bounded ones. The next result shows several possibilities. It is very important here that we work with probability measures. Even for finite measures, it may fail if they do not have the same total mass.

THEOREM 1.13 (Restriction of test functions). Let μ_n, μ be probability measures on (E, d) , then the following are equivalent:

- (i) $\mu_n(f) \rightarrow \mu(f)$ for all f continuous and bounded,
- (ii) $\mu_n(f) \rightarrow \mu(f)$ for all f uniformly continuous and bounded,
- (iii) $\mu_n(f) \rightarrow \mu(f)$ for all f Lipschitz and bounded,
- (iv) $\mu_n(f) \rightarrow \mu(f)$ for all f continuous μ -almost surely and bounded.

For the last item, we should make sure first that the continuity set of a function is measurable.

EXERCISE 1.14. Let $f : (E, d) \rightarrow (E', d')$ be a function. Prove that the set of continuity points $\{x \in E : f \text{ is continuous at } x\}$ belongs to $\mathcal{B}(E)$. Its complement, the set of discontinuities, thus also belongs to $\mathcal{B}(E)$.

Proof. Clearly, (i) $\implies$ (ii) $\implies$ (iii). Suppose (iii) holds and let us prove (i). Recall the sequence of functions

$$f_k(x) = \max(0, 1 - kd(x, C)),$$

where $k \geq 1$ and C is a fixed closed set. Then this function is k -Lipschitz and the argument from the proof of Theorem 1.10 shows that the convergence $\mu_n(f_k) \rightarrow \mu(f_k)$ for all k implies $\limsup_n \mu_n(C) \leq \mu(C)$, which implies (i) by this theorem.

Since clearly (iv) $\implies$ (i) it only remains to prove the converse implication. Suppose that (i) holds, fix f continuous μ -almost surely and bounded and fix $\varepsilon > 0$. Let $K > 0$ be such that $|f| \leq K$ and note that as in the proof of Theorem 1.10, for any $k \geq 1$ there can only be at most k different values of $t \in \mathbb{R}$ such that $\mu(f^{-1}(t)) \geq 1/k$ so the set $D = \{t \in \mathbb{R} : \mu(f^{-1}(t)) > 0\} \subset [-K, K]$ is at most countable. For $k \geq 1$ large enough we can fix some values $a_0 < \dots < a_k$ such that: $a_0 < -K$, $a_k > K$, and $a_i - a_{i-1} \leq \varepsilon$ and $a_i \in D^c$ for all $i \leq k$. Define then $A_i = \{x \in \mathbb{R}^d : a_{i-1} < f(x) \leq a_i\}$ so $\partial A_i \subset \{x \in \mathbb{R}^d : f(x) \in \{a_{i-1}, a_i\}\} \cup C_f^c$ where we let $C_f = \{x \in \mathbb{R}^d : f \text{ is continuous at } x\}$. By our assumption $\mu(C_f^c) = 0$ and by our construction of A_i we get $\mu(\partial A_i) = 0$ so $\mu_n(A_i) \rightarrow \mu(A_i)$ for each $i \leq k$ and so

$$\sum_{i=1}^k a_i \mu_n(A_i) \xrightarrow{n \rightarrow \infty} \sum_{i=1}^k a_i \mu(A_i).$$

Finally, from the construction we have $f \leq \sum_{i=1}^k a_i \mathbb{1}_{A_i} \leq f + \varepsilon$ so we infer that

$$\limsup_{n \rightarrow \infty} \mu_n(f) \leq \limsup_{n \rightarrow \infty} \sum_{i=1}^k a_i \mu_n(A_i) = \sum_{i=1}^k a_i \mu(A_i) \leq \mu(f) + \varepsilon,$$

and

$$\liminf_{n \rightarrow \infty} \mu_n(f) \geq \liminf_{n \rightarrow \infty} \sum_{i=1}^k a_i \mu_n(A_i) = \sum_{i=1}^k a_i \mu(A_i) \geq \mu(f) - \varepsilon,$$

so $\mu_n(f) \rightarrow \mu(f)$ since $\varepsilon > 0$ is arbitrary. $\square$

REMARK 1.15. It is easy to see that if $\mu_n(f)$ converges to $\mu(f)$ for all functions f in some set, then this also holds for functions in the closure of that set of the uniform norm, simply use that if two functions are ε -close in uniform norm, then their integrals against any probability measure are ε -close as well. For example, in $E = \mathbb{R}$, bounded functions that are infinitely differentiable with bounded derivatives are dense in continuous and bounded functions.

DEFINITION 1.16. Let μ be a probability measure on E and $f : (E, d) \rightarrow (E', d')$ be a measurable function. The *push-forward* or *image measure* of μ by f is the probability measure on E' given by

$$f_*\mu(B) = \mu(\{f \in B\}), \quad B \in \mathcal{B}(E').$$

It is clear from the definition that if $\mu_n \Rightarrow \mu$ and if f is continuous, then $f_*\mu_n \Rightarrow f_*\mu$. The last item of the previous theorem allows to generalise this fact.

LEMMA 1.17 (Continuous mapping). Suppose that $\mu_n \Rightarrow \mu$ in E and that $f : (E, d) \rightarrow (E', d')$ is measurable and continuous μ -almost surely, then $f_*\mu_n \Rightarrow f_*\mu$ in E' .

In terms of random variables, this means that if $X_n \rightarrow X$ in distribution, then $f(X_n) \rightarrow f(X)$ in distribution provided that f is measurable and continuous $\mathbb{P}_X$ -almost surely.

Proof. Let g be a continuous and bounded function, then $g \circ f$ is bounded and continuous μ -almost surely so by Theorem 1.13 we have:

$$(f_*\mu_n)(g) = \mu_n(g \circ f) \xrightarrow{n \rightarrow \infty} \mu(g \circ f) = (f_*\mu)(g).$$

Since this holds for any $g \in C_b$, then $f_*\mu_n \Rightarrow f_*\mu$. $\square$

EXAMPLE 1.18. Let us anticipate Chapter 7.

Let E denote the space of continuous functions from $[0, 1]$ to $\mathbb{R}$ equipped with the supremum norm. Then $f \mapsto \max f$ is continuous so if X_n are random continuous functions that converge in distribution to a random continuous function X , then $\max X_n \rightarrow \max X$ in distribution. The law of $\max X_n$ may be non explicit, but for standard Brownian motion, this maximum is distributed as the absolute value of a standard Gaussian random variable (as shown by the reflection principle for example).

The function $\operatorname{argmax} : f \mapsto \inf\{t \in [0, 1] : f(t) = \max f\}$ is *not* continuous (prove it!) but it is almost surely continuous with respect to the Brownian motion (see the course about stochastic calculus), so if X is a Brownian motion, then $\operatorname{argmax} X_n \rightarrow \operatorname{argmax} X$ in distribution. Again, the law of $\operatorname{argmax} X_n$ may be non explicit, but that of $\operatorname{argmax} X$ is very explicit: it is called the ‘arcsine law’, with density $1/(\pi\sqrt{x(1-x)})$ on $(0, 1)$.

1.5 JOINT CONVERGENCE, INDEPENDENCE

If (X_n, Y_n) are pairs of random variables that converge in distribution to a pair (X, Y) , then by continuity of the projections $(x, y) \mapsto x$ and $(x, y) \mapsto y$, we have $X_n \rightarrow X$ in distribution and $Y_n \rightarrow Y$ in distribution. In general, these two individual convergences do not conversely imply the joint convergence $(X_n, Y_n) \rightarrow (X, Y)$ in distribution. It does so when X_n and Y_n are independent, in which case X and Y are independent as well. Let us discuss this more formally. Recall that for measurability issues, when considering probabilities on product spaces, we work with separable spaces.

THEOREM 1.19. Assume that $E = E' \times E''$ is the product of separable metric spaces. Let μ_n and μ be probability measures on E . Then $\mu_n \Rightarrow \mu$ if and only if $\mu_n(A' \times A'') \rightarrow \mu(A' \times A'')$ for every $A' \in \mathcal{B}(E')$ and $A'' \in \mathcal{B}(E'')$ such that $\mu(\partial A' \times E'') = \mu(E' \times \partial A'') = 0$.

Proof. The direct implication is easy, it relies on the inclusion of boundaries:

$$\partial(A \cap B) \subset (\partial A) \cup (\partial B)$$

for every subsets A and B of a metric space. Since $A' \times A'' = (A' \times E'') \cap (E' \times A'')$, then under the assumption we have $\mu(\partial(A' \times A'')) = 0$, so the claim follows from the Portmanteau theorem 1.10.

Conversely, let $\mathcal{P}$ denote the set of products $A' \times A''$ with $A' \in \mathcal{B}(E')$ and $A'' \in \mathcal{B}(E'')$ such that $\mu(\partial A' \times E'') = \mu(E' \times \partial A'') = 0$. Assume that $\mu_n(A' \times A'') \rightarrow \mu(A' \times A'')$ for every $A' \times A'' \in \mathcal{P}$ and let us prove that $\mu_n \Rightarrow \mu$. Fix $x' \in E'$ and $x'' \in E''$. Let us equip $E = E' \times E''$ with the supremum norm, then the open ball in E centred at (x', x'') with radius r is the product of the open balls with radius r centred at x' and at x'' in E' and in E'' respectively. These balls have disjoint boundaries when r varies, so there can only be at most countably many values of r for which these boundaries have nonzero measure. Therefore, for every $(x', x'') \in E$, the set $R(x', x'')$ of those values of $r > 0$ for which the corresponding open ball does not belong to $\mathcal{P}$ is at most countable.

Since E is separable, then there exists a countable dense set of pairs $(x'_i, x''_i)_{i \geq 1}$. Let Q be any countable set of $(0, \infty) \setminus \bigcup_i R(x'_i, x''_i)$ that is dense in $(0, \infty)$. Then every open set O can be written as the countable union of the open balls with radius in Q and center in $\bigcup_i (x'_i, x''_i)$ that are entirely contained in O . Since these balls all belong to $\mathcal{P}$, then the claim follows from the next lemma. $\square$

We used the following lemma in the previous proof.

LEMMA 1.20. Let μ_n and μ be probability measures on E and suppose that there exists a π -system $\mathcal{P}$ (closed under finite intersections) such that every open set is a countable union of sets in $\mathcal{P}$ and such

that $\mu_n(A) \rightarrow \mu(A)$ for every $A \in \mathcal{P}$. Then $\mu_n \Rightarrow \mu$.

Proof. If A and B belong to $\mathcal{P}$, then so does $A \cap B$, hence,

$$\mu_n(A \cup B) = \mu_n(A) + \mu_n(B) - \mu_n(A \cap B) \xrightarrow{n \rightarrow \infty} \mu(A) + \mu(B) - \mu(A \cap B) = \mu(A \cup B).$$

This extends to any finite union of elements of $\mathcal{P}$ by the inclusion-exclusion formula. Now let O be an open set, from our assumption $O = \bigcup_{i \geq 1} A_i$ with $A_i \in \mathcal{P}$ for every $i \geq 1$. By monotonicity, for every $N \geq 1$, we have

$$\mu_n(O) \geq \mu_n\left(\bigcup_{i=1}^N A_i\right) \xrightarrow{n \rightarrow \infty} \mu\left(\bigcup_{i=1}^N A_i\right) \xrightarrow{N \rightarrow \infty} \mu(O).$$

Hence $\liminf_n \mu_n(O) \geq \mu(O)$ for every open set O and the claim follows from the Portmanteau theorem 1.10. $\square$

Here is the aforementioned claim on independent random variables. Recall that at the level of measures, independence translates into product measures. Precisely, if X' has law μ' on E' and X'' has law μ'' on E'' , then the pair (X', X'') has the law $\mu' \otimes \mu''$ on $E' \times E''$, which is the probability measure on $\mathcal{B}(E') \otimes \mathcal{B}(E'')$ uniquely characterised by

$$\mu' \otimes \mu''(A' \times A'') = \mu'(A')\mu''(A''), \quad A' \in \mathcal{B}(E'), A'' \in \mathcal{B}(E'').$$

COROLLARY 1.21. Assume that $E = E' \times E''$ is the product of separable metric spaces. Let μ'_n and μ' be probability measures on E' and μ''_n and μ'' be probability measures on E'' . Then

$$\mu'_n \otimes \mu''_n \Rightarrow \mu' \otimes \mu'' \quad \text{if and only if} \quad \mu'_n \Rightarrow \mu' \quad \text{and} \quad \mu''_n \Rightarrow \mu''.$$

Proof. By Theorem 1.19 $\mu'_n \otimes \mu''_n \Rightarrow \mu' \otimes \mu''$ if and only if $\mu'_n(A')\mu''_n(A'') \rightarrow \mu'(A')\mu''(A'')$ for every $A' \in \mathcal{B}(E')$ and $A'' \in \mathcal{B}(E'')$ such that $\mu'(\partial A') = \mu''(\partial A'') = 0$. On the other hand, by the Portmanteau theorem 1.10 $\mu'_n \Rightarrow \mu'$ if and only if $\mu'_n(A') \rightarrow \mu'(A')$ for every $A' \in \mathcal{B}(E')$ such that $\mu'(\partial A') = 0$, and similarly for μ'' . The converse implication is thus immediate, whilst the direct implication follows by simply taking either $A'' = E''$ or $A' = E'$. $\square$

1.6 PROVING WEAK CONVERGENCE

Recall the following easy exercise.

EXERCISE 1.22. In a topological space, a sequence $(x_n)_{n \geq 1}$ converges to some x if and only if every subsequence has a further sub-subsequence that converges to x .

Weak convergence of probability measures does correspond to a topology. The functional analysis enthusiasts have indeed observed that probability measures form a subset of the dual of the space of continuous and bounded functions, so one can use the trace of the weak-* topology. Letting $\mathcal{M}_1(E)$ denote the set of all probability measures on E , a basis of open sets is given by the collection of sets:

$$\{\mu \in \mathcal{M}_1(E) : |\mu(f_i) - \nu(f_i)| < \varepsilon_i \text{ for all } 1 \leq i \leq n\}$$

for all $\nu \in \mathcal{M}_1(E)$, $n \geq 1$, $\varepsilon_1, \dots, \varepsilon_n > 0$, and $f_1, \dots, f_n \in C_b(E)$. We shall however not work in this general setting; indeed we shall see later that when (E, d) is separable, then this weak-* topology can be induced by a distance that turns $\mathcal{M}_1(E)$ into a metric space; in addition this space is separable, and is it complete, resp. compact, if and only if (E, d) is.

For the moment, let us argue in a simple way that the conclusion of the exercise holds for weak convergence of probability measures. We also split the second statement in two.

LEMMA 1.23. A sequence $(\mu_n)_{n \geq 1}$ of probability measures converges weakly to a probability measure μ if and only if the following two properties both hold:

- (i) Every subsequence has a further sub-subsequence that converges weakly to *some probability measure*, which could depend on this subsequence,
- (ii) If a subsequence is convergent, then its limit is μ .

Proof. Clearly the combination of the two properties is equivalent to the fact that every subsequence has a further sub-subsequence that converges weakly to μ . Suppose first that $\mu_n \Rightarrow \mu$ and consider a subsequence $(\mu_{n_k})_{k \geq 1}$. Then for every continuous and bounded function f , since the real-valued sequence $(\mu_n(f))_{n \geq 1}$ converges to $\mu(f)$, then so does every subsequence, and in particular $\mu_{n_k}(f) \rightarrow \mu(f)$. Suppose on the contrary that μ_n does not converge weakly to μ , then there exists a continuous and bounded function f such that $\mu_n(f)$ does not converge to $\mu(f)$. This means that there exists $\varepsilon > 0$ such that $|\mu_n(f) - \mu(f)| > \varepsilon$ for infinitely many indices n , and hence no further sub-subsequence can converge to μ . $\square$

We chose to split the two properties because they typically require different types of arguments, and depending on the context, one can be simpler than the other, but there is no general rule. The second property asks to characterise in a unique way a probability measure, which is often done relying on the monotone class lemma, by checking that the subsequential limits coincide with μ on a π -system that generates the σ -algebra. This depends on the context and on the space E that we consider.

DEFINITION 1.24. A collection $\mathcal{P}$ of probability measures on E is called:

- (sequentially) *relatively compact* when every sequence in $\mathcal{P}$ has a subsequence that converges weakly to some probability measure, which may or may not belong to $\mathcal{P}$.
- *tight* when for every $\varepsilon > 0$, there exists a compact K of E such that $\mu(K) > 1 - \varepsilon$ for every $\mu \in \mathcal{P}$.

The following result links these two notions; it is key to proving weak convergence.

THEOREM 1.25 (Prohorov). Let (E, d) be any metric space and $\mathcal{P}$ be a collection of probability measures on E .

- (i) If the collection $\mathcal{P}$ is tight, then it is sequentially relatively compact.
- (ii) The converse implication holds if (E, d) is separable and complete.

The converse part may be useful in some cases but the most useful part of the statement is the first one since it provides a way of proving relative compactness of a collection of probability measures.

From these results, the standard scheme to prove that $\mu_n \Rightarrow \mu$, very often used in practice, is to prove that (a) the sequence $(\mu_n)_{n \geq 1}$ is tight and (b) all subsequential limits agree.

Prohorov's theorem is one of the big goals of the first part of the course. There are several different approaches to its proof, all of them non trivial. Here are at least four:

- (i) by extracting diagonally a subsequence that converges to some linear form and using the Riesz representation theorem to conclude that the limit is a probability measure (see e.g. in Dudley)

- (ii) by extracting diagonally again a subsequence that converges to an outer measure and using then Caratheodory's extension theorem (see e.g. in Billingsley 2nd edition, or Klenke)
- (iii) by proving it in $\mathbb{R}^d$ with d finite by hands, then extending to $\mathbb{R}^N$, and then embedding a general space in $\mathbb{R}^N$ (see e.g. in Billingsley 1st edition, or Kallenberg)
- (iv) by coupling methods, first describing the topology of weak convergence (see e.g. in Ethier & Kurtz).

The first route is probably the shortest for those familiar with the Riesz representation theorem and has more functional analysis flavour. The second one is more measure theory-oriented and it may please those of you who sweated to prove the existence of the Lebesgue measure in an earlier course to see that some of the arguments are used again here. The third option offers a natural path, starting from the easy case of real-valued random variables, with which you might be already familiar, and extending step by step to more and more complicated cases. The last one offers many advantages and leads also to other results; it is also perhaps the most probabilistic approach.

These notes will follow the third approach. This will be the occasion to review and maybe complement the theory of weak convergence in $\mathbb{R}$ and extend it to $\mathbb{R}^d$ before the general case.

1.7 EXERCISES

The results from the first exercise will be used at several occasions.

EXERCISE 1. Let (E, d) be any metric space and let μ be a probability measure.

- (i) Prove that for every measurable set $A \in \mathcal{B}(E)$ and every $\varepsilon > 0$, there exist an open set O and a closed set C such that

$$C \subset A \subset O \quad \text{and} \quad \mu(O \setminus C) < \varepsilon.$$

In other words,

$$\begin{aligned} \mu(A) &= \inf\{\mu(O) : O \text{ open and } A \subset O\} \\ &= \sup\{\mu(C) : C \text{ closed and } C \subset A\}. \end{aligned}$$

Hint: a direct proof may not work, try proving that the collection of such A 's is a σ -algebra that contains all closed sets.

- (ii) When do we have

$$\begin{aligned} \mu(A) &= \inf\{\mu(C) : C \text{ closed and } A \subset C\} \\ &= \sup\{\mu(O) : O \text{ open and } O \subset A\}? \end{aligned}$$

- (iii) Prove that if μ is tight, then for every $A \in \mathcal{B}(E)$, we have the stronger approximation:

$$\mu(A) = \sup\{\mu(K) : K \text{ compact and } K \subset A\}.$$

- (iv) Prove that if (E, d) is separable and complete, then every singleton is tight.

The next exercise discusses what happens when the space is a topological space, not a metric space.

EXERCISE 2. Consider the collection of subsets of $[0, 1]$ given by

$$\mathcal{T} = \{A \subset [0, 1] : A = \emptyset \text{ or } A^c \text{ is at most countable}\}.$$

- (i) Prove that $\mathcal{T}$ is a topology on $[0, 1]$ in which two nonempty open sets are never disjoint. In particular this topology cannot be induced by a metric.
- (ii) Describe the elements in the Borel σ -algebra $\sigma(\mathcal{T})$.
- (iii) Let μ denote the Lebesgue measure restricted to $\sigma(\mathcal{T})$. Prove that it is not true that for any $A \in \sigma(\mathcal{T})$ and any $\varepsilon > 0$, there exist an open set O and a closed set C (for the topology $\mathcal{T}$) such that $C \subset A \subset O$ and $\mu(O \setminus C) < \varepsilon$.

EXERCISE 3. Let μ be a probability on the separable metric space (E, d) and let $(X_n)_{n \geq 1}$ be i.i.d. random variables distributed according to μ and defined a common probability space. For each $n \geq 1$, the *empirical distribution* is the random (!) probability measure M_n defined by

$$M_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \quad \text{i.e.} \quad M_n(A) = \frac{\#\{1 \leq i \leq n : X_i \in A\}}{n},$$

for every $A \in \mathcal{B}(E)$. Prove that

$$M_n \Rightarrow \mu \quad \text{almost surely.}$$

(Hint: Lemma 1.20 can be handy.)

EXERCISE 4. Let (E, d) be a metric space.

- (i) Prove that if $\mathcal{P}_1, \dots, \mathcal{P}_n$ are tight collections of probability measures, then so is $\bigcup_{i=1}^n \mathcal{P}_i$. Give an example that shows that it may fail for infinite countable unions.
- (ii) Let $\mathcal{P}$ be a collection of probability measures on E and let $f : (E, d) \rightarrow (E', d')$ be a continuous function and define $\mathcal{P}'$ the collection of probability measures on E' given by the push-forward by f of the measures in $\mathcal{P}$. Prove that if $\mathcal{P}$ is tight, then so is $\mathcal{P}'$.

The following easy exercise is used all the time to slightly simplifies the definition of tightness for sequences in a Polish space.

EXERCISE 5. Prove that if every single measure is tight, then a sequence of probability measures $(\mu_n)_{n \geq 1}$ is tight if and only if for every $\varepsilon > 0$, there exists a compact set K such that $\limsup_n \mu_n(K^c) < \varepsilon$.

EXERCISE 6 (Finite measures). The definition of weak convergence extends to measures μ_n and μ on (E, d) that have any finite total mass m_n and m respectively:

$$\mu_n \Rightarrow \mu \quad \text{when} \quad \int f \, d\mu_n \rightarrow \int f \, d\mu \quad \text{for all continuous bounded } f.$$

We say that $(\mu_n)_{n \geq 1}$ is tight when for every $\varepsilon > 0$, there exists a compact set K such that $\mu_n(K^c) < \varepsilon$ for all $n \geq 1$. Finally, we say that $(\mu_n)_{n \geq 1}$ is relatively compact when every subsequence has a weakly convergent sub-subsequence.

- (i) Prove that if μ is the null measure, then $\mu_n \Rightarrow \mu$ if and only if $m_n \rightarrow 0$.
- (ii) Prove that if μ is not the null measure, then $\mu_n \Rightarrow \mu$ if and only if both $m_n \rightarrow m$ and we have the weak convergence of probability measures: $m_n^{-1} \mu_n \Rightarrow m^{-1} \mu$.
- (iii) Prove that Prohorov's theorem 1.25 extends as follows:
 - (a) If $(\mu_n)_{n \geq 1}$ is tight and $\sup_n m_n < \infty$, then $(\mu_n)_{n \geq 1}$ is relatively compact.
 - (b) The converse holds when (E, d) is separable and complete.

CHAPTER 2

CONVERGENCE OF RANDOM VECTORS

Back to Earth! We focus on $E = \mathbb{R}^d$ for some finite dimension $d \geq 1$, with the usual Euclidean distance and Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$. Two very important tools are the distribution function and the characteristic function. The case of $\mathbb{R}$ should be very familiar at this point, and $\mathbb{R}^d$ brings some technical details, but everything works in just the same way. We try to emphasise in this chapter tools to prove weak convergence in practice.

CONTENTS

2.1	Distribution functions on the line	19
2.2	Distribution functions in higher dimension	21
2.3	Tightness	24
2.4	Characteristic functions	26
2.5	Characteristic functions and weak convergence	29
2.6	The moment method	31
2.7	Exercises	33

2.1 DISTRIBUTION FUNCTIONS ON THE LINE

Let us consider first recall the case of real-valued random variables and their laws, i.e. we take in this section $E = \mathbb{R}$.

DEFINITION 2.1. A *distribution function* is a function $F : \mathbb{R} \rightarrow [0, 1]$ that satisfies the following three properties:

- (i) It is nondecreasing: $x \leq y \implies F(x) \leq F(y)$.
- (ii) It is right-continuous: For any $x \in \mathbb{R}$ and any sequence $x_n \downarrow x$, we have $F(x_n) \downarrow F(x)$.
- (iii) It has the limits $F(x) \rightarrow 1$ as $x \rightarrow \infty$ and $F(x) \rightarrow 0$ as $x \rightarrow -\infty$.

THEOREM 2.2. If μ is a probability measure on $\mathbb{R}$, then the function defined by

$$F_\mu(x) = \mu((-\infty, x]), \quad x \in \mathbb{R}$$

is a distribution function. Conversely, given any distribution function F on $\mathbb{R}$, there exists a unique probability measure μ on $\mathbb{R}$ such that $F = F_\mu$.

Proof. The first claim is easy and follows from monotonicity and continuity of measures. Let F be a distribution function and let us prove that it takes the form F_μ for some probability μ . For every $u \in [0, 1]$,

let

$$\underline{F}(u) = \inf\{z \in \mathbb{R} : F(z) \geq u\}.$$

Using that F is nondecreasing and tends to 0 at $-\infty$ and to 1 at ∞ , we see that for each fixed $u \in (0, 1)$, the set on the right is a nonempty interval, with unbounded right edge, i.e. either of the form $[\underline{F}(u), \infty)$ or $(\underline{F}(u), \infty)$. It follows that if $u \leq F(x)$, then x belongs to that interval so $\underline{F}(u) \leq x$. Conversely if $u > F(x)$, then since F is right-continuous at x , then there exists $\delta > 0$ such that $u > F(x + \delta)$, and therefore $x + \delta$ does not belong to this interval so $x + \delta \leq \underline{F}(u)$. We have therefore proved that $u \leq F(x)$ if and only if $\underline{F}(u) \leq x$. Now let Leb denote the Lebesgue measure on $[0, 1]$ and let $\mu = \underline{F}_* \text{Leb}$ denote the push-forward, then by definition we have

$$F_\mu(x) = \text{Leb}(\{u \in [0, 1] : \underline{F}(u) \leq x\}) = \text{Leb}(\{u \in [0, 1] : u \leq F(x)\}) = F(x).$$

Finally this measure μ is unique by the monotone class lemma since the collection of intervals $\{(-\infty, x], x \in \mathbb{R}\}$ is closed under finite intersections and generates the Borel σ -algebra $\mathcal{B}(\mathbb{R})$. $\square$

Our next task is to link weak convergence of probability measures on $\mathbb{R}$ and their distribution function. For any function F we let

$$C_F = \{x \in \mathbb{R} : F \text{ is continuous at } x\}$$

denote its continuity set.

DEFINITION 2.3 (Weak convergence of distribution functions). A sequence $(F_n)_{n \geq 1}$ of functions is said to converge *weakly* to a function F when

$$F_n(x) \xrightarrow{n \rightarrow \infty} F(x) \quad \text{for all } x \in C_F.$$

PROPOSITION 2.4. We have $X_n \rightarrow X$ in distribution if and only if $F_{X_n} \rightarrow F_X$ weakly.

Proof. Note that for every $x \in \mathbb{R}$ we have $\partial(-\infty, x] = \{x\}$ and $\mathbb{P}(X = x) = F(x) - F(x-)$ so the direct implication is a particular case of the Portmanteau theorem 1.10. For the converse implication, suppose that $F_{X_n} \rightarrow F_X$ weakly and let $a < b$. Then

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in (a, b)) &= \liminf_{n \rightarrow \infty} (F_{X_n}(b-) - F_{X_n}(a)) \\ &\geq \liminf_{n \rightarrow \infty} F_{X_n}(b-) - \limsup_{n \rightarrow \infty} F_{X_n}(a) \\ &\geq F_X(b-) - F_X(a) = \mathbb{P}(X \in (a, b)). \end{aligned}$$

Recall that any open set of $\mathbb{R}$ is a countable union of disjoint open interval, say $O = \bigcup_k (a_k, b_k)$, then

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in O) &= \liminf_{n \rightarrow \infty} \sum_k \mathbb{P}(X_n \in (a_k, b_k)) \\ &\geq \sum_k \liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in (a_k, b_k)) \\ &\geq \sum_k \mathbb{P}(X \in (a_k, b_k)) \\ &= \mathbb{P}(X \in O). \end{aligned}$$

By the Portmanteau theorem 1.10 again, this is equivalent to $X_n \rightarrow X$ in distribution. $\square$

The next theorem is useful in practice, but one should not misunderstand its statement. Convergence in distribution of random variables is really just the convergence of their laws; Theorem 2.5 now shows

that we can realise copies of all the random variables in the same probability space (possibly a different one from the original one) so that they converge in the strong almost sure sense.

THEOREM 2.5 (Skorokhod's representation). Suppose that F_n, F are distribution functions such that $F_n \rightarrow F$ weakly. Then on the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \text{Leb})$ we can construct random variables X_n, X which have distribution function F_n, F respectively and such that $X_n \rightarrow X$ almost surely.

The proof is guided in Exercise 7.

REMARK 2.6. If one starts with random variables X_n, X in the first place, defined on possibly different probability spaces, and with $X_n \rightarrow X$ in distribution, then the theorem states that there exists a probability space with new random variables X'_n, X' defined on it, with the same law as X_n and X respectively and such that $X'_n \rightarrow X'$ almost surely. This does *not* mean that $X_n \rightarrow X$ almost surely! As a matter of fact, the correlations between the X_n 's are completely modified in this procedure.

EXAMPLE 2.7. To see how this reasoning works, let us give another proof of Lemma 1.17. Suppose that $X_n \rightarrow X$ in distribution and that $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous $\mathbb{P}_X$ -almost surely. Then there exist X'_n, X' with the same law as X_n, X and such that $X'_n \rightarrow X'$ almost surely. Then $f(X'_n) \rightarrow f(X')$ almost surely by Lemma 5.4 and so $f(X'_n) \rightarrow f(X')$ in distribution. Since $f(X'_n), f(X')$ have the same law as $f(X_n), f(X)$ respectively, we conclude that $f(X_n) \rightarrow f(X)$ in distribution.

EXAMPLE 2.8. Here is another example of application of this theorem. Recall that the dominated convergence theorem can be pushed to the maximum as follows. Let X and $(X_n)_{n \geq 1}$ be random variables, and assume that $X_n \rightarrow X$ in probability and that $(X_n)_{n \geq 1}$ is uniformly integrable, namely that

$$\lim_{K \rightarrow \infty} \sup_{n \geq 1} \mathbb{E}[|X_n| \mathbb{1}_{|X_n| > K}] = 0.$$

Then $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|X_n - X|] \rightarrow 0$; in particular $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$.

Theorem 2.5 allows to recover this last convergence when we only have $X_n \rightarrow X$ in distribution: just replace each X_n by X'_n and X by X' with $X'_n \rightarrow X'$ almost surely (so in probability as well) and notice that uniform integrability is only about the individual law of each X_n , so $(X'_n)_n$ is also uniformly integrable. Hence $\mathbb{E}[X_n] = \mathbb{E}[X'_n] \rightarrow \mathbb{E}[X'] = \mathbb{E}[X]$. However, we also have $\mathbb{E}[|X'_n - X'|] \rightarrow 0$, but not $\mathbb{E}[|X_n - X|] \rightarrow 0$ in general because this depends on the joint distributions.

Similarly, any theorem which assumes that $X_n \rightarrow X$ almost surely and conclude about the behaviour of quantities of the form $\mathbb{E}[f(X_n)]$ also generalises to assuming only $X_n \rightarrow X$ in distribution.

2.2 DISTRIBUTION FUNCTIONS IN HIGHER DIMENSION

From now on, we consider laws and random variables in $\mathbb{R}^d$ for some finite $d \geq 1$. We shall write $x \leq y$ for the partial order $x_i \leq y_i$ for every $1 \leq i \leq d$ and similarly $x < y$ when $x_i < y_i$ for every $1 \leq i \leq d$ (in particular $x < y$ is more restrictive than the combination of $x \leq y$ and $x \neq y$). We shall write $x_n \downarrow x$ to mean that x_n converges to x and $x \leq x_n$ for each n .

Fix $x \leq y$; we shall also use the interval notation:

$$(x, y] = \prod_{i=1}^d (x_i, y_i] = \{u \in \mathbb{R}^d : x < u \leq y\},$$

which we call a *box*. Let then $V(x, y] = \{u \in \mathbb{R}^d : \forall 1 \leq i \leq d, u_i \in \{x_i, y_i\}\}$ denote the 2^d vertices of this box and, for any function F defined on $\mathbb{R}^d$, denote the *increment* of F over this box by:

$$\Delta_{(x,y]}F = \sum_{u \in V(x,y]} (-1)^{\sum_{i=1}^d \mathbb{1}_{u_i=x_i}} F(u).$$

In dimension $d = 1$, we have simply $\Delta_{(x,y]}F = F(y) - F(x)$. In dimension $d = 2$, we have $\Delta_{(x,y]}F = F(y_1, y_2) - F(x_1, y_2) - F(y_1, x_2) + F(x_1, x_2)$.

If μ is a probability measure on $\mathbb{R}^d$, then let

$$F_\mu(x) = \mu((-\infty, x]), \quad x \in \mathbb{R}^d. \quad (2.1)$$

Then applying the inclusion-exclusion principle, we calculate

$$\mu((x, y]) = \Delta_{(x,y]}F_\mu, \quad x < y.$$

DEFINITION 2.9. A *distribution function* F on $\mathbb{R}^d$ is a function defined on $\mathbb{R}^d$ to $[0, 1]$ that satisfies the following properties:

- (i) It has nonnegative increments: for any $x < y$, we have $\Delta_{(x,y]}F \geq 0$.
- (ii) It is right-continuous: For any $x \in \mathbb{R}^d$ and any sequence $x_n \downarrow x$, we have $F(x_n) \rightarrow F(x)$.
- (iii) It has the limits $F(x) \rightarrow 0$ as $\min_{1 \leq i \leq d} x_i \rightarrow -\infty$ and $F(x) \rightarrow 1$ as $\min_{1 \leq i \leq d} x_i \rightarrow \infty$.

EXERCISE 2.10. Prove that the nonnegative increment property implies that the function F is non-decreasing in each variable. Give an example of a function that is nondecreasing in each variable, continuous and with the above limits but does not satisfy the nonnegative increment property.

THEOREM 2.11. If μ is a probability measure on $\mathbb{R}^d$, then F_μ as defined by (2.1) is a distribution function. Conversely, given any distribution function F on $\mathbb{R}^d$, there exists a unique probability measure μ such that $F = F_\mu$.

Proof. As in dimension 1, it is easy and left as an exercise to check that for any probability measure μ , the function F_μ satisfies the three properties of a distribution function. Let us focus on the second claim. Uniqueness again follows from the monotone class lemma since the collection of finite boxes $\{(-x, y], x \leq y\}$ is closed under finite intersections and generates the Borel σ -algebra. The key point is to prove that given F , there exists a measure μ such that $F_\mu = F$.

One possible proof is simply to redo the construction of the Lebesgue measure on $\mathbb{R}^d$ (which would correspond to $F(x) = x$): Start with the set function μ defined on all finite boxes by $\mu((x, y]) = \Delta_{(x,y]}F$; extend it to the algebra of sets that are finite unions of boxes and eventually use Caratheodory's extension theorem to extend it to a measure on $\mathbb{R}^d$, which will be a probability due to the limit of F at ∞ . Instead, as in dimension 1, we rely again on the existence of the Lebesgue measure on the unit interval.

Let F be a distribution function. Fix $n \geq 1$ and consider the discrete space $2^{-n}\mathbb{Z}^d$. We claim that the formula

$$\mu_n(2^{-n}k) = \Delta_{(2^{-n}(k-1), 2^{-n}k]}F, \quad k \in \mathbb{Z}^d,$$

defines a probability measure on $2^{-n}\mathbb{Z}^d$. Indeed, combining disjoint boxes, we see that for $k \leq \ell$, we can decompose $(2^{-n}(k-1), 2^{-n}\ell]$ as a disjoint union of boxes B in such a way that $\Delta_{(2^{-n}(k-1), 2^{-n}\ell]}F$ equals the sum of the $\Delta_B F$'s. Letting $\min_i k_i \rightarrow -\infty$ and $\min_i \ell_i \rightarrow \infty$, using the limits of F then shows that

$\sum_{k \in \mathbb{Z}^d} \mu_n(2^{-n}k) = 1$. Subdividing boxes further shows that for $k \in \mathbb{Z}^d$ and $m < n$, it holds:

$$\mu_m(2^{-m}(k-1, k]) = \mu_m(2^{-m}k) = \mu_n(2^{-m}(k-1, k]).$$

Using again the limit 0 of F at $-\infty$, we deduce that for every $m < n$ and every $k \in \mathbb{Z}^d$, it holds:

$$\mu_n((-\infty, 2^{-m}k]) = F(2^{-m}k).$$

Take the unit interval $(0, 1]$ and for each $n \geq 1$, let us partition it into subintervals I_k^n (open on the left and closed on the right) with length given by $\mu_n(2^{-n}k)$, where k ranges over $\mathbb{Z}^d$. Define a random vector X^n on the probability space $(\Omega, \mathcal{A}, \mathbb{P}) = ((0, 1], \mathcal{B}((0, 1]), \text{Leb})$ by $X^n(\omega) = 2^{-n}k$ for the unique $k \in \mathbb{Z}^d$ such that $\omega \in I_k^n$. Then X^n has law μ_n . The previous display allows to make this construction for all $n \geq 1$ in a consistent way such that

$$X^m - 2^{-m} < X^n \leq X^m, \quad m < n,$$

where we recall that $x \leq y$ when $x_i \leq y_i$ for each $1 \leq i \leq d$, and $x < y$ when $x_i < y_i$ for each $1 \leq i \leq d$. Considering each coordinate i of X^n for ω fixed shows that they are nondecreasing as n increases, and bounded below by $X_i^1(\omega) - 2^{-1}$, so they converge. Hence the entire vector $X^n(\omega)$ converges to some $X(\omega)$. Let $\mu = X_* \text{Leb}$ denote its law.

From $X^n \downarrow X$, we infer that for every $x \in \mathbb{R}^d$, we have $X < x$ if and only if $X^n < x$ for all n large enough; formally $\{X < x\} = \bigcup_{N \geq 1} \bigcap_{n \geq N} \{X^n < x\}$. On the right the union is that of nondecreasing events that are each contained in $\{X^N < x\}$. It then follows from monotonicity of measures that

$$\mathbb{P}(X < x) = \mathbb{P}(X^n < x \text{ for all } n \text{ large enough}) \leq \liminf_{n \geq 1} \mathbb{P}(X^n < x).$$

Similarly, we have $\{X \leq x\} \supset \bigcap_{N \geq 1} \bigcup_{n \geq N} \{X^n \leq x\}$, which is now a nonincreasing union of events that each contain $\{X^N \leq x\}$ so

$$\mathbb{P}(X \leq x) \geq \mathbb{P}(X^n \leq x \text{ infinitely often}) \geq \limsup_{n \geq 1} \mathbb{P}(X^n \leq x).$$

Recall that if $x = 2^{-m}k$ for some $m \geq 1$ and $k \in \mathbb{Z}^d$, then $\mathbb{P}(X^n \leq x) = F(x)$ for all $n \geq m$. For such x and any $N \geq 1$, if we let $x^N = x + 2^{-N}1$, then

$$F(x) = \limsup_{n \geq 1} \mathbb{P}(X^n \leq x) \leq \mathbb{P}(X \leq x) \leq \mathbb{P}(X < x^N) \leq \liminf_{n \geq 1} \mathbb{P}(X^n < x^N) \leq F(x^N).$$

By right-continuity, the rightmost term tends to $F(x)$ as $N \rightarrow \infty$, which shows that $\mathbb{P}(X \leq x) = F(x)$ for every $x \in \{2^{-m}k : m \geq 1, k \in \mathbb{Z}^d\}$ but then further for every $x \in \mathbb{R}^d$ using right-continuity of both sides and density of this set in $\mathbb{R}^d$. $\square$

Our next task is to link weak convergence of probability measures on $\mathbb{R}^d$ and that of their distribution function, just as in dimension 1.

DEFINITION 2.12 (Weak convergence of distribution functions). A sequence $(F_n)_{n \geq 1}$ of functions from $\mathbb{R}^d$ to $\mathbb{R}$ is said to converge *weakly* to a function F when

$$F_n(x) \xrightarrow{n \rightarrow \infty} F(x) \quad \text{for all } x \in C_F,$$

where $C_F = \{x \in \mathbb{R}^d : F \text{ is continuous at } x\}$ denotes the continuity set of F .

Weak convergence of probability measures coincides again with weak convergence of their distribution

functions.

THEOREM 2.13. We have $\mu_n \Rightarrow \mu$ if and only if $F_{\mu_n} \rightarrow F_\mu$ weakly.

Proof. The function F_μ is continuous at $x \in \mathbb{R}^d$ if and only if μ gives no mass to the boundary of the infinite box $(-\infty, x]$ so the direct implication follows from the Portmanteau theorem 1.10. Suppose that $F_{\mu_n} \rightarrow F_\mu$ weakly. Let us fix an open set O of $\mathbb{R}^d$ and let us prove that $\liminf_n \mu_n(O) \geq \mu(O)$. This will prove that $\mu_n \Rightarrow \mu$ by the Portmanteau theorem 1.10 again.

In general F_μ can have uncountably many discontinuity points, but, for each $1 \leq i \leq d$, there can only be countably many values $a_i \in \mathbb{R}$ such that $\mu(\{x \in \mathbb{R}^d : x_i = a_i\}) > 0$. Let B denote the set of all boxes $(-x, y]$ such that both x and y have no such coordinate. Then all vertices of such a box B are continuity points of F_μ , so the weak convergence $F_{\mu_n} \rightarrow F_\mu$ implies (recall the definition of $\Delta_B F$) that

$$\mu_n(B) = \Delta_B F_{\mu_n} \xrightarrow{n \rightarrow \infty} \Delta_B F_\mu = \mu(B).$$

The inclusion-exclusion formula then shows more generally the convergence

$$\mu_n(B_1 \cup \dots \cup B_K) \xrightarrow{n \rightarrow \infty} \mu(B_1 \cup \dots \cup B_K),$$

for any finite collection of such boxes.

By density, the open set O of $\mathbb{R}^d$ can be written as a countable union of such boxes, say $O = \bigcup_{k \geq 1} B_k$. We may now conclude from the previous display that for any $K \geq 1$, we have

$$\liminf_n \mu_n(O) \geq \liminf_n \mu_n(B_1 \cup \dots \cup B_K) = \mu(B_1 \cup \dots \cup B_K) \xrightarrow{K \rightarrow \infty} \mu(O),$$

which is what we want. □

2.3 TIGHTNESS

It turns out that distribution functions always converge weakly along a subsequence. It is important to note however that the limit function F may fail to obey the limits:

$$\lim_{\min_i x_i \rightarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{\min_i x_i \rightarrow \infty} F(x) = 1, \quad (2.2)$$

and in this case it is not a distribution function.

LEMMA 2.14 (Helly's selection principle). Let F_n be a distribution function on $\mathbb{R}^d$ for every $n \geq 1$. Then there exists a function $F : \mathbb{R}^d \rightarrow [0, 1]$ which is right-continuous with nonnegative increments and an increasing sequence of integers $(n_k)_{k \geq 1}$ such that $F_{n_k} \rightarrow F$ weakly.

Proof. Let us enumerate $\mathbb{Q}^d$ as $(r_k)_{k \geq 1}$. Since the sequence $(F_n(r_1))_{n \geq 1}$ takes values in the compact interval $[0, 1]$, then it admits a convergent subsequence, say $(F_{n_{1,k}}(r_1))_{k \geq 1}$ and we denote by $F^\circ(r_1) \in [0, 1]$ its limit. Similarly, the sequence $(F_{n_{1,k}}(r_2))_{k \geq 1}$ has a further convergent subsequence, say $F_{n_{2,k}}(r_2) \rightarrow F^\circ(r_2) \in [0, 1]$ as $k \rightarrow \infty$. By induction, for any $j \geq 1$, there exists $F(r_j) \in [0, 1]$ and a subsequence $F_{n_{j,k}}(r_j) \rightarrow F^\circ(r_j)$ as $k \rightarrow \infty$. Let us write $n_k = n_{k,k}$, then in particular we have

$$F_{n_k}(r_j) \xrightarrow{k \rightarrow \infty} F^\circ(r_j) \quad \text{for all } j \geq 1.$$

Let us extend the limit to $\mathbb{R}^d$ by setting

$$F(x) = \inf\{F^\circ(r) : r \in \mathbb{Q}^d, x \leq r\}.$$

Then F coincides with F° on $\mathbb{Q}^d$ and is right-continuous. In dimension $d = 1$, each function F_n is

nondecreasing, thus so is F° defined on $\mathbb{Q}$ and then extended to F on $\mathbb{R}$. In any dimension d , for every box $(r, s]$ with $r < s$ both in $\mathbb{Q}^d$, we have $\Delta_{(r,s]}F^\circ = \lim_k \Delta_{(r,s]}F_{n_k} \geq 0$ since each F_{n_k} is a distribution function. Now for any given box $(x, y]$ with $x < y$ in $\mathbb{R}^d$, we can find a box $(r, s]$ with $r < s$ both in $\mathbb{Q}^d$ and such that every vertex v of the first box satisfies $v < q$ for the corresponding vertex of the second box, so $q \in \mathbb{Q}^d$, and these vertices are close enough so $|F(v) - F^\circ(q)|$ is arbitrarily small. This implies that $\Delta_{(x,y]}F$ is arbitrarily close to $\Delta_{(r,s]}F^\circ$ and thus nonnegative as well.

It remains to prove that F_{n_k} converges weakly to F . Let $x \in \mathbb{R}^d$ be a continuity point of F . Then there exist $q, r \in \mathbb{Q}^d$ such that $x \in (q, r]$ and close enough to x so $F(q) \leq F(x) \leq F(r)$ are arbitrarily close. Since $F_{n_k}(q) \leq F_{n_k}(x) \leq F_{n_k}(r)$ and the extremities converge to $F(q)$ and $F(r)$ respectively, then we conclude that $F_{n_k}(x) \rightarrow F(x)$ as desired. $\square$

In order that the limit F be a distribution function, it must satisfy the limits (2.2). The problem that can arise in general is that some of the mass escapes to infinity. The next assumption will forbid this issue.

DEFINITION 2.15 (Tightness). A sequence of probability measures $(\mu_n)_{n \geq 1}$ on $\mathbb{R}^d$ is said to be *tight* when for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset \mathbb{R}^d$ such that

$$\mu_n(K_\varepsilon^c) \leq \varepsilon \quad \text{for every } n \text{ large enough.}$$

A sequence of real-valued random variables $(X_n)_{n \geq 1}$ is said to be tight when their laws $(\mathbb{P}_{X_n})_{n \geq 1}$ are, and in this case we also say that their distribution functions $(F_{X_n})_{n \geq 1}$ are tight.

Prohorov's theorem 1.25 in $\mathbb{R}^d$ reads as follows.

THEOREM 2.16. A sequence of distribution functions $(F_n)_{n \geq 1}$ is tight if and only if from every subsequence, one can extract a further subsequence along which it converges weakly to some distribution function.

Proof. Recall from Lemma 2.14 that we may always extract from $(F_n)_n$ a subsequence $(F_{n_k})_k$ which converges weakly, and in fact converges pointwise on all $\mathbb{Q}^d$, to some function F which is right-continuous with nonnegative increments, so we only need to prove that the limit F is a distribution function, that is, that it satisfies (2.2) if and only if the tightness property holds.

First assume tightness of the sequence. Then for every $\varepsilon > 0$, there exists $M_\varepsilon > 0$ such that for all n large enough, we have $\mu_n(\{x \in \mathbb{R}^d : |x| \geq M_\varepsilon\}) < \varepsilon$. Let $x \in \mathbb{R}^d$ be such that $\min_i x_i > M_\varepsilon$. Let $r > x$ be in $\mathbb{Q}^d$, then $F(q) = \lim_k F_{n_k}(q) \geq 1 - \varepsilon$. Letting r tend to x , we obtain by right-continuity that $F(x) \geq 1 - \varepsilon$. On the contrary, if $\min_i x_i < -M_\varepsilon$, then we can find $r \in \mathbb{Q}^d$ such that $x < r$ and $\min_i r_i < -M_\varepsilon$. This implies that $F_{n_k}(r) \leq \varepsilon$ for n_k large enough, so $F(r) \leq \varepsilon$ and then $F(x) \leq \varepsilon$ by letting r tend to x . Hence tightness implies the limits (2.2).

Conversely, assume that the sequence is not tight: there exists $\varepsilon > 0$ such that for every $M > 0$, it holds $\mu_n(\{x \in \mathbb{R}^d : |x| \geq M\}) > \varepsilon$ for infinitely many indices n . Then let M run through the integers and build an increasing sequence of integers $(n_M)_{M \geq 1}$ that satisfies this condition. Let F be any limit along a subsequence of those indices and let us prove that F is not a distribution function. Assume by contradiction that it does satisfy the limits (2.2). Then there exists $K > 0$ such that $F(x) \geq 1 - \varepsilon/8$ if $\min_i x_i \geq K$ and $F(x) \leq \varepsilon 2^{-d}/8$ if $\min_i x_i \leq -K$ and such that the vertices of the box B with coordinates $\pm K$ are all continuity points of F . It follows that $\Delta_B F = \lim_n \Delta_B F_{n_M}$ along our sub-subsequence and also that $\Delta_B F \geq 1 - \varepsilon/8 - (2^d - 1)\varepsilon 2^{-d}/8 \geq 1 - \varepsilon/4$. Consequently $\mu_n(B) = \Delta_B F_{n_M} \geq 1 - \varepsilon/2$ for all n large enough along our sub-subsequence. Take $M_n > K$ for an index n large enough in this sub-subsequence, then we also require that $\mu_n(\{x \in \mathbb{R}^d : |x| > K\}) > \varepsilon$ which adds to more than 1 with $\mu_n(\mathbb{R}^d)$. $\square$

One may reformulate the statement as: A sequence of random vectors $(X_n)_{n \geq 1}$ is tight if and only if

from every subsequence, one can extract a further subsequence along which it converges in distribution to some random variable. In practice, this leads to the following two step scheme to prove convergence in distribution.

COROLLARY 2.17. A sequence μ_n converges weakly to μ if and only if the following two properties hold:

- (i) It is tight,
- (ii) If $(\mu_{n_k})_k$ is any convergent subsequence, then its limit is μ .

Proof. By the previous lemma, tightness is equivalent to the possibility to extract a convergent subsequence from any subsequence. Then it is clear that if the entire sequence converges, then it is tight and any subsequence also converges to the same limit. Conversely, if μ_n does not converge weakly to μ , then there exists a continuous and bounded function f such that $\mu_n(f)$ does not converge to $\mu(f)$ so there exists $\varepsilon > 0$ and infinitely many indices such that $|\mu_n(f) - \mu(f)| > \varepsilon$.

the sequence satisfies the two properties, then for every continuous and bounded function f , the sequence $(\mu_n(f))_n$ in $\mathbb{R}$ is such that $\square$

2.4 CHARACTERISTIC FUNCTIONS

Let us continue the study of weak convergence in $\mathbb{R}^d$ with a very powerful tool, known as the characteristic function, or Fourier transform.

We shall denote the scalar product in $\mathbb{R}^d$ just as the usual product in $\mathbb{R}$, namely we write $ux = \sum_{i=1}^d u_i x_i$ whenever $u = (u_1, \dots, u_d)$ and $x = (x_1, \dots, x_d)$.

We also denote by $|\cdot|$ the associated square norm (as well as the modulus in $\mathbb{C}$). We use t to denote a real number and u for a vector in $\mathbb{R}^d$. We shall denote by 0 the vector $(0, \dots, 0) \in \mathbb{R}^d$. We let $(u_1, \dots, u_d)$ denote the coordinates of u and $(u^n)_{n \geq 1}$ denote a sequence of vectors $(u_1^n, \dots, u_d^n)$. We stress that we use the line notation for vectors when writing in the text, but we think of them as columns when it comes to matrix operations.

DEFINITION 2.18. The *characteristic function* of a random vector X in $\mathbb{R}^d$ is the function $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ defined by:

$$\varphi_X(u) = \mathbb{E}[e^{iuX}] = \mathbb{E}[\cos uX] + i \mathbb{E}[\sin uX].$$

It is well-defined since $x \mapsto e^{iux}$ is continuous and bounded for every given $u \in \mathbb{R}^d$.

The characteristic function corresponds to the Fourier transform of the law of the random vector. It generalises the Fourier transform of a function when the law has a density with respect to the Lebesgue measure. Indeed, for explicit calculations, by the transfer lemma, if X admits a density f_X with respect to the Lebesgue measure in $\mathbb{R}^d$, then

$$\varphi_X(u) = \int_{\mathbb{R}^d} e^{iux} f_X(x) dx,$$

whereas if X takes countably many values, say in $\mathbb{Z}^d$ for example, then

$$\varphi_X(u) = \sum_{x \in \mathbb{Z}^d} e^{iux} \mathbb{P}(X = x).$$

We leave the proof of the basic properties as an exercise.

PROPOSITION 2.19. The characteristic function of any random vector X satisfies the following properties:

- (i) $\varphi_X(0) = 1$,
- (ii) $\varphi_X(-u) = \overline{\varphi_X(u)}$ for every $u \in \mathbb{R}^d$,
- (iii) $|\varphi_X(u)| \leq 1$ for every $u \in \mathbb{R}^d$,
- (iv) $|\varphi_X(u+h) - \varphi_X(u)| \leq \mathbb{E}[|e^{ihX} - 1|]$ for every $u, h \in \mathbb{R}^d$ so φ_X is uniformly continuous.
- (v) If X has dimension d_2 , then for every $d_1 \times d_2$ matrix C and vectors $u, v \in \mathbb{R}^{d_1}$, we have

$$\varphi_{CX+v}(u) = \mathbb{E}[e^{iu(CX+v)}] = \mathbb{E}[e^{iuv+i(C^\top u)X}] = e^{iuv} \varphi_X(C^\top u).$$

In particular, if we let I_d denote the identity matrix of dimension d , then:

- (a) If $d_1 = d_2 = d$ and $C = aI_d$ with $a \in \mathbb{R}$, then $\varphi_{aX+v}(u) = e^{iuv} \varphi_X(au)$.
- (b) Also, if $d_1 = 1$, then $C = c \in \mathbb{R}^{d_2}$ and $CX = cX$, and so $\varphi_{c(X+s)}(t) = e^{ist} \varphi_X(tc)$ for every $s, t \in \mathbb{R}$.
- (vi) If $X^1, \dots, X^n$ are independent random vectors with dimension $d_1, \dots, d_n$ respectively, then for all $u^1 \in \mathbb{R}^{d_1}, \dots, u^n \in \mathbb{R}^{d_n}$, we have

$$\varphi_{(X^1, \dots, X^n)}(u^1, \dots, u^n) = \mathbb{E}\left[\prod_{k=1}^n e^{iu^k X^k}\right] = \prod_{k=1}^n \varphi_{X^k}(u^k).$$

Combined with the previous item, if $d_1 = \dots = d_n = d$, then for every $u \in \mathbb{R}^d$, we have

$$\varphi_{X^1 + \dots + X^n}(u) = \prod_{k=1}^n \varphi_{X^k}(u).$$

Indeed, $X^1 + \dots + X^n = CX$ where C is the $d \times dn$ matrix with all entries equal to 1 and $X = (X^1, \dots, X^n)$ is the concatenation of the X^k 's.

EXERCISE 2.20. Compute the characteristic function of a Gaussian random variable $Z \sim \mathcal{N}(m, \sigma^2)$: for every $t \in \mathbb{R}$, we have

$$\varphi_Z(t) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(itz - \frac{(z-m)^2}{2\sigma^2}\right) dz = \exp\left(itm - \frac{t^2\sigma^2}{2}\right).$$

Consequently, if $Z_1, \dots, Z_d$ are i.i.d. with the law $\mathcal{N}(0, \sigma^2)$, by independence we get for every $u \in \mathbb{R}^d$:

$$\varphi_{Z_1, \dots, Z_d}(u) = \int_{\mathbb{R}^d} \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(iuz - \frac{|z|^2}{2\sigma^2}\right) dz = \exp\left(-\frac{|u|^2\sigma^2}{2}\right). \quad (2.3)$$

Let us see a consequence of this identity which will prove useful.

LEMMA 2.21. Let $Z = (Z_1, \dots, Z_d)$ be i.i.d. standard Gaussian random variables and let X be a random vector independent of Z . Then for every $c \in \mathbb{R}^*$, the vector $Z + cZ$ has density

$$f_c(u) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_X(z) \exp\left(-iuz - \frac{c^2|z|^2}{2}\right) dz$$

with respect to the Lebesgue measure on $\mathbb{R}^d$.

Proof. Fix $g : \mathbb{R}^d \rightarrow [0, \infty)$ measurable and $c \neq 0$. Then since we integrate nonnegative terms,

$$\begin{aligned} \mathbb{E}[g(X + cZ)] &\stackrel{\text{transfer}}{=} \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x + cz) \mathbb{P}_X(dx) \otimes \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|z|^2}{2}\right) dz \\ &\stackrel{\text{Fubini}}{=} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(x + cz) \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|z|^2}{2}\right) dz \right) \mathbb{P}_X(dx) \\ &\stackrel{u=x+cz}{=} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(u) \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|x-u|^2}{2c^2}\right) |c|^{-d} du \right) \mathbb{P}_X(dx). \end{aligned}$$

By (2.3) the exponential term can be rewritten as

$$\begin{aligned} \exp\left(-\frac{|x-u|^2}{2c^2}\right) &= \int_{\mathbb{R}^d} \frac{1}{(2\pi c^{-2})^{d/2}} \exp\left(i(x-u)z - \frac{c^2|z|^2}{2}\right) dz \\ &= \int_{\mathbb{R}^d} \frac{|c|^d}{(2\pi)^{d/2}} \exp(izx) \exp\left(-iuz - \frac{c^2|z|^2}{2}\right) dz. \end{aligned}$$

Using Fubini's theorem for integrable functions now, we get:

$$\begin{aligned} \mathbb{E}[g(X + cZ)] &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(u) \left(\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(izx) \exp\left(-iuz - \frac{c^2|z|^2}{2}\right) dz \right) du \right) \mathbb{P}_X(dx) \\ &= \int_{\mathbb{R}^d} g(u) \left(\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_X(z) \exp\left(-iuz - \frac{c^2|z|^2}{2}\right) dz \right) du \\ &= \int_{\mathbb{R}^d} g(u) f_c(u) du. \end{aligned}$$

Since this holds true for any $g \geq 0$ measurable, then f_c is indeed a density for $X + cZ$. $\square$

Let us end with a useful computation: the moments of a random vector can be obtained by differentiating the characteristic function at 0.

PROPOSITION 2.22 (Moments). Suppose that $\mathbb{E}[|X|^n] < \infty$, then $\varphi_X \in C^n$ and every $u \in \mathbb{R}^d$, every $k \leq n$, and for every $j_1, \dots, j_k \in \{1, \dots, d\}$, it holds

$$\frac{\partial^k}{\partial u_{j_1} \dots \partial u_{j_k}} \varphi_X(u) = i^k \mathbb{E}[X_{j_1} \dots X_{j_k} e^{iuX}].$$

In particular, for $u = 0$, the right hand side equals $i^k \mathbb{E}[X_{j_1} \dots X_{j_k}]$. In dimension $d = 1$, we have in particular

$$\left. \frac{d^k}{du^k} \varphi_X(u) \right|_{u=0} = i^k \mathbb{E}[X^k],$$

for every $k \leq n$.

Proof. It is a matter of exchanging derivation and expectation; we use that $\mathbb{E}[|X|^k] < \infty$ to get the domination $|i^k \prod_{\ell=1}^k X_{j_\ell} e^{iuX}| \leq \prod_{\ell=1}^k |X_{j_\ell}| \in L^1$. $\square$

As the name suggests, the characteristic function characterises the law of a random vector. This is an easy consequence of Lemma 2.21.

THEOREM 2.23. If X and Y have the same characteristic function then they have the same law.

Proof. Let $Z = (Z_1, \dots, Z_d)$ be i.i.d. standard Gaussian random variables and assume that they are independent from X and Y . According to Lemma 2.21, if $\varphi(z) = \mathbb{E}[\exp(izX)] = \mathbb{E}[\exp(izY)]$ denotes the characteristic function of X and of Y , then for every $k \geq 1$, the vector $X + Z/k$ has density

$$u \mapsto \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi(z) \exp\left(-iuz - \frac{|z|^2}{2k^2}\right) dz,$$

on $\mathbb{R}^d$. Since this density only depends on φ , then $Y + Z/k$ has the same density and thus the same law as $X + Z/k$. To conclude, observe that $X + Z/k \rightarrow X$ almost surely and so in distribution; similarly $Y + Z/k \rightarrow Y$ in distribution and since $X + Z/k$ and $Y + Z/k$ have the same law, then the limit law is the same: X has the same law as Y . $\square$

REMARK 2.24. One can recover more explicitly the law of a random variable with a given characteristic function φ . In dimension $d = 1$ to simplify, one always has:

$$\frac{1}{2\pi} \int_{-K}^K \frac{e^{-iat} - e^{-ibt}}{it} \varphi(t) dt \xrightarrow{K \rightarrow \infty} \frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2},$$

where F is the corresponding distribution function. Moreover, if $\int |\varphi| < \infty$, then the law admits a continuous density given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \varphi(t) dt.$$

For a proof, look for “Lévy’s inversion formula” or “inverse Fourier transform”.

A consequence of the previous theorem is that the characteristic function of a vector characterises the independence in the following way, stronger than in Proposition 2.19.

COROLLARY 2.25. Let $X^1, \dots, X^n$ be random vectors in $\mathbb{R}^d$. They are independent if and only if for every $u^1, \dots, u^n \in \mathbb{R}^d$ it holds

$$\mathbb{E}\left[\exp\left(i \sum_{k=1}^n u^k X^k\right)\right] = \prod_{k=1}^n \mathbb{E}[\exp(iu^k X^k)].$$

Proof. The direct implication is clear since if the vectors X^k are independent, then so are the $\exp(iu^k X^k)$ ’s and thus

$$\mathbb{E}\left[\prod_{k=1}^n \exp(iu^k X^k)\right] = \prod_{k=1}^n \mathbb{E}[\exp(iu^k X^k)].$$

Conversely, suppose that this identity holds for all u^k ’s; because the characteristic function characterises the law, then the vector $(X^1, \dots, X^n)$ in $\mathbb{R}^{dn}$ has the same law as the concatenation of n independent vectors so indeed these vectors are independent. $\square$

2.5 CHARACTERISTIC FUNCTIONS AND WEAK CONVERGENCE

In addition to characterising the law and independence, another reason of the success of characteristic functions is that they characterise the convergence in distribution.

THEOREM 2.26. We have $X^n \rightarrow X$ in distribution if and only if $\varphi_{X^n} \rightarrow \varphi_X$ pointwise.

Proof. The direct implication follows from the fact that the function $x \mapsto e^{iux}$ is continuous and bounded for every fixed $u \in \mathbb{R}^d$. For the converse one, we argue as in the proof of Theorem 2.23 and we consider $X^n + Z/k$ where Z is an independent random vector which has the same law as d i.i.d. standard Gaussian

random variables. Recall from Lemma 2.21 that $X^n + Z/k$ has a density

$$f_k^n(u) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_{X^n}(z) \exp\left(-iuz - \frac{|z|^2}{2k^2}\right) dz.$$

If $\varphi_{X^n} \rightarrow \varphi_X$ pointwise, then by dominated convergence, f_k^n converges pointwise to f_k , the density of $X + Z/k$. By a second application of the dominated convergence theorem we infer that for any continuous and bounded function g ,

$$\mathbb{E}[g(X^n + Z/k)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[g(X + Z/k)],$$

i.e. that $X^n + Z/k \rightarrow X + Z/k$ in distribution for any $k \geq 1$ fixed.

Fix g bounded and L -Lipschitz, then for every $k \geq 1$,

$$\begin{aligned} & |\mathbb{E}[g(X^n)] - \mathbb{E}[g(X)]| \\ & \leq \mathbb{E}[|g(X^n + Z/k) - g(X^n)|] + |\mathbb{E}[g(X^n + Z/k)] - \mathbb{E}[g(X + Z/k)]| + \mathbb{E}[|g(X + Z/k) - g(X)|] \\ & \leq \frac{2L}{k} \mathbb{E}[|Z|] + |\mathbb{E}[g(X^n + Z/k)] - \mathbb{E}[g(X + Z/k)]|. \end{aligned}$$

In the last line, the second term tends to 0 as $n \rightarrow \infty$ and further the first one tends to 0 as $k \rightarrow \infty$. Hence $\mathbb{E}[g(X^n)] \rightarrow \mathbb{E}[g(X)]$ for every Lipschitz and bounded function, which is equivalent to the convergence in distribution by Theorem 1.13. $\square$

This theorem shows that, given a sequence $(X^n)_n$ and a candidate X for its limit in distribution, in order to prove the convergence one may rely on the characteristic functions. However sometimes one does not have an a priori candidate X . Suppose that we prove that the characteristic function of X^n does converge pointwise to some limit function φ . According to Theorem 2.16, a way to prove

THEOREM 2.27 (Lévy). Let X^n have characteristic function φ^n for every $n \geq 1$ and suppose that there exists a function $\varphi : \mathbb{R}^d \rightarrow \mathbb{C}$ which is continuous at 0 and such that $\varphi^n \rightarrow \varphi$ pointwise. Then there exists X whose characteristic function is φ and such that $X^n \rightarrow X$ in distribution.

Recall that the characteristic function of a random vector is always continuous at 0 so if φ_n does not converge, or converges to a limit which is not continuous at 0, then X_n does not converge in distribution.

Proof. Let us first consider the dimension $d = 1$. The key point is to prove under the assumptions that the sequence is tight in the sense of Definition 2.15. Indeed, Theorem 2.16 then implies that there exists a random variable X and a subsequence $(X^{n_k})_k$ which converges in distribution to X . According to Theorem 2.26 the characteristic function of X^{n_k} thus converges pointwise to that of X . Since we assume that it converges pointwise to φ , then φ is the characteristic function of X and a second application of Theorem 2.26 shows that X^n converges in distribution to X .

It remains to prove tightness of the sequence $(X^n)_n$. Let us consider the real part of the characteristic function $\operatorname{Re} \varphi^n(t) = \mathbb{E}[\cos(tX^n)]$ in order to deal with real-valued functions. By Fubini's theorem, for every $\delta > 0$ and $n \geq 1$, it holds

$$\delta^{-1} \int_{-\delta}^{\delta} (1 - \mathbb{E}[\cos(tX^n)]) dt = \mathbb{E} \left[\delta^{-1} \int_{-\delta}^{\delta} (1 - \cos(tX^n)) dt \right] = \mathbb{E} \left[2 \left(1 - \frac{\sin(\delta X^n)}{\delta X^n} \right) \right].$$

The function $x \mapsto \sin(x)/x$ is continuous and takes the value 1 at $x = 0$, but is strictly less than 1 in absolute value otherwise. In particular, there exists a constant $c > 0$ such that $1 - \sin(x)/x \geq c/2$ as soon as $|x| \geq 1/2$, and $1 - \sin(x)/x \geq 0$ for any x . We deduce the lower bound:

$$\delta^{-1} \int_{-\delta}^{\delta} (1 - \operatorname{Re} \varphi^n(t)) dt = \delta^{-1} \int_{-\delta}^{\delta} (1 - \mathbb{E}[\cos(tX^n)]) dt \geq \mathbb{E}[c \mathbb{1}_{|\delta X^n| \geq 1/2}] = c \mathbb{P}(|X^n| \geq 1/(2\delta)).$$

Recall that we assume that $\varphi^n \rightarrow \varphi$ pointwise. Then by dominated convergence,

$$\limsup_{n \rightarrow \infty} c \mathbb{P}(|X^n| \geq 1/(2\delta)) \leq \limsup_{n \rightarrow \infty} \delta^{-1} \int_{-\delta}^{\delta} (1 - \operatorname{Re} \varphi^n(t)) dt = \delta^{-1} \int_{-\delta}^{\delta} (1 - \operatorname{Re} \varphi(t)) dt.$$

Note that we have not used yet the assumption that φ is continuous at 0. This continuity (and that of the real part Re), together with the fact that $\varphi(0) = \lim \varphi^n(0) = 1$, imply that the right hand side above further tends to 0 as $\delta \rightarrow 0$. Since $c > 0$ is some absolute constant, this implies that (take $K = 1/(2\delta)$)

$$\limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|X^n| \geq K) = 0,$$

which is the desired tightness.

Now to extend the result to higher dimension, recall that if X^n has characteristic function $u \in \mathbb{R}^d \mapsto \varphi^n(u)$, then for every $1 \leq i \leq d$, the random variable X_i^n has characteristic function $t \in \mathbb{R} \mapsto \varphi_i^n(te_i)$, where e_i is the i 'th vector in the orthonormal basis of $\mathbb{R}^d$. The assumption that φ^n converges to some function φ continuous at 0 implies the same property for φ_i^n , so the laws of $(X_i^n)_{n \geq 1}$ are tight: for every $\varepsilon > 0$, there exists $K_i > 0$ such that $\mathbb{P}(|X_i^n| > K_i) < \varepsilon/d$ for all n large enough. It follows that $\mathbb{P}(|X_i^n| \notin \prod_{i=1}^d [-K_i, K_i]) < \varepsilon$ for all n large enough, i.e. that $(X^n)_{n \geq 1}$ is tight. $\square$

2.6 THE MOMENT METHOD

Let us go back to dimension $d = 1$. We have seen in Proposition 2.22 that we can express the moments of a random variables in terms of its characteristic function. A natural question is: suppose that $\mathbb{E}[|X|^n] < \infty$ for every $n \geq 1$, can we express the characteristic function in terms of the moments and thus characterise entirely the law of X ?

DEFINITION 2.28. A real-valued random variable X has a law that is *characterised by its moments* when $\mathbb{E}[|X|^n] < \infty$ for each $n \geq 1$ and for any random variable Y , if $\mathbb{E}[Y^n] = \mathbb{E}[X^n]$ for all $n \geq 1$, then Y has the same law as X .

Not all (laws of) random variables are characterised by their moments, as shown by the following example.

EXERCISE 2.29. Let Z have the standard Gaussian distribution $\mathcal{N}(0, 1)$, then $X_0 = e^Z$ has density:

$$f_0(x) = \frac{1}{x\sqrt{2\pi}} e^{-\ln(x)^2/2}, \quad x > 0.$$

For every $a \in (-1, 1)$, define the function:

$$f_a(x) = f_0(x)(1 + a \sin(2\pi \log x)), \quad x > 0.$$

Prove that f_a is a probability density. Let X_a have the corresponding law. Prove that $\mathbb{E}[|X_a|^n] < \infty$ for each $n \geq 1$ and then prove that $\mathbb{E}[X_a^n] = \mathbb{E}[X_0^n]$ for all $n \geq 1$.

An easy case where moments do characterise the entire distribution is for bounded random variables.

EXERCISE 2.30. Prove that if X and Y are bounded and satisfy $\mathbb{E}[X^n] = \mathbb{E}[Y^n]$ for all n , then they have the same law.

Let us next provide a weaker sufficient criterion.

PROPOSITION 2.31. Assume that $\mathbb{E}[|X|^n] < \infty$ for all $n \geq 1$ and that the power series

$$\sum_{n \geq 1} \frac{\mathbb{E}[|X|^n]}{n!} z^n$$

has a nonzero radius of convergence. Then X is characterised by its moments.

The idea is to recover the characteristic function from the moments.

Proof. For any complex number z with small enough radius, we may apply Fubini's theorem and write

$$\mathbb{E}[e^{zX}] = \mathbb{E}\left[\sum_{n \geq 0} \frac{(zX)^n}{n!}\right] = \sum_{n \geq 0} \frac{z^n \mathbb{E}[X^n]}{n!}.$$

Both terms on the left and on the right are analytic functions on a strip $\{z \in \mathbb{C} : |\operatorname{Re} z| < \delta\}$ for some small $\delta > 0$, where they must therefore agree. This holds in particular for $\operatorname{Re} z = 0$: the sequence of moments allow to recover the characteristic function. $\square$

THEOREM 2.32. Let X and $(X_n)_{n \geq 1}$ be random variables such that $\mathbb{E}[X_n^k] \rightarrow \mathbb{E}[X^k]$ for each $k \geq 1$ and that X is characterised by its moments, then $X_n \rightarrow X$ in distribution.

Proof. Let us apply the scheme from Corollary 2.17. We first prove that the sequence $(X_n)_{n \geq 1}$ is tight: for every $K > 0$, since $(\mathbb{E}[X_n^2])_{n \geq 1}$ is a convergent sequence, then it is bounded and thus, by the Markov inequality,

$$\mathbb{P}(|X_n| > K) \leq \frac{\sup_{n \geq 1} \mathbb{E}[X_n^2]}{K^2} \xrightarrow{K \rightarrow \infty} 0.$$

According to Theorem 2.16, tightness implies that every subsequence of $(X_n)_{n \geq 1}$ has a convergent sub-subsequence in distribution, and it remains to prove that the only possible limit is X . Let us therefore consider a subsequence $(X_{n_j})_{j \geq 1}$ that converges in distribution to some Y , and let us prove that Y and X have the same law.

According to Skorokhod's representation theorem 2.5, we may construct on some probability space random variables X'_{n_j} with the same law as X_{n_j} for each j , and Y' with the same law as Y , and such that $X'_{n_j} \rightarrow Y'$ almost surely. In addition, for every $k \geq 1$, since $\mathbb{E}[(X'_{n_j})^{2k}] = \mathbb{E}[X_{n_j}^{2k}]$ converges (to $\mathbb{E}[X^{2k}]$), then it is bounded. It follows that $(X'_{n_j})^k \rightarrow (Y')^k$ almost surely and that $((X'_{n_j})^k)_{j \geq 1}$ is uniformly integrable, so $\mathbb{E}[(X'_{n_j})^k] \rightarrow \mathbb{E}[(Y')^k]$. However it also converges to $\mathbb{E}[X^k]$, so we infer that $\mathbb{E}[(Y')^k] = \mathbb{E}[X^k]$ for each $k \geq 1$. Since X is characterised by its moments, then Y' and X have the same law and so finally Y and X have the same law. $\square$

EXAMPLE 2.33. Here are some examples of distributions that are often obtained as limit laws by considering the sequence of moments:

- Gaussian laws in CLT-like results, but also e.g. the number of prime divisors of a random integers (Erdős–Kac theorem)
- Poisson distributions: number of occurrences of subgraphs in a random Erdős–Rényi graph, or number of fixed points in a random permutation,
- The *Rayleigh distribution* for the height of nodes in random rooted trees,
- The *semi-circular law*, with density $\frac{1}{2\pi} \sqrt{4 - x^2} \mathbb{1}_{|x| \leq 2}$, which arises as the limit of the empirical measure on the eigenvalues of some random matrices.

In general, this method is widely used when dealing with combinatorial objects, for which computing

moments can often be achieved by enumeration.

2.7 EXERCISES

EXERCISE 7 (Skorokhod representation in $\mathbb{R}$). Let μ be a probability on $\mathbb{R}$ and let $F : t \mapsto \mu((-\infty, t])$ be its distribution function. Define its generalised inverse $G : [0, 1] \rightarrow [-\infty, \infty]$ by:

$$G(u) = \inf\{t \in \mathbb{R} : F(t) > u\}, \quad u \in [0, 1].$$

- (i) Prove that for each $u \in [0, 1]$, the set $\{t \in \mathbb{R} : F(t) > u\}$ is an interval (or is empty), which has left edge $G(u)$ and right edge ∞ . Prove that G is non-decreasing and left-continuous at every $u \in (0, 1)$.
- (ii) Prove that for every $t \in \mathbb{R}$ and $u \in [0, 1]$, we have $G(u) \leq t \iff F(t) \geq u$. Deduce that if U has the uniform distribution on $[0, 1]$, then $G(U)$ has distribution function F , and hence law μ .
- (iii) For every $n \geq 1$, let μ_n be a probability measure on $\mathbb{R}$, with distribution function F_n whose generalised inverse is G_n . Suppose that $\mu_n \rightarrow \mu$ weakly. Prove that for every $u \in (0, 1)$, it holds that

$$G(u) \leq \liminf_{n \rightarrow \infty} G_n(u) \leq \limsup_{n \rightarrow \infty} G_n(u) \leq G(u+),$$

where $G(u+)$ denotes the right-limit of G at u .

- (iv) Construct on some probability space a random variable X with law μ and random variables X_n with law μ_n for each $n \geq 1$ such that $X_n \rightarrow X$ almost surely.

EXERCISE 8 (Yet another property of Gaussian random variables). Let X and Y be i.i.d. real-valued random variables with $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$ and such that $X - Y$ and $X + Y$ are independent. Denote by ϕ the characteristic function of X (and Y) and let $\psi : t \mapsto \phi(t)/\phi(-t)$.

- (i) Recall the series expansion of ϕ at 0, deduce that of ψ .
- (ii) Show that for any $t \in \mathbb{R}$, it holds $\phi(2t) = \phi(t)^3 \phi(-t)$.
- (iii) Show that for any $n \geq 1$ and any $t \in \mathbb{R}$, it holds $\psi(t) = \psi(t/2^n)^{2^n}$. Deduce that $\psi(t) = 1$ for any $t \in \mathbb{R}$.
- (iv) Conclude that X and Y are standard Gaussian random variables.

EXERCISE 9 (Gaussian stability).

- (i) Let X have the standard Gaussian law; recall its characteristic function ϕ_X and the series expansion at 0 of the latter. Deduce the moments $\mathbb{E}[X^k]$ for all $k \geq 0$.
- (ii) Let $(X_n)_{n \geq 1}$ be a sequence of Gaussian random variables with mean m_n and variance σ_n^2 respectively and assume that it converges to some real random variable X that is almost surely finite. Show successively that:
 - (a) The sequence $(\sigma_n^2)_{n \geq 1}$ converges to some limit $\sigma^2 \in [0, \infty)$.
 - (b) The sequence $(m_n)_{n \geq 1}$ is bounded.
 - (c) The sequence $(m_n)_{n \geq 1}$ converges to some limit $m \in \mathbb{R}$.
 - (d) The law of X is the Gaussian law with mean m and variance σ^2 when $\sigma^2 > 0$ and the Dirac mass at m when $\sigma^2 = 0$.

EXERCISE 10. Prove that a collection $(\mu_i : i \in I)$ of probability measures on $\mathbb{R}^d$ is tight if and only if there exists a function $f : \mathbb{R}^d \rightarrow [0, \infty)$ such that both

$$\lim_{|x| \rightarrow \infty} f(x) = \infty \quad \text{and} \quad \sup_{i \in I} \int f \, d\mu_i < \infty.$$

EXERCISE 11. Let $(X_n)_{n \geq 1}$ be random vectors in $\mathbb{R}^d$. Prove that this sequence is tight if and only if $c_n X_n \rightarrow 0$ in probability for every sequence $(c_n)_{n \geq 1}$ of positive real numbers such that $c_n \rightarrow 0$

EXERCISE 12. Let $(\mu_n)_{n \geq 1}$ and μ be probability measures on $\mathbb{R}^d$, with characteristic function $\hat{\mu}_n$ and $\hat{\mu}$.

- (i) Suppose that $(\mu_n)_{n \geq 1}$ is tight. Prove that the characteristic function $(\hat{\mu}_n)_{n \geq 1}$ are uniformly equicontinuous, namely that: for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $n \geq 1$ we have for every $u, v \in \mathbb{R}^d$:

$$|u - v| < \delta \implies |\hat{\mu}_n(u) - \hat{\mu}_n(v)| < \varepsilon.$$

- (ii) Suppose that $\mu_n \Rightarrow \mu$. Prove that $\hat{\mu}_n$ converges to $\hat{\mu}$ uniformly on compact sets.

- (iii) Give an example in which $\mu_n \Rightarrow \mu$ but $\hat{\mu}_n$ does not converge to $\hat{\mu}$ uniformly on $\mathbb{R}^d$.

EXERCISE 13 (Glivenko–Cantelli’s theorem). Let $(X_n)_{n \geq 1}$ be i.i.d. real-valued random variables with common distribution μ and distribution function $F(t) = \mathbb{P}(X_1 \leq t)$. Recall from Exercise 3 the *empirical measure*, which is the random probability measure μ_n defined by

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}.$$

We saw there that μ_n converges weakly to μ and the goal of this exercise is to prove a stronger convergence. For this, Let F_n denote the distribution function of μ_n , defined by:

$$F_n(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k \leq t}, \quad t \in \mathbb{R}.$$

The *Glivenko–Cantelli theorem* states that, whatever μ , we have

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = 0 \quad \text{almost surely.}$$

- (i) Prove (without referring to Exercise 3) the weaker result that $F_n(t) \rightarrow F(t)$ almost surely for each fixed $t \in \mathbb{R}$.
- (ii) Prove the uniform convergence when μ is the uniform law on $[0, 1]$.
- (iii) Prove the uniform convergence in the general case.

EXERCISE 14 (On the moment method). If X is a real-valued random variable such that $\mathbb{E}[|X|^k] < \infty$, we define its k ’th *factorial moment* as:

$$\mathbb{E}[(X)_k] = \mathbb{E}[X(X-1) \cdots (X-k+1)].$$

- (i) Compute all factorial moments of a Poisson distributed random variable with parameter $\lambda > 0$ and prove that this law is characterised by its moments.
- (ii) Let $X = \sum_{i \in I} \mathbb{1}_{A_i}$ be a random variable which takes the form of the sum of indicator functions of events. Prove that

$$\mathbb{E}[(X)_k] = \sum_{i_1, \dots, i_k \in I} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}),$$

where the sum is over *pairwise distinct* indices.

- (iii) Let us see some applications. Use these questions (and the course) to prove the following convergences.
- (a) Random variables with the binomial law with parameters n and p_n converge in distribution to a Poisson distributed random variable with parameter $\lambda > 0$ when $np_n \rightarrow \lambda$.
 - (b) Let $\sigma_n \in \mathfrak{S}_n$ be a permutation of $\{1, \dots, n\}$ sampled uniformly at random, then the number of fixed points (those $i \in \{1, \dots, n\}$ such that $\sigma(i) = i$) converges in distribution to a Poisson distributed random variable with parameter 1.
 - (c) Let $G(n, p_n)$ denote the graph on the vertices $1, \dots, n$ in which every possible edge is present independently of the others with probability p_n (i.e. the bond percolation on the complete graph). Then the number of triangles, i.e. triples of vertices $i < j < k$ which are linked together by edges, converges in distribution to a Poisson distributed random variable with parameter $\lambda > 0$ when $np_n \rightarrow \lambda$.

CHAPTER 3

INTERMEZZO: PRODUCT SPACES

Before going further, let us first review some facts about product spaces. The entire discussion is valid for product of different spaces, of the form $\prod_{i \in I} E_i$, but we chose to restrict the presentation to the case $E_i = E$ of a fixed space since only this case will be used. We shall present the product σ -algebra and the product topology (and their relations), as well as weak convergence in the space of sequence $E^{\mathbb{N}}$.

CONTENTS

3.1	Product σ -algebra	36
3.2	Extension theorems	39
3.3	Product topology	42
3.4	Weak convergence of sequences	43
3.5	Exercises	44

Henceforth I is any (nonempty) set and, for the moment, we consider $(E, \mathcal{E})$ an arbitrary measurable space (with at least two elements). We denote by $E^I = \prod_{i \in I} E$ the product set, that is, the set of collections $x = (x_i)_{i \in I}$ where $x_i \in E$ for every $i \in I$, or equivalently, the set of functions $x : I \rightarrow E$.

3.1 PRODUCT σ -ALGEBRA

DEFINITION 3.1. For every subset $J \subset I$, we let $\pi_J : E^I \rightarrow E^J$ denote the projection

$$\pi_J((x_i)_{i \in I}) = (x_i)_{i \in J}.$$

In the case of a singleton $J = \{t\}$, we simply write π_t for $\pi_{\{t\}}$. These projections are also called *evaluation* maps in the context of functions.

DEFINITION 3.2. Let $J \subset I$. A *cylinder* indexed by J is a subset of E^I of the form:

$$\pi_J^{-1}(A_J) = \{x = (x_i)_{i \in I} : (x_i)_{i \in J} \in A_J\} \quad \text{where} \quad A_J \subset E^J.$$

A *rectangle* indexed by J is the particular case of a cylinder, of the form:

$$\pi_J^{-1}\left(\prod_{j \in J} A_j\right) \quad \text{where} \quad A_j \subset E \text{ for every } j \in J.$$

A cylinder or a rectangle is called *countable*, resp. *finite*, when J is (at most) countable, resp. finite. When J is a singleton, cylinders and rectangles coincide and we speak of *one-dimensional* cylinders/rectangles.

In a sense, we equip the product space with the smallest σ -algebra that contains the measurable sets of the base space, seen as embedded in the product space.

DEFINITION 3.3. The *product σ -algebra* on E^I is the σ -algebra generated by the measurable one-dimensional cylinders, namely

$$\mathcal{E}^{\otimes I} = \sigma(\pi_i^{-1}(A) : i \in I, A \in \mathcal{E}).$$

It is by definition the smallest σ -algebra that makes the projection π_i measurable for every $i \in I$.

DEFINITION 3.4. For every $J \subset I$, we let

$$\mathcal{R}_J = \left\{ \pi_J^{-1} \left(\prod_{j \in J} A_j \right) : A_j \in \mathcal{E} \text{ for every } j \in J \right\} \quad \text{and} \quad \mathcal{C}_J = \left\{ \pi_J^{-1}(A_J) : A_J \in \mathcal{E}^{\otimes J} \right\},$$

denote respectively the set of *measurable rectangles* and *cylinders* indexed by J . We also let

$$\mathcal{R}^{\text{fi}} = \bigcup_{J \subset I \text{ finite}} \mathcal{R}_J, \quad \mathcal{R}^{\text{co}} = \bigcup_{J \subset I \text{ countable}} \mathcal{R}_J, \quad \mathcal{C}^{\text{fi}} = \bigcup_{J \subset I \text{ finite}} \mathcal{C}_J, \quad \text{and} \quad \mathcal{C}^{\text{co}} = \bigcup_{J \subset I \text{ countable}} \mathcal{C}_J,$$

denote respectively the set of measurable finite and countable rectangles and cylinders.

The collection of all finite measurable rectangles is a π -system (stable under finite intersections) and the next proposition shows that it generates the product σ -algebra.

PROPOSITION 3.5. It holds:

$$\mathcal{E}^{\otimes I} = \sigma(\mathcal{R}^{\text{fi}}) = \sigma(\mathcal{R}^{\text{co}}).$$

In particular, if I is countable, then

$$\mathcal{E}^{\otimes I} = \sigma \left(\prod_{i \in I} A_i; A_i \in \mathcal{E} \text{ for each } i \in I \right).$$

Proof. Recall that $\mathcal{E}^{\otimes I}$ is generated by the one-dimensional rectangles, of the form $\pi_i^{-1}(A_i)$ where $A_i \in \mathcal{E}$ and i runs over I , so clearly

$$\mathcal{E}^{\otimes I} \subset \sigma(\mathcal{R}^{\text{fi}}) \subset \sigma(\mathcal{R}^{\text{co}}).$$

In order to prove the converse inclusions, it suffices to show that $\mathcal{E}^{\otimes I}$ contains $\mathcal{R}^{\text{co}}$. This follows from the identity:

$$\pi_J^{-1} \left(\prod_{j \in J} A_j \right) = \bigcap_{j \in J} \pi_j^{-1}(A_j).$$

If $A_j \in \mathcal{E}_j$ for every $j \in J$ and if J is countable, then each $\pi_j^{-1}(A_j) \in \mathcal{E}^{\otimes I}$ and so does their intersection. $\square$

The situation is simpler for cylinders: let us now prove that the product σ -algebra is actually made of the measurable countable cylinders. Therefore $\mathcal{E}^{\otimes I}$ is the σ -algebra of sets that only depend on countably many coordinates.

PROPOSITION 3.6. It holds:

$$\mathcal{E}^{\otimes I} = \mathcal{C}^{\text{co}}.$$

Proof. Since $\mathcal{C}^{\text{co}}$ clearly contains every one-dimensional rectangles, then we need to prove that it is a σ -algebra and it is contained in $\mathcal{E}^{\otimes I}$. Let us first prove this inclusion. Fix a countable subset $J \subset I$ and notice that the collection

$$\{A_J \subset E^J : \pi_J^{-1}(A_J) \in \mathcal{E}^{\otimes I}\},$$

is a σ -algebra on E^J since

$$\pi_J^{-1}(A_J)^c = \pi_J^{-1}(A_J^c) \quad \text{and} \quad \pi_J^{-1}\left(\bigcup_{n \geq 1} A_J^n\right) = \bigcup_{n \geq 1} \pi_J^{-1}(A_J^n).$$

By Proposition 3.5, it contains all the sets of the form $A_J = \prod_{j \in J} A_j$ with $A_j \in \mathcal{E}_j$ for every $j \in J$, and this collection generates $\mathcal{E}^{\otimes J}$. Hence it contains this σ -algebra, that is:

$$\pi_J^{-1}(A_J) \in \mathcal{E}^{\otimes I} \quad \text{for every} \quad A_J \in \mathcal{E}^{\otimes J},$$

which is exactly saying $\mathcal{C}_J \subset \mathcal{E}^{\otimes I}$; since this holds for every countable $J \subset I$, then

$$\mathcal{C}^{\text{co}} = \bigcup_{J \subset I \text{ countable}} \mathcal{C}_J \subset \mathcal{E}^{\otimes I}.$$

It remains to prove that $\mathcal{C}^{\text{co}}$ is a σ -algebra. It contains the empty set and is stable by complement since each $\mathcal{C}_J$ is, but when considering unions, as opposed to what we had above, the A_J^n 's may have different indices. For every n , let $J_n \subset I$ and let $A_{J_n}^n \subset E^{J_n}$, then for $J = \bigcup_n J_n$, we have

$$\bigcup_{n \geq 1} \pi_{J_n}^{-1}(A_{J_n}^n) = \pi_J^{-1}(\pi_{J \rightarrow J_n}^{-1}(A_{J_n}^n)),$$

where $\pi_{J \rightarrow J_n}$ is the projection from J to J_n . If each J_n is countable, then so is J , so the previous argument applied to $I = J$ and $J = J_n$ shows that if each $A_{J_n}^n \in \mathcal{E}^{\otimes J_n}$, then $\pi_{J \rightarrow J_n}^{-1}(A_{J_n}^n) \in \mathcal{E}^{\otimes J}$, so finally their union over n also belongs to $\mathcal{E}^{\otimes J}$ and the above display thus belongs to $\mathcal{C}_J$. $\square$

The importance of this product σ -algebra stems from the following result. Recall that a random variable is a measurable function on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The first part of the following proposition shows that a collection of random variables can alternatively be seen as a single random variable on the space of collections. This allows to make sense of the law of the entire collection. The second part of the statement shows that this law is entirely characterised by the joint law of the evaluations on *finite* subsets, which is much more tractable than the entire collection. The very last item reduces further the effort by evaluating these joint laws at product sets only.

THEOREM 3.7. Let $X : \Omega \rightarrow E^I$ be a function, or equivalently, a collection $X = (X_i)_{i \in I}$ of functions $X_i : \Omega \rightarrow E$. Then the following are equivalent:

- (i) $X = (X_i)_{i \in I}$ is measurable for $\mathcal{E}^{\otimes I}$
- (ii) $(X_i)_{i \in J}$ is measurable for $\mathcal{E}^{\otimes J}$ for every $J \subset I$ finite (or also at most countable).
- (iii) X_i is measurable for $\mathcal{E}$ for every $i \in I$.

Also, if $X = (X_i)_{i \in I}$ and $Y = (Y_i)_{i \in I}$ are two such random variables, then the following are equivalent:

- (i) the law of $(X_i)_{i \in I}$ and $(Y_i)_{i \in I}$ coincide on $\mathcal{E}^{\otimes I}$
- (ii) the law of $(X_i)_{i \in J}$ and $(Y_i)_{i \in J}$ coincide on $\mathcal{E}^{\otimes J}$ for every $J \subset I$ finite (or also at most countable)
- (iii) the law of $(X_i)_{i \in J}$ and $(Y_i)_{i \in J}$ coincide on $\mathcal{R}^{\text{fi}}$.

Proof. For the first part, that (i) implies each other assertion follows by measurability of the projections π_i and π_J and composition of measurable maps. Conversely, if each X_i is measurable for $\mathcal{E}$, then for every $B \in \mathcal{E}$,

$$\{X \in \pi_i^{-1}(B)\} = \{X_i \in B\} \in \mathcal{A}.$$

This is enough to that X is measurable for $\sigma(\pi_i^{-1}(B) : i \in I, B \in \mathcal{E}) = \mathcal{E}^{\otimes I}$.

For the second part, that (i) holds means that $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ for every $B \in \mathcal{E}^{\otimes I}$. Then (ii) is a particular case when $B \in \pi_J^{-1}(A)$ with $A \in \mathcal{E}^{\otimes J}$, and (iii) is a further particular case of product sets $A = \prod_{i \in J} A_i$. Conversely, since the collection of measurable finite rectangles is a π -system that generates $\mathcal{E}^{\otimes I}$, then a probability measure on $\mathcal{E}^{\otimes I}$ is uniquely characterised by its values on this collection. $\square$

DEFINITION 3.8. A random variable $X = (X_i)_{i \in I}$ with values in $(E^I, \mathcal{E}^{\otimes I})$ is called a *stochastic process* indexed by I and with values in E . The family of laws $(X_i)_{i \in J}$ for all finite subsets of indices $J \subset I$ is called the collection of the *finite-dimensional marginal laws* of the process.

REMARK 3.9. In terms of stochastic processes, the second part in the previous proposition shows that for $(X_i)_{i \in I}$ and $(Y_i)_{i \in I}$ to have the same law, it is necessary and sufficient that they have the same finite-dimensional marginal laws, and one can even restrict to product sets, i.e.

$$\mathbb{P}(X_i \in A_i \text{ for every } i \in J) = \mathbb{P}(Y_i \in A_i \text{ for every } i \in J)$$

for every finite $J \subset I$ and every $A_i \in \mathcal{E}$ for each $i \in J$.

The product σ -algebra has many advantages, but it also has limitation. Take as index set $I = [0, 1]$ and as state space $E = \mathbb{R}$, then Exercise 18 shows that $\mathcal{B}(\mathbb{R})^{\otimes [0,1]}$ does not contain the set of continuous functions. Therefore, given a real-valued stochastic process $X = (X_t)_{t \in [0,1]}$, it does not make sense to ask the probability that it is continuous! For this reason, we shall deal with continuous paths processes in a different way.

3.2 EXTENSION THEOREMS

Another advantage of the product σ -algebra is to provide a priori the existence of probability measures on the product set with prescribed projections on each finite product. Precisely, suppose that for every finite subset $J \subset I$, we are given a probability measure μ^J on $(E^J, \mathcal{E}^{\otimes J})$, does there exists a probability measure μ on $(E^I, \mathcal{E}^{\otimes I})$ such that the image measure of μ by the projection π_J coincides with μ^J for every J finite? Namely such that

$$\mu(\pi_J^{-1}(A_J)) = \mu^J(A_J) \quad (3.1)$$

for every $J \subset I$ finite and every $A_J \in \mathcal{E}^{\otimes J}$?

For this to hold, the measures μ^J must necessarily satisfy the following consistency relation: for every finite sets of indices $K \subset J$ and every $A_K \in \mathcal{E}^{\otimes K}$, we must have

$$\mu^J(\pi_K^{-1}(A_K)) = \mu^K(A_K). \quad (3.2)$$

This condition is not sufficient in general and Andersen & Jessen produced in 1948 an example in which I is countable (but E is wild!). However it is sufficient under some regularity conditions on the space E . This is known as the *(Daniell-)Kolmogorov extension theorem*. The statement is more general and applies to product of possibly different spaces E_i , with each of them being a Polish space; the proof is exactly the same as the one we give.

THEOREM 3.10 (Kolmogorov's extension). If E is a separable and complete metric space and $\mathcal{E} = \mathcal{B}(E)$ is its Borel σ -algebra, then under the consistency relation (3.2), there exists a unique probability measure μ on the infinite product space $(E^I, \mathcal{E}^{\otimes I})$ that solves (3.1).

Recall that an *algebra* over a set S is a collection $\mathcal{A}$ of subsets of S satisfying:

- (i) $S \in \mathcal{A}$, (ii) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$, (iii) if $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$.

The last property extends by induction to any *finite* union, which distinguishes an algebra from a σ -algebra, for which it holds for countable unions. For such an algebra, a function $\mu : \mathcal{A} \rightarrow [0, \infty]$ is said to be *finitely additive* when $\mu(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mu(A_k)$ for every $n \geq 1$ and disjoint $A_1, \dots, A_n \in \mathcal{A}$. The function μ is said to be σ -*additive* when for every countable collection $(A_n)_{n \geq 1}$ of disjoint elements of $\mathcal{A}$ such that $\bigcup_n A_n \in \mathcal{A}$, we have $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$. The function μ is said to be σ -finite when there exists a countable collection $(A_n)_{n \geq 1}$ of elements of $\mathcal{A}$ such that $\bigcup_n A_n = S$ and $\mu(A_n) < \infty$ for every $n \geq 1$. The key result to extension problems is the following that we state without proof.

THEOREM 3.11 (Carathéodory's extension). Let $\mathcal{A}$ be an algebra on S and let $\mu : \mathcal{A} \rightarrow [0, \infty]$ be σ -additive and with $\mu(\emptyset) = 0$. Then there exists a measure μ^* on $\sigma(\mathcal{A})$ that agrees with μ on $\mathcal{A}$. If μ is σ -finite, then μ^* is unique and σ -finite as well.

This result is typically used to prove the existence of the Lebesgue measure, by extending the notion of length, or volume in higher dimension, from intervals (more precisely the algebra of the sets that are finite unions of intervals) to the Borel σ -algebra they generate.

Proof of Kolmogorov's extension theorem 3.10. We shall argue by applying Carathéodory's extension theorem 3.11. The algebra we shall use is the collection $\mathcal{C}^{\text{fi}}$ of finite measurable cylinders, that indeed generates $\mathcal{E}^{\otimes I}$. Then define μ on $\mathcal{C}^{\text{fi}}$ by (3.1), namely: if $B \in \mathcal{C}^{\text{fi}}$ takes the form $B = \pi_J^{-1}(A_J)$ for some $J \subset I$ finite and $A_J \in \mathcal{E}^{\otimes J}$, then set

$$\mu(B) = \mu^J(\pi_J^{-1}(A_J)).$$

The representation of B as $B = \pi_J^{-1}(A_J)$ may not be unique. However if $B = \pi_J^{-1}(A_J) = \pi_{J'}^{-1}(A_{J'})$, then let $A_{J \cup J'} = \pi_{J \cup J'}(B)$ and observe that A_J is the image of $A_{J \cup J'}$ by the projection from $E^{J \cup J'}$ to E^J , and similarly for $A_{J'}$. The consistency relation (3.2) then implies that

$$\mu^J(A_J) = \mu^{J \cup J'}(A_{J \cup J'}) = \mu^{J'}(A_{J'}).$$

This shows that the definition of μ does not depend on the choice of the representation of $B \in \mathcal{C}^{\text{fi}}$ as $B = \pi_J^{-1}(A_J)$, so μ is well-defined on $\mathcal{C}^{\text{fi}}$. In addition, since each μ^J is a probability measure, then $\mu(E^I) = 1$ and $\mu(\emptyset) = 0$. Should we extend μ to a probability on $\sigma(\mathcal{C}^{\text{fi}}) = \mathcal{E}^{\otimes I}$, then it would solve (3.1) by construction.

The function μ inherits finite additivity from the μ^J 's, but σ -additivity requires more work. Let $(B_n)_{n \geq 1}$ be disjoint elements of $\mathcal{C}^{\text{fi}}$ such that $B = \bigcup_{n \geq 1} B_n$ also belongs to $\mathcal{C}^{\text{fi}}$. For every $n \geq 1$, let $C_n = B \setminus \bigcup_{k=1}^n B_k$. Since $\mathcal{C}^{\text{fi}}$ is an algebra, then $C_n \in \mathcal{C}^{\text{fi}}$, these sets are non-increasing, with limit $\bigcap_{n \geq 1} C_n = \emptyset$. Finite additivity of μ shows that

$$\mu(B) = \sum_{k=1}^n \mu(B_k) + \mu(C_n).$$

Therefore all it remains to prove is that $\mu(C_n) \rightarrow 0$ as $n \rightarrow \infty$. Assume for the sake of contradiction that it this fails, so we may, and do, fix some $\varepsilon > 0$ such that $\mu(C_n) > \varepsilon$ for all $n \geq 1$.

Recall that we assume that the space E is separable and complete; then the same holds for E^J for any finite J . This implies by Exercise 1 that the probability measure μ^J on E^J satisfies the inner regularity:

$$\mu^J(A) = \sup\{\mu^J(K) : K \subset A \text{ compact of } E^J\}, \quad A \in \mathcal{E}^{\otimes J}.$$

Since $C_n \in \mathcal{C}^{\text{fi}}$, then it takes the form $C_n = \pi_{J_n}^{-1}(A_n)$; replacing J_n by $\bigcup_{k=1}^n J_k$ if necessary, let us assume that they are nondecreasing. Since the C_n 's are nonincreasing, then $A_{n+1} \subset \pi_{J_n+1 \rightarrow J_n}^{-1}(A_n)$. Inner regularity

provides for each ℓ a compact set $K_\ell \subset A_\ell$ such that

$$\mu^{J_\ell}(K_\ell) \geq \mu^{J_\ell}(A_\ell) - \varepsilon 2^{-\ell-1}.$$

The set

$$K'_n = \bigcap_{\ell=1}^n \pi_{J_n \rightarrow J_\ell}^{-1}(K_\ell) \subset K_n \subset A_n$$

is compact. In addition, we have $A_n \subset \bigcap_{\ell=1}^n \pi_{J_n \rightarrow J_\ell}^{-1}(A_\ell)$ so

$$\begin{aligned} \mu^{J_n}(A_n) - \mu^{J_n}(K'_n) &\leq \mu^{J_n} \left(\bigcap_{\ell=1}^n \pi_{J_n \rightarrow J_\ell}^{-1}(A_\ell) \setminus \bigcap_{\ell=1}^n \pi_{J_n \rightarrow J_\ell}^{-1}(K_\ell) \right) \\ &\leq \mu^{J_n} \left(\bigcup_{\ell=1}^n (\pi_{J_n \rightarrow J_\ell}^{-1}(A_\ell) \setminus \pi_{J_n \rightarrow J_\ell}^{-1}(K_\ell)) \right) \\ &\leq \sum_{\ell=1}^n (\mu^{J_\ell}(A_\ell) - \mu^{J_\ell}(K_\ell)) \\ &\leq \sum_{\ell=1}^n \varepsilon 2^{-\ell-1} \leq \varepsilon/2. \end{aligned}$$

Recall that we assume that $\mu^{J_n}(A_n) \geq \varepsilon$, then

$$\mu^{J_n}(K'_n) \geq \varepsilon/2,$$

and therefore $K'_n \neq \emptyset$. Note also that $K'_{n+1} \subset \pi_{J_{n+1} \rightarrow J_n}^{-1}(K'_n)$. Define finally $L_n = \pi_{J_n}^{-1}(K'_n)$, which is a subset of C_n , and therefore $\bigcap_{n \geq 1} L_n \subset \bigcap_{n \geq 1} C_n = \emptyset$. Choose for each $n \geq 1$ an element $x_n \in L_n$. Note that $\pi_{J_1}(x_n) \in K'_1$ for all $n \geq 1$ and since the latter set is compact then we extract a subsequence $(x_{1,n})_{n \geq 1}$ whose projection on J_1 converges in K'_1 . Then $\pi_{J_2}(x_{1,n}) \in K'_1$ for all $n \geq 2$ so, similarly, we extract a further subsequence $(x_{2,n})_{n \geq 1}$ whose projections on both J_1 and J_2 converge. A diagonal argument allows to build a subsequence $(y_n)_{n \geq 1}$ of elements of E^I whose projections on any J_ℓ converge to a limit $(z_i)_{i \in J_\ell}$ that belongs to K'_ℓ . This constructs an element $(z_i)_{i \in \bigcup_{\ell \geq 1} J_\ell}$ in the product set $E^{\bigcup_{\ell \geq 1} J_\ell}$. For those indices $i \in I \setminus \bigcup_{\ell \geq 1} J_\ell$, define z_i as an arbitrary element of E . Then $z = (z_i)_{i \in I}$ belongs to L_n for every n by construction, which contradicts the fact that $\bigcap_{n \geq 1} L_n = \emptyset$. $\square$

A particular case of consistent probability measures are given by product measures, which translate in probability into independence. Theorem 3.10 thus proves in particular the existence of the law of an arbitrary collection of independent random variables, each with a prescribed individual laws, provided the space is separable and complete. It turns out that with product measures, no assumption is needed for the space, and, again, this space could even depend on the index. We shall not prove the following result, although a rather simple and short proof was given by Sadahiro Saeki in an article from 1996, also based on Carathéodory's extension theorem.

THEOREM 3.12. Let $(E, \mathcal{E})$ be any measurable space and for every $i \in I$, let μ^i be any probability measure on this space. Then there exists a unique probability measure μ on the product space $(E^I, \mathcal{E}^{\otimes I})$ such that

$$\mu \left(\pi_J^{-1} \left(\prod_{i \in J} A_i \right) \right) = \prod_{i \in J} \mu^i(A_i)$$

for every $J \subset I$ finite and every $A_i \in \mathcal{E}$ for each $i \in J$.

This probability is usually denoted by $\mu = \bigotimes_{i \in I} \mu^i$ and it represents the law of independent random variables X_i , each one having law μ^i respectively.

3.3 PRODUCT TOPOLOGY

In order to speak about weak convergence of probability measures on E^I , we need a topology on this space. Suppose then that E is endowed with a metric d . The product topology is defined in a similar way to the product σ -algebra.

DEFINITION 3.13. The *product topology* is the coarsest topology on E^I that makes each projection $\pi_i : E^I \rightarrow E$ continuous. A basis of open sets is the collection of all finite open rectangles:

$$\left\{ \pi_J^{-1} \left(\prod_{j \in J} A_j \right) : J \subset I \text{ finite, } A_j \text{ open for every } j \in J \right\}.$$

REMARK 3.14. As for the product σ -algebra, this topology has the property that a function f from a topological space to E^I is continuous if and only if for every $i \in I$, its i 'th component $f_i = \pi_i \circ f$ is continuous.

Now let us equip E with the Borel σ -algebra $\mathcal{B}(E) = \sigma(\text{open sets})$. The product space E^I carries then two natural σ -algebras: its Borel σ -algebra $\mathcal{B}(E^I)$ induced by the product topology and the product σ -algebra $\mathcal{B}(E)^{\otimes I}$. Recall that a topological space is said to be separable when it contains a countable dense set.

PROPOSITION 3.15. The inclusion

$$\mathcal{B}(E)^{\otimes I} \subset \mathcal{B}(E^I)$$

always holds. The converse inclusion holds as soon as both E is separable and I is finite or countable.

Proof. The inclusion $\mathcal{B}(E)^{\otimes I} \subset \mathcal{B}(E^I)$ simply follows from the fact that the projections π_i are continuous for the product topology and so measurable for $\mathcal{B}(E^I)$, while $\mathcal{B}(E)^{\otimes I}$ is the smallest σ -algebra that makes them measurable.

Let us now prove the converse inclusion under the assumption that I is countable and E is separable. Let $(x^k)_{k \geq 1}$ be a dense sequence in E and consider the countable collection $\mathcal{U} = \{B(x^k, r), k \geq 1, r \in \mathbb{Q} \cap (0, \infty)\}$ of open balls centred at these points and with rational radius. Then every open set O of E can be written as $O = \bigcup \{B \in \mathcal{U} : B \subset O\}$. Each $B \in \mathcal{U}$ is open so it belongs to $\mathcal{B}(E)$ and further $\pi_i^{-1}(B) \in \mathcal{B}(E)^{\otimes I}$ for every $i \in I$. Now let O be an open set of E^I , then O can be written as a countable union of sets of the form $\pi_J^{-1}(\prod_{i \in J} B_i) = \bigcap_{i \in J} \pi_i^{-1}(B_i)$ where $J \subset I$ and each $B_i \in \mathcal{U}$. All these sets belong to $\mathcal{B}(E)^{\otimes I}$ and thus their countable union as well. $\square$

REMARK 3.16. Recall from Proposition 3.6 that $\mathcal{B}(E)^{\otimes I}$ coincides with the measurable countable cylinders. Then a non-separable space can be so large that the inclusion $\mathcal{B}(E^2) \subset \mathcal{B}(E)^{\otimes 2}$ may fail, for example if its cardinal is larger than continuum, then the diagonal $\{(x, x) : x \in E\}$, although closed, does not belong to the product σ -algebra. As for uncountable products, if E contains at least two elements, the inclusion $\mathcal{B}(E^I) \subset \mathcal{B}(E)^{\otimes I}$ again fails: any singleton $(x_i)_{i \in I} \in E^I$ is a closed set but not a countable cylinder.

This strict inclusion of σ -algebras causes some problems since a collection of random variables $(X_i)_{i \in I}$ on $(E, \mathcal{B}(E))$ is a priori only measurable for $\mathcal{B}(E)^{\otimes I}$ and not the largest Borel σ -algebra $\mathcal{B}(E^I)$. For the moment, let us therefore only consider countable products of separable spaces. For a finite product, the metric d on E naturally induces metrics on $E^n = E^{\{1, \dots, n\}}$, such as

$$d_n((x_1, \dots, x_n), (y_1, \dots, y_n)) = \max_{1 \leq i \leq n} d(x_i, y_i).$$

One could also consider ℓ^p type distances, but this ℓ^∞ one enjoys the property that the ball of radius r centred at $(x_1, \dots, x_n)$ in (E^n, d_n) is precisely the product of the balls of radius r centred at x_i for each $1 \leq i \leq n$. Let us now consider infinite countable index sets and without loss of generality, take $I = \mathbb{N}$.

PROPOSITION 3.17. For any metric space (E, d) , the function

$$d_{\mathbb{N}}((x_i)_{i \in \mathbb{N}}, (y_i)_{i \in \mathbb{N}}) = \sum_{i \in \mathbb{N}} \frac{d(x_i, y_i) \wedge 1}{2^i},$$

is a distance on $E^{\mathbb{N}}$ that induces the product topology. The product space $(E^{\mathbb{N}}, d_{\mathbb{N}})$ is separable, resp. complete, if and only if (E, d) is.

Of course, the last claim is valid for finite products. Again, these considerations fail for uncountable products: unless E is a singleton, if I is uncountable, then there is no metric on E^I that induces the product topology.

The proof of this theorem is guided in Exercise 15.

3.4 WEAK CONVERGENCE OF SEQUENCES

Recall that a probability measure on $\mathcal{B}(E)^{\otimes \mathbb{N}}$ is entirely characterised by the collection of its projections on finitely many coordinates. Let us strengthen this result by showing that weak convergence of all these marginal laws implies weak convergence of the entire laws. Below we put $[k] = \{1, \dots, k\}$.

PROPOSITION 3.18. Assume that (E, d) is separable. Let μ and $(\mu_n)_{n \geq 1}$ be probability measures on $E^{\mathbb{N}}$. Then $\mu_n \Rightarrow \mu$ if and only if $\pi_{[k]*}\mu_n \Rightarrow \pi_{[k]*}\mu$ for every $k \geq 1$.

Proof. The assumption that E is separable allows to identify $\mathcal{B}(E)^{\otimes \mathbb{N}} = \mathcal{B}(E^{\mathbb{N}})$. The direct implication simply follows from continuity of the projection. For the converse implication, according to Theorem 1.13, it is sufficient to prove that $\mu_n(f) \rightarrow \mu(f)$ for every f uniformly continuous and bounded function over $E^{\mathbb{N}}$. Fix such a function f , fix $\varepsilon > 0$, and let $\delta > 0$ be such that for all $x, y \in E^{\mathbb{N}}$, if $d_{\mathbb{N}}(x, y) < \delta$ then $|f(x) - f(y)| < \varepsilon$. Fix an arbitrary element $a \in E$ and for every $k \geq 1$ and every $x = (x_n)_{n \geq 1} \in E^{\mathbb{N}}$, let $x^{(k)} \in E^{\mathbb{N}}$ be given by $x_i^{(k)} = x_i$ if $i \leq k$ and $x_i^{(k)} = a$ if $i \geq k + 1$. Then

$$d_{\mathbb{N}}(x, x^{(k)}) = \sum_{i \geq k+1} 2^{-i} (d(x_i, a) \wedge 1) \leq 2^{-k}$$

for any $x \in E^{\mathbb{N}}$. Choose then k large enough so $2^{-k} < \delta$, it follows that $|f(x) - f(x^{(k)})| < \varepsilon$ for every $x \in E^{\mathbb{N}}$. Let $\mu^{(k)}$ denote the image of μ by the function $x \mapsto x^{(k)}$ and similarly for μ_n , then

$$|\mu_n(f) - \mu(f)| \leq |\mu_n^{(k)}(f) - \mu^{(k)}(f)| + 2\varepsilon.$$

Note that $\mu^{(k)}$ is also the image of $\pi_{[k]*}\mu$ by the continuous function $(x_1, \dots, x_k) \mapsto x^{(k)}$, so the convergence of these projections on the first k indices implies that of $\mu_n^{(k)}$. $\square$

3.5 EXERCISES

EXERCISE 15. For each $i \geq 1$, let (E_i, d_i) be a metric space. Let $E = \prod_{i=1}^{\infty}$ denote their product set and define for every $x = (x_i)_{i \geq 1} \in E$ and every $y = (y_i)_{i \geq 1} \in E$ the quantity:

$$d(x, y) = \sum_{i=1}^{\infty} \frac{d_i(x_i, y_i) \wedge 1}{2^i}.$$

- (i) Prove that d is a metric on E .
- (ii) Prove that d is the metric of coordinate-wise convergence, namely that $x^n = (x_i^n)_{i \geq 1}$ converges to $x = (x_i)_{i \geq 1}$ for the metric d if and only if x_i^n converges to x_i for the metric d_i for each $i \geq 1$.
- (iii) Prove that if each E_i is separable, then so is E .
- (iv) Prove that if each E_i is complete, then so is E .
- (v) Prove that if each E_i is compact, then so is E .

EXERCISE 16. For each $i \geq 1$, let (E_i, d_i) be a separable and complete metric space and let $E = \prod_{i=1}^{\infty}$ denote their product set, equipped with the metric d from Exercise 15. For every $i \geq 1$, let

$$\pi_i : x \in E \mapsto x_i \in E_i$$

be the natural projection and for every probability measure μ on E , let $\pi_{i*}\mu$ denote the probability on E_i given by the push-forward of the projection π_i . Let $\mathcal{P}$ be a collection of probability measures on E and let $\mathcal{P}_i = \{\pi_{i*}\mu : \mu \in \mathcal{P}\}$ denote the collection of their images. Prove that

$$\mathcal{P} \text{ is tight} \quad \text{if and only if} \quad \mathcal{P}_i \text{ is tight for every } i \geq 1.$$

EXERCISE 17. Find a sequence $(f_n)_{n \geq 1}$ of functions from $[0, 1]$ to $\mathbb{R}$ such that every subsequence has a further sub-subsequence that converges almost everywhere (for the Lebesgue measure) to the constant 0, but the entire sequence does *not* converge almost everywhere to 0. What does this mean about convergence almost everywhere?

EXERCISE 18 (Continuity is not measurable). For every $u \in [0, 1]$, let $I^u : [0, 1] \rightarrow \{0, 1\}$ denote the function given by $I_u^u = 1$ and $I_t^u = 0$ for every $t \neq u$. Let also 0 denote the null function $0_t = 0$ for every $t \in [0, 1]$.

- (i) Prove that $u \mapsto I^u$ and $u \mapsto 0$ are measurable from $([0, 1], \mathcal{B}([0, 1]))$ to $(\mathbb{R}^{[0,1]}, \mathcal{B}(\mathbb{R})^{\otimes [0,1]})$.
- (ii) Let U be a random variable with the uniform law on $[0, 1]$. Prove that, as random variables in $(\mathbb{R}^{[0,1]}, \mathcal{B}(\mathbb{R})^{\otimes [0,1]})$, the function I^U has the same law as the null function 0.
- (iii) Conclude that the set of continuous functions $[0, 1] \rightarrow \mathbb{R}$ does not belong to $\mathcal{B}(\mathbb{R})^{\otimes [0,1]}$.
- (iv) Prove that no singleton $\{x\}$ belongs to $\mathcal{B}(\mathbb{R})^{\otimes [0,1]}$, for arbitrary $x \in \mathbb{R}^{[0,1]}$.

CHAPTER 4

THE TOPOLOGY OF WEAK CONVERGENCE

Let us go back to the general setting of a metric space (E, d) equipped with its Borel σ -algebra $\mathcal{B}(E)$. We prove in this chapter Prohorov's theorem before diving into the topological structure of weak convergence of probability measures.

CONTENTS

4.1	Prohorov's theorem	45
4.2	Metric on probability measures	48
4.3	Topology of the Lévy–Prohorov distance	50
4.4	Exercises	51

4.1 PROHOROV'S THEOREM

Recall that a collection $\mathcal{P}$ of probability measures on (E, d) is tight when for every $\varepsilon > 0$, there exists a compact K of E such that $\mu(K) > 1 - \varepsilon$ for every $\mu \in \mathcal{P}$. It is sequentially relatively compact when every sequence in $\mathcal{P}$ has a subsequence that converges weakly to some probability measure, which may not belong to $\mathcal{P}$ in general. The first goal of this chapter is to prove Prohorov's theorem 1.25 that relates these two notions as follows by showing that:

- (i) if a collection $\mathcal{P}$ is tight, then this collection is also sequentially relatively compact,
- (ii) the converse implication holds if (E, d) is separable and complete.

When $E = \mathbb{R}^d$ is a finite-dimensional Euclidean space, this was proved in Theorem 2.16. In this chapter, we shall build upon this result and extend it to the space $\mathbb{R}^{\mathbb{N}}$ of sequences and then to a general metric space.

THE CONVERSE PART

Let us start with the second item, which is the simplest to prove.

PROPOSITION 4.1 (Part (ii) of Prohorov's theorem 1.25). If (E, d) is separable and complete, then any sequentially relatively compact collection of probability measures is tight.

Proof. Consider the following property (P): For every $\varepsilon, \delta > 0$, there exists finitely many open balls, say $B_1, \dots, B_k$, with radius δ and such that $\mu(\bigcup_{i=1}^k B_i) > 1 - \varepsilon$ for all $\mu \in \mathcal{P}$. Let us prove that if (P) is satisfied then $\mathcal{P}$ is tight and if not then $\mathcal{P}$ is not sequentially relatively compact. This indeed shows that sequential relative compactness implies Property (P) and thus tightness.

Assume that (P) holds. Fix $\varepsilon > 0$. For each $n \geq 1$, use (P) to construct a set A_n which is the union of finitely many open balls with radius $1/n$ and such that $\mu(A_n) > 1 - \varepsilon 2^{-n}$ for all $\mu \in \mathcal{P}$. The set $\bigcap_{n \geq 1} A_n$ is pre-compact by construction and since E is complete, then $K = \overline{\bigcap_{n \geq 1} A_n}$ is both pre-compact and

complete, i.e. compact. Finally for any $\mu \in \mathcal{P}$, we have $\mu(K^c) \leq \mu(\bigcup_{n \geq 1} A_n^c) \leq \sum_{n \geq 1} \varepsilon 2^{-n} = \varepsilon$.

Next assume that (P) fails: fix $\varepsilon, \delta > 0$ such that if A is any finite union of open balls with radius δ , then there exists $\mu \in \mathcal{P}$ with $\mu(A) \leq 1 - \varepsilon$. Since E is separable then it may be entirely covered by countably many open balls with radius δ , say $(B_i)_{i \geq 1}$. For each $n \geq 1$, let $A_n = \bigcup_{i=1}^n B_i$; then there exists $\mu_n \in \mathcal{P}$ such that $\mu_n(A_m) \leq \mu_n(A_n) \leq 1 - \varepsilon$ for all $m \leq n$. Therefore, for each $m \geq 1$ fixed, the open set A_m satisfies $\liminf \mu_n(A_n) \leq 1 - \varepsilon$. The Portmanteau theorem 1.10 shows that if $(\mu_n)_n$ had a convergent subsequence, the limit μ would satisfy $\mu(A_m) \leq 1 - \varepsilon$ for all $m \geq 1$, and so $\mu(E) = \lim_m \mu(A_m) \leq 1 - \varepsilon$. $\square$

The next sections are devoted to proving the direct implication, which holds in any metric space.

REAL-VALUED SEQUENCES

Let us specialise the results from Chapter 3 on general product spaces to $\mathbb{R}^{\mathbb{N}}$. We equip this space with the product topology, which is induced by the distance

$$d_{\mathbb{N}}((x_i)_{i \in \mathbb{N}}, (y_i)_{i \in \mathbb{N}}) = \sum_{i \in \mathbb{N}} \frac{|x_i - y_i| \wedge 1}{2^i}.$$

This is the coarsest topology that makes continuous all the projections $\pi_{[k]} : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^k$ on finitely many components. A sequence of sequences then converges whenever the sequence of the first k elements does for every $k \geq 1$ fixed. The space $(\mathbb{R}^{\mathbb{N}}, d_{\mathbb{N}})$ inherits separability and completeness from $\mathbb{R}$. Its Borel σ -algebra coincides with the product σ -algebra $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ and a probability measure μ on $\mathbb{R}^{\mathbb{N}}$ is uniquely characterised by all its projections $(\pi_{[k]*}\mu)_{k \geq 1}$. In addition, Proposition 3.18 further shows that weak convergence of probability measures on $\mathbb{R}^{\mathbb{N}}$ is equivalent to the convergence of all these projections

The direct part of Prohorov's theorem 1.25 in $\mathbb{R}^{\mathbb{N}}$ then follows as a corollary.

PROPOSITION 4.2. Any tight collection of probability measures in $\mathbb{R}^{\mathbb{N}}$ is sequentially relatively compact.

Proof. Let $\mathcal{P}$ be a tight collection of probability measures on $\mathbb{R}^{\mathbb{N}}$. For every $k \geq 1$, the canonical projection $\pi_{[k]} : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^k$ is continuous (by definition of the product topology). Since the image by a continuous function of a compact set is compact, then the collection of the image measures $\{\pi_{[k]*}\mu : \mu \in \mathcal{P}\}$ is tight in $\mathbb{R}^k$. By Theorem 2.16 this collection is then relatively compact in $\mathbb{R}^k$. A diagonal extraction then allows to extract from every sequence in $\mathcal{P}$ a subsequence $(\mu_n)_{n \geq 1}$ such that $(\pi_{[k]*}\mu_n)_{n \geq 1}$ converges in $\mathbb{R}^k$ simultaneously for every $k \geq 1$. Let us call $(\mu^k)_{k \geq 1}$ the limits. These laws inherit from the sequence $(\pi_{[k]*}\mu_n)_{n \geq 1}$ the consistency property (3.2) so Kolmogorov's extension theorem 3.10 ensures the existence of a unique probability measure μ on $\mathbb{R}^{\mathbb{N}}$ such that $\pi_{[k]*}\mu = \mu^k$ for every $k \geq 1$. Then the subsequence $(\mu_n)_{n \geq 1}$ satisfies $\pi_{[k]*}\mu_n \Rightarrow \pi_{[k]*}\mu$ for every $k \geq 1$. According to Proposition 3.18, this implies $\mu_n \Rightarrow \mu$. $\square$

PROHOROV'S THEOREM ON SUBSPACES

We shall extend Prohorov's theorem using embeddings into $\mathbb{R}^{\mathbb{N}}$. To make this work, we need to extend it from $\mathbb{R}^{\mathbb{N}}$ to a subset. Recall that if A is any subset of E , the restriction of the distance d makes it a metric space on its own; it can then be equipped with a Borel σ -algebra $\mathcal{B}(A)$. Since any open set of A takes the form $O \cap A$ for some open set O of E , then $\mathcal{B}(A) = \mathcal{B}(E) \cap A$. When A is itself an element of $\mathcal{B}(E)$, this reduces to $\mathcal{B}(A) = \{B \in \mathcal{B}(E) : B \subset A\}$.

PROPOSITION 4.3. Suppose that on a metric space (E, d) , any tight collection of probability measures is sequentially relatively compact. Then the same holds for any Borel set $A \in \mathcal{B}(E)$.

Proof. Let $\mathcal{P}$ be a tight collection of probability measures on $(A, d, \mathcal{A})$. For every $\mu \in \mathcal{P}$, let $\bar{\mu}$ denote the probability measure on $(E, d, \mathcal{B}(E))$ given by $\bar{\mu}(B) = \mu(B \cap A)$ for every $B \in \mathcal{B}(E)$. The collection $\{\bar{\mu} : \mu \in \mathcal{P}\}$ is tight since a compact set of A is a compact set of E . Hence $\{\bar{\mu} : \mu \in \mathcal{P}\}$ is sequentially relatively compact. Let $(\mu_n)_{n \geq 1}$ be a sequence in $\mathcal{P}$ and let $(\mu_{n_k})_{k \geq 1}$ be a subsequence such that $\bar{\mu}_{n_k}$ converges to some probability measure ν on E .

We claim that $\nu(A) = 1$. If A was closed, this would be an easy consequence of the Portmanteau theorem 1.10 since $\bar{\mu}_{n_k}(A) = \mu_{n_k}(A) = 1$ for each k . Let us argue using tightness on A : for every $\varepsilon > 0$, there is a compact set K of A such that $\bar{\mu}_{n_k}(K) = \mu_{n_k}(K) > 1 - \varepsilon$. This compact set of A is closed in $\mathbb{R}^N$; the Portmanteau theorem 1.10 then shows that $\nu(A) \geq \nu(K) \geq \limsup_k \bar{\mu}_{n_k}(K) > 1 - \varepsilon$. Since this holds true for every $\varepsilon > 0$, then indeed $\nu(A) = 1$.

Define then a probability μ on A by $\mu(B) = \nu(B)$ for every $B \in \mathcal{B}(A)$. Then for any open set O of A , there exists an open set U of E such that $O = U \cap A$ and then by definition of $\bar{\mu}_{n_k}$ and then the Portmanteau theorem 1.10, we have

$$\liminf_k \mu_{n_k}(O) = \liminf_k \bar{\mu}_{n_k}(U) \geq \nu(U).$$

Since $\nu(A) = 1$, then $\nu(U) = \nu(U \cap A) = \mu(U \cap A) = \mu(O)$. We conclude from the Portmanteau theorem 1.10 again that μ_{n_k} converges weakly to μ . $\square$

Combined with Proposition 4.2, we obtain the following result.

COROLLARY 4.4. For any Borel set A of $\mathbb{R}^N$, any tight collection of probability measures on A is sequentially relatively compact.

THE GENERAL CASE

Let us finally prove the first part of Prohorov's theorem 1.25 in a general metric space. We shall use the following embedding of a separable metric space into $\mathbb{R}^N$.

EXERCISE 4.5. Suppose that (E, d) is a separable metric space, with $(x_n)_{n \geq 1}$ a dense sequence. For every $x \in E$, let

$$h(x) = (d(x, x_n))_{n \geq 1} \in \mathbb{R}^N.$$

Prove that:

- (i) The function h is injective, so it defines a bijection from E to $h(E)$.
- (ii) This bijection is a homeomorphism: both h and h^{-1} are continuous.
- (iii) *Harder.* If (E, d) is complete, then $h(E)$ is a countable intersection of open sets, so in particular it is a Borel set of $\mathbb{R}^N$.

In general, the set $h(E)$ may not be a Borel set of $\mathbb{R}^N$. In the following lines, it will follow from another reason than completeness. Call a space E σ -compact when it is a countable union of compact sets. Since the image of a compact set by a continuous function is compact, then E σ -compact implies that $h(E)$ is σ -compact, and in particular Borel measurable.

PROPOSITION 4.6 (Part (i) of Prohorov's theorem 1.25). If (E, d) is any metric space, then any tight collection of probability measures is sequentially relatively compact.

Proof. Suppose that $\mathcal{P}$ is a tight collection of probability measures on (E, d) . Then for every $n \geq 1$, there exists a compact set K_n such that $\mu(K_n^c) \leq 1/n$ for all $\mu \in \mathcal{P}$. Let $E^0 = \bigcup_{n \geq 1} K_n$, which is a measurable set with $\mu(E^0) = 1$ for all $\mu \in \mathcal{P}$. For every $\mu \in \mathcal{P}$, let μ^0 be the restriction of μ on E^0 , then the collection $\mathcal{P}^0 = \{\mu^0 : \mu \in \mathcal{P}\}$ is tight as probability measures on E^0 : simply use the compact sets K_n . In addition, if it is relatively compact, then so is $\mathcal{P}$: given a convergent subsequence $(\mu_{n_k}^0)_k$ in $\mathcal{P}$, with limit ν^0 , define ν on E by $\nu(\cdot) = \nu^0(\cdot \cap E^0)$, and then μ_{n_k} converges weakly to ν by continuity of the map $x \in E^0 \mapsto x \in E$.

What this shows is that it is sufficient to prove that tightness implies relative compactness for the collection $\mathcal{P}^0$ on E^0 . Then apply the homeomorphism h from E^0 to $h(E^0) \subset \mathbb{R}^N$; then $h(E^0)$ is σ -compact as well and in particular Borel-measurable. Tightness and weak convergence are both preserved by homeomorphisms so if $\mathcal{P}^0$ is a tight collection of probability measures on E^0 , then their image by h form a tight collection in $h(E^0)$; according to Corollary 4.4 it is then relatively compact so finally is $\mathcal{P}^0$. $\square$

4.2 METRIC ON PROBABILITY MEASURES

Let $\mathcal{M}_1(E)$ denote the set of all probability measures on E . A natural question is: can we define a topology on $\mathcal{M}_1(E)$ that corresponds to weak convergence? The answer is positive: the functional analysis enthusiasts have observed that probability measures form a subset of the dual of the space of continuous and bounded functions, so one can use the trace of the weak-* topology. A base of open sets is given by the collection of sets

$$\{\mu \in \mathcal{M}_1(E) : |\mu(f_i) - \nu(f_i)| < \varepsilon_i \text{ for all } 1 \leq i \leq n\}$$

for all $\nu \in \mathcal{M}_1(E)$, $n \geq 1$, $\varepsilon_1, \dots, \varepsilon_n > 0$, and $f_1, \dots, f_n \in C_b(E)$.

We shall however not work in this general setting and prefer metric spaces. Lévy introduced a distance between distribution functions on the real line that Prohorov later generalised to a distance between any probability measures on a metric space.

DEFINITION 4.7. The Lévy–Prohorov distance between two probability measures μ and ν on E is defined by:

$$d_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \mu(C) \leq \nu(C^\varepsilon) + \varepsilon \text{ for all closed sets } C\}.$$

Note that d_{LP} depends on the distance on d on E through C^ε .

PROPOSITION 4.8. The Lévy–Prohorov distance is indeed a distance between probability measures.

Proof. The definition is not obviously symmetric but it actually is for suppose that for some $a > 0$, we have $\mu(C) \leq \nu(C^a) + a$ for all closed sets C . Fix any closed set C and let $D = E \setminus C^a$, which is closed such that $D^a \subset E \setminus C$. Then

$$\nu(C) = 1 - \nu(E \setminus C) \leq 1 - \nu(D^a) \leq 1 - \mu(D) + a = \mu(E \setminus D) + a = \mu(C^a) + a.$$

Hence, for every $a \geq d_{LP}(\mu, \nu)$, we have $d_{LP}(\nu, \mu) \leq a$, which means that $d_{LP}(\nu, \mu) \leq d_{LP}(\mu, \nu)$. Exchanging the role of μ and ν shows the equality.

Next, using this symmetry, $d_{LP}(\mu, \nu) = 0$ means that for every closed set C , for every $\varepsilon > 0$, we have both $\mu(C) \leq \nu(C^\varepsilon) + \varepsilon$ and $\nu(C) \leq \mu(C^\varepsilon) + \varepsilon$. Recall that any closed set C can be written as $C = \bigcap_n C^{1/n}$ so $\nu(C) = \lim_n \nu(C^{1/n})$ and hence $\mu(C) \leq \nu(C)$ and symmetrically $\nu(C) \leq \mu(C)$. Thus μ and ν agree on all closed sets and therefore everywhere.

Finally, let us check the triangle inequality: let μ, ν, π be three probability measures with $d_{LP}(\mu, \nu) \leq a$ and $d_{LP}(\nu, \pi) \leq b$, so for every closed set C , we have both $\mu(C) \leq \nu(C^a) + a$ and $\nu(C^a) \leq \pi(C^{a+b}) + b$ since $(C^a)^b = C^{a+b}$. Hence $\mu(C) \leq \pi(C^{a+b}) + a + b$, that is, $d_{LP}(\mu, \pi) \leq a + b$ for every $a \geq d_{LP}(\mu, \nu)$ and every $b \geq d_{LP}(\nu, \pi)$, from which $d_{LP}(\mu, \pi) \leq d_{LP}(\mu, \nu) + d_{LP}(\nu, \pi)$ follows. $\square$

Now to link the Lévy–Prohorov distance and weak topology.

PROPOSITION 4.9. If $d_{LP}(\mu_n, \mu) \rightarrow 0$, then $\mu_n \Rightarrow \mu$. The converse holds when E is separable.

Proof. If $d_{LP}(\mu_n, \mu) \rightarrow 0$, then for every $\varepsilon > 0$, it holds $\limsup_n \mu_n(C) \leq \mu(C^\varepsilon) + \varepsilon$ for all closed sets C . Letting $\varepsilon \downarrow 0$, we obtain $\limsup_n \mu_n(C) \leq \mu(C)$ for all closed sets C , so $\mu_n \Rightarrow \mu$ by the Portmanteau theorem 1.10.

Suppose conversely that E is separable and $\mu_n \Rightarrow \mu$. Fix $\varepsilon > 0$, since E is separable we may cover it by countably many open balls with radius ε , say $(B_i)_{i \geq 1}$. By monotonicity, there exists $N \geq 1$ such that $\mu(\bigcup_{i \leq N} B_i) \geq 1 - \varepsilon$. For any closed set C , let $C_N = \bigcup \{B_i : 1 \leq i \leq N, B_i \cap C \neq \emptyset\}$. Note that it is an open set, so the Portmanteau theorem 1.10 shows that $\liminf_n \mu_n(C_N) \geq \mu(C_N)$. This implies that for $\mu(C_N) \leq \mu_n(C_N) + \varepsilon$ for every n larger than some n_0 which depends on C_N a priori, and this would be an issue. However there are only finitely possibilities for C_N as C varies, so we can take the largest of the corresponding n_0 's and infer that for every n large enough, we have $\mu(C_N) \leq \mu_n(C_N) + \varepsilon$ simultaneously for every closed set C . To conclude, observe that since each B_i has radius ε , then $C_N \subset C^{2\varepsilon}$ so finally, for every n large enough, for every closed set C , we have $\mu(C_N) \leq \mu_n(C^{2\varepsilon}) + \varepsilon$ and so $d_{LP}(\mu, \mu_n) \leq \varepsilon$ for n large enough. The claim follows since $\varepsilon > 0$ can be arbitrarily small. $\square$

Thus, on a separable space, we can equivalently use d_{LP} to speak about the weak convergence. The next proposition is a first step towards the topology of this space.

PROPOSITION 4.10. If (E, d) is separable, then so is $(\mathcal{M}_1(E), d_{LP})$.

Proof. Fix $D \subset E$ a countable dense set and let $M(D)$ denote the countable set of probability measures of the form $\sum_{i \in I} q_i \delta_{x_i}$ with I finite, $q_i \in \mathbb{Q}$ and $x_i \in D$ for each $i \in I$. Then $M(D)$ is countable and we claim that it is dense. Fix μ a probability and $\varepsilon > 0$ and let us construct an element of $M(D)$ that lies at distance at most 2ε from μ . The argument is close to the previous proof.

First, we may cover E by countably many balls with diameter ε , with center the elements of D , say $(x_k)_{k \geq 1}$. Define then $A_1^\varepsilon = B(x_1, \varepsilon)$ and then recursively for every $k \geq 2$, let $A_k^\varepsilon = B(x_k, \varepsilon) \setminus \bigcup_{i=1}^{k-1} B(x_i, \varepsilon)$. The sets $(A_i^\varepsilon)_{i \geq 1}$ form a countable partition of E into disjoint measurable sets with diameter at most ε . Let $N \geq 1$ be large enough so $\mu(\bigcup_{i \leq N} A_i^\varepsilon) \geq 1 - \varepsilon$. Then let $(q_i^\varepsilon)_{1 \leq i \leq N}$ be nonnegative rational numbers such that $q_i^\varepsilon \leq \mu(A_i^\varepsilon) \leq q_i^\varepsilon + \varepsilon/N$ for each $1 \leq i \leq N$. Then let $q_0^\varepsilon = 1 - \sum_{i=1}^N q_i^\varepsilon$ be such that $q_0^\varepsilon \geq 0$ and $\sum_{i=0}^N q_i^\varepsilon = 1$. Let $x_0^\varepsilon \in D$ be arbitrary and for each $1 \leq i \leq N$, fix an arbitrary element $x_i^\varepsilon \in A_i^\varepsilon \cap D$ and let $\mu^\varepsilon = \sum_{i=0}^N q_i^\varepsilon \delta_{x_i^\varepsilon} \in M(D)$. Fix C a closed set. Since the A_i^ε 's have diameter at most ε , then $A_i^\varepsilon \cap C \neq \emptyset$ implies that $x_i^\varepsilon \in C^{2\varepsilon}$. Hence

$$\mu(C) \leq \sum_{i=1}^N \mu(A_i^\varepsilon) \mathbb{1}_{A_i^\varepsilon \cap C \neq \emptyset} + \sum_{i > N} \mu(A_i^\varepsilon) \leq \sum_{i=1}^N (q_i^\varepsilon + \varepsilon/N) \mathbb{1}_{x_i^\varepsilon \in C^{2\varepsilon}} + \varepsilon \leq \mu^\varepsilon(C^{2\varepsilon}) + 2\varepsilon.$$

Hence $d_{LP}(\mu, \mu^\varepsilon) \leq 2\varepsilon$. $\square$

One can naturally identify any element of E as the probability on E given by the Dirac mass at this element. The next lemma shows that this identification is a homeomorphism. The proof is guided in the exercise section.

LEMMA 4.11. For every $x, y \in E$, we have

$$d_{LP}(\delta_x, \delta_y) = \min(d(x, y), 1).$$

Consequently, the map $x \mapsto \delta_x$ is a homeomorphism between E and the set $\Delta = \{\delta_x; x \in E\}$ of Dirac measures. Finally Δ is closed in $\mathcal{M}_1(E)$.

Next, let us relate more properties of E and $\mathcal{M}_1(E)$. Recall from Section 1.1 the notions of compactness and pre-compactness (or total boundedness), and that a metric space is compact if and only if it is both pre-compact and complete. Note also that a pre-compact space is separable.

COROLLARY 4.12. For each property amongst pre-compactness, compactness, and completeness, if $(\mathcal{M}_1(E), d_{LP})$ satisfies this property, then so do (Δ, d_{LP}) and (E, d) .

Proof. If a metric space is pre-compact, then so is every subset; compactness and completeness are transferred to closed subsets. Hence, for any of the three properties, Lemma 4.11 implies that if $(\mathcal{M}_1(E), d_{LP})$ satisfies this property, then so does (Δ, d_{LP}) . Further so does (E, d) by the same lemma: compactness is preserved under continuous functions, and from the form of the homeomorphism here, a finite covering of Δ by balls with radius smaller than 1 is mapped to the same covering of E , and similarly the homeomorphism preserves Cauchy sequences and so completeness. $\square$

What this corollary shows is that if one wants a good property to hold in $\mathcal{M}_1(E)$, it is necessary that this property holds in E . For practical use, the important question is whether this is sufficient. We shall see in the next section that it is, at least for separable spaces.

4.3 TOPOLOGY OF THE LÉVY–PROHOROV DISTANCE

Prohorov's theorem 1.25 showed that if a collection of probability measures concentrates on a compact set of E , then this collection is sequentially compact. The Lévy–Prohorov distance provides an analogous results for pre-compact sets.

PROPOSITION 4.13. Let (E, d) be any metric space and $\mathcal{P}$ be a collection of probability measures such that for every $\varepsilon > 0$, there exists a pre-compact set K of E such that $\mu(K) > 1 - \varepsilon$ for every $\mu \in \mathcal{P}$. Then $\mathcal{P}$ is pre-compact for d_{LP} .

Proof. Fix $\varepsilon > 0$ and let K be a pre-compact set such that $\mu(K) > 1 - \varepsilon$ for every $\mu \in \mathcal{P}$. Let $N \geq 1$ and $x_1, \dots, x_N \in K$ be such that $K \subset \bigcup_{i=1}^N B(x_i, \varepsilon)$. Add an arbitrary $x_0 \in E$, take $m \geq N/\varepsilon$, and let $\mathcal{Q}_m \subset \mathcal{P}$ denote the finite set of probability measures of the form

$$\nu = \sum_{i=0}^N \frac{\alpha_i}{m} \delta_{x_i} \quad \text{where} \quad \alpha_i \in \{0, \dots, m\} \quad \text{with} \quad \sum_{i=0}^N \alpha_i = m.$$

We claim that $\mathcal{P}$ is covered by the d_{LP} -balls centred at the elements of $\mathcal{Q}_m$ and with radius 2ε .

Indeed, fix an arbitrary probability μ in $\mathcal{P}$. For each $1 \leq i \leq N$, let $A_i = B(x_i, 2\varepsilon) \setminus \bigcup_{j=1}^{i-1} B(x_j, 2\varepsilon)$ and then let $\alpha_i = \lfloor m\mu(A_i) \rfloor$. Let further $\alpha_0 = m - \sum_{i=1}^N \alpha_i$. Let ν denote the corresponding probability as in the previous display and let C be a closed set. First, since $\mu(\bigcup_{i=1}^N A_i) \geq 1 - \varepsilon$, then

$$\mu(C) \leq \sum_{A_i \cap C \neq \emptyset} \mu(A_i) + \varepsilon.$$

Recall that $\mu(A_i) \leq (\alpha_i + 1)/m$ and $m \geq N/\varepsilon$, then further

$$\mu(C) \leq \sum_{A_i \cap C \neq \emptyset} \frac{\alpha_i}{m} + 2\varepsilon.$$

Finally, we have $A_i \cap C \neq \emptyset$ if and only if $x_i \in C^{2\varepsilon}$, so

$$\mu(C) \leq \nu(C^{2\varepsilon}) + 2\varepsilon,$$

and thus $d_{LP}(\mu, \nu) \leq 2\varepsilon$ as we claimed. $\square$

Recall from Corollary 4.12 that a good property on (E, d) holds true as soon as it does for $(\mathcal{M}_1(E), d_{LP})$. More useful in practice is the converse that we now provide. Recall that compactness implies pre-compactness, which implies separability. Recall also from Proposition 4.9 that in a separable space, convergence for d_{LP} is equivalent to weak convergence.

THEOREM 4.14. If (E, d) is pre-compact, respectively compact, then so is $(\mathcal{M}_1(E), d_{LP})$. If (E, d) is both separable and complete, then so is $(\mathcal{M}_1(E), d_{LP})$.

Proof. The pre-compact case follows directly from the previous proposition. We also proved separability in Proposition 4.10. Finally compactness amounts to the combination of pre-compactness (which implies separability) and completeness. It thus only remains to prove that if (E, d) is both separable and complete, then $(\mathcal{M}_1(E), d_{LP})$ is complete.

For this, let $(\mu_n)_{n \geq 0}$ be Cauchy sequence in $(\mathcal{M}_1(E), d_{LP})$ and let us prove that it admits a convergent subsequence, then the entire Cauchy sequence must converge. Fix $\varepsilon > 0$ arbitrary. Since we may extract from any Cauchy sequence $(\mu_n)_n$ a subsequence $(\mu_{n_k})_k$ such that $d_{LP}(\mu_{n_k}, \mu_{n_{k+1}}) \leq \varepsilon 2^{-k-1}$ for every $k \geq 1$, we may as well assume that $d_{LP}(\mu_n, \mu_{n+1}) \leq \varepsilon 2^{-n-1}$ for every $n \geq 1$ in the first place. This will reduce the notation. By the triangle inequality, it follows that $d_{LP}(\mu_n, \mu_{n+m}) \leq \varepsilon \sum_{i=1}^m 2^{-n-i} \leq \varepsilon 2^{-n}$ for every $n, m \geq 1$.

Since the space E is Polish, then any single probability measure is tight: for each $n \geq 1$, there exists a compact set K_n such that $\mu_n(K_n) \geq 1 - \varepsilon 2^{-n}$. By definition of the Lévy–Prohorov distance, for any $n, m \geq 1$ we have then

$$1 - \varepsilon 2^{-n} \leq \mu_n(K_n) \leq \mu_{n+m}(K_n^{\varepsilon 2^{-n}}) + \varepsilon 2^{-n}.$$

Exchanging the role of μ_n and μ_{n+m} , we have also:

$$1 - \varepsilon 2^{-n-m} \leq \mu_{n+m}(K_{n+m}) \leq \mu_n(K_{n+m}^{\varepsilon 2^{-n-m}}) + \varepsilon 2^{-n-m}.$$

We deduce that for any indices $i, j \geq 1$, we have

$$\mu_i(K_j^{\varepsilon 2^{-j}}) \geq 1 - 2\varepsilon 2^{-j}.$$

Let $K = \overline{\bigcap_j K_j^{\varepsilon 2^{-j}}}$, which is pre-compact by construction, and closed, so complete since E is; it is then a compact set. In addition, by construction, for any $i \geq 1$, it holds:

$$\mu_i(K^c) \geq \mu_i\left(\bigcup_{j \geq 1} (K_j^{\varepsilon 2^{-j}})^c\right) \leq 2\varepsilon \sum_{j \geq 1} 2^{-j} = \varepsilon.$$

The collection $(\mu_i)_{i \geq 1}$ is therefore tight and we conclude from Prohorov's theorem 1.25 that it admits convergent subsequences. $\square$

4.4 EXERCISES

EXERCISE 19 (Lemma 4.11). Let (E, d) be a metric space and let $\Delta = \{\delta_x; x \in E\}$ denote the set of Dirac measures.

(i) Prove that for all $x, y \in E$, we have

$$d_{LP}(\delta_x, \delta_y) = \min(d(x, y), 1).$$

(ii) Let $x_n \in E$ for every $n \geq 1$.

(a) Prove that if $x_n \rightarrow x$ for some $x \in E$, then $d_{LP}(\delta_{x_n}, \delta_x) \rightarrow 0$.

- (b) Assume that $(x_n)_n$ has no convergent subsequence and prove that any subset $F \subset \{x_n : n \geq 1\}$ in the range of values is closed.
- (c) Prove that in this case δ_{x_n} does not converge weakly to any probability measure.

Conclude by contradiction that that Δ is closed for d_{LP} and for the weak convergence.

EXERCISE 20. Let $(X_n)_{n \geq 1}$ and $(Y_{n,k})_{n,k \geq 1}$ be random variables all defined on the same probability space and with values in the separable and complete metric space (E, d) . Suppose that for each fixed $k \geq 1$, the sequence $(Y_{n,k})_{n \geq 1}$ converges in distribution to some random variable Y_k . Suppose also that for every $\varepsilon > 0$,

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(d(X_n, Y_{n,k}) > \varepsilon) = 0.$$

- (i) Prove that the sequence $(Y_k)_{k \geq 1}$ converges in distribution to some limit Y .
- (ii) Prove that $(X_n)_{n \geq 1}$ converges in distribution to the same limit Y .

EXERCISE 21. Let the random variables be defined on the same probability space and with values in separable and complete metric spaces. Let us provide another argument that if $X_n \rightarrow X$ in distribution, if $Y_n \rightarrow Y$ in distribution, and if X_n and Y_n are independent for each n , then the pair (X_n, Y_n) converges in distribution to a pair of independent random variables, distributed as X and Y respectively.

- (i) Prove that $(X_n, Y_n)_{n \geq 1}$ is tight (this does not require independence).
- (ii) Prove that any subsequential limit has the law of the pair (X, Y) of independent random variables.

EXERCISE 22. Let $\mathcal{M}_1(\mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d$, which we view as a (separable and complete) metric space, equipped with the Lévy–Prohorov distance d_{LP} . Let $(\Lambda_n)_{n \geq 0}$ be a sequence of random probability measures on $(\mathcal{M}_1(\mathbb{R}^d), d_{LP})$.

- (i) Prove that the sequence $(\Lambda_n)_{n \geq 0}$ is tight in the space $(\mathcal{M}_1(\mathbb{R}^d), d_{LP})$ if and only if for every $\varepsilon > 0$ there exists $A > 0$ such that:

$$\sup_{n \geq 0} \mathbb{P}(\Lambda_n([-A, A]^d) \geq 1 - \varepsilon) \geq 1 - \varepsilon.$$

- (ii) Assume now that each Λ_n has a (random!) continuous and bounded density f_n with respect to the Lebesgue measure on $\mathbb{R}^d$ and that f_n converges almost surely to a (random) density function f in the uniform norm over $\mathbb{R}^d$. Prove that Λ_n converges in distribution to the random probability measure with density f .
- (iii) Would the convergence in distribution $\Lambda_n \rightarrow \Lambda$ still hold under a weaker convergence of $f_n \rightarrow f$ such as the convergence in distribution $(f_n(x_1), \dots, f_n(x_k)) \rightarrow (f(x_1), \dots, f(x_k))$ for any given $k \geq 1$ and any k -tuple $x_1, \dots, x_k \in \mathbb{R}^d$?

CHAPTER 5

OTHER MODES OF CONVERGENCE & COUPLING

Let us discuss in this chapter convergence of random variables rather than their laws, namely almost sure convergence and convergence in probability, in a general separable metric space. We shall relate these notions to weak convergence and discuss a stronger convergence of probability measures, known as the total variation distance.

CONTENTS

5.1	Convergence of random variables	53
5.2	The Skorokhod representation theorem	55
5.3	Topology of convergence in probability	56
5.4	Coupling	57
5.5	Total variation distance	58
5.6	Exercises	62

Convergence in distribution of a sequence of random variables is really the convergence of their laws. Let us here discuss notions of convergence of the random variables themselves. For this we shall always assume in this chapter that

all random variables are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and take values in a separable metric space (E, d) .

5.1 CONVERGENCE OF RANDOM VARIABLES

The following definition should be well-known, at least when $E = \mathbb{R}$.

DEFINITION 5.1. For random variables X and $(X_n)_{n \geq 1}$ in a separable metric space (E, d) , we say that X_n converges to X *almost surely* and *in probability* respectively when

$$\mathbb{P}(X_n \rightarrow X) = 1, \quad \text{resp.} \quad \forall \varepsilon > 0, \quad \mathbb{P}(d(X_n, X) > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

Almost sure convergence is simply the pointwise convergence of the function $X_n(\omega) \rightarrow X(\omega)$, just not for every $\omega \in \Omega$, but on a set with probability 1. It is simple to manipulate just by using deterministic reasonings: simply fix $\omega \in \{X_n \rightarrow X\}$, then the sequence $(X_n(\omega))_n$ converges to $X(\omega)$ in E so for example, if one applies a continuous function f , then $(f(X_n(\omega)))_n$ converges to $f(X(\omega))$. Since this holds true for every $\omega \in \{X_n \rightarrow X\}$ and the latter has probability 1, then $\{f(X_n) \rightarrow f(X)\}$ has probability 1 as well, i.e. $f(X_n) \rightarrow f(X)$ almost surely.

Convergence in probability is very natural as well since it raises the question: for a given $\varepsilon > 0$ of margin and a given n , what is the probability that X_n lies at distance more than ε from X ? An advantage of

this convergence, as shown in the next result, is that it is weaker than almost sure convergence, so it holds in more contexts.

PROPOSITION 5.2. If X_n and X are random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and with values in a separable metric space (E, d) , then the following properties hold:

- (i) If $X_n \rightarrow X$ almost surely then also in probability.
- (ii) If $X_n \rightarrow X$ and $X_n \rightarrow Y$ in probability then $X = Y$ almost surely. In particular the same conclusion holds if we replace one or both convergences by an almost sure one.
- (iii) $X_n \rightarrow X$ in probability if and only if every subsequence has a further subsequence that converges almost surely to X .

Proof. (i) Fix $\varepsilon > 0$ and note that if $X_n \rightarrow X$ almost surely then $\mathbb{1}_{d(X_n, X) > \varepsilon} \rightarrow 0$ almost surely and it is bounded by 0 and 1 so $\mathbb{P}(d(X_n, X) > \varepsilon) = \mathbb{E}[\mathbb{1}_{d(X_n, X) > \varepsilon}] \rightarrow 0$ by dominated convergence.

(ii) By the triangle inequality, for every $\varepsilon > 0$, we have

$$\mathbb{P}(d(X, Y) > \varepsilon) \leq \mathbb{P}(d(X_n, X) > \varepsilon/2) + \mathbb{P}(d(X_n, Y) > \varepsilon/2) \xrightarrow{n \rightarrow \infty} 0.$$

Hence $\mathbb{P}(d(X, Y) > 1/k) = 0$ for any given integer $k \geq 1$. Thus

$$\mathbb{P}(X \neq Y) = \mathbb{P}\left(\bigcup_{k \geq 1} \{d(X, Y) > 1/k\}\right) \leq \sum_{k \geq 1} \mathbb{P}(d(X, Y) > 1/k) = 0.$$

(iii) Define $n_1 = 1$ and iteratively for every $k \geq 1$,

$$n_{k+1} = \inf\{j > n_k : \mathbb{P}(d(X_j, X) > 2^{-(k+1)}) \leq 2^{-(k+1)}\}.$$

These are all well-defined since $X_n \rightarrow X$ in probability. Observe that, by construction, we have $\sum_{k \geq 1} \mathbb{P}(d(X_{n_k}, X) > 2^{-k}) < \infty$ so by the Borel–Cantelli lemma, with probability 1, for every k large enough, we have $d(X_{n_k}, X) \leq 2^{-k}$ and so $X_{n_k} \rightarrow X$ with probability 1.

Conversely, if X_n does not converge to X in probability, then there exists $\varepsilon > 0$ and an increasing sequence of integers $(n_k)_{k \geq 1}$ such that $\mathbb{P}(d(X_{n_k}, X) > \varepsilon) > \varepsilon$ for all $k \geq 1$ so this subsequence has no further subsequence that converges in probability to X and so not almost surely. $\square$

EXERCISE 5.3. Prove that $X_n \rightarrow X$ almost surely if and only if $\sup_{k \geq n} d(X_k, X) \rightarrow 0$ in probability.

As for weak convergence of measures, these notions of convergence are stable under application of almost surely continuous functions.

LEMMA 5.4 (Continuous mapping). Let (E, d) and (E', d') be separable metric spaces and $f : (E, d) \rightarrow (E', d')$ be measurable and continuous $\mathbb{P}_X$ -almost surely. Then:

- (i) If $X_n \rightarrow X$ almost surely then $f(X_n) \rightarrow f(X)$ almost surely.
- (ii) If $X_n \rightarrow X$ in probability then $f(X_n) \rightarrow f(X)$ in probability.

Proof. (i) Let $C_f = \{x \in E : f \text{ is continuous at } x\} \in \mathcal{B}(E)$. Then we have

$$\{X \in C_f\} = \{\omega \in \Omega : f \text{ is continuous at } X(\omega)\} \in \mathcal{A}$$

and by assumption $1 = \mathbb{P}_X(C_f) = \mathbb{P}(X \in C_f)$ so if we let $A = \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}$,

then $\mathbb{P}(A) = 1$ and so $\mathbb{P}(A \cap \{X \in C_f\}) = 1$. Finally for every $\omega \in A \cap \{X \in C_f\}$ we have $f(X_n(\omega)) \rightarrow f(X(\omega))$.

- (ii) If $X_n \rightarrow X$ in probability then every subsequence $(n_k)_k$ has a further subsequence $(n_{k_j})_j$ such that $X_{n_{k_j}} \rightarrow X$ almost surely so by the first point $f(X_{n_{k_j}}) \rightarrow f(X)$ almost surely. By the last item in Proposition 5.2 this is equivalent to $f(X_n) \rightarrow f(X)$ in probability. $\square$

5.2 THE SKOROKHOD REPRESENTATION THEOREM

Recall that Theorem 2.5 allowed to represent the weak convergence of probability measures on $\mathbb{R}$ as the almost sure convergence of suitable random variables on some probability space. This extends to any separable metric space.

THEOREM 5.5 (Skorokhod's representation). Suppose that (E, d) is separable and $\mu_n \Rightarrow \mu$. Then there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and E -valued random variables X_n and X defined on it such that $X_n \sim \mu_n$, $X \sim \mu$ and $X_n \rightarrow X$ almost surely.

Let us mention that the probability space will be abstract; when (E, d) is also complete, we may always take it as the interval $[0, 1]$ equipped with the Lebesgue measure.

Proof of Theorem 5.5. Since E is separable, then for every $k \geq 1$, we may partition $E = \bigcup_{i=0}^{N_k} E_{k,i}$ into disjoint measurable sets such that $\mu(E_{k,0}) < 2^{-k}$ and for each $1 \leq i \leq N_k$, the diameter of $E_{k,i}$ is less than 2^{-k} and $\mu(E_{k,i}) > 0$ and $\mu(\partial E_{k,i}) = 0$. Indeed, start by covering E by the collection of balls centred at the elements $(x_i)_{i \geq 1}$ of a countable dense set and each with a radius $r_i < 2^{-k-1}$ such that $\mu(\partial B(x_i, r_i)) = 0$, which exists since $\mu(\partial B(x_i, r)) > 0$ can only hold for at most countably many values of r . Then let $E_{k,1} = B(x_1, r_1)$ and for $i \geq 2$, let $E_{k,i} = B(x_i, r_i) \cap (\bigcap_{j=1}^{i-1} B(x_j, r_j))^c$ to turn these balls into a partition in sets with null boundary. By monotonicity, we have $1 = \mu(E) = \lim_N \mu(\bigcup_{i=1}^N E_{k,i})$ so for each $k \geq 1$, we may choose $N_k \geq 1$ such that $\mu(\bigcup_{i \geq N_k+1} E_{k,i}) < 2^{-k}$. We let then $E_{k,0}$ be $\bigcup_{i \geq N_k+1} E_{k,i}$ to which we add each $E_{k,i}$ for $1 \leq i \leq N_k$ such that $\mu(E_{k,i}) = 0$. We then relabel to others $E_{k,i}$'s and change the value of N_k accordingly.

Fix such partition for some $k \geq 1$. Since $\mu_n \Rightarrow \mu$ and E is separable, then by the Portmanteau theorem 1.10, for each $1 \leq i \leq N_k$, we have $\mu_n(E_{k,i}) \rightarrow \mu(E_{k,i})$. It follows that there exists $n_k \geq 1$ such that for any $n \geq n_k$, it holds:

$$\mu_n(E_{k,i}) \geq (1 - 2^{-k})\mu(E_{k,i}) > 0.$$

We shall assume without loss of generality that the n_k 's are increasing.

Let us construct on some probability space independent random variables (for example, on the product space from Theorem 3.12) with the following laws:

- U with the uniform distribution over $[0, 1]$,
- X with law μ and Y_n with law μ_n for each $n \geq 1$,
- for every $n \geq n_1$ and $i \geq 1$, let $Y_{n,i}$ have the law $\mu_n(\cdot \mid E_{k,i})$ for the index $k \geq 1$ such that $n_k \leq n \leq n_{k+1}$,
- for every $n \geq n_1$, let Z_n have the law, with $k \geq 1$ such that $n_k \leq n \leq n_{k+1}$, given by:

$$2^k \sum_{i=0}^{N_k} \mu_n(\cdot \mid E_{k,i}) (\mu_n(E_{k,i}) - (1 - 2^{-k})\mu(E_{k,i})).$$

Let us finally define for all $n \geq 1$:

$$X_n = \begin{cases} Y_n & \text{if } n < n_1 \text{ (provided } n_1 \geq 2), \\ Y_{n,i} & \text{if both } U > 2^{-k} \text{ and } X \in E_{k,i} \text{ with } 0 \leq i \leq N_k \\ Z_n & \text{if } U \leq 2^{-k}. \end{cases}$$

Let us check that X_n has law μ_n . This is plain if $n < n_1$ so assume that $n_k \leq n < n_{k+1}$ for some $k \geq 1$. Then for every event A , we have by independence:

$$\begin{aligned} \mathbb{P}(X_n \in A) &= \mathbb{P}(U \leq 2^{-k}, Z_n \in A) + \sum_{i=0}^{N_k} \mathbb{P}(U > 2^{-k}, X \in E_{k,i}) \mathbb{P}(Y_{n,i} \in A) \\ &= \sum_{i=0}^{N_k} \mu_n(A \mid E_{k,i}) (\mu_n(E_{k,i}) - (1 - 2^{-k})\mu(E_{k,i})) + \sum_{i=0}^{N_k} (1 - 2^{-k})\mu(E_{k,i})\mu_n(A \mid E_{k,i}) \\ &= \sum_{i=0}^{N_k} \mu_n(A \mid E_{k,i})\mu_n(E_{k,i}) \\ &= \mu_n(A), \end{aligned}$$

by the law of total probability.

Next, since $\mathbb{P}(U \leq 2^{-k} \text{ or both } U > 2^{-k} \text{ and } X \in E_{k,0}) \leq 2^{-k}\mu(E_{k,0}) \leq 2^{-k+1}$ has finite sum, then the Borel–Cantelli lemma shows that the event $\{U > 2^{-k} \text{ and } X \in \bigcup_{i=1}^{N_k} E_{k,i} \text{ for all } k \text{ large enough}\}$ has probability 1. On this event, for all k large enough, both X_n and X belong to the same $E_{k,i}$, that has diameter less than 2^{-k} , so $X_n \rightarrow X$, with probability 1. $\square$

5.3 TOPOLOGY OF CONVERGENCE IN PROBABILITY

As we mentioned earlier, almost sure convergence does not correspond to any topology, because Proposition 5.2 shows that it does not satisfy Exercise 1.22. However convergence in probability does correspond to a topology, and even a metric space.

DEFINITION 5.6. Let X and Y be random variables defined the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and with values in a separable metric space (E, d) . Define their *Ky Fan distance* as:

$$\alpha(X, Y) = \inf\{\varepsilon \geq 0 : \mathbb{P}(d(X, Y) \geq \varepsilon) \leq \varepsilon\}.$$

REMARK 5.7. If X and Y are almost surely constant random variables, equal to some x and y respectively, then

$$\alpha(X, Y) = d(x, y).$$

From now on we regard two random variables X and Y as equal if they agree almost surely.

LEMMA 5.8. The function α is a distance and $\alpha(X_n, X) \rightarrow 0$ if and only if $X_n \rightarrow X$ in probability.

Proof. Obviously $\alpha(X, Y) = \alpha(Y, X) \geq 0$. Also $\alpha(X, Y) = 0$ if and only if $\mathbb{P}(d(X, Y) \geq \varepsilon) = 0$ for every $\varepsilon > 0$, which means that $X = Y$ almost surely. For the triangle inequality, let $\varepsilon \geq \alpha(X, Y)$ and

$\delta \geq \alpha(Y, Z)$, then

$$\begin{aligned} \mathbb{P}(d(X, Z) \geq \varepsilon + \delta) &\leq \mathbb{P}(d(X, Y) + d(Y, Z) \geq \varepsilon + \delta) \\ &\leq \mathbb{P}(d(X, Y) \geq \varepsilon \text{ or } d(Y, Z) \geq \delta) \\ &\leq \varepsilon + \delta. \end{aligned}$$

Letting $\varepsilon \downarrow \alpha(X, Y)$ and $\delta \downarrow \alpha(Y, Z)$ yields the result. Finally, $X_n \rightarrow X$ in probability means that for every $\varepsilon > 0$, for every n large enough, we have $\mathbb{P}(d(X_n, X) \geq \varepsilon) \leq \varepsilon$, which means $\alpha(X_n, X) \leq \varepsilon$. $\square$

PROPOSITION 5.9. If (E, d) is a complete metric space if and only if α is a complete distance.

Proof. Suppose that (E, d) is complete and let $(X_n)_{n \geq 1}$ be a Cauchy sequence for the distance α , i.e. that satisfies

$$\sup_{m \geq n} \alpha(X_m, X_n) \xrightarrow{n \rightarrow \infty} 0.$$

Let us prove that it admits a convergent subsequence; then it must be convergent. Let us extract a subsequence $(X_{n_k})_{k \geq 1}$ that satisfies

$$\sup_{\ell \geq k} \alpha(X_{n_\ell}, X_{n_k}) \leq 2^{-k}.$$

In particular $\mathbb{P}(d(X_{n_{k+1}}, X_{n_k}) \geq 2^{-k}) \leq 2^{-k}$ for every $\ell \geq k$ so by the Borel–Cantelli lemma, almost surely, we have $d(X_{n_{k+1}}, X_{n_k}) \leq 2^{-k}$ for all k large enough. On this event we have for every $\ell \geq k$:

$$d(X_{n_\ell}, X_{n_k}) \leq \sum_{i=k}^{\ell-1} d(X_{n_{i+1}}, X_{n_i}) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{-k-1}.$$

Hence on the event $d(X_{n_{k+1}}, X_{n_k}) \leq 2^{-k}$ for all k large enough, that has probability 1, the sequence $(X_{n_k})_k$ is Cauchy in E , and hence convergent if E is complete. This sequence therefore converges almost surely and thus in probability, so equivalently for α .

Conversely, suppose that α is a complete metric and let $(x_n)_{n \geq 1}$ be a Cauchy sequence in E . Then the almost surely constant random variables $X_n = x_n$ for each n form a Cauchy sequence for α since $\alpha(X_n, X_m) = d(x_n, x_m)$. It follows that X_n converges for α (i.e. in probability) to some random variable X and since each X_n is almost surely constant, then so is X (recall from Lemma 4.11 that the set of Dirac masses is closed for weak convergence). Hence there exists some $x \in E$ such that $X = x$ almost surely and we conclude that $d(x, x_n) = \alpha(X, X_n) \rightarrow 0$. $\square$

More information about convergence in probability can be found in the exercise section below.

5.4 COUPLING

Skorokhod's representation theorem 5.5 realises on a common probability space random variables each with a specified law. This is an example of *coupling*, which is a useful approach to compare two (or more) probability measures.

DEFINITION 5.10. Let μ_1 be a probability on $(E_1, \mathcal{E}_1)$ and μ_2 a probability on $(E_2, \mathcal{E}_2)$. A *coupling* of μ_1 and μ_2 is a probability measure μ on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ such that for every $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$, it holds

$$\mu(A_1 \times E_2) = \mu_1(A_1) \quad \text{and} \quad \mu(E_1 \times A_2) = \mu_2(A_2).$$

We shall denote by $\Gamma(\mu_1, \mu_2)$ the set of all the couplings of μ_1 and μ_2 . In terms of random variables, a coupling of X and Y is a pair (X', Y') defined on a common probability space such that X' has the same

law as X and Y' the same as Y .

EXAMPLE 5.11. If X and Y both have the Bernoulli law with parameter $1/2$, we can take either X' and Y' independent with this law, or $X' = Y'$, or $X' = 1 - Y'$. This provides three (very) different couplings.

Note that $\Gamma(\mu_1, \mu_2)$ is never empty since one can always consider the product $\mu = \mu_1 \otimes \mu_2$; in terms of random variables, this means taking X' and Y' to be independent. This coupling is however rarely the best suited for applications.

EXAMPLE 5.12. Fix $n \geq 1$ and $0 \leq p \leq q \leq 1$ and consider the binomial distributions with parameters (n, p) and (n, q) . We can couple them as follows: let $U_1, \dots, U_n$ be i.i.d. random variables with the uniform distribution on $[0, 1]$ and let $X = \#\{i \leq n : U_i \leq p\}$ and $Y = \#\{i \leq n : U_i \leq q\}$. Then they have the prescribed binomial laws and in addition it holds $X \leq Y$.

EXERCISE 5.13. Find similarly a coupling of two Poisson random variables $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\lambda')$ with $X \leq Y$ when $\lambda \leq \lambda'$. Find a coupling between two exponential random variables $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\lambda')$ with $X \leq Y$ when $\lambda \geq \lambda'$.

The following result provides a probabilistic interpretation of the Lévy–Prohorov distance by relating it to couplings and the Ky Fan metric.

THEOREM 5.14 (Strassen). Assume that (E, d) is separable and let μ and ν be two probability measures. Then

$$d_{LP}(\mu, \nu) = \inf_{\text{coupling}(X, Y)} \alpha(X, Y),$$

where X has law μ and Y has law ν .

Proof of $d_{LP} \leq \inf \alpha$. This inequality is simple and useful in practice, to control the Lévy–Prohorov distance. Let $\alpha > 0$ and assume that there exists a coupling such that $\alpha(X, Y) < \alpha$. Let C be a closed set, then

$$\mathbb{P}(X \in C) \leq \mathbb{P}(X \in C, d(X, Y) < \alpha) + \mathbb{P}(d(X, Y) \geq \alpha) \leq \mathbb{P}(Y \in C^\alpha) + \alpha,$$

so $d_{LP}(\mu, \nu) \leq \alpha$. Since this holds for every $\alpha > \alpha(X, Y)$, then $d_{LP}(\mu, \nu) \leq \alpha(X, Y)$. $\square$

The proof of the converse inequality is much harder and technical, and this makes sense since one has to construct a suitable coupling by hand. We omit it and refer e.g. to the book of Ethier & Kurtz or Billingsley for a complete proof of Theorem 5.14. This construction has the same flavour as the proof of Skorokhod's representation theorem 5.5 but is more delicate.

5.5 TOTAL VARIATION DISTANCE

The notion of coupling is also related to the following distance on probability measures. Note that here we work on a general measurable space $(E, \mathcal{E})$, not necessarily a metric space.

DEFINITION 5.15. Let μ_1 and μ_2 be two probability measures on the measurable space $(E, \mathcal{E})$. The quantity:

$$d_{TV}(\mu_1, \mu_2) = \sup_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)|$$

is called the *total variation distance* between μ_1 and μ_2 . We define the total variation distance between

two random variables on the same probability space and with values in E as the distance between their laws:

$$d_{TV}(X, Y) = \sup_{A \in \mathcal{E}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

Then random variables X_n converge to another one X for the total variation distance when their total variation distance tends to 0.

Convergence in total variation distance intuitively corresponds to uniform convergence if one sees a measure as a function from $\mathcal{E}$ to $[0, 1]$. It is stronger than convergence in law which is nearly a pointwise convergence by the portmanteau theorem 1.10.

EXERCISE 5.16. Prove that the total variation distance is indeed a distance between probability measures on the same space $(E, \mathcal{E})$.

Total variation distance is intimately related to couplings as follows. Recall that $\Gamma(\mu_1, \mu_2)$ is the set of distributions on pairs whose coordinates have law μ_1 and μ_2 respectively.

THEOREM 5.17. Let μ_1 and μ_2 be two probability measures on the same space $(E, \mathcal{E})$. Then

$$d_{TV}(\mu_1, \mu_2) = \inf_{\text{coupling } (X, Y)} \mathbb{P}(X \neq Y),$$

where X has law μ_1 and Y has law μ_2 . In addition, there exists a coupling (X, Y) of (μ_1, μ_2) that reaches the bound.

Proof of $d_{TV} \leq \inf \mathbb{P}(X \neq Y)$. Let us first prove this simple inequality which, again, is the one useful in practice. For any coupling $\gamma \in \Gamma(\mu_1, \mu_2)$ and any event $A \in \mathcal{E}$ it holds:

$$\begin{aligned} \mathbb{P}_\gamma(X \neq Y) &\geq \mathbb{P}_\gamma(X \in A, Y \in A^c) \\ &\geq \mathbb{P}_\gamma(X \in A, Y \in A^c) - \mathbb{P}_\gamma(X \in A^c, Y \in A) \\ &= \mathbb{P}_\gamma(X \in A, Y \in A) + \mathbb{P}_\gamma(X \in A, Y \in A^c) - \mathbb{P}_\gamma(X \in A, Y \in A) - \mathbb{P}_\gamma(X \in A^c, Y \in A) \\ &= \mathbb{P}_\gamma(X \in A) - \mathbb{P}_\gamma(Y \in A) \\ &= \mu_1(A) - \mu_2(A). \end{aligned}$$

Since this holds true for any $A \in \mathcal{E}$, then $\mathbb{P}_\gamma(X \neq Y) \geq d_{TV}(\mu_1, \mu_2)$ for any $\gamma \in \Gamma(\mu_1, \mu_2)$. $\square$

This result is useful in practice since it shows that if one can find a coupling for which $\mathbb{P}(X \neq Y)$ is small, then the total variation distance is small. Nevertheless it is true in general that the infimum is actually achieved by a certain coupling, and that it equals $d_{TV}(\mu_1, \mu_2)$ as we shall prove below. Before that, let us give an alternative expression for this distance when that measures have a density with respect to the same reference measure. Recall that μ is said to have density f with respect to ν when

$$\forall A \in \mathcal{E}, \quad \mu(A) = \int_A f \, d\nu.$$

The Radon–Nikodým theorem states that for σ -finite measures, the existence of a density is equivalent to the absolute continuity property:

$$\forall A \in \mathcal{E}, \quad \nu(A) = 0 \implies \mu(A) = 0.$$

LEMMA 5.18. Let two probability measures μ_1 and μ_2 have a density f_1 and f_2 respectively with respect

to the same measure ν . Then

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \int |f_1 - f_2| d\nu = \max_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)|.$$

As usual, if ν is the Lebesgue measure on $E = \mathbb{R}^d$, then the densities are the usual ‘probability densities’. Informally, the integral in the middle then represents the positive volume between the curves of f_1 and f_2 . When E is a discrete countable set, we may take ν to be the counting measure, so we read:

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)| = \max_{A \subseteq E} |\mu_1(A) - \mu_2(A)|.$$

Proof. For the first equality, there is an easy upper bound. For any $A \in \mathcal{E}$, we have since μ_1 and μ_2 are both probability measures:

$$\mu_1(A) - \mu_2(A) = (1 - \mu_1(A^c)) - (1 - \mu_2(A^c)) = \mu_2(A^c) - \mu_1(A^c).$$

Thus

$$\begin{aligned} 2|\mu_1(A) - \mu_2(A)| &= |\mu_1(A) - \mu_2(A)| + |\mu_2(A^c) - \mu_1(A^c)| \\ &= \left| \int_A (f_1 - f_2) d\nu \right| + \left| \int_{A^c} (f_1 - f_2) d\nu \right| \\ &\leq \int_A |f_1 - f_2| d\nu + \int_{A^c} |f_1 - f_2| d\nu \\ &= \int |f_1 - f_2| d\nu. \end{aligned}$$

Hence $d_{TV}(\mu_1, \mu_2) \leq \frac{1}{2} \int |f_1 - f_2| d\nu$ by taking the supremum over all $A \in \mathcal{E}$. Further, notice that for the particular measurable set

$$A = \{x \in E : f_1(x) \geq f_2(x)\},$$

we have precisely $f_1 - f_2 = |f_1 - f_2|$ on A and $f_1 - f_2 = -|f_1 - f_2|$ on A^c , so that in the previous calculation, the inequality at the third line is an equality. This proves both that the supremum in d_{TV} is a maximum and that this maximal values equals $\frac{1}{2} \int |f_1 - f_2| d\nu$. $\square$

We may now prove the equality in Theorem 5.17

Proof of Theorem 5.17. Let us provide a coupling of μ_1 and μ_2 that reaches the lower bound. Notice that μ_1 and μ_2 are absolutely continuous with respect to the finite measure $\nu = \mu_1 + \mu_2$. According to the Radon–Nikodým theorem, we may therefore, and will, assume without loss of generality that μ_1 and μ_2 admit a density f_1 and f_2 respectively with respect to the same measure ν . Let us use the same notation $A = \{x \in E : f_1(x) \geq f_2(x)\}$ as in the previous proof and let

$$p = 1 - \int \min(f_1, f_2) d\nu = 1 - (\mu_2(A) + \mu_1(A^c)) = \mu_1(A) - \mu_2(A) = d_{TV}(\mu_1, \mu_2),$$

where the last equality was derived in the previous proof. Let then ξ have the Bernoulli law with parameter p . If $\xi = 0$, then let $X = Y$ have density with respect to ν given by:

$$g^0(x) = \frac{1}{1-p} \min(f_1(x), f_2(x)) = \frac{1}{1-p} (f_1(x) \mathbb{1}_{x \in A^c} + f_2(x) \mathbb{1}_{x \in A}).$$

It is indeed the density of a probability since $g^0 \geq 0$ is measurable and

$$\int g^0 d\nu = \frac{1}{1-p} \int (f_1(x) \mathbb{1}_{x \in A^c} + f_2(x) \mathbb{1}_{x \in A}) d\nu = \frac{1}{1-p} (\mu_1(A^c) + \mu_2(A)) = 1.$$

If $\xi = 1$, then let X and Y be independent and with density with respect to ν given respectively by:

$$g_X^1(x) = \frac{1}{p}(f_1(x) - f_2(x)) \mathbb{1}_{x \in A} \quad \text{and} \quad g_Y^1(y) = \frac{1}{p}(f_2(y) - f_1(y)) \mathbb{1}_{y \in A^c}.$$

One checks as above that these are indeed densities.

By conditioning with respect to ξ , we infer that for any $B \in \mathcal{E}$, we have:

$$\begin{aligned} \mathbb{P}(X \in B) &= p \mathbb{P}(X \in B \mid \xi = 1) + (1 - p) \mathbb{P}(X \in B \mid \xi = 0) \\ &= p \int_B g_X^1 d\nu + (1 - p) \int_B g^0 d\nu \\ &= \left(\int_{A \cap B} f_1 d\nu - \int_{A \cap B} f_2 d\nu \right) + \left(\int_{A^c \cap B} f_1 d\nu + \int_{A \cap B} f_2 d\nu \right) \\ &= \int_{A \cap B} f_1 d\nu + \int_{A^c \cap B} f_1 d\nu \\ &= \mu_1(A \cap B) + \mu_1(A^c \cap B) \\ &= \mu_1(B). \end{aligned}$$

Thus X has the law μ_1 . A similar calculation shows that Y the law μ_2 so we have indeed constructed a coupling. Finally, if $\xi = 0$ then $X = Y$ and if $\xi = 1$, then $X \in A$ and $Y \in A^c$ so $X \neq Y$ if and only if $\xi = 1$, which occurs with probability $p = d_{TV}(\mu_1, \mu_2)$. $\square$

Finally, the convergence of the densities implies the convergence in total variation. The proof of the next result is left as an exercise.

LEMMA 5.19 (Scheffé). Let ν be a σ -finite measure and let μ and $(\mu_n)_{n \geq 1}$ be probability measures with densities f and f_n respectively with respect to ν . If $f_n(x) \rightarrow f(x)$ for ν -almost every x , then

$$\int |f_n - f| d\nu \xrightarrow{n \rightarrow \infty} 0$$

and consequently $\mu_n \rightarrow \mu$ in total variation.

Consider once again the two typical examples of applications. If $\nu = \text{Leb}$ is the Lebesgue measure on $\mathbb{R}^d$, then f_n and f are usual ‘probability densities’. This is the original framework of Henry Scheffé. If E is a discrete countable set, such as $\mathbb{Z}^d$, and ν is the counting measure then X_n and X are usual ‘discrete random variables’ and in this case $f(x) = \mathbb{P}(X = x)$. Actually in this case we have proved the following equivalence.

PROPOSITION 5.20. Let E be a countable discrete space and let X and X_n be random variables on E . The following assertions are equivalent:

- (i) $\mathbb{P}(X_n = x) \rightarrow \mathbb{P}(X = x)$ for every $x \in E$,
- (ii) we have $X_n \rightarrow X$ in total variation distance,
- (iii) we have $X_n \rightarrow X$ in distribution.

Proof. Scheffé’s lemma 5.19 with ν the counting measure on E proves that (i) $\implies$ (ii). We know that (ii) $\implies$ (iii) and we have proved already in Corollary 1.11 that (iii) $\iff$ (i). $\square$

Let us mention for the curious reader that there exist other distances on probabilities that relate to couplings in a different way, such as the *Wasserstein distance*, especially the L^1 and L^2 versions, which is extensively used for example in optimal transport and its application in machine learning.

5.6 EXERCISES

EXERCISE 23. Let X_n and X be random variables defined on the same probability space and with values in a separable metric space E . Prove that:

- (i) If $X_n \rightarrow X$ in probability, then also in distribution.
- (ii) The converse holds when X is almost surely constant.

EXERCISE 24 (Slutsky). Let X_n, Y_n, X , and Y be random variables defined on the same probability space, where X_n and X take values in a separable metric space E_1 and Y_n and Y take values in a separable metric space E_2 . Suppose that $X_n \rightarrow X$ and $Y_n \rightarrow Y$ both in distribution and that Y is almost surely constant. Prove that $(X_n, Y_n) \rightarrow (X, Y)$ in distribution in $E_1 \times E_2$.

EXERCISE 25. For any pair of random variables (X, Y) in some separable metric space (E, d) , let

$$\beta(X, Y) = \mathbb{E} \left[\frac{d(X, Y)}{d(X, Y) + 1} \right].$$

As always, let us not distinguish random variable that are almost surely equal.

- (i) Prove that β is a distance on random variables (up to almost sure equality)
- (ii) Prove that $\beta(X_n, X) \rightarrow 0$ if and only if $X_n \rightarrow X$ in probability.

Recall that α stands for the Ky Fan metric from Definition 5.6. We just proved that α and β induce the same topology; the next questions investigate quantitative links between them.

- (iii) Prove that there exist constants $c_1, c_2 < \infty$ such that

$$\beta \leq c_1 \alpha \quad \text{and} \quad \alpha \leq c_2 \beta^{1/2}.$$

- (iv) Prove that there is no $c < \infty$ such that $\alpha \leq c\beta$.

EXERCISE 26. Let $\mathbb{Q}$ be a probability on $(\Omega, \mathcal{A})$ which is equivalent to $\mathbb{P}$, in the sense that for every $A \in \mathcal{A}$, we have $\mathbb{Q}(A) = 0$ if and only if $\mathbb{P}(A) = 0$. Prove that convergence in probability for $\mathbb{P}$ is equivalent to convergence in probability for $\mathbb{Q}$ but that in general, there is no universal constant $c < \infty$ such that

$$\alpha_{\mathbb{P}} \leq c \alpha_{\mathbb{Q}}.$$

Here $\alpha_{\mathbb{P}}$ and $\alpha_{\mathbb{Q}}$ refer to the Ky Fan metric with the corresponding underlying probability on Ω .

EXERCISE 27 (Scheffé's lemma). Prove Lemma 5.19, namely that if ν is a σ -finite measure on $(E, \mathcal{E})$ and if f_n and f are nonnegative functions with ν -integral equal to 1, then

$$f_n \xrightarrow[n \rightarrow \infty]{} f \text{ } \nu\text{-almost everywhere} \implies \int |f_n - f| d\nu \xrightarrow[n \rightarrow \infty]{} 0.$$

Hint: you may write $|f_n - f| = f_n + f - 2 \min(f_n, f)$.

PART II

CONVERGENCE OF STOCHASTIC
PROCESSES

CHAPTER 6

CONTINUOUS STOCHASTIC PROCESSES

Back to following Billingsley. We focus in this chapter on random variables with values in the space of continuous functions from an interval to a Polish space. After equipping this space with a suitable Polish topology, we shall develop in this framework tools to prove tightness and weak convergence of such random functions.

CONTENTS

6.1	Stochastic processes vs. random functions	64
6.2	Separability and completeness	65
6.3	Marginal laws	66
6.4	The compact sets	68
6.5	Tightness criteria	70
6.6	Exercises	75

6.1 STOCHASTIC PROCESSES VS. RANDOM FUNCTIONS

STOCHASTIC PROCESSES

Our aim in this part is to study convergence in distribution of stochastic processes. Let us first give a formal definition of these objects.

DEFINITION 6.1. A *stochastic process* is a collection of random variables $X = (X_i)_{i \in I}$, all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in the same *state space* $(E, \mathcal{E})$, and with some *index set* I . Equivalently, it is a function $X : \Omega \times I \rightarrow E$ such that for every $i \in I$ fixed, the function $\omega \in \Omega \mapsto X_i(\omega)$ is measurable. For every $\omega \in \Omega$, the function $t \in I \mapsto X_t(\omega)$ is called a *sample path* or *realisation* of X .

We shall denote functions from I to E by x, y, z and then random functions by X, Y, Z . We shall denote by s, t indices in I , so $x_t \in E$ is the value of the function x at position t .

In this course, we shall only be concerned with the cases $I = [0, T]$ for some $T > 0$ and $I = [0, \infty)$, in which case the process X is thought of as the evolution over time of a single random variable. However, there are many interesting applications with more general sets I (even for cases in which I is itself a random set...), in which case we rather speak of *random fields* over I . Many of the results presented here carry over with little changes to this more general setting.

The state space E will be a separable and complete metric space and, although we shall here illustrate the results with $E = \mathbb{R}$ or $\mathbb{R}^d$, there are similarly many interests in studying stochastic processes with values in more abstract spaces, such as the space of (probability) measures on a metric space, or a subspace of the space of functions with values in a metric space, namely each X_t is a random measure on some space, or a random function from a set to another one.

CONTINUOUS FUNCTIONS

Recall from Chapter 3 that a natural framework to study stochastic processes is the product σ -algebra, which has several advantages, such as Kolmogorov's extension theorem. However it also has disadvantages and for example, we saw in Exercise 18 that it does not contain the set of continuous functions so it does not make sense a priori to speak in this framework of the probability that a given stochastic process is continuous... In another direction, theory of weak convergence is developed in the context of the Borel σ -algebra, and the one induced by the natural product topology is in general strictly bigger than the product σ -algebra, and in fact too big. For these reasons, we first focus in this chapter and the next one on continuous functions only, which eliminates these problems and provides a very efficient framework to use the tools of weak convergence developed in the first part.

From now on, the index set I will be either a compact interval $[0, T]$, or the half-line $[0, \infty)$. We shall denote by d_E the metric on E , whilst the notation d will refer to a distance on functions from I to E . We assume henceforth that (E, d_E) is separable and complete.

DEFINITION 6.2. We let $\mathcal{C} = \mathcal{C}(I, E) \subset E^I$ denote the set of continuous and bounded functions from I to E . When $I = [0, T]$ is compact, we equip $\mathcal{C}$ with the distance

$$d(x, y) = \sup_{t \in [0, T]} d_E(x_t, y_t).$$

When $I = [0, \infty)$, we use instead:

$$d(x, y) = \sum_{n \geq 1} \frac{\sup_{t \in [0, n]} d_E(x_t, y_t) \wedge 1}{2^n}.$$

In both cases, we further equip $(\mathcal{C}, d)$ with the associated Borel σ -algebra $\mathcal{B}(\mathcal{C})$.

REMARK 6.3. When $I = [0, T]$ is a compact interval, convergence for the distance d corresponds to uniform convergence; when $I = [0, \infty)$, it is the uniform convergence on compact intervals.

6.2 SEPARABILITY AND COMPLETENESS

First, let us check that we are in the best position to apply our general results on weak convergence.

PROPOSITION 6.4. The space $(\mathcal{C}, d)$ is separable and complete.

REMARK 6.5. Exercise 28 shows in the case $I = [0, \infty)$ that the uniform distance over the entire half-line however *does not* yield a separable space in general. This proposition explains why in this case we consider the uniform convergence on compact sets.

Proof for compact intervals. Let us consider here the case $I = [0, T]$. Let us first prove completeness. Suppose that $(x^n)_{n \geq 1}$ is a Cauchy sequence in $(\mathcal{C}, d)$. For every $t \in [0, T]$, since $d_E(x_t^n, x_t^m) \leq d(x^n, x^m)$, then the sequence $(x_t^n)_{n \geq 1}$ is a Cauchy sequence in the complete space E , so it converges to some limit x_t . In other words the functions x^n converge pointwise to a function x . To see that it converges uniformly to x , note that for every $\varepsilon > 0$, there exists $n_0 \geq 1$ such that for all $n, m \geq n_0$ and all $t \in [0, T]$, we have $d_E(x_t^n, x_t^m) \leq d(x^n, x^m) \leq \varepsilon$. Keep n fixed and let $m \rightarrow \infty$ to obtain $d_E(x_t^n, x_t) \leq \varepsilon$. Since this holds for every $t \in [0, T]$, then $d(x^n, x) \leq \varepsilon$, with $\varepsilon > 0$ arbitrary. Continuity of x then follows easily from continuity of x^n and this bound.

Let us next prove that $\mathcal{C}$ is separable, again when $I = [0, T]$. For every $n, m \geq 1$, let $\mathcal{C}_{n,m}$ denote the set of those continuous functions x such that

$$\sup_{1 \leq k \leq m} \sup_{s, t \in [\frac{(k-1)T}{m}, \frac{kT}{m}]} d_E(x_s, x_t) < \frac{1}{n}.$$

Fix a countable dense subset D of E and for every $n, m \geq 1$ and every $a_0, \dots, a_m \in D$, choose one arbitrary function y in $\mathcal{C}_{n,m}$ that satisfies $d_E(y_{kT/m}, a_k) < 1/n$ for every $0 \leq k \leq m$, provided there exists any. Let $\mathcal{D}_{n,m}$ denote the countable collection of those functions y as the a_i 's vary in D . Let $x \in \mathcal{C}$ be any function and fix $n \geq 1$ arbitrary. Since x is uniformly continuous, then there exists $m \geq 1$ such that $x \in \mathcal{C}_{n,m}$, and since D is dense in E , then for each $0 \leq k \leq m$, there exists $a_k \in D$ such that $d_E(x_{kT/m}, a_k) < 1/n$. Consequently $\mathcal{D}_{n,m}$ is not empty and contains a function y . For every $t \in [0, T]$, there exists $0 \leq k \leq m$ such that $|t - kT/m| < 1/m$, so we infer from the triangle inequality that

$$d_E(x_t, y_t) \leq d_E(x_t, x_{kT/m}) + d_E(x_{kT/m}, a_k) + d_E(a_k, y_{kT/m}) + d_E(y_{kT/m}, y_t) < 4/n,$$

since both $x, y \in \mathcal{C}_{n,m}$. We have thus shown that for every $x \in \mathcal{C}$, for every $n \geq 1$, there exists $y \in \bigcup_m \mathcal{D}_{n,m}$ such that $d(x, y) < 1/n$. The countable set $\bigcup_{n,m} \mathcal{D}_{n,m}$ is therefore dense. $\square$

Let us next extend the argument to $I = [0, \infty)$.

Proof for the half-line. Take now $I = [0, \infty)$. The proof of completeness is exactly the same: a Cauchy sequence converges pointwise and further uniformly on compact intervals, hence for the distance d . As for separability, the previous argument shows that for every $N \geq 1$, there exists a countable set of functions on $[0, N]$ that approximates any function $x \in \mathcal{C}$ uniformly on $[0, N]$. Extend any such function on $[0, \infty)$ by assigning its value at N on the entire interval $[N, \infty)$, then the union of these functions over N approximates any continuous function uniformly on compact intervals. $\square$

A consequence of this result is that we can apply the ‘standard scheme’: if $(\mu_n)_n$ and μ are probability measures on $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, then $\mu_n \Rightarrow \mu$ if and only if both

- (i) the sequence $(\mu_n)_n$ is tight and (ii) any subsequence that converges weakly has μ as limit.

Let us first discuss the second point before turning to tightness.

6.3 MARGINAL LAWS

Recall the product σ -algebra $\mathcal{B}(E)^{\otimes I}$ on the set of all functions $I \rightarrow E$, which is the one generated by the evaluation maps $x \mapsto x_t$ for every $t \in I$. Although $\mathcal{C}$ is not a measurable set for this σ -algebra (recall Exercise 18), we can consider its trace $\mathcal{B}'(\mathcal{C}) = \{\mathcal{C} \cap A : A \in \mathcal{B}(E)^{\otimes I}\}$, which is the σ -algebra generated again by the evaluation maps $x \mapsto x_t$ for every $t \in I$, but viewed as projections from $\mathcal{C}$ to E , i.e. restricted to continuous functions. This σ -algebra on $\mathcal{C}$ enjoys the same properties as in Theorem 3.7, with the exact same proof, namely:

- measurability of $X : \Omega \rightarrow \mathcal{C}$ for $\mathcal{B}'(\mathcal{C})$ is equivalent to measurability of $X_t : \Omega \rightarrow E$ for every $t \in I$,
- the law of a random variable X with values in $(\mathcal{C}, \mathcal{B}'(\mathcal{C}))$ is entirely characterised by the joint law of $(X_t)_{t \in J}$ for every finite subset $J \subset I$, and one can even just consider finite product of sets in $\mathcal{B}(E)$.

Recall however that in order to consider weak convergence, we want to work with the Borel σ -algebra $\mathcal{B}(\mathcal{C})$, and not the product σ -algebra $\mathcal{B}'(\mathcal{C})$. Hopefully, what fails in general for the set of all functions (recall Proposition 3.15) holds true for continuous functions.

PROPOSITION 6.6. The two σ -algebras $\mathcal{B}(\mathcal{C})$ and $\mathcal{B}'(\mathcal{C})$ coincide.

Consequently, the above properties are valid for the Borel σ -algebra $\mathcal{B}(\mathcal{C})$. The essential idea is that a continuous function is characterised by its values on a countable dense set.

Proof. For every $t \in I$, the evaluation map $x \in \mathcal{C} \mapsto x_t \in E$ is clearly continuous and hence measurable for $\mathcal{B}(\mathcal{C})$, which thus contains $\mathcal{B}'(\mathcal{C})$. Let us prove the converse inclusion.

Let us start with $I = [0, T]$. Fix arbitrary $x \in \mathcal{C}$ and $\varepsilon > 0$. Let $\mathcal{Q}_T = \mathcal{Q} \cap [0, T]$, then by continuity,

$$\begin{aligned} B(x, \varepsilon) &= \{y \in \mathcal{C} : d(x, y) < \varepsilon\} = \{y \in \mathcal{C} : \sup_{t \in [0, T]} d_E(x_t, y_t) < \varepsilon\} \\ &= \{y \in \mathcal{C} : \sup_{t \in \mathcal{Q}_T} d_E(x_t, y_t) < \varepsilon\} \\ &= \bigcap_{t \in \mathcal{Q}_T} \pi_t^{-1}(B_E(x_t, \varepsilon)), \end{aligned}$$

where $B_E(x_t, \varepsilon)$ is the open ball in E centred at x_t and with radius ε . It belongs to $\mathcal{E}$, so $\pi_t^{-1}(B_E(x_t, \varepsilon)) \in \mathcal{B}'(\mathcal{C})$ and finally the open ball $B(x, \varepsilon)$ in $\mathcal{C}$ belongs to $\mathcal{B}'(\mathcal{C})$ since $\mathcal{Q}_T$ is countable. This σ -algebra thus contains every open ball of $\mathcal{C}$. Since $\mathcal{C}$ is separable by Proposition 6.4, then any open set can be written as a countable union of open balls: for example the balls centred at each element of a countable dense set, with rational radius, and which are entirely contained in the open set. Consequently, the σ -algebra $\mathcal{B}'(\mathcal{C})$ contains every open set of $\mathcal{C}$, and hence the entire Borel σ -algebra $\mathcal{B}(\mathcal{C})$.

In the case $I = [0, \infty)$, let

$$B^n(x, \varepsilon) = \{y \in \mathcal{C} : \sup_{t \in [0, n]} d_E(x_t, y_t) < \varepsilon\}.$$

Then $B(x, \varepsilon 2^{-n}) \subset B^n(x, \varepsilon) \subset B(x, \varepsilon)$ for every $x \in \mathcal{C}$, every $n \geq 1$, and every $\varepsilon \in (0, 1)$. Again, by separability, any open set O of $\mathcal{C}$ can be written as the union of those balls $B(x, \varepsilon)$ with x in a countable dense set D of E , $\varepsilon \in \mathcal{Q} \cap (0, 1)$, and $B(x, \varepsilon) \subset O$. One can also restrict the values of ε to $\mathcal{Q} \cap (0, 1/2)$, so by the previous inclusions, any open set O can be written as the union of the sets $B^n(x, \varepsilon)$ with $x \in D$, $\varepsilon \in \mathcal{Q} \cap (0, 1)$, $n \geq 1$, and $B^n(x, \varepsilon) \subset O$. The argument in the case $I = [0, T]$ shows that each such set $B^n(x, \varepsilon)$ belongs to $\mathcal{B}'(\mathcal{C})$, so we conclude again that this σ -algebra contains all the open sets and thus the Borel σ -algebra $\mathcal{B}(\mathcal{C})$. $\square$

DEFINITION 6.7. Let X and $(X^n)_{n \geq 1}$ be random variables in $\mathcal{C}$. We say that X^n converges in distribution to X in the sense of finite-dimensional marginals when for every finite subset $J \subset I$, it holds

$$(X_t^n)_{t \in J} \xrightarrow[n \rightarrow \infty]{(d)} (X_t)_{t \in J} \quad \text{in} \quad E^J.$$

Convergence in this sense is much simpler to check in practice than weak convergence of the entire process and happens to characterise uniquely the weak limit, provided we know that the sequence does converge weakly in the first place. Precisely, the following result, made simple thanks to our previous work, is (almost always) used in practice to prove convergence of continuous stochastic processes.

THEOREM 6.8 (Proving weak convergence). Let X and $(X^n)_{n \geq 1}$ be random variables in $\mathcal{C}$. Then X^n converges in distribution to X in $\mathcal{B}(\mathcal{C})$ if and only if both the sequence of the laws of X^n is tight and $X^n \rightarrow X$ in distribution the sense of finite-dimensional marginal laws.

Proof. Suppose first that $X^n \rightarrow X$ in distribution in $\mathcal{B}(\mathcal{C})$. Then the convergence holds in the sense of finite-dimensional marginals by continuity of the projections $\pi_J : x \mapsto (x_t)_{t \in J}$ for $J \subset I$ finite. In

addition, the set $\{X^n : n \geq 1\} \cup \{X\}$ of (the laws of) these random variables is compact. This implies that it is tight by Prohorov's theorem 1.25 since $\mathcal{C}$ is separable and complete by Proposition 6.4. Finally, the sub-collection $\{X^n : n \geq 1\}$ is then also tight.

Conversely assume that $(X^n)_n$ is tight and X^n converges to X in distribution in the sense of finite-dimensional marginals. By Prohorov's theorem 1.25, every sub-sequence has a further subsequence that converges in distribution to some random variable Y . But this sub-sequence also converges to X in the sense of finite-dimensional marginals. We infer that X and Y have the same finite-dimensional marginals, so their laws agree on $\mathcal{B}(\mathcal{C})$ by Proposition 6.6. $\square$

EXAMPLE 6.9. Consider the function $x^n : [0, 1] \rightarrow \mathbb{R}$ which is constant equal to 0 on $[0, 1/2] \cup [1/2 + 1/(2n), 1]$, which is equal to 1 at $1/2 + 1/n$, and which is linear on $[1/2, 1/2 + 1/n]$ and on $[1/2 + 1/n, 1/2 + 1/(2n)]$. It clearly converges to the function 0 constant equal to zero on $[0, 1]$ in the sense of finite-dimensional marginals since every finite set of $[0, 1]$ eventually lies outside the open interval $(1/2, 1/2 + 1/(2n))$. However $d(x^n, 0) = 1$ so the convergence does not hold in the uniform sense: the sequence $(x^n)_n$ is not relatively compact in $\mathcal{C}$.

There is no general scheme to prove the convergence in the sense of finite-dimensional marginals, the argument depends on the context. However, proving tightness requires to characterise the compact subsets of $\mathcal{C}$, which is the goal of the next section.

6.4 THE COMPACT SETS

COMPACT INTERVALS

Let us explicit the compact sets of $\mathcal{C}(I, E)$ first when $I = [0, T]$ is a compact interval.

DEFINITION 6.10. For any function $x : [0, T] \rightarrow E$ (not necessarily continuous), we define its *modulus of continuity* at any $\delta > 0$ by

$$\omega(x, \delta) = \sup\{d_E(x_s, x_t) : s, t \in [0, T], |t - s| < \delta\}.$$

A collection $\mathcal{K}$ of functions on $[0, T]$ is said to be (uniformly) *equi-continuous* when for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sup_{x \in \mathcal{K}} \omega(x, \delta) < \varepsilon.$$

REMARK 6.11. For a single function x , equi-continuity simply means $\omega(x, \delta) \rightarrow 0$ as $\delta \rightarrow 0$, which is equivalent to uniform continuity, which amounts to continuity since $[0, T]$ is compact. For a collection of functions, we require that this holds uniformly over these functions. In particular, any finite collection of functions in $\mathcal{C}([0, T], E)$ is equi-continuous. Also note that, for any $\delta > 0$, the mapping $x \in \mathcal{C}([0, T], E) \mapsto \omega(x, \delta)$ is continuous and so measurable for the Borel σ -algebras.

THEOREM 6.12 (Arzelà–Ascoli). A collection $\mathcal{K}$ of functions in $\mathcal{C}([0, T], E)$ has compact closure if and only if it is equi-continuous and there exists a countable and dense subset $D \subset [0, T]$ such that for every $t \in D$, the set $\pi_t(\mathcal{K}) = \{x_t : x \in \mathcal{K}\}$ has compact closure in E .

Here is a simplification in some nice cases, such as $E = \mathbb{R}^d$ for some $d \geq 1$ that allows to replace D by a singleton, usually 0 in practice.

COROLLARY 6.13. If E has the property that all bounded sets have compact closure, then a collection $\mathcal{K}$ of functions in $\mathcal{C}([0, T], E)$ has compact closure if and only if it is equi-continuous and there exists $t \in [0, T]$ such that the set $\pi_t(\mathcal{K}) = \{x_t : x \in \mathcal{K}\}$ is bounded in E .

Proof of Corollary 6.13. For every $\delta > 0$, let us cover the interval $[0, T]$ by finitely many, say $k_\delta \geq 1$, intervals with length $\delta/2$. It follows from the triangle inequality that for any $s, t \in [0, T]$, we have $d_E(x_s, x_t) \leq (k_\delta + 2) \times \omega(x, \delta)$. In an equi-continuous collection $\mathcal{K}$, taking $\varepsilon = 1$ and the corresponding δ , this quantity is bounded uniformly over $x \in \mathcal{K}$. Consequently, if the set $\{x_t : x \in \mathcal{K}\}$ is bounded for some $t \in [0, T]$, then it is bounded for every $t \in [0, T]$, and hence relatively compact in E . $\square$

Proof of Theorem 6.12. $\implies$: Suppose that $\mathcal{K}$ is relatively compact in $\mathcal{C}([0, T], E)$. Fix any $t \in [0, T]$, since the evaluation map $\pi_t : x \mapsto x_t$ is continuous and the closure $\overline{\mathcal{K}}$ is compact, then its image $\pi_t(\overline{\mathcal{K}})$ is compact. Consequently the set $\{x_t : x \in \mathcal{K}\}$ is included in the compact set $\{x_t : x \in \overline{\mathcal{K}}\}$: it is relatively compact. Let us next prove that $\mathcal{K}$ is equi-continuous. Fix $\varepsilon > 0$ arbitrary. First, by the triangle inequality, for any $x, y \in \mathcal{C}([0, T], E)$ and any $s, t \in [0, T]$, we have

$$d_E(y_s, y_t) \leq d_E(y_s, x_s) + d_E(x_s, x_t) + d_E(y_t, x_t),$$

from which it follows that for any $\delta > 0$, we have

$$\omega(y, \delta) \leq \omega(x, \delta) + 2d(x, y).$$

Since $\overline{\mathcal{K}}$ is compact, then it can be covered by finitely many balls with radius $\varepsilon/2$, centred at, say, $x^1, \dots, x^k$. Then $\mathcal{K}$ is covered by the balls with the same centers and radius ε . Applying the previous bound to any function $y \in \mathcal{C}$ and x any of the x^i 's, we obtain that for any $\delta > 0$, we have

$$\sup_{y \in \mathcal{K}} \omega(y, \delta) \leq \sup_{1 \leq i \leq k} \omega(x^i, \delta) + 2\varepsilon.$$

Since any finite collection of continuous functions is equi-continuous, then the right-hand side is less than 3ε for δ small enough, which proves that $\mathcal{K}$ is equi-continuous.

$\impliedby$: Suppose conversely that $\mathcal{K}$ is an equi-continuous collection and that there is a countable and dense subset $D \subset [0, T]$ such that for every $t \in D$, the set $\pi_t(\mathcal{K}) = \{x_t : x \in \mathcal{K}\}$ is relatively compact in E . Let $(x^n)_n$ be a sequence in $\mathcal{K}$ and let us prove that it admits a subsequence that converges in $\mathcal{C}([0, T], E)$. For every $t \in D$, the sequence $(x_t^n)_n$ lies in the relatively compact set $\pi_t(\mathcal{K})$, so it admits a subsequence that converges to some limit $x_t \in E$. By a diagonal extraction argument, since D is countable, there is a subsequence $(y^n)_n$ of $(x^n)_n$ such that $y_t^n \rightarrow y_t$ for every $t \in D$. We claim that this subsequence $(y^n)_n$ is a Cauchy sequence in $\mathcal{C}([0, T], E)$; since the latter is complete, then y^n must converge.

Since $[0, T]$ is compact, then for every $\delta > 0$, it can be covered by finitely many intervals with length δ , say $I_1, \dots, I_{k_\delta}$. Since D is dense, then each such interval I_j contains an element of D ; choose one arbitrary t_j for each interval and then for every $t \in [0, T]$, let t_δ be one of these t_j 's for an interval such that $t \in I_j$ (make an arbitrary choice if there are several I_j 's). Since $|t - t_\delta| < \delta$, then for every $n, k \geq 1$, we have:

$$\begin{aligned} d_E(y_t^n, y_t^{n+k}) &\leq d_E(y_t^n, y_{t_\delta}^n) + d_E(y_{t_\delta}^n, y_{t_\delta}^{n+k}) + d_E(y_{t_\delta}^{n+k}, y_t^{n+k}) \\ &\leq d_E(y_{t_\delta}^n, y_{t_\delta}^{n+k}) + \omega(y^n, \delta) + \omega(y^{n+k}, \delta). \end{aligned}$$

Since $\mathcal{K}$ is equi-continuous, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\omega(y^n, \delta) \leq \varepsilon$ uniformly over n . Also, since there are only finitely values of t_δ , and since $y_{t_\delta}^n \rightarrow y_{t_\delta}$ for each of them, then $d_E(y_{t_\delta}^n, y_{t_\delta}^{n+k}) \leq \varepsilon$ for every n large enough, uniformly over $k \geq 1$ and t_δ . Hence

$$\sup_{k \geq 1} \sup_{t \in [0, T]} d_E(y_t^n, y_t^{n+k}) \leq 3\varepsilon$$

for every n large enough, namely $(y^n)_n$ is a Cauchy sequence in $\mathcal{C}([0, T], E)$ as we claimed. $\square$

THE HALF LINE

There are two ways to deal with $\mathcal{C}([0, \infty), E)$. One is to extend the notion of modulus of continuity to a local one: for any function $x : [0, \infty) \rightarrow E$, any $\delta > 0$, and any $N \geq 1$, let

$$\omega_N(x, \delta) = \sup\{d_E(x_s, x_t) : s, t \in [0, N], |t - s| < \delta\}.$$

Then a collection $\mathcal{K}$ of functions is said to be *locally (uniformly) equi-continuous* when for every $\varepsilon > 0$ and every $N \geq 1$, there exists $\delta > 0$ such that

$$\sup_{x \in \mathcal{K}} \omega_N(x, \delta) < \varepsilon.$$

The Arzelà–Ascoli theorem 6.12 then extends to show that a collection $\mathcal{K}$ of functions in $\mathcal{C}([0, \infty), E)$ is relatively compact if and only if it is locally equi-continuous and there exists a countable and dense subset $D \subset [0, \infty)$ such that for every $t \in D$, the set $\pi_t(\mathcal{K}) = \{x_t : x \in \mathcal{K}\}$ is relatively compact in E . The proof is almost exactly the same as for compact intervals.

Another way is the following useful observation that relative compactness is a local property. For a function $x \in \mathcal{C}([0, \infty), E)$ and for $T > 0$, we let $x|_{[0, T]}$ denote the restriction of x to $[0, T]$. For a collection $\mathcal{K} \subset \mathcal{C}([0, \infty), E)$ of functions, we let $\mathcal{K}|_{[0, T]} \subset \mathcal{C}([0, T], E)$ denote the collection of their restrictions.

LEMMA 6.14. A subset $\mathcal{K} \subset \mathcal{C}([0, \infty), E)$ is relatively compact if and only if for every $N \geq 1$, the subset $\mathcal{K}|_{[0, N]} \subset \mathcal{C}([0, N], E)$ is relatively compact.

Proof. The direct implication is an immediate consequence of the continuity of $x \mapsto x|_{[0, N]}$. For the converse implication, suppose that $\mathcal{K} \subset \mathcal{C}([0, \infty), E)$ is such that $\mathcal{K}|_{[0, N]}$ is relatively compact in $\mathcal{C}([0, N], E)$ for every $N \geq 1$, fix a sequence $(x^n)_n$ in $\mathcal{K}$ and let us prove that it admits a convergent subsequence in $\mathcal{C}([0, \infty), E)$. For every $N \geq 1$, the sequence of the restrictions $(x^n|_{[0, N]})_n$ does admit a convergent subsequence in $\mathcal{C}([0, N], E)$. The limits y^N are consistent in the sense that $y_{[0, N]}^{N+k} = y^N$ for every $N, k \geq 1$, so we can define $y \in \mathcal{C}([0, \infty), E)$ such that $y|_{[0, N]} = y^N$ for every $N \geq 1$. Finally, a diagonal extraction argument shows that $(x^n)_n$ admits a subsequence whose restriction to $[0, N]$ converges to $y|_{[0, N]}$ in $\mathcal{C}([0, N], E)$ for every $N \geq 1$. This means that this subsequence converges to y in $\mathcal{C}([0, \infty), E)$. $\square$

6.5 TIGHTNESS CRITERIA

Recall from Theorem 6.8 that convergence in distribution of random variables in $\mathcal{C}(I, E)$ is equivalent to the combination of tightness and convergence in distribution of all the finite-dimensional marginals. Notice that when considering the finite dimensional distributions, say at time $0 \leq t_1 < \dots < t_k$, it is sufficient to restrict the processes to the compact interval $[0, t_k]$. Similarly, according to Lemma 6.14, proving relative compactness can be achieved by considering the restrictions to any compact interval. As a consequence, so does tightness.

COROLLARY 6.15. A collection of probability measures in $\mathcal{C}([0, \infty), E)$ is tight if and only if for every $N \geq 1$, their restrictions to the interval $[0, N]$ are tight when seen as measures on $\mathcal{C}([0, N], E)$.

Proof. If a collection $\mathcal{P}$ of probability measures in $\mathcal{C}([0, \infty), E)$ is tight, then for every $\varepsilon > 0$, there exists a compact set $\mathcal{K}$ of $\mathcal{C}([0, \infty), E)$ such that $\mu(\mathcal{K}) \geq 1 - \varepsilon$ for every $\mu \in \mathcal{P}$. Fix $N \geq 1$, then by Lemma 6.14 the restriction $\mathcal{K}|_{[0, N]}$ is relatively compact in $\mathcal{C}([0, N], E)$, and it is also closed since $\mathcal{K}$ is. Hence $\mathcal{K}|_{[0, N]}$ is compact, and if we let μ_N denote the restriction of μ to $\mathcal{C}([0, N], E)$, then $\mu_N(\mathcal{K}|_{[0, N]}) \geq \mu(\mathcal{K}) \geq 1 - \varepsilon$ for every $\mu \in \mathcal{P}$.

Conversely, suppose that for every $N \geq 1$, the restrictions $\{\mu_N : \mu \in \mathcal{P}\}$ are tight, so for every

$\varepsilon > 0$, there exists a compact set $\mathcal{K}_N$ of $\mathcal{C}([0, N], E)$ such that $\mu_N(\mathcal{K}_N) \geq 1 - \varepsilon 2^{-N}$ for every $\mu \in \mathcal{P}$. Let $\mathcal{K} = \{x \in \mathcal{C}([0, \infty), E) : x_{|[0, N]} \in \mathcal{K}_N \text{ for every } N \geq 1\}$. Then by a union bound

$$\mu(\mathcal{K}^c) \leq \sum_{N \geq 1} \mu_N(\mathcal{K}_N^c) \leq \varepsilon.$$

In addition $\mathcal{K}$ is compact by Lemma 6.14. □

For this reason, we only consider in this section the case of a compact interval, and precisely, from now on, we assume that

$$I = [0, 1] \quad \text{and} \quad \mathcal{C} = \mathcal{C}([0, 1], E),$$

in order to slightly lighten the notation, since the general case $[0, T]$ is exactly the same.

THEOREM 6.16. A collection $\mathcal{P}$ of stochastic processes in $\mathcal{C}$ is tight if and only if the following properties hold:

- (i) there exists a countable and dense subset $D \subset [0, 1]$ such that for every $t \in D$, the collection of the laws of $\{X_t : X \in \mathcal{P}\}$ is tight in E and
- (ii) for every $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\sup_{X \in \mathcal{P}} \mathbb{P}(\omega(X, \delta) > \eta) < \varepsilon.$$

If $\mathcal{P} = (X^n)_{n \geq 1}$ is a sequence, then we may replace the supremum in the second requirement by:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\omega(X^n, \delta) > \eta) < \varepsilon.$$

Finally, if E has the property that all bounded sets are relatively compact, then we may replace the first requirement by the existence of one $t \in [0, 1]$ such that $\{X_t : X \in \mathcal{P}\}$ is tight.

Proof. $\implies$: Suppose first that $\mathcal{P}$ is tight and let $\varepsilon, \eta > 0$ be arbitrary. Then first, there exists a compact set $\mathcal{K} \subset \mathcal{C}$ such that $\mathbb{P}(X \notin \mathcal{K}) < \varepsilon$ for all $X \in \mathcal{P}$. Second, by the Arzelà–Ascoli theorem 6.12 $\mathcal{K}$ is equi-continuous, so there exists $\delta > 0$ such that $\omega(x, \delta) \leq \eta$ for every $x \in \mathcal{K}$. We infer that for every $X \in \mathcal{P}$, we have

$$\mathbb{P}(\omega(X, \delta) > \eta) \leq \mathbb{P}(X \notin \mathcal{K}) < \varepsilon.$$

In addition, for every $t \in [0, 1]$, by continuity of the evaluation $\pi_t : x \mapsto x_t$, the set $\pi_t(\mathcal{K})$ is compact in E and for every $X \in \mathcal{P}$, we have $\mathbb{P}(X_t \notin \pi_t(\mathcal{K})) \leq \mathbb{P}(X \notin \mathcal{K}) < \varepsilon$, so $\{X_t : X \in \mathcal{P}\}$ is tight.

$\impliedby$: Suppose conversely that both properties hold, let $\varepsilon > 0$ be arbitrary and let us find a compact set $\mathcal{K} \subset \mathcal{C}$ such that $\mathbb{P}(X \notin \mathcal{K}) < \varepsilon$ for every $X \in \mathcal{P}$. Let $(t_n)_{n \geq 1}$ be an enumeration of D . By the first property, for every $n \geq 1$, there exists a compact set $K_n \subset E$ such that $\mathbb{P}(X_{t_n} \notin K_n) < \varepsilon/2^{n+1}$ for every $X \in \mathcal{P}$. By the second property, for every $n \geq 1$, there is also $\delta_n > 0$ such that $\mathbb{P}(\omega(X, \delta_n) > 2^{-n}) < \varepsilon 2^{-(n+1)}$ for every $X \in \mathcal{P}$. Define then

$$\mathcal{K} = \{x \in \mathcal{C} : x_{t_n} \in K_n \text{ and } \omega(x, \delta_n) \leq 2^{-n} \text{ for every } n \geq 1\}.$$

Observe that it is closed: if $x \notin \mathcal{K}$, then there exists $n \geq 1$ with either $x_{t_n} \notin K_n$ or $\omega(x, \delta_n) > 2^{-n}$, but in any case, any $y \in \mathcal{C}$ that is close enough to x will also satisfy this so $\mathcal{K}^c$ is open. Also, by a union bound, for every $X \in \mathcal{P}$, we have

$$\mathbb{P}(X \notin \mathcal{K}) \leq \sum_{n \geq 1} \left(\frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2^{n+1}} \right) = \varepsilon.$$

Notice that $\mathcal{K}$ is equi-continuous. For every $n \geq 1$, the set $\pi_{t_n}(\mathcal{K}) = \{x_{t_n} : x \in \mathcal{K}\}$ is relatively compact in E , and the set $D = \{t_n : n \geq 1\}$ is dense in $[0, 1]$. The Arzelà–Ascoli theorem 6.12 then shows that $\mathcal{K}$ is relatively compact in $\mathcal{C}$, and thus compact since it is closed.

In the case of a sequence, the bound for the lim sup means that it holds uniformly over $n \geq n_0$, for some n_0 (that depends on $\varepsilon, \eta, \delta$). Since ω decreases as δ decreases, then this bound remains true for any smaller δ . We can then decrease δ if necessary in such a way that the bound also holds for all the finitely many $n < n_0$.

The last claimed is proved just as Corollary 6.13 and is left for reader. $\square$

This theorem shows that in order to prove tightness, one can consider the modulus of continuity $\omega(x, \delta)$, which seems easier to control than considering general compact spaces of $\mathcal{C}$. It is still a complicated object since it depends on the entire trajectory $(x_t)_{t \in [0, 1]}$. One would prefer a criterion that can be checked using only the finite-dimensional marginals for example.

SIMPLER CRITERIA

From now on, we focus on sequences of random processes rather than general collections since this really is what we are interested in; the reader should be able to adapt the statements and proofs to the general case. Here is a first simpler sufficient condition to ensure tightness.

PROPOSITION 6.17. A sequence $(X^n)_{n \geq 1}$ of stochastic processes in $\mathcal{C}$ is tight as soon as both:

- (i) there exists a countable and dense subset $D \subset [0, 1]$ such that for every given $t \in D$, the sequence $(X_t^n)_{n \geq 1}$ is tight in E and
- (ii) for every $\varepsilon, \eta > 0$, there exists $k \geq 1$ such that

$$\limsup_{n \rightarrow \infty} k \max_{0 \leq i \leq k-1} \mathbb{P} \left(\sup_{s \in [0, 1/k]} d_E(X_{i/k}^n, X_{s+i/k}^n) > \eta \right) < \varepsilon.$$

Also, if E has the property that all bounded sets are relatively compact, then we may replace the first requirement by the existence of a single $t \in [0, 1]$ such that $(X_t^n)_{n \geq 1}$ is tight.

Proof. The key is to notice that for every $x \in \mathcal{C}$ and $k \geq 1$, we have

$$\omega(x, 1/k) \leq 3 \max_{0 \leq i \leq k-1} \sup_{s \in [0, 1/k]} d_E(x_{i/k}, x_{s+i/k}).$$

Indeed, the modulus of continuity on the left only involves times s, t which are less than $1/k$ apart, so either they belong to the same interval $[i/k, (i+1)/k]$ for some $0 \leq i \leq k-1$ and then

$$\begin{aligned} d_E(x_s, x_t) &\leq d_E(x_s, x_{i/k}) + d_E(x_t, x_{i/k}) \\ &\leq 2 \sup_{r \in [0, 1/k]} d_E(x_{i/k}, x_{r+i/k}), \end{aligned}$$

or one of them (say s) belongs to some interval $[i/k, (i+1)/k]$ and the other one (say t) in the next one $[(i+1)/k, (i+2)/k]$, and then

$$\begin{aligned} d_E(x_s, x_t) &\leq d_E(x_s, x_{i/k}) + d_E(x_t, x_{(i+1)/k}) + d_E(x_{i/k}, x_{(i+1)/k}) \\ &\leq 3 \max_{0 \leq i \leq k-1} \sup_{r \in [0, 1/k]} d_E(x_{i/k}, x_{r+i/k}). \end{aligned}$$

We then infer from a union bound that

$$\begin{aligned} \mathbb{P}(\omega(X^n, 1/k) > 3\eta) &\leq \mathbb{P}\left(\max_{0 \leq i \leq k-1} \sup_{s \in [0, 1/k]} d_E(X_{i/k}^n, X_{s+i/k}^n) > \eta\right) \\ &\leq k \max_{0 \leq i \leq k-1} \mathbb{P}\left(\sup_{s \in [0, 1/k]} d_E(X_{i/k}^n, X_{s+i/k}^n) > \eta\right), \end{aligned}$$

and we conclude from Theorem 6.16. $\square$

The next result is of high importance since it allows to prove tightness of a process by only studying the marginal laws of dimension 2! The conditions are strong and may fail, but it is often a good idea to try this first.

THEOREM 6.18 (Kolmogorov). Let $(X^n)_{n \geq 1}$ be stochastic processes in $\mathcal{C}$ and assume that:

- (i) There exists a countable and dense subset $D \subset [0, 1]$ such that for every $t \in D$, the sequence $(X_t^n)_{n \geq 1}$ is tight in E ; if E has the property that all bounded sets are relatively compact, then only assume the existence of one $t \in [0, \infty)$ such that $(X_t^n)_{n \geq 1}$ is tight.
- (ii) There exist constants $\beta, p, C > 0$ such that for every $n \geq 1$, for every $s, t \in [0, 1]$, we have

$$\mathbb{E}[d_E(X_s^n, X_t^n)^p] \leq C |t - s|^{1+\beta}.$$

Then $(X^n)_{n \geq 1}$ is tight and, even more, for every $\alpha \in (0, \beta/p)$, for every $\varepsilon > 0$, there exists $K > 0$ such that for every $n \geq 1$,

$$\mathbb{P}\left(\sup_{\substack{s, t \in [0, 1] \\ s \neq t}} \frac{d_E(X_s^n, X_t^n)}{|t - s|^\alpha} \geq K\right) \leq \varepsilon.$$

REMARK 6.19. The quantity

$$H^\alpha(x) = \sup_{\substack{s, t \in [0, 1] \\ s \neq t}} \frac{d_E(x_s, x_t)}{|t - s|^\alpha} = \sup_{\delta > 0} \frac{\omega(x, \delta)}{\delta^\alpha}$$

is called the α -Hölder norm of x on $[0, 1]$. It is bounded whenever x is α -Hölder continuous on this interval.

The idea of the proof is to cut $[0, 1]$ at the dyadic numbers and control the process at these numbers. For every $k \geq 0$, we let $Q_k = \{i/2^k : 0 \leq i \leq 2^k\}$ denote the dyadic numbers of depth k and $Q = \bigcup_k Q_k$. We will use the following lemma.

LEMMA 6.20. Let $x : [0, 1] \rightarrow E$ be any function and assume that there exist $K, \alpha > 0$ such that for every $k \geq 0$ and every $0 \leq i \leq 2^k - 1$, it holds that

$$d_E(x_{i2^{-k}}, x_{(i+1)2^{-k}}) \leq K 2^{-k\alpha}.$$

Then for any $s, t \in Q$, we have

$$d_E(x_s, x_t) \leq \frac{2K}{1 - 2^{-\alpha}} |t - s|^\alpha.$$

If $x \in \mathcal{C}$, then

$$H_1^\alpha(x) \leq \frac{2K}{1 - 2^{-\alpha}}.$$

Let differ the proof of this lemma and first deduce Kolmogorov's theorem from it.

Proof of Theorem 6.18. By a union bound and the Markov inequality,

$$\begin{aligned} & \mathbb{P}\left(\max_{0 \leq i \leq 2^k-1} d_E(X_{i2^{-k}}^n, X_{(i+1)2^{-k}}^n) > K2^{-k\alpha}\right) \\ & \leq 2^k \max_{0 \leq i \leq 2^k-1} \mathbb{P}(d_E(X_{i2^{-k}}^n, X_{(i+1)2^{-k}}^n) > K2^{-k\alpha}) \\ & \leq 2^k \max_{0 \leq i \leq 2^k-1} \frac{\mathbb{E}[d_E(X_{i2^{-k}}^n, X_{(i+1)2^{-k}}^n)^p]}{K^p 2^{-kp\alpha}} \\ & \leq 2^k \frac{C 2^{-k(1+\beta)}}{K^p 2^{-kp\alpha}} = CK^{-p} 2^{-k(\beta-p\alpha)}, \end{aligned}$$

from our assumption. Taking $\alpha \in (0, \beta/p)$, the upper bound has a finite sum over k . We infer from the lemma that, uniformly over $n \geq 1$,

$$\begin{aligned} \mathbb{P}\left(H_1^\alpha(X^n) > \frac{2K}{1-2^{-\alpha}}\right) & \leq \mathbb{P}\left(\bigcup_{k \geq 0} \left\{ \max_{0 \leq i \leq 2^k-1} d_E(X_{i2^{-k}}^n, X_{(i+1)2^{-k}}^n) > K2^{-k\alpha} \right\}\right) \\ & \leq CK^{-p} \sum_{k \geq 0} 2^{-k(\beta-p\alpha)}, \end{aligned}$$

and the sum is finite, so the right-hand side converges to 0 as $K \rightarrow \infty$, and this proves the last identity in the claim. Tightness easily follows since for any $x \in \mathbb{C}$ and every $\delta > 0$, we have $\omega(x, \delta) \leq H_1^\alpha(x)\delta^\alpha$. $\square$

ANOTHER RESULT OF KOLMOGOROV (★)

Recall that Theorem 3.10 ensures the existence of a stochastic process in E^I given its finite-dimensional distributions. However it does not ensure that such a process has continuous paths, and continuity is not a measurable property. The following result, whose statement and proof are very close to the previous tightness theorem, shows that, under some moment assumption, one can find a continuous-path process with the same law as the original process.

THEOREM 6.21 (Kolmogorov–Chentsov). Let $X = (X_t)_{t \in [0,1]}$ be a stochastic process, not necessarily continuous, and assume that there exist constants $\beta, p, C > 0$ such that for every $s, t \in [0, 1]$, we have

$$\mathbb{E}[d_E(X_s, X_t)^p] \leq C |t - s|^{1+\beta}.$$

Then there exists a process $X' = (X'_t)_{t \in [0,1]}$ whose sample paths $t \mapsto X'_t$ are α -Hölder continuous for every $\alpha < \beta/p$ and such that $\mathbb{P}(X_t = X'_t) = 1$ for every $t \in [0, 1]$.

A process Y such that $\mathbb{P}(Y_t = X_t) = 1$ for every t is called a *modification*, or a *version* of X . Two such processes have the same finite-dimensional distributions, and hence the same law. Since X is regarded as a random variable in the product space $E^{[0,1]}$ and continuity is not a measurable property, then this modification is the correct way to proceed: we replace X by X' which can alternatively be regarded as a random variable in the polish space $\mathbb{C}$.

If X is defined on $[0, \infty)$ in the first place, then we may apply this theorem on each interval $[n-1, n]$, with $n \geq 1$, to deduce the existence of a modification X' that is locally Hölder continuous (and thus continuous everywhere).

Proof. Just as in the preceding proof, we deduce from the assumption that for $\alpha \in (0, \beta/p)$,

$$\sum_{k \geq 0} \mathbb{P}\left(\max_{0 \leq i \leq 2^k-1} d_E(X_{i2^{-k}}^n, X_{(i+1)2^{-k}}^n) > 2^{-k\alpha}\right) \leq C \sum_{k \geq 0} 2^{-k(\beta-p\alpha)} < \infty$$

By the Borel–Cantelli lemma, almost surely, the events on the left eventually all fail: on an event A of

probability 1, there exists $k_0 \geq 1$ such that for every $k \geq k_0$, we have

$$\max_{0 \leq i \leq 2^k - 1} \frac{d_E(X_{i2^{-k}}^n, X_{(i+1)2^{-k}}^n)}{2^{-k\alpha}} \leq 1.$$

Adding the finitely many terms corresponding to $0 \leq k < k_0$, we deduce that the left-hand side is bounded by some (random) value $K < \infty$ uniformly over all $k \geq 0$. Lemma 6.20 allows to conclude that on the event A with probability 1,

$$d_E(X_s, X_t) \leq \frac{2K}{1 - 2^{-\alpha}} |t - s|^\alpha,$$

for every dyadic numbers $s, t \in Q$. Still on this event, we may extend uniquely $(X_t)_{t \in Q}$ continuously to a process $(X'_t)_{t \in [0,1]}$, which is an α -Hölder continuous function. On the event A^c , we define X' as a deterministic arbitrary value in E , so it is also an α -Hölder continuous function on $[0, 1]$.

It remains to prove that $\mathbb{P}(X'_t = X_t) = 1$ for every fixed $t \in [0, 1]$. This is plain by construction if $t \in Q$ is a dyadic number; otherwise, let $(t_n)_n$ be a sequence of dyadic numbers with $t_n \rightarrow t$. From the assumption we have

$$\mathbb{E}[d_E(X_{t_n}, X_t)^p] \leq C |t - t_n|^{1+\beta}.$$

Also $X_{t_n} \rightarrow X'_t$ almost surely since X' is continuous, so by Fatou's lemma,

$$\mathbb{E}[d_E(X'_t, X_t)^p] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[d_E(X_{t_n}, X_t)^p] = 0,$$

and so $X'_t = X_t$ almost surely. □

REMARK 6.22. Recall that tightness implies relative compactness by Prohorov's theorem 1.25. Under the assumption of Theorem 6.18, let X be the limit in distribution of a subsequence $(X^{n_k})_k$ of $(X^n)_n$. We claim that X falls into the framework of Theorem 6.21 so this process admits a locally α -Hölder continuous version.

Indeed, using Skorokhod's representation theorem 5.5 to replace the convergence in distribution by an almost sure one, and applying Fatou's lemma, we deduce that for every $s, t \in [0, 1]$, we have

$$\mathbb{E}[d_E(X_s, X_t)^p] \leq \liminf_{k \rightarrow \infty} \mathbb{E}[d_E(X_s^{n_k}, X_t^{n_k})^p] \leq C |t - s|^{1+\beta}.$$

Alternatively, one can use the identity $\mathbb{E}[d_E(X_s, X_t)^p] = \int_0^\infty \mathbb{P}(d_E(X_s, X_t)^p > r) dr$ by Fubini's theorem, and then that the convergence in distribution implies by the Portmanteau theorem 1.10 that $\mathbb{P}(d_E(X_s, X_t)^p > r) \leq \liminf_n \mathbb{P}(d_E(X_s^{n_k}, X_t^{n_k})^p > r)$, and finally apply Fatou's lemma for the integrals.

6.6 EXERCISES

EXERCISE 28. Prove that the space of continuous and bounded functions from $[0, \infty)$ to $[0, 1]$ equipped with the uniform norm is not separable. *Hint:* you can construct an uncountable set of such functions that all lie at distance 1 from each other.

EXERCISE 29. Let E be a metric space. For every $n \geq 1$, let $X^n = (X_t^n)_{t \in I}$ and $Y^n = (Y_t^n)_{t \in I}$ be two random variables in $C([0, 1], E)$. Assume that the collection of the laws of $(X^n)_{n \geq 1}$ is tight in $C([0, 1], E)$, and so is that of the laws of $(Y^n)_{n \geq 1}$. We wonder whether the collection of the laws of the pair (X^n, Y^n) is tight, in the following spaces:

- (i) when the pair (X^n, Y^n) is viewed as a random variable in the space $(C([0, 1], E)^2, \mathcal{B}(C([0, 1], E)^2))$,

- (ii) when the pair (X^n, Y^n) is viewed as a random variable in the space $C([0, 1], E^2)$, that is, when we see it as a single process $t \mapsto (X_t^n, Y_t^n) \in E^2$,
- (iii) when the pair (X^n, Y^n) is viewed as a random variable in the space $C([0, 1]^2, E^2)$, that is, when we see it as a process of two variables $(s, t) \mapsto (X_s^n, Y_t^n) \in E^2$.

EXERCISE 30. Let $C = C([0, 1], \mathbb{R})$, equipped with its Borel σ -algebra $\mathcal{B}(C)$.

- (i) Prove that the subset of nondecreasing continuous functions is measurable.
- (ii) Let X^n, X be random variables in $C([0, 1], \mathbb{R})$ which are almost surely nondecreasing; prove that $X^n \rightarrow X$ in distribution in C if and only if $X^n \rightarrow X$ in distribution in the sense of the finite-dimensional marginal laws.
- (iii) What if X^n is the linear interpolation of a nondecreasing random walk, with i.i.d. nonnegative increments, and rescaled by some factor $a_n \rightarrow \infty$?

EXERCISE 31. Let E be separable and complete metric space and for every $n \geq 1$, let X^n be a random variable in $C = C([0, 1], E)$. Suppose that X^n converges in distribution in C to some X . Let $f : E \rightarrow \mathbb{R}$ be a continuous function and define real-valued random variables:

$$Y_n = \int_0^1 f(X_t^n) dt \quad \text{and} \quad Y = \int_0^1 f(X_t) dt.$$

Prove that the pair $(X^n, Y_n) \rightarrow (X, Y)$ in distribution in $C \times \mathbb{R}$.

EXERCISE 32. Let E be a separable and complete metric space and let $C = C([0, \infty), E)$. Consider for every $n \geq 1$ a continuous process X^n in C , which converges in distribution to some X in C . Assume that for each $n \geq 1$, the process X^n is a martingale, namely $X_t^n \in L^1$ for each $t \geq 0$ and

$$\mathbb{E}[X_{t+s}^n | \mathcal{F}_t^n] = X_t^n$$

for every $s, t \geq 0$, where $\mathcal{F}^n = (\mathcal{F}_t^n)_{t \geq 0}$ is some filtration. Assume in addition that for every $t \geq 0$, the sequence $(X_t^n)_{n \geq 1}$ is uniformly integrable, namely:

$$\lim_{K \rightarrow \infty} \sup_{n \geq 1} \mathbb{E}[X_t^n \mathbb{1}_{|X_t^n| > K}] = 0.$$

Prove that X is also a martingale, at least for the natural filtration $\mathcal{F}_t = \sigma(X_s, s \leq t)$.

CHAPTER 7

RANDOM WALKS & BROWNIAN MOTION

This short chapter is dedicated to a fundamental result in probability. We shall apply the results from the previous chapter to the case of random functions obtained by (linearly interpolating and) rescaling random walks with finite variance to make sense as a limiting object of the celebrated Brownian motion, thus extending the central limit theorem in a functional version, without any extra assumption.

CONTENTS

7.1	Donsker's theorem & Brownian motion	77
7.2	Finite-dimensional marginals	78
7.3	Tightness	79
7.4	Some applications	80
7.5	Exercises	82

7.1 DONSKER'S THEOREM & BROWNIAN MOTION

We consider here random continuous functions with values in $E = \mathbb{R}^d$ for some $d \geq 1$. Specifically, let $(\xi_n)_{n \geq 1}$ be i.i.d. random variables in $\mathbb{R}^d$ that have

$$\mathbb{E}[\xi_1] = 0 \quad \text{and} \quad \text{Cov}(\xi_1) = I_d,$$

where I_d is the identity matrix of dimension d . We let $S_0 = 0$ and for every $n \geq 1$, we let $S_n = \xi_1 + \dots + \xi_n$.

The central limit theorem in $\mathbb{R}^d$ shows that

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}(0, I_d),$$

where the limit is the law of a Gaussian vector with mean 0 and covariance matrix I_d , i.e. a collection of d i.i.d. standard Gaussian random variables. Now we want to consider the entire path $(S_k)_{0 \leq k \leq n}$. We thus extend this sequence to a continuous function $(S_t)_{t \in [0, \infty)}$ by linear interpolation between integer times; an explicit formula is:

$$S_t = \sum_{i=1}^{\lfloor t \rfloor} \xi_i + (t - \lfloor t \rfloor) \xi_{\lfloor t \rfloor + 1} \quad t \geq 0.$$

Finally, let us rescale this function in time and space as follows:

$$S_t^n = \frac{S_{nt}}{\sqrt{n}}, \quad n \geq 1, t \geq 0.$$

Therefore each S^n belongs to $\mathcal{C}([0, \infty), \mathbb{R}^d)$.

Let us introduce the limit process before stating the convergence.

DEFINITION 7.1. The (standard) *Brownian motion* in $\mathbb{R}^d$ is the continuous stochastic process $B = (B_t)_{t \in [0, \infty)}$ whose law is characterised by the fact that $B_0 = 0$ and for any $k \geq 1$ and any $0 = t_0 < t_1 < \dots < t_k$, the increments $B_{t_i} - B_{t_{i-1}}$ for $1 \leq i \leq k$ are independent Gaussian vectors with mean 0 and covariance matrix $(t_i - t_{i-1})I_d$ respectively.

One can note that the law of the increments characterises the law of the finite-dimensional marginals of B , which indeed characterises uniquely the law of the entire process by Proposition 6.6. Hence there exists at most one such process (in law). However, existence is not easy. The Kolmogorov extension theorem 3.10 shows that there exists a stochastic process in the product space $\mathbb{R}^{[0, \infty)}$ that has these marginals, and the Kolmogorov continuity theorem 6.21 further shows the existence of a continuous-path modification. This approach is used in the stochastic calculus course; here, we provide with the next theorem an alternative and independent proof of existence in a single step, and adds the important fact that this process appears as the limit of random walks with a finite variance.

THEOREM 7.2 (Donsker). Brownian motion B exists and, under the above assumptions, we have the convergence in distribution in $\mathcal{C}([0, \infty), \mathbb{R}^d)$:

$$S^n \xrightarrow[n \rightarrow \infty]{(d)} B.$$

Specifically, we shall prove that $(S^n)_{n \geq 1}$ is tight in $\mathcal{C}([0, \infty), \mathbb{R}^d)$ and converges in the sense of finite-dimensional marginal distributions to the marginal laws of B (which do exist in $\mathbb{R}^k$!). Then S^n must converge in distribution in $\mathcal{C}([0, \infty), \mathbb{R}^d)$ and its limit must have these marginal laws.

7.2 FINITE-DIMENSIONAL MARGINALS

Let us prove that S^n converges in the sense of finite-dimensional marginal distributions. This is the easy part of the theorem. Fix $k \geq 1$ and $0 = t_0 < t_1 < \dots < t_k$ and let $(G_i)_{1 \leq i \leq k}$ be independent Gaussian vectors with mean 0 and covariance matrix $(t_i - t_{i-1})I_d$ respectively. We claim that

$$(S^n_{t_i} - S^n_{t_{i-1}})_{1 \leq i \leq k} \xrightarrow[n \rightarrow \infty]{(d)} (G_1 + \dots + G_i)_{1 \leq i \leq k} \quad (7.1)$$

in $\mathbb{R}^k$. Since the ξ_k 's are i.i.d. then the vectors

$$S_{[nt_i]} - S_{[nt_{i-1}]} = \sum_{k=[nt_{i-1}]+1}^{[nt_i]} \xi_k$$

are independent as i varies, and have the same law respectively as

$$S_{[nt_i]-[nt_{i-1}]} = \sum_{k=1}^{[nt_i]-[nt_{i-1}]} \xi_k.$$

The standard central limit theorem in $\mathbb{R}^d$ asserts that for each $1 \leq i \leq k$, if N_i is a standard Gaussian vector, then

$$\frac{S_{[nt_i]} - S_{[nt_{i-1}]}}{\sqrt{[nt_i] - [nt_{i-1}]}} \xrightarrow[n \rightarrow \infty]{(d)} N_i.$$

Since the random variables on the left are independent, then the convergence in distribution holds jointly for all of them, with independent limits. Since $n^{-1}([nt_i] - [nt_{i-1}]) \rightarrow t_i - t_{i-1}$, and since G_i has the same law as $\sqrt{t_i - t_{i-1}} \cdot N_i$, then

$$\left(\frac{S_{[nt_i]} - S_{[nt_{i-1}]}}{\sqrt{n}} \right)_{1 \leq i \leq k} \xrightarrow[n \rightarrow \infty]{(d)} (G_i)_{1 \leq i \leq k}.$$

By continuity of the sum, we further infer that

$$\left(\frac{S_{\lfloor nt_i \rfloor}}{\sqrt{n}} \right)_{1 \leq i \leq k} \xrightarrow[n \rightarrow \infty]{(d)} (G_1 + \dots + G_i)_{1 \leq i \leq k}.$$

To conclude, we want to remove the integer parts. For this, observe that for any $t > 0$,

$$\frac{|S_{nt} - S_{\lfloor nt \rfloor}|}{\sqrt{n}} \leq \frac{|\xi_{\lfloor nt \rfloor + 1}|}{\sqrt{n}},$$

which converges in L^2 to 0 since the ξ_k 's have the same law, which is square integrable. Consequently, the vector $((S_{nt_i} - S_{\lfloor nt_i \rfloor})/\sqrt{n})_{1 \leq i \leq k}$ converges in probability to 0 and the convergence in distribution

$$\left(\frac{S_{nt_i}}{\sqrt{n}} \right)_{1 \leq i \leq k} \xrightarrow[n \rightarrow \infty]{(d)} (G_1 + \dots + G_i)_{1 \leq i \leq k},$$

follow from Slutsky's theorem.

7.3 TIGHTNESS

Let us now prove that the sequence $(S^n)_{n \geq 1}$ is tight. Recall from Corollary 6.15 that it is sufficient to work on compact intervals and prove that the sequence of processes $(S_t^n)_{t \in [0, N]}$ for $n \geq 1$ is tight for each fixed $N \geq 1$. Let us prove it for $N = 1$ to simplify the notation, as the general case is similar. Since $S_0^n = 0$ for every $n \geq 1$, then this sequence is tight in $\mathbb{R}$, so by Proposition 6.17 it suffices to prove that for every $\varepsilon, \eta > 0$, there exists $k \geq 1$ such that

$$\limsup_{n \rightarrow \infty} k \max_{0 \leq i \leq k-1} \mathbb{P} \left(\sup_{s \in [0, n/k]} |S_{s+in/k} - S_{in/k}| > \eta \sqrt{n} \right) < \varepsilon.$$

As before, it is more comfortable to work with integer times. For this, notice that, since the value S_r is a linear interpolation between $S_{\lfloor r \rfloor}$ and $S_{\lfloor r \rfloor + 1}$, then for every $s \in [0, n/k]$,

$$\begin{aligned} |S_{s+in/k} - S_{in/k}| &\leq |S_{s+in/k} - S_{\lfloor in/k \rfloor}| + |S_{in/k} - S_{\lfloor in/k \rfloor}| \\ &\leq \max\{|S_{\lfloor s+in/k \rfloor} - S_{\lfloor in/k \rfloor}\}, |S_{\lfloor s+in/k \rfloor + 1} - S_{\lfloor in/k \rfloor}|\} + |S_{\lfloor in/k \rfloor + 1} - S_{\lfloor in/k \rfloor}|. \end{aligned}$$

Since $\lfloor s + in/k \rfloor + 1 \leq \lfloor (i+1)n/k \rfloor + 1 \leq \lfloor in/k \rfloor + \lfloor n/k \rfloor + 2$, then

$$\sup_{s \in [0, n/k]} |S_{s+in/k} - S_{in/k}| \leq 2 \max_{0 \leq j \leq \lfloor n/k \rfloor + 2} |S_{j+\lfloor in/k \rfloor} - S_{\lfloor in/k \rfloor}|,$$

and the right-hand side has the same law as $2 \max_{0 \leq j \leq \lfloor n/k \rfloor + 2} |S_j|$ for every i .

Fix an integer $M \geq 1$ and a real number $K > 0$. Let $\tau = \inf\{j \geq 1 : |S_j| > 2K\}$. Since $S_M - S_\ell$ is independent of $\{\tau = \ell\}$ and has the same law as $S_{M-\ell}$, then

$$\begin{aligned} \mathbb{P} \left(\max_{0 \leq j \leq M} |S_j| > 2K \right) &\leq \mathbb{P}(|S_M| > K) + \sum_{\ell=1}^{M-1} \mathbb{P}(|S_M| \leq K \text{ and } \tau = \ell) \\ &\leq \mathbb{P}(|S_M| > K) + \sum_{\ell=1}^{M-1} \mathbb{P}(|S_M - S_\ell| > K \text{ and } \tau = \ell) \\ &= \mathbb{P}(|S_M| > K) + \sum_{\ell=1}^{M-1} \mathbb{P}(|S_{M-\ell}| > K) \mathbb{P}(\tau = \ell) \\ &\leq \max_{0 \leq j \leq M} \mathbb{P}(|S_j| > K) + \mathbb{P}(\tau \leq M-1) \max_{0 \leq j \leq M} \mathbb{P}(|S_j| > K) \\ &\leq 2 \max_{0 \leq j \leq M} \mathbb{P}(|S_j| > K). \end{aligned}$$

We infer that for every $\eta > 0$ and every $n, k \geq 1$, we have

$$\max_{0 \leq i \leq k-1} \mathbb{P} \left(\sup_{s \in [0, n/k]} |S_{s+in/k} - S_{in/k}| > \eta \sqrt{n} \right) \leq 2 \max_{0 \leq j \leq \lfloor n/k \rfloor + 2} \mathbb{P}(|S_j| > \eta \sqrt{n}/4).$$

Since $\mathbb{E}[|S_j|^2] = jd$, then the Markov inequality yields for each $J \geq 1$:

$$\max_{1 \leq j \leq J} \mathbb{P}(|S_j| > \eta \sqrt{n}/4) \leq \max_{1 \leq j \leq J} \frac{16j}{\eta^2 n} = \frac{16J}{\eta^2 n}.$$

In order to apply Proposition 6.17, we want this quantity, once multiplied by k , to be arbitrarily small for n large enough, which is true for any $J \ll n$, but not for $J = \lfloor n/k \rfloor + 2$. In order to deal with the large values of j , notice that for $3 \leq j \leq \lfloor n/k \rfloor + 2$, we have $n \geq k(j-2) \geq kj/3$. Since $j^{-1/2}S_j$ converges to a Gaussian vector G , with zero mean and identity covariance matrix (and since the law of $|G|$ has a density), then letting $j \rightarrow \infty$ by keeping η, k fixed, we have

$$\mathbb{P}(|S_j| > \eta \sqrt{n}/4) \leq \mathbb{P}(|S_j| > \eta \sqrt{kj}/(4\sqrt{3})) \xrightarrow{j \rightarrow \infty} \mathbb{P}(|G| > \eta \sqrt{k}/(4\sqrt{3})).$$

The Markov inequality applied to $|G|^4$, whose expectation is some constant depending only on d , then shows that there exists a constant C and a threshold $J \geq 3$, which depends on η and k , such that, for every n large enough,

$$\max_{J \leq j \leq \lfloor n/k \rfloor + 2} \mathbb{P}(|S_j| > \eta \sqrt{n}/4) \leq \frac{C}{\eta^4 k^2},$$

which is larger than the bound for $j \leq J$ when n is large enough so we conclude that for every $\eta > 0$ and $k \geq 1$, there exists $J \geq 3$ such that for every n large enough,

$$\begin{aligned} k \max_{0 \leq i \leq k-1} \mathbb{P} \left(\sup_{s \in [0, n/k]} |S_{s+in/k} - S_{in/k}| > \eta \sqrt{n} \right) &\leq 2k \max_{0 \leq j \leq \lfloor n/k \rfloor + 2} \mathbb{P}(|S_j| > \eta \sqrt{n}/4) \\ &\leq 2k \max \left\{ \frac{16J}{\eta^2 n}, \frac{C}{\eta^4 k^2} \right\} \\ &\leq \frac{2C}{\eta^4 k}. \end{aligned}$$

The value of $\eta > 0$ being fixed, this upper bound can be made arbitrarily small by taking k large enough. This implies tightness by Proposition 6.17.

7.4 SOME APPLICATIONS

Let us discuss some consequences of Donsker's theorem, here in dimension 1. We assume some familiarities with Brownian motion and refer to the stochastic calculus course for details.

FROM CONTINUOUS TO DISCRETE. This theorem allows to derive asymptotic properties about discrete random walks from the study of Brownian motion. For example, an application of the strong Markov property for Brownian motion shows that almost surely, its maximum over two disjoint compact intervals are different. If $(S_n)_n$ is any random walk whose increments have zero mean and finite variance, then we infer that with probability tending to 1, its maximum over two disjoint discrete time intervals with length of order n cannot be $o(\sqrt{n})$.

FROM DISCRETE TO CONTINUOUS. Donsker's theorem also permits to derive some properties about Brownian motion by studying a particular model of discrete random walk with finite variance, such as the simple random walk, with increments ± 1 with probability $1/2$ each. The exercises below contain two examples that we state here.

THEOREM 7.3 (Arcsine laws). Consider the following four quantities about Brownian motion up to time 1:

$$Z = \sup\{t \in [0, 1] : B_t = 0\}, \quad P = \text{Leb}(\{t \in [0, 1] : B_t > 0\}),$$

where Leb denotes the Lebesgue measure, as well as

$$I = \inf\{t \in [0, 1] : B_t = \inf_{s \in [0, 1]} B_s\}, \quad \text{and} \quad M = \sup\{t \in [0, 1] : B_t = \sup_{s \in [0, 1]} B_s\}.$$

These four random variables all have the same law, which is the one on $[0, 1]$ with distribution function and density respectively given by:

$$\mathbb{P}(\cdot \leq t) = \frac{2}{\pi} \arcsin \sqrt{t}, \quad \text{or equivalently,} \quad \mathbb{P}(\cdot \in dt) = \frac{1}{\pi \sqrt{t(1-t)}} dt.$$

This common distribution is called the *arcsine law*. It is the law of $(\sin \Theta)^2$ where Θ is an angle uniformly distributed over $[0, 2\pi]$.

Although this theorem can be proved directly in the continuous setting (see the stochastic calculus course), Exercise 38 below proves it by approximation using the simple random walk, for which one can express explicitly the law of the analogous quantities.

REMARK 7.4 (From discrete to continuous ... to discrete again). By combining approximations, Donsker's theorem allows to deduce asymptotic properties about any random walk with finite variance from an exact property about a single model. For example, once one expresses explicitly the law of one the analogue of the random variables in the previous theorem for the simple random walk and passes to the limit to prove the theorem, then one deduces that the same limit holds for any random walk with finite variance, even if there is no hope to express exactly the law of the corresponding quantity for a general random walk for a fixed n .

The arcsine law contains curious information. Observe indeed that the density is maximal near the edges 0 and 1 of the interval, and minimal at the midpoint $1/2$. When considering the quantity P for example, this means that for Brownian motion (and then for any random walk with zero mean and finite variance), it is much, much, more likely to spend either almost the entire time on the positive side or the negative side, than to spend close to half the time on each side. Similarly for Z , it is more likely to reach a side, either positive or negative, early and never change after that, or to change side almost at the end, than to last change the side near the middle of the interval. Let us refer to Feller, Volume I, Chapter III, for discussion and tables on this.

A second property of Brownian motion that can be proved directly in the continuous (once again, in the stochastic calculus course) is yet another theorem attributed to Lévy.

THEOREM 7.5 (Lévy's identity). If $B = (B_t)_{t \in [0, \infty)}$ is the standard Brownian motion, then the processes

$$(B_t - \inf_{s \in [0, t]} B_s)_{t \in [0, \infty)} \quad \text{and} \quad (|B_t|)_{t \in [0, \infty)}$$

have the same law.

Exercise 39 below provides a proof based again on the approximation of Brownian motion by the simple random walk (which is not straightforward since the analogous identity fails for the simple random walk).

As concluding remarks, first note that the arcsine law for Z and for I are equivalent to each other by Lévy's identity. Second, a theorem proved by Pitman¹ by relying on discrete approximations with random

¹One-dimensional Brownian motion and the three-dimensional Bessel process. In *Advances in Applied Probability*, 1975.

walks states that the processes

$$(B_t - 2 \inf_{s \in [0, t]} B_s)_{t \in [0, \infty)} \quad \text{and} \quad (|B_t^{(3)}|)_{t \in [0, \infty)}$$

have the same law, where $B^{(3)}$ is a three-dimensional Brownian motion. Its norm is known as the *Bessel process with dimension 3*. It turns out that this Bessel process can be seen (in ways explained in Pitman's paper) as Brownian motion conditioned to stay positive forever.

7.5 EXERCISES

Throughout this section, we focus on the dimension 1. We let $B = (B_t)_{t \geq 0}$ denote a standard Brownian motion. We use the notation $S_n = \xi_1 + \dots + \xi_n$, with also $S_0 = 0$, where the ξ_i 's are i.i.d. The sequence $(S_n)_{n \geq 0}$ is always extended to $(S_t)_{t \in [0, \infty)}$ by linear interpolation. The “simple random walk” corresponds to the case where the ξ_i 's are equal to ± 1 with probability $1/2$.

EXERCISE 33 (Donsker's theorem with higher moments). Suppose that $\mathbb{E}[\xi^4] < \infty$ and $\mathbb{E}[\xi_1] = 0$ and $\text{Var}(\xi_1) = \sigma^2$. Defined the (linearly interpolated and) rescaled processes:

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n\sigma^2}}.$$

- (i) Prove that there exists a constant $c > 0$ such that $\mathbb{E}[|S_{ns} - S_{nt}|^4] \leq cn^2(t - s)^2$ for every $n \geq 1$ and every $s, t \geq 0$.
- (ii) Simplify the proof of the convergence in distribution of $S^{(n)}$ to the standard Brownian motion.

EXERCISE 34 (Skorokhod's embedding). This exercise proposes completely different proof of the convergence of the simple random walk to the Brownian motion. Consider the sequence of stopping times defined recursively by $\tau_0 = 0$ and then

$$\tau_{i+1} = \inf\{t \geq \tau_i : |B_t - B_{\tau_i}| = 1\}.$$

- (i) Prove that $(\tau_{i+1} - \tau_i, B_{\tau_{i+1}} - B_{\tau_i})_{i \geq 1}$ is a sequence of i.i.d. random variables with the same law as (τ_1, B_{τ_1}) .
- (ii) Express the law of B_{τ_1} .
- (iii) Prove that $\mathbb{E}[\tau] = 1$. *Hint*: use the stopping theorem.
- (iv) For each $n \geq 1$ and $t \geq 0$, let

$$S_t^{(n)} = \frac{B_{\tau_{[nt]}} + (nt - [nt])(B_{\tau_{[nt]+1}} - B_{\tau_{[nt]}})}{\sqrt{n}} \quad \text{and} \quad B_t^{(n)} = \frac{B(nt)}{\sqrt{n}}.$$

- (v) Prove that $B^{(n)}$ is a standard Brownian motion.
- (vi) Prove that for any $K > 0$, it holds

$$\sup_{t \in [0, K]} |S_t^{(n)} - B_t^{(n)}| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

- (vii) Deduce Donsker's theorem for the simple random walk.

EXERCISE 35. Suppose that $E[\xi_1] = 0$ and $\text{Var}(\xi_1) = 1$. Let $A_0 = 0$ and $A_n = \sum_{i=1}^n S_i = \sum_{i=1}^n \sum_{k=1}^i \xi_k$ for every $n \geq 1$. Extend both $(S_n)_{n \geq 0}$ to $(S_t)_{t \in [0, \infty)}$ and $(A_n)_{n \geq 0}$ to $(A_t)_{t \in [0, \infty)}$ by linear interpolation and finally let

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n}} \quad \text{and} \quad A_t^{(n)} = \frac{A_{nt}}{n^{3/2}}.$$

Prove that the sequence of continuous processes $(S^{(n)}, A^{(n)})_{n \geq 1}$ converges in distribution and express the limit.

EXERCISE 36. For every $n \geq 1$, let X^n be a random variable in $C = C([0, \infty), \mathbb{R})$ with $X_0^n = 0$ and assume that $X^n \rightarrow B$ in distribution.

- (i) Fix an arbitrary $T > 0$ and let $\bar{X}_T^n = \sup_{t \leq T} X_t^n$ and similarly let $\bar{B}_T = \sup_{t \leq T} B_t$. Do we have $\bar{X}_T^n \rightarrow \bar{B}_T$ in distribution? What is the law of $\bar{B}_T$?
- (ii) Fix an arbitrary $a > 0$ and let $T_a^n = \inf\{t \geq 0 : X_t^n \geq a\}$ and similarly let $T_a = \inf\{t \geq 0 : B_t \geq a\}$. Do we have $T_a^n \rightarrow T_a$ in distribution? (This requires some extra notions on the Brownian motion.)

EXERCISE 37. For every $c > 0$, let $B^{(c)}$ denote the process given by

$$B_t^{(c)} = \frac{1}{\sqrt{c}} B_{ct}.$$

Let B' be another standard Brownian motion on $\mathbb{R}$, independent of B .

- (i) Show that $B^{(c)}$ is also a standard Brownian motion on $\mathbb{R}$ for any $c > 0$.
- (ii) Fix $c \in (0, 1)$, prove that the pairs

$$((B_s^{(c)})_{0 \leq s \leq 1}, (B_{c+t})_{t \geq 0}) \quad \text{and} \quad ((B_s)_{0 \leq s \leq 1}, (\sqrt{c}B_1 + B'_t)_{t \geq 0})$$

have the same law on $C([0, 1], \mathbb{R}) \times C([0, \infty), \mathbb{R})$.

- (iii) Deduce the convergence in distribution:

$$((B_s^{(c)})_{0 \leq s \leq 1}, (B_t)_{t \geq 0}) \xrightarrow[c \rightarrow 0]{(d)} ((B_s)_{0 \leq s \leq 1}, (B'_t)_{t \geq 0}).$$

Is it surprising?

- (iv) What happens if instead we let $c \rightarrow \infty$?

ARCSINE LAWS. Consider the following four quantities about Brownian motion up to time 1:

$$Z = \sup\{t \in [0, 1] : B_t = 0\}, \quad P = \text{Leb}(\{t \in [0, 1] : B_t > 0\}),$$

where Leb denotes the Lebesgue measure, as well as

$$I = \inf\{t \in [0, 1] : B_t = \inf_{s \in [0, 1]} B_s\}, \quad \text{and} \quad M = \sup\{t \in [0, 1] : B_t = \sup_{s \in [0, 1]} B_s\}.$$

Let us recall some properties of Brownian motion that hold almost surely:

- (i) it does not make a local extremum at any given height, so in particular if $B_t = 0$ then it is not a local extremum,
- (ii) local extrema are all distinct, so in particular times such that $B_t = \inf_{s \in [0, 1]} B_s$ or such that $B_t = \sup_{s \in [0, 1]} B_s$ are unique,

(iii) we have $\text{Leb}(\{t \in [0, 1] : B_t = 0\}) = 0$ so $P = \text{Leb}(\{t \in [0, 1] : B_t \geq 0\})$.

The *arcsine law* is the law of a random variable $T \in [0, 1]$ with distribution function and density respectively given by:

$$\mathbb{P}(T \leq t) = \frac{2}{\pi} \arcsin \sqrt{t}, \quad \text{or equivalently,} \quad \mathbb{P}(T \in dt) = \frac{1}{\pi \sqrt{t(1-t)}} dt.$$

It is the law of $(\sin \Theta)^2$ where Θ is an angle uniformly distributed over $[0, 2\pi]$.

EXERCISE 38 (Arcsine laws). The goal of this exercises is to show that

$$Z, P, I, M \quad \text{all have the arcsine law!}$$

We propose to do so by approximation with the discrete model of the simple random walk $S_n = \xi_1 + \dots + \xi_n$. Define thus accordingly

$$Z_n = \sup\{k \in \{0, \dots, n\} : S_{2k} = 0\}, \quad P_n = \#\{k \in \{1, \dots, n\} : S_k > 0\},$$

as well as

$$I_n = \inf\{k \in \{0, \dots, n\} : S_k = \min_{0 \leq j \leq n} S_j\}, \quad \text{and} \quad M_n = \sup\{k \in \{0, \dots, n\} : S_k = \max_{0 \leq j \leq n} S_j\}.$$

Let us start with some preliminary steps.

- (i) Prove that each of these four random variables divided by n converges in distribution to the analogous quantity on Brownian motion.
- (ii) Establish that

$$\mathbb{P}(S_{2n} = 0) = \mathbb{P}\left(\bigcap_{k=1}^{2n} S_k \geq 0\right) = \mathbb{P}\left(\bigcap_{k=1}^{2n-1} S_k \geq 0\right) = 2 \mathbb{P}\left(\bigcap_{k=1}^{2n} S_k > 0\right) = \mathbb{P}\left(\bigcap_{k=1}^{2n} S_k \neq 0\right).$$

Remark: The first identity can be proved by establishing a bijection between the corresponding paths; the other equalities are simpler.

Let us first consider Z .

- (iii) Prove that for any $0 \leq k \leq n$, we have

$$\mathbb{P}(Z_n = k) = \mathbb{P}(S_{2k} = 0) \mathbb{P}(S_{2(n-k)} = 0).$$

Deduce that for every $t \in (0, 1)$, we have

$$n \mathbb{P}(Z_n = \lfloor tn \rfloor) \xrightarrow{n \rightarrow \infty} \frac{1}{\pi \sqrt{t(1-t)}}.$$

- (iv) Observe that both sides above define probability densities on $[0, 1]$. Prove that if $\tilde{Z}_n$ and $\tilde{Z}$ have these densities respectively, then

$$\tilde{Z}_n \xrightarrow[n \rightarrow \infty]{(d)} \tilde{Z}.$$

- (v) Prove that Z_n has the same law as $\lfloor n \tilde{Z}_n \rfloor$ and conclude that

$$n^{-1} Z_n \xrightarrow[n \rightarrow \infty]{(d)} \tilde{Z}.$$

- (vi) Conclude that Z has the arcsine law.

Let us next turn to I_n and M_n .

(vii) Prove that for any $0 \leq k \leq n$, we have

$$\mathbb{P}(I_n = k) = \frac{1}{2} \mathbb{P}(S_{k^{\leftarrow}} = 0) \mathbb{P}(S_{(n-k)^{\rightarrow}} = 0),$$

where $k^{\leftarrow} = k$ if k is even and $k^{\leftarrow} = k - 1$ if k is odd and $(n - k)^{\rightarrow} = n - k$ if $n - k$ is even and $(n - k)^{\rightarrow} = n - k + 1$ if $n - k$ is odd.

(viii) Adapt the argument used for Z_n to prove that $n^{-1}I_n$ converges in distribution to the arcsine law.

(ix) Conclude about M_n .

Let us finally consider P . Let $\mathbf{z} = (z_k)_{k \geq 1} \in \{-1, +1\}^{\mathbb{N}}$ be an infinite sequence and for each $n \geq 1$, let $\mathbf{z}_n = (z_1, \dots, z_n)$ denote the first n entries. For every $x \in \{-1, +1\}$, we let

$$x \cdot \mathbf{z}_n = (x, z_1, \dots, z_n) \in \{-1, +1\}^{n+1} \quad \text{and} \quad \mathbf{z}_n \cdot x = (z_1, \dots, z_n, x) \in \{-1, +1\}^{n+1}.$$

Let us then define recursively a permutation $\widehat{\mathbf{z}}_n$ of $\mathbf{z}_n$ as follows. First, let $\widehat{\mathbf{z}}_1 = \mathbf{z}_1 = z_1$ and for every $n \geq 2$, let

$$\widehat{\mathbf{z}}_n = \begin{cases} \mathbf{z}_n \cdot \widehat{\mathbf{z}}_{n-1} & \text{if } z_1 + \dots + z_n > 0, \\ \widehat{\mathbf{z}}_{n-1} \cdot \mathbf{z}_n & \text{otherwise.} \end{cases}$$

(x) Just to understand the construction: Prove that if $\mathbf{z}_4 = (-1, +1, +1, -1)$, then $\widehat{\mathbf{z}}_4 = (+1, -1, +1, -1)$.

(xi) Prove by induction that if we write $\widehat{\mathbf{z}}_n = (\widehat{z}_{n,1}, \dots, \widehat{z}_{n,n})$, then (an empty sum being null),

$$\#\{k \in \{1, \dots, n\} : z_1 + \dots + z_k > 0\} = \min\{k \in \{0, \dots, n\} : \widehat{z}_{n,1} + \dots + \widehat{z}_{n,k} = \max_{0 \leq j \leq n} \widehat{z}_{n,1} + \dots + \widehat{z}_{n,j}\}.$$

Let $\Xi = (\xi_n)_{n \geq 1}$ be i.i.d. random variables that are uniformly distributed over $\{-1, +1\}$ and for each $n \geq 1$, let $\widehat{\Xi}_n = (\widehat{\xi}_{n,1}, \dots, \widehat{\xi}_{n,n})$ denote the permutation of $(\xi_1, \dots, \xi_n)$ constructed as above. Let $S_n = \xi_1 + \dots + \xi_n$.

(xii) Prove that $\widehat{\Xi}_n$ has the same law as Ξ_n for every $n \geq 1$.

(xiii) Conclude that $n^{-1}P_n$ converges to the arcsine law.

EXERCISE 39 (Lévy's identity). Let $\underline{B}_t = \inf_{s \in [0, t]} B_s$. A famous theorem of Lévy (again!) shows that the processes

$$(B_t - \underline{B}_t)_{t \in [0, \infty)} \quad \text{and} \quad (|B_t|)_{t \in [0, \infty)}$$

have the same law. Just as for the arcsine laws, this exercise proposes a proof based on the approximation by simple random walk. Define therefore accordingly $\underline{S}_n = \min_{0 \leq i \leq n} S_i$. Let further $\theta_0 = \theta_0^{-1} = 0$, and

$$\theta_n = \#\{1 \leq k \leq n : S_k - \underline{S}_k \neq S_{k-1} - \underline{S}_{k-1}\} \quad \text{and} \quad \theta_k^{-1} = \inf\{n \geq 1 : \theta_n = k\}.$$

Finally, let $X_n = S_{\theta_n^{-1}} - \underline{S}_{\theta_n^{-1}}$ for every $n \geq 0$ and extend all sequences by linear interpolation.

(i) Prove that the sequence $(S_n - \underline{S}_n)_{n \geq 0}$ is a Markov chain on $\mathbb{Z}_+$ and that the sequences $(X_n)_{n \geq 0}$ and $(|S_n|)_{n \geq 0}$ have the same law (in the product space $\mathbb{Z}_+^{\mathbb{Z}_+}$).

(ii) Prove that

$$\frac{1}{n} \#\{k \leq n : S_k - \underline{S}_k = 0\} \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Deduce that $(\theta_n^{-1}/n)_{n \in [0, \infty)}$ converges almost surely to the identity function.

(iii) Conclude about Lévy's identity.

CHAPTER 8

DISCONTINUOUS STOCHASTIC PROCESSES

This chapter briefly describes how to deal with stochastic processes with non-continuous sample paths, which in many contexts is more natural than continuous functions. This will be rapidly covered in class; the idea is to adapt the tools from continuous functions in order to accomodate discontinuities. We shall mostly follow the presentation of Billingsley again.

CONTENTS

8.1	Cadlag functions	86
8.2	The Skorokhod distance	88
8.3	A complete distance	90
8.4	Marginal laws	90
8.5	Compact sets and tightness	93
8.6	Exercises	94

You might have already encountered in previous courses discontinuous processes, such as the Poisson process (presented in Chapter 10 below). More generally, Chapter 10 will introduce the family of Lévy processes, which are all discontinuous except, among them: deterministic affine functions and affine transformations of Brownian motion, but that's it! It is therefore of great importance to be able to work with discontinuous processes, but the uniform distance (on compact sets if one wants to deal with processes defined on $[0, \infty)$) is not well adapted to this. For example, it feels natural to require that if $a_n \rightarrow a$ in $\mathbb{R}$, then $\mathbb{1}_{[a_n, \infty)} \rightarrow \mathbb{1}_{[a, \infty)}$, but, unless $a_n = a$, these two functions lie at uniform distance 1 from each other, so it is not even a Cauchy sequence! (and this shows some separability issue as well.)

The natural idea to deal with discontinuities is to allow for small time-distortions. The picture one may have in mind is that two functions x and y are ε -close to each other in the uniform distance (on compact sets) when the graph of y belongs to the 'strip' given by the graph of x enlarged vertically by $\pm\varepsilon$. Then now enlarge also the graph of x horizontally by $\pm\varepsilon$, so if x has a jump, then y will stay close to x whenever it also has a jump of approximately the same size, at approximately, but not necessarily exactly, the same time.

This chapter will only be superficially covered in class; some proofs are still missing from these notes.

8.1 CADLAG FUNCTIONS

We will not treat all stochastic processes, but instead we will impose a less restrictive behaviour on the trajectory than continuity by considering càdlàg processes. Here, the discontinuities will cause some (light) issues when considering functions on $[0, \infty)$ and we only focus on functions defined on $[0, 1]$, but it could be any compact interval with obvious changes.

DEFINITION 8.1. Let (E, d) be a metric space. A function $x : [0, 1] \rightarrow E$ is said to be *càdlàg* when:

- For every $t \in (0, 1]$, the left-limit $x_{t-} = \lim_{s \uparrow t} x_s$ exists,
- For every $t \in [0, 1)$, the right-limit $x_{t+} = \lim_{s \downarrow t} x_s$ exists and $x_{t+} = x_t$.

We henceforth denote by $\mathcal{D}$ the set of càdlàg functions $x : [0, 1] \rightarrow E$.

Let henceforth assume that (E, d) is separable and complete.

Recall the modulus of continuity of a function $x : [0, 1] \rightarrow E$ at $\delta > 0$, given by

$$\omega(x, \delta) = \sup\{d_E(x_s, x_t) : s, t \in [0, 1], |t - s| < \delta\}.$$

Recall that x is continuous if and only if $\omega(x, \delta) \rightarrow 0$ as $\delta \rightarrow 0$. Let us adapt the definition to accommodate the discontinuities. First for $0 \leq s < t \leq 1$, let

$$\omega(x, [s, t)) = \sup_{u, v \in [s, t)} d_E(x_u, x_v).$$

Next, for $\delta > 0$, let S_δ denote the set of finite sequences $0 = t_0 < \dots < t_k = 1$ for some $k \geq 1$, that satisfy $t_i - t_{i-1} > \delta$ for every $1 \leq i \leq k$. Finally, let

$$\omega'(x, \delta) = \inf_{(t_i)_{i \in S_\delta}} \max_{1 \leq i \leq k} \omega(x, [t_{i-1}, t_i)). \quad (8.1)$$

LEMMA 8.2. A function $x : [0, 1] \rightarrow E$ lies in $\mathcal{D}$ if and only if $\omega'(x, \delta) \rightarrow 0$ as $\delta \rightarrow 0$.

Proof. $\Rightarrow$: Suppose that $x \in \mathcal{D}$ and fix $\varepsilon > 0$. Since x is right-continuous at 0, then there exists $t > 0$ such that $\omega(x, [0, t)) < \varepsilon$. Therefore, if we let $\bar{t}$ denote the supremum of those $t \in [0, 1]$ such that the interval $[0, t)$ can be decomposed into finitely many sub-intervals $[t_{i-1}, t_i)$ with $\max_{1 \leq i \leq k} \omega(x, [t_{i-1}, t_i)) < \varepsilon$, then $\bar{t} > 0$. In particular x admits a left-limit at $\bar{t}$ and this implies that the interval $[0, \bar{t})$ can be decomposed like so. If $\bar{t} < 1$, by right-continuity, a slightly larger interval could also be decomposed like so, which would contradict the very definition of $\bar{t}$, so $\bar{t} = 1$. Taking $\delta > 0$ smaller than $\min_i(t_i - t_{i-1})$, this decomposition belongs to S_δ so $\omega'(x, \delta) < \varepsilon$.

$\Leftarrow$: Suppose conversely that for every $\varepsilon > 0$, there exists $\delta > 0$ and a sequence $(t_i)_{0 \leq i \leq k}$ in S_δ such that $\max_{1 \leq i \leq k} \omega(x, [t_{i-1}, t_i)) < \varepsilon$. Fix any $0 \leq t < 1$, let $\varepsilon > 0$ be arbitrary and let $(t_i)_{0 \leq i \leq k}$ be such a sequence for some $\delta > 0$. Then t belongs to some interval $[t_{i-1}, t_i)$ so $d_E(x_t, x_s) < \varepsilon$ for every $s \in [t, t_i)$. Since $\varepsilon > 0$ is arbitrary, then x is right-continuous at t . For any $0 < t \leq 1$, this time now belongs to some interval $(t_{i-1}, t_i]$, so any sequence $s_n \uparrow t$ is eventually contained in (t_{i-1}, t_i) , so for every n, m large enough, we have $d_E(x_{s_n}, x_{s_m}) < \varepsilon$. The sequence $(x_{s_n})_n$ is thus a Cauchy sequence, and hence a convergent sequence so x admits a left-limit at t . $\square$

We may derive several useful results from this lemma. Let us say that $t \in (0, 1]$ is a *jump time* of x when $d_E(x_t, x_{t-}) > 0$, the value of $d_E(x_t, x_{t-})$ is called the *magnitude* of this jump.

- For any $\varepsilon > 0$, a function in $\mathcal{D}$ only has finitely many jump times of magnitude larger than ε , namely $\#\{t \in (0, 1] : d_E(x_t, x_{t-}) > \varepsilon\} < \infty$.
- As a consequence (taking rational values of ε), a function in $\mathcal{D}$ only has at most countably many jump times at all.
- For any function in $x \in \mathcal{D}$, the closure of its range $x([0, 1]) = \{x_t : t \in [0, 1]\}$ is a compact set in E . Indeed, let $(t_n)_n$ be any sequence in $[0, 1]$, then it admits a convergent subsequence $(t_{n_i})_{i}$, with some

limit t , and taking a further sub-sequence if necessary, we may assume that it is either increasing or nonincreasing. We have $x_{t_{n_i}} \rightarrow x_t$ in the latter case and $x_{t_{n_i}} \rightarrow x_{t-}$ in the former case. In any case, the limit belongs to the closure of the range.

- (iv) Any function $x \in \mathcal{D}$ is measurable for the Borel σ -algebra on $[0, 1]$ and E . Indeed, for every ε , there exists $\delta > 0$ and a sequence $(t_i)_i \in S_\delta$ such that the function constant equal to $x_{t_{i-1}}$ on each interval $[t_{i-1}, t_i)$ approximates x uniformly, up to an ε of margin. Such step functions are measurable thus so is their limit as $\varepsilon \rightarrow 0$.

8.2 THE SKOROKHOD DISTANCE

Let us formalise the idea that two possibly discontinuous functions x and y lie at distance less than ε from each other when the graph of y belongs to the vertical and horizontal ε -enlargement of the graph of x . The horizontal distortion is captured in the following set of homeomorphisms:

$$\Lambda = \{\lambda : [0, 1] \rightarrow [0, 1] \text{ continuous, strictly increasing, and with } \lambda(0) = 0 \text{ and } \lambda(1) = 1\}.$$

Let $I \in \Lambda$ denote the identity. Then for every functions $x, y : [0, 1] \rightarrow E$, let

$$d(x, y) = \inf_{\lambda \in \Lambda} \left(\sup_{t \in [0, 1]} |\lambda(t) - t| \vee \sup_{t \in [0, 1]} d_E(x_{\lambda(t)}, y_t) \right). \quad (8.2)$$

Exercise 41 is to show that d defines a distance on $\mathcal{D}$. The topology induced by this distance is called the *Skorokhod topology*. Note that for functions x and x^n in $\mathcal{D}$, we have $d(x^n, x) \rightarrow 0$ if and only if there exists a sequence λ^n of functions in Λ such that both $\lambda^n \rightarrow I$ and $x^n \circ \lambda^n \rightarrow x$ uniformly.

Exercise 42 discusses the links between convergence in this topology and uniform convergence. A consequence of this exercise is that the topology induced on $\mathcal{C}$ by the Skorokhod topology coincides with the uniform topology. A consequence of Exercise 43 is that the space $\mathcal{C}$ of continuous functions is closed in $\mathcal{D}$.

REMARK 8.3. A sequence of continuous functions cannot converge in the Skorokhod topology to a discontinuous function. So for example, the function constant equal to 0 on $[0, 1/2 - 1/n]$, constant equal to 1 on $[1/2, 1]$, and linear on $[1/2 - 1/n, 1/2]$ does not converge at all in $(\mathcal{D}, d)$ since it converges pointwise to the discontinuous function $\mathbb{1}_{[1/2, 1]}$.

REMARK 8.4. Actually, the convergence in the previous example does hold in a weaker topology: Skorokhod defined four topologies on real-valued càdlàg functions, called J_1 , J_2 , M_1 , and M_2 . The one we present here is the J_1 -topology, which is the strongest and most predominant one. The example does converge in the weaker M_1 topology, which in some sense is made for such a convergence to hold. In a very loose sense, this topology consists in ‘augmenting’ the graph of a càdlàg function by adding, for each jump time t the vertical line from x_{t-} to x_t to the graph, and saying again that two functions are close to each other if, after a small time- and space-distortion, the two augmented graphs are close to each other. We will not face these issues here and only speak of the J_1 topology hereafter. The interested reader can find more information on the other topologies in the book *Stochastic-Process Limits* by Ward Whitt. We stress however that the ‘vertical line from x_{t-} to x_t ’ does not make sense in a general metric space E , so when using the M_1 topology, one usually works in a Banach linear space.

REMARK 8.5. Another issue when dealing with càdlàg functions is that the summation is not continuous! Take $x^n = \mathbb{1}_{[1/2 - 1/(2n), 1]}$ and $y^n = \mathbb{1}_{[1/2 - 1/n, 1]}$. Then both $x^n, y^n \rightarrow x = \mathbb{1}_{[1/2, 1]}$ for the distance d and

$x_t^n + y_t^n \rightarrow 2x_t$ for every $t \neq 1/2$. However $x^n + y^n$ does not converge to $2x$ for d , and thus does not converge at all in $(\mathcal{D}, d)$. This is due to the small part $[1/2 - 1/(2n), 1/2 - 1/n]$ where $x^n + y^n = 1$, which cannot be matched in $2x$, whatever homeomorphism λ^n we take. (Exercise: fill-in the details.) Thus, although $\mathcal{D}$ is a vector space, $(\mathcal{D}, d)$ is not a topological vector space, as opposed to $\mathbb{C}$ equipped with the uniform distance.

(We do have $x^n + y^n \rightarrow 2x$ in the weaker M_1 topology alluded in the previous remark, but not $x^n - y^n \rightarrow 0$, so this weaker topology again fails to give a topological vector space.)

Recall that we developed the theory of weak convergence preferentially on separable and complete spaces.

PROPOSITION 8.6. The space $(\mathcal{D}, d)$ is separable but is in general not complete.

Proof of separability. Let F denote a countable dense set of E . For every $k \geq 1$, let B_k denote the set of functions x that are constant on each interval $[(i-1)/k, i/k]$ for $1 \leq i \leq k$, where they assume a value in F , and with also $x(1) \in F$. Then $B = \bigcup_k B_k$ is countable and we claim that it is dense. Indeed, fix $x \in \mathcal{D}$ and $\varepsilon > 0$ arbitrary and let $k > 1/\varepsilon$ be such that $\omega'(x, 1/k) < \varepsilon$. For each $1 \leq i \leq k+1$, choose a value $a_{i-1} \in F$ such that $d_E(x_{(i-1)/k}, a_{i-1}) < 1/\varepsilon$ and let $y \in B$ denote the function constant equal to a_{i-1} on $[(i-1)/k, i/k]$ for $1 \leq i \leq k$ and with $y_1 = a_k$. Let us upper bound the distance between x and y by constructing an appropriate homeomorphism $\lambda \in \Lambda$.

Since $\omega'(x, 1/k) < \varepsilon$, then there exists a finite sequence $0 = t_0 < \dots < t_\ell = 1$ such that $t_j - t_{j-1} > 1/k$ and $\sup_{u,v \in [t_{j-1}, t_j]} d_E(x_u, x_v) < \varepsilon$ for each $1 \leq j \leq \ell$. Let $\lambda(0) = 0$ and for each $1 \leq j \leq \ell$, let $\lambda(t_j) = i/k$ for the unique index i such that $(i-1)/k < t_j \leq i/k$. Since $t_j - t_{j-1} > 1/k$ then no two t_j 's lie in the same interval $((i-1)/k, i/k]$ and thus the sequence $(\lambda(t_j))_j$ is strictly increasing. Then the linear interpolation defines a function $\lambda \in \Lambda$ that satisfies $\sup_t |\lambda(t) - t| \leq 1/k$.

Fix $t \in (0, 1)$, so there exists $1 \leq i \leq k$ such that $(i-1)/k \leq t < i/k$. Let $1 \leq j \leq \ell$ be the unique index such that $t_{j-1} \leq (i-1)/k < t_j$. We claim that $\lambda^{-1}(t)$ belongs to the same interval $[t_{j-1}, t_j]$. Indeed, first $t_j > (i-1)/k$, so $\lambda(t_j) \geq i/k > t$, and further $t_j > \lambda^{-1}(t)$. Next $t_{j-1} \leq (i-1)/k$, so $\lambda(t_{j-1}) \leq (i-1)/k \leq t$ so $t_{j-1} \leq \lambda^{-1}(t)$.

We have shown that if $t \in [(i-1)/k, i/k]$ for some $1 \leq i \leq k$, then $(i-1)/k$ and $\lambda^{-1}(t)$ belong to the same interval $[t_{j-1}, t_j]$. By the very definition of this interval, we have $d_E(x_{(i-1)/k}, x_{t_{j-1}}) < \varepsilon$ and $d_E(x_{\lambda^{-1}(t)}, x_{t_{j-1}}) < \varepsilon$. In addition, by construction of the function y , we have $y_t = y_{(i-1)/k} = a_{i-1}$ with $d_E(x_{(i-1)/k}, a_{i-1}) < 1/\varepsilon$. We conclude that

$$d_E(x_{\lambda^{-1}(t)}, y_t) \leq d_E(x_{\lambda^{-1}(t)}, x_{t_{j-1}}) + d_E(x_{(i-1)/k}, x_{t_{j-1}}) + d_E(x_{(i-1)/k}, y_t) < 3\varepsilon.$$

This concludes the proof that the countable set B of such functions y is dense in $\mathcal{D}$. $\square$

Proof of incompleteness. Let us exhibit a Cauchy sequence that does not converge. Fix two distinct points $a \neq b$ in E (if E is a singleton, then $\mathcal{D}$ is complete...) and let $\Delta = d_E(a, b) > 0$. For each $n \geq 1$, let $x_t^n = a$ if $t \in [1/2, 1/2 + 1/2^n]$ and $x_t^n = b$ otherwise and let $\lambda^n \in \Lambda$ be any increasing homeomorphism such that $\lambda^n(1/2) = 1/2$ and $\lambda^n(1/2 + 1/2^n) = 1/2 + 1/2^{n+1}$, i.e. that maps $[1/2, 1/2 + 1/2^n]$ to $[1/2, 1/2 + 1/2^{n+1}]$. Then $x^{n+1} \circ \lambda^n = x^n$. If we further take the λ^n that is linear on the three intervals $[0, 1/2]$, $[1/2, 1/2 + 1/2^n]$, and $[1/2 + 1/2^n, 1]$, then we can compute exactly $\sup_t |\lambda^n(t) - t| = 2^{-(n+1)}$. Hence $d(x^n, x^{n+1}) \leq 2^{-(n+1)}$ and this implies that it is a Cauchy sequence.

Suppose that $x^n \rightarrow x$ in the Skorokhod topology for some $x \in \mathcal{D}$. Then according to Exercise 42, we have $x_t^n \rightarrow x_t$ for every t at which x is continuous. In addition, by Lemma 8.2 the function x can only have at most countably many jump times, so continuity points are dense in $[0, 1]$. However we see that $x_t^n \rightarrow b$ for every $t \neq 1/2$ so we have $x_t = b$ for a dense set of $[0, 1]$, and since x is right-continuous,

then $x_t = b$ for all $t \in [0, 1]$. Hence, if x^n converges in the Skorokhod topology to some càdlàg limit, this limit is the function constant equal to b . But $d(x^n, b) = \Delta$ for every n , which contradicts the fact that x^n converges at all in the Skorokhod topology. $\square$

8.3 A COMPLETE DISTANCE

Actually, let us recall that the general theory of weak convergence does not require a separable and complete metric space, but only a *Polish space*, that is, a separable topological space whose topology can be induced by complete distance. Just as the open interval $(0, 1)$ is Polish by replacing the Euclidean distance by an equivalent one that is complete (inherited from $\mathbb{R}$), we are going to construct another distance d° on $\mathcal{D}$ that is equivalent to d and is complete (assuming that E is separable and complete). This shows that $\mathcal{D}$ is a Polish space and we can still use the distance d in practice, which is simpler to manipulate than d° .

The idea is to restrict more the set of homeomorphisms Λ in the definition (8.2) of d by requiring that, not only λ is uniformly close to the identity, but it is also close to be 1-Lipschitz, i.e. that $(\lambda(t) - \lambda(s))/(t - s)$ is close to 1. This makes sense if we think of time distortion: we want the flow of time to be close to unit speed. It will be simpler to require instead the logarithm be close to 0. More precisely, for every $\lambda \in \Lambda$, define

$$|\lambda|^\circ = \sup_{0 \leq s < t \leq 1} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|.$$

That this quantity is finite implies that the ratios in the log are bounded away from both 0 and ∞ . Note that the upper bound shows that λ is Lipschitz, and the lower bound that it is strictly increasing. Next we simply define for every functions $x, y : [0, 1] \rightarrow E$:

$$d^\circ(x, y) = \inf_{\lambda \in \Lambda} \left(|\lambda|^\circ \vee \sup_{t \in [0, 1]} d_E(x_{\lambda(t)}, y_t) \right). \quad (8.3)$$

Exercise 44 is to show that d° defines a distance on $\mathcal{D}$.

The next lemma proves that d and d° are equivalent metrics.

LEMMA 8.7. Fix any $x, y \in \mathcal{D}$. Then

$$d(x, y) \leq e^{d^\circ(x, y)} - 1,$$

and for every $\delta \in (0, 1/2)$,

$$\text{if } d(x, y) < \delta^2, \quad \text{then } d^\circ(x, y) \leq 4\delta + \omega'(x, \delta).$$

Proof. To do! $\square$

Separability is conserved from d to d° , but we gain completeness. For a sanity check, we encourage the reader to go back to the example of a non convergent Cauchy sequence for d and check that it is not a Cauchy sequence for d° .

PROPOSITION 8.8. The space $(\mathcal{D}, d^\circ)$ is complete.

Proof. To do! $\square$

8.4 MARGINAL LAWS

Recall the evaluation map

$$\pi_J : x \mapsto (x_t)_{t \in J}$$

for every subset $J \subset [0, 1]$ and every function $x : [0, 1] \rightarrow E$. Let us restrict to functions $x \in \mathcal{D}$. Note that π_0 and π_1 are continuous for the Skorokhod topology since any $\lambda \in \Lambda$ has $\lambda(0) = 0$ and $\lambda(1) = 1$. However, as opposed to continuous functions, for $t \in (0, 1)$, the evaluation π_t is not continuous in general. It is thus not as immediate whether it is Borel-measurable and how the σ -algebra generated by the projections compares with the Borel σ -algebra $\mathcal{B}(\mathcal{D})$.

THEOREM 8.9. The following holds.

- (i) For every $t \in (0, 1)$, the evaluation π_t is continuous at a function $x \in \mathcal{D}$ if and only if the function x is continuous at t .
- (ii) For every finite subset $J \subset [0, 1]$, say with k elements, the evaluation π_J is measurable for the Borel σ -algebras $\mathcal{B}(\mathcal{D})$ and $\mathcal{B}(E^k) = \mathcal{B}(E)^{\otimes k}$ (since E is separable).
- (iii) If $J \subset [0, 1]$ is dense and contains 1, then $\mathcal{B}(\mathcal{D}) = \sigma(\pi_t : t \in J)$. In addition, two probability measures on this σ -algebra that agree on the collection $\{\pi_K^{-1}(\prod_{t \in K} A_t) : A_t \in \mathcal{C} \text{ for every } t \in K \text{ and } K \subset J \text{ is finite}\}$ agree on $\mathcal{B}(\mathcal{D})$.

Proof. Let x^n and x be functions in $\mathcal{D}$.

- (i) If x is continuous at t , then Exercise 42 shows that, if $x^n \rightarrow x$ in $\mathcal{D}$, then $\pi_t(x^n) = x_t^n \rightarrow x_t = \pi_t(x)$, so π_t is continuous at x . Conversely, if x is not continuous at t , then let simply $\lambda^n \in \Lambda$ be linear on $[0, t]$ and on $[t, 1]$ and with $\lambda^n(t) = t - n^{-1}$, so $x^n = x \circ \lambda^n \rightarrow x$ in $\mathcal{D}$ but $x_t^n = x_{t-n^{-1}} \rightarrow x_{t-} \neq x_t$, so π_t is not continuous at x .
- (ii) Since measurability for $\mathcal{B}(E)^{\otimes k}$ amounts to the individual measurability for $\mathcal{B}(E)$ of each component, then it is sufficient to treat the case $k = 1$. For $t = 0$ and $t = 1$, the evaluation π_t is continuous and so Borel measurable. Fix $t \in (0, 1)$ and $f : E \rightarrow \mathbb{R}$ a continuous and bounded function. For every $\varepsilon > 0$, define a function $f_t^\varepsilon : \mathcal{D} \rightarrow \mathbb{R}$ by

$$f_t^\varepsilon(x) = \varepsilon^{-1} \int_t^{t+\varepsilon} f(x_s) ds.$$

Since $x^n \rightarrow x$ in $\mathcal{D}$ implies $x_s^n \rightarrow x_s$ at each continuity point s of x , whose complement is at most countable, then $f(x_s^n) \rightarrow f(x_s)$ Lebesgue almost everywhere, so f_t^ε is a continuous function on $\mathcal{D}$ by dominated convergence. In particular it is Borel measurable. Since $s \mapsto f(x_s)$ is right-continuous at t , then $f_t^\varepsilon(x) \rightarrow f(x_t)$ as $\varepsilon \rightarrow 0$, so $f \circ \pi_t$ is Borel measurable for every continuous and bounded function $f : E \rightarrow \mathbb{R}$. Let $F \subset E$ be a closed set and recall that the indicator function $\mathbb{1}_F$ can be approximated by the continuous functions $a \mapsto (1 - nd_E(a, F))_+$, so $\mathbb{1}_F \circ \pi_t$ is Borel measurable for every closed set F of E . Consequently, for every closed set F ,

$$\pi_t^{-1}(F) = \{x \in \mathcal{D} : x_t \in F\} = \{x \in \mathcal{D} : \mathbb{1}_F(x_t) = 1\} = (\mathbb{1}_F \circ \pi_t)^{-1}(\{1\}),$$

belongs to $\mathcal{B}(\mathcal{D})$ and this extends to any set in the σ -algebra they generate, namely $\mathcal{B}(E)$.

- (iii) The last part follows from the monotone class lemma since the collection is a π -system that generates $\sigma(\pi_t : t \in J)$. It remains to prove the equality of the two σ -algebras. Since we have shown that each π_t is Borel measurable, then

$$\sigma(\pi_t : t \in J) \subset \sigma(\pi_t : t \in [0, 1]) \subset \mathcal{B}(\mathcal{D}),$$

and it remains to prove the converse inclusions. For the first one, since J is dense, then for every $t \in [0, 1)$, there exists a sequence $(t_n)_n$ in J with $t_n \downarrow t$, which implies that $\pi_{t_n} \rightarrow \pi_t$. With

the additional assumption that $1 \in J$, we conclude that for every $t \in [0, 1]$, the evaluation π_t is $\sigma(\pi_t : t \in J)$ -measurable, so

$$\sigma(\pi_t : t \in J) = \sigma(\pi_t : t \in [0, 1]).$$

Now to prove that $\mathcal{B}(\mathcal{D}) \subset \sigma(\pi_t : t \in [0, 1])$, let us prove that the identity $x \mapsto x$ is measurable when seen as a function $(\mathcal{D}, \sigma(\pi_t : t \in [0, 1])) \rightarrow (\mathcal{D}, \mathcal{B}(\mathcal{D}))$. The argument is quite close to the one used in the proof of Proposition 8.6.

Fix $k \geq 1$ and let $J = \{i/k : 0 \leq i \leq k\}$. Then $\pi_J : \mathcal{D} \rightarrow E^{k+1}$ is measurable for the σ -algebras $\sigma(\pi_t : t \in J)$ and $\mathcal{B}(E)^{\otimes k+1}$ since each component is clearly measurable for $\sigma(\pi_t : t \in J)$ and $\mathcal{B}(E)$. For any $A_k = (a_0, \dots, a_k) \in E^{k+1}$, let $(xA_k) \in \mathcal{D}$ be constant equal to $a_{(i-1)/k}$ on each interval $[(i-1)/k, i/k)$ for $1 \leq i \leq k$, and with $(xA_k)_1 = a_k$. Then $A_k \mapsto xA_k$ is a continuous function, so it is measurable from $\mathcal{B}(E)^{\otimes k+1}$ to $\mathcal{B}(\mathcal{D})$. By composition, the mapping $F_k : (x \mapsto A_k) \circ \pi_J$ is measurable from $\sigma(\pi_t : t \in J)$ to $\mathcal{B}(\mathcal{D})$. Now observe that $F_k(x)$ is the function constant equal to $x_{(i-1)/k}$ on each interval $[(i-1)/k, i/k)$ for $1 \leq i \leq k$, and with $x_1^k = x_1$. It is nearly the function y from the proof of Proposition 8.6, except that y was defined by replacing each value $x_{(i-1)/k}$ by a close one in a dense set of E . The proof there shows that $F_k(x) \rightarrow x$ in $\mathcal{D}$, so this limit, the identity, is indeed measurable from $\sigma(\pi_t : t \in J)$ to $\mathcal{B}(\mathcal{D})$, which means $\mathcal{B}(\mathcal{D}) \subset \sigma(\pi_t : t \in [0, 1])$. $\square$

Thanks to this theorem, if X and X^n are random variables in $\mathcal{D}$, then X_t and X_t^n are random variables in E for every $t \in [0, 1]$. In addition, if $X^n \rightarrow X$ in distribution, then by the continuous mapping theorem $X_t^n \rightarrow X_t$ in distribution whenever t belongs to the set

$$C_X = \{0\} \cup \{1\} \cup \{t \in (0, 1) : \mathbb{P}(X_t \neq X_{t-}) = 0\},$$

of times at which X is almost surely continuous. Recall that if $x \in \mathcal{D}$, then $\{t \in (0, 1) : x_t \neq x_{t-}\}$ is at most countable. The next lemma shows that, analogously, the complement of C_X is also at most countable.

LEMMA 8.10. For any random function $X \in \mathcal{D}$, the set $\{t \in (0, 1) : \mathbb{P}(X_t \neq X_{t-}) > 0\}$ is at most countable.

Proof. For every $\varepsilon > 0$ and every $t \in (0, 1)$, let $J_t(\varepsilon) = \{x \in \mathcal{D} : d_E(x_t, x_{t-}) > \varepsilon\}$. Recall that any function $x \in \mathcal{D}$ can only belong to $J_t(\varepsilon)$ for at most finitely many values of $t \in (0, 1)$. In particular, for any sequence $(t_n)_n$ of distinct values in $(0, 1)$, we have

$$0 = \mathbb{P}(X \in \limsup_{n \rightarrow \infty} J_{t_n}(\varepsilon)) \geq \inf_{n \geq 1} \mathbb{P}(X \in J_{t_n}(\varepsilon)).$$

Hence, for any $\varepsilon, \delta > 0$, there can only be at most finitely many distinct values of $t \in (0, 1)$ such that $\mathbb{P}(X \in J_t(\varepsilon)) \geq \delta$. We infer by letting $\varepsilon \downarrow 0$ that for any $\delta > 0$, there can only be at most countably many distinct values of $t \in (0, 1)$ such that $\mathbb{P}(X_t \neq X_{t-}) \geq \delta$. Letting further $\delta \downarrow 0$ yields our claim. $\square$

The analogue of Theorem 6.8 for proving weak convergence in $\mathcal{C}$ is then the following.

THEOREM 8.11 (Proving weak convergence in $\mathcal{D}$). Let X and $(X^n)_{n \geq 1}$ be random variables in $\mathcal{D}$. Then X^n converges in distribution to X in $\mathcal{D}$ if and only if both the sequence of the laws of X^n is tight and $(X_t^n)_{t \in J} \rightarrow (X_t)_{t \in J}$ for every finite subset $J \subset C_X$.

Proof. The direct implication is clear since convergence in distribution implies tightness in a Polish space and the continuous mapping theorem implies the convergence of the projections on any finite subset $J \subset C_X$. Let us prove the converse implication. Since tightness implies relative compactness by Prohorov's theorem 1.25 it remains to check that the second assumption characterises uniquely the limit of any convergent subsequence.

Suppose that $(Y^n)_n$ is a subsequence of $(X^n)_n$ that converges in distribution to some Y in $\mathcal{D}$. Fix any finite subset $J \subset C_X$, so $(Y_t^n)_{t \in J} \rightarrow (X_t)_{t \in J}$ in distribution. If $J \subset C_Y$, then by the continuous mapping theorem, we also have $(Y_t^n)_{t \in J} \rightarrow (Y_t)_{t \in J}$ in distribution. The distributions of $(X_t)_{t \in J}$ and $(Y_t)_{t \in J}$ thus coincide on every finite subset $J \subset C_X \cap C_Y$. By the previous lemma the complement of C_X is at most countable and so is that of C_Y , hence so is that of their intersection. In particular $C_X \cap C_Y$ is dense in $[0, 1]$, and it also contains both 0 and 1. We then infer from Theorem 8.9 that two distributions that agree on every finite subset $J \subset C_X \cap C_Y$ agree everywhere, so Y has the same law as X . $\square$

8.5 COMPACT SETS AND TIGHTNESS

In order to derive tightness criteria, we shall now characterise the relatively compact sets of $\mathcal{D}$ in a very similar way to those of $\mathcal{C}$ in the Arzelà–Ascoli theorem 6.12. Recall the modulus of continuity ω' from (8.1).

THEOREM 8.12 (Arzelà–Ascoli analogue in $\mathcal{D}$). A collection $\mathcal{K}$ of functions in $\mathcal{D}$ has compact closure if and only if

$$\lim_{\delta \rightarrow 0} \sup_{x \in \mathcal{K}} \omega'(x, \delta) = 0$$

and there exists a countable dense subset $Q \subset [0, 1]$ such that for every $t \in Q$, the set $\pi_t(\mathcal{K}) = \{x_t : x \in \mathcal{K}\}$ has compact closure in E .

Proof. To do! $\square$

As for Corollary 6.13, if E has the property that all bounded sets have compact closure, such as $E = \mathbb{R}^d$, then we may replace the second requirement by the existence of one value of $t \in [0, 1]$ such that $\pi_t(\mathcal{K})$ is bounded in E .

This result then yields the analogue of the tightness criterion in $\mathcal{C}$ in Theorem 6.16.

THEOREM 8.13. A collection $\mathcal{P}$ of stochastic processes in $\mathcal{D}$ is tight if and only if the following properties both hold:

- (i) there exists a countable and dense subset $Q \subset [0, 1]$ such that for every $t \in Q$, the collection of the laws of $\{X_t : X \in \mathcal{P}\}$ is tight in E and
- (ii) for every $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\sup_{X \in \mathcal{P}} \mathbb{P}(\omega'(X, \delta) > \eta) < \varepsilon.$$

If $\mathcal{P} = (X^n)_{n \geq 1}$ is a sequence, then we may replace the supremum in the second requirement by:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\omega'(X^n, \delta) > \eta) < \varepsilon.$$

Finally, if E has the property that all bounded sets have compact closure, then we may replace the first requirement by the existence of one $t \in [0, 1]$ such that $\{X_t : X \in \mathcal{P}\}$ is tight in E .

Proof. To do! $\square$

As for $\mathcal{C}$, further tightness criteria exist in $\mathcal{D}$ and we refer the reader to e.g. the books of Billingsley or Ethier & Kurtz.

8.6 EXERCISES

Recall that $\mathcal{D}$ denotes the set of càdlàg functions $x : [0, 1] \rightarrow E$, where E is a separable and complete metric space. For any function $x \in \mathcal{D}$, denote by

$$j(x) = \sup_{t \in (0,1]} d_E(x_t, x_{t-})$$

the magnitude of its largest jump.

EXERCISE 40. Recall the modified modulus of continuity from (8.1). Prove that for every $\delta \in (0, 1/2)$, we have

$$\omega'(x, \delta) \leq \omega(x, 2\delta) \leq 2\omega'(x, 2\delta) + j(x).$$

EXERCISE 41. Prove that d defined in (8.2) is a distance on $\mathcal{D}$.

EXERCISE 42. Let x and x^n be functions in $\mathcal{D}$. Prove the following properties.

- (i) If $x^n \rightarrow x$ uniformly, then also in the Skorokhod topology.
- (ii) If $x^n \rightarrow x$ in the Skorokhod topology, then $x_t^n \rightarrow x_t$ for every $t \in [0, 1]$ such that x is continuous at t (i.e. such that $d_E(x_t, x_{t-}) = 0$).
- (iii) If $x^n \rightarrow x$ in the Skorokhod topology and x is (uniformly) continuous on $[0, 1]$, then $x^n \rightarrow x$ uniformly.

EXERCISE 43. Prove that the largest jump j is continuous on $\mathcal{D}$. Consequently, the subset $\mathcal{C} = j^{-1}(\{0\})$ of continuous functions is closed in $\mathcal{D}$.

EXERCISE 44. Prove that d° in (8.3) is a distance on $\mathcal{D}$.

PART III

INFINITELY DIVISIBLE LAWS & LÉVY
PROCESSES

CHAPTER 9

INFINITELY DIVISIBLE DISTRIBUTIONS

Back to $\mathbb{R}^d$! This chapter characterises all the possible distributions in $\mathbb{R}^d$, beyond Gaussian distributions, that may appear as the limit of sums of n i.i.d. random variables, with a common distribution that may depend on n . We shall eventually end the chapter with a convergence theorem that provides necessary and sufficient conditions for such a convergence to hold, and to which limit law. This chapter is inspired by Breiman (extended to any finite dimension).

CONTENTS

9.1	Poisson convergence	96
9.2	Infinitely divisible distributions	98
9.3	Limits of compound Poisson distributions	101
9.4	The Lévy–Khintchine formula	104
9.5	A convergence criterion	106
9.6	Exercises	109

For each $n \geq 1$, let $(\xi_k^{(n)})_{k \geq 1}$ be i.i.d. random variables in $\mathbb{R}^d$, but whose law may vary with n , and let

$$S_n = \xi_1^{(n)} + \dots + \xi_n^{(n)} \quad \text{and further} \quad S_t^n = \xi_1^{(n)} + \dots + \xi_{[nt]}^{(n)}, \quad t \geq 0.$$

The central limit theorem proves that when the law of each $\xi_k^{(n)}$ is that of $\xi / \sqrt{n}$ where ξ is a random variable with (nonzero and) finite variance, then S_n converges to a Gaussian distribution. Donsker's theorem 7.2 extends this convergence to that of the processes to a Brownian motion. Our aim in this chapter is to describe all the possible limit laws for S_n . We shall then consider processes in the next chapter.

This part heavily relies on the characteristic functions.

9.1 POISSON CONVERGENCE

Let us start with a behaviour that is drastically different from the central limit theorem. Consider indeed the case where for each $n \geq 1$, the $(\xi_k^{(n)})_{k \geq 1}$ are i.i.d. with the Bernoulli distribution with parameter p_n such that np_n converges to some value $\lambda \in (0, \infty)$. Then S_n has the binomial law and it is straightforward to check that it converges to the Poisson distribution with parameter λ . By independence and stationarity of the increments of the walk, this extends to a finite-dimensional convergence: for every $0 \leq t_1 < \dots < t_k \leq 1$, we have

$$(S_{[nt_1]}^{(n)}, \dots, S_{[nt_k]}^{(n)}) \xrightarrow[n \rightarrow \infty]{(d)} (X_{t_1}, \dots, X_{t_k}),$$

where the increments $(X_{t_i} - X_{t_{i-1}})_{1 \leq i \leq k}$ (where $t_0 = 0$ and $X_0 = 0$) are independent and Poisson distributed with parameter $\lambda(t_i - t_{i-1})$ respectively.

Let us generalise the previous model as follows: let μ be a probability measure on $\mathbb{R}^d$, let Z have the law μ and independently, let ϵ_n have the Bernoulli law with parameter p_n such that $np_n \rightarrow \lambda \in (0, \infty)$. Then let

$(\xi_k^{(n)})_{k \geq 1}$ be i.i.d. with the same law as $\epsilon_n Z$. Let $\varphi_Z(u) = \mathbb{E}[e^{iuZ}]$ denote the characteristic function of Z . We compute the characteristic functions

$$\mathbb{E}[e^{iu\xi_1^{(n)}}] = (1 - p_n) + p_n \varphi_Z(u)$$

so

$$\mathbb{E}[e^{iuS_{[nt]}^{(n)}}] = (1 - p_n(1 - \varphi_Z(u)))^{[nt]} \xrightarrow{n \rightarrow \infty} e^{-\lambda t(1 - \varphi_Z(u))}.$$

The limit function is continuous at 0, so by Lévy's theorem 2.27 it is the characteristic function of some random variable X_t which is the limit in distribution of $S_{[nt]}^{(n)}$. Just as before, this extends to the joint convergence

$$(S_{[nt_1]}^{(n)}, \dots, S_{[nt_k]}^{(n)}) \xrightarrow[n \rightarrow \infty]{(d)} (X_{t_1}, \dots, X_{t_k}),$$

where the increments $(X_{t_i} - X_{t_{i-1}})_{1 \leq i \leq k}$ (where $t_0 = 0$ and $X_0 = 0$) are independent and distributed as $X_{t_i - t_{i-1}}$ respectively. We can be more explicit and describe the law of X_t as follows: recall that $\varphi_Z(u) = \mathbb{E}[e^{iuZ}]$, then by Fubini's theorem,

$$e^{-\lambda t(1 - \varphi_Z(u))} = \sum_{k \geq 0} \varphi_Z(u)^k \frac{(\lambda t)^k e^{-\lambda t}}{k!} = \mathbb{E}[\varphi_Z(u)^N] = \mathbb{E}[e^{iu \sum_{j=1}^N Z_j}],$$

where $(Z_j)_{j \geq 1}$ are i.i.d. with law μ and independent of N that has the Poisson distribution with parameter λt . These laws will turn out to be *very* important so we name them. They depend on the distribution μ and the value λ , which can alternatively view as a single parameter by considering the finite measure $\nu(\cdot) = \lambda \mu(\cdot)$.

DEFINITION 9.1. Let ν be a finite (and nonzero) measure on $\mathbb{R}^d$, let $\lambda = \nu(\mathbb{R}^d) > 0$ be its total mass, so $\mu(\cdot) = \lambda^{-1} \nu(\cdot)$ is a probability measure. The *compound Poisson* distribution with intensity ν is the law of $Z_1 + \dots + Z_N$ where:

- N has the Poisson distribution with parameter λ ,
- $(Z_n)_{n \geq 1}$ are i.i.d. with the law μ ,
- N and $(Z_n)_{n \geq 1}$ are independent.

We denote this law by $\text{CP}(\nu)$. Its characteristic function equals for every $u \in \mathbb{R}^d$:

$$\exp\left(\lambda \int (e^{iux} - 1) \mu(dx)\right) = \exp\left(\int (e^{iux} - 1) \nu(dx)\right).$$

Let us make some observations about this compound Poisson distribution and the approximation by $S^{(n)}$. First, for any Borel set $B \in \mathcal{B}(\mathbb{R}^d)$, by independence, the value $\nu(B) = \lambda \mu(B)$ can be seen in the distribution as:

$$\nu(B) = \mathbb{E}[N] \mathbb{P}(Z \in B) = \mathbb{E}[\#\{1 \leq k \leq N : Z_k \in B\}].$$

It is the average number of elements in the random sum $Z_1 + \dots + Z_N$ that fall in B . Similarly in the discrete random walk, we have

$$\mathbb{E}[\#\{1 \leq k \leq n : \xi_k^{(n)} \in B\}] = np_n \mu(B),$$

which converges to the expectation for the limit law.

Second, from the characteristic function, we see that if X and X' are independent compound Poisson random variables with intensity ν and ν' respectively, then for every $u \in \mathbb{R}^d$,

$$\mathbb{E}[e^{iu(X+X')}] = \exp\left(\int (e^{iux} - 1) \nu(dx)\right) \exp\left(\int (e^{iux} - 1) \nu'(dx)\right) = \exp\left(\int (e^{iux} - 1) (\nu + \nu')(dx)\right),$$

so $X + X'$ has the compound Poisson distribution with intensity $\nu + \nu'$.

Finally, if we let $(B_k)_{k \leq K}$ be a partition of $\mathbb{R}^d$ with Borel sets with $\nu(B_k) > 0$ for all $k \leq K$, then

$$\begin{aligned} \exp(-\lambda t(1 - \mathbb{E}[e^{iuZ}])) &= \prod_{k=1}^K \exp\left(-\lambda t(\mu(B_k) - \mathbb{E}[e^{iuZ} \mathbb{1}_{Z \in B_k}])\right) \\ &= \prod_{k=1}^K \exp\left(-\lambda \mu(B_k) t(1 - \mathbb{E}[e^{iuZ} | Z \in B_k])\right). \end{aligned}$$

Hence the compound Poisson distribution with intensity ν has the same law as the sum of independent compound Poisson distributions with intensity $\nu(\cdot \cap B_k)$. In a sense, the contributions of the Z_k 's in disjoint Borel sets do not interact: their number are Poisson distributed, weighted by the mass $\mu(B_k)$ of the set, and the values they take are distributed according to the law μ conditioned on B_k .

9.2 INFINITELY DIVISIBLE DISTRIBUTIONS

If $S_n = \xi_1^{(n)} + \dots + \xi_n^{(n)}$ does converge in distribution to some law, then we can expect this limit to be itself the law of the sum of n i.i.d. random variables, for every $n \geq 1$. We are going to see that it is indeed the case, and provide more insight, all thanks to the following lemma.

Instead of random variables, let us consider probability measures only. Let μ_n denote the common distribution of the $(\xi_k^{(n)})_{k \geq 1}$, so S_n has the law μ_n^{*n} of the n fold convolution. Let

$$\hat{\mu}_n(u) = \int_{\mathbb{R}^d} e^{iux} \mu_n(dx) = \mathbb{E}[e^{iu\xi_1^{(n)}}]$$

denote the characteristic function of μ_n , then the characteristic function of S_n is $\mathbb{E}[e^{iuS_n}] = \hat{\mu}_n(u)^n$.

LEMMA 9.2. Let $(\mu_n)_n$ be any sequence of probabilities on $\mathbb{R}^d$ and let $\hat{\mu}_n$ denote their respective characteristic functions. Then, we have for the pointwise convergence:

$$(\hat{\mu}_n)^n \text{ has a continuous limit } \phi \iff n(1 - \hat{\mu}_n) \text{ has a continuous limit } \psi$$

and in this case $\phi = \exp(-\psi)$ is the characteristic function of some probability μ , which is the weak limit of μ_n^{*n} , and the convergences hold uniformly on compact sets.

Proof. Since $(\hat{\mu}_n)^n$ is a characteristic function, then Lévy's continuity theorem 2.27 asserts that if it converges pointwise to a limit function which is continuous at 0, then this limit is the characteristic function of some probability μ , and then μ_n^{*n} converges weakly to μ , and finally the convergence in fact holds uniformly on compact sets (Theorem 2.26).

Let us turn to the equivalence. The easiest part is the converse implication. Suppose indeed that $n(1 - \hat{\mu}_n) \rightarrow \psi$. Then $\hat{\mu}_n \rightarrow 1$, so we may write it as $\hat{\mu}_n = \exp(-\psi_n)$ with $\psi_n \rightarrow 0$, and consequently $\hat{\mu}_n = 1 - \psi_n + o(\psi_n)$. Then $n(1 - \hat{\mu}_n) = n\psi_n(1 + o(1))$. By our assumption, the left-hand side converges to ψ , hence $n\psi_n \rightarrow \psi$, so finally

$$(\hat{\mu}_n)^n = \exp(-n\psi_n) \rightarrow \exp(-\psi).$$

Suppose conversely that $(\hat{\mu}_n)^n$ has a continuous limit ϕ . As mentioned above, ϕ must be the characteristic function of some distribution and the convergence holds uniformly on compact sets. By continuity, since $\phi(0) = 1$, then ϕ does not vanish in some compact neighbourhood of 0, say $U = [-\varepsilon, \varepsilon]^d$ with $\varepsilon > 0$. Let us restrict for the moment to this neighbourhood. That ϕ does not vanish ensures that there exists a priori a continuous function ψ on U that lifts it as $\phi = \exp(-\psi)$, and which is uniquely determined by the condition $\psi(0) = 0$. Since the convergence $(\hat{\mu}_n)^n \rightarrow \phi$ holds uniformly in U , then $\hat{\mu}_n$ does not vanish in it for n large enough, and hence takes the form $\hat{\mu}_n = \exp(-\psi_n)$ for a unique

continuous function ψ_n with $\psi_n(0) = 0$. Further, we have $n\psi_n \rightarrow \psi$ (uniformly) in U , from which it follows (take the modulus) that $\psi_n \rightarrow 0$ and finally that $n(1 - \hat{\mu}_n) \sim n\psi_n \rightarrow \psi$.

In order to conclude about the direct implication, it remains to prove that the limit ϕ never vanishes, so the previous limits hold everywhere. Recall that ϕ is the characteristic function of some random variable X . Then if X' has the same law as X and is independent of it, the difference $X - X'$ has characteristic function $|\phi|^2$. Similarly $|\hat{\mu}_n|^2$ is the characteristic function of the symmetrised random variable with characteristic function $\hat{\mu}_n$. Hence, using this symmetry if necessary, we may assume that $\hat{\mu}$ and $\hat{\mu}_n$ are real-valued and nonnegative (and bounded by 1). Let $\varepsilon > 0$ be such that ϕ does not vanish in $[-\varepsilon, \varepsilon]^d$ and let us prove that it does not vanish in $[-2\varepsilon, 2\varepsilon]^d$. Then by induction it never vanishes.

The convergence $(\hat{\mu}_n)^n \rightarrow \phi$ uniformly on compact sets and the fact that ϕ does not vanish in $[-\varepsilon, \varepsilon]^d$ imply that $n(1 - \hat{\mu}_n) \rightarrow \psi$ uniformly in this interval so in particular $n(1 - \hat{\mu}_n)$ is bounded uniformly over n in this interval. Observe that for any random variable X with a symmetric distribution, its characteristic function equals $\mathbb{E}[e^{iuX}] = \mathbb{E}[\cos(uX)]$ and since $1 - \cos(2x) = 1 - \cos^2 x + \sin^2 x = 2(1 - \cos^2 x) = 2(1 + \cos x)(1 - \cos x) \leq 4(1 - \cos x)$, then

$$1 - \mathbb{E}[e^{i2uX}] = 1 - \mathbb{E}[\cos(2uX)] \leq 4(1 - \mathbb{E}[\cos(uX)]) = 4(1 - \mathbb{E}[e^{iuX}]).$$

In our case, we infer that

$$n(1 - \hat{\mu}_n(2u)) \leq 4n(1 - \hat{\mu}_n(u)).$$

Since the right side is bounded in $[-\varepsilon, \varepsilon]^d$, then so is the left side, namely $n(1 - \hat{\mu}_n)$ is bounded in $[-2\varepsilon, 2\varepsilon]^d$ uniformly in n . Consequently $\phi = \lim_n \exp(-n(1 - \hat{\mu}_n))$ stays bounded away from 0 in $[-2\varepsilon, 2\varepsilon]^d$ and this completes the proof. $\square$

Let us derive important consequences of this lemma. Recall for this the compound Poisson distribution with intensity a finite measure $\nu = \lambda\mu$ where μ is a probability, denoted by $\text{CP}(\nu)$, whose characteristic function is $\exp(-\lambda(1 - \hat{\mu}))$.

THEOREM 9.3. Let μ be a probability measure on $\mathbb{R}^d$. Then the following assertions are equivalent:

- (i) There exists a sequence of probability measures $(\mu_n)_{n \geq 1}$ such that $\mu = \mu_n^{*n}$ for every $n \geq 1$.
- (ii) There exists a sequence of probability measures $(\nu_n)_{n \geq 1}$ such that μ is the weak limit of ν_n^{*n} .
- (iii) With the same sequence $(\nu_n)_{n \geq 1}$ the compound Poisson distribution $\text{CP}(n\nu_n)$ converges weakly to μ .

In addition, when this holds, the sequence $(\mu_n)_{n \geq 1}$ in the first item is unique, and the characteristic function of μ never vanishes and takes the form $\hat{\mu} = \exp(-\psi)$ for a unique continuous function ψ with $\psi(0) = 0$.

DEFINITION 9.4. A probability measure μ that can be written in the form μ_n^{*n} for every $n \geq 1$ for some probability measures μ_n is called *infinitely divisible*. The function ψ such that $\hat{\mu} = \exp(-\psi)$ is called its *characteristic exponent*.^a

^aBeware some authors write $\hat{\mu} = \exp(\psi)$ and call this ψ the characteristic exponent of μ .

Proof of Theorem 9.3. Let us first prove the last claims before turning our attention to the equivalence of the three items. Suppose that μ is infinitely divisible and write it as $\mu = \mu_n^{*n}$ for every $n \geq 1$ where μ_n is a probability. Then $\hat{\mu} = (\hat{\mu}_n)^n$ and Lemma 9.2 (and its proof) applied with $\phi = \hat{\mu}$ shows that $\hat{\mu}$ never vanishes and takes the form $\hat{\mu} = \exp(-\psi)$ for a unique continuous function ψ with $\psi(0) = 0$. In addition, these laws μ_n are unique. Indeed, start again from $\hat{\mu} = (\hat{\mu}_n)^n$. Since $\hat{\mu}$ does not vanish then $\hat{\mu}_n$ does not

either so it can as well be written as $\hat{\mu}_n = \exp(-\psi_n)$ for some continuous function ψ_n that vanishes at 0. Further we have the identity $\exp(-\psi) = \exp(-n\psi_n)$. Since both ψ and $n\psi_n$ are continuous and null at 0, then $\psi_n = \psi/n$. This determines uniquely the law μ_n .

Now the three items. Clearly (i) $\implies$ (ii) with $\nu_n = \mu_n$. Let us prove that (ii) $\implies$ (iii). The weak convergence $\nu_n^{*n} \Rightarrow \mu$ implies $(\hat{\nu}_n)^n \rightarrow \hat{\mu}$, which by Lemma 9.2 implies that $n(1 - \hat{\nu}_n) \rightarrow \psi$ and further that $\exp(-n(1 - \hat{\nu}_n)) \rightarrow \exp(-\psi) = \hat{\mu}$, that is, that $\text{CP}(n\nu_n) \Rightarrow \mu$.

We may now prove that (iii) $\implies$ (i). Suppose indeed that $\text{CP}(n\nu_n) \Rightarrow \mu$. From the point of view of characteristic functions, we read $\exp(-n(1 - \hat{\nu}_n)) \rightarrow \hat{\mu}$. Then for any $k \geq 1$, this implies that $\hat{\mu}^{1/k} = \lim_n \exp(-\frac{n}{k}(1 - \hat{\nu}_n))$, but this is the characteristic exponent of the compound Poisson distribution with intensity $\frac{n}{k}\nu_n$. Since the limit $\hat{\mu}^{1/k}$ is continuous then Lévy's theorem 2.27 implies that it is the characteristic function of some probability μ_k . Then μ_k^{*k} has characteristic function $\hat{\mu}$ so indeed $\mu = \mu_k^{*k}$ for every $k \geq 1$. $\square$

EXERCISE 9.5. Suppose that $S_n = \xi_1^{(n)} + \dots + \xi_n^{(n)}$ converges in distribution to some X , where $(\xi_i^{(n)})_{i \geq 1}$ are i.i.d. Fix $k \geq 2$. For every $j \geq 1$, consider $n = jk$: write $S_{jk} = Y_{j,1} + \dots + Y_{j,k}$ for some i.i.d. random variables $Y_{j,1}, \dots, Y_{j,k}$. Using that $(S_{jk})_{j \geq 1}$ is tight, prove that $(Y_{j,i})_{j \geq 1}$ is tight as well and deduce that X has the same law as $Y_1 + \dots + Y_k$ for some i.i.d. random variables $Y_1, \dots, Y_k$. This gives a short proof that limits in distribution of sums of i.i.d. random variables are exactly the infinitely divisible laws.

REMARK 9.6. If a sequence of infinitely divisible distribution converges weakly, then the limit is infinitely divisible as well: simply write the n 'th term of such a sequence in the form ν_n^{*n} and use (ii) $\implies$ (i) in the theorem.

REMARK 9.7. If μ is infinitely divisible and can be decomposed as $\mu = \mu_n^{*n}$ for every $n \geq 1$, with probability measures μ_n , then using these μ_n as ν_n in (ii), we deduce from (iii) that μ is the weak limit of the compound Poisson distribution $\text{CP}(n\mu_n)$.

REMARK 9.8. In terms of random variables, that (ii) $\iff$ (iii) shows that if for every $n, k \geq 1$, we let $S_k^{(n)} = \xi_1^{(n)} + \dots + \xi_k^{(n)}$ be the sum of k i.i.d. random variables and if, independently, we let $N(n)$ have the Poisson distribution with parameter n , then $S_{N(n)}^{(n)}$ converges in distribution if and only if $S_{N(n)}^{(n)}$ does, and with the same limit. Replacing a deterministic parameter by an independent random Poisson number is a powerful technique called *Poissonisation*; it is often the case that the Poissonised version is simpler to analyse (with convenient independence properties for example), and going from a random parameter to a deterministic one can be challenging.

This theorem shows the importance of infinitely divisible distributions since, not only are they the laws of sums of i.i.d. random variables, but more generally they are exactly the possible limits of such sums, in a generalisation of the central limit theorem. It also provides a strategy to describe all infinitely divisible distributions by searching to express all the possible limits of compound Poisson distributions. Finally, it also provides a criterion for the convergence of sums of i.i.d. random variables, by considering the associated compound Poisson distributions. We shall develop these ideas in this chapter.

REMARK 9.9. If μ_n^{*n} converges weakly to some law μ , then by Lemma 9.2 we have $\hat{\mu}_n \rightarrow 1$, that is, μ_n converges weakly to the Dirac mass at 0. In terms of random variables, if $S_n = \xi_1^{(n)} + \dots + \xi_n^{(n)}$ has a limit in distribution, then $\xi_1^{(n)}$ (and thus any $\xi_k^{(n)}$) converges in probability to 0.

9.3 LIMITS OF COMPOUND POISSON DISTRIBUTIONS

Not all distributions are infinitely divisible and for example it should be clear that a Bernoulli law cannot be the law of even just the sum of two i.i.d. random variables: how can such a sum assume only the values 0 and 1? More generally, Exercise 45 claims if a random variable is almost surely bounded and has an infinitely divisible distribution, then it is almost surely constant.

In order to describe all infinitely divisible distributions, we rely on a consequence of Theorem 9.3 that every infinitely divisible distribution is the weak limit of compound Poisson distribution with intensity $n\mu_n$ where each μ_n is the probability such that $\mu = \mu_n^{*n}$.

Fix henceforth a sequence $(\mu_n)_{n \geq 1}$ of probability measures, assume that the compound Poisson distribution $\text{CP}(n\mu_n)$ converges weakly to some μ , and let us see what form takes the function ψ in $\hat{\mu} = \exp(-\psi)$. The starting point is thus the characteristic function of the law $\text{CP}(n\mu_n)$ itself, which is given by $\exp(-\psi_n)$ where

$$\psi_n(u) = n(1 - \hat{\mu}_n(u)) = \int_{\mathbb{R}^d} (1 - e^{iux}) n\mu_n(dx), \quad u \in \mathbb{R}^d. \quad (9.1)$$

According to Lemma 9.2, we have $\psi_n \rightarrow \psi$ uniformly on compact sets.

We are guided by a first intuition. Recall that μ is also the limit of the sum of n i.i.d. random variables with law μ_n , then the number of terms in this sum whose norm exceeds any fixed $K > 0$ should stay finite, otherwise the sum could go to infinity. Only the small increments accumulate. This intuition is true and we can also control these small increments.

LEMMA 9.10. If $\text{CP}(n\mu_n)$ converges weakly then the following bounds hold:

$$\limsup_{n \rightarrow \infty} n\mu_n(\mathbb{R}^d \setminus [-K, K]^d) < \infty \quad \text{for all } K > 0$$

and it further converges to 0 as $K \rightarrow \infty$. In addition,

$$\limsup_{n \rightarrow \infty} n \int_{[-1,1]^d} |x|^2 \mu_n(dx) < \infty.$$

Proof. To simplify the notation, let us denote by $C_r = [-r, r]^d$ the d -dimensional cube with side $2r$.

The first limsup is bounded using a d -dimensional extension of the bound used in the proof of Lévy's continuity theorem 2.27. Precisely, let X be any random variable in $\mathbb{R}^d$ and let $\delta > 0$ be arbitrary, then, by Fubini's theorem,

$$\int_{C_\delta} \cos(uX) du = \mathbb{E} \left[\prod_{k=1}^d \int_{-\delta}^{\delta} \cos(u_k X_k) du_k \right] = \mathbb{E} \left[\prod_{k=1}^d 2\delta \frac{\sin(\delta X_k)}{\delta X_k} \right].$$

Observe that $|\sin(t)/t| \leq 1$ for all $t \in \mathbb{R}$ and $|\sin(t)/t| \leq 1/2$ for all $|t| \geq 2$ so

$$\frac{1}{(2\delta)^d} \int_{C_\delta} (1 - \cos(uX)) du = \mathbb{E} \left[1 - \prod_{k=1}^d \frac{\sin(\delta X_k)}{\delta X_k} \right] \geq \mathbb{E} \left[\left(1 - \prod_{k=1}^d \frac{\sin(\delta X_k)}{\delta X_k} \right) \mathbb{1}_{X \notin C_{2/\delta}} \right] \geq \frac{1}{2} \mathbb{P}(X \notin C_{2/\delta}).$$

We deduce in our case that

$$n\mu_n(C_K^c) \leq \frac{2n}{(4/K)^d} \int_{C_{2/K}} (1 - \text{Re } \hat{\mu}_n(u)) du.$$

Recall from Lemma 9.2 that $n(1 - \hat{\mu}_n) \rightarrow \psi$ uniformly on compact sets, so after taking the real part, the right-side above converges to the finite limit

$$\frac{2}{(4/K)^d} \int_{C_{2/K}} \text{Re } \psi(u) du,$$

which further tends to 0 as $K \rightarrow \infty$ since $\psi(0) = 0$.

For the second bound, observe that $\cos(1x) \leq 1$ for all $x \in \mathbb{R}^d$ and there exists a constant $a > 0$ such that $\cos(1x) \leq 1 - a|x|^2$ for all $x \in C_1$. It follows that

$$an \int_{C_1} |x|^2 \mu_n(dx) \leq n \int_{C_1} (1 - \cos(1x)) \mu_n(dx) \leq n \int_{\mathbb{R}^d} (1 - \cos(1x)) \mu_n(dx) = n(1 - \operatorname{Re} \hat{\mu}_n(1)),$$

which converges to $1 - \operatorname{Re} \psi(1) < \infty$. $\square$

Note that in the characteristic exponent (9.1) of our compound Poisson law, the integrand is a continuous and bounded function, but it behaves only linearly near 0, and not as a, much smaller, quadratic function. The idea then is to subtract the linear term and write for all $u \in \mathbb{R}^d$:

$$\psi_n(u) = \int_{\mathbb{R}^d} (1 + iuh(x) - e^{iux}) n\mu_n(dx) - iu \int_{\mathbb{R}^d} h(x) n\mu_n(dx), \quad (9.2)$$

where $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a continuous and bounded function such that $h(x) = x + o(|x|^2)$ at $x = 0$. A typical example is $h(x) = x/(1+|x|^2)$; in dimension $d = 1$, we also often find $h(x) = \sin x$. In this way the integrand in the first integral on the right remains continuous and bounded, but now it behaves as $(ux)^2/2 = O(|x|^2)$ as $x \rightarrow 0$.

Define a measure ρ_n with a density $|x|^2/(1+|x|^2)$ with respect to $n\mu_n$, namely for every f measurable and bounded on $\mathbb{R}^d$, we have

$$\int_{\mathbb{R}^d} f(x) \rho_n(dx) = \int_{\mathbb{R}^d} f(x) \frac{|x|^2}{1+|x|^2} n\mu_n(dx).$$

Lemma 9.10 implies that $\limsup \rho_n(\mathbb{R}^d) < \infty$. Suppose that, along some subsequence, we have $\rho_n(\mathbb{R}^d) \rightarrow 0$. Then along this subsequence, we also have

$$\int_{\mathbb{R}^d} (1 + iuh(x) - e^{iux}) n\mu_n(dx) \xrightarrow{n \rightarrow \infty} 0.$$

Then, passing to the limit in (9.2), we obtain:

$$\psi(u) = -iu\Gamma \quad \text{for some} \quad \Gamma \in \mathbb{R}^d.$$

This corresponds to $\mu = \delta_\Gamma$, which is indeed infinitely divisible.

We henceforth exclude this case and only restrict to subsequences along which $\lambda_n = \rho_n(\mathbb{R}^d)$ converges to some limit $\lambda \in (0, \infty)$. Then, for n large enough, the finite measure ρ_n can be normalised into a probability measure $\eta_n(\cdot) = \rho_n(\cdot)/\lambda_n$. Lemma 9.10 shows that the sequence $(\eta_n)_{n \geq 1}$ is tight so we may extract a further sub-subsequence along which it converges weakly to some probability η . Below all limits are taken along this subsequence.

Then for any function f continuous and bounded, it holds along this subsequence:

$$\int_{\mathbb{R}^d} f(x) \frac{|x|^2}{1+|x|^2} n\mu_n(dx) = \int_{\mathbb{R}^d} f(x) \lambda_n \eta_n(dx) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) \lambda \eta(dx).$$

Fix $u \in \mathbb{R}^d$ and recall $h(x) = x + o(|x|^2)$ at 0; in dimension $d = 1$, we may take f as the real and imaginary part respectively of:

$$(1 + iuh(x) - e^{iux}) \frac{1 + |x|^2}{|x|^2},$$

which is well-defined for $x \neq 0$ and extended by continuity at $x = 0$ with the value $\lim_{x \rightarrow 0} (ux)^2/(2|x|^2) = u^2/2$. This fails in higher dimension and the previous function is not continuous at $x = 0$. To deal with this, we subtract one more term and write for all $u \in \mathbb{R}^d$:

$$\psi_n(u) = \int_{\mathbb{R}^d} \left(1 + iuh(x) - \frac{(uh(x))^2}{2} - e^{iux} \right) n\mu_n(dx) + \int_{\mathbb{R}^d} \frac{(uh(x))^2}{2} n\mu_n(dx) - iu \int_{\mathbb{R}^d} h(x) n\mu_n(dx). \quad (9.3)$$

Then now the integrand of the first integral is continuous, bounded and $o(|x|^2)$ at 0 so

$$\int_{\mathbb{R}^d} \left(1 + iuh(x) - \frac{(uh(x))^2}{2} - e^{iux} \right) n\mu_n(dx) = \int_{\mathbb{R}^d} \left(1 + iuh(x) - \frac{(uh(x))^2}{2} - e^{iux} \right) \frac{1 + |x|^2}{|x|^2} \rho_n(dx),$$

and the integrand on the right is continuous (equal to 0 at $x = 0$) and bounded. Along our subsequence, this integral therefore converges to the same one against the measure $\lambda\eta$.

Let us finally consider the other two integrals in (9.3). Let $h(x)^\top$ denote the transpose vector of $h(x)$ and define a $d \times d$ matrix Λ_n and a vector $\Gamma_n \in \mathbb{R}^d$ respectively by:

$$\Lambda_n = \int_{\mathbb{R}^d} h(x)h(x)^\top n\mu_n(dx), \quad \text{and} \quad \Gamma_n = \int_{\mathbb{R}^d} h(x)n\mu_n(dx).$$

Then (9.3) reads

$$\psi_n(u) - \int_{\mathbb{R}^d} \left(1 + iuh(x) - \frac{(uh(x))^2}{2} - e^{iux} \right) n\mu_n(dx) = \frac{1}{2}(\Lambda_n u)u - iu\Gamma_n,$$

and the left side converges for every $u \in \mathbb{R}^d$ along the previous subsequence. Since the right side is a degree 2 polynomial in u , then so is its limit (possibly a degenerate one). We conclude that there exist a vector $\Gamma \in \mathbb{R}^d$ and a $d \times d$ matrix Λ such that, along the same subsequence as above,

$$\Lambda_n \xrightarrow{n \rightarrow \infty} \Lambda \quad \text{and} \quad \Gamma_n \xrightarrow{n \rightarrow \infty} \Gamma.$$

We conclude by passing to the limit along this subsequence in (9.3) that

$$\begin{aligned} \psi(u) &= \lim_{n \rightarrow \infty} \psi_n(u) = \int_{\mathbb{R}^d} \left(1 + iuh(x) - \frac{(uh(x))^2}{2} - e^{iux} \right) \frac{1 + |x|^2}{|x|^2} \lambda\eta(dx) + \frac{1}{2}(\Lambda u)u - iu\Gamma \\ &= \int_{\mathbb{R}^d} (1 + iuh(x) - e^{iux}) \pi(dx) + \frac{1}{2}(\tilde{\Lambda} u)u - iu\Gamma, \end{aligned} \quad (9.4)$$

where π is the measure with density $\lambda \frac{1+|x|^2}{|x|^2} \mathbb{1}_{x \neq 0}$ with respect to η (recall that the integrand in the first line vanishes at $x = 0$ so we may as well integrate over $\mathbb{R}^d \setminus \{0\}$) and then $\tilde{\Lambda}$ is the $d \times d$ matrix given by $\tilde{\Lambda} = \Lambda - \int_{\mathbb{R}^d} h(x)h(x)^\top \pi(dx)$.

A natural question following these calculations is: conversely, does any ψ of the form (9.4) correspond to an infinitely divisible distribution μ via $\hat{\mu} = \exp(-\psi)$? Some conditions on the measure π and the matrix $\tilde{\Lambda}$ are needed. Indeed, we can see some necessary conditions already. First notice that, since η is a probability measure and $0 < \lambda < \infty$, then π must enjoy the properties:

$$\pi(\{0\}) = 0, \quad \text{and both} \quad \pi(\mathbb{R}^d \setminus [-r, r]^d) < \infty \quad \text{and} \quad \int_{[-r, r]^d} |x|^2 \pi(dx) < \infty \quad \text{for all } r > 0.$$

As for the matrix $\tilde{\Lambda}$, by construction Λ_n is symmetric (meaning $\Lambda_n^\top = \Lambda_n$) and positive (meaning $(\Lambda_n u)u \geq 0$ for every $u \in \mathbb{R}^d$). Then so is Λ after taking the limit, as well as $\int_{\mathbb{R}^d} h(x)h(x)^\top \pi(dx)$. Then their difference $\tilde{\Lambda}$ is symmetric and we claim that it is positive as well. Indeed, fix $u \in \mathbb{R}^d$, then, along a subsequence,

$$(\tilde{\Lambda} u)u = \left(\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} (uh(x))^2 n\mu_n(dx) \right) - \int_{\mathbb{R}^d} (uh(x))^2 \pi(dx).$$

Using the finite measures ρ_n and ρ as above, recall that $\int f n d\mu_n$ converges to $\int f d\pi$ for every function f which is continuous, bounded and $f(x) = o(|x|^2)$ as $x \rightarrow 0$. This is not the case of $f(x) = (uh(x))^2$ but for every $k \geq 1$, it is the case of $f_k(x) = i_k(x)(uh(x))^2$ where $i_k : \mathbb{R}^d \rightarrow [0, 1]$ is continuous, constant equal to 1 on the set $\{x \in \mathbb{R}^d : |x| \geq 2/k\}$ and constant equal to 0 on the set $\{x \in \mathbb{R}^d : |x| \leq 1/k\}$. Note that $i_k(x) \uparrow 1$ for every $x \neq 0$. Since $\pi(\{0\}) = 0$, then for every $k \geq 1$, it holds

$$\liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d} (uh(x))^2 n\mu_n(dx) \geq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}^d \setminus \{0\}} f_k n d\mu_n = \int_{\mathbb{R}^d} f_k d\pi \xrightarrow{k \rightarrow \infty} \int_{\mathbb{R}^d} (uh(x))^2 \pi(dx),$$

by monotone convergence. Hence $\tilde{\Lambda}$ is symmetric positive. Recall that any such matrix can be decomposed as $\tilde{\Lambda} = \Sigma^2$ for some symmetric matrix Σ . Then $(\tilde{\Lambda} u)u = (\Sigma^2 u)u = (\Sigma u)^2$.

9.4 THE LÉVY–KHINTCHINE FORMULA

The main result of this section gives a complete characterisation of infinitely divisible distributions by showing that the previous necessary conditions on the measure π and the matrix Σ are sufficient to ensure that (9.4) defines such a distribution. Let us go over these conditions again.

First, let us fix a *truncation function* $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$, which is measurable, but not necessarily continuous, bounded, and such that $h(x) = x + o(|x|^2)$ near 0; the most common choice in the literature nowadays is certainly

$$h(x) = x \mathbb{1}_{|x| < 1},$$

but the function $h(x) = x/(1 + |x|^2)$ that we used previously has the advantage of being continuous, just as $h(x) = \sin x$ in dimension $d = 1$. Next, let $\Gamma \in \mathbb{R}^d$ be arbitrary, let Σ^2 be a symmetric positive $d \times d$ matrix, and finally, let π be a σ -finite measure on $\mathbb{R}^d$ that satisfies:

$$\pi(\{0\}) = 0 \quad \text{and} \quad \int (|x|^2 \wedge 1) \pi(dx) < \infty.$$

Such a measure is called a *Lévy measure*.

THEOREM 9.11 (Lévy–Khintchine). Fix a truncation function h . A probability measure μ is infinitely divisible if and only if there exists a triple (Γ, Σ^2, π) as above such that the characteristic function $\hat{\mu}$ takes the form:

$$\hat{\mu}(u) = \exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2 + \int_{\mathbb{R}^d} (e^{iux} - 1 - iuh(x))\pi(dx)\right), \quad (9.5)$$

for every $u \in \mathbb{R}^d$. In addition, this triple is unique.

On a side note, the assumption $\pi(\{0\}) = 0$ is only to guaranty uniqueness in the theorem, since adding a constant times δ_0 to π does not affect (9.5).

REMARK 9.12. Let $\psi_{(\Gamma, \Sigma^2, \pi, h)}(u)$ be such that $\hat{\mu}(u) = \exp(-\psi_{(\Gamma, \Sigma^2, \pi, h)}(u))$ in (9.5), namely:

$$\psi_{(\Gamma, \Sigma^2, \pi, h)}(u) = -i\Gamma u + \frac{1}{2}(\Sigma u)^2 - \int_{\mathbb{R}^d} (e^{iux} - 1 - iuh(x))\pi(dx). \quad (9.6)$$

- (i) The integrability condition on π implies that for any neighbourhood V of 0, its complement has finite mass. Indeed, let $C = 1 \vee \sup\{|x|^{-2} : x \in V^c\} < \infty$, then $\mathbb{1}_{x \in V^c} \leq C(|x|^2 \wedge 1)$ and so

$$\pi(V^c) \leq C \int (|x|^2 \wedge 1) d\pi < \infty.$$

The total mass of π may however be infinite due to the possibility of $\pi(V) = \infty$ for every open neighbourhood V of 0.

- (ii) This implies further that for every $u \in \mathbb{R}^d$, the integral in (9.6) is well-defined since the integrand is bounded and is $O(|x|^2)$ near 0. By dominated convergence, the function $\psi_{(\Gamma, \Sigma^2, \pi, h)}$ is a continuous function.
- (iii) Finally, if we replace h by another such function $\tilde{h}$, then $h(x) - \tilde{h}(x) = o(|x|^2)$ near 0 and is bounded on $\mathbb{R}^d$ so it is π -integrable and hence

$$\psi_{(\Gamma, \Sigma^2, \pi, h)}(u) = \psi_{(\Gamma, \Sigma^2, \pi, \tilde{h})}(u) - iu \int (\tilde{h} - h) d\pi = \psi_{(\tilde{\Gamma}, \Sigma^2, \pi, \tilde{h})}(u),$$

where $\tilde{\Gamma} = \Gamma + \int (\tilde{h} - h) d\pi$ but neither Σ^2 nor π are affected.

We have shown the (difficult) part that any infinitely divisible distribution assumes the form (9.5) for such a triple. Let us now prove conversely that any such triple gives rise to an infinitely divisible distribution. Uniqueness of the triple is not trivial and requires a proof, but we leave that for later. Indeed, analytical proofs in the same vein as the previous section exist, but it will be an immediate corollary of a forthcoming important result in the next chapter.

Proof of existence of infinitely divisible distributions. Let us fix the truncation function $h(x) = x \mathbb{1}_{|x| < 1}$; recall that another choice only changes the value of Γ . Fix a triple (Γ, Σ^2, π) and let us prove that there exists an infinitely divisible distribution μ with characteristic function (9.5). First

$$\exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2\right)$$

is the characteristic function of a Gaussian vector X with mean vector Γ and covariance matrix Σ^2 . Let $\pi_1(\cdot) = \pi(\cdot \cap \{|x| \geq 1\})$, which is a finite measure. Let Y have the compound Poisson distribution with intensity π_1 , which has characteristic function:

$$\exp\left(\int (e^{iux} - 1) \mathbb{1}_{|x| \geq 1} \pi(dx)\right).$$

It remains to consider the unit ball. For every $n \geq 2$, let $\pi_n(\cdot) = \pi(\cdot \cap \{1/n \leq |x| < 1\})$, which is again a finite measure, and let Y_n be a compound Poisson random variable with intensity π_n . Recall that Y_n has the same law as $\sum_{i=1}^{N_n} \xi_{n,i}$ where N_n has the Poisson distribution with parameter $\pi_n(\mathbb{R}^d)$ and, independently, the $(\xi_{n,i})_{i \geq 1}$ are i.i.d. with the law $\pi_n(\cdot)/\pi_n(\mathbb{R}^d)$. Note that $1/n \leq |\xi_{n,i}| < 1$, so in particular Y_n is integrable and we have

$$\mathbb{E}[Y_n] = \mathbb{E}[N_n] \mathbb{E}[\xi_{n,1}] = \pi_n(\mathbb{R}^d) \int x \frac{\pi_n(dx)}{\pi_n(\mathbb{R}^d)} = \int x \mathbb{1}_{1/n \leq |x| < 1} \pi(dx).$$

Define then $Z_n = Y_n - \int x \mathbb{1}_{1/n \leq |x| < 1} \pi(dx)$, which has characteristic function:

$$\exp\left(\int (e^{iux} - 1 - iux) \mathbb{1}_{1/n \leq |x| < 1} \pi(dx)\right).$$

By dominated convergence, this converges to the integral over the entire unit ball as $n \rightarrow \infty$. Since this limit is continuous at 0, then Lévy's theorem 2.27 shows that it is the characteristic function of a random variable Z which is the limit in distribution of Z_n . Since each Z_n is infinitely divisible, then so is Z .

To conclude, let the Gaussian vector X , the compound Poisson Y and the limit Z be independent, then their sum $X + Y + Z$ is infinitely divisible and has characteristic function precisely (9.5). $\square$

REMARK 9.13. We saw in the proof that a Gaussian vector with mean vector Γ and covariance matrix Σ^2 satisfied the Lévy–Khintchine formula with triple $(\Gamma, \Sigma^2, 0)$. As for compound Poisson distribution with finite intensity measure ν , the characteristic function is

$$\exp\left(\int (e^{iux} - 1) \nu(dx)\right) = \exp\left(iu \int h d\nu + \int (e^{iux} - 1 - iuh(x)) \nu(dx)\right),$$

so the corresponding triple is $(\int h d\nu, 0, \nu)$.

REMARK 9.14. If the measure π satisfies the stronger integrability assumption $\int (|x| \wedge 1) \pi(dx) < \infty$ instead of only a square, then both $e^{iux} - 1$ and $uxh(x)$ become integrable and we may rewrite the

Lévy–Khintchine formula as:

$$\hat{\mu}(u) = \exp\left(iu\left(\Gamma - \int h \, d\pi\right) - \frac{1}{2}(\Sigma u)^2 + \int (e^{iux} - 1)\pi(dx)\right), \quad u \in \mathbb{R}^d.$$

Beware that this is *not* the sum of a compound Poisson distribution and an independent Gaussian vector, unless π has finite mass.

9.5 A CONVERGENCE CRITERION

Recall that we motivated the study of infinitely divisible distributions by the fact, according to Theorem 9.3, that they appear as the possible limit laws of sums $S_n = \sum_{i=1}^n \xi_{n,i}$ where for each $n \geq 1$, the $(\xi_{n,i})_{i \geq 1}$ are i.i.d. random variables, with some common distribution μ_n that may depend on n . Let us denote by ξ_n a generic random variable with law μ_n .

THE GENERAL CASE

Now that the Lévy–Khintchine formula (9.5) characterises infinitely divisible distributions using a unique triple,¹ we have access to conditions on μ and μ_n for this μ to appear at the limit of μ_n^{*n} . When dealing with weak convergence, it is important to use the Lévy–Khintchine formula with a *continuous* truncation function h , which we recall is also bounded and such that $h(x) = x + o(|x|^2)$ near 0.

THEOREM 9.15. With the above notation, the sequence S_n converges in distribution to some limit μ if and only if μ is infinitely divisible with associated Lévy triple (Γ, Σ^2, π) for a given *continuous* truncation function h and the following convergences hold true:

- (i) $n \mathbb{E}[h(\xi_n)] \rightarrow \Gamma$ in $\mathbb{R}^d$,
- (ii) $\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} |n \mathbb{E}[(u\xi_n)^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] - (\Sigma u)^2| = 0$ for each $u \in \mathbb{R}^d$,
- (iii) $n \mathbb{E}[f(\xi_n)] \rightarrow \int f \, d\pi$ for any continuous and bounded function f with $f(x) = o(|x|^2)$ as $x \rightarrow 0$.

Let us start with the proof that these three items imply that S_n converges in distribution to the associated infinitely divisible law, which is the useful part in practice.

Proof of the sufficiency. Suppose that all three items hold true and let μ have characteristic function

$$\begin{aligned} \hat{\mu}(u) &= \exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2 + \int_{\mathbb{R}^d} (e^{iux} - 1 - iuh(x))\pi(dx)\right) \\ &= \exp\left(i\Gamma u - \frac{1}{2}\left((\Sigma u)^2 + \int_{\mathbb{R}^d} (uh)^2 \, d\pi\right) + \underbrace{\int_{\mathbb{R}^d} \left(e^{iux} - 1 - iuh(x) - \frac{(uh(x))^2}{2}\right)\pi(dx)}_{= f_u(x)}\right). \end{aligned}$$

Recall from Theorem 9.3 that S_n converges weakly to μ if and only if the compound Poisson distribution with intensity $n\mu_n$ converges weakly to the same limit μ . Remark 9.13 specialises the Lévy–Khintchine formula to compound Poisson distributions. Recall that ξ_n is a random variable with law μ_n , then this compound Poisson distribution has characteristic function given for all $u \in \mathbb{R}^d$ by

$$\exp(iun \mathbb{E}[h(\xi_n)] + n \mathbb{E}[e^{iu\xi_n} - 1 - iuh(\xi_n)]) = \exp\left(iun \mathbb{E}[h(\xi_n)] - \frac{n}{2} \mathbb{E}[(uh(\xi_n))^2] + n \mathbb{E}[f_u(\xi_n)]\right).$$

On the right, the first term converges to $i\Gamma u$ as desired; in the last expectation, the function f_u is

¹OK, we did not yet prove that it was unique, but we promise that the proof would come and be independent of what follows, so there is no circular reasoning.

continuous, bounded, and $o(|x|^2)$ near 0, so by the third assumption the last term converges to the corresponding integral against π . In order to conclude, it only remains to show that for each $u \in \mathbb{R}^d$,

$$n \mathbb{E}[(uh(\xi_n))^2] \xrightarrow{n \rightarrow \infty} (\Sigma u)^2 + \int (uh)^2 d\pi.$$

Let us fix $u \in \mathbb{R}^d$. Recall that $h(x) = x + o(|x|^2)$ at 0 so $(uh)^2$ is not $o(|x|^2)$ at 0 in general and we cannot apply the third assumption. We are going to use instead the second assumption via a cut-off. Let $\varepsilon > 0$. Using uniform approximations of $\mathbb{1}_{|x| > \varepsilon}$ by continuous and bounded functions that vanish on $\{|x| \leq \varepsilon/2\}$, we deduce from the third assumption that

$$n \mathbb{E}[(uh(\xi_n))^2 \mathbb{1}_{|\xi_n| > \varepsilon}] \xrightarrow{n \rightarrow \infty} \int_{|x| > \varepsilon} (uh(x))^2 \pi(dx). \quad (9.7)$$

For the converse indicator function, using similarly uniform approximations of $\mathbb{1}_{|x| \leq \varepsilon}$ by continuous and bounded functions that vanish on $\{|x| > 2\varepsilon\}$, and since $(uh(x))^2 - (ux)^2 = (u(h(x) - x)) \times (u(h(x) + x))$, is continuous and $o(|x|^2)$ near 0 since $h(x) - x$ is, we obtain:

$$n \mathbb{E}[(uh(\xi_n))^2 - (u\xi_n)^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] \xrightarrow{n \rightarrow \infty} \int_{|x| \leq \varepsilon} ((uh(x))^2 - (ux)^2) \pi(dx). \quad (9.8)$$

Combining the two displays yields:

$$\begin{aligned} n \mathbb{E}[(uh(\xi_n))^2] &= n \mathbb{E}[(uh(\xi_n))^2 \mathbb{1}_{|\xi_n| > \varepsilon}] + n \mathbb{E}[(uh(\xi_n))^2 - (u\xi_n)^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] + n \mathbb{E}[(u\xi_n)^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] \\ &= \int (uh(x))^2 \pi(dx) + \int_{|x| \leq \varepsilon} (ux)^2 \pi(dx) + n \mathbb{E}[(u\xi_n)^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] + o(1). \end{aligned}$$

By the second item, the last term converges to $(\Sigma u)^2$ as first $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$ and by dominated convergence, the second integral converges to 0 as $\varepsilon \rightarrow 0$. $\square$

Proof of the necessity. Assume now that S_n , or equivalently, the compound Poisson laws converge to μ and let us check that the three conditions are fulfilled, where (Γ, Σ^2, π) is related to μ by the Lévy–Khintchine formula. We constructed infinitely divisible distributions and found the triple (Γ, Σ^2, π) by taking subsequential limits of such compound Poisson distributions. Uniqueness of this triple then implies that these subsequential limits must agree. Let us recall the main steps in Section 9.4.

Let us define a measure ρ_n with density $|x|^2/(1+|x|^2)$ with respect to $n\mu_n$, with finite total mass. By Lemma 9.10, these measures are tight. Along a suitable subsequence, we thus have $\rho_n \rightarrow \rho$ weakly, in the sense that $\int f d\rho_n \rightarrow \int f d\rho$, for some measure ρ with finite total mass, possibly 0. Then define the Lévy measure π as the measure with density $\frac{1+|x|^2}{|x|^2} \mathbb{1}_{x \neq 0}$ with respect to ρ . Consequently, along the previous subsequence,

$$n \mathbb{E}[f(\xi_n)] = \int f n d\mu_n = \int f(x) \frac{(1+|x|^2)}{|x|^2} \rho_n(dx) \xrightarrow{n \rightarrow \infty} \int f(x) \frac{(1+|x|^2)}{|x|^2} \rho(dx) = \int f d\pi,$$

whenever f is continuous, bounded, and $f(x)/|x|^2 \rightarrow 0$ as $x \rightarrow 0$ to deal with the fact that π forgets the possible atom of ρ at 0.

Next let us decompose the characteristic exponent ψ_n of our compound Poisson law just as previously:

$$\psi_n(u) = n \mathbb{E} \left[1 + iuh(\xi_n) - \frac{(uh(\xi_n))^2}{2} - e^{iu\xi_n} \right] + \frac{n}{2} \mathbb{E}[(uh(\xi_n))^2] - iun \mathbb{E}[h(\xi_n)].$$

Then the function in the first expectation is continuous, bounded, and $o(|x|^2)$ as $x \rightarrow 0$, so this term converges to the integral against π . Just as in Section 9.4, by Lemma 9.2, the fact that the compound Poisson laws converge to μ implies that $\psi_n \rightarrow \psi$, even uniformly on compact sets. Hence the sum of the

last two terms in the above display must converge at every $u \in \mathbb{R}^d$. Since it is a quadratic function of u , then so is the limit: there exists a vector Γ in $\mathbb{R}^d$ and a symmetric matrix Λ such that

$$n \mathbb{E}[h(\xi_n)] \xrightarrow{n \rightarrow \infty} \Gamma \quad \text{and} \quad \frac{n}{2} \mathbb{E}[(uh(\xi_n))^2] \xrightarrow{n \rightarrow \infty} \frac{1}{2}(\Lambda u)^2,$$

for each $u \in \mathbb{R}^d$. Recall that $h(x) = x + o(|x|^2)$ as $x \rightarrow 0$. It remains to check the second of the three conditions in the claim. Fix once and for all $u \in \mathbb{R}^d$; fix also $\varepsilon > 0$. The limits (9.7) and (9.8) in the first part of the proof based on the third item remain valid here and yield the same conclusion, just presented differently:

$$n \mathbb{E}[(u\xi_n)^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] \xrightarrow{n \rightarrow \infty} (\Lambda u)^2 - \int (uh(x))^2 \pi(dx) - \int_{|x| \leq \varepsilon} (uh(x))^2 \pi(dx).$$

The last integral further tends to 0 as $\varepsilon \rightarrow 0$ by dominated convergence. The remaining part is $(\Sigma u)^2$ for some symmetric matrix Σ , as was shown at the end of Section 9.4. Then all the three items in the claim hold true for this triple (Γ, Σ^2, π) .

In order to relate this triple to the law μ , let us apply the sufficient part of the theorem that we proved earlier, that shows that S_n converges in distribution to the infinitely divisible law associated with this triple (Γ, Σ^2, π) . Since it also converges to μ , then by uniqueness of the triple, it must be the one corresponding to μ . then the previous limits that were taken along a subsequence hold for the entire sequences. $\square$

THE DIMENSION 1

Let us conclude with the important special case of the dimension $d = 1$. We use small letters γ and σ instead of Γ and Σ . As we mentioned, the standard truncation function used in the Lévy–Khintchine formula is nowadays:

$$h(x) = x \mathbb{1}_{|x| < 1}.$$

It is not continuous, but the function $\tilde{h}$ given by the identity on $[-1, 1]$ and extended as the constant -1 on $(-\infty, -1]$ and the constant 1 on $[1, \infty)$ is continuous. Recall that replacing h by $\tilde{h}$ in the Lévy–Khintchine formula only amounts to replacing the triple (γ, σ, π) by $(\tilde{\gamma}, \sigma, \pi)$ where

$$\tilde{\gamma} = \gamma + \int (\tilde{h} - h) d\pi = \gamma + \pi([1, \infty)) - \pi((-\infty, -1]),$$

due to the choice of $\tilde{h}$ that makes $\tilde{h} - h = \mathbb{1}_{[1, \infty)} - \mathbb{1}_{(-\infty, -1]}$.

THEOREM 9.16. Let the dimension be $d = 1$ and choose the (discontinuous) truncation function $h(x) = x \mathbb{1}_{|x| < 1}$. Then S_n converges in distribution to some limit μ if and only if μ is infinitely divisible with associated Lévy triple (γ, σ, π) and the following convergences hold true:

$$\gamma = \lim_{n \rightarrow \infty} n(\mathbb{E}[\xi_n \mathbb{1}_{|\xi_n| < 1}] + \mathbb{P}(\xi_n \geq 1) - \mathbb{P}(\xi_n \leq -1)) - (\pi([1, \infty)) - \pi((-\infty, -1])), \quad (9.9)$$

and

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} |n \mathbb{E}[\xi_n^2 \mathbb{1}_{|\xi_n| \leq \varepsilon}] - \sigma^2| = 0, \quad (9.10)$$

as well as

$$\pi([x, \infty)) = \lim_{n \rightarrow \infty} n \mathbb{P}(\xi_n \geq x), \quad \text{and} \quad \pi((-\infty, -y]) = \lim_{n \rightarrow \infty} n \mathbb{P}(\xi_n \leq -y), \quad (9.11)$$

for every $x, y > 0$ such that $\pi(\{x\}) = \pi(\{-y\}) = 0$.

Proof. We rely of course on Theorem 9.15. The first assumption here is the same as the first assumption there after replacing the discontinuous h by the continuous $\tilde{h}$. The second assumptions are the same in both theorems. Finally, let us check that the second and third assumptions in Theorem 9.15 combined are equivalent to the combination of (9.10) and (9.11).

First, the third assumption in Theorem 9.15 implies (9.11) simply by approximating the indicator functions $\mathbb{1}_{[x,\infty)}$ and $\mathbb{1}_{(-\infty,-y]}$ for every $x, y > 0$ by continuous and bounded functions that vanish in a neighbourhood of 0. Suppose conversely that both (9.10) and (9.11) hold. Let us fix a continuous and bounded function f such that $f(u) = o(u^2)$ as $u \rightarrow 0$. Let $\varepsilon > 0$ be such that $\pi(\{-\varepsilon\}) = \pi(\{\varepsilon\}) = 0$ (which are dense since the complement is countable) and let us split

$$n \mathbb{E}[f(\xi_n)] = n \mathbb{E}[f(\xi_n) \mathbb{1}_{\xi_n < -\varepsilon}] + n \mathbb{E}[f(\xi_n) \mathbb{1}_{\xi_n > \varepsilon}] + n \mathbb{E}[f(\xi_n) \mathbb{1}_{|\xi_n| \leq \varepsilon}].$$

Let μ_n denote the law of ξ_n ; then the measures $n\mu_n$ and π restricted to the interval $[\varepsilon, \infty)$ are finite measures and (9.11) applied with $x = \varepsilon$ shows that the total mass of the former converges to that of the latter. In addition, for any $x > \varepsilon$ such that $\pi(\{x\}) = 0$, by applying (9.11) to both ε and x , we deduce that $n\mu_n([\varepsilon, x]) \rightarrow \pi([\varepsilon, x])$. If $\pi([\varepsilon, \infty)) > 0$, then the probability given by dividing $n\mu_n$ on $[\varepsilon, \infty)$ by its total mass satisfies:

$$\frac{n\mu_n([\varepsilon, x])}{n\mu_n([\varepsilon, \infty))} \xrightarrow{n \rightarrow \infty} \frac{\pi([\varepsilon, x])}{\pi([\varepsilon, \infty))},$$

for all $x > \varepsilon$ such that the limit is continuous at x , that is, by Proposition 2.4, weak convergence of these probabilities. It follows then that the finite measures obtained without rescaling converge weakly. If $\pi([\varepsilon, \infty)) = 0$, then the convergence of the total mass of $n\mu_n$ implies weak convergence as well. The symmetric argument applies to the measures restricted to the interval $(-\infty, -\varepsilon]$ so we conclude that for any $\varepsilon > 0$ such that $\pi(\{-\varepsilon\}) = \pi(\{\varepsilon\}) = 0$ and for any continuous and bounded function f on $(-\infty, -\varepsilon] \cup [\varepsilon, \infty)$, we have

$$n \mathbb{E}[f(\xi_n) \mathbb{1}_{|\xi_n| > \varepsilon}] \xrightarrow{n \rightarrow \infty} \int_{|x| > \varepsilon} f d\pi.$$

Now suppose f is a continuous and bounded function on $\mathbb{R}$ and such that $f(x) = o(x^2)$ at 0. Then by (9.10) we have

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} |n \mathbb{E}[f(\xi_n) \mathbb{1}_{|\xi_n| \leq \varepsilon}]| = 0.$$

Combining the last two displays and letting $\varepsilon \rightarrow 0$ along a sequence such that $\pi(\{-\varepsilon\}) = \pi(\{\varepsilon\}) = 0$ yields the third assumption in Theorem 9.15. $\square$

REMARK 9.17. Note that if (9.11) holds at $x = y = 1$, then (9.9) reduces to

$$\gamma = \lim_{n \rightarrow \infty} n \mathbb{E}[\xi_n \mathbb{1}_{|\xi_n| < 1}].$$

In particular, if the limit law we are after has a Lévy measure π with $\pi(\{x\}) = 0$ for every $x \in \mathbb{R}$ (for example if π has a density with respect to the Lebesgue measure), then the weak convergence takes place if and only if the above display holds as well as (9.10) and (9.11) at every positive point.

9.6 EXERCISES

EXERCISE 45. Prove that an infinitely divisible law μ with compact support, that is $\mu(\mathbb{R}^d \setminus [-K, K]^d) = 0$ for some $K > 0$, is necessarily a Dirac mass.

EXERCISE 46. Prove that the exponential laws are infinitely divisible and express their characteristic triple.

EXERCISE 47. Prove that the geometric laws (say on $\{0, 1, 2, \dots\}$) are infinitely divisible and express their

characteristic triple.

EXERCISE 48. Let $(\xi_n)_n$ be i.i.d. random variables and let $S_n = \xi_1 + \cdots + \xi_n$ for every $n \geq 0$. Let τ be independent and with the geometric law with parameter $p \geq 1$. Is the compound distribution S_τ infinitely divisible? What do you think about the case where the ξ_n 's have a Bernoulli distribution?

EXERCISE 49. Let X be a random variable with values in $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. Prove that the following assertions are equivalent:

- (i) For every $n \geq 1$, the random variable X has the same law as the sum of n i.i.d. random variables with values in $\mathbb{Z}_+$,
- (ii) The system of equations:

$$k P(X = k) = \sum_{j=1}^k j \lambda_j P(X = k - j)$$

for all $k \geq 1$ has a solution $\lambda_k \geq 0$ for all $k \geq 1$,

- (iii) X has the same law as $\sum_{j=1}^{\infty} j Z_j$ where the Z_j 's are independent with the Poisson distribution with parameter $\lambda_j > 0$ respectively, with $\sum_j \lambda_j = \lambda < \infty$ (so, necessarily, $\lambda = -\log P(X = 0)$),
- (iv) X has the same law as $\sum_{j=1}^N Y_j$ where the Y_j 's are i.i.d. with the law $P(Y_1 = k) = \lambda_k / \lambda$ and N is independent and has the Poisson distribution with parameter $\lambda = -\log P(X = 0)$.

In particular, this shows that every infinitely divisible on $\mathbb{Z}_+$ is a compound Poisson distribution.

CHAPTER 10

LÉVY PROCESSES

This chapter lifts the discussion from the previous chapter to the level of processes. We shall indeed construct (càdlàg) stochastic process corresponding to infinitely divisible distribution in the same way Brownian motion corresponds to the standard Gaussian distribution. These processes are thus the expected limits of rescaled random walks in the generalisation of Donsker's theorem (although we will not prove it). We believe that these processes also unravel some of the mystery of the Lévy–Khintchine formula.

CONTENTS

10.1	Lévy processes	111
10.2	Poisson processes	112
10.3	Compound Poisson processes	115
10.4	Lévy–Itô decomposition	118
10.5	Construction of the martingale	119
10.6	Poisson random measure of jumps	121
10.7	Exercises	123

10.1 LÉVY PROCESSES

Here are the main characters of this chapter.

DEFINITION 10.1. A Lévy process is a càdlàg stochastic process $(X_t)_{t \geq 0}$ in $\mathbb{R}^d$ that has *independent and stationary increments*, namely:

- for every $0 = t_0 < t_1 < \dots < t_k$, the random variables $(X_{t_i} - X_{t_{i-1}})_{1 \leq i \leq k}$ are independent,
- for every $s, t > 0$, the increment $X_{t+s} - X_t$ has the same law as $X_s - X_0$.

Unless otherwise specified, we shall here always assume that $X_0 = 0$.

Lévy processes are the continuous-time analogues of random walks, given by sums of i.i.d. random variables, and indeed such sums satisfy the independent and stationary increments assumption when we restrict the t_i 's to integers. These processes are named after Paul Lévy, as many theorems in probability, but at his time, Lévy referred to them as 'processus additifs', which is a broader class by only assuming independence of increments; they were then simply referred to as processes with independent and stationary increments until the 1970's where the denomination of Lévy processes grew to become standard.

Lévy processes generalise the Brownian motion, and indeed, examples are $(t\Gamma + \Sigma B_t)_{t \geq 0}$ for any vector $\Gamma \in \mathbb{R}^d$ and any symmetric matrix Σ , where B is the standard Brownian motion in $\mathbb{R}^d$. The space $\mathcal{D}$ of càdlàg functions is important here as we shall see that linear transforms of Brownian motion $(t\Gamma + \Sigma B_t)_{t \geq 0}$

are the *only* Lévy processes with continuous paths. Let us denote by

$$\Delta X_t = X_t - X_{t-}$$

the, possibly null, ‘jump’ of X at time t .

Lévy processes are intimately related to infinitely divisible distributions. Recall from Definition 9.4 the characteristic exponent ψ associated with any such distribution.

LEMMA 10.2. If $X = (X_t)_{t \geq 0}$ is a Lévy process, then for every $t \geq 0$, the law of X_t is infinitely divisible, with characteristic exponent $t\psi$ where ψ is the characteristic exponent of X_1 . In addition, if $Y = (Y_t)_{t \geq 0}$ is another Lévy process such that X_1 and Y_1 have the same law, then the entire processes have the same law.

Proof. Recall that the law of a stochastic process $(X_t)_t$ is characterised by its finite-dimensional marginal distributions, namely the joint law of $(X_{t_i})_{1 \leq i \leq k}$ for every $k \geq 1$ and $t_1 < \dots < t_k$. Equivalently, one can consider the increments between the consecutive t_i 's and by the very definition of a Lévy process, since these increments are independent and stationary, then the law of Lévy process $(X_t)_t$ is characterised by its one-dimensional marginals, that is, the law of X_t for any given $t \geq 0$.

Now observe that for every $n \geq 1$ this law can be decomposed as:

$$X_t = \sum_{i=1}^n (X_{it/n} - X_{(i-1)t/n}),$$

where the summands are i.i.d. with the same law as $X_{t/n}$. Hence X_t has an infinitely divisible distribution. Let us consider characteristic functions. Let $t = p/q$ be a rational number, where $p, q \geq 1$ are integers, then $X_{p/q}$ has the law of the sum of p i.i.d. copies of $X_{1/q}$ and X_1 has the law of the sum of q i.i.d. copies of $X_{1/q}$ so

$$\mathbb{E}[e^{iuX_{p/q}}] = \mathbb{E}[e^{iuX_{1/q}}]^p \quad \text{and} \quad \mathbb{E}[e^{iuX_1}] = \mathbb{E}[e^{iuX_{1/q}}]^q.$$

This implies that

$$\mathbb{E}[e^{iuX_t}] = \mathbb{E}[e^{iuX_1}]^t = e^{t\psi(u)},$$

for every rational number $t \geq 0$. Since $t \mapsto X_t$ is right-continuous, then both sides are right-continuous and then the identity extends to all real numbers $t \geq 0$. This proves that the law of X_1 characterises that of X_t for any $t \geq 0$ and therefore the entire law of the process X . $\square$

We are going to see now that, conversely, any infinitely divisible distribution can be seen as the law at time 1 of a (necessarily unique in distribution) Lévy process. Recall that the Lévy–Khintchine formula expresses the characteristic function of any infinitely divisible distribution by (9.5). It suffices then to construct for any such characteristic function a Lévy process with this law at time 1. This will in return shed some lights on this formula.

10.2 POISSON PROCESSES

Let us start with particular cases of Lévy processes, which are related to the Poisson and compound Poisson distributions.

DEFINITION 10.3. A Lévy process in which X_1 has the Poisson distribution with parameter $\lambda > 0$ is called the *Poisson process with intensity λ* .

Existence of such a process is given by the theorem below, which relates them to the following notion.

DEFINITION 10.4. A *counting process* is a càdlàg stochastic process $(N_t)_{t \geq 0}$ with values in $\mathbb{Z}_+$, which is constant except only for jumps of size exactly 1. It takes the form

$$N_t = \#\{n \geq 1 : T_n \leq t\}, \quad t > 0,$$

where T_n is the time of the n 'th jump for each $n \geq 1$. We shall also denote by $T_0 = 0$.

The formula explains the name: think of the T_n 's representing instants at which events of interest occur, then N_t counts the number of such events that have occurred so far.

THEOREM 10.5. Let $X = (X_t)_{t \geq 0}$ be a càdlàg process in $\mathbb{R}$.

- (i) If X is a Lévy process and X_1 follows a Poisson distribution then X is a counting process.
- (ii) If X is both a counting process and a Lévy process, then the time intervals between the jumps $(T_n - T_{n-1})_{n \geq 1}$, where $T_0 = 0$, are i.i.d. and exponentially distributed.
- (iii) The counting process $(X_t)_{t \geq 0}$ associated with times such that $(T_n - T_{n-1})_{n \geq 1}$ are i.i.d. and exponentially distributed, with some parameter λ , is a Lévy process and X_t has the Poisson distribution with parameter λt for each $t > 0$.

REMARK 10.6. The sample paths of a Poisson process are thus extremely simple: it stays constant equal to 0 for a random amount of time that follows an exponential distribution, at which point it jumps to 1 and stays then constant for another random amount of time with the same exponential distribution and independent of the first one, etc.

Let us only prove the last assertion, which shows that Poisson processes do exist. We leave the proof of the first two assertions in Exercise 51; the first item shows that Poisson processes are necessarily counting processes, and the second item shows that they are the only ones.

Proof of (iii). Let us first prove that X_t has the Poisson law with parameter λt for each fixed $t > 0$. Indeed

$$\mathbb{P}(X_t = n) = \mathbb{P}(T_n \leq t < T_{n+1}) = \mathbb{P}(U_1 + \dots + U_n \leq t < U_1 + \dots + U_{n+1}),$$

where $U_1 = T_1$ and $U_i = T_i - T_{i-1}$ for $i \geq 1$. We just argued that the U_i 's are i.i.d. with the exponential law with parameter λ so we may compute:

$$\begin{aligned} \mathbb{P}(T_n \leq t < T_{n+1}) &= \int_{[0, \infty)^{n+1}} \mathbb{1}_{u_1 + \dots + u_n \leq t < u_1 + \dots + u_{n+1}} \prod_{i=1}^{n+1} \lambda e^{-\lambda u_i} du_i \\ &= \int_{[0, t]^n} \mathbb{1}_{u_1 + \dots + u_n \leq t} \left(\underbrace{\int_{t - (u_1 + \dots + u_n)}^{\infty} \lambda e^{-\lambda u_{n+1}} du_{n+1}}_{= e^{-\lambda(t - (u_1 + \dots + u_n))}} \right) \prod_{i=1}^n \lambda e^{-\lambda u_i} du_i \\ &= \lambda^n e^{-\lambda t} \int_{[0, t]^n} \mathbb{1}_{u_1 + \dots + u_n \leq t} \prod_{i=1}^n du_i. \end{aligned}$$

Let us change the variables by $t_k = t(u_1 + \dots + u_k)$ to get:

$$\mathbb{P}(X_t = n) = (\lambda t)^n e^{-\lambda t} \int_{[0, 1]^n} \mathbb{1}_{0 < t_1 < \dots < t_n \leq 1} \prod_{i=1}^n dt_i.$$

The integral is the probability that n successive i.i.d. uniform random times on $[0, 1]$ appear in increasing order, which by symmetry equals $1/n!$ which concludes the proof that X_t has the desired Poisson law.

Let us conclude with the proof that X is a Lévy process; it is càdlàg by construction so let us fix $0 = t_0 < t_1 < \dots < t_k$ and prove that the increments $(X_{t_i} - X_{t_{i-1}})_{1 \leq i \leq k}$ are independent and Poisson distributed with parameter $\lambda(t_i - t_{i-1})$ respectively. Fix $\ell_1, \dots, \ell_k \geq 0$ and, in order to ease notation, let us write $L_i = \ell_1 + \dots + \ell_i$ and $U_{L_i} = u_1 + \dots + u_{L_i}$. Then by a similar computation as above,

$$\begin{aligned}
& \mathbb{P}(\forall 1 \leq i \leq k : X_{t_i} - X_{t_{i-1}} = \ell_i) \\
&= \mathbb{P}(T_{\ell_1} \leq t_1 < T_{\ell_1+1} \leq T_{\ell_1+\ell_2} \leq t_2 < \dots < T_{\ell_1+\dots+\ell_k} \leq t_k < T_{\ell_1+\dots+\ell_{k+1}}) \\
&= \int_{[0,\infty)^{L_{k+1}}} \prod_{j=1}^k \mathbb{1}_{U_{L_j} \leq t_j < U_{L_{j+1}}} \prod_{i=1}^{L_{k+1}} \lambda e^{-\lambda u_i} du_i \\
&= \int_{[0,\infty)^{L_k}} \prod_{j=1}^{k-1} \mathbb{1}_{U_{L_j} \leq t_j < U_{L_{j+1}}} \mathbb{1}_{U_{L_k} \leq t_k} e^{-\lambda(t_k - U_{L_k})} \prod_{i=1}^{L_k} \lambda e^{-\lambda u_i} du_i \\
&= \lambda^{L_k} e^{-\lambda t_k} \int_{[0,\infty)^{L_k}} \prod_{j=1}^{k-1} \mathbb{1}_{U_{L_j} \leq t_j < U_{L_{j+1}}} \mathbb{1}_{U_{L_k} \leq t_k} \prod_{i=1}^{L_k} du_i \\
&= \lambda^{L_k} e^{-\lambda t_k} \int_{[0,\infty)^{L_{k-1}}} \prod_{j=1}^{k-2} \mathbb{1}_{U_{L_j} \leq t_j < U_{L_{j+1}}} \mathbb{1}_{U_{L_{k-1}} \leq t_{k-1}} \left(\int_{[0,\infty)^{\ell_k}} \mathbb{1}_{t_{k-1} < U_{L_{k-1}+1} \leq U_{L_k} \leq t_k} \prod_{i=L_{k-1}+1}^{L_k} du_i \right) \prod_{i=1}^{L_{k-1}} du_i.
\end{aligned}$$

Just as before the integral in parenthesis, divided by $(t_k - t_{k-1})^{\ell_k}$, is the probability that ℓ_k i.i.d. random variables uniformly distributed in $[t_{k-1}, t_k]$ are ranked in increasing order, which equals $1/(\ell_k)!$ so

$$\mathbb{P}(\forall 1 \leq i \leq k : X_{t_i} - X_{t_{i-1}} = \ell_i) = \lambda^{L_k} e^{-\lambda t_k} \frac{(t_k - t_{k-1})^{\ell_k}}{\ell_k!} \int_{[0,\infty)^{L_{k-1}}} \prod_{j=1}^{k-2} \mathbb{1}_{U_{L_j} \leq t_j < U_{L_{j+1}}} \mathbb{1}_{U_{L_{k-1}} \leq t_{k-1}} \prod_{i=1}^{L_{k-1}} du_i.$$

This integral is exactly the same as in the penultimate line in the first computation, with $k-1$ instead of k . We thus infer by induction that

$$\mathbb{P}(\forall 1 \leq i \leq k : X_{t_i} - X_{t_{i-1}} = \ell_i) = \lambda^{L_k} e^{-\lambda t_k} \prod_{j=1}^k \frac{(t_j - t_{j-1})^{\ell_j}}{\ell_j!} = \prod_{j=1}^k \frac{(t_j - t_{j-1})^{\ell_j} \lambda^{\ell_j} e^{-\lambda(t_j - t_{j-1})}}{\ell_j!},$$

so indeed the $X_{t_i} - X_{t_{i-1}}$'s are independent with the desired Poisson distribution. $\square$

REMARK 10.7. By a similar reasoning, one can prove that, conditionally on the event $N_t = n$, the vector $(T_1, \dots, T_n)$ is distributed as the increasing rearrangement of n i.i.d. uniform random times on $[0, t]$.

The following two results will be useful. The first one shows that the superposition of independent Poisson processes is again a Poisson process.

LEMMA 10.8 (Summation). Fix $K \geq 1$ and for each $1 \leq i \leq K$, let $(N_{t,i})_{t \geq 0}$ be a Poisson process with intensity λ_i respectively. Assume that these processes are independent, then

$$N_t = \sum_{i=1}^K N_{t,i}, \quad t > 0$$

is a Poisson process with intensity $\sum_{i=1}^K \lambda_i$.

Proof. It is clear that the sum of independent Lévy processes is a Lévy process. In addition, an easy exercise shows that the sum of independent Poisson random variables has the Poisson distribution with parameter the sum of the parameters. We conclude from Lemma 10.2. $\square$

The next result is a kind of converse property, showing that when sampling a Poisson process and

assigning types (or colours) randomly to the jumps, then the number of jumps of each type follows independent Poisson processes.

LEMMA 10.9 (Colouring/Thinning). Let $(N_t)_{t \geq 0}$ be a Poisson process with intensity λ and, independently, let $(Y_n)_{n \geq 1}$ be i.i.d. random variables with values in a finite set $\{1, \dots, J\}$. Then the processes defined by

$$N_{t,j} = \#\{n \geq 1 : T_n \leq t \text{ and } Y_n = j\}, \quad 1 \leq j \leq J, t \geq 0$$

are independent Poisson processes with intensity $\lambda \mathbb{P}(Y_1 = j)$ respectively.

Proof. Clearly each process is càdlàg. Since the law of a stochastic process is characterised by the joint law of arbitrary finitely many increments, then we must prove that such increments for these different processes are all independent: from a process to another as well as the increments of the same process, and that they are stationary. By splitting further each time interval, it is sufficient to consider the same increments for all processes (draw a picture), namely we must prove that for any $0 = t_0 < t_1 < \dots < t_k$, the random variables

$$(N_{t_i,j} - N_{t_{i-1},j})_{1 \leq i \leq k, 1 \leq j \leq J}$$

are independent and Poisson distributed with parameter $\lambda(t_i - t_{i-1})p_j$ respectively, where $p_j = \mathbb{P}(Y_1 = j)$. For this, let us decompose according to the increments of the entire process $(N_t)_{t \geq 0}$. Fix nonnegative integers $(\ell_{i,j})_{1 \leq i \leq k, 1 \leq j \leq J}$ and let $L_i = \sum_{j=1}^J \ell_{i,j}$ for each i , then

$$\begin{aligned} & \mathbb{P}(\forall 1 \leq i \leq k, 1 \leq j \leq J : N_{t_i,j} - N_{t_{i-1},j} = \ell_{i,j}) \\ &= \mathbb{P}(\forall 1 \leq i \leq k : N_{t_i} - N_{t_{i-1}} = L_i \text{ and } \forall 1 \leq j \leq J : \#\{n \in \{1 + \sum_{a=1}^{i-1} L_a, \dots, \sum_{a=1}^i L_a\} : Y_n = j\} = \ell_{i,j}) \\ &= \prod_{i=1}^k \frac{(\lambda(t_i - t_{i-1}))^{L_i} e^{-\lambda(t_i - t_{i-1})}}{L_i!} \frac{L_j!}{\prod_{j=1}^J \ell_{i,j}!} \prod_{j=1}^J p_j^{\ell_{i,j}} \\ &= \prod_{i=1}^k \prod_{j=1}^J \frac{(\lambda(t_i - t_{i-1})p_j)^{\ell_{i,j}} e^{-\lambda(t_i - t_{i-1})p_j}}{\ell_{i,j}!}. \end{aligned}$$

This proves the claim. □

10.3 COMPOUND POISSON PROCESSES

Now we may use the Poisson process to define more general Lévy processes, whose law at time 1 is a compound Poisson distribution as in Definition 9.1.

DEFINITION 10.10. Let ν be a finite (and nonzero) measure on $\mathbb{R}^d$. A Lévy process in which X_1 has the compound Poisson distribution with intensity $\nu > 0$ is called the *compound Poisson process with intensity ν* . We denote the law of this process by $\text{CPP}(\nu)$.

THEOREM 10.11. Let ν be a finite (and nonzero) measure on $\mathbb{R}^d$ and let

- $(N_t)_{t \geq 0}$ be a Poisson process with intensity $\lambda = \nu(\mathbb{R}^d) > 0$,
- $(Z_n)_{n \geq 1}$ be i.i.d. with the law $\mu(\cdot) = \nu(\cdot)/\lambda$,
- and such that $(N_t)_{t \geq 1}$ and $(Z_n)_{n \geq 1}$ are independent.

Then the process

$$X_t = Z_1 + \cdots + Z_{N_t}, \quad t > 0,$$

is a compound Poisson process with intensity ν .

Proof. This process is càdlàg by construction, and its increments inherit stationarity and independence from the Poisson process and the i.i.d. jump values: for every $0 = t_0 < t_1 < \cdots < t_k$, and every Borel sets $B_1, \dots, B_k$ of $\mathbb{R}^d$, we have

$$\begin{aligned} & \mathbb{P}(\forall 1 \leq i \leq k : X_{t_i} - X_{t_{i-1}} \in B_i) \\ &= \sum_{n_1, \dots, n_k \geq 0} \mathbb{P}\left(\forall 1 \leq i \leq k : N_{t_i} - N_{t_{i-1}} = n_i \text{ and } \sum_{j=n_1+\dots+n_{i-1}+1}^{n_1+\dots+n_i} Z_j \in B_i\right) \\ &= \sum_{n_1, \dots, n_k \geq 0} \prod_{i=1}^k \mathbb{P}(N_{t_i-t_{i-1}} = n_i) \mathbb{P}\left(\sum_{j=1}^{n_i} Z_j \in B_i\right) \\ &= \prod_{i=1}^k \sum_{n_i \geq 0} \mathbb{P}\left(N_{t_i-t_{i-1}} = n_i, \sum_{j=1}^{n_i} Z_j \in B_i\right) \\ &= \prod_{i=1}^k \sum_{n_i \geq 0} \mathbb{P}(X_{t_i-t_{i-1}} \in B_i), \end{aligned}$$

which proves that the increments are independent and stationary. In addition, the law X_1 is the construction of the compound Poisson distribution with intensity ν . $\square$

REMARK 10.12. A compound Poisson process has a simple sample path structure: it is piecewise constant, with isolated jumps. We may then write:

$$X_t = \sum_{s \leq t} \Delta X_s, \quad t \geq 0,$$

where the sum only has finitely many nonzero terms for each $t \geq 0$.

REMARK 10.13. Just as Poisson processes are the only Lévy processes that are also counting processes, one can prove that compound Poisson processes are the only Lévy processes that are piecewise constant. Indeed, if a Lévy process $X = (X_t)_{t \geq 0}$ is piecewise constant, then we may order the jump times in increasing order $(T_n)_{n \geq 1}$ and associate with it the values of the jumps $Z_n = \Delta X_{T_n}$. The counting process $N_t = \#\{n \geq 1 : T_n \leq t\}$ then inherits from X stationary and independent increments. Hence it is a Poisson process. By decomposing according to N_t and using that X has stationary and independent increments, one can further prove that the jumps $(Z_n)_{n \geq 1}$ are i.i.d. and independent from $(N_t)_{t \geq 0}$. We will give another proof in the next chapter (see Proposition 11.2), based on the forthcoming result of the present chapter.

In Section 9.1, we observed that the summands in a compound Poisson random variable that fall in disjoint sets are independent. This picture is clearer at the level of processes. Precisely, the summation and thinning properties of Poisson processes extend to compound Poisson processes in a straightforward way.

PROPOSITION 10.14. Compound Poisson processes enjoy the following properties:

- (i) Fix $K \geq 1$ and let $((X_{t,i})_{t \geq 0})_{1 \leq i \leq K}$ be independent compound Poisson processes with intensity ν_i

respectively, then

$$X_t = \sum_{i=1}^K X_{t,i}, \quad t > 0$$

is a compound Poisson process with intensity $\sum_{i=1}^K \nu_i$.

- (ii) Let $(X_t)_{t \geq 0}$ be a Poisson process with intensity ν and let $(B_k)_{k \leq K}$ be partition of $\mathbb{R}^d$ into Borel sets with $\nu(B_k) > 0$ for each $k \leq K$ and decompose the process $X = \sum_{k=1}^K X^k$ where for each $k \leq K$, we let

$$X_t^k = \sum_{s \leq t} \Delta X_s \mathbb{1}_{\Delta X_s \in B_k} = \sum_{i=1}^{N_t} Z_i \mathbb{1}_{Z_i \in B_k}, \quad t > 0.$$

Then the processes X^k are independent compound Poisson processes with intensity $\nu(\cdot \cap B_k)$ respectively.

Proof. First note that all processes we consider are càdlàg by construction.

- (i) The sum of independent Lévy processes is a Lévy process. In addition, we saw in Section 9.1 that the sum of independent compound Poisson random variables has the compound Poisson distribution with intensity the sum of intensities. Hence $(X_t)_{t \geq 0}$ is a Lévy process and X_1 has the compound Poisson distribution with intensity $\sum_{i=1}^K \nu_i$, which fully characterises the law of the process by Lemma 10.2.
- (ii) That the X^k 's are compound Poisson processes with the claimed intensity follows by combining Theorem 10.11 with the thinning property of Poisson processes from Lemma 10.9 since the $(Z_i \mathbb{1}_{Z_i \in B_k})_{i \geq 1}$ are i.i.d. random variables with law $\nu(\cdot \cap B_k)/\nu(B_k)$ and the process $N_{t,k} = \sum_{n \geq 1} \mathbb{1}_{T_n \leq t, Z_n \in B_k}$ is a Poisson process with intensity $\nu(\mathbb{R}^d) \times \nu(B_k)/\nu(\mathbb{R}^d) = \nu(B_k)$.

Independence can be proved by a similar calculation to the proof of the thinning property of Poisson processes. Let us propose another, indirect argument. Indeed, let $(Y^k)_{k \leq K}$ be independent compound Poisson process with intensity $\nu(\cdot \cap B_k)$ respectively and let $Y_t = \sum_{k=1}^K Y_t^k$. By the first item, the process Y is a compound Poisson process with intensity $\sum_{k=1}^K \nu(\cdot \cap B_k) = \nu$, just as X is. In addition, the Y^k 's may be retrieved from Y in the same way as the X^k 's from X , namely

$$Y_t^k = \sum_{s \leq t} \Delta Y_s \mathbb{1}_{\Delta Y_s \in B_k}, \quad 1 \leq k \leq K, t > 0.$$

The processes $(Y, Y^1, \dots, Y^K)$ and $(X, X^1, \dots, X^K)$ thus have the same law, and since the Y^k 's are independent, then the X^k 's must be independent. $\square$

Let us record here a result that will shall need later. Indeed, we are used to define a discrete-time martingale from the sum of i.i.d. random variables by subtracting the mean, and this extends in this context.¹

LEMMA 10.15. Let X be a compound Poisson process with intensity some finite measure ν . Suppose that $\int |x| \nu(dx) < \infty$, then the process defined by

$$M_t = X_t - t \int x \nu(dx), \quad t \geq 0,$$

is a martingale. If in addition $\int |x|^2 \nu(dx) < \infty$, then it is a square-integrable martingale.

¹A martingale in $\mathbb{R}^d$ is a process whose coordinates are martingales in $\mathbb{R}$.

Proof. We use the natural filtration $\mathcal{F}_t = \sigma(X_s, s \leq t)$. Then $(M_t)_t$ is adapted. Next, let us write $X_t = \sum_{i=1}^{N_t} Z_i$ with $(N_t)_t$ a Poisson process with intensity $\lambda = \nu(\mathbb{R}^d)$ and, independently, $(Z_i)_{i \geq 1}$ i.i.d. random variables with law $\nu(\cdot)/\lambda$. Our assumption implies that $\mathbb{E}[|Z_1|] < \infty$ so, by Fubini's theorem,

$$\mathbb{E}[|X_t|] \leq \mathbb{E}\left[\sum_{i=1}^{N_t} |Z_i|\right] = \mathbb{E}[N_t] \mathbb{E}[|Z_1|] < \infty.$$

This integrability justifies the same identity without norm:

$$\mathbb{E}[X_t] = \mathbb{E}[N_t] \mathbb{E}[Z_1] = t \nu(\mathbb{R}^d) \int x \frac{\nu(dx)}{\nu(\mathbb{R}^d)} = t \int x \nu(dx) < \infty.$$

Integrability of $(X_t)_t$ implies that $(M_t)_n$ is integrable. Finally, recall that a compound Poisson process, just as any Lévy process, is a Markov process, and precisely for any $s, t > 0$, the random increment $X_{t+s} - X_t$ is independent from $\mathcal{F}_s$ and has the same law as X_s . It follows that

$$\mathbb{E}[M_{t+s} - M_t \mid \mathcal{F}_t] = \mathbb{E}[(X_{t+s} - X_t) - s \int x \nu(dx) \mid \mathcal{F}_t] = \mathbb{E}[X_s] - s \int x \nu(dx) = 0.$$

Assume now in addition that $\int |x|^2 \nu(dx) < \infty$ and let us prove that X_t is square-integrable. This assumption now implies that the Z_i 's are square-integrable, so we may justify the identity

$$\mathbb{E}[|X_t|^2] \leq \mathbb{E}\left[\sum_{i=1}^{N_t} |Z_i|^2 + 2 \sum_{1 \leq i < j \leq N_t} |Z_i Z_j|\right] = \mathbb{E}[N_t] \mathbb{E}[|Z_1|^2] + 2 \mathbb{E}[N_t(N_t - 1)] \mathbb{E}[|Z_1 Z_2|] < \infty.$$

Square integrability of $(X_t)_t$ is then transferred to $(M_t)_t$. □

10.4 LÉVY-ITÔ DECOMPOSITION

Let us recall the Lévy-Khintchine formula from Theorem 9.11. For simplicity, we fix the truncation function $h(x) = x \mathbb{1}_{|x| < 1}$. Then this formula establishes a one-to-one correspondence between infinitely divisible distributions μ and triples (Γ, Σ^2, π) where $\Gamma \in \mathbb{R}^d$ is any vector, Σ is any symmetric $d \times d$ matrix, and π is a Lévy measure, that is, a measure on $\mathbb{R}^d$ such that

$$\pi(\{0\}) = 0 \quad \text{and} \quad \int (|x|^2 \wedge 1) \pi(dx) < \infty.$$

The correspondence is precisely given by the fact that the characteristic function of μ is given for all $u \in \mathbb{R}^d$ by:

$$\widehat{\mu}(u) = \exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2 + \int_{\mathbb{R}^d} (e^{iux} - 1 - iux \mathbb{1}_{|x| < 1}) \pi(dx)\right).$$

Let us point out that we did prove that any such triple gives rise to an infinitely divisible distribution and that conversely, any infinitely divisible distribution takes such a form. However we did not prove that this was a one-to-one correspondence, in that the triple is unique in this formula. This will be a consequence of the following result that extends the Lévy-Khintchine formula to the level of processes.

THEOREM 10.16 (Lévy-Itô). A càdlàg process X with $X_0 = 0$ is a Lévy process if and only if there exists a triple (Γ, Σ^2, π) as above such that X has the same law as the sum of three independent processes, given respectively as follows:

- (i) $G_t = t\Gamma + \Sigma B_t$ where B is the d -dimensional Brownian motion,
- (ii) Z_t is a compound Poisson process with intensity the finite measure $\nu(\cdot) = \pi(\cdot \cap \{|x| \geq 1\})$,

- (iii) M_t is a càdlàg square-integrable martingale and a Lévy process, with only jumps of norm less than 1 and with characteristic function given for each $t > 0$ and $u \in \mathbb{R}^d$ by:

$$\mathbb{E}[e^{iuM_t}] = \exp\left(t \int_{|x|<1} (e^{iux} - 1 - iux)\pi(dx)\right).$$

In addition, this decomposition is unique.

In the sequel we shall keep the notation G , Z , and M to refer to these independent processes that add up to X .

REMARK 10.17. Note that we have the characteristic functions:

$$\mathbb{E}[e^{iuG_1}] = \exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2\right) \quad \text{and} \quad \mathbb{E}[e^{iuZ_1}] = \exp\left(\int_{|x|\geq 1} (e^{iux} - 1)\pi(dx)\right).$$

By independence, the sum $X_1 = G_1^1 + Z_1^2 + M_1^3$ has therefore characteristic function:

$$\mathbb{E}[e^{iuX_1}] = \exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2 + \int_{\mathbb{R}^d} (e^{iux} - 1 - iux \mathbb{1}_{|x|<1})\pi(dx)\right).$$

Combining the Lévy–Khintchine formula for infinitely divisible distributions and the Lévy–Itô decomposition for Lévy processes, we thus infer that:

- given any admissible triple (Γ, Σ^2, π) , there exists a Lévy process X whose law at time 1, which is infinitely divisible, satisfies the Lévy–Khintchine formula with this triple,
- conversely any infinitely divisible distribution satisfies the Lévy–Khintchine formula with some triple, and therefore is the law at time 1 of some Lévy process,
- since the decomposition of the Lévy process is unique, then so is the triple.

We proved the Lévy–Khintchine formula without referring to Lévy processes (uniqueness of the triple can also be proved using analytic arguments). Similarly, the Lévy–Itô decomposition can be proved without referring to infinitely divisible distributions, but we shall not do so. Precisely, we construct in the next section the process M as above given a Lévy measure π , thus showing the existence of the Lévy process $X = G + Z + M$ associated with any triple. Recall that for infinitely divisible distributions, this was the easy part. The more difficult part is, just as for infinitely divisible distributions, given any Lévy process X , to decompose its law as that of $G + Z + M$. We refer to e.g. the book of Kallenberg for a proof of this; here we simply deduce it from the Lévy–Khintchine formula since X_1 is infinitely divisible and thus has a characteristic function in the form above, which allows to construct G, Z, M in the previous way.

10.5 CONSTRUCTION OF THE MARTINGALE

Let us construct the process M as in Theorem 10.16. Recall that for infinitely divisible distributions, this part of the formula was obtained by a limiting procedure, by considering compound Poisson distributions with intensity $\pi(\cdot \cap \{1/n \leq |x| < 1\})$ and recentred. Then M will be similarly obtained as limit of centred compound Poisson processes. Precisely, for each $k \geq 1$, let $Y^k = (Y_t^k)_{t \geq 0}$ be a compound Poisson process with intensity

$$\pi_k(\cdot) = \pi(\cdot \cap A_k) \quad \text{where} \quad A_k = \left\{x \in \mathbb{R}^d : \frac{1}{k+1} \leq |x| < \frac{1}{k}\right\},$$

and assume that they are independent. Let then

$$M_t^n = \sum_{k=1}^{n-1} \left(Y_t^k - t \int_{A_k} x \pi(dx) \right) = \sum_{k=1}^{n-1} Y_t^k - t \int_{1/n \leq |x| < 1} x \pi(dx), \quad n \geq 2, t > 0.$$

Since $\int |x|^2 \pi_n(dx) < \infty$, then according to Lemma 10.15, each summand in the middle is a square-integrable martingale, so the processes $(M_t^n)_{t \geq 0}$ are martingales for the same filtration $\mathcal{F}_t = \sigma(Y_s^k, s \leq t, k \geq 1)$ since the Y^k 's are independent.

LEMMA 10.18. There exists a stochastic process M such that for each $t > 0$, we have the convergence:

$$\mathbb{E} \left[\sup_{s \leq t} |M_s - M_s^n|^2 \right] \xrightarrow{n \rightarrow \infty} 0.$$

Proof. Fix $t > 0$. Since the random variables $(Y_t^k)_{k \geq 1}$'s are independent and have mean $t \int_{A_k} x \pi(dx)$ respectively (recall the proof of Lemma 10.15), then M_t^n is the sum of independent zero-mean random variables so the sequence $(M_t^n)_{n \geq 2}$ is a discrete-time martingale! In addition,

$$\mathbb{E}[|M_t^n|^2] = \sum_{k=1}^{n-1} \mathbb{E} \left[\left| Y_t^k - t \int_{A_k} x \pi(dx) \right|^2 \right] = t \sum_{k=1}^{n-1} \int_{A_k} |x|^2 \pi(dx) = t \int_{1/n \leq |x| < 1} |x|^2 \pi(dx),$$

which converges by monotone convergence to $t \int_{0 \leq |x| < 1} |x|^2 \pi(dx) < \infty$. Hence the martingale is bounded in L^2 and so it converges almost surely and in L^2 to some limit M_t .

Now for n fixed, as t varies, the processes $(M_t^n)_{t \geq 0}$ are martingales for the same filtration $(\mathcal{F}_t)_{t \geq 0}$. The L^2 , and thus L^1 , convergence for each t implies that $(M_t)_{t \geq 0}$ is a $(\mathcal{F}_t)_{t \geq 0}$ -martingale. Indeed for any $s, t > 0$ and any event $A \in \mathcal{F}_t$, it holds

$$\mathbb{E}[M_{t+s} \mathbb{1}_A] = \lim_{n \rightarrow \infty} \mathbb{E}[M_{t+s}^n \mathbb{1}_A] = \lim_{n \rightarrow \infty} \mathbb{E}[M_t^n \mathbb{1}_A] = \mathbb{E}[M_t \mathbb{1}_A].$$

As a consequence, for any $n \geq 1$, the process $(M_t - M_t^n)_{t \geq 0}$ is a square-integrable martingale. Doob's maximal inequality implies then that for any $t > 0$, we have

$$\mathbb{E} \left[\sup_{s \leq t} |M_s - M_s^n|^2 \right] \leq 4 \mathbb{E}[|M_t - M_t^n|^2] \xrightarrow{n \rightarrow \infty} 0,$$

as we claimed. $\square$

This convergence propagates some properties of the M^n 's, showing that this limit M is the process that appears in Theorem 10.16.

PROPOSITION 10.19. The limit process M is both a square-integrable martingale and a Lévy process, and its characteristic function is given by:

$$\mathbb{E}[e^{iuM_t}] = \exp \left(t \int_{|x| < 1} (e^{iux} - 1 - iux) \pi(dx) \right).$$

Proof. We saw that M was a martingale in the previous proof. Let us prove that it is a Lévy process. The L^2 -convergence implies the almost sure convergence along a subsequence. By applying the previous lemma to each integer time $t \in \mathbb{N}$, a diagonal extraction argument then provides a subsequence along which almost surely, we have $\sup_{s \leq t} |M_s - M_s^n| \rightarrow 0$ for any $t > 0$. Since each M^n is càdlàg, then so is M by this local uniform convergence (more precisely, it admits a càdlàg modification).

The summation property from Proposition 10.14 implies that for each $n \geq 1$, the martingale M^n is a compound Poisson process with intensity $\sum_{k=1}^{n-1} \pi(\cdot \cap A_k) = \pi(\cdot \cap \bigcup_{k=1}^{n-1} A_k)$ and centred by subtracting

the mean. Independent and stationary increments for M is then inherited from the M^n 's by letting $n \rightarrow \infty$ along a subsequence that converges almost surely. Finally, we know the characteristic functions of compound Poisson processes so, taking into account the linear term,

$$\mathbb{E}[e^{iuM_t^n}] = \exp\left(t \int_{1/n \leq |x| < 1} (e^{iux} - 1 - iux)\pi(dx)\right),$$

which converges to the same expression in which we integrate over the entire unit ball by dominated convergence, using that $\int_{|x| < 1} |x|^2 \pi(dx) < \infty$. $\square$

10.6 POISSON RANDOM MEASURE OF JUMPS

The description of the jumps of a Lévy process can be made very precise with the notion of *Poisson random measures*. More on this notion is provided in Chapter A.

DEFINITION 10.20. A *random measure* on a measurable space $(E, \mathcal{E})$ is a random variable in the space $\mathcal{M}$ of measures on E , for the σ -algebra $\mathcal{M} = \sigma(\mu \mapsto \mu(A) \in [0, \infty] : A \in \mathcal{E})$ generated by the evaluations on measurable sets. For any such a random measure M , the formula

$$m(A) = \mathbb{E}[M(A)], \quad A \in \mathcal{E}$$

defines a (deterministic) measure on E called the *intensity measure* of M .

Let us extend Poisson distributions to include the null and infinite parameters, in which case a Poisson distributed random variable is constant equal to 0 and constant equal to infinity respectively.

DEFINITION 10.21. Let $(E, \mathcal{E})$ be measurable space and m a measure on this space. A *Poisson random measure* with intensity m is a random measure M on E with the property that for every finite collection $A_1, \dots, A_k$ of disjoint measurable sets, their masses $M(A_1), \dots, M(A_k)$ are independent and Poisson distributed, with parameter $m(A_1), \dots, m(A_k)$ respectively.

Note that with the previous convention, if $A \in \mathcal{E}$ is a set such that $m(A) = 0$, then $M(A) = 0$ almost surely, whereas if $m(A) = \infty$, then $M(A) = \infty$ almost surely. In fact, for any set $A \in \mathcal{E}$,

$$m(A) = 0 \iff M(A) = 0 \text{ almost surely} \quad \text{and} \quad m(A) = \infty \iff M(A) = \infty \text{ almost surely.}$$

Let us denote by dt the Lebesgue measure on $[0, \infty)$.

THEOREM 10.22. The jump measure of a Lévy process X , defined for every $A \in \mathcal{B}([0, \infty)) \otimes \mathcal{B}(\mathbb{R}^d)$ by:

$$M^X(A) = \sum_{t \geq 0} \delta_{(t, X_t)}(A),$$

is a Poisson random measure on $[0, \infty) \times \mathbb{R}^d$ with intensity $dt \otimes \pi$.

Proof. Let us first prove the claim when X is a compound Poisson process with intensity a finite measure π . Fix Borel sets $B_1, \dots, B_k$ of $\mathbb{R}^d$ and half open intervals $[c_1, d_1), \dots, [c_k, d_k)$ of $[0, \infty)$. By considering pairwise intersections if necessary, we may assume that the intervals $[c_i, d_i)$ are disjoint and the sets B_j are disjoint as well. Then the processes $(X_{c_i+t} - X_{c_i} : t \in [0, d_i - c_i))_{1 \leq i \leq k}$ are independent and distributed as $(X_t : t \in [0, d_i - c_i))$ respectively. According to Proposition 10.14, the processes restricted to each B_j are independent compound Poisson processes with intensity $\pi(\cdot \cap B_j)$. In particular, the total number of jumps made by the process X in a set B_j during a time interval $[c_i, d_i)$ is Poisson distributed with

parameter $(d_i - c_i)\pi(B_j)$ and these numbers are independent as i and j vary.

This shows that if M is a Poisson random measure with intensity $dt \otimes \pi$, then the joint law of the vectors

$$(M^X([c_i, d_i] \times B_j))_{1 \leq i, j \leq k} \quad \text{and} \quad (M([c_i, d_i] \times B_j))_{1 \leq i, j \leq k}$$

are the same for all disjoint intervals $[c_i, d_i]$ and disjoint measurable sets B_j . Since the collection of such products is stable under finite intersection and generates $\mathcal{B}([0, \infty)) \otimes \mathcal{B}(\mathbb{R}^d)$, then we conclude from the monotone class lemma that the random measures have the same law.

Now let us consider a general Lévy process X . Let us fix again disjoint intervals $[c_1, d_1], \dots, [c_k, d_k]$ of $[0, \infty)$ and disjoint measurable sets $B_1, \dots, B_k$ of $\mathbb{R}^d$. Since the property is only about the jumps, we may just forget about the Gaussian part G , or equivalently assume that Γ and Σ vanish. In this case X has the same law as $Z + M$ where Z is a compound Poisson process and is independent of M , which is the limit, in the sense of Lemma 10.18, of the compound Poisson processes with drift M^n . Assume that there exists $n \geq 1$ such that $B_j \subset \{x \in \mathbb{R}^d : |x| \geq 1/n\}$ for each $1 \leq j \leq k$. Define $Z^n = Z + M^n$, which is a compound Poisson process with intensity $\pi(\cdot \cap \{|x| \geq 1/n\})$, minus a linear term with slope $\int_{1/n \leq |x| < 1} x \pi(dx)$. Notice that Z^n only makes jumps with norm greater than or equal to $1/n$, while $X - Z^n = M - M^n$ is the limit of $M^{n+p} - M^p$ as $p \rightarrow \infty$, and as such it only makes jumps with norm smaller than $1/n$. Our assumption on the B_j 's then implies that

$$M^X([c_i, d_i] \times B_j) = M^{Z^n}([c_i, d_i] \times B_j), \quad 1 \leq i, j \leq k.$$

By the first part of the proof, the right sides are independent and Poisson distributed with parameter $(d_i - c_i)\pi(B_j)$ respectively.

This holds true as long as there exists $n \geq 1$ such that $B_j \subset \{x \in \mathbb{R}^d : |x| \geq 1/n\}$ for each $1 \leq j \leq k$, so it holds for all closed sets that do not contain the origin. Since the collection of products $\{[c, d] \times B : 0 \leq c < d < \infty, B \text{ closed and } 0 \notin B\}$ is closed under finite intersections and generates $\mathcal{B}([0, \infty)) \otimes \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, then we infer as above from a monotone class argument as above that, the measure M^X and a Poisson random measure M with intensity $dt \otimes \pi$ have the same law on $[0, \infty) \times (\mathbb{R}^d \setminus \{0\})$. Since $\pi(\{0\}) = 0$, then $M([0, \infty) \times \{0\}) = 0$ almost surely, and $M^X([0, \infty) \times \{0\}) = 0$ almost surely by definition of the jumps, so finally the two random measures have the same law. $\square$

REMARK 10.23. The Lévy measure π describes the law of the jumps of the process X , and for example for every $T > 0$ and any Borel set $A \in \mathcal{B}(\mathbb{R}^d)$, the quantity

$$\#\{t \leq T : \Delta X_t \in A \setminus \{0\}\}$$

is Poisson distributed with parameter $T\pi(A)$. In particular either $\pi(A) = 0$ and then almost surely X will never make any jump of size in A , or $\pi(A) > 0$ and then almost surely X will make infinitely many jumps of size in A (since there is the same positive probability to make at least 1 in each interval $[n, n+1)$).

As a first corollary, let us at last prove that the characteristic triple of an infinitely divisible distribution, or of a Lévy process, is unique.

COROLLARY 10.24. The triple (Γ, Σ^2, π) in the Lévy–Khintchine formula from Theorem 9.11 and in the Lévy–Itô decomposition from Theorem 10.16 is unique.

Proof. Consider an infinitely divisible distribution μ constructed from two triples (Γ, Σ^2, π) and (Γ', Σ', π') . Construct also two Lévy processes from these triples, their laws at time 1 is μ , so the entire processes have the same law by Lemma 10.2. According to Theorem 10.22, the jump measures of these Lévy

processes is a Poisson random measure with intensity both $dt \otimes \pi$ and $dt \otimes \pi'$. Clearly the intensity measure of a random measure is unique, hence these two measures coincide and so $\pi = \pi'$. The law of the Lévy process thus characterises uniquely the Lévy measure, and so the law of processes Z and M which are constructed using π only. We conclude that if two Lévy processes have the same law, then their Gaussian parts G have the same law, and thus same mean vector $\Gamma = \Gamma'$ and same covariance matrix $\Sigma^2 = (\Sigma')^2$. $\square$

10.7 EXERCISES

EXERCISE 50. Let $(N_t)_{t \geq 0}$ be a Poisson process with intensity $\lambda > 0$ and let T_n denote the time of the n 'th jump.

- (i) In words, what does T_{N_t} and T_{N_t+1} represent?
- (ii) Compute the joint law of $t - T_{N_t}$ and $T_{N_t+1} - t$.
- (iii) What do you observe about the increment $T_{N_t+1} - T_{N_t}$? What is the limit in distribution as $t \rightarrow \infty$?

EXERCISE 51. The goal of this exercise is to characterise Poisson processes among all real-valued Lévy processes.

- (i) Let $X = (X_t)_{t \geq 0}$ be a Lévy process in $\mathbb{R}$ such that X_1 follows a Poisson distribution, say with parameter $\lambda > 0$.

- (a) Prove that X_t has the Poisson distribution with parameter λt for any $t > 0$.
- (b) Prove that almost surely, X takes values in $\mathbb{Z}_+$ and is nondecreasing.
- (c) Prove that almost surely,

$$\sup_{t \in (0,1]} \Delta X_t = \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} (X_{k/n} - X_{(k-1)/n}),$$

and use this to prove that, almost surely, it holds $\Delta X_t \in \{0, 1\}$ for all $t > 0$.

- (d) Conclude that X is necessarily a counting process.

- (ii) Now let us assume that $N = (N_t)_{t \geq 0}$ is both a Lévy process and a counting process, with jump times $(T_n)_{n \geq 1}$. Assume that T_1 is almost surely finite (i.e. that N is not almost surely constant null).

- (a) Prove that T_1 is exponentially distributed.
- (b) Let $M_t = N_{T_1+t} - N_{T_1}$ for any $t \geq 0$. Fix $0 < t_0 < \dots < t_k$ and $u, u_1, \dots, u_k \geq 0$. By decomposing according to the events $(n-1)\delta < T_1 \leq n\delta$ with $\delta > 0$ sufficiently small, prove that

$$e^{-u\delta} \mathbb{E}[e^{-uT_1}] \prod_{i=1}^k \mathbb{E}[e^{-u_i N_{t_i - t_{i-1} + \delta}}] \leq \mathbb{E}[e^{-uT_1 - \sum_{i=1}^k u_i (M_{t_i} - M_{t_{i-1}})}] \leq e^{u\delta} \mathbb{E}[e^{-uT_1}] \prod_{i=1}^k \mathbb{E}[e^{-u_i N_{t_i - t_{i-1} - \delta}}].$$

- (c) Deduce that the entries of the vector $(T_1, M_{t_1} - M_{t_0}, \dots, M_{t_k} - M_{t_{k-1}})$ are independent and each $M_{t_i} - M_{t_{i-1}}$ has the same law as $N_{t_i - t_{i-1}}$.
- (d) Infer that $(M_t)_{t \geq 0}$ is independent of T_1 and distributed as $(N_t)_{t \geq 0}$ and conclude finally $(T_n - T_{n-1})_{n \geq 1}$ are i.i.d.

EXERCISE 52. Prove that the coordinates of a Lévy process in $\mathbb{R}^d$ are Lévy processes in $\mathbb{R}$. Prove conversely that a vector of *independent* real-valued Lévy process is a multidimensional Lévy process. Provide an example of two dependent Lévy process whose pair is not a Lévy process.

EXERCISE 53. Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two compound Poisson processes. Let ψ_X and ψ_Y denote their respective characteristic exponent.

- (i) Prove that if X and Y are independent, then almost surely, they never jump at the same time.
- (ii) Construct an example in which the processes are not independent but never jump at the same time anyway.

The goal of this exercise is to prove that if $(X_t, Y_t)_{t \geq 0}$ is a Lévy process and if X and Y never jump at the same time, then they are independent.

- (iii) Prove that for every $u, v \in \mathbb{R}$, the processes defined for each $t \geq 0$ by

$$M_t = \exp(iuX_t + t\psi_X(u)) \quad \text{and} \quad N_t = \exp(ivY_t + t\psi_Y(v)),$$

are complex-valued martingales for the filtration $\mathcal{F}_t = \sigma((X_s, Y_s) : s \leq t)$ and that they are uniformly bounded on every bounded interval.

- (iv) Let us use the generic notation $\Delta Z_{k/n} = Z_{k/n} - Z_{(k-1)/n}$. Establish that for each $n \geq 1$, we have

$$M_1 N_1 - 1 = \sum_{k=1}^n M_{(k-1)/n} \Delta N_{k/n} + \sum_{k=1}^n N_{(k-1)/n} e^{\psi_Y(v)/n} \Delta M_{k/n} + \sum_{k=1}^n (N_{k/n} - N_{(k-1)/n} e^{\psi_Y(v)/n}) \Delta M_{k/n},$$

and that the first two sums have zero expectation.

- (v) Prove that the expectation of the last sum above tends to 0 provided that X and Y have no simultaneous jump.
- (vi) Conclude that if X and Y have no simultaneous jump, then they are independent.

CHAPTER 11

SOME PROPERTIES OF LÉVY PROCESSES

This chapter shows how the characteristic triple (Γ, Σ^2, π) can give precise information about the behaviour of a Lévy process. It can encode for example whether the jump times are isolated, or dense, whether the sample paths have finite or infinite variation, whether, in dimension 1, the process is nondecreasing, or also if the one-dimensional marginal laws are integrable, square integrable, etc. Our goal is mostly to exhibit how useful to Lévy–Itô decomposition can be.

CONTENTS

11.1	Structure of the jumps	125
11.2	Finite and infinite variation	126
11.3	Increasing Lévy processes	128
11.4	Moments of a Lévy process	130
11.5	Exercises	134

11.1 STRUCTURE OF THE JUMPS

The Lévy–Itô decomposition $X = G + Z + M$ from Theorem 10.16 provides information about the jumps of X . Recall that Z is a compound Poisson process with intensity π restricted to $\{x \in \mathbb{R}^d : |x| \geq 1\}$ and that M is the limit of the processes M^n which are centred compound Poisson processes with intensity π restricted to $\{x \in \mathbb{R}^d : 1/n \leq |x| < 1\}$ respectively. Then for each $n \geq 1$, let us decompose our Lévy process in law as

$$X = G + Z^n + Y^n,$$

where:

- $G_t = t\Gamma + \Sigma B_t$ is a scaled Brownian motion with a drift, which has continuous paths
- $Z^n = Z + M^n$ is, by independence, according to Proposition 10.14, a compound Poisson process with intensity π restricted to $\{|x| \geq 1/n\}$, to which we subtract the linear drift with slope $\int_{1/n \leq |x| < 1} x\pi(dx)$; it is piecewise linear, with only isolated jumps with norm greater than or equal to $1/n$, which arrive according to a Poisson process with intensity $\pi(\{|x| \geq 1/n\}) < \infty$, so they are spaced out by i.i.d. exponential random times with mean $1/\pi(\{|x| \geq 1/n\})$ by Theorem 10.5,
- $Y^n = M - M^n = \lim_{k \rightarrow \infty} (M^{n+k} - M^n)$ is the limit of the recentred compound Poisson processes with intensity π restricted to $\{1/(n+k) \leq |x| < 1/n\}$ as $k \rightarrow \infty$; each $M^{n+k} - M^n$ only has jumps smaller than $1/n$, hence so does Y^n .

We infer that Z^n is just the collection of the jumps of X larger than $1/n$ plus a linear drift, namely for all $t > 0$:

$$Z_t^n = \sum_{s \leq t} \Delta X_s \mathbb{1}_{|\Delta X_s| \geq 1/n} - t \int_{1/n \leq |x| < 1} x\pi(dx), \quad (11.1)$$

where the sum contains only finitely many nonzero terms. In general however, the process Y^n cannot be written similarly as the sum of the increments of X that are smaller than $1/n$ plus a drift, simply because these jumps may not be summable, this is why we need the construction of Section 10.5 (see Section 11.2 below for more information).

COROLLARY 11.1. A Lévy process has continuous path if and only if π is the null measure, which is the case if and only if $X_t = G_t = t\Gamma + \Sigma B_t$ is a scaled Brownian motion with a drift.

Proof. If π is the null measure, then for every $n \geq 1$, the process Z^n is constant equal to 0, so X can only have jump with norm less than $1/n$. Since this holds true for any $n \geq 1$, then X has no jump at all. On the contrary, if π is not the null measure, and since $\pi(\{0\}) = 0$, then there must exist $n \geq 1$ such that $\pi(\{x \in \mathbb{R}^d : |x| \geq 1/n\}) > 0$. For such a value of n , the process Z^n is not constant equal to 0 since it makes a Poisson number of jumps with parameter $\pi(\{x \in \mathbb{R}^d : |x| \geq 1/n\})$ on any interval with unit length; hence X is discontinuous. $\square$

Let us remove the Brownian component and now focus on the ‘jump’ part. The simple image of a Lévy process with $\Sigma = 0$ as piecewise linear with slope dictated by Γ and discrete, isolated jumps dictated by π is a bit naive.

PROPOSITION 11.2. When $\Sigma = 0$, we have the dichotomy:

- Either π has finite total mass, and then X is almost surely piecewise linear and in fact reduces to a compound Poisson process with intensity π and a linear drift, in which case the time-intervals between two jumps are i.i.d. and exponentially distributed with expectation $1/\pi(\mathbb{R}^d) > 0$,
- Or π has infinite total mass, and then X has almost surely (countably) infinitely many jumps in any open time-interval.

Proof. Suppose first that π has infinite mass and fix an open interval with length $\ell > 0$. Let us write $X = Z^n + Y^n$ for each $n \geq 1$ and recall that Z^n has during this interval a Poisson number of jumps with parameter $\ell\pi(\{x \geq 1/n\}) < \infty$. Since $\pi(\{x \geq 1/n\}) \rightarrow \pi(\mathbb{R}^d) = \infty$, then this Poisson random variable tends to infinity almost surely as $n \rightarrow \infty$. Note that the process X makes at least as many jumps as Z^n in this interval, and this number is arbitrarily large when n is large.

Next suppose that π has finite mass. In this case, the function $ux \mathbb{1}_{|x|<1}$ is integrable so we may rewrite the Lévy–Khintchine formula as (since we assume $\Sigma = 0$):

$$\mathbb{E}[e^{iuX_1}] = \exp\left(iu\left(\Gamma - \int_{|x|<1} x\pi(dx)\right) + \int_{\mathbb{R}^d} (e^{iux} - 1)\pi(dx)\right),$$

to see that X reduces to a compound Poisson process with intensity π plus a linear drift with slope $\Gamma - \int_{|x|<1} x\pi(dx)$, which is indeed piecewise linear. The jump times of X are then distributed as a Poisson process with intensity $\pi(\mathbb{R}^d) < \infty$ so the conclusion follows from Theorem 10.5. $\square$

11.2 FINITE AND INFINITE VARIATION

Recall that the martingale M was obtained as the limit of the compound Poisson processes of the jumps whose norm lies between $1/n$ and 1 and recentred by their mean. One can wonder if this recentring is really required to pass to the limit. Letting $n \rightarrow \infty$ in $\sum_{s \leq t} \Delta X_s \mathbb{1}_{|\Delta X_s| \geq 1/n}$ a priori requires that the series of the jumps of X be absolutely convergent. Recall that the Lévy measure in general integrates $|x|^2$ near 0.

LEMMA 11.3. The series $\sum_{s \leq t} |\Delta X_s|$ is either almost surely finite for every $t > 0$ or almost surely infinite for every $t > 0$ according to whether the integral $\int_{|x| < 1} |x| \pi(dx)$ is finite or not.

Proof. Recall that for every $t > 0$, the series $\sum_{s \leq t} |\Delta X_s| \mathbb{1}_{|\Delta X_s| \geq 1}$ is almost surely finite since there are only finitely many summands, so we focus on $\sum_{s \leq t} |\Delta X_s| \mathbb{1}_{|\Delta X_s| < 1}$. For every $n \geq 1$, the process given by $(\sum_{s \leq t} |\Delta X_s| \mathbb{1}_{1/n \leq |\Delta X_s| < 1})_{t \geq 0}$ is a compound Poisson process with intensity the image by the norm of π restricted to $\{1/n \leq |x| < 1\}$. Then the argument from Section 10.5 shows that

$$\sum_{s \leq t} |\Delta X_s| \mathbb{1}_{1/n \leq |\Delta X_s| < 1} - \int_{1/n \leq |x| < 1} |x| \pi(dx)$$

converges locally uniformly in L^2 and thus almost surely along a subsequence to a finite limit. By monotone convergence applied to each term, this limit equals

$$\sum_{s \leq t} |\Delta X_s| \mathbb{1}_{|\Delta X_s| < 1} - \int_{|x| < 1} |x| \pi(dx).$$

Hence one term is finite if and only if the other one is. $\square$

REMARK 11.4. By the same argument, since a Lévy measure satisfies $\int_{|x| < 1} |x|^2 \pi(dx) < \infty$, then it is always true that $\sum_{s \leq t} |\Delta X_s|^2 < \infty$.

We immediately infer the following special form of the Lévy–Itô decomposition.

COROLLARY 11.5. When $\int (|x| \wedge 1) \pi(dx) < \infty$, the Lévy–Itô decomposition of X takes the form:

$$X_t = t \left(\underbrace{\Gamma - \int_{|x| < 1} x \pi(dx)}_{= \tilde{\Gamma}} \right) + \Sigma B_t + \sum_{s \leq t} \Delta X_s.$$

Its characteristic function at $t = 1$ equals:

$$\hat{\mu}(u) = \exp \left(iu\tilde{\Gamma} - \frac{1}{2}(\Sigma u)^2 + \int (e^{iux} - 1) \pi(dx) \right), \quad u \in \mathbb{R}^d.$$

This summation property of the jumps actually relates to the behaviour of the sample paths of X .

DEFINITION 11.6. The (total) *variation* over an interval $[s, t]$ of a function $x : [0, \infty) \rightarrow \mathbb{R}^d$ is:

$$V(x; [s, t]) = \sup \left\{ \sum_{i=1}^k |x_{u_i} - x_{u_{i-1}}| : k \geq 1, s = u_0 < u_1 < \dots < u_k = t \right\}.$$

We say that x has finite variation when $V(x; [s, t]) < \infty$ for every $s < t$ and has infinite variation when $V(x; [s, t]) = \infty$ for every $s < t$.

REMARK 11.7. Clearly, if x is nondecreasing, then its total variation is $V(x; [s, t]) = x_t - x_s < \infty$. In particular the difference of two nondecreasing functions has finite variation and it turns out the converse is true: notice that $t \mapsto V(x; [s, t])$ is nondecreasing and so is $t \mapsto V(x; [s, t]) - x_t$.

EXAMPLE 11.8. Any linear drift has finite variation, as well as any compound Poisson process since it only assumes finitely many different values in a bounded interval. However Brownian motion has infinite variation, see e.g. the course on stochastic calculus.

Since a Lévy process is the independent sum of a Brownian motion with a linear term, a compound Poisson process with only jumps larger than 1, and of a square-integrable martingale M with only jumps smaller than 1, then the study of the variation of the entire process reduces to that of this martingale M .

PROPOSITION 11.9 (total variation). A Lévy process has almost surely finite variation when both

$$\Sigma = 0 \quad \text{and} \quad \int (|x| \wedge 1) \pi(dx) < \infty,$$

and almost surely infinite variation otherwise.

Proof. First note that if the two conditions are fulfilled, then by the previous corollary, $X_t = t\tilde{\Gamma} + \sum_{s \leq t} \Delta X_s$ and $\sum_{s \leq t} |\Delta X_s| < \infty$ almost surely so X has finite variation almost surely.

Suppose instead that X has finite variation with positive probability. On this event it holds $\sum_{s \leq t} |\Delta X_s| < \infty$ and thus $\int (|x| \wedge 1) \pi(dx) < \infty$. This further implies that $\sum_{s \leq t} |\Delta X_s| < \infty$ almost surely and the process X takes the form $X_t = \Sigma B_t + \tilde{X}_t$ where $\tilde{X}_t = t\tilde{\Gamma} + \sum_{s \leq t} \Delta X_s$ has finite variation almost surely. Hence the process $\Sigma B_t = X_t - \tilde{X}_t$ has finite variation with positive probability. Since Brownian motion has infinite variation almost surely, then necessarily $\Sigma = 0$, and so finally $X = \tilde{X}$ has finite variation almost surely. $\square$

Combining this proposition and the previous corollary, we infer that the Lévy–Itô decomposition of a Lévy process X with finite variation takes the form:

$$X_t = t\tilde{\Gamma} + \sum_{s \leq t} \Delta X_s \quad \text{where} \quad \tilde{\Gamma} = \Gamma - \int_{|x| < 1} x \pi(dx),$$

and its characteristic function at $t = 1$ equals:

$$\hat{\mu}(u) = \exp\left(iu\tilde{\Gamma} + \int (e^{iux} - 1) \pi(dx)\right), \quad u \in \mathbb{R}^d.$$

It thus resembles a compound Poisson process with a drift, but here the Lévy measure can have infinite total mass. In this case X is not a compound Poisson process with a drift and is not piecewise linear with some jumps: the small jumps still accumulate, but they are (absolutely) summable, as opposed to only square-summable for general Lévy processes.

11.3 INCREASING LÉVY PROCESSES

Let us consider Lévy processes in dimension 1, with triple (γ, σ^2, π) .

LEMMA 11.10. The process X has no negative jump if and only if $\pi((-\infty, 0)) = 0$.

Proof. Let us decompose $X = G + Z^n + Y^n$ for each $n \geq 1$. If $\pi((-\infty, 0)) = 0$, then Z^n only has positive jumps (if any), so X can only make negative jumps that are smaller in absolute value than $1/n$, so no negative jump at all by letting $n \rightarrow \infty$. Conversely if X has no negative jump, then Z^n only has positive jumps (if any) and so $\pi((-\infty, -1/n)) = 0$ for each $n \geq 1$, and thus $\pi((-\infty, 0)) = \lim_n \pi((-\infty, -1/n)) = 0$ by monotonicity. $\square$

However, even without Brownian part and when $\gamma \geq 0$, the condition $\pi((-\infty, 0)) = 0$ is not sufficient to conclude that X has nondecreasing paths. This is due to the martingale part constructed by a limit procedure after removing a (nonnegative in this case) drift. In words, small jumps may accumulate and be compensated by a decreasing part in this construction. Preventing such an accumulation turns out to require a stronger integrability condition than $\int (x^2 \wedge 1)\pi(dx) < \infty$ and indeed, recall that nondecreasing functions have finite variation.

THEOREM 11.11. A real-valued Lévy process X has nondecreasing paths if and only if

$$\sigma = 0, \quad \pi((-\infty, 0)) = 0, \quad \text{and} \quad \int_{(0, \infty)} (x \wedge 1)\pi(dx) < \infty \quad \text{and} \quad \gamma \geq \int_{(0, 1)} x\pi(dx).$$

REMARK 11.12. Again, we stress that when $\int_{(0, 1)} x\pi(dx) = \infty$, even if $\sigma = 0$ and $\pi((-\infty, 0)) = 0$, no matter how large γ is, the process X does *not* have nondecreasing paths. These are not at all pathological examples and such processes with one-sided jumps but still not monotone are basic objects in insurance theory, queuing theory, branching processes, and much more.

Proof. In order that X have nondecreasing paths, it is necessary that it has finite variation and no negative jumps. According to Proposition 11.9 and the previous lemma, the first three conditions are necessary and sufficient for this. In addition, in this case it takes the form

$$X_t = \tilde{\gamma}t + \sum_{s \leq t} \Delta X_s, \quad \text{where} \quad \tilde{\gamma} = \gamma - \int_{(0, 1)} x\pi(dx).$$

Since $\Delta X_s \geq 0$ for any $s \geq 0$, then the process X has nondecreasing paths as soon as $\tilde{\gamma} \geq 0$. In order to see that the converse implication holds, observe that if X is nondecreasing, then it remains so after removing finitely many jumps, so

$$X_t^n = X_t - \sum_{s \leq t} \Delta X_s \mathbb{1}_{\Delta X_s \geq 1/n}$$

remains increasing, and in particular nonnegative. Letting $n \rightarrow \infty$ shows that

$$t\tilde{\gamma} = X_t - \sum_{s \leq t} \Delta X_s \geq 0,$$

so indeed $\tilde{\gamma} \geq 0$. □

Nondecreasing Lévy processes play an important role and have deserved a name.

DEFINITION 11.13. A Lévy process in $\mathbb{R}$ with nondecreasing paths is called a *subordinator*. In this case, the Lévy–Itô decomposition takes the form:

$$X_t = \delta t + \sum_{s \leq t} \Delta X_s,$$

where $\delta \geq 0$ and $\Delta X_s \geq 0$ for all $s \geq 0$. Its characteristic function at $t = 1$ takes the form:

$$\mathbb{E}[e^{iuX_1}] = \exp\left(i\delta u + \int_{\mathbb{R}} (e^{iux} - 1)\pi(dx)\right),$$

where π satisfies $\pi((-\infty, 0]) = 0$ and $\int (x \wedge 1)\pi(dx) < \infty$.

Because of the above Lévy–Itô decomposition, the nonnegative real number δ is called the *drift* of the

subordinator. If $\delta = 0$, the subordinator is said to be ‘pure jump’. On the contrary, note that a subordinator with continuous paths is necessarily just a deterministic linear drift $t \mapsto \delta t$.

REMARK 11.14. In addition of being an interesting class of Lévy processes on its own, subordinators play a very important role in the theory of Markov. Precisely, Bochner developped in the 1950’s a theory of ‘subordination’ of processes in which one takes a Lévy process (or more generally, a Markov process) X and an independent subordinator S , and one considers the time-changed process X_{S_t} , which is again a Lévy process (resp. a Markov process); one can express the characteristic function of X_{S_t} in terms of those of X and S simply by conditioning.

11.4 MOMENTS OF A LÉVY PROCESS

The previous sections have shown how integrability of the Lévy measure π near 0 affects the behaviour of the Lévy process. Let us turn to integrability at ∞ .

Indeed, one might suspect that existence of moments of a Lévy process, that is, the finiteness of $\mathbb{E}[|X_t|^\alpha]$ for a given $\alpha > 0$, or more generally of $\mathbb{E}[g(X_t)]$ with $g \geq 0$, has to do with large jumps of the process, and thus with the integrability properties at infinity of the Lévy measure. The type of functions $g : \mathbb{R}^d \rightarrow \mathbb{R}$ we can treat is quite large: we assume that g is measurable, bounded on every compact set, and that there exists $c > 0$ such that for every $x, y \in \mathbb{R}^d$, we have

$$g(x + y) \leq cg(x)g(y).$$

Such functions are called *submultiplicative*.

EXAMPLE 11.15. For any $u \in \mathbb{R}^d$, the function $\exp(ux)$ is (sub)multiplicative. For any $0 < a \leq 1$, the function $\exp(|x|^a)$ for any $0 < a \leq 1$ is submultiplicative. So are $\log(|x| \vee e)$ and $\log \log(|x| \vee e^e)$. The norm is *not* submultiplicative (take y very small), but $|x| \vee 1$ is. Also, if g is submultiplicative, then so is $g(ax + b)^c$ for any $a \in \mathbb{R}$, $b \in \mathbb{R}^d$ and $c > 0$.

THEOREM 11.16 (moments). Let g be a submultiplicative function which is measurable and bounded on every compact set. Then a Lévy process X satisfies $\mathbb{E}[g(X_t)] < \infty$ for some $t > 0$ if and only if

$$\int_{|x| \geq 1} g(x) \pi(dx) < \infty,$$

and in this case $\mathbb{E}[g(X_t)] < \infty$ for every $t > 0$.

Why considering submultiplicative functions? From a technical point of view, because we would like to consider only exponential functions, due to their multiplicative property that goes well with independence. Then submultiplicative functions allow to consider more generally the norm raised to any positive power. It turns out one can link submultiplicative functions and exponential functions. Indeed, if $g \geq 0$ satisfies $g(x + y) \leq cg(x)g(y)$ for all $x, y \in \mathbb{R}^d$, then for every $n \geq 1$, we have by induction:

$$g(x) = g(n^{-1}x + (1 - n^{-1})x) \leq cg(n^{-1}x)g(n^{-1}x + (1 - 2n^{-1})x) \leq c^{n-1}g(n^{-1}x)^n.$$

If in addition g is bounded on every compact set, then let $a = \max\{c^{-1}, \sup_{|x| \leq 1} g(x)\} < \infty$ and observe that for every $x \in \mathbb{R}^d$, for every $n \geq 1$ such that $n - 1 \leq |x| < n$, we have

$$g(x) \leq c^{n-1}g(n^{-1}x)^n \leq c^{n-1}a^n \leq a(ac)^{|x|}.$$

Combined with the definition of submultiplicativity, we infer that for every $x, y \in \mathbb{R}^d$, we have

$$g(x + y) \leq c g(x) g(y) \leq (ac)^{|x|+1} g(y). \quad (11.2)$$

Proof for exponential functions. Let us first prove the theorem when

$$g(x) = e^{ax} \quad \text{for some} \quad a \in \mathbb{R}^d \setminus \{0\}.$$

Assume first that $\mathbb{E}[e^{aX_t}] < \infty$ for some $t > 0$ and fix such a t . Let us decompose the process as usual as $X = G + Z + M$. By independence, we have

$$\mathbb{E}[e^{aX_t}] = \mathbb{E}[e^{a(G_t + M_t)}] \mathbb{E}[e^{aZ_t}],$$

and since $\mathbb{E}[e^{aX_t}] < \infty$ and $\mathbb{E}[e^{a(G_t + M_t)}] > 0$ (since $G_t + M_t$ is finite), then $\mathbb{E}[e^{aZ_t}] < \infty$. Recall that Z is a compound Poisson process with intensity $\pi(\cdot \cap \{|x| \geq 1\})$. Let $\lambda = \pi(\{|x| \geq 1\})$ and $\eta(\cdot) = \pi(\cdot \cap \{|x| \geq 1\})/\lambda$ denote the law of its jumps, then

$$\begin{aligned} \mathbb{E}[e^{aZ_t}] &= \sum_{n \geq 0} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \int e^{a(x_1 + \dots + x_n)} \prod_{i=1}^n \eta(dx_i) \\ &= \sum_{n \geq 0} \frac{t^n e^{-\lambda t}}{n!} \int \prod_{i=1}^n e^{ax_i} \mathbb{1}_{|x_i| \geq 1} \pi(dx_i) \\ &= \sum_{n \geq 0} \frac{t^n e^{-\lambda t}}{n!} \left(\int_{|x| \geq 1} e^{ax} \pi(dx) \right)^n. \end{aligned}$$

Recall that $\mathbb{E}[e^{aZ_t}] < \infty$ so the last series is finite, and in particular the summand for $n = 1$ is finite, so the integral is finite.

Suppose conversely that this integral is finite, then we may compute the last sum and find

$$\mathbb{E}[e^{aZ_t}] = \exp\left(t \int_{|x| \geq 1} (e^{ax} - 1) \pi(dx)\right),$$

which is finite for all $t > 0$. It only remains to prove that $\mathbb{E}[e^{a(G_t + M_t)}] < \infty$ for all $t > 0$, or equivalently, that $\mathbb{E}[e^{aX_t}] < \infty$ for all $t > 0$ when π is supported by the unit ball. Let us prove that in this case, we have in fact $\mathbb{E}[e^{aX_t}] < \infty$ for all $t > 0$. Note that $(aX_t)_{t \geq 0}$ is a linear combination of the components of X , and as such it is a Lévy process and, if ψ is the characteristic exponent of X , then that ψ_a of aX equals for all $z \in \mathbb{R}$:

$$\begin{aligned} \psi_a(z) &= \psi(za) \\ &= -iz\Gamma a + \frac{z^2}{2}(\Sigma a)^2 - \int_{|x| < 1} (e^{izax} - 1 - izax)\pi(dx) \\ &= -iz \underbrace{\left(\Gamma a + \int ax(\mathbb{1}_{|ax| < 1} - \mathbb{1}_{|x| < 1})\pi(dx) \right)}_{=\gamma} + \underbrace{\frac{z^2}{2}(\Sigma a)^2}_{=\sigma^2} - \int_{(-1,1)} (e^{izr} - 1 - izr)\pi_a(dr), \end{aligned}$$

where π_a is the image of π by the multiplication by a . Since π is supported by the unit ball, then π_a has bounded support as well. It follows that the formula above makes sense for any $z \in \mathbb{C}$ and yields an analytic function. To see this, write the integrand as a power series and use that π_a has bounded support to exchange summation and integration. This implies that the characteristic function of aX_t can be extended to an analytic function on $\mathbb{C}$, and in particular is infinitely differentiable. This further implies that all moments of aX_t are finite and we can write this characteristic function as the power series with infinite radius of convergence:

$$\mathbb{E}[e^{izaX_t}] = \sum_{n \geq 0} \frac{i^n}{n!} \mathbb{E}[(aX_t)^n] z^n, \quad z \in \mathbb{C}.$$

Now use that for any $b \in \mathbb{R}$, we have $|b|^{2k+1} \leq b^{2k} + b^{2k+2}$ (consider separately $|b| \geq 1$ and $|b| < 1$), hence for any $k \geq 0$, we have $\mathbb{E}[|aX_t|^{2k}] = \mathbb{E}[(aX_t)^{2k}]$ and $\mathbb{E}[|aX_t|^{2k+1}] \leq \mathbb{E}[(aX_t)^{2k}] + \mathbb{E}[(aX_t)^{2k+2}]$, so finally

$$\mathbb{E}[e^{|aX_t|}] = \sum_{n \geq 0} \frac{\mathbb{E}[|aX_t|^n]}{n!} \leq \sum_{k \geq 0} \frac{\mathbb{E}[(aX_t)^{2k}]}{(2k)!} + \sum_{k \geq 0} \frac{\mathbb{E}[(aX_t)^{2k}] + \mathbb{E}[(aX_t)^{2k+2}]}{(2k+1)!} < \infty.$$

This completes the proof for exponential functions. $\square$

Proof in the general case. Let us now consider a general nonnegative and submultiplicative function g that is bounded on compact sets. We shall rely on the case of exponential functions thanks to the bound (11.2).

Indeed, assume first that $\mathbb{E}[g(X_t)] < \infty$ for some $t > 0$ and let us fix such a t . Let us decompose again $X = G + Z + M$, then by independence and Fubini's theorem, we have:

$$\mathbb{E}[g(X_t)] = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(u+v) \mathbb{P}(Z_t \in du) \right) \mathbb{P}(G_t + M_t \in dv).$$

Since the left-side is finite, then the integral in parenthesis must be finite for some $v \in \mathbb{R}^d$. Fix such a v , then just as in the case of an exponential function, we infer that

$$\begin{aligned} \mathbb{E}[g(Z_t + v)] &= \sum_{n \geq 0} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \int g((x_1 + \dots + x_n) + v) \prod_{i=1}^n \eta(dx_i) \\ &= \sum_{n \geq 0} \frac{t^n e^{-\lambda t}}{n!} \int g((x_1 + \dots + x_n) + v) \prod_{i=1}^n \mathbb{1}_{|x_i| \geq 1} \pi(dx_i), \end{aligned}$$

where $\lambda = \pi(\{|x| \geq 1\})$ and $\eta(\cdot) = \pi(\cdot \cap \{|x| \geq 1\})/\lambda$. Apply the bound (11.2) with $x = -v$ and $y = x_1 + \dots + x_n + v$ to obtain:

$$\sum_{n \geq 0} \frac{t^n e^{-\lambda t}}{n!} \int g(x_1 + \dots + x_n) \prod_{i=1}^n \mathbb{1}_{|x_i| \geq 1} \pi(dx_i) \leq (ac)^{|v|+1} \sum_{n \geq 0} \frac{t^n e^{-\lambda t}}{n!} \int g((x_1 + \dots + x_n) + v) \prod_{i=1}^n \mathbb{1}_{|x_i| \geq 1} \pi(dx_i).$$

The series on the right is finite, hence so is that on the left, and in particular the summand for $n = 1$ is finite, hence

$$\int_{|x| \geq 1} g(x) \pi(dx) < \infty.$$

Suppose conversely that this integral is finite. Since g is submultiplicative, then

$$\begin{aligned} \int g(x_1 + \dots + x_n) \prod_{i=1}^n \mathbb{1}_{|x_i| \geq 1} \pi(dx_i) &\leq \int \prod_{i=1}^n c^{n-1} g(x_i) \mathbb{1}_{|x_i| \geq 1} \pi(dx_i) \\ &= c^{n-1} \left(\int_{|x| \geq 1} g(x) \pi(dx) \right)^n. \end{aligned}$$

It follows that for every $t > 0$, we have that

$$\begin{aligned} \mathbb{E}[g(Z_t)] &= \sum_{n \geq 0} \frac{t^n e^{-\lambda t}}{n!} \int g((x_1 + \dots + x_n)) \prod_{i=1}^n \mathbb{1}_{|x_i| \geq 1} \pi(dx_i) \\ &\leq c^{-1} \sum_{n \geq 0} \frac{e^{-\lambda t}}{n!} \left(tc \int_{|x| \geq 1} g(x) \pi(dx) \right)^n \\ &\leq c^{-1} \exp \left(t \int_{|x| \geq 1} (cg(x) - 1) \pi(dx) \right), \end{aligned}$$

which is finite. Applying (11.2) with $x = G_t + M_t$ and $y = Z_t$ yields, using independence:

$$\mathbb{E}[g(X_t)] \leq \mathbb{E}[g(Z_t)] \mathbb{E}[(ac)^{|G_t+M_t|+1}].$$

We just argued that $\mathbb{E}[g(Z_t)] < \infty$ and the proof in the case of exponential functions showed that, since $G + M$ has a Lévy measure with bounded support, then $\mathbb{E}[\alpha^{|G_t+M_t|}] < \infty$ for every $\alpha > 0$ and every $t > 0$. $\square$

EXAMPLE 11.17. In particular, for any $u \in \mathbb{R}^d$, we have

$$\mathbb{E}[e^{uX_t}] < \infty \iff \int_{|x| \geq 1} e^{ux} \pi(dx) < \infty$$

and for any $c > 0$, we have

$$\mathbb{E}[|X_t|^c] < \infty \iff \int_{|x| \geq 1} |x|^c \pi(dx) < \infty.$$

For the latter, the function $|x|^c$ is not submultiplicative, but $|x|^c \vee 1$ is and obviously $\mathbb{E}[|X_t|^c] < \infty$ if and only if $\mathbb{E}[|X_t|^c \vee 1] < \infty$.

COROLLARY 11.18. In dimension 1, if $\pi((-\infty, 0]) = 0$, then $\mathbb{E}[e^{-uX_t}] < \infty$ for every $u \geq 0$.

Hence, in the case of positive jumps only, we may work with the Laplace transform instead of the Fourier transform. This applies in particular to subordinators, i.e. nondecreasing Lévy processes, and subordinators plus a negative drift, but recall that they are Lévy processes that only make positive jumps and that are much more complicated than just a subordinators plus a negative drift (due to the required compensation in the martingale M)!

REMARK 11.19. When $\int (|x|^2 \wedge |x|) \pi(dx) < \infty$, that is, when X_t is integrable for any $t > 0$, we can compute its expectation by differentiating its characteristic function at 0 to get

$$\mathbb{E}[X_t] = t \left(\Gamma + \int x \mathbb{1}_{|x| \geq 1} \pi(dx) \right).$$

Similarly, when $\int |x|^2 \pi(dx) < \infty$, that is, when X_t is square-integrable for any $t > 0$, then its covariance matrix is

$$\text{Cov}(X_t) = t \left(\Sigma + \int xx^\top \pi(dx) \right).$$

As a last remark, recall that a Lévy process is a Markov process, and precisely $(X_{t+s} - X_t)_{s \geq 0}$ is independent of $\mathcal{F}_t = \sigma(X_u, u \leq t)$ and has the same law as X . When X_t is integrable for any $t > 0$, then we have that $\mathbb{E}[X_{t+s} | \mathcal{F}_t] = X_t + \mathbb{E}[X_s]$. Hence a Lévy process is a martingale if and only if it has 0 mean, that is, when

$$\int (|x|^2 \wedge |x|) \pi(dx) < \infty \quad \text{and} \quad \Gamma + \int x \mathbb{1}_{|x| \geq 1} \pi(dx) = 0.$$

Another way to visualise this is to recall the Lévy–Itô decomposition $X = G + Z + M$ where M is a martingale, $G_t = t\Gamma + \Sigma B_t$ where B is the Brownian motion, which is a martingale, and Z is compound Poisson process. Hence we require to match the parameter Γ with the expectation of the compound Poisson process to make the sum a martingale.

11.5 EXERCISES

EXERCISE 54. Let $X = (X_t)_{t \geq 0}$ be a subordinator, i.e. a one-dimensional, nondecreasing, Lévy process.

- (i) Prove that its Laplace transform is well-defined and given by:

$$\mathbb{E}[e^{-uX_t}] = \exp\left(-t\left(\delta u + \int_{\mathbb{R}} (1 - e^{-ux})\pi(dx)\right)\right), \quad t, u \geq 0$$

for some $\delta \geq 0$ and where $\pi((-\infty, 0]) = 0$ and $\int (x \wedge 1)\pi(dx) < \infty$.

- (ii) Compute $\mathbb{E}[X_t]$ for each $t > 0$.
 (iii) Give an example that satisfies $\mathbb{E}[X_t] = \infty$ for each $t > 0$.
 (iv) Prove that $X_t/t \rightarrow \delta$ in probability as $t \rightarrow \infty$.

EXERCISE 55. Let $B = (B_t)_{t \geq 0}$ be a one-dimensional standard Brownian motion and let τ be an independent exponential random time, say with parameter 1. Prove that B_τ is infinitely divisible and provides its characteristic triple.

EXERCISE 56. Let $B = (B_t^1, B_t^2)_{t \geq 0}$ be a two-dimensional standard Brownian motion. For every $a \geq 0$, let

$$T_a^1 = \inf\{t \geq 0 : B_t^1 = a\}.$$

- (i) Express the law of T_a^1 for each $a > 0$. (Use stochastic calculus!)
 (ii) Prove that $(T_a^1)_{a \geq 0}$ is a subordinator.
 (iii) Prove that $(B_{T_a^1}^2)_{a \geq 0}$ is a Lévy process; do you recognise it?

CHAPTER 12

STABLE LAWS & PROCESSES

We went very far from the central limit theorem to infinitely divisible distributions. Let us consider in this chapter those infinitely divisible distributions in $\mathbb{R}^d$ that appear as limits of rescaled sums of n i.i.d. random variables with a distribution that is fixed once and for all, just as in the central limit theorem (but assuming infinite variance). We shall characterise all the possible limit laws, with a much more explicit characteristic function than general infinitely divisible distributions. This chapter and the next one are mostly inspired by Feller and Sato.

CONTENTS

12.1	Stable laws	135
12.2	Stable Lévy processes	138
12.3	Domain of attraction	139
12.4	The Lévy–Khintchine formula	141
12.5	Moments and variation	145
12.6	An explicit formula for the characteristic function	146
12.7	Exercises	147

In this chapter, we focus on a subcase of infinitely divisible laws, namely we fix once and for all some random variable X in $\mathbb{R}^d$, we let $(X_n)_{n \geq 1}$ be i.i.d. copies of X and $S_n = X_1 + \dots + X_n$ for all $n \geq 1$. Now we ask about limits of sequences of the form

$$\frac{S_n}{a_n} - b_n \quad \text{where} \quad a_n > 0, b_n \in \mathbb{R}^d.$$

Do they exist? Can we characterise the possible limit laws? Can we find conditions on the law of X to ensure such a convergence? The d -dimensional central limit theorem provides a complete answer when X is square-integrable so we shall focus on the infinite variance regime.

In this chapter, we rather speak of random variables than probability measures, and we use throughout the chapter the notation $S_n = X_1 + \dots + X_n$ where the X_i 's are i.i.d. with law μ , and we let also X be a random variable with this law. We also write

$$X \stackrel{(d)}{=} Y$$

to mean that the random variables X and Y have the same law.

12.1 STABLE LAWS

Let us define the first main notion of this chapter.

DEFINITION 12.1. A random variable X has a *stable law* if it is not almost surely constant and for every $n \geq 1$, there exist $a_n > 0$ and $b_n \in \mathbb{R}^d$ such that

$$\frac{S_n}{a_n} - b_n \stackrel{(d)}{=} X, \quad \text{or equivalently,} \quad S_n \stackrel{(d)}{=} a_n X + c_n,$$

where $c_n = a_n b_n$. The law is said to be *strictly stable* if $b_n = c_n = 0$.

Clearly, stable laws are infinitely divisible, but not all infinitely divisible laws are stable.

EXERCISE 12.2. Prove that the Gaussian laws are stable, and strictly stable if and only if their expectation is 0, that the Cauchy distribution is strictly stable, but that Poisson laws or exponential laws are not stable.

It turns out that in a stable law, the scaling factor a_n is necessarily a power of n , with a limited range of values. In addition, outside the, quite special, case of the exponent 1, we can always shift the distribution by a constant to make it strictly stable, that is, with $b_n = 0$. We shall see a second proof later, relying on the theory of infinitely divisible distributions, but we can prove this directly by hand just from the definition.

PROPOSITION 12.3. In Definition 12.1, there exists $\alpha \in (0, 2]$ such that $a_n = n^{1/\alpha}$ for all $n \geq 1$. In addition, when $\alpha \neq 1$, there exists $b \in \mathbb{R}$ such that the law of $X + b$ is stable with the same $a_n = n^{1/\alpha}$ but now $b_n = 0$.

To get a feeling about this, notice that if X has finite variance, with covariance matrix Σ^2 , and if X has a stable law, then S_n has finite variance, and covariance matrix $n\Sigma^2$, but it also has the same law as $a_n X + c_n$ and so also covariance matrix $a_n^2 \Sigma^2$. This shows that stable laws with finite variance necessarily have $a_n = \sqrt{n}$. This argument will be extended in the proof of the proposition to exclude the possibility of a_n being a power of n smaller than $1/2$ by showing that this would yield a finite-variance law.

As for the shift property, suppose that X is stable, fix any $b \in \mathbb{R}^d$ and let $X' = X + b$ and S'_n the corresponding sum, then

$$\frac{S'_n}{a_n} = \frac{S_n}{a_n} + \frac{nb}{a_n} \stackrel{(d)}{=} X + b_n + \frac{nb}{a_n} = X' + b_n - b \frac{n - a_n}{a_n}.$$

If X has a well-defined and finite expectation, then for $b = -\mathbb{E}[X]$, both X' and then S'_n have zero expectation, which implies necessarily that the deterministic term on the right is zero, so $a_n b_n / (n - a_n) = -\mathbb{E}[X]$ does not depend on n , and X' is strictly stable. It turns out that such a shift also exists when X has no expectation, at least when $a_n \neq n$ (otherwise the previous ratio $a_n b_n / (n - a_n)$ is ill-defined), so when $\alpha \neq 1$.

DEFINITION 12.4. The real number $\alpha \in (0, 2]$ such that a stable law has $a_n = n^{1/\alpha}$ for all $n \geq 1$ is called the *index of stability* of the law. A stable random variable and its law are then said to be α -stable.

EXAMPLE 12.5. The index of a Gaussian law is 2 and that of the Cauchy distribution is 1.

We will see later that all values of $\alpha \in (0, 2]$ are possible and further that:

- stable laws with a finite variance are necessarily Gaussian laws,
- and for every $\alpha \in (0, 2)$, a random variable X with an α -stable law has $\mathbb{E}[|X|^c] < \infty$ if and only if $c \in (0, \alpha)$; in particular non Gaussian stable laws all have infinite variance and the index $\alpha = 1$ is the threshold to have a well-defined expectation,

- in the case $\alpha = 1$, we have $b_n = b \log n$ for some $b \in \mathbb{R}^d$, and $b = 0$ if and only if X is a scaled Cauchy random variable.

Let us now prove Proposition 12.3.

Proof of Proposition 12.3. Let us first consider symmetric distributions, i.e. when X and $-X$ have the same law. In this case we must have $b_n = 0$, i.e. $S_n = a_n X$ in distribution.

First, for every $n_1, \dots, n_k \geq 1$, the sum $S_{n_1+\dots+n_k}$, which has the same law as $a_{n_1+\dots+n_k} X$, also has the same law as the sum of k independent random variables with law $S_{n_1}, \dots, S_{n_k}$ respectively, and hence the same law as $a_{n_1} X_1 + \dots + a_{n_k} X_k$. We thus have the identity in law:

$$a_{n_1+\dots+n_k} X \stackrel{(d)}{=} a_{n_1} X_1 + \dots + a_{n_k} X_k.$$

For $n_1 = \dots = n_k = n$, this yields

$$a_{kn} X \stackrel{(d)}{=} a_n S_k \stackrel{(d)}{=} a_n a_k X.$$

This implies that $a_{kn} = a_n a_k$ for every $n, k \geq 1$, for if it were not the case, then X would have the same law as cX for some $0 < c < 1$ (the ratio $\min\{a_{kn}, a_n a_k\} / \max\{a_{kn}, a_n a_k\}$), and by induction X would have the same law as $c^n X \rightarrow 0$ almost surely, but we excluded the case $X = 0$ almost surely. We infer by induction that

$$a_{n^\ell} = a_n^\ell \quad \text{for all } n, \ell \geq 1.$$

Remark 9.9 shows that $X/a_n \rightarrow 0$ in probability, so in particular $a_n \rightarrow \infty$. We infer that for every $n \geq 1$ fixed, we must have $a_{n^\ell} = a_n^\ell \rightarrow \infty$ as $\ell \rightarrow \infty$, and therefore $a_n > 1$. We may thus write $a_n = n^{1/\alpha_n}$ where $\alpha_n = \log n / \log a_n > 0$ and our goal is to show that α_n is constant equal to some $\alpha \in (0, 2]$.

Let us now prove that the ratios a_n/a_{n+m} are uniformly bounded over all $n, m \geq 1$. We may write S_{n+m} as the sum of two independent random variables, one with law S_n , same as $a_n X_1$ and one with law S_m , same as $a_m X_2$. For every $x > 0$, since X_2 has a symmetric distribution and is independent of X_1 , which has the same law as X , then

$$\mathbb{P}(X > x) \leq 2 \mathbb{P}(a_n X_1 > x a_n, X_2 \geq 0) \leq 2 \mathbb{P}(a_n X_1 + a_m X_2 > x a_n) = 2 \mathbb{P}\left(X > x \frac{a_n}{a_{n+m}}\right).$$

Since X is not almost surely null, then for some $x > 0$, the left side is some positive value; this implies that the ratio a_n/a_{n+m} cannot be made arbitrarily large, i.e. that these are uniformly bounded over all $n, m \geq 1$.

For every m , then for every $m \neq n$, both different from 1 and for every $k \geq 1$, there exists $\ell \geq 1$ such that $n^\ell < m^k \leq n^{\ell+1}$ and then, since $\alpha_m > 0$,

$$\frac{a_{m^k}}{a_{n^{\ell+1}}} = \frac{a_m^k}{a_n^{\ell+1}} = \frac{m^{k/\alpha_m}}{n^{(\ell+1)/\alpha_n}} \geq \frac{n^{\ell/\alpha_m}}{n^{(\ell+1)/\alpha_n}}.$$

Recall that the ratios on the left are uniformly bounded whilst, $n, m \geq 2$ being fixed, if k is arbitrary large, then so is ℓ and therefore, in order that the ratio on the right be bounded, we must have $\alpha_m \geq \alpha_n$. The symmetric argument shows that $\alpha_m = \alpha_n = \alpha$ for some $\alpha > 0$.

At this point, we have shown that, for any symmetric stable distribution, we have $a_n = n^\alpha$ for every $n \geq 1$, where $\alpha > 0$ is some positive value. Let us now prove that $\alpha \leq 2$. We saw earlier that if X had finite variance, then necessarily $a_n = \sqrt{n}$. To conclude that $\alpha \leq 2$, we will prove by contradiction that if we had $\alpha > 2$, then X would have finite variance. Fix $n \geq 2$ and let $I_n = \inf\{1 \leq i \leq n : |X_i| = \max_{1 \leq j \leq n} |X_j|\}$. Since the law of X is symmetric, then the pairs $(|X_{I_n}|, S_n - |X_{I_n}|)$ and $(|X_{I_n}|, -(S_n - |X_{I_n}|))$ have the same

law, so for every $x > 0$,

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq j \leq n} |X_j| > xa_n\right) &\leq \mathbb{P}(|X_{I_n}| > xa_n, S_n - |X_{I_n}| \geq 0) + \mathbb{P}(|X_{I_n}| > xa_n, S_n - |X_{I_n}| \leq 0) \\ &\leq 2 \mathbb{P}(|X_{I_n}| > xa_n, S_n - |X_{I_n}| \geq 0) \\ &= 2 \mathbb{P}(S_n > xa_n). \end{aligned}$$

Since S_n/a_n has the same law as X , then by choosing $x > 0$ such that $\mathbb{P}(X > x) \leq 1/4$, we infer that

$$\mathbb{P}\left(\max_{1 \leq j \leq n} |X_j| > xa_n\right) \leq 2 \mathbb{P}(X > x) \leq 1/2.$$

Then we infer together with the bound $e^{-t} \geq 1 - t$ for all $t \in [0, 1]$ that

$$\exp(-n \mathbb{P}(|X| > xa_n)) \geq \mathbb{P}(|X| \leq xa_n)^n = \mathbb{P}\left(\max_{1 \leq j \leq n} |X_j| \leq xa_n\right) \geq 1/2.$$

Hence $\sup_n n \mathbb{P}(|X| > xn^{1/\alpha}) < \infty$ and then $\sup_{t>x} t^\alpha \mathbb{P}(|X| > t) < \infty$. Let M be larger than this supremum, then for every $k \geq 1$, we have

$$\mathbb{E}[X^2 \mathbb{1}_{2^{k-1} < |X| \leq 2^k}] \leq 2^{2k} \mathbb{P}(|X| > 2^{k-1}) \leq 2^{2k} 2^{-(k-1)\alpha} M \leq 2^{(2-\alpha)k} 2^\alpha M.$$

If $\alpha > 2$, then

$$\mathbb{E}[X^2] = \sum_{k \geq 1} \mathbb{E}[X^2 \mathbb{1}_{2^{k-1} < |X| \leq 2^k}] < \infty.$$

Hence X has finite variance, which implies that $a_n = \sqrt{n}$ and contradicts that $\alpha > 2$.

Let us finally remove the assumption of symmetry and consider a general stable distribution. The claim $a_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$ remains true after replacing X by the difference of two i.i.d. copies of X , let us focus on b_n . For every $n, k \geq 2$, the sum S_{nk} has the same law as the sum of k i.i.d. random variables with the same law as S_n , and hence:

$$S_{nk} \stackrel{(d)}{=} \sum_{i=1}^k a_n(X_i + b_n) = a_n S_k + k a_n b_n \stackrel{(d)}{=} a_n a_k X + a_n a_k b_k + k a_n b_n.$$

By exchanging the role of k and n , we obtain for all $n, k \geq 2$:

$$a_n a_k b_k + k a_n b_n = a_n a_k b_n + n a_k b_k, \quad \text{or equivalently,} \quad \frac{a_k b_k}{a_k - k} = \frac{a_n b_n}{a_n - n},$$

where the assumption $\alpha \neq 1$ ensures that the denominators do not vanish. We conclude that there exists $b \in \mathbb{R}$ such that $a_n b_n = b(a_n - n)$ for all $n \geq 2$ and so finally

$$\sum_{i=1}^n (X_i + b) = S_n + nb \stackrel{(d)}{=} a_n(X + b_n) + nb = a_n(X + b),$$

and the proof is complete. □

12.2 STABLE LÉVY PROCESSES

Recall that there is a one-to-one correspondance between infinitely divisible laws and Lévy processes.

DEFINITION 12.6. A Lévy process $X = (X_t)_{t \geq 0}$ is said to be *stable* when its law at time $t = 1$ is *strictly stable* (!). Recall that a stable law has an index $\alpha \in (0, 2]$, then we say that the Lévy process is α -stable.

EXAMPLE 12.7. In stochastic calculus, you find that the process of hitting times of a one-dimensional Brownian motion is a strictly stable subordinator, and so is its inverse local time.

Stable processes generalise the well-known *scaling property* of Brownian motion.

PROPOSITION 12.8. Let $(X_t)_{t \geq 0}$ be a Lévy process and let $\alpha \in (0, 2]$. The following assertions are equivalent:

- (i) it is α -stable,
- (ii) the law of X_t is strictly α -stable for any $t > 0$,
- (iii) for any $c > 0$, the process $(c^{-1/\alpha} X_{ct})_{t \geq 0}$ has the same law as $(X_t)_{t \geq 0}$.

Proof. Let us denote by $\hat{\mu}_t$ the characteristic function of X_t . That X_1 is strictly α -stable means that

$$\hat{\mu}_1(un^{-1/\alpha})^n = \hat{\mu}_1(u)$$

for every $n \geq 1$ and every $u \in \mathbb{R}^d$. Recall that in a Lévy process, we have $\hat{\mu}_1 = \exp(-\psi)$ for some function ψ , and further $\hat{\mu}_t = \exp(-t\psi)$. The previous identity then extends to any $t > 0$, showing that (i) implies (ii). Obviously Property (ii) implies conversely (i). Let us prove that (ii) and (iii) are equivalent.

Condition (iii) implies in particular that for any integer $c = n$ and any fixed $t > 0$, the random variables X_t and $n^{-1/\alpha} X_{nt} = n^{-1/\alpha} \sum_{i=1}^n (X_{it} - X_{(i-1)t})$ have the same law, and since the summands are i.i.d. with the same law as X_t , then the latter has a strictly α -stable law. Hence (iii) implies (ii). To see that (ii) implies (iii), let us fix two integers $p, q \geq 1$. Property (ii) with $t = 1/q$ shows the identity in distribution:

$$X_{p/q} \stackrel{(d)}{=} \sum_{i=1}^p (X_{i/q} - X_{(i-1)/q}) \stackrel{(d)}{=} p^{1/\alpha} X_{1/q}$$

since the summands in the middle are i.i.d. with the same law as $X_{1/q}$. Taking $p = q$, we read

$$X_1 \stackrel{(d)}{=} q^{1/\alpha} X_{1/q},$$

and the combination of these displays shows that

$$c^{-1/\alpha} X_c \stackrel{(d)}{=} X_1$$

for every rational value of $c = p/q$. It extends to irrational values by considering a rational sequence that decreases to c and using right-continuity of X . To conclude about Property (iii), observe that the process $(c^{-1/\alpha} X_{ct})_{t \geq 0}$ is a Lévy process: the transformation does not affect the càdlàg property and the increments remain independent and stationary. Since this Lévy process has the same law at time $t = 1$ as $(X_t)_{t \geq 0}$, then the entire processes have the same law. $\square$

We can note already that when $\alpha = 2$, Property (iii) above is the usual scaling property of Brownian motion. We shall see indeed that in this case X is an affine transform of a Brownian motion.

12.3 DOMAIN OF ATTRACTION

Just as for infinitely divisible laws, let us weaken the definition by considering limits of normalised sums.

DEFINITION 12.9. The (law of a) random variable X belongs to the *domain of attraction* of the (law of a) random variable Y when there exist sequences $(\alpha_n)_n$ in $(0, \infty)$ and $(\beta_n)_n$ in $\mathbb{R}^d$ such that

$$\frac{S_n}{\alpha_n} - \beta_n \xrightarrow[n \rightarrow \infty]{(d)} Y.$$

If Y has a stable law, then it belongs to its own domain of attraction. It turns out only stable laws admit a domain of attraction.

THEOREM 12.10. If the law of X belongs to the domain of attraction of the law of Y , then Y has a stable law.

Proof. Fix $k \geq 1$, we want to prove that if $Y_1, \dots, Y_k$ are independent random variables with the same law as Y , then their sum has the same law as a linear function of Y . Let us use the notation $S_n^j = \sum_{i=1}^n X_{(j-1)n+i}$, which are independent copies of S_n whose sum over $j = 1, \dots, k$ equals S_{nk} . Then we let us decompose:

$$\frac{1}{\alpha_{nk}} S_{nk} - \beta_{nk} = \frac{\alpha_n}{\alpha_{nk}} \sum_{j=1}^k \left(\frac{1}{\alpha_n} S_n^j - \beta_n \right) + \frac{k\alpha_n\beta_n}{\alpha_{nk}} - \beta_{nk}.$$

The term on the left converges in distribution to Y as $n \rightarrow \infty$, and so do each term in parenthesis on the right and since they are i.i.d. as j varies, then the entire sum converges in distribution to $Y_1 + \dots + Y_k$. The claim then follows from the next lemma which shows that the scaling factor α_n/α_{nk} converges to some $1/A_k$ and the shift term $k\alpha_n\beta_n/\alpha_{nk} - \beta_{nk}$ to some $-B_k$, so indeed Y has the same law as $A_k^{-1}(Y_1 + \dots + Y_k) - B_k$, for every $k \geq 1$. $\square$

The following lemma is sometimes called the *convergence of type theorem*. In this context, the *type* of the law of a random variable X is the collection of the law of all random variables $aX + b$ with $a \neq 0$ and $b \in \mathbb{R}$. Note that stability of a distribution is in fact a property on its type, in the sense that X has a stable law if and only if every random variable with the same type has.

LEMMA 12.11 (Convergence of types). Let Z and Z' be two random variables in $\mathbb{R}^d$ that are not almost surely constant. Suppose that $Z_n \rightarrow Z$ in distribution and that $\alpha_n > 0$ and $c_n \in \mathbb{R}^d$ are such that $aZ_n + c_n \rightarrow Z'$ in distribution. Then there exist $a = \lim_n \alpha_n \in (0, \infty)$ and $c = \lim_n c_n \in \mathbb{R}^d$ such that Z' has the same law as $aZ + c$.

Proof. From the perspective of characteristic functions, the assumptions read:

$$\mathbb{E}[e^{iuZ_n}] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[e^{iuZ}] \quad \text{and} \quad \mathbb{E}[e^{i\alpha_n u Z_n}] e^{ic_n u} \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[e^{iuZ'}],$$

for every $u \in \mathbb{R}^d$ and even uniformly on compact sets. In particular, we have

$$|\mathbb{E}[e^{iuZ_n}]| \xrightarrow[n \rightarrow \infty]{} |\mathbb{E}[e^{iuZ}]| \quad \text{and} \quad |\mathbb{E}[e^{i\alpha_n u Z_n}]| \xrightarrow[n \rightarrow \infty]{} |\mathbb{E}[e^{iuZ'}]|, \quad (12.1)$$

uniformly on compact sets.

Let us extract a subsequence along which α_n converges to some $a \in [0, \infty]$. Let us prove that a cannot be infinite or null, before proving that it does not depend on the subsequence, so in fact $\alpha_n \rightarrow a \in (0, \infty)$ without the need to extract a subsequence. Suppose first that $a = \infty$ and let us take limit along this subsequence only. Then for every $v \in \mathbb{R}^d$, we have $v/\alpha_n \rightarrow 0$ so this sequence takes values in a compact set and so we may apply (12.1) with $u = v$ on the left and $u = v/\alpha_n$ on the right to get both:

$$|\mathbb{E}[e^{ivZ_n}]| \xrightarrow[n \rightarrow \infty]{} |\mathbb{E}[e^{ivZ}]| \quad \text{and} \quad |\mathbb{E}[e^{ivZ_n}]| \xrightarrow[n \rightarrow \infty]{} |\mathbb{E}[e^{i0Z'}]| = 1.$$

It follows that $|\mathbb{E}[e^{iuZ}]| = 1$ for every $u \in \mathbb{R}^d$, which implies that Z is almost surely constant, which we excluded. Hence $a < \infty$. If instead $a = 0$, then symmetrically, applying (12.1) with $u = v\alpha_n$ on the left and $v = u$ on the right yields both

$$|\mathbb{E}[e^{i\alpha_n v Z_n}]| \xrightarrow{n \rightarrow \infty} |\mathbb{E}[e^{ivZ'}]| \quad \text{and} \quad |\mathbb{E}[e^{i\alpha_n v Z_n}]| \xrightarrow{n \rightarrow \infty} |\mathbb{E}[e^{i0Z}]| = 1,$$

so $|\mathbb{E}[e^{ivZ'}]| = 1$ for every $v \in \mathbb{R}^d$, which implies that Z' is almost surely constant, which we excluded as well.

It follows that any subsequential limit a of α_n lies in $(0, \infty)$. Along such a subsequence, by uniform convergence again, for every $u \in \mathbb{R}^d$, the sequence $|\mathbb{E}[e^{i\alpha_n u Z_n}]|$ converges both to $|\mathbb{E}[e^{iauZ}]|$ and to $|\mathbb{E}[e^{iuZ'}]|$, so we must have

$$|\mathbb{E}[e^{iauZ}]| = |\mathbb{E}[e^{iuZ'}]|, \quad u \in \mathbb{R}^d.$$

This holds for any subsequential limit a of α_n . Suppose that $0 < a' < a$ is another one, then for every $u \in \mathbb{R}^d$, we have $|\mathbb{E}[e^{iauZ}]| = |\mathbb{E}[e^{ia'uZ}]|$. Replacing u by u/a and iterating n times reads:

$$|\mathbb{E}[e^{iauZ}]| = |\mathbb{E}[e^{i(a'/a)uZ}]| = |\mathbb{E}[e^{i(a'/a)^n uZ}]|, \quad n \geq 1.$$

However the right side converges to $|\mathbb{E}[e^{i0uZ}]| = 1$, so we find again $|\mathbb{E}[e^{iauZ}]| = 1$ for all u , which is excluded. It follows that no two subsequential limits differ, hence α_n converges to some $a \in (0, \infty)$ without extraction. In addition, for this limit a , we have

$$\mathbb{E}[e^{i\alpha_n u Z_n}] \xrightarrow{n \rightarrow \infty} \mathbb{E}[e^{iauZ}], \quad u \in \mathbb{R}^d.$$

Recall that we also have

$$\mathbb{E}[e^{i\alpha_n u Z_n}] e^{ic_n u} \xrightarrow{n \rightarrow \infty} \mathbb{E}[e^{iuZ'}], \quad u \in \mathbb{R}^d,$$

then $e^{ic_n u}$ must converge for all u in a small neighbourhood of the origin, where $|\mathbb{E}[e^{iauZ}]| > 0$, to a limit which is continuous at $u = 0$. This implies (exercise!) that c_n must converge to some $c \in \mathbb{R}^d$. $\square$

EXERCISE 12.12. Prove that if a random variable X in $\mathbb{R}^d$ has $|\varphi_X| = 1$, then X is constant almost surely.

REMARK 12.13. Note that in Definition 12.9 of the domain of attraction, the scaling sequence satisfies $\alpha_n \rightarrow \infty$. Indeed, if $\beta_n = 0$ for each n , then $\alpha_n^{-1} S_n = \sum_{i=1}^n \xi_{n,i}$ where $\xi_{n,1} = \alpha_n^{-1} X_i$ and Remark 9.9 implies that $\xi_{n,i} = \alpha_n^{-1} X_1$ tends to 0, so indeed $\alpha_n \rightarrow \infty$.

In the general case, observe that if $(X'_n)_{n \geq 1}$ are i.i.d. with the same law as X and are independent of $(X_n)_{n \geq 1}$, then the corresponding sum satisfies $\alpha_n^{-1} S'_n - \beta_n \rightarrow Y'$ in distribution, where Y' has the same law as Y and is independent of it. It follows that $\alpha_n^{-1} (S_n - S'_n) \rightarrow (Y - Y')$ in distribution, and then the previous argument applies to this symmetric version.

12.4 THE LÉVY-KHINTCHINE FORMULA

Clearly, a stable law is infinitely divisible, and thus satisfies the Lévy-Khintchine formula, namely

$$\hat{\mu}(u) = \exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2 + \int_{\mathbb{R}^d} (e^{iux} - 1 - iux \mathbb{1}_{|x|<1}) \pi(dx)\right),$$

for some $\Gamma \in \mathbb{R}^d$, some symmetric matrix Σ , and a Lévy measure π , which satisfies $\pi(\{0\}) = 0$ and $\int (|x|^2 \wedge 1) \pi(dx) < \infty$. The next theorem identifies explicitly the characteristic triple of a stable law. In particular, stable laws have either a nonzero measure π or a nonzero Gaussian parameter Σ , but not both

at the same time! Although we already proved the first item, we shall now provide another, independent, argument.

THEOREM 12.14. Let μ be a stable law with characteristic triple (Γ, Σ^2, π) . The following assertions hold.

- (i) There exists $\alpha \in (0, 2]$ such that $a_n = n^{1/\alpha}$ for all $n \geq 1$.
- (ii) We have $\alpha = 2$ if and only if μ is a Gaussian law, with mean vector Γ and nonzero covariance matrix Σ^2 . In particular $\pi = 0$.
- (iii) When $\alpha \in (0, 2)$, we have $\Sigma = 0$ and the Lévy measure takes the form:

$$\pi(A) = \int_S \lambda(dz) \int_0^\infty \frac{1}{r^{1+\alpha}} \mathbb{1}_{rz \in A} dr, \quad (12.2)$$

where $S = \{x \in \mathbb{R}^d : |x| = 1\}$ is the unit sphere and λ is a finite measure on S different from the null measure.

- (iv) Let $\bar{\lambda} = \int_S z \lambda(dz) \in \mathbb{R}^d$. Then we have $b_n = \bar{\lambda} \log n$ when $\alpha = 1$, and when $\alpha \in (0, 1) \cup (1, 2)$, the shifted random variable $X - b$ is strictly stable, with the same $a_n = n^{1/\alpha}$ but now $b_n = 0$, where $b = \Gamma - (1 - \alpha)^{-1} \bar{\lambda}$.

Proof. If X is stable, then S_n and $a_n X + a_n b_n$ have the same law and both are infinitely divisible. It follows that their triple are equal. It is immediate to check that the characteristic triple of S_n is $(n\Gamma, \sqrt{n}\Sigma, n\pi)$ and that of $a_n X + a_n b_n$ is $(\Gamma_{a_n, b_n}, a_n \Sigma, \pi_{a_n})$ where π_{a_n} is the image of π by the multiplication $x \mapsto a_n x$ and

$$\Gamma_{a_n, b_n} = a_n \left(\Gamma + b_n - \int_{1/a_n \leq |x| < 1} x \pi(dx) \right).$$

We thus have

$$n\Gamma = \Gamma_{a_n, b_n}, \quad \sqrt{n}\Sigma = a_n \Sigma, \quad n\pi = \pi_{a_n}.$$

We already observe that either $\Sigma = 0$ or $a_n = \sqrt{n}$ (or both).

STEP 1: GAUSSIAN LAWS. The identity $\pi = n^{-1} \pi_{a_n}$ implies that

$$\int (|x|^2 \wedge 1) \pi(dx) = \int \frac{(a_n |x|)^2 \wedge 1}{n} \pi(dx) = \int \frac{(a_n |x|)^2 \wedge 1}{n(|x|^2 \wedge 1)} (|x|^2 \wedge 1) \pi(dx).$$

(Since $\pi(\{0\}) = 0$, we may exclude 0 and so $|x|^2 \wedge 1 > 0$.) Suppose that $a_n/\sqrt{n}$ does not converge to infinity, then extracting a subsequence if necessary, we may assume that it is bounded by some constant $C < \infty$, and then the ratio in the last integral is bounded by C^2 . In addition, it converges to 0 for each $x \neq 0$ fixed, since $a_n \rightarrow \infty$ as observed in Remark 12.13. Since π integrates $x^2 \wedge 1$, then by dominated convergence, the integral on the right tends to 0, which means that π is the constant null measure in this case. Since we exclude Dirac masses, then we must have $\Sigma \neq 0$ and so $a_n = \sqrt{n}$ due to the identity $a_n \Sigma = n \Sigma$. In addition, in this case, by matching the first elements in the triples, we have $n\Gamma = \Gamma_{a_n, a_n b_n} = a_n \Gamma + a_n b_n$, and therefore

$$a_n b_n = (n - a_n) \Gamma = (n - \sqrt{n}) \Gamma.$$

We have thus shown that if $(a_n)_n$ admits a subsequence that remains bounded by some constant times $\sqrt{n}$, then in fact $a_n = \sqrt{n}$ for all n , also $b_n = (\sqrt{n} - 1) \Gamma$, and μ is a Gaussian law with mean vector Γ and nonzero covariance matrix Σ^2 . Note in particular that $X - \Gamma$ has a strictly stable law.

STEP 2: $a_n = n^{1/\alpha}$ FOR SOME $\alpha \in (0, 2)$. Now suppose that $a_n/\sqrt{n} \rightarrow \infty$. The identity $a_n \Sigma = n \Sigma$ implies then that $\Sigma = 0$. Since we exclude Dirac masses, then π is not the null measure. For every $r > 0$,

let $F(r) = \pi(\{x \in \mathbb{R}^d : |x| \geq r\})$. Note that F is a nonincreasing function on $(0, \infty)$ that tends to 0 at infinity and since π is not the null measure, then there exists $r_0 > 0$ such that $F(r_0) > 0$. The identity $n\pi = \pi_{a_n}$ implies that $nF(r) = F(r/a_n)$ for every $r > 0$. By induction, we get:

$$F\left(\left(\frac{a_{n+1}}{a_n}\right)^k r_0\right) = \left(\frac{n}{n+1}\right)^k F(r_0), \quad k \in \mathbb{Z}.$$

Letting $k \rightarrow \infty$, then right-hand side converges to 0, which implies that $a_{n+1} > a_n$. Define then

$$\alpha_n = \frac{\log((n+1)/n)}{\log(a_{n+1}/a_n)} > 0, \quad \text{so} \quad \frac{a_{n+1}}{a_n} = \left(\frac{n+1}{n}\right)^{1/\alpha_n}.$$

The previous identity then reads

$$F\left(\left(\frac{a_{n+1}}{a_n}\right)^k r_0\right) = \left(\left(\frac{a_{n+1}}{a_n}\right)^k\right)^{-\alpha_n} F(r_0), \quad k \in \mathbb{Z}.$$

Now for any $r > 0$, there exists $k \in \mathbb{Z}$ such that $(a_{n+1}/a_n)^k r_0 < r \leq (a_{n+1}/a_n)^{k+1} r_0$. Since F is nonincreasing, then

$$\frac{F(r)}{F(r_0)} \leq \frac{F((a_{n+1}/a_n)^k r_0)}{F(r_0)} = \left(\left(\frac{a_{n+1}}{a_n}\right)^k\right)^{-\alpha_n} \leq \left(\frac{a_n}{a_{n+1}} \frac{r}{r_0}\right)^{-\alpha_n} = \frac{n+1}{n} \left(\frac{r}{r_0}\right)^{-\alpha_n}.$$

Similarly, for every $m \geq 1$ we have

$$\frac{F(r)}{F(r_0)} \geq \frac{m}{m+1} \left(\frac{r}{r_0}\right)^{-\alpha_m}.$$

Consequently,

$$\frac{mn}{(m+1)(n+1)} \leq \left(\frac{r}{r_0}\right)^{\alpha_m - \alpha_n},$$

for every $m, n \geq 1$ and every $r > 0$. A symmetric argument shows that the left side is also upper bounded by $(r/r_0)^{\alpha_n - \alpha_m}$ and therefore this exponent must be null. Hence all the α_n 's are equal to some $\alpha > 0$, and by their very definition, we have $a_{n+1}/a_n = ((n+1)/n)^{1/\alpha}$. Since $a_1 = 1$, then this implies that

$$a_n = n^{1/\alpha},$$

and since we assume that $a_n/\sqrt{n} \rightarrow \infty$, then $\alpha \in (0, 2)$.

STEP 3: EXPRESSION OF π . Up to this point, we have proved that $a_n = n^{1/\alpha}$ with $\alpha \in (0, 2]$, that μ is a Gaussian law $\mathcal{N}(\Gamma, \Sigma)$ when $\alpha = 2$, and that on the contrary, when $\alpha < 2$, we have $\Sigma = 0$. Let us now express π in this case. Recall the bounds for every $m, n \geq 1$ and $r > 0$:

$$\frac{m}{m+1} \left(\frac{r}{r_0}\right)^{-\alpha} \leq \frac{F(r)}{F(r_0)} \leq \frac{n+1}{n} \left(\frac{r}{r_0}\right)^{-\alpha}, \quad \text{so} \quad F(r) = \left(\frac{r_0}{r}\right)^\alpha F(r_0),$$

by letting both $m, n \rightarrow \infty$. In particular F never vanishes and we may a posteriori take $r_0 = 1$ and finally, we obtain:

$$\pi(\{|x| \geq r\}) = \frac{F(1)}{r^\alpha}, \quad r > 0.$$

Recall from the beginning of this proof that for any $n \geq 1$, the measure $n\pi$ coincides with the image of π by the multiplication by $a_n = n^{1/\alpha}$. This means that for any set $A \in \mathcal{B}(\mathbb{R}^d)$, we have $n\pi(A) = \pi(n^{-1/\alpha}A)$ for every $n \geq 1$. By applying this identity to two different indices $n, m \geq 1$ and to the set $n^{1/\alpha}A$ instead of A , we obtain:

$$\pi(A) = \frac{n}{m} \pi((n/m)^{1/\alpha}A), \quad A \in \mathcal{B}(\mathbb{R}^d), n, m \geq 1.$$

For any subset $B \subset \mathbb{S}$ of the unit sphere and any $s > 0$, let $(s, \infty) \cdot B = \{rz : r > s, z \in B\}$ denote the union of all rays started from distance s from the origin and whose projection on the sphere lies in B . Note that $(s, \infty) \cdot B = s((1, \infty) \cdot B)$. Then for any set $B \in \mathcal{B}(\mathbb{S})$ and any positive rational number s , we have with $A = (1, \infty) \cdot B$:

$$\pi((s^{1/\alpha}, \infty) \cdot B) = s^{-1} \pi((1, \infty) \cdot B) = \alpha \pi((1, \infty) \cdot B) \int_{s^{1/\alpha}}^{\infty} \frac{1}{r^{1+\alpha}} dr.$$

Note that

$$\lambda(B) = \alpha \pi((1, \infty) \cdot B), \quad B \in \mathcal{B}(\mathbb{S}),$$

defines a finite measure λ on the unit sphere $\mathbb{S}$, and the previous identities show that the measures

$$\pi(A) \quad \text{and} \quad \int_{\mathbb{S}} \lambda(dz) \int_0^{\infty} \frac{1}{r^{1+\alpha}} \mathbb{1}_{rz \in A} dr$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$ agree on the collection of sets A of the form $(t, \infty) \cdot B$ where B varies in $\mathcal{B}(\mathbb{S})$ and t varies in $\{s^{1/\alpha} : s \in \mathbb{Q} \cap (0, \infty)\}$. Fix any $\varepsilon > 0$ and consider only those sets for which $t > \varepsilon$. Then they are closed under finite intersection and generate the Borel σ -algebra on the complement of the closed ball with radius ε in $\mathbb{R}^d$. In addition, outside this ball, the above measures have the same finite mass, so the monotone class lemma implies that they agree on the entire complement of the ball with radius ε . Since this holds for any ε and since both measures give zero mass to the origin, then, they must agree everywhere.

STEP 4: VALUE OF b_n WHEN $\alpha \in (0, 2)$. Recall from the beginning of the proof that the parameter Γ associated with $a_n X + a_n b_n$ is Γ_{a_n, b_n} given by

$$\Gamma_{a_n, b_n} = a_n \left(\Gamma + b_n - \int_{1/a_n \leq |x| < 1} x \pi(dx) \right).$$

Since it equals $n\Gamma$, the parameter of S_n , then

$$b_n = \left(\frac{n}{a_n} - 1 \right) \Gamma + \int_{1/a_n \leq |x| < 1} x \pi(dx).$$

We can compute the integral with the previous identity for π , namely

$$\int_{1/a_n \leq |x| < 1} x \pi(dx) = \int_{\mathbb{S}} \lambda(dz) \int_0^{\infty} \frac{1}{r^{1+\alpha}} \mathbb{1}_{n^{-1/\alpha} \leq r < 1} rz dr = \int_{\mathbb{S}} z \lambda(dz) \int_{1/a_n}^1 \frac{1}{r^\alpha} dr.$$

The expression of the integral depends on α and precisely, since $a_n = n^{1/\alpha}$,

$$\int_{1/a_n}^1 \frac{1}{r^\alpha} dr = \begin{cases} \log n & \text{if } \alpha = 1, \\ \frac{1 - a_n^{-(1-\alpha)}}{1-\alpha} = \frac{a_n - n}{a_n(1-\alpha)} & \text{if } \alpha \in (0, 1) \cup (1, 2). \end{cases}$$

Let us denote by $\bar{\lambda} = \int_{\mathbb{S}} z \lambda(dz) \in \mathbb{R}^d$. Then

$$b_n = \left(\frac{n}{a_n} - 1 \right) \Gamma + \bar{\lambda} \int_{1/a_n}^1 \frac{1}{r^\alpha} dr = \begin{cases} \bar{\lambda} \log n & \text{if } \alpha = 1, \\ \frac{n - a_n}{a_n} \left(\Gamma - \frac{\bar{\lambda}}{1-\alpha} \right) & \text{if } \alpha \in (0, 1) \cup (1, 2). \end{cases}$$

Take $\alpha \in (0, 1) \cup (1, 2)$ and let $b = \Gamma - \frac{\bar{\lambda}}{1-\alpha}$; then for $X' = X - b$, the corresponding sum S'_n satisfies

$$S'_n = S_n - nb \stackrel{(d)}{=} a_n(X + b_n) - nb = a_n X' + a_n b_n - (n - a_n)b = a_n X'.$$

Hence $X - b$ is strictly stable. This concludes the proof. $\square$

REMARK 12.15 (Most important!). In dimension $d = 1$, the unit sphere $\mathbb{S}$ reduces to $\{-1, +1\}$ so λ takes the form $c_+ \delta_{+1} + c_- \delta_{-1}$ for some $c_+, c_- \geq 0$ with at least one of them nonzero. The Lévy measure of a non Gaussian stable law thus takes the form:

$$\pi(dx) = \frac{c_+}{x^{1+\alpha}} \mathbb{1}_{x>0} dx + \frac{c_-}{|x|^{1+\alpha}} \mathbb{1}_{x<0} dx. \quad (12.3)$$

Note that in the last item of the theorem, in this case $\bar{\lambda} = c_+ - c_-$.

REMARK 12.16. For any $\alpha \in (0, 2)$, the finite measure λ in the representation (12.2) of π , and in dimension 1, the constants c_+, c_- , can be any measure different from the constant null measure. Indeed, for any λ , the formula

$$\pi(A) = \int_{\mathbb{S}} \lambda(dz) \int_0^\infty \frac{1}{r^{1+\alpha}} \mathbb{1}_{rz \in A} dr,$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$ defines a measure with $\pi(\{0\}) = 0$ and

$$\int (|x|^2 \wedge 1) \pi(dx) = \int_{\mathbb{S}} \lambda(dz) \int_0^\infty \frac{1}{r^{1+\alpha}} (r^2 \wedge 1) dr < \infty.$$

Thus π is a Lévy measure and therefore there exists an infinitely divisible law with triple $(\Gamma, 0, \pi)$ for any $\Gamma \in \mathbb{R}^d$. Then it is easy to check using the characteristic functions that this law in turn is stable with $a_n = n^{1/\alpha}$ and b_n given as above.

12.5 MOMENTS AND VARIATION

Many properties and exponents related to stable laws depend on the index of stability. Here is one example.

REMARK 12.17. Finiteness of $\mathbb{E}[g(X)]$ for a submultiplicative function $g \geq 0$ when X is α -stable can now be read from Theorem 11.16 which stated (take $X = X_1$ there) that

$$\mathbb{E}[g(X)] < \infty \iff \int_{|x|>1} g(x) \pi(dx) < \infty.$$

Using the representation (12.2) we obtain for any $\alpha \in (0, 2)$:

$$\mathbb{E}[g(X)] < \infty \iff \int_{\mathbb{S}} \lambda(dz) \int_1^\infty \frac{1}{r^{1+\alpha}} g(rz) dr < \infty.$$

In particular, for $g(x) = |x|^c \vee 1$, we have $\mathbb{E}[|X|^c] < \infty \iff \mathbb{E}[|X|^c \vee 1] < \infty$ and therefore

$$\mathbb{E}[|X|^c] < \infty \iff \int_{\mathbb{S}} \lambda(dz) \int_1^\infty \frac{r^c}{r^{1+\alpha}} dr < \infty \iff \alpha > c.$$

Recall that $\alpha = 2$ is a Gaussian case, in which moments of all order exist.

As another example, the index α dictates whether the paths of the associated Lévy process have finite or infinite variation.

COROLLARY 12.18. A stable Lévy process with index $\alpha \in (0, 2)$ has finite variation paths when $\alpha < 1$ and infinite variation paths when $\alpha \geq 1$.

Proof. We rely here on the criterion from Proposition 11.9. For any $\alpha \in (0, 2)$, we have $\Sigma = 0$ so we are

in the finite variation case if and only if the following integral is finite:

$$\int (|x| \wedge 1) \pi(dx) = \int_{\mathbb{S}} \lambda(dz) \int_0^\infty \frac{1}{r^{1+\alpha}} (r \wedge 1) dr,$$

which is indeed the case if and only if $\alpha < 1$. $\square$

REMARK 12.19. Notice on the other hand that

$$\pi(\mathbb{R}^d) = \int_{\mathbb{S}} \lambda(dz) \int_0^\infty \frac{1}{r^{1+\alpha}} dr = \infty$$

for all $\alpha \in (0, 2)$ so stable-Lévy processes are not compound Poisson processes and have dense jump times.

12.6 AN EXPLICIT FORMULA FOR THE CHARACTERISTIC FUNCTION

To conclude this section, let us give without proof a more explicit expression of the characteristic function of a stable law μ . In the case $\alpha = 2$, since μ is a Gaussian law, then its characteristic function is simply given by

$$\exp\left(i\Gamma u - \frac{1}{2}(\Sigma u)^2\right).$$

In the case $\alpha \in (0, 2)$, we have seen that $\Sigma = 0$ and π takes the form (12.2). Although not straightforward, one can compute the integral against π in the representation (12.2)¹ and derive the more explicit formula:

$$\widehat{\mu}(u) = \begin{cases} \exp(i\eta u - \int_{\mathbb{S}} |uz|^\alpha (1 - i \tan \frac{\pi\alpha}{2} \operatorname{sgn}(uz)) \rho(dz)) & \text{when } \alpha \neq 1, \\ \exp(i\eta u - \int_{\mathbb{S}} |uz| (1 + i \frac{2}{\pi}(uz) \log |uz|) \rho(dz)) & \text{when } \alpha = 1, \end{cases} \quad (12.4)$$

where $\operatorname{sgn}(t) = \mathbb{1}_{t>0} - \mathbb{1}_{t<0}$, $\eta \in \mathbb{R}^d$ and ρ is a finite measure (different from the null measure) on the sphere $\mathbb{S}$. Conversely, for any η and ρ , these formulas define a stable law. In addition, the law is strictly stable if and only if $\eta = 0$ in the case $\alpha \neq 1$ and $\bar{\rho} = \int_{\mathbb{S}} z \rho(dz) = 0$ in the case $\alpha = 1$. The measure ρ here relates to λ in (12.2) via

$$\begin{cases} \rho(\cdot) = -\Gamma(-\alpha) \cos \frac{\pi\alpha}{2} \lambda(\cdot) & \text{when } \alpha \neq 1, \\ \rho(\cdot) = \frac{\pi}{2} \lambda(\cdot) & \text{when } \alpha = 1, \end{cases}$$

where $\Gamma(w) = \int_0^\infty t^{w-1} e^{-t} dt$ is Euler's gamma function. The vector η relates to Γ via:

$$\begin{cases} \eta = \Gamma + \frac{\bar{\lambda}}{\alpha-1} & \text{when } \alpha \neq 1, \\ \eta = \Gamma + \bar{\lambda} \int_0^\infty t^{-2} (\sin t - t \mathbb{1}_{0<t<1}) dt & \text{when } \alpha = 1. \end{cases}$$

In the case $\alpha \neq 1$, this η coincides with the shift b from Theorem 12.14 that makes the law strictly stable.

In dimension 1, the measure λ simply gives a mass c_+ to $+1$ and c_- to -1 . Then (12.4) can be written as:

$$\widehat{\mu}(t) = \begin{cases} \exp(i\eta t - c|t|^\alpha (1 - i\beta \tan \frac{\pi\alpha}{2} \operatorname{sgn}(t))) & \text{when } \alpha \neq 1, \\ \exp(i\eta t - c|t| (1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \log |t|)) & \text{when } \alpha = 1, \end{cases} \quad (12.5)$$

with $\eta \in \mathbb{R}$, $c > 0$, and $\beta \in [-1, 1]$ and these parameters relate to γ and the Lévy measure π and its two constants c_+ and c_- by:

$$\beta = \frac{c_+ - c_-}{c_+ + c_-}, \quad \text{and} \quad c = \begin{cases} -(c_+ + c_-) \Gamma(-\alpha) \cos \frac{\pi\alpha}{2} & \text{when } \alpha \neq 1, \\ (c_+ + c_-) \frac{\pi}{2} & \text{when } \alpha = 1, \end{cases}$$

¹See Theorem 14.10 in Sato in dimension d or Exercise 1.4 in Kyprianou, or Section XII.3 in Feller for the computation.

and finally,

$$\eta = \begin{cases} \gamma + \frac{c_+ - c_-}{\alpha - 1} & \text{when } \alpha \neq 1, \\ \gamma + (c_+ - c_-) \int_0^\infty t^{-2} (\sin t - t \mathbb{1}_{0 < t < 1}) dt & \text{when } \alpha = 1. \end{cases}$$

Then the distribution is strictly stable when $\eta = 0$ in the case $\alpha \neq 1$ and when $\beta = 0$ in the case $\alpha = 1$. In both cases, the distribution is symmetric when $\eta = \beta = 0$, in which case it reduces simply to

$$\hat{\mu}(u) = \exp(-c|u|^\alpha)$$

for all $\alpha \in (0, 2)$, as well as for $\alpha = 2$.

One can note that the characteristic function $\hat{\mu}$ in (12.5) is integrable, which implies that stable laws admit a density with respect to the Lebesgue measure, given by the inverse Fourier transform:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \hat{\mu}(t) dt.$$

There are only three closed form of densities: the Gaussian laws ($\alpha = 2$), the scaled and shifted Cauchy distributions

$$f(x) = \frac{1}{\pi} \frac{c}{c^2 + (x - x_0)^2},$$

with $c > 0$ and $x_0 \in \mathbb{R}$, for which $\alpha = 1$, and the functions defined on $(0, \infty)$ by

$$f(x) = \frac{a}{\sqrt{2\pi}x^{3/2}} e^{-a^2/(2x)},$$

for which $\alpha = 1/2$, and which is the law of the time it takes to a standard Brownian motion to reach the value a (see the course on stochastic calculus). The inverse Fourier transform nevertheless allows to obtain plenty of results on these densities without explicit formulae, see e.g. the books by Ibragimov & Linnik or Zolotarev.

12.7 EXERCISES

Recall that X has a stable law when for every $n \geq 2$, there exist a_n, c_n such that

$$S_n \stackrel{(d)}{=} a_n X + c_n, \quad (12.6)$$

where $S_n = X_1 + \dots + X_n$ and the X_i 's are i.i.d. with the same law as X . Recall that we exclude constant random variables.

EXERCISE 57. Prove that X has a stable law if and only if (12.6) holds for $n = 2$ and $n = 3$.

Hint: the set $\{2^k 3^\ell : k, \ell \in \mathbb{Z}\}$ is dense in $[0, \infty)$; you may also want to work with the characteristic function, and the equation corresponding to (12.6).

EXERCISE 58. Prove that X has a stable law if and only if for every $a, b \in [0, \infty)$, there exist $c \in [0, \infty)$ and $d \in \mathbb{R}$ such that

$$aX_1 + bX_2 \stackrel{(d)}{=} cX + d.$$

In addition we have precisely $c = (a^\alpha + b^\alpha)^{1/\alpha}$.

CHAPTER 13

DOMAIN OF ATTRACTION OF STABLE LAWS

Now that we have characterised all the stable laws, we provide in this chapter necessary and sufficient conditions for a normalised sum of n i.i.d. random variables to converge to a stable law, and which one. The exercises should convince the reader that these conditions can be used in practice. This chapter is mostly inspired by Feller and Sato.

CONTENTS

13.1	Slow variation	148
13.2	Limits of normalised random walks	150
13.3	Proof of the sufficiency	152
13.4	Proof of the necessity	155
13.5	Exercises	155

Recall that the (law of a) random variable X is said to belong to the *domain of attraction* of the (law of a) random variable Y when there exist sequences $(\alpha_n)_n$ in $(0, \infty)$ and $(\beta_n)_n$ in $\mathbb{R}^d$ such that

$$\frac{S_n}{\alpha_n} - \beta_n \xrightarrow[n \rightarrow \infty]{(d)} Y.$$

Recall from Theorem 12.10 that a law Y possesses a nonempty domain of attraction if and only if it is stable. The goal of this chapter is to provide explicit necessary and sufficient conditions for a random variable X to belong to the domain of attraction of a given stable law, and explicit in this case the scaling a_n and the shift b_n .

To simplify the exposition, we restrict to dimension 1 only.

13.1 SLOW VARIATION

Recall from Theorem 12.14 that in the definition 12.1 a stable law, the scaling factor is necessarily $a_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$. In the definition 12.9 of the domain of attraction of a stable law, the scaling α_n will not necessarily be exactly $n^{1/\alpha}$, but close to this, namely of the form $\alpha_n = n^{1/\alpha} L(n)$ where $(L(n))_n$ is a *slowly varying sequence* as defined below.

DEFINITION 13.1. A strictly positive function f , resp. a strictly positive sequence $(u_n)_{n \geq 0}$, is said to be *slowly varying*, when for every $c > 0$, it satisfies

$$\frac{f(cx)}{f(x)} \xrightarrow{x \rightarrow \infty} 1, \quad \text{resp.} \quad \frac{u_{[cn]}}{u_n} \xrightarrow{n \rightarrow \infty} 1.$$

A function of the form $x^\rho f(x)$ with f slowly varying is called *regularly varying* with index $\rho \in \mathbb{R}$. Similarly for sequences.

One can prove that any slowly varying function f is ‘small’, in the sense that $x^{-\varepsilon} \ll f(x) \ll x^\varepsilon$ at infinity for any $\varepsilon > 0$ (easy if f is monotone, not straightforward in the general case). These functions and sequences will thus appear as lower order corrections.

EXAMPLE 13.2. Importantly, functions that have a finite and nonzero limit are slowly varying. Other typical examples are log, powers of log and iterations of log. Products and sums of slowly varying functions are slowly varying, and so is any power of a slowly varying function.

In our search of necessary and sufficient conditions for a random variable X to belong to the domain of attraction of a random variable Y , the following functions will appear. Recall that Y is necessarily stable, with some index $\alpha \in (0, 2]$. Then for every $\alpha \in (0, 2]$ and every $x > 0$, let us put:

$$L(x) = x^{-(2-\alpha)} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}] \quad \text{and} \quad L_1(x) = x^\alpha \mathbb{P}(|X| > x). \quad (13.1)$$

THEOREM 13.3. Fix $\alpha \in (0, 2)$. If L and L_1 are as in (13.1), then one is slowly varying if and only if the other one is, which is also equivalent to

$$\frac{x^2 \mathbb{P}(|X| > x)}{\mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]} = \frac{L_1(x)}{L(x)} \xrightarrow{x \rightarrow \infty} \frac{2 - \alpha}{\alpha}. \quad (13.2)$$

For $\alpha = 2$, the convergence (13.2) holds if and only if L is slowly varying, and this occurs as soon as L_1 is slowly varying; however it does not imply conversely that L_1 is slowly varying, but only that $L_1 = o(L)$.

Working with $\mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]$ can be cumbersome, and in practice one would prefer working with the $\mathbb{P}(|X| > x)$ only. In our convergence criterion, the natural assumption will be (13.2), and the previous theorem shows that, when $\alpha < 2$, this is equivalent to checking that L_1 is slowly varying. When $\alpha = 2$, the latter condition is more restrictive, so it is good to try it first, and if it fails, to try to prove that L is slowly varying.

REMARK 13.4. About the last part of the statement, Feller provides the following example:

$$\mathbb{P}(X = 2^n) = \frac{c}{n4^n}, \quad n \geq 1.$$

Then

$$\mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}] = \sum_{n \geq [x]} \frac{c}{n} \sim c \log x$$

is slowly varying, so here $\alpha = 2$. However one can check that

$$\frac{\mathbb{P}(X > 2 \times 2^{n-1})}{\mathbb{P}(X > 2^{n-1})} \xrightarrow{n \rightarrow \infty} \frac{1}{4},$$

and thus $L_1 : x \mapsto x^2 \mathbb{P}(X > x)$ is not slowly varying. We will see that, suitably normalised, the sum of such random variables converges anyway to a Gaussian law.

Proof of Theorem 13.3. Let us use the notation $U(x) = \mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]$. It is an easy exercise to prove that

$$U(x) = 2 \int_0^x s \mathbb{P}(|X| > s) ds - x^2 \mathbb{P}(|X| > x) \quad \text{and} \quad \mathbb{P}(|X| > x) = -\frac{U(x)}{x^2} + 2 \int_x^\infty \frac{U(s)}{s^3} ds.$$

(Hint: use Fubini’s theorem to exchange integral and expectation.) Then the functions L and L_1 in (13.1)

are related by

$$\frac{L(x)}{L_1(x)} = 2 \underbrace{\frac{\int_0^x s^{1-\alpha} L_1(s) ds}{x^{2-\alpha} L_1(x)}}_{= 1/F_\alpha(x)} - 1 \quad \text{and} \quad \frac{L_1(x)}{L(x)} = 2 \underbrace{\frac{\int_x^\infty s^{-1-\alpha} L(s) ds}{x^{-\alpha} L(x)}}_{= 1/G_\alpha(x)} - 1. \quad (13.3)$$

The convergence $L_1/L \rightarrow (2 - \alpha)/\alpha$ in (13.2) is therefore equivalent to $F_\alpha \rightarrow 2 - \alpha$ and to $G_\alpha \rightarrow \alpha$.

Slowly varying functions do not interfere much with the integral of power functions, and indeed as shown e.g. in Seneta, Theorem 2.1 and Exercises 2.1 and 2.2, or Feller, Theorem VIII.9.1, a function ℓ is slowly varying if and only if there exists $p < -1$ such that

$$\frac{x^{p+1} \ell(x)}{\int_x^\infty s^p \ell(s) ds} \xrightarrow{x \rightarrow \infty} -(p+1),$$

and then it holds for every $p \leq -1$; however this convergence for $p = -1$ does not implies conversely that ℓ is slowly varying. Similarly, ℓ is slowly varying if and only if there exists $q > -1$ such that

$$\frac{x^{q+1} \ell(x)}{\int_0^x s^q \ell(s) ds} \xrightarrow{x \rightarrow \infty} q+1,$$

and in this case this holds actually for every $q \geq -1$, but again this for $q = -1$ does not implies conversely that ℓ is slowly varying.

For any $\alpha \in (0, 2]$, since $p = -1 - \alpha < -1$, then L is slowly varying if and only if G_α in (13.3) tends to α , which is equivalent to (13.2). Similarly, if L_1 is slowly varying, since $q = 1 - \alpha \geq -1$, then F_α in (13.3) tends to $2 - \alpha$, which is equivalent to (13.2). However, we can conversely deduce from this convergence that L_1 is slowly varying only when $\alpha < 2$ since we need $q = 1 - \alpha > -1$. $\square$

13.2 LIMITS OF NORMALISED RANDOM WALKS

We may now state our desired convergence criterion. Recall that the law of X is said to belong to the domain of attraction of the law of Y when there exists sequences α_n and β_n such that

$$\frac{S_n - \beta_n}{\alpha_n} \xrightarrow[n \rightarrow \infty]{(d)} Y.$$

We saw in Theorem 12.10 that, in this particular case, if this converges holds, then the limit Y necessarily has a stable law, with some index $\alpha \in (0, 2]$. Recall the functions L and L_1 from (13.1).

THEOREM 13.5. The law of X belongs to the domain of attraction of an α -stable law for some $\alpha \in (0, 2]$ if and only if the convergence (13.2) holds and, when $\alpha < 2$, there exists $p \in [0, 1]$ such that the balancing condition

$$\frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} \xrightarrow{x \rightarrow \infty} p \quad \text{and} \quad \frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} \xrightarrow{x \rightarrow \infty} 1 - p$$

is satisfied.

The proof will provide the suitable normalising sequences α_n and β_n .

REMARK 13.6. This recovers the central limit theorem since, if X has a finite and nonzero variance, then L has a finite and nonzero limit and is thus slowly varying, which is equivalent to (13.2) and $\alpha = 2$ in this case. However, the Gaussian case is possible also when X has an infinite variance, but in this case the scaling factor satisfies $\alpha_n/\sqrt{n} \rightarrow \infty$ (but not fast: this sequence is slowly varying).

REMARK 13.7. Recall that if Y has an α -stable law, then $\mathbb{E}[|Y|^c] < \infty$ if and only if $c < \alpha$. When X is in the domain of attraction of this stable law, then $\mathbb{E}[|X|^c] < \infty$ for every $c < \alpha$ and, provided that $\alpha < 2$, we have $\mathbb{E}[|X|^c] = \infty$ for every $c > \alpha$. Indeed, this follows from the fact that in this case, we have $\mathbb{P}(|X| > x) = x^{-\alpha} L_1(x)$ with L_1 slowly varying (or $L_1 = o(L)$ with L slowly varying in the case $\alpha = 2$) and that, as we mentioned, such a slowly varying function satisfies $x^{-\varepsilon} \ll L_1(x) \ll x^\varepsilon$ at infinity for any $\varepsilon > 0$. We note that for $c = \alpha$, the moment $\mathbb{E}[|X|^\alpha]$ can be finite or infinite.

We shall rely on the criterion from Theorem 9.16 applied to the random variables

$$X^n = \frac{X}{\alpha_n} - \frac{\beta_n}{n} \quad \text{so} \quad \sum_{i=1}^n X_i^n = \frac{S_n}{\alpha_n} - \beta_n,$$

where the X_i^n 's are i.i.d. copies of X^n . As we mentioned already, in this case, the limit, if it exists, must be stable, and the corresponding Lévy triple is thus given by Theorem 12.14. In particular, if $\alpha = 2$, then it is a Gaussian law, with variance $\sigma^2 > 0$ and $\pi = 0$, and on the contrary if $\alpha < 2$, then $\sigma^2 = 0$ and there exist $c_+, c_- \geq 0$ with at least one of them nonzero such that

$$\pi(dx) = \frac{c_+}{x^{1+\alpha}} \mathbb{1}_{x>0} dx + \frac{c_-}{|x|^{1+\alpha}} \mathbb{1}_{x<0} dx.$$

In particular π never has atoms. Then the necessary and sufficient conditions from Theorem 9.16 for this sum to converge in distribution to some limit μ , with associated Lévy triple (γ, σ^2, π) , simplify as follows:

$$\lim_{n \rightarrow \infty} n \mathbb{E}[X^n \mathbb{1}_{|X^n| < 1}] = \gamma, \quad (13.4)$$

and for every $x > 0$,

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X^n \geq x) = \pi([x, \infty)) \quad \text{and} \quad \lim_{n \rightarrow \infty} n \mathbb{P}(X^n \leq -x) = \pi((-\infty, -x]), \quad (13.5)$$

as well as

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} |n \mathbb{E}[(X^n)^2 \mathbb{1}_{|X^n| \leq \varepsilon}] - \sigma^2| = 0, \quad (13.6)$$

Again, when $\alpha = 2$, the limits in (13.5) are 0 and that in (13.6) is $\sigma^2 > 0$, whereas when $\alpha < 2$, we have $\sigma^2 = 0$ and we may calculate the integrals:

$$\pi([x, \infty)) = \frac{c_+}{\alpha} x^{-\alpha}, \quad \pi((-\infty, -x]) = \frac{c_-}{\alpha} x^{-\alpha}, \quad \text{and} \quad \int_{[-x, x]} t^2 \pi(dt) = \frac{c_+ + c_-}{2 - \alpha} x^{2-\alpha},$$

for every $x > 0$.

Recall that the characteristic function of a stable laws is more explicitly expressed in (12.5), using three parameters $\eta \in \mathbb{R}$, $c > 0$, and $\beta \in [-1, 1]$, which one can relate to α , γ , c_+ , and c_- . The proof of the sufficiency will provide a suitable scaling factor α_n and will show that, when $\alpha \in (0, 1)$, the convergence holds with $\beta_n = 0$, and the limit law has $\eta = 0$ in the expression (12.5). When $\alpha \in (1, 2]$, we know that $\mathbb{E}[X]$ is well-defined and finite, so the choice $\beta_n = n\alpha_n^{-1} \mathbb{E}[X]$ that makes $\mathbb{E}[\alpha_n^{-1} S_n - \beta_n] = 0$ seems natural, and we will see that indeed, this leads to the convergence with again a limit law with $\eta = 0$. The case $\alpha = 1$ is more complicated, but we will again obtain the limit law with $\eta = 0$ using $\beta_n = n \mathbb{E}[\sin(\alpha_n^{-1} X)]$. In all cases, we can observe that:

- the scaling takes the form $\alpha_n = n^{1/\alpha} L^*(n)$ for some slowly varying function L^* ,
- the centring takes the form $\beta_n = o(n)$.

13.3 PROOF OF THE SUFFICIENCY

Fix $\alpha \in (0, 2]$ and let us assume that

$$\frac{x^2 \mathbb{P}(|X| \geq x)}{\mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]} \xrightarrow{x \rightarrow \infty} \frac{2 - \alpha}{\alpha}.$$

Recall that this is equivalent to assuming that L is slowly varying, where

$$L(x) = x^{-(2-\alpha)} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}] \quad \text{and} \quad L_1(x) = x^\alpha \mathbb{P}(|X| > x).$$

In addition, when $\alpha < 2$, this is also equivalent to assuming that L_1 is slowly varying. In this regime $\alpha < 2$, we also assume that

$$\frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} \xrightarrow{x \rightarrow \infty} p \quad \text{and} \quad \frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} \xrightarrow{x \rightarrow \infty} 1 - p$$

for some $p \in [0, 1]$.

STEP 1: DEFINITION OF α_n

Let us fix a constant $C > 0$; since L is slowly varying, then we have the sequence $(\alpha_n)_n$ defined by

$$n\alpha_n^{-2} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq \alpha_n}] \xrightarrow{n \rightarrow \infty} C \quad \text{where} \quad \alpha_n = \inf\{x > 0 : nx^{-2} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}] \leq C\}.$$

Let us fix this sequence (but any sequence satisfying this asymptotic behaviour would do). Since L is slowly varying, then we have further

$$n\alpha_n^{-2} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq \alpha_n x}] \xrightarrow{n \rightarrow \infty} Cx^{2-\alpha},$$

for every $x > 0$. We then infer from our assumption that

$$\lim_{n \rightarrow \infty} n \mathbb{P}(|X| \geq \alpha_n x) = \frac{2 - \alpha}{\alpha} Cx^{-\alpha},$$

for every $x > 0$. Together with the balancing condition, when $\alpha < 2$, we have

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X \geq \alpha_n x) = \frac{2 - \alpha}{\alpha} Cpx^{-\alpha} \quad \text{and} \quad \lim_{n \rightarrow \infty} n \mathbb{P}(X \leq -\alpha_n x) = \frac{2 - \alpha}{\alpha} Cqx^{-\alpha},$$

for every $x > 0$.

REMARK 13.8. Let $L^*(n) = C^{-1/\alpha} L(\alpha_n)^{1/\alpha}$, then by definition of α_n we have

$$\alpha_n \sim n^{1/\alpha} L^*(n).$$

The sequence $(L^*(n))_n$ is slowly varying and is related to the notion of *conjugate* of the slowly varying function of $1/L$, see e.g. the book of Seneta for more information.

STEP 2: CONVERGENCE OF TAILS

Let $\delta_n = \beta_n/n$ (which may be positive or negative) and recall that $X^n = \alpha_n^{-1} X - \delta_n$. We claim that as long as $\delta_n \rightarrow 0$, the previous limits which apply to $\alpha_n^{-1} X$ apply equivalently to X^n . Indeed, consider for example the right tails. Fix an arbitrary $\varepsilon > 0$ so for n large enough, we have $|\delta_n| < \varepsilon$, and hence

$$\mathbb{P}(X \geq \alpha_n(x - \varepsilon)) \leq \mathbb{P}(X \geq \alpha_n x + \alpha_n \delta_n) \leq \mathbb{P}(X \geq \alpha_n(x + \varepsilon)).$$

The extremities converge to $\pi([x - \varepsilon, \infty))$ and $\pi([x + \varepsilon, \infty))$ respectively which, if one takes ε arbitrarily small, are arbitrarily close to $\pi([x, \infty))$. Therefore, the convergence without δ_n implies that with δ_n and a similar reasoning shows the converse implication. The same holds for the left tails and for the integral of the square, namely that $n \mathbb{E}[(X^n)^2 \mathbb{1}_{|X^n| \leq x}]$ has a finite limit for each $x > 0$ if and only if $n \mathbb{E}[(\alpha_n^{-1} X)^2 \mathbb{1}_{|\alpha_n^{-1} X| \leq x}]$ has the same limit for every $x > 0$. We conclude that, as long as $\delta_n = \beta_n/n \rightarrow 0$, we have for every $x > 0$:

$$\lim_{n \rightarrow \infty} n \mathbb{E}[(X^n)^2 \mathbb{1}_{|X^n| \leq x}] = Cx^{2-\alpha}, \quad \lim_{n \rightarrow \infty} n \mathbb{P}(X^n \geq x) = \frac{C_\alpha p}{x^\alpha}, \quad \lim_{n \rightarrow \infty} n \mathbb{P}(X^n \leq -x) = \frac{C_\alpha(1-p)}{x^\alpha},$$

where $C_\alpha = C(2 - \alpha)/\alpha$. This exactly corresponds to (13.5) and (13.6) where $\sigma^2 = C$ when $\alpha = 2$ and, when $\alpha < 2$, the two constants c_+ and c_- solve the system:

$$C = \frac{c_+ + c_-}{2 - \alpha} \quad \text{and} \quad p = \frac{c_+}{c_+ + c_-}.$$

STEP 3: THE CENTRING TERM β_n

In order to conclude that $\alpha_n^{-1} S_n - \beta_n/n$ converges to the corresponding stable law, with an extra parameter η in the formulation (12.5), it remains to prove (13.4).

Consider first the case $\alpha \in (0, 1)$. We claim that with $\beta_n = 0$, the convergence (13.4) holds with the limit $\gamma = (c_+ - c_-)/(1 - \alpha)$, and hence $\eta = 0$ in (12.5). To see this, apply Fubini's theorem to get

$$\begin{aligned} n \mathbb{E}[\alpha_n^{-1} X \mathbb{1}_{|\alpha_n^{-1} X| < 1}] &= \int_0^1 n \mathbb{E}[\mathbb{1}_{X > 0} \mathbb{1}_{t < \alpha_n^{-1} X < 1} - \mathbb{1}_{X < 0} \mathbb{1}_{t < -\alpha_n^{-1} X < 1}] dt \\ &= \int_0^1 n (\mathbb{P}(X > t\alpha_n) - \mathbb{P}(X \geq \alpha_n) - \mathbb{P}(X < -t\alpha_n) - \mathbb{P}(X \leq -\alpha_n)) dt \\ &\xrightarrow{n \rightarrow \infty} \int_0^1 \frac{2-\alpha}{\alpha} C(p-q)(t^{-\alpha} - 1) dt \\ &= \frac{2-\alpha}{1-\alpha} C(p-q) = \frac{c_+ - c_-}{1-\alpha}. \end{aligned}$$

Consider next the case $\alpha \in (1, 2)$. We claim that the same result holds provided that we subtract the expectation, so again $\eta = 0$. Indeed, first recall that $L_1(x) = x^\alpha \mathbb{P}(|X| > x)$ is slowly varying, and thus much smaller than x^ε for any $\varepsilon > 0$. For $\varepsilon < \alpha - 1$, this implies that $\mathbb{E}[|X|] = \int_0^\infty \mathbb{P}(|X| > x) dx < \infty$. Therefore X has a well-defined and finite expectation, and taking $\beta_n = n \mathbb{E}[X]/\alpha_n$ amounts to replacing X by $X - \mathbb{E}[X]$, so instead and in order to simplify the notation, let us assume that $\mathbb{E}[X] = 0$ and $\beta_n = 0$. This implies that

$$\begin{aligned} n \mathbb{E}[\alpha_n^{-1} X \mathbb{1}_{|\alpha_n^{-1} X| < 1}] &= -n \mathbb{E}[\alpha_n^{-1} X \mathbb{1}_{|\alpha_n^{-1} X| \geq 1}] \\ &= -n \mathbb{E}\left[\mathbb{1}_{X \geq \alpha_n} \int_0^\infty \mathbb{1}_{t < \alpha_n^{-1} X} dt - \mathbb{1}_{X \leq -\alpha_n} \int_0^\infty \mathbb{1}_{t < -\alpha_n^{-1} X} dt\right] \\ &= -n \mathbb{E}\left[\mathbb{1}_{X \geq \alpha_n} + \int_1^\infty \mathbb{1}_{t < \alpha_n^{-1} X} dt - \mathbb{1}_{X \leq -\alpha_n} - \int_1^\infty \mathbb{1}_{t < -\alpha_n^{-1} X} dt\right] \\ &= -n \left(\mathbb{P}(X \geq \alpha_n) + \int_1^\infty \mathbb{P}(X > \alpha_n t) dt - \mathbb{P}(X \leq -\alpha_n) - \int_1^\infty \mathbb{P}(X < -\alpha_n t) dt \right) \\ &\xrightarrow{n \rightarrow \infty} -\frac{2-\alpha}{\alpha} C(p-q) \left(1 + \int_1^\infty t^{-\alpha} dt \right) \\ &= -\frac{2-\alpha}{\alpha} C(p-q) \frac{\alpha}{\alpha-1} = -\frac{c_+ - c_-}{\alpha-1}. \end{aligned}$$

This also holds for $\alpha = 2$, in which case the limit is simply 0.

It remains to consider the case $\alpha = 1$. Here, and just as for the definition of the stable law, we cannot get rid of the centring term. In order to conclude about the convergence with $\eta = 0$ in the representation (12.5) in this case as well, we must have

$$n \mathbb{E}[X^n \mathbb{1}_{|X^n| < 1}] \xrightarrow{n \rightarrow \infty} -(c_+ - c_-) \int_0^\infty t^{-2} (\sin t - t \mathbb{1}_{0 < t < 1}) dt.$$

The centring,

$$\beta_n = n \mathbb{E}[\sin(\alpha_n^{-1} X)]$$

achieves this. Note that, since $\alpha_n \rightarrow \infty$, then $\beta_n/n \rightarrow 0$ as we wanted. To see that this works, recall that our convergence criterion comes from the consequence of Lemma 9.2 that a sum of n i.i.d. random variables with a law μ_n converges in distribution if and only if the compound Poisson distribution Z_n obtained by summing instead an independent random Poisson number with parameter n of them converges in distribution, and the limits are the same. Let μ_n denote the law of $\alpha_n^{-1} X$ and recall that this compound Poisson random variable Z_n has a Lévy triple $(\gamma_n, 0, n\mu_n)$, where $\gamma_n = \int_{(-1,1)} x n\mu_n(dx) = n \mathbb{E}[\alpha_n^{-1} X \mathbb{1}_{|\alpha_n^{-1} X| < 1}]$. Then for any sequence β_n , the random variable $Z_n - \beta_n$ has Lévy triple $(\gamma_n - \beta_n, 0, n\mu_n)$. Our previous calculations show the convergence of the measure ρ_n with density $|x|^2/(1+|x|^2)$ with respect to $n\mu_n$ (recall for this the proof of Theorem 9.16). Then, if β_n is such that $\gamma_n - \beta_n$ converges to some finite limit γ' , then $Z_n - \beta_n$ converges in distribution to the infinitely divisible with triple $(\gamma', 0, \pi)$. By the convergence of ρ_n , we infer, using approximations by continuous functions, that

$$\begin{aligned} \lim_{n \rightarrow \infty} n(\mathbb{E}[\alpha_n^{-1} X \mathbb{1}_{|\alpha_n^{-1} X| < 1}] - \mathbb{E}[\sin(\alpha_n^{-1} X)]) &= - \int_{\mathbb{R}} (\sin r - r \mathbb{1}_{|r| < 1}) \pi(dr) \\ &= -(c_+ - c_-) \int_0^\infty r^{-2} (\sin r - r \mathbb{1}_{r < 1}) dr. \end{aligned}$$

With the choice $\beta_n = n \mathbb{E}[\sin(\alpha_n^{-1} X)]$, the left side equals $\gamma_n - \beta_n$, so, as we claimed, the sequence $Z_n - \beta_n$ converges in distribution to the law with characteristic function (12.5) with $\eta = 0$.

In order to conclude that $\alpha_n^{-1} S_n - \beta_n$ converges to the same limit, we compare its characteristic function to that of $Z_n - \beta_n$, or simply those of $\alpha_n^{-1} S_n$ and Z_n . Let φ_X denote the characteristic function of X , then that of $\alpha_n^{-1} S_n$ is given for any $u \in \mathbb{R}$ by:

$$\varphi_X(\alpha_n^{-1} u)^n = \exp(n(\varphi_X(\alpha_n^{-1} u) - 1) + O(n|\varphi_X(\alpha_n^{-1} u) - 1|^2)).$$

We saw previously when $\alpha > 1$ that X had finite moment of any order less than α . The argument remains true for $\alpha = 1$ (and less), so $\mathbb{E}[|X|^c] < \infty$ for any $c \in (0, 1)$. Using that $|e^{it} - 1| \leq 2|t|^c$, we infer that

$$|\varphi_X(\alpha_n^{-1} u) - 1|^2 \leq 4 \mathbb{E}[|X|^c]^2 |u|^{2c} \alpha_n^{-2c}.$$

Recall that we have defined α_n to satisfy:

$$n\alpha_n^{-2} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq \alpha_n}] \xrightarrow{n \rightarrow \infty} C,$$

and $L(x) = x^{-1} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]$ is slowly varying so $L(x) \gg x^{1-c}$ for any $c \in (0, 1)$. If $c < 1$ is close enough to 1, then

$$n\alpha_n^{-2c} = n\alpha_n^{-2} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq \alpha_n}] L(\alpha_n)^{-1} \alpha_n^{1-2c} \xrightarrow{n \rightarrow \infty} 0,$$

and consequently

$$\varphi_X(\alpha_n^{-1} u)^n \sim \exp(n(\varphi_X(\alpha_n^{-1} u) - 1)).$$

We conclude that

$$\mathbb{E}[e^{iu\alpha_n^{-1} S_n - \beta_n}] \sim \mathbb{E}[e^{iuZ_n - \beta_n}] \xrightarrow{n \rightarrow \infty} \hat{\mu}(u),$$

as defined in (12.5) with $\eta = 0$.

13.4 PROOF OF THE NECESSITY

Let us keep the same notation and suppose conversely that there exist sequences α_n and β_n such that $\alpha_n^{-1}S_n - \beta_n$ converges in distribution, necessarily to some α -stable law, and let us prove that

Recall from Remark 12.13 that $\alpha_n \rightarrow \infty$; further, we saw in the proof of Lemma 9.2 that the characteristic function of X^n must tend to 1, namely for every $u \in \mathbb{R}$,

$$\mathbb{E}[e^{iuX^n}] = \mathbb{E}[e^{i(u/\alpha_n)X}] e^{-iu\beta_n/n} \xrightarrow{n \rightarrow \infty} 1.$$

Since $\alpha_n \rightarrow \infty$, then the first expectation in the middle tends to 1, so we must have $\beta_n/n \rightarrow 0$.

The convergence of our sum implies the convergence of the tail probability of X^n as in (13.5), and since $\beta_n/n \rightarrow 0$, we saw in the previous section that this is equivalent to the convergence of the tail probability of $\alpha_n^{-1}X$, without the centring term, to the same limit. Suppose first that the limit law has index $\alpha < 2$. Then we have

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X \geq \alpha_n x) = \frac{c_+}{\alpha} x^{-\alpha} \quad \text{and} \quad \lim_{n \rightarrow \infty} n \mathbb{P}(X \leq -\alpha_n x) = \frac{c_-}{\alpha} x^{-\alpha}$$

for every $x > 0$, and therefore

$$\lim_{n \rightarrow \infty} n \mathbb{P}(|X| \geq \alpha_n x) = \frac{c_+ + c_-}{\alpha} x^{-\alpha}.$$

Using that the tail probability is monotone, it is not difficult to deduce that it is regularly varying, with exponent $-\alpha$, i.e. that L_1 in (13.1) is slowly varying. According to Theorem 13.3, this is equivalent to the convergence

$$\lim_{x \rightarrow \infty} \frac{x^2 \mathbb{P}(|X| \geq x)}{\mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]} = \frac{2 - \alpha}{\alpha}.$$

In addition, we have the balancing equations:

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(X \geq x)}{\mathbb{P}(|X| \geq x)} = \frac{c_+}{c_+ + c_-} \quad \text{and} \quad \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X \leq -x)}{\mathbb{P}(|X| \geq x)} = \frac{c_-}{c_+ + c_-}.$$

Indeed, these limits hold along any sequence $(\alpha_n x)_n$ for any $x > 0$, so they hold along any sequence $x_n \rightarrow \infty$. In the case $\alpha = 2$, where π is the null measure and $\sigma^2 > 0$, the convergences (13.5) and (13.6) without the term $\beta_n/n \rightarrow 0$ simply mean that

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X \geq \alpha_n x) = \lim_{n \rightarrow \infty} n \mathbb{P}(X \leq -\alpha_n x) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n \alpha_n^{-2} \mathbb{E}[X^2 \mathbb{1}_{|X| \leq \alpha_n x}] = \sigma^2 > 0$$

for every $x > 0$. We conclude that the convergence

$$\lim_{x \rightarrow \infty} \frac{x^2 \mathbb{P}(|X| \geq x)}{\mathbb{E}[X^2 \mathbb{1}_{|X| \leq x}]} = 0,$$

holds, since it follows from the previous display that it holds along every sequence of the form $(\alpha_n x)_n$ for all $x > 0$.

13.5 EXERCISES

EXERCISE 59. Let X be a random variable with a density which is symmetric, continuous, and nonzero at 0. Prove that $1/X$ belongs to the domain of a Cauchy distribution, with an explicit renormalisation.

EXERCISE 60. Let $X_{n,1}, \dots, X_{n,n}$ be i.i.d. uniformly sampled on $[-n, n]$, let $p > 0$, and let

$$F_n = \sum_{i=1}^n \frac{\text{sgn}(X_{n,i})}{|X_{n,i}|^p}.$$

Prove that:

- (i) if $p > 1/2$, then F_n converges in distribution to a stable law with index $\alpha = 1/p$,
- (ii) if $p < 1/2$, then $F_n/n^{1/2-p}$ converges in distribution to a Gaussian law,
- (iii) and if $p = 1/2$, then $F_n/\sqrt{\log n}$ converges in distribution to a Gaussian law.

EXERCISE 61. Let $(S_n)_n$ denote the simple random walk on $\mathbb{Z}$, given by $S_0 = 0$ and $S_n = \xi_1 + \dots + \xi_n$ where the ξ_k 's are i.i.d. with $\mathbb{P}(\xi_1 = +1) = \mathbb{P}(\xi_1 = -1) = 1/2$. Prove that $\tau = \inf\{n \geq 0 : S_{2n+1} = 1\}$ and $\sigma = \inf\{n \geq 1 : S_{2n} = 0\}$ belong to the domain of attraction of the same stable law with index $\alpha = 1/2$, with the same scaling factor and with shifting term 0. The limit is one of the few with an explicit density:

$$\frac{1}{\sqrt{2\pi x^3}} \exp(-1/(2x)) \quad \text{for all } x > 0.$$

Hint: You may use the identity $\mathbb{P}(\bigcap_{k=1}^n \{S_{2k} \neq 0\}) = \mathbb{P}(S_{2n} = 0)$.

PART IV

APPENDIX

APPENDIX A

POISSON RANDOM MEASURES

This chapter is more personal, but we can refer to Kallenberg, Kingman, Last–Penrose, or Kyprianou. This is pure extra content to discuss further the theory of Poisson random measures that was mentioned in relation with Lévy processes, but has applications in many other subjects.

CONTENTS

A.1	Random measures	159
A.2	Poisson random measures	161
A.3	Stability properties	163
A.4	Integration against a Poisson measure	165
A.5	Compensated integrals	167
A.6	Space-time measures	169
A.7	Exercises	173

If we have a finite measure μ on a space $(E, \mathcal{E})$, then we can sample i.i.d. points according to the probability measure $\mu(\cdot)/\mu(E)$. Now what to do with an infinite measure? How to sample points uniformly at random in $\mathbb{R}^d$? Although we cannot make sense of one uniform random point in $\mathbb{R}^d$, we can make sense of a random cloud of points $\{X_i : i \in I\}$ uniformly spread out in $\mathbb{R}^d$.

Here is an idea: for any fixed $n \geq 1$, sample n^d i.i.d. points uniformly at random in the cube $C_n = [-n/2, n/2]^d$. Then for every bounded set $A \in \mathcal{B}(\mathbb{R}^d)$, for n large enough we have $A \subset C_n$, and then the probability of such a random point falling in A equals $p_n(A) = n^{-d}|A|$, where $|A|$ is the Lebesgue measure of A . Hence the total number of such points has the binomial law $\text{Bin}(n^d, p_n(A)) \Rightarrow \text{Poi}(|A|)$ and conditionally on these numbers, the points are indeed uniformly distributed on A (this is the *rejection sampling* algorithm). In addition, these numbers for disjoint sets are independent since the correlation of the associated binomial laws tends to 0.

Can we make sense, formally, of a collection of random points in $\mathbb{R}^d$, with I countable, that has the property that the number of points in each bounded set $A \in \mathcal{B}(\mathbb{R}^d)$ has the Poisson law $\text{Poi}(|A|)$ and these numbers are independent for disjoint sets? Instead of collections of random points $(X_i)_{i \in I}$, it will actually be easier to manipulate random measures, given by $M = \sum_{i \in I} \delta_{X_i}$, so $M(A) = \#\{i \in I : X_i \in A\}$ for every $A \in \mathcal{B}(\mathbb{R}^d)$. As the previous idea shows, Poisson random measures, in which the $M(A)$'s are Poisson distributed and independent for disjoint A 's, will play a predominant role.

As a last preliminary remark, although the case of $\mathbb{R}^d$ is already very interesting for applications e.g. in random geometry, communication networks, etc. more general spaces are also very natural to consider. In addition, the specific nature of $\mathbb{R}^d$ adds nothing to the theory so we will work on a fairly general measurable space $(E, \mathcal{E})$. To give only one example, to be studied in the second semester, if we take a real-valued Brownian motion $(B_t)_{t \in [0, \infty)}$, the set of times where it sits at 0 is pretty complicated (with no isolated point, yet with Lebesgue measure 0). Yet it is almost surely a closed set of $[0, \infty)$, so its complement is a countable union of open intervals: we can write $\{t \in [0, \infty) : B_t \neq 0\} = \bigcup_{i \in I} (a_i, b_i)$. The part of the trajectory $B^i = (B_t)_{t \in [a_i, b_i]}$ is called the *excursion* of B on the interval (a_i, b_i) . A famous result due to Itô

shows that the collection of these paths $(B^i)_{i \in I}$ forms such a Poisson random cloud of points, or alternatively $M = \sum_{i \in I} \delta_{B^i}$ forms a Poisson random measure, on the space of real-valued continuous functions defined on compact intervals.

A.1 RANDOM MEASURES

Let us fix the general setting. Let $(E, \mathcal{E})$ be any measurable space. We shall consider in this chapter random variables with values, not in E , but in the set $\mathcal{M}$ of (nonnegative) measures on E equipped with the σ -algebra generated, as usual, by the elementary projections/evaluations:

$$\mathcal{M} = \sigma(\mu \in \mathcal{M} \mapsto \mu(A) \in [0, \infty], A \in \mathcal{E}).$$

As usual $\mathcal{M}$ is equivalently generated by the evaluations on finitely many sets, namely

$$\mathcal{M} = \sigma(\mu \in \mathcal{M} : \mu(A_1) \in B_1, \dots, \mu(A_k) \in B_k \text{ for every } k \geq 1, A_1, \dots, A_k \in \mathcal{E}, B_1, \dots, B_k \in \mathcal{B}([0, \infty))).$$

EXERCISE A.1. Let $\mathcal{C}$ denote a collection of sets in $\mathcal{E}$ and let $\mathcal{P}$ denote the collection of subsets of $\mathcal{M}$ of the form

$$\{\mu \in \mathcal{M} : \mu(A_i) \in B_i \text{ for every } 1 \leq i \leq k\},$$

where $B_1, \dots, B_k$ are Borel sets of $[0, \infty]$ and $A_1, \dots, A_k$ are disjoint elements of $\mathcal{C}$. Prove that if $\mathcal{C}$ is a π -system that generates $\mathcal{E}$, then $\mathcal{P}$ is a π -system that generates $\mathcal{M}$. As a consequence, two random measures have the same law as soon as their joint distributions on every finite collection of disjoint elements of $\mathcal{C}$ have the same law.

We shall denote by $M : (\Omega, \mathcal{A}) \rightarrow (\mathcal{M}, \mathcal{M})$ a random measure. For every $A \in \mathcal{E}$, we denote by $M(A)$ the mass of M at A , which is a well-defined random variable in $[0, \infty]$ by construction. Then we can consider its expectation.

DEFINITION A.2. For any random measure M , its *intensity measure* is defined as the mapping $m = \mathbb{E}[M] : \mathcal{E} \rightarrow [0, \infty]$ given by $m(A) = \mathbb{E}[M(A)]$ for every $A \in \mathcal{E}$.

LEMMA A.3. The intensity measure m is indeed a measure on $\mathcal{E}$.

Proof. Clearly $m(\emptyset) = \mathbb{E}[M(\emptyset)] = \mathbb{E}[0] = 0$. Next, if $(A_n)_{n \geq 1}$ are disjoint elements of $\mathcal{E}$, then $M(\bigcup_n A_n) = \sum_n M(A_n) = \uparrow \lim_N \sum_{n \leq N} M(A_n)$ so

$$m\left(\bigcup_n A_n\right) = \mathbb{E}\left[\uparrow \lim_N \sum_{n \leq N} M(A_n)\right] = \uparrow \lim_N \sum_{n \leq N} \mathbb{E}[M(A_n)] = \uparrow \lim_N \sum_{n \leq N} m(A_n) = \sum_n m(A_n),$$

by linearity and monotone convergence. $\square$

Let us denote by $M(f) = \int f dM$ the integral of a measurable function $f : E \rightarrow \mathbb{R}$, whenever it is well-defined, namely when f is nonnegative or when either f^- or f^+ has finite integral. That $M(f)$ is a well-defined random variable follows by the usual approximation argument from *simple functions*, that is, functions that are finite linear combinations of indicator functions of measurable sets: $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$.

PROPOSITION A.4 (Campbell's formula). Let M be a random measure with intensity $m = \mathbb{E}[M]$.

(i) For every measurable and nonnegative function f , the mapping $M(f)$ is measurable and we have

$$\mathbb{E}[M(f)] \stackrel{\text{def}}{=} \mathbb{E}\left[\int f(x)M(dx)\right] = \int f(x)m(dx) \stackrel{\text{def}}{=} m(f).$$

(ii) This extends to measurable functions f of any sign that satisfy $m(|f|) = \mathbb{E}[M(|f|)] < \infty$.

Proof. If $f = \mathbb{1}_A$ for some $A \in \mathcal{E}$, then $M(f) = M(A)$ is measurable by construction. By composition $M(f)$ is measurable for any nonnegative simple function f and the identity $m(f) = \mathbb{E}[M(f)]$ extends by linearity. If f is a measurable and nonnegative function, then it can be written as the nondecreasing limit of nonnegative simple functions f_n , and the monotone convergence theorem shows that $M(f) = \uparrow \lim_n M(f_n)$ is measurable, and monotone convergence also implies that $m(f) = \uparrow \lim_n m(f_n) = \uparrow \lim_n \mathbb{E}[M(f_n)] = \mathbb{E}[M(f)]$. Finally, if f is any measurable function such that $m(|f|) = \mathbb{E}[M(|f|)] < \infty$, then $M(f) = M(f^+) - M(f^-)$ is well-defined, measurable, and has $\mathbb{E}[M(f)] = m(f)$ again. $\square$

Recall that the law of a real-valued random variable is uniquely characterised by its characteristic function, or Fourier transform; for nonnegative random variables, one often prefers working with the Laplace transform. The analogue for random measures is the following.

DEFINITION A.5. Let M be a random measure. For any measurable function f such that the integral $M(f)$ is well-defined almost surely, we let

$$L_M(f) = \mathbb{E}[e^{-M(f)}] \quad \text{and} \quad \Phi_M(f) = \mathbb{E}[e^{iM(f)}].$$

They are called respectively the *Laplace functional* and *Fourier functional* of M at f .

LEMMA A.6. The law of a random measure M is uniquely characterised by each of the following quantities;

- (i) the distribution of $M(f)$ for every nonnegative simple function f ,
- (ii) the value of $L_M(f)$ for every nonnegative simple function f ;
- (iii) the value of $\Phi_M(f)$ for every simple function f .

Proof. That (i) and (ii) or (i) and (iii) are equivalent is the standard property that the law of a real-valued random variable is characterised by its characteristic function or its Laplace transform.

Suppose that $M(f)$ and $M'(f)$ have the same law for all nonnegative simple functions f . Then $L_M(f) = L_{M'}(f)$, namely for any $n \geq 1$ and any $A_1, \dots, A_n \in \mathcal{E}$, we have

$$\mathbb{E}[e^{-\sum_{k=1}^n a_k M(A_k)}] = \mathbb{E}[e^{-\sum_{k=1}^n a_k M'(A_k)}],$$

for all $a_1, \dots, a_n \geq 0$. This implies that $(M(A_1), \dots, M(A_n))$ and $(M'(A_1), \dots, M'(A_n))$ have the same law. Since this holds for every $n \geq 1$ and every $A_1, \dots, A_n \in \mathcal{E}$, then M and M' have the same law. The proof for the Fourier functional is similar. $\square$

Let us make sure that the usual operations on measures are measurable.

LEMMA A.7. Let M be a random measure with intensity m on E and let $A \in \mathcal{E}$. Then the restriction $M|_A(\cdot) = M(A \cap \cdot)$ is a random measure on E with intensity $m|_A(\cdot) = m(A \cap \cdot)$.

Proof. The intensity of $M|_A$ is clearly $m|_A$, but we need to make sure first that the mapping $\mu \mapsto \mu|_A$ is measurable. Fix some $B \in \mathcal{E}$ and some $c \geq 0$, since $A \cap B \in \mathcal{E}$ and $\mu|_A(B) = \mu(A \cap B)$, then $\{\mu \in \mathcal{M} : \mu|_A(B) \leq c\} \in \mathcal{M}$. This implies the desired measurability property since this collection of sets generates $\mathcal{M}$. $\square$

LEMMA A.8. Let M be a random measure with intensity m on $(E, \mathcal{E})$. For any measurable mapping $f : (E, \mathcal{E}) \rightarrow (E', \mathcal{B}(E'))$, the image measure $M \circ f^{-1}$ is a random measure on $(E', \mathcal{B}(E'))$ with intensity $m \circ f^{-1}$.

Proof. Let $\mathcal{M}(E)$ and $\mathcal{M}(E')$ denote the set of measures on E and on E' respectively. As in the previous proof, fix any $B \in \mathcal{B}(E')$ and any $c \geq 0$ and let $A = \{\mu \in \mathcal{M}(E') : \mu(B) \leq c\}$. Since $f^{-1}(B) \in \mathcal{E}$, then $\{\mu \in \mathcal{M}(E) : \mu \circ f^{-1} \in A\} = \{\mu \in \mathcal{M}(E) : \mu(f^{-1}(B)) \leq c\} \in \mathcal{M}(E)$. Since the collection of such sets A generates $\mathcal{M}(E')$, then $\{\mu \in \mathcal{M}(E) : \mu \circ f^{-1} \in A\} \in \mathcal{M}(E)$ for every $A \in \mathcal{M}(E')$, so the mapping $\mu \mapsto \mu(f^{-1}(\cdot))$ is measurable and so if M is a random measure on E , then $M \circ f^{-1}$ is a random measure on E' . Its intensity is clearly the image of that of M by f . $\square$

LEMMA A.9. Let I be a finite or countable set and let $(M_i)_{i \in I}$ be random measures on E with intensity m_i respectively. Then $M = \sum_{i \in I} M_i$ is a random measure with intensity $m = \sum_{i \in I} m_i$.

Proof. Recall that $S : [0, \infty]^I \rightarrow [0, \infty]$ defined by $S((a_i)_{i \in I}) = \sum_{i \in I} a_i$ is continuous, and hence measurable from $\mathcal{B}([0, \infty])^{\otimes I} = \mathcal{B}([0, \infty]^I)$ to $\mathcal{B}([0, \infty])$. By composition, for every $A \in \mathcal{E}$, since each $M_i(A) : \Omega \rightarrow [0, \infty]$ is measurable, then $M(A) = S((M_i(A))_{i \in I}) : \Omega \rightarrow [0, \infty]$ is measurable and it has $\mathbb{E}[M(A)] = m(A)$. $\square$

A.2 POISSON RANDOM MEASURES

The Poisson distributions play a special role because they appear as laws of rare and independent events, in the sense e.g. that the Poisson law with parameter (or *rate*) $\lambda \in (0, \infty)$ is the limit of the binomial law with parameters n and p_n when $np_n \rightarrow \lambda$. Let N_λ have this law, then recall the mass function:

$$\mathbb{P}(N_\lambda = n) = \frac{\lambda^n e^{-\lambda}}{n!} \quad \text{for all } n \geq 0.$$

Then one consider the Laplace transform on the entire complex plane and the formula

$$\mathbb{E}[e^{-zN_\lambda}] = e^{-\lambda(1-e^{-z})},$$

is valid for every $z \in \mathbb{C}$. Taking $z = u \in \mathbb{R}$ recovers the real Laplace transform, whilst $z = -iu$ with $u \in \mathbb{R}$ recovers the Fourier transform. One can also easily express the factorial moments of the Poisson law and we have:

$$\mathbb{E}[N_\lambda(N_\lambda - 1) \cdots (N_\lambda - k + 1)] = \lambda^k,$$

for all $k \geq 1$, so in particular

$$\mathbb{E}[N_\lambda] = \text{Var}(N_\lambda) = \lambda.$$

We shall agree that the constant random variables equal to 0 and to ∞ have the Poisson law with rate 0 and ∞ respectively, which is consistent with the fact that $N_\lambda \rightarrow 0$ in probability as $\lambda \rightarrow 0$, whilst $N_\lambda \rightarrow \infty$ in probability as $\lambda \rightarrow \infty$.

DEFINITION A.10. Let m be a measure on a measurable space $(E, \mathcal{E})$. A *Poisson random measure* with intensity m is a random measure M satisfying the following two properties:

- (i) for every $n \geq 1$ and every $A_1, \dots, A_n$ disjoint sets in $\mathcal{E}$, the random variables $M(A_1), \dots, M(A_n)$ are independent,
- (ii) each $M(A_i)$ has the Poisson law with parameter $m(A_i) \in [0, \infty]$.

The name is justified by the fact that for any $A \in \mathcal{E}$, we have $\mathbb{E}[M(A)] = m(A)$, so indeed $m = \mathbb{E}[M]$ is

the intensity measure of M .

We are going to give a constructive proof that Poisson random measures exist. This will be based on the following two properties of Poisson distributions.

LEMMA A.11 (Summation). Let $(N_i)_{i \geq 1}$ be independent Poisson random variables with rate λ_i respectively. Then $\sum_{i \geq 1} N_i$ has the Poisson distribution with rate $\sum_{i \geq 1} \lambda_i$.

Proof. Let us set $\Lambda_n = \sum_{i=1}^n \lambda_i$ for every $n \geq 1$ and let us first prove that $\sum_{i=1}^n N_i$ has the Poisson distribution with rate Λ_n . If $\lambda_i = \infty$ for some $i \leq n$, then $N_i = \infty$ so $\sum_{i=1}^n N_i = \infty$ and we are done. Suppose thus that $\Lambda_n < \infty$, then for every $k \geq 0$, the multinomial theorem shows that

$$\mathbb{P}\left(\sum_{i=1}^n N_i = k\right) = \sum_{\substack{k_1, \dots, k_n \geq 0 \\ k_1 + \dots + k_n = k}} \prod_{i=1}^n \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!} = \frac{(\sum_{i=1}^n \lambda_i)^k e^{-\sum_{i=1}^n \lambda_i}}{k!},$$

so $\sum_{i=1}^n N_i$ indeed has the Poisson distribution with rate $\Lambda_n = \sum_{i=1}^n \lambda_i$. Let $\Lambda = \sum_{i \geq 1} \lambda_i = \uparrow \lim_n \Lambda_n$. Then when $n \rightarrow \infty$, the right-hand side converges to $\Lambda^k e^{-\Lambda} / k!$ if $\Lambda < \infty$ and to 0 otherwise. This shows that $\sum_{i=1}^n N_i$ converges in distribution to a Poisson law with rate Λ in any case, and since it converges almost surely to $\sum_{i \geq 1} N_i$, then the latter has this Poisson law. $\square$

The second property is a kind of converse property, showing that when sampling a Poisson number of elements and assigning them types (or colours) randomly, then the number of elements of each type is again Poisson distributed.

LEMMA A.12 (Colouring). Let N have the Poisson distribution with rate λ and, independently, let $(Z_n)_{n \geq 1}$ be i.i.d. with values in a finite or countable set I . Write $p_i = \mathbb{P}(Z_1 = i)$ for each $i \in I$. Then the random variables

$$N_i = \sum_{n=1}^N \mathbb{1}_{Z_n=i} \quad \text{for each } i \in I$$

are independent and have the Poisson distribution with rate λp_i respectively.

Proof. Recall that independence of infinitely many random variables is independence of any finite subcollection. Let $J \subset I$ be a finite subset and for every $i \in J$, let $k_i \geq 0$ be some integer. Conditionally on N , the N_i 's have the multinomial distribution: let $k_J = \sum_{i \in J} k_i$ and $p_J = \sum_{i \in J} p_i$, then

$$\begin{aligned} \mathbb{P}(N_i = k_i \text{ for every } i \in J) &= \sum_{K \geq k_J} \mathbb{P}(N = K) \mathbb{P}\left(\sum_{n=1}^K \mathbb{1}_{Z_n=i} = k_i \text{ for every } i \in J\right) \\ &= \sum_{K \geq k_J} \frac{\lambda^K e^{-\lambda}}{K!} \frac{K!}{\prod_{i \in J} k_i! (K - k_J)!} \prod_{i \in J} p_i^{k_i} (1 - p_J)^{K - k_J} \\ &= \prod_{i \in J} \frac{(\lambda p_i)^{k_i} e^{-\lambda p_i}}{k_i!} \underbrace{\sum_{K \geq k_J} \frac{\lambda^{K - k_J} e^{-\lambda(1 - p_J)}}{(K - k_J)!} (1 - p_J)^{K - k_J}}_{= 1}, \end{aligned}$$

which is indeed the probability that independent Poisson random variables with rate λp_i are equal to k_i for each $i \in J$. $\square$

We may now prove the existence of Poisson random measures. Recall that a measure m is said to be σ -finite when there exists a countable partition $E = \bigcup_{n \geq 1} E_n$ of E into disjoint elements $E_n \in \mathcal{E}$ such that $m(E_n) < \infty$ for every n .

THEOREM A.13. For every σ -finite measure m on E there exists a Poisson random measure M with intensity m . In addition, its law is unique, and we may construct it in the form:

$$M = \sum_{i=1}^N \delta_{X_i}$$

where the X_i 's are random variables with values in E and where $N = \infty$ almost surely if $m(E) = \infty$, and otherwise N has the Poisson distribution with rate $m(E) < \infty$.

Proof. Uniqueness of the distribution of M is plain since we prescribe the joint distribution of any finite disjoint collection of sets. Let us construct M first in the case where m is a finite measure. First, if $m(E) = 0$, then the null measure works. Assume then that $m(E) > 0$.

In this case, the measure $\mu(\cdot) = m(\cdot)/m(E)$ is a probability, let then $(X_n)_{n \geq 1}$ be i.i.d. random variables with this law and, independently, let N have the Poisson distribution with parameter $m(E)$. Define then the measure

$$M = \sum_{n=1}^N \delta_{X_n}.$$

Let $A_1, \dots, A_k$ be disjoint sets in $\mathcal{E}$. Lemma A.12 with $Z_n = i$ if and only if $X_n \in A_i$ shows that the $M(A_i)$'s are independent and Poisson distributed with rate $m(A_i)$ respectively.

Now if m is only a σ -finite measure, then there exists a partition $E = \bigcup_{i \geq 1} E_i$ of disjoint elements $E_i \in \mathcal{E}$ such that $0 < m(E_i) < \infty$ for each $i \geq 1$. Then for every $i \geq 1$, let $m_i(\cdot) = m(\cdot \cap E_i)$, which is a finite measure. Construct on some probability space independent random variables $(N_i, X_{n,i} : n, i \geq 1)$ such that:

- each N_i has the Poisson law with rate $m_i(E) = m(E_i)$,
- each $X_{n,i}$ has the law $m_i(\cdot)/m_i(E) = m(\cdot \cap E_i)/m(E_i)$.

then let

$$M_i = \sum_{n=1}^{N_i} \delta_{X_{n,i}} \quad \text{and then} \quad M = \sum_{i \geq 1} M_i.$$

By the first part of the proof, the M_i 's are independent Poisson random measures with intensity $m_i(\cdot) = m(\cdot \cap E_i)$ respectively. Lemma A.11 easily implies that M is a Poisson random measure with intensity $\sum_i m_i = m$. $\square$

REMARK A.14. We call the representation $M = \sum_{n=1}^N \delta_{X_n}$ the *canonical form* of the Poisson random measure M . When considering the law of a Poisson random measure, we can always assume that it takes this form (which we will do). Beware that in this representation, the X_n 's are in general identically distributed only when $m(E) < \infty$.

A.3 STABILITY PROPERTIES

Let us derive simple but very useful properties of Poisson random measures.

RESTRICTION. Let M be a Poisson random measure with intensity m and let $(A_i)_{i \in I}$ any collection of disjoint sets in $\mathcal{E}$. Then the restrictions $M_{|A_i} = M(\cdot \cap A_i)$ are independent Poisson random measures with intensity $m_{|A_i} = m(\cdot \cap A_i)$ respectively.

Proof. First, recall from Lemma A.7 that the restriction is a measurable operation, so the $M_{|A_i}$'s are indeed random measures. Recall that independence of infinitely many random variables is independence of any finite subcollection, therefore it suffices to prove the claim when I is finite. Let $B_1, \dots, B_n$ be disjoint sets in $\mathcal{E}$, then the sets $(B_k \cap A_i : i \in I, 1 \leq k \leq n)$ are also disjoint sets in $\mathcal{E}$, so the random variables $(M(B_k \cap A_i) : i \in I, 1 \leq k \leq n)$ are independent and Poisson distributed with parameter $m(B_k \cap A_i)$ respectively. By grouping them, the collections $(M_{|A_i}(B_k), 1 \leq k \leq n)_{i \in I}$ are independent for every $B_1, \dots, B_n$ disjoint sets in $\mathcal{E}$, so the measures $(M_{|A_i})_{i \in I}$ are independent. Finally since $M_{|A_i}(B_k)$ has the Poisson distribution with parameter $m_{|A_i}(B_k)$ for every $B_1, \dots, B_n$ disjoint sets in $\mathcal{E}$, then $M_{|A_i}$ is a Poisson random measure with intensity $m_{|A_i}$. $\square$

IMAGE. Let M be a Poisson random measure with intensity m on $(E, \mathcal{E})$ and let $f : (E, \mathcal{E}) \rightarrow (E', \mathcal{B}(E'))$ be a measurable mapping. Then the image measure of M by f is a Poisson random measure on $(E', \mathcal{B}(E'))$, with intensity the image measure of m by f .

Proof. First, recall from Lemma A.8 that the image of M by f is indeed a random measure on E' , with intensity the image of m . That this measure is a Poisson random measure easily follows from the definition: if the A_k 's are disjoint elements of $\mathcal{B}(E')$, then so are the $f^{-1}(A_k)$ in $\mathcal{E}$, so the $M'(A_k) = M(f^{-1}(A_k))$'s are independent and have the Poisson distribution with rate $m(f^{-1}(A_k))$. $\square$

Note that the image of a σ -finite measure may not be σ -finite, such as the image of the Lebesgue measure on $\mathbb{R}^2$ by the projection on the first coordinate. Nevertheless there still exists a Poisson random measure with intensity $m \circ f^{-1}$ provided there exists one with intensity m in the first place. If m is σ -finite, we saw that M can be written in the canonical form $\sum_{n=1}^N \delta_{X_n}$ with $N \sim \text{Po}(m(E))$ and then, even if the image of M is not σ -finite, it can still be written as $\sum_{n=1}^N \delta_{f(X_n)}$ with $N \sim \text{Po}(m(E) = m(f^{-1}(F)))$.

SUMMATION. Let I be a finite or countable set and let $(M_i)_{i \in I}$ be independent Poisson random measures on E with intensity m_i respectively. Then $M = \sum_{i \in I} M_i$ is a Poisson random measure with intensity $m = \sum_{i \in I} m_i$.

Proof. Recall from Lemma A.9 that M is indeed a random measure with intensity m . That it is a Poisson random measure follows easily from Lemma A.11. $\square$

As for the previous property, the infinite countable sum of σ -finite measures may not be σ -finite: just sum the same (nonzero...) even finite measure infinitely many times, but there still exists a Poisson random measure with intensity m .

MARKING. Suppose that M is a Poisson random measure with intensity m that takes the form $M = \sum_{n=1}^N \delta_{X_n}$ for some random variable N with values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ and random variables $(X_n)_{n \geq 1}$ with values in E , as constructed in Theorem A.13. Let $(Y_n)_{n \geq 1}$ be i.i.d. random variables with distribution π on some space $(E', \mathcal{B}(E'))$ and assume that this sequence is independent of $(N, X_1, X_2, \dots)$. Then

$$\tilde{M} = \sum_{n=1}^N \delta_{(X_n, Y_n)}$$

is a Poisson random measure on $(E \times E', \mathcal{E} \otimes \mathcal{B}(E'))$, with intensity $m \otimes \pi$.

Proof. Fix any $A \in \mathcal{E}$ and $B \in \mathcal{B}(E')$ and let us prove that $\tilde{M}(A \times B)$ has the Poisson distribution with rate $m(A)\pi(B)$. Since $\mathcal{E} \otimes \mathcal{B}(E')$ is generated by these product sets, then this shows that $\tilde{M}(C)$ has the Poisson distribution with rate $(m \otimes \pi)(C)$ for every $C \in \mathcal{E} \otimes \mathcal{B}(E')$. Suppose first that $0 < m(A)\pi(B) < \infty$.

Fix $k \geq 0$, since the Y_n 's are i.i.d. and are independent of N and the X_n 's, then

$$\begin{aligned} \mathbb{P}(\tilde{M}(A \times B) = k) &= \sum_{K \geq k} \mathbb{P}(M(A) = K) \mathbb{P}(\#\{1 \leq i \leq K : Y_i\} = k) \\ &= \sum_{K \geq k} \frac{m(A)^K e^{-m(A)}}{K!} \binom{K}{k} \pi(B)^k (1 - \pi(B))^{K-k} \\ &= \frac{(m(A)\pi(B))^k e^{-m(A)}}{k!} \sum_{K \geq k} \frac{m(A)^{K-k}}{(K-k)!} (1 - \pi(B))^{K-k} \\ &= \frac{(m(A)\pi(B))^k e^{-m(A)\pi(B)}}{k!}, \end{aligned}$$

so $\tilde{M}(A \times B)$ has the Poisson law with rate $m(A)\pi(B)$. This also holds when $m(A)\pi(B) = \infty$ since in this case $m(A) = \infty$ and $\pi(B) > 0$, so $M(A) = \infty$ almost surely and finally $\tilde{M}(A \times B) = \infty$ almost surely. Similarly, if $m(A)\pi(B) = 0$, then either $m(A) = 0$ and then $M(A) = 0$ almost surely so further $\tilde{M}(A \times B) = \infty$ almost surely, or $\pi(B) = 0$ and then $\tilde{M}(A \times B) = 0$ almost surely.

It remains to check that the measures of disjoint sets are independent. Again, it suffices to consider product sets $A_1 \times B_1, \dots, A_n \times B_n$ and this follows easily independence of the $M(A_i)$'s and of the Y_i 's. $\square$

THINNING. Suppose that M is a Poisson random measure with intensity m that takes the form $M = \sum_{n=1}^N \delta_{X_n}$ for some random variable N with values in $\{0, 1, 2, \dots\} \cup \{\infty\}$ and random variables $(X_n)_{n \geq 1}$ with values in E , as constructed in Theorem A.13. Let $(Y_n)_{n \geq 1}$ be i.i.d. random variables with values in a finite or countable set I , with law π , and assume that $(Y_n)_{n \geq 1}$ is independent of $(N, X_1, X_2, \dots)$. Then the measures

$$\sum_{\substack{1 \leq n \leq N \\ Y_n = i}} \delta_{X_n} \quad \text{for each } i \in I$$

are independent Poisson random measures with intensity $\pi_i m$ respectively.

Proof. First, by the marking property, the measure $\sum_{n=1}^N \delta_{(X_n, Y_n)}$ is a Poisson random measure on $E \times I$ with intensity $m \otimes \pi$. Since the sets $E \times \{i\}$ are disjoint for different values of $i \in I$, then the restriction property shows that the measures

$$\sum_{n=1}^N \delta_{(X_n, Y_n)}(\cdot \cap (E \times \{i\}))$$

are independent Poisson random measures with intensity $m \otimes \pi(\cdot \cap (E \times \{i\}))$. The claim then follows from the image property, with $f : E \times I \rightarrow E$ the natural projection. $\square$

A.4 INTEGRATION AGAINST A POISSON MEASURE

Recall from Lemma A.6 that the law of a random measure is characterised by its Laplace functional or its Fourier functions. Those of Poisson random measures are fairly explicit.

PROPOSITION A.15. A random measure M is a Poisson random measure with intensity m if and only if for every measurable and nonnegative function f , we have:

$$\mathbb{E}[e^{-M(f)}] = \exp\left(-\int (1 - e^{-f(x)})m(dx)\right).$$

This is also equivalent to having

$$\mathbb{E}[e^{iM(f)}] = \exp\left(\int (e^{if(x)} - 1)m(dx)\right),$$

for every measurable function f which is m -integrable.

REMARK A.16. Many calculations here and in the next chapter are based on the following simple, but useful, observation:

$$1 - e^{-x} \mathbb{1}_{x \in A} = (1 - e^{-x}) \mathbb{1}_{x \in A}.$$

Proof. We only do it for the Laplace functional since the Fourier one is very similar. Suppose first that M is a Poisson random measure with intensity m . If $f = c \mathbb{1}_A$ for arbitrary $c > 0$ and $A \in \mathcal{E}$, then $M(f) = cM(A)$ and $M(A)$ has the Poisson distribution with rate $m(A)$ so

$$\exp\left(-\int (1 - e^{-f(x)})m(dx)\right) = \exp\left(-\int_A (1 - e^{-c})m(dx)\right) = e^{-m(A)(1-e^{-c})} = \mathbb{E}[e^{-cM(A)}].$$

Suppose more generally that $f = \sum_{k=1}^n c_k \mathbb{1}_{A_k}$ is a simple function, with the c_k 's nonnegative and the A_k 's disjoint, so the $M(A_k)$'s are independent and Poisson distributed with rate $m(A_k)$ respectively. The previous formula extends then easily:

$$\begin{aligned} \exp\left(-\int (1 - e^{-f(x)})m(dx)\right) &= \exp\left(-\int \sum_{k=1}^n (1 - e^{-c_k}) \mathbb{1}_{x \in A_k} m(dx)\right) \\ &= \prod_{k=1}^n \exp\left(-\int_{A_k} (1 - e^{-c_k})m(dx)\right) \\ &= \prod_{k=1}^n \mathbb{E}[e^{-c_k M(A_k)}] \\ &= \mathbb{E}[e^{-M(f)}]. \end{aligned}$$

If f is any measurable and nonnegative function, then there exist simple functions $f_n \geq 0$ such that $f_n \uparrow f$. Then for every n , we have

$$\exp\left(-\int (1 - e^{-f_n(x)})m(dx)\right) = \mathbb{E}[e^{-M(f_n)}].$$

By monotone convergence, the integral on the left converges to $\int (1 - e^{-f(x)})m(dx)$ and $M(f_n) \uparrow M(f)$, leading to the identity for f by dominated convergence on the right.

The converse follows from Lemma A.6. $\square$

EXERCISE A.17. Give an alternative proof of the stability properties of Poisson random measures based on the Laplace formula from Proposition A.15.

Recall from Proposition A.4 that for a measurable and nonnegative function f and a random measure M with intensity m , we have $\mathbb{E}[M(f)] = \int f dm$. One can recover this formula when M is a Poisson random measure and also compute the second moment.

COROLLARY A.18. Let M be a Poisson random measure with intensity m . For every measurable and nonnegative function f such that $\int f dm < \infty$, we have

$$\mathbb{E}[M(f)] = \int f dm \quad \text{and} \quad \text{Var}(M(f)) = \int f^2 dm.$$

Proof. The first formula was proved in Proposition A.4 for general random measures. In the case of a Poisson random measure, we may recover it by applying the Laplace formula since

$$\mathbb{E}[M(f)] = -\lim_{t \downarrow 0} \frac{\partial}{\partial t} \mathbb{E}[e^{-tM(f)}]$$

by dominated convergence. On the other hand

$$\frac{\partial}{\partial t} \exp\left(-\int (1 - e^{-tf}) dm\right) = -\left(\int f e^{-tf} dm\right) \exp\left(-\int (1 - e^{-tf}) dm\right),$$

which converges to $-\int f dm$ as $t \downarrow 0$. Similarly, by taking the second derivative,

$$\mathbb{E}[M(f)^2] = \left(\int f^2 dm\right) + \left(\int f dm\right)^2,$$

so indeed $\text{Var}(M(f)) = \int f^2 dm$. $\square$

We see that for a measurable and nonnegative function f and a Poisson random measure M with intensity m , we have $\mathbb{E}[M(f)] < \infty$ if and only if $\int f dm < \infty$. Here is now a necessary and sufficient condition for $M(f)$ to be finite.

COROLLARY A.19. Let M be a Poisson random measure with intensity m . For every measurable function $f : E \rightarrow \mathbb{R}_+$, we have the following dichotomy:

- either $\int (f \wedge 1) dm < \infty$ and then $M(f) < \infty$ almost surely,
- or $\int (f \wedge 1) dm = \infty$ and then $M(f) = \infty$ almost surely.

Proof. Observe that $e^{-tM(f)} = 0$ for any $t > 0$ when $M(f) = \infty$, but $e^{-tM(f)} \rightarrow 1$ as $t \rightarrow 0$ when $M(f) < \infty$. We then deduce from dominated convergence that

$$\mathbb{P}(M(f) < \infty) = \mathbb{E}[\mathbb{1}_{M(f) < \infty}] = \lim_{t \downarrow 0} \mathbb{E}[e^{-tM(f)}] = \lim_{t \downarrow 0} \exp\left(-\int (1 - e^{-tf(x)}) m(dx)\right),$$

by Proposition A.15. We will use the following elementary inequalities: if $c = 1 - e^{-1}$, then $c(u \wedge 1) \leq 1 - e^{-u} \leq u \wedge 1$ for every $u \geq 0$. As a consequence, since $(1 - e^{-tf}) \rightarrow 0$ as $t \downarrow 0$, then if $\int (f \wedge 1) dm < \infty$, then the right-hand side in the previous display tends to 1 by dominated convergence. On the contrary, if $\int (f \wedge 1) dm = \infty$, then for $0 < t < 1$,

$$\int (1 - e^{-tf}) dm \geq \int c(1 \wedge (tf)) dm \geq \int ct(1 \wedge f) dm = \infty,$$

and so $\mathbb{E}[e^{-tM(f)}] = 0$ for every $0 < t < 1$, which means $M(f) = \infty$ almost surely. $\square$

A.5 COMPENSATED INTEGRALS

Let $f : E \rightarrow [-1, 1]$ be measurable. By the previous result, it is necessary that $\int |f| dm < \infty$ in order that $\int |f| dM < \infty$ with non zero probability, and then it actually holds with probability 1, and even more $\mathbb{E}[\int |f| dM] < \infty$. In this case we can make sense of

$$\int f d(M - m) = \int f dM - \int f dm,$$

which has mean zero and variance $\int f^2 dm < \infty$ since $f^2 \leq |f|$. We are now going to make sense of the left-hand side under the weaker integrability assumption $\int f^2 dm < \infty$, under which both integrals on the right may not make sense. We use that m is σ -finite, so there exists a partition $E = \bigcup_{k \geq 1} E_k$ into measurable sets such that $m(E_k) < \infty$ for each $k \geq 1$, so

$$X_k = \int_{E_k} f dM - \int_{E_k} f dm$$

does make sense for every $k \geq 1$.

LEMMA A.20 (Compensated Poisson integral). Let $f : E \rightarrow [-1, 1]$ be a measurable function. With the previous notation, the series $\sum_k X_k$ converges almost surely if and only if $\int f^2 dm < \infty$. In this case the convergence also holds in L^2 and the limit, which we denote by $\int f d(M - m)$, satisfies

$$\mathbb{E}\left[\int f d(M - m)\right] = 0 \quad \text{and} \quad \mathbb{E}\left[\left(\int f d(M - m)\right)^2\right] = \int f^2 dm,$$

as well as, for every $u \in \mathbb{R}$,

$$\mathbb{E}[e^{iu \int f d(M - m)}] = \exp\left(\int (e^{iuf} - 1 - iuf) dm\right).$$

Finally, the law of $\int f d(M - m)$ does not depend on the partition $E = \bigcup_{k \geq 1} E_k$.

Proof. Suppose first that $\int f^2 dm < \infty$. Since the E_k 's are disjoint, then the Poisson random measures restricted to each of them are independent, so the X_k are independent. By the previous corollary, they have zero mean and variance $\int_{E_k} f^2 dm$ respectively. Since the sum of these variances is finite, then the cumulative sum $(\sum_{k \leq n} X_k)_{n \geq 1}$ is a martingale bounded in L^2 , so it converges almost surely and in L^2 to a finite limit, which we denote by

$$\int f d(M - m) = \sum_{k \geq 1} X_k,$$

which has mean zero and variance $\int f^2 dm$. Finally, according to Proposition A.15, for every $u \in \mathbb{R}$, each X_k has characteristic function:

$$\begin{aligned} \mathbb{E}[e^{iuX_k}] &= \mathbb{E}[e^{iuM(f \mathbb{1}_{E_k})}] e^{-i u m(f \mathbb{1}_{E_k})} = \exp\left(\int (e^{iuf \mathbb{1}_{E_k}} - 1 - iuf \mathbb{1}_{E_k}) dm\right) \\ &= \exp\left(\int_{E_k} (e^{iuf} - 1 - iuf) dm\right). \end{aligned}$$

By independence, the characteristic function of their sum $\int f d(M - m)$ is the product of these characteristic functions. This characteristic function, and thus the law of $\int f d(M - m)$, does not depend on the partition.

Suppose conversely that $\sum_k X_k$ to converge almost surely and let M' be another Poisson random measure with intensity m , independent of M . Then the corresponding series of $X'_k = \int_{E_k} f dM' - \int_{E_k} f dm$ has the same law so is also converges almost surely. The $Y_k = X_k - X'_k$'s are independent, with characteristic function

$$\mathbb{E}[e^{iuY_k}] = \exp\left(-2 \int_{E_k} (1 - \cos(uf)) dm\right),$$

so by independence,

$$\mathbb{E}[e^{iu \sum_{k=1}^n Y_k}] = \exp\left(-2 \int_{\bigcup_{k=1}^n E_k} (1 - \cos(uf)) dm\right).$$

Since $\sum_k Y_k$ converges almost surely, then its limit has characteristic function:

$$\mathbb{E}[e^{iu \sum_k Y_k}] = \exp\left(-2 \int_E (1 - \cos(uf)) dm\right).$$

The left-hand side is continuous in u , and equal to 1 at $u = 0$, so it is nonzero on a neighbourhood of 0. Thus, for every u close enough to 0, we have

$$\int_E (1 - \cos(uf)) dm < \infty.$$

Since $1 - \cos(uf) \geq (uf)^2/4$ for $|u| \leq 1$ (recall that $|f| \leq 1$), then we must have $\int f^2 dm < \infty$. $\square$

Let us extend the compensated integral to unbounded functions. Let $f : E \rightarrow \mathbb{R}$ satisfy $\int (f^2 \wedge 1) dm < \infty$. Let us decompose the space as $E = \{|f| \leq 1\} \cup \{|f| > 1\}$. Since these sets are disjoint, then the Poisson random measures restricted to each of them are independent. The previous lemma applies on $\{|f| \leq 1\}$ to construct the compensated integral $\int_{|f| \leq 1} f d(M - m)$, with characteristic function:

$$\mathbb{E}[e^{iu \int_{|f| \leq 1} f d(M-m)}] = \exp\left(\int_{|f| \leq 1} (e^{iuf} - 1 - iuf) dm\right).$$

Let us focus now on the subspace $\{|f| > 1\}$. Here no integrability is assumed on f but m restricted to this space is a finite measure, so the Poisson random measure M restricted to this takes the form $M = \sum_{n=1}^N \delta_{X_n}$ where N is a Poisson random variable with parameter $m(\{|f| > 1\}) < \infty$, and, independently, the X_n 's are i.i.d. with the law $m(\cdot \cap \{|f| > 1\})/m(\{|f| > 1\})$. Then we simply have $\int_{|f| > 1} f dM = \sum_{n=1}^N f(X_n)$ which is a compound Poisson random variable! The intensity ν is the image by f of the measure $m(\cdot \cap \{|f| > 1\})$ so the characteristic function is:

$$\mathbb{E}[e^{iu \int_{|f| > 1} f dM}] = \exp\left(\int (e^{iux} - 1)\nu(dx)\right) = \exp\left(\int_{|f| > 1} (e^{iuf} - 1) dm\right).$$

If we have $\int (f^2 \wedge |f|) dm < \infty$, then $\int_{|f| > 1} f dm$ makes sense, and letting further $\int_{|f| > 1} f d(M - m) = \int_{|f| > 1} f dM - \int_{|f| > 1} f dm$, we get:

$$\mathbb{E}[e^{iu \int_{|f| > 1} f d(M-m)}] = \exp\left(\int_{|f| > 1} (e^{iuf} - 1 - iuf) dm\right).$$

Then the random variable $\int f d(M - m) = \int_{|f| \leq 1} f d(M - m) + \int_{|f| > 1} f d(M - m)$ has characteristic function

$$\mathbb{E}[e^{iu \int f d(M-m)}] = \exp\left(\int (e^{iuf} - 1 - iuf) dm\right). \quad (\text{A.1})$$

A.6 SPACE-TIME MEASURES

Let μ be a σ -finite measure (and not identically zero) on a measurable space $(E, \mathcal{E})$ and let Leb denote the Lebesgue measure on $[0, \infty)$. Let us consider a Poisson random measure M on the product space $([0, \infty) \times E, \mathcal{B}([0, \infty)) \otimes \mathcal{E})$ with intensity $m = \text{Leb} \otimes \mu$. This intensity is again σ -finite, and with infinite total mass, so one can canonically represent such a Poisson random measure in the form

$$M = \sum_{n \geq 1} \delta_{(t_n, x_n)}.$$

One can interpret M as a rain of points (the x_n 's), falling at random times (the t_n 's). The next lemma shows that two points never drop simultaneously.

LEMMA A.21. Almost surely, for every $t \geq 0$, we have $M(\{t\} \times E) \in \{0, 1\}$.

Proof. Since μ is σ -finite, then there exists a partition $E = \bigcup_{i \geq 1} E_i$ with $\mu(E_i) < \infty$ for every $i \geq 1$, and the collection of sets $E_{k,i} = [k-1, k) \times E_i$ for $k, i \geq 1$ is a measurable partition of $[0, \infty) \times E$ into sets of finite measure $m(E_{k,i}) = \mu(E_i)$. The proof of Theorem A.13 shows that for every $k, i \geq 1$, the restriction of the Poisson random measure M to $[k-1, k) \times E_i$ may be written as $\sum_{n=1}^{N_{k,i}} \delta_{(t_n^{k,i}, x_n^{k,i})}$ where $N_{k,i}$ has the Poisson distribution with rate $\mu(E_i)$ and the $(t_n^{k,i}, x_n^{k,i})$'s are i.i.d. with the law $m(\cdot \cap E_{k,i})/m(E_{k,i})$, independently of $N_{k,i}$. In particular, the $t_n^{k,i}$'s have the uniform distribution on $[k-1, k)$ so almost surely, all they are different as n, k , and i vary. $\square$

The random measure $M = \sum_{n \geq 1} \delta_{(t_n, x_n)}$ on $[0, \infty) \times E$ is called a *Poisson point process* with intensity μ , although this denomination is also sometimes used as a synonym of a Poisson random measure on E with intensity μ , and also for the stochastic process $(e_t)_{t \geq 0}$ defined as follows:

$$e_t = \begin{cases} x_n & \text{if there exists } n \geq 1 \text{ such that } t = t_n, \\ \dagger & \text{otherwise,} \end{cases} \quad (\text{A.2})$$

where $\dagger$ is an extra element that does not belong to E and is called a *cemetery point*. Note that this process is well-defined thanks to the previous lemma that shows that, almost surely, for every $t \geq 0$, there can be at most one n such that $t = t_n$, so $e_t = x_n$. Note that for each fixed $t \geq 0$, we have $m(\{t\} \times E) = 0$ and therefore $M(\{t\} \times E) = 0$ almost surely, so $e_t = \dagger$ almost surely. Therefore this process point of view is not informative since the times $t \geq 0$ such that $e_t \neq \dagger$ are an exception and the constant process equal to $\dagger$ is always a modification of $(e_t)_t$. The ‘correct’ point of view on this process is the underlying Poisson random measure.

PROPOSITION A.22. If $\mu(E) = \infty$, then $\{t_n : n \geq 1\}$ is dense in $[0, \infty)$. If $\mu(E) < \infty$, then almost surely the set $\{t_n : n \geq 1\}$ can be ordered in an increasing sequence $0 < T_1 < T_2 < \dots$ with the following properties (put $T_0 = 0$):

- (i) the differences $(T_k - T_{k-1})_{k \geq 1}$ are i.i.d. with the exponential law with rate $\mu(E)$,
- (ii) the values $(x_n)_{n \geq 1}$ are i.i.d. with the law $\mu(\cdot)/\mu(E)$,
- (iii) the sequence $(x_n)_{n \geq 1}$ and $(T_k)_{k \geq 1}$ are independent.

Proof. If $\mu(E) = \infty$, then $m((a, b) \times E) = (b - a)\mu(E) = \infty$ for every $0 < a < b < \infty$. Consequently $M((a, b) \times E) = \infty$ almost surely for every fixed $0 < a < b < \infty$, and hence almost surely simultaneously for all $a < b$ in $\mathbb{Q}$. We conclude that almost surely $\{t_n : n \geq 1\}$ intersects each interval with rational extremities, and is therefore dense.

Suppose now that $\mu(E) < \infty$. It suffices to prove the claim when restricting to the time interval $[0, T]$ for any fixed $T > 0$. Since $m([0, T] \times E) = T\mu(E) < \infty$, then the measure M restricted to $[0, T] \times E$ can be written as $M = \sum_{n=1}^N \delta_{(t_n, x_n)}$ where N has the Poisson distribution with parameter $T\mu(E)$ and $(t_n, x_n)_{n \geq 1}$ is a sequence of i.i.d. random pairs with the law

$$\frac{1}{T\mu(E)} \text{Leb}(\cdot \cap [0, T]) \otimes \mu(\cdot) = \left(\frac{1}{T} \text{Leb}(\cdot \cap [0, T]) \right) \otimes \left(\frac{1}{\mu(E)} \mu(\cdot) \right)$$

and is independent of N . In particular the t_n ’s are uniformly distributed on $[0, T]$ and the x_n ’s have the law $\mu(\cdot)/\mu(E)$, and all the random variables $N, (t_n)_n, (x_n)_n$ are independent. We conclude that $(t_n)_{n \leq N}$ has the same law as the claimed $(T_n)_{n \leq N}$. $\square$

In the finite mass case, the description is reminiscent of the compound Poisson process that we next recall.

DEFINITION A.23. Let $\lambda \in (0, \infty)$. A (homogeneous) *Poisson process* with intensity λ is a process $(N_t)_t$ defined by

$$N_t = \sum_{i \geq 1} \mathbb{1}_{T_i \leq t},$$

where $(T_i - T_{i-1})_{i \geq 1}$ are i.i.d. with the exponential law with parameter λ (where $T_0 = 0$).

DEFINITION A.24. Let ν be a finite (and not identically zero) measure on $\mathbb{R}$. Let $(N_t)_t$ be a Poisson process with intensity $\lambda = \nu(\mathbb{R})$ and independently, let $(\xi_i)_{i \geq 1}$ be i.i.d. random variables with the law

$\nu(\cdot)/\nu(\mathbb{R})$ and let $S_n = \xi_1 + \dots + \xi_n$ denote the associated random walk. Then a *compound Poisson process* with intensity ν is a process $(X_t)_t$ defined by

$$X_t = S_{N_t} = \sum_{i \geq 1} \xi_i \mathbb{1}_{T_i \leq t},$$

where the T_i 's are the times of jump of $(N_t)_t$.

In order to transport our setting from a general space E to $\mathbb{R}$, let $f : E \rightarrow \mathbb{R}$ be a measurable function, then for every $t \geq 0$, provided the integral is well-defined, we may consider:

$$X_t^f = \int_{[0,t] \times E} f(x) M(dt, dx).$$

Recall that we may write $M = \sum_{n \geq 1} \delta_{(t_n, x_n)}$. Then the process X^f makes a jump of size $f(x_n)$ at time t_n . It is often written as

$$X_t^f = \sum_{s \leq t} f(e_s),$$

where $(e_t)_t$ is the process defined in (A.2), if we agree that $f(\dagger) = 0$, so the above sum is actually restricted to the (countably many) t_n 's that lie in $[0, t]$.

By the image property, we may alternatively write

$$X_t^f = \int_{[0,t] \times \mathbb{R}} x M^f(dt, dx) = \sum_{s \leq t} e_s^f,$$

where $M^f = \sum_{n \geq 1} \delta_{(t_n, y_n)}$ is a Poisson random measure on $[0, \infty) \times \mathbb{R}$ with intensity $\text{Leb} \otimes \nu$ where $\nu = \mu \circ f^{-1}$ and e^f is defined accordingly. The process now makes a jump of size y_n at time t_n for each atom (t_n, y_n) of M^f .

COROLLARY A.25. If $f \in L^1(\mu)$, then the process $(X_t^f)_t$ is a compound Poisson process with intensity the finite measure $\nu = \mu \circ f^{-1}$. For every $t > 0$ and $u \in \mathbb{R}$, its characteristic function is given by:

$$\mathbb{E}[e^{iuX_t^f}] = \exp\left(-t \int (1 - e^{iuf}) d\mu\right).$$

Proof. That X^f is such a compound Poisson process follows from the previous proposition. The expression of the characteristic function follows from Proposition A.15 $\square$

REMARK A.26. The Poisson process with intensity $\lambda > 0$ corresponds to X^f where $f = 1$.

It is interesting to note a converse statement. Let $X_t = \sum_{i \geq 1} \xi_i \mathbb{1}_{T_i \leq t}$ be a compound Poisson process with intensity some finite measure ν on $\mathbb{R}$. Note that X makes finitely many jumps in any bounded time interval. For every $A \in \mathcal{B}([0, \infty) \times \mathbb{R})$, let

$$M^X(A) = \sum_{i \geq 1} \delta_{(T_i, \xi_i)}(A),$$

denote the number of ‘time-space jump points’ that belong to A .

PROPOSITION A.27. The measure M^X is a Poisson random measure with intensity $\text{Leb} \otimes \nu$.

Proof. Let $\lambda = \nu(\mathbb{R})$ and $\pi(\cdot) = \mu(\cdot)/\lambda$. Fix some $t > 0$. Then the number of jumps $M^X([0, t] \times \mathbb{R})$ of X during the interval $(0, t]$ is Poisson distributed with parameter λt , and the times and values of these jumps are independent, the values are distributed according to π while the times are uniformly

distributed over $[0, t]$. Fix $B_1, \dots, B_k \in \mathcal{B}(\mathbb{R})$ disjoint Borel sets and $(s_1, t_1], \dots, (s_\ell, t_\ell]$ disjoint sub-intervals of $(0, t]$. We infer from the colouring lemma A.12 applied to i.i.d. random variables Z_n with the law

$$\mathbb{P}(Z_n = (i, j)) = \pi(B_i) \frac{t_j - s_j}{t}, \quad i \in \{1, \dots, k\}, \quad j \in \{1, \dots, \ell\}$$

that the random variables $M^X((s_j, t_j] \times B_i)$ are independent and Poisson distributed with parameter

$$\lambda t \cdot \pi(B_i) \frac{t_j - s_j}{t} = \mu(B_i)(t_j - s_j) = m((s_j, t_j] \times B_i)$$

respectively. Since the collection of such sets is a π -system that generates $\mathcal{B}([0, \infty) \times \mathbb{R})$, this characterises the law of M^X by Exercise A.1. $\square$

A compound Poisson process is in some sense a discrete-time random walk turned into a continuous-time process. A simple way to construct a martingale from a random walk is to subtract its mean, provided it is well-defined. The next result is the analogue.

LEMMA A.28. Suppose that μ is a finite measure and f is a μ -integrable function. Then the process

$$X_t^f - t \int_E f \, d\mu$$

is a càdlàg martingale for the filtration $\mathcal{F}_t = \sigma(M_{[0, t] \times E})$.

Proof. For each $s, t \geq 0$, the Poisson random measure $M_{[(t, t+s] \times E]}$ is independent of $\mathcal{F}_t$, from which follows that $X_{t+s}^f - X_t^f$ is independent of $\mathcal{F}_t$. In addition it has the same law as X_s^f , which has expectation $s \int f \, d\mu$ and the claim follows. $\square$

Recall that when μ is only a σ -finite measure and $|f| \leq 1$ only satisfies $\int f^2 \, d\mu < \infty$, then we can define the compensated integral

$$\begin{aligned} \tilde{X}_t^f &= \int_{[0, t] \times E} f(x)(M(dt, dx) - t \, \nu(dx)) \\ &= \sum_{n=1}^{\infty} \left(\int_{[0, t] \times E_n} f(x) M(dt, dx) - t \int_{E_n} f \, d\nu \right), \end{aligned}$$

where the series converges almost surely and in L^2 for every given $t \geq 0$ and where $E = \bigcup_{n \geq 1} E_n$ is any measurable partition such that $\mu(E_n) < \infty$, for every $n \geq 1$.

PROPOSITION A.29. Suppose that μ is a σ -finite measure and $|f| \leq 1$ is such that $\int f^2 \, d\mu < \infty$. Then the process $\tilde{X}^f$ is a càdlàg and L^2 martingale for the filtration $\mathcal{F}_t = \sigma(M_{[0, t] \times E})$. In addition, for every $s, t \geq 0$, the increment $\tilde{X}_{t+s}^f - \tilde{X}_t^f$ is independent of $\mathcal{F}_t$ and has the same law as $\tilde{X}_s^f$. Finally, for every $t > 0$ and $u \in \mathbb{R}$, its characteristic function is given by:

$$\mathbb{E}[e^{iu\tilde{X}_t^f}] = \exp\left(t \int (e^{if} - 1 - if) \, d\mu\right).$$

Proof. Let $\tilde{X}_t^N = \sum_{n=1}^N (\int_{[0, t] \times E_n} f(x) M(dt, dx) - t \int_{E_n} f \, d\mu)$. According to the previous lemma, it is a càdlàg martingale for the filtration $\mathcal{F}_t$ as the sum of such processes. As $N \rightarrow \infty$, for any fixed t it converges almost surely to $\tilde{X}_t^f$, which is thus $\mathcal{F}_t$ -measurable, and it also converges in L^2 so $\tilde{X}_t^f \in L^2$. The martingale property is also inherited from $\tilde{X}_t^N$ by the L^2 , and thus L^1 , convergence: simply let $A \in \mathcal{F}_t$, then

$$\mathbb{E}[\tilde{X}_{t+s}^N \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[\tilde{X}_{t+s}^N | \mathcal{F}_t] \mathbb{1}_A] = \mathbb{E}[\tilde{X}_t^N \mathbb{1}_A],$$

and the left-hand side converges to $\mathbb{E}[\tilde{X}_{t+s}^f \mathbb{1}_A]$ whilst the right-hand side converges to $\mathbb{E}[\tilde{X}_t^f \mathbb{1}_A]$.

It follows that $\tilde{X}^f - \tilde{X}^N$ is also a martingale in L^2 . Doob's maximal inequality implies then that for any $t > 0$, we have

$$\mathbb{E} \left[\sup_{s \leq t} (\tilde{X}_s^f - \tilde{X}_s^N)^2 \right] \leq 4 \mathbb{E}[(\tilde{X}_t^f - \tilde{X}_t^N)^2] \xrightarrow{n \rightarrow \infty} 0.$$

This L^2 -convergence implies the almost sure convergence along a subsequence: for any $t > 0$, there exists a sequence $N_k \rightarrow \infty$ along which $\sup_{s \leq t} |\tilde{X}_s^f - \tilde{X}_s^{N_k}| \rightarrow 0$ almost surely, that is, along which $\tilde{X}^{N_k}$ converges almost surely to $\tilde{X}^f$ uniformly over $[0, t]$. Since each path $\tilde{X}^N$ is càdlàg, then so is $\tilde{X}^f$ over $[0, t]$. Finally for any $s < t$, the increment $\tilde{X}_t^f - \tilde{X}_s^f$ is the limit of $\tilde{X}_t^{N_k} - \tilde{X}_s^{N_k}$, which is independent of $\mathcal{F}_s$ and with the same law as $\tilde{X}_{t-s}^{N_k}$ by the previous proof. This is preserved at the limit. The characteristic function is obtained as the limit of that of $\tilde{X}^N$. $\square$

Let us recall that, as any càdlàg function, the process $\tilde{X}^f$ has at most countably many discontinuities, namely let $\Delta \tilde{X}_t^f = \tilde{X}_t^f - \tilde{X}_{t-}^f$ for every $t > 0$, then the set $\mathcal{J} = \{t > 0 : \Delta \tilde{X}_t^f \neq 0\}$ is at most countable. In addition, one can recover the Poisson random measure M by

$$M(A) = \sum_{t \in \mathcal{J}} \delta_{(t, \Delta \tilde{X}_t^f)}(A),$$

for every $A \in \mathcal{B}([0, \infty) \times \mathbb{R})$. Indeed, each process $\int_{[0, t] \times E_n} f(x) M(dt, dx) - t \int_{E_n} f dv$ is a compound Poisson processes plus a linear deterministic function, so Proposition A.27 shows that the analogous jump measure is a Poisson random measure with intensity $\text{Leb} \otimes \mu(\cdot \cap E_n)$. In addition, these measure are independent since the processes are (since the E_n 's are disjoint), and the jump measure M is obtained as their sum, so by stability, it is a Poisson random measure with intensity $\sum_{n \geq 1} \text{Leb} \otimes \mu(\cdot \cap E_n) = \text{Leb} \otimes \mu$.

A.7 EXERCISES

EXERCISE 62 (stability properties). Give an alternative proof of the stability of Poisson random measures under image, summation, and marking based on the characterisation by the Laplace functional: for all $f : E \rightarrow \mathbb{R}_+$,

$$\mathbb{E}[e^{-\int f dM}] = \exp\left(-\int (1 - e^{-f}) dm\right).$$

EXERCISE 63 (percolation). Let $M = \sum_{n \geq 1} \delta_{X_n}$ be a Poisson random measure with intensity the Lebesgue measure on $\mathbb{R}^d$. Let $(R_n)_{n \geq 1}$ be i.i.d. random variables in $(0, \infty)$ independent of M . For every $n \geq 1$, let B_n denote the open ball centred at X_n and with radius R_n . We aim at studying the random set $A = \bigcup_{n \geq 1} B_n$.

(i) Express the law of $\min_n |X_n|$. Compute its expectation in dimension $d = 2$.

(ii) Fix $x \in \mathbb{R}^d$. Express the law of the number of balls that cover x :

$$N(x) = \#\{n \geq 1 : x \in B_n\}$$

(iii) Assume $\mathbb{E}[R_1^d] = \infty$. Prove that for every $x \in \mathbb{R}^d$, we have $x \in A$ almost surely. Prove further that $A = \mathbb{R}^d$ almost surely.

(iv) What happens if $\mathbb{E}[R_1^d] < \infty$?

EXERCISE 64 (a particle system). Let μ be a probability measure on $\mathbb{R}$ with finite first moment $v_m = \int |v| \mu(dv) < \infty$ and let

$$M = \sum_{n \in \mathbb{Z}} \delta_{(X_n, V_n)}$$

be a Poisson random measure on $\mathbb{R}^2$ with intensity $\lambda dx \otimes \mu(dv)$. We view the X_n 's as the positions of particles on $\mathbb{R}$ and the V_n 's as their velocity, both at time 0. The velocity can be either positive or negative,

simply indicating the direction. We assume that the velocity stays constant at all time $t \in \mathbb{R}$ (time also can be negative), so that for every $t \in \mathbb{R}$, the particles lie at position $X_n + tV_n$. We can label canonically these particles using their position at time $t = 0$, e.g. by

$$\cdots < X_{-n} < \cdots < X_{-1} < 0 \leq X_0 < X_1 < \cdots < X_n < \cdots$$

For every $n \in \mathbb{Z}$, we let $T_n = X_n/V_n \in \mathbb{R}$ denote the instant the particle number n lies at 0.

- (i) Prove that for every $t \neq 0$, the measure $M_t = \sum_{n \in \mathbb{Z}} \delta_{(X_n + tV_n, V_n)}$ has the same law as $M = M_0$.
- (ii) Prove that the measure $M' = \sum_{n \in \mathbb{Z}} \delta_{(T_n, V_n)}$ is a Poisson random measure and express its intensity.
- (iii) Fix $a < b$ two real numbers and compute for each $t \geq 0$:

$$\mathbb{P}(\forall s \in [0, t] \forall n \in \mathbb{Z} : X_n + sV_n \notin [a, b])$$

the probability that the segment $[a, b]$ remains empty of particle during the time interval $[0, t]$.

- (iv) Express the law of $\inf\{t \geq 0 : \exists s \in [0, t] \exists n \in \mathbb{Z} : X_n + sV_n \in [a, b]\}$.

EXERCISE 65 (atom extraction). Let M and M' be two i.i.d. Poisson random measures with intensity the Lebesgue measure on $[-1, 1]$. For every $n \geq 1$, let $M'_n(\cdot) = M'(\cdot \cap [-1/n, 1/n])$ and $M^n(\cdot) = M(\cdot \cap ([-1, 1] \setminus [-1/n, 1/n]))$ denote their restriction to $[-1/n, 1/n]$ and to its complement respectively.

- (i) Argue that M^n and M'_n are independent and $M^n + M'_n$ has the same law as M .
- (ii) Prove that $M^n \rightarrow M$ almost surely for the topology of weak convergence of finite measures on $[-1, 1]$, defined by $m_n \rightarrow m$ when $\int f dm_n \rightarrow \int f dm$ for all continuous and bounded function f .
- (iii) Prove that

$$\mathbb{P}(M'([-1/n, 1/n]) > 1 \mid M'([-1/n, 1/n]) \geq 1) \xrightarrow{n \rightarrow \infty} 0$$

and that the conditional probability given $\{M'([-1/n, 1/n]) \geq 1\}$ that the measure M'_n has a unique atom tends to 1. Deduce that M'_n conditionally on $\{M'([-1/n, 1/n]) \geq 1\}$ converges in distribution to δ_0 (for the topology of weak convergence).

- (iv) Deduce that M conditionally on $\{M([-1/n, 1/n]) \geq 1\}$ converges in distribution (again for the topology of weak convergence) to $M + \delta_0$.

We interpret this conclusion by saying that, conditionally on M having an atom at 0, its law is the same as that of the unconditional law plus an atom at 0. This holds true in a much more general context, and is known as ‘Palm formula’.

EXERCISE 66 (Rényi characterisation). Let $M = \sum_{n \geq 1} \delta_{X_n}$ be a random measure on $[0, \infty)$ that satisfies the following properties almost surely: for every $t \geq 0$, we have $M(\{t\}) \in \{0, 1\}$ (i.e. the X_n ’s are all distinct), for every compact set $K \subset [0, \infty)$, we have $M(K) < \infty$, and finally for every $A \in \mathcal{B}([0, \infty))$, we have

$$\mathbb{P}(M(A) = 0) = \exp(-|A|),$$

where $|A|$ is the Lebesgue measure of A . The goal of this exercise is to prove that M necessarily is a Poisson random measure, with intensity the Lebesgue measure.

- (i) Prove that the events $\{M(A_1) = 0\}, \dots, \{M(A_k) = 0\}$ are independent whenever $A_1, \dots, A_k$ are disjoint measurable sets.

- (ii) Fix $I \subset [0, \infty)$ a bounded interval with nonempty interior. Let $D_{n,k} = [k2^{-n}, (k+1)2^{-n})$ for every $n, k \geq 0$ and let $K_n(I) = \#\{k \geq 0 : D_{n,k} \subset I\}$. Express the law of

$$M_n(I) = \sum_{k \geq 0 : D_{n,k} \subset I} \mathbb{1}_{M(D_{n,k}) \neq 0}$$

in terms of n and $K_n(I)$ and deduce the law of $M(I)$.

- (iii) Conclude that M is indeed a Poisson random measure, with intensity the Lebesgue measure.

EXERCISE 67 (Description of Poisson point processes). Let $M = \sum_{n \geq 1} \delta_{(T_n, X_n)}$ be a Poisson point process on E with intensity μ a σ -finite measure on E . Let $J = \{T_n : n \geq 1\}$ denote the set of ‘jump times’.

- (i) Prove that if $\mu(E) = \infty$ then J is dense in $[0, \infty)$, whereas if $\mu(E) < \infty$, then J is made of distinct, isolated points, that can be ordered in an increasing way.

We assume for the rest of this exercise that $\mu(E) < \infty$ and that $0 < T_1 < T_2 < \dots$. We let $\gamma_n = T_n - T_{n-1}$ for every $n \geq 1$, where $T_0 = 0$.

- (ii) Prove further that:

- (a) $(\gamma_n)_{n \geq 1}$ are i.i.d. with the exponential law with rate $\mu(E)$,
- (b) $(X_n)_{n \geq 1}$ are i.i.d. with the law $\mu(\cdot)/\mu(E)$,
- (c) the sequences $(X_n)_{n \geq 1}$ and $(T_n)_{n \geq 1}$ are independent.

Hint: you may first describe the law of the atoms in $[0, T] \times E$ for a fixed time horizon T and check in particular that the sums of i.i.d. exponential law on this interval would yield the same law.

When $\mu(E) < \infty$, the process defined for every $t \geq 0$ by

$$N_t^\lambda = M([0, t] \times E) = \sum_{n \geq 1} \mathbb{1}_{T_n \leq t},$$

is called the *homogeneous Poisson process* with intensity λ , and the one defined for every $t \geq 0$ by

$$X_t^\lambda = \int_{[0, t] \times E} x M(dt, dx) = \sum_{n \geq 1} X_n \mathbb{1}_{T_n \leq t} = \sum_{n=1}^{N_t^\lambda} X_n,$$

is called the *compound Poisson process* with intensity $\lambda\mu$.

EXERCISE 68. Let

$$M = \sum_{n \geq 1} \delta_{(T_n, X_n)}$$

denote a Poisson random measure on $[0, 1] \times (0, \infty)$ with intensity

$$m(dt, dx) = \mathbb{1}_{t \in [0, 1]} dt \otimes x^{-3} \mathbb{1}_{x > 0} dx.$$

Define inductively

$$m_1 = \sup\{X_n : n \geq 1\}, \quad m_k = \sup\{X_n : n \geq 1, X_n < m_{k-1}\}.$$

- (i) Prove that m_k is well-defined for all $k \geq 1$, strictly decreasing, and that almost surely there exists a unique $n \geq 1$ such that $X_n = m_k$.
- (ii) Find the law of m_1 .
- (iii) Prove that $M' = \sum_{k \geq 1} \delta_{m_k}$ is a Poisson random measure and express its intensity.
- (iv) Prove that, conditionally given m_1 , the measure $M'' = \sum_{k \geq 2} \delta_{m_k} = M' - \delta_{m_1}$ is also a Poisson random measure and express its intensity.